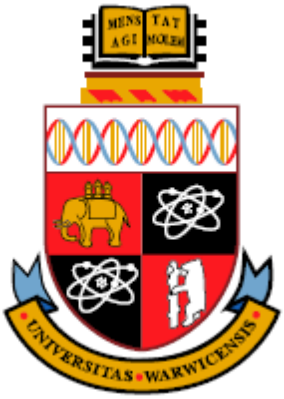

Life or Cell Death: Identifying c-Myc Regulated Genes in Two Distinct Tissues

by

Samuel Charles Robson BSc MSc

November 21, 2008

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

MOAC Doctoral Training Centre, University of Warwic

Table of Contents

Figures

Chapter 1: Introduction

Chapter 2: Materials and Methods

Chapter 3: *Envisage*: Significance Analysis of Microarray Data Using Linear Models

Chapter 4: Comparison of Transcriptional Response to MycER^{TAM} Activation in Suprabasal Keratinocytes and Pancreatic β-Cells

Chapter 5: General Discussion

Tables

Chapter 2: Materials and Methods

Chapter 3: *Envisage*: Significance Analysis of Microarray Data Using Linear Models

Chapter 4: Comparison of Transcriptional Response to MycER^{TAM} Activation in Suprabasal Keratinocytes and Pancreatic β-Cells

Acknowledgements

First and foremost, I would like to thank my supervisors Dr. Michael Khan, Prof. David Epstein and Dr. Stella Pelengaris for taking me in as a lowly mathematics graduate and moulding me into a biostatistician. Your constant support and kind words have always inspired me, and I will be forever grateful. I would also like to thank Prof. Alison Rodger and the rest of the MOAC cadre for giving me the opportunity to be a part of something very special. I hope that I make you all proud!

Thanks must go to the girls at the Molecular Biology Service, Dr. Helen Brown, Lesley Ward and Sue Davis, who have graciously put up with my constant presence and offered support and advice throughout my time here. I am especially grateful to Lesley, who suffered the majority of the work load, and Helen, whose constant support, advice and patience throughout my project has been invaluable. I am also indebted to Dr. Heather Turner for her assistance with programming and statistical issues, without which this may have ended up a pure biology PhD, and to Dr. Ewan Hunter from Agilent Technologies for giving me the opportunity to add something to the wider world of data analysis.

To my family and friends, whose constant encouragement and support has kept me sane throughout my PhD, I am also extremely grateful. I would particularly like to thank my loving fiancé Jen for putting up with me through this difficult period; something which I know can be difficult at the best of times!

This work was funded by the Engineering and Physical Sciences Research Council (EPSRC) through the Molecular Organisation and Assembly in Cells (MOAC) Doctoral Training Centre. The microarray portion of the project was additionally funded by the Biotechnology and Biological Sciences Research Council (BBSRC), the Association for International Cancer Research (AICR), Eli Lilly (Indianapolis, IL) and Amylin Pharmaceuticals Inc. (San Diego, CA).

Declaration

The author declares that, to the best of his knowledge, the data contained within this thesis is original and his own work under the supervision of his supervisors, Dr. Michael Khan, Prof. David Epstein and Dr. Stella Pelengaris.

The material in this thesis is submitted for the degree of PhD to the University of Warwick only and has not been submitted to any other university. All sources of information have been specifically acknowledged in the form of references.

Abbreviations

%P	Percent Present
2a-cRNA	Double-Amplified Biotin-Labelled cRNA
4OHT	4-Hydroxytamoxifen
A	Adenine
AIC	Akaike Information Criterion
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
ATP	Adenosine Triphosphate
BH	Bcl2 Homology Domain
bHLH-LZ	Basic-Helix-Loop-Helix Leucine-Zipper
BLAST	Basic Local Alignment Search Tool
BSA	Bovine Serum Albumin
C	Cytosine
CDK	Cyclin-dependent Kinase
cDNA	Complementary Deoxyribonucleic Acid
CGH	Comparative Genome Hybridisation
ChIP	Chromatin Immunoprecipitation
c-Myc	Cellular Myelocytomatosis
CPU	Central Processing Unit
cRNA	Complementary Ribonucleic Acid

CTD	Carboxyl-Terminal Domain
DEPC	Diethylpyrocarbonate
DF	Degrees of Freedom
DISC	Death-Inducing Signalling Complex
DNA	Deoxyribonucleic Acid
DNase	Deoxyribonuclease
dNTP	Deoxynucleotide Triphosphate
DPX	P-xylene-bis-pyridinium Bromide
EDTA	Ethylenediaminetetra acetic acid
Envisage	Enables Numerous Variables In Significance Analysis of Gene-expression
ERα	Estrogen Receptor Alpha
FAM	6-Carboxyfluorescein
FDR	False Discovery Rate
FITC	Fluorescein Isothiocyanate
FRET	Fluorescence Resonance Energy Transfer
FWER	Family-Wise Error Rate
G	Guanine
GB	Gigabyte
GCOS	GeneChip Operating System
GC-RMA	GeneChip Robust multi-chip averaging
GHz	Gigahertz

GO	Gene Ontology
GS-GX	GeneSpring Gene-expression Analysis Suite
H&E	Haematoxylin and Eosin
HDAC	Histone Deacetylase
hnRNA	Heterogeneous Nuclear Ribonuclease
IAP	Inhibitor of Apoptosis Protein
IgG	Immunoglobulin G
IID	Independent and Identically Distributed
IP	Intraperitoneal
IVT	In Vitro Transcription
kb	Kilobase
kDa	KiloDalton
LCM	Laser Capture Microdissection
MAC	Mitochondrial Apoptosis-induced Channel
MAS	Microarray Suite
MB	Myc Homology Box
MBEI	Model-Based Expression Index
MCM	Minichromosome Maintenance
MES	2-(*N*-morpholino) Ethanesulfonic acid
MGB	Minor-Groove Binding
MGED	Microarray Gene-expression Database

MIAME	Minimum Information About a Microarray Experiment
MMP	Matrix metalloproteinase
MOMP	Mitochondrial Outer Membrane Permeability
MTC	Multiple Testing Correction
NCBI	National Centre for Biotechnology Information
NOD	Non-Obese Diabetic
NSB	Non-Specific Binding
NTD	Amino-Terminal Domain
OCT	Optimal Cutting Temperature
Oligo(dT)	Oligo-Deoxythymidine
PBS	Phosphate Buffered Saline
PBSt	Phosphate Buffered Saline + 0.1 % Tween
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PFA	Paraformaldehyde
PLM	Probe-Level Model
PP	Pancreatic Polypeptide
QC	Quality Control
qRT-PCR	Quantitative Real-Time Reverse-Transcriptase Polymerase Chain Reaction
RAM	Random Access Memory
RIN	RNA Integrity Number

RIP	Rat Insulin Promoter
RKE	Rat Kidney Epithelial
RM	*RIP7-Bcl$_{XL}$/pins-mycER^{TAM}* Double Transgenic Mouse
RMA	Robust multi-chip averaging
RNA	Ribonucleic Acid
RNase	Ribonuclease
ROS	Reactive Oxygen Species
rpm	Revolutions Per Minute
rRNA	Ribosomal Ribonucleic Acid
RSS	Residual Sum of Squares
RT	Room Temperature
SAGE	Serial Analysis Of Gene-expression
SAPE	Streptavidin-phycoerythin
SED	Standard Error of the Difference
SNP	Single Nucleotide Polymorphism
SOM	Self-Organising Map
SS	Sum of Squares
SSPE	Saline-Sodium Phosphate-EDTA
T	Thymine
TA	Transit Amplifying
TBE	Tris/Borate/EDTA buffer

TNF	Tumour Necrosis Factor
U	Uracil
VIC	2′-Chloro-7′-Phenyl-1,4-Dichloro-6-Carboxyfluorescein
VT	Vehicle Treated
WT	Wild Type

Conventions

The following conventions have been used throughout this thesis to ensure consistency throughout:

1. Protein names are written in standard font, in lower case with a capitalised first letter (e.g. Bcl2, Myc and Insulin).
2. Gene names are written in *italic* font, in lower case (e.g. *bcl2*, *c-myc*, *ins2*).
3. All gene and protein names and abbreviations used throughout the main text are given in Appendix A: Gene abbreviations
4. Relevant terms of interest are highlighted in *italic* font.
5. Whilst every care is taken to clearly differentiate between the terms 'probe set' and 'gene', the two terms are used interchangeably for brevity when no ambiguity exists.
6. The term 'suprabasal keratinocytes' will be replaced by the more general term 'skin' when no ambiguity exists.
7. The term 'β-cells' will be replaced by the more general term 'pancreas' when no ambiguity exists.
8. The term 'significance' will generally be used to refer to 'statistical significance' (i.e. p-value < 0.05).
9. Measurement errors are given as $mean \pm standard\ deviation$.

Abstract

The *c-myc* oncogene is over-expressed or deregulated in many human cancers. *c-myc* encodes a transcription factor, the oncoprotein c-Myc (Myc), which acts as a master regulator of genes involved in such diverse cellular processes as replication and growth, loss of differentiation, invasion, and angiogenesis. Myc can also act as its own tumour suppressor by promoting cell death in the form of apoptosis. Thus, for putative cancer cells to arise, apoptosis must be blocked. Conditional MycER^{TAM} transgenic mice allow regulated activation of Myc in distinct cell populations (skin suprabasal keratinocytes and pancreatic islet β-cells) and have highlighted contrasting behaviour between these two adult tissues *in vivo*: proliferation in the skin, and apoptosis in the pancreas.

Given the crucial dependence on tissue location *in vivo*, we still do not know enough about the key divergence in Myc-regulated genes and proteins under conditions favouring opposing outcomes. To address this, we performed high-throughput transcriptome analysis using oligonucleotide microarrays. The *in vivo* transcriptional response to deregulated Myc was analysed for skin keratinocytes and laser-captured pancreatic islets following a time-course of MycER^{TAM} activation. Due to the multi-factorial nature of the experimental design, novel statistical tools were developed allowing the use of linear models for inference of changes in gene-expression based on multiple experimental variables.

Comparison of the transcriptional response between the two tissues identified potential signalling pathways which may promote apoptosis of β-cells or survival of skin keratinocytes: the DNA damage response pathway, and the Insulin-like growth factor 1 (Igf1) signalling pathway respectively. In addition, a marked change in expression was detected in members of the steroid hormone-regulated Kallikrein serine protease family in suprabasal keratinocytes but not for β-cells. These have been found to play an important role in regulating Igf1/Igf1-receptor ligation through proteolysis of the Igf1 binding proteins, are previously categorised markers for several human cancers, and may indicate a tissue-specific regulatory mechanism for determining ultimate Myc function *in vivo*.

Chapter 1 Introduction

1.1 The *c-myc* oncogene

Strain 29 of the avian myelocytomatosis retrovirus (MC29), a replication-defective acute leukaemia transforming virus, was found to induce a number of diseases in chickens, particularly cancers such as carcinomas, leukemias and sarcomas (Mladenov *et al.*, 1980). This oncogenic potential is due to a single gene, the *v-myc* oncogene, centrally located in the 5.5-kilobase (kb) viral genome of MC29 and the related retroviruses MH2, CMII and OK10 (Duesberg and Vogt, 1979; Roussel *et al.*, 1979). The *myc* (myelocytomatosis) family of genes, comprising *c-myc* (cellular), *n-myc* (neuronal) and *l-myc* (lung), were subsequently identified due to their homology to the transforming gene *v-myc*, and are amongst the earliest examples of *oncogenes* – transforming genes with the potential to cause cancer – to be identified. The cellular homolog of the viral oncogene, *c-myc*, was first discovered due to its homology to a 1,500 nucleotide region of *v-myc*, with an additional 1,100 base pair (bp) intron-like region (Vennstrom *et al.*, 1982). Further members *n-myc* and *l-myc* were later discovered in the amplified sequences of neuroblastoma cells (Schwab *et al.*, 1983) and small cell lung tumours (Nau *et al.*, 1985; Ryan and Birnie, 1996) respectively.

1.1.1 Myc protein function

The protein product of *c-myc* – Myc – is a transcription factor localised in the cell nucleus that is conserved throughout vertebrate evolution (Persson and Leder, 1984). Myc has been found to regulate the expression of a wide range of cellular

targets involved in a variety of diverse functions including cell growth, proliferation, loss of cell-cell contact, and angiogenesis (Amati *et al.*, 1998; Dang, 1999; Coller *et al.*, 2000; de Alboran *et al.*, 2001; Eisenman, 2001; Trumpp *et al.*, 2001; Pelengaris *et al.*, 2002a; Pelengaris and Khan, 2003).

During normal embryonic development of mouse and humans, Myc is expressed ubiquitously in growing tissues (Schmid *et al.*, 1989; Hirvonen *et al.*, 1990). It has been shown to be crucial for normal embryonic development, as deletion of both alleles results in embryonic lethality in mice (Davis *et al.*, 1993). In adult mice, it is found highly expressed in tissue compartments possessing high cellular turnover (e.g. skin epidermis and gut), whilst it is undetectable in cells that have exited the cell cycle.

The predominant role of Myc under normal physiological conditions is promotion of proliferation (Persson *et al.*, 1984; Henriksson and Luscher, 1996; Amati *et al.*, 1998) and inhibition of terminal differentiation (Freytag, 1988; Hoffman-Liebermann and Liebermann, 1991; Selvakumaran *et al.*, 1993). Expression of *c-myc* is kept under tight control by the presence of growth and survival factors. Disruption of these normally tightly regulated processes, such as through mutations in the *c-myc* gene, can lead to unchecked proliferation. In fact, deregulated or up-regulated expression of *c-myc* is a frequent feature of human cancers (Hueber *et al.*, 1997; Amati *et al.*, 1998; Nesbit *et al.*, 1999; Schlagbauer-Wadl *et al.*, 1999), which will be discussed in more detail in Section 1.1.3. The *c-myc* gene is therefore termed a *proto-oncogene*, indicating that it has oncogenic potential but is kept under tight control by mitogenic stimuli.

Deregulated expression of Myc leads to unchecked proliferation, but has also been shown to sensitise cells to apoptotic stimuli in several cell types under conditions of depleted growth and survival factors (Askew *et al.*, 1991; Evan *et al.*, 1992; Shi *et al.*, 1992). A role for oncogenic Myc in sensitising cells to external apoptotic stimuli through the tumour necrosis factor (TNF) death receptor CD95/Fas has also been suggested (Janicke *et al.*, 1994; Klefstrom *et al.*, 1994; Hueber *et al.*, 1997).

This tight linkage between growth and death pathways creates a defence mechanism for the body against progressive diseases such as cancer; upon aberrant induction of cell cycle progression by deregulated Myc, cells become sensitised to a wide range of stimuli (such as DNA damage and hypoxia) which promote apoptosis and provide an in-built tumour suppressor mechanism both *in vitro* (Wagner *et al.*, 1994; Tanaka *et al.*, 2002; Vafa *et al.*, 2002) and *in vivo (Alarcon et al., 1996)*. The link between these two seemingly opposing functions – proliferation and apoptosis – is not limited to *c-myc*. Several other cell cycle-associated genes, such as *e2f*, *e1a* and *c-fos*, are also found to show apoptotic functionality under certain conditions (Harrington *et al.*, 1994b).

The exact mechanisms by which Myc elicits the vast host of biological responses for which it appears to be responsible are not yet fully understood. Observations of changing gene-expression following over-expression of Myc in primary rat fibroblasts (Coller *et al.*, 2000), conditional *c-myc* expression in human cell cultures (Schuhmacher *et al.*, 2001; Fernandez *et al.*, 2003) and comparison of *c-myc*-null rat cell lines with normal cells (Guo *et al.*, 2000; Watson *et al.*, 2002; O'Connell *et al.*, 2003) have identified a wide range of potential transcriptional targets for deregulated Myc. Currently, around 1,700 genes have been classified as putative Myc targets (Zeller *et al.*, 2003) using methods such as serial analysis of gene-expression (SAGE) (Menssen and Hermeking, 2002), DNA microarrays (Coller *et al.*, 2000) and subtractive hybridisation (Lewis *et al.*, 1997). It has been hypothesised that Myc may have the potential to regulate up to 15 % of the entire genome (Patel *et al.*, 2004), leading to it being described as a 'master regulator' of gene-expression.

It is still not clear if this regulatory role is due to direct transcription, or if there is in fact a smaller subset of direct Myc targets responsible for activation of other genes downstream. Experiments using chromatin immunoprecipitation (ChIP) have confirmed several putative genes as direct Myc targets by identifying the association of Myc with promoter regions (Fernandez *et al.*, 2003; Li *et al.*, 2003). Further studies have shown Myc to be a relatively weak transcription factor, promoting changes in expression in the order of only 2- to 5-fold (Grandori and

Eisenman, 1997; Cole and McMahon, 1999), and this weak transcriptional activity can make identification of putative target genes difficult.

Recent studies suggest an additional non-transcriptional role for Myc in gene regulation. In the work of Dominguez-Sola *et al*. (2007), Myc over-expression in human and mouse cells with no RNA transcriptional activity showed that Myc interacts directly with the pre-replicative complex, playing a significant non-transcriptional role in DNA replication. Also, Cowling and Cole (2007) showed that Myc mutants, modified to have no DNA binding affinity, were able to promote cell growth in *c-myc* (-/-) mice through binding of the Myc transactivation domain to the transcriptional initiation sites of target genes, increasing mRNA cap methylation and increasing the rate of translation. It has been noted that a number of previously characterised Myc-target genes may be regulated by this Myc-induced translational mechanism, suggesting a further level of complexity to the Myc-regulatory web. It is important to note that such effects are not identifiable through transcription-focused analysis methods, such as SAGE and microarrays.

Many of the studies into oncogenic Myc function thus far have involved a combination of *in vitro* and *in silico* work. Whilst such studies have undoubtedly provided an understanding of the functional role of this enigmatic transcription factor, it is not clear which of the many potential Myc-targets are in fact responsible for the ultimate development of specific phenotypes. Importantly, such studies fail to identify the divergence in Myc-induced gene-expression *in vivo* under circumstances where Myc may promote opposing outcomes, such as proliferation and apoptosis. It is clear that tissue context plays a major part in determining the ultimate role of Myc (Section 1.2.4), so it is of great interest for therapeutic intervention to understand the complex interactions involved in Myc-regulated tumourigenesis *in vivo*.

1.1.2 Myc protein structure and transcriptional control

The Myc protein contains several distinct regions (Figure 1.1.1). The carboxy-terminal domain (CTD) contains a basic-helix-loop-helix-leucine-zipper (bHLH-LZ) which associates with a similar domain in another nuclear protein, Max, to form a heterodimeric complex. This Myc-Max heterodimer binds to canonical (CACGTG) and non-canonical E-box sequences within the promoter region of target genes to initiate transcription (Blackwell *et al.*, 1990; Amati *et al.*, 1992; Kretzner *et al.*, 1992; Blackwell *et al.*, 1993; Prochownik and VanAntwerp, 1993). The amino-terminal domain (NTD) of the Myc protein contains the prominent evolutionarily conserved domains; the Myc homology boxes (MBs).

Once bound to the target gene DNA, the conserved transactivation regions MBI and MBII recruit the highly conserved co-activating protein Trrap (transformation/transcription domain associated protein) (McMahon *et al.*, 1998), a common feature in many complexes including histone acetyltransferases (HATs), kinases, basal factors and ubiquitin ligases. Association of further cofactors, such as the F-box protein Skp2 (Kim *et al.*, 2003; von der Lehr *et al.*, 2003) and the cAMP-response-element-binding protein Cbp (Vervoorts *et al.*, 2003), stabilises the complex and allows recruitment of the HAT Gcn5 (general control of amino-acid synthesis 5) to the E-box site. This acetylates nucleosomal histones H4 and, to a lesser extent, H3 in adjacent regions (Bouchard *et al.*, 2001; Frank *et al.*, 2001) and allows binding of ribonucleic acid (RNA) polymerases I and III to initiate transcription of the target gene (White, 2005). A further conserved element within the central region of Myc, MBIII, has also been identified (Herbst *et al.*, 2004) which is important in stability of the Myc protein and Myc-mediated transcriptional repression. Disruption of MBIII increases the apoptotic potential of Myc both *in vitro* and in mouse models, indicating an anti-apoptotic role for the MBIII domain (Herbst *et al.*, 2005). A final Myc-box element – MBIV – has been identified, which shows both mild transactivational activity, and a role in Myc-induced apoptosis (Cowling *et al.*, 2006).

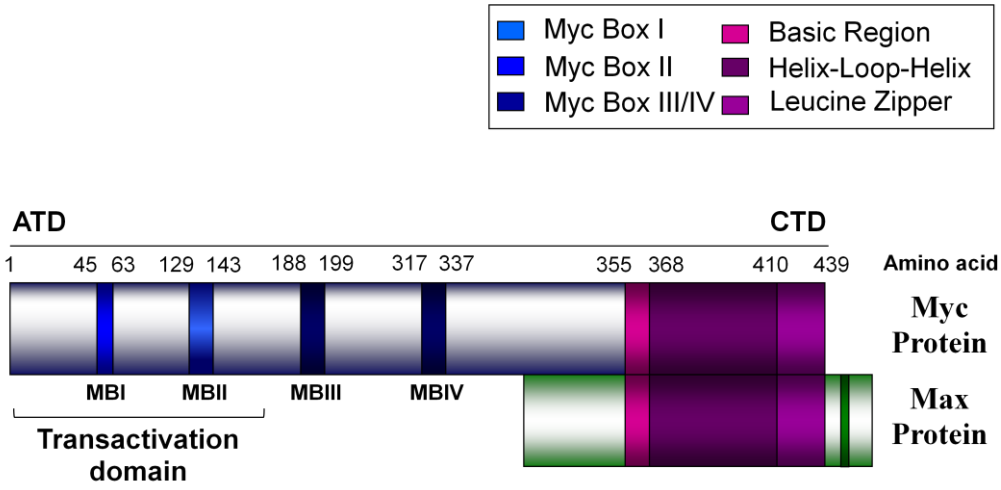

Figure 1.1.1: Structure of the Myc transcription factor and hetero-dimerisation with Max

The Myc protein interacts with its partner protein Max at the carboxy-terminal basic-helix-loop-helix leucine zipper (bHLH-LZ) to form a heterodimeric complex that regulates transcription of target genes. Myc-Max heterodimers bind to E-box sequences in the promoter regions of target genes and recruit various cofactors (such as Trrap, Skp2 and Cbp) to the highly conserved Myc Box regions, MBI and MBII. These associate with histone acetyltransferases such as Gcn5, which acetylate nucleosomal histones and allow binding of RNA polymerases I and III and initiation of transcription of the target gene. Further conserved regions, MBIII and MBIV, have been shown to be involved in protein stability and Myc-induced apoptosis. Adapted from Pelengaris et al. (2002a).

Transcriptional regulation of target genes by Myc is entirely dependent on dimerisation to its partner protein Max (Amati *et al.*, 1993a; Amati *et al.*, 1993b). However Max is also found to bind with other members of the Mad and Mnt protein families, including binding with itself in homodimeric complexes (Grandori *et al.*, 2000). Mad-Max heterodimeric complexes also bind to E-box promoter sequences, and associate with the co-repressor Sin3 at the NTD of Max. This complex recruits class T histone deacetylases (HDACs) Hdac1 and Hdac2 which deacetylate histones H3 and H4, promoting transcriptional repression (Sommer *et al.*, 1997). Thus Max also acts as an antagonist to Myc transcriptional activity in partnership with Mad, and transcription rates are dependent on interactions between the Myc, Mad and Max proteins (Ayer *et al.*, 1993).

As well as its ability to promote transcription of target genes, Myc is also able to repress the expression of target genes. Since the Myc-Max complex does not appear to associate directly with repressors or co-repressors, it has been suggested that Myc represses gene-expression not by direct binding to target gene DNA, but in an indirect manner through antagonistic associations with other positively acting transcription factors (Orian and Eisenman, 2001). Miz1 (Myc-interacting zinc finger protein-1) is a transcription factor that binds directly to the promoter region of target genes, such as the cyclin-dependent kinase inhibitor (CDKI) $p15^{Ink4b}$ (*cdkn2b*). Miz1-induced expression of *cdkn2b* inhibits the cyclin-dependent kinase (CDK) Cdk4 and promotes cell cycle arrest (Section 1.1.4). The initiator region of Miz1 also associates directly with the bHLH-LZ region of Myc when in partnership with Max, preventing association with the co-activator p300 and inhibiting transcription (Peukert *et al.*, 1997; Staller *et al.*, 2001; Herold *et al.*, 2002). The specific role of Miz1 in apoptosis was found when it was shown that inhibition of transcription through association of Myc with Miz1 is essential for Myc-mediated apoptosis (Patel and McMahon, 2006). Co-immunoprecipitation analyses identified direct interactions between Myc and Sp1/Sp3, and glutathione S-transferase/Sp1 fusion protein experiments identified direct association between the central region of the Myc protein and the c-terminal DNA binding zinc finger of Sp1. Sequestration of these factors prevents transcription of target genes involved in promoting cell cycle arrest, such as *cdkn1a* which encodes the CDKI

$p21^{Waf1/Cip1}$ (Gartel *et al.*, 2001). Thus association of Myc with factors involved in the expression of inhibitors of cell cycle progression, leading to indirect downregulation, may further contribute to cell proliferation (Section 1.1.4).

1.1.3 Myc and cancer

The term *cancer* can be applied to over 100 diseases and can affect any part of the body. It is currently one of the leading causes of death in humans, and accounts for around 13 % of all deaths worldwide.[1] Cancer occurs when genomic insults result in the bypass of normal proliferative controls and cells are allowed to propagate without the requirement of exogenous mitogenic stimulation from the micro-environment. Tumourigenesis is a multistage procedure that begins with some genetic lesion that confers a growth advantage over other cells and inhibits regulatory control of proliferation. In comparison to normal cells, cancer cells no longer require exogenous mitogenic stimulation from their tissue micro-environment in order to proliferate. These cells begin to grow uncontrollably, whilst simultaneously losing their differentiated state and losing contact with each other and the surrounding environment. Expansion of cancer cells leads to the formation of tumours, which recruit vasculature to the site to provide oxygen and nutrients required for growth. Often, cancer cells are able to break away from the primary tumour site into the lymphatic or vascular system, where they travel to secondary sites in the body and may form further tumours in a process termed *metastasis*.

The cell replicating machinery utilises a number of defensive mechanisms to prevent aberrant proliferative activity, which work to avoid the development of cancers that would otherwise develop over time. During normal cell cycle control, oncogenic factors cooperate with one another to protect against uncontrolled

[1] World Health Organisation, February 2006 – www.who.int

growth. Due to this tight control of proliferative activity, mutations must occur in several cooperating oncogenes and tumour suppressor genes in order to provide the circumstances necessary for neogenesis. The complex web of genetic events involved in cancer formation means that inactivation of a single genetic product involved in tumourigenesis may often be insufficient to reverse cancer formation. This can make identification of putative targets for cancer therapy difficult. Current treatments consist of removal of primary tumours through surgery and destruction of metastatic cancer cells using 'shotgun' approaches such as radiotherapy and chemotherapy. These treatments are non-specific in their action, leading to the destruction of both healthy and malignant cells.

The *c-myc* oncogene was first implicated in human cancer in the 1980's when it was shown to be found on a small region of chromosome 8 (q24 – qter) that is translocated to chromosome 2, 14 or 22 in Burkitt's lymphoma (Dalla-Favera *et al.*, 1982). These regions contain antibody-encoding genes and lead to deregulated expression of *c-myc*. Constitutive over-expression of Myc was shown to immortalise rat fibroblasts and to prevent cell cycle withdrawal (Land *et al.*, 1983; Mougneau *et al.*, 1984), whilst cellular transformation *in vitro* was found to require additional lesions (Land *et al.*, 1986; Lugo and Witte, 1989; Reed *et al.*, 1990; Fanidi *et al.*, 1992; Morgenbesser and DePinho, 1994; Pelengaris *et al.*, 2002a).

Deregulation of Myc can occur due to numerous mechanisms, both directly via stabilisation of Myc messenger RNA (mRNA) transcripts or increased initiation of transcription (up-regulation) due to mutations in the internal ribosome entry site (Bernasconi *et al.*, 2000; Chappell *et al.*, 2000), or indirectly via activation of upstream mitogenic signalling cascades (Barone and Courtneidge, 1995; Kolligs *et al.*, 2000; Bowman *et al.*, 2001; Chiariello *et al.*, 2001). Genomic alterations in the *c-myc* locus, such as that seen in Burkitt's lymphoma, can also play a part in deregulated Myc activity, as can stabilisation of the Myc protein through mutations resulting in inefficient ubiquitination-mediated protein degradation (Salghetti *et al.*, 1999; Gregory and Hann, 2000). It has been shown that deregulated expression of *c-myc*, specifically up-regulation, is a common feature

of a large number of human cancers (Ryan and Birnie, 1996; Amati *et al.*, 1998; Nesbit *et al.*, 1999; Schlagbauer-Wadl *et al.*, 1999) and is often associated with aggressive poorly differentiated tumours.

However, given that many human tumours are advanced at the time of detection and may contain many genetic alterations, it is difficult to ascertain the point at which *c-myc* deregulation occurred. Conventional Myc transgenic models allow constitutive expression of the oncogene in target tissues using specific promoters, and have supported the view that deregulated Myc is important to the formation of certain cancers (Adams *et al.*, 1985; Blyth *et al.*, 1995). However, since expression of the oncoprotein occurs throughout development, such models preclude the observation of the proposed apoptotic tumour suppressive function of Myc.

Regulatable transgenic mice, such as the MycER^{TAM} model described in Section 1.2, allow controlled activation and deactivation within target tissues. This provides a model for deregulated expression of proteins such as Myc within the adult organism, allowing observation of subsequent downstream effects. Whilst such models of widespread activation within all cells of the target tissue do not precisely represent the process of sporadic mutations seen during tumour formation, the ability to switch oncogenes on and off at will has vastly improved physiological cancer models by offering the experimenter exquisite control of the 'time 0' of genetic alteration. Such models also allow analysis of tumour reversal following deactivation of the transgenic oncogene, allowing testing of the therapeutic benefits of targeting specific oncogenes (Felsher, 2003; Giuriato *et al.*, 2004).

Inactivation of the single initiating oncogene in many cancer models is sufficient not only for reversal of the primary tumour, but also for reversal of invasive and metastatic lesions (Pelengaris *et al.*, 1999; Moody *et al.*, 2002; Pelengaris *et al.*, 2002b; Karlsson *et al.*, 2003; Felsher, 2004; Marinkovic *et al.*, 2004). Given that such lesions may contain a number of further genetic mutations subsequent to oncogene activation, this is an exciting prospect for oncogene-targeted cancer therapies. Jain *et al.* (2002) showed that brief inactivation (10 days) of Myc was

sufficient for regression of Myc-induced osteogenic sarcomas in transgenic mice, and that subsequent reactivation of Myc led not to restoration of the neoplastic phenotype, but to extensive apoptosis.

However, in contrast to this, brief inactivation of Myc in tumour onset is not sufficient to maintain tumour regression in suprabasal keratinocytes and pancreatic β-cells, and reactivation of Myc leads to restoration of the oncogenic properties of Myc (Pelengaris *et al.*, 2004). A similar study by Shachaf *et al.* (2004) used a lineage-tracking approach to demonstrate that inactivation of Myc in Myc-induced hepatocellular carcinomas resulted in regression, but that some tumour cells remained dormant (often for prolonged periods of time) until Myc expression was reinitiated, after which these dormant cells contributed to cancer progression. It is therefore clear that further study into the environmental context and genetic basis of particular tumours is required to determine the possibility of using oncogene-targeted therapies.

The role of Myc in tumour progression is clearly complex, and its presence in a broad range of human cancers, particularly those with a poor prognosis, suggests that Myc deregulation may be used as a diagnostic marker for cancer in certain tumour types. Furthermore, inactivation of Myc or downstream targets of Myc may provide important therapeutic strategies in the fight against cancer, and several such strategies are currently under investigation (Robson *et al.*, 2006).

1.1.4 Myc and proliferation

As previously described, one of the key biological functions of Myc is its ability to promote cell cycle progression (Amati *et al.*, 1998; Dang, 1999; Eilers, 1999; Amati, 2001). Myc expression is virtually undetectable in quiescent cells *in vitro.* However, upon mitogenic or serum stimulation, Myc levels are rapidly induced at both the mRNA and protein levels, and cells begin the proliferative procedure (Persson *et al.*, 1984; Hann *et al.*, 1985). The levels of *c-myc* mRNA in

proliferating cells subsequently decline to low but detectable steady-state levels. Removal of mitogenic signalling, by reducing serum or growth factor levels, results in a decline in Myc protein levels to undetectable levels and arrest of the cell cycle (Davis *et al.*, 1993). Myc expression is well correlated with active proliferation during development, and down-regulation accompanies mitotic arrest and onset of differentiation. Myc is essential for embryonic development, as shown by embryonic lethality at day 9.5-10.5 following targeted gene disruption of both Myc alleles in embryonic stem cells (Schmid *et al.*, 1989; Hirvonen *et al.*, 1990).

Myc was first associated with cell cycle control after correlations were seen between Myc and the expression of the rate-limiting eukaryotic translation initiation factors Eif4e and Eif2α (Rosenwald *et al.*, 1993) now known to be direct Myc targets (Jones *et al.*, 1996; Coller *et al.*, 2000). Further studies identified a direct role for Myc in the growth of invertebrate (Johnston *et al.*, 1999) and mammalian cells (Iritani and Eisenman, 1999; Schuhmacher *et al.*, 1999; Beier *et al.*, 2000).

The cell cycle of eukaryotic organisms consists of two main phases – the *interphase* and *mitosis*. The interphase is the stage during which the cell prepares itself for proliferation by synthesising proteins required for DNA replication (G_1 growth phase), duplicating their DNA (synthesis, or S-phase), and finally by producing the proteins required for mitosis (G_2 growth phase). When the cell is prepared, it undergoes mitosis (M-phase) and splits to form two identical daughter cells. Non-proliferating or post-mitotic cells enter a third resting phase, G_0, and remain quiescent until acted upon by mitotic stimuli. Myc-mediated promotion of cell proliferation is a result of activation and repression of key cell cycle regulatory genes, particularly those involved in G_1/S-phase progression (Figure 1.1.2).

Progression through the cell cycle is largely controlled by the interactions of a group of proteins – the cyclins – and their respective kinases. G_1/S-phase progression of eukaryotic cells is controlled by the activities of the CDK complexes, Cyclin D-Cdk4/6 and Cyclin E-Cdk2, which together promote hyper-

phosphorylation of the Rb tumour suppressor protein. This results in the release of E2f family transcription factors from the phosphorylated Rb protein, which activate expression of genes required for S-phase.

Myc induces Cyclin E-Cdk2 activity early in the G_1 phase of the cell cycle, which is regarded as an essential event in Myc-induced G_1/S-phase progression (Steiner *et al.*, 1995; Berns *et al.*, 1997). This occurs via direct activation of the Myc target genes c*cnd2* (Cyclin D2) (Bouchard *et al.*, 2001) and *cdk4* (Hermeking *et al.*, 2000). Increased expression of the Cyclin D2 and Cdk4 proteins leads to sequestration of the CDKI $p27^{Kip1}$ by the Cyclin D2-Cdk4 complex (Bouchard *et al.*, 1999; Perez-Roger *et al.*, 1999). This association leads to the subsequent proteosomal degradation of $p27^{Kip1}$ by the products of two further Myc target genes, c*ul1* and c*ks* (O'Hagan *et al.*, 2000). By preventing the binding of the CDKI $p27^{Kip1}$ with Cyclin E-Cdk2 complexes, they are free to become phosphorylated by cyclin activating kinase (Cak), a further Myc target (Muller *et al.*, 1997; Perez-Roger *et al.*, 1999). This allows Rb hyper-phosphorylation, subsequent release of E2f, and G_1/S-phase cell cycle progression.

Although Cyclin D2 is known to be an essential downstream effector of Myc in promoting cell proliferation (Bouchard *et al.*, 1999; O'Hagan *et al.*, 2000), the D cyclins do not appear to be required for Myc-induced apoptosis *in vitro* (Berns *et al.*, 1997). This indicates that the two major functions of Myc – proliferation and apoptosis – may involve distinct sets of downstream mediators.

Myc has also been found to promote cell cycle progression by repressing transcription of inhibitors of cell cycle progression, such as $p15^{Ink4b}$ (*cdkn2b*) (Staller *et al.*, 2001) and $p21^{Cip1}$ (c*dkn1a*) (Gartel *et al.*, 2001; Herold *et al.*, 2002), both of which are involved in cell cycle arrest. Association of the Myc-Max heterodimeric complex with Miz1 blocks association of Miz1 with its co-activator p300, preventing Miz1-mediated expression of c*dkn2b* ($p15^{Ink4b}$), and direct association of Myc with the Sp1 and Sp3 transcription factors prevents expression of the Sp1/Sp3 target c*dkn1a* ($p21^{Cip1}$).

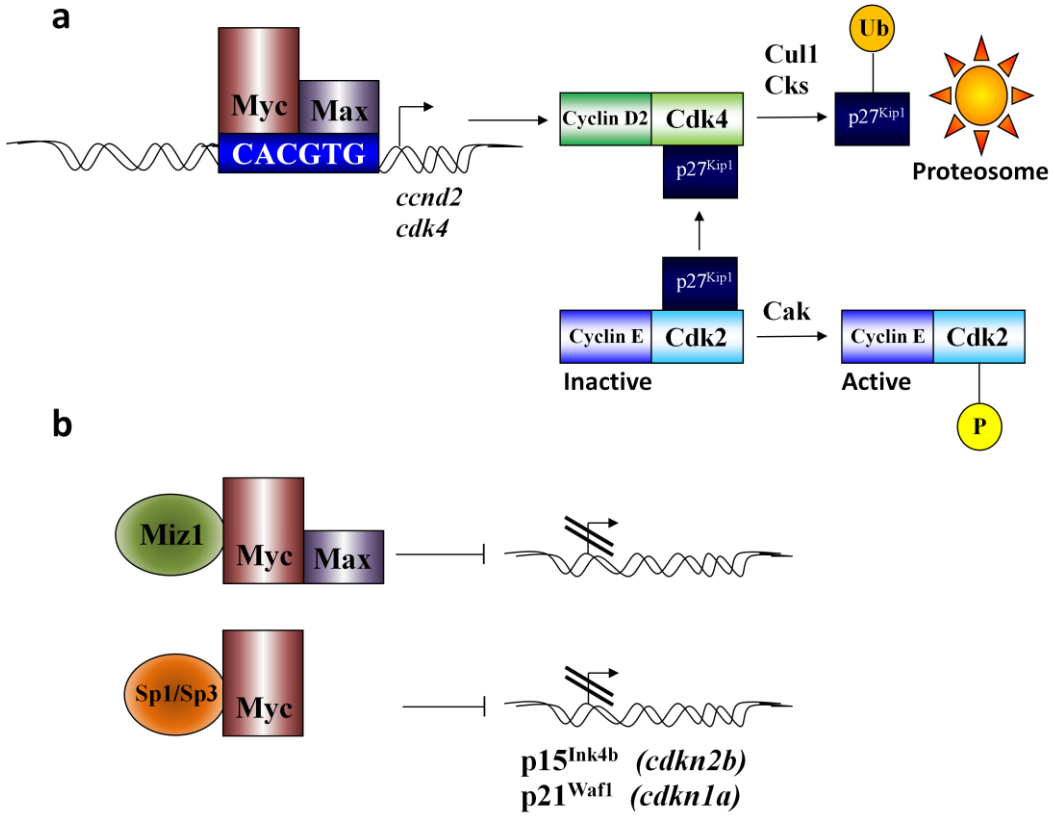

Figure 1.1.2: Activation and repression of target genes by Myc results in cell cycle progression

a) Myc-Max heterodimers promote transcription of target genes for Cyclin D2 (*ccnd2*) and Cyclin-dependent kinase 4 (c*dk4*). Cyclin D2-Cdk4 complexes sequester the CDK inhibitor $p27^{Kip1}$, which is subsequently ubiquitinated by products of two further Myc-target genes, *cul1* and *cks*, leading to proteosomal degradation. $p27^{Kip1}$ is thus prevented from binding to and inhibiting Cyclin E-Cdk2 complexes, allowing phosphorylation by cyclin activating kinase (Cak), a further Myc-target. Activated Cyclin E-Cdk2 complexes hyperphosphorylate the Rb tumour suppressor protein, allowing release of the E2f transcription factor and G_1/S-phase cell cycle progression. b) Myc can also indirectly repress activation of the CDK inhibitors, $p15^{Ink4B}$ and $p21^{Cip1}$, which are involved in cell cycle arrest. By association of Myc-Max heterodimers with the transcription factor Miz1, or by direct interactions with transcription factors Sp1 or Sp3, Myc prevents transactivation of $p15^{Ink4b}$ (*cdkn2b*) and $p21^{Cip1}$ (*cdkn1a*), preventing cell cycle arrest and allowing proliferation to continue. Adapted from (Robson *et al.*, 2006).

1.1.5 Myc and apoptosis

1.1.5.1 Apoptosis

Apoptosis is a form of programmed cell death that generally confers advantages to the organism, such as through the destruction of damaged cells (Kerr *et al.*, 1972). Putative cancer cells are no longer subject to mitotic control from external signals, and must avoid apoptosis in order for tumours to arise. The net expansion of a clone of transformed cells following oncogenic deregulation comes about through a combination of increased proliferation of cells and decreased apoptosis. The signal for a cell to become apoptotic may come from external signals (*extrinsic* apoptotic pathway) or internally due to stress responses through the mitochondria (*intrinsic* apoptotic pathway) (Figure 1.1.3).

The effectors of apoptosis are a family of cysteine proteases known as *caspases* (Yuan *et al.*, 1993; Alnemri *et al.*, 1996). Initiator caspases, such as Caspase 9, are sequentially activated through proteolytic cleavage in a cascade of caspase activation, until so-called executioner/effector caspases, such as Caspase 3, are activated. These lead to proteolytic degradation of cellular components and further morphological changes, such as chromatin condensation, fragmentation of DNA, nucleus fragmentation and formation of nucleosomal units, blebbing of the cell membrane, and separation of the cell into apoptotic bodies which are removed by phagocytes to avoid inflammation (Wyllie *et al.*, 1980).

The intrinsic pathway is governed through signalling from the mitochondria, a cellular organelle responsible for aerobic respiration. Internal stresses such as hypoxia, DNA damage, viral infection, nutrient deprivation or radiation result in association of the pro-apoptotic Bcl2 (Beta-Cell Leukaemia/Lymphoma) family members to the mitochondrial outer membrane (Wolter *et al.*, 1997; Luo *et al.*, 1998). This results in an increase in mitochondrial outer-membrane permeability (MOMP) (Kluck *et al.*, 1999) and release of Cytochrome c (Cyto c) through the mitochondrial apoptosis-induced channel (MAC) into the cytosol (Pavlov *et al.*,

2001). Release of Cytochrome c through this channel is governed by Bcl2 family members residing in the mitochondrial outer membrane, and is a key process in the onset of apoptosis. The Bcl2 protein family all share one or more of the Bcl2 homology (BH) domains, BH1, BH2, BH3, and BH4 (Yin *et al.*, 1994; Chittenden *et al.*, 1995; Gibson *et al.*, 1996; Zha *et al.*, 1996a). Many also contain a transmembrane domain, allowing functional control over the release of apoptotic factors from the mitochondria (see below).

Broadly, Bcl2 protein family members can be split into three main groups: pro-apoptotic members such as Bax and Bak which contain BH regions BH1-BH3, anti-apoptotic members such as Bcl2, Bcl_{XL} and Bcl_{W} which contain BH1-BH4 regions, and pro-apoptotic members such as tBid, Bim, Bad, Puma and Noxa which contain only the BH3 region. Bcl2 members act competitively to control MOMP. BH3-only proteins have been shown to assist in the activation of Bax via oligomerisation of the proteins, leading to migration from the cytoplasm to the outer mitochondrial membrane (Gross *et al.*, 1998; Khaled *et al.*, 1999; Eskes *et al.*, 2000). Once localised to the membrane, Bax forms the MAC and induces release of Cytochrome c (Dejean *et al.*, 2005; Dejean *et al.*, 2006). Anti-apoptotic members Bcl2 and Bcl_{XL} prevent formation of the MAC by competitively forming heterodimers with Bax (Dejean *et al.*, 2006). A further role of BH3-only Bcl2 members in promoting apoptosis is via association with anti-apoptotic Bcl2 members, neutralising their ability to bind and inhibit Bax (O'Connor *et al.*, 1998; Strasser *et al.*, 2000; Adams and Cory, 2007).

Once released, cytosolic Cytochrome c binds to apoptotic protease activating factor 1 (Apaf1) in a cyclic structure with 7-fold symmetry known as the *apoptosome*, or "wheel of death" (Acehan *et al.*, 2002). Binding of adenosine triphosphate (ATP) results in a conformational change in the apoptosome, allowing binding of Pro-Caspase 9 (Li *et al.*, 1997). Once bound to this complex, intrinsic autocatalytic cleavage of Pro-Caspase 9 at Asp-315 produces large and small subunits, including an active Caspase 9 subunit molecule with proteolytic ability (Srinivasula *et al.*, 1998). Active Caspase 9 in turn activates the executioner Caspase 3 in a caspase cascade, leading to cellular degradation and

initiation of apoptosis. MOMP and formation of the MAC also result in the release of pro-apoptotic factors into the cytosol. Such factors include apoptosis-inducing factors (AIFs) and Endonuclease G (Endog), which act to induce apoptotic changes such as chromatin condensation and DNA fragmentation (Susin *et al.*, 1996; Susin *et al.*, 1997), and second mitochondria-derived activator of caspases (Smac), which binds to and inhibits the inhibitors of apoptosis proteins (IAP) allowing apoptosis to proceed (Du *et al.*, 2000).

The extrinsic pathway occurs predominantly through the signalling of the TNF and Fas receptors. Tnf is a cytokine produced by active macrophages that binds to the TNF receptors Tnfr1 and Tnfr2 (Carswell *et al.*, 1975). Association with Tnfr1 results in association with the TNF receptor-associated death domain (Tradd) and the Fas-associated death domain protein (Fadd), which recruits several Pro-Caspase 8 molecules in close proximity, facilitating autocatalytic cleavage and formation of proteolytically active large Caspase 8 subunits (reviewed in Chen and Goeddel, 2002; Wajant *et al.*, 2003). The Fas transmembrane receptor (also known as CD95 or Apo1), aggregates to form transmembrane Fas trimers upon binding of the trimeric Fas ligand (Schneider *et al.*, 1997). This aggregation results in internalization of the complex, association of the intracellular Fas death domain with Fadd and formation of the death-inducing signalling complex (DISC) (Kischkel *et al.*, 1995). The DISC binds to and activates Pro-Caspase 8 via self-proteolytic cleavage (Medema *et al.*, 1997).

In both cases, active Caspase 8 (also known as Fadd-like interleukin-1 beta-converting enzyme, or Flice) instigates a cascade of caspase activation, resulting in initiation of apoptosis. Alternatively, active Caspase 8 can cleave the pro-apoptotic Bcl2 member BH3-interacting domain death agonist (Bid), resulting in a truncated molecule (tBid) that translocates from the cytosol to the mitochondria (Li *et al.*, 1998b). This BH3-only Bcl2 protein family member can trigger Bax-mediated permeabilisation of the outer mitochondrial membrane and instigation of the intrinsic apoptotic pathway (Wang *et al.*, 1996; Eskes *et al.*, 2000; Korsmeyer *et al.*, 2000; Wei *et al.*, 2000). This represents one of the ways in which the intrinsic and extrinsic pathways can interact with one another.

1.1.5.2 Apoptosis and Myc

Intriguingly, it was found that oncogenes such as c-*myc*, whose cellular function largely involves cell cycle regulation, also have apoptotic properties (Askew *et al.*, 1991; Evan *et al.*, 1992; Shi *et al.*, 1992). In the study of Evan *et al.* (1992), ectopic expression of the Myc protein in Rat-1 fibroblasts cultured under conditions of depleted survival factors induced apoptosis and the eventual loss of the entire population. A widely held view of oncoprotein-induced apoptosis is that the induction of cell cycle entry 'sensitises' the cell to apoptotic stimuli, indicating a coupling between the two processes of cell proliferation and apoptosis. However, the relative availability of appropriate survival factors providing anti-apoptotic signals suppresses the apoptotic pathway. The predominant outcome of these contradictory processes is therefore dependent on the availability pro- and anti- apoptotic factors.

This indicates that cells acquiring growth-deregulating mutations *in vivo* possess an 'in-built' tumour suppressor function, which prevents expansion of potentially malignant cells. This in-built defensive mechanism must therefore be evaded for such a cell population to outgrow its environment. The synergy between oncoproteins and apoptosis-suppressing mechanisms (such as over-expression of anti-apoptotic proteins Bcl2 or Bcl_{XL}, or loss of $p19^{Arf}$ or p53 tumour suppressors – see below), suggests the importance of cooperation between oncogenic stress and evasion of tumour suppressor mechanisms in tumour expansion (Adams *et al.*, 1985; Strasser *et al.*, 1990; Blyth *et al.*, 1995; Elson *et al.*, 1995; Eischen *et al.*, 1999; Jacobs *et al.*, 1999).

Stimulation of apoptosis can arise not only through direct links with the cell cycle, but also through indirect actions that may culminate in DNA damage. Myc has been linked to the accumulation of reactive oxygen species (ROS) via E2f1-mediated inhibition of Nf-κb *in vitro* (Tanaka *et al.*, 2002; Vafa *et al.*, 2002) (Figure 1.1.3f). The consequence of this, either apoptosis or growth arrest, may be critically dependent on cell type.

Deregulated expression of Myc sensitises cells to a wide range of pro-apoptotic stimuli such as hypoxia, DNA damage and depleted survival factors (Strasser *et al.*, 1990; Askew *et al.*, 1991; Evan *et al.*, 1992; Debbas and White, 1993; Qin *et al.*, 1994). Myc has also been found to enhance sensitivity to external signalling through the Fas/CD95 (Hueber *et al.*, 1997), Tnf (Klefstrom *et al.*, 1994) and Trail (TNF-related apoptosis inducing ligand) (Lutz *et al.*, 1998) death receptors (Figure 1.1.3d), and has been shown to result in down-regulation of the inhibitor of caspase activation Flice inhibitory protein (Flip) (Ricci *et al.*, 2004) (Figure 1.1.3e). Flip inhibits the extrinsic pathway of apoptosis by competing with Pro-Caspase 8 for binding to the DISC, so Myc-induced sensitisation to death receptor stimuli may be explained in part by this down-regulation.

Early studies indicating the role of Myc in apoptosis showed that it could induce the release of Cytochrome c from the mitochondria, and ectopic addition of Cytochrome c sensitised cells to apoptosis (Juin *et al.*, 1999). Myc can also promote apoptosis through more direct effects on the expression of members of the Bcl2 protein family, such as by repressing expression of *bcl2* and bcl_{XL} (Figure 1.1.3h), both of which have anti-apoptotic protein products, or by inducing expression of pro-apoptotic members, such as *bim* (Figure 1.1.3g) and *bax* (Figure 1.1.3a) (Gross *et al.*, 1998; Khaled *et al.*, 1999; Eskes *et al.*, 2000; Martinou and Green, 2001; Soucie *et al.*, 2001; Juin *et al.*, 2002; Morrish *et al.*, 2003).

The survival factor Igf1 has been shown to inhibit Myc-induced apoptosis *in vitro* by blocking Cytochrome c release (Lowe *et al.*, 2004). Survival signals mediated via the Igf1 receptor or activated Ras can lead to activation of the Akt1 (thymoma viral proto-oncogene) serine/threonine kinase pathway (Kauffmann-Zeh *et al.*, 1997). Activated Akt1 phosphorylates the pro-apoptotic BH3-only Bcl2 family protein Bad resulting in its sequestration and inactivation by the cytosolic 14-3-3 proteins (Zha *et al.*, 1996b) (Figure 1.1.3j). This prevents Bad-mediated inhibition of the anti-apoptotic Bcl2 family member Bcl_{XL}, leading to prevention of Bax-mediated release of Cytochrome c from the mitochondria (Zha *et al.*, 1996b; del Peso *et al.*, 1997; Schurmann *et al.*, 2000). Elevated signalling through the Igf1

pathway is found in many tumours, and genetic mutations involved in activation of the Pi3k pathway collaborate with Myc during tumour progression.

One key process in which Myc has been associated is through the indirect activation of the p53 tumour suppressor via $p19^{Arf}$ (Zindy *et al.*, 1998). The $p19^{Arf}$ tumour suppressor protein acts in a checkpoint that guards against unscheduled cellular proliferation in response to oncogenic signalling. $p19^{Arf}$ also prevents hyper-proliferation and transformation caused by Myc and enhances Myc-induced apoptosis independently of p53 (Qi *et al.*, 2004). p53, sometimes referred to as the "Guardian Angel" of the cell, is a transcription factor whose targets are key factors in promoting apoptosis and cell cycle arrest (Vogelstein *et al.*, 2000; Zhao *et al.*, 2000). Key p53 target genes include the cell cycle arrest promoting $p21^{Cip1}$, and pro-apoptotic members of the Bcl2 superfamily Bax, Puma (*bbc3*), Noxa, Bid and Bim. However, it has also been shown that p53 has a more direct role in promoting MOMP, via direct activation of Bax or inhibition of Bcl_{XL} and Bcl2 in the cytosol (Mihara *et al.*, 2003; Chipuk *et al.*, 2004; Erster and Moll, 2005).

Levels of p53 are kept low in normal cells due to the p53 regulator Mdm2, which inhibits transactivation (Momand *et al.*, 1992; Chen *et al.*, 1995) and facilitates translocation of the protein to the cytosol for proteosomal degradation (Maki *et al.*, 1996; Haupt *et al.*, 1997). In this way, Mdm2 acts as a suppressor of p53-mediated apoptosis and cell cycle arrest (Chen *et al.*, 1996). $p19^{Arf}$ regulates p53 activity by binding to and inhibiting Mdm2, allowing p53 stabilisation and accumulation (Stott *et al.*, 1998; Zhang *et al.*, 1998). It can also act as a tumour suppressor independently of p53 (Weber *et al.*, 2000), e.g. through Bax-mediated release of Cytochrome c from the mitochondria (Suzuki *et al.*, 2003). $p19^{Arf}$ has also been found to bind to and inhibit the Myc protein itself, resulting in a feedback loop (Datta *et al.*, 2004; Qi *et al.*, 2004).

p53 is often associated with DNA damage sensing pathways, which may be activated due to the effects of ROS, spontaneous mutations, cytochemicals, mutagenic chemicals and radiation on DNA structure. Upon sensing of DNA damage, Rad1, Rad9 and Hus1 form the 9-1-1 complex (Burtelow *et al.*, 2001; Lindsey-Boltz *et al.*, 2001) which binds to chromatin in a Rad17-dependent

manner (Zou *et al.*, 2002). This leads to initiation of cell cycle checkpoint pathways that promote cell cycle arrest to allow DNA repair. These checkpoints are primarily regulated by the serine/threonine-specific protein kinases ataxia telangiectasia mutated (Atm) and ataxia telangiectasia and Rad3-related kinase (Atr), which phosphorylate and modify cell cycle and checkpoint proteins leading to cell cycle arrest, DNA repair, and induction of apoptosis if DNA cannot be suitably repaired. Atm is primarily activated early in the cell cycle in response to double strand DNA breaks, whilst Atr is activated later due to UV radiation-induced DNA damage and blocks in transcription (Shiloh, 2001).

Both Atm and Atr are able to phosphorylate p53 directly at serine-15 at the N-terminus, preventing association with the p53-suppressor Mdm2, accumulation of stable p53 protein, and subsequent induction of p53-mediated pro-apoptotic factors (Banin *et al.*, 1998; Canman *et al.*, 1998; Tibbetts *et al.*, 1999). Atm is able to activate p53 indirectly by phosphorylating and activating the p53-activating checkpoint kinases Chk1 (Shieh *et al.*, 2000; Gatei *et al.*, 2003) and Chk2 (Chaturvedi *et al.*, 1999; Ahn *et al.*, 2000; Chehab *et al.*, 2000; Hirao *et al.*, 2000), and the polo-like kinase 3 (Plk3) (Xie *et al.*, 2001), and also phosphorylates and activates further tumour suppressors, such as Brca1 (Cortez *et al.*, 1999), Smc1 (Kim *et al.*, 2002; Yazdi *et al.*, 2002) and E2f1 (Lin *et al.*, 2001). The main role of Atr is in activating Chk1 (Liu *et al.*, 2000; Zhao and Piwnica-Worms, 2001), which leads to cell cycle arrest by phosphorylating the cell division cycle protein Cdc25a, resulting in binding to 14-3-3 proteins and prevention of Cdc2 activation, which promotes cell cycle arrest (Zeng *et al.*, 1998; Lopez-Girona *et al.*, 1999).

The process of Myc-induced apoptosis is complex, and may occur through a variety of pathways. It is widely believed that apoptosis is the key function for limiting oncogenic Myc-mediated tumourigenesis, and the favoured mechanism for Myc-induced apoptosis may be ultimately dictated by cell type and tissue location. *In vivo* studies by Pusapati *et al.* (2006) showed that over-expression of Myc in mouse squamous epithelial cells resulted in p53-induced apoptosis in order to inhibit tumour formation. However, this was dependent on the presence of functional Atm, indicating that Myc-induced apoptosis involves DNA damage

pathways *in vivo*. Studies by Dansen *et al*. (2006) using the *pins-mycER*^{TAM} transgenic mouse model for Myc over-expression in pancreatic β-cells (described in Section 1.2.3) identified Bax as an essential factor in Myc-mediated apoptosis *in vivo*, with deletion of *bax* allowing Myc to drive the formation of angiogenic and invasive β-cell tumours. This suggests that Myc-related apoptosis occurs through Bax-mediated release of Cytochrome c from the mitochondria. This is further evidenced given the ability of Bcl2 and Bcl_{XL} to prevent Myc-induced apoptosis, allowing the tumourigenic potential of Myc to be observed (Bissonnette *et al.*, 1992; Fanidi *et al.*, 1992; Pelengaris *et al.*, 2002b; Finch *et al.*, 2006; Lawlor *et al.*, 2006).

Other methods for circumventing Myc-induced apoptotic activity *in vivo*, such as over-expression of Bcl_{XL}, or loss of $p19^{Arf}$ and p53, were found to operate through distinct mechanisms (Finch *et al.*, 2006). While Bcl_{XL} over-expression results in a reduction in Myc-induced apoptosis, loss of $p19^{Arf}$ results in an increase in both apoptosis and proliferation, resulting in balanced cell mass with a greatly increased cell turnover. This suggests that the role for $p19^{Arf}$ in Myc-induced apoptosis may be related to its proliferative activity, such as through cell cycle arrest. Loss of p53 resulted in a similar increase in proliferation together with a loss in apoptosis, leading to formation of highly aggressive tumours. This suggests that oncogenic cooperation of Myc may take many forms, and these mechanisms may themselves cooperate to further augment the lesion.

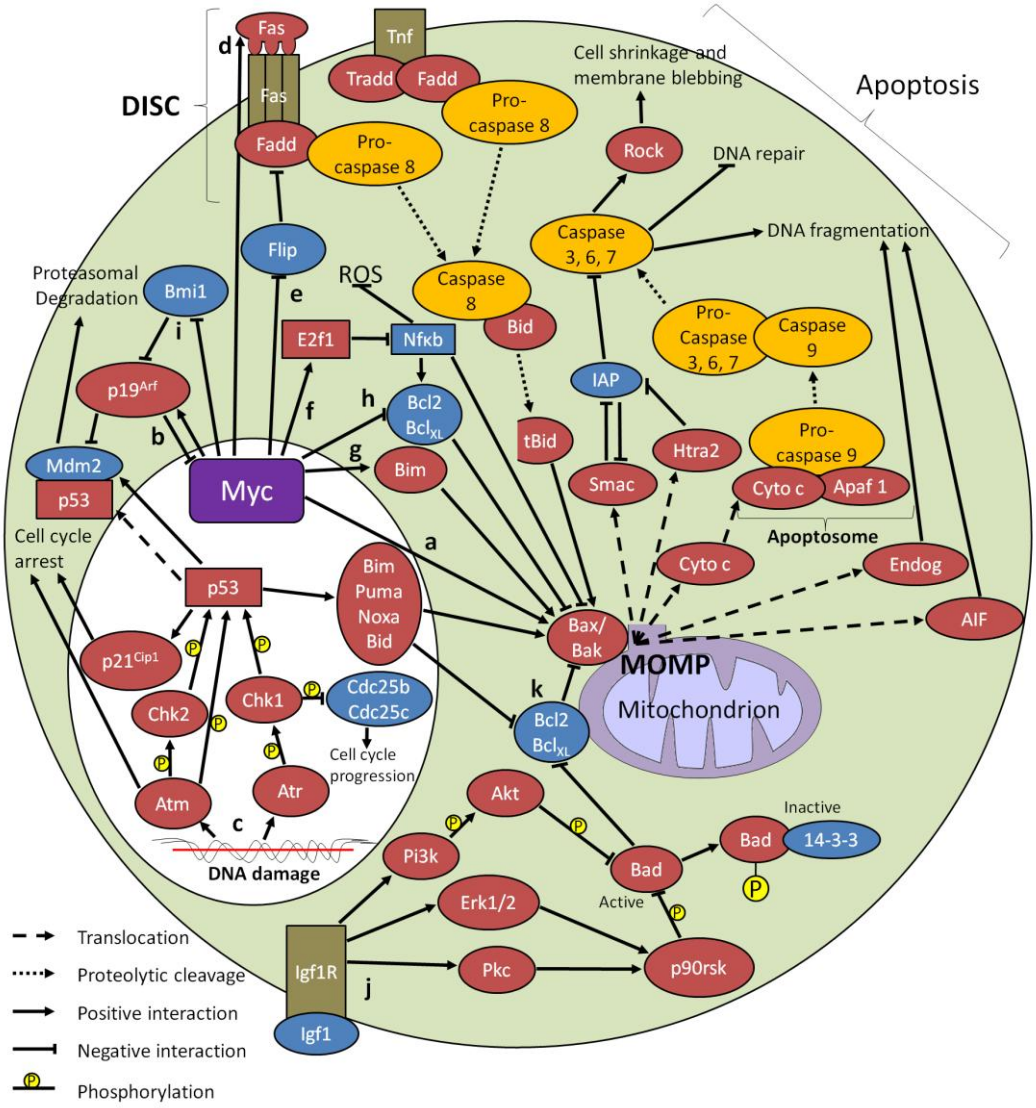

Figure 1.1.3: Pathways involving Myc and apoptosis

Myc sensitises cells to a wide range of pro-apoptotic stimuli, such as hypoxia, DNA damage, depleted survival factors, and signalling through Fas, Tnf and Trail death receptors. Shown here is a schematic representation of factors involved in Myc-mediated control of apoptosis. Pro-apoptotic factors are shown in red, anti-apoptotic factors in blue, caspases in orange, and receptors in brown. Transcription factors are indicated by rectangles. a) One of the key events in intrinsic apoptosis pathways is alteration of mitochondrial membrane pores via activation of pro-apoptotic Bcl2 members such as Bax and Bak. This results in activation of MOMP and release of Cytochrome c into the cytosol. Cytochrome c associates with Apaf1 and Pro-Caspase 9 to form the apoptosome ('wheel of death'). Caspase 9 is auto-catalytically activated, resulting in proteolytic cleavage of downstream effector caspases, including Caspase 3, leading to degradation of cell components. Further pro-apoptotic factors are also released, including Smac and Htra2 which inhibit IAPs, and

Endog and AIFs which assist in fragmentation of DNA. b) Myc also affects the release of Cytochrome c through indirect activation of the p53 tumour suppressor by stimulating expression of $p19^{Arf}$, which associates with and inhibits the p53 inhibitory protein Mdm2. This leads to stabilisation of p53, which transcribes cell cycle arrest factors such as $p21^{Cip1}$, and BH3-only Bcl2 members involved in Bax activation at the mitochondrial membrane. $P19^{Arf}$ blocks Myc-induced transcription of genes involved in growth, whilst exerting no effect on Myc-mediated gene repression, creating a feedback mechanism that results in increased apoptosis. c) p53 is also activated through the DNA damage pathways in response to activated Atr/Atm, and the checkpoint kinases Chk1 and Chk2. d) The extrinsic pathway for apoptosis involves ligation of death receptors such as Fas with their respective cytokines, resulting in association with the intracellular adaptor protein Fadd to form the DISC. Fadd recruits Pro-Caspase 8 resulting in autocatalytic activation of the pro-caspase, which begins a caspase cascade leading to activation of downstream executioner caspases and apoptosis. e) Myc inhibits Flip, which competes for binding to the DISC, preventing activation and leading to sensitisation of the cell to extrinsic apoptosis signals. Caspase 8 may also activate the pro-apoptotic protein, Bid, which promotes MOMP by association with Bax at the mitochondria, thereby linking the extrinsic and intrinsic pathways of apoptosis. f) Recently Myc has been reported to mediate apoptosis through suppression of Nfκb, leading to accumulation of ROS. g, h) Myc can also induce expression of the BH3-only pro-apoptotic protein Bim and suppress expression of the anti-apoptotic proteins Bcl2 and Bcl_{XL}. i) Expression of $p19^{Arf}$ is not usually detected during normal cell replication, so $p19^{Arf}$ may occur through inhibition of the $p19^{Arf}$ inhibitor Bmi1. j) Cell survival occurs through survival signals such as Igf1, which blocks Myc-induced apoptosis by association with the Igf1 receptor and activation of the Akt signalling pathway. Activation of Akt1 through Pi3k leads to phosphorylation of the pro-apoptotic protein Bad which is sequestrated and inactivated by cytosolic 14-3-3 proteins. k) Anti-apoptotic proteins, such as Bcl2 and Bcl_{XL}, reside in the outer mitochondrial membrane and block Cytochrome c release through sequestration of Bax. Adapted from Robson *et al* (2006).

1.2 MycER^{TAM} – Switchable transgenic model

1.2.1 The MycER^{TAM} switchable protein

In vitro experiments can only reveal so much about molecular interactions in cellular biology. Whilst we can observe the effects of a particular drug on a cell population, we fail to take into account the complex interactions between different cell species that can occur within tissues as a whole. Factors such as cell type, differentiation state and tissue location can play a significant role in determining cellular response to Myc deregulation. Lipinski and Hopkins point out that "the nature and properties of a chemical tool cannot be considered independently of the system it is to be tested in" (Lipinski and Hopkins, 2004). For this reason, *in vivo* studies using living organisms are vitally important to ensure that the complex cellular interactions that occur during everyday life are taken into account. *In vitro* studies show results given a very specific set of conditions that can be greatly affected by the experimenter (e.g. by changing the levels of growth and survival factors in the serum) – *in vivo* studies allow the greater role of local tissue environment and context to be considered.

Animal models are often used to observe the effects of scientific intervention on an organism as a whole, whilst allowing experimenters to control the environment in which samples are studied and reduce sample-to-sample variation. Experiments using mice in which target genes have been *knocked out* (altered at the genetic level such that the protein product is no longer produced) can answer important questions regarding the role of particular genes in cellular function. The development of genetically altered mice in which the expression or activation of a given gene or protein can be switched on or off *in vivo* has provided a means to assess the physiological roles of genes and proteins in tumourigenesis within the context of the adult organism as a whole (reviewed in Maddison and Clarke, 2005). These conditional transgenic mouse models allow direct control of the

'time 0' of the genetic mutation under study, allowing dissection of the pathways involved in the early stages of tumourigenesis.

One such conditional transgenic model utilises a modified estrogen receptor to provide a switch for the regulation of heterologous proteins. A human complementary DNA (cDNA) transgene is constructed encoding for a chimeric Myc protein fused at the CTD to the hormone-binding domain of a 4-hydroxytamoxifen (4OHT)-responsive murine mutant estrogen receptor (ER^{TAM}) (Littlewood *et al.*, 1995). The transgenic construct is introduced to the genome of a fertilised mouse oocyte by microinjection, such that expression is placed under the control of a tissue-specific promoter (see below). The resulting fusion protein is constitutively synthesised within target cells, but remains inactive by steric hindrance due to binding of heat shock protein 90 (Hsp90) to the ligand binding domain. Administration of the ligand 4OHT to the animal leads to preferential binding of 4OHT to the ligand binding domain, allowing association of the chimeric $MycER^{TAM}$ protein to its partner protein Max (Figure 1.2.1). The $MycER^{TAM}$-Max heterodimer is then able to carry out its transcriptional role, which has shown similar (if not identical) functions to that of the native Myc protein (Remondini *et al.*, 2005).

Expression of the $MycER^{TAM}$ protein can be targeted to cells of interest *in vivo* by placing transcription under the control of a cell-specific promoter. Two well categorised tissues for targeted $MycER^{TAM}$ expression are the suprabasal keratinocytes of the epidermis and pancreatic β-cells, which are discussed in the following sections.

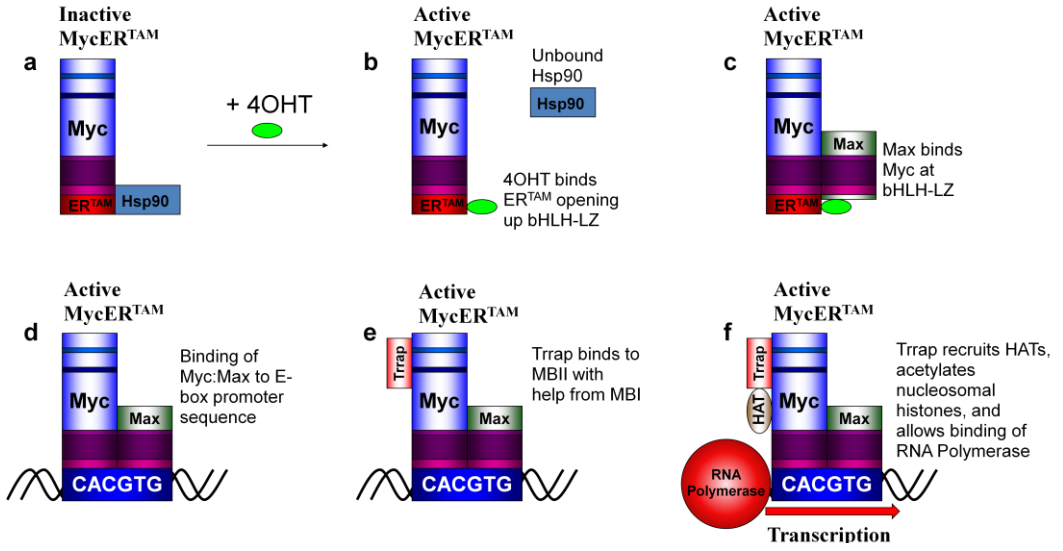

Figure 1.2.1: Activation of the MycER^{TAM} fusion protein by 4OHT

Activation of the MycER^{TAM} fusion protein by administration of 4OHT leads to transcriptional activation of target genes. a) The chimeric MycER^{TAM} protein is synthesised within target cells, but remains inactive due to binding of heat shock protein 90 (Hsp90) to the modified murine estrogen receptor ligand binding domain (ER^{TAM}). b) Administration of 4OHT molecules to target cells results in preferential binding to the ER^{TAM} domain, displacing the bound Hsp90. c) This allows association of the bHLH-LZ regions of the Myc and Max molecules and formation of the active Myc-Max heterodimeric complex. d) Myc-Max heterodimers bind to specific DNA sequences in target genes, such as the E-box sequence; CACGTG. e) Myc homology boxes I and II (MBI/II) recruit co-activators such as Trrap to the target gene promoter site. f) Trrap binds a histone acetyl transferase (HAT) which promotes chromatin remodelling and access of RNA polymerase to the activation site, leading to initiation of transcription.

1.2.2 Activation of MycER^{TAM} in skin suprabasal keratinocytes

The skin is the largest organ in the human body, and is comprised of two main layers: the *epidermis*, a stratified layer of epithelial cells (predominantly keratinocytes) that provides a protective barrier from the surrounding environment, and the thicker *dermis*, made up predominantly of collagen and other connective tissues that provides a protective cushion surrounding the delicate muscles beneath. Basal epidermal cells are anchored to a basement membrane that segregates the two layers.

The outermost epidermal layer maintains dynamic homeostasis throughout adult life through continuous proliferation of pluripotent stem cells in the basal epidermal cell compartment (Watt, 1998; Taylor *et al.*, 2000; Watt and Hogan, 2000; Oshima *et al.*, 2001) (Figure 1.2.2). The predominant model for epidermal cell replenishment is the idea that basal stem cells divide infrequently to form daughter cells with limited proliferative capacity, known as transit-amplifying (TA) cells (Mackenzie, 1970; Potten, 1974). TA cells remain within the basal layer and undergo a small number of divisions before withdrawing from the cell cycle, breaking away from the basement membrane, migrating into the suprabasal epidermis and committing to a program of terminal differentiation (Lajtha, 1979; Alonso and Fuchs, 2003). However, the role of TA cells in skin homeostasis has recently been called into question, with recent studies suggesting that proliferation of epidermal keratinocytes is attributable to a single progenitor cell (Clayton *et al.*, 2007; Jones *et al.*, 2007).

Terminal differentiation involves migration towards the skin surface in conjunction with accumulation of tough keratin filaments, which provide resilience to mechanical stress. Keratinocytes within the top layer of the suprabasal layer produce small basophilic granules, such as membrane-coated lamellar granules and keratohyalin granules, in their cytoplasm. These assist in strengthening the cell. Throughout migration, keratinocytes gradually begin to die due to a lack of nutrients and oxygen, and by the time they reach the outer level of the epidermis, they become dead, flattened, denucleated, hyperkeratinised

squames, which provide a waterproof barrier from the environment. Squamous cells are continuously sloughed off from the skin surface to be replaced by new differentiating suprabasal keratinocytes.

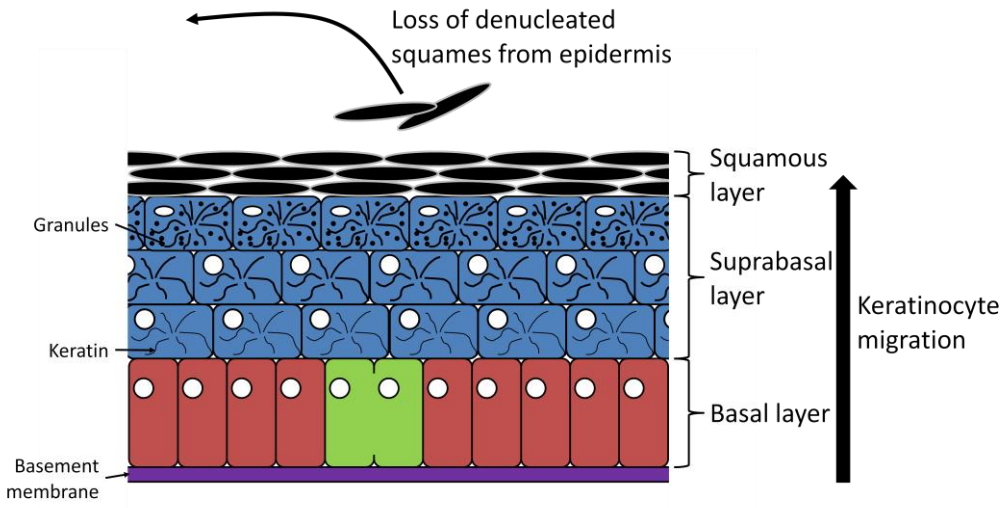

Figure 1.2.2: Homeostasis in skin epidermis is maintained by continuous proliferation of pluripotent stem cells in the basal layer

Homeostasis in the skin epidermis is maintained by dynamic equilibrium between keratinocyte proliferation and cell loss. In this way, the skin is kept in a constant state of renewal, which is particularly important for rapid wound healing response. The epidermis is a stratified epithelial tissue, made up of several distinct layers. Cells in the basal layer (red) are attached to the basement membrane, which separates the epidermis from the protective dermis beneath. Keratinocyte stem cells (or TA cells) in the basal layer replicate to form daughter cells (green), which detach from the basement membrane and migrate into the suprabasal layer, (blue) where they become committed to a program of terminal differentiation. Cells accumulate tough structural keratin proteins, which protect the cell from mechanical stress. Further accumulation of keratin and formation of membrane-coated lamellar granules and keratohyalin granules accompanies further migration, until cells become denucleated hyperkeratinised squames. These provide a protective waterproof barrier, and are continuously shed from the epidermis surface to be replenished by further differentiating keratinocytes.

In vivo models that target Myc activation to epidermal cell compartments allow more accurate studies of the effects of Myc deregulation, whilst taking the all-important context of the specific tissue into account. In particular, the MycER^{TAM} switchable transgenic model allows controlled activation (and deactivation) of deregulated Myc in suprabasal keratinocytes using the suprabasal keratinocyte specific Involucrin promoter (*inv-mycER^{TAM}*). Activation of MycER^{TAM} through daily topical administration of 4OHT results in loss of differentiation and induction of proliferation in suprabasal keratinocytes that are committed to a program of terminal differentiation. Continued activation of MycER^{TAM} leads to the formation of benign tumours resembling papillomas, or pre-malignant lesions that resemble actinic keratosis (Pelengaris *et al.*, 1999) (Figure 1.2.3).

This also leads to a loss in cell-cell contact and increased angiogenic growth (Pelengaris *et al.*, 1999; Knies-Bamforth *et al.*, 2004). Interestingly, the resulting phenotype is entirely dependent on continued administration of 4OHT, and is fully reversed upon cessation of 4OHT administration (Pelengaris *et al.*, 1999). However, despite showing many of the hallmarks of cancerous growth, over-representation of the Myc protein alone is insufficient to result in malignancy. This may be due to the fact that the keratinocyte population is self-limiting, since completion of terminal differentiation results in the ultimate loss of affected cells (Flores *et al.*, 2004).

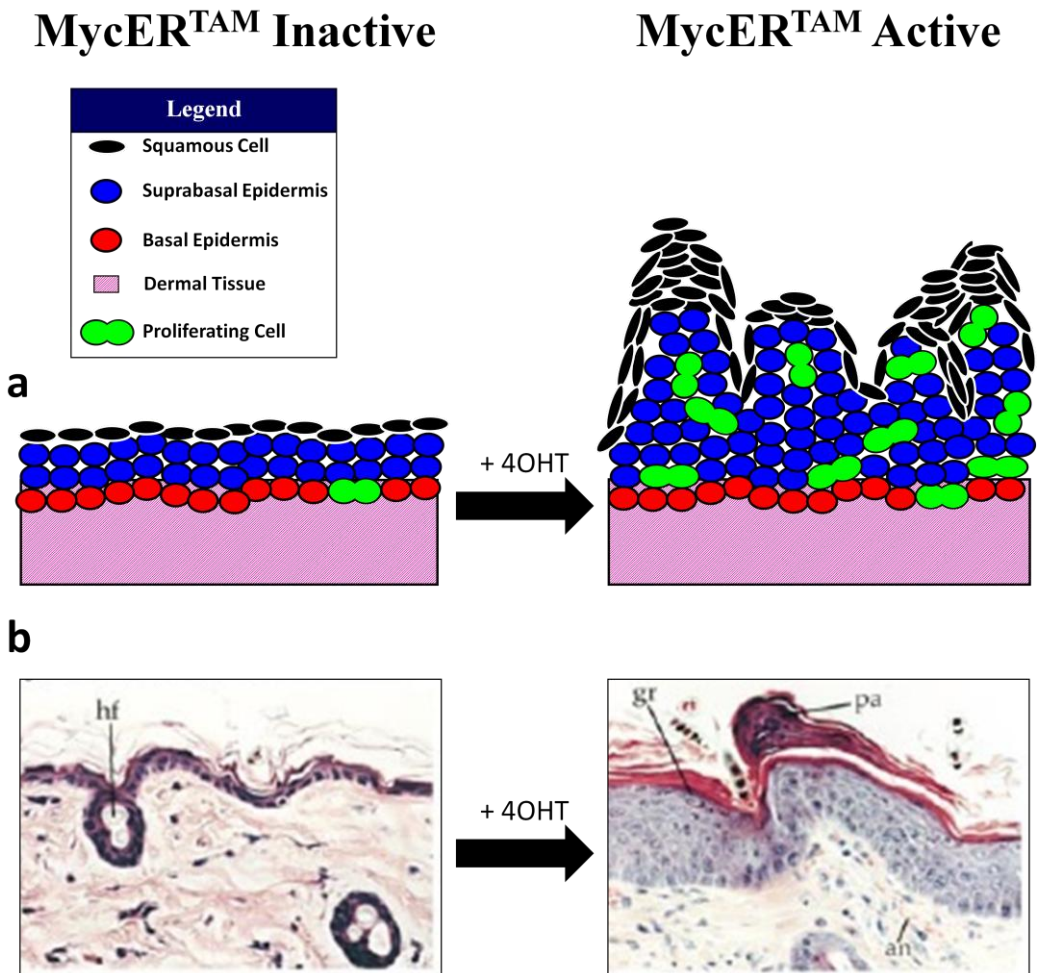

Figure 1.2.3: Activation of MycER^{TAM} in suprabasal keratinocytes results in increased proliferation and formation of benign papilloma-like tumours

The MycER^{TAM} transgenic construct can be targeted to suprabasal keratinocytes via the Involucrin promoter. In wild type (WT) and vehicle-treated (VT) transgenic animals, proliferation is confined to the basal compartment. Post-mitotic cells enter a program of terminal differentiation and migrate towards the tissue surface. Activation of the MycER^{TAM} protein in terminally differentiating suprabasal cells results in disruption of differentiation and induction of proliferation, leading to formation of benign papilloma-like growths. a) Cartoon representation of induction of proliferation in suprabasal keratinocytes. b) H&E staining for skin tissue from a transgenic untreated mouse, and a mouse treated with 4OHT for 7 days (taken with permission from Pelengaris, Littlewood *et al.*, 1999). hf, hair follicle; gr, granular cell; pa, parakeratosis.

1.2.3 Activation of MycER^{TAM} in pancreatic β-cells

The pancreas is a gland organ within the abdomen with roles in the digestive and endocrine systems of vertebrates. The β-cells of the pancreas are the sole source of the Glucose regulatory hormone Insulin which is released upon the sensing of high Glucose levels within the blood. The hormone is released into the blood via pancreatic ducts and associates with Insulin receptors on the cellular surface to stimulate uptake of Glucose for metabolism. Insulin also informs cells in the liver and the muscles to convert Glucose into glycogen for storage. β-cells are isolated within self-contained groupings of endocrine cells in the pancreas known as 'Islets of Langerhans', and make up the majority of the islet mass (Langerhans, 1869). Islets are evenly spaced throughout the pancreas and make up only a small amount (roughly 2 %) of the total pancreas mass (Elayat *et al.*, 1995), with exocrine tissue such as connective tissue, enzyme producing cells, and ducts for transportation making up the remainder. Other cells contained within the islets include: glucagon-producing α-cells, responsible for turning stored glycogen back into Glucose in times of low blood Glucose (Gaede *et al.*, 1950; Ferner, 1951); somatostatin-producing δ-cells, which inhibit Insulin and glucagon production (Alberti *et al.*, 1973; Iversen, 1974; Koerker *et al.*, 1974); and pancreatic polypeptide (PP) producing cells, which are thought to be involved in the regulation of gastrointestinal secretions (Lin and Chance, 1974; Larsson *et al.*, 1975).

In contrast to the skin epidermis which maintains a high rate of cellular turnover, β-cell turnover is much slower with an estimated life span of around 58 days for rats (Finegood *et al.*, 1995), and a far lower replication rate detected for humans (Bouwens *et al.*, 1997). This demonstrates the need for a stable supply of Insulin within the body due to the delicate control of Glucose levels required to prevent excess Glucose from accumulating in the blood (hyperglycaemia). Insulin production and release is closely linked to Glucose levels, as well as effects on β-cell replication, size and apoptosis. Failure of β-cells to regulate blood Glucose levels can lead to a condition known as diabetes mellitus. The two most common

forms of diabetes are Type I and Type II; Type I diabetes results from auto-immune destruction of β-cells, whilst Type II diabetes can occur due to resistance of cells to the action of Insulin in the blood, or due to β-cell Insulin production being insufficient (or a combination of the two).

Expression of MycER^{TAM} has been specifically targeted to pancreatic β-cells using the β-cell specific Insulin promoter (Pelengaris *et al.*, 2002b). Activation of the transgene (*pins-mycER^{TAM}*) through intraperitoneal (IP) administration of 4OHT leads to G_0/G_1 cell cycle entry of target cells, as seen in the suprabasal keratinocytes. However, in stark contrast to the skin, this response is soon overshadowed by a significant apoptotic response (Pelengaris *et al.*, 2002b) (Figure 1.2.4). This leads to reduction in islet mass, and the onset of diabetes within 3 days. Deactivation of MycER^{TAM} by halting 4OHT administration results in return to normal β-cell function and mass as remaining β-cells re-differentiate and β-cell numbers are replenished (Pelengaris *et al.*, 2002b).

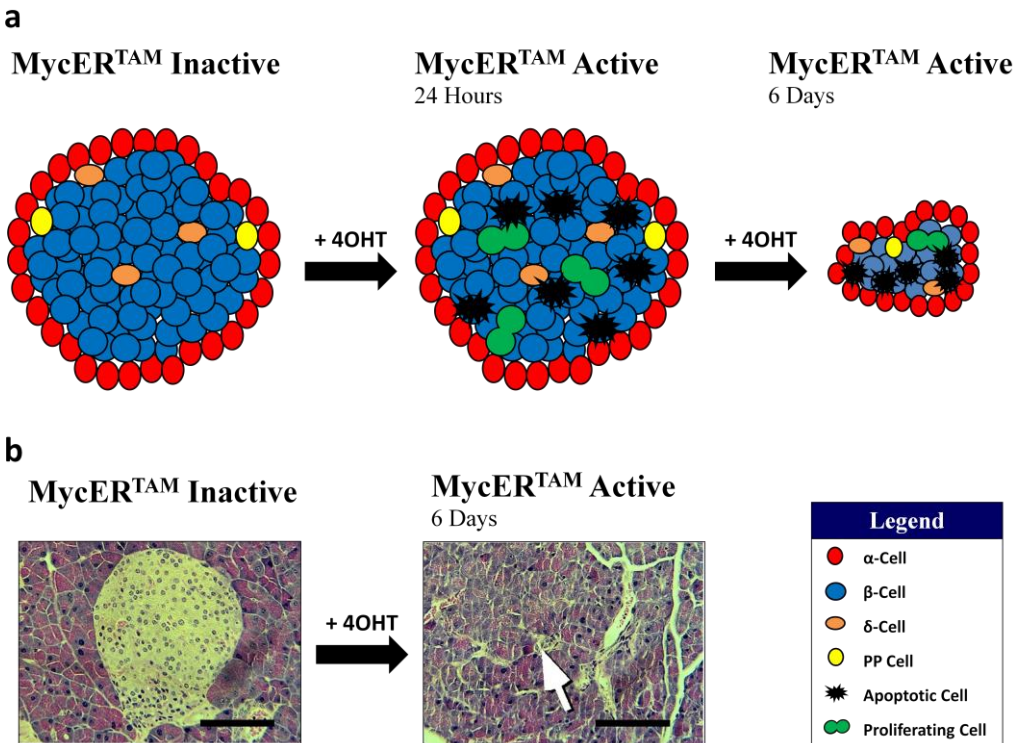

Figure 1.2.4: Activation of MycER^{TAM} in pancreatic β-cells results in apoptosis and involution of islet mass

The *mycER^{TAM}* transgenic construct can be targeted to the pancreatic β-cells via the Insulin promoter. In WT and VT animals, β-cell turnover is slow. Upon activation of MycER^{TAM}, proliferation of β-cells is induced across all islets. However, before β-cells become fully mitotic, they are driven to apoptosis resulting in involution of islet mass and induction of diabetes. a) Cartoon representation of MycER^{TAM} activation in pancreatic β-cells. Upon activation of the chimeric protein, proliferation and apoptosis are detected in β-cells. The apoptotic response soon overshadows proliferation, leading to loss of islet mass. b) H&E staining for mouse pancreas tissue from transgenic untreated sample and a mouse treated with 4OHT for 6 days, identifying severe loss in islet mass after 6 days of MycER^{TAM} activation (taken with permission from Pelengaris and Khan, 2002).

This mouse model shows the ability of Myc to act as its own tumour suppressor, which may be dependent on cell type (β-cells) and/or tissue. The β-cells appear to be only mildly buffered against cell death *in vivo* by survival signals and intrinsic anti-apoptotic mechanisms. The Myc-induced apoptotic phenotype can be suppressed by introduction of a second transgene, the anti-apoptotic protein β-cell lymphoma, extra large (Bcl_{XL}) (Zhou *et al.*, 2000; Pelengaris *et al.*, 2002b). Mice expressing the *pins-mycER^{TAM}* transgene are crossed with those expressing transgenic *bcl_{XL}* under the control of the rat Insulin promoter (RIP7), producing double transgenic *RIP7-Bcl_{XL}/pins-mycER^{TAM}* (RM) mice. Activation of MycER^{TAM} in double transgenic RM mice results in the rapid entry of nearly all β-cells into the cell-cycle with no discernable Myc-induced apoptosis (Pelengaris *et al.*, 2002b). Sustained MycER^{TAM} activation through continuous 4OHT administration results in grossly hyperplastic islets.

1.2.4 Comparison of the transcriptional response to MycER^{TAM} activation in the pancreas and skin

The two *in vivo* models for Myc activation previously described indicate a vital role for tissue context in determining the ultimate fate of the cell. Deregulation of a single oncogenic factor, the transcription factor Myc, is sufficient to elicit several of the hallmarks of cancer (Robson *et al.*, 2006) in the subrabasal keratinocytes, leading to the production of pre-cancerous papillomas and induction of angiogenic growth. However, deregulation of the same oncogenic factor in a distinct tissue, the β-cells of the islets of Langerhans, results in a vastly different outcome as β-cells ultimately follow an apoptotic pathway. This indicates a crucial role for tissue context and the surrounding micro-environment in determining cell fate. In this instance, it would appear that the suprabasal keratinocytes are largely buffered against the pro-apoptotic functions of oncogenic Myc.

The reasons for this are unclear, though it has been suggested that terminal differentiation in skin epidermis may in itself act as a tumour suppressive function, ultimately leading to the loss of cells showing irregular growth (Flores *et al.*, 2004). The divergence of Myc potentiality between the two tissues highlights the need for a real understanding of the pathways involved in these two functions. Developing further understanding of the processes by which cells are able to circumvent dangerous proliferative activity, and the conditions under which these in-built tumour suppressor functions may be avoided, will allow the identification of possible functional targets for therapeutic intervention in diseases such as cancer.

1.3 Microarrays: High-throughput transcriptomics

The completion of the human genome project in 2003 (International Human Genome Sequencing Consortium, 2004) ushered in a new age of research allowing analysis of cellular function at a previously unforeseen molecular level. Tools are now available allowing researchers to analyse the many complex changes at the cellular level in a single experiment. Of these tools, perhaps the most familiar is the microarray – a single chip, about the size of a microscope slide, which allows analysis of thousands of mRNA transcripts in a single experiment. In this way, researchers can compare the transcriptional fingerprint of cells between conditions of interest to find key genes whose transcriptional activation or repression may be involved in the divergence in physiology between the conditions. Using this technology, researchers are able to find possible genetic targets for drug treatment against diseases, or genetic markers – gene signatures that may be used to identify specific genotypes and improve the early diagnosis of diseases such as cancer. In this section, the development of the microarray and its roles in the advancement of molecular biology are discussed.

1.3.1 The microarray

Whilst it is the proteins themselves that determine ultimate cellular function, the levels of mRNA – the intermediary molecule in the protein synthesis pathway – provide unique information regarding the levels of protein synthesis within a cell at a given time. Microarrays utilise the unique association of base pairs (adenine (A) with thymine (T) or uracil (U), and cytosine (C) with guanine (G)) to measure the abundance of mRNA transcripts within a specific sample, often across many thousands of genes in parallel. This gives a quantifiable measurement of the level of transcription of particular genes within cells of interest at a given time – the *transcriptome*. The knowledge of levels of gene-expression can be used to make

inferences regarding the ultimate levels of the protein products of those genes, under the assumption that an increase in mRNA production correlates with an increase in protein synthesis.

Unlike other methods for determining gene-expression levels, such as quantitative real-time reverse-transcriptase polymerase chain reaction (qRT-PCR) (Heid *et al.*, 1996), northern blot (Alwine *et al.*, 1977) and *in situ* hybridisation (Jin and Lloyd, 1997), the microarray allows analysis of the expression levels of a large number of genes simultaneously with only a fraction of the starting material. Microarray analysis has the additional bonus of not requiring prior knowledge of the transcript sequence under study, allowing for a more exploratory discovery-based approach to transcriptomic analysis. Microarrays are also relatively cheap in comparison to other methods of high-throughput gene-expression analysis, such as serial analysis of gene-expression (SAGE) (Velculescu *et al.*, 1995) which incurs a relatively high cost in sequencing of mRNA.

It must be noted that the transcriptome does not give a full account of the state of the cellular machinery. Not all mRNA transcripts are necessarily translated into proteins, and one transcript may produce a variety of proteins, each with diverse functions (Pandey and Mann, 2000). Also, further modification of mRNA transcripts, such as degradation by microRNAs (untranslated short single stranded RNA molecules that target mRNA molecules for degradation) (Lee *et al.*, 1993; Ruvkun, 2001), and other epigenetic events (heritable modifications to DNA and chromatin structure that affect gene-expression, but which do not affect DNA sequence) may occur prior to translation. In addition, post-translational modifications, such as activation of proteins by phosphorylation or degradation of proteins by proteolytic enzymes (proteases), cannot be identified using transcription-level analyses. However, given the central role of RNA in the cellular machinery of the cell, transcriptional profiling remains a powerful approach for functional analysis of cellular function.

A microarray is essentially a small array of DNA probes, each designed to recognise a specific sequence of nucleotides (mRNA or DNA). The unique pairing of nucleotide bases that provides the mechanism for DNA and RNA

replication is utilised; each probe is made up of a large number of nucleotide sequences complementary to the target sequence. Starting mRNA or DNA molecules are processed and labelled with a fluorescent molecule. After hybridisation to the array, fluorescence scanning determines the relative abundance of molecules bound to their respective probes. The first such experiment was performed by Schena et al. (1995), looking at the expression levels of 45 Arabidopsis genes. Realising the potential of this approach, many more studies soon followed, each increasing the number of genes queried (DeRisi *et al.*, 1996; Schena *et al.*, 1996; Shalon *et al.*, 1996). In just a few short years, the microarray has developed far beyond the relatively small-scale production of these pioneering experiments to a stage whereby the expression of tens of thousands of genes can now be analysed simultaneously.

Microarrays for gene-expression analysis can be largely separated into two classes; those designed for 2-colour hybridisations and those designed for 1-colour hybridisations (Figure 1.3.1).

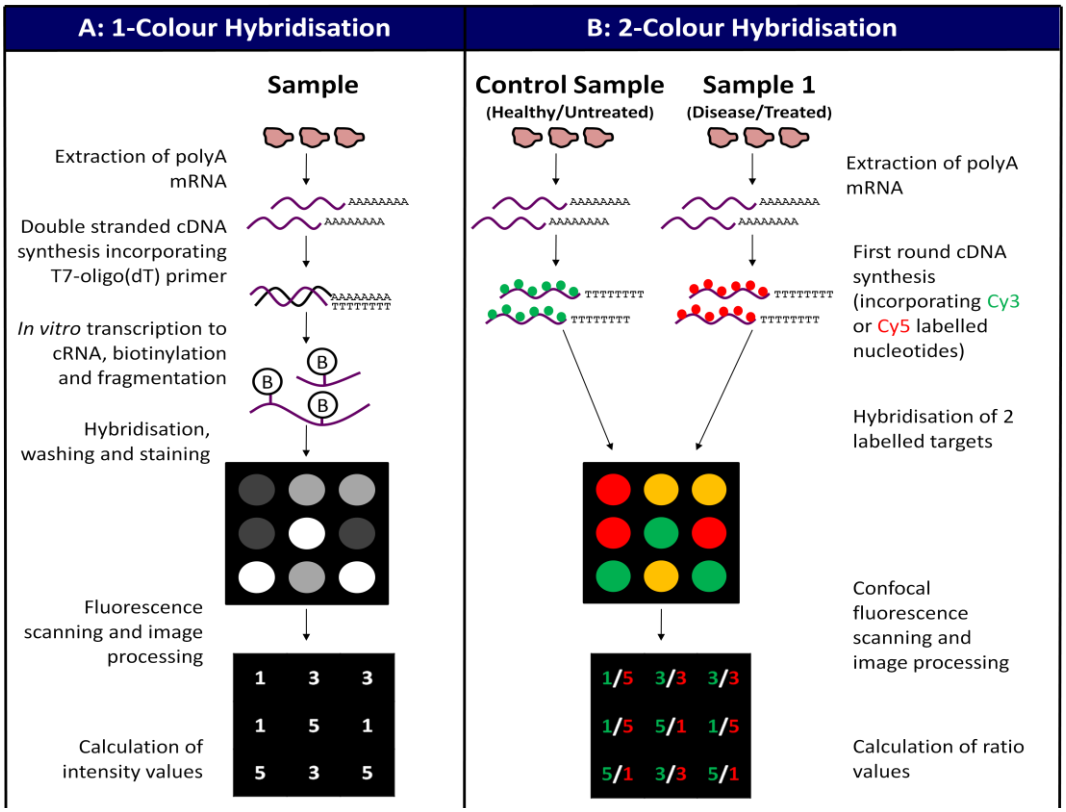

Figure 1.3.1: Schematic representation of 1-colour and 2-colour hybridisation assays

A) In 1-colour experiments, transcript abundance of each sample is represented by a single array. mRNA is extracted and used to generate double stranded cDNA with a transcriptional start site for T7 DNA polymerase. This is used to initiate reverse transcription, incorporating biotin labelled nucleotides. The resulting biotin labelled complementary RNA (cRNA) is fragmented, hybridised to a microarray, washed and stained with a streptavidin-coupled fluorescent dye. Scanning produces a fluorescence signal relating to the abundance of corresponding mRNA transcripts. Comparison of probe signal intensities across arrays is used to measure relative mRNA abundance. B) In 2-colour experiments, transcript abundance of each sample is measured relative to a single control sample of mRNA on a single array. Single stranded cDNA is synthesised from sample mRNA, incorporating nucleotides labelled with one of two different fluorescent molecules, often Cy3 (green) and Cy5 (red). Both differentially labelled samples are combined and hybridised to a single array, and confocal fluorescence scanning provides ratio values of relative abundance of mRNA transcripts between samples. Adapted from Trevino *et al.* (2007).

1.3.1.1 Spotted 2-colour microarrays

2-colour microarrays are designed to determine comparative levels of hybridisation of transcripts from two unique sources on a single array (Figure 1.3.1B). The first type of arrays to be developed, known as cDNA microarrays, were designed in this way with probes pre-fabricated from known DNA sequences and spotted onto a glass support platform (Schena *et al.*, 1995; DeRisi *et al.*, 1996; Schena *et al.*, 1996; Shalon *et al.*, 1996). These probes are usually PCR products amplified from cDNA clones, and are spotted onto specific regions of the array using robotic printing techniques.

Often, a test sample is compared directly to a control sample (e.g. treated vs. untreated). mRNA is collected from each of the two samples of interest, reverse transcribed to give cDNA, and each is labelled with a different dye (often Cy3, which fluoresces in the red part of the spectrum, and Cy5, which fluoresces in the far-red part of the spectrum). Typically these dyes are assigned false colours of green and red for visualisation (e.g. Cy3 – green, Cy5 – red). These labelled cDNA molecules, designated the *target*, are hybridised directly to the array, where they hybridise to their specific probes. Laser scanning of the array at the specific frequencies for the fluorophores produces two values for each probe; one representing the intensity of fluorescence for Cy3-labelled targets, and one for fluorescence of Cy5 targets. Combining these values into a pseudo-image shows relative signals of the two fluorophores for each probe, with probes containing a higher abundance of Cy3-labelled target appearing green, a higher abundance of Cy5-labelled target appearing red, and an equal abundance of Cy3- and Cy5-labelled target appearing yellow. Thus, relative change in target abundance between the two samples can be identified by calculating the ratio of Cy3 signal to Cy5 signal for each probe.

Since cDNA arrays can be easily produced by experimenters ‘in-house’, they are often chosen due to the relatively low costs involved, and the level of specification afforded in the array design. They are of particular use when performing experiments in which the change in expression of only a small number

of genes is of interest. However, the use of long PCR products in cDNA arrays, often 500-1000 bp, means that the levels of cross-hybridisation of targets with similar sequences may be high, making analysis of changes in similar or overlapping genes and splice variants difficult. Also, probes on cDNA arrays are sometimes of varying lengths and with highly varying GC contents within the bases. Since base pair interactions between guanine and cytosine (3 hydrogen bonds) are stronger than those between adenine and thymine (2 hydrogen bonds), this results in high variation in the individual probe-target affinities, causing variations in signal strength not related to transcript abundance.

1.3.1.2 Oligonucleotide 1-colour microarrays

Some microarrays combat the problem of varying probe-affinity by using much shorter oligonucleotide probes – sequences of nucleotides synthesised directly to the microarray support (Lockhart *et al.*, 1996). By using smaller probes for target sequences, probe-specific variation is minimised and cross-hybridisation is reduced, increasing probe specificity. This process also allows analysis of splice variants and similar or overlapping genes, since smaller, more specific subsequences can be queried. To ensure that specificity is not lost by looking at only a small region of the target transcript, a number of oligonucleotide probes are designed to represent a single transcript. The process of producing these arrays lends itself well to automated procedures, and this type of array is readily available 'off the shelf' from a variety of vendors. One of the most well known commercial vendors of microarrays is Affymetrix (Santa Clara, CA), whose range of GeneChip gene-expression arrays is one of the largest (Section 1.4).

The Affymetrix GeneChip is a 1-colour microarray system, whereby the gene-expression of a single sample is represented on a single array (Figure 1.3.1A). In many instances, this may be preferable to the 2-colour approach, as signal values represent estimations of absolute gene-expression levels rather than relative values. This means that information is known individually for all samples, and a single outlying sample cannot affect the raw expression values of any other

sample. Also, since the commercial 'production-line' manufacturing process is more reproducible as compared to the production of cDNA arrays, between-array variation is greatly reduced even across different experiments. The main downside of commercial products over home-made is the comparatively greater cost involved, since a greater number of arrays must be used and manufacturing costs are often higher. However, due to improvements in the technology, the cost-benefit gap for using commercial arrays is quickly closing, and more experimenters are moving towards off-the shelf products. Many commercially available technologies, such as the Agilent Dual Mode gene-expression arrays (Agilent Technologies, Santa Clara, CA) also allow both 1-colour and 2-colour hybridisation approaches. The relative merits of each type of microarray are discussed further in Section 1.5.1.2.

1.3.1.3 Microarray applications

The ability to analyse changes in gene transcription across the entire genome simultaneously provides researchers with a powerful tool for identifying cellular event that may be linked to physiological status. Microarray gene-expression studies can largely be identified as class comparison, class discovery or class prediction analyses (Miller *et al.*, 2002; Olson, 2006):

- **Class comparison**

  In class comparison studies, the aim is to identify genes that are differentially expressed between samples that fall into 2 or more groups. These groups could be treated samples versus untreated samples, diseased samples versus healthy samples, etc., or they could be samples at various time points of a time course experiment (for example, see Spellman *et al.*, 1998; Brachat *et al.*, 2000; Coller *et al.*, 2000; Guo *et al.*, 2000; De Leon *et al.*, 2006; Lawlor *et al.*, 2006).
- **Class discovery**

  Class discovery experiments aim to find possible groupings that exist within the data, either between samples or between genes. This can be

used to identify gene signatures, also known as *biomarkers*, able to distinguish between healthy and diseased samples. Biomarkers have been identified for a number of diseases, such as acute lymphoblast leukemia (Golub *et al.*, 1999), breast cancer (Perou *et al.*, 2000; van 't Veer *et al.*, 2002), prostate cancer (Singh *et al.*, 2002), lung squamous cell carcinoma (Wang *et al.*, 2000), colon cancer (Alon *et al.*, 1999), Alzheimer's disease (Ginsberg *et al.*, 2000), schizophrenia (Mirnics *et al.*, 2000), and HIV infection (Geiss *et al.*, 2000).

- **Class prediction**

  Class prediction aims to use biomarkers and specific gene signatures to accurately predict the membership of samples to specific groups. This is often used to test the efficacy of using biomarkers to diagnose illnesses and disease (for example, see Golub *et al.*, 1999; Ramaswamy *et al.*, 2001; Tibshirani *et al.*, 2002; van't Veer *et al.*, 2002).

Whilst gene-expression studies are by far the most common use of microarray technology, this technique can also be used in many different ways for high-throughput measurements of changes in the genome. Some examples of studies that can be performed using microarrays include:

1. **Array based comparative genome hybridisation**

   High-density DNA microarrays with probes spanning the entire human genome can be used to observe gains, losses and amplifications in copy number of genomic DNA as compared to that of a reference genome (Pollack *et al.*, 1999). For comparative genome hybridisation (CGH), test and reference genomic DNA are labelled with different fluorescent molecules in a similar fashion to 2-colour gene-expression analyses, and labelled DNA samples are hybridised to the array. Fluorescence signals for the two dyes are taken sequentially along the length of the chromosome, and ratios corresponding to variation in copy number between healthy and diseased cells are taken. Detection of copy number variants – regions of DNA $> 1\ kb$ in length that are present at variable copy numbers within a population (Feuk *et al.*, 2006) – can be used to identify genetic loci that

differ between healthy and diseased cell samples (Natrajan *et al.*, 2006; Pierga *et al.*, 2007; Suela *et al.*, 2007).

2. **Single nucleotide polymorphisms**

   As well as copy number variations, the human genome contains over 10 million single nucleotide polymorphisms (SNPs) – regions of the genome varying by a single nucleotide in at least 1 % of the population (Li and Grauer, 1991; Piotrowski *et al.*, 2006). High-density arrays with probes to measure SNPs along the whole genome are used to genotype thousands of SNPs simultaneously (Kennedy *et al.*, 2003), allowing the development of ultra-high-density SNP maps (International HapMap Consortium, 2005; Frazer *et al.*, 2007). SNP arrays can also be used, as with CGH, to measure gains and losses in copy number, and also to detect loss of heterozygocity – a measure of allelic imbalance due to the loss of or gain in copy number of one allele in comparison to the other (Yamamoto *et al.*, 2007). This is often used to detect disease loci within the genome, particularly for diseases such as cancer (Gorringe *et al.*, 2007; Heinrichs and Look, 2007; Hunter *et al.*, 2007; Lips *et al.*, 2007). One advantage of SNP arrays over CGH is the ability to detect copy-neutral losses of heterozygocity, such as detection of uniparental disomy, whereby both chromosomal copies are received from a single parent (reviewed in Walker and Morgan, 2006).

3. **Methylation studies**

   DNA methylation is one of the major sources of epigenetic modification (changes to gene-expression that do not alter the DNA sequence), and acts to silence gene-expression through addition of a methyl group to genomic DNA. Typically, around 80 % of all CpG-dinucleotides (cytosine nucleotide linked to a guanine nucleotide through a phosphate) are methylated at the fifth position of the cytosine pyrimidine ring (Braude *et al.*, 2006). Methylation of cytosine nucleotides in promoter regions can result in inactivation of genes, even in the presence of transcription factors (Bird, 1986). The methylation state of genomic DNA is therefore of great importance in the understanding of the cellular machinery. In array-based methylation analyses, genomic DNA is cleaved using methylation-

sensitive restriction enzymes to form methylated CpG fragments. These can be hybridised to whole-genome microarrays, and a comparison of the signal between two conditions can be used to identify regions of increased/decreased methylation (Schumacher *et al.*, 2006). Methylation state has been particularly linked to cancer, and imbalances in methylation state may serve as a prognostic tool for detecting neoplastic growth (Lodygin *et al.*, 2005; Gebhard *et al.*, 2006; Schumacher *et al.*, 2006; Wei *et al.*, 2006; Zhang *et al.*, 2006; Shi *et al.*, 2007a).

4. **ChIP-on-chip**

   Mutations and alterations in transcription factor expression have been identified in several diseases (reviewed in Moreno Rocha *et al.*, 1999). ChIP-on-chip combines the analysis technique of chromatin immunoprecipitation (ChIP) with the high-density genomic coverage of analysis of microarrays. ChIP is used to investigate interactions between proteins and DNA, and is typically used to identify transcription factor binding sites. The protein of interest is cross-linked to DNA, which is sheared to produce short fragments. DNA fragments bound to the protein are isolated using antibodies bound to a solid surface (e.g. magnetic beads). After filtering, cross-linking of protein-bound DNA fragments is reversed, and the fragments are amplified and denatured. Single stranded DNA is fluorescently labelled and hybridised to a genome-spanning high-density microarray and scanned. Fluorescence levels correspond to levels of transcript abundance, with greater abundance implying greater binding with the protein of interest. In this way, protein binding sites across the genome can be found in a single experiment (Ren *et al.*, 2000; Iyer *et al.*, 2001; Lieb *et al.*, 2001; Lee *et al.*, 2002; Sandmann *et al.*, 2006).

1.4 Affymetrix oligonucleotide GeneChips

The oligonucleotide GeneChips® developed by Affymetrix (Santa Clara, CA) (Lockhart *et al.*, 1996; Lipshutz *et al.*, 1999) are amongst the most widely used commercially available microarrays. Each array is split up into unique spatial regions known as *cells*, each corresponding to a single transcript. Each cell contains a large number of identical 25-mer oligonucleotide *probes* made up of 25 nucleotides. These are bound to the array using a photolithographic procedure, whereby ultraviolet light is used to allow binding of nucleotides to only selected cells on the array. This process is repeated for each oligonucleotide species in an automated cyclic procedure (Affymetrix, 2004).

This process is highly reproducible, resulting in less array-to-array variability as compared to 'home-made' 2-colour cDNA arrays. The use of shorter probes on the array as compared to cDNA arrays also means that specific regions of the target transcript can be queried allowing detection of splice variants, and also that transcripts with highly similar sequences can be distinguished. Each oligonucleotide sequence on the array is designed to be identical to a short sequence located towards the 3' end of a specific mRNA transcript (Affymetrix, 2003). The 3'-bias of these arrays compensates for degradation that may occur in mRNA molecules, which occurs primarily at the 5'-end of the molecule.

Thousands of these 25-mer oligonucleotides are localised on a single cell on the array, which are termed *probes* for the target sequence. To ensure suitable coverage of the target sequence, a number of probes (typically 11-20) are designed to span the transcript of interest (Figure 1.4.1). Together, these probes make up the *probe set* for the transcript. The terms 'probe set' and 'gene' may often be used interchangeably, however a single gene may be represented by several probe sets on an array. Each probe in the probe set is spatially separated from the others on the array to ensure that, if a problem with hybridisation is seen in a particular area of the chip, entire probe sets are not compromised. Probes are randomly distributed throughout the array. Target preparation for hybridisation is based on the Eberwine procedure (Van Gelder *et al.*, 1990), whereby cDNA is

synthesised from sample RNA from polyadenylation (polyA) tails using an oligo(dT) primer. An *in vitro* transcription reaction (IVT), utilising incorporated T7 promoter sequences and biotin-labelled ribonucleotides, generates amplified biotin-labelled cRNA targets for hybridisation. The IVT procedure has an inherent 3' bias, so probes on the array are designed to preferentially cover the 3' region of the transcript of interest.

Biotinylated cRNA molecules are fragmented and hybridised to the array, where they bind specifically to their corresponding 25-mer sequence. A streptavidin-phycoerythin (SAPE) fluorescent dye is washed over the chip where it binds to the biotin label. A scan of the levels of fluorescence for each probe of the array is recorded, and gene-expression changes are analysed under the assumption that higher fluorescence signal for a particular probe indicates a higher abundance of the corresponding mRNA molecule in the sample.

Whilst the specificity of probes on the array is high, it is entirely possible for target transcripts to bind to probes with similar sequences (though with lower affinity). This process is termed non-specific binding (NSB), and may result in noisy data. In an effort to reduce background noise due to NSB, Affymetrix use a method to directly detect levels of background binding. Each probe, made of thousands of 25-mers that match the target sequence of interest exactly (the *Perfect Match* probe – PM) has another partner probe associated with it which is identical except for a modified nucleotide, replaced by its Watson-Crick complement, in the central (13^{th}) position of the 25-mer oligonucleotide sequence (the *Mismatch* probe – MM). Each probe set contains 11-20 pairs of PM and MM probes which are spatially separated in a random fashion on the array (Figure 1.4.1). By removing the signal shown on the MM probes, which represents hybridisation of nucleotide sequences not specific for the probe, from the signal shown on the PM probes, the true signal can be seen. However, as will be discussed in the following section, the use of MM probes as measures of background hybridisation is widely criticised.

Before data from Affymetrix GeneChip arrays can be successfully analysed for changes in gene-expression, a number of low level analyses must be performed to

combine the probe level signal information to give a single value for expression of each transcript on the array. These include extraction of fluorescence intensity values for each probe on the array, summary of transcript intensity values across the probe set to obtain a single intensity value for each target, and normalisation of data to allow comparison between targets and across individual arrays (Schadt *et al.*, 2000; Irizarry *et al.*, 2003a). The wealth of information available from a single GeneChip hybridisation also provides the experimenter with a number of metrics for assessing hybridisation quality and noise. Standard methods for low-level analyses and data quality control are discussed in the following sections.

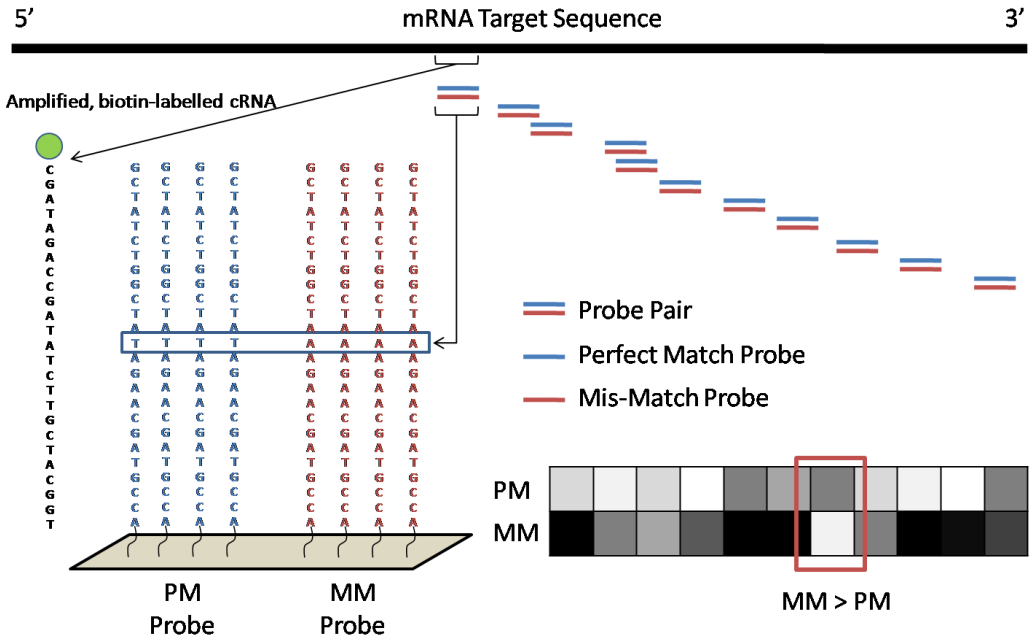

Figure 1.4.1: Probe design for Affymetrix GeneChip Microarrays

Affymetrix GeneChip microarrays are designed such that each mRNA target sequence is represented by 11 pairs of probes on the array, each only 25-nucleotides in length. The perfect match (PM; blue) probe matches the target sequence exactly and binds cRNA molecules created from the target mRNA exactly. The mis-match (MM; red) probe is identical to the PM probe except for a single nucleotide at the 13th position. This probe measures signal due to NSB. Probe pairs are designed to recognise sequences at the 3'-end of mRNA targets to avoid 5'-bias RNA degradation, and the 11 probe pairs span this region, occasionally with some cross-over. MM probes act as direct measures of noise due to NSB for PM probes which can be subtracted before analysis. However, MM probes often detect real signal as well as NSB, and can show higher signal (brighter fluorescence) than PM probes, resulting in problems with NSB subtraction. In cases such as this, where the signal of the MM probe is greater than the signal from the PM probe, a modified MM probe intensity is used (the ideal mismatch; IM) that is never greater than the respective PM probe intensity (see Equation 1-4). Adapted from Affymetrix Data Analysis Fundamentals (Affymetrix, 2002b).

1.4.1 Image pre-processing

After hybridisation of biotin-labelled cRNA target molecules, a fluorescent molecule is washed onto the array and a fluorescent scan taken. The assumption here is that fluorescence is in direct proportion to the abundance of cRNA molecules bound to the probes on the array. The first step in analysing for transcript abundance is therefore to convert the fluorescence levels of this scanned image into numerical values that can be compared to assess relative abundance between samples.

Each pixel of the raw fluorescence image is assigned a 2-byte unsigned little-endian integer value between 0 and 65,535 based on the level of fluorescence from the scanner. At this stage, it is not known which pixel values apply to each probe. Each array contains a number of control B2 oligonucleotide probes (Section 1.4.3.6) which mark the boundaries of the hybridisation area. These features allow automatic alignment of the scanned image with a grid used to segment the image into cells, each corresponding to individual probe features on the array. Each probe on the array is represented by a cell on the grid. A single intensity value is calculated for each cell by removing the outer-most pixels (to ensure there are no problems due to mis-alignment of the grid), and the 75^{th}-percentile of the remaining pixels is calculated (Figure 1.4.2) as an estimate for the probe cell intensity (Affymetrix, 2004). This information, together with information regarding the standard deviation of measurements, is reported in an Affymetrix-specific data format with file extension *.CEL*. This file contains the pre-processed probe-level data.

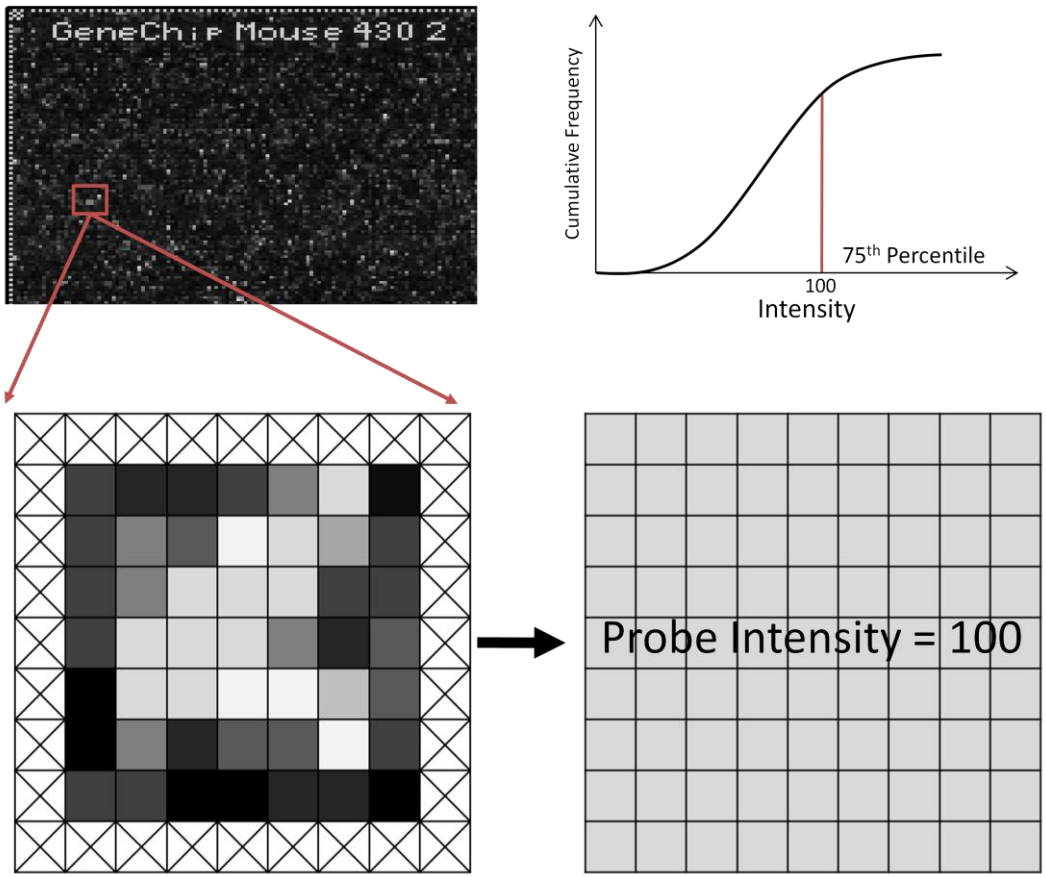

Figure 1.4.2: Image processing of scanned Affymetrix GeneChip images

The scanned fluorescence image of each array is processed to convert pixel intensity values into a single probe signal value for each probe on the array. The image is segmented by a grid that separates pixels into cells corresponding to individual probes. The outer pixels for each cell are ignored to prevent problems due to cell alignment, and a single intensity value is calculated for the cell by calculating the 75th percentile. This produces a matrix of probe-level intensity values for further processing.

1.4.2 Background subtraction, normalisation and summary

The image processing described in the previous section results in a file containing information on the signal strength for each probe on the array. Further low-level data manipulation is required to organise the data into a structure whereby each probe set on the array is represented by a single value relating to the relative abundance of the corresponding mRNA transcript such that comparisons can be made between different arrays (i.e. comparing abundance of a particular transcript between conditions) and between probe sets (e.g. looking for similarly expressed transcripts).

The first step in this procedure is to remove the NSB signal to ensure that background signal bias is removed and that measured signal values relate specifically to the transcript of interest. Probe-level signal values are then summarised across the probe set to give a single value. Finally, to ensure that comparisons can be made both across arrays and across probe sets, a normalisation procedure is used to remove systematic variation in the data by scaling signal values across samples and probe sets such that they are comparable on the same scale. This is often done such that all probe sets have mean 1 across the samples. A number of methods are available for performing these transformations, and several of the most common algorithms are discussed in this section.

1.4.2.1 MAS 4.0

Early results suggested that the subtraction of MM signal from PM signal was linear with RNA concentration (Lockhart *et al.*, 1996). The earliest editions of Microarray Suite (MAS 4.0), the Affymetrix supplied software for analysis of GeneChip microarray data, used a simple average difference method to remove signal information from the MM probes from the 'real' signal of the PM probes and summarise set (Affymetrix, 1999). For a given probe set $n = 1, \dots, N$ on array $i = 1, \dots I$, the 'Average Difference' is defined as:

$$AvDiff_{in} = \frac{1}{|A_{in}|} \sum_{j \in A_{in}} \left(PM_{ijn} - MM_{ijn} \right) \quad \textbf{1-1}$$

Where $j = 1, \ldots, J$ is the physical position of the probe pair within the probe set, and A_{in} is the subset of probes for which $d_{ijn} = (PM_{ijn} - MM_{ijn})$ is within 3 standard deviations of the average of $d_{i2n}, \ldots, d_{i(J-1)n}$. This calculation is based on the underlying model for probe level correction:

$$PM_{ijn} - MM_{ijn} = \theta_{in} + \varepsilon_{ijn} \quad \textbf{1-2}$$

θ_{in} represents the mean expression of the target transcript n on array i, and ε_{ijn} represents the probe-level error. The summary described in Equation 1-1 assumes that the error terms ε_{ijn} have equal variance for all probes in the probe set. However, it has been shown that this assumption does not hold for GeneChip data since probes with a higher mean-intensity also have a larger variance in their errors (Irizarry *et al.*, 2003b). Another problem that arises with this method for background subtraction is that often (1/3 of all probes in some cases) the signal for the MM probes is higher than that of the PM probes, indicating that the MM probes are sensitive to targets of the PM probes (Affymetrix, 2002a; Irizarry *et al.*, 2003b). This may result in the loss of real signal and not just background. More worryingly, the correction $PM - MM$ produces negative values for these probe pairs, precluding the use of a log transformation to account for the multiplicative errors, and producing negative expression values for roughly 5 % of probe sets (Wu *et al.*, 2004). The loss of signal by subtracting MM probe signal can also result in a large amount of noise, particularly at lower intensity levels, reducing accuracy and making prediction of differential expression difficult.

1.4.2.2 MAS 5.0

To avoid the problems of noise seen at lower intensity levels using the MAS 4.0 algorithm, a log transformation was used to reduce the dependence of the variance

of the error terms on the mean (Hubbell *et al.*, 2002), and a robust estimator – the Tukey biweight (Hoaglin *et al.*, 2000) – was introduced to down-weight the effects of outlying probes on the summary signal over the probe set. For some cutoff value *c* chosen in advance, the Tukey biweight function is defined as:

$$\psi(x) = \begin{cases} x\left(1 - \frac{x^2}{c^2}\right)^2 & for\ |x| < c \\ 0 & for\ |x| > c \end{cases} \quad \textbf{1-3}$$

To minimise the introduction of noise due to removal of MM signal, the concept of the ideal mismatch (IM) was introduced. If the MM probe signal is lower than the PM signal for a particular probe, the MM signal is assumed to be informative for NSB with no cross hybridisation, and the MM value is taken as the ideal mismatch value. If MM probe values are generally lower than PM values across the probe set, except for a small number of probes, then the IM for these uninformative probes is imputed from the biweight mean of the PM and MM ratio. If however the MM probe signals are generally higher than the PM probe signals across the probe set, the IM value is taken as a value slightly below that of the corresponding PM signal (Affymetrix, 2002b). Therefore, for probe pair *j* of probe set n on array i, the MAS5.0 signal is computed as:

$$MAS\ 5.0\ signal_{in} = \psi(\log_2(PM_{ijn} - IM_{ijn})) \quad \textbf{1-4}$$

This summary method is currently employed in the GeneChip Operating System (GCOS) supplied by Affymetrix. However, despite the addition of the robust Tukey biweight estimator, data calculated using the MAS 5.0 algorithm are still noisy, particularly at lower intensity levels (for instance, see Figure 1.4.3) (Irizarry *et al.*, 2003b). A strong probe effect, additive on the log scale, is detected even after removal of MM signal (Li and Wong, 2001; Irizarry *et al.*, 2003b). This indicates that subtraction of MM signal alone is insufficient to remove probe-specific effects.

1.4.2.3 Model based expression index

Due to the reproducibility of arrays produced using photolithographic and inkjet techniques, individual probe affinities can be modelled well. Li and Wong (2001) suggested a multiplicative model-based approach to estimate expression for each probe set using probe-specific affinities. This approach is termed the model based expression index (MBEI), and is implemented in the analysis package DNA-Chip Analyser (dChip) (Li and Wong, 2001):

$$PM_{ij} - MM_{ij} = \theta_i \phi_j + \varepsilon_{ij} \quad \textbf{1-5}$$

Where PM_{ij} and MM_{ij} represent the detected PM and MM signal for the probe in the j^{th} $(j = 1,..,J)$ position of the probe sets for array $i = 1,\dots,I$, ϕ_j represents the probe specific affinities for the j^{th} probe in each probe set which can be estimated from the multiple arrays in the analysis, θ_i are the estimates of the expression for each probe set on array i, and the ε_{ij} are error terms assumed to be independent and identically distributed (IID) across the arrays. Estimates for θ_i are calculated by iteratively fitting the model with variable ϕ_j, aiming to minimise the sum of the squared residuals.

This process corrects expression estimates for the effects of individual probe affinities improving precision. However, since this procedure still removes MM signal for NSB correction, the problems of noise are still present, albeit reduced. Also, it was found that this procedure results in underestimates of the predicted values for higher concentrations of RNA in spike-in studies (Irizarry *et al.*, 2003b).

1.4.2.4 Robust multi-chip averaging

By performing extensive statistical analyses on a spike in study using known concentrations of 16 probe sets on the Affymetrix HGU95A GeneChip (Affymetrix, 2002c), Irizarry *et al.* (2003b) concluded that the probe signal

strength increases linearly on the normal scale, but not on the log scale. This indicates that NSB is additive and not multiplicative as suggested by Li and Wong (Li and Wong, 2001). Given that probe effects appear to be additive on the log scale, this led several researchers to suggest the need for a method that modelled background in an additive fashion, and the error in a multiplicative fashion (additive on the log scale) (Durbin *et al.*, 2002; Huber *et al.*, 2002; Cui *et al.*, 2003). Given the problems seen with removing MM signal in NSB correction, an improved method for probe-level normalisation was suggested based on multi-variate linear models estimated using only PM signal (Irizarry *et al.*, 2003a). This measure was termed the Robust Multi-chip Averaging (RMA).

Model based estimates of the NSB probability density function negates the need for including the MM signal in the NSB estimation. Assuming the additive background model $PM_{ijn} = s_{ijn} + bg_{ijn}$, background corrected signal is defined as:

$$B(PM_{ijn}) \equiv E(s_{ijn} | PM_{ijn}) \quad \textbf{1-6}$$

Computation of the background adjusted signal is performed by using a kernel density estimate over the detected PM signals to produce a smooth probability density curve, allowing estimation of the expected signal given that the PM signal PM_{ijn} is detected. Background adjusted values are log transformed (typically base 2) and are normalised using quantile normalisation (Bolstad *et al.*, 2003) to remove systematic differences between arrays and ensure that the distribution of the log-transformed values more closely approximates a normal distribution ($\sim N(0, \delta^2)$). A linear additive model is fitted to the background adjusted, normalised and log transformed PM signal, Y_{ijn}, for array $i = 1, \dots I$, probe $j = 1, \dots, J$, and probe set $n = 1, \dots N$:

$$Y_{ijn} = \mu_{in} + \alpha_{jn} + \varepsilon_{ijn} \quad \textbf{1-7}$$

Where α_{jn} represents the individual probe affinity effect, μ_{in} represents the real log scale expression for array i, and ε_{ijn} represents the error term, assumed to be IID with normal distribution $\sim N(0, \sigma^2)$. It is also assumed that the probes on the array were designed such that the probe intensities are on average representative of the corresponding gene-expression, such that $\sum_j \alpha_j = 0$. Finally, median polishing (Holder *et al.*, 2001) is applied to the estimates μ_i of log scale expression levels for each array i to protect against the effect of outlying probes.

The main benefits of using RMA stem from the fact that background correction is not reliant on removal of MM data, which may measure actual signal as well as NSB. Figure 1.4.3 shows a comparison of the GC-RMA summary method with the Affymetrix standard MAS 5.0 method. This figure shows a clear reduction in variance using GC-RMA as compared to MAS 5.0, particularly at lower expression levels, indicating increased precision in the expression estimates. GC-RMA also results in improved sensitivity and specificity for fold change estimation, reducing the number of false positives (Irizarry *et al.*, 2003a). It is also interesting to note that this figure indicates that the signal intensity of each probe appears to be higher when using GC-RMA than when using MAS 5.0. This may be due to lower levels of background signal detected for all probes using GC-RMA than MAS 5.0.

One problem with the RMA probe-level normalisation procedure is that the use of only a global background adjustment does not adjust well for NSB. Although RMA reduces the number of false positives, robust estimation of the expression of some genes can result in an increase in the number of false negatives during analysis for differential expression, particularly for lower abundance targets (Wu *et al.*, 2004), indicating that accuracy of fold change estimates is sacrificed for precision.

The GeneChip Robust Multi-chip Averaging (GC-RMA) method of probe level normalisation proposed by Wu *et al.* (2004) improves upon the background correction portion of the RMA algorithm by using a probe-specific weighting of the MM probe signal that is dependent on the content and position of higher

affinity guanine and cytosine nucleotides within the oligonucleotide sequence (Naef and Magnasco, 2003). This prevents losing 50 % of the data by using a more sophisticated MM subtraction method that is dependent on MM probe sequence. The probe affinity is modelled as the sum of the individual effects of the bases in the probe sequence:

$$\alpha = \sum_{k=1}^{25} \sum_{j \in \{G,C,A,T\}} \mu_{j,k} 1_{b_k=j} \quad \textbf{1-8}$$

$$\mu_{j,k} = \sum_{l=0}^{3} \beta_{j,l} k^l \quad \textbf{1-9}$$

$$1_{b_k=j} = \begin{cases} 1 \; if \; b_k = j \\ 0 \; if \; b_k \neq j \end{cases} \quad \textbf{1-10}$$

Where $k = 1, \ldots, 25$ is the position along the 25-mer oligonucleotide probe, b_k is the base content at the k^{th} nucleotide position, and $\mu_{j,k}$ is the contribution that base j has on the overall affinity when in position k. This estimate is used to correct the MM values for their individual affinities to estimate NSB. These estimates were found to model NSB almost as well as the MM signal, with the advantage that computed estimates do not detect real signal. This process retains the benefits in the precision of the results as compared to MAS 5, but does not suffer from the loss in accuracy that is seen in RMA.

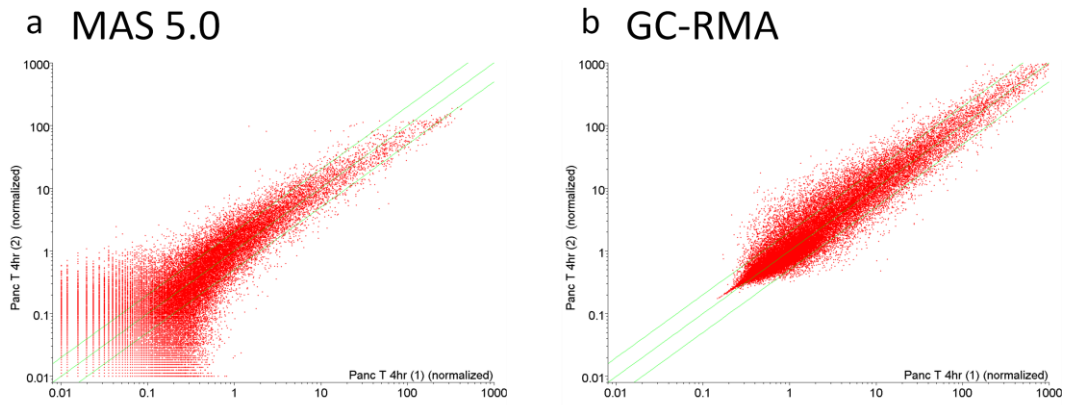

Figure 1.4.3: Comparison of MAS 5.0 and GC-RMA probe-level summary methods

Summary and normalisation of probe-level data is required to provide a single intensity signal for each probe-set. Two widely used algorithms are MAS 5.0 and GC-RMA. These procedures improve concordance between replicate samples, as can be seen by the scatterplots shown here. The signal for each probe set for two replicate samples from the main experiment (Panc T 4hr (1) and (2)) were plotted against each other on a simple Cartesian plot. Perfect similarity between the two replicates would be identified by points lying along the 45° identity line. a) While application of the MAS 5.0 algorithm to the data resulted in relatively high similarity between the two replicates, a large amount of variability was detected for probe-sets with lower signals. This may result in a large number of probe sets called as false positives. b) This region of high variability was not present after application of the GC-RMA algorithm, resulting in a tighter fit along the identity line. This 'squashing' of the highly variable region greatly reduces the number of false positive calls, but may also inadvertently reduce the fold change of real low abundance biological variation, resulting in an increase in the false negative rate.

1.4.3 Quality control procedures for Affymetrix oligonucleotide microarrays

Microarray data can suffer from large amounts of noise, and it can be extremely difficult to draw conclusions that represent true biological events. Also, systematic bias in the data due to technical procedures may confound results. This may be due to the effects of unavoidable nuisance variables, or due to limits on sample size, often as a result of cost implications limiting the number of replicate samples that can be analysed. The Affymetrix GeneChip platform was chosen for the analyses described in this body of work due to the reduced chip-to-chip variation seen as compared to homemade cDNA microarrays (Rogojina *et al.*, 2003; Jarvinen *et al.*, 2004; Yauk *et al.*, 2004). However, despite the improved reproducibility in microarray production provided by off-the-shelf manufactured arrays, it is still of great importance to ensure that the quality of starting materials is suitable and that hybridisation of cRNA prepared from sample RNA is efficient.

GCOS provides a number of quality control checks to ensure hybridisation of cRNA to the arrays has been efficient. These range from analysis of control probes on the array, to the calculation of statistical measures designed to test the efficiency of hybridisation across the individual probes in the probe sets on the arrays. These measures, and their use for determining data quality, are discussed in the following sections.

1.4.3.1 Percent present calls

As part of the scanning and summary process performed by the MAS 5.0 algorithm in GCOS (Section 1.4.2.2), a detection p-value is calculated for each probe set $n \in (1, \ldots, N)$ and for each array $i \in (1, \ldots, I)$ by comparing the discrimination scores R (Equation 1-11) of all probe sets against a user-definable detection threshold value τ ($default = 0.015$).

$$R_{in} = \frac{(PM_{in} - MM_{in})}{(PM_{in} + MM_{in})} \qquad \textbf{1-11}$$

The discrimination score provides a measure of the background corrected target-specific intensity of the probe pair relative to the total hybridisation intensity. Higher values of R imply confidence in the detection rate of the probe pair, resulting in a lower p-value assignment. Probe sets containing more probe pairs with $R > \tau$ are thus more reliable than those containing probe pairs with $R < \tau$. Probe sets are assigned flag values dependent on the p-values of their discrimination scores; *Present* (P; p-value < 0.04), *Marginal* (M; 0.04 < p-value < 0.06), A*bsent* (A; p-value > 0.06) or *Unknown* (U).

The percentage of probe sets on an array with a *Present* flag relative to the total number of probe sets (%P) can give an indication as to the quality of hybridisation. The expected %P for any hybridisation can be dependent on a variety of factors, such as tissue type, biological and environmental stimuli, array type and quality of starting material. However, replicate samples would be expected to show a similar %P.

1.4.3.2 Background hybridisation

The probe-pair design of the Affymetrix GeneChips allows analysis of the levels of background hybridisation on the array. Whilst the efficiency of the MM probes in identifying background binding signal is questionable (Section 1.4.2), the levels of signal in the MM probes can be used to identify arrays with significantly high background signal. 100 % efficiency of hybridisation, with zero cross hybridisation, would result in an observed average background signal of zero. High average background signal intensity over the MM probes for a particular array can indicate that these probes may detect real signal, and not just non-specific background signal. An average MM probe signal intensity of between 20 and 100 is considered within normal bounds (Data Analysis Fundamentals,

Affymetrix, 2002b), whilst arrays with higher levels of overall background may indicate problems with the efficiency of microarray hybridisation procedures.

1.4.3.3 Raw Q noise

Scanning an Affymetrix GeneChip produces a raw fluorescence image indicating the abundance of biotin-labelled target levels across the array. This image is processed to obtain a single fluorescence value for each probe on the array by taking the 75^{th} percentile as a summary measure of pixels corresponding to a single feature (Section 1.4.1). The raw Q noise value for each probe is a measure of the pixel-to-pixel variation for each probe. The two main sources of noise on scanned GeneChips are electrical noise (often due to problems with the scanner itself) and sample quality (particularly the amount of cRNA hybridised to the array). Due to the sensitivity of the scanning step, electrical noise is often the major contributor to noise in the data, with scans taken using different scanners often showing high variability (Bammler *et al.*, 2005; Dobbin *et al.*, 2005; Irizarry *et al.*, 2005). However, samples scanned with a single scanner should display comparable levels of noise, with variation likely being attributable to sample quality.

1.4.3.4 Scale factor

In a well designed microarray experiment it is expected that only a relatively small number of transcripts will change across the conditions. Therefore, the majority of transcripts will remain constant across the samples, with a roughly comparable average signal across all microarrays. A simple normalisation technique employed by the GCOS MAS 5.0 software is to scale the average signal values for each array to a single target intensity such that all arrays have the same mean across all probes. If all samples show roughly identical average signal, then this scale factor should be low (~1) for all arrays. Outlying samples with particularly low average signal intensity as compared to the other arrays will be identified by a large scale factor. Many core facilities consider that the absolute

value of the scale factors are not important, but that they should lie within 3-fold of each other between the samples of an experiment (Helen Brown, personal communication).

1.4.3.5 Raw probe images (.DAT files)

Fluorescence scanning of the microarray produces a raw fluorescence image (*.DAT* files), which represents the abundance of fluorescently tagged cRNA species that have hybridised to the array. This file shows microarray data in its most raw form, and analysis of the images can ensure that:

1. Hybridisation has occurred correctly across the entire array
2. Fluorescence intensity is of a suitable strength
3. There are no obvious regions of poor/increased hybridisation compared to the rest of the array
4. Artefacts (such as dust, grit, scratches or air bubbles) have not interfered with hybridisation in some way

Since the probes in a probe set are not arranged contiguously on the array, Affymetrix recommends that up to 10 % of the features on the array can show compromised (erroneously high or low due to the presence of some obscuring feature such as those described above) signal intensities before data analysis will be significantly affected (Affymetrix, 2002b; Affymetrix, 2004). Such features can be identified and somewhat corrected by fitting a probe-level model to the data with a term for the individual probe-affinities, allowing identification of regions of the array showing consistently high or low intensities. Use of the RMA probe-level summary methods therefore correct for such artefacts to some degree.

1.4.3.6 Hybridisation control probes

Affymetrix GeneChips are designed with a number of control probes used to ensure the efficiency of hybridisation. The signal of these probes can be used to identify outlying samples.

1) B2 oligonucleotides

Biotinylated B2 oligonucleotides are included in the hybridisation cocktail to bind to a series of high-affinity control features on the array that are readily identifiable in the .DAT images (Figure 1.4.4). These features are:

1. A dotted line of probes around the edge of the area that should show as alternating black and white on the image
2. A 2x2 checker-board motif of probes at each corner of the array
3. The trade name of the microarray (MOE430 PLUS2.0) spelled out in the upper left portion of the array

2) Hybridisation controls

The first set of hybridisation control probes represent prokaryotic RNA transcripts that should not be present in eukaryotic RNA samples hybridised to the array. These genes are *bioB, bioC* and *bioD*, which are involved in the biotin synthesis pathway of *Escherichia coli*, and *cre*, the recombinase gene from bacteriophage P1. Biotin-labelled cRNA transcripts specific for these control probes are added to the hybridisation cocktail at known concentrations of 1.5 pM, 5 pM, 25 pM and 100 pM for *bioB, bioC, bioD* and *cre* respectively. The signal for each spike-in control probe should therefore by consistent across all arrays in the experiment, and all control probes should show P calls across all samples with increasing signal depending on spike-in concentration.

3) House-keeping controls

Genes encoding proteins involved in basic cellular functions that are highly conserved, such as glyceraldehyde 3-phosphate dehydrogenase (*gapdh*), involved in glycolysis, and β-Actin (a*ctb*), one of the major components of the cytoskeleton, are often used in gene-expression analysis as so-called 'house-keeping' controls. These genes are constitutively expressed within all cells and thus provide a positive control for any gene-expression analysis. Affymetrix GeneChips contain probes for measuring expression of the genes for Gapdh and β-Actin, with probes spanning 3', midpoint and 5' sequences. By looking at the

3’/5’ or 3’/midpoint ratios for these internal control genes, RNA integrity can be assessed. Due to the inherent 3’ bias of probes on the array, the 3’/5’ ratio will typically be > 1. The Tumour Analysis Best Practices working group (2004) suggests that the 3’/5’ ratio for *gapdh* should be no greater than 6. By comparing the ratio across all samples in the experiment, samples with particularly high ratios (indicating poor RNA integrity) can be found.

4) Unlabeled poly-A controls

Further control probes on the array are designed to analyze the efficiency of sample preparation from the RNA level. *Bacillus subtilis* gene transcripts *dap, lys, phe, thr* and *trp,* which are modified by the addition of poly-A tails, are cloned into pBluescript vectors containing both T3 and T7 promoter sequences, and amplified with T3 RNA polymerase to give polyadenylated sense RNAs. These are spiked into the total RNA samples and carried through the sample preparation process to act in a similar fashion to internal control genes. Low abundance signals for any of these genes may indicate inefficiency in the *in vitro* transcription or 2-round amplification stages of the protocol.

Figure 1.4.4: B2 Oligonucleotide GeneChip control probes

B2 oligonucleotide control features on Affymetrix GeneChips allow confirmation of hybridisation efficacy. Scanned intensity images (.DAT files) of hybridised microarrays identify three features common to all arrays; a checkerboard pattern of high and low intensity probes at the four corners, alternating high and low intensity probes around the edge of the array, and the product name of the GeneChip array highlighted in high intensity probes. If these features are not readily identifiable, this indicates either a problem with the hybridisation of the targets to the array, or a problem with the fluorescence scanning of the array.

1.4.3.7 Probe-level models

Fitting a model to the probe-level data is the approach taken by a number of normalisation algorithms (such as dChip, RMA, GC-RMA) and allows correction for individual probe effects. A probe-level model (PLM) for the background adjusted, normalised probe-level data Y_{gij} may be of the form:

$$\log(y_{gij}) = \theta_{gi} + \phi_{gi} + \epsilon_{gij} \qquad \textbf{1-12}$$

Where θ_{gi} is the log-scale expression value for gene g on array i (the value of interest), ϕ_{gi} is the probe-specific effect of probe j for gene g, and ϵ_{gij} is the measurement error. The probe-effect ϕ has a large effect on the variability of probe intensities, making it more difficult to judge regions of high or low intensity when observing the raw probe intensity image of the .DAT file. The PLM in Equation 1-12 corrects for these probe-effects, and viewing pseudo-images of various model parameters such as the residual (the difference between the measured value and the fitted value) can often show up regions of high or low intensity that may otherwise be masked by the probe effects in the .DAT images.

1.4.3.8 Probe-level data distribution

Comparing the distribution of the probe-level intensities across all samples can be a very effective method for identifying outlying samples. Given that the majority of transcripts would be expected to be unchanging across the arrays, replicate samples would be expected to show roughly similar distributions of probe intensities, which are roughly normal on the log scale (Li and Wong, 2001; Giles and Kipling, 2003; Irizarry *et al.*, 2003b; Wu *et al.*, 2004). Therefore, samples that show dissimilar distributions compared to their replicates may represent poor quality samples.

One method of viewing the distribution of the probe intensities for each sample is to use a box plot. For each sample, five summary statistics are calculated:

minimum value, lower quartile (25^{th} percentile), median (50^{th} percentile), upper quartile (75^{th} percentile) and maximum value. The lower quartile and upper quartile are plotted as the borders of a box which is bisected by the median value, providing a visual representation of the interquartile-range of the data. The maximum and minimum values are plotted as 'whiskers' extending from either side of the box, indicating the range of the data. In this way, an overall view of the distribution of the data can be seen. Replicate samples should have comparable boxplots, particularly when comparing the medians. Hence boxplots provide a graphical means by which poor quality arrays can be easily identified.

Another way of viewing the distribution of the data is to plot the density of the intensity values on a density plot. A density plot is produced for each sample by plotting the density (number of probes with a given intensity) against all possible intensity values to form a continuous graph. As with the boxplots, this graphical representation of the distribution of the probe intensities can be used to compare replicate samples and identify poor quality samples.

1.4.3.9 Probe-level analysis of RNA degradation

One of the largest sources of error in any microarray experiment is due to deterioration of RNA molecules by ribonuclease (RNase) enzymes. Great care must be taken throughout the experimental procedure to ensure that this detrimental effect is reduced. However in most experiments it would be impossible to completely prevent RNase activity and this is particularly true when using RNA isolated from the enzyme-producing pancreas. It is therefore useful to know to what extent RNA degradation has affected the final microarray data.

Degradation by RNases occurs from the 5'- to the 3'-end of the RNA transcript. By comparing the 3'/5' ratio of each sample, we can identify samples whose RNA has become substantially degraded at the 5'-end compared to the 3'-end. A degraded RNA sample will produce data with a greater 3'/5' ratio than that of an intact sample. The probes making up the probe sets for each gene transcript on Affymetrix GeneChips are ordered sequentially from the 3'- to the 5'-end of the

sequence. As described in Section 1.4, probes are biased towards the 3'-end of the RNA molecule to limit the effect of RNase degradation on downstream data analysis. An RNA degradation plot can be made by sequentially ordering the 11 probes in each probe set from the 5'- to the 3'-end, then finding the average across all probe sets (i.e. label each probe '1'-'11' from the 5'- to the 3'-end, then take the average over probes labelled '1', then '2', etc.) to obtain 11 sequentially ordered average probe intensities. Plotting these values gives a graph showing a visual representation of the change in average intensity from the 5'- to the 3'-end over all transcripts. By comparing replicate samples against one another, those whose RNA appears to have undergone significant degradation can be found.

1.5 Microarray data analysis

1.5.1 Experimental design and statistical considerations

As with any experiment, great care must be taken throughout the procedure to ensure that the data obtained from a microarray experiment are valid and accurately represents the biology of the system under study. Experiments using RNA molecules, particularly microarrays, are prone to errors due to the delicate nature of the RNA species. However, with careful planning and implementation, these problems can be overcome. Nevertheless, a great deal of thought must be put into the design of the experiment to ensure that the resulting data are suitably set up to address the identified hypotheses. While microarrays can be used to give a general overview of the changes in transcriptional activity under varying conditions, they are at their most powerful when directed towards a specific question: e.g. "Are there any genes whose expression can be used to accurately predict cancer early in patients?", "Does treatment with drug X have a positive regulatory effect on the function of β-cells in diabetes sufferers?", etc. However, if the samples used for the analysis are of a poor quality, the resulting data will not accurately represent the biology of the system: garbage in → garbage out. In this section, considerations for experimental design and statistical issues which may affect the overall conclusions of the study are discussed.

1.5.1.1 Sources of error

Typical experiments are designed to analyse for differences between independent populations of individuals, such as a population of patients treated with some drug versus a population of untreated patients. However even with the best designed experiments, it is impossible to limit the sources of variation to only those of primary interest to the experimenter. Obscuring variation can be introduced into the data during sample preparation, array manufacture and sample processing

(Hartemink *et al.*, 2001; Bolstad *et al.*, 2003), along with sample-to-sample variation that cannot be avoided (Jin *et al.*, 2001). It is therefore important to be aware of the number of variables that may affect the experiment outcome, and the design of the experiment must address these and aim to minimise them as much as is feasibly possible. Sometimes, the effects of these unavoidable sources of error are negligible, but often not taking care to prevent such errors can have disastrous effects on the outcome of the experiment. In general, the sources of error in any experiment fall into two categories:

1. **Random errors**

   Random errors occur when some factor results in random fluctuation in the measurement of a random variable. This adds variability to the distribution of measurements for a population of samples, increasing the variance but leaving the overall mean unaffected. This is often termed *noise* and can occur due to technical limitations, such as limits in the resolution of the scanner used to capture the raw intensity image after microarray hybridisations. The effects of noise can be reduced by increasing replicate numbers, which reduce the effects of noise associated with each individual observation on the population mean leading to more stable estimates.

2. **Systematic errors**

   Systematic errors occur when some factor results in a consistent fluctuation (up or down) in the measurement of a variable for a population of samples. This has no effect on the variability of the measurement, but will result in an increase or decrease of the mean value for the sample population. This is often termed *bias* and can occur as a result of environmental or physiological differences between sample groups (for instance between males and females, or due to differences in sample preparation between groups), or due to differences in sample preparation between sample cohorts. It can be difficult to detect systematic errors, but careful experimental design, such as randomisation of samples for processing (Kerr and Churchill, 2001a; Yang and Speed, 2002; Kerr, 2003), can reduce their effects by spreading the errors amongst the whole population rather than a subset that represent a single condition.

Despite the constant refinement and improvement of the technology, microarray experiments are still seen to be highly prone to errors. Parmigiani *et al.* (2003) classify sources of error based on the five phases of microarray data acquisition:

1. Microarray manufacturing – Manufacturing errors are specific to the technology. Mass produced oligonucleotide arrays minimise manufacturing errors by using a high-throughput standardised process for all arrays.
2. Isolation of mRNA from target cells – Variability during sample preparation can be dependent on the protocols and platform being used. Sources of error include labelling procedures (particularly in 2-colour systems where individual dye effects can play a part) and RNA degradation (discussed in more detail in Section 1.5.1.5).
3. Hybridisation – Errors in the hybridisation procedure can include physical artefacts on the array (dust, air bubbles, scratches, etc.), variability in environmental conditions (such as humidity and temperature) and 'edge effects' whereby preferential hybridisation is seen at the edges of the array. Also, hybridisation of non-specific target sequences to probes on the array can result in variable background staining (NSB).
4. Scanning – Due to limits in resolution, scanning of hybridised arrays can be noisy, with rescanned images or images scanned using two different scanners giving differing intensity values. This noise may be particularly detrimental for low fluorescence signals.
5. Imaging and pre-processing – Imaging procedures for converting scanned intensity signal to gene-expression values, such as those described for Affymetrix arrays in Section 1.4.1, often require user defined parameters which can drastically affect the outcome.

These sources of error are often termed *technical variation* and refer to limitations in the current technology. Since we are dealing with biologically active species, we must also be aware of the effect that *biological variation* between samples can play on the response in gene-expression. The cells under study may be undergoing a variety of complex processes at any time, all of which interact with each other in

extremely complex ways. Gene-expression signatures can be very different between individuals, and even between different cells of a single individual (Bengtsson *et al.*, 2005). Between-sample variation can occur due to environmental factors (temperature, humidity, local environment, etc.) and physiological differences between samples (age, height, gender, etc.).

Given the sheer amount of obscuring variation that can be introduced throughout a microarray experiment, it is important to minimise errors as much as possible throughout the experiment. This will ensure the validity of the data and hence the conclusions drawn from the resulting analyses.

1.5.1.2 Type of assay

A question that must first be addressed by the experimenter is whether to use a 1-colour or 2-colour approach to microarray hybridisation. A previous analysis by the Microarray Quality Control project determined that the results of 1- and 2-colour analyses were largely identical (Patterson *et al.*, 2006). Thus, typically, the choice between the two approaches is largely dependent on the aims of the experiment, and on the questions to be addressed in the study.

Another early decision that must be made is whether to produce microarrays for the study in-house (e.g. cDNA microarrays), or whether to opt for a commercially available chip (typically consisting of oligonucleotide probes). cDNA microarrays have the advantage of being made specifically for the experiment in question. If the experiment is focused on analyzing the changes in expression of only a subset of genes in the genome, using a cDNA microarray would allow the array to be optimally designed for this purpose. Also, cDNA microarrays have much lower running costs when compared to commercially available chips, meaning that for an experiment considering a large number of samples, it would be far cheaper to run these on cDNA microarrays than on commercially purchased arrays. However, the production techniques of companies such as Agilent, Affymetrix and Illumina mean that between-array variation is greatly reduced when compared to home-made arrays, where slight variations in feature production can affect

overall data quality (Rogojina *et al.*, 2003; Jarvinen *et al.*, 2004; Yauk *et al.*, 2004), reducing the need for replicates. This may be particularly important when starting material is limited.

1.5.1.3 Sample replication

The decision as to which type of array to use in an experiment may ultimately depend on the budget available. One important aspect to be aware of is replication of observations, since running replicate samples is an essential aspect of any experiment for which random variation between samples may be expected. When observing the results for a single sample, the experimenter assumes that the results hold for the global population of all possible samples. For instance, in an experiment looking for biomarkers in disease, if a particular gene is found to be differentially expressed in the disease sample as compared to the healthy sample, the experimenter may conclude that this gene is differentially expressed in all diseased patients. However, to make such a claim from a single sample would be foolish, as the experimenter has no way of knowing for sure if the sample under scrutiny is representative of the global population.

Instead, gene-expression measures from a number of diseased patients is compared to those of a number of healthy patients to ensure that obscuring variation can be identified. Standard analysis techniques such as the t-test and analysis of variance (ANOVA; see Section 1.5.3.2) compare the means of different treatment populations to assess whether the variation between means is large compared to the pooled within-treatment variation. By using replicate samples, more realistic estimates of the within-treatment variation of the global population can be used to assess the significance of the difference in population means.

The decision as to how many replicates are necessary to ensure accuracy is non-trivial. It is a generally accepted rule that microarray experiments should be run using at least triplicate samples to ensure accuracy of the data (Lee *et al.*, 2000; Kerr and Churchill, 2001b; Kerr and Churchill, 2001a; Nadon and Shoemaker,

2002; Kerr, 2003), although others have suggested that a minimum number of 4 or 5 replicate samples should be used (Pavlidis, 2003). The number of replicates required to identify significant differences in population means at some pre-determined level (e.g. 5 %) may also be calculated by using a power analysis. Given some estimate of the expected size of the differences between population means and some estimate of the variability between replicate observations, the number of samples required to successfully reject the null hypothesis if it is indeed false can be estimated.

However, such considerations are often largely dependent on the budget for the experiment, and on the type of experiment being conducted. Between-sample variation may be expected to be high for clinical patients under different care around the country, so many replicates may be required to isolate interesting changes in gene-expression. However the signal-to-noise level is much lower for inbred transgenic animals housed under environmentally controlled conditions, and so fewer replicates may reasonably be used.

It is important to be aware of the difference between technical replication and biological replication when designing a microarray experiment. A technical replicate is used to gain information as to the precision of the technology independent of the samples hybridised. This may be done by having several different probes for a single gene on each array, which can be compared to ensure correct hybridisation has occurred at each probe, or by hybridising a single RNA sample to several arrays to confirm that all show the same result. A biological replicate is used to gain information about the differences between individual sources of biological material. This requires hybridisation of a single RNA sample to a single array. Ideally, both types of replication should be performed to allow for separation of biological and technical variablity, although budgetary limits often prevent the use of additional arrays as technical replicates. In such a case, the decision must be made of whether to run biological replicates (hybridising each RNA sample to a single array) or technical replicates (pooling the biologically replicated RNA samples together and running the pooled RNA on several arrays). Given the efficiency and reproducibility of modern microarray

production, it may be assumed that biological errors are greater than technical errors, and biological replicates are generally preferred to technical replicates (Lee *et al.*, 2000; Kerr and Churchill, 2001b; Yang and Speed, 2002; Simon and Dobbin, 2003).

1.5.1.4 In vitro vs. in vivo

The question of replication is also difficult to address since between-sample variation can vary significantly from experiment to experiment. In an *in vitro* experiment on cultured *Escherichia coli* (*E. coli*), biological variation between samples may be very small providing that standard laboratory procedures are adhered to throughout. However, in a clinical study analyzing the effects of a drug on human patients, samples may be collected from a wide range of locations, from patients with highly varying backgrounds. Even with great care in matching patients to reduce such variation, the variance between replicate samples may be much higher than in the *in vitro* experiment. The number of replicates would clearly need to be much higher to account for this increased variation. Once again, careful planning of the experiment, such as matching replicate samples (same gender same age, same weight, etc.), using inbred lines in animal experiments and keeping environmental conditions constant, can minimise these biological errors. An important source of biological error to be aware of is the effect of circadian rhythms on the transcriptome. For instance, rodents are far more active at night than during the day. The transcriptome will vary between these times. It is therefore important to ensure that all RNA species represent the same time period during the day for each individual sample.

1.5.1.5 RNA degradation

One major limiting factor for microarray experiments is the stability of the initial RNA molecules, particularly for clinical studies where RNA extraction may involve collaborative efforts between surgeons, pathologists, nurses and researchers. Degradation of RNA molecules plays a large factor in the regulation

of gene-expression (Liebhaber, 1997; Mitchell and Tollervey, 2000; Guhaniyogi and Brewer, 2001). RNases are enzymes inherent in cells that catalytically degrade RNA molecules, and many classes of RNase molecule exist (Mian, 1997). RNase II in particular is a ubiquitous exoribonuclease involved in turnover and quality control of RNA molecules as a component of the exosome RNA degrading complex (Mitchell *et al.*, 1997) or in independent complexes (Frazao *et al.*, 2006). Degradation occurs in the 5' to 3' direction, sequentially hydrolysing RNA and releasing 5' monophosphates (Mian, 1997).

It is important that great care is taken when collecting and processing RNA to minimise RNA degradation. RNase enzymes may continue to act even after RNA has been extracted unless samples are protected by fixing, freezing at -80 °C, or by using an RNA-stabilising agent such as RNAlater (Ambion, Foster City, CA). Due to the 3'-bias of GeneChip probes (Section 1.4), the Affymetrix microarray design is somewhat tolerant to degradation by RNases, and also the 5'-truncation that occurs during the IVT reaction (Luzzi *et al.*, 2003; Schoor *et al.*, 2003; Cope *et al.*, 2006). However, it has been shown that samples that show significant loss in RNA integrity, yet still pass quality thresholds, show a significant reduction in sensitivity and an increase in false positives (Thompson *et al.*, 2007). It is therefore important to ensure that RNA quality is maximised for all samples by minimising thawing of RNA samples, ensuring that procedures are carried out quickly, keeping all instruments and work surfaces clean from RNases using RNase-degrading agents such as diethylpyrocarbonate- (DEPC) treated water and RNaseZap (Ambion, Foster City, CA), and by following good laboratory practices.

mRNA makes up only 1-3 % of total RNA samples, so assessment of mRNA quantity and quality can be difficult (Palmer and Prediger, 2004). Instead, the quality of ribosomal RNA (rRNA) – which makes up over 80 % of the total RNA volume – is used to assess the level of degradation that has occurred throughout sample processing. The Agilent Bioanalyzer 2100 (Agilent, Santa Cruz, CA), which uses a combination of microfluidics, capillary electrophoresis, and fluorescent staining of nucleic acids, is often used for this purpose as it is able to

assess both RNA quality and quantity from only a small quantity of starting material (Agilent Technologies, 2007).

The highest abundance rRNA species in mammalian cells are the 18S and 28S rRNA species. Typically these are around 2 kb and 5 kb in size respectively. A 28S:18S ratio of approximately 2.7:1 is expected when RNA is intact, although a ratio of 2:1 is often used as a benchmark due to the unstable nature of 28S rRNA *in vivo*. The software accompanying the Agilent Bioanalyzer provides software for calculating an objective measure of RNA quality based on factors such as the state of the 18S and 28S rRNA peaks, termed the RNA integrity number (RIN) (Schroeder *et al.*, 2006). This is a value between 1 and 10, with a RIN of 1 representing degraded RNA and a RIN of 10 representing completely intact RNA. This measure is often used by researchers as a measure determining mRNA quality, under the assumption that the level of degradation of the rRNA is indicative of the level of degradation of mRNA species. The Bioanalyzer software also provides an electropherogram of fluorescence against time (equivalent to prevalence of individual RNA species against their size) that can be used to subjectively determine RNA quality (Figure 1.5.1). Typically, sharp 18S and 28S peaks indicate good quality RNA, while unclear 18S and 28S peaks, together with an abundance of RNA species detected towards the lower end of the time/size scale (cleaved RNA species) indicates poor quality RNA. The use of the RIN has been demonstrated to give good prediction of poor quality RNA, which can be used to assess the likelihood of developing poor quality data (Copois *et al.*, 2007).

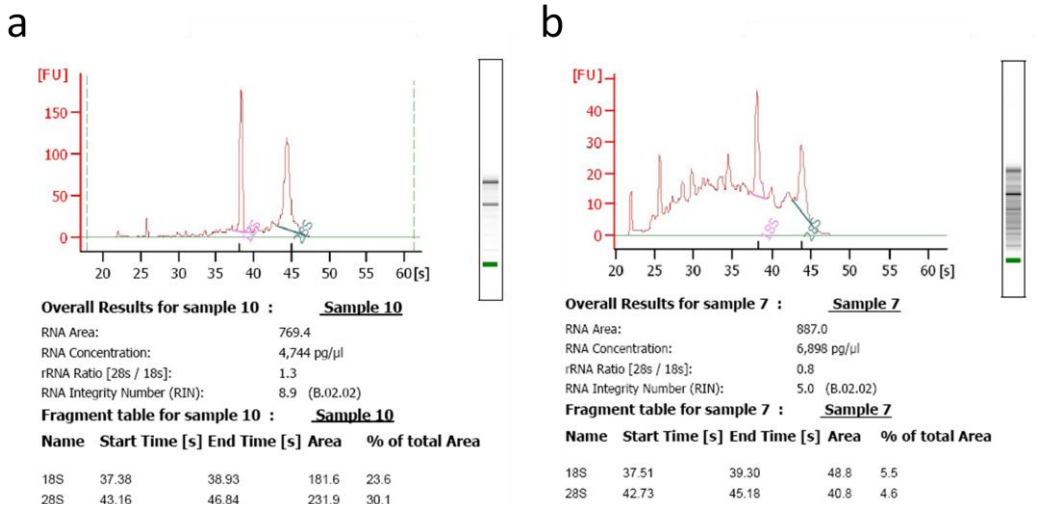

Figure 1.5.1: Quantification of RNA integrity with the Agilent Bioanalyzer spectrophotometer

Shown here are two output graphs from the Agilent Bioanalyzer spectrophotometer, representing good quality RNA (a) and poor quality RNA (b). The traces shown here represent the relative frequency of RNA species against the size of the species (measured in relation to the time taken to pass through the micro-capillary). The 18S and 28S rRNA species are typically highly abundant in total RNA samples, and are identified by peaks at around 38 secs and 44 secs respectively. Good quality RNA is identified by predominant peaks for the 18S and 28S rRNA species, with few peaks observed elsewhere. Degraded RNA samples are identified by a large number of smaller RNA species, identified by a large number of peaks at the lower end of the scale (b). Measurements made by the Bioanalyzer include the RNA area (the area under the graph) and the RNA concentration which can be used to quantify RNA levels, and the ratio of the rRNA peaks (28S/18S) and the RNA integrity number which provide a measurement of RNA integrity.

1.5.1.6 Standards in microarray experiments

The potential for using microarray technologies in drug development, disease diagnostics and discovery are clear, however there is often concern regarding the reliability and consistency of microarray results. Tan *et al.* (2003) compared biological and technical replicates across 3 commercially available microarray platforms and found little overlap in the resulting lists of differentially expressed genes between the three. Bammler *et al.* (2005) studied the variance between microarray results conducted independently across seven laboratories and found that, while reproducibility was generally high within each laboratory, reproducibility of data across different laboratories was poor. However, when protocols for sample processing, hybridisation, data acquisition and data normalisation were standardised across all laboratories, reproducibility was markedly improved indicating a lack of consistency between available platforms, protocols and analysis techniques. Perhaps more worrying are the studies of Mecham *et al.* (2004) which found that a large number (> 19 %) of the probes on Affymetrix mammalian arrays failed to match up to their intended mRNA reference sequence.

Given the poor reproducibility seen in these studies, the reliability of microarray analysis is frequently questioned (Marshall, 2004; Miklos and Maleszka, 2004; Frantz, 2005). However, similar meta-analyses indicate high reproducibility between different platforms and laboratories (Barnes *et al.*, 2005; Dobbin *et al.*, 2005; Irizarry *et al.*, 2005; Larkin *et al.*, 2005; Petersen *et al.*, 2005; Shi *et al.*, 2006), indicating the effect that non-standardised procedures can have on the consistency of microarray results. To allow collaborative work between groups, a process of key importance to allow use of the technology in clinical practice, such problems with consistency must be resolved (Shi *et al.*, 2005; Frueh, 2006; Fuscoe *et al.*, 2007).

To this end, standards have been set up by groups such as the MicroArray Quality Control group (Shi *et al.*, 2007b) describing standardised protocols and analysis techniques for microarray studies. By standardising the procedure across different

sites of operation, collaboration between groups becomes achievable. In the case of clinical studies, where tissue samples are collected from many independent sites, consistent handling and preparation of samples is essential.

To allow comparison of data sets between groups, it is also important to standardise analysis techniques. The Minimum Information About a Microarray Experiment (MIAME), put forward by the microarray gene-expression database (MGED) group (Brazma *et al.*, 2001) dictates the minimum amount of information that should be presented with any publicly available data set to allow other investigators to repeat the analysis and understand the biological context. The MGED have also developed the XML-based MicroArray Gene-expression markup language data exchange format (MAGE-ML) and object model (MAGE-OM) which provide a consistent framework for describing microarray experiments that is portable between databases (Spellman *et al.*, 2002). More recently, a spreadsheet based version of this microarray data format was developed and is now the currently recommended format (Rayner *et al.*, 2006).

A standardised list of terms for defining gene function, the gene ontology (GO), is also in wide use (Schulze-Kremer, 1997; Schulze-Kremer, 1998; Ashburner *et al.*, 2000). This provides a consistent vocabulary for biological terms throughout the gene-expression community, which allows consistent annotation across the wide range of species for which microarrays are commonly used and aids in cross-laboratory comparisons. GO classification terms are defined for biological process, cellular component and molecular function. This process is widely used for functional analysis of microarray data. In particular, data mining for GO terms within significant differentially expressed genes can identify GO classifications that are over-represented, or *enriched*, suggesting biological significance.

Hypothesis testing (Section 1.5.3.1) can be used to identify the significance of enrichment, providing a statistical measure of the number of genes from a given GO class that would be expected to appear in any given gene list purely through chance. Many statistical tests are used to calculate such statistics, and available tools for analysis of GO enrichment, such as DAVID (Dennis *et al.*, 2003), BiNGO (Maere *et al.*, 2005) and GOstat (Beissbarth and Speed, 2004), utilise

different statistical tests to determine GO significance. However, it is not entirely clear which test is most suitable for determining biological significance (Rivals *et al.*, 2007).

Consistent annotation is essential for database management of microarray datasets. These practices are in use amongst many of the largest microarray databases, such as the ArrayExpress (Brazma *et al.*, 2003) database of the European Bioinformatics Institute (EBI), the Gene-expression Omnibus (GEO) of the National Centre for Biotechnology Information (NCBI) (Edgar *et al.*, 2002), and the Microarray Mining Resource (MiMiR) of the Clinical Sciences Centre and Imperial College (CRC-IC) Microarray Centre (Navarange *et al.*, 2005). Together, these standard practices provide guidelines allowing researchers to be assured of the consistency of their data, a process that is necessary to enable larger scale analyses across multiple data sets that may have been developed at different sites.

1.5.2 Background correction, summary and normalisation of probe level microarray data

1.5.2.1 Probe level summary

The multi-probe design of Affymetrix GeneChips requires that individual probe-level signals are combined to give a single intensity value for each transcript. Typical probe level summary methods for Affymetrix GeneChip microarrays were discussed in Section 1.4.2. The choice of summary method has been found to play a significant role in determining the comparability of the results of data analysis, and was found to be the largest source of error in meta-analyses between data sets produced by different laboratories (Bammler *et al.*, 2005; Irizarry *et al.*, 2005). A lot of thought must therefore be given to the choice of algorithm to use to summarise and normalise the probe-level data. It is necessary to be aware of the negative effects of using each of the available methods.

For instance, while RMA is more precise, specific and sensitive than MAS 5.0, resulting in reduced noise at the lower expression level values and fewer false positives (Figure 1.4.3), it also suffers from a reduction in accuracy and an increase in false negatives (Irizarry *et al.*, 2003a; Wu *et al.*, 2004). This results in fewer transcripts identified as differentially expressed when using RMA as compared to MAS 5.0. The information lost by using RMA can never be recovered; false positives can be confirmed or rejected with validation studies, whilst false negative results are gone forever. GC-RMA maintains the improvements seen with RMA, but uses a weighting based on the probe affinity to estimate NSB, which increases accuracy to levels comparable to MAS 5.0.

GC-RMA is typically considered to be one of the best probe level normalisation techniques, with other model-based probe-level normalisation methods such as dChip from Li and Wong (2001) performing well also. Due to the reliance on MM signal estimates for NSB, MAS 5.0 is often considered the least suitable of the techniques, however studies by Choe *et al.* (2005) have found that MAS 5.0 outperformed GC-RMA in determining differential expression for their spike in data set (although the data set and experimental design leading to these conclusions have been criticised (Dabney and Storey, 2006; Irizarry *et al.*, 2006)). Also, Pepper *et al.* (2007) found that false positives in MAS 5.0 can be greatly minimised when used alongside the detection calls described in Section 1.4.3.1. Interestingly, it has recently been discovered that GC-RMA can result in severe artefacts in the data, leading to overestimation of pairwise correlation and inaccuracies in the calculation of network structures (Lim *et al.*, 2007). Thus, in this context, MAS 5.0 may prove to be more reliable.

The choice of algorithm can depend largely on the data being analysed. Typically, if differential expression is suspected to be low, MAS 5.0 may be preferred in order to avoid the loss of interesting changes. However, if differential expression is suspected to be mainly of a high level, or if inter-replicate variation is high, GC-RMA is generally preferred in order to reduce false positives.

1.5.2.2 Normalisation

To allow analysis of relative expression of genes across arrays, it is important to perform normalisation to remove systematic errors from the data and ensure that data distributions are comparable. Often, this is achieved by scaling such that the distribution of each gene is centred with a mean of 0 and a standard deviation of 1. In a gene-expression context, it is the log gene-expression that is scaled, as microarray data is thought to follow a roughly log-normal distribution (Giles and Kipling, 2003). This process removes bias and noise from the data that may be present due to technical variation that may obscure the interesting biological variation under study. The method used by Affymetrix is to simply scale the expression values such that all arrays have the same mean. Another often used method is to normalise to a reference array, which can be constructed by taking the median gene-expression across all arrays (Parmigiani *et al.*, 2002).

In a comparison test of 5 commonly used normalisation procedures, Bolstad *et al.* (2003), quantile normalisation was found to perform preferably in terms of speed, minimising variance and reducing bias. The process of quantile normalisation transforms the data such that the distribution of gene abundance is roughly equal across all arrays. Typically, the pooled distribution across all arrays is used as the reference distribution (F_{ref}) to which each individual array's signal distribution (F_i for arrays $i = 1, \dots I$) should be scaled. Thus for each array *i*, points are taken regularly at intervals along the cumulative distribution function (quantiles), and for each quantile x the transformation $x_{norm} = \mathrm{F_i^{-1}}\left(\mathrm{F_{ref}(x)}\right)$ is applied. The resulting set of normalised quantiles is used to build up the normalised signal distribution for array *i*. That is to say that the intensity values across the arrays are scaled in such a way as to be equal at specified intervals on the cumulative frequency plots. In a graphical sense, this can be thought of as adjusting the distribution of the probe intensities such that the I-dimensional quantile-quantile plot (a plot of the discretised cumulative distributions of 2 or more data distributions for comparison) approaches the identity as closely as possible. Model based normalisation techniques such as RMA, GC-RMA and dChip

incorporate the quantile normalisation procedure into their algorithms to allow calculation of summarised, normalised and background corrected expression data from probe-level intensity values.

1.5.3 Testing for significant differential expression

The goal of many microarray experiments is to find genes whose expression is altered, or *differentially expressed*, as a result of some experimental factor, such as disease state, drug treatment, or time. The dynamics of these changes can also be observed by measuring changes in gene-expression at different time points, making up a time course of gene-expression. In the earliest microarray analyses, so-called differential expression was calculated by looking at the average change in expression of genes between experimental conditions, often termed the *fold change* (Schena *et al.*, 1995; DeRisi *et al.*, 1996; Schena *et al.*, 1996; Eisen *et al.*, 1998). However, this approach fails to take into account the variability seen between the replicates and so may be unreliable (Chen *et al.*, 1997). The average expression value for a gene may appear to be higher for one subset of samples as compared to another, but if the samples used to estimate this average show high variability amongst themselves, this difference in expression may be unreliable.

1.5.3.1 Hypothesis testing

The statistical significance of a test statistic is usually expressed in terms of the *p-value* – the probability under the null hypothesis of observing a test statistic value equal to or greater than that observed. A statistically significant difference is said to exist between two groups if the p-value is below some significance level α, often taken as 0.05. In this case, we define the amount of evidence required to reject the null hypothesis in advance. The significance level α can also be regarded as the probability of rejecting the null hypothesis when it is actually true, or observing an effect when in fact there is none. This is known as a Type I error,

or *false positive*. Thus, if $\alpha = 0.05$, we expect to see false positive results 5 % of the times. On the other hand, if the null hypothesis is not rejected on the evidence of the samples, but is in fact false, this is known as a Type II error, or *false negative*. Again, this can be described as not observing an effect when there is one. The probability of this type of error is denoted β. Typically, the reliability of such tests is measured on three criteria:

1. **Specificity** – The ability of a statistical test to correctly determine true negative outcomes $(1 - \alpha)$. Can be increased by reducing the significance level α, although this may result in an increase in false negatives.
2. **Sensitivity/Power** – A measure of a test's ability to accurately reject the null hypothesis when it is false $(1 - \beta)$. Can be increased by increasing the sample size.

1.5.3.2 ANOVA and the t-test

Statistical hypothesis testing can be used to quantify the significance of observed differential expression to determine whether observed changes in expression across groups is likely related to a biological effect, or purely due to chance. Hypothesis tests can be *parametric* or *non-parametric*. For parametric tests, the distribution of the data is assumed *a priori*, whilst for non-parametric tests no such assumptions are made. An assumption often made of log-transformed microarray data, and for many parametric tests, is that errors follow a normal distribution. Whilst this has yet to be conclusively tested, evidence suggests that in some instances it may be a valid assumption (Giles and Kipling, 2003), particularly for *in vitro* and transgenic studies whereby within-group variation would be expected to be small (Olson, 2006).

Replicate microarray data are collected for each condition of interest and the variance in the signal between replicates for a particular gene is compared with the variance between the conditions to give some idea of the reliability of differentially expressed genes (Lee *et al.*, 2000; Kerr and Churchill, 2001a; Kerr and Churchill, 2001b; Nadon and Shoemaker, 2002; Kerr, 2003). Many studies

focus on the conditions of a single experimental variable, such as drug treatment, disease state, treatment time, etc. For such experimental designs, significance analysis is routinely performed by calculating a t-test statistic (two classes) or one-way ANOVA F-statistic (multiple classes) (Wolfinger *et al.*, 2001; Cui and Churchill, 2003; Pavlidis, 2003; Churchill, 2004). The role of these test statistics is to identify significant differences between the means of the different sample groups – that is, differences in the means that would not be expected by chance alone. These tests therefore consider the variance between group means relative to the pooled variance of observations within each group.

ANOVA can be extended to analyse for significant changes in gene-expression in response to more than one experimental variable. Experiments comparing the effects of two or more variables on gene-expression will often be designed with a factorial treatment structure, such that all combinations of experimental conditions are represented (e.g. male & treated; female & treated; male & untreated; female & untreated) (Fisher, 1926). Such analyses consider not just the main effects of the experiment variables, but also their interactions; modifications of the combined main effects caused by interdependencies between the variables. For instance we may see that a drug under study imposes a stronger effect on males than on females.

Whilst t-test statistics can be calculated for cases when the variance in the two groups is not equal, the ANOVA F-test statistic assumes equal variance across the groups and relies on the parametric assumption of a normal distribution of error terms. As previously discussed, such an assumption may not hold for gene-expression data, so resulting p-values should be treated cautiously. Also, this procedure performs tens-of-thousands of hypothesis tests simultaneously across the genes on the arrays. Each test is assumed to be independent, however given the complex interactions between genes within the cell due to co-expression (simultaneous expression due to related function), this assumption is likely false.

However, the ANOVA F-test is often used as a starting point for analyses in order to gain an understanding of data structure and dependencies before progressing to more sophisticated techniques that consider the relatedness of the individual

genes, such as empirical Bayes approaches (Baldi and Long, 2001; Efron and Tibshirani, 2002), or to reduce the dimensionality prior to higher order analyses such as hierarchical clustering (Section 1.5.4).

1.5.3.3 Analysis of gene-expression using linear models

Linear model analysis is used to fit a relevant model that can be used to identify statistically significant effects. The variable of interest, or *response* variable, is related to the *predictor* variables through a linear model. For some transcript $g \in (1, \dots, G)$ with gene-expression $Y_g = (y_{g1}, \dots, y_{gn})$ over n samples, a linear model can be applied to y_{gi} with experiment variables $x_1 = (x_{11}, \dots, x_{1n})$, $x_2 = (x_{21}, \dots, x_{2n})$, $etc.$ as predictor variables:

$$y_{gi} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \epsilon_{gi} \qquad \textbf{1-13}$$

Where β_0 is the intercept term for transcript g across all samples $i \in (1, \dots, n)$, $\beta = (\beta_1, \dots, \beta_p)$ are the regression coefficients (see below) for the predictor variables $x_1, \dots, x_p$, $X = (x_1^T, \dots, x_p^T)$ is the design matrix of observed values for the predictor variables $x_1, \dots, x_p$ for each observation i, and ϵ_{gi} is some error term assumed to be IID $\sim N(0, \sigma^2)$.

It is assumed that the response variable $Y_g = (y_{g1}, \dots, y_{gn})$ for transcript g is made up of n independently observed values, and that each value of the response variable is observed for some designed value of the predictor variables. These are typically considered as valid assumptions for microarray analyses. The relationship of the response variable y_{gi} to the predictor variables can also be written using the R-specific notation $y_{gi} \sim x_1 + x_2 + \cdots + x_p$, where the $\sim$ symbol implies that "y_{gi} is modelled by the additive main effects of x_1, x_2, etc". If interaction terms are of interest (for instance, we may suspect that the effect of drug treatment depends on age), these terms can be included in the model also:

$$y_{gi} \sim x_1 + x_2 + \cdots + x_p + x_1 : x_2 + \cdots + x_{(p-1)} : x_p \quad \textbf{1-14}$$

Where $x_n : x_m$ indicates the 1st order interaction between terms x_n and x_m for $n, m \in (1, \ldots, p)$. Here we consider only 1st order interaction terms, although higher order interactions may also be of interest. For 2-colour microarray experiments, the gene-expression data y_{gi} are typically normalised log-ratios, and for 1-colour experiments are typically normalised log-signals.

The explanatory variables x_k for $k \in (1, \ldots, p)$ can take many forms, including both categorical variables which take one of a finite number of levels, and numerical variables which take any value within a continuous range. These variables can represent a wide range of experimental conditions. The ANOVA model is a special case of the linear model in which all model terms are taken from a restricted set of designed factor levels. Analysis of covariance (ANCOVA) is an extension of ANOVA including both factors and continuous explanatory variables showing a linear relationship to the response variable (covariates). Such covariates may influence the response of the factor terms on the response variable, and ANCOVA models allow the removal of such nuisance covariate effects.

The error terms, or residuals, ϵ_{gi} are assumed to be IID such that $\epsilon_{gi} \in N(0, \sigma^2)$. Whilst the assumption of normality in microarray data is contentious, it is generally accepted to hold after transformation of the data to the log scale, and it has been suggested that such a transformation may in fact be unnecessary to ensure normality (Giles and Kipling, 2003).

For numerical variables $x_1, \ldots, x_p$, the coefficients $\beta = (\beta_1, \ldots, \beta_p)$ can be calculated using regression analysis, or curve fitting. One such method is ordinary least squares fitting, whereby the *residual sum of squares* (RSS), $\sum_{i=1}^{n} \epsilon_i^2$, is minimised. The linear model described in Equation 1-13 can be considered the equation of a curve within a $p + 1$ dimensional space. Thus regression analysis aims to fit a p-dimensional surface such that the residuals are minimised. For instance, consider the simplest case of $p = 1$ – e.g. analysing the effect of the

numerical variable age on the expression of a single gene (the subscript g is removed for convenience). This would be modelled by the equation:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad \textbf{1-15}$$

All values y_i and x_{1i} ($i \in (1, \ldots, n)$) can be plotted on Cartesian co-ordinates, with the response variable y on the vertical-axis and the explanatory variable x on the horizontal-axis. By assuming a linear relationship between *x* and *y*, one such model is to fit a straight line with y-intercept $\hat{\beta}_0$ and slope $\hat{\beta}_1$, such that the RSS is minimised. Note however that "linear model" does not imply a linear relationship. A linear model is defined as a model where the explanatory variables are related to the response variable through a linear combination of terms (for instance $y \sim 3.1 + 2.7x + 1.4x^2$ is a linear model despite the second order term x^2, with explanatory variables *x* and x^2). The solution to this simple linear relation regression is given by:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\,(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^{\,2}}$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad \textbf{1-16}$$

Where $\bar{x}$ and $\bar{y}$ are the means of the x and y variables respectively. Fitting this model to a random sample of the global population results in estimates of the model parameters $\hat{\beta}$. The residual error ε_i for each observation *i* is defined as the difference between the fitted value (μ) and the observed value (y_i), and the residuals are assumed to be IID and $\sim N(0, \sigma^2)$. However, this imposes constraints on the number of measured values that are free to vary. That is, if 100 measurements are randomly sampled from the population with residuals $\epsilon_1, \ldots, \epsilon_{100}$, the final measurement will necessarily be defined as $\epsilon_{100} = -\sum_{i=1}^{99} \epsilon_i$. In this case, we say that estimation of this statistic has 99 *degrees of freedom* (DF).

For factorial variables (such as with ANOVA), regression coefficients can be calculated for each level of the factor by assigning dummy variables to each

factor. If the resulting coefficient estimate $\hat{\beta}_1$ is high, this suggests an effect on the response variable. For a higher number of explanatory variables, the procedure is the same over a higher dimensional space, with the size of the regression coefficients relating to the size of the effect.

An F-test statistic can be used to test for significant differences between nested models to judge for improvements in model fit, by giving a measure of the significance of the difference in the resulting change in RSS. For nested models $Y_1 \sim t_1 + \cdots + t_{p1}$ and $Y_2 \sim s_1 + \cdots + s_{p2}$, where $p2 < p1$ and the terms $t_i \in \left(x_1, \ldots, x_p, x_1{:}\,x_2, \ldots, x_{p-1}{:}\,x_p\right)$ for $i \in (1, \ldots, p1)$, and terms $s_j \in \left(t_1, \ldots, t_{p1}\right)$ for $j \in (1, \ldots, p2)$, the F-test statitic is given by:

$$F = \frac{(RSS_2 - RSS_1)/(p1 - p2)}{RSS_1/(n - p1)} \qquad \textbf{1-17}$$

The F-statistic is used to test the null hypothesis that the *p1-p2* term(s) that differ between Y_1 and Y_2 have no effect on the response variable. This approach can be used in an iterative manner to test the significance of each coefficient term in the model by comparing the model fit with and without each term. Thus for two models *U* and *R* with *n* observations, where model *U* has *k* unrestricted coefficients and model *R* restricts *m* of the coefficients to zero, the F-test statistic is defined as:

$$F = \frac{(n - k)(RSS_R - RSS_U)}{m.RSS_U} \qquad \textbf{1-18}$$

The fraction of the total SS explained by each of the terms in the model is calculated sequentially to account for the inclusion of previous terms in the model. Traditionally, one of three methods can be used to determine the explained SS for each model term (Yates, 1934; Speed *et al.*, 1978; Herr, 1986; Langsrud, 2003). In a Type I sum of squares (SS) method, the significance of each term in the model is calculated by sequentially adding a term, recalculating the SS, and comparing the models before and after (Overall and Spiegel, 1969). This is

repeated until the model becomes saturated. If the design of the experiment is not orthogonal or is unbalanced, the resulting SS can be greatly influenced by the order in which terms appear in the model, and different permutations can give vastly different results (Langsrud, 2003). In contrast, Type II and Type III SS are not reliant on the order of the terms, and are therefore better suited for unbalanced designs. In a Type II SS method, each term under consideration is adjusted for the terms in the model that do not contain the term of interest. In a Type III SS, each term is adjusted for all other terms in the model. Many statistical packages offer the Type III SS as a default, although this has been criticised due to the fact that this this can lead to inclusion of interaction terms without the inclusion of corresponding main effect terms, and Type II SS has been found to have higher power when analysing unbalanced designs (Langsrud, 2003).

Often, a common approach in significance analysis of gene-expression changes across multiple variables is to fit a saturated model for each gene incorporating all variables and their interactions. However, it is preferable and more relevant to fit a single model for each gene to account for per-gene variability (Jin *et al.*, 2001; Wolfinger *et al.*, 2001; Smyth, 2004). Several approaches to significance analysis of gene-expression data using linear models have been previously described (Kerr *et al.*, 2000; Jin *et al.*, 2001; Wolfinger *et al.*, 2001; Chu *et al.*, 2002; Smyth, 2004), and one of the most widely used is the *limma* (linear models for microarray data) package in R (Smyth, 2004; Smyth, 2005).

1.5.3.4 Multiple testing

The use of hypothesis testing for gene-expression analysis is popular due to simplicity, and the large number of pre-existing methods available. However, problems may arise due to the large number of tests performed at any one time. The cutoff for significance in many experimental procedures is $\alpha = 0.05$, which indicates that we can expect to see a false positive for 1 out of every 20 tests. Thus the simultaneous testing of tens-of-thousands of genes may result in hundreds, or even thousands, of false positive results. The multiplicative nature of probabilities

indicates that the p-values cannot be calculated in isolation for multiple tests. This is termed the *multiple testing problem.* To account for the multiplicative nature when testing multiple hypotheses, multiple testing corrections (MTC) must be applied to correct the p-values for the number of concurrent tests performed (Dudoit *et al.*, 2003; Reiner *et al.*, 2003).

One class of MTC are the family-wise error rate (FWER) corrections, which gives the probability of making one or more false discoveries across a family of tests. p-values for each gene are adjusted based on the number of individual tests performed, which accounts for the multiplicity of hypothesis testing. One of the best known FWER MTCs is the Bonferroni procedure (Bonferroni, 1936), which is often found to be very conservative in its p-value estimates. A second class of MTCs are the false discovery rate (FDR) procedures which control the proportion of incorrectly rejected null hypotheses, and are generally considered to be less conservative than FWER procedures. The modified p-value gives the expected proportion of false positives that can be expected in a set of tests at a given confidence level.

FDR and FWER MTCs reduce the number of false positive results, but also reduce the power of the statistical test for individual genes. The FDR also gives fewer false negatives than the FWER, but increases power at the cost of the specificity and is often seen to be less stringent (Reiner *et al.*, 2003). One of the most widely used FDR corrections is that of Benjamini and Hochberg (Benjamini and Hochberg, 1995). An improvement on this correction, the q value of Storey *et al.* (Storey, 2002; Storey and Tibshirani, 2003) improves the power of the test and eliminates the need to set the error rate before-hand. Since FDR corrections such as these require test statistics for each gene to be independent, or at most weakly dependent, there remains criticism as to the relevance of their application to the field of microarrays where dependence exists between many genes (Jung and Jang, 2006; Gordon *et al.*, 2007), although further MTCs that account for positive regression dependency between hypothesis test statistics are available (Benjamini and Yekutieli, 2001). Regardless, given the huge number of tests performed simultaneously when observing differential expression using hypothesis testing,

the use of some form of correction is required to account for and minimise false positive results.

1.5.4 Clustering for co-regulation of gene-expression

Gene-expression in cells is not independent, and many genes may be involved in similar cellular functions. Genes that are involved in the complex pathways altered by treatment in the microarray experiment may therefore be expected to show similar traits in their change in expression. For instance, two or more genes may be activated by the same transcription factor resulting in concomitant, or *co-expression*. Alternatively, activation of one gene may result in simultaneous repression of another, resulting in a negative correlation between the two profiles. Thus observing gene-expression profiles can allow identification of genes relating to similar functions, which can provide information for determining the complex pathways involved.

One method for seeking co-expression in microarray data is to use clustering methods to find genes showing similar profiles in their gene-expression response. Clustering uses some algorithm to divide objects up in such a way that the similarity between the objects within the groups is greater than that between the groups themselves. There are two main types of clustering algorithm:

1. **Hierarchical clustering**

   Hierarchical clustering produces a nested tree, or dendogram, with more similar objects connected by shorter branches. Clustering can be either agglomerative (bottom up) or divisive (top down). In an agglomerative clustering procedure, each object is first assigned to an individual cluster. A distance or similarity metric is used to combine the two most similar clusters into a single cluster, and this process is repeated until a single cluster is produced containing all objects. In a divisive clustering procedure, the algorithm begins with all objects assigned to a single

cluster. At the first stage, the cluster is divided into two clusters based on a similarity or distance metric. This is repeated until each object has been assigned to an individual cluster. The choice of agglomerative or divisive, and the similarity metric used, can greatly affect the resulting dendogram.

2. **Non-hierarchical (partitional) clustering**

   Non-hierarchical clustering techniques partition objects into a number of exclusive groups. Unlike hierarchical clustering, this produces independent groups that are not nested. Typically used methods include self organising maps (SOMs) (Kohonen, 1982; Kohonen, 1995; Tamayo *et al.*, 1999), k-means clustering (Hartigan and Wong, 1979), and quality threshold (QT) clustering (Heyer *et al.*, 1999). SOMs are a form of artificial neural network that use machine learning algorithms to reduce the dimensionality of the input data. This reduction in dimensionality highlights similarity between objects allowing similar objects to be grouped. In k-means clustering, objects are randomly assigned to each of the specified clusters, and an iterative procedure aims to reduce the within-cluster variance, but increase the between-cluster variance. In both cases, the user must specify the number of clusters required *a priori*, which can affect the clustering outcome. QT clustering improves on k-means for gene-expression analysis by allowing users instead to specify the minimum size of clusters, and the level of relatedness between genes.

Many measures of similarity can be used to identify similarity between gene-expression signatures, and the resulting dendogram formed by applying this clustering method to the data can be greatly affected by the choice of similarity metric. One often used method is the Pearson cross-correlation, or product-moment correlation. For two vectors $X = (x_1, \dots, x_n)$ and $Y = (y_1, \dots, y_n)$ with means $\bar{X}$ and $\bar{Y}$ respectively, the Pearson coefficient is:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{X})^2 \sum_{i=1}^{n}(y_i - \bar{Y})^2}} \qquad \textbf{1-19}$$

Such clustering procedures are often used to identify genes showing similar expression profiles over time (co-expression), or to identify groups of individuals showing similar expression over all genes (subgroup classification). Heatmaps combine hierarchical clustering of genes and samples in a single image, typically with genes as rows and samples as columns. By assigning a colour representing gene-expression levels to each element of the resulting array, clusters can be identified as blocks of similar colour (Figure 1.5.2). This provides an informative visual representation of the association between samples and genes (Eisen *et al.*, 1998).

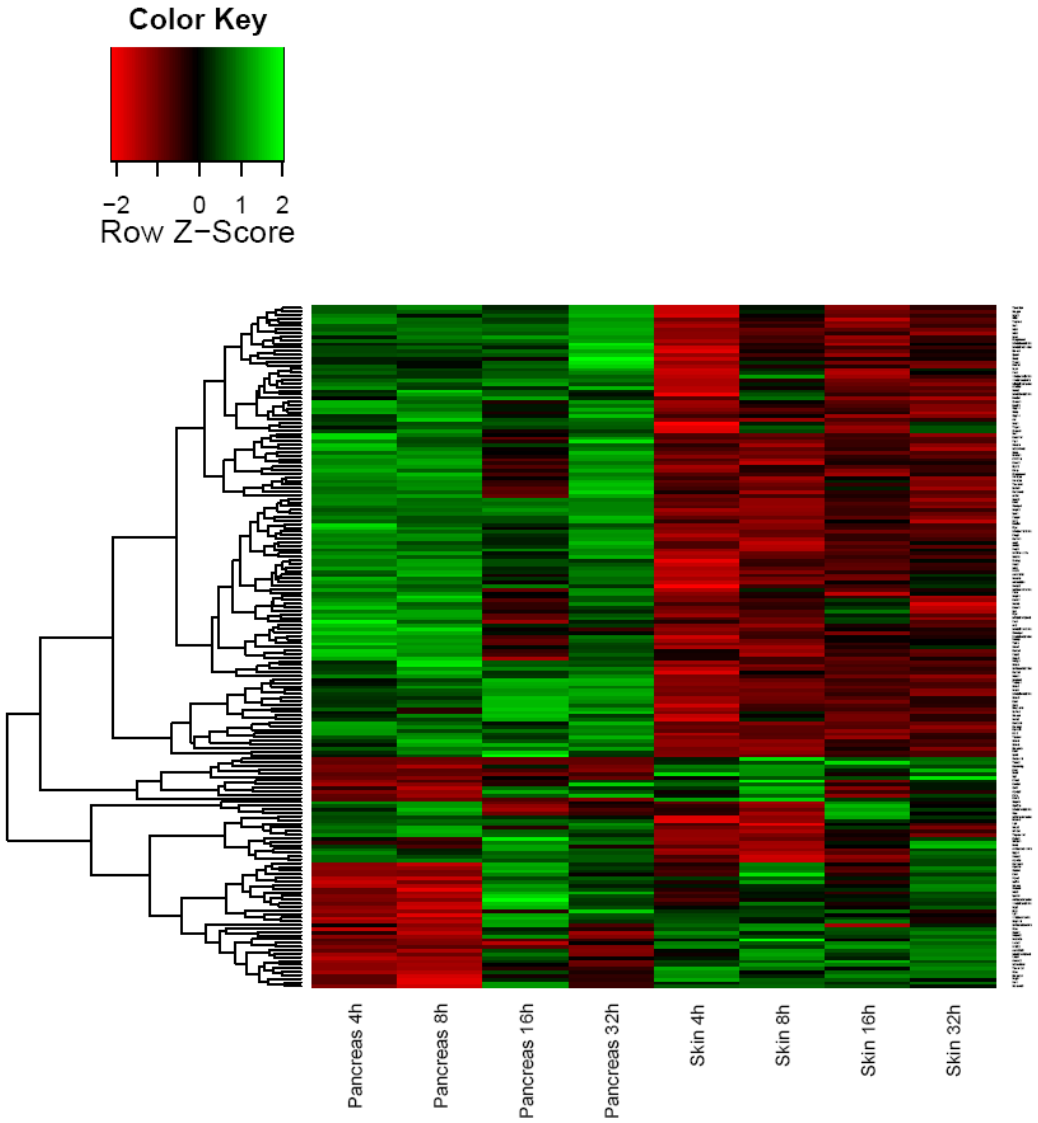

Figure 1.5.2: Visualising the association between genes and samples using a heatmap

Hierarchical clustering can be used to rank genes based on their similarity across the sample set. Some similarity measure (such as the Pearson correlation coefficient) is used to identify genes showing similar expression profiles across the experiment conditions. Genes are coloured based on their expression for a particular condition (green; high expression, red; low expression) producing a heatmap of the data, and allowing visual inspection of relatedness between genes and samples. A hierarchical tree, shown to the left of the above heatmap, can be used to identify clusters within the data. Blocks of colour within the heatmap may represent interesting clusters of genes whose expression is strongly correlated, perhaps indicating co-regulation and biological similarities.

1.5.5 Validation of results

Before making conclusions regarding microarray results, investigators must be certain that the results provide an accurate representation of the biological system under study. Validation of results can largely fall into three categories – quality control of data, independent confirmation of results and comparison of results with the biological system under study (reviewed in Chuaqui *et al.*, 2002). Quality control of Affymetrix data has been previously described (Section 1.4.3). Independent confirmation of the results can include laboratory based confirmation of mRNA levels using more sensitive gene-expression analysis methods, and *in silico* analysis comparing results with previous studies to confirm certain trends in the data. Finally, these results must be tied in to the biology of the system to confirm that they make sense within the context of the biological system under study.

The method of statistical hypothesis testing used to test for differential expression often produces false positive results. Microarray experiments typically produce a large number of potentially interesting results. To confirm that these results are in fact an accurate representation of the biology of the system, more sensitive methods of gene-expression analysis are often used to validate the results. It is thus important to minimise errors (particularly Type I errors) in the initial analysis to minimise the costs of validation.

One method typically used in microarray validation is qRT-PCR (Rajeevan *et al.*, 2001), which allows real time quantification of the change in PCR product throughout the PCR amplification. In the TaqMan qRT-PCR assay (Heid *et al.*, 1996), a non-extendible oligonucleotide probe is engineered for the transcript of interest, and is labelled at the 5' end with a fluorescent reporter dye (e.g. 6-carboxyfluorescein (FAM) or 2′-chloro-7′-phenyl-1,4-dichloro-6-carboxyfluorescein (VIC)). A non-fluorescent minor-groove binding (MGB) quencher molecule is also attached to the probe at the 3' end, and serves to quench the emission spectra of the reporter through fluorescence resonance energy transfer (FRET) (Forster, 1948; Applied Biosystems, 2005). The close proximity

of the fluorophore to the quenching molecule ensures that no fluorescence is detectable. During the PCR reaction, the probe binds to the transcript of interest, resulting in cleavage of the reporter dye by the 5' nuclease activity of the Taq polymerase during the extension phase. This allows the reporter to escape the quenching activity, resulting in an increase in fluorescence. This cleavage also removes the probe from the target strand, allowing PCR to continue unabated. Thus fluorescence levels increase following each PCR cycle in proportion to the amount of amplicon produced in the PCR reaction.

Assuming complete efficiency of the PCR reaction, amplicon abundance is doubled every cycle. However, the observed changes in fluorescence during the early cycles of the PCR reaction are low and not detectable by the scanner. A threshold value for signal detection is set, and the (fractional) cycle number at which fluorescence levels first exceed this value is termed the threshold cycle (C_T). This value is used to estimate transcript abundance, since a higher copy number in the starting sample will require fewer cycles (and hence a lower C_T) for amplicon abundance to reach suitable levels for detection. The threshold for detection is set such that, at the point of detection, amplification is in the exponential stage, which is limited by reagent availability within the assay. This ensures that the C_T value can be used to quantify transcript abundance, either absolutely by extrapolating from a previously produced standard curve of known transcript concentrations, or comparatively by comparing directly between two samples. qRT-PCR is sensitive over a larger dynamic range than microarrays, is able to accurately identify lower fold changes, and requires only a small quantity of starting material making it ideal for validation of microarray results.

1.6 Thesis overview

The first section of this introductory chapter (Section 1.1) has given an overview of the current state of knowledge about the *c-myc* oncogene, its role in normal cellular function, and also its role in diseases such as cancer. It is clear that, whilst much is known about the myriad cellular functions of Myc, there is still much to be learnt about the role played by tissue context *in vivo*. Section 1.2 discusses a well characterised *in vivo* model for deregulated Myc, and in particular the difference in the phenotypic outcomes in two diverse tissues; skin and pancreas. The divergence of Myc potentiality between the two tissues, leading to suppression of tissue expansion via apoptosis in one but not the other, highlights the need for a real understanding of the pathways involved in these two functions. In Section 1.3, a method for high-throughput analysis of transcriptional response – the microarray – is discussed, which provides a method for analysing the key divergences in Myc-regulated processes at the transcriptional level between the skin and the pancreas. In Section 1.4, the Affymetrix system of GeneChip microarrays is discussed, and Section 1.5 describes methods and considerations for experimental design, data analysis and identification of significantly changing genes. Lastly, the aims for this project are formally expressed in Section 1.7.

Chapter 2 discusses materials and methods used throughout this project, including many of the careful experimental techniques that were employed to ensure validity of the results.

High-throughput analyses such as microarrays produce large amounts of data and successful hypothesis testing requires suitably designed statistical methods. Chapter 3 discusses the motivation behind the creation of a new package for the Bioconductor project (Gentleman *et al.*, 2004), a collection of libraries for analysis of gene-expression data in the statistical programming language R. This package, *Envisage* (Enables Numerous Variables In Significant Analysis of Gene-expression), was specifically designed for the present analysis, but is also widely applicable to a wide range of data.

Chapter 4 is the main chapter of the thesis and discusses the results of a microarray-based transcriptional comparison between Myc activation in the skin and Myc activation in the pancreas using the MycER^{TAM} transgenic model.

Finally, Chapter 5 gives a final discussion of the results seen throughout the project and how they relate to the initial aims and hypotheses.

1.7 Project aims and hypotheses

The project described within this thesis has been separated into two main chapters: Chapter 3: *Envisage*: Significance Analysis of Microarray Data and Chapter 4: Comparison of Transcriptional Response to MycERTAM Activation in Suprabasal Keratinocytes and Pancreatic β-Cells. The aims and hypotheses for each of these are described here.

1.7.1 *Envisage*: Significance Analysis of Microarray Data Using Linear Models

Many sources of variation exist within microarray experiments, and sample-to-sample variation in phenotype and sample preparation can elicit a significant effect on gene-expression. Analysis of differential expression may therefore identify genes whose expression varies in response to variables that are not of primary interest to the investigator. The aim of this chapter was to develop statistical methods, based around linear models, allowing analysis of significant differential expression across a large number of experimental variables, in order to allow biological context to be factored into the analysis. Given the large size of microarray-derived data sets and the problems implicit in their analysis, this was designed to be automated and simple to use in order to limit user-errors. To ensure standards were maintained throughout, this procedure was designed to integrate with standard analysis packages: GeneSpring GX 7.3.1 (GS-GX; Agilent Technologies, Santa Clara, CA) and the Bioconductor library packages in R (Gentleman *et al.*, 2004).

1.7.2 Comparison of Transcriptional Response to MycERTAM Activation in Suprabasal Keratinocytes and Pancreatic β-Cells

It is clear from Section 1.2.4 that tissue context plays a large role in determining the ultimate role of deregulated Myc. The divergence of Myc potentiality between conditions that favour opposing outcomes – proliferation and apoptosis – is of key importance to understand the role of oncogenic Myc in circumventing normal cell function. The aim of this chapter was to analyse gene-expression changes following a time-course of Myc activation in the skin and pancreas to identify downstream targets of deregulated Myc that promote cell replication/survival and apoptotic cell death. Comparative analysis of these results, utilising novel multi-variable statistical analysis techniques based on linear models for identification of significant changes in gene-expression in complex experimental design, identified key genes whose expression profiles between the tissues may delineate the seemingly contradictory phenotypes.

Chapter 2 Materials and Methods

2.1 Treatment of transgenic animals

This study utilised the switchable *mycER^{TAM}* transgenic mouse model, described in Section 1.2, to allow direct control of aberrant Myc activity *in vivo*. The *inv-mycER^{TAM}* (Pelengaris *et al.*, 1999) and *pins-mycER^{TAM}* (Pelengaris *et al.*, 2002b) transgenic mouse models were used, whereby the *mycER^{TAM}* transgene is targeted specifically to the suprabasal keratinocytes and pancreatic beta cells respectively. This model allowed exquisite control of the time 0 of ectopic Myc activity, allowing analysis of downstream events at the molecular level.

2.1.1 Genotyping of *mycER^{TAM}* transgenics

For both *inv-mycER^{TAM}* (Pelengaris *et al.*, 1999) and *pins-mycER^{TAM}* (Pelengaris *et al.*, 2002b) transgenic mice, tissue was collected for genotyping using an ear-punch technique. DNA was extracted by lysing tissue in 75 µl 'Hotshot' reagent (25 mM NaOH, 0.2 mM disodium ethylenediaminetetra acetic acid (EDTA) and pH 12) for 30 mins. Samples were neutralised by adding equal volumes of neutralising reagent (40 mM Tris-HCl, pH 5) and cooling to 4 °C overnight. DNA samples were stored at -20 °C for long-term storage. For each DNA sample, a PCR cocktail was made up containing 3 µl DNA, 17.25 µl sterile H_2O, 2.5 µl x10 PCR buffer (Invitrogen, Carlsbad, CA), 0.75 µl 50 mM $MgCl_2$ (Invitrogen, Carlsbad, CA), 0.5 µl MYC5 3'-primer (10 pm/µl; 5'-AGGGTCAAGTTGGACAGTGTCAGAGTC-3'), 0.5 µl MERTM 5'-primer (10 pm/µl; 5'-CCAAAGGTTGGCAGCCCTCATGTC-3'), 0.25 µl Taq polymerase (5 U/µl; Invitrogen, Carlsbad, CA) and 0.25 µl deoxynucleotide triphosphate (dNTPs; 10 mM; Invitrogen, Carlsbad, CA). PCR cocktails were run using a PTC-

100 Programmable Thermal Controller (MJ Research, Inc., Waltham, MA) using the following program:

1. 2 mins at 94 °C
2. 30 cycles of: 1 min at 94 °C, 1 min at 57 °C, 2 mins at 72 °C
3. 10 mins at 72 °C

PCR products, positive and negative control samples (*mycER^{TAM}*-positive DNA and water respectively) and a 1 kb DNA ladder were loaded on a 1 % agarose gel with Tris/borate/EDTA buffer (TBE; 0.89 M Tris base, 0.02 M EDTA-Na_2-salt, 0.89 M boric acid). with a 6x loading buffer (0.25 % bromophenol blue, 0.25 % xylene cyanol FF and 30 % glycerol in water at 4 °C). The gel was run for 2 hours at 90 V and migration was captured using the Gene Genius Bio Imaging System with the GeneSnap Image Capture Suite (Syngene, Frederick, MD).

2.1.2 Administration of 4-hydroxytamoxifen (4OHT)

Activation of the MycER^{TAM} protein in adult transgenic mice was achieved through daily administration of 4OHT (Sigma-Aldrich, St. Louis, MO). For *inv-mycER^{TAM}* mice expressing MycER^{TAM} in suprabasal keratinocytes, 4OHT was dissolved in ethanol (1 mg/0.2 ml) and 200 µl was applied topically to a shaved area of dorsal skin daily. For *pins-mycER^{TAM}* mice expressing MycER^{TAM} in pancreatic β-cells, 4OHT was sonicated in peanut oil (1 mg/0.1 ml) and 100 µl was administered through daily IP injection. Control VT mice received equal volumes of their respective vehicle (ethanol or peanut oil respectively) without 4OHT daily.

2.1.3 Tissue excision and preparation

8-12 week old male *inv-mycER^{TAM}* and *pins-mycER^{TAM}* mice were sacrificed by cervical dislocation 4, 8, 16 or 32 hours after initial 4OHT/vehicle dose (time 0). Pancreata and dorsal skin were immediately dissected subsequent to sacrifice. Isolated tissue was bisected; one tissue sample was immediately embedded in optimum cutting temperature (OCT) medium on dry ice at an orientation to ensure optimum size for tissue sections. A second sample was snap frozen in liquid nitrogen for future functional validation studies. Tissue samples were stored at -80 °C.

2.1.4 Sample labelling

This study consisted of 16 individual experimental conditions; 4 conditions over 4 time points. Each condition was represented by 3 independent replicates for a total of 48 independent samples. Labelling of samples consisted of an identifier for each of the 4 main conditions followed by the time point in hours and a number used to identify independent replicate samples; skin with MycER^{TAM} active (Skin T), skin with MycER^{TAM} inactive (Skin U), pancreas with MycER^{TAM} active (Panc T) and pancreas with MycER^{TAM} inactive (Panc U) across the four time points.. For example, ‘Panc T 16hr (3)’ represents the third replicate of pancreas samples treated with 4OHT (MycER^{TAM} active) for 16 hours, and ‘Skin U 32hr (1)’ represents the first replicate of skin samples treated with vehicle only (MycER^{TAM} inactive) for 32 hours.

2.2 Laser capture microdissection

The *pins-mycER*^{TAM} transgene is expressed only in the Insulin-producing β-cells which exist within islets; homogenous clusters of endocrine cells making up roughly 2 % of the total mass of the mouse pancreas (Elayat *et al.*, 1995). β-cells are by far the most abundant of the islet cells, yet make up only a small proportion of the total pancreas cell mass. Therefore, activation of MycER^{TAM} results in changes in expression in only a minority of cells within the pancreas. It is therefore important to remove non-MycER^{TAM}-producing cells to ensure that changes in gene-expression following MycER^{TAM} activation in the β-cells are not overshadowed by intrinsic changes in gene-expression within non-β-cells due to normal cellular homeostasis over the time course. Laser capture microdissection (LCM) is a microscopic technique allowing isolation of pure cell populations from heterogenous tissue sections whilst leaving morphology intact (Figure 2.2.1). Tissue sections are cut and mounted on a special membrane slide consisting of a metal support frame with a transparent transfer film. Sections are fixed, stained and dehydrated, before being placed on a microscope platform for morphological identification. Cells of interest are selected, and a low powered infra-red laser is used to isolate selected cells from the surrounding tissue. This process results in minimal damage to cell morphology and leaves DNA and RNA molecules intact, making this procedure excellent for isolation of RNA molecules for gene-expression analysis.

This method was employed to isolate homogenous populations of endocrine islet cells from the surrounding exocrine tissue in pancreas. This ensured that the majority of RNA species originated from MycER^{TAM}-encoding β-cells and reduced contamination by non-MycER^{TAM}-encoding exocrine cells.

Standard protocols for LCM, such as the Arcturus Histogene LCM staining protocol (Arcturus, Mountain View, CA) and the P.A.L.M. Microlaser Systems protocol (Zeiss, Bernried, Germany) use similar methods for sample preparation prior to LCM. Sections are cut and allowed to air dry at room temperature before fixing in ethanol (usually 70 %). Several staining protocols are suggested, such as

a haematoxylin and eosin stain (as described in Materials and Methods, Section 2.8.1), and sections are dehydrated prior to laser cutting using a series of graded ethanol washes (50 % – 100 %) followed by a final wash using xylene. Evaporation dries the tissue sections allowing the laser to accurately cut the tissue.

However, due to the high number of RNase species within the pancreas (Chirgwin *et al.*, 1979; Gill *et al.*, 1996; Mullin *et al.*, 2006), these protocols produced RNA of a low quality, unsuitable for microarray hybridisation. Optimisation of the protocol, as described in Section 4.2.2.1, produced a procedure allowing isolation of RNA from pancreatic islets of a suitable quality for hybridisation to Affymetrix GeneChips (Materials and Methods, Section 2.2.1). However, similar attempts to optimise the protocol for isolation of RNA from skin keratinocytes proved impractical (Section 4.2.2.2). Given that the level of contamination from non-MycER^{TAM}-encoding cells was more favourable for the skin than the pancreas, RNA was isolated from whole skin tissue prior to microarray hybridisation (Materials and Methods, Section 2.2.2).

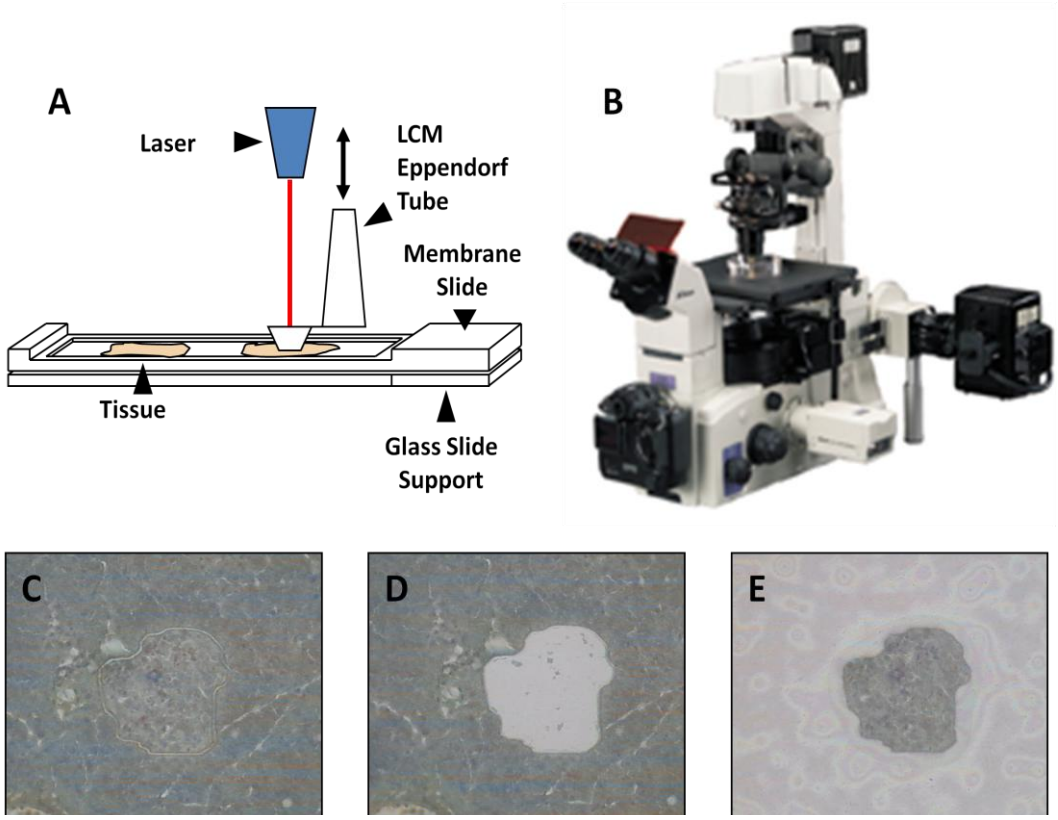

Figure 2.2.1: Laser capture microdissection allows isolation of pure cell populations while maintaining tissue morphology

Tissue sections are bound to the thermal membrane of an LCM slide and mounted on a glass slide for support (A). Slides are transported to the platform of an LCM microscope (B) and the cap of an LCM eppendorf tube is lowered onto the membrane. A low powered infra-red laser is used to cut around cell populations of interest (C) and lifted away with the sticky eppendorf cap (D). The isolated tissue (E) can be used for downstream processing, such as RNA extraction.

2.2.1 Optimised laser capture microdissection and RNA isolation from pancreatic islets

Fresh frozen pancreas sections were cut to a thickness of 15 µm using a Bright 5040 cryostat (Jencons Scientific, Bridgeville, PA), bound to a MMI MembraneSlide (Molecular Machines and Industries, Rockledge, FL) and fixed in ice-cold 100 % ethanol for 2 minutes. Sections were stained briefly (10 secs) with a 1 % Toluidine Blue dye in 100 % ethanol to allow identification of islet morphology whilst preserving RNA quality. Stained sections were dehydrated in 2 changes of 100 % ethanol and 2 changes of xylene for 1 minute each, airdried for 2 minutes and finally left in a vacuum dessicator for 5 minutes before transportation to the laser capture platform.

The SL µCut laser capture microdissection system (Molecular Machines and Industries, Rockledge, FL) was used to isolate islets of Langerhans from surrounding exocrine tissue. Isolated islets were collected on the lid of a MMI IsolationCap (Molecular Machines and Industries, Rockledge, FL) and RNA was homogenised in a solution of buffer RLT (Qiagen, Valencia, CA), a guanidine-isothiocyanate-containing lysis buffer, with β-mercaptoethanol (100:1) as described in the RNA Microkit protocol. The laser capture procedure was repeated on freshly cut pancreas sections to collect a total area of islet cells equal to roughly 1.5 x 10^6 μm^2 for each sample. Time on the microscope platform was strictly limited to 15 minutes to minimise RNA degradation. Lysed cell samples collected from a single tissue source were pooled and stored at -80 ºC prior to RNA extraction.

2.2.2 RNA isolation from skin keratinocytes

Five fresh frozen skin sections (20 µm) were collected across several levels of OCT-embedded tissue using a Bright 5040 cryostat (Jencons Scientific,

Bridgeville, PA) and lysed directly in a solution of buffer RLT (Qiagen, Valencia, CA) and β-mercaptoethanol (100:1). Samples were vortexed to homogenise and stored at -80 ºC prior to RNA extraction.

2.3 RNA extraction

Total RNA was isolated from homogenised pancreata and skin lysates using the Qiagen RNeasy Micro Kit (Qiagen, Valencia, CA) for small sample RNA preparation, incorporating a deoxyribonuclease (DNase) I treatment step to remove DNA molecules that may copurify with the silica-gel columns. RNA integrity was protected by cleaning work areas and instruments with RNaseZap® (Ambion, Austin, TX) and DEPC-treated water to inactivate RNases. Nuclease-free water was used throughout to ensure that no RNases were introduced, and good laboratory practices were observed at all times. RNA was isolated from the forty-eight tissue samples described here, as well as a further thirty-six pancreatic tissue samples as part of a parallel study not discussed within this thesis, resulting in a total of eighty-four individual RNA samples. Samples were randomly assigned into seven batches of twelve for RNA extraction, in a batching scheme chosen such that each batch filled a single plate for RNA quantification as described below (see below). It is noted however that, if RNA was extracted solely from tissue samples described for the current experiment, a more natural scheme may have been to divide samples into eight batches to allow sensible allocation of treatment sets to blocks of treatment and time point.

Lysed cell samples were made up to 350 µl in buffer RLT with β-mercaptoethanol (100:1) and vortexed for 30 secs. 1 volume (350 µl) 70 % ethanol was added to the homogenised lysate and mixed thoroughly by pipetting. The mixture was added to an RNeasy MinElute silica-gel column (Qiagen, Valencia, CA) in a 2 ml collection tube and centrifuged at $\geq 8000 \times g$ for 15 secs. The flow-through was discarded, and 350 µl buffer RW1 (a guanidinium thiocyanate-containing buffer; Qiagen, Valencia, CA) was added to the spin column before centrifuging at $\geq 8000 \times g$ for 15 secs to wash the column. The flow-through was discarded, and a DNase I incubation mix (10 µl DNase I stock solution, 70 µl buffer RDD (Qiagen, Valencia, CA)) was applied directly to the MinElute silica-gel membrane for 15 mins at RT to remove genomic DNA. The column was washed by adding 350 µl buffer RW1 and centrifuging for 15 secs at $\geq 8000 \times g$. The MinElute

column was transferred into a new 2 ml collection tube, 500 µl buffer RPE (Qiagen, Valencia, CA), diluted from fresh in 4 volumes 80 % ethanol, was added to the column and centrifuged for 2 mins at $\geq 8000 \times g$ to dry the silica-gel membrane. The MinElute column was transferred to a new 2 ml collection tube and centrifuged with an open cap at full speed for 5 mins to further dry the membrane. RNA was eluted by applying 14 µl nuclease-free water directly onto the silica-gel membrane and centrifuging at maximum speed for 1 min.

RNA integrity was analysed by the Molecular Biology Service (University of Warwick) using an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA) and RNA was quantified using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, DE). RNA quality and yield is discussed further in Section 4.2.2.3. All samples were included in the microarray hybridisation procedures as it was decided that the introduction of systematic errors by repeating poor quality samples in isolation would be more detrimental to the resulting data than using degraded starting material.

2.4 Affymetrix GeneChip protocols

2.4.1 Optimised *in vitro* transcription (IVT) protocol

Sample preparation for microarray hybridisation was conducted using the GeneChip Two-Cycle Target Labelling and Control Reagents kit (Affymetrix, Santa Clara, CA) by Lesley Ward at the Molecular Biology Service at the University of Warwick. The two-cycle target labelling reaction was used instead of the standard one-cycle reaction due to the increased amplification that this protocol offers. This technique is recommended for RNA isolated from cells collected using LCM, due to the low yields of RNA often produced when using this technique. As described in Section 4.2.3, use of standard protocols produced a low yield of biotin-labelled cRNA. Personal communication with Giorgia Riboldi-Tunnicliffe from Affymetrix allowed the development of a modified protocol, producing yields of labelled cRNA suitable for microarray hybridisation.

Briefly, in the first cycle, poly-A controls were added to 10 ng sample RNA in 6 µl nuclease-free water (twice the recommended volume) and used as a template for synthesis of 1st strand antisense cDNA. T7-oligo-deoxythymidine (T7-oligo(dT)) primers were bound to the poly-A tail of mRNA species and DNA reverse transcription was initiated using Superscript II reverse transcriptase (Affymetrix, Santa Clara, CA) to produce an antisense cDNA strand. Sense RNA was degraded using RNase H leaving the antisense cDNA template, and 2nd strand sense cDNA was synthesised using *E. coli* DNA polymerase I. In order to improve first round cRNA yield, double volumes of polyA control probes and first cycle reagents were used. *In vitro* transcription was performed using the MEGAscript T7 kit (Ambion, Austin, TX) to produce cRNA from the T7-promoter-containing antisense cDNA strand. First round cRNA clean up was performed using the Affymetrix GeneChip sample cleanup module (Qiagen, Valencia, CA).

In the second cycle, polyadenylated random primers were used in 1st strand cDNA synthesis to produce short polyadenylated sense cDNA. cRNA was degraded using RNase H, and 2nd strand cDNA synthesis, incorporating T7-oligo(dT) primers, was performed using *E. coli* DNA polymerase I to produce a second T7-primed antisense cDNA strand. Double stranded cDNA cleanup was performed using the Affymetrix GeneChip sample cleanup module (Qiagen, Valencia, CA).

I*n vitro* transcription was performed using the Affymetrix GeneChip IVT labelling kit (Affymetrix, Santa Clara, CA) incorporating biotin labelled ribonucleotides. This produced double-amplified biotin labelled RNA from the T7-promoter-containing antisense cDNA strand complementary to the starting mRNA (double-amplified biotin-labelled cRNA; 2a-cRNA). 2a-cRNA cleanup was performed using the Affymetrix GeneChip sample cleanup module (Qiagen, Valencia, CA), and samples were fragmented prior to hybridisation to the array (Affymetrix, 2004).

The resulting yields of 2a-cRNA are described in Section 4.2.3.2. Five samples were found to produce a yield less than the 10 μg recommended for microarray hybridisation. Given that the effects of confounding variation (such as changes in gene-expression relating to circadian rhythms and those relating to environmental effects) were avoided by running treated and untreated time courses in parallel, it was decided to hybridise these sub-optimal samples regardless to avoid introducing confounding variation relating to temporal batch effects by processing samples at a later date (although such effects could be minimised by ensuring that treatments occurred at times comparable to those used for the original samples). The choice between the introduction of bias due to low signal levels inherent with the hybridisation of low levels of 2a-cRNA, and the introduction of temporal batch effects, was also influenced by the appreciable time and resources required to run additional samples. This would also allow observation of the quality of data obtainable when pushing the boundaries beyond those suggested by Affymetrix.

2.4.2 Microarray hybridisation and scanning

Forty-eight 2a-cRNA samples were randomly and independently hybridised to Affymetrix MOE430 Plus 2.0 GeneChip microarrays, which have been previously shown to give accurate reproducible results (Affymetrix, 2003). Hybridisation was performed by Lesley Ward at the Molecular Biology Service facility at the University of Warwick following standard protocols (Affymetrix, 2004).

2.4.2.1 Hybridisation

A hybridisation cocktail was made up for each 2a-cRNA sample, containing 10 µg fragmented 2a-cRNA, 5 µl control B2 oligonucleotides (3nM), 15 µl 20x eukaryotic hybridisation controls (heated to 65 °C for 5 mins before making aliquots), 3 µl herring sperm DNA (10 mg/ml), 3 µl acetylated bovine serum albumin (BSA; 50 mg/ml), 150 µl 2x hybridisation buffer and 109 µl nuclease-free water. 200 µl of this cocktail was heated to 99 °C for 5 mins, followed by 45 °C for 5 mins in a 0.5 ml RNase-free tube. Hybridisation cocktail was centrifuged at maximum speed for 5 mins to remove insoluble material. 200 µl 1x hybridisation buffer was incubated on the array for 10 mins at 45 °C with rotation at 60 revolutions per minute (rpm). Hybridisation buffer was removed, and 200 µl clarified hybridisation cocktail was incubated on the wet array at 45 °C with rotation for 16 hours at 60 rpm.

2.4.2.2 Washing

Arrays were washed with non-stringent wash buffer (300 ml 20x Saline-Sodium Phosphate-EDTA (SSPE) buffer (0.75 M NaCl, 50 mM NaH_2PO_4, 5 mM EDTA, made up to pH 7.0 with NaOH), 1 ml Tween-20 (10 %) and 699 ml nuclease-free water for 1 L stock solution) and stringent wash buffer (83.3 ml 12x 2-(*N*-morpholino) Ethanesulfonic buffer (MES) stock, 5.2 ml NaCl, 1 ml Tween-20,

910.5 ml nuclease-free water for 1 L stock solution) using the Affymetrix Fluidics Station 450 (Affymetrix, Santa Clara, CA).

2.4.2.3 Staining

Arrays were stained using the Affymetrix Fluidics Station 450 (Affymetrix, Santa Clara, CA), loaded with 2x stain buffer (41.7 ml 12x MES stock, 92.5 ml NaCl (5 M), 2.5 ml 10 % Tween-20 and 113.5 ml nuclease free water for a 1 L stock solution), SAPE solution (600 µl 2x MES stain buffer, 48 µl acetylated BSA (50 mg/ml), 12 µl SAPE (1 mg/ml), 540 µl nuclease-free water per sample), and antibody solution (300 µl 2x MES stain buffer, 24 µl acetylated BSA (50 mg/ml), 6 µl normal goat Immunoglobulin G (IgG; 10 mg/ml), 3.6 µl biotinylated antibody (0.5 mg/ml), 266.4 µl nuclease-free water per sample).

2.4.2.4 Scanning

Arrays were scanned using the Affymetrix GeneChip Scanner 3000 7G (Affymetrix, Santa Clara, CA), and samples were analysed using GCOS. Pre-processing of .DAT raw fluorescent images was performed as described in Section 1.4.1 to produce a series of .CEL data files, which were used in further analyses. Data quality control metrics were calculated and analysed as described in Section 1.4.3.

2.5 Experimental design

The need for careful experimental design in microarray experiments was discussed in Section 1.5.1, and these considerations were taken into account when planning and implementing this experiment. To ensure that gene-expression values accurately represented those of the global population, all conditions were represented by three independent hybridisations, resulting in a total of 48 independent hybridisations. Whilst three replicates represents the minimum number that should be used in a microarray experiment (Section 1.5.1.3), this allowed the maximum number of time points to be analysed given the limited budget for the experiment. Also, many sources of variation were limited through design: between-sample variation was reduced by using inbred transgenic lines to limit animal-to-animal genetic variation, mice were housed under environmentally controlled conditions to limit environmental effects on gene-expression, and to limit further effects all animals chosen for this study were of the same gender (male) and aged between 8-12 weeks. A great deal of care was also taken to limit variation in sample preparation, such as by following standardised protocols at all times and by ensuring all replicates were treated concomitantly to limit circadian-based variation in gene-expression response.

A further important experimental design feature was the randomisation of samples during processing to limit the effects of systematic errors on the resulting data. The typical approach to experimental design to avoid the effects of confounding covariates is to create groups (or blocks) of samples with similar covariate values. In this way, such covariate effects become orthogonal to the treatment effects, allowing them to be more easily accounted for in the analysis. Further covariate effects can then be minimised by randomising sample treatments within these blocks. However, given the limited availability of transgenic animals from similar litters, and given that the most significant covariates (e.g. gender, age, temperature, humidity, and experimenter) were largely controlled by the laboratory settings, blocking was not performed in this experiment. To account for covariate effects, samples were re-randomised during key stages of processing

(LCM, RNA extraction, and microarray hybridisation). The thinking behind this process was to ensure that, when logistics required separation of samples into batches for processing, this elicited no significant systematic effect on the response of the gene-expression ('batch effects'). However, it is noted that such continuous re-randomisation may actually introduce error variability rather than reduce it.

To analyse the differences in transcriptional response to deregulated Myc function in the two tissues of interest (skin and pancreas), time courses were set up following activation of MycER^{TAM} via 4OHT administration using the *pins-mycER^{TAM}* and *inv-mycER^{TAM}* transgenic models described in Section 1.2. Time points considered were 4, 8, 16 and 32 hours following initial activation of the MycER^{TAM} protein, with each time point and condition represented by unique transgenic animals. These were chosen to allow analysis of early gene-expression changes when Myc transactivation is maximal (Wu *et al.*, 1999), and also to allow analysis of the gene-expression signature of the two phenotypes at later times when transcriptional response had diverged to elicit the ultimate phenotypic outcome (unchecked proliferation in the skin, apoptosis in the pancreas).

To ensure that observed changes in gene-expression were related to Myc-function, and not due to a response to stress factors, circadian rhythms, natural cellular function, etc., control time courses were set up in parallel, with animals treated with their respective vehicle (peanut oil for pancreas, ethanol for skin) as with their 4OHT counterparts (Materials and Methods, Section 2.1.3). Vehicle-treated (VT) samples acted as direct inactive MycER^{TAM} controls to their active MycER^{TAM} counterparts. Direct comparison between the two produced relative gene-expression values relating to changes induced by activation of the MycER^{TAM} chimeric protein.

This experimental design represented a complete factorial design across the three main experimental factors of: tissue type (skin or pancreas), 4OHT treatment (treated or untreated) and time point following initial 4OHT dose (4, 8, 16 or 32 hours) as illustrated in Figure 2.5.1. This experiment represented 16 unique conditions; skin with MycER^{TAM} active (Skin T), skin with MycER^{TAM} inactive

(Skin U), pancreas with MycER^{TAM} active (Panc T) and pancreas with MycER^{TAM} inactive (Panc U), all across four time points.

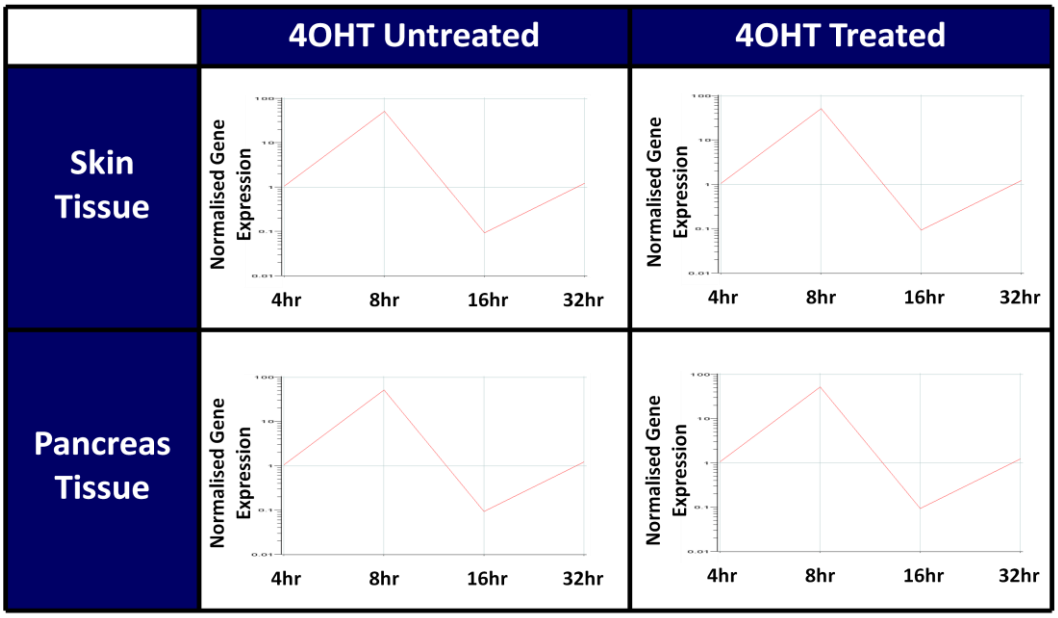

Figure 2.5.1: Experimental design for skin vs. pancreas study

Cartoon representation of the experimental design for the skin vs. pancreas microarray experiment, comparing the transcriptomic response to activation of the MycER^{TAM} chimeric protein of the skin and pancreas. 4OHT untreated samples (MycER^{TAM} inactive) acted as direct controls to their 4OHT treated (MycER^{TAM} active) counterparts. The gene-expression profile represented here is merely for illustration, and should not be taken as representative of the data for each condition.

2.6 Microarray Data Analysis

Analysis of microarray data was performed using a combination of the GS-GX gene-expression analysis suite, version 7.3.1 (Agilent Technologies, Santa Clara, CA) and the Bioconductor library packages (Version 2.1) in Version 2.6.2 of R (Gentleman *et al.*, 2004).

2.6.1 Normalisation of data

Sample data were background corrected, normalised and summarised using the GC-RMA probe-level normalisation procedure (Section 1.4.2.4**Error! Reference source not found.**). This was chosen due to the increased sensitivity afforded as compared to MAS 5.0 (see Figure 1.4.3). To ensure that no tissue bias was introduced, pancreas and skin samples were GC-RMA normalised as separate experiments, then combined prior to data analysis. The decision to normalise the data separately was based on discussions on the Bioconductor discussion forums. Summarised replicate data were further processed in GS-GX by normalising each of the 4OHT treated samples to the median of their respective controls (e.g. the median of the log-signals for the pancreas VT 4 hour samples was subtracted from the log-signal of each of the pancreas samples treated for 4 hours with 4OHT; Materials and Methods, Section 2.5). Data were further normalised by subtracting the normalised data for each probe for each sample by the median across samples.

Normalised values for each probe set therefore represented the fold change of the relevant transcript upon MycER^{TAM}-activation. This process served to both remove changes in gene-expression that were unrelated to Myc activity (normal cellular function, circadian rhythms, stress response to IP injection, etc.), and also to assign a specific meaning to the normalised expression values that could be instantly understood; namely, the fold-change in response to MycER^{TAM} activation.

2.6.2 Gene curation

The MOE430 PLUS 2.0 GeneChip contains probe sets measuring the expression of a total of 45,101 transcripts (representing ~34,000 genes) from the mouse transcriptome. However, it is unlikely that all of these probe sets will show significant changes across the experimental conditions since the majority of the transcripts will be similarly expressed in all samples. Since the main goal of this microarray experiment was to identify genes showing significant differential expression between the experimental conditions, gene curation was performed to remove probe sets that did not measure interesting changes in gene-expression. Curation of the probe set list consisted of the removal of the following probe sets:

1. **Control probe sets**

   Affymetrix GeneChips contain a number of control probes (Section 1.4.3.6) designed to measure expression changes in housekeeping genes and spike-in control transcripts. These should show identical results across samples, and were removed from future analyses. Analysis of the signal intensity data for these probes was incorporated into the QC analysis (described in Section 4.2.4), with poor-quality hybridisation resulting in non-comparable signal intensity for these probes as compared to other samples within the study.

2. **Absent probe sets**

   MAS 5.0 software assigns diagnostic flags to each probe set identifying the confidence in the detected signal (Section 1.4.3.1). Probe sets representing gene transcripts whose expression was not detected (*Absent*) in all samples were removed from further analyses (based on a comparison of the discrimination score R against the default detection threshold value of $\tau = 0.015$).

3. **Non-changing probe sets**

   Normalised expression values for each condition represent fold-change subsequent to treatment. Probe sets representing genes whose expression was altered less than 2-fold across all conditions were deemed uninteresting and were removed from further analyses. The choice of 2-fold as a threshold was arbitrary and was chosen due to convention.

4. **Highly variable probe sets**

   To ensure that only significant changes in gene-expression were considered (i.e. changes that would not be expected purely due to chance given the sample set), probe sets showing high variability across replicates (SD > 1 for all but 2 conditions) were removed.

Curation of the gene lists produced a list of 12,349 probe sets (representing 8,946 genes) for the experiment which were used in all subsequent analyses. It is noted that the removal of genes showing less than 2-fold change may result in the loss of potentially interesting genes, particularly since the transcriptional response to Myc is typically low (of the order of 2-3-fold change). This filter was included due to convention and to ensure that the number of genes to be analysed was kept to a managable number, although the use of statistical methods such as *Envisage* may have provided a more objective analysis. Also, the removal of highly variable probes may result in removal of probes whose mean signal change is also large, and hence may still show significant effects. However, it is noted that the number of genes removed due to high variation was low, so this loss of potentially interesting genes would have been low.

2.6.3 Identification of significant differential expression

The multi-factor nature of this experimental design inspired the creation of the gene-expression analysis package *Envisage* in R (Chapter 3). This package allows

analysis for significant differential expression across a number of experimentally-controlled variables and other sources of variation that may influence the response of the gene-expression to the main experimental variables (covariates). This allows identification of probe sets representing genes whose transcripts show significant response to the three main factors:

1. **4OHT treatment** (Treated or Untreated)
2. **Time after initial 4OHT treatment** (4, 8, 16 and 32 hours)
3. **Tissue type** (Skin or Pancreas)

Since the present study was conducted on inbred transgenic mouse lines, between-sample phenotypic variability was minimal. Animals were housed under controlled conditions, limiting environmental effects on gene-expression. Also, a great deal of effort was spent ensuring that all possible sources of variation in sample processing were reduced or removed (Materials and Methods, Section 2.5). However, given the wide range in the quality of starting material within this study (Section 4.2.2), it was of great interest to observe the effect that the RNA quality played on differential gene-expression. Also, given the necessity for randomised batching of samples to prevent introducing systematic bias into the data set, the severity of the role played by batching variables in determining the response in gene-expression was of interest.

Envisage was used to assess for the effects of the three main factors, whilst allowing for the impact of additional covariates for the RNA integrity (RIN), yield of 2a-cRNA from the IVT reaction, and RNA extraction batch identifier number (batch). It is noted that since sample batching followed a continuous re-randomisation scheme rather than a blocking scheme (randomising within treatment blocks), the impact of the batch on gene-expression response may be more difficult to deconvoluted.

The results for this analysis are presented in Section 3.4.2. Significant lists of genes found using *Envisage* were used in subsequent stages of the analysis.

2.6.4 Clustering of gene-expression data

Genes showing significant differential expression in response to variables of interest were clustered to identify possible co-regulation, indicative of functional relationships. Quality threshold (QT) clustering (Section 1.5.4) was performed in GS-GX, with a minimum cluster size of 14 and minimum correlation 0.9. Correlation between expression profiles was calculated using the Pearson cross-correlation coefficient.

2.6.5 Gene ontology (GO) classification of gene-expression data

Standard GO classifications were used to define the role of genes of interest in biological processes, cellular components, and molecular function (Schulze-Kremer, 1997; Schulze-Kremer, 1998; Ashburner *et al.*, 2000). Enrichment of GO terms in lists of significantly changing genes was tested using the GO browser in GS-GX, and the *annaffy* (Smith, 2007), *mouse4302* and *mouse4302cdf* (Ting-Yuan, ChenWei, *et al.*, 2007), *annotate* (Gentleman, 2007), *GOstats* (Gentleman and Falcon, 2007) and *GO* (Ting-Yuan, ChenWei, *et al.*, 2007) packages in Bioconductor. This allowed identification of interesting biological processes that were over-represented within the data set.

2.7 Quantitative real-time reverse transcription polymerase chain reaction (qRT-PCR)

TaqMan qRT-PCR (Section 2.7.4) was performed on original total RNA samples to validate gene-expression results for genes of interest. qRT-PCR for skin and pancreas 4OHT- and vehicle-treated samples was performed over the three replicates for early time-points 4 hrs and 8 hrs, and for the later 32 hrs time-point. Given constraints on resources towards the end of the project, the 16 hour time point was chosen as the least informative of the four time points and was left out of the qRT-PCR validation to allow analysis of a larger number of genes. Due to the limited nature of the samples, an amplification step was included to increase the abundance of transcripts for the genes of interest. This pre-amplification was carried out in a multiplexed reaction to minimise between-assay variability in amplification rates (see below). As with the microarray analysis, relative quantitative measures of gene-expression upon Myc-activation were calculated by comparing 4OHT- and vehicle-treated samples directly for each condition (Materials and Methods, Section 2.5). Assays were normalised using the endogenous 18s rRNA control probe (Applied Biosystems, Foster City, CA) to correct for differences in starting RNA concentrations.

2.7.1 Reverse transcription of RNA

20 ng total RNA was reverse transcribed to cDNA using a high-capacity cDNA reverse transcription kit (Applied Biosystems, Foster City, CA) specifically designed for use with small volumes of RNA such as those collected from laser captured tissue. Briefly, 10 µl dilute total RNA (2 ng/µl) was made up in a PCR reaction mix with 2 µl 10x RT (reverse transcription) buffer, 0.8 µl 25x dNTPs (100 mM), 2 µl 10x RT random oligo(dT) primers, 1 µl MultiScribe reverse transcriptase and 4.2 µl nuclease-free water. PCR cocktails were run on a PTC-

100 Programmable Thermal Controller (MJ Research, Inc., Waltham, MA) using the following program:

1. 25 °C for 10 mins
2. 37 °C for 120 mins
3. 85 °C for 5 secs.

cDNA was stored at 4 °C for short term storage (up to 24 hours) or at -20 °C for long term storage.

2.7.2 Pooling gene-expression assays

The 22 TaqMan qRT-PCR gene-expression assays (20x; Applied Biosystems, Foster City, CA; Table 2.7.1) were pooled for use in the pre-amplification multiplex reactions. The 22 assays were pooled in 1x TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7.5) such that each individual assay in the pool was at a final concentration of 0.2x. e.g. 5 µl of each 20x assay were pooled and made up with 390 µl TE buffer to a final volume of 500 µl).

2.7.3 Pre-amplification of cDNA

cDNA transcripts were pre-amplified prior to the qRT-PCR reaction by preparing multiplexed amplification cocktails for each sample consisting of 25 µl 2x TaqMan preAmp mastermix (Applied Biosystems, Foster City, CA), 12.5 µl pooled TaqMan qRT-PCR gene-expression assays (see above), 7.5 µl nuclease-free water, and 5 µl cDNA sample. Pre-amplification cocktail was run on a thermal cycler using the following program:

1. Hold at 95 °C for 10 mins
2. 14 cycles of 95 °C for 15 secs (denature) and 60 °C for 4 mins (anneal/extension)

Pre-amplified cDNA samples were diluted 1:20 in 1x TE buffer and stored at -20 °C prior to use.

2.7.4 qRT-PCR

A PCR reaction was set up on a 96-well PCR plate. For each well, the PCR master mix consisted of 25 µl TaqMan 2x gene-expression Master Mix, 12.5 µl diluted pre-amplified cDNA sample, 10 µl nuclease-free water, and 2.5 µl TaqMan qRT-PCR gene-expression assay or 18S rRNA control. Individual gene-expression assays, 18s rRNA endogenous positive control probe (Applied Biosystems, Foster City, CA) and water negative control were run in triplicate wells. qRT-PCR was performed using an ABI Prism 7000 scanner (Applied Biosystems, Foster City, CA), using the following program:

1. Hold for 2 mins at 50 °C for activation of Uracil-DNA glycosylase
2. Hold for 10 mins at 95 °C for activation of AmpliTaq Gold enzyme
3. 40 cycles of 95 °C for 15 secs (denature) and 60 °C for 1 min (anneal/extension)

Table 2.7.1: Gene-expression assays for quantitative real-time qRT-PCR.

Gene Assay	Catalogue Number	Product Code	Amplicon Length
18s rRNA	4319413E	NA	187
akt1	4331182	Mm00437443_m1	76
atr	4351372	Mm01223637_m1	132
ccna2	4331182	Mm00438064_m1	90
ccnb1	4331182	Mm00838401_g1	122
ccnd1	4331182	Mm00432359_m1	58
ccnd2	4331182	Mm00438071_m1	70
ccne1	4331182	Mm00432367_m1	63
ccne2	4331182	Mm00438077_m1	88
cdc2a	4331182	Mm00772471_m1	75
cdk4	4331182	Mm00726334_s1	54
cdkn1a	4331182	Mm00432448_m1	96
cdkn1b	4331182	Mm00438168_m1	81
cdkn2a	4331182	Mm00494449_m1	55
cdkn2b	4331182	Mm00483241_m1	112
cdkn2c	4331182	Mm00483243_m1	85
chk1	4331182	Mm00432485_m1	130
chk2	4331182	Mm00443844_m1	74
cycs	4351372	Mm01621044_g1	144
endog	4331182	Mm00468248_m1	104
fas	4331182	Mm00433237_m1	107
igf1	4331182	Mm00439561_m1	69
igf1r	4331182	Mm00802831_m1	106

2.7.5 Analysis of qRT-PCR data

As described in Section 1.5.5, the threshold cycle C_T is the (fractional) cycle at which a significant exponential amplification is detected above background signal. A smaller C_T value indicates higher transcript abundance in the original RNA sample. The C_T value can be used to compute an absolute quantitative value for transcript abundance using a previously calculated standard curve of known concentrations, or can be compared between samples to calculate relative abundance between two samples. Given the design of the experiment, relative quantitation was used to determine the change in gene-expression following MycER^{TAM} activation for both tissues across the time points. The C_T value for each transcript (averaged over all replicate wells) was compared to the C_T value of the endogenous control 18S rRNA (which should be equally expressed in all cells) to normalise all values to a common reference:

$$\Delta C_{T\ (Transcript\ )} = \ mean(C_{T\ (Transcript\ )}) - mean(C_{T\ (18s\ rRNA)}) \qquad \textbf{2-1}$$

The change in expression of a particular transcript at each time point and for each tissue type was calculated by comparing the ΔC_T values across the experimental conditions:

$$\Delta\Delta C_T \ = \ \Delta C_{T\ (MycER^{TAM}\ Active\ )} - \ \Delta C_{T\ (MycER^{TAM}\ Inactive\ )} \qquad \textbf{2-2}$$

2.7.6 Confirmation of uniformity for qRT-PCR pre-amplification

To ensure the efficacy of the pre-amplification reaction, and to confirm uniform amplification across all multiplexed gene-expression assays, qRT-PCR of non-limited high quality pancreas RNA ($RIN \cong 8$; kindly supplied by Luxian Zhou) was compared to qRT-PCR of dilute non-limited high quality RNA (1:500)

incorporating a pre-amplification step as described for all target genes considered. ΔC_T values were calculated for each transcript for the two conditions (each run in triplicate), and $\Delta\Delta C_T$ values for each transcript were calculated as:

$$\Delta\Delta C_T = \Delta C_{T\,(Pre-amp)} - \Delta C_{T\,(No\ pre-amp)} \quad \textbf{2-3}$$

The results of these calculations are shown in Table 2.7.2. Assays showing $|\Delta\Delta C_T| > 1.5$ (highlighted in red) were considered to show non-uniform amplification and were not used in the multiplexed gene-expression assay pool. Gene-expression assays for *ccne1* ($\Delta\Delta C_T = -2.921$), *cdkn2b* ($p15^{Ink4b}$; $\Delta\Delta C_T = 2.144$) and *chk1* ($\Delta\Delta C_T = -7.799$) were found to amplify in a non-uniform manner and were thus not included in the multiplexed pre-qRT-PCR amplification reaction. The gene-expression assay for *ccnd2* ($\Delta\Delta C_T = 1.556$) was borderline with a threshold value of $\Delta\Delta C_T = \pm 1.5$, so it was decided to include this in the assay pool. Amplification uniformity was not confirmed for skin RNA due to budgetary constraints. Uniformity was assumed to be similar for both pancreas- and skin-derived RNA.

Table 2.7.2: Confirmation of uniformity for multiplexed pre-amplification qRT-PCR assay

	Non-Preamplified RNA		Preamplified RNA		
Gene assay	**Average C_T**	**ΔC_T**	**Average C_T**	**ΔC_T**	**$\Delta\Delta C_T$**
18S rRNA	27.920	NA	21.019	NA	NA
akt1	34.892	6.972	27.493	6.474	0.498
atr	37.873	9.953	32.221	11.202	-1.249
ccna2	38.426	10.506	31.432	10.413	0.093
ccnb1	37.139	9.219	31.378	10.359	-1.140
ccnd1	35.773	7.853	28.366	7.347	0.506
ccnd2	35.661	7.741	27.204	6.185	1.556
ccne1	37.492	9.572	33.512	12.493	-2.921
ccne2	37.394	9.474	30.869	9.850	-0.376
cdc2a	38.294	10.374	30.414	9.395	0.979
cdk4	34.113	6.193	25.929	4.910	1.283
cdkn1a	35.153	7.233	27.236	6.217	1.016
cdkn1b	35.994	8.074	27.817	6.798	1.276
cdkn2a	40.000	12.080	34.217	13.198	-1.118
cdkn2b	38.083	10.163	29.038	8.019	2.144
cdkn2c	35.744	7.824	27.759	6.740	1.084
chk1	39.102	11.182	40.000	18.981	-7.799
chk2	35.736	7.816	28.676	7.657	0.159
cycs	35.697	7.777	28.122	7.103	0.674
endog	36.948	9.028	29.011	7.992	1.036
fas	35.858	7.938	30.174	9.155	-1.217
igf1	34.804	6.884	28.218	7.199	-0.315
igf1r	38.454	10.534	31.053	10.034	0.500

2.8 Immunohistochemical staining of tissue

2.8.1 Haematoxylin & Eosin staining

Frozen OCT-embedded tissue sections were cut to 10 μm using a Bright 5040 cryostat (Jencons Scientific, Bridgeville, PA) and sequentially washed in Haematoxylin solution (Surgipath, Richmond, IL) for 135 secs, tap water for 45 secs, acid alcohol solution (70 % ethanol, 0.5 % HCl) for 45 secs, tap water for 45 secs, Scott's solution (40 g $MgSO_4$, 7 g $NaHCO_3$ and Thymol crystals in 2 L tap water), 1 % aqueous Eosin solution (VWR International Ltd., Leicestershire, England) for 135 secs and tap water for 45 secs. Sections were dried by sequential washing in 5 ethanol solutions graded from 50 % to 100 % for 1 min each, and finally washed in two changes of xylene for 1 min each. Sections were mounted immediately in p-xylene-bis-pyridinium bromide (DPX) mounting medium (Agar Scientific Ltd., Essex, England) and viewed using an Axiostar Plus light microscope (Zeis, Oberkochen, Germany). Images were captured using a Powershot G5 digital camera (Canon, Tokyo, Japan).

2.8.2 Immunohistological staining

Tissue sections were stained for proliferating cells using antibodies for the cell cycle marker Ki67 (rabbit; Novocastra, UK), apoptotic cells using antibodies for the executioner Caspase 3 (rabbit; Cell Signalling Technology, Inc., Boston, MA), β-cells using antibodies for the β-cell-specific Insulin hormone (guinea pig; DAKO, Glostrup, Denmark), and suprabasal keratinocytes using antibodies for the suprabasal-specific Keratin 1 (rabbit; BabCo, Berkeley, CA). Fluorescently labelled secondary antibodies were chosen to match the species in which primary antibodies were raised (Fluorescein Isothiocyanate (FITC) anti-rabbit, Vektor Co.,

Germany; ALEXA 633 anti-guinea pig, Invitrogen, Carlsbad, CA; ALEXA 633 anti-rabbit, Invitrogen, Carlsbad, CA).

Frozen OCT-embedded tissue sections were cut to 10 μm with 3 sections per glass slide using a Bright 5040 cryostat (Jencons Scientific, Bridgeville, PA). Frozen sections were fixed with 4 % paraformaldehyde (PFA) at RT for 10 mins, washed in PBS for 5 mins, and incubated at RT in a humidifying temperature for 30 mins in 10 % BSA. Antibody staining was then performed as described below.

2.8.2.1 Pancreas tissue

Pancreas tissue sections were double stained for Ki67 and Insulin, or Caspase 3 and Insulin. Sections were incubated at 4 °C overnight in primary antibodies diluted in 1 % BSA (Insulin, 1:100; Ki67, 1:200; Caspase 3, 1:200). One section on each slide was treated as a negative control and was incubated with 1 % BSA only (no primary antibody). Sections were washed twice in PBS with 0.1 % tween (PBSt) for 5 mins each and incubated for 30 mins at RT in a humidifying chamber with secondary antibodies diluted in 1 % BSA (1:200). Finally, samples were washed in two changes of PBSt for 5 mins each. Slides were mounted in Vectashield (Vector Labs, Burlingame, CA) mounting medium containing 4',6-diamidino-2-phenylindole (DAPI) and viewed using a Leica Sp2 confocal microscope (Leica, Wetzlar, Germany).

2.8.2.2 Skin tissue

Skin tissue sections were sequentially stained for Keratin 1 and Ki67, or Keratin 1 and Caspase 3. Sections were incubated for 1 hour in Ki67 or Caspase 3 primary antibodies diluted in 1 % BSA (1:200). One section on each slide was treated as a negative control and was incubated with 1 % BSA only (no primary antibody). Sections were washed twice with PBSt for 5 mins each and incubated for 30 mins at RT in a humidifying chamber with FITC anti-rabbit secondary antibodies diluted in 1 % BSA (1:200). Finally, samples were washed in two changes of

PBSt for 5 mins each. This cycle was repeated using Keratin 1 primary antibodies diluted in 1 % BSA (1:100) and ALEXA633 anti-rabbit secondary antibodies diluted in 1 % BSA (1:200). Slides were mounted in Vectashield (Vector Labs, Burlingame, CA) mounting medium containing 4',6-diamidino-2-phenylindole (DAPI) and viewed using a Leica Sp2 confocal microscope (Leica, Wetzlar, Germany).

Chapter 3 *Envisage*: Significance Analysis of Microarray Data Using Linear Models

3.1 Introduction

Standard procedures for the analysis of microarray data, along with considerations for experimental design, were previously discussed in Section 1.5. Microarray analysis is often used to identify candidate genes showing significant differential expression between two or more biological conditions. There are two types of variation that may elicit a response in gene-expression that must be considered; variation fixed by the experimenter (typically known as explanatory variables or factors, but henceforth termed *parameters* to follow the conventions of GS-GX) and further sources of variation that may influence the relationship between the gene-expression and experiment parameters (which may be expressed as blocking factors or as continuous *covariates*).

Described in this chapter is an analysis package designed to allow researchers to identify the effects (if any) that such covariates may elicit on the expression of genes in a microarray experiment, allowing biological context to be taken into account. Whilst methods such as ANOVA, ANCOVA and regression analysis are available in many existing Bioconductor packages for the analysis of microarray data, *Envisage* provides a simple approach (particularly for non-statisticians) to the analysis of experiments where covariate effects are not well controlled. This method is particularly important in clinical experiments, where many such sources of variation between samples exist (e.g. differences in phenotype, environmental factors, and technical variation in sample processing). In this way, researchers can ensure that changes in gene-expression found in the analysis are a product of the effects of the parameter variable under consideration

3.2 Envisage

Envisage (Robson, S., Turner, H., Brown, H., Hunter, E., paper in preparation) was developed as a package in R (Ihaka and Gentleman, 1996), utilising the extensive bioinformatics packages of Bioconductor (Gentleman *et al.*, 2004), allowing significance analysis of gene-expression data across a broad range of experimental variables. These can include experimental parameters (e.g. drug treatment, disease state, time, etc.), phenotypic covariates (e.g. gender, age, weight, etc.), environmental covariates (e.g. temperature, humidity, light levels, etc.) and nuisance covariates (such as batch effects). Typically, such covariate terms are controlled by designing them into the experimental design through the use of block designs, ensuring that treatments are applied to blocks of samples with similar covariate effects. In this way, the effects of nuisance variables are made to be orthogonal to the effects of treatments, allowing these to be separated. The development of *Envisage* was largely influenced by the need to account for covariate terms that cannot easily be designed into the experiment structure. In particular, *Envisage* is well suited for the analysis of clinical studies, where between-sample variation is largely high due to differences in age, gender, weight, methods of sample processing, geographical location, etc., which cannot easily be designed into the experiment.

Such unwarranted variation must be accounted for to draw accurate biological conclusions from the analysis of gene-expression data. The most important aspect of the technique employed by *Envisage* is its ability to mine genes that show significant differential expression across the conditions of experimentally varied parameters, whilst also taking into account and correcting for effects on the gene-expression that are attributable to further sources of variation within the experiment. In this way, the biological context of the samples can be taken into account. By considering all variables in the analysis and not just those fixed by the experimenter, significant effects on gene-expression may be found for unexpected variables.

As well as offering a further level of insight into the biological context of the system of interest, this method also provides a way of ensuring that superfluous variables do not convey any undue significance to the observed changes in gene-expression. For instance, in Section 1.5.1.1, the need for randomisation of samples to avoid the introduction of systematic errors in sample processing was discussed. By recording some batch identifier number for each sample during sample processing, this information can be included as a categorical covariate in the linear model to allow observation of the effect (if any) that batching samples has on the overall data. For instance, one batch of samples may accidentally be treated with twice the volume of reagents, resulting in those samples appearing to be more correlated with one another than with samples from other batches.

This technique can thus be used in several ways:

1. As an exploratory tool to find experimental parameters and covariates that elicit a significant effect on the expression of genes
2. As a quality control tool to ensure that batch effects and other such nuisance covariates have no detrimental effect on gene-expression data
3. As a tool for the detection of significantly changing gene-expression across variables of interest

Linear models have been utilised previously in microarray data analysis algorithms, most notably within the *limma* Bioconductor package (Smyth, 2004; Smyth, 2005). *Envisage* differs from this method by focusing on identifying and correcting for unwanted variation in the data set, which may have been introduced due to the experimental design. Matsui *et al.* (2007) describe a similar multi-variable linear modelling algorithm, designed to allow the inclusion of between-sample phenotypic characteristics as factors into the analysis for clinical studies. However, this method requires variables to be categorical factors, which fails to consider many sources of variation within clinical studies that may be numerical in nature. *Envisage* allows users to include both numerical and factorial explanatory variables into the model, ensuring that variables relating to environment, phenotype, and technical aspects of the experiment are included in the analysis. Superfluous sources of error can thus be detected and corrected for,

allowing users to draw their conclusions within the context of the biological system under analysis.

3.3 The *Envisage* package

3.3.1 Introduction

Envisage is a package written in R version 2.6.2 (Robson *et al.*, manuscript in preparation). It is available from the author's website [2] and is soon to be submitted to version 2.2 of Bioconductor. A Tcl/Tk front-end GUI interface, produced using the widgets available through the *tcltk* Bioconductor package in R (Dalgaard, 2001), was implemented to ensure ease of use, particularly for non-statisticians (Figure 3.3.1). This allows users to specify which of the variables defined within their data set are required for analysis, and what type of data they contain (categorical or numerical). Further arguments, such as the p-value cutoff for significance (default $p = 0.05$), the MTC to be used (default = Benjamini & Hochberg; Section 1.5.3.4) and whether or not to include interaction terms in the analysis, can also be specified. Currently *Envisage* supports the analysis of only first-order interaction terms, although it is hoped that higher order interaction terms will be supported in the future. *Envisage* also allows users to specify a minimal model to be used when fitting a candidate model to the genes. This allows the specification of terms that the user may consider to be of practical importance – for instance, one may wish to ensure that the term for the drug treatment is always included, since that is the most interesting of the variables under consideration. In particular, for designed experiments this allows the inclusion into the model selection process of treatment and block terms that were included in the experimental design. Arguments can also be defined by using the command line interface of R, although care must be taken to ensure that variable information (particularly class definition) is expressed correctly.

[2] http://www2.warwick.ac.uk/fac/sci/moac/currentstudents/2003/sam_robson/linear_models/downloads/

Envisage was developed in close collaboration with Agilent Technologies, and was designed to integrate with the GS-GX gene-expression analysis suite. *Envisage* is best suited for 1-colour microarray data, although can also be used to analyse 2-colour microarray data with a reference-based design. While the present discussion focuses primarily on gene-expression analysis, this technique can be equally well applied to other high-throughput sources of biological data (such as mass spectroscopic results and data from methylation-based analyses and comparative genomic hybridisation studies).

Input data for gene-expression analyses consist of normalised, curated, log-transformed gene-expression data from either 1-colour (log-intensity values) or 2-colour (log-ratio values) microarray experiments. Such data are stored, together with information on experimental variables and other important experiment data (e.g. MIAME-compliant information; Section 1.5.1.6), in the form of an object of class *ExpressionSet* in R. This object can be created by the user directly by using the packages available from Bioconductor, or can be created automatically from the user's experiment interpretation in the gene-expression analysis suite GS-GX by using the *GeneSpring* Bioconductor package and external program interface (de Boer, 2007).

Results are output as tab delimited text files containing gene lists of significantly changing genes for each variable and interaction in the analysis. Output lists can be annotated and analysed further using a variety of Bioconductor packages, or can be imported into any suitable software suite, such as GS-GX. The modelling procedure utilised by *Envisage* is based on the linear model described in Section 1.5.3.3, and is outlined in the following section.

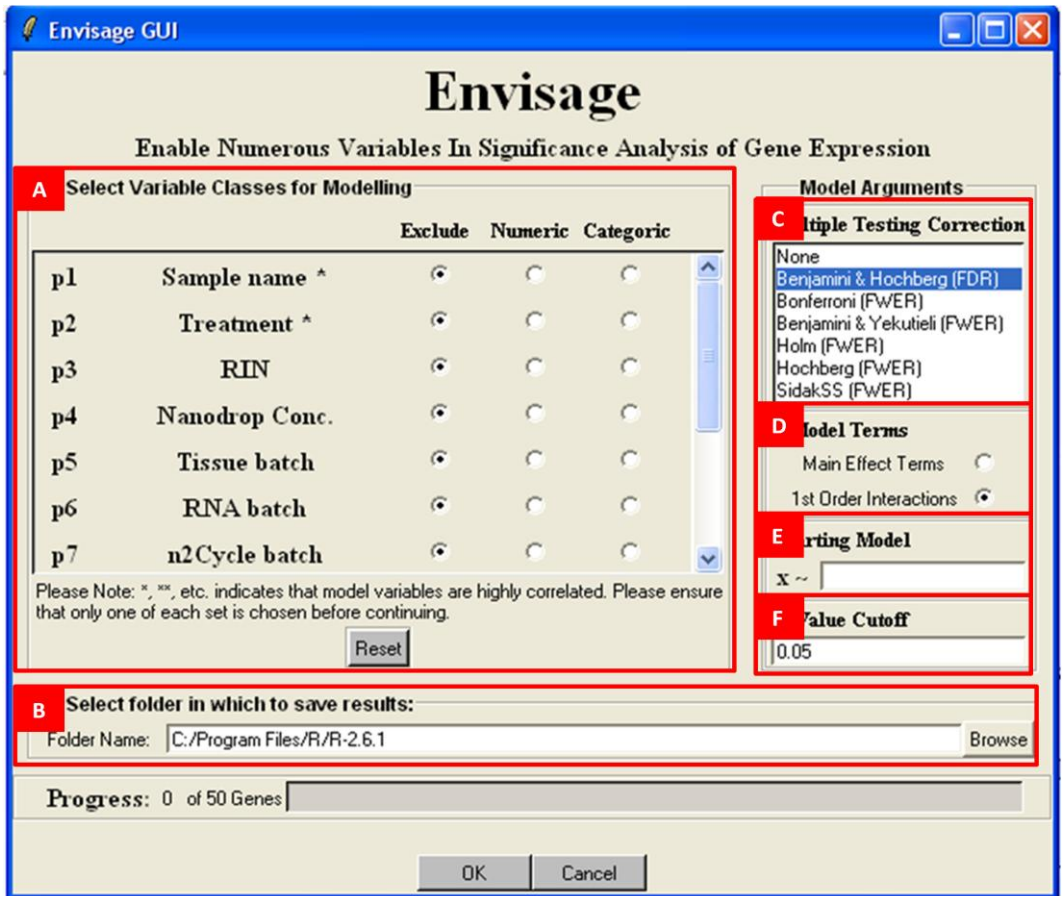

Figure 3.3.1: The *Envisage* graphical user interface.

Arguments for the *Envisage* package can be specified through a simple to use graphical user interface to avoid errors in data entry. This widget allows users to specify; A) the variables to use in the modelling procedure and their respective classes (categorical or numerical), B) the file location for the output gene lists to be saved to, C) the multiple testing correction to use, D) whether to include interactions or just model main effect terms, E) a minimum model to force terms into the overall model, F) the threshold value for the significance p-value.

3.3.2 Model selection

Linear models and their use in gene-expression analysis are discussed in Section 1.5.3.3. For analysis using *Envisage*, a candidate linear model is selected for each gene by first considering the saturated model, defined in Equation 1-14, which contains all available variable terms and their first order interactions as model terms. A candidate model is then selected by removing terms that are not found to provide a notable increase in the explanatory power of the model. An automated stepwise process is used to compare the explanatory power of the model with and without a particular term – if the term confers no additional explanatory power, it is deemed unnecessary and is removed. This is performed by sequential use of the functions *add1()* and *drop1()*, which compute the model terms that can be added to or subtracted from the model (ensuring that model hierarchy is respected such that interactions are only included for main effect terms already included in the model, and main effect terms contained within an interaction term are not removed). Removed terms are then considered for addition back into the model. This process is iterated until a candidate model containing only those terms of practical importance is found.

The criterion used for keeping or removing a term from the model is the Akaike information criterion (AIC) (Akaike, 1974). In the case of least squares estimation with normally distributed errors, the AIC can be described as:

$$AIC = n \ln\left(\frac{RSS}{n}\right) + 2k \qquad \textbf{3-1}$$

where k represents the number of variables in the fitted model, n is the number of observations and RSS is the residual sum of squares for the model as previously described. The first term in this equation gives an estimate of the goodness of fit of the model, with a lower value indicating a model with improved explanatory power. As additional terms are added to the model, the additional information available will inevitably improve the fit to some extent. However, if the

improvement in fit brought about by the addition of a term is only small, we may prefer to leave the term out for the sake of parsimony. Thus the second term acts to penalise models with higher numbers of free variables. A model with a lower AIC is preferred to a model with a higher AIC, and any change to the model must reduce the deviance by at least 2 in order for the AIC to decrease. This limits the number of terms in the model, and ensures that only terms of practical importance are considered for addition. The AIC was chosen since it is a well established criterion for model selection, and allows automation of the model selection procedure, which is essential given the large number of genes considered.

3.3.3 Significance analysis

For each gene, a candidate model is fitted to the data containing only those terms found to confer some improvement to the model fit. This model selection procedure was chosen, as opposed to fitting all model terms to each gene in a saturated model (as with ANOVA), to account for per-gene variability in the effects of the treatment and covariate terms (Jin *et al.*, 2001; Wolfinger *et al.*, 2001; Smyth, 2004).

Fitted model terms may be main effects or interaction terms up to the first order (the inclusion of higher order interaction terms may be included at a later date to account for joint effects for multiple experimental variables). To analyse the statistical significance of the role that each of these terms plays on the expression of the gene of interest, an F-test statistic (Equation 1-18) is calculated to compare the fit of the selected model to the fit of the model with the term of interest removed. This is repeated for all model terms in the selected model for each gene to calculate a series of significance p-values, which can be used to identify terms that play a significant role in the final model. Since the model-selection process may result in only a subset of the overall model terms being selected for inclusion in the model for each gene, some genes will have no associated p-value for these

excluded terms, and the matrix of p-values will contain a number of NA values. This can make it more difficult to screen or order the genes against the treatment effects.

These values are compared to some user defined threshold value to discern variables that elicit a significant response in the expression of the gene. By performing this analysis over all genes, lists of genes whose expression is significantly affected by each of the model terms is produced. These lists can be used to identify candidate genes whose expression varies in response to particular experiment variables.

Envisage utilises a Type II SS in its model fitting procedure, which is more powerful when looking at unbalanced experiment designs since it is not reliant on constraints on the parameters (Langsrud, 2003). Use of a Type II SS prevents significance analyses from being dependent on the order of the terms within the model, which allows for automation of the process. The SS calculation is implemented through the package *car* in R (Fox, 2002).

3.3.4 Multiple testing correction

The problem of multiple hypothesis testing is described in Section 1.5.3.4. FWER and FDR multiple testing corrections are implemented in *Envisage* through the *multtest* package in R (Pollard *et al.*). However, since the matrix of p-values contains missing values due to the model fitting procedure, there may be some bias in the multiple testing procedures due to the variable number of non-missing p-values for each model term. This means that p-values for model terms found to have a significant effect for only a small number of genes may be overly down-weighted by the applied MTC.

3.3.5 Model aliasing

The structure of the gene-expression data set used to fit the model may cause adverse effects on the modelling procedure. If the experiment is not designed carefully, aliasing may occur between variables such that they become indistinguishable. Aliasing may occur for one of two reasons:

1. *Extrinsic* aliasing is due to the structure of the data such that there is no relevant data for a particular interaction group (e.g. there are no diseased patients in a particular age group, so the effects for the interaction diseased:age cannot be estimated).
2. *Intrinsic* aliasing is due to the relationship between model terms such that particular effects cannot be separated adequately (e.g. all diseased individuals in a particular age group also have the same blood group. In this case we cannot separate the effects of the interactions diseased:age and diseased:blood).

Extrinsic aliasing will generally occur due to the use of an unbalanced data set, but will not cause a problem with the significance calculations. Intrinsic aliasing is indicative of correlation between explanatory variables, so the model will fit one variable and then try to fit the same information again with the second variable. This may then pose a problem since the significance results will depend on the order in which the terms are fitted.

If the data used for model fitting is not adequate to fit the required effects, aliasing will be detected and output is provided to the user in the form of a list of aliased terms for affected genes. This allows the user to identify variables that may be showing high correlation. This may require that the user reformulate the data set to remove inter-dependencies between the variables, or may require that fewer terms are included in the model fitting procedure.

3.4 Results

Envisage was run on curated gene lists, as described in Materials and Methods, Section 2.6.2. Samples were normalised using GC-RMA (Section 1.4.2.4) and experiment interpretations were set up as described in Materials and Methods, Section 2.5, using GS-GX. *Envisage* was run using the GS-GX/R external program interface to transfer experiment interpretations into R 2.6.2 in the form of an object of class ExpressionSet. Analyses were first performed using main experiment parameters only for comparison with standard ANOVA analyses (Section 3.4.1), followed by a complete analysis considering experiment parameters together with further covariates. This allowed observation of the effects (if any) of variables such as RNA quality and batching on the resulting change in gene-expression (Section 3.4.2). All analyses were corrected for multiple testing using the Benjamini and Hochberg FDR correction (Benjamini and Hochberg, 1995), and significance was determined using a p-value threshold of $p = 0.05$. All analyses were performed using a Toshiba Satellite S2450-201 laptop under Windows XP (Microsoft, Redmond, WA) with a Pentium 4 (Intel, Santa Clara, CA) 2.40 GHz CPU and 2 GB RAM.

3.4.1 Comparison with ANOVA

As described in Section 1.5.3.2, ANOVA is often used to identify genes whose expression shows significant change based on variables of interest using a Type I SS F-test. Whilst ANOVA is by no means the best method for determining significance of gene-expression, it is a well established method implemented within GS-GX that provides a simple comparison for establishing the validity of the *Envisage* modelling procedure. In particular, ANOVA is well suited for the analysis of balanced factorial designed experiments where the effects of treatments are independent and orthogonal.

As a first test for *Envisage*, results were compared to those seen using standard ANOVA Type I SS. As described in Materials and Methods, Section 2.5, the experiment design was a balanced factorial design across the 3 main parameters; 4OHT-treatment ('4OHT'), tissue type ('Tissue'), and time following MycER^{TAM} activation ('Time'). A 3-way ANOVA model, consisting of the 3 main effect variable terms and their interactions, was fitted to the data and an F-test statistic, utilising a Type I SS, was used to identify significantly changing genes for each term. These results were compared to a similar analysis, fitting the main effect and interaction terms using the *Envisage* modelling procedure. Note that for this comparison, the *Envisage* package was modified specifically to allow inclusion of the higher order 3-way interaction term, 4OHT:Tissue:Time. However, as described in Section 4.2.5, quality control analysis of the microarray data identified seven samples with poor hybridisation which were removed from the final data set. For this reason, the data set considered for this comparison was not balanced, and hence we would expect the choice of the SS method (Type I for ANOVA, Type II for *Envisage*) to have an effect on the resulting F-ratio estimates for the model terms, resulting in potentially different results between the two analyses.

The sizes of the gene lists resulting from the *Envisage* analysis for each model term (main effect and interaction terms) are shown in Table 3.4.1. Also shown are the comparative results for the 3-way ANOVA analysis on the three main parameters and their interactions. It was clear from this comparison that, whilst not identical, there was a high level of concordance of the results of the *Envisage* analysis with those from ANOVA, with almost all genes deemed to be significant at a given confidence level by *Envisage* also deemed significant by ANOVA (> 90 % at $p = 0.05$).

Table 3.4.1: Comparison of significance analysis using *Envisage* and ANOVA.

	No. Genes ANOVA	**No. Genes *Envisage***	**Intersect of *Envisage* and ANOVA**
Main Effect Terms			
4OHT Treatment	5346	5296	5111
Tissue	3850	3851	3458
Time	4830	4883	4664
Interaction Terms			
4OHT:Tissue	4013	4088	3736
4OHT:Time	5063	4984	4809
Tissue:Time	5666	5666	5666
4OHT:Tissue:Time	5788	5788	5788

In general, the discrepancy between the two methods resulted in genes that fell just within the p-value threshold for one method, but which were assigned a p-value just above the threshold for the other. Genes found significant by *Envisage* but not by ANOVA showed average p-values of $p = 0.07 \pm 0.03$ for the ANOVA analysis, whilst genes found significant by ANOVA but not by *Envisage* showed average p-values of $p = 0.08 \pm 0.03$ for the *Envisage* analysis. This indicated that many of these outlying genes showed p-values very close to the threshold value of $p = 0.05$.

It was also possible that the process of selecting an individual model for each gene may have resulted in removal of terms from the model at the selection stage that may have, nevertheless, been statistically significant. The model fitting procedure results in the selection of *practically* significant terms, and significant terms that offered no additional benefit to the selected model given the terms already included may not have been included. On the other hand, ANOVA included all terms in a saturated model for all genes. *Envisage* allows users to specify a minimal model for analysis, allowing terms to be 'forced' into the selected model. It was therefore decided to observe the effects of specifying a saturated model as the minimal model, forcing the saturated model to be fitted for all genes, making the procedure more comparable with ANOVA.

The results of this analysis are shown in Table 3.4.2. It appeared that fitting a saturated model to all genes using the *Envisage* modelling procedure did little to improve comparability with the ANOVA results, and in fact in some instances the number of significant genes was actually reduced. This was likely due to the fact that the presence of the additional terms of the saturated model, as compared to the model selected by the stepwise regression, reduced the significance of those terms added in the model fitting procedure. For many genes, the inclusion of additional terms in the fitted model may have affected the calculated SS, resulting in slightly modified p-values for each term. In fact, genes that differed between *Envisage* when a saturated model was fitted, and *Envisage* when no model was specified, showed an average p-value of $p = 0.06 \pm 0.01$, indicating that these genes showed p-values very close to the threshold value of 0.05. For these genes,

fitting a saturated model resulted in an increase in the p-value for some terms above the threshold value, resulting in it being excluded from the resulting gene list.

Table 3.4.2: Comparison of significance analysis using ANOVA and using *Envisage* with a saturated minimal model.

	No. Genes ANOVA	No. Genes *Envisage*	Intersect of *Envisage* and ANOVA
Main Effect Terms			
4OHT Treatment	5346	5249	5089
Tissue	3850	3804	3436
Time	4830	4811	4652
Interaction Terms			
4OHT:Tissue	4013	4040	3721
4OHT:Time	5063	4897	4784
Tissue:Time	5666	5666	5666
4OHT:Tissue:Time	5788	5788	5788

Figure 3.4.1 shows a comparison of the p-value assignments for each gene between ANOVA and *Envisage* for each of the 3-way model terms. To ensure that p-value estimates were obtained for all genes and for all terms to allow direct comparison, ANOVA results were compared to those of fitting *Envisage* with a saturated model as described above. In general, p-value assignments were comparable between the two procedures, as can be seen by the close approximation to the identity line and the high Pearson correlation coefficient between the two ($r > 0.863$ for all terms).

It is interesting to note the close similarity between the two methods for all terms including the 'Time' variable, in particular for the interaction terms '4OHT:Tissue:Time' and 'Tissue:Time'. However, it is worth noting that treatment of Time as a categorical variable may not be an accurate representation of the data set as it fails to take into account the dependent nature of the variable.

Points on the graphs are coloured dependent on the p-value for each method (see figure caption). This allows the discrimination of genes whose significance is determined differentially between the two procedures. It appears from these images that the number of genes shown to be significant using ANOVA but not with *Envisage* (blue) is larger than those shown to be significant using *Envisage* but not with ANOVA (red). It is unclear as to the precise reason behind this, but it may indicate a higher number of false positive calls for ANOVA than for *Envisage* due to the use of a saturated model for all genes.

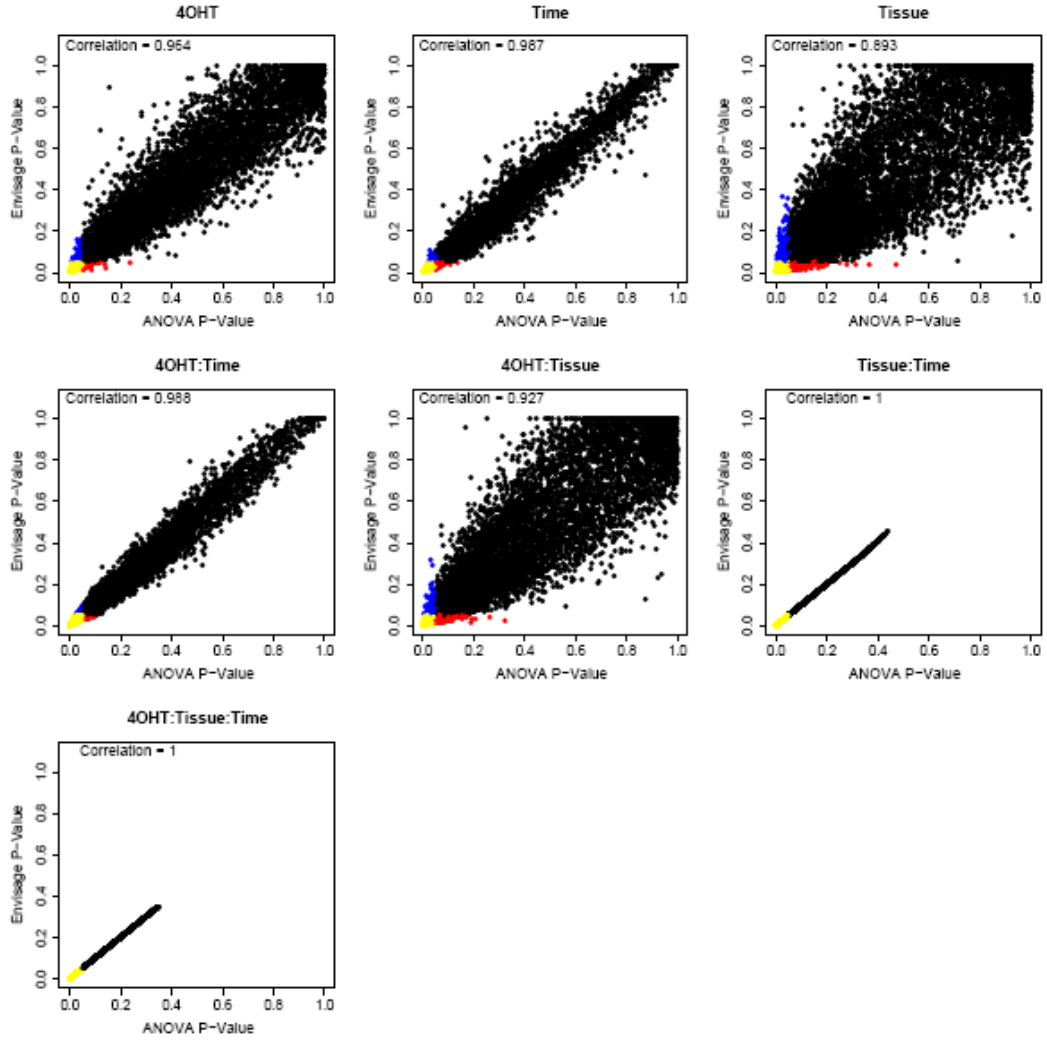

Figure 3.4.1: Comparison of p-values for significance analysis of gene-expression using ANOVA and *Envisage*

A 3-way analysis was performed across the 3 main experimental parameters of 4OHT treatment ('4OHT'), tissue type ('Tissue'), and time following MycER^{TAM} activation ('Time'), using ANOVA and *Envisage* with a saturated model specified. In general, the concordance between the two procedures was high, with Pearson correlations > 0.8 for all main effect and interaction terms. Yellow points indicate genes showing a p-value < 0.05 for both *Envisage* and ANOVA, blue points indicate genes showing a p-value < 0.05 for ANOVA but not for *Envisage*, red points indicate genes showing a p-value < 0.05 for *Envisage* but not for ANOVA, and black points indicate genes showing a p-value > 0.05 for both ANOVA and *Envisage*.

3.4.2 Analysis with inclusion of covariates

The previous section shows that the modelling procedure used by *Envisage* produces results comparable to those using a simple ANOVA-based Type I SS analysis. However, the inclusion of covariate information into the analysis, such as by using an ANCOVA model structure, is necessary to encompass additional sources of variation which may illicit some effect on the main variables of the study. This is particularly true of poorly-designed experiments and clinical studies where confounding variation is difficult to avoid. One of the key features of *Envisage* is the ability to include numerical variables into the modelling procedure using a step-wise approach, allowing analysis of experiments were such variables have not been designed into the structure of the experimental design. By including all observed sources of variation into the test for significance, a more accurate model of the biology of the system can be estimated.

Much of the variation typically seen in microarray experiments was minimised in these studies, as described in Materials and Methods, Section 2.5. One of the main points is that, since gene-expression studies were carried out on age- and gender-controlled inbred transgenic mice housed under environmentally controlled conditions, animal-to-animal genetic variation was greatly reduced. Thus physiological and environmental variables were generally highly comparable across samples. However, there were two main sources of variation in sample processing that should be accounted for during analyses; the quality of the starting RNA material, and batching effects.

As described in Sections 4.2.2.3 and 4.2.3.2, RNA integrity and resulting yield of 2a-cRNA (respectively) were variable across the experiment. RNA integrity is a measure of the level of degradation of RNA species prior to processing which, as described in Section 1.5.1.5, can have a detrimental effect on resulting data. Also, whilst the concentration of 2a-cRNA was adjusted to a constant level across all samples prior to hybridisation, the yield of 2a-cRNA after the IVT reaction provides a good measure of initial RNA quality, since low yields may be

indicative of highly degraded or contaminated RNA samples preventing correct amplification.

The two variables, '2a-cRNA' and 'RIN' were used as additional covariates in the *Envisage* analysis to account for RNA quality within the analysis. Both measures are numerical in nature. Also, due to the large sample size, samples were separated into 7 randomised batches for sample processing. To ensure that no batch effects existed within the data, the batch identifier number (arbitrarily labelled 1-7) was used as a further categorical variable ('Batch number') in the *Envisage* analysis. Due to the number of DF required to fit the interactions of this 7-level factor compared to the number of free DF (dependent on sample size), batch interactions were not included in the model fitting procedure to avoid overfitting of the data. Given the extensive measures taken to ensure that covariates elicit minimal effects on the gene-expression, it was hoped that they would show only minimal effects.

The number of genes found to be significant for the experiment variables and their interactions are shown in Table 3.4.3. It is clear from these figures that more genes were found to be significantly affected by parameter terms than by covariate terms, indicating that the parameter terms were responsible for the majority of the variation in the gene-expression data (although there may still be genes for which the covariate effects were more significant than the effects of the design factors). This indicated that the effects of the covariates were minimised by the experimental design such that, generally speaking, changes in gene-expression were attributable to the variables of interest and their interactions. This was good news for the analysis, and suggested that the experimental design and implementation was good. Also, by including these terms in the analysis using *Envisage*, the effects of the covariates (small though they may be) were corrected for, improving the relationship of the results to the biology of the system.

Of the covariate main effects, the RNA quality measures ('RIN' and '2a-cRNA yield') showed significant effects for the smallest number of genes, with only 291 (2.36 %) and 9 (0.07 %) of the tested genes showing a significant response respectively. This was an interesting result as it indicated that the quality of the

starting RNA may have had a negligible effect on gene-expression. Given the wide range of RNA quality within this experiment (Sections 4.2.2.3 and 4.2.3.2), this was an important result as it indicated that RNA quality may not be an adequate predictor of ultimate microarray data quality. This point was also discussed in more detail in Section 4.2.3.2.

Covariate interactions with the experimental factors were also fitted in this model up to the first order only (due to the limitations of the *Envisage* package). Within the ANCOVA framework, covariate terms are assumed to show a linear relationship to the dependent variable, and the slopes of the regression lines of the covariates are assumed to be homogeneous between treatment factor levels. Linearity of the covariates '2a-cRNA' and 'RIN' was varied, with the median p-value of a fitted linear regression, $Y_g \sim X_g$ (where Y_g is the log-normalised expression and X_g is the value of the covariate term across all samples $i \in 1, \ldots, 41$ for gene $g \in 1, \ldots, 12{,}349$), equal to 0.23 ± 0.298 for 'RIN' and 0.38 ± 0.304 for 'cRNA'. Also, Table 3.4.3 shows that the number of genes showing significant effects for the covariate interaction terms relating to 'Tissue' and '4OHT' was low, suggesting that the assumption of homogeneity holds for these treatment factors. On the other hand, the covariate interaction terms in the model relating to 'Time' ('2a-cRNA:Time and RIN:Time) appeared to show a significant effect for a larger number of genes, indicating that there may be some confounding effect between RNA quality (particularly 2a-cRNA yield) and time. However, the number of genes showing such interactions was typically small in comparison to the number of genes showing significant effects for the main experiment variables (877 (7.1 %) genes for 2a-cRNA:Time and 196 (1.6 %) genes for RIN:Time).

This confounding effect of the covariate terms with the treatment terms can be seen in Figure 3.4.2, and was further suggested by the change in the number of significantly changing genes for the main treatment variables following inclusion of covariate terms into the model fitting process. In particular, inclusion of covariate terms in the model resulted in an increase in the number of significant genes for treatment main effect terms, but a decrease in the number of significant genes for treatment interaction terms. These confounding effects were particularly

noticeable for ‘Tissue’ and ‘Time’ variables, as examplified by the ~4-fold reduction in the number of genes classed as significant for the interaction between the two (‘Tissue:Time’).

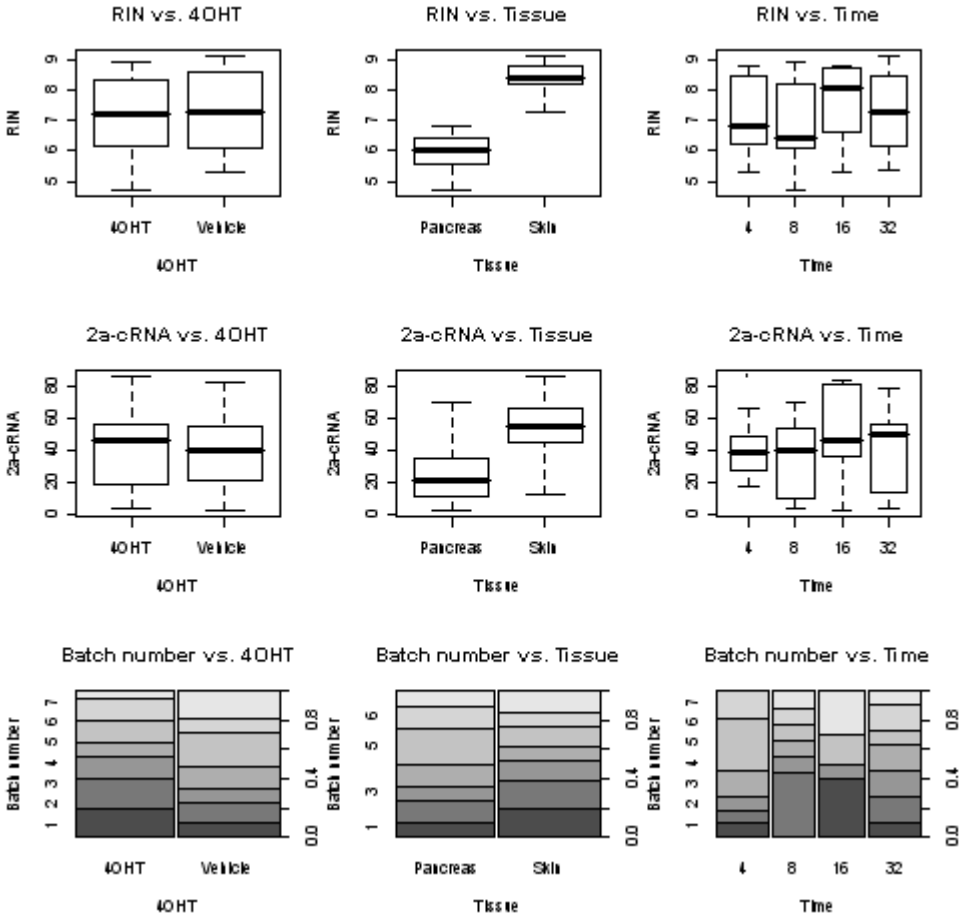

Figure 3.4.2: Relation of covariate terms to treatment terms

To check for confounding effects of the covariate terms with the treatment effects, covariate values were plotted against the corresponding treatment factors for each sample. To ensure efficiency of covariates, covariate terms should be homogeneous between treatment factor levels. A clear confounding effect was detected between RNA quality covariates (2a-cRNA and RIN) and the Tissue treatment factor. This effect likely relates to the clear difference in the quality of RNA isolated from the pancreas as compared to the skin due to the presence of RNases (Section 4.2.2.3 and Section 4.2.3.2). A smaller confounding effect was also detected with the Time treatment factor, but no such confoundment was detected for the 4OHT factor. Also, non-homogeneity of batch number between time points suggests that there may also be some confoundment between the two variables, suggesting the introduction of some batche effect during processing.

The number of genes showing significant effects relating to the batching of samples was greater than the number detected relating to the RNA quality, with 626 (5.07 %) of the genes showing significant change in expression. Interaction terms were not fit for the batch variable due to the large number of batch classes that would need to be assessed. Given the randomised approach to the selection of samples for particular treatment schemes, interaction effects between batch and treatment variables were reduced. However, a preferable approach to experimental design would have been to design batches for processing into the experiment structure in a block design, such that each batch contained an equal number of individuals with roughly similar covariate effects for each treatment condition. In this way, covariate and batch effects would be orthogonal to the treatment effects, reducing confounding effects between covariates and treatment variables. It is worth noting that if such a design were implemented, the need for inclusion of covariate interactions in the model fitting would be reduced. However, due to limits on the number of available mice for this experiment, together with the fact that these samples were but a subset of a larger experiment (as previously described), such blocking was not performed.

These results suggest that there are a small number of genes in the study for which the considered covariates showed a significant effect. This could be seen by observing the difference in the resulting gene lists between using models containing parameter and covariate terms, and those containing only terms of the main parameters (Table 3.4.3). One important change of note was the relatively small number of genes showing significant effects based on the 'Tissue' variable between the model with covariate terms included, and the main effects-only model. This further indicated a counfounding effect between the covariate terms and the 'Tissue' term. Given the difference in RNA quality seen between the two tissues (Section 4.2.2.3 and Section 4.2.3.2), this suggests that these variables are perhaps not well suited for inclusion in the model, as it may be difficult to distinguish 'Tissue' effects from covariate effects.

These data suggest that inclusion of these covariate terms may lead to difficult to analyse effects, where it is difficult to discern significant changes over time from

nuisance effects relating to RNA quality. In particular, there seems to be some confounding effect between the covariate terms and the 'Tissue' and 'Time' treatment variables. These effects may relate to the difference in RNA quality detected between the two tissues, and the cumulative effects of treatment over time (e.g. stress responses) which may be more evident for mice treated for longer time points in comparison to those culled after only a short time of treatment. However, on a more positive note, these data also suggested that the majority of genes appeared to show changes in expression due primarily to the more interesting variables of the experiment. Despite the potential confounding effects, these covariate terms were nevertheless included in the model fitting procedure.

Table 3.4.3: Significantly changing genes found by running *Envisage* using parameters and covariates as model terms

	No. Of Genes for Multi-variable *Envisage*	**No. Of Genes for 3-variable *Envisage***	**Intersect of multi-variable and 3-variable *Envisage***
Parameter Main Effect Terms			
4OHT treatment	5841	5296	4524
Tissue	4387	3851	2076
Time	6012	4883	3867
Covariate Main Effect Terms			
Batch number	626	NA	NA
2a-cRNA yield	9	NA	NA
RIN	291	NA	NA
Parameter Interaction terms			
4OHT:Tissue	2939	4088	1335
4OHT:Time	4941	4984	3160
Tissue:Time	1449	5666	957
Covariate Interaction Terms			
2a-cRNA:4OHT	86	NA	NA
2a-cRNA:Tissue	24	NA	NA
2a-cRNA:Time	877	NA	NA
RIN:4OHT	6	NA	NA
RIN:Tissue	13	NA	NA
RIN:Time	196	NA	NA
RIN:2a-cRNA	154	NA	NA

3.4.3 Comparison of 'Time' variable as numerical and categorical

In Section 3.4.1, the results of significance analysis of gene-expression based on the 'Time' variable were found to be highly similar between ANOVA and *Envisage*. However, given the non-independent nature of this temporal variable, it was not clear how best to treat it. Whilst each time-point replicate corresponded to RNA derived from tissue extracted from a unique animal (and hence time points were independent, as opposed to if RNA had been extracted repeatedly from a single animal), the ordered nature of the variable implies relatedness between signals at subsequent time points. To observe the effects of treating 'Time' in an ordered way, *Envisage* was rerun with 'Time' classified as numerical. The number of genes found to be statistically significant ($p < 0.05$) for variables in the model fitting procedure, along with comparisons to gene lists produced treating 'Time' as a factor, are shown in Table 3.4.4.

The most striking thing to notice in this comparison is the large reduction in genes found to be significant for model terms related to 'Time', with only 1,241 genes (10 %) found to change significantly for the main effect 'Time' term as compared to 6,012 genes (49 %) previously. However, this is most likely a result of the difference in model assumptions between the two methods. Treating 'Time' as a numerical variable in this model assumes a linear response with 1 DF (straight line), whilst treating 'Time' as a categorical variable allows for differences between the four time points in a 3 DF model (curve). It is therefore likely that the loss in the number of genes detected as significant under the assumption of a linear relationship of 'Time' relates to the loss of genes whose expression, whilst not changing significantly across the time course as a whole, may change significantly between individual time points. However, another striking feature is the dramatic increase in the number of genes related to covariates 'Batch number' (3,447 genes (28 %) as compared to 626 genes (5 %)) and '2a-cRNA' (1,044 genes (8 %) as compared to 9 genes (0.07 %)). However, whilst the gene lists for 'Time' and the covariates changes, the gene lists of interest ('4OHT' and '4OHT:Tissue') remain relatively invariant. These data suggest that there may be

some confounding effect between the time-point to time-point effects (detected in the "categorical" model but not in the "numerical" model) and these covariates. In particular, the batch covariate appears to show a high level of counfoundment with 'Time', suggesting that the batching approach used here introduced some systematic bias between time points for some genes. The fact that there is no such change detected for the '4OHT' and '4OHT:Tissue' terms suggests that these variables are orthogonal to time effects.

In the study of Fischer *et al.* (2007), a variety of methods for differential expression were applied to a time-series experiment to identify the best test able to identify significant changes in expression across a temporal dimension. It was found that the use of ANOVA (in particular, employing a type II SS test) performed the best out of the tests considered, particularly on background-corrected data. Also, Park *et al. (2003)* developed a statistical test for time-series gene-expression data based upon the ANOVA model, although this model was able to account for the ordered nature of the variable. This suggests that treatment of time as a factor in a model-based statistical test for differential expression is statistically sound, particularly when using *Envisage* where a type II SS is used to calculate significance.

Given the larger number of genes found to be significant due to the less constrained model assumptions when treating 'Time' as a factor, and given the apparent orthogonality of the main variables '4OHT' and '4OHT:Tissue', it was decided to proceed with the analysis treating 'Time' in a factorial way. One possible alternative approach, which was not examined further, would be to include Time as an explanatory variate with a non-linear relationship to the level of gene-expression to account for between-time effects not considered with a linear-relationship for time. The decision to treat time as a factor was made to maximise the number of genes of primary interest ('4OHT'- and '4OHT:Tissue'-related).

Table 3.4.4: Significantly changing genes found by running *Envisage* with 'Time' as a numerical variable compared to running *Envisage* with 'Time' as a categorical factor.

	'Time' as Numeric	'Time' as Factor	Intersect
Parameter Main Effect Terms			
4OHT treatment	4041	5841	3692
Tissue	1479	4387	1347
Time	1241	6012	1096
Covariate Main Effect Terms			
Batch number	3447	626	407
2a-cRNA yield	1044	9	3
RIN	57	291	20
Parameter Interaction terms			
4OHT:Tissue	3112	2939	1437
4OHT:Time	585	4941	394
Tissue:Time	817	1449	193
Covariate Interaction Terms			
2a-cRNA:4OHT	481	86	7
2a-cRNA:Tissue	539	24	5
2a-cRNA:Time	118	877	14
RIN:4OHT	118	6	0
RIN:Tissue	54	13	2
RIN:Time	0	196	0
RIN:2a-cRNA	2169	154	57

3.5 Summary

In this chapter, the new analysis package *Envisage*, designed and written by the author, was described. This procedure allows the inclusion of data relating to a wide range of experimental variation into analysis for significant changes of gene-expression. Whilst an experiment may be focused on identifying key genes whose expression varies due to some condition altered by the experimenter, it is naive to believe that this will be the only source of variation within an experimental setting. Between-sample differences in phenotype, environmental factors and technical processing add noise to expression data that must be accounted for to ensure that any conclusions drawn are representative of the biological system under study.

Standard experimental design methods account for such covariation by applying treatments within blocks of samples with similar covariate effects, ensuring that the effects of covariation can be separated from treatment affects during analysis. However, it is not always possible to design nuisance variation into the experiment structure, making it difficult to discern real variation relating to variables of interest from variation due to nuisance effects. This is particularly true for clinical studies, where such between-sample variation cannot be easily controlled, and sample preparation is often performed at different sites, using different methods, by different people. The utility of *Envisage* lies predominately with such experiments, where it is not practical to design such variation into the experiment structure. In particular, the author envisages that this package will provide a useful tool in the analysis of clinical microarray data, particularly in the identification of spurious sources of gene-expression change.

Given the weak transcriptional activity exhibited by Myc (Grandori and Eisenman, 1997; Cole and McMahon, 1999), it is particularly important to remove the noise caused by superfluous sources of variation to ensure that significant changes in gene-expression as a result of $MycER^{TAM}$ activation can be identified. Many sources of error were prevented through careful experimental design. In particular, environmental variables were controlled by consistent laboratory

conditions in the animal house, variation in sample processing was controlled by ensuring common experimental procedures (such as ensuring consistent start times for all treatments to remove effects of circadian rhythms), and some phenotypic variables were controlled by the selection of age- and gender-matched animals. The main source of unaccounted nuisance variation between samples was genotypic difference relating to the use of mice from different litters (although the use of inbred transgenic lines ensured that these effects were minimised). However, due to limits on animal numbers for this study, a batched design was not used to ensure independence of treatment and nuisance variables. Instead, randomisation was performed over all samples to remove systematic bias between treatment groups.

However, inclusion of the batching variable as a covariate in the *Envisage* analysis, together with further covariates relating to RNA quality obtained from each sample prior to hybridisation, identified flaws with this design. Analysis of the effects of treating the 'Time' treatment variable as a factor, compared to as a linear variate, identified a clear confounding effect between 'Time' and the covariates. In particular, the batch covariate appeared to show a significant relationship with 'Time', suggesting that the batching approach utilised introduced some systematic bias between time points, and may not be suitable for inclusion into the model. This time-effect appeared to be orthogonal to the treatment variables '4OHT' and '4OHT:Tissue', which were of interest to the analysis. Also, inclusion of covariate terms into the *Envisage* model identified a confounding effect with 'Tissue', highlighting the fact that RNA quality varies based on the tissue from which it is isolated (Section 4.2.2.3 and Section 4.2.3.2). This suggests that the covariate terms suggested for inclusion here may not be suitable for inclusion in this analysis.

This analysis indicates that the *Envisage* analysis package may not be best suited for the analysis of designed factorised experiments such as this, but may be more suited for studies where covariate effects cannot be assessed prior to analysis and designed into the experiment structure. Linear model approaches to model fitting, such as those utilised in *Envisage*, allow for a wide range of analytical

approaches. There are currently many analytical packages available that provide such tools for the analysis of gene-expression data, and in particular to allow the inclusion of nuisance variables into the analysis. It is hoped that *Envisage* will find a place among these, allowing access to this powerful class of models for users with limited knowledge of statistics and programming, ensuring that the greater biological context can be taken into account.

Chapter 4 Comparison of Transcriptional Response to MycER^{TAM} Activation in Suprabasal Keratinocytes and Pancreatic β-Cells

4.1 Introduction

Section 1.1 provided a review of the known cellular functions of Myc, particularly the molecular pathways involved in promoting proliferation and apoptosis. However, many of the studies cited in this section derive their conclusions from *in vitro* analyses. Whilst these allow us to understand in greater detail *how* Myc is able to drive cellular function, it is not always clear *why* any single pathway is favoured over another in the context of the organism as a whole. The MycER^{TAM} transgenic model for Myc deregulation, described in Section 1.2, provides exquisite control over Myc-induced proliferative and apoptotic phenotypes *in vivo*. This model allows direct control over the 'time 0' of Myc deregulation, allowing tracking of the changes in gene-expression subsequent to activation of ectopic Myc activity. By observing changes in gene-expression that occur in the early stages following MycER^{TAM} activation, direct targets of Myc may be distinguished from those changes occurring downstream of Myc deregulation.

As described in Section 1.2.4, it is clear that tissue context plays a significant role in determining ultimate cell fate in response to deregulated Myc function. Upon activation of the MycER^{TAM} chimeric protein in the suprabasal keratinocytes of the skin, target cells enter the cell cycle and undergo mitosis, leading to an increase in cell numbers within the suprabasal layer, and formation of papilloma-like growths. In stark contrast to this, whilst activation of MycER^{TAM} in pancreatic β-cells results in entry of target cells into the early stages of the cell cycle (G_1/S-phase), the predominant phenotypic outcome is acute apoptosis, leading to destruction of β-cells and involution of islet mass. This indicates that suprabasal skin cells are able to bypass apoptosis-related tumour suppressor functions in order to expose the oncogenic potential of the Myc transcription

factor, while the tumour suppressive functions remain intact in pancreatic β-cells. This fact is corroborated by the proliferative response to MycER^{TAM} activation seen when apoptosis is blocked by concomitant expression of the anti-apoptotic Bcl2 homolog Bcl$_{XL}$ (Pelengaris *et al.*, 2002b).

In this chapter, the divergence in the response to deregulated Myc activity between the skin and pancreatic cell populations is analysed at the transcriptome level using Affymetrix MOE430 Plus 2 mouse GeneChips (Section 1.4), together with statistical tools designed specifically for this type of multi-variable analysis (Chapter 3). The MycER^{TAM} conditional *in vivo* model for Myc activation (Section 1.2) was utilised to allow identification of early responses to deregulated Myc activity, with an aim to identify putative target genes that may explain the seemingly dichotomous function of Myc. Full lists of genes discussed within this thesis can be found in Appendix B: Gene lists.

4.2 Results

4.2.1 Evidence of MycER^{TAM} activation following 4OHT administration

4.2.1.1 Pancreatic islet β-cells

As previously described in Section 1.2.3, 4OHT treatment of *pins-mycER^{TAM}* transgenic mice leads to entry of β-cells into the cell cycle, but also results in a predominantly apoptotic response (Pelengaris *et al.*, 2002b). The short time course of MycER^{TAM} activation considered within this experiment (maximum 32 hours) was insufficient to result in an observable change in islet morphology for target tissue (as seen by staining with H&E), but showed an increase in cell proliferation as evidenced by immunofluorescent staining of Ki67, a marker for cell cycle entry (Figure 4.2.1a).

Samples were stained concomitantly for β-cell-specific Insulin (red) and Ki67 (green), together with a nucleus-specific DAPI stain (blue), to confirm that cells whose proliferation status was altered upon MycER^{TAM} activation were Insulin-producing β-cells. VT control animals (32 hours) showed no proliferating cells within their islets, but after only 4 hours of 4OHT treatment, Ki67 positive β-cells were detectable. By 32 hours of 4OHT treatment, Ki67-positive β-cells were still detectable, but possibly at lower levels. This may be indicative of the promotion of apoptosis.

Apoptotic cells were identified by immunofluorescent staining for the executioner Caspase 3 molecule. The antibody chosen was able to bind both the full length Pro-Caspase 3 protein (35 kiloDaltons (kDa); cytoplasmic) and the large fragment of the cleaved Caspase 3 protein (17/19 kDa; perinuclear). This method was chosen over the often used terminal uridine deoxynucleotidyl transferase-mediated dUTP nick-end labelling (TUNEL) assay (which identifies fragmented

DNA) (Gavrieli *et al.*, 1992), due to its ability to identify cells in the early stages of apoptosis (Duan *et al.*, 2003).

Pancreas tissue was stained concomitantly for the executioner Caspase 3 protein (green), β-cell-specific Insulin (red) and a nuclear specific stain (blue). Staining of VT control tissue sections identified no Caspase 3-positive cells, suggesting that no Caspase 3 (active or inactive) was present. However, by 4 hours of MycER^{TAM} activation, staining for cleaved Caspase 3 was detected around nuclear DAPI throughout the islets. Staining for the active Caspase 3 subunit was also detected 32 hours subsequent to MycER^{TAM} activation, indicating significant apoptotic activity,

Previous studies using an antibody specific for the large subunit (17/19 kDa) of cleaved Caspase 3 have identified no cleaved Caspase 3 staining within WT β-cells (Ladiges *et al.*, 2005) and β-cells of non-obese diabetic (NOD) mice (Reddy *et al.*, 2003a; Reddy *et al.*, 2003b). However, given that identification of apoptosis using TUNEL identifies only 4-7 % of β-cells as apoptotic within 72 hours (Pelengaris *et al.*, 2002b), it is perhaps surprising that such wide-spread apoptosis was detected following only 4 hours of MycER^{TAM} activation. This is particularly true given that this implies that Pro-Caspase 3 was produced and cleaved to its active form within this short time period. Whilst the staining performed here identified a clear increase in positively stained cells, it was not clear exactly how this correlated with Caspase 3 activation.

It is possible that the discrepancy in the data may have been due to so called 'leaky Myc', whereby MycER^{TAM} molecules were activated independently of 4OHT treatment. However, with no WT samples to compare against, this hypothesis could not be examined. It is also possible that these discrepancies were due to some problem with the staining protocol used. Unfortunately, due to constraints on time, repeating the staining was not possible. However, the induction of apoptosis in *pins-mycER^{TAM}* mice following MycER^{TAM} activation is well documented (Pelengaris *et al.*, 2002b; Pelengaris *et al.*, 2004; Lawlor *et al.*, 2006; Cano *et al.*, 2007), and was clear given the change in gene-expression detected (Section 4.2.9.1).

4.2.1.2 Suprabasal keratinocytes

As described in Section 1.2.2, 4OHT treatment of *inv-mycER^{TAM}* transgenic mice leads to entry of suprabasal keratinocytes into the cell cycle, resulting in an increase in proliferation and eventual formation of papilloma-like growths (Pelengaris *et al.*, 1999). The short time course of MycER^{TAM} activation considered within this experiment (maximum 32 hours) was insufficient to produce a noticeable change in keratinocyte morphology, with the suprabasal epidermal layer remaining only 1 or 2 cells thick. However, immunofluorescent staining for the cell cycle marker Ki67 indicated an increase in cell proliferation following MycER^{TAM} activation (Figure 4.2.1b).

Samples were stained sequentially for the proliferation marker Ki67 (green) and the suprabasal keratinocyte-specific Keratin 1 (red), together with the nuclear-specific DAPI stain (blue) to discern between proliferating basal keratinocytes and suprabasal keratinocytes that have begun to proliferate due to activation of MycER^{TAM}. VT control tissue sections (32 hours) showed proliferating cells only within the basal layer of the epidermis. After 32 hours of MycER^{TAM} activation, Ki67 staining was also detected in Keratin 1-positive cells, indicating that activation of MycER^{TAM} in suprabasal keratinocytes resulted in cell cycle entry of suprabasal cells undergoing a program of terminal differentiation as previously described (Pelengaris *et al.*, 1999).

Whilst conventional apoptosis does not feature in the maintenance of epidermis homeostasis, it has been suggested that terminal differentiation of skin keratinocytes may itself act to control against aberrant growth, since affected cells will ultimately be shed and removed from the surrounding micro-environment (Jensen and Watt, 2006). Sequential staining of skin sections for the Caspase 3 executioner protein (green), and suprabasal-specific Keratin 1 (red), together with a nuclear-specific DAPI stain (blue) identified no apoptotic cells in either VT control tissue samples or samples taken following 32 hours of MycER^{TAM} activation. This indicated that activation of MycER^{TAM} did not promote apoptosis

in suprabasal keratinocytes, suggesting that suprabasal keratinocytes are able to circumvent Myc-induced cell death.

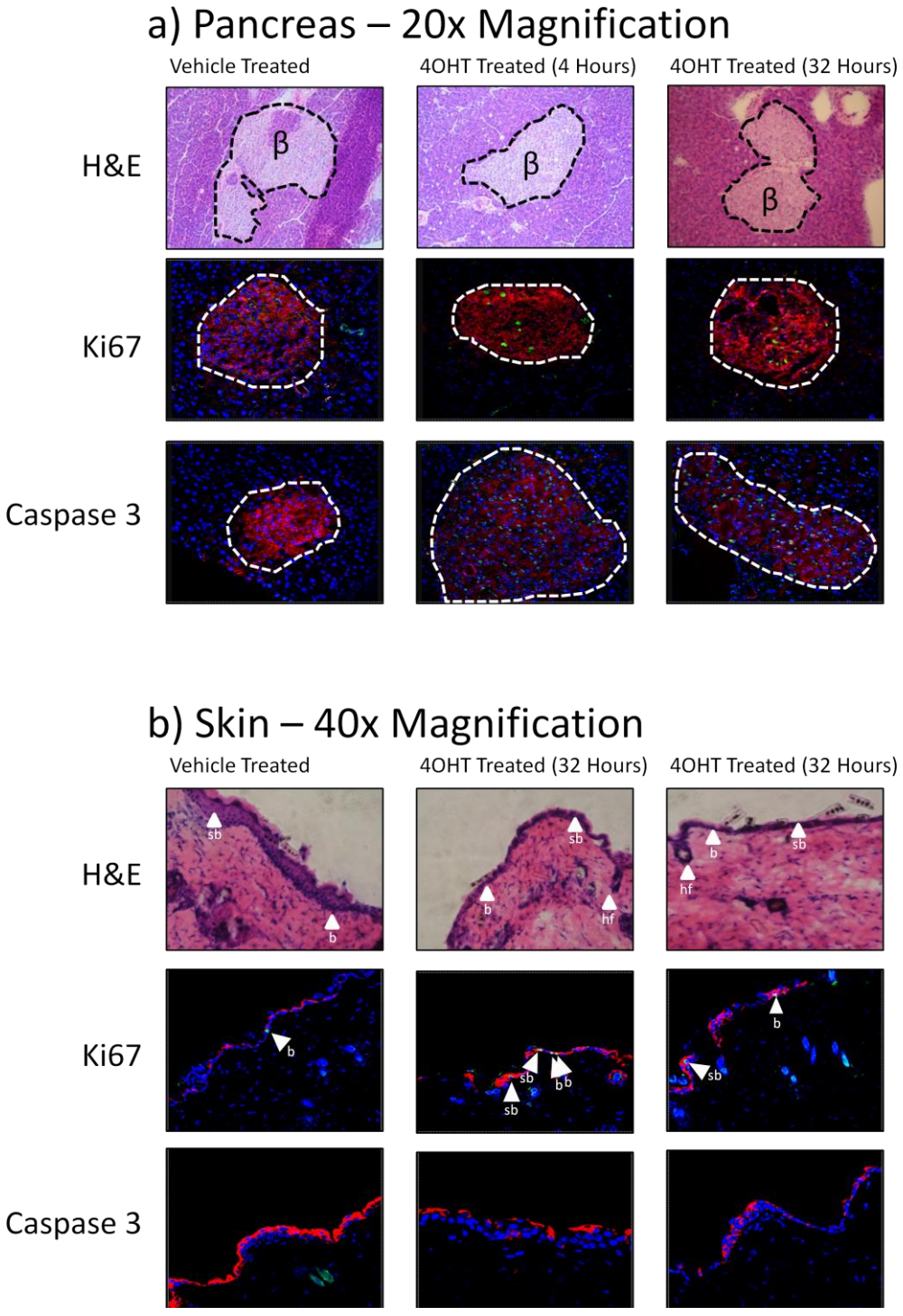

Figure 4.2.1: Short term activation of MycER^{TAM} in pancreatic β-cells and suprabasal keratinocytes (p.t.o)

a) Staining of pancreas tissue sections with H&E identified tissue morphology. The short time scale of activation within this experiment (maximum 32 hours) was insufficient to allow an observable change in tissue morphology between VT controls (32 hours) and 4OHT-treated samples. However, immunofluorescent staining for the cell cycle marker Ki67 (green) identified an increase in β-cell proliferation (identified by concomitant staining with β-cell-specific Insulin; red) following activation of MycER^{TAM}. Staining for the early apoptosis marker Caspase 3 (green) identified a significant increase in apoptotic β-cells (identified by concomitant staining with Insulin; red) following MycER^{TAM} activation. This indicated that activation of MycER^{TAM} in *pins-mycER^{TAM}* mice resulted in promotion of cell cycle and apoptosis in β-cells, as previously described (Pelengaris *et al.*, 2002b).

b) Staining of skin tissue sections with H&E identified epidermis morphology, although the thin nature of murine epidermis made distinguishing epidermal layers difficult. Staining with H&E showed that the short time scale of activation within this experiment (maximum 32 hours) was insufficient to allow an observable change in tissue morphology between VT controls (32 hours) and 4OHT-treated samples. Immunofluorescent staining for the cell cycle marker Ki67 (green) identified positively stained cells in the basal layer of the epidermis in VT control sections. However, cells in the suprabasal layers (identified by sequential staining with suprabasal keratinocyte-specific Keratin 1; red) showed no such staining. Following 32 hours of MycER^{TAM} activation, Ki67 staining was detected in both basal and suprabasal cells, indicating promotion of cell proliferation in suprabasal cells in response to MycER^{TAM} activity. Staining for the early apoptotic marker Caspase 3 (green) identified no apoptotic cells in either basal or suprabasal layers, either before or after MycER^{TAM} activation. This indicates that normal apoptosis does not play a role in the MycER^{TAM}-initiated phenotype in suprabasal keratinocytes, and that activation of MycER^{TAM} in *inv-mycER^{TAM}* mice results in promotion of cell cycle with no discernable apoptosis detected, as previously described (Pelengaris *et al.*, 1999). Note that 4 hour MycER^{TAM} active skin sections were not stained due to constraints on time.

Images for each condition are representative sections from different levels through replicate tissue samples. β, β-cell; b, basal keratinocyte; sb, suprabasal keratinocyte; hf, hair follicle.

4.2.2 Optimisation of laser capture microdissection (LCM) protocol

4.2.2.1 LCM and RNA extraction for pancreatic islet cells

Due to the high number of RNase species, RNA extracted from the pancreas is very prone to rapid degradation (Chirgwin *et al.*, 1979; Gill *et al.*, 1996; Mullin *et al.*, 2006). Freshly isolated pancreas RNA from tissue immediately snap frozen after sacrifice is often of a poorer quality than that of other tissues, resulting in a smaller 28S:18S rRNA ratio. For example, Figure 4.2.2A shows the Bioanalyzer trace for a control sample of freshly isolated good quality RNA from the skin (RIN = 9.2), whilst Figure 4.2.2B shows the Bioanalyzer trace for a freshly isolated sample of "good" quality RNA extracted from the pancreas (RIN = 7.9). Existing protocols described in Materials and Methods, Section 2.2, were not suitable for isolation of RNA of an adequate quality for microarray hybridisation. RNases remain inactive at low temperatures; hence it is important to ensure careful storage of RNA samples to avoid degradation over time. However, the RNases in pancreas are so active that air drying the section at room temperature (RT) for 10 secs is sufficient for RNA to become completely degraded by RNases (Figure 4.2.2C; RIN = 2.3).

Laser capture microdissection of islet cell populations using the PixCell I and II LCM system (Arcturus Engineering, Mountain View, CA) has previously been described (Ahn *et al.*, 2007), using the protocol defined in Sgroi, Teng *et al.* (1999). One key step in this protocol is to ensure that frozen sections are not allowed to air-dry prior to the procedure. To ensure maximum RNA quality, it was necessary to prepare a modified LCM protocol that further preserved RNA integrity. This protocol was optimised in a sequential manner such that RNA integrity was preserved throughout the fixing, staining and dehydrating procedures, and a compromise between RNA quality and RNA yield was made when performing the laser capture procedure. The complete optimised protocol is described in Materials and Methods, Section 2.2.1.

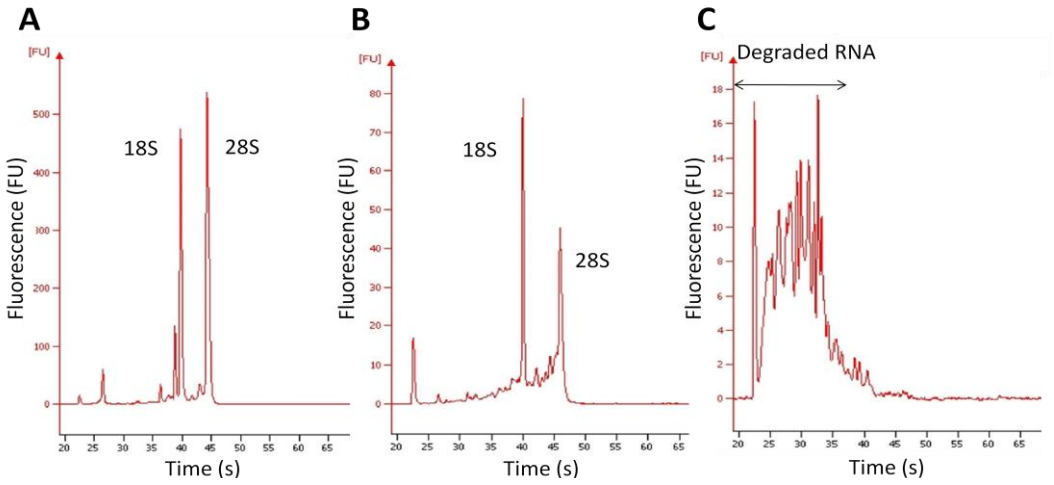

Figure 4.2.2: RNA degradation by ribonucleases is rapid in pancreatic tissue

The high abundance of RNase species in pancreatic tissue means that RNA is rapidly degraded. RNA quality is often measured using the Agilent 2100 Bioanalyzer, which produces an electropherogram trace of fluorescence against time, which also indicates RNA abundance against molecule size. A) RNA isolated from skin tissue is robust and of a good quality. This is evidenced in the Bioanalyzer trace by two distinct peaks representing the 18S and 28S rRNA species. B) RNA isolated from pancreas tissue is generally of a poorer quality due to the presence of ribonuclease molecules. This can be seen by the reduction in size of the 28S peak which is susceptible to RNA degradation. C) Briefly air drying pancreas tissue sections for only 10 secs is sufficient for ribonucleases to completely degrade isolated RNA. This is clear from the lack of any discernible 18S and 28S peaks, and the abundance of peaks at the lower end of the time scale, indicating the presence of mainly smaller cleaved RNA molecules.

4.2.2.2 LCM and RNA extraction from suprabasal keratinocyte cells

In order to extract RNA from a pure population of epidermal cells rather than whole skin (which also contains dermis), we attempted LCM. However, due to the nature of skin epidermal tissue as a protective barrier, intercellular bonds are very strong and isolation of sufficient good quality RNA for microarray hybridisation from LCM of suprabasal keratinocytes proved difficult using the protocol described above, a problem also noted by Agar et al. (2003).

To overcome the strong bonds between cells in the skin, the power of the laser was increased to its maximum setting, and the speed of movement of the piezo mount was decreased to its slowest setting. Even with such drastic changes in the laser parameters, it often took 10 or more complete passes of the laser before the cells of interest became fully separated from the surrounding tissue (as compared to 2 passes for the pancreas). Also, strong electrostatic forces between the skin tissue and the glass slide support made it difficult to consistently lift tissue from the LCM platform which confounded the problem. With the slow setting, this meant that only one or two laser captured tissue segment could be collected within the time scale of 15 minutes, and these were not guaranteed to lift away. Given that the isolated suprabasal tissue segment was generally only 1 or 2 cells in thickness, this resulted in collection of a negligible quantity of cells for RNA isolation.

Several steps were taken to improve isolated cell numbers. Firstly, given that RNA isolated from skin cells is generally more stable than that collected from pancreas, the strict limit of 15 mins for the laser capture could safely be disregarded. However, even by increasing the time spent collecting cells, it was only possible to isolate a small area of suprabasal cells. The use of the serine protease trypsin was considered, to break down cell adhesion proteins and reduce inter-cellular attraction. Trypsin was made up in solution with DEPC-treated phosphate buffered saline (PBS) and added to the section before fixing. Comparing the effects at a variety of concentrations indicated that any beneficiary

effects on the cell adhesion were counteracted by the effect on cell morphology, making it difficult to isolate intact segments of suprabasal keratinocytes.

As suggested by Agar *et al.* (2003), the strong electrostatic forces between the skin sections to the glass slide support were prevented by using non-charged Superfrost glass slides (Menzel-Glaser, Braunschweig, Germany) coated in RNase-free glycerol. A comparison of different times spent in the glycerol/ethanol solution and different concentrations indicated that this process had little effect under the SL μCut laser capture system used in this study.

Given the failure of these optimisation procedures to produce an adequate number of cells for RNA-extraction, and due to the time constraints of the doctorate program, it was decided to forego the use of LCM for the skin samples. Instead, RNA was isolated from whole dorsal skin sections (Materials and Methods, Section 2.2.2). Given that the major constituent of whole skin is the dermis, which consists predominantly of connective tissues, it was believed that there would be fewer cells showing gene-expression changes over time throughout the time course of MycER^{TAM} activation due to normal homeostasis in the skin as compared to the pancreas. One issue that remained however was the presence of basal keratinocytes, which undergo proliferation as a part of normal homeostasis. Changes in gene-expression in basal cells may therefore confound changes in gene-expression driven by MycER^{TAM} activation in the suprabasal keratinocytes. However, this problem would be difficult to avoid with typical approaches to isolation of the epidermis (such as trypsin-based degradation of the dermis), and given the thin nature of murine epidermis would have been impossible to avoid using the SL μ-cut LCM platform.

4.2.2.3 Quality of isolated RNA from optimised laser captured pancreatic islets and whole skin sections

RNA quality was measured using the RNA integrity number (RIN) as defined by the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA), and RNA was quantified using a Nanodrop ND-1000 spectrophotometer (Nanodrop

Technologies, Wilmington, DE). RNA quality was generally good, with mean RINs of 6.1 ± 0.85 and 8.1 ± 0.90, and overall yields of 76 ± 46.7 ng and 80 ± 49.0 ng total RNA for the pancreas and skin respectively. Overall RNA quality for all samples in the study is shown in Figure 4.2.3. In general, skin samples produced RNA of a higher quality than pancreas samples, which is likely a result of the increased RNase activity present in pancreas tissue. This suggests that there is in fact some confounding effect between tissue and RNA quality.

Both tissues produced RNA that was determined to be of a suitable quality for use in microarray analyses, with only one out of forty-eight samples producing RNA with a RIN less than 5; Panc T 8hr (2) ($RIN = 4.7$). However, whilst the RIN provides a useful objective measure of RNA quality, there are no commonly accepted threshold values to determine the relative merit of the observed value. For the present analysis, an arbitrary threshold of RIN = 5 was used to determine which RNA samples were of a suitable quality for microarray hybridisation, and this cutoff was chosen based on the conventions of the Molecular Biology Service Department at the University of Warwick. However, it must be noted that selection of samples for removal based on this threshold is similarly arbitrary.

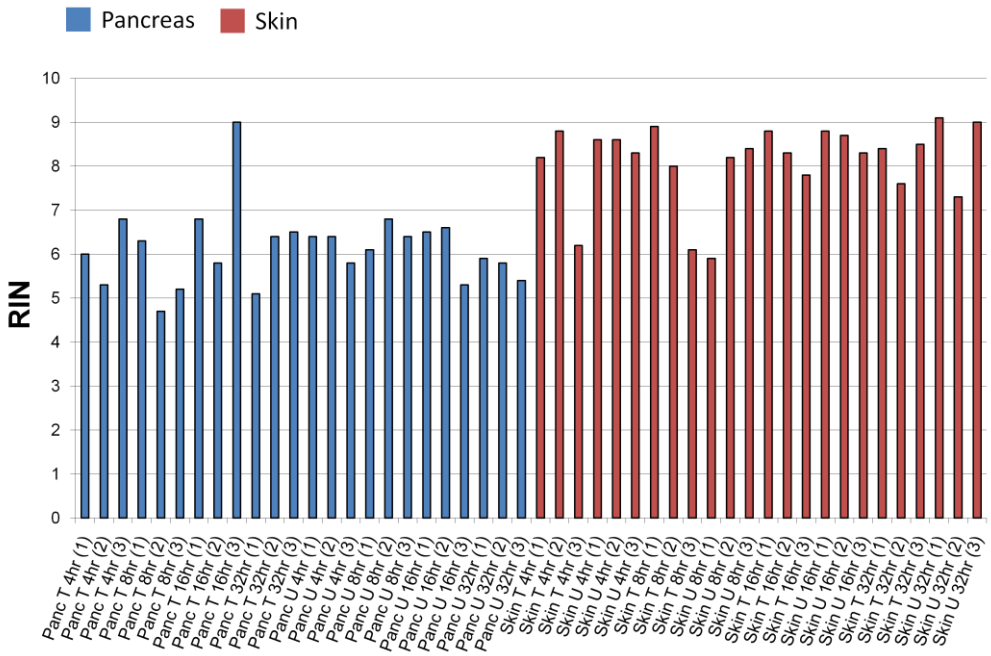

Figure 4.2.3: Quality of isolated RNA for pancreas and skin tissue

Quality of RNA following isolation from laser captured islets and whole skin sections was measured using the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA). This produced a quantitative value between 1 and 10, the RNA integrity number (RIN), defining the quality of the RNA sample. A RIN of 5 or above was considered to be suitable for microarray hybridisation. In general, RNA quality was good, with only one out of forty-eight samples analysed showing a RIN below this limit (Panc T 8hr (2)). There was a clear difference in quality of RNA between the two tissues, with skin samples generally producing RNA of a higher quality than pancreas samples. This is indicative of the higher number of RNases present in the pancreas, and also of the difference in sample preparation between the two tissues, and suggests a confounding effect between tissue and RNA quality.

4.2.3 Optimisation of *in vitro* transcription (IVT) protocol

4.2.3.1 Protocol optimisation

The standard protocols for the Affymetrix IVT protocol defined in the GeneChip Expression Analysis Technical Manual (Affymetrix, 2004) resulted in very low yields of 2a-cRNA for six samples in the first batch processed (Figure 4.2.4). To avoid loss of precious RNA samples, analysis of further batches was halted while this problem was studied further. Personal communication with Giorgia Riboldi-Tunnicliffe from Affymetrix raised the possibility of the presence of some inhibitor within the RNA samples that may inhibit the amplification of cDNA.

Following Giorgia's recommendations, the process was repeated with the following modifications to the protocol. Starting RNA samples were diluted to 6 µl in nuclease-free water instead of 3 µl, and double volumes of polyA control probes and first cycle reagents were used to improve first round cRNA yield. After first cycle clean up, cRNA was eluted in normal volumes nuclease-free water and second cycle cRNA synthesis was performed as normal. This modified protocol (described in detail in Materials and Methods, Section 2.4.1) greatly improved 2a-cRNA yields, with only 2 of 14 samples tested failing to meet the recommended minimum yield of 10 µg (Figure 4.2.4). Only one of these samples produced a yield significantly below this cutoff.

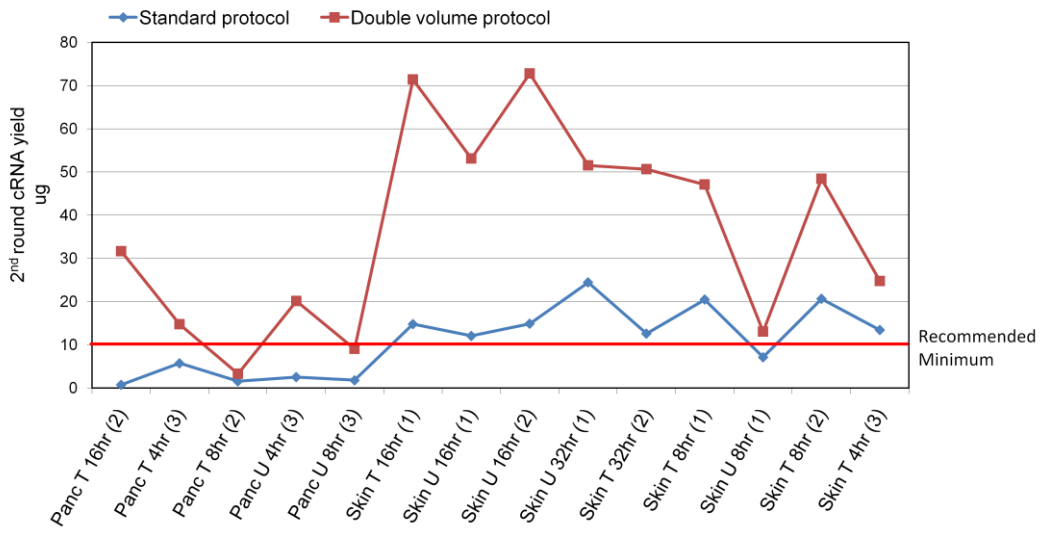

Figure 4.2.4: Using double volumes of reagents in the second cycle of the Affymetrix IVT reaction produced higher yields of 2a-cRNA for microarray hybridisation

Following the standard protocols for cRNA synthesis using the Affymetrix GeneChip Two-Cycle Target Labelling and Control Reagents kit produced low yields of 2a-cRNA. 6 out of the 14 samples tested produced a yield below the 10 µg minimum required for hybridisation to the arrays. Using double volumes of all reagents and polyA controls in the first cycle cRNA synthesis step of the protocol produced greater yields for all samples after second cycle synthesis of 2a-cRNA, with only 2 of 14 samples failing to produce yields suitable for hybridisation.

4.2.3.2 Yield of 2a-cRNA from optimised IVT protocol

The resulting yields of 2a-cRNA following the modified 2-round linear amplification are shown in Figure 4.2.5. The modified 2-round linear amplification method yielded 32 ± 17.1 µg and 52 ± 21.7 µg high-quality, 2a-cRNA for pancreas and skin respectively for 43 of the 48 independent samples. Five samples produced less than the recommended 10 µg 2a-cRNA after 2-round linear-amplification: Panc T 8hr (2) ($23.8\ \mu g$), Panc T 8hr (3) ($4.1\ \mu g$), Panc T 32hr (1) ($0\ \mu g$), Panc U 16hr (3) ($1.9\ \mu g$) and Panc U 32hr (1) ($3.2\ \mu g$). Of these 5 samples, only 1 also showed a low RIN (Panc T 8hr (2)). The yield for the remaining samples was in general good, indicating that the poor yield of 2a-cRNA may be due to the presence of some remaining inhibitor that may prevent amplification. Of particular note was the sample Panc T 32hr (1) which produced an almost undetectable yield of 2a-cRNA. As with the RNA quality, the yield of 2a-cRNA was in general higher for skin samples than for pancreas samples, suggesting a confounding effect with the tissue variable. The high variability seen for these yields most likely relates to the fact that, due to limitations of the LCM procedure, 2a-cRNA was typically produced from low volumes of RNA.

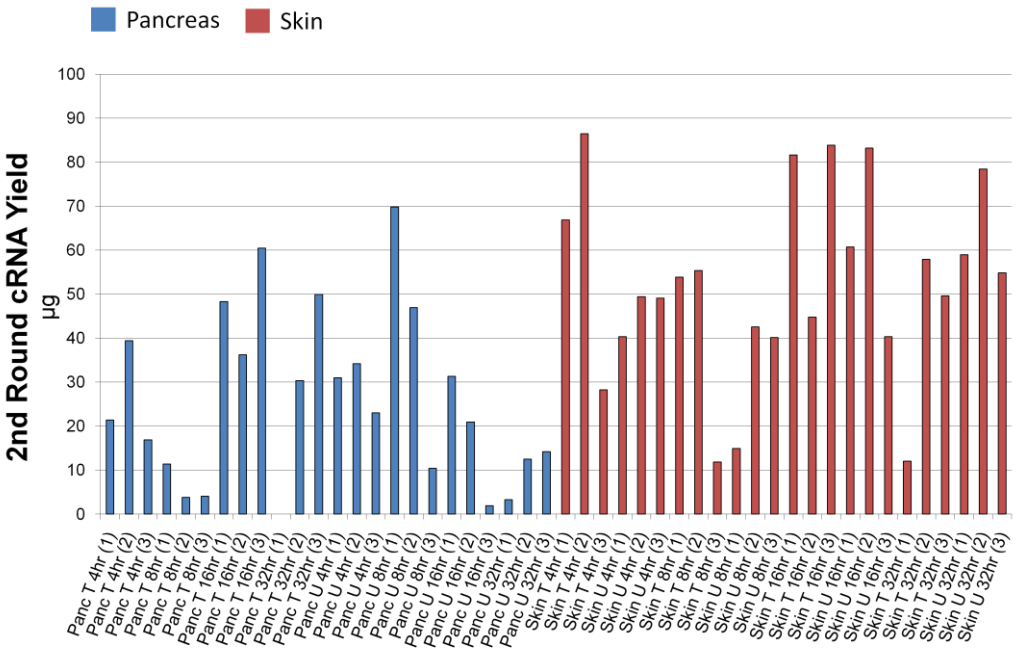

Figure 4.2.5: Yield of 2a-cRNA following two-cycle *in vitro* transcription

A two-round IVT reaction was used to prepare double-amplified biotin-labelled cRNA (2a-cRNA) for microarray hybridisation. The protocol was modified as described in Section 4.2.3.1 to ensure maximum yield for hybridisation. A minimum of 10 µg 2a-cRNA is suggested for hybridisation to Affymetrix GeneChips. In general, 2a-cRNA yield was good, with skin samples producing higher yields on average than pancreas samples (suggesting confoundment of 2a-cRNA yield with tissue). Five of the forty-eight samples in the study produced a yield of 2a-cRNA lower than the recommended limit, with one sample – Panc T 32hr (1) – producing an undetectable yield of cRNA. All samples were hybridised to avoid introduction of systematic errors by repeating samples in isolation.

4.2.4 Microarray data quality control

The quality of the data was assessed using a large number of diagnostic quality control (QC) metrics as described in Section 1.4.3. Despite the importance of quality control in microarray experiments, guidelines for acceptable quality in these analyses are vague. The approach used here was inspired by the comprehensive QC analysis of Jones *et al.* (2006), where pre-hybridisation QC measures for Affymetrix microarrays were compared to post-hybridisation QC measures to identify possible diagnostics for the identification of samples that will perform poorly before hybridisation to expensive arrays. Here, QC diagnostic tests were used to confirm the quality of sample data at all stages of the microarray process, including pre-hybridisation measures to confirm the quality of the starting material, post-hybridisation measures to confirm that the hybridisation procedure introduced no errors, and measures of the quality of the resulting data (at both the probe-level and at the summarised probe set level).

For brevity, the results from each individual test are not included here. As an example, Figure 4.2.6 shows the distribution of gene-expression data for each sample represented as boxplots (as described in Section 1.4.3.8). Replicate samples (shown using identical colouring) should show comparable distributions, allowing identification of poor quality samples (ringed red). In general, samples showed very good comparison, particularly for the skin.

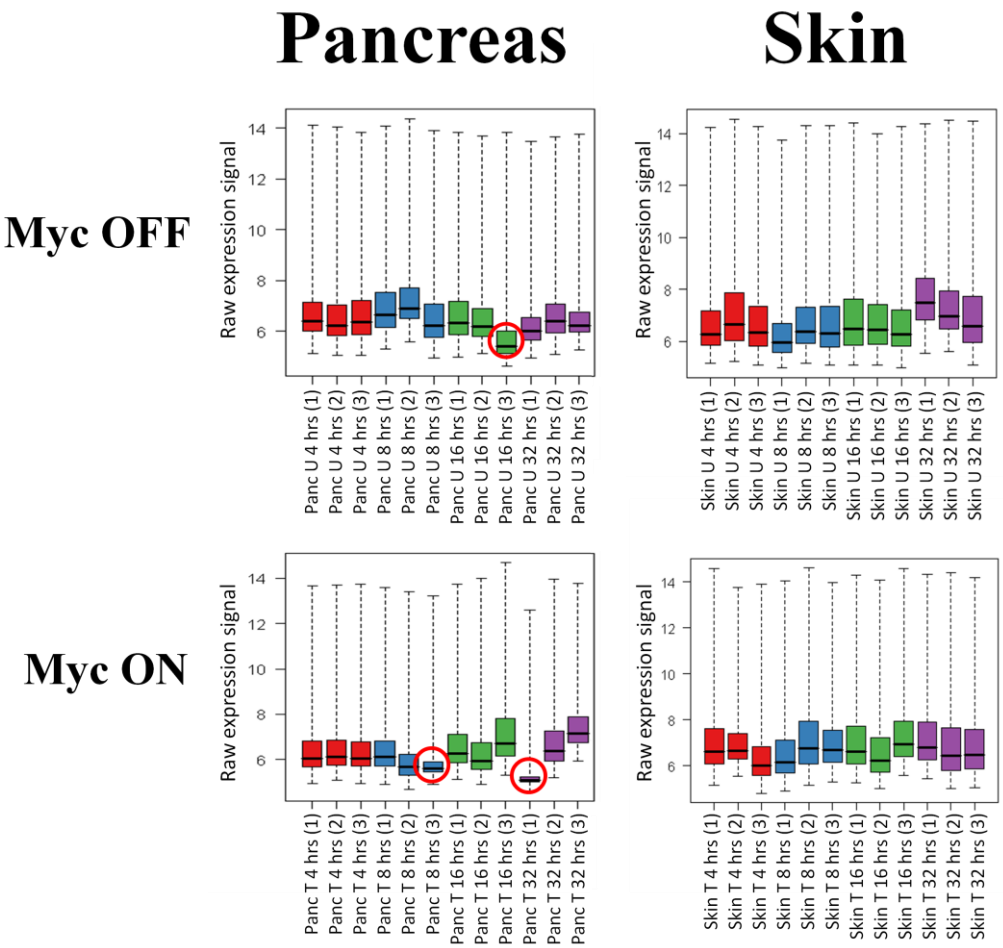

Figure 4.2.6: Graphical display of the distribution of gene-expression data using boxplots allows identification of outlying samples

Boxplots of gene-expression data sets allow visual representation of the distribution of the data. Comparing individual boxplots for un-normalised data allows identification of outlying samples whose data is not comparable to its replicates. Triplicate samples are coloured similarly to assist in comparisons. In general, the distribution of replicates is comparable, and samples showing non-comparable data distribution can be identified (ringed red). These represent samples showing low yields of 2a-cRNA, indicating that hybridising an array with a small amount of starting material can result in poor signal. This is especially true of sample Panc T 32 hrs (1), which showed a yield of almost 0 µg of 2a-cRNA in the IVT reaction, and appears to show generally low expression signals throughout.

Samples identified as poor quality in the QC analyses also showed low 2a-cRNA yields, suggesting that low quantities of starting material may result in low levels of hybridisation to the array and poor data quality. Of particular note is sample Panc T 32 hrs (1), which showed a yield of almost 0 μg 2a-cRNA using IVT, and generally showed very low levels of hybridisation.

A simple penalty scoring system was used to penalise samples that performed poorly in the QC metrics and the resulting QC score was used to identify poor quality samples in the sample set. Samples were assigned a score of 0 if the results of the QC test were satisfactory, 1 if the results of the QC test were below recommendations but still tolerable, or 2 if the results of the QC test were poor. This penalty score was used to identify poor quality samples instead of a single QC metric to combine information regarding sample and data quality from a large number of metrics relating to various aspects of sample processing. In particular, Jones *et al.* (2006) showed that the quality of data was not well predicted by pre-hybridisation measures such as the RIN, suggesting that no single measure captures the full quality of samples for determining poorly hybridised samples.

Figure 4.2.7 shows the QC penalty scores for each of the 48 samples. Bars on the graph are coloured if they represent samples that were hybridised with a low quantity of 2a-cRNA (red), had a low RIN (green) or both (yellow). From this graph, it appears that hybridising poorer quality RNA with a low RIN had no clear effect on determining ultimate data quality, whilst hybridising samples with a lower quantity of 2a-cRNA resulted in a high QC penalty score for several samples.

This process identified seven of the forty-eight samples as being of poor quality: Panc T 16hr (3), Panc T 32hr (1), Panc U 8hr (2), Panc U 16hr (1), Skin T 4hr (3), Skin T 8hr (3), and Skin U 8hr (1). The poor quality of these samples in comparison to their replicates would result in an increase in the variance of residual estimates, leading to lower p-value estimates during model fitting. These samples were therefore removed from all further analyses to increase the chance of detecting significant effects. However, removal of these samples leads to the introduction of bias into the data due to the resulting unbalanced treatment groups

(since some groups will have three replicates while others will have fewer). The resulting unbalanced design also increases the complexity of the required model since the treatments are no longer orthogonal. It was lucky to note that samples removed from the analysis represented unique conditions such that one replicate at most was removed from each of the triplicate sets.

The relationship between the QC penalty score and both the RNA integrity and 2a-cRNA yield is shown in Figure 4.2.8. It is clear that both low and high RIN samples produced poor quality data, and in fact no samples that were identified as being of poor quality showed a $RIN < 5$. This indicated that an observed low RIN is not sufficient to determine samples that may produce poor quality data. Similarly, the correlation of the QC penalty score with the 2a-cRNA yield indicated that only one of the five samples that were hybridised with less than 10 μg 2a-cRNA were removed. This indicates that good quality data can be produced from low 2a-cRNA yields and that the limit of 10 μg may be overly conservative. These pre-hybridisation QC metrics may therefore be insufficient to determine the quality of the resulting data, a conclusion also drawn by Jones *et al.* (2006).

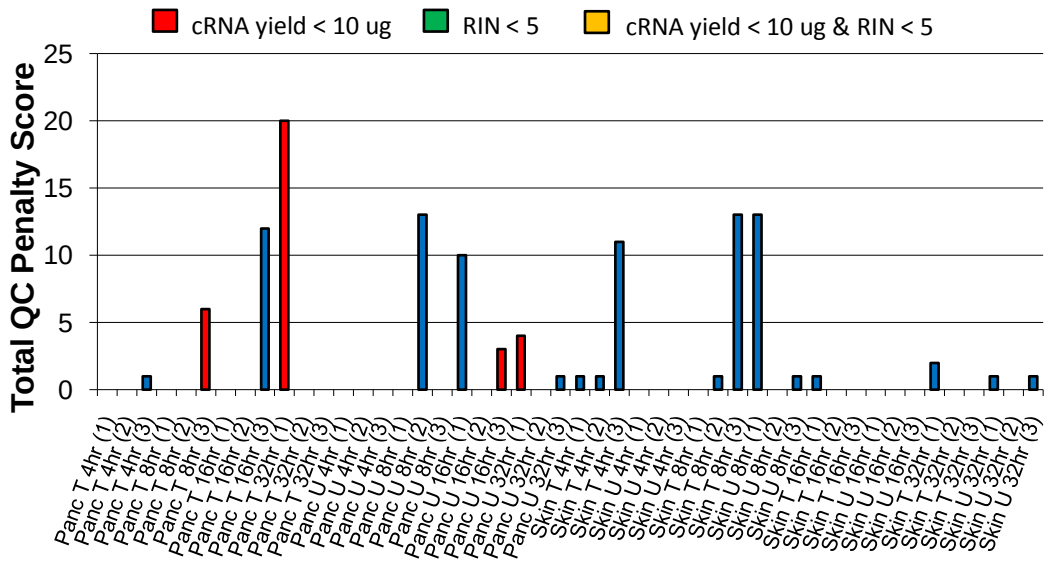

Figure 4.2.7: Quality control penalty scores for array data

A simple penalty scoring system, based on a number of QC metrics at all stages of microarray analysis, was used to identify outlying samples that may bias the microarray gene-expression data. In general, samples that performed poorly for one QC test performed poorly across all QC tests. These samples produced large QC penalty scores and were removed if this was greater than a cutoff value, chosen arbitrarily as 10. Bars on the graph are coloured if they represent samples that were hybridised with a low quantity of 2a-cRNA (red), had a low RIN (green) or both (yellow). From this graph, it appears that low RIN had no clear effect on determining ultimate data quality, whilst only one sample with a low 2a-cRNA yield resulted in a high QC penalty score. This indicates that pre-hybridisation measures of sample quality may not be suitable metrics for determining resulting data quality.

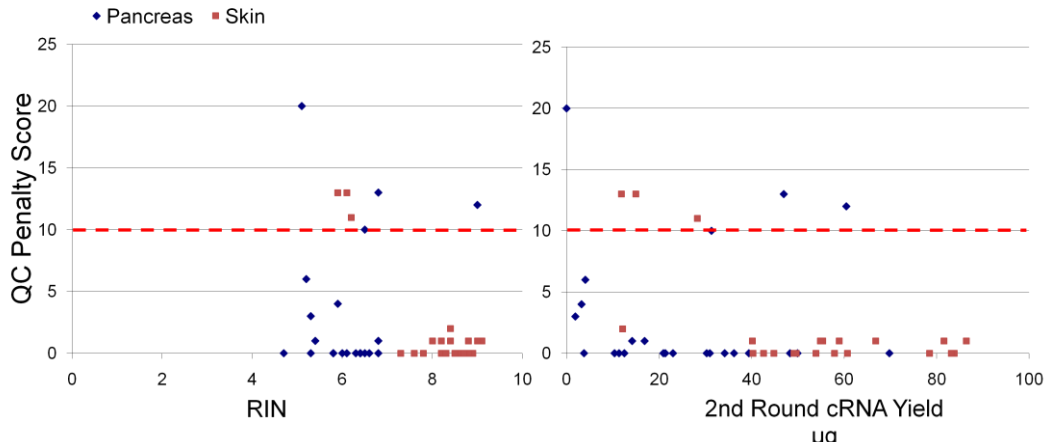

Figure 4.2.8: Comparison of QC penalty score with RNA integrity and 2a-cRNA yield

A simple penalty score system was implemented based on a number of quality control metrics to determine outlying samples within the data. In general, samples that performed poorly in one QC test performed poorly across all of the QC tests. A cutoff of 10 was used to determine poor quality samples that should be removed prior to analysis. The quality of samples prior to hybridisation was a poor estimator of ultimate data quality. Samples removed from the analysis showed a wide range of RINs and produced varying yields of 2a-cRNA. In fact, good quality, reproducible data was produced from less than the recommended minimum starting amount of 2a-cRNA, indicating that the suggested minimum of 10 µg 2a-cRNA may be conservative.

4.2.5 Data quality analysis

Data were clustered by sample in GS-GX using hierarchical clustering based on the Pearson correlation (Section 1.5.4), such that samples showing similar expression profiles across the gene set were ranked as being more similar than those showing more varied patterns of expression. A heatmap showing clustering of similar samples, together with a hierarchical tree structure identifying the level of similarity between samples, can be seen in Figure 4.2.9. When clustering was performed across all samples in the data set, the majority of samples clustered based on the tissue of origin (Figure 4.2.9a). However, six samples did not fall into their associated tissue class, indicating that these samples represented outlying expression sets.

These samples represented all but one of the seven samples identified as having poor quality hybridisation in Section 4.2.3.2. This may indicate that sample quality was poor, preventing correct clustering with related samples of a higher quality. It is also possible that these samples may indicate a mix-up in sample labelling, and the author notes the high RIN (RIN = 9) of sample Pancreas T 16hr (3), which may be more consistent with skin derived RNA than pancreas. However, the poor performance of these samples across a wide range of QC metrics, and the general care that was taken in preventing such labelling errors, may indicate that poor sample hybridisation is more likely the case. Removal of these samples from further analyses produced a quality controlled data set that showed perfect clustering based on tissue of origin (Figure 4.2.9b).

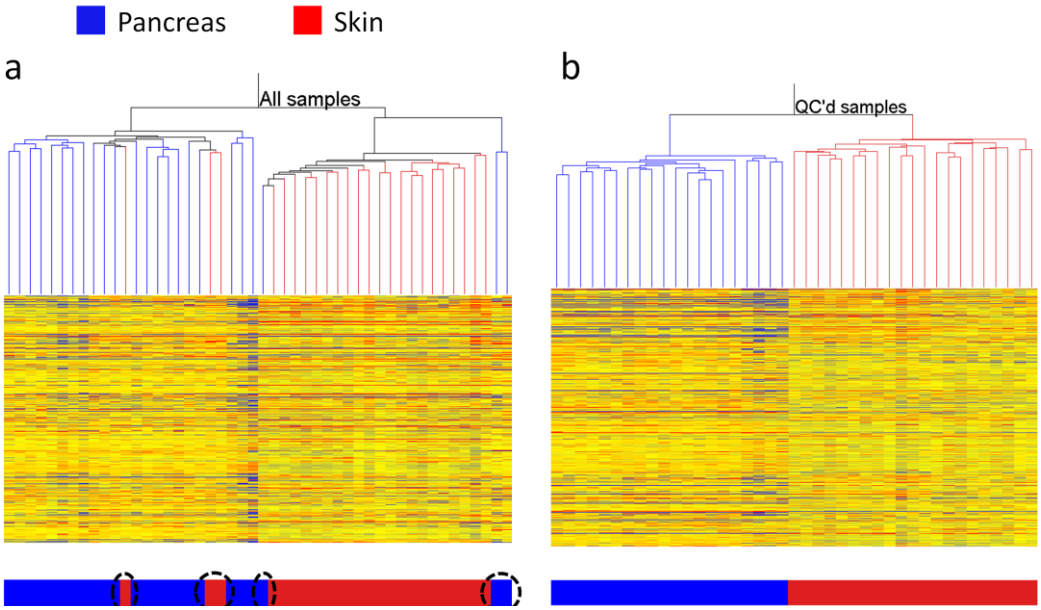

Figure 4.2.9: Clustering of normalised expression signal identifies a clear separation between skin and pancreas

Probe level data was summarised and normalised, and expression levels were compared across samples (using the Pearsson correlation as a measure of similarity) to identify gene-expression profiles that do not correlate well between replicates. Hierarchical clustering of samples identified a clear separation between skin and pancreas samples, which was expected given the difference in sample quality and processing between these cohorts. This shows the need to normalise each cohort individually (as described in Materials and Methods, Section 2.6.1). a) Clustering of all samples in the data set identified 6 samples (highlighted) that clustered within the wrong cohort, indicating poor data quality. These samples represent six of the seven samples shown to be of poor quality using a combination of QC metrics. b) Removal of these samples, along with the 7th poor quality sample defined in Section 4.2.3.2, resulted in complete clustering of samples by tissue.

Quality controlled data were normalised as described in Section 2.6.1, producing time courses for both the skin and the pancreas such that the expression signal at a particular time point represented the fold change detected in the probe set following MycER^{TAM} activation. The resulting expression profiles of the normalised data (described in Section 2.6.1) for the 12,349 curated probe sets (described in Section 2.6.2) are shown in Figure 4.2.10. The profile for each probe set is coloured based on the normalised signal at the 4 hour time point for the respective tissue (red, up-regulated; yellow, no change; blue, down-regulated).

For the pancreas, it appears that early changes in gene-expression remain stable, with probe sets showing an increase or decrease in expression at 4 hours remaining up- or down-regulated at 8 hours. By 16 hours, changes in expression had become more varied, illustrating the dynamic nature of gene-expression.

For the skin, changes in expression across the time points appears more varied, with many genes showing initial down-regulation at 4 hours also showing a large increase in expression at 8 hours. It also appeared that the expression of genes showed peaks at the 8 and 32 hour time points. Given that these time points represent 8 hours following daily 4OHT administration, it is possible that this represents a delay in MycER^{TAM} activation following topical application of 4OHT. The structure of the gene-expression data is discussed further in the following chapters.

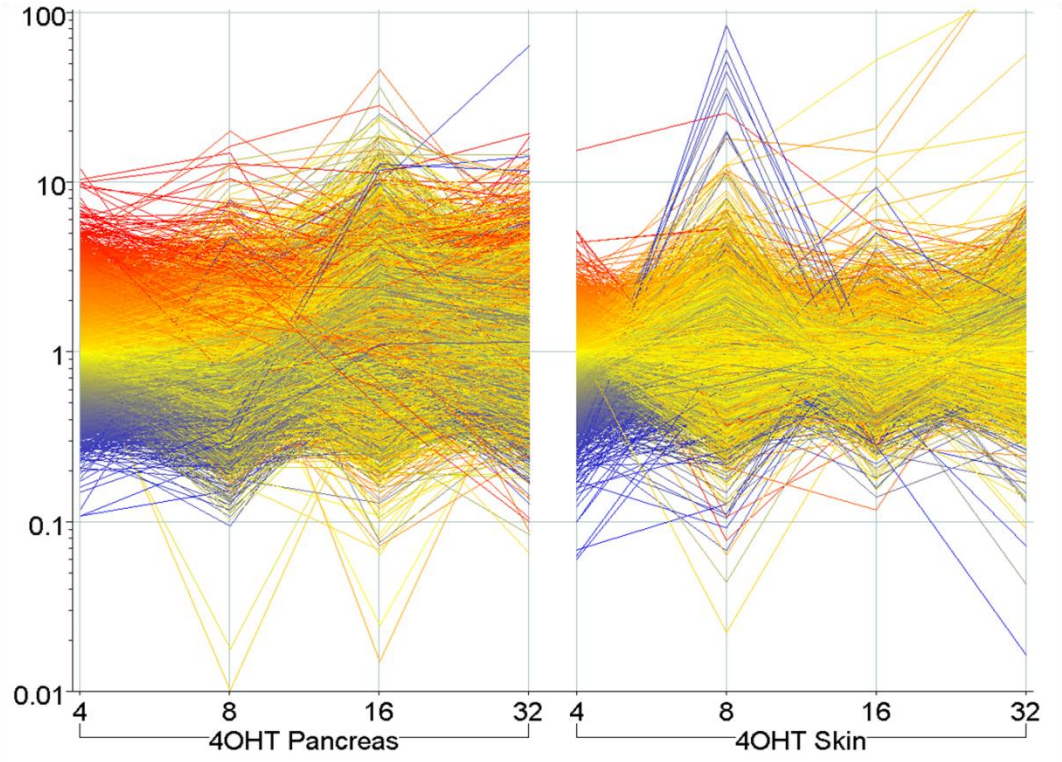

Figure 4.2.10: Normalised expression profile for curated genes following a time course of MycER^{TAM} activation in pancreatic β-cells and suprabasal keratinocytes

As described in Section 2.5, the design of this experiment allowed normalisation of 4OHT-treated mice to their vehicle-treated counterparts. The GC-RMA normalised data for each 4OHT-treated sample was normalised to the median of the vehicle-treated GC-RMA normalised replicates within the same treatment group. This was done for each time point in the experiment for the two tissues. Data were further normalised using a median normalisation across samples. The normalised signals for each condition are shown here for the 12,349 curated probe sets described in Section 2.6.2. The expression profile for each probe set is coloured based on its expression level at 4 hours for the respective tissue (red, up-regulated; yellow, no change; blue, down-regulated).

In the following sections, results were taken from the *Envisage* model fitted with covariate terms as described in Section 3.4.2. Model terms '4OHT', 'Tissue', 'Time', '2a-cRNA yield', 'RIN' and 'Batch number', and their associated interactions (up to the first order), were used in the model fitting procedure for each gene. Significance p-values for each term were calculated for each gene that included the term in its fitted model, as described in Section 3.3.3. For the first stage of the analysis, genes showing a significant effect for 4OHT treatment were of primary interest, as these represented genes showing response to activation of $MycER^{TAM}$.

Both tissues were analysed together within the fitted model, although the results for each tissue are presented separately in the following sections to make clear distinctions between the responses to $MycER^{TAM}$ activation in the two cell types. The values expressed within these tables represent the normalised data (representing fold-change following $MycER^{TAM}$ activation), as described in Materials and Methods, Section 2.6.1. 4OHT p-values defined in each table represent the p-values estimated for the '4OHT' variable using *Envisage*. Genes of interest were identified by a significant effect in response to 4OHT treatment (p-value ≤ 0.05, chosen due to convention) in this covariate model, and a 2-fold change in the normalised gene-expression at either 4 or 8 hours following $MycER^{TAM}$ activation. Focus was placed on these early time points to identify genes whose change in expression may relate to direct Myc-activation, since Myc transactivation is shown to be maximal within 8 hours (Wu *et al.*, 1999).

The magnitude of the fold changes for each tissue and time point was identified by colouring the tables to identify genes showing $\geq$ 2-fold up-regulation (red) or down-regulation (blue). t-test-based flags indicate the significance of gene-expression changes following activation of $MycER^{TAM}$ within individual tissue/time point conditions ("*" for p-value ≤ 0.05, "**" for p-value ≤ 0.01). Briefly, for each gene and for each condition in the experiment (i.e. some specific time point and tissue), a t-statistic was calculated and used to estimate the significance of the difference in the means of the fitted data Y_T for the 4OHT-

treated samples and the means of the fitted data Y_U for the vehicle-treated samples given the fitted model. The t-statistic is defined as:

$$t = \frac{Y_T - Y_U}{SED} \quad \textbf{4-1}$$

Where SED is the standard error of the difference:

$$SED = \sqrt{\left(\frac{1}{n_T} + \frac{1}{n_U}\right)(SSE/df_{res})} \quad \textbf{4-2}$$

Where SSE is the error sum of squares $\sum_i^n (y_i - \hat{y}_i)^2$, df_{res} is the residual degrees of freedom from the full fitted model, and n_T and n_U are the number of observed values for the 4OHT-treated and VT groups (for the given time point and tissue) respectively. This statistic was compared to the t-distribution with df_{res} DF to estimate the probability that such a t-statistic would be found purely by chance. These contrast p-values therefore specify the significance of the change in expression detected at each experimental condition.

4.2.6 The transcriptional response upon activation of MycER^{TAM} in the skin was delayed in comparison to the pancreas

Analysis of gene-expression using *Envisage* identified 5,841 probe sets (representing 4,624 unique genes) as being significantly altered following activation of MycER^{TAM} by administration of 4OHT (Section 3.4.2). Of these, the expression levels of 1,309 probe sets (1,101 genes) were altered greater than 2-fold (up- or down-regulated) after only 4 hours of 4OHT treatment for the pancreatic β-cells, while only 373 probe sets (348 genes) were similarly affected for the suprabasal keratinocytes. However, after 8 hours following initial 4OHT treatment, the expression levels of 1,711 (1,445 genes) and 1,567 (1,389 genes) probe sets were altered greater than 2-fold for the pancreas and the skin

respectively (Figure 4.2.11). This suggests that the transcriptional effect of MycER^{TAM} activation was more rapid in the β-cells than for the suprabasal keratinocytes. The transcriptional delay in skin compared to the pancreas may be the result of different methods of 4OHT administration used for the two tissues (topical and IP administration for the skin and pancreas respectively). Topical administration of 4OHT may require a longer time to perfuse into keratinocytes of the skin than for β-cells of the pancreas. MycER^{TAM}-initiated changes in gene-expression were therefore detected at a later time point for the skin than for the pancreas.

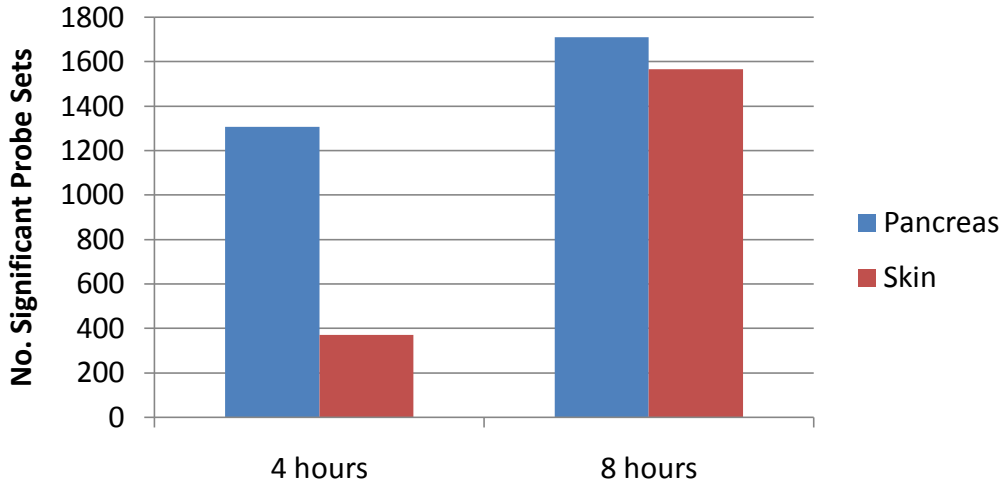

Figure 4.2.11: Transcriptional response to MycER^{TAM} activation was delayed in the skin

Activation of MycER^{TAM} led to a significant change in expression of a large number of genes within only 4 hours for the pancreas (> 2-fold change, $p < 0.05$). However, the number of genes showing a significant change in expression was much lower for the skin, and a similar transcriptional response to the pancreas was not detected until 8 hours post-MycER^{TAM} activation. This indicated that topical application of 4OHT may be a slower means of 4OHT administration than IP injection, possibly due to the need for perfusion through the outer skin layers.

4.2.7 Activation of MycER^{TAM} in the skin and the pancreas mediated the transcription of genes involved in a wide range of cellular functions

As described in Section 1.1, Myc is involved in a wide range of cellular functions, including cell growth, proliferation, apoptosis and differentiation. Analysis for enrichment of GO terms within the list of 5,841 probe sets (4,624 unique genes) found to be significantly altered following activation of MycER^{TAM} identified genes involved in myriad cellular functions. Figure 4.2.12 shows gene list numbers for the biological functions of most relevance to this study. The majority of genes found to be significantly altered following MycER^{TAM} activation in either tissue (green bars) appeared to be related to metabolism, which may not be surprising given the role of Myc as a global transcription factor. This was also indicated by the large number of genes relating to transcription, transport and signal transduction.

Interestingly, genes involved in post-transcriptional (mRNA processing, RNA splicing) and post-translational modification (phosphorylation, protein folding, proteolysis and ubiquitin cycle) were also present. This suggested that Myc may also exhibit regulatory effects on cellular function through non-transcriptional means, a field of work that has recently received much attention (Cowling and Cole, 2007; Dominguez-Sola *et al.*, 2007). As expected, genes involved in cell cycle progression and apoptosis were also present, along with genes involved in related functions such as cell adhesion and cytoskeletal organisation. Genes involved in angiogenesis, cellular growth, DNA damage and DNA repair were also represented.

Independent analysis of GO classifications for genes showing significant early changes in expression ($\geq$ 2-fold change within the first 8 hours of MycER^{TAM} activation and $p \leq 0.05$ for model term '4OHT') for the pancreas (blue bars) and skin (red bars) identified no clear difference in gene numbers between the two. This indicates that MycER^{TAM} activation resulted in modulation of a wide range of functions in both tissues. In general, activation of MycER^{TAM} in the pancreas

resulted in a slightly higher proportion of genes being affected than for the skin, although this is likely a result of the delayed transcriptional response seen in the skin as compared to the pancreas (Section 4.2.2).

Table 4.2.1 shows a hypergeometric analysis of GO term enrichment (Materials and Methods, Section 2.6.5) performed in GS-GX for the genes of interest. p-values represent the probability of seeing the observed proportion of genes from a particular GO term by chance (based on a hypergeometric distribution), with a low value suggesting significant enrichment of genes relating to the specific biological process. From this, it is clear that genes involved in RNA processing and metabolism, DNA replication, cell cycle, apoptosis and DNA damage repair are all significantly enriched within this list. These data therefore fit well with the notion of Myc as a pleiotropic transcription factor, and particularly with its known role in proliferation and apoptosis.

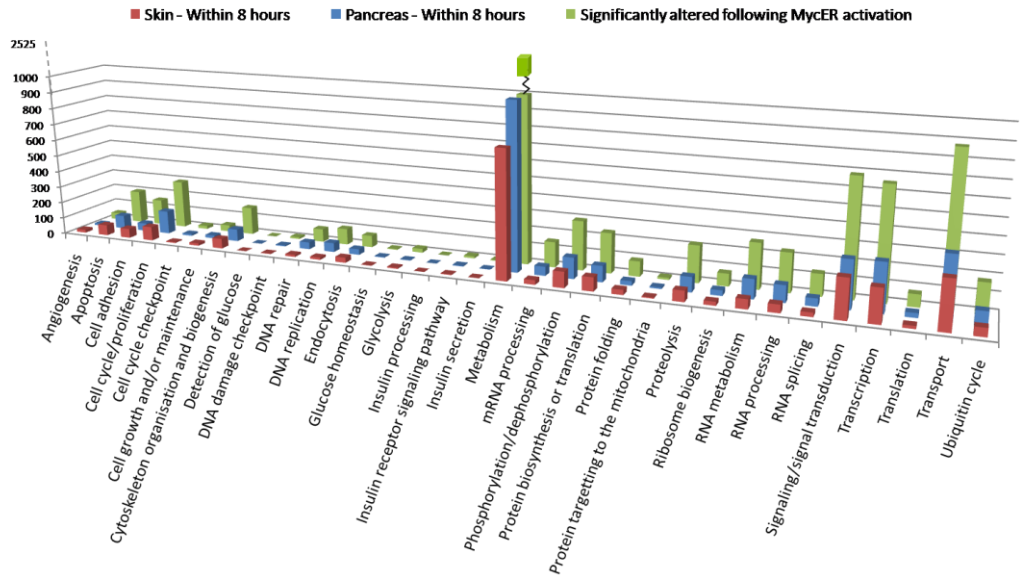

Figure 4.2.12: Activation of MycER^{TAM} resulted in significant changes in expression of genes involved in a wide range of cellular functions

Analysis of 5,841 probe sets (representing 4,624 unique genes), whose expression was altered upon activation of MycER^{TAM} in pancreas or skin (green bars), identified a wide range of Myc-mediated functions. Myc is known to have pleiotropic properties, and this was revealed in the microarray data. The majority of genes found to be regulated by Myc were involved in metabolism and transcriptional activity. However, Myc also appeared to play a part in regulating genes involved in post-transcriptional and post-translational regulatory functions, indicating a possible role for Myc in non-transcriptional modification of gene regulation. Genes involved in cell proliferation and apoptosis, and related functions such as cell growth, cell adhesion and cytoskeletal organisation were also affected by deregulated Myc. Independent analyses of genes whose expression was altered greater than 2-fold within 8 hours for the skin (red bars) and the pancreas (blue bars) identified no significant variation between the levels of response for the two tissues.

Table 4.2.1: Gene ontology enrichment for MycER^{TAM}-mediated gene-expression

GO Term	Expression altered in response to MycER^{TAM} for either tissue	Pancreas 4 or 8 hours	Skin 4 or 8 hours	p-value
mRNA processing	151	54	27	1.05E-26
Cell cycle/proliferation	292	141	85	5.19E-18
RNA processing	231	100	48	5.67E-18
RNA metabolism	269	115	57	6.92E-18
RNA splicing	129	46	25	1.11E-11
Metabolism	2525	1052	763	2.69E-11
Ribosome biogenesis	74	32	23	1.21E-09
DNA replication	97	55	14	1.94E-09
Protein targeting to the mitochondria	12	9	6	9.64E-07
Protein biosynthesis or translation	239	96	82	2.75E-03
Translation	73	27	17	4.15E-03
Cell cycle checkpoint	19	11	6	5.13E-03
Protein folding	90	26	30	6.68E-03
DNA damage checkpoint	11	6	3	2.04E-02
Proteolysis	219	89	65	5.44E-02
Glycolysis	23	3	7	5.46E-02
DNA repair	79	42	13	7.52E-02
Apoptosis	199	81	64	8.56E-02
Insulin secretion	4	2	1	0.179
Cytoskeleton organisation and biogenesis	168	73	60	0.329
Ubiquitin cycle	177	79	50	0.815
Glucose homeostasis	5	4	2	0.868
Cell growth and/or maintenance	39	18	15	0.901
Endocytosis	71	34	31	0.951
Phosphorylation/ dephosphorylation	293	126	95	0.979
Insulin receptor signalling pathway	12	5	4	0.995
Angiogenesis	40	12	18	0.996
Transport	875	362	286	0.996
Transcription	655	284	199	0.999
Cell adhesion	157	47	57	1
Signalling/signal transduction	685	280	233	1

4.2.8 Activation of MycER^{TAM} promoted cell cycle entry in pancreatic β-cells and suprabasal keratinocytes

As discussed in Section 1.1.4, one of the key functions of the Myc proto-oncoprotein is promotion of cell cycle progression, particularly G_1/S-phase transition (Amati *et al.*, 1998; Dang, 1999; Eilers, 1999). The role of MycER^{TAM} activation in the promotion of G_1/S transition in the skin and the pancreas has been previously noted (Pelengaris *et al.*, 1999; Pelengaris *et al.*, 2002b; Lawlor *et al.*, 2006). Here we show that this previously identified phenotype is identified in the gene-expression analysis for both the skin and the pancreas.

4.2.8.1 MycER^{TAM} activation-induced cell cycle entry of pancreatic β-cells

The effect of Myc on the cell cycle at the transcriptional level was apparent from the number of key cell cycle regulatory genes whose expression was altered upon Myc activation after 4 to 8 hours, when Myc transactivation was maximal (Wu *et al.*, 1999). Of the 2,482 probe sets (2,032 genes) found to change significantly upon MycER^{TAM} activation for the pancreatic β-cells within the first 8 hours, 213 probe sets (171 genes) were designated as relating to cell cycle and proliferation by GO classification in GS-GX. Of these, 116 probe sets (88 genes) showed an increase in expression (Supplementary Table 1) and 101 probe sets (88 genes) showed a decrease in expression (Supplementary Table 2). Note that here, and elsewhere, there appears to be some discrepency in the number of detected genes between the up- and down-regulated gene lists. However, this is due to the presence of genes whose expression is detected as up-regulated at one of the early time points, but down-regulated at the other (or vice-versa). These genes will therefore be included in both the up-regulated and down-regulated gene lists. Genes of interest described in this section are shown in Table 4.2.2, and are coloured in relation to their expression (≥ 2-fold, red; ≤ 2-fold, blue).

Cyclin proteins and their associated kinases, key regulators of cell cycle progression, were well represented within this group. Cyclin genes *ccnd1* (Figure

4.2.15), *ccnd2* (Figure 4.2.16) and *ccne2* (Figure 4.2.17), whose products are necessary for G_1/S-phase transition in the cell cycle, were significantly up-regulated greater than 2-fold within 4 hours of Myc activation, indicating a direct transcriptional role for Myc in promoting G_1/S-phase progression. Cyclin genes *ccna2* (Figure 4.2.18), *ccnb1* (Figure 4.2.19) and *ccne1*, whose products are involved in later G_1/S-phase and G_2/M-phase cell cycle events, were up-regulated greater than 3-fold subsequently at 8 hours, with highly significant (contrast p-value ≤ 0.01) changes at almost all subsequent time points. The Myc target gene database [3] maintains up-to-date information regarding putative Myc-target genes. Of these cyclin genes, *ccna2* (Jansen-Durr *et al.*, 1993) and *ccnd1* (Philipp *et al.*, 1994; Guo *et al.*, 2000) have been previously designated as putative Myc targets through high-throughput screening, and *ccnb1* (Yin *et al.*, 2001) and *ccnd2* (Bouchard *et al.*, 1999; Perez-Roger *et al.*, 1999; Bouchard *et al.*, 2001) have been previously confirmed as Myc targets using ChIP analysis.

The gene for the Cyclin D2-related *cdk4*, also a previously characterised direct Myc target gene (Hermeking *et al.*, 2000), showed a significant increase in expression of 2-fold after 4 hours of $MycER^{TAM}$ activation, with a significant > 6-fold increase detected subsequently at 16 hours (contrast p-value ≤ 0.01). This was detected for three independent probe sets on the array designed to query the *cdk4* transcript. Myc-induced cell cycle progression in β-cells was further aided by > 3-fold down-regulation at 8 hours of another known Myc target gene, the CDKI *cdkn1b* ($p27^{Kip1}$), which inhibits G_1/S-phase transition by association with the Cyclin E-Cdk2 complex (Yang *et al.*, 2001) (Figure 4.2.22). Also, the expression of *cks2*, a Myc target gene whose product is involved in degradation of $p27^{Kip1}$, was found to increase significantly from 8 hours following $MycER^{TAM}$ activation (contrast p-values ≤ 0.01 for all subsequent time points), with a particularly large increase in expression of > 12-fold detected at the 16 hour time point. These data fit well with the role for Myc in early cell cycle progression described in Section 1.1.4.

[3] http://www.myc-cancer-gene.org/site/mycTargetDB.asp

Interestingly, the *cdc2a* gene, whose product Cdk1 is essential for mammalian cell division (Th'ng *et al.*, 1990; Itzhaki *et al.*, 1997), was also found to be highly up-regulated; 3-fold at 8-hours (contrast p-value ≤ 0.05), 12-fold at 16 hours (contrast p-value ≤ 0.05), and 3-fold at 32 hours (contrast p-value ≤ 0.05; see Figure 4.2.20). Cdk1 has been found to substitute for other CDKs to drive cell cycle progression (Santamaria *et al.*, 2007), and is particularly associated with Cdk4 in G_1/S-phase progression (reviewed in Kaldis and Aleem, 2005). This may indicate a significant role for Cdk1 in the promotion of cell cycle progression following $MycER^{TAM}$ activation in β-cells. Alternatively, it has been shown that premature activation of Cdk1 can lead to mitotic catastrophe in G_2/M-phase and apoptosis (upstream of p53-induced MOMP) in neurons (Castedo *et al.*, 2002). Given that this CDK was detected at later time points, this may indicate a possible role for Cdk1 in the $MycER^{TAM}$-induced apoptosis pathways.

In addition to this, the CDKI *cdkn2c* ($p18^{Ink4c}$), which inhibits G_1/S-phase transition via interactions with Cdk4 (Serrano *et al.*, 1993) and Cdk6 (Guan *et al.*, 1994), was down-regulated 3-fold at 4 hours. However, by 16 hours the expression of *cdkn2c* ($p18^{Ink4c}$) had risen dramatically by 6-fold . In addition, the CDKI *cdkn1a* ($p21^{Cip1}$) – a downstream target of the tumour suppressor p53 – was up-regulated 2-fold at 8 hours (Figure 4.2.23), together with significant changes in expression less than 2-fold also detected at 4 hours and 8 hours (contrast p-value $\leq$ 0.01). These results could indicate that cell cycle arrest is associated with promotion of apoptosis by $MycER^{TAM}$, or that p53-induced *cdkn1a* expression results from activation of the DNA damage response pathway in relation to apoptosis, as discussed in Section 4.2.9.1.

Table 4.2.2: Genes relating to cell cycle showing significant change in expression following activation of MycER^{TAM} (p-value ≤ 0.05 for 4OHT term) in pancreatic β-cells. Red = ≥ 2-fold up-regulation; Blue = ≥ 2-fold down-regulation. '*' = t-test p-value ≤ 0.05, '' = t-test p-value ≤ 0.01.**

		Fold change from control					
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**	**Biological role**
ccna2	X75483	0.91 **	5.41 *	14.48 **	3.11 **	0.01087	G1/S and G2/M cell cycle
ccnb1	NM_007629	1.07	3.37 **	11.07 **	3.49 **	0.00214	G2/M cell cycle
ccnd1	NM_007631	2.02 **	2.25 *	0.84 *	1.93	0.00163	G1/S cell cycle
ccnd1	NM_007631	1.92 **	3.93 *	1.19 *	3.21	0.01484	G1/S cell cycle
ccnd1	NM_007631	3.59 **	4.3 *	1.68 *	3.41	0.00387	G1/S cell cycle
ccnd2	NM_009829	1.68 *	0.72	2.18	3.01 *	0.02235	G1/S cell cycle
ccnd2	NM_009829	2.3 *	2.02	1.3	2.29 *	0.02619	G1/S cell cycle
ccnd2	AK007904	2.05 *	0.67	1.99	2.52 *	0.01481	G1/S cell cycle
ccne1	NM_007633	1.8 **	3.35 **	8.21 **	1.82 **	0.00128	G1/S cell cycle
ccne1	BB293079	1.82 **	7.54 **	1.86 **	2.13 **	2.73E-05	G1/S cell cycle
ccne2	AF091432	3.04 **	8.16 **	7.78 **	8.86 **	0.00087	G1/S cell cycle
cdc2a	NM_007659	1.06	4.54 **	14.85 **	3.45 **	0.00419	G1/S and G2/M cell cycle
cdk2	AV303171	1.075	1.331	0.912	0.902	NA	G1/S and G2/M cell cycle
cdk2	NM_016756	1.024	0.846	1.67	1.225	NA	G1/S and G2/M cell cycle
cdk4	NM_009870	2.24	0.94	10.97 **	1.17	0.04959	G1/S cell cycle
cdk4	NM_009870	2.03	0.74	6.52 **	1.63	0.02607	G1/S cell cycle
cdk4	NM_009870	2.18	0.9	6.34 **	1.38	0.02168	G1/S cell cycle
cdkn1a	AK007630	1.56 **	2.12 **	1.62 **	1.49	0.00501	Cell cycle arrest
cdkn1b	NM_009875	1.23	0.29 **	0.76 *	0.96 **	0.01016	Cell cycle arrest
cdkn2c	BC027026	0.32 **	1.15	7.05 **	0.56 *	0.02466	Cell cycle arrest
cks2	NM_025415	0.91	3.26 **	13.58 **	4.86 **	0.004375	G1/S cell cycle
cks2	NM_025415	0.96	2.86 **	12.67 **	2.33 **	0.011831	G1/S cell cycle

4.2.8.2 MycER^{TAM} activation-induced cell cycle entry of suprabasal keratinocytes

The transcriptional role of Myc on the cell cycle was less pronounced for suprabasal keratinocytes when compared to β-cells. Of the 1,821 probe sets (1,617 genes) found to change significantly upon MycER^{TAM} activation within the first 8 hours for the suprabasal keratinocytes, 144 probe sets (129 genes) were designated as relating to cell cycle and proliferation by GO classifications. Of these, 73 probe sets (68 genes) showed an increase in expression (Supplementary Table 3), and 74 probe sets (65 genes) showed a decrease in expression (Supplementary Table 4). Several of the genes defined in these lists are more commonly classified as relating to angiogenesis or cell survival, so subheadings have been used to make this clear. Genes of interest described in this section are shown in Table 4.2.3, and are coloured in relation to their expression (≥ 2-fold, red; ≤ 2-fold, blue).

Cell Cycle Genes

As with the pancreas, the cyclin gene *ccnd2* (Figure 4.2.16) was up-regulated greater than 2-fold at 8 hours. No significant change was detected at 4 hours, and by 16 hours, expression levels had returned to normal. However, a 2-fold increase was once again detected at 32 hours, suggesting peaks in expression at 8 hrs and 32 hrs. This pattern of expression in the suprabasal keratinocytes was detected for many genes, and is likely due to the delayed activity of MycER^{TAM} following topical application, as described in Section 4.2.6. This indicates that there may be a delay of roughly 8 hours following each daily dose of 4OHT. However, this expression profile was detected for only one of three probe sets specific for the *ccnd2* gene, and the time point specific contrast p-values were not found to be significant.

The cyclin D3 gene *ccnd3* also showed a significant increase in expression of 2-fold at 8 hours (contrast p-value ≤ 0.01), with significant but low magnitude changes also detected at 4 hours and 16 hours (contrast p-values ≤ 0.01). However, later cyclin genes, such as *ccna* and *ccne*, showed no significant change

within the suprabasal keratinocytes. Also, the *ccnb1* gene, whose product is involved predominantly in later cell cycle events, was shown to have a significant change within the skin. However in contrast to the pancreas, the detected change was a significant 2-fold down-regulation at 8 hours (contrast p-values ≤ 0.05; Figure 4.2.19).

The CDK gene *cdk4*, whose product forms a complex with Cyclin D2 to promote cell cycle entry, was highly up-regulated at 8 hours, with a fold change of almost 12 (Figure 4.2.21), although this change was not seen at the later time points. This expression profile was seen across several probe sets suggesting this accurately represented the change in gene-expression. Intriguingly, however, this was not confirmed by qRT-PCR (Section 4.2.12.4). Also, as with the pancreas, no change was detected for the Cyclin E-associated CDK gene, *cdk2*. However *cdk7*, which has a role in both activating cyclin complexes and regulating transcription, was also 2-fold up-regulated at 8 hours (contrast p-values ≤ 0.01), although no change was detected at the later time points.

As with the pancreas, the CDKI gene *cdkn1b* ($p27^{Kip1}$) (Figure 4.2.22) was down-regulated greater than 2-fold throughout the time course (with a < 2-fold change in expression at 8 hours, contrast p-value ≤ 0.01), although a lower change in expression was detected using qRT-PCR (Section 4.2.12.6). This may indicate a loss in cell cycle inhibitory proteins during $MycER^{TAM}$-induced proliferation. Oddly, the loss of expression of *cdc25a*, which is required for G_1/S-phase cell cycle transition by activating Cdk1 (Galaktionov and Beach, 1991), seemed incompatible with the proliferative response seen in the cells.

A significant loss in expression, particularly at 8 hours following $MycER^{TAM}$ activation (contrast p-value ≤ 0.01), was detected for the epidermal growth factor gene *egf* and its receptor *egfr*, whose products are involved in stimulating epidermal growth (Cohen and Elliott, 1963), although the relevance of the *egf* pathway to $MycER^{TAM}$-promotion of proliferation is not clear.

Angiogenesis Genes

The loss of expression at 8 hours seen for the epidermal growth factor *egf* and its receptor *egfr*, whose products are involved in stimulating epidermal growth (Cohen and Elliott, 1963), seemed incompatible to the observed phenotype. The placental growth factor gene, *pgf*, showed a marked increase throughout the time course, increasing in expression by 2-fold at 8 hours and 16 hours (contrast p-values ≤ 0.01), and 8-fold at 32 hours (contrast p-value ≤ 0.01). Pgf is a member of the vascular endothelial growth factor (VEGF) family, and has been shown to result in increased numbers, branching and size of dermal blood vessels following over-expression in basal keratinocytes of adult mice (Odorisio *et al.*, 2002). This may indicate a role in the development of neovasculature seen following MycER^{TAM} activation in the suprabasal keratinocytes (Pelengaris *et al.*, 1999).

vegfc, a further member of the VEGF family, showed down-regulation of 2-fold at 8 hours, although this change was not classed as significant from the contrast p-values and was not seen at any other time point. Further VEGF genes, *vegfa* and *vegfb*, were not classed as showing significant change upon MycER^{TAM} activation by *Envisage*, although *vegfa* did show a significant 2-fold increase in expression at 8 hours (contrast p-value ≤ 0.05). These results may indicate a transcriptional response upon MycER^{TAM} activation for genes relating to neovascular growth. However, given that prominent angiogenesis is not detected in the skin until 3-4 days following MycER^{TAM} activation (Pelengaris *et al.*, 1999), it is likely that the short time course considered here is too early to identify a transcriptional response in genes relating to vascularisation.

DNA Damage Genes

The checkpoint kinase gene *chk1*, whose product is involved in Atr-related DNA damage control, was down-regulated by greater than 2-fold at 8 hours. However, this change was seen for only one of the three probe sets specific for the *chk1* gene, and was not detected as significant in the contrasts. In general, the change in expression for *chk1* was limited, which indicates that DNA damage checkpoints were not activated within the suprabasal keratinocytes upon MycER^{TAM}

activation, allowing proliferation to continue unchecked. As shown below in Section 4.2.9, this is in contrast to the DNA damage response seen in the pancreas.

Cell Survival Genes

akt1 and *akt2*, members of the Akt survival pathway, showed greater than 2-fold change in expression at 8 hours and 32 hours (contrast p-values ≤ 0.01) suggesting a possible role for the Akt pathway in promoting survival in the skin, however this result was not confirmed by using qRT-PCR (Figure 4.2.25). This is discussed in greater detail below in Section 4.2.9.2.

Summary

Activation of MycER^{TAM} in *inv-mycER^{TAM}* mice results in entry of suprabasal keratinocytes into the cell cycle (Pelengaris *et al.*, 1999), and this phenotype was confirmed through immunofluorescence staining with the cell cycle marker Ki67 (Figure 4.2.1). However, many genes that would be expected to be involved in MycER^{TAM}-regulated promotion of suprabasal keratinocyte proliferation (e.g. *ccnd2*, *cdk4*, and *cdkn1b*) did not appear to show expected changes in expression that were seen in the pancreas (Section 4.2.8.1).

As previously described, turnover of cells in the pancreas is far slower than that of the skin. Whilst it is important for the skin to be in a constant state of dynamic equilibrium between cell proliferation and cell loss from the skin surface, β-cells must remain more stable in order to maintain the delicate balance of Insulin in the blood. Therefore, whilst islet cells from VT control mice are largely quiescent, the basal stem cells of the skin epidermis are undergoing cell cycle progression during normal homeostasis (Section 1.2). Given that the number of suprabasal cells induced to proliferate within 32 hrs of MycER^{TAM} activation was small, changes in expression of cell cycle genes likely show less significance than for the islets.

Table 4.2.3: Genes relating to cell cycle showing significant change in expression following activation of MycER^{TAM} (p-value ≤ 0.05 for 4OHT term) in suprabasal keratinocytes. Red = ≥ 2-fold up-regulation; Blue = ≥ 2-fold down-regulation. '*' = t-test p-value ≤ 0.05, '' = t-test p-value ≤ 0.01.**

		Fold change from control					
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**	**Biological role**
akt1	NM_009652	0.79	3.33 **	1.07	2.72 **	0.003781	Cell survival
akt2	NM_007434	1.24	2.3 **	1.1	2.37 **	0.007853	Cell survival
ccnb1	NM_007629	0.69	0.44 *	1.36	0.66	0.002143	G2/M cell cycle
ccnd2	NM_009829	1.32	1.52	1.01	1.42	0.022348	G1/S cell cycle
ccnd2	NM_009829	0.89	0.9	1.3	0.87	0.026191	G1/S cell cycle
ccnd2	AK007904	1.63	3	1.03	2.4	0.014806	G1/S cell cycle
ccnd3	NM_007632	0.84 **	2.64 **	0.85 **	3.86	0.000512	G1/S cell cycle
cdc25a	C76119	1.19	0.44	1.08	0.81	0.00184	Activation of Cdk1
cdk2	AV303171	1.136	1.05	1.022	0.926	NA	G1/S cell cycle
cdk2	NM_016756	0.885	0.882	1.032	0.945	NA	G1/S cell cycle
cdk4	NM_009870	1.83	12.73 **	0.88	1.93	0.049585	G1/S cell cycle
cdk4	NM_009870	1.5	11.77 **	1.09	1.89	0.026067	G1/S cell cycle
cdk4	NM_009870	1.64	11.53 **	0.91	1.63	0.021676	G1/S cell cycle
cdk7	U11822	0.91	2.24 **	1.05	1.57 *	0.002113	Cell cycle regulation
cdkn1a	AK007630	1.01	1.4 *	1.1	0.91	0.005006	Cell cycle arrest
cdkn1b	NM_009875	0.2	0.51 **	0.5	1.06 *	0.010159	Cell cycle arrest
chk1	BB298208	0.88	1	1.36	1.28	0.004048	DNA damage checkpoint
chk1	C85740	0.83	0.43	1.45	1.15	0.004993	DNA damage checkpoint
chk1	NM_007691	0.7	0.8	1.47	1.4	0.005829	DNA damage checkpoint
egf	NM_010113	1.64 **	0.47 **	0.93	0.91	0.027391	Growth factor
egfr	AV369812	1.15	0.4 **	1.37 *	0.49 **	0.000479	Growth factor receptor
pgf	NM_008827	1.1	2.33 **	2.23 **	8.19 **	0.031444	Angiogenesis
vegfa	U50279	0.785	1.998 *	1.16	1.026	NA	Angiogenesis
vegfa	NM_009505	1.086	2.139 *	1.155	0.734	NA	Angiogenesis
vegfb	U48800	0.553 *	1.796 *	0.908	1.287	NA	Angiogenesis
vegfc	BB089170	0.79 **	0.39	0.85	0.79	0.016314	Angiogenesis

4.2.9 Activation of MycER^{TAM} *in vivo* leads to up-regulation of apoptotic death pathways in pancreatic β-cells but not in suprabasal keratinocytes

In stark contrast to the proliferation seen in the suprabasal keratinocytes, the overwhelming phenotypic response to activation of transgenic MycER^{TAM} in pancreatic β-cells is apoptosis (Pelengaris *et al.*, 2002b). In this section, the difference in transcriptional regulation of genes relating to apoptosis and cell survival are discussed.

4.2.9.1 MycER^{TAM} activation-induced apoptosis of pancreatic β-cells

Of the 2,482 probe-sets found to change significantly upon Myc activation for the pancreas within the first 4 to 8 hours, 92 (representing 79 unique genes) were designated as relating to cell death and apoptosis by GO classification. Of these, 42 probe sets (32 genes) showed an increase in expression (Supplementary Table 5), and 50 probe sets (47 genes) showed a decrease in expression (Supplementary Table 6). Genes of interest described in this section are shown in Table 4.2.4, and are coloured in relation to their expression (≥ 2-fold, red; ≤ 2-fold, blue).

Early activation of key regulators of apoptosis featured prominently in these data. The tumour suppressor *cdkn2a*, which encodes for the CDKI p16^{Ink4a} and the alternative reading frame tumour suppressor p19^{Arf} (p14^{Arf} in humans), was significantly up-regulated 2-fold at 4 hours and remained at an elevated level throughout the time course (Figure 4.2.26). This indicated a possible role for the p19^{Arf}/p53/Mdm2 tumour suppressor pathway in Myc-mediated apoptosis.

The role of Myc in this pathway has previously been identified (Zindy *et al.*, 1998; Eischen *et al.*, 1999). Also, in the study of Lawlor *et al.* (2006), the RM double transgenic model for Myc deregulation was used to identify genes whose expression was altered immediately upon Myc activation and subsequently reversed upon deactivation of MycER^{TAM}, thereby suggesting a key role in Myc-induced tumour maintenance. No change was observed in the study for the gene

cdkn2a ($p19^{Arf}$), nor in the p53-target genes *cdkn1a* ($p21^{Cip1}$) and *mdm2*, indicating that concomitant over-expression of Bcl_{XL} may protect against apoptosis by limiting Myc-induced expression of *cdkn2a* ($p19^{Arf}$), which has also been noted by (Nilsson and Cleveland, 2003).

It is also worth noting the work of Finch *et al.* (2006), who showed no role for $p19^{Arf}$ in Myc-induced apoptosis in β-cells, as loss of $p19^{Arf}$ in the $MycER^{TAM}$ transgenic model resulted in mainly increased proliferation, not suppression of apoptosis. However, the role of $p19^{Arf}$ in defending against aberrant oncogenic Myc-induced hyper-proliferation may be related to cell cycle arrest, and not directly to apoptosis pathways.

Expression of genes involved in the DNA damage response appeared to play a large part in Myc-induced apoptosis in this model (Supplementary Table 7). A large increase in expression was detected after 8 hours of Myc activation for *rad51* and *h2afx*, whose protein products are involved in homologous recombination and repair of DNA. Also, significant up-regulation of *hus1* and *rad1* – whose products form the 9-1-1 DNA damage sensing machinery with Rad9 – indicated that oncogenic stress through deregulation of Myc resulted in the induction of DNA double strand breaks.

The gene for the DNA-damage mediator Atr was also found to be up-regulated by 2-fold throughout the time course from 4 hours (Figure 4.2.27), and the associated checkpoint kinases *chk1* and *chk2* were up-regulated 2-fold from 8 hours (contrast p-values ≤ 0.05 at 8 hours, ≤ 0.01 at 16 and 32 hours; Figure 4.2.28). The gene for the double-strand break-related DNA damage mediator Atm showed no significant change in expression, although it has been shown that Atm plays a significant role in Myc-induced apoptosis in lymphomagenesis in mice (Maclean *et al.*, 2007). The genes *cdkn2a* ($p19^{Arf}$) (Zindy *et al.*, 1998; Dang, 1999) and *atr* (Schlosser *et al.*, 2003) are previously categorised Myc target genes. The checkpoint kinase genes *chk1* and *chk2* have not been previously classified as Myc target genes, indicating that the observed changes in expression may be downstream of Atr.

The anti-apoptotic function of Bcl_{XL} seen in RM double transgenic mice (Pelengaris *et al.*, 2002b) indicates that $MycER^{TAM}$-induced apoptosis is related to Bax/Bak-mediated intrinsic mitochondrial pathway. Activation of this intrinsic apoptotic pathway was evident at the transcriptional level by 2-fold increased expression of *bax* and the somatic Cytochrome c gene, *cycs* (Figure 4.2.29) at 16 hours after $MycER^{TAM}$ activation, together with lower significant fold-changes at all other time points. Both *bax* (Mitchell *et al.*, 2000; Fernandez *et al.*, 2003) and *cycs* (Guo *et al.*, 2000; Morrish *et al.*, 2003) have been previously shown to be putative direct Myc targets due to the presence of non-canonical E-box Myc-Max binding sites, and association of Myc with the *bax* promoter has been demonstrated through ChIP (Fernandez *et al.*, 2003). However, given that the increased expression of these genes is detected at 16 hours, it is unlikely that these are direct $MycER^{TAM}$ targets in this case, and it is likely that changes in expression occur downstream of $MycER^{TAM}$ activation.

$MycER^{TAM}$-mediated expression of *bax* may be a direct consequence of p53 stabilisation, also suggested by *cdkn1a* ($p21^{Cip1}$) up-regulation. However, no significant change in expression was detected in other pro-apoptotic p53-target genes such as the Bcl2 family members *bbc3, noxa, bim* and *bid*. This indicates that p53-dependent activation of Bax may occur not through transcriptional regulation of pro-apoptotic Bcl2 members, but through direct associations of p53 within the cytosol (Mihara *et al.*, 2003; Chipuk *et al.*, 2004; Erster and Moll, 2005). The mitochondrial respiratory gene for Endonuclease G, *endog* (Figure 4.2.30), was also found to be up-regulated after only 4 hours of $MycER^{TAM}$ activation, further indicating the promotion of MOMP in response to $MycER^{TAM}$ activation. This short time scale also supports the early increase in perinuclear Caspase 3 staining shown in Figure 4.2.1a.

As well as the intrinsic mitochondrial apoptotic pathway, Myc activation *in vitro* sensitises cells to extrinsic signals through the TNF death receptor Fas (Janicke *et al.*, 1994; Klefstrom *et al.*, 1994; Hueber *et al.*, 1997). Expression of the *fas* gene was increased dramatically (~ 6-fold) after only 4 hours of Myc activation in the pancreas, and remained highly expressed throughout the 32 hour time course. This

indicated a possible direct transcriptional role for Myc in the extrinsic Fas pathway, although these changes were not detected as significant through the use of contrast t-tests. Also *flip* (also known as *cflar*), whose product acts to inhibit association of Fadd with the Fas ligand/receptor complex, was down-regulated 2-fold at 32 hours. However, the large change in expression seen for *fas* in the microarray study was not confirmed using qRT-PCR (Figure 4.2.31). As described in Section 4.2.12.10, this was due to a problem relating to probe annotations.

The expression of both *cdc2a* and *birc5* showed almost identical profiles of up-regulation upon activation of MycER^{TAM} – 3-fold at 8-hours, 12-fold at 16 hours, and 3-fold at 32 hours. Under normal conditions, Cyclin B-Cdk1 complexes phosphorylate and activate the IAP Survivin (*birc5*), which inhibits apoptosis and allows proliferation to continue. However, given that apoptosis is not inhibited in this case, it is likely that the role of these genes is in proliferation, and not survival. Survivin is essential for normal mitotic regulation in human cell lines due to its role in chromatid segregation and microtubule assembly in late mitosis (Yang *et al.*, 2004), and expression of *birc5* is often seen in cells undergoing rapid proliferation (Ambrosini *et al.*, 1997). The high increase in expression of both of these genes at the later 16 hour time point may therefore be linked to the predominant role of *birc5* in the G_2/M transition phase of the cell cycle (Li *et al.*, 1998a). Also, premature activation of Cdk1, prior to G_2/M-phase cell cycle, has previously been shown to promote mitotic catastrophe and apoptosis, and this has been tied to the effects of DNA damage (Castedo *et al.*, 2002). The increase in expression detected within 8 hours of MycER^{TAM}-activation may therefore indicate a possible role for this CDK in oncogenic Myc-induced apoptosis.

Together, these data suggest a cooperative role for a variety of Myc targets in promoting apoptosis. We propose a model showing that sustained Myc activation results in DNA damage (as previously reported), which correlates with up-regulation of genes involved in the DNA-damage response. This leads to activation of p53 and Bax-mediated release of Cytochrome c from the mitochondrion, resulting in apoptosis. Given the lack of transcriptional changes

seen in p53-regulated pro-apoptotic members of the Bcl2 superfamily, it is possible that this occurred through direct activation of Bax, or inhibition of Bcl_{XL} and Bcl2, by accumulated active p53 in the cytosol. However, this cannot be shown from the microarray data, and further proteomics analysis would be required to confirm or refute this.

Activation of MOMP was further evidenced by an increase in expression of the pro-apoptotic mitochondrial factors *cycs* and e*ndog*, which may indicate replenishment of proteins lost from the mitochondria during apoptotic signalling. Release of Cytochrome c from the mitochondria leads to a caspase activation cascade, resulting in cell death. Increased expression of the tumour suppressor *cdkn2a* ($p19^{Arf}$) was also detected, which may have been involved in stabilisation of the p53 tumour suppressor by inhibition of Mdm2, or in promoting cell cycle arrest prior to apoptosis together with up-regulation of the p53-target gene *cdkn1a* ($p21^{Cip1}$).

Table 4.2.4: Genes relating to apoptosis showing significant change in expression following activation of MycER^{TAM} (p-value ≤ 0.05 for 4OHT term) in pancreatic β-cells. Red = ≥ 2-fold up-regulation; Blue = ≥ 2-fold down-regulation. '*' = t-test p-value ≤ 0.05, '' = t-test p-value ≤ 0.01.**

		Fold change from control					
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**	**Biological role**
atm	AK021102	0.787	0.545	1.173	0.669	NA	DNA damage response
atm	NM_007499	0.84	0.442 **	1.038	1.11	NA	DNA damage response
atr	AF236887	2.31	4.04 **	2.57	2.87	4.44E-05	DNA damage response
bax	BC018228	1.17	1.7 **	2.55 **	1.83 **	0.00104	Induction of MOMP
bbc3	AW489168	1.401	0.701	1.065	1.154	NA	Induction of MOMP
bid	AV376592	1.165	0.834	1.137	1.14	NA	Induction of MOMP
bid	NM_007544	1.097	1.063	1.006	1.03	NA	Induction of MOMP
bid	NM_007544	0.986	0.852	1.163	1.283	NA	Induction of MOMP
birc5	BC004702	0.542 *	2.034 **	17.28 **	1.886 **	NA	Inhibitor of apoptosis
cdc2a	NM_007659	1.06	4.54 **	14.85 **	3.45 **	0.00419	Cell cycle
cdkn1a	AK007630	1.56 **	2.12 **	1.62 **	1.49	0.00501	Cell cycle arrest
cdkn2a	NM_009877	2.54 **	2 **	1.49 **	2.7 **	0.00026	cell cycle arrest
cflar	AK020765	0.81 *	0.79 *	1.2	0.49 **	0.00268	Inhibitor of Fas signalling
cflar	BE284491	0.83 *	0.88 *	0.55	0.43 **	0.0443	Inhibitor of Fas signalling
cflar	BE284491	1.02 *	0.98 *	1	0.94 **	0.00663	Inhibitor of Fas signalling
chk1	NM_007691	1.52	2.19 *	3.12 **	1.65 **	0.00405	DNA damage response
chk1	C85740	2.05	5.34 *	6.01 **	3.86 **	0.00499	DNA damage response
chk1	BB298208	1.15	3.84 *	2.76 **	2.05 **	0.00583	DNA damage response
chk2	NM_016681	1.22	2.17 **	3.81 **	2.29 **	0.01976	DNA damage response
cycs	NM_007808	1.15 *	2.04 **	1.64 **	2.06 **	0.04759	Activation of Caspase 9
cycs	NM_007808	0.96 *	1.02 **	2.72 **	2.42 **	0.00164	Activation of Caspase 9
endog	AV104666	2.94 **	3.02 **	1.48 **	1.85 **	1.19E-05	DNA fragmentation
endog	NM_007931	2.66 **	3.12 **	2.27 **	2.07 **	7.61E-06	DNA fragmentation
fas	BG976607	2.74	1.87	2.58	2.58	0.00014	Death signal receptor
fas	BG976607	9.9	13.59	9.74	11.38	0.00632	Death signal receptor
fas	BG976607	4	7.72	2.13	4.96	0.00112	Death signal receptor
fas	BG976607	6.57	6.48	3.57	14.03	0.00221	Death signal receptor
h2afx	NM_010436	1.16	2.42 **	5.22 **	1.4 **	0.00117	DNA damage repair
hus1	AF076845	1.45	1.1	2.35 **	1.3	0.01463	DNA damage marker
hus1	NM_008316	0.93	2.13	2.07 **	1.67	0.01852	DNA damage marker
pmaip1	NM_021451	1.13	1.15	1	0.88	0.02258	Induction of MOMP
rad1	NM_011232	2.08 **	2.87 **	2.2 *	1.76	0.00377	DNA damage marker

4.2.9.2 Survival of suprabasal keratinocytes following MycER^{TAM} activation

The primary phenotype upon activation of MycER^{TAM} in the suprabasal keratinocytes is proliferation (Pelengaris *et al.*, 1999), as suprabasal cells that have begun a process of terminal differentiation re-enter the cell cycle (Section 1.2.2). Whilst conventional apoptosis does not feature in the maintenance of epidermis homeostasis, it has been suggested that terminal differentiation of skin keratinocytes may itself act to control against aberrant growth, since affected cells will ultimately be shed and removed from the surrounding micro-environment (Jensen and Watt, 2006).

Of the 1,821 probe sets (1,617 genes) found to change significantly upon MycER^{TAM} activation within the first 8 hours for the suprabasal keratinocytes, 73 (66 genes) were found to relate to apoptosis and cell death by GO classification. Of these, 42 probe sets (37 genes) showed an increase in expression (Supplementary Table 8), and 31 probe sets (29 genes) showed a decrease in expression (Supplementary Table 9). Genes of interest described in this section are shown in Table 4.2.5, and are coloured in relation to their expression ($\geq$ 2-fold, red; $\leq$ 2-fold, blue).

Of particular interest were the two probe sets for the survival factor gene *igf1* (Figure 4.2.32), whose product has been shown to inhibit Myc-induced apoptosis *in vitro* by blocking Cytochrome c release from the mitochondria through the Akt1 tumour suppressor pathway (Lowe *et al.*, 2004). *igf1* was significantly up-regulated greater than 2-fold by 8 hours and remained up-regulated at the later 32 hour time point for the skin (contrast p-values $\leq$ 0.01), suggesting a possible survival pathway allowing the suprabasal keratinocytes to bypass the Myc-induced apoptosis seen in the β-cells. In conjunction with this, the thymoma viral proto-oncogene *akt1* was significantly up-regulated 3-fold at 8 hours and 2-fold at 32 hours (Figure 4.2.25), and a similar expression profile was also seen for *akt2*, a second member of the Akt protein kinase family. This suggested that Igf1-

mediated survival signalling following MycER^{TAM} activation in the suprabasal keratinocytes occurred through the Akt protein kinase pathway.

However, as can be seen in Figure 4.2.32 and Figure 4.2.25, the change in expression identified using the microarrays was not replicated with qRT-PCR for *igf1* and *akt1*. Whilst a change in expression was detected for Igf1, it was below the 2-fold threshold. However, given the increased sensitivity of the qRT-PCR procedure, this may still represent a significant change in expression. Interestingly, qRT-PCR also identified a significant increase in expression for *igf1* in the β-cells that was not detected in the microarrays, suggesting that increase of *igf1* expression may occur in response to MycER^{TAM} for both tissues. Validation of the results seen for *akt1* using qRT-PCR identified no change in expression, suggesting that this may represent a false positive result. This is discussed in more detail in Section 4.2.12.1.

In comparison to results seen for the pancreatic β-cells, no expression change was detected for genes involved in the DNA damage response, such as *atr, chk1* or *chk2*. However, activation of such proteins (e.g. through phosphorylation events) within MycER^{TAM}-activated keratinocytes would need to be investigated further before ruling out the DNA damage pathway. Interestingly, the growth arrest and DNA-damage-inducible 45 gamma gene *gadd45g*, whose product is involved in the G_2/M DNA damage checkpoint, showed a highly significant 4-fold change in expression at 4 hours, and remained significantly up-regulated throughout the time course (contrast p-values ≤ 0.01). The pro-apoptotic Bcl2 family member *pmaip1* (Noxa) showed a decrease in expression of 2-fold at 8 hours. Oddly, a 2-fold loss of expression was also seen for the inhibitor of apoptosis protein Birc4 within 8 hours of MycER^{TAM} activation.

Several members of the extrinsic apoptosis pathway were also evident in this list, including a probe set representing the *fas* TNF death receptor gene showing a 5-fold up-regulation at 8 hours. However, this was not maintained throughout the time course, and a 2-fold down-regulation was detected at 32 hours for this and two other probe sets. However, as described in Section 4.2.12.10, it was later

discovered that these probe sets did not in fact relate to the Fas receptor, but that an error in probe annotation had occurred.

Also seen were several members of the TNF superfamily of apoptosis-inducing receptors. *tnfrsf12a*, which has been found to be involved in inducing both apoptosis and angiogenesis (Wiley *et al.*, 2001; Wiley and Winkles, 2003), showed a significant increase in expression of 3-fold at 8 hours. *tnfrsf4*, whose product has been implicated in promoting survival through induction of Bcl2 and Bcl_{XL} expression in CD4 T cells (Rogers *et al.*, 2001), similarly showed an increase in expression of 2-fold at 8 hours.

It appears from these data that the response in apoptosis-related genes for the suprabasal keratinocytes was quite different from that of the β-cells. In particular, in contrast to the skin, various members of the DNA damage checkpoint pathway were up-regulated in the β-cells following $MycER^{TAM}$ activation. Induction of survival related genes *igf1*, *akt1* and *akt2* suggested a role for the Igf1-mediated Akt pathway in keratinocyte survival, although it is not clear whether apoptosis is predominantly inhibited via the Igf1 signalling pathways by preventing MOMP at the mitochondria, preventing activation of DNA damage response, or both. Since this analysis focuses on RNA expression levels, it is not possible to conclusively say whether or not DNA damage pathways are active in the skin, however it is likely that damaged DNA accumulates as the tumour progresses, but apoptosis is inhibited environmentally (such as through the Igf1 receptor survival pathways).

Table 4.2.5: Genes relating to apoptosis showing significant change in expression following activation of MycER^{TAM} (p-value ≤ 0.05 for 4OHT term) in suprabasal keratinocytes. Red = ≥ 2-fold up-regulation; Blue = ≥ 2-fold down-regulation. '*' = t-test p-value ≤ 0.05, '' = t-test p-value ≤ 0.01.**

		Fold change from control					
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**	**Biological role**
akt1	NM_009652	0.79	3.33 **	1.07	2.72 **	0.00378	Cellular survival
akt2	NM_007434	1.24	2.3 **	1.1	2.37 **	0.007853	Insulin signalling
atr	AF236887	1.216	1.09	1.5	1.496	NA	DNA damage response
birc4	BF134200	0.51 **	0.34 **	1.12	0.87	0.005291	Inhibitor of caspases
chk1	NM_007691	0.7	0.797	1.464	1.395	NA	DNA damage response
chk2	NM_016681	0.84	0.607 *	1.112	1.37	NA	DNA damage response
fas	BG976607	0.84	1.13	2.12 **	1.37	0.00014	Death signal receptor
fas	BG976607	1.35	0.58	1.52 **	0.45	0.00632	Death signal receptor
fas	BG976607	0.85	5.35	0.86 **	0.49	0.00112	Death signal receptor
fas	BG976607	0.8	1.37	1.15 **	0.38	0.00221	Death signal receptor
gadd45g	AK007410	4.44 **	5.44 **	3.27 **	2.87 **	1.30E-05	DNA damage and growth arrest
igf1	NM_010512	1.09 *	2.5 **	1.61	2.35 **	0.00325	Proliferation and survival
igf1	AF440694	1.8 *	2.31 **	1.13	3.17 **	0.00013	Proliferation and survival
igf1	BG075165	1.04	1.12	2.43	4	0.00084	Proliferation and survival
igf1r	BB446952	1.10	0.52 **	0.848	0.74 *	NA	Proliferation and survival
pmaip1	NM_021451	1.43 *	0.44 **	0.82 *	0.9	0.022582	Activation of Bax
tnfrsf12a	NM_013749	1.14	3.21 **	1.44 **	1.18	0.013343	Apoptosis and angiogenesis receptor
tnfrsf12a	NM_013749	1.22	2.9 **	1.4 **	1.14	0.010767	Apoptosis and angiogenesis receptor
tnfrsf4	NM_011659	1.25	2.24 **	1.94 **	1.04	0.005782	Death signal receptor

4.2.10 Activation of MycER^{TAM} resulted in loss of differentiation markers in pancreatic β-cells and suprabasal keratinocytes

4.2.10.1 MycER^{TAM} activation-induced differentiation of pancreatic β-cells

Activation of Myc is often associated with loss of differentiation of cells and has been found to block terminal differentiation in a variety of cell types (Coppola and Cole, 1986; Maruyama *et al.*, 1987; Freytag, 1988). Ectopic expression of Myc has been shown to repress the expression of the Insulin gene, *ins2* (Kaneto *et al.*, 2002; Laybutt *et al.*, 2002; Pelengaris *et al.*, 2002b). Genes related to β-cell differentiation and function are shown in Table 4.2.6, and are coloured in relation to their expression (≥ 2-fold, red; ≤ 2-fold, blue).

Activation of MycER^{TAM} in the pancreatic β-cells resulted in down-regulation of *ins2* by > 6-fold at 4 hours, indicating a loss in Insulin production within a short time period following Myc-deregulation. However, the expression level of *ins2* subsequently increased dramatically, showing a significant 12-fold increase in expression at 16 hours and 14-fold at 32 hours following MycER^{TAM} activation (Figure 4.2.13A). Since the number of β-cells used for RNA extraction was roughly identical for each sample, this indicates acute increase in the levels of Insulin production within the β-cells in response to continuous MycER^{TAM} activation, and not in response to an increase in β-cell mass.

Although a paradox at first glance, this response may be the result of a positive feedback loop. We have observed a period of hypoglycaemia in mice within the first 24 hours of MycER^{TAM} activation (Figure 4.2.13B), which correlates with an increase in Insulin release into the bloodstream (manuscript in preparation). As our microarray data show an increase in Insulin at the transcript level, this may also contribute to the onset of hypoglycaemia. The experiment described within this thesis was conducted in parallel with a further microarray study, observing the effects of the Glucagon-like peptide 1 (Glp1) analog, Exenatide (Amylin Pharmaceuticals, San Diego, CA; Eli Lilly and Company, Indianapolis, IN), on the diabetic phenotype associated with MycER^{TAM} activation in the pancreatic β-

cells. This branch of the study (which will not be discussed further) considered a further time point at 72 hours following MycER^{TAM} activation. Observation of *ins2* levels at this later time point indicated that this period of high Insulin production is limited, as gene-expression subsequently returned to low levels (> 3-fold down-regulated) indicative of loss of β-cell differentiation. This leads to a time window within the first two days where a balance is struck between an increased rate of Insulin production and the simultaneous loss of cells due to MycER^{TAM}-driven apoptosis.

It has recently been postulated that this period of hypoglycaemia is due to sudden ablation of β-cells, leading to synchronous release of large amounts of excess Insulin into the pancreatic ducts (Cano *et al.*, 2007). However, recent work within the Michael Khan group at the University of Warwick has confirmed that this period of hypoglycaemia is also detected within non-apoptotic RM mice, indicating that it is unlikely to be attributable to β-cell apoptosis (manuscript in preparation). Also, this hypothesis does not account for the observed increase in Insulin production at the transcriptional level. This phenomenon is currently being studied by members of the group.

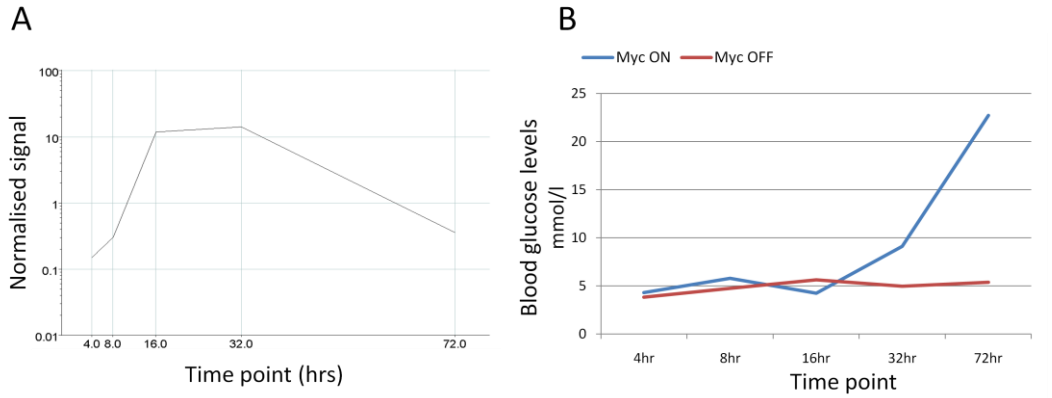

Figure 4.2.13: Activation of MycER^{TAM} in β-cells leads to loss of insulin production and increased blood glucose levels, with a brief window of hypoglycaemia in the first 24 hours

A) Activation of MycER^{TAM} in pancreatic β-cells led to a rapid decrease in expression of Insulin (*ins2*), indicating a loss of differentiation correlating with increased β-cell proliferation. However, by 16 hours following MycER^{TAM} activation, *ins2* expression had risen dramatically to > 12-fold indicating acute increase in Insulin production in β-cells showing deregulated Myc activity. By 72 hours, *ins2* expression had returned to low levels, indicating that this period of increased Insulin production is limited to only a few days, and soon returns to levels indicative of loss of β-cell differentiation. B) This change in expression of Insulin correlates with a period of hypoglycaemia observed in *pIns-mycER^{TAM}* (for example, indicated here at time 16 hrs) and in RM double transgenic mice (data not shown) following MycER^{TAM} activation, following acute release of Insulin into the bloodstream.

Members of the homeodomain transcription factors Pdx1, Pax4, Hb9, Nkx2.2 and Nkx6.1 are essential in pancreatic development (Chakrabarti and Mirmira, 2003). Probe sets for the pancreatic and duodenal homeobox gene *pdx1* (or *ipf1*), whose product activates transcription of the Insulin gene as well as a number of genes involved in Glucose-sensing (Hui and Perfetti, 2002), were also found to show a significant loss in expression at 8 hours following MycER^{TAM} activation (as well as significant loss in expression < 2-fold at all other time points), which correlated with the early reduction seen in Insulin production.

The transcription factor *nkx6.1*, whose product is essential for β-cell differentiation (Sander *et al.*, 2000), also showed significant down-regulation in the early stages of MycER^{TAM} activation, although expression of this gene was shown to increase during later stages, possibly linking with the induced hypoglycaemia. Further Pdx1-regulated genes *slc2a2* (Glut2; previously classed as a putative Myc target gene) and *gck* (Glucokinase), both part of the Glucose-sensing machinery and involved in membrane transport and phosphorylation of Glucose respectively, also followed similar expression profiles, as did the gene for the Glucagon-like peptide, *glp1*, which acts to stimulate Insulin release and block Glucagon release (Nauck *et al.*, 1993; Elahi *et al.*, 1994; Kjems *et al.*, 2003). These data indicated a loss in β-cell differentiation and carbohydrate metabolism function following activation of MycER^{TAM}.

In the study of Gu *et al.* (2004), the transcriptional profile of murine pancreas was categorised at various points during development. A group of 217 mature islet-specific genes were defined, providing markers for fully differentiated β-cells. These 217 genes were represented by 378 probe sets on the MOE 430 Plus 2 GeneChip arrays. Of these probe sets, 84 (60 genes) showed significant change in the pancreas upon activation of MycER^{TAM}. Thirty eight of these probe sets (30 genes) showed a 2-fold decrease within 8 hours of MycER^{TAM} activation (Supplementary Table 10), while only 7 probe sets (6 genes) showed a 2-fold increase (Supplementary Table 11). This indicated that activation of MycER^{TAM} resulted predominantly in the loss of mature β-cell markers, and loss of cell differentiation.

Of the seven probe sets that showed an increase in gene-expression in the present study, four represented genes that were classified as being related to immune response through GO ontologies, including histocompatability two genes *h2-l, h2-aa* and *h2-d1* which were up-regulated greater than 2-fold throughout the time course. Other genes included *pcnt*, which is involved in centrosomal microtubule formation during mitosis; the E1A antagonist *creg1,* which acts to control regulation through the cell cycle (Veal *et al.*, 1998; Flory *et al.*, 2000); and *pcsk2,* which encodes a serine protease that plays a role in the processing of neuroendocrine precursors into mature hormones together with Pcsk1 and Pcsk3 (Seidah and Chretien, 1997; Steiner, 1998). Pcsk2 plays an important role in converting Pro-Insulin to mature Insulin through proteolytic cleavage (Furuta *et al.*, 1997; Furuta *et al.*, 2001; Zhu *et al.*, 2002).

Of the thirty eight probe sets that showed a decrease in gene-expression upon $MycER^{TAM}$ activation, the majority were involved in metabolism. The solute carrier family member Glut2 is a transmembrane protein responsible for the transport of Glucose through the cell membrane of β-cells and is an essential part of the Glucose sensing machinery. The corresponding gene, *Slc2a2*, was down-regulated 2-fold at 8 hours indicating a reduction in the Glucose sensing capacity of the cells as previously described. The interleukin receptors *il1r1* and *il6ra* were both down-regulated greater than 2-fold from 4 hours, indicating a reduced response to cytokine signalling through Il-1 and Il-6. Il-1 has been previously implicated in lipid metabolism by regulating Insulin levels and lipase activity under physiological conditions (Matsuki *et al.*, 2003). The protein product of the *kl* gene (Klotho) is a hormone that acts to repress signalling by Insulin and Igf1 to prevent aging (Kurosu *et al.*, 2005), and this was seen to be down-regulated greater than 2-fold at 8 hours.

These data suggest that activation of $MycER^{TAM}$ resulted in loss of function of β-cells by interrupting the Glucose sensing machinery. Insulin production was also directly affected, resulting in poor glycaemic regulation, and onset of hyperglycaemia and diabetes as previously described. However, a short period of increased Insulin production was seen within the first day following onset of β-

cell ablation, which corresponded to a previously identified 'window' of hypoglycaemia seen in both *pins-mycER*^{TAM} and RM transgenic mice. It is worth noting that this change in Insulin expression would similarly affect the expression of the c-*mycER*^{TAM} transgene, leading to a feedback loop in the system. However, the levels of inactive MycER^{TAM} in β-cells is high prior to initial 4OHT treatment, and further expression of MycER^{TAM} through the short time course considered here contributes only minimally to transient MycER^{TAM} levels.

Table 4.2.6: Genes relating to differentiation showing significant change in expression following activation of MycER^{TAM} (p-value ≤ 0.05 for 4OHT term) in pancreatic β-cells. Red = ≥ 2-fold up-regulation; Blue = ≥ 2-fold down-regulation. '*' = t-test p-value ≤ 0.05, '' = t-test p-value ≤ 0.01.**

		Fold change from control					
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**	**Biological role**
creg1	BC027426	2.09 **	2.25 **	1.52 **	2.31 **	0.00171	E1A antagonist
gck	BC011139	0.62	0.63	1.27	1.19	NA	Glucose metabolism
gck	L38990	0.59 *	0.14 **	2.32 **	0.87	NA	Glucose metabolism
glp1	AF276754	0.838	0.544	1.138	1.033	NA	Glucose metabolism
h2-aa	AV086906	7.66	1.52 **	0.74 *	4.26	0.00605	Chromosome structure
h2-d1	M34962	4.55	2.25	1.59	5.37 *	0.00231	Chromosome structure
h2-l	M86502	4.93	1.99	1.94	5.96 **	0.00514	Chromosome structure
h2-l	M69068	4.72	1.91	1.9	7.18 **	0.00165	Chromosome structure
il1r1	NM_008362	0.53 **	0.43 **	1.7 *	0.6 *	0.00289	Cytokine signalling
il6ra	X53802	0.36 **	0.45 *	1	0.47 **	0.0047	Cytokine signalling
ins2	NM_008387	0.15 *	0.32	11.8 **	14.15 **	NA	Glucose metabolism
kl	BQ175355	0.84	0.37 **	0.98	0.89	0.0213	Repression of Insulin and Igf1 signalling
nkx2-2	NM_010919	0.62	0.74 *	1.78	0.47 **	0.02222	Pancreas development
nkx6-1	AF357883	0.57 *	0.35 **	2.63 **	1.29	0.03537	Pancreas development
pcnt	NM_008787	2.63	4.27	4.98 **	4.78	5.43E-06	Centrosomal microtubule formation
pcsk2	BB357975	2.41 **	0.56 **	0.78 **	1.09 **	0.01453	Insulin processing
pcsk2	AI839700	1 **	0.48 **	0.74 **	0.96 **	0.03789	Insulin processing
pcsk2	NM_008792	0.57 **	0.39 **	1.99 **	2.68 **	0.04146	Insulin processing
pdx1	AK020261	0.76 **	0.22 **	0.76 *	0.47 **	0.00821	Insulin production
pdx1	AK020261	0.69 **	0.38 **	0.71 *	0.74 **	0.0414	Insulin production
slc2a2	NM_031197	0.89 *	0.42 **	0.97	0.77 **	0.00248	Glucose sensing

4.2.10.2 MycER^{TAM} activation-induced differentiation of suprabasal keratinocytes

Activation of MycER^{TAM} in suprabasal keratinocytes has previously been shown to result in loss of differentiation of suprabasal keratinocytes undergoing terminal differentiation (Pelengaris *et al.*, 2002b). Analysis of genes showing significant changes in expression upon MycER^{TAM} activation identified key genes relating to keratinocyte differentiation. These are shown in Table 4.2.7, and are coloured based on their change in expression ($\geq$ 2-fold, red; $\leq$ 2-fold, blue).

Most notable within this list is the Involucrin gene *inv*, which encodes a key factor in the progression of differentiation of keratinocytes. Involucrin works together with its substrate transglutaminase to cross link with membrane proteins and provide support to the cell (Eckert and Green, 1986). *inv* expression was decreased significantly throughout much of the time course, indicating that activation of MycER^{TAM} results in loss of differentiation in suprabasal keratinocytes. Despite the down-regulation seen in *inv*, the transglutaminase gene *tgm2*, which is involved in the formation of covalent bonds throughout keratinocyte differentiation, seemed to show a slight increase in expression from 8 hours.

Key processes in epidermal homeostasis include regulation of the actin cytoskeleton of differentiating cells and tight control of cell adhesion to allow cell migration to the surface. Integrin genes *itga7*, *itgb2*, and *itgb6*, whose products are involved in signal transduction and mediation of cell adhesion, showed predominantly up-regulation following MycER^{TAM} activation. However, *itga7* also showed a 2-fold down-regulation at 4 hours. Probe sets representing the Plectin gene (*plec1*), whose product is involved in formation of the cytoskeleton and maintenance of structural integrity (Svitkina *et al.*, 1996; Wiche, 1998), were also found to show an increase in expression, with one probe set showing high levels of expression change throughout the time course. However, contrast t-tests did not identify these changes as being statistically significant.

Cystatin A (*csta*) is a cysteine protease inhibitor that is involved in development of epidermal keratinocytes through its action as a precursor to the cornified cell envelope of differentiated keratinocytes (Rasanen *et al.*, 1978; Pernu *et al.*, 1990). As with *inv*, this gene shows down-regulation of 2-fold at 4 hours and 16 hours, suggesting a loss in the differentiated status of cells. Cystatin F (*cst7*), a further cysteine protease inhibitor involved in keratinocyte development (Ni *et al.*, 1998), showed up-regulation of 2-fold at 8 hours; although this was not seen throughout the whole time course.

Perhaps surprisingly, differentiation-related keratins such as suprabasal Keratins 1 and 10, and basal Keratins 5 and 14, were not identified as changing significantly due to 4OHT treatment by *Envisage* (and hence have 4OHT p-value = NA). However, t-test contrast analysis of individual conditions identified a significant loss in expression of 2-fold (contrast p-value $\leq$ 0.01) for the suprabasal keratinocyte-specific Keratin 1 (*krt1*) at 4 hours. This suggested an initial loss in suprabasal keratinocyte differentiation. Interestingly however, an increase in expression was detected at the later 8 hour time point (contrast p-value $\leq$ 0.01), which was also shown for the basal keratinocyte-specific Keratin 14, although this change was not maintained throughout the time course. This may suggest that, whilst early suprabasal keratinocyte differentiation markers show decreased expression, the short time course of MycER^{TAM} activation considered here may not be sufficient to see a loss in expression of key Keratins.

These data show that activation of MycER^{TAM} in suprabasal keratinocytes results in down-regulation of genes important in keratinocyte differentiation. These include genes involved in maintaining structural integrity of the cell and cross-linking in the cornified epidermal layer. It is interesting to note that several of these genes showed down-regulation within 4 hours of MycER^{TAM} activation, despite the apparent ~ 8 hour delay seen in MycER^{TAM} activity (Section 4.2.6). It is also worth noting that the early down-regulation seen in the suprabasal keratinocyte differentiation-related *inv* gene results in concomitant loss of MycER^{TAM} expression in *inv-mycER^{TAM}* transgenic mice. This may partially explain the minimal proliferation detected within these cells throughout this short

time-course. These results suggest that induction of proliferation in suprabasal keratinocytes is incompatible with normal terminal differentiation.

Table 4.2.7: Genes relating to differentiation showing significant change in expression following activation of MycER^{TAM} (p-value ≤ 0.05 for 4OHT term) in suprabasal keratinocytes. Red = ≥ 2-fold up-regulation; Blue = ≥ 2-fold down-regulation. '*' = t-test p-value ≤ 0.05, '' = t-test p-value ≤ 0.01.**

		Fold change from control					
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**	**Biological role**
cst7	NM_009977	1.03	2.02 **	1.02	1.46 **	0.03966	Cysteine protease inhibitor
csta	C89521	0.48 **	1.07	0.37 **	0.53 **	0.00051	Cysteine protease inhibitor
itga7	NM_008398	0.49 *	1.64 *	0.86	2.63 **	3.05E-02	Adhesion and cell signalling
itgb2	NM_008404	1.3	3.05 **	0.62	4.39 **	0.00118	Adhesion and cell signalling
itgb6	NM_021359	1.11	2.24 **	1.33 **	1.89 **	0.00059	Adhesion and cell signalling
inv	AV009441	0.36 **	0.64 **	0.26 **	0.84	0.00556	Keratinocyte differentiation
krt1	NM_008473	0.45 **	4.25 **	0.73 *	0.84	0.0405	Suprabasal keratinocyte differentiation
krt5	BC006780	1.67 **	1.26 *	1.32	0.69 **	NA	Basal keratinocyte differentiation
krt10	AK014360	0.86	0.877	0.90	1.00	NA	Suprabasal keratinocyte differentiation
krt14	BC011074	0.91	2.17 **	0.92	1.00	NA	Basal keratinocyte differentiation
krt14	BC011074	0.89	2.02 **	0.976	0.951	NA	Basal keratinocyte differentiation
plec1	BM210485	0.50	1.41	1.26	3.07 *	0.03399	Cytoskeleton formation
plec1	BM232239	3.34	5.39	2.83	0.82 *	0.00209	Cytoskeleton formation
tgm2	BC016492	0.88	2 **	1.92 **	5.34	0.00154	Structural integrity of cell
tgm2	AW321975	1.18	0.77 **	1.83 **	1.08	0.0006	Structural integrity of cell
tgm2	BB550124	1.34	0.71 **	1.82 **	1.09	0.00056	Structural integrity of cell
tgm2	BB041811	1.06	0.66 **	2.23 **	1.29	0.00075	Structural integrity of cell

4.2.11 Comparison between the skin and the pancreas

The *Envisage* model fitting procedure produced a list of 1,564 probe sets (1,434 genes) whose expression was found to change significantly within the first 8 hours due to the joint effects of $MycER^{TAM}$ activation and the type of tissue under study (≥ 2-fold change in expression at 8 hours, and $p < 0.05$ for model term '4OHT:Tissue'). These genes therefore represent a differential transcriptional response to $MycER^{TAM}$ activation between the two tissues, and may indicate functional diversity indicative of the opposing outcomes seen in the phenotypes. QT clustering of the expression profiles of genes across the time points for both tissues together (Materials and Methods, Section 2.6.4) identified 27 clusters of genes whose expression profiles across the time courses for the two tissues were correlated between genes, indicating possible co-regulation and functional similarity (Figure 4.2.14: 1-27). Relaxing the criteria to allow clusters containing fewer than 10 genes had no effect on those shown below, but identified 18 further clusters, including one which was included due to its high significance to the comparison of the skin and the pancreas (Figure 4.2.14: 28).

GO enrichment was performed to identify likely roles for co-regulated genes within each cluster. As expected, many probe sets within the clusters related to processes such as metabolism and transcription. Several clusters showing particularly varied expression profiles between the two tissues were of particular interest, and GO analysis identified functionally related classes of genes involved in key cellular events in $MycER^{TAM}$-induced proliferation and apoptosis.

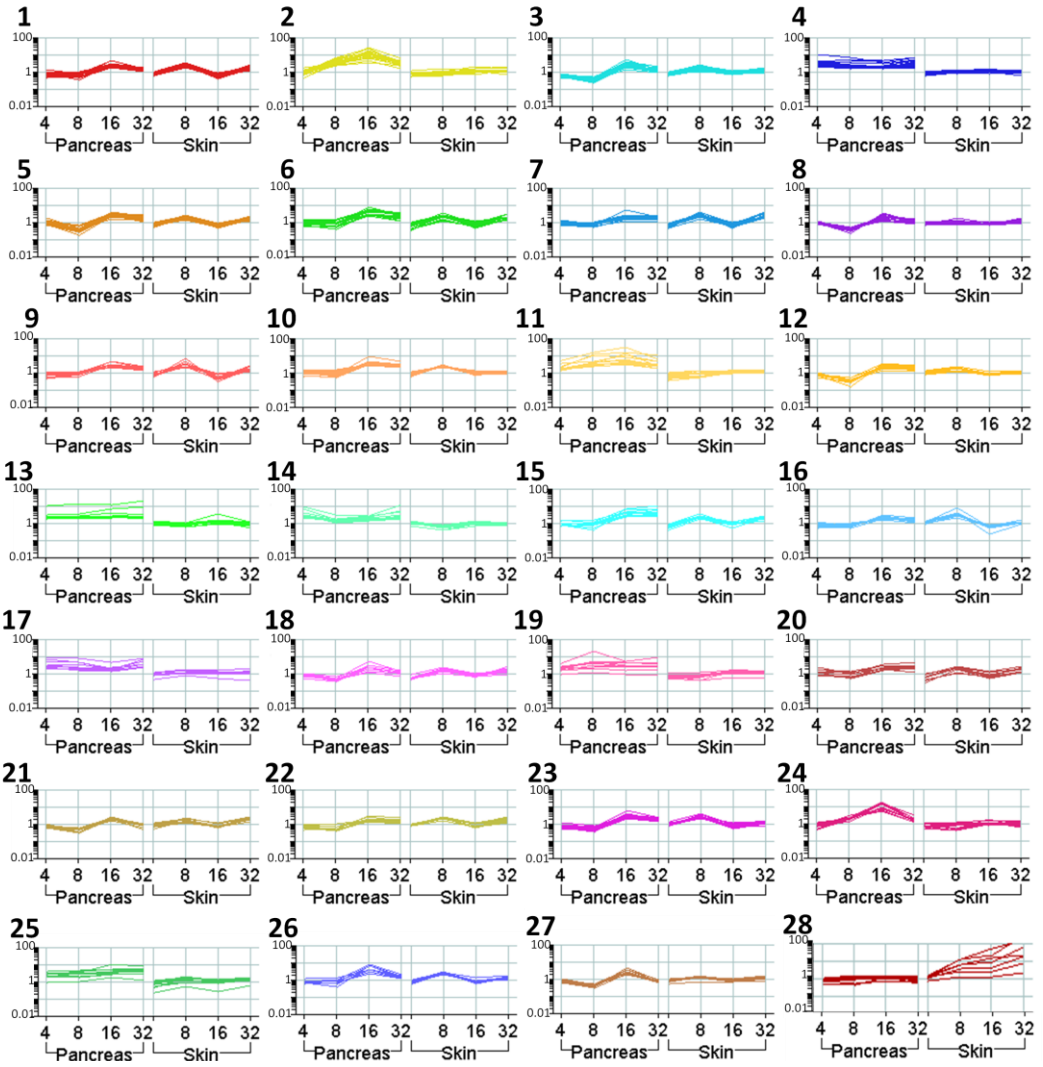

Figure 4.2.14: Quality threshold clustering of genes altered due to the joint effects of MycER^{TAM} activation and tissue type identified functionally related gene clusters

Genes identified as showing significant differential expression due to the joint effects of MycER^{TAM} activation and tissue were clustered to identify functionally related genes. QT clustering was performed using the Pearson correlation, and identified 28 unique clusters across the time courses of MycER^{TAM} activation for the skin and the pancreas. Shown here are the normalised signal intensities (representing fold-change following activation of MycER^{TAM}) for these genes over time for the two tissues. Of particular interest were clusters showing varied expression profiles between the two tissues, indicating disparate functional roles that may explain the dichotomous phenotypic outcomes of MycER^{TAM} activation in the two tissues.

Cluster 2 consisted of probe sets representing genes whose expression continued to rise steadily within the first 16 hours, but began to level off after 32 hours for the pancreas, but showed no change for the skin. Probe sets within this cluster were largely found to be involved with the M-phase of the cell cycle, particularly DNA replication and repair, which may explain why these genes were not significantly up-regulated within the first 4 hours. This cluster identified co-expression of probe sets for Cdk1 and Cyclin A2, suggesting a possible role for Cdk1 in G_2/M-phase following MycER^{TAM} activation. However, it is also possible that Cdk1 is related to mitotic catastrophe and apoptosis in β-cells following deregulation of Myc, as suggested by Castedo *et al.* (2002). These genes together indicate a significant increase in replication activity upon MycER^{TAM} activation.

The probe sets of cluster 24 shared similar expression profiles with those of cluster 2, and analysis of GO term enrichment again identified cell cycle and DNA replication as the predominant cellular functions. In particular, genes involved in cytoskeleton organisation and biogenesis, organelle localisation and cell division were well represented, indicating that cellular mitosis is one of the predominant responses following MycER^{TAM}-activation in β-cells.

Cluster 11, whose members were significantly increased throughout the time course for the pancreas but not for the skin, contained probe sets relating primarily to DNA replication and cell cycle, again linking with the role for MycER^{TAM} in initiating cell cycle progression. The presence of probe sets relating to the minichromosome maintenance (MCM) deficient genes *mcm6* and *mcm7*, whose products make up part of the MCM complex involved in DNA unwinding (Ishimi, 1997), the *cdt1* gene, whose product is involved in association of the MCM complex with chromatin, and various helicase related genes indicated that this cluster represents genes relating primarily to DNA replication.

Probe sets in cluster 19 also showed significant increase throughout the time course for the pancreas and not for the skin, and GO enrichment identified DNA replication, DNA damage checkpoint and cell cycle as being well represented within the list. This was particularly obvious due to the correlation between the DNA damage checkpoint related genes *atr* and *chk1*, and further members of the

MCM complex, *mcm2*, *mcm5* and *mcm7*. Similarity between clusters 11 and 19 indicated a key role for DNA damage response and repair in MycER^{TAM}-induced apoptosis in the β-cells. Association of Myc with MCM proteins has previously been described (Koch *et al.*, 2007). This close co-expression between genes whose products are involved in the formation of the DNA replication machinery may link to the recent work of Dominguez-Sola *et al.* (2007), who identified a non-transcriptional role for Myc in DNA replication. While such associations with Myc at the protein level are not determinable from these expression data, the synergy indicated here suggests functional cooperation.

Cluster 4 identified genes whose expression was significantly up-regulated at a steady rate upon MycER^{TAM} activation throughout the time course for the pancreas, but showed no change for the skin. Probe sets within this cluster appeared to be particularly related to carbohydrate metabolism, and the most obvious of these was that of the pyruvate dehydrogenase kinase gene *pdk1*. Mice with Pdk1 knocked out specifically within the β-cells have been shown to develop severe hyperglycaemia due to loss of β-cell mass, indicating a role for Pdk1 in maintaining Glucose homeostasis (Hashimoto *et al.*, 2006). It is therefore surprising to find that, in a model in which β-cell mass is ultimately reduced, expression of *pdk1* actually appeared to increase. This may relate to the changes seen in the Insulin gene in Section 4.2.10.1. Other probe sets within this cluster represented genes involved in RNA processing, indicating a transcriptional response to MycER^{TAM} activation.

Cluster 28 was one of 18 additional clusters formed when the criteria were relaxed to allow for clusters with fewer members. It was included here due to the striking difference between the expression profiles for the skin and for the pancreas. The placental growth factor gene *pgf*, whose product is a member of the VEGF family of angiogenesis-related genes, showed 2-fold increase in expression at 8 hours, which increased further throughout the time course to 6-fold at 32 hours. Over-expression of Pgf in basal keratinocytes of adult mice has been previously shown to result in increased numbers, branching and size of dermal blood vessels (Odorisio *et al.*, 2002), indicating that this may play a key role in the induction of

angiogenesis previously associated with activation of MycER^{TAM} in suprabasal keratinocytes (Pelengaris *et al.*, 1999).

However, the predominant cellular function of these genes was found to be regulation of proteolysis, with the majority of probe sets (5 of 7) relating to members of the kallikrein family of serine proteases. These proteins have been implicated in the regulation of tissue micro-environment, particularly through degradation of the extra-cellular matrix, often via activation of matrix metalloproteinases (MMPs). This is required to allow the growth of new vasculature, indicating a functional relationship between kallikreins and Pgf. As with *pgf*, expression of *klk* genes was found to increase significantly within 8 hours following activation of MycER^{TAM}, and continued to increase dramatically throughout the time course. *klk9*, *klk21*, *klk24* and *klk27* showed increases in expression of 57-, 18-, 20- and 143-fold respectively at 32 hours in the skin, which were particularly high in comparison to other genes in this experiment. The Kallikrein 1 gene, *klk1*, in particular showed an increase of expression of 274-fold at 32 hours, indicating a highly significant change in expression following Myc-deregulation. The fact that these expression changes occurred at later time points following MycER^{TAM} activation, indicates that they are not related to direct transcriptional regulation by Myc.

Klk1 and Klk9 have previously been found to be expressed throughout the epidermis of normal human skin (Komatsu *et al.*, 2003), and play a role in degradation of the extra-cellular matrix and loss of squamous cells during keratinocyte differentiation. Deregulated expression of Klk family members has also been implicated in many cancer types (Yu *et al.*, 1996; Bhattacharjee *et al.*, 2001; Iacobuzio-Donahue *et al.*, 2003; Chung *et al.*, 2004; Yousef *et al.*, 2004), and several human Klks are used as biomarkers in the screening, diagnosis and prognosis of cancers – e.g. Kallikrein 3 for prostate cancer (Stamey *et al.*, 1987) and Kallikrein 9 for ovarian cancer (Yousef *et al.*, 2001).

Of particular note is the role of Klk1 in Igf1-regulated tumour survival through degradation of the Igf binding protein, Igfbp3, in humans. This prevents Igfbp3 from antagonising Igf1-Igf1r interactions, allowing Igf1 to bind to its receptor and

initiate tumour survival through the Akt pathway as previously described (Rehault *et al.*, 2001). The mouse kallikreins Klk21, Klk24 and Klk27 have also been shown to be functionally active within the testes, both in degradation of the extra-cellular matrix and initiation of survival through degradation of Igf1bp3 (Matsui *et al.*, 2000; Matsui and Takahashi, 2001; Matsui *et al.*, 2005).

Together with changes in expression of Igf1 and Akt genes shown previously in Section 4.2.9.2, these data suggest a possible significant role for Igf1 and the kallikrein family of serine proteases in promoting cell survival in the keratinocytes upon MycER^{TAM} activation, allowing the tumourigenic potential of Myc to be realised.

4.2.12 Validation of gene-expression using qRT-PCR

Due to the often highly variable nature of microarray experiments, it is important to validate the results of gene-expression analysis to avoid false positive results. The sensitivity of qRT-PCR to even small changes in gene-expression between samples makes it one of the most commonly used procedures for validation of microarray results. 22 genes of interest were selected based on the microarray gene-expression results, and comparative qRT-PCR was performed to estimate changes in gene-expression between 4OHT-treated and vehicle-treated samples at 4 hrs, 8 hrs and 32 hrs for the skin and the pancreas. For qRT-PCR, total RNA was taken from original samples – those used for the microarray hybridisation – to ensure comparability with the microarray gene-expression results.

Due to the generally low concentration of starting RNA (Section 4.2.2.3), samples were amplified for the transcript of interest prior to qRT-PCR in a multiplexed reaction (Materials and Methods, Section 2.7.3). Because of the non-uniform amplification seen for several of the gene-expression assays, qRT-PCR could not be performed for the genes *ccne1*, *cdkn2b* (p15^{Ink4b}) and *chk1* (Materials and Methods, Section 2.7.6). Results for the pre-amplified qRT-PCR reactions and

their comparison with the microarray gene-expression results are shown in Figure 4.2.15 to Figure 4.2.33. Each figure shows a comparison between the normalised gene-expression data from the microarray analysis and from subsequent qRT-PCR analysis, and an image of the microarray gene-expression profiles for the skin and the pancreas, captured from the GS-GX analysis suite. Bars in the barplot indicate the mean of the observed normalised data for replicate samples (as described in Materials and Methods, Section 2.6.1), which represents the mean fold change following activation of MycER^{TAM}. Error bars in each case represent the standard deviation of the replicate data.

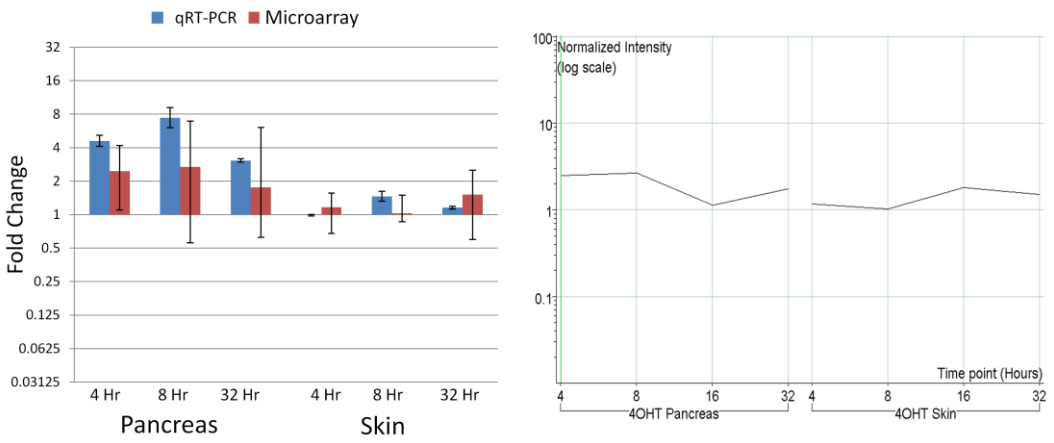

Figure 4.2.15: qRT-PCR and microarray results for Cyclin D1 (*ccnd1*)

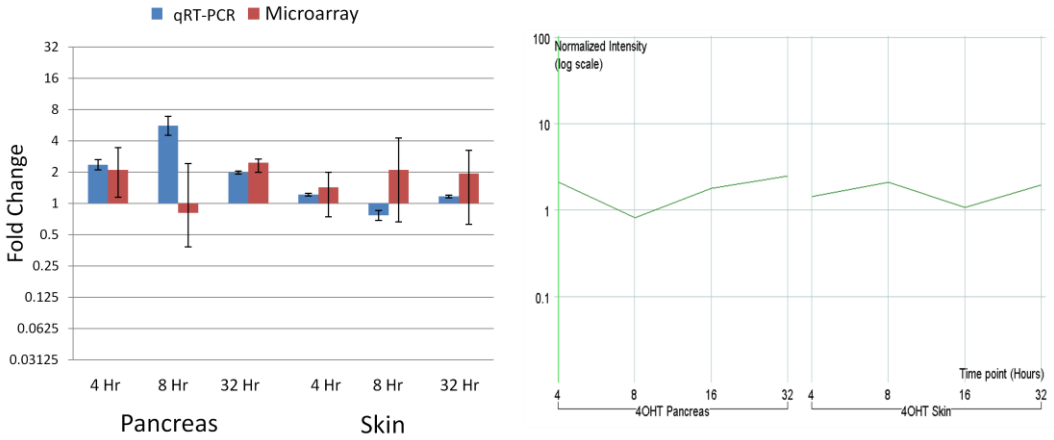

Figure 4.2.16: qRT-PCR and microarray results for Cyclin D2 (*ccnd2*)

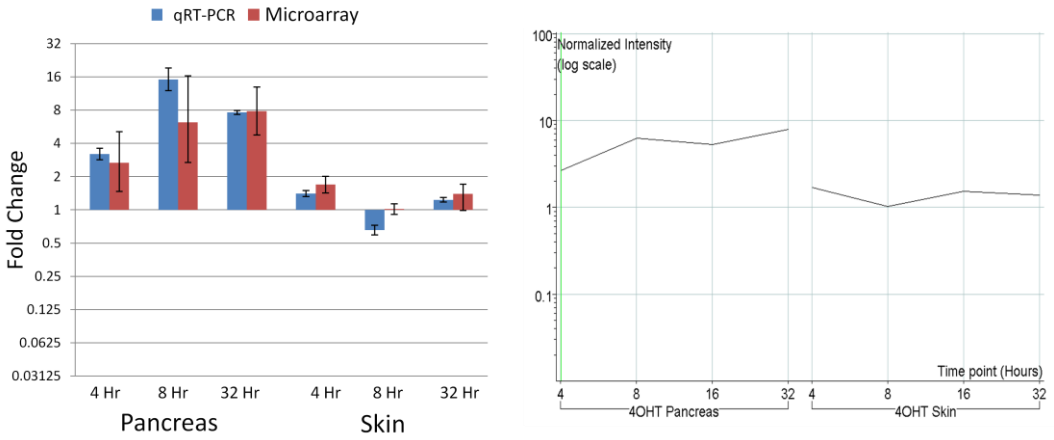

Figure 4.2.17: qRT-PCR and microarray results for Cyclin E2 (*ccne2*)

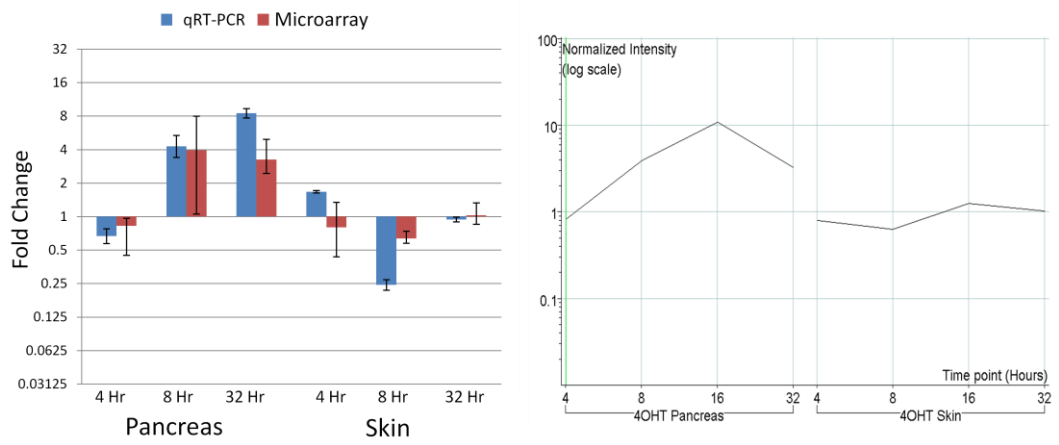

Figure 4.2.18: qRT-PCR and microarray results for Cyclin A2 (*ccna2*)

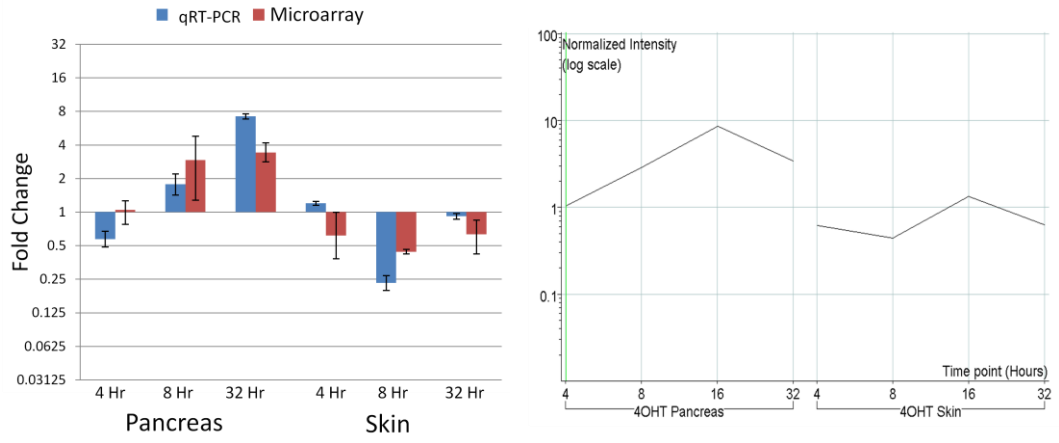

Figure 4.2.19: qRT-PCR and microarray results for Cyclin B1 (*ccnb1*)

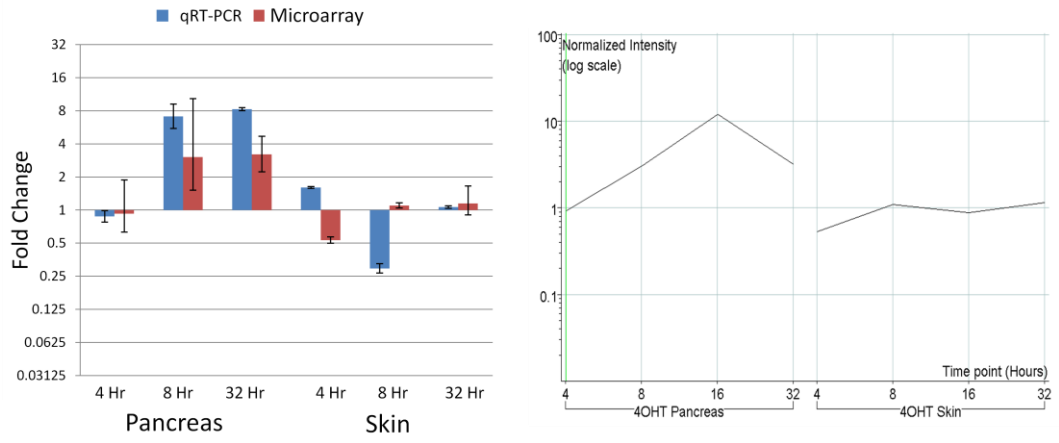

Figure 4.2.20: qRT-PCR and microarray results for Cell division cycle 2a (*cdc2a*)

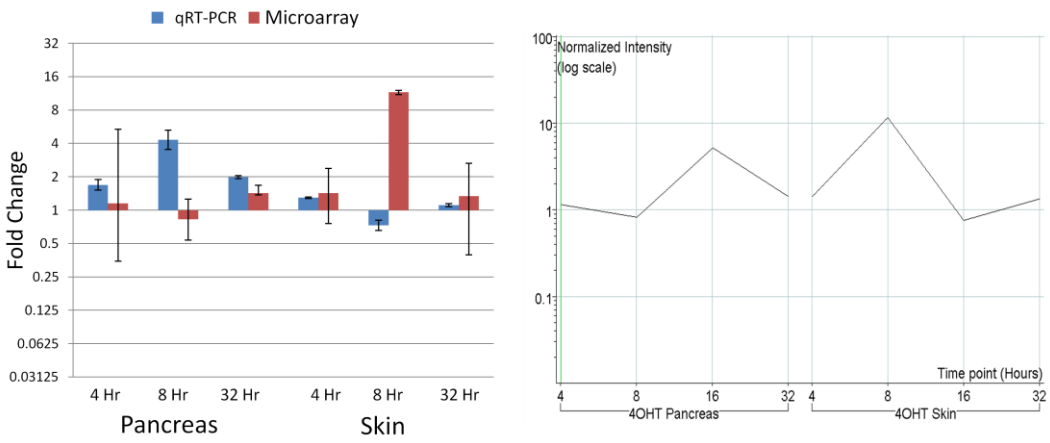

Figure 4.2.21: qRT-PCR and microarray results for Cyclin dependent kinase 4 (*cdk4*)

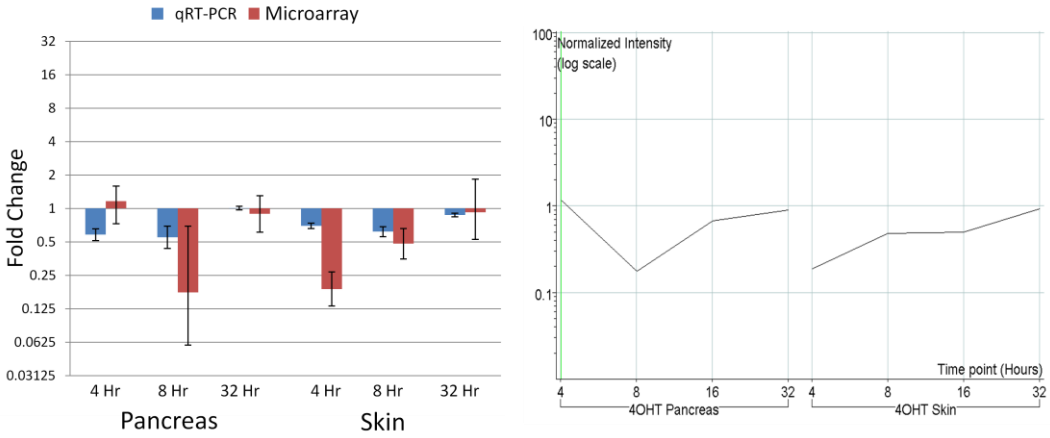

Figure 4.2.22: qRT-PCR and microarray results for p27^{Kip1} (*cdkn1b*)

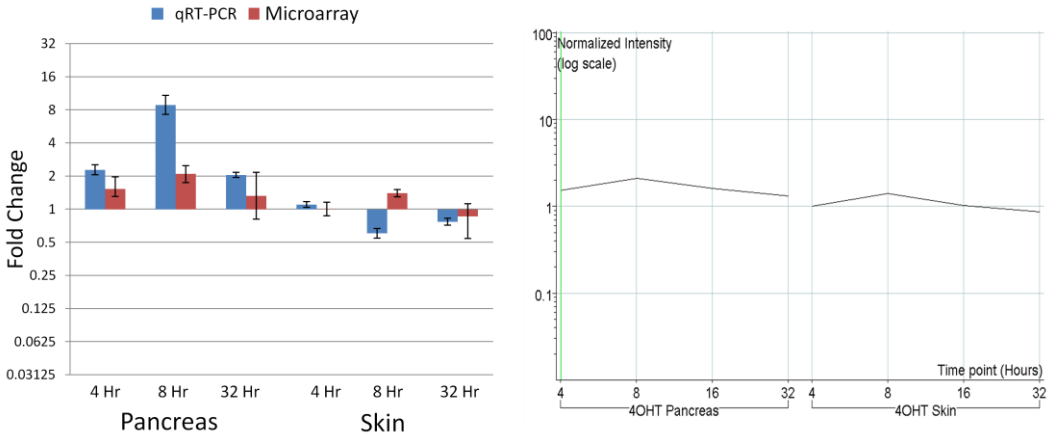

Figure 4.2.23: qRT-PCR and microarray results for p21^{Cip1} (*cdkn1a*)

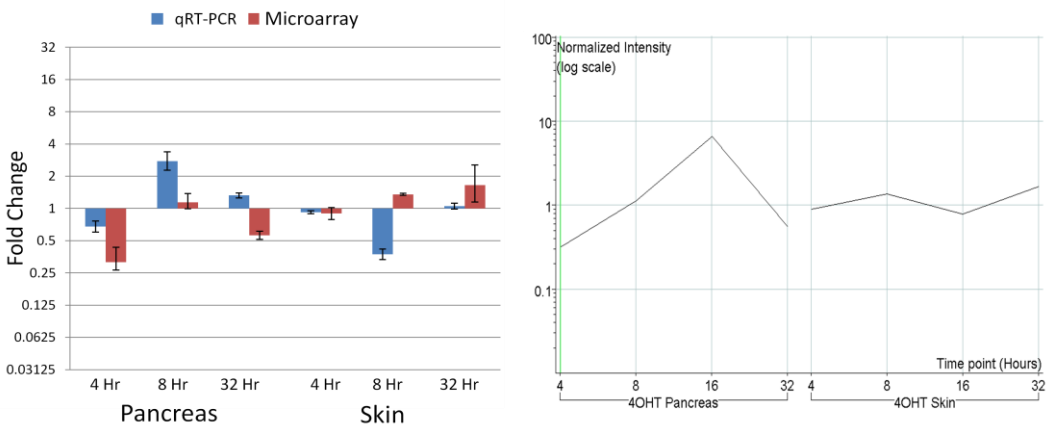

Figure 4.2.24: qRT-PCR and microarray results for p18^{Ink4c} (*cdkn2c*)

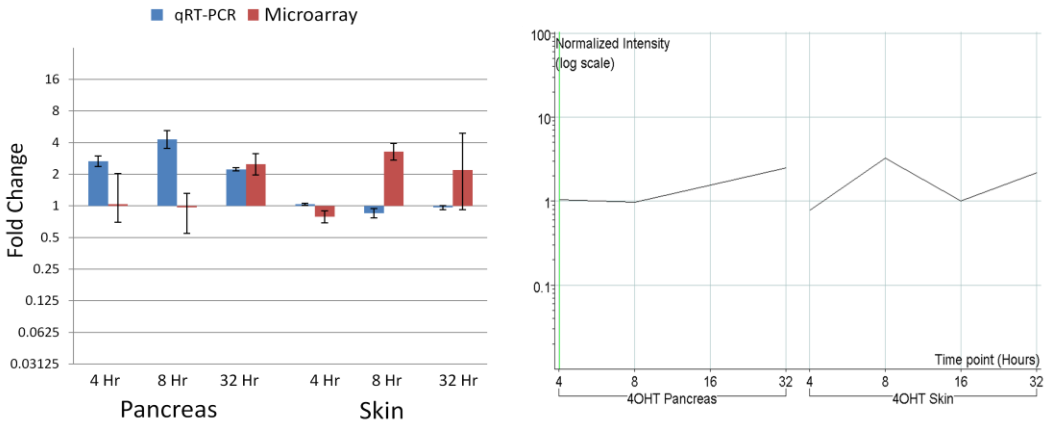

Figure 4.2.25: qRT-PCR and microarray results for Thymoma viral proto-oncogene 1 (*akt1*)

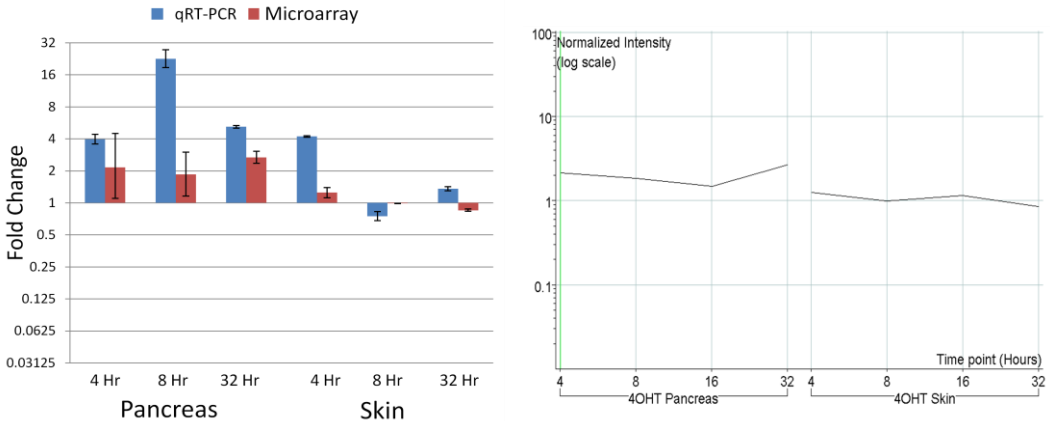

Figure 4.2.26: qRT-PCR and microarray results for p19^{Arf} (*cdkn2a*)

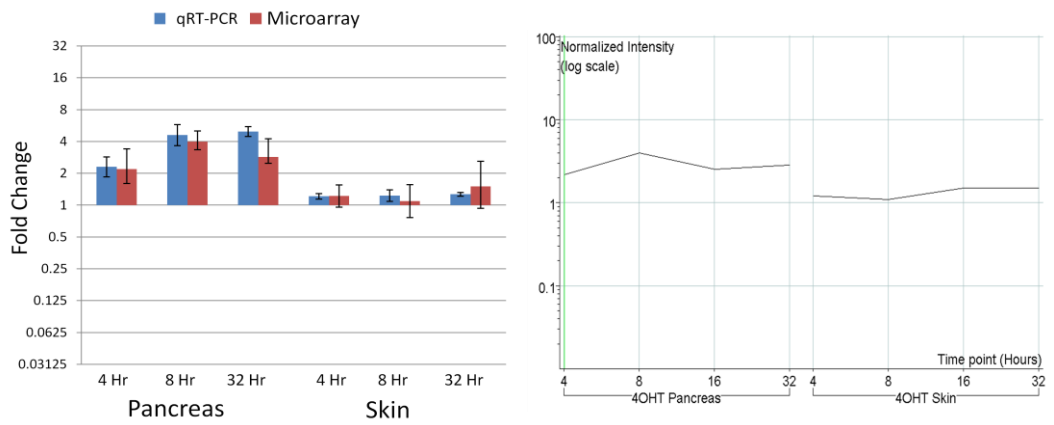

Figure 4.2.27: qRT-PCR and microarray results for Ataxia telangiectasia and Rad3 related (*atr*)

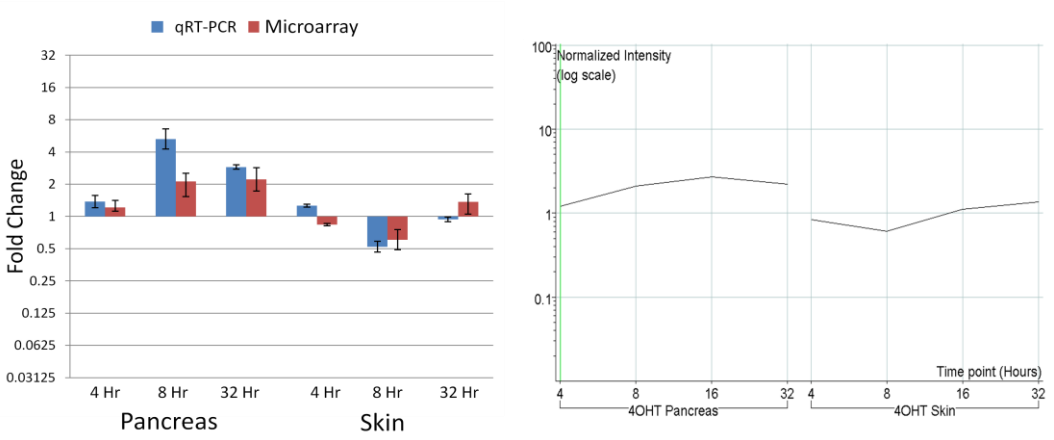

Figure 4.2.28: qRT-PCR and microarray results for Checkpoint kinase 2 (*chk2*)

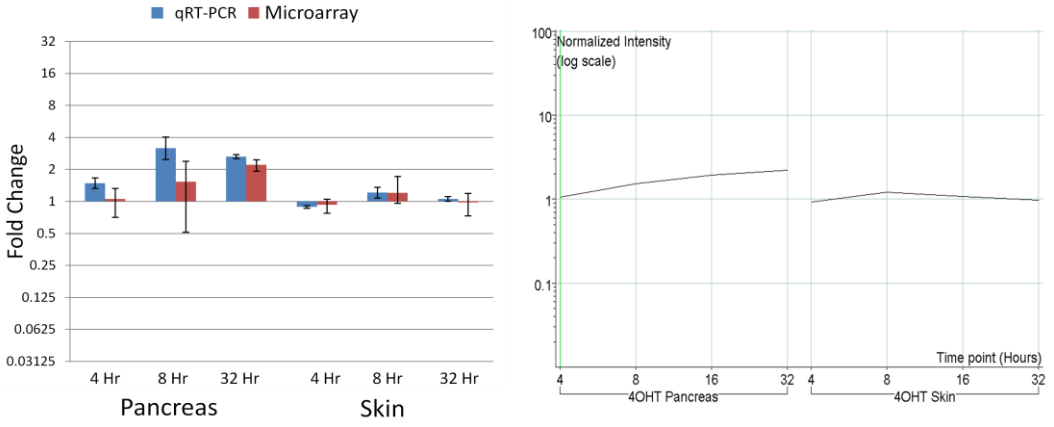

Figure 4.2.29: qRT-PCR and microarray results for Cytochrome c, somatic (*cycs*)

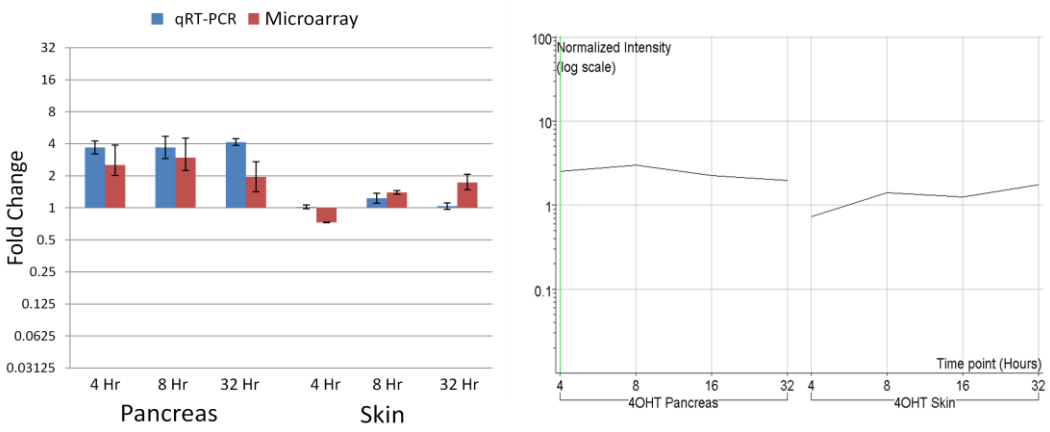

Figure 4.2.30: qRT-PCR and microarray results for Endonuclease G (*endog*)

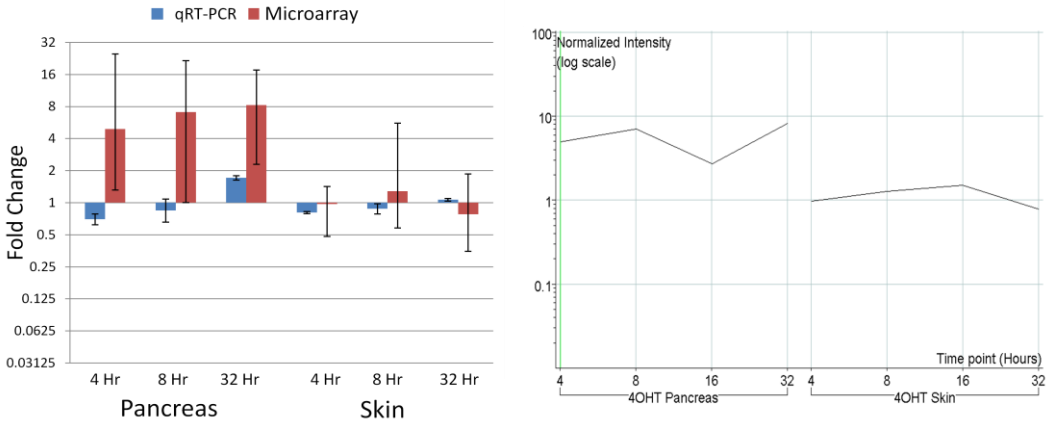

Figure 4.2.31: qRT-PCR and microarray results for Fas/CD95 receptor (*fas*)

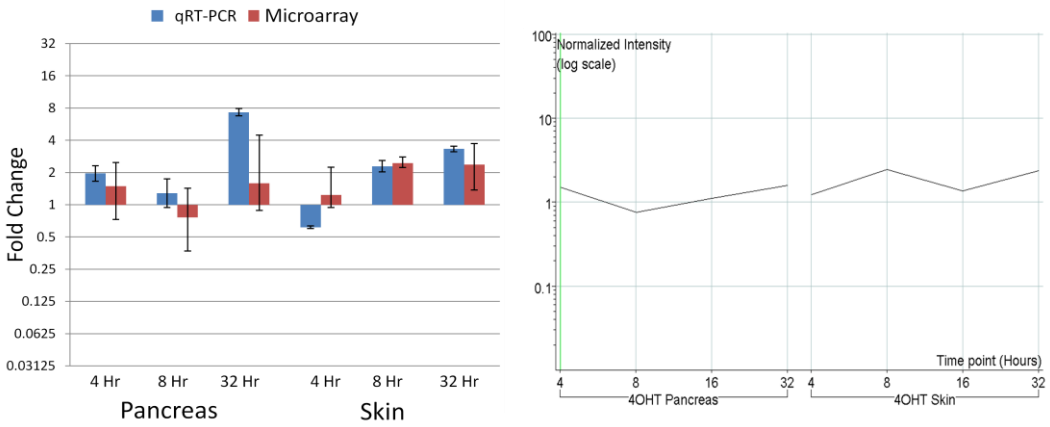

Figure 4.2.32: qRT-PCR and microarray results for Insulin-like growth factor 1 (*igf1*)

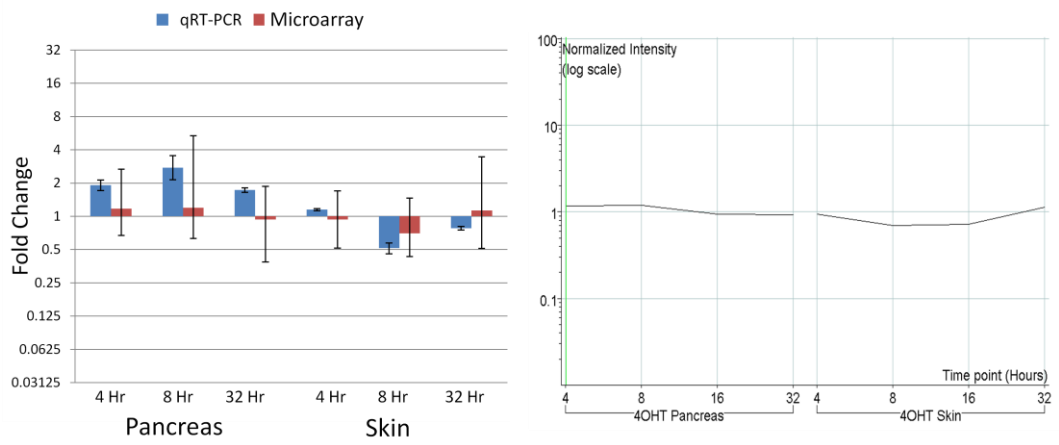

Figure 4.2.33: qRT-PCR and microarray results for Insulin-like growth factor 1 receptor (*igf1r*)

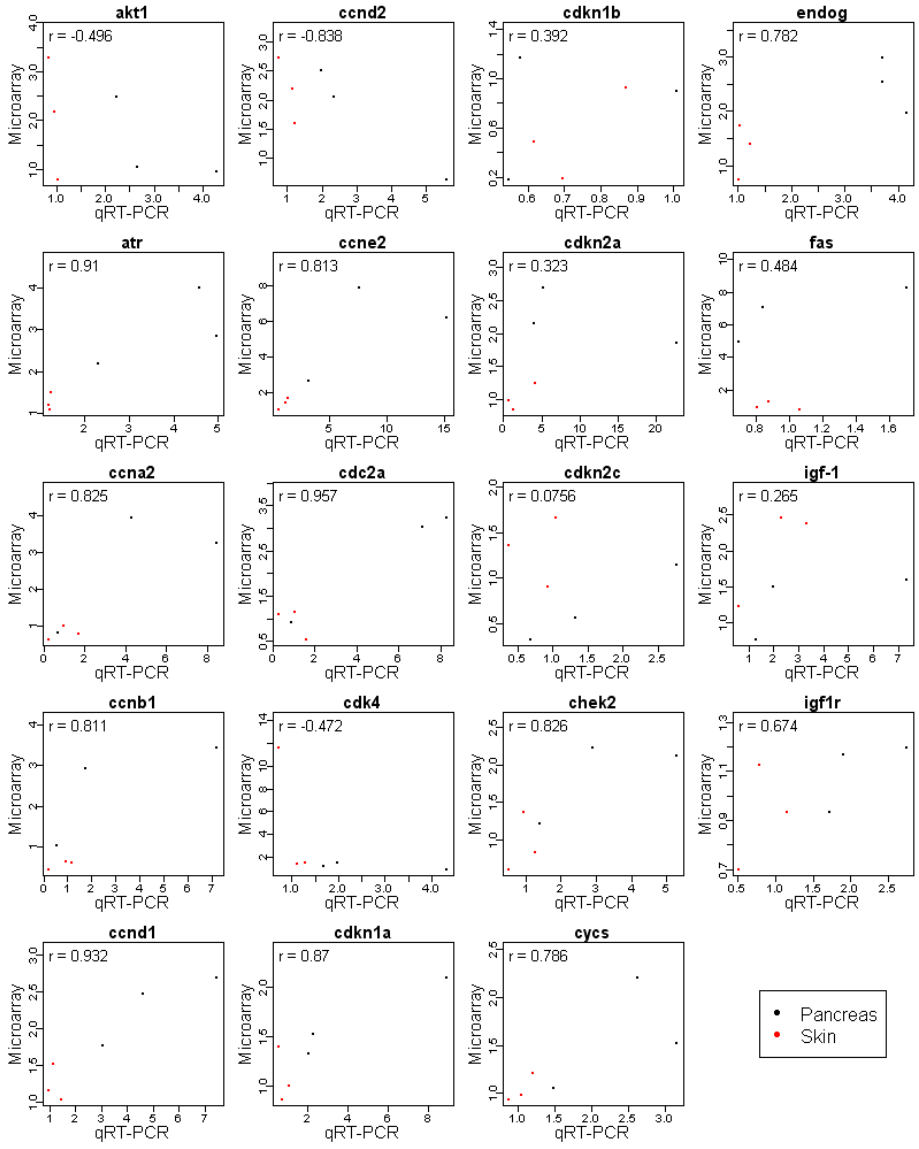

Figure 4.2.34: Correlation between gene-expression results from microarray and qRT-PCR analyses.

The similarity of the normalised gene-expression results between the microarray and qRT-PCR analyses for the 19 genes of interest are shown by plotting a scatterplot of the mean over replicates for the skin (red) and pancreas (black) over the three time points considered. Also shown is the Pearson correlation (*r*) between the two methods for the combined skin and pancreas data, indicating the similarity between the two approaches. Validation using qRT-PCR was generally comparable with microarray results ($r = 0.66 \pm 0.259$), although there were several genes where microarray results were not replicated (e.g. *akt1*, *cdkn1b*, *cdkn2a*, *fas*, *igf1*, *cdk4*).

Validation of microarray gene-expression results using qRT-PCR in general showed good reproducibility between the two methods, particularly for the β-cells. It is clear that the results from the microarray analysis are in general more variable than those of the qRT-PCR results, as evidenced by the large standard deviation error bars. This is due to the increased sensitivity of the qRT-PCR procedure in detecting changes in gene-expression. The correlation between the two methods for each of the 19 genes is also shown in Figure 4.2.34. This figure shows a scatterplot of the mean normalised gene-expression values for the two methods for each time point, and for each tissue (pancreas = black, skin = red). Also shown is the Pearson correlation (*r*) between the two methods of the combined skin and pancreas data, which provides an objective value defining the level of validation. In general, the qRT-PCR method validated the results from the microarray experiment well ($r = 0.66 \pm 0.259$). There were however several genes showing disparate results between the two methods, and these are discussed in the following sections.

4.2.12.1 Thymoma viral proto-oncogene (akt1)

The *akt1* gene showed a low level of correlation between the two methods ($r = -0.496$). No change in expression was detected for the early time points in the microarray analysis (1-fold change for both 4 and 8 hours) for the pancreas (Figure 4.2.25). However, qRT-PCR analysis identified a significant increase in expression for both the 4 hour (greater than 2-fold) and 8 hour (greater than 4-fold) time points. In contrast to this, the microarray analysis identified a significant increase in expression for the skin samples at the 8 hour (3-fold) and 32 hour (2-fold) time points. However, qRT-PCR analysis identified no such increase, indicating a possible false positive result. Sequence alignment using the basic local alignment search tool (BLAST) (Altschul *et al.*, 1990) from the NCBI confirmed that both the qRT-PCR probe and the microarray probe set identified the same *akt1* transcript. The reason behind the discrepancy between the two methods is thus unknown, and further validation is necessary to confirm the microarray gene-expression.

4.2.12.2 Cyclin D2 (ccnd2)

The correlation between the qRT-PCR and microarray results for the *ccnd2* gene was high (*r = -0.838*), lending support to the validity of this result. However, a notable difference between the two methods was seen for the pancreas 8 hour time point (Figure 4.2.16), where microarray analysis failed to identify a significant change in gene-expression, but qRT-PCR identified a large increase in expression (greater than 5-fold) that fit in well with the observed proliferative phenotype. In contrast, the 2-fold change in gene-expression seen in the skin samples at 8 hours and 32 hours was not identified in the qRT-PCR results.

The expression level for *ccnd2* in the microarray experiment was calculated as an average across three probe sets that were each found to be significant using *Envisage*. One of these probe sets was designated as detecting alternative transcripts from the same gene, so may affect the levels of detected expression. Removing this probe set resulted in levels of gene-expression remaining at normal levels throughout the time course for the skin, which matched to the results found using qRT-PCR. Comparing the expression levels of the two remaining probe sets, it appeared that while one showed no change in expression at 8 hours for the pancreas, the other showed an increase in expression greater than 2-fold throughout the time course.

Sequence alignment using BLAST indicated that the non-specific probe set and the probe set showing no change in expression at 8 hours were located towards the 5' end of the mRNA sequence, suggesting that the discrepancy in the signals may be related to RNA degradation. It also appeared that the length of the *ccnd2* transcript interrogated by these probe sets was shorter than that of the transcript identified by qRT-PCR, possibly leading to an increased effect of RNA degradation on the signal. If the probe set that identifies transcripts within the mid region of the *ccnd2* mRNA was considered alone, the detected levels of expression change were very similar for the microarray and qRT-PCR data in both the skin and the pancreas.

4.2.12.3 Cell division cycle 2a (cdc2a)

Observed changes in gene-expression for *cdc2a* were generally well validated by qRT-PCR (*r = 0.957*), with the possible note that the detected changes in gene-expression were higher using qRT-PCR than for the microarray experiments (Figure 4.2.20). This is due to the higher sensitivity in the qRT-PCR procedure. The microarray results identified no change in expression through the time course for the skin, and the qRT-PCR results confirmed this at the 4 and 32 hour time points. However, qRT-PCR detected almost 4-fold down-regulation of *cdc2a* at 8 hours. Sequence analysis of the qRT-PCR probes identified no obvious difference between the selected primer and the microarray probe set, indicating that further validation is necessary.

4.2.12.4 Cyclin-dependent kinase 4 (cdk4)

Despite the apparent proliferative phenotype of the pancreatic β-cells, the c*dk4* gene was found to show no change in expression upon MycER^{TAM} activation in the pancreas microarray experiment at 4, 8 or 32 hours (Figure 4.2.21). An increase in expression of almost 9-fold was detected at 16 hours, but this time point was not included in the validation studies due to budget constraint. qRT-PCR validation for the *cdk4* gene showed poor correlation with the detected microarray results (*r = -0.472*), and identified significant increases in *cdk4* expression at 8 hours (4-fold) and 32 hours (2-fold). The skin showed the opposite effect, with microarray results identifying a large increase in *cdk4* expression at 8 hours (11-fold) which was not identified using qRT-PCR. Given the small error bars in the detected microarray signal, which indicate closely matching results for all replicates, this is surprising. This discrepancy appeared to be confined to the 8 hour samples for the skin and the pancreas.

As with *ccnd2*, the *cdk4* gene-expression was calculated as the average of 3 probe sets found to be significant using *Envisage*. Of these probe sets, only one was designed to match a single unique transcript for the *cdk4* gene. One probe set

recognised several alternative transcripts from the *cdk4* gene, while the other contained probes that may cross-react with transcripts from other genes. This may indicate that some form of cross-reaction occurred such that the microarray signal was no longer specific to the transcript of interest. Also, the transcript sequences targeted by these probe sets were very short, indicating that they may be susceptible to cross-reaction and RNA degradation. Despite this, it is interesting to note that the 3 probe sets showed almost identical expression profiles to one another.

Another explanation is that the qRT-PCR probes selected for this analysis were designed within a single exon of the *cdk4* transcript, indicating that genomic DNA may be detected. This problem was unavoidable with the off-the-shelf gene-expression assays used in this study, although may be avoided in the future by designing primers specifically across exon boundaries. Detection of genomic DNA as well as the specific transcript in VT 8 hour samples may explain the apparent down-regulation detected here.

4.2.12.5 Cyclin-dependent kinase inhibitor 1a ($p21^{Cip1}$)

The results from the qRT-PCR analysis closely match those of the microarrays (Figure 4.2.23; $r = 0.870$). However, the changes in gene-expression detected with qRT-PCR were in general larger as compared to the microarrays for the pancreas. This showed that the expression of the cell cycle arrest-promoting CDKI $p21^{Cip1}$ is influenced by $MycER^{TAM}$ activation even at 4 hours, which may indicate inherent tumour suppressor activity in the cell following $MycER^{TAM}$-induced proliferation. Given that proliferating β-cells are detected throughout the time course, it is clear that these changes are not sufficient to arrest cell cycle progression following ectopic activation of $MycER^{TAM}$.

4.2.12.6 Cyclin-dependent kinase inhibitor 1b (p27^{Kip1})

The CDKI p27^{Kip1} showed down-regulation in both the skin and the pancreas tissue, indicating Myc-initiated loss of cell cycle inhibitors in cell cycle progression (Figure 4.2.22). Validation using qRT-PCR showed poor correlation with the microarray results (*r = 0.392*), and whilst qRT-PCR identified loss of expression of *cdkn1b* (p27^{Kip1}) for both the pancreas and the skin, this change was much smaller than that seen using microarrays. Due to the sensitivity of the technique, the gene-expression estimates using qRT-PCR are more accurate and these changes, while less than 2-fold, may still be significant.

4.2.12.7 Cyclin-dependent kinase inhibitor 2a (p19^{Arf})

In general, Figure 4.2.26 showed that the results for qRT-PCR matched up well with the microarray results, although this correlation did not seem to be represented by the detected correlation in Figure 4.2.34 (*r = 0.323*). In general, qRT-PCR identified larger fold-changes for the pancreas than the microarray, possibly due to the increased sensitivity of the assay, suggesting that the relationship between the two sets of results may not be linear and this may explain the low correlation score detected for this gene. The qRT-PCR results also indicate a more significant increase in *cdkn2a* (p19^{Arf}) expression for the pancreas at 4 hours as compared to the microarray results, with no discernible change at the later time points. This may indicate that the p19^{Arf}-related apoptotic pathways are immediately activated by Myc activation within the skin, but that apoptosis is inhibited by survival signalling pathways, such as Igf1.

4.2.12.8 Cyclin-dependent kinase inhibitor 2c (p18^{Ink4c})

The expression of the Cdk4 and Cdk6 inhibitor p18^{Ink4C} was not well validated by the qRT-PCR results (Figure 4.2.24; *r = 0.076*). BLAST sequence analysis indicated that the microarray probe set recognised a sequence located primarily at

the 3' end of the *cdkn2c* ($p18^{Ink4c}$) transcript, whilst the qRT-PCR probes spanned the majority of the whole transcript. This may indicate a reason for the discrepancy seen between the two methods.

4.2.12.9 Cytochrome c, somatic (cycs)

The results of qRT-PCR and microarray studies in general matched well for the *cycs* gene (Figure 4.2.29; $r = 0.786$). The change in expression detected in the microarrays for the pancreas at 8 hours following $MycER^{TAM}$ activation showed high variation, but validation using qRT-PCR confirmed a significant increase in expression of 3-fold. However, as with *cdk4*, the increased signal detected in the qRT-PCR results may be a result of detection of genomic DNA due to the probe used.

4.2.12.10 Fas death receptor (fas)

The results of the microarray analysis indicated a very large statistically significant increase in expression of *fas* in the pancreas across 4 probe sets on the array. However, validation using qRT-PCR revealed no such increase in expression (Figure 4.2.31), and in general correlation between the two data sets was poor ($r = 0.484$). Given the size of the observed change in expression (~ 8 fold across the time course) and the fact that this was seen in 4 out of 5 probe sets for the Fas receptor, this result was surprising and unfortunate. To understand why the two procedures produced such vastly different results, BLAST was used to compare the probe set sequences to the gene transcript sequence queried by the qRT-PCR probe. When the sequences were compared, no significant overlap was discovered between the sequences, indicating that their target transcripts were in fact different. The only Fas-specific probe set that matched to the qRT-PCR probe was the probe set showing no change in expression across the time course for both the skin and the pancreas. This indicates that there is no observed change in expression for *fas* in either the microarray study or the qRT-PCR study, and that

the probe sets that previously showed significant fold changes are not specific to the *fas* transcript.

Confirmation of the probe annotation using the Affymetrix GeneChip annotation database, NetAffx [4], identified only the non-changing probe set as specific for Fas. Two of the remaining probe sets were identified with the X-linked myotubular myopathy gene *mtm1*, and the remaining two had no annotation. This indicates that annotations were incorrect in both GS-GX (annotation updated in March 2007 using standard annotations from the NCBI RefSeq[5], GenBank[6] and Entrez Gene[7] sequence databases) and Bioconductor (annotation updated using the October 2007 build of the *mouse4302* (Liu *et al.*) and *mouse4302cdf* packages). This discovery highlights the need to confirm probe annotations, and to ensure that annotations are kept up to date.

[4] https://www.affymetrix.com/analysis/netaffx/index.affx

[5] http://www.ncbi.nlm.nih.gov/RefSeq/

[6] http://www.ncbi.nlm.nih.gov/Genbank/index.html

[7] http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene

4.3 Summary

This chapter has discussed the results of a large microarray experiment performed to analyse the differences in transcriptional response to deregulated Myc activity in two distinct tissues – suprabasal keratinocytes and pancreatic β-cells. To ensure that changes in gene-expression were attributable to activation of the MycER^{TAM} protein in β-cells, and not to effects within the exocrine pancreas, LCM was used to isolate homogenous islet tissue for RNA extraction. Despite the large detrimental effect of RNases on RNA integrity seen in pancreas derived samples, the optimised protocol discussed in Section 4.2.2.1 produced good quality RNA suitable for microarray hybridisation. Due to the nature of murine epidermis, LCM of suprabasal cells was found to be impractical given the constraints of the project, and RNA was instead isolated from sections of whole skin tissue.

The overall quality of RNA and subsequent yields of 2a-cRNA were good, with skin-derived samples generally showing superior quality. Extensive quality control throughout the microarray hybridisation procedure identified outlying samples that may bias subsequent analyses. In general, there appeared to be no correlation between initial sample quality and ultimate data quality, indicating that hybridisation of low quality samples can still produce good quality data. There are limits however, and one sample hybridised with a very low volume of 2a-cRNA produced particularly poor quality data.

Gene-expression analysis identified a clear proliferative response in the β-cells, evidenced by the change in expression of key cell cycle genes following MycER^{TAM} activation. This was also shown by immunofluorescent staining with the cell cycle marker protein Ki67 (Figure 4.2.1). Ki67 staining in the skin also identified increased expression in MycER^{TAM}-encoding suprabasal keratinocytes, however the levels of Ki67-positive cells was less conclusive than for the β-cells. Ki67 staining also identified proliferating cells in the basal epidermis layer in both VT and 4OHT-treated animals, which may have acted to mask differential expression of key cell cycle genes within the suprabasal keratinocytes. This may

also have been detrimentally affected by the apparent delay in MycER^{TAM} response to 4OHT treatment when administered topically (Section 4.2.6).

Analysis of changes in expression of genes relating to apoptosis and survival identified increased expression of key members of the DNA damage pathway as being correlated with MycER^{TAM} activation and subsequent apoptosis in the pancreatic β-cells. It is hypothesised that this culminates in activation of Bax at the mitochondria (although it is not clear whether this occurs via activation of of the tumour suppressor p53), resulting in release of Cytochrome c from the mitochondria and subsequent apoptotic signalling in the β-cells. However, further analyses would be required to establish a causal role for Myc in these pathways.

Increased expression of the Igf1 survival factor in the suprabasal keratinocytes suggests that avoidance of Myc-induced apoptosis may occur through the Igf1 receptor (e.g. through the Akt survival pathway). Previous proteomics studies within the Mike Khan research group at the University of Warwick using the same transgenic system have identified levels of the Igf1 receptor protein as being increased following activation of MycER^{TAM} in suprabasal keratinocytes (manuscript in preparation), further suggesting a role for the Igf pathway in survival of suprabasal keratinocytes. However, no such change was detected for the Igf1 receptor at the transcript level (Figure 4.2.33) suggesting that this change is not a transcriptional response to Myc. A large increase in expression of key Kallikrein genes in the skin, whose products are involved in both vascularisation (specific to the keratinocyte MycER^{TAM} model) and also in augmentation of the Igf1 receptor pathway through degradation of the Igf1r antagonising Igf binding proteins, suggests a tissue specific pathway for determining ultimate cell fate following MycER^{TAM} activation.

Chapter 5 General Discussion

5.1 Discussion

The Myc proto-oncogene has previously been implicated in the control of many cellular functions, particularly those involved with promoting proliferation. The myriad putative targets of the Myc transcription factor, found using global studies for gene-expression such as microarrays, have led to the belief that Myc may act as a global transcription factor, regulating the expression of up to 15 % of the genome. An alternative hypothesis is that only a small subset of these genes is in fact direct targets of Myc, with further changes occurring downstream.

Of particular interest has been the discovery that oncogenes such as *c-myc*, whose deregulation can lead to unchecked proliferation and formation of tumours, have the potential to promote apoptosis under certain conditions. This acts as an inherent tumour suppressor function to limit the formation of cancers. The formation of malignant growths thus requires the cooperation of several somatic lesions to not only deregulate proliferative control, but also to inhibit protective apoptotic function and allow the oncogenic potential of proto-oncogenes such as *c-myc* to be realised.

However, whilst *in vitro* studies identifying the conditions under which Myc can promote apoptosis have provided key insights into the diverse range of Myc-induced functionality, they fail to take into account the all-important effect of tissue context *in vivo*. The MycER^{TAM} switchable *in vivo* model for Myc-deregulation has provided a number of insights into the dichotomy of Myc-functionality. The most intriguing of these is the fact that deregulation of Myc activity alone can result in vastly different outcomes depending solely on tissue location – unchecked proliferation in the suprabasal keratinocytes of the skin and apoptosis in the pancreatic β-cells.

To identify the processes by which circumvention of tumour suppression is accomplished within the suprabasal keratinocytes, transcriptional analysis using high-density microarrays was performed to identify the differential transcriptional response between the two tissues in a time course following Myc deregulation. The switchable MycER^{TAM} transgenic mouse model allowed precise control over the 'Time 0' of aberrant Myc activity, allowing analysis of transcriptional changes at the very early stages of the Myc-induced phenotype. In this way, changes in gene-expression that occur as a direct consequence of Myc deregulation can be distinguished from those that occur further downstream.

The linear model fitting algorithm *Envisage* was designed and written by the author to identify statistically significant effects in the gene-expression data in complex experimental designs such as this. In particular, this process allowed identification of effects on gene-expression of not just experimental variables, but also of superfluous variation related to experimental design, non-homogeneity of samples, and fluctuations in environmental conditions. This allowed identification and correction of problems within the experimental design that may otherwise have resulted in false assignment of significance during the analysis, such as batch effects and sample quality.

5.1.1 Quality control

One of the key rules for microarray analyses is the saying: "garbage in → garbage out". It was therefore essential to ensure that the quality of starting materials and the resulting data were of a suitable quality to ensure accurate conclusions could be drawn. Quality control was performed throughout the experiment to ensure that no errors were introduced prior to data analysis. This included performing pre-hybridisation (RNA and 2a-cRNA quality), post-hybridisation (%P, control probe signal, scale factor, etc.), probe-level (PLMs, RNA degradation plots, boxplots, etc.) and post-normalisation (boxplots, PCA, scatterplots, etc.) control checks.

Throughout the experiment, many precautions were taken to ensure that introduction of errors was minimised, and analysis of these QC metrics identified no significant introduction of error throughout the protocol. Given the extensive processing required to extract RNA from pancreatic islets using LCM, this was particularly good news.

In general, RNA quality was good, indicating that the optimised LCM protocol described in Section 4.2.2.1 is suitable for isolation of RNA from homogenous sources with high levels of RNA degradation, such as the pancreatic islets. Throughout the procedure, the difference in quality between the pancreas and the skin was evident, with skin samples generally performing better in tests than pancreas samples. This was evidence of both the difference in RNA degradation levels in the two tissues, and also the disparity in the RNA isolation methods used. However, this also indicated that there was a level of confoundment between RNA quality and tissue, suggesting that this was not a suitable variable for inclusion as a covariate in the *Envisage* model.

Poor quality samples were detected throughout the experiment by their poor performance in the various QC tests (Section 4.2.4). A large number of QC metrics were used to identify sample quality at all stages of the analysis. A simple scoring system was used to combine the various metric scores, allowing information relating to all possible aspects of microarray hybridisation to be considered. A simple score was assigned to each sample depending on whether the sample passed the recommended thresholds for the QC metric (QC penalty = 0), fell outside of the threshold but remained within suitable bounds (QC penalty = 1), or fell well outside of the threshold value (QC penalty = 2). Whilst this approach was in some ways not as objective as using, say, the RIN to define poor quality samples, it allowed the large number of QC analyses performed to be included in the decision of which samples to remove from the analysis. In particular, Jones *et al.* (2006) showed that pre-hybridisation metrics such as RIN may not be an accurate predictor of ultimate data quality. The decision as to which samples should be removed was, in general, straightforward. In particular, it was positive to see that the sample hybridised with almost 0 μg 2a-cRNA was

identified as the sample of poorest quality, indicating that this process was capable of identifying particularly poor samples. However, the four other samples hybridised with less than 10 μg 2a-cRNA were not identified as having poor quality hybridisation, and in fact performed well across the QC tests. This may suggest that the recommended minimum of 10 μg 2a-cRNA is overly conservative, or alternatively may suggest that quantification was inaccurate for low yields.

It was interesting to note that sample Panc T 16hr (3), which was identified as a poor quality sample with a QC penalty of 12, corresponded to a pancreas-derived RNA sample with a particularly high RIN of 9. Given that the goal of many of the QC tests was the identification of outliers from the triplicate samples, it is possible that this sample was identified not because it represented a poor replicate, but instead because the sample quality was superior to the other replicates (with RINs of 6.8 and 5.8 for replicates (1) and (2) respectively). However, the fact that this sample appeared to be of a much greater quality than all other pancreas RNA samples called its authenticity into question. The RIN for this sample was greater than that for RNA extracted from freshly excised pancreas, and in fact seemed more in line with RINs seen in skin-derived RNA samples. Since the reason for the extremely high RIN was not well understood, and the two remaining replicate samples performed extremely well across the QC tests (with both showing QC penalty scores of 0), this sample was removed.

Surprisingly, there appeared to be no significant relationship between sample quality prior to microarray hybridisation and the quality of the resulting data (Figure 4.2.8). Poor quality RNA produced reproducible data, and poor quality data were produced from good quality RNA. This indicates that screening samples based on pre-hybridisation measures such as the RIN may not be a suitable approach to the removal of poor quality samples, which may otherwise have been a useful way for cutting costs given the high price of running replicate samples. A similar conclusion was also drawn in a more thorough study of this kind performed by Jones *et al*. (2006). This indicates that seemingly poor quality or

low yields of starting material may yet produce good quality microarray data from which conclusions may be drawn.

5.1.2 *Envisage*: Significance Analysis of Microarray Data

Within an experimental setting, it is impossible to ensure that the condition under consideration is the only source of variation within the experiment. Random measurement errors, technical errors, and effects due to variation in phenotype and environment can elicit significant responses in gene-expression that may shroud the true changes of interest. The dynamic nature of gene-expression means that these effects can never be truly accounted for in an *in vivo* system, particularly in a clinical setting where laboratory conditions and standards cannot be maintained. Such unwanted variation can introduce bias into standard significance analysis (such as ANOVA), assigning significance to genes whose change in expression may be in response to variables other than the main experimental parameters.

A method for correcting for the effects of superfluous variation in analyses for significant changes in expression is therefore required, and this was the inspiration behind the development of *Envisage* (Chapter 3) (Robson *et al.*, manuscript in preparation). By including a wide range of variables in the model, variation in gene-expression can be better attributed to the different sources, thus reducing the estimated residual variability, and allowing a more precise assessment of treatment effects. This allows the effects of otherwise ignored sources of variation on gene-expression to be considered. *Envisage* provides a frame for the inclusion of such extenuating factors into standard analyses for significant gene-expression changes, thus providing a more accurate representation of the true biology of the system under scrutiny.

The techniques ustilised in the *Envisage* modelling package are by no means unique, and many other packages provide tools allowing for similar analyses (e.g.

the *limma* package available from Bioconductor (Smyth *et al.*), the microarray ANOVA (MAANOVA) package available as a package in R or Matlab (Wu *et al.*), and the Significance Analysis of Microarrays (samr) package in R (Tibshirani *et al.*)). One of the main advantages of *Envisage* is the simplicity of the interface, which allows analysis of complex data sets in a simple point and click manner. Also, the interface of *Envisage* with GS-GX makes this package particularly useful for use by non-statistical based users, who may not be familiar or comfortable with using other R-based packages for analysis of their data.

It is fair to say, however, that for more complex analyses, other packages that allow more complex data structures may be preferable. In particular, *Envisage* provides no means for providing contrast information, allowing direct comparisons between specific conditions in the analysis. However, *Envisage* is well suited for broad analyses of the significant effects in experiments where the effects of additional confounding variables cannot be well controlled (in particular, for clinical studies). In such experiments, where confounding effects cannot be designed into the structure of the data (such as though the use of experiment blocks), *Envisage* allows the inclusion of such terms into the analysis. However, for designed experiments, where covariate effects are designed into the experiment structure and treatment effects are balanced and orthogonal, other analysis methods (such as ANOVA) may be preferable.

Section 3.4.1 shows the results of a comparison between *Envisage*, and an ANOVA F-test across three variables. The limitations of ANOVA were described previously in Section 1.5.3.3, although ANOVA can be considered a 'gold standard' approach for the analysis of balanced and orthogonal experimental designs (with ANCOVA a natural extension when there are potential additional covariates). This offered a simple way to test that the automated model fitting procedure used in *Envisage* was suitable. In particular, *Envisage* was tested against ANOVA with a full saturated model specified, which should show identical results between the two for a balanced experimental design.

The removal of poor quality samples (Section 4.2.4) ensured that the data set was no longer balanced, and hence the different approaches (Type I SS for ANOVA,

Type II SS for *Envisage*) would no longer be expected to show identical results. Also, the model-selection procedure of *Envisage* meant that the number of genes showing p-values for each model term was often lower for *Envisage* than for ANOVA, although this was avoided by fitting a saturated model for both ANOVA and *Envisage*. For each model term, comparison between genes with associated p-values for both techniques showed good correlation between the two. As described, the majority of the differences were attributable to slight differences in the p-value assignment of each method, as erroneous genes showed p-values very close to the cutoff of $p = 0.05$. This indicates that the model fitting procedure produces accurate models to describe the data.

However, the main benefit of *Envisage* is its ability to control for the effects on gene-expression of variables not of primary interest to the analysis. The typical approach to avoid the effect of confounding sources of variation is to design these into the structure of the experimental setup, such that treatments are applied to blocks of samples showing similar covariate effects. In this way, covariate effects are made to be orthogonal to treatment effects, and effects relating to variables of interest can be distinguished from nuisance variables. Whilst phenotypic and genotypic discrepancies were largely avoided in this study through the use of age- and gender-controlled inbred mouse lines, and environmental effects were minimised by ensuring consistent humidity, temperature, lighting, etc. in the animal housing facility, variation existed in the technical processing of samples. Despite measures taken to minimise the detrimental effects of RNases during sample processing (for example, see Section 4.2.2), RNA extracted for microarray hybridisation was found to be of varying quality – particularly for the pancreas samples (Section 4.2.2.3). The yield of 2a-cRNA from the optimised IVT protocol was also found to be variable throughout the samples (Section 4.2.3.2), and may indicate initial sample quality. Also, given the logistics of processing forty eight samples, batching of samples was necessary.

The limited number of genotypically viable mice available for this study meant that design of experiment blocks (in particular to create blocks of litter mates) was difficult, and instead randomisation was performed across the entire sample set.

The choice of splitting samples into seven batches was based on the need for RNA to be processed on the Agilent Bioanalyzer in groups of twelve samples. The forty-eight samples considered in this experiment were a subset of eighty-four samples crossing two experiments which were processed in parallel, providing a natural choice of seven batches of twelve samples for processing. Had the forty-eight samples been processed independently, the design of the experiment may have leant itself to treatment in 8 batches, with each batch consisting of two sets of triplicate samples (for either skin and pancreas samples with the same treatment, or for 4OHT and vehicle treated samples for a single tissue). This experimental design would have allowed block design of nuisance variables such as the litter number, ensuring that such blocked effects were not confounded with the treatment effects of interest. To limit the introduction of systematic variation related to the treatment of samples within batches, randomisation was performed across all samples. Samples were continuously re-randomised at each stage of sample processing in an attempt to prevent the introduction of batch-specific effects, although it is possible that this approach may have actually worked to increase confounding variation within the data set.

RNA quality, 2a-cRNA yield and batch number were included in the modelling procedure in order to identify significant effects. RNA quality and 2a-cRNA yield showed significant effects on the expression of a relatively small number of genes, and this agreed with results seen during the data QC stage showing that the pre-hybridisation sample quality was not significantly correlated with data quality (Section 4.2.3.2). This is likely due to the 3'-bias of the probes on the microarrays, since RNA degradation occurs from the 5' to the 3' end. A larger effect was detected due to batching, although this was not deemed to be detrimental to the data. However, genes found to show a significant effect based on the batch were treated with caution. Importantly, any covariate effects were corrected by including these additional variables within the analysis. The majority of variation in the gene-expression was explained by the main variables of interest, indicating that this was a well designed and well implemented experiment, and that the results and conclusions drawn were relevant to the biological reality of the system.

However, this analysis identified several problems with the design of this experiment. Firstly, the detected batch effects indicated that the continuous re-randomisation approach used throughout this analysis was not the most suitable to remove nuisance batch effects. A block design may have been more suitable in this case. Randomisation, whilst important to remove systematic effects, is best performed within this block structure ("block what you can, randomise what you cannot" – Sir George E. P. Box). However, due to the small number of samples available for this experiment with the required genotype, blocking of littermates (to reduce genotypic difference between samples, a key source of confounding bias) proved to be difficult. Since many additional sources of confounding variation (gender, age, temperature, humidity, feeding times, etc.) were largely controlled by the design of the experiment and the conditions within the animal housing, blocking was not performed, and randomisation was performed across the entire sample set. The use of continuous re-randomisation was suggested as a method of ensuring that batch effects would add no significant bias to the data (personal communications with members of the UKAffy discussion group), but may have actually had an adverse effect on the introduction of nuisance variation in the data. In particular, the ability to remove such effects is relinquished by using this method.

Secondly, the use of covariate terms relating to RNA quality may not be suitable for including in this analysis. As seen in Figure 4.2.3 and Figure 4.2.5, there is a clear difference in RNA quality between the two tissues indicating that there may be a confounding effect between the tissue variable and the RNA quality covariates. This was particularly clear from the large reduction in the number of genes found to show a significant response to model interaction terms relating to tissue (4OHT:Tissue and Tissue:Time) when covariate effects were also included in the model.

Comparing the results of *Envisage* analysis when the 'Time' variable is treated as a factor compared to when it is treated as a numerical variable with a linear relationship identified further confounding effects in the data, particularly with the covariate terms included in the model. The model assumptions are inherently

different between these two methods, since numerical treatment of 'Time' assumes a linear relationship between gene-expression and 'Time' with 1 DF, whilst treating 'Time' as a factor allows for changes in gene-expression between time points with a 3 DF model. This results in a decrease in the number of significantly detected genes for the linear 'Time' model as it is unable to identify genes showing significant effects between time points, but not across the time course as a whole. Changing the model assumption for the 'Time' variable also resulted in an increase in the number of genes showing a significant effect for the covariate terms, suggesting that there may be some confoundment between these variables and 'Time'. However, no such confounding effect was detected for the 4OHT and 4OHT:Tissue terms which were of primary interest.

These results suggest that there were perhaps issues with the experimental design, although the decisions behind the design choices remain valid. These data also suggest that there were various confounding effects between the experimental variables, and in particular for the covariate terms. With this in mind, it may perhaps have been unwise to include these terms in the *Envisage* model fitting process. This perhaps highlights that the *Envisage* modelling procedure is not best suited for the analysis of designed experiments such as this, and in particular for studies where nuisance variables are accounted for in the block structure of the experimental design. One particular point to be aware of for such designed experiments is that the treatment and blocking variables included in the experimental design should always be included in the model. This can be ensured by specifying these terms as a minimal model when running the program, (although it must be noted that this was not done for the present analysis). *Envisage* therefore remains most useful for the analysis of experiments where nuisance variables cannot be easily designed into the structure of the experiment, such as for clinical studies where the effects of such covariates cannot be analysed ahead of time.

Despite these potential issues however, the *Envisage* analysis identified clear biologically relevant changes in expression in response to activation of the $MycER^{TAM}$ transcript in both the pancreatic β-cells and (to a lesser extent) the

suprabasal keratinocytes. These responses in gene-expression are described in the following sections.

5.1.3 Comparison of Transcriptional Response to MycERTAM Activation in Suprabasal Keratinocytes and Pancreatic β-Cells

5.1.3.1 Activation of MycERTAM promoted cell cycle entry in pancreatic β-cells and suprabasal keratinocytes

The proliferative response to MycER^{TAM} activation within the two distinct tissues is well documented, and was confirmed in this study through immunohistochemical staining for the proliferation marker Ki67 (Figure 4.2.1). The resulting expression changes detected in cell cycle related genes for the two tissues are discussed in the following sections.

Pancreatic β-cells

The transcriptional response of cell cycle genes in the pancreas were indicative of a strong proliferative response as described in Section 1.1.4, with key G_1/S-phase genes (e.g. Cyclin D2, Cyclin E and Cdk4) showing significant increases in expression within the first 4 hours and later G_2/M-phase genes (e.g. Cyclin A and Cyclin B) increasing subsequently from 8 hours and remaining up-regulated throughout the time course. Many of these changes represent putative Myc target genes, and were confirmed using qRT-PCR (Section 4.2.12).

Interestingly, the Cdk1 gene (*cdc2a*) showed a significant increase in gene-expression from 8 hours though to 32 hours (3-12 fold), and this was validated with qRT-PCR. However, the role for Cdk1 in this case remains unclear, since it has been implicated in: promoting proliferation by substituting with other CDKs, such as Cdk2 which showed no change in expression (Santamaria *et al.*, 2007); inhibiting apoptosis in Myc-transformed cell lines through phosphorylation of

Survivin (*birc5*) by Cyclin B/Cdk1 complexes (Goga *et al.*, 2007); and mitotic catastrophe and promotion of apoptosis following early activation of Cdk1 prior to G_2/M-phase (Castedo *et al.*, 2002).

The dramatic down-regulation seen in the CDKI *cdkn1b* ($p27^{Kip1}$) from 8 hours in the pancreas pointed towards Myc-mediated induction of proliferation. However, this result was intriguing, since loss of $p27^{Kip1}$ normally occurs through degradation of the protein when complexed with Cyclin E/Cdk2 (Section 1.1.4). Down-regulation at 4 hours of the CDKI *cdkn2c* ($p18^{Ink4c}$), which inhibits Cdk4 and Cdk6, also pointed towards G_1/S phase cell cycle entry. However, the subsequent 6-fold increase by 16 hours may indicate induction of cell cycle arrest prior to apoptosis. Importantly, a link has been shown previously between $p18^{Ink4c}$ and the Atr/Atm DNA damage response pathways described below (Park *et al.*, 2005). This study shows direct interaction between $p18^{Ink4c}$ and the Atr/Atm kinases, resulting in increased levels of p53, leading to growth arrest or apoptosis. An important role for p53 is detected in our data (shown below) following $MycER^{TAM}$ activation, including up-regulation of the p53 cell cycle arrest target *cdkn1a* ($p21^{Cip1}$).

These data support the idea that Myc-induced cell cycle progression occurs through direct activation of key cell cycle genes such as *ccnd1, ccnd2* and *ccne2*. Whilst it is not possible to infer specific Myc-induced transcription from the microarray data, the short time point at which these changes were seen (4 hours) would suggest a causal effect of Myc. Further changes in cell cycle related genes such as *ccna2, ccnb1, cdc2a* and *cdkn1b* ($p27^{Kip1}$), which showed expression changes from 8 hours of $MycER^{TAM}$ activation, may therefore occur downstream of Myc.

Also of interest was the increase in expression of genes such as *mcm2, mcm7* and *cdt1*, which are involved directly in DNA replication, within only 4 hours of $MycER^{TAM}$ activation. The level of co-expression seen in these genes following Myc expression may link to recent work suggesting a non-transcriptional role for Myc in the control of DNA replication (Dominguez-Sola *et al.*, 2007). However,

it is not possible to make any firm assertions regarding protein interactions from the expression data, so this remains to be seen.

Suprabasal keratinocytes

Despite the clear transcriptional role for MycER^{TAM} in proliferation in the β-cells, the gene-expression response for cell cycle-related genes in the suprabasal keratinocytes was less pronounced. The cyclin gene *ccnd2* was up-regulated greater than 2-fold at 8 hours and 32 hours, with no significant change detected for the other time points. This pattern of expression was detected for a number of genes, and can be seen in the normalised expression profiles of Figure 4.2.10. Peaks appeared to occur at 8 hours following daily 4OHT administration, which may indicate a delay in 4OHT treatment through the use of topical administration (Section 4.2.6).

Other cell cycle-related genes showing significant expression included *ccnd3* and *cdk4*, which both showed up-regulation at 8 hours. Also, as with the pancreas, a significant down-regulation was detected for the CDKI gene *cdkn1b* (p27^{Kip1}) throughout the time course. Given that changes to p27^{Kip1} in cell cycle progression normally occur at the protein level through ubiquitin-mediated protein degradation, this result was surprising. Validation using qRT-PCR identified a similar loss in expression; however this was of a lower magnitude than for the microarray.

No change in expression was detected at later time points, and further cell cycle related genes (e.g. *ccna*, *ccne* and *cdk2*) remained unchanged, which seemed contrary to the proliferative phenotype shown in Figure 4.2.1. This problem may have arisen due to the presence of RNA from non-*mycER^{TAM}*-expressing cells. Given the thin nature of the mouse epidermis (one or two cells thick only), this was unavoidable. Isolation of suprabasal keratinocytes using LCM proved impractical, and other methods (such as trypsin-based degradation of the dermis) were deemed too damaging to RNA. Given that the dermis contains very few cells in comparison to the epidermis (for example, see Figure 4.2.1b), and that these

cells likely undergo limited changes in gene-expression, it was decided that inclusion of dermal-derived total RNA would not prove problematic.

In fact, the problem likely arose due to the presence of basal-derived total RNA, since basal cells undergo cell cycle progression as part of normal homeostasis even in VT transgenic animals (Figure 4.2.1b). In the β-cells, no proliferating cells were detected in VT islets (Figure 4.2.1a), meaning that changes in gene-expression relating to cell-cycle progression were more specific to those cells induced to proliferate following $MycER^{TAM}$ activation. However in comparison, the presence of proliferating cells in the basal layer meant that it was difficult to attribute detected gene-expression changes solely to activation of $MycER^{TAM}$ in suprabasal keratinocytes. Given that 2-fold change in expression was typically used to identify differentially expressed genes, detected changes in the expression of key cell cycle genes may have fallen below cutoff values. Given the high levels of noise inherent to microarray studies, particularly with a maximum of only 3 replicate samples for each condition (a necessity due to budget constraints), identification of differentially expressed genes relating to the cell cycle between $MycER^{TAM}$-active and $MycER^{TAM}$-inactive samples was difficult.

Despite this, it may be argued that the presence of RNA from proliferating basal cells should not be detrimental to the detection of differential expression for genes relating to apoptosis and survival, as no apoptosis was detected within the basal or suprabasal cells of VT animals (Figure 4.2.1b). Thus comparing the transcriptional response to Myc-deregulation in the two tissues should still identify genes relating to apoptosis and survival whose response to $MycER^{TAM}$ activation is different for the two tissues. Such variation may relate to transcriptional events responsible for the dichotomous phenotypic response to Myc-deregulation. Comparison between the two systems was therefore still considered a valid analysis.

5.1.3.2 Activation of MycERTAM in vivo leads to up-regulation of apoptotic death pathways in pancreatic β-cells but not in suprabasal keratinocytes

To confirm the presence of apoptotic cells within the tissue samples of interest, sections were stained for the effector Caspase 3. No staining was detected for active Caspase 3 within the epidermis, indicating that apoptosis is not initiated within the suprabasal keratinocytes following activation of MycER^{TAM}. However, staining for Caspase 3 in pancreatic β-cells identified a clear apoptotic response after only 4-hours of MycER^{TAM} activation. The transcriptional response of each tissue to MycER^{TAM} activation for genes relating to apoptosis and cell survival is discussed here.

Pancreatic β-cells

Promotion of apoptosis in the β-cells was clear from the number of apoptosis genes identified as being significantly affected upon MycER^{TAM} activation. Of particular interest was the change seen in genes involved in the DNA damage checkpoint pathway, which has been previously implicated in Myc-induced apoptosis (Felsher and Bishop, 1999; Mai and Mushinski, 2003; Bartkova *et al.*, 2005; Gorgoulis *et al.*, 2005; Dominguez-Sola *et al.*, 2007). Genes whose products are involved in DNA repair, such as *rad51* and *h2afx*, and members of the 9-1-1 DNA repair marker complex, *rad1* and *hus1*, showed significant changes in expression early following MycER^{TAM} activation, indicating that early events in Myc-mediated apoptosis may occur as a result of DNA strand breaks.

Subsequent progression of apoptosis through the DNA damage response checkpoints appeared to follow the Atr pathway, and not the Atm pathway as has previously been described (Pusapati *et al.*, 2006; Maclean *et al.*, 2007). However, since the microarray data corresponds to transcriptional events, protein level activation of these pathways cannot be confirmed or rejected in our study, and it is not possible to exclude a role for Atm. *atr* and the related checkpoint kinase genes, *chk1* and *chk2*, showed significant up-regulation throughout the time course for both the microarray and qRT-PCR, suggesting a role for Myc in direct

transcriptional regulation of DNA response genes. *atr* has previously been identified as being responsive to Myc in human cells (Schlosser *et al.*, 2003), although further work is required to confirm these genes as direct Myc targets. Activation of the Atr-related DNA checkpoint pathways seen in these data corroborates current hypotheses of DNA damage as one of the defining events in Myc-induced apoptosis.

The apoptotic function of Myc likely occurs through Bax-mediated permeabilisation of the mitochondria, leading to release of Cytochrome c and activation of effector caspases such as Caspase 3. This is evident from the anti-apoptotic role of Bcl_{XL} in RM double transgenic mice (Pelengaris *et al.*, 2002b; Lawlor *et al.*, 2006), and from studies that have shown Bax to be essential for Myc-induced apoptosis (Eischen *et al.*, 2001; Soucie *et al.*, 2001; Juin *et al.*, 2002; Dansen *et al.*, 2006). In fact, loss of Bax is sufficient to allow rapid progression of invasive, angiogenic tumours following Myc deregulation (Dansen *et al.*, 2006).

A recent study using a transgenic $MycER^{TAM}$ mouse model in which Myc is expressed in the basal layer of the epidermis under the Keratin 5 promoter (K5-Myc) showed that deregulated Myc resulted in an increase of p53 resulting (at least to some degree) from activated Atm (Pusapati *et al.*, 2006). Loss of Atm in this model resulted in a significant decrease in apoptosis, and near-complete inhibition of apoptosis was detected in a p53-null background. This shows that deregulated Myc-induced apoptosis can be driven through DNA damage-related activation of the tumour suppressor p53, which promotes formation of the MAC by Bax and release of Cytochrome c into the cytosol. The data presented in this thesis fit well with this model, indicating that Myc induces apoptosis by resulting in activation of the DNA damage pathway. This leads to activation of Atr/Atm and checkpoint kinases Chk1 and Chk2, which ultimately phosphorylate and activate the tumour suppressor p53. However, we are aware that for such claims to be made, we would need to confirm activation of the Atm/Atr proteins, as well as downstream proteins such as Chk1, Chk2 and H2ax.

Gene-expression analysis suggests that active p53 likely has two complementary roles in this case: arresting cell cycle progression though CDKIs $p21^{Cip1}$ and $p18^{Ink4c}$, and initiating apoptosis through Bax. Whilst expression of *bax* was found to increase within 8 hours of $MycER^{TAM}$ activation, this change was not maintained throughout the time course. Also, other pro-apoptotic p53 target genes (such as the Bcl2 family members *bbc3, noxa, bid* and *bim*) showed no significant change in expression, suggesting that p53 transcriptional activity was not induced. Previous studies have shown that, while Bax is essential for Myc-induced apoptosis (Section 1.1.5), this is not related to altered *bax* expression (Soucie *et al.*, 2001; Juin *et al.*, 2002). This indicates that activation of Bax at the mitochondria does not occur as a result of p53-mediated transcription, but rather through some p53-dependent non-transcriptional mechanism.

This may occur due to the ability of activated p53 in the cytosol to act in a similar way to BH3-only Bcl2 family members, leading to direct activation of Bax at the mitochondrial membrane and/or release of pro-apoptotic factors from sequestration by Bcl_{XL} and Bcl2 (Mihara *et al.*, 2003; Chipuk *et al.*, 2004; Erster and Moll, 2005). The loss of membrane potential and the subsequent release of apoptotic factors from the mitochondria may also explain the subsequent up-regulation of mitochondrial proteins such as Cytochrome c and Endog throughout the time course, which may occur as a response to replenish lost mitochondrial stores.

These data also suggest a possible role for $p19^{Arf}$ in the Myc-induced apoptotic response, as expression of *cdkn2a* ($p19^{Arf}$) was found to be significantly increased throughout the time course. $p19^{Arf}$ may act to stabilise the active p53 tumour suppressor in the cytosol by inhibiting the p53-antagonist Mdm2. $p19^{Arf}$ has previously been shown to be implicated in Myc-induced apoptosis in the K5-Myc mouse model (Russell *et al.*, 2002), and has been shown to work in cooperation with DNA damage-related activation of p53 upon oncogenic stress (Pauklin *et al.*, 2005). However, the role of $p19^{Arf}$ in the suppression of Myc-induced tumourigenesis has more recently been linked to its ability to arrest cell cycle progression (Finch *et al.*, 2006). In this study, suppression of $p19^{Arf}$ in the *pins*-

mycER^{TAM} transgenic model identified no loss in apoptosis, but instead showed an increase in β-cell replication. This suggests that the primary role for p19^{Arf} following deregulation of Myc may be in suppression of proliferation. The fact that loss of p53 in this model resulted in marked reduction in Myc-induced β-cell apoptosis suggests that there may be another upstream promoter of p53 activity other than p19^{Arf} that is yet to be identified for this mouse model. One strong possibility is the Atm/Atr DNA damage pathways described above.

It originally appeared as if the Fas TNF pathway played a significant role in MycER^{TAM} induced apoptosis, suggesting that the extrinsic death pathways may work in conjunction with intrinsic mitochondrial pathways during Myc-induced apoptosis. A significantly large increase in the expression of several probe sets specific for the Fas receptor was seen, suggesting that Myc may sensitise cells to apoptotic signalling by increasing the number of available Fas receptors in the cell membrane. Unfortunately, this result turned out to be incorrect due to an error in probe annotation (Section 4.2.12.10). In fact, of the 5 probe sets on the array shown to relate to the Fas receptor, the single probe set that showed no change in expression upon MycER^{TAM} activation was the only one to be specific for the *fas* transcript. This error in annotation appears to be recent, since up-to-date annotation files for both Bioconductor and GS-GX were used during analysis. This highlights the need to ensure correct and up-to-date annotations are maintained at all times. Also, this suggests that the Fas death receptor is not involved in MycER^{TAM}-induced apoptosis in the pancreatic β-cell model, and instead that the intrinsic mitochondrial pathway is the key route to Myc-induced cell death.

Suprabasal keratinocytes

In contrast to the islets, no change in expression was detected in genes involved in DNA damage response (e.g. *atr, chk1* or *chk2)* within suprabasal keratinocytes. Also, no change in expression was detected for the p53-pathway mediator $p19^{Arf}$, although a significant increase of 4-fold was detected using qRT-PCR at 4 hours following $MycER^{TAM}$ activation. Since these data correspond to changes at the transcriptional level, further analyses would need to be performed to confirm or discount Atr/Atm pathway protein activation within these cells. In particular, the data described above for the K5-Myc basal epidermis model indicates both the $p19^{Arf}$ (Russell *et al.*, 2002) and Atm (Pusapati *et al.*, 2006) pathways are activated to induce keratinocyte apoptosis *in vivo* following Myc-deregulation. Although apoptosis is not evident in the *inv-mycER*^{TAM} model, it is possible that Myc over-expression leads to activation of $p19^{Arf}$ and/or Atm/Atr at the protein level, leading to p53 activation. One indication that p53 is indeed active in the skin model is the significant and sustained increase in expression of the G_2/M DNA damage checkpoint gene *gadd45g*, a known p53 target which induces growth arrest in the G_2/M transition stage.

Of particular interest to this study were the cell survival genes *igf1*, *akt1*, and *akt2*. Increased expression of these genes through the time course as compared to the β-cells suggests that evasion of apoptosis may occur through the Igf1-Igf1r-mediated Akt pathway. Activation of this pathway results in inactivation of the pro-apoptotic Bcl2 protein Bad, allowing Bcl2 and Bcl_{XL} to bind to and inhibit Bax at the mitochondria (Figure 1.1.3). This prevents Bax-mediated release of Cytochrome c from the mitochondria and allows cells to continue to proliferate unchecked.

These data provide some evidence towards the role of the Igf1 survival pathway in suprabasal keratinocytes following $MycER^{TAM}$ activation, with 2-fold up-regulation of Igf1 detected in the skin from 8 hours through to 32 hours. Confirmation of results using qRT-PCR identified a lesser response in gene-expression change for *igf1*, although given the increased sensitivity of the technique this may still be of a significant magnitude. Although no change in

expression was seen for the Igf1 receptor, proteomic studies on the *inv-mycER*^{TAM} model within the group have identified an increase in the Igf1r protein in suprabasal epidermis (data not shown).

Increased expression was also detected for the survival factor genes *akt1* and *akt2*, which showed up-regulation in the keratinocytes from 8 hours. However, this was not confirmed using qRT-PCR (Figure 4.2.25). Despite this, it is activation of the Akt protein that would indicate activation of the Igf1 signalling pathway, and this is currently under investigation.

Apoptosis is prevented in suprabasal keratinocytes following deregulation of Myc. However, it is possible that terminal differentiation of keratinocytes itself serves as a tumour suppressive function. It is possible that Myc induces the Atm/Atr DNA damage response pathway to promote cell cycle arrest to control against aberrant proliferation in order to allow repair of damaged DNA. Cells in which DNA repair is insufficient remain maintained within the epidermis. However, whilst such an event would be deleterious for many internal tissues such as the pancreas, this is not the case for skin epidermis, where Myc-activated suprabasal keratinocytes will ultimately be sloughed off the skin surface as denucleated squames (Figure 1.2.2).

5.1.3.3 Activation of MycERTAM resulted in loss of differentiation markers in pancreatic β-cells and suprabasal keratinocytes

Activation of MycER^{TAM} resulted in a clear loss in expression of differentiation marker genes for both the skin and the pancreas. Most obvious amongst these were the cell-specific genes for Insulin (*ins2*) in the β-cells and Involucrin (*inv*) in the suprabasal keratinocytes, which both showed an early loss in expression following MycER^{TAM} activation. It is interesting to note that, since the MycER^{TAM} construct is placed under the control of the *ins2* and *inv* gene promoters for the pancreas and the skin respectively, expression of the *pins-mycER*^{TAM} and *inv-mycER*^{TAM} transgenes may also be similarly affected. In this way, activation of MycER^{TAM} may ultimately lead to a reduction in MycER^{TAM} production, resulting

in a feedback loop. MycER^{TAM} levels are typically high in the cells of these transgenic mice (Littlewood *et al.,* 1995). Despite the fact that the MycER^{TAM} protein has only a short half-life, it was assumed that cellular levels of the MycER^{TAM} protein remained suitably high throughout the short time course even with a reduction in *mycER^{TAM}* expression (although this was not analysed explicitly). It was believed that this change in MycER^{TAM} production did not have any significant negative effect on the data, however such a feedback loop may explain the bimodal expression profile typically seen of genes throughout this experiment, particularly for the skin (Figure 4.2.10).

Pancreatic β-cells

Comparison of the pancreas data set with the gene markers for pancreas development identified by Gu *et al.* (2004) showed that several genes relating to β-cell function were down-regulated. Many of these related to Glucose sensing and Insulin secretion, indicating that MycER^{TAM} plays a role not just in decreasing β-cell numbers through apoptosis, but also in inhibiting β-cell function. These effects cooperate to deregulate Glucose metabolism, leading to hyperglycaemia. Interestingly however, after the initial drop in Insulin expression, levels proceeded to increase dramatically for the 16 hour time point. A similar change in expression was also seen in various genes relating to Glucose sensing and Insulin secretion. Further analysis of a later time point at 72 hours showed that after 3 days of MycER^{TAM} activation, Insulin expression levels once again fell to below those of VT controls (Figure 4.2.13). This indicated that this period of high *ins2* expression was limited to the first few days of MycER^{TAM} activation.

A short period of hypoglycaemia within the first 24 hours or so of MycER^{TAM}, which correlates with increased *ins2* expression levels, has previously been noted (manuscript in preparation). The reasons behind this remain unclear, although the suggested hypothesis is that the sudden onset of apoptosis within the β-cells results in an influx of Insulin into the blood stream from acutely apoptotic cells (Cano *et al.*, 2007). However, the results seen here identified an increase not only in Insulin levels in the blood, but also in *ins2* expression at the transcriptional level. This hypoglycaemic window is also seen in 4OHT-treated RM mice

(manuscript in preparation), suggesting that this phenotype is not related to MycER^{TAM}-induced apoptosis. This indicates that the proposed hypothesis is incorrect, and work is currently underway to understand this phenomenon.

Suprabasal keratinocytes

The role for Myc in modulating both proliferation (Hashiro *et al.*, 1991) and terminal differentiation (Gandarillas and Watt, 1997) in the epidermis has been previously described. The microarray study of Frye *et al.* (2003) showed that activation of Myc in the basal keratinocytes using the *k14-mycER^{TAM}* transgenic model results in basal cells leaving the stem cell compartment and entering a program of terminal differentiation. However, activation of Myc in the suprabasal keratinocytes – which are already in the process of terminal differentiation – results in cells exiting from their differentiated state and re-entering the cell cycle (Pelengaris *et al.*, 1999). This can be seen in the present study by early down-regulation of the key differentiation marker *inv*, the suprabasal-specific *krt1*, and genes relating to cell adhesion, cytoskeleton formation and structural integrity such as *csta* and the Integrin genes *itga7*, *itgb2* and *itgb6*.

However, increased expression was also detected in several genes relating to keratinocyte differentiation, including the *plec1* gene, whose product is involved in formation of the cytoskeleton and maintenance of structural integrity (Svitkina *et al.*, 1996; Wiche, 1998), and *tgm2*, whose product is involved in the formation of covalent bonds throughout keratinocyte differentiation. Also, the suprabasal-specific Keratin 1 gene, *krt1*, actually showed an increase in expression at later time points. However, it is possible that these discrepancies are due to the small number of cells induced to proliferate within the short time course considered. These data indicate that activation of MycER^{TAM} results in a decrease in the differentiated state of keratinocytes, possibly indicating that Myc-induced suprabasal keratinocyte proliferation is incompatible with terminal differentiation.

5.1.3.4 Comparison between the skin and the pancreas

Analysis of the temporal pattern of gene-expression following MycER^{TAM} activation allowed identification of functionally related genes by using clustering algorithms (Section 1.5.4). By identifying co-regulated genes showing varying profiles between the two tissues, it was hoped that key functional divergences in the functional role of Myc may be discovered. QT clustering identified a number of co-regulated genes involved in a number of cellular functions (Section 4.2.11).

A high level of similarity was seen in genes involved in key cell cycle events, particularly DNA replication, spindle formation, organelle localisation, cytoskeleton organisation and division. DNA damage genes *atr* and *chk1* were found to have similar expression profiles, and these were clustered well with MCM genes involved in DNA unwinding. The expression of these genes was also closely linked to similarly expressed genes involved in DNA repair, indicating close linkage between the proliferative action of Myc and the DNA damage response mechanism. These changes were detected for the pancreas, but not for the skin. However, it is not clear whether this represents a fundamental divergence in function between the two tissues, or whether the presence of replicating basal keratinocytes may influence the detection of changes in DNA damage response genes in suprabasal keratinocytes.

Of particular interest was the co-regulation seen in several kallikrein genes, which were found to be highly expressed in the skin but not in the pancreas. This change in expression was very large in comparison to other genes, so clearly represented a significant effect. The role of these proteins in cell survival – specifically in the Igf1 pathway through degradation of the Igf1 inhibitor, Igfbp3 – suggests a possible role in survival for the suprabasal keratinocytes. The role of Igf1 in the prevention of Myc-induced apoptosis in fibroblasts has been previously noted (Harrington *et al.*, 1994a), and Igf1 is well described as a factor in cell survival during tumour growth, particularly in breast cancer (Bonneterre *et al.*, 1990; Ellis *et al.*, 1998; Subramanian *et al.*, 2007) and pancreatic cancer (Ohmura *et al.*, 1990; Bergmann *et al.*, 1995; Min *et al.*, 2003; Stoeltzing *et al.*, 2003; Zeng *et al.*,

2003). Therefore, the increase in expression of Igf1 detected in the suprabasal keratinocytes following MycER^{TAM} activation, together with proteomic evidence of an increase in the Igf1 receptor (data not shown), suggests a key role in determining response to MycER^{TAM} activation.

Kallikreins such as these also play a role in control of tissue micro-environment, degradation of the extracellular matrix through activation of MMPs. Similar co-expression was detected for the angiogenesis inducing placental growth factor gene *pgf*, which requires extracellular matrix degradation to allow vascular growth. These results suggest that prevention of apoptosis may occur due to an increase in the levels of kallikrein proteins in the extracellular space, which may also link to induction of vascularisation seen in the *inv-mycER^{TAM}* transgenic model. These may facilitate tumourigenesis and neovascularisation by degrading the extracellular matrix, and may also act to degrade Igf1-inhibiting proteins such as Igfbp3. This allows Igf1 to bind to its receptor and initiate survival pathways (e.g. through Akt), leading to inhibition of Bax-mediated Cytochrome c release from the mitochondria. The fact that up-regulation of Igf1 was seen for both the skin and the pancreas (as evidenced in the qRT-PCR results – Figure 4.2.32) may suggest that this plays a major role in determining cell survival, but is kept in check through inhibitory mechanisms that must be avoided to prevent apoptosis.

However, this proposed mechanism raises two important questions:

1. If suppression of Myc-mediated apoptosis occurs at the mitochondria through inactivation of Bax, why do we not see earlier apoptosis-related expression changes (such as DNA-damage related genes) in the skin?
2. Why are the kallikrein genes up-regulated in the skin but not in the pancreas?

If the point at which apoptosis was evaded occurred downstream of p53 activation, evidence of activation of the DNA damage response would be expected in the suprabasal keratinocytes. However, whilst no significant change in expression was detected for the DNA damage response genes in the skin, this does not rule out activation of the pathway at the protein level, through

phosphorylation of Atr/Atm and subsequent checkpoint kinases Chk1 and Chk2. Protein-level analysis remains to be performed to confirm or reject the role of the DNA damage pathway in suprabasal keratinocytes.

The question as to why kallikreins were activated in the skin but not in the pancreas may well be central to the dichotomous phenotypic outcome of MycER^{TAM} activation between the two tissues. Given that Myc usually acts as a rather weak transactivator, with changes in expression usually of only a few fold changes (Grandori and Eisenman, 1997; Cole and McMahon, 1999), it is likely that the kallikrein genes (which showed very large changes in expression) are under the influence of some other regulatory factor. One possible explanation for the difference in activity between the tissues may come from the role of steroid hormones, such as estrogen, in regulating kallikrein expression (Smith *et al.*, 1992; Borgono and Diamandis, 2004; Rajapakse *et al.*, 2007). Estrogen is known to be very active within the skin, with roles in wound healing and regulation of epidermal thickness (Bolognia, 1995; Azzi *et al.*, 2005). It may be the case that estrogen activity within the β-cells is insufficient to protect against Myc-induced apoptosis. In fact, it has previously been shown that estrogen can protect against oxidative stress-induced apoptosis in β-cells in mice by associating with estrogen receptor alpha (ERα) (Le May *et al.*, 2006).

The idea that estrogen levels may influence the phenotypic outcome may also be supported by observations that male mice suffer more severe hyperglycaemia than female mice following MycER^{TAM} activation (data not shown). This may suggest that the higher levels of estrogen in female mice alleviate MycER^{TAM}-induced apoptosis somewhat, resulting in fewer apoptotic cells. However, studies are yet to confirm that the number of apoptotic β-cells is greater for male mice than for females, and all mice used throughout this study were male to ensure that between-sample variation was minimised.

However, one further confounding issue with using the 4OHT driven MycER^{TAM} transgenic model is the fact that synthetic analogues of estrogen such as 4OHT have been shown to activate the endogenous estrogen receptor (Smith *et al.*, 1997; Dudley *et al.*, 2000). It is therefore possible that these skin-specific changes in

gene-expression relate to activation of estrogen-responsive pathways by ectopically administered 4OHT as opposed to endogenous estrogen. To rule out this effect, further studies would need to be performed, comparing the resulting changes in gene-expression following activation of MycER^{TAM} by 4OHT with those in 4OHT-treated WT littermates.

These data suggest a possible mechanism by which the tissue-specific environment may play a role in determining the fate of cells following Myc deregulation (Figure 5.1.1). It is important to note that changes at the transcriptional level make up only a small part of the many changes occurring within the cell in response to Myc deregulation. Details of protein-level changes, post-translational modifications, or epigenetic modifications of DNA are beyond the scope of this analysis. However, these results provide several avenues for future research, and further analysis into the roles of estrogen, kallikreins and Igf1 may lead to improved understanding of the survival pathways that can serve to augment the oncogenic role of Myc.

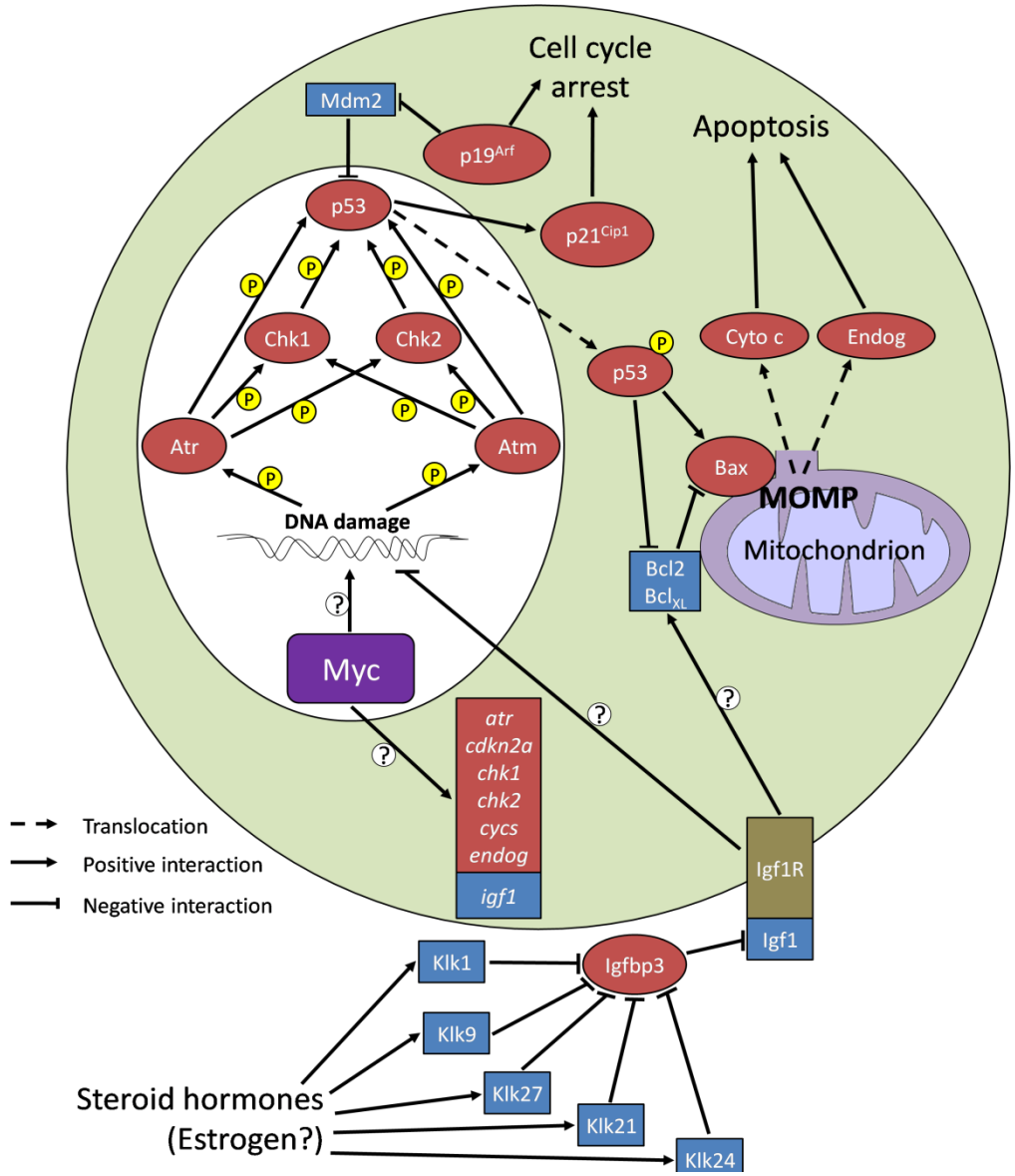

Figure 5.1.1: Proposed mechanism for Myc-induced apoptosis and survival

Activation of MycER^{TAM} leads to unchecked proliferation in suprabasal keratinocytes, but predominantly apoptosis in pancreatic β-cells. Shown here is a proposed mechanism for the tissue-specific determination of cell fate following Myc deregulation. Initiation of the apoptotic machinery following activation of MycER^{TAM} appears to occur through activation of the DNA damage checkpoint pathway, leading to activation of the p53 tumour suppressor. The route by which Myc is able to promote DNA damage is not clear, but the data may suggest a direct role for Myc in transcriptional regulation of key DNA damage genes *atr*, *chk1* and *chk2*. Gene-expression results suggest there is no transcriptional response in *atm*, however activation of Atm may be involved at the protein level. Activation of p53 prevents association with the p53-inhibitor Mdm2, allowing accumulation of the stable protein. This process may be further assisted by p19^{Arf} (a previously categorised Myc target gene), which binds to and inhibits the Mdm2 protein. It is also

possible that the role for $p19^{Arf}$ in Myc-induced apoptosis is through its ability to promote cell cycle arrest, along with the p53-activated $p21^{Cip1}$. As well as activating cell cycle, p53 may be involved in Bax-activated MOMP. The lack of expression change seen for pro-apoptotic p53-target genes suggests that the predominant role for activated p53 is in direct activation of Bax or inhibition of Bcl2 or Bcl_{XL} at the mitochondria. Activated Bax forms the MAC, allowing release of apoptotic factors such as Endog and Cytochrome c into the cytosol. This is highlighted by the increase in expression of the corresponding genes, *endog* and *cycs*, which may suggest transcriptional replenishment of lost mitochondrial proteins. Given the rapid change in expression of these genes following $MycER^{TAM}$ activation, this may suggest a direct role in transcriptional regulation of these genes by Myc. Cytochrome c induces apoptosis through activation of a caspase cascade, culminating in activation of the effector Caspase 3, and destruction of the cell. This is further aided by Endog-mediated degradation of DNA. Activation of Bax is regulated by both pro- and anti-apoptotic members of the Bcl2 superfamily. Bcl2 and Bcl_{XL} act to prevent formation of the MAC by associating with and inhibiting Bax proteins. Increased expression of Igf1 in both tissues suggests that survival pathways are also activated by Myc. This may occur through the Akt pathway, culminating in inhibition of Bcl2 and Bcl_{XL}. Whilst no change in these key DNA damage-responsive genes was detected in the skin, the activation of the DNA damage response at the protein level cannot be discounted. However, it is also possible that survival signals act to prevent Myc-induced DNA damage. The Igf1 protein is prevented from binding to its receptor and activating survival pathways by the Igf binding protein Igfbp3. It is suggested that Igfbp3 prevents Igf1 from activating survival signals in the pancreas, allowing suppression of tumour growth through apoptosis, whilst activation of $MycER^{TAM}$ in the skin leads to a large response in kallikrein genes, whose products are known to degrade Igfbp3. Kallikrein genes are transcriptionally activated by steroid hormones such as estrogens, which are found in large numbers within the epidermis, suggesting a possible tissue environment-specific divergence in transcriptional response following $MycER^{TAM}$ activation.

5.2 Conclusions

The fate of a cell following deregulation of Myc is dependent on cell type and location. However, the pathways responsible for deciding the ultimate fate of the cell between 'life' (uncontrolled proliferation) and 'death' (i.e. apoptosis) *in vivo* are not yet clear. The MycER^{TAM} transgenic mouse model allows controlled deregulation of Myc in adult mice, enabling tracking of early changes downstream of aberrant Myc activity. High throughput transcriptional profiling was used to identify transcriptional events that may explain the disparity in the phenotypic response to MycER^{TAM} activation in the suprabasal keratinocytes (proliferation) and pancreatic β-cells (apoptosis). Gene-expression profiling of the pancreatic β-cells identified the DNA damage checkpoint pathway as the likely route by which Myc-mediates apoptosis in this system, resulting in activation of p53 and release of the pro-apoptotic factor Cytochrome c from the mitochondria. A role for the tumour suppressor p19^{Arf} was also detected, although recent studies have suggested that this may relate to Myc-induced cell cycle arrest, and not to p53 stabilisation and apoptosis as previously hypothesised. Whilst a direct transcriptional role for these expression changes cannot be inferred from these data, the suggested mechanism fits well with previous studies in the field, and suggests a key role for Myc in inducing DNA damage and abrogating the apoptotic response through induction of further pro-apoptotic factors.

Comparative analysis between the two tissues suggests an important role for the Igf1 survival pathway in determining ultimate cell fate. The key discrepancy between the two systems appears to be the presence of members of the kallikrein serine protease family, which were dramatically up-regulated throughout the time course. These enzymes have been implicated both in degradation of the extracellular matrix facilitating angiogenic growth (a consequence of MycER^{TAM} activation present in the suprabasal keratinocytes and not in the β-cells), and in degrading Igf1r antagonists such as Igfbp3. In this way, Klks may work to promote survival signalling through Igf1, leading to inhibition of apoptosis at the mitochondria and allowing the tumourigenic potential of deregulated Myc to be

realised. Whilst at this stage we can only speculate, it is proposed that this difference in Klk activity may occur as a result of high levels of steroid hormones, such as estrogen, which are responsible for regulating Klk expression. A mechanism is therefore proposed by which the tissue-specific environment may influence the role of Myc in determining cell fate (Figure 5.1.1), however this remains to be investigated further in future studies.

The decision for a cell to become apoptotic depends on the complex interactions of many pro- and anti-apoptotic factors. The comparative levels of these factors may ultimately determine the fate of a cell. Different tissues may exhibit varying levels of these factors, resulting in a seesaw effect as pro- and anti-apoptotic factors compete for dominance. However, it also appears that tissue-specific environmental characteristics can affect the interaction between these factors, having a decisive effect on cell fate. The cellular events that can tip the balance one way or another during oncogenic stress are thus of great importance, as understanding the circumstances under which cells can bypass this defensive apoptotic response and allow tumour growth may lead to the discovery of new targets for therapeutic intervention.

This study has aimed to identify tissue-specific transcriptional events that may allow cells to evade the apoptotic effects of tumour suppression. We have sought to determine the factors that allow a cell to determine whether to continue life despite irregularity in Myc function, or to destroy itself to protect the organism. In some sense, we have aimed to determine the 'suicide note' for the β-cells – why have they resigned to destroy themselves whilst the suprabasal keratinocytes are happy to continue to survive despite the detrimental effect of Myc?

To answer this, a complete and thorough microarray experiment was designed and implemented to study the function of the cells at the transcriptional level. Experimental protocols were significantly optimised to ensure data quality was maximised; a thoroughly considered quality control procedure was utilised to ensure erroneous data were identified and removed; the analysis tool *Envisage* was designed and employed to specifically answer the questions central to the experiment; a thorough analysis of the resulting data was performed to identify

potential mechanisms indicative of the discrepancy in the phenotypes; and several results of interest were analysed (and in many cases, successfully validated) using qRT-PCR.

A possible mechanism whereby tissue-specific environmental factors may influence cell fate following Myc deregulation has been proposed, hypothesising that the decision to live or die may relate to tissue-specific environmental factors. However, this remains speculation as the approach taken here gives an insight into only one aspect of the changes occurring within the cell in response to Myc deregulation. Much remains to be learnt from analysis of protein-level changes, post-translational modifications, or epigenetic modifications of DNA. This study has identified several new lines of investigation for future analysis into the dual roles of Myc in apoptosis and survival. It is hoped that such studies will prove fruitful and provide further insight into the complex role of this enigmatic protein.

5.3 Further work

The design and implementation of *Envisage* largely focused on the ability to identify and correct for effects on gene-expression of sources of variation within an experiment that are not of primary interest to the researcher. This was influenced by the highly variable biological background of replicate samples in clinical studies. However, in its current form, the algorithm used in the model fitting procedure remains quite simple and may yet be improved.

Currently, one problem with this method is that estimates of the variance for each group may be sensitive to outliers, particularly if only a few replicate samples are used (as is often common with microarray experiments). The number of arrays used in an experiment will typically be much fewer than the number of genes analysed. Inclusion of methods such as the Empirical Bayes approach utilised by the *limma* package in R (Smyth, 2004; Smyth, 2005), or a non-parametric re-sampling method such as the bootstrap or jack-knife methods (Efron, 1981), would allow more stable estimates for the variance by "borrowing" information on variance across the genes as well as across the samples.

There are also several issues with the model fitting approach which must be considered and improved. Firstly, since the model fitting approach of *Envisage* results in only a subset of the model terms included in the selected model for some genes, the multiple testing correction is applied to a different number of genes for each model term. This may result in biased estimates for the adjusted p-values. Another issue with the *Envisage* package is that it is currently quite simple, and is not able to account for more complex experiment designs or multiple error strata. This may thus limit the utility of *Envisage* to only a limited number of experimental designs. In addition, there are currently no methods for the inclusion of contrasts into the analysis approach, so whilst it is possible to analyse the variables that show significant effects, it is not possible to further analyse the effects of each specific factor level. This makes *Envisage* useful for broad analyses of the most significant variables within a dataset, but not for more

specific analyses of the conditions under which significant changes are detected. Such contrast specification utilities may yet be added to a future release.

Also, the use of the AIC in the model fitting procedure may be overly liberal, adding terms with small regression coefficients that may lead to unnecessarily complex models. A model fitting criterion such as the least absolute shrinkage and selection operator, or LASSO (Tibshirani, 1996; Tibshirani, 1997), may instead be preferable. This minimises the SS, as described in Section 1.5.3.3, whilst imposing a limit on the absolute sum of the coefficients of Equation 1-13 such that $\sum_{m=1}^{p} |\beta_m| < s$ for some threshold value s. This method essentially shrinks the coefficients of some terms while setting others to zero, limiting the number of terms and ensuring parsimony in the fitted model.

Another addition to the package would be to include some method of clustering results within the model-fitting framework. It may be assumed that genes showing similar effects from the model terms are functionally related, providing a method of clustering to infer functionally related genes based on similarities between the fitted models. This clustering method may allow identification of genes showing similar regression coefficients, which may indicate co-expression in the system under analysis. Such improvements were unfortunately not implemented due to constraints on time, however remain viable options for the future.

Much still remains to be discovered within the microarray data set. The temporal nature of the experiment was originally designed with identification of network structure in mind. Whilst it was decided that such an analysis was outside of the scope of this project, current collaborations with members of the Engineering Department at the University of Warwick will hopefully yield interesting results in the future.

The role of the DNA damage response pathways in MycER^{TAM}-induced apoptosis remains to be studied in more detail. Validation studies using qRT-PCR have certainly strengthened the evidence pertaining to a DNA damage response being instrumental in Myc-induced apoptosis, and this corroborates the work of other researchers within the field. However, further functional studies may yet be

performed to confirm these results within this system, and in particular to identify the presence/absence of the DNA damage response in suprabasal keratinocytes. The central role of Atr and the checkpoint kinases, and the role of the p53-stabilising tumour suppressor $p19^{Arf}$ in particular may provide interesting targets for gene knockdown approaches such as RNA interference, which may functionally validate their roles in this process. It is of particular interest to understand the role played by $p19^{Arf}$ in this system, whether as a stabiliser of p53 activity, or as an inducer of cell cycle arrest.

Possible routes for the bypass of the apoptotic functions of Myc may involve the survival factor Igf1, and the kallikrein family of serine proteases, which have all been shown to exhibit significant changes in their expression within this experiment. Further functional validation must be performed in order to make firm assertions regarding their role in the dichotomy of $MycER^{TAM}$-activation in these diverse tissues. Studies are currently underway regarding the role of the Igf1 and Igf1 receptor in the prevention of Myc-induced apoptosis, which have also been identified through proteomics analyses performed within the Michael Khan group (data not shown). Of particular interest will be the role (if any) of estrogens and the kallikrein family members in determining cell fate. Functional studies, perhaps using kallikrein knockout mice or RNA interference techniques to suppress specific mRNAs, may help to elucidate the mechanisms involved in the regulation of these pathways. In particular, the specific role played by Myc in these pathways remains to be seen.

Bibliography

Acehan, D., Jiang, X., Morgan, D. G., Heuser, J. E., Wang, X. and Akey, C. W. (2002). *Three-dimensional structure of the apoptosome: implications for assembly, procaspase-9 binding, and activation.* Mol Cell, **9**(2): 423-32.

Adams, J. M. and Cory, S. (2007). *Bcl-2-regulated apoptosis: mechanism and therapeutic potential.* Curr Opin Immunol, **19**(5): 488-96.

Adams, J. M., Harris, A. W., Pinkert, C. A., Corcoran, L. M., Alexander, W. S., Cory, S., Palmiter, R. D. and Brinster, R. L. (1985). *The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice.* Nature, **318**(6046): 533-8.

Affymetrix (1999). *Microarray Suite User Guide: Version 4*. Affymetrix, Santa Clara, CA: http://www.affymetrix.com/support/technical/manuals.affx.

Affymetrix (2002a). *Data Analysis Fundamentals*. Affymetrix,

Affymetrix (2002b). *GeneChip® Expression Analysis: Data Analysis Fundamentals*. Affymetrix, Santa Clara, CA: http://www.affymetrix.com/support/downloads/manuals/data_analysis_fundamentals_manual.pdf.

Affymetrix (2002c). *Human genome U95 data set*. Affymetrix, Santa Clara, CA: http://www.affymetrix.com/support/technical/sample_data/datasets.affx.

Affymetrix (2003). *Technical Note: Array Design and Performance of the GeneChip® Mouse Expression Set 430*. Affymetrix, Santa Clara, CA: http://www.affymetrix.com/support/technical/technotes/mouse430_technote.pdf.

Affymetrix (2004). *GeneChip® Expression Analysis Technical Manual*. Affymetrix, Santa Clara, CA: https://www.affymetrix.com/support/file_download.affx?onloadforward=/support/downloads/manuals/expression_ever_manual.zip.

Agar, N. S., Halliday, G. M., Barnetson, R. S. and Jones, A. M. (2003). *A novel technique for the examination of skin biopsies by laser capture microdissection.* J Cutan Pathol, **30**(4): 265-70.

Agilent Technologies (2007). *Agilent 2100 Bioanalyzer*: http://www.chem.agilent.com/temp/rad9E880/00000445.PDF.

Ahn, J. Y., Schwarz, J. K., Piwnica-Worms, H. and Canman, C. E. (2000). *Threonine 68 phosphorylation by ataxia telangiectasia mutated is required for efficient activation of Chk2 in response to ionizing radiation.* Cancer Res, **60**(21): 5934-6.

Ahn, Y. B., Xu, G., Marselli, L., Toschi, E., Sharma, A., Bonner-Weir, S., Sgroi, D. C. and Weir, G. C. (2007). *Changes in gene expression in beta cells after islet isolation and transplantation using laser-capture microdissection.* Diabetologia, **50**(2): 334-42.

Akaike, H. (1974). *A new look at the statistical model identification.* IEEE Transactions on Automatic Control, **19**(6): 716-723.

Alarcon, R. M., Rupnow, B. A., Graeber, T. G., Knox, S. J. and Giaccia, A. J. (1996). *Modulation of c-Myc activity and apoptosis in vivo.* Cancer Res, **56**(19): 4315-9.

Alberti, K. G., Christensen, N. J., Christensen, S. E., Hansen, A. P., Iversen, J., Lundbaek, K., Seyer-Hansen, K. and Orskov, H. (1973). *Inhibition of insulin secretion by somatostatin.* Lancet, **2**(7841): 1299-301.

Alnemri, E. S., Livingston, D. J., Nicholson, D. W., Salvesen, G., Thornberry, N. A., Wong, W. W. and Yuan, J. (1996). *Human ICE/CED-3 protease nomenclature.* Cell, **87**(2): 171.

Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D. and Levine, A. J. (1999). *Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays.* Proc Natl Acad Sci U S A, **96**(12): 6745-50.

Alonso, L. and Fuchs, E. (2003). *Stem cells of the skin epithelium.* Proc Natl Acad Sci U S A, **100 Suppl 1**: 11830-5.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). *Basic local alignment search tool.* J Mol Biol, **215**(3): 403-10.

Alwine, J. C., Kemp, D. J. and Stark, G. R. (1977). *Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes.* Proc Natl Acad Sci U S A, **74**(12): 5350-4.

Amati, B. (2001). *Integrating Myc and TGF-beta signalling in cell-cycle control.* Nat Cell Biol, **3**(5): E112-3.

Amati, B., Alevizopoulos, K. and Vlach, J. (1998). *Myc and the cell cycle.* Front Biosci, **3**: d250-68.

Amati, B., Brooks, M. W., Levy, N., Littlewood, T. D., Evan, G. I. and Land, H. (1993a). *Oncogenic activity of the c-Myc protein requires dimerization with Max.* Cell, **72**(2): 233-45.

Amati, B., Dalton, S., Brooks, M. W., Littlewood, T. D., Evan, G. I. and Land, H. (1992). *Transcriptional activation by the human c-Myc oncoprotein in yeast requires interaction with Max.* Nature, **359**(6394): 423-6.

Amati, B., Littlewood, T. D., Evan, G. I. and Land, H. (1993b). *The c-Myc protein induces cell cycle progression and apoptosis through dimerization with Max.* Embo J, **12**(13): 5083-7.

Ambrosini, G., Adida, C. and Altieri, D. C. (1997). *A novel anti-apoptosis gene, survivin, expressed in cancer and lymphoma.* Nat Med, **3**(8): 917-21.

Applied Biosystems (2005). *Real-Time PCR Systems: Applied Biosystems 7900HT fast real-time PCR system and 7300/7500 real-time PCR systems chemistry guide*. Applied Biosystems, Foster City, CA: http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_042681.pdf.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. and Sherlock, G. (2000). *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, **25**(1): 25-9.

Askew, D. S., Ashmun, R. A., Simmons, B. C. and Cleveland, J. L. (1991). *Constitutive c-myc expression in an IL-3-dependent myeloid cell line suppresses cell cycle arrest and accelerates apoptosis.* Oncogene, **6**(10): 1915-22.

Ayer, D. E., Kretzner, L. and Eisenman, R. N. (1993). *Mad: a heterodimeric partner for Max that antagonizes Myc transcriptional activity.* Cell, **72**(2): 211-22.

Azzi, L., El-Alfy, M., Martel, C. and Labrie, F. (2005). *Gender differences in mouse skin morphology and specific effects of sex steroids and dehydroepiandrosterone.* J Invest Dermatol, **124**(1): 22-7.

Baldi, P. and Long, A. D. (2001). *A Bayesian framework for the analysis of microarray expression data: regularized t -test and statistical inferences of gene changes.* Bioinformatics, **17**(6): 509-19.

Bammler, T., Beyer, R. P., Bhattacharya, S., Boorman, G. A., Boyles, A., Bradford, B. U., Bumgarner, R. E., Bushel, P. R., Chaturvedi, K., Choi, D., Cunningham, M. L., Deng, S., Dressman, H. K., Fannin, R. D., Farin, F. M., Freedman, J. H., Fry, R. C., Harper, A., Humble, M. C., Hurban, P., Kavanagh, T. J., Kaufmann, W. K., Kerr, K. F., Jing, L., Lapidus, J. A., Lasarev, M. R., Li, J., Li, Y. J., Lobenhofer, E. K., Lu, X., Malek, R. L., Milton, S., Nagalla, S. R., O'Malley J, P., Palmer, V. S., Pattee, P., Paules, R. S., Perou, C. M., Phillips, K., Qin, L. X., Qiu, Y., Quigley, S. D., Rodland, M., Rusyn, I., Samson, L. D., Schwartz, D. A., Shi, Y., Shin, J. L., Sieber, S. O., Slifer, S., Speer, M. C., Spencer, P. S., Sproles, D. I., Swenberg, J. A., Suk, W. A., Sullivan, R. C., Tian, R., Tennant, R. W., Todd, S. A., Tucker, C. J., Van Houten, B., Weis, B. K., Xuan, S. and Zarbl, H. (2005). *Standardizing global gene expression analysis between laboratories and across platforms.* Nat Methods, **2**(5): 351-6.

Banin, S., Moyal, L., Shieh, S., Taya, Y., Anderson, C. W., Chessa, L., Smorodinsky, N. I., Prives, C., Reiss, Y., Shiloh, Y. and Ziv, Y. (1998). *Enhanced phosphorylation of p53 by ATM in response to DNA damage.* Science, **281**(5383): 1674-7.

Barnes, M., Freudenberg, J., Thompson, S., Aronow, B. and Pavlidis, P. (2005). *Experimental comparison and cross-validation of the Affymetrix and Illumina gene expression analysis platforms.* Nucleic Acids Res, **33**(18): 5914-23.

Barone, M. V. and Courtneidge, S. A. (1995). *Myc but not Fos rescue of PDGF signalling block caused by kinase-inactive Src.* Nature, **378**(6556): 509-12.

Bartkova, J., Horejsi, Z., Koed, K., Kramer, A., Tort, F., Zieger, K., Guldberg, P., Sehested, M., Nesland, J. M., Lukas, C., Orntoft, T., Lukas, J. and Bartek, J. (2005). *DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis.* Nature, **434**(7035): 864-70.

Beier, R., Burgin, A., Kiermaier, A., Fero, M., Karsunky, H., Saffrich, R., Moroy, T., Ansorge, W., Roberts, J. and Eilers, M. (2000). *Induction of cyclin E-cdk2 kinase activity, E2F-dependent transcription and cell growth by Myc are genetically separable events.* Embo J, **19**(21): 5813-23.

Beissbarth, T. and Speed, T. P. (2004). *GOstat: find statistically overrepresented Gene Ontologies within a group of genes.* Bioinformatics, **20**(9): 1464-5.

Bengtsson, M., Stahlberg, A., Rorsman, P. and Kubista, M. (2005). *Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels.* Genome Res, **15**(10): 1388-92.

Benjamini, Y. and Hochberg, Y. (1995). *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* Journal of the Royal Statistical Society, **Ser. B 57**: 289-300.

Benjamini, Y. and Yekutieli, D. (2001). *The control of the false discovery rate in multiple testing under dependency.* Ann. Stat., **29**(4): 1165–1188.

Bergmann, U., Funatomi, H., Yokoyama, M., Beger, H. G. and Korc, M. (1995). *Insulin-like growth factor I overexpression in human pancreatic cancer: evidence for autocrine and paracrine roles.* Cancer Res, **55**(10): 2007-11.

Bernasconi, N. L., Wormhoudt, T. A. and Laird-Offringa, I. A. (2000). *Post-transcriptional deregulation of myc genes in lung cancer cell lines.* Am J Respir Cell Mol Biol, **23**(4): 560-5.

Berns, K., Hijmans, E. M. and Bernards, R. (1997). *Repression of c-Myc responsive genes in cycling cells causes G1 arrest through reduction of cyclin E/CDK2 kinase activity.* Oncogene, **15**(11): 1347-56.

Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., Loda, M., Weber, G., Mark, E. J., Lander,

E. S., Wong, W., Johnson, B. E., Golub, T. R., Sugarbaker, D. J. and Meyerson, M. (2001). *Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses.* Proc Natl Acad Sci U S A, **98**(24): 13790-5.

Bird, A. P. (1986). *CpG-rich islands and the function of DNA methylation.* Nature, **321**(6067): 209-13.

Bissonnette, R. P., Echeverri, F., Mahboubi, A. and Green, D. R. (1992). *Apoptotic cell death induced by c-myc is inhibited by bcl-2.* Nature, **359**(6395): 552-4.

Blackwell, T. K., Huang, J., Ma, A., Kretzner, L., Alt, F. W., Eisenman, R. N. and Weintraub, H. (1993). *Binding of myc proteins to canonical and noncanonical DNA sequences.* Mol Cell Biol, **13**(9): 5216-24.

Blackwell, T. K., Kretzner, L., Blackwood, E. M., Eisenman, R. N. and Weintraub, H. (1990). *Sequence-specific DNA binding by the c-Myc protein.* Science, **250**(4984): 1149-51.

Blyth, K., Terry, A., O'Hara, M., Baxter, E. W., Campbell, M., Stewart, M., Donehower, L. A., Onions, D. E., Neil, J. C. and Cameron, E. R. (1995). *Synergy between a human c-myc transgene and p53 null genotype in murine thymic lymphomas: contrasting effects of homozygous and heterozygous p53 loss.* Oncogene, **10**(9): 1717-23.

Bolognia, J. L. (1995). *Aging skin.* Am J Med, **98**(1A): 99S-103S.

Bolstad, B. M., Irizarry, R. A., Astrand, M. and Speed, T. P. (2003). *A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.* Bioinformatics, **19**(2): 185-93.

Bonferroni, C. E. (1936). *Teoria statistica delle classi e calcolo delle probabilità.* Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, **8**: 3-62.

Bonneterre, J., Peyrat, J. P., Beuscart, R. and Demaille, A. (1990). *Prognostic significance of insulin-like growth factor 1 receptors in human breast cancer.* Cancer Res, **50**(21): 6931-5.

Borgono, C. A. and Diamandis, E. P. (2004). *The emerging roles of human tissue kallikreins in cancer.* Nat Rev Cancer, **4**(11): 876-90.

Bouchard, C., Dittrich, O., Kiermaier, A., Dohmann, K., Menkel, A., Eilers, M. and Luscher, B. (2001). *Regulation of cyclin D2 gene expression by the Myc/Max/Mad network: Myc-dependent TRRAP recruitment and histone acetylation at the cyclin D2 promoter.* Genes Dev, **15**(16): 2042-7.

Bouchard, C., Thieke, K., Maier, A., Saffrich, R., Hanley-Hyde, J., Ansorge, W., Reed, S., Sicinski, P., Bartek, J. and Eilers, M. (1999). *Direct induction of cyclin*

D2 by Myc contributes to cell cycle progression and sequestration of p27. Embo J, **18**(19): 5321-33.

Bouwens, L., Lu, W. G. and De Krijger, R. (1997). *Proliferation and differentiation in the human fetal endocrine pancreas.* Diabetologia, **40**(4): 398-404.

Bowman, T., Broome, M. A., Sinibaldi, D., Wharton, W., Pledger, W. J., Sedivy, J. M., Irby, R., Yeatman, T., Courtneidge, S. A. and Jove, R. (2001). *Stat3-mediated Myc expression is required for Src transformation and PDGF-induced mitogenesis.* Proc Natl Acad Sci U S A, **98**(13): 7319-24.

Brachat, A., Pierrat, B., Brungger, A. and Heim, J. (2000). *Comparative microarray analysis of gene expression during apoptosis-induction by growth factor deprivation or protein kinase C inhibition.* Oncogene, **19**(44): 5073-82.

Braude, I., Vukovic, B., Prasad, M., Marrano, P., Turley, S., Barber, D., Zielenska, M. and Squire, J. A. (2006). *Large scale copy number variation (CNV) at 14q12 is associated with the presence of genomic abnormalities in neoplasia.* BMC Genomics, **7**: 138.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C. A., Causton, H. C., Gaasterland, T., Glenisson, P., Holstege, F. C., Kim, I. F., Markowitz, V., Matese, J. C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J. and Vingron, M. (2001). *Minimum information about a microarray experiment (MIAME)-toward standards for microarray data.* Nat Genet, **29**(4): 365-71.

Brazma, A., Parkinson, H., Sarkans, U., Shojatalab, M., Vilo, J., Abeygunawardena, N., Holloway, E., Kapushesky, M., Kemmeren, P., Lara, G. G., Oezcimen, A., Rocca-Serra, P. and Sansone, S. A. (2003). *ArrayExpress--a public repository for microarray gene expression data at the EBI.* Nucleic Acids Res, **31**(1): 68-71.

Burtelow, M. A., Roos-Mattjus, P. M., Rauen, M., Babendure, J. R. and Karnitz, L. M. (2001). *Reconstitution and molecular analysis of the hRad9-hHus1-hRad1 (9-1-1) DNA damage responsive checkpoint complex.* J Biol Chem, **276**(28): 25903-9.

Canman, C. E., Lim, D. S., Cimprich, K. A., Taya, Y., Tamai, K., Sakaguchi, K., Appella, E., Kastan, M. B. and Siliciano, J. D. (1998). *Activation of the ATM kinase by ionizing radiation and phosphorylation of p53.* Science, **281**(5383): 1677-9.

Cano, D. A., Rulifson, I. C., Heiser, P. W., Swigart, L. B., Pelengaris, S., German, M., Evan, G. I., Bluestone, J. A. and Hebrok, M. (2007). *Regulated -cell regeneration in the adult mouse pancreas.* Diabetes.

Carswell, E. A., Old, L. J., Kassel, R. L., Green, S., Fiore, N. and Williamson, B. (1975). *An endotoxin-induced serum factor that causes necrosis of tumors.* Proc Natl Acad Sci U S A, **72**(9): 3666-70.

Castedo, M., Perfettini, J. L., Roumier, T. and Kroemer, G. (2002). *Cyclin-dependent kinase-1: linking apoptosis to cell cycle and mitotic catastrophe.* Cell Death Differ, **9**(12): 1287-93.

Chakrabarti, S. K. and Mirmira, R. G. (2003). *Transcription factors direct the development and function of pancreatic beta cells.* Trends Endocrinol Metab, **14**(2): 78-84.

Chappell, S. A., LeQuesne, J. P., Paulin, F. E., deSchoolmeester, M. L., Stoneley, M., Soutar, R. L., Ralston, S. H., Helfrich, M. H. and Willis, A. E. (2000). *A mutation in the c-myc-IRES leads to enhanced internal ribosome entry in multiple myeloma: a novel mechanism of oncogene de-regulation.* Oncogene, **19**(38): 4437-40.

Chaturvedi, P., Eng, W. K., Zhu, Y., Mattern, M. R., Mishra, R., Hurle, M. R., Zhang, X., Annan, R. S., Lu, Q., Faucette, L. F., Scott, G. F., Li, X., Carr, S. A., Johnson, R. K., Winkler, J. D. and Zhou, B. B. (1999). *Mammalian Chk2 is a downstream effector of the ATM-dependent DNA damage checkpoint pathway.* Oncogene, **18**(28): 4047-54.

Chehab, N. H., Malikzay, A., Appel, M. and Halazonetis, T. D. (2000). *Chk2/hCds1 functions as a DNA damage checkpoint in G(1) by stabilizing p53.* Genes Dev, **14**(3): 278-88.

Chen, G. and Goeddel, D. V. (2002). *TNF-R1 signaling: a beautiful pathway.* Science, **296**(5573): 1634-5.

Chen, J., Lin, J. and Levine, A. J. (1995). *Regulation of transcription functions of the p53 tumor suppressor by the mdm-2 oncogene.* Mol Med, **1**(2): 142-52.

Chen, J., Wu, X., Lin, J. and Levine, A. J. (1996). *mdm-2 inhibits the G1 arrest and apoptosis functions of the p53 tumor suppressor protein.* Mol Cell Biol, **16**(5): 2445-52.

Chen, Y., Dougherty, E. R. and Bittner, M. L. (1997). *Ratio-based decisions and quantitative analysis of cDNA microarray images.* J Biomed Opt, **2**(4): 364-367.

Chiariello, M., Marinissen, M. J. and Gutkind, J. S. (2001). *Regulation of c-myc expression by PDGF through Rho GTPases.* Nat Cell Biol, **3**(6): 580-6.

Chipuk, J. E., Kuwana, T., Bouchier-Hayes, L., Droin, N. M., Newmeyer, D. D., Schuler, M. and Green, D. R. (2004). *Direct activation of Bax by p53 mediates mitochondrial membrane permeabilization and apoptosis.* Science, **303**(5660): 1010-4.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. and Rutter, W. J. (1979). *Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease.* Biochemistry, **18**(24): 5294-9.

Chittenden, T., Flemington, C., Houghton, A. B., Ebb, R. G., Gallo, G. J., Elangovan, B., Chinnadurai, G. and Lutz, R. J. (1995). *A conserved domain in Bak, distinct from BH1 and BH2, mediates cell death and protein binding functions.* Embo J, **14**(22): 5589-96.

Choe, S. E., Boutros, M., Michelson, A. M., Church, G. M. and Halfon, M. S. (2005). *Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset.* Genome Biol, **6**(2): R16.

Chu, T. M., Weir, B. and Wolfinger, R. (2002). *A systematic statistical linear modeling approach to oligonucleotide array experiments.* Math Biosci, **176**(1): 35-51.

Chuaqui, R. F., Bonner, R. F., Best, C. J., Gillespie, J. W., Flaig, M. J., Hewitt, S. M., Phillips, J. L., Krizman, D. B., Tangrea, M. A., Ahram, M., Linehan, W. M., Knezevic, V. and Emmert-Buck, M. R. (2002). *Post-analysis follow-up and validation of microarray experiments.* Nat Genet, **32 Suppl**: 509-14.

Chung, C. H., Parker, J. S., Karaca, G., Wu, J., Funkhouser, W. K., Moore, D., Butterfoss, D., Xiang, D., Zanation, A., Yin, X., Shockley, W. W., Weissler, M. C., Dressler, L. G., Shores, C. G., Yarbrough, W. G. and Perou, C. M. (2004). *Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression.* Cancer Cell, **5**(5): 489-500.

Churchill, G. A. (2004). *Using ANOVA to analyze microarray data.* Biotechniques, **37**(2): 173-5, 177.

Clayton, E., Doupe, D. P., Klein, A. M., Winton, D. J., Simons, B. D. and Jones, P. H. (2007). *A single type of progenitor cell maintains normal epidermis.* Nature, **446**(7132): 185-9.

Cohen, S. and Elliott, G. A. (1963). *The stimulation of epidermal keratinization by a protein isolated from the submaxillary gland of the mouse.* J Invest Dermatol, **40**: 1-5.

Cole, M. D. and McMahon, S. B. (1999). *The Myc oncoprotein: a critical evaluation of transactivation and target gene regulation.* Oncogene, **18**(19): 2916-24.

Coller, H. A., Grandori, C., Tamayo, P., Colbert, T., Lander, E. S., Eisenman, R. N. and Golub, T. R. (2000). *Expression analysis with oligonucleotide microarrays reveals that MYC regulates genes involved in growth, cell cycle, signaling, and adhesion.* Proc Natl Acad Sci U S A, **97**(7): 3260-5.

Cope, L., Hartman, S. M., Gohlmann, H. W., Tiesman, J. P. and Irizarry, R. A. (2006). *Analysis of Affymetrix GeneChip data using amplified RNA.* Biotechniques, **40**(2): 165-6, 168, 170.

Copois, V., Bibeau, F., Bascoul-Mollevi, C., Salvetat, N., Chalbos, P., Bareil, C., Candeil, L., Fraslon, C., Conseiller, E., Granci, V., Maziere, P., Kramar, A., Ychou, M., Pau, B., Martineau, P., Molina, F. and Del Rio, M. (2007). *Impact of RNA degradation on gene expression profiles: assessment of different methods to reliably determine RNA quality.* J Biotechnol, **127**(4): 549-59.

Coppola, J. A. and Cole, M. D. (1986). *Constitutive c-myc oncogene expression blocks mouse erythroleukaemia cell differentiation but not commitment.* Nature, **320**(6064): 760-3.

Cortez, D., Wang, Y., Qin, J. and Elledge, S. J. (1999). *Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks.* Science, **286**(5442): 1162-6.

Cowling, V. H., Chandriani, S., Whitfield, M. L. and Cole, M. D. (2006). *A conserved Myc protein domain, MBIV, regulates DNA binding, apoptosis, transformation, and G2 arrest.* Mol Cell Biol, **26**(11): 4226-39.

Cowling, V. H. and Cole, M. D. (2007). *The Myc transactivation domain promotes global phosphorylation of the RNA polymerase II carboxy-terminal domain independently of direct DNA binding.* Mol Cell Biol, **27**(6): 2059-73.

Cui, X. and Churchill, G. A. (2003). *Statistical tests for differential expression in cDNA microarray experiments.* Genome Biol, **4**(4): 210.

Cui, X., Kerr, M. K. and Churchill, G. A. (2003). *Transformations for cDNA microarray data.* Stat Appl Genet Mol Biol, **2**: Article4.

Dabney, A. R. and Storey, J. D. (2006). *A reanalysis of a published Affymetrix GeneChip control dataset.* Genome Biol, **7**(3): 401.

Dalgaard, P. (2001). *The R-Tcl/Tk Interface.* Proceedings of the 2nd International Workshop on Distributed Statistical Computing, Vienna, Austria.

Dalla-Favera, R., Bregni, M., Erikson, J., Patterson, D., Gallo, R. C. and Croce, C. M. (1982). *Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells.* Proc Natl Acad Sci U S A, **79**(24): 7824-7.

Dang, C. V. (1999). *c-Myc target genes involved in cell growth, apoptosis, and metabolism.* Mol Cell Biol, **19**(1): 1-11.

Dansen, T. B., Whitfield, J., Rostker, F., Brown-Swigart, L. and Evan, G. I. (2006). *Specific requirement for Bax, not Bak, in Myc-induced apoptosis and tumor suppression in vivo.* J Biol Chem, **281**(16): 10890-5.

Datta, A., Nag, A., Pan, W., Hay, N., Gartel, A. L., Colamonici, O., Mori, Y. and Raychaudhuri, P. (2004). *Myc-ARF (alternate reading frame) interaction inhibits the functions of Myc.* J Biol Chem, **279**(35): 36698-707.

Davis, A. C., Wims, M., Spotts, G. D., Hann, S. R. and Bradley, A. (1993). *A null c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice.* Genes Dev, **7**(4): 671-82.

de Alboran, I. M., O'Hagan, R. C., Gartner, F., Malynn, B., Davidson, L., Rickert, R., Rajewsky, K., DePinho, R. A. and Alt, F. W. (2001). *Analysis of C-MYC function in normal cells via conditional gene-targeted mutation.* Immunity, **14**(1): 45-55.

de Boer, T. (2007). *GeneSpring: GeneSpring R Integration Functions R package version 2.12.0.*

De Leon, D. D., Farzad, C., Crutchlow, M. F., Brestelli, J., Tobias, J., Kaestner, K. H. and Stoffers, D. A. (2006). *Identification of transcriptional targets during pancreatic growth after partial pancreatectomy and exendin-4 treatment.* Physiol Genomics, **24**(2): 133-43.

Debbas, M. and White, E. (1993). *Wild-type p53 mediates apoptosis by E1A, which is inhibited by E1B.* Genes Dev, **7**(4): 546-54.

Dejean, L. M., Martinez-Caballero, S., Guo, L., Hughes, C., Teijido, O., Ducret, T., Ichas, F., Korsmeyer, S. J., Antonsson, B., Jonas, E. A. and Kinnally, K. W. (2005). *Oligomeric Bax is a component of the putative cytochrome c release channel MAC, mitochondrial apoptosis-induced channel.* Mol Biol Cell, **16**(5): 2424-32.

Dejean, L. M., Martinez-Caballero, S., Manon, S. and Kinnally, K. W. (2006). *Regulation of the mitochondrial apoptosis-induced channel, MAC, by BCL-2 family proteins.* Biochim Biophys Acta, **1762**(2): 191-201.

del Peso, L., Gonzalez-Garcia, M., Page, C., Herrera, R. and Nunez, G. (1997). *Interleukin-3-induced phosphorylation of BAD through the protein kinase Akt.* Science, **278**(5338): 687-9.

Dennis, G., Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C. and Lempicki, R. A. (2003). *DAVID: Database for Annotation, Visualization, and Integrated Discovery.* Genome Biol, **4**(5): P3.

DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A. and Trent, J. M. (1996). *Use of a cDNA microarray to analyse gene expression patterns in human cancer.* Nat Genet, **14**(4): 457-60.

Dobbin, K. K., Beer, D. G., Meyerson, M., Yeatman, T. J., Gerald, W. L., Jacobson, J. W., Conley, B., Buetow, K. H., Heiskanen, M., Simon, R. M., Minna, J. D., Girard, L., Misek, D. E., Taylor, J. M., Hanash, S., Naoki, K., Hayes, D. N., Ladd-Acosta, C., Enkemann, S. A., Viale, A. and Giordano, T. J. (2005).

Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays. Clin Cancer Res, **11**(2 Pt 1): 565-72.

Dominguez-Sola, D., Ying, C. Y., Grandori, C., Ruggiero, L., Chen, B., Li, M., Galloway, D. A., Gu, W., Gautier, J. and Dalla-Favera, R. (2007). *Non-transcriptional control of DNA replication by c-Myc.* Nature, **448**(7152): 445-51.

Du, C., Fang, M., Li, Y., Li, L. and Wang, X. (2000). *Smac, a mitochondrial protein that promotes cytochrome c-dependent caspase activation by eliminating IAP inhibition.* Cell, **102**(1): 33-42.

Duan, W. R., Garner, D. S., Williams, S. D., Funckes-Shippy, C. L., Spath, I. S. and Blomme, E. A. (2003). *Comparison of immunohistochemistry for activated caspase-3 and cleaved cytokeratin 18 with the TUNEL method for quantification of apoptosis in histological sections of PC-3 subcutaneous xenografts.* J Pathol, **199**(2): 221-8.

Dudley, M. W., Sheeler, C. Q., Wang, H. and Khan, S. (2000). *Activation of the human estrogen receptor by the antiestrogens ICI 182,780 and tamoxifen in yeast genetic systems: implications for their mechanism of action.* Proc Natl Acad Sci U S A, **97**(7): 3696-701.

Dudoit, S., Shaffer, J. P. and Boldrick, J. C. (2003). *Multiple Hypothesis Testing in Microarray Experiments.* Statistical Science, **18**(1): 71-103.

Duesberg, P. H. and Vogt, P. K. (1979). *Avian acute leukemia viruses MC29 and MH2 share specific RNA sequences: evidence for a second class of transforming genes.* Proc Natl Acad Sci U S A, **76**(4): 1633-7.

Durbin, B. P., Hardin, J. S., Hawkins, D. M. and Rocke, D. M. (2002). *A variance-stabilizing transformation for gene-expression microarray data.* Bioinformatics, **18 Suppl 1**: S105-10.

Eckert, R. L. and Green, H. (1986). *Structure and evolution of the human involucrin gene.* Cell, **46**(4): 583-9.

Edgar, R., Domrachev, M. and Lash, A. E. (2002). *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository*. Nucleic Acids Res, **30**(1): 207-10.

Efron, B. (1981). *Nonparametric Estimates of Standard Error: The Jackknife, the Bootstrap and Other Methods.* Biometrika, **68**(3): 589-599.

Efron, B. and Tibshirani, R. (2002). *Empirical bayes methods and false discovery rates for microarrays.* Genet Epidemiol, **23**(1): 70-86.

Eilers, M. (1999). *Control of cell proliferation by Myc family genes.* Mol Cells, **9**(1): 1-6.

Eischen, C. M., Roussel, M. F., Korsmeyer, S. J. and Cleveland, J. L. (2001). *Bax loss impairs Myc-induced apoptosis and circumvents the selection of p53 mutations during Myc-mediated lymphomagenesis.* Mol Cell Biol, **21**(22): 7653-62.

Eischen, C. M., Weber, J. D., Roussel, M. F., Sherr, C. J. and Cleveland, J. L. (1999). *Disruption of the ARF-Mdm2-p53 tumor suppressor pathway in Myc-induced lymphomagenesis.* Genes Dev, **13**(20): 2658-69.

Eisen, M. B., Spellman, P. T., Brown, P. O. and Botstein, D. (1998). *Cluster analysis and display of genome-wide expression patterns.* Proc Natl Acad Sci U S A, **95**(25): 14863-8.

Eisenman, R. N. (2001). *Deconstructing myc.* Genes Dev, **15**(16): 2023-30.

Elahi, D., McAloon-Dyke, M., Fukagawa, N. K., Meneilly, G. S., Sclater, A. L., Minaker, K. L., Habener, J. F. and Andersen, D. K. (1994). *The insulinotropic actions of glucose-dependent insulinotropic polypeptide (GIP) and glucagon-like peptide-1 (7-37) in normal and diabetic subjects.* Regul Pept, **51**(1): 63-74.

Elayat, A. A., el-Naggar, M. M. and Tahir, M. (1995). *An immunocytochemical and morphometric study of the rat pancreatic islets.* J Anat, **186 (Pt 3)**: 629-37.

Ellis, M. J., Jenkins, S., Hanfelt, J., Redington, M. E., Taylor, M., Leek, R., Siddle, K. and Harris, A. (1998). *Insulin-like growth factors in human breast cancer.* Breast Cancer Res Treat, **52**(1-3): 175-84.

Elson, A., Deng, C., Campos-Torres, J., Donehower, L. A. and Leder, P. (1995). *The MMTV/c-myc transgene and p53 null alleles collaborate to induce T-cell lymphomas, but not mammary carcinomas in transgenic mice.* Oncogene, **11**(1): 181-90.

Erster, S. and Moll, U. M. (2005). *Stress-induced p53 runs a transcription-independent death program.* Biochem Biophys Res Commun, **331**(3): 843-50.

Eskes, R., Desagher, S., Antonsson, B. and Martinou, J. C. (2000). *Bid induces the oligomerization and insertion of Bax into the outer mitochondrial membrane.* Mol Cell Biol, **20**(3): 929-35.

Evan, G. I., Wyllie, A. H., Gilbert, C. S., Littlewood, T. D., Land, H., Brooks, M., Waters, C. M., Penn, L. Z. and Hancock, D. C. (1992). *Induction of apoptosis in fibroblasts by c-myc protein.* Cell, **69**(1): 119-28.

Fanidi, A., Harrington, E. A. and Evan, G. I. (1992). *Cooperative interaction between c-myc and bcl-2 proto-oncogenes.* Nature, **359**(6395): 554-6.

Felsher, D. W. (2003). *Cancer revoked: oncogenes as therapeutic targets.* Nat Rev Cancer, **3**(5): 375-80.

Felsher, D. W. (2004). *Reversibility of oncogene-induced cancer.* Curr Opin Genet Dev, **14**(1): 37-42.

Felsher, D. W. and Bishop, J. M. (1999). *Transient excess of MYC activity can elicit genomic instability and tumorigenesis.* Proc Natl Acad Sci U S A, **96**(7): 3940-4.

Fernandez, P. C., Frank, S. R., Wang, L., Schroeder, M., Liu, S., Greene, J., Cocito, A. and Amati, B. (2003). *Genomic targets of the human c-Myc protein.* Genes Dev, **17**(9): 1115-29.

Ferner, H. (1951). *[The insular alpha and beta cells as the source of the antagonistic principles glucagon and insulin.].* Mikroskopie, **6**(1-2): 1-8.

Feuk, L., Marshall, C. R., Wintle, R. F. and Scherer, S. W. (2006). *Structural variants: changing the landscape of chromosomes and design of disease studies.* Hum Mol Genet, **15 Spec No 1**: R57-66.

Finch, A., Prescott, J., Shchors, K., Hunt, A., Soucek, L., Dansen, T. B., Swigart, L. B. and Evan, G. I. (2006). *Bcl-xL gain of function and p19 ARF loss of function cooperate oncogenically with Myc in vivo by distinct mechanisms.* Cancer Cell, **10**(2): 113-20.

Finegood, D. T., Scaglia, L. and Bonner-Weir, S. (1995). *Dynamics of beta-cell mass in the growing rat pancreas. Estimation with a simple mathematical model.* Diabetes, **44**(3): 249-56.

Fischer, E. A., Friedman, M. A. and Markey, M. K. (2007). *Empirical comparison of tests for differential expression on time-series microarray experiments.* Genomics, **89**(4): 460-70.

Fisher, R. A. (1926). *The Arrangement of Field Experiments.* J Min Agric Gr Br, **33**: 503-513.

Flores, I., Murphy, D. J., Swigart, L. B., Knies, U. and Evan, G. I. (2004). *Defining the temporal requirements for Myc in the progression and maintenance of skin neoplasia.* Oncogene, **23**(35): 5923-30.

Flory, M. R., Moser, M. J., Monnat, R. J., Jr. and Davis, T. N. (2000). *Identification of a human centrosomal calmodulin-binding protein that shares homology with pericentrin.* Proc Natl Acad Sci U S A, **97**(11): 5919-23.

Forster, V. T. (1948). *Zwischenmolekulare energiewanderung und fluoreszenz.* Ann. Physics (Leipzig), 2:55-75.

Fox, J. (2002). *An R and S-PLUS Companion to Applied Regression*. SAGE Publications Ltd, Thousand Oaks, CA.

Frank, S. R., Schroeder, M., Fernandez, P., Taubert, S. and Amati, B. (2001). *Binding of c-Myc to chromatin mediates mitogen-induced acetylation of histone H4 and gene activation.* Genes Dev, **15**(16): 2069-82.

Frantz, S. (2005). *An array of problems.* Nat Rev Drug Discov, **4**(5): 362-3.

Frazao, C., McVey, C. E., Amblar, M., Barbas, A., Vonrhein, C., Arraiano, C. M. and Carrondo, M. A. (2006). *Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex.* Nature, **443**(7107): 110-4.

Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S. B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R. C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M. M., Tsui, S. K., Xue, H., Wong, J. T., Galver, L. M., Fan, J. B., Gunderson, K., Murray, S. S., Oliphant, A. R., Chee, M. S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J. F., Phillips, M. S., Roumy, S., Sallee, C., Verner, A., Hudson, T. J., Kwok, P. Y., Cai, D., Koboldt, D. C., Miller, R. D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L. C., Mak, W., Song, Y. Q., Tam, P. K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C. P., Delgado, M., Dermitzakis, E. T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B. E., Whittaker, P., Bentley, D. R., Daly, M. J., de Bakker, P. I., Barrett, J., Chretien, Y. R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D. J., Sabeti, P., Saxena, R., Schaffner, S. F., Sham, P. C., Varilly, P., Altshuler, D., Stein, L. D., Krishnan, L., Smith, A. V., Tello-Ruiz, M. K., Thorisson, G. A., Chakravarti, A., Chen, P. E., Cutler, D. J., Kashuk, C. S., Lin, S., Abecasis, G. R., Guan, W., Li, Y., Munro, H. M., Qin, Z. S., Thomas, D. J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L. R., Clarke, G., Evans, D. M., Morris, A. P., Weir, B. S., Tsunoda, T., Mullikin, J. C., Sherry, S. T., Feolo, M., Skol, A., Zhang, H., Zeng, C., Zhao, H., Matsuda, I., Fukushima, Y., Macer, D. R., Suda, E., Rotimi, C. N., Adebamowo, C. A., Ajayi, I., Aniagwu, T., Marshall, P. A., Nkwodimmah, C., Royal, C. D., Leppert, M. F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I. F., Knoppers, B. M., Foster, M. W., Clayton, E. W., Watkin, J., Gibbs, R. A., Belmont, J. W., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G. M., Wheeler, D. A., Yakub, I., Gabriel, S. B., Onofrio, R. C., Richter, D. J., Ziaugra, L., Birren, B. W., Daly, M. J., Altshuler, D., Wilson, R. K., Fulton, L. L., Rogers, J., Burton, J., Carter, N. P., Clee, C. M., Griffiths, M., Jones, M. C., McLay, K., Plumb, R. W., Ross, M. T., Sims, S. K., Willey, D. L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J. C., L'Archeveque, P., Bellemare, G., Saeki, K., Wang, H., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A. L., Brooks, L. D., McEwen, J. E., Guyer, M. S., Wang, V. O., Peterson, J. L., Shi, M.,

Spiegel, J., Sung, L. M., Zacharia, L. F., Collins, F. S., Kennedy, K., Jamieson, R. and Stewart, J. (2007). *A second generation human haplotype map of over 3.1 million SNPs.* Nature, **449**(7164): 851-61.

Freytag, S. O. (1988). *Enforced expression of the c-myc oncogene inhibits cell differentiation by precluding entry into a distinct predifferentiation state in G0/G1.* Mol Cell Biol, **8**(4): 1614-24.

Frueh, F. W. (2006). *Impact of microarray data quality on genomic data submissions to the FDA.* Nat Biotechnol, **24**(9): 1105-7.

Frye, M., Gardner, C., Li, E. R., Arnold, I. and Watt, F. M. (2003). *Evidence that Myc activation depletes the epidermal stem cell compartment by modulating adhesive interactions with the local microenvironment.* Development, **130**(12): 2793-808.

Furuta, M., Yano, H., Zhou, A., Rouille, Y., Holst, J. J., Carroll, R., Ravazzola, M., Orci, L., Furuta, H. and Steiner, D. F. (1997). *Defective prohormone processing and altered pancreatic islet morphology in mice lacking active SPC2.* Proc Natl Acad Sci U S A, **94**(13): 6646-51.

Furuta, M., Zhou, A., Webb, G., Carroll, R., Ravazzola, M., Orci, L. and Steiner, D. F. (2001). *Severe defect in proglucagon processing in islet A-cells of prohormone convertase 2 null mice.* J Biol Chem, **276**(29): 27197-202.

Fuscoe, J. C., Tong, W. and Shi, L. (2007). *QA/QC issues to aid regulatory acceptance of microarray gene expression data.* Environ Mol Mutagen, **48**(5): 349-53.

Gaede, K., Ferner, H. and Kastrup, H. (1950). *[Second carbohydrate metabolism hormone of the pancreas (glucagon) and its origin in the alpha cells.].* Klin Wochenschr, **28**(23-24): 388-93.

Galaktionov, K. and Beach, D. (1991). *Specific activation of cdc25 tyrosine phosphatases by B-type cyclins: evidence for multiple roles of mitotic cyclins.* Cell, **67**(6): 1181-94.

Gandarillas, A. and Watt, F. M. (1997). *c-Myc promotes differentiation of human epidermal stem cells.* Genes Dev, **11**(21): 2869-82.

Gartel, A. L., Ye, X., Goufman, E., Shianov, P., Hay, N., Najmabadi, F. and Tyner, A. L. (2001). *Myc represses the p21(WAF1/CIP1) promoter and interacts with Sp1/Sp3.* Proc Natl Acad Sci U S A, **98**(8): 4510-5.

Gatei, M., Sloper, K., Sorensen, C., Syljuasen, R., Falck, J., Hobson, K., Savage, K., Lukas, J., Zhou, B. B., Bartek, J. and Khanna, K. K. (2003). *Ataxia-telangiectasia-mutated (ATM) and NBS1-dependent phosphorylation of Chk1 on Ser-317 in response to ionizing radiation.* J Biol Chem, **278**(17): 14806-11.

Gavrieli, Y., Sherman, Y. and Ben-Sasson, S. A. (1992). *Identification of programmed cell death in situ via specific labeling of nuclear DNA fragmentation.* J Cell Biol, **119**(3): 493-501.

Gebhard, C., Schwarzfischer, L., Pham, T. H., Schilling, E., Klug, M., Andreesen, R. and Rehli, M. (2006). *Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia.* Cancer Res, **66**(12): 6118-28.

Geiss, G. K., Bumgarner, R. E., An, M. C., Agy, M. B., van 't Wout, A. B., Hammersmark, E., Carter, V. S., Upchurch, D., Mullins, J. I. and Katze, M. G. (2000). *Large-scale monitoring of host cell gene expression during HIV-1 infection using cDNA microarrays.* Virology, **266**(1): 8-16.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y. and Zhang, J. (2004). *Bioconductor: open software development for computational biology and bioinformatics.* Genome Biol, **5**(10): R80.

Gibson, L., Holmgreen, S. P., Huang, D. C., Bernard, O., Copeland, N. G., Jenkins, N. A., Sutherland, G. R., Baker, E., Adams, J. M. and Cory, S. (1996). *bcl-w, a novel member of the bcl-2 family, promotes cell survival.* Oncogene, **13**(4): 665-75.

Giles, P. J. and Kipling, D. (2003). *Normality of oligonucleotide microarray data and implications for parametric statistical analyses.* Bioinformatics, **19**(17): 2254-62.

Gill, S. S., Aubin, R. A., Bura, C. A., Curran, I. H. and Matula, T. I. (1996). *Ensuring recovery of intact RNA from rat pancreas.* Mol Biotechnol, **6**(3): 359-62.

Ginsberg, S. D., Hemby, S. E., Lee, V. M., Eberwine, J. H. and Trojanowski, J. Q. (2000). *Expression profile of transcripts in Alzheimer's disease tangle-bearing CA1 neurons.* Ann Neurol, **48**(1): 77-87.

Giuriato, S., Rabin, K., Fan, A. C., Shachaf, C. M. and Felsher, D. W. (2004). *Conditional animal models: a strategy to define when oncogenes will be effective targets to treat cancer.* Semin Cancer Biol, **14**(1): 3-11.

Goga, A., Yang, D., Tward, A. D., Morgan, D. O. and Bishop, J. M. (2007). *Inhibition of CDK1 as a potential therapy for tumors over-expressing MYC.* Nat Med, **13**(7): 820-7.

Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D. and

Lander, E. S. (1999). *Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.* Science, **286**(5439): 531-7.

Gordon, A., Glazko, G., Qiu, X. and Yakovlev, A. (2007). *Control of the mean number of false discoveries, bonferroni and stability of multiple testing.* The Annals of Applied Statistics, **1**(1): 179-190.

Gorgoulis, V. G., Vassiliou, L. V., Karakaidos, P., Zacharatos, P., Kotsinas, A., Liloglou, T., Venere, M., Ditullio, R. A., Jr., Kastrinakis, N. G., Levy, B., Kletsas, D., Yoneta, A., Herlyn, M., Kittas, C. and Halazonetis, T. D. (2005). *Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions.* Nature, **434**(7035): 907-13.

Gorringe, K. L., Jacobs, S., Thompson, E. R., Sridhar, A., Qiu, W., Choong, D. Y. and Campbell, I. G. (2007). *High-resolution single nucleotide polymorphism array analysis of epithelial ovarian cancer reveals numerous microdeletions and amplifications.* Clin Cancer Res, **13**(16): 4731-9.

Grandori, C., Cowley, S. M., James, L. P. and Eisenman, R. N. (2000). *The Myc/Max/Mad network and the transcriptional control of cell behavior.* Annu Rev Cell Dev Biol, **16**: 653-99.

Grandori, C. and Eisenman, R. N. (1997). *Myc target genes.* Trends Biochem Sci, **22**(5): 177-81.

Gregory, M. A. and Hann, S. R. (2000). *c-Myc proteolysis by the ubiquitin-proteasome pathway: stabilization of c-Myc in Burkitt's lymphoma cells.* Mol Cell Biol, **20**(7): 2423-35.

Gross, A., Jockel, J., Wei, M. C. and Korsmeyer, S. J. (1998). *Enforced dimerization of BAX results in its translocation, mitochondrial dysfunction and apoptosis.* Embo J, **17**(14): 3878-85.

Gu, G., Wells, J. M., Dombkowski, D., Preffer, F., Aronow, B. and Melton, D. A. (2004). *Global expression analysis of gene regulatory pathways during endocrine pancreatic development.* Development, **131**(1): 165-79.

Guan, K. L., Jenkins, C. W., Li, Y., Nichols, M. A., Wu, X., O'Keefe, C. L., Matera, A. G. and Xiong, Y. (1994). *Growth suppression by p18, a p16INK4/MTS1- and p14INK4B/MTS2-related CDK6 inhibitor, correlates with wild-type pRb function.* Genes Dev, **8**(24): 2939-52.

Guhaniyogi, J. and Brewer, G. (2001). *Regulation of mRNA stability in mammalian cells.* Gene, **265**(1-2): 11-23.

Guo, Q. M., Malek, R. L., Kim, S., Chiao, C., He, M., Ruffy, M., Sanka, K., Lee, N. H., Dang, C. V. and Liu, E. T. (2000). *Identification of c-myc responsive genes using rat cDNA microarray.* Cancer Res, **60**(21): 5922-8.

Hann, S. R., Thompson, C. B. and Eisenman, R. N. (1985). *c-myc oncogene protein synthesis is independent of the cell cycle in human and avian cells.* Nature, **314**(6009): 366-9.

Harrington, E. A., Bennett, M. R., Fanidi, A. and Evan, G. I. (1994a). *c-Myc-induced apoptosis in fibroblasts is inhibited by specific cytokines.* Embo J, **13**(14): 3286-95.

Harrington, E. A., Fanidi, A. and Evan, G. I. (1994b). *Oncogenes and cell death.* Curr Opin Genet Dev, **4**(1): 120-9.

Hartemink, A. J., Gifford, D. K., Jaakola, T. S. and Young, R. A. (2001). *Maximum likelihood estimation of optimal scaling factors for expression array normalization.* In: SPIE BiOS.

Hartigan, J. A. and Wong, M. A. (1979). *A k-means clustering algorithm.* Applied Statistics, **28**: 100-108.

Hashimoto, N., Kido, Y., Uchida, T., Asahara, S., Shigeyama, Y., Matsuda, T., Takeda, A., Tsuchihashi, D., Nishizawa, A., Ogawa, W., Fujimoto, Y., Okamura, H., Arden, K. C., Herrera, P. L., Noda, T. and Kasuga, M. (2006). *Ablation of PDK1 in pancreatic beta cells induces diabetes as a result of loss of beta cell mass.* Nat Genet, **38**(5): 589-93.

Hashiro, M., Matsumoto, K., Okumura, H., Hashimoto, K. and Yoshikawa, K. (1991). *Growth inhibition of human keratinocytes by antisense c-myc oligomer is not coupled to induction of differentiation.* Biochem Biophys Res Commun, **174**(1): 287-92.

Haupt, Y., Maya, R., Kazaz, A. and Oren, M. (1997). *Mdm2 promotes the rapid degradation of p53.* Nature, **387**(6630): 296-9.

Heid, C. A., Stevens, J., Livak, K. J. and Williams, P. M. (1996). *Real time quantitative PCR.* Genome Res, **6**(10): 986-94.

Heinrichs, S. and Look, A. T. (2007). *Identification of structural aberrations in cancer by SNP array analysis.* Genome Biol, **8**(7): 219.

Henriksson, M. and Luscher, B. (1996). *Proteins of the Myc network: essential regulators of cell growth and differentiation.* Adv Cancer Res, **68**: 109-82.

Herbst, A., Hemann, M. T., Tworkowski, K. A., Salghetti, S. E., Lowe, S. W. and Tansey, W. P. (2005). *A conserved element in Myc that negatively regulates its proapoptotic activity.* EMBO Rep, **6**(2): 177-83.

Herbst, A., Salghetti, S. E., Kim, S. Y. and Tansey, W. P. (2004). *Multiple cell-type-specific elements regulate Myc protein stability.* Oncogene, **23**(21): 3863-71.

Hermeking, H., Rago, C., Schuhmacher, M., Li, Q., Barrett, J. F., Obaya, A. J., O'Connell, B. C., Mateyak, M. K., Tam, W., Kohlhuber, F., Dang, C. V., Sedivy,

J. M., Eick, D., Vogelstein, B. and Kinzler, K. W. (2000). *Identification of CDK4 as a target of c-MYC.* Proc Natl Acad Sci U S A, **97**(5): 2229-34.

Herold, S., Wanzel, M., Beuger, V., Frohme, C., Beul, D., Hillukkala, T., Syvaoja, J., Saluz, H. P., Haenel, F. and Eilers, M. (2002). *Negative regulation of the mammalian UV response by Myc through association with Miz-1.* Mol Cell, **10**(3): 509-21.

Herr, D. G. (1986). *On the History of ANOVA in Unbalanced Designs: The First 30 Years.* The American Statistician, **40**(4): 265-270.

Heyer, L. J., Kruglyak, S. and Yooseph, S. (1999). *Exploring expression data: identification and analysis of coexpressed genes.* Genome Res, **9**(11): 1106-15.

Hirao, A., Kong, Y. Y., Matsuoka, S., Wakeham, A., Ruland, J., Yoshida, H., Liu, D., Elledge, S. J. and Mak, T. W. (2000). *DNA damage-induced activation of p53 by the checkpoint kinase Chk2.* Science, **287**(5459): 1824-7.

Hirvonen, H., Makela, T. P., Sandberg, M., Kalimo, H., Vuorio, E. and Alitalo, K. (1990). *Expression of the myc proto-oncogenes in developing human fetal brain.* Oncogene, **5**(12): 1787-97.

Hoaglin, D. C., Mosteller, F. and Tukey, J. W. (2000). *Understanding robust and exploratory data analysis*. Wiley, New York.

Hoffman-Liebermann, B. and Liebermann, D. A. (1991). *Interleukin-6- and leukemia inhibitory factor-induced terminal differentiation of myeloid leukemia cells is blocked at an intermediate stage by constitutive c-myc.* Mol Cell Biol, **11**(5): 2375-81.

Holder, D., Raubertas, R. F., Pikounis, V. B., Svetnik, V. and Soper, K. (2001). *Statistical analysis of high density oligonucleotide arrays: a SAFER approach.* Proceedings of the ASA Annual Meeting, Atlanta, GA.

Hubbell, E., Liu, W. M. and Mei, R. (2002). *Robust estimators for expression analysis.* Bioinformatics, **18**(12): 1585-92.

Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A. and Vingron, M. (2002). *Variance stabilization applied to microarray data calibration and to the quantification of differential expression.* Bioinformatics, **18 Suppl 1**: S96-104.

Hueber, A. O., Zornig, M., Lyon, D., Suda, T., Nagata, S. and Evan, G. I. (1997). *Requirement for the CD95 receptor-ligand pathway in c-Myc-induced apoptosis.* Science, **278**(5341): 1305-9.

Hui, H. and Perfetti, R. (2002). *Pancreas duodenum homeobox-1 regulates pancreas development during embryogenesis and islet cell function in adulthood.* Eur J Endocrinol, **146**(2): 129-41.

Hunter, D. J., Kraft, P., Jacobs, K. B., Cox, D. G., Yeager, M., Hankinson, S. E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A., Wang, J., Yu, K., Chatterjee, N., Orr, N., Willett, W. C., Colditz, G. A., Ziegler, R. G., Berg, C. D., Buys, S. S., McCarty, C. A., Feigelson, H. S., Calle, E. E., Thun, M. J., Hayes, R. B., Tucker, M., Gerhard, D. S., Fraumeni, J. F., Jr., Hoover, R. N., Thomas, G. and Chanock, S. J. (2007). *A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer.* Nat Genet, **39**(7): 870-4.

Iacobuzio-Donahue, C. A., Ashfaq, R., Maitra, A., Adsay, N. V., Shen-Ong, G. L., Berg, K., Hollingsworth, M. A., Cameron, J. L., Yeo, C. J., Kern, S. E., Goggins, M. and Hruban, R. H. (2003). *Highly expressed genes in pancreatic ductal adenocarcinomas: a comprehensive characterization and comparison of the transcription profiles obtained from three major technologies.* Cancer Res, **63**(24): 8614-22.

Ihaka, R. and Gentleman, R. (1996). *R: A Language for Data Analysis and Graphics.* Journal of Computational and Graphical Statistics, **5**(3): 299-314.

International HapMap Consortium (2005). *A haplotype map of the human genome.* Nature, **437**(7063): 1299-320.

International Human Genome Sequencing Consortium (2004). *Finishing the euchromatic sequence of the human genome.* Nature, **431**(7011): 931-45.

Iritani, B. M. and Eisenman, R. N. (1999). *c-Myc enhances protein synthesis and cell size during B lymphocyte development.* Proc Natl Acad Sci U S A, **96**(23): 13180-5.

Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B. and Speed, T. P. (2003a). *Summaries of Affymetrix GeneChip probe level data.* Nucleic Acids Res, **31**(4): e15.

Irizarry, R. A., Cope, L. M. and Wu, Z. (2006). *Feature-level exploration of a published Affymetrix GeneChip control dataset.* Genome Biol, **7**(8): 404.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U. and Speed, T. P. (2003b). *Exploration, normalization, and summaries of high density oligonucleotide array probe level data.* Biostatistics, **4**(2): 249-64.

Irizarry, R. A., Warren, D., Spencer, F., Kim, I. F., Biswal, S., Frank, B. C., Gabrielson, E., Garcia, J. G., Geoghegan, J., Germino, G., Griffin, C., Hilmer, S. C., Hoffman, E., Jedlicka, A. E., Kawasaki, E., Martinez-Murillo, F., Morsberger, L., Lee, H., Petersen, D., Quackenbush, J., Scott, A., Wilson, M., Yang, Y., Ye, S. Q. and Yu, W. (2005). *Multiple-laboratory comparison of microarray platforms.* Nat Methods, **2**(5): 345-50.

Ishimi, Y. (1997). *A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex.* J Biol Chem, **272**(39): 24508-13.

Itzhaki, J. E., Gilbert, C. S. and Porter, A. C. (1997). *Construction by gene targeting in human cells of a "conditional' CDC2 mutant that rereplicates its DNA.* Nat Genet, **15**(3): 258-65.

Iversen, J. (1974). *Inhibition of pancreatic glucagon release by somatostatin: in vitro.* Scand J Clin Lab Invest, **33**(2): 125-9.

Iyer, V. R., Horak, C. E., Scafe, C. S., Botstein, D., Snyder, M. and Brown, P. O. (2001). *Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF.* Nature, **409**(6819): 533-8.

Jacobs, J. J., Scheijen, B., Voncken, J. W., Kieboom, K., Berns, A. and van Lohuizen, M. (1999). *Bmi-1 collaborates with c-Myc in tumorigenesis by inhibiting c-Myc-induced apoptosis via INK4a/ARF.* Genes Dev, **13**(20): 2678-90.

Jain, M., Arvanitis, C., Chu, K., Dewey, W., Leonhardt, E., Trinh, M., Sundberg, C. D., Bishop, J. M. and Felsher, D. W. (2002). *Sustained loss of a neoplastic phenotype by brief inactivation of MYC.* Science, **297**(5578): 102-4.

Janicke, R. U., Lee, F. H. and Porter, A. G. (1994). *Nuclear c-Myc plays an important role in the cytotoxicity of tumor necrosis factor alpha in tumor cells.* Mol Cell Biol, **14**(9): 5661-70.

Jansen-Durr, P., Meichle, A., Steiner, P., Pagano, M., Finke, K., Botz, J., Wessbecher, J., Draetta, G. and Eilers, M. (1993). *Differential modulation of cyclin gene expression by MYC.* Proc Natl Acad Sci U S A, **90**(8): 3685-9.

Jarvinen, A. K., Hautaniemi, S., Edgren, H., Auvinen, P., Saarela, J., Kallioniemi, O. P. and Monni, O. (2004). *Are data from different gene expression microarray platforms comparable?* Genomics, **83**(6): 1164-8.

Jensen, K. B. and Watt, F. M. (2006). *Single-cell expression profiling of human epidermal stem and transit-amplifying cells: Lrig1 is a regulator of stem cell quiescence.* Proc Natl Acad Sci U S A, **103**(32): 11958-63.

Jin, L. and Lloyd, R. V. (1997). *In situ hybridization: methods and applications.* J Clin Lab Anal, **11**(1): 2-9.

Jin, W., Riley, R. M., Wolfinger, R. D., White, K. P., Passador-Gurgel, G. and Gibson, G. (2001). *The contributions of sex, genotype and age to transcriptional variance in Drosophila melanogaster.* Nat Genet, **29**(4): 389-95.

Johnston, L. A., Prober, D. A., Edgar, B. A., Eisenman, R. N. and Gallant, P. (1999). *Drosophila myc regulates cellular growth during development.* Cell, **98**(6): 779-90.

Jones, L., Goldstein, D. R., Hughes, G., Strand, A. D., Collin, F., Dunnett, S. B., Kooperberg, C., Aragaki, A., Olson, J. M., Augood, S. J., Faull, R. L., Luthi-Carter, R., Moskvina, V. and Hodges, A. K. (2006). *Assessment of the*

relationship between pre-chip and post-chip quality measures for Affymetrix GeneChip expression data. BMC Bioinformatics, **7**: 211.

Jones, P. H., Simons, B. D. and Watt, F. M. (2007). *Sic Transit Gloria: Farewell to the epidermal transit amplifying cell?* Cell Stem Cell, **1**(4): 371-381.

Jones, R. M., Branda, J., Johnston, K. A., Polymenis, M., Gadd, M., Rustgi, A., Callanan, L. and Schmidt, E. V. (1996). *An essential E box in the promoter of the gene encoding the mRNA cap-binding protein (eukaryotic initiation factor 4E) is a target for activation by c-myc.* Mol Cell Biol, **16**(9): 4754-64.

Juin, P., Hueber, A. O., Littlewood, T. and Evan, G. (1999). *c-Myc-induced sensitization to apoptosis is mediated through cytochrome c release.* Genes Dev, **13**(11): 1367-81.

Juin, P., Hunt, A., Littlewood, T., Griffiths, B., Swigart, L. B., Korsmeyer, S. and Evan, G. (2002). *c-Myc functionally cooperates with Bax to induce apoptosis.* Mol Cell Biol, **22**(17): 6158-69.

Jung, S. H. and Jang, W. (2006). *How accurately can we control the FDR in analyzing microarray data?* Bioinformatics, **22**(14): 1730-6.

Kaldis, P. and Aleem, E. (2005). *Cell cycle sibling rivalry: Cdc2 vs. Cdk2.* Cell Cycle, **4**(11): 1491-4.

Kaneto, H., Sharma, A., Suzuma, K., Laybutt, D. R., Xu, G., Bonner-Weir, S. and Weir, G. C. (2002). *Induction of c-Myc expression suppresses insulin gene transcription by inhibiting NeuroD/BETA2-mediated transcriptional activation.* J Biol Chem, **277**(15): 12998-3006.

Karlsson, A., Giuriato, S., Tang, F., Fung-Weier, J., Levan, G. and Felsher, D. W. (2003). *Genomically complex lymphomas undergo sustained tumor regression upon MYC inactivation unless they acquire novel chromosomal translocations.* Blood, **101**(7): 2797-803.

Kauffmann-Zeh, A., Rodriguez-Viciana, P., Ulrich, E., Gilbert, C., Coffer, P., Downward, J. and Evan, G. (1997). *Suppression of c-Myc-induced apoptosis by Ras signalling through PI(3)K and PKB.* Nature, **385**(6616): 544-8.

Kennedy, G. C., Matsuzaki, H., Dong, S., Liu, W. M., Huang, J., Liu, G., Su, X., Cao, M., Chen, W., Zhang, J., Liu, W., Yang, G., Di, X., Ryder, T., He, Z., Surti, U., Phillips, M. S., Boyce-Jacino, M. T., Fodor, S. P. and Jones, K. W. (2003). *Large-scale genotyping of complex DNA.* Nat Biotechnol, **21**(10): 1233-7.

Kerr, J. F., Wyllie, A. H. and Currie, A. R. (1972). *Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics.* Br J Cancer, **26**(4): 239-57.

Kerr, M. K. (2003). *Design considerations for efficient and effective microarray studies.* Biometrics, **59**(4): 822-8.

Kerr, M. K. and Churchill, G. A. (2001a). *Experimental design for gene expression microarrays.* Biostatistics, **2**(2): 183-201.

Kerr, M. K. and Churchill, G. A. (2001b). *Statistical design and the analysis of gene expression microarray data.* Genet Res, **77**(2): 123-8.

Kerr, M. K., Martin, M. and Churchill, G. A. (2000). *Analysis of variance for gene expression microarray data.* J Comput Biol, **7**(6): 819-37.

Khaled, A. R., Kim, K., Hofmeister, R., Muegge, K. and Durum, S. K. (1999). *Withdrawal of IL-7 induces Bax translocation from cytosol to mitochondria through a rise in intracellular pH.* Proc Natl Acad Sci U S A, **96**(25): 14476-81.

Kim, S. T., Xu, B. and Kastan, M. B. (2002). *Involvement of the cohesin protein, Smc1, in Atm-dependent and independent responses to DNA damage.* Genes Dev, **16**(5): 560-70.

Kim, S. Y., Herbst, A., Tworkowski, K. A., Salghetti, S. E. and Tansey, W. P. (2003). *Skp2 regulates Myc protein stability and activity.* Mol Cell, **11**(5): 1177-88.

Kischkel, F. C., Hellbardt, S., Behrmann, I., Germer, M., Pawlita, M., Krammer, P. H. and Peter, M. E. (1995). *Cytotoxicity-dependent APO-1 (Fas/CD95)-associated proteins form a death-inducing signaling complex (DISC) with the receptor.* Embo J, **14**(22): 5579-88.

Kjems, L. L., Holst, J. J., Volund, A. and Madsbad, S. (2003). *The influence of GLP-1 on glucose-stimulated insulin secretion: effects on beta-cell sensitivity in type 2 and nondiabetic subjects.* Diabetes, **52**(2): 380-6.

Klefstrom, J., Vastrik, I., Saksela, E., Valle, J., Eilers, M. and Alitalo, K. (1994). *c-Myc induces cellular susceptibility to the cytotoxic action of TNF-alpha.* Embo J, **13**(22): 5442-50.

Kluck, R. M., Esposti, M. D., Perkins, G., Renken, C., Kuwana, T., Bossy-Wetzel, E., Goldberg, M., Allen, T., Barber, M. J., Green, D. R. and Newmeyer, D. D. (1999). *The pro-apoptotic proteins, Bid and Bax, cause a limited permeabilization of the mitochondrial outer membrane that is enhanced by cytosol.* J Cell Biol, **147**(4): 809-22.

Knies-Bamforth, U. E., Fox, S. B., Poulsom, R., Evan, G. I. and Harris, A. L. (2004). *c-Myc interacts with hypoxia to induce angiogenesis in vivo by a vascular endothelial growth factor-dependent mechanism.* Cancer Res, **64**(18): 6563-70.

Koch, H. B., Zhang, R., Verdoodt, B., Bailey, A., Zhang, C. D., Yates, J. R., 3rd, Menssen, A. and Hermeking, H. (2007). *Large-scale identification of c-MYC-associated proteins using a combined TAP/MudPIT approach.* Cell Cycle, **6**(2): 205-17.

Koerker, D. J., Ruch, W., Chideckel, E., Palmer, J., Goodner, C. J., Ensinck, J. and Gale, C. C. (1974). *Somatostatin: hypothalamic inhibitor of the endocrine pancreas.* Science, **184**(135): 482-4.

Kohonen, T. (1982). *Analysis of a simple self-organising process.* Biological Cybernetics, **43**: 59-69.

Kohonen, T. (1995). *Self organising maps*. Springer-Verlag, Berlin.

Kolligs, F. T., Kolligs, B., Hajra, K. M., Hu, G., Tani, M., Cho, K. R. and Fearon, E. R. (2000). *gamma-catenin is regulated by the APC tumor suppressor and its oncogenic activity is distinct from that of beta-catenin.* Genes Dev, **14**(11): 1319-31.

Komatsu, N., Takata, M., Otsuki, N., Toyama, T., Ohka, R., Takehara, K. and Saijoh, K. (2003). *Expression and localization of tissue kallikrein mRNAs in human epidermis and appendages.* J Invest Dermatol, **121**(3): 542-9.

Korsmeyer, S. J., Wei, M. C., Saito, M., Weiler, S., Oh, K. J. and Schlesinger, P. H. (2000). *Pro-apoptotic cascade activates BID, which oligomerizes BAK or BAX into pores that result in the release of cytochrome c.* Cell Death Differ, **7**(12): 1166-73.

Kretzner, L., Blackwood, E. M. and Eisenman, R. N. (1992). *Myc and Max proteins possess distinct transcriptional activities.* Nature, **359**(6394): 426-9.

Kurosu, H., Yamamoto, M., Clark, J. D., Pastor, J. V., Nandi, A., Gurnani, P., McGuinness, O. P., Chikuda, H., Yamaguchi, M., Kawaguchi, H., Shimomura, I., Takayama, Y., Herz, J., Kahn, C. R., Rosenblatt, K. P. and Kuro-o, M. (2005). *Suppression of aging in mice by the hormone Klotho.* Science, **309**(5742): 1829-33.

Ladiges, W. C., Knoblaugh, S. E., Morton, J. F., Korth, M. J., Sopher, B. L., Baskin, C. R., MacAuley, A., Goodman, A. G., LeBoeuf, R. C. and Katze, M. G. (2005). *Pancreatic beta-cell failure and diabetes in mice with a deletion mutation of the endoplasmic reticulum molecular chaperone gene P58IPK.* Diabetes, **54**(4): 1074-81.

Lajtha, L. G. (1979). *Stem cell concepts.* Differentiation, **14**(1-2): 23-34.

Land, H., Chen, A. C., Morgenstern, J. P., Parada, L. F. and Weinberg, R. A. (1986). *Behavior of myc and ras oncogenes in transformation of rat embryo fibroblasts.* Mol Cell Biol, **6**(6): 1917-25.

Land, H., Parada, L. F. and Weinberg, R. A. (1983). *Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes.* Nature, **304**(5927): 596-602.

Langerhans, P. (1869). *Beitrag zur mikroskopischen: Anatomie der Bauchspeicheldrüse*. Inaugural dissertation, G. Lange, Berlin.

Langsrud, O. (2003). *ANOVA for unbalanced data: Use Type II instead of Type III sums of squares.* Statistics and Computing, **13**(2): 163-167.

Larkin, J. E., Frank, B. C., Gavras, H., Sultana, R. and Quackenbush, J. (2005). *Independence and reproducibility across microarray platforms.* Nat Methods, **2**(5): 337-44.

Larsson, L. I., Sundler, F. and Hakanson, R. (1975). *Immunohistochemical localization of human pancreatic polypeptide (HPP) to a population of islet cells.* Cell Tissue Res, **156**(2): 167-71.

Lawlor, E. R., Soucek, L., Brown-Swigart, L., Shchors, K., Bialucha, C. U. and Evan, G. I. (2006). *Reversible kinetic analysis of Myc targets in vivo provides novel insights into Myc-mediated tumorigenesis.* Cancer Res, **66**(9): 4591-601.

Laybutt, D. R., Weir, G. C., Kaneto, H., Lebet, J., Palmiter, R. D., Sharma, A. and Bonner-Weir, S. (2002). *Overexpression of c-Myc in beta-cells of transgenic mice causes proliferation and apoptosis, downregulation of insulin gene expression, and diabetes.* Diabetes, **51**(6): 1793-804.

Le May, C., Chu, K., Hu, M., Ortega, C. S., Simpson, E. R., Korach, K. S., Tsai, M. J. and Mauvais-Jarvis, F. (2006). *Estrogens protect pancreatic beta-cells from apoptosis and prevent insulin-deficient diabetes mellitus in mice.* Proc Natl Acad Sci U S A, **103**(24): 9232-7.

Lee, M. L., Kuo, F. C., Whitmore, G. A. and Sklar, J. (2000). *Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations.* Proc Natl Acad Sci U S A, **97**(18): 9834-9.

Lee, R. C., Feinbaum, R. L. and Ambros, V. (1993). *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.* Cell, **75**(5): 843-54.

Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., Zeitlinger, J., Jennings, E. G., Murray, H. L., Gordon, D. B., Ren, B., Wyrick, J. J., Tagne, J. B., Volkert, T. L., Fraenkel, E., Gifford, D. K. and Young, R. A. (2002). *Transcriptional regulatory networks in Saccharomyces cerevisiae.* Science, **298**(5594): 799-804.

Lewis, B. C., Shim, H., Li, Q., Wu, C. S., Lee, L. A., Maity, A. and Dang, C. V. (1997). *Identification of putative c-Myc-responsive genes: characterization of rcl, a novel growth-related gene.* Mol Cell Biol, **17**(9): 4967-78.

Li, C. and Wong, W. H. (2001). *Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection.* Proc Natl Acad Sci U S A, **98**(1): 31-6.

Li, F., Ambrosini, G., Chu, E. Y., Plescia, J., Tognin, S., Marchisio, P. C. and Altieri, D. C. (1998a). *Control of apoptosis and mitotic spindle checkpoint by survivin.* Nature, **396**(6711): 580-4.

Li, H., Zhu, H., Xu, C. J. and Yuan, J. (1998b). *Cleavage of BID by caspase 8 mediates the mitochondrial damage in the Fas pathway of apoptosis.* Cell, **94**(4): 491-501.

Li, P., Nijhawan, D., Budihardjo, I., Srinivasula, S. M., Ahmad, M., Alnemri, E. S. and Wang, X. (1997). *Cytochrome c and dATP-dependent formation of Apaf-1/caspase-9 complex initiates an apoptotic protease cascade.* Cell, **91**(4): 479-89.

Li, W. H. and Grauer, D. (1991). *Fundamentals of molecular evolution*. Sinauer Associates, Sunderland, MA.

Li, Z., Van Calcar, S., Qu, C., Cavenee, W. K., Zhang, M. Q. and Ren, B. (2003). *A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells.* Proc Natl Acad Sci U S A, **100**(14): 8164-9.

Lieb, J. D., Liu, X., Botstein, D. and Brown, P. O. (2001). *Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association.* Nat Genet, **28**(4): 327-34.

Liebhaber, S. A. (1997). *mRNA stability and the control of gene expression.* Nucleic Acids Symp Ser(36): 29-32.

Lim, W. K., Wang, K., Lefebvre, C. and Califano, A. (2007). *Comparative analysis of microarray normalization procedures: effects on reverse engineering gene networks.* Bioinformatics, **23**(13): i282-8.

Lin, T. M. and Chance, R. E. (1974). *Candidate hormones of the gut. VI. Bovine pancreatic polypeptide (BPP) and avian pancreatic polypeptide (APP).* Gastroenterology, **67**(4): 737-8.

Lin, W. C., Lin, F. T. and Nevins, J. R. (2001). *Selective induction of E2F1 in response to DNA damage, mediated by ATM-dependent phosphorylation.* Genes Dev, **15**(14): 1833-44.

Lindsey-Boltz, L. A., Bermudez, V. P., Hurwitz, J. and Sancar, A. (2001). *Purification and characterization of human DNA damage checkpoint Rad complexes.* Proc Natl Acad Sci U S A, **98**(20): 11236-41.

Lipinski, C. and Hopkins, A. (2004). *Navigating chemical space for biology and medicine.* Nature, **432**(7019): 855-61.

Lips, E. H., de Graaf, E. J., Tollenaar, R. A., van Eijk, R., Oosting, J., Szuhai, K., Karsten, T., Nanya, Y., Ogawa, S., van de Velde, C. J., Eilers, P. H., van Wezel, T. and Morreau, H. (2007). *Single nucleotide polymorphism array analysis of chromosomal instability patterns discriminates rectal adenomas from carcinomas.* J Pathol, **212**(3): 269-77.

Lipshutz, R. J., Fodor, S. P., Gingeras, T. R. and Lockhart, D. J. (1999). *High density synthetic oligonucleotide arrays.* Nat Genet, **21**(1 Suppl): 20-4.

Littlewood, T. D., Hancock, D. C., Danielian, P. S., Parker, M. G. and Evan, G. I. (1995). *A modified oestrogen receptor ligand-binding domain as an improved switch for the regulation of heterologous proteins.* Nucleic Acids Res, **23**(10): 1686-90.

Liu, L. Y., Lin, C., Falcon, S., Zhang, J. and MacDonald, J. W. (2007). *mouse4302: Affymetrix Mouse Genome 430 2.0 Array Annotation Data. R package version 2.0.1.*

Liu, Q., Guntuku, S., Cui, X. S., Matsuoka, S., Cortez, D., Tamai, K., Luo, G., Carattini-Rivera, S., DeMayo, F., Bradley, A., Donehower, L. A. and Elledge, S. J. (2000). *Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint.* Genes Dev, **14**(12): 1448-59.

Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. and Brown, E. L. (1996). *Expression monitoring by hybridization to high-density oligonucleotide arrays.* Nat Biotechnol, **14**(13): 1675-80.

Lodygin, D., Epanchintsev, A., Menssen, A., Diebold, J. and Hermeking, H. (2005). *Functional epigenomics identifies genes frequently silenced in prostate cancer.* Cancer Res, **65**(10): 4218-27.

Lopez-Girona, A., Furnari, B., Mondesert, O. and Russell, P. (1999). *Nuclear localization of Cdc25 is regulated by DNA damage and a 14-3-3 protein.* Nature, **397**(6715): 172-5.

Lowe, S. W., Cepero, E. and Evan, G. (2004). *Intrinsic tumour suppression.* Nature, **432**(7015): 307-15.

Lugo, T. G. and Witte, O. N. (1989). *The BCR-ABL oncogene transforms Rat-1 cells and cooperates with v-myc.* Mol Cell Biol, **9**(3): 1263-70.

Luo, X., Budihardjo, I., Zou, H., Slaughter, C. and Wang, X. (1998). *Bid, a Bcl2 interacting protein, mediates cytochrome c release from mitochondria in response to activation of cell surface death receptors.* Cell, **94**(4): 481-90.

Lutz, W., Fulda, S., Jeremias, I., Debatin, K. M. and Schwab, M. (1998). *MycN and IFNgamma cooperate in apoptosis of human neuroblastoma cells.* Oncogene, **17**(3): 339-46.

Luzzi, V., Mahadevappa, M., Raja, R., Warrington, J. A. and Watson, M. A. (2003). *Accurate and reproducible gene expression profiles from laser capture microdissection, transcript amplification, and high density oligonucleotide microarray analysis.* J Mol Diagn, **5**(1): 9-14.

Mackenzie, I. C. (1970). *Relationship between mitosis and the ordered structure of the stratum corneum in mouse epidermis.* Nature, **226**(5246): 653-5.

Maclean, K. H., Kastan, M. B. and Cleveland, J. L. (2007). *Atm deficiency affects both apoptosis and proliferation to augment Myc-induced lymphomagenesis.* Mol Cancer Res, **5**(7): 705-11.

Maddison, K. and Clarke, A. R. (2005). *New approaches for modelling cancer mechanisms in the mouse.* J Pathol, **205**(2): 181-93.

Maere, S., Heymans, K. and Kuiper, M. (2005). *BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks.* Bioinformatics, **21**(16): 3448-9.

Mai, S. and Mushinski, J. F. (2003). *c-Myc-induced genomic instability.* J Environ Pathol Toxicol Oncol, **22**(3): 179-99.

Maki, C. G., Huibregtse, J. M. and Howley, P. M. (1996). *In vivo ubiquitination and proteasome-mediated degradation of p53(1).* Cancer Res, **56**(11): 2649-54.

Marinkovic, D., Marinkovic, T., Mahr, B., Hess, J. and Wirth, T. (2004). *Reversible lymphomagenesis in conditionally c-MYC expressing mice.* Int J Cancer, **110**(3): 336-42.

Marshall, E. (2004). *Getting the noise out of gene arrays.* Science, **306**(5696): 630-1.

Martinou, J. C. and Green, D. R. (2001). *Breaking the mitochondrial barrier.* Nat Rev Mol Cell Biol, **2**(1): 63-7.

Maruyama, K., Schiavi, S. C., Huse, W., Johnson, G. L. and Ruley, H. E. (1987). *myc and E1A oncogenes alter the responses of PC12 cells to nerve growth factor and block differentiation.* Oncogene, **1**(4): 361-7.

Matsui, H., Moriyama, A. and Takahashi, T. (2000). *Cloning and characterization of mouse klk27, a novel tissue kallikrein expressed in testicular Leydig cells and exhibiting chymotrypsin-like specificity.* Eur J Biochem, **267**(23): 6858-65.

Matsui, H. and Takahashi, T. (2001). *Mouse testicular Leydig cells express Klk21, a tissue kallikrein that cleaves fibronectin and IGF-binding protein-3.* Endocrinology, **142**(11): 4918-29.

Matsui, H., Takano, N., Moriyama, A. and Takahashi, T. (2005). *Single-chain tissue-type plasminogen activator is a substrate of mouse glandular kallikrein 24.* Zoolog Sci, **22**(10): 1105-11.

Matsui, S., Ito, M., Nishiyama, H., Uno, H., Kotani, H., Watanabe, J., Guilford, P., Reeve, A., Fukushima, M. and Ogawa, O. (2007). *Genomic characterization of multiple clinical phenotypes of cancer using multivariate linear regression models.* Bioinformatics, **23**(6): 732-8.

Matsuki, T., Horai, R., Sudo, K. and Iwakura, Y. (2003). *IL-1 plays an important role in lipid metabolism by regulating insulin levels under physiological conditions.* J Exp Med, **198**(6): 877-88.

McMahon, S. B., Van Buskirk, H. A., Dugan, K. A., Copeland, T. D. and Cole, M. D. (1998). *The novel ATM-related protein TRRAP is an essential cofactor for the c-Myc and E2F oncoproteins.* Cell, **94**(3): 363-74.

Mecham, B. H., Wetmore, D. Z., Szallasi, Z., Sadovsky, Y., Kohane, I. and Mariani, T. J. (2004). *Increased measurement accuracy for sequence-verified microarray probes.* Physiol Genomics, **18**(3): 308-15.

Medema, J. P., Scaffidi, C., Kischkel, F. C., Shevchenko, A., Mann, M., Krammer, P. H. and Peter, M. E. (1997). *FLICE is activated by association with the CD95 death-inducing signaling complex (DISC).* Embo J, **16**(10): 2794-804.

Menssen, A. and Hermeking, H. (2002). *Characterization of the c-MYC-regulated transcriptome by SAGE: identification and analysis of c-MYC target genes.* Proc Natl Acad Sci U S A, **99**(9): 6274-9.

Mian, I. S. (1997). *Comparative sequence analysis of ribonucleases HII, III, II PH and D.* Nucleic Acids Res, **25**(16): 3187-95.

Mihara, M., Erster, S., Zaika, A., Petrenko, O., Chittenden, T., Pancoska, P. and Moll, U. M. (2003). *p53 has a direct apoptogenic role at the mitochondria.* Mol Cell, **11**(3): 577-90.

Miklos, G. L. and Maleszka, R. (2004). *Microarray reality checks in the context of a complex disease.* Nat Biotechnol, **22**(5): 615-21.

Miller, L. D., Long, P. M., Wong, L., Mukherjee, S., McShane, L. M. and Liu, E. T. (2002). *Optimal gene expression analysis by microarrays.* Cancer Cell, **2**(5): 353-61.

Min, Y., Adachi, Y., Yamamoto, H., Ito, H., Itoh, F., Lee, C. T., Nadaf, S., Carbone, D. P. and Imai, K. (2003). *Genetic blockade of the insulin-like growth factor-I receptor: a promising strategy for human pancreatic cancer.* Cancer Res, **63**(19): 6432-41.

Mirnics, K., Middleton, F. A., Marquez, A., Lewis, D. A. and Levitt, P. (2000). *Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex.* Neuron, **28**(1): 53-67.

Mitchell, K. O., Ricci, M. S., Miyashita, T., Dicker, D. T., Jin, Z., Reed, J. C. and El-Deiry, W. S. (2000). *Bax is a transcriptional target and mediator of c-myc-induced apoptosis.* Cancer Res, **60**(22): 6318-25.

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. and Tollervey, D. (1997). *The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases.* Cell, **91**(4): 457-66.

Mitchell, P. and Tollervey, D. (2000). *mRNA stability in eukaryotes.* Curr Opin Genet Dev, **10**(2): 193-8.

Mladenov, Z., Nedyalkov, S., Ivanov, I. and Toshkov, I. (1980). *Neoplastic growths in chickens treated with cell and cell-free material from transplantable hepatoma induced by virus strain MC-29.* Neoplasma, **27**(2): 175-82.

Momand, J., Zambetti, G. P., Olson, D. C., George, D. and Levine, A. J. (1992). *The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation.* Cell, **69**(7): 1237-45.

Moody, S. E., Sarkisian, C. J., Hahn, K. T., Gunther, E. J., Pickup, S., Dugan, K. D., Innocent, N., Cardiff, R. D., Schnall, M. D. and Chodosh, L. A. (2002). *Conditional activation of Neu in the mammary epithelium of transgenic mice results in reversible pulmonary metastasis.* Cancer Cell, **2**(6): 451-61.

Moreno Rocha, J. C., Revol de Mendoza, A. and Barrera Saldana, H. A. (1999). *[Genetic transcription in eukaryotes: from transcriptional factors to disease].* Rev Invest Clin, **51**(6): 375-84.

Morgenbesser, S. D. and DePinho, R. A. (1994). *Use of transgenic mice to study myc family gene function in normal mammalian development and in cancer.* Semin Cancer Biol, **5**(1): 21-36.

Morrish, F., Giedt, C. and Hockenbery, D. (2003). *c-MYC apoptotic function is mediated by NRF-1 target genes.* Genes Dev, **17**(2): 240-55.

Mougneau, E., Lemieux, L., Rassoulzadegan, M. and Cuzin, F. (1984). *Biological activities of v-myc and rearranged c-myc oncogenes in rat fibroblast cells in culture.* Proc Natl Acad Sci U S A, **81**(18): 5758-62.

Muller, D., Bouchard, C., Rudolph, B., Steiner, P., Stuckmann, I., Saffrich, R., Ansorge, W., Huttner, W. and Eilers, M. (1997). *Cdk2-dependent phosphorylation of p27 facilitates its Myc-induced release from cyclin E/cdk2 complexes.* Oncogene, **15**(21): 2561-76.

Mullin, A. E., Soukatcheva, G., Verchere, C. B. and Chantler, J. K. (2006). *Application of in situ ductal perfusion to facilitate isolation of high-quality RNA from mouse pancreas.* Biotechniques, **40**(5): 617-21.

Nadon, R. and Shoemaker, J. (2002). *Statistical issues with microarrays: processing and analysis.* Trends Genet, **18**(5): 265-71.

Naef, F. and Magnasco, M. O. (2003). *Solving the riddle of the bright mismatches: labeling and effective binding in oligonucleotide arrays.* Phys Rev E Stat Nonlin Soft Matter Phys, **68**(1 Pt 1): 011906.

Natrajan, R., Williams, R. D., Hing, S. N., Mackay, A., Reis-Filho, J. S., Fenwick, K., Iravani, M., Valgeirsson, H., Grigoriadis, A., Langford, C. F., Dovey, O., Gregory, S. G., Weber, B. L., Ashworth, A., Grundy, P. E., Pritchard-Jones, K.

and Jones, C. (2006). *Array CGH profiling of favourable histology Wilms tumours reveals novel gains and losses associated with relapse.* J Pathol, **210**(1): 49-58.

Nau, M. M., Brooks, B. J., Battey, J., Sausville, E., Gazdar, A. F., Kirsch, I. R., McBride, O. W., Bertness, V., Hollis, G. F. and Minna, J. D. (1985). *L-myc, a new myc-related gene amplified and expressed in human small cell lung cancer.* Nature, **318**(6041): 69-73.

Nauck, M. A., Bartels, E., Orskov, C., Ebert, R. and Creutzfeldt, W. (1993). *Additive insulinotropic effects of exogenous synthetic human gastric inhibitory polypeptide and glucagon-like peptide-1-(7-36) amide infused at near-physiological insulinotropic hormone and glucose concentrations.* J Clin Endocrinol Metab, **76**(4): 912-7.

Navarange, M., Game, L., Fowler, D., Wadekar, V., Banks, H., Cooley, N., Rahman, F., Hinshelwood, J., Broderick, P. and Causton, H. C. (2005). *MiMiR: a comprehensive solution for storage, annotation and exchange of microarray data.* BMC Bioinformatics, **6**: 268.

Nesbit, C. E., Tersak, J. M. and Prochownik, E. V. (1999). *MYC oncogenes and human neoplastic disease.* Oncogene, **18**(19): 3004-16.

Ni, J., Fernandez, M. A., Danielsson, L., Chillakuru, R. A., Zhang, J., Grubb, A., Su, J., Gentz, R. and Abrahamson, M. (1998). *Cystatin F is a glycosylated human low molecular weight cysteine proteinase inhibitor.* J Biol Chem, **273**(38): 24797-804.

Nilsson, J. A. and Cleveland, J. L. (2003). *Myc pathways provoking cell suicide and cancer.* Oncogene, **22**(56): 9007-21.

O'Connell, B. C., Cheung, A. F., Simkevich, C. P., Tam, W., Ren, X., Mateyak, M. K. and Sedivy, J. M. (2003). *A large scale genetic analysis of c-Myc-regulated gene expression patterns.* J Biol Chem, **278**(14): 12563-73.

O'Connor, L., Strasser, A., O'Reilly, L. A., Hausmann, G., Adams, J. M., Cory, S. and Huang, D. C. (1998). *Bim: a novel member of the Bcl-2 family that promotes apoptosis.* Embo J, **17**(2): 384-95.

O'Hagan, R. C., Ohh, M., David, G., de Alboran, I. M., Alt, F. W., Kaelin, W. G., Jr. and DePinho, R. A. (2000). *Myc-enhanced expression of Cul1 promotes ubiquitin-dependent proteolysis and cell cycle progression.* Genes Dev, **14**(17): 2185-91.

Odorisio, T., Schietroma, C., Zaccaria, M. L., Cianfarani, F., Tiveron, C., Tatangelo, L., Failla, C. M. and Zambruno, G. (2002). *Mice overexpressing placenta growth factor exhibit increased vascularization and vessel permeability.* J Cell Sci, **115**(Pt 12): 2559-67.

Ohmura, E., Okada, M., Onoda, N., Kamiya, Y., Murakami, H., Tsushima, T. and Shizume, K. (1990). *Insulin-like growth factor I and transforming growth factor*

alpha as autocrine growth factors in human pancreatic cancer cell growth. Cancer Res, **50**(1): 103-7.

Olson, N. E. (2006). *The microarray data analysis process: from raw data to biological significance.* NeuroRx, **3**(3): 373-83.

Orian, A. and Eisenman, R. N. (2001). *TGF-beta flips the Myc switch.* Sci STKE, **2001**(88): PE1.

Oshima, H., Rochat, A., Kedzia, C., Kobayashi, K. and Barrandon, Y. (2001). *Morphogenesis and renewal of hair follicles from adult multipotent stem cells.* Cell, **104**(2): 233-45.

Overall, J. E. and Spiegel, D. K. (1969). *Concerning Least Squares Analysis of Experimental Data.* Psychological Bulletin, **72**(5): 311-322.

Palmer, M. and Prediger, E. (2004). *Assessing RNA Quality*. Ambion Technotes **11**(1): http://www.ambion.com/jp/techlib/tn/111/8.html.

Pandey, A. and Mann, M. (2000). *Proteomics to study genes and genomes.* Nature, **405**(6788): 837-46.

Park, B. J., Kang, J. W., Lee, S. W., Choi, S. J., Shin, Y. K., Ahn, Y. H., Choi, Y. H., Choi, D., Lee, K. S. and Kim, S. (2005). *The haploinsufficient tumor suppressor p18 upregulates p53 via interactions with ATM/ATR.* Cell, **120**(2): 209-21.

Park, T., Yi, S. G., Lee, S., Lee, S. Y., Yoo, D. H., Ahn, J. I. and Lee, Y. S. (2003). *Statistical tests for identifying differentially expressed genes in time-course microarray experiments.* Bioinformatics, **19**(6): 694-703.

Parmigiani, G., Garrett, E. S., Anbazhagan, R. and Gabrielson, E. (2002). *A statistical framework for expression-based molecular classification in cancer.* Journal of the Royal Statistical Society, Series B, **64**: 717-736.

Parmigiani, G., Garrett, E. S., Irizarry, R. A. and Zeger, S. L. (2003). *The Analysis of Gene Expression Data: An Overview of Methods and Software*. Springer, New York.

Patel, J. H., Loboda, A. P., Showe, M. K., Showe, L. C. and McMahon, S. B. (2004). *Analysis of genomic targets reveals complex functions of MYC.* Nat Rev Cancer, **4**(7): 562-8.

Patel, J. H. and McMahon, S. B. (2006). *Targeting of Miz-1 is essential for Myc-mediated apoptosis.* J Biol Chem, **281**(6): 3283-9.

Patterson, T. A., Lobenhofer, E. K., Fulmer-Smentek, S. B., Collins, P. J., Chu, T. M., Bao, W., Fang, H., Kawasaki, E. S., Hager, J., Tikhonova, I. R., Walker, S. J., Zhang, L., Hurban, P., de Longueville, F., Fuscoe, J. C., Tong, W., Shi, L. and Wolfinger, R. D. (2006). *Performance comparison of one-color and two-color*

platforms within the MicroArray Quality Control (MAQC) project. Nat Biotechnol, **24**(9): 1140-50.

Pauklin, S., Kristjuhan, A., Maimets, T. and Jaks, V. (2005). *ARF and ATM/ATR cooperate in p53-mediated apoptosis upon oncogenic stress.* Biochem Biophys Res Commun, **334**(2): 386-94.

Pavlidis, P. (2003). *Using ANOVA for gene selection from microarray studies of the nervous system.* Methods, **31**(4): 282-9.

Pavlov, E. V., Priault, M., Pietkiewicz, D., Cheng, E. H., Antonsson, B., Manon, S., Korsmeyer, S. J., Mannella, C. A. and Kinnally, K. W. (2001). *A novel, high conductance channel of mitochondria linked to apoptosis in mammalian cells and Bax expression in yeast.* J Cell Biol, **155**(5): 725-31.

Pelengaris, S., Abouna, S., Cheung, L., Ifandi, V., Zervou, S. and Khan, M. (2004). *Brief inactivation of c-Myc is not sufficient for sustained regression of c-Myc-induced tumours of pancreatic islets and skin epidermis.* BMC Biol, **2**: 26.

Pelengaris, S. and Khan, M. (2003). *The many faces of c-MYC.* Arch Biochem Biophys, **416**(2): 129-36.

Pelengaris, S., Khan, M. and Evan, G. (2002a). *c-MYC: more than just a matter of life and death.* Nat Rev Cancer, **2**(10): 764-76.

Pelengaris, S., Khan, M. and Evan, G. I. (2002b). *Suppression of Myc-induced apoptosis in beta cells exposes multiple oncogenic properties of Myc and triggers carcinogenic progression.* Cell, **109**(3): 321-34.

Pelengaris, S., Littlewood, T., Khan, M., Elia, G. and Evan, G. (1999). *Reversible activation of c-Myc in skin: induction of a complex neoplastic phenotype by a single oncogenic lesion.* Mol Cell, **3**(5): 565-77.

Pepper, S. D., Saunders, E. K., Edwards, L. E., Wilson, C. L. and Miller, C. J. (2007). *The utility of MAS5 expression summary and detection call algorithms.* BMC Bioinformatics, **8**: 273.

Perez-Roger, I., Kim, S. H., Griffiths, B., Sewing, A. and Land, H. (1999). *Cyclins D1 and D2 mediate myc-induced proliferation via sequestration of p27(Kip1) and p21(Cip1).* Embo J, **18**(19): 5310-20.

Pernu, H., Rasanen, O., Salo, T., Rinne, A., Herva, R. and Jarvinen, M. (1990). *Cystatin A and B in the development of human squamous epithelia.* Acta Histochem, **88**(1): 53-7.

Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A. L., Brown, P. O. and Botstein, D. (2000). *Molecular portraits of human breast tumours.* Nature, **406**(6797): 747-52.

Persson, H., Hennighausen, L., Taub, R., DeGrado, W. and Leder, P. (1984). *Antibodies to human c-myc oncogene product: evidence of an evolutionarily conserved protein induced during cell proliferation.* Science, **225**(4663): 687-93.

Persson, H. and Leder, P. (1984). *Nuclear localization and DNA binding properties of a protein expressed by human c-myc oncogene.* Science, **225**(4663): 718-21.

Petersen, D., Chandramouli, G. V., Geoghegan, J., Hilburn, J., Paarlberg, J., Kim, C. H., Munroe, D., Gangi, L., Han, J., Puri, R., Staudt, L., Weinstein, J., Barrett, J. C., Green, J. and Kawasaki, E. S. (2005). *Three microarray platforms: an analysis of their concordance in profiling gene expression.* BMC Genomics, **6**(1): 63.

Peukert, K., Staller, P., Schneider, A., Carmichael, G., Hanel, F. and Eilers, M. (1997). *An alternative pathway for gene regulation by Myc.* Embo J, **16**(18): 5672-86.

Philipp, A., Schneider, A., Vasrik, I., Finke, K., Xiong, Y., Beach, D., Alitalo, K. and Eilers, M. (1994). *Repression of cyclin D1: a novel function of MYC.* Mol Cell Biol, **14**(6): 4032-43.

Pierga, J. Y., Reis-Filho, J. S., Cleator, S. J., Dexter, T., Mackay, A., Simpson, P., Fenwick, K., Iravani, M., Salter, J., Hills, M., Jones, C., Ashworth, A., Smith, I. E., Powles, T. and Dowsett, M. (2007). *Microarray-based comparative genomic hybridisation of breast cancer patients receiving neoadjuvant chemotherapy.* Br J Cancer, **96**(2): 341-51.

Piotrowski, A., Benetkiewicz, M., Menzel, U., de Stahl, T. D., Mantripragada, K., Grigelionis, G., Buckley, P. G., Jankowski, M., Hoffman, J., Bala, D., Srutek, E., Laskowski, R., Zegarski, W. and Dumanski, J. P. (2006). *Microarray-based survey of CpG islands identifies concurrent hyper- and hypomethylation patterns in tissues derived from patients with breast cancer.* Genes Chromosomes Cancer, **45**(7): 656-67.

Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D. and Brown, P. O. (1999). *Genome-wide analysis of DNA copy-number changes using cDNA microarrays.* Nat Genet, **23**(1): 41-6.

Pollard, K. S., Ge, Y., Taylor, S. and Dudoit, S. (2007). *multtest: Resampling-based multiple hypothesis testing. R package version 1.18.0.*

Potten, C. S. (1974). *The epidermal proliferative unit: the possible role of the central basal cell.* Cell Tissue Kinet, **7**(1): 77-88.

Prochownik, E. V. and VanAntwerp, M. E. (1993). *Differential patterns of DNA binding by myc and max proteins.* Proc Natl Acad Sci U S A, **90**(3): 960-4.

Pusapati, R. V., Rounbehler, R. J., Hong, S., Powers, J. T., Yan, M., Kiguchi, K., McArthur, M. J., Wong, P. K. and Johnson, D. G. (2006). *ATM promotes*

apoptosis and suppresses tumorigenesis in response to Myc. Proc Natl Acad Sci U S A, **103**(5): 1446-51.

Qi, Y., Gregory, M. A., Li, Z., Brousal, J. P., West, K. and Hann, S. R. (2004). *p19ARF directly and differentially controls the functions of c-Myc independently of p53.* Nature, **431**(7009): 712-7.

Qin, X. Q., Livingston, D. M., Kaelin, W. G., Jr. and Adams, P. D. (1994). *Deregulated transcription factor E2F-1 expression leads to S-phase entry and p53-mediated apoptosis.* Proc Natl Acad Sci U S A, **91**(23): 10918-22.

Rajapakse, S., Yamano, N., Ogiwara, K., Hirata, K., Takahashi, S. and Takahashi, T. (2007). *Estrogen-dependent expression of the tissue kallikrein gene (Klk1) in the mouse uterus and its implications for endometrial tissue growth.* Mol Reprod Dev, **74**(8): 1053-63.

Rajeevan, M. S., Vernon, S. D., Taysavang, N. and Unger, E. R. (2001). *Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR.* J Mol Diagn, **3**(1): 26-31.

Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C. H., Angelo, M., Ladd, C., Reich, M., Latulippe, E., Mesirov, J. P., Poggio, T., Gerald, W., Loda, M., Lander, E. S. and Golub, T. R. (2001). *Multiclass cancer diagnosis using tumor gene expression signatures.* Proc Natl Acad Sci U S A, **98**(26): 15149-54.

Rasanen, O., Jarvinen, M. and Rinne, A. (1978). *Localization of the human SH-protease inhibitor in the epidermis. Immunofluorescent studies.* Acta Histochem, **63**(2): 193-6.

Rayner, T. F., Rocca-Serra, P., Spellman, P. T., Causton, H. C., Farne, A., Holloway, E., Irizarry, R. A., Liu, J., Maier, D. S., Miller, M., Petersen, K., Quackenbush, J., Sherlock, G., Stoeckert, C. J., Jr., White, J., Whetzel, P. L., Wymore, F., Parkinson, H., Sarkans, U., Ball, C. A. and Brazma, A. (2006). *A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB.* BMC Bioinformatics, **7**: 489.

Reddy, S., Bradley, J., Ginn, S., Pathipati, P. and Ross, J. M. (2003a). *Immunohistochemical study of caspase-3-expressing cells within the pancreas of non-obese diabetic mice during cyclophosphamide-accelerated diabetes.* Histochem Cell Biol, **119**(6): 451-61.

Reddy, S., Bradley, J. and Ross, J. M. (2003b). *Immunolocalization of caspase-3 in pancreatic islets of NOD mice during cyclophosphamide-accelerated diabetes.* Ann N Y Acad Sci, **1005**: 192-5.

Reed, J. C., Cuddy, M., Haldar, S., Croce, C., Nowell, P., Makover, D. and Bradley, K. (1990). *BCL2-mediated tumorigenicity of a human T-lymphoid cell*

line: synergy with MYC and inhibition by BCL2 antisense. Proc Natl Acad Sci U S A, **87**(10): 3660-4.

Rehault, S., Monget, P., Mazerbourg, S., Tremblay, R., Gutman, N., Gauthier, F. and Moreau, T. (2001). *Insulin-like growth factor binding proteins (IGFBPs) as potential physiological substrates for human kallikreins hK2 and hK3.* Eur J Biochem, **268**(10): 2960-8.

Reiner, A., Yekutieli, D. and Benjamini, Y. (2003). *Identifying differentially expressed genes using false discovery rate controlling procedures.* Bioinformatics, **19**(3): 368-75.

Remondini, D., O'Connell, B., Intrator, N., Sedivy, J. M., Neretti, N., Castellani, G. C. and Cooper, L. N. (2005). *Targeting c-Myc-activated genes with a correlation method: detection of global changes in large gene expression network dynamics.* Proc Natl Acad Sci U S A, **102**(19): 6902-6.

Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T. L., Wilson, C. J., Bell, S. P. and Young, R. A. (2000). *Genome-wide location and function of DNA binding proteins.* Science, **290**(5500): 2306-9.

Ricci, M. S., Jin, Z., Dews, M., Yu, D., Thomas-Tikhonenko, A., Dicker, D. T. and El-Deiry, W. S. (2004). *Direct repression of FLIP expression by c-myc is a major determinant of TRAIL sensitivity.* Mol Cell Biol, **24**(19): 8541-55.

Rivals, I., Personnaz, L., Taing, L. and Potier, M. C. (2007). *Enrichment or depletion of a GO category within a class of genes: which test?* Bioinformatics, **23**(4): 401-7.

Robson, S., Pelengaris, S. and Khan, M. (2006). *c-Myc and downstream targets in the pathogenesis and treatment of cancer.* Recent Patents Anticancer Drug Discov, **1**(3): 305-26.

Rogers, P. R., Song, J., Gramaglia, I., Killeen, N. and Croft, M. (2001). *OX40 promotes Bcl-xL and Bcl-2 expression and is essential for long-term survival of CD4 T cells.* Immunity, **15**(3): 445-55.

Rogojina, A. T., Orr, W. E., Song, B. K. and Geisert, E. E., Jr. (2003). *Comparing the use of Affymetrix to spotted oligonucleotide microarrays using two retinal pigment epithelium cell lines.* Mol Vis, **9**: 482-96.

Rosenwald, I. B., Rhoads, D. B., Callanan, L. D., Isselbacher, K. J. and Schmidt, E. V. (1993). *Increased expression of eukaryotic translation initiation factors eIF-4E and eIF-2 alpha in response to growth induction by c-myc.* Proc Natl Acad Sci U S A, **90**(13): 6175-8.

Roussel, M., Saule, S., Lagrou, C., Rommens, C., Beug, H., Graf, T. and Stehelin, D. (1979). *Three new types of viral oncogene of cellular origin specific for haematopoietic cell transformation.* Nature, **281**(5731): 452-5.

Russell, J. L., Powers, J. T., Rounbehler, R. J., Rogers, P. M., Conti, C. J. and Johnson, D. G. (2002). *ARF differentially modulates apoptosis induced by E2F1 and Myc.* Mol Cell Biol, **22**(5): 1360-8.

Ruvkun, G. (2001). *Molecular biology. Glimpses of a tiny RNA world.* Science, **294**(5543): 797-9.

Ryan, K. M. and Birnie, G. D. (1996). *Myc oncogenes: the enigmatic family.* Biochem J, **314 (Pt 3)**: 713-21.

Salghetti, S. E., Kim, S. Y. and Tansey, W. P. (1999). *Destruction of Myc by ubiquitin-mediated proteolysis: cancer-associated and transforming mutations stabilize Myc.* Embo J, **18**(3): 717-26.

Sander, M., Sussel, L., Conners, J., Scheel, D., Kalamaras, J., Dela Cruz, F., Schwitzgebel, V., Hayes-Jordan, A. and German, M. (2000). *Homeobox gene Nkx6.1 lies downstream of Nkx2.2 in the major pathway of beta-cell formation in the pancreas.* Development, **127**(24): 5533-40.

Sandmann, T., Jakobsen, J. S. and Furlong, E. E. (2006). *ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in Drosophila melanogaster embryos.* Nat Protoc, **1**(6): 2839-55.

Santamaria, D., Barriere, C., Cerqueira, A., Hunt, S., Tardy, C., Newton, K., Caceres, J. F., Dubus, P., Malumbres, M. and Barbacid, M. (2007). *Cdk1 is sufficient to drive the mammalian cell cycle.* Nature, **448**(7155): 811-5.

Schadt, E. E., Li, C., Su, C. and Wong, W. H. (2000). *Analyzing high-density oligonucleotide gene expression array data.* J Cell Biochem, **80**(2): 192-202.

Schena, M., Shalon, D., Davis, R. W. and Brown, P. O. (1995). *Quantitative monitoring of gene expression patterns with a complementary DNA microarray.* Science, **270**(5235): 467-70.

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O. and Davis, R. W. (1996). *Parallel human genome analysis: microarray-based expression monitoring of 1000 genes.* Proc Natl Acad Sci U S A, **93**(20): 10614-9.

Schlagbauer-Wadl, H., Griffioen, M., van Elsas, A., Schrier, P. I., Pustelnik, T., Eichler, H. G., Wolff, K., Pehamberger, H. and Jansen, B. (1999). *Influence of increased c-Myc expression on the growth characteristics of human melanoma.* J Invest Dermatol, **112**(3): 332-6.

Schlosser, I., Holzel, M., Murnseer, M., Burtscher, H., Weidle, U. H. and Eick, D. (2003). *A role for c-Myc in the regulation of ribosomal RNA processing.* Nucleic Acids Res, **31**(21): 6148-56.

Schmid, P., Schulz, W. A. and Hameister, H. (1989). *Dynamic expression pattern of the myc protooncogene in midgestation mouse embryos.* Science, **243**(4888): 226-9.

Schneider, P., Bodmer, J. L., Holler, N., Mattmann, C., Scuderi, P., Terskikh, A., Peitsch, M. C. and Tschopp, J. (1997). *Characterization of Fas (Apo-1, CD95)-Fas ligand interaction.* J Biol Chem, **272**(30): 18827-33.

Schoor, O., Weinschenk, T., Hennenlotter, J., Corvin, S., Stenzl, A., Rammensee, H. G. and Stevanovic, S. (2003). *Moderate degradation does not preclude microarray analysis of small amounts of RNA.* Biotechniques, **35**(6): 1192-6, 1198-201.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M. and Ragg, T. (2006). *The RIN: an RNA integrity number for assigning integrity values to RNA measurements.* BMC Mol Biol, **7**: 3.

Schuhmacher, M., Kohlhuber, F., Holzel, M., Kaiser, C., Burtscher, H., Jarsch, M., Bornkamm, G. W., Laux, G., Polack, A., Weidle, U. H. and Eick, D. (2001). *The transcriptional program of a human B cell line in response to Myc.* Nucleic Acids Res, **29**(2): 397-406.

Schuhmacher, M., Staege, M. S., Pajic, A., Polack, A., Weidle, U. H., Bornkamm, G. W., Eick, D. and Kohlhuber, F. (1999). *Control of cell growth by c-Myc in the absence of cell division.* Curr Biol, **9**(21): 1255-8.

Schulze-Kremer, S. (1997). *Adding semantics to genome databases: towards an ontology for molecular biology.* Proc Int Conf Intell Syst Mol Biol, **5**: 272-5.

Schulze-Kremer, S. (1998). *Ontologies for molecular biology.* Pac Symp Biocomput: 695-706.

Schumacher, A., Kapranov, P., Kaminsky, Z., Flanagan, J., Assadzadeh, A., Yau, P., Virtanen, C., Winegarden, N., Cheng, J., Gingeras, T. and Petronis, A. (2006). *Microarray-based DNA methylation profiling: technology and applications.* Nucleic Acids Res, **34**(2): 528-42.

Schurmann, A., Mooney, A. F., Sanders, L. C., Sells, M. A., Wang, H. G., Reed, J. C. and Bokoch, G. M. (2000). *p21-activated kinase 1 phosphorylates the death agonist bad and protects cells from apoptosis.* Mol Cell Biol, **20**(2): 453-61.

Schwab, M., Alitalo, K., Klempnauer, K. H., Varmus, H. E., Bishop, J. M., Gilbert, F., Brodeur, G., Goldstein, M. and Trent, J. (1983). *Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour.* Nature, **305**(5931): 245-8.

Seidah, N. G. and Chretien, M. (1997). *Eukaryotic protein processing: endoproteolysis of precursor proteins.* Curr Opin Biotechnol, **8**(5): 602-7.

Selvakumaran, M., Liebermann, D. and Hoffman-Liebermann, B. (1993). *Myeloblastic leukemia cells conditionally blocked by myc-estrogen receptor chimeric transgenes for terminal differentiation coupled to growth arrest and apoptosis.* Blood, **81**(9): 2257-62.

Serrano, M., Hannon, G. J. and Beach, D. (1993). *A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4.* Nature, **366**(6456): 704-7.

Sgroi, D. C., Teng, S., Robinson, G., LeVangie, R., Hudson, J. R., Jr. and Elkahloun, A. G. (1999). *In vivo gene expression profile analysis of human breast cancer progression.* Cancer Res, **59**(22): 5656-61.

Shachaf, C. M., Kopelman, A. M., Arvanitis, C., Karlsson, A., Beer, S., Mandl, S., Bachmann, M. H., Borowsky, A. D., Ruebner, B., Cardiff, R. D., Yang, Q., Bishop, J. M., Contag, C. H. and Felsher, D. W. (2004). *MYC inactivation uncovers pluripotent differentiation and tumour dormancy in hepatocellular cancer.* Nature, **431**(7012): 1112-7.

Shalon, D., Smith, S. J. and Brown, P. O. (1996). *A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization.* Genome Res, **6**(7): 639-45.

Shi, H., Guo, J., Duff, D. J., Rahmatpanah, F., Chitima-Matsiga, R., Al-Kuhlani, M., Taylor, K. H., Sjahputera, O., Andreski, M., Wooldridge, J. E. and Caldwell, C. W. (2007a). *Discovery of novel epigenetic markers in non-Hodgkin's lymphoma.* Carcinogenesis, **28**(1): 60-70.

Shi, L., Perkins, R. G., Fang, H. and Tong, W. (2007b). *Reproducible and reliable microarray results through quality control: good laboratory proficiency and appropriate data analysis practices are essential.* Curr Opin Biotechnol.

Shi, L., Reid, L. H., Jones, W. D., Shippy, R., Warrington, J. A., Baker, S. C., Collins, P. J., de Longueville, F., Kawasaki, E. S., Lee, K. Y., Luo, Y., Sun, Y. A., Willey, J. C., Setterquist, R. A., Fischer, G. M., Tong, W., Dragan, Y. P., Dix, D. J., Frueh, F. W., Goodsaid, F. M., Herman, D., Jensen, R. V., Johnson, C. D., Lobenhofer, E. K., Puri, R. K., Schrf, U., Thierry-Mieg, J., Wang, C., Wilson, M., Wolber, P. K., Zhang, L., Amur, S., Bao, W., Barbacioru, C. C., Lucas, A. B., Bertholet, V., Boysen, C., Bromley, B., Brown, D., Brunner, A., Canales, R., Cao, X. M., Cebula, T. A., Chen, J. J., Cheng, J., Chu, T. M., Chudin, E., Corson, J., Corton, J. C., Croner, L. J., Davies, C., Davison, T. S., Delenstarr, G., Deng, X., Dorris, D., Eklund, A. C., Fan, X. H., Fang, H., Fulmer-Smentek, S., Fuscoe, J. C., Gallagher, K., Ge, W., Guo, L., Guo, X., Hager, J., Haje, P. K., Han, J., Han, T., Harbottle, H. C., Harris, S. C., Hatchwell, E., Hauser, C. A., Hester, S., Hong, H., Hurban, P., Jackson, S. A., Ji, H., Knight, C. R., Kuo, W. P., LeClerc, J. E., Levy, S., Li, Q. Z., Liu, C., Liu, Y., Lombardi, M. J., Ma, Y., Magnuson, S. R., Maqsodi, B., McDaniel, T., Mei, N., Myklebost, O., Ning, B., Novoradovskaya, N., Orr, M. S., Osborn, T. W., Papallo, A., Patterson, T. A., Perkins, R. G., Peters, E. H., Peterson, R., Philips, K. L., Pine, P. S., Pusztai, L., Qian, F., Ren, H., Rosen, M., Rosenzweig, B. A., Samaha, R. R., Schena, M., Schroth, G. P., Shchegrova, S., Smith, D. D., Staedtler, F., Su, Z., Sun, H., Szallasi, Z., Tezak, Z., Thierry-Mieg, D., Thompson, K. L., Tikhonova, I., Turpaz, Y., Vallanat, B., Van, C., Walker, S. J., Wang, S. J., Wang, Y., Wolfinger, R., Wong, A., Wu, J., Xiao, C., Xie, Q., Xu, J., Yang, W., Zhang, L., Zhong, S., Zong, Y. and Slikker, W., Jr.

(2006). *The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements.* Nat Biotechnol, **24**(9): 1151-61.

Shi, L., Tong, W., Fang, H., Scherf, U., Han, J., Puri, R. K., Frueh, F. W., Goodsaid, F. M., Guo, L., Su, Z., Han, T., Fuscoe, J. C., Xu, Z. A., Patterson, T. A., Hong, H., Xie, Q., Perkins, R. G., Chen, J. J. and Casciano, D. A. (2005). *Cross-platform comparability of microarray technology: intra-platform consistency and appropriate data analysis procedures are essential.* BMC Bioinformatics, **6 Suppl 2**: S12.

Shi, Y., Glynn, J. M., Guilbert, L. J., Cotter, T. G., Bissonnette, R. P. and Green, D. R. (1992). *Role for c-myc in activation-induced apoptotic cell death in T cell hybridomas.* Science, **257**(5067): 212-4.

Shieh, S. Y., Ahn, J., Tamai, K., Taya, Y. and Prives, C. (2000). *The human homologs of checkpoint kinases Chk1 and Cds1 (Chk2) phosphorylate p53 at multiple DNA damage-inducible sites.* Genes Dev, **14**(3): 289-300.

Shiloh, Y. (2001). *ATM and ATR: networking cellular responses to DNA damage.* Curr Opin Genet Dev, **11**(1): 71-7.

Simon, R. M. and Dobbin, K. (2003). *Experimental design of DNA microarray experiments.* Biotechniques, **Suppl**: 16-21.

Singh, D., Febbo, P. G., Ross, K., Jackson, D. G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A. A., D'Amico, A. V., Richie, J. P., Lander, E. S., Loda, M., Kantoff, P. W., Golub, T. R. and Sellers, W. R. (2002). *Gene expression correlates of clinical prostate cancer behavior.* Cancer Cell, **1**(2): 203-9.

Smith, C. L., Nawaz, Z. and O'Malley, B. W. (1997). *Coactivator and corepressor regulation of the agonist/antagonist activity of the mixed antiestrogen, 4-hydroxytamoxifen.* Mol Endocrinol, **11**(6): 657-66.

Smith, M. S., Lechago, J., Wines, D. R., MacDonald, R. J. and Hammer, R. E. (1992). *Tissue-specific expression of kallikrein family transgenes in mice and rats.* DNA Cell Biol, **11**(5): 345-58.

Smyth, G. K. (2004). *Linear models and empirical bayes methods for assessing differential expression in microarray experiments.* Stat Appl Genet Mol Biol, **3**: Article3.

Smyth, G. K. (2005). Limma: linear models for microarray data. In: *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. R. Gentleman, V. Carey, S. Dudoit, R. Irizarry and W. Huber (Ed.), New York, Springer**:** 397-420.

Sommer, A., Hilfenhaus, S., Menkel, A., Kremmer, E., Seiser, C., Loidl, P. and Luscher, B. (1997). *Cell growth inhibition by the Mad/Max complex through recruitment of histone deacetylase activity.* Curr Biol, **7**(6): 357-65.

Soucie, E. L., Annis, M. G., Sedivy, J., Filmus, J., Leber, B., Andrews, D. W. and Penn, L. Z. (2001). *Myc potentiates apoptosis by stimulating Bax activity at the mitochondria.* Mol Cell Biol, **21**(14): 4725-36.

Speed, F. M., Hocking, R. R. and Hackney, O. P. (1978). *Methods of Analysis of Linear Models with Unbalanced Data.* Journal of the American Statistical Association, **73**(361): 105-112.

Spellman, P. T., Miller, M., Stewart, J., Troup, C., Sarkans, U., Chervitz, S., Bernhart, D., Sherlock, G., Ball, C., Lepage, M., Swiatek, M., Marks, W. L., Goncalves, J., Markel, S., Iordan, D., Shojatalab, M., Pizarro, A., White, J., Hubley, R., Deutsch, E., Senger, M., Aronow, B. J., Robinson, A., Bassett, D., Stoeckert, C. J., Jr. and Brazma, A. (2002). *Design and implementation of microarray gene expression markup language (MAGE-ML).* Genome Biol, **3**(9): RESEARCH0046.

Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D. and Futcher, B. (1998). *Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization.* Mol Biol Cell, **9**(12): 3273-97.

Srinivasula, S. M., Ahmad, M., Fernandes-Alnemri, T. and Alnemri, E. S. (1998). *Autoactivation of procaspase-9 by Apaf-1-mediated oligomerization.* Mol Cell, **1**(7): 949-57.

Staller, P., Peukert, K., Kiermaier, A., Seoane, J., Lukas, J., Karsunky, H., Moroy, T., Bartek, J., Massague, J., Hanel, F. and Eilers, M. (2001). *Repression of p15INK4b expression by Myc through association with Miz-1.* Nat Cell Biol, **3**(4): 392-9.

Stamey, T. A., Yang, N., Hay, A. R., McNeal, J. E., Freiha, F. S. and Redwine, E. (1987). *Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate.* N Engl J Med, **317**(15): 909-16.

Steiner, D. F. (1998). *The proprotein convertases.* Curr Opin Chem Biol, **2**(1): 31-9.

Steiner, P., Philipp, A., Lukas, J., Godden-Kent, D., Pagano, M., Mittnacht, S., Bartek, J. and Eilers, M. (1995). *Identification of a Myc-dependent step during the formation of active G1 cyclin-cdk complexes.* Embo J, **14**(19): 4814-26.

Stoeltzing, O., Liu, W., Reinmuth, N., Fan, F., Parikh, A. A., Bucana, C. D., Evans, D. B., Semenza, G. L. and Ellis, L. M. (2003). *Regulation of hypoxia-inducible factor-1alpha, vascular endothelial growth factor, and angiogenesis by an insulin-like growth factor-I receptor autocrine loop in human pancreatic cancer.* Am J Pathol, **163**(3): 1001-11.

Storey, J. D. (2002). *A direct approach to false discovery rates.* J. R. Statist. Soc. B, **64**(3): 479-498.

Storey, J. D. and Tibshirani, R. (2003). *Statistical significance for genomewide studies.* Proc Natl Acad Sci U S A, **100**(16): 9440-5.

Stott, F. J., Bates, S., James, M. C., McConnell, B. B., Starborg, M., Brookes, S., Palmero, I., Ryan, K., Hara, E., Vousden, K. H. and Peters, G. (1998). *The alternative product from the human CDKN2A locus, p14(ARF), participates in a regulatory feedback loop with p53 and MDM2.* Embo J, **17**(17): 5001-14.

Strasser, A., Harris, A. W., Bath, M. L. and Cory, S. (1990). *Novel primitive lymphoid tumours induced in transgenic mice by cooperation between myc and bcl-2.* Nature, **348**(6299): 331-3.

Strasser, A., Puthalakath, H., Bouillet, P., Huang, D. C., O'Connor, L., O'Reilly, L. A., Cullen, L., Cory, S. and Adams, J. M. (2000). *The role of bim, a proapoptotic BH3-only member of the Bcl-2 family in cell-death control.* Ann N Y Acad Sci, **917**: 541-8.

Subramanian, A., Sharma, A. K., Banerjee, D., Jiang, W. G. and Mokbel, K. (2007). *Evidence for a tumour suppressive function of IGF1-binding proteins in human breast cancer.* Anticancer Res, **27**(5B): 3513-8.

Suela, J., Alvarez, S. and Cigudosa, J. C. (2007). *DNA profiling by arrayCGH in acute myeloid leukemia and myelodysplastic syndromes.* Cytogenet Genome Res, **118**(2-4): 304-9.

Susin, S. A., Zamzami, N., Castedo, M., Daugas, E., Wang, H. G., Geley, S., Fassy, F., Reed, J. C. and Kroemer, G. (1997). *The central executioner of apoptosis: multiple connections between protease activation and mitochondria in Fas/APO-1/CD95- and ceramide-induced apoptosis.* J Exp Med, **186**(1): 25-37.

Susin, S. A., Zamzami, N., Castedo, M., Hirsch, T., Marchetti, P., Macho, A., Daugas, E., Geuskens, M. and Kroemer, G. (1996). *Bcl-2 inhibits the mitochondrial release of an apoptogenic protease.* J Exp Med, **184**(4): 1331-41.

Suzuki, H., Kurita, M., Mizumoto, K., Nishimoto, I., Ogata, E. and Matsuoka, M. (2003). *p19ARF-induced p53-independent apoptosis largely occurs through BAX.* Biochem Biophys Res Commun, **312**(4): 1273-7.

Svitkina, T. M., Verkhovsky, A. B. and Borisy, G. G. (1996). *Plectin sidearms mediate interaction of intermediate filaments with microtubules and other components of the cytoskeleton.* J Cell Biol, **135**(4): 991-1007.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E. S. and Golub, T. R. (1999). *Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation.* Proc Natl Acad Sci U S A, **96**(6): 2907-12.

Tan, P. K., Downey, T. J., Spitznagel, E. L., Jr., Xu, P., Fu, D., Dimitrov, D. S., Lempicki, R. A., Raaka, B. M. and Cam, M. C. (2003). *Evaluation of gene*

expression measurements from commercial microarray platforms. Nucleic Acids Res, **31**(19): 5676-84.

Tanaka, H., Matsumura, I., Ezoe, S., Satoh, Y., Sakamaki, T., Albanese, C., Machii, T., Pestell, R. G. and Kanakura, Y. (2002). *E2F1 and c-Myc potentiate apoptosis through inhibition of NF-kappaB activity that facilitates MnSOD-mediated ROS elimination.* Mol Cell, **9**(5): 1017-29.

Taylor, G., Lehrer, M. S., Jensen, P. J., Sun, T. T. and Lavker, R. M. (2000). *Involvement of follicular stem cells in forming not only the follicle but also the epidermis.* Cell, **102**(4): 451-61.

Th'ng, J. P., Wright, P. S., Hamaguchi, J., Lee, M. G., Norbury, C. J., Nurse, P. and Bradbury, E. M. (1990). *The FT210 cell line is a mouse G2 phase mutant with a temperature-sensitive CDC2 gene product.* Cell, **63**(2): 313-24.

Thompson, K. L., Pine, P. S., Rosenzweig, B. A., Turpaz, Y. and Retief, J. (2007). *Characterization of the effect of sample quality on high density oligonucleotide microarray data using progressively degraded rat liver RNA.* BMC Biotechnol, **7**: 57.

Tibbetts, R. S., Brumbaugh, K. M., Williams, J. M., Sarkaria, J. N., Cliby, W. A., Shieh, S. Y., Taya, Y., Prives, C. and Abraham, R. T. (1999). *A role for ATR in the DNA damage-induced phosphorylation of p53.* Genes Dev, **13**(2): 152-7.

Tibshirani, R. (1996). *Regression Shrinkage and Selection via the LASSO.* J. R. Statist. Soc., **58**(1): 267-288.

Tibshirani, R. (1997). *The lasso method for variable selection in the Cox model.* Stat Med, **16**(4): 385-95.

Tibshirani, R., Hastie, T., Narasimhan, B. and Chu, G. (2002). *Diagnosis of multiple cancer types by shrunken centroids of gene expression.* Proc Natl Acad Sci U S A, **99**(10): 6567-72.

Trevino, V., Falciani, F. and Barrera-Saldana, H. A. (2007). *DNA microarrays: a powerful genomic tool for biomedical and clinical research.* Mol Med, **13**(9-10): 527-41.

Trumpp, A., Refaeli, Y., Oskarsson, T., Gasser, S., Murphy, M., Martin, G. R. and Bishop, J. M. (2001). *c-Myc regulates mammalian body size by controlling cell number but not cell size.* Nature, **414**(6865): 768-73.

Tumor Analysis Best Practices Working Group (2004). *Expression profiling--best practices for data generation and interpretation in clinical trials.* Nat Rev Genet, **5**(3): 229-37.

Vafa, O., Wade, M., Kern, S., Beeche, M., Pandita, T. K., Hampton, G. M. and Wahl, G. M. (2002). *c-Myc can induce DNA damage, increase reactive oxygen*

species, and mitigate p53 function: a mechanism for oncogene-induced genetic instability. Mol Cell, **9**(5): 1031-44.

van't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R. and Friend, S. H. (2002). *Gene expression profiling predicts clinical outcome of breast cancer.* Nature, **415**(6871): 530-6.

van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R. and Friend, S. H. (2002). *Gene expression profiling predicts clinical outcome of breast cancer.* Nature, **415**(6871): 530-6.

Van Gelder, R. N., von Zastrow, M. E., Yool, A., Dement, W. C., Barchas, J. D. and Eberwine, J. H. (1990). *Amplified RNA synthesized from limited quantities of heterogeneous cDNA.* Proc Natl Acad Sci U S A, **87**(5): 1663-7.

Veal, E., Eisenstein, M., Tseng, Z. H. and Gill, G. (1998). *A cellular repressor of E1A-stimulated genes that inhibits activation by E2F.* Mol Cell Biol, **18**(9): 5032-41.

Velculescu, V. E., Zhang, L., Vogelstein, B. and Kinzler, K. W. (1995). *Serial analysis of gene expression.* Science, **270**(5235): 484-7.

Vennstrom, B., Sheiness, D., Zabielski, J. and Bishop, J. M. (1982). *Isolation and characterization of c-myc, a cellular homolog of the oncogene (v-myc) of avian myelocytomatosis virus strain 29.* J Virol, **42**(3): 773-9.

Vervoorts, J., Luscher-Firzlaff, J. M., Rottmann, S., Lilischkis, R., Walsemann, G., Dohmann, K., Austen, M. and Luscher, B. (2003). *Stimulation of c-MYC transcriptional activity and acetylation by recruitment of the cofactor CBP.* EMBO Rep, **4**(5): 484-90.

Vogelstein, B., Lane, D. and Levine, A. J. (2000). *Surfing the p53 network.* Nature, **408**(6810): 307-10.

von der Lehr, N., Johansson, S., Wu, S., Bahram, F., Castell, A., Cetinkaya, C., Hydbring, P., Weidung, I., Nakayama, K., Nakayama, K. I., Soderberg, O., Kerppola, T. K. and Larsson, L. G. (2003). *The F-box protein Skp2 participates in c-Myc proteosomal degradation and acts as a cofactor for c-Myc-regulated transcription.* Mol Cell, **11**(5): 1189-200.

Wagner, A. J., Kokontis, J. M. and Hay, N. (1994). *Myc-mediated apoptosis requires wild-type p53 in a manner independent of cell cycle arrest and the ability of p53 to induce p21waf1/cip1.* Genes Dev, **8**(23): 2817-30.

Wajant, H., Pfizenmaier, K. and Scheurich, P. (2003). *Tumor necrosis factor signaling.* Cell Death Differ, **10**(1): 45-65.

Walker, B. A. and Morgan, G. J. (2006). *Use of single nucleotide polymorphism-based mapping arrays to detect copy number changes and loss of heterozygosity in multiple myeloma.* Clin Lymphoma Myeloma, **7**(3): 186-91.

Wang, K., Yin, X. M., Chao, D. T., Milliman, C. L. and Korsmeyer, S. J. (1996). *BID: a novel BH3 domain-only death agonist.* Genes Dev, **10**(22): 2859-69.

Wang, T., Hopkins, D., Schmidt, C., Silva, S., Houghton, R., Takita, H., Repasky, E. and Reed, S. G. (2000). *Identification of genes differentially over-expressed in lung squamous cell carcinoma using combination of cDNA subtraction and microarray analysis.* Oncogene, **19**(12): 1519-28.

Watson, J. D., Oster, S. K., Shago, M., Khosravi, F. and Penn, L. Z. (2002). *Identifying genes regulated in a Myc-dependent manner.* J Biol Chem, **277**(40): 36921-30.

Watt, F. M. (1998). *Epidermal stem cells: markers, patterning and the control of stem cell fate.* Philos Trans R Soc Lond B Biol Sci, **353**(1370): 831-7.

Watt, F. M. and Hogan, B. L. (2000). *Out of Eden: stem cells and their niches.* Science, **287**(5457): 1427-30.

Weber, J. D., Jeffers, J. R., Rehg, J. E., Randle, D. H., Lozano, G., Roussel, M. F., Sherr, C. J. and Zambetti, G. P. (2000). *p53-independent functions of the p19(ARF) tumor suppressor.* Genes Dev, **14**(18): 2358-65.

Wei, M. C., Lindsten, T., Mootha, V. K., Weiler, S., Gross, A., Ashiya, M., Thompson, C. B. and Korsmeyer, S. J. (2000). *tBID, a membrane-targeted death ligand, oligomerizes BAK to release cytochrome c.* Genes Dev, **14**(16): 2060-71.

Wei, S. H., Balch, C., Paik, H. H., Kim, Y. S., Baldwin, R. L., Liyanarachchi, S., Li, L., Wang, Z., Wan, J. C., Davuluri, R. V., Karlan, B. Y., Gifford, G., Brown, R., Kim, S., Huang, T. H. and Nephew, K. P. (2006). *Prognostic DNA methylation biomarkers in ovarian cancer.* Clin Cancer Res, **12**(9): 2788-94.

White, R. J. (2005). *RNA polymerases I and III, growth control and cancer.* Nat Rev Mol Cell Biol, **6**(1): 69-78.

Wiche, G. (1998). *Role of plectin in cytoskeleton organization and dynamics.* J Cell Sci, **111 (Pt 17)**: 2477-86.

Wiley, S. R., Cassiano, L., Lofton, T., Davis-Smith, T., Winkles, J. A., Lindner, V., Liu, H., Daniel, T. O., Smith, C. A. and Fanslow, W. C. (2001). *A novel TNF receptor family member binds TWEAK and is implicated in angiogenesis.* Immunity, **15**(5): 837-46.

Wiley, S. R. and Winkles, J. A. (2003). *TWEAK, a member of the TNF superfamily, is a multifunctional cytokine that binds the TweakR/Fn14 receptor.* Cytokine Growth Factor Rev, **14**(3-4): 241-9.

Wolfinger, R. D., Gibson, G., Wolfinger, E. D., Bennett, L., Hamadeh, H., Bushel, P., Afshari, C. and Paules, R. S. (2001). *Assessing gene significance from cDNA microarray expression data via mixed models.* J Comput Biol, **8**(6): 625-37.

Wolter, K. G., Hsu, Y. T., Smith, C. L., Nechushtan, A., Xi, X. G. and Youle, R. J. (1997). *Movement of Bax from the cytosol to mitochondria during apoptosis.* J Cell Biol, **139**(5): 1281-92.

Wu, K. J., Grandori, C., Amacker, M., Simon-Vermot, N., Polack, A., Lingner, J. and Dalla-Favera, R. (1999). *Direct activation of TERT transcription by c-MYC.* Nat Genet, **21**(2): 220-4.

Wu, Z., Irizarry, R. A., Gentleman, R., Murillo, F. M. and Spencer, F. (2004). *A Model Based Background Adjustment for Oligonucleotide Expression Arrays.* Johns Hopkins University Department of Biostatistics Working Papers.

Wyllie, A. H., Kerr, J. F. and Currie, A. R. (1980). *Cell death: the significance of apoptosis.* Int Rev Cytol, **68**: 251-306.

Xie, S., Wu, H., Wang, Q., Cogswell, J. P., Husain, I., Conn, C., Stambrook, P., Jhanwar-Uniyal, M. and Dai, W. (2001). *Plk3 functionally links DNA damage to cell cycle arrest and apoptosis at least in part via the p53 pathway.* J Biol Chem, **276**(46): 43305-12.

Yamamoto, G., Nannya, Y., Kato, M., Sanada, M., Levine, R. L., Kawamata, N., Hangaishi, A., Kurokawa, M., Chiba, S., Gilliland, D. G., Koeffler, H. P. and Ogawa, S. (2007). *Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of affymetrix single-nucleotide-polymorphism genotyping microarrays.* Am J Hum Genet, **81**(1): 114-26.

Yang, D., Welm, A. and Bishop, J. M. (2004). *Cell division and cell survival in the absence of survivin.* Proc Natl Acad Sci U S A, **101**(42): 15100-5.

Yang, W., Shen, J., Wu, M., Arsura, M., FitzGerald, M., Suldan, Z., Kim, D. W., Hofmann, C. S., Pianetti, S., Romieu-Mourez, R., Freedman, L. P. and Sonenshein, G. E. (2001). *Repression of transcription of the p27(Kip1) cyclin-dependent kinase inhibitor gene by c-Myc.* Oncogene, **20**(14): 1688-702.

Yang, Y. H. and Speed, T. (2002). *Design issues for cDNA microarray experiments.* Nat Rev Genet, **3**(8): 579-88.

Yates, F. (1934). *The analysis of multiple classifications with unequal numbers in the different classes.* Journal of the American Statistical Association, **29**(185): 51-66.

Yauk, C. L., Berndt, M. L., Williams, A. and Douglas, G. R. (2004). *Comprehensive comparison of six microarray technologies.* Nucleic Acids Res, **32**(15): e124.

Yazdi, P. T., Wang, Y., Zhao, S., Patel, N., Lee, E. Y. and Qin, J. (2002). *SMC1 is a downstream effector in the ATM/NBS1 branch of the human S-phase checkpoint.* Genes Dev, **16**(5): 571-82.

Yin, X. M., Oltvai, Z. N. and Korsmeyer, S. J. (1994). *BH1 and BH2 domains of Bcl-2 are required for inhibition of apoptosis and heterodimerization with Bax.* Nature, **369**(6478): 321-3.

Yin, X. Y., Grove, L., Datta, N. S., Katula, K., Long, M. W. and Prochownik, E. V. (2001). *Inverse regulation of cyclin B1 by c-Myc and p53 and induction of tetraploidy by cyclin B1 overexpression.* Cancer Res, **61**(17): 6487-93.

Yousef, G. M., Kyriakopoulou, L. G., Scorilas, A., Fracchioli, S., Ghiringhello, B., Zarghooni, M., Chang, A., Diamandis, M., Giardina, G., Hartwick, W. J., Richiardi, G., Massobrio, M., Diamandis, E. P. and Katsaros, D. (2001). *Quantitative expression of the human kallikrein gene 9 (KLK9) in ovarian cancer: a new independent and favorable prognostic marker.* Cancer Res, **61**(21): 7811-8.

Yousef, G. M., Yacoub, G. M., Polymeris, M. E., Popalis, C., Soosaipillai, A. and Diamandis, E. P. (2004). *Kallikrein gene downregulation in breast cancer.* Br J Cancer, **90**(1): 167-72.

Yu, H., Diamandis, E. P., Levesque, M., Giai, M., Roagna, R., Ponzone, R., Sismondi, P., Monne, M. and Croce, C. M. (1996). *Prostate specific antigen in breast cancer, benign breast disease and normal breast tissue.* Breast Cancer Res Treat, **40**(2): 171-8.

Yuan, J., Shaham, S., Ledoux, S., Ellis, H. M. and Horvitz, H. R. (1993). *The C. elegans cell death gene ced-3 encodes a protein similar to mammalian interleukin-1 beta-converting enzyme.* Cell, **75**(4): 641-52.

Zeller, K. I., Jegga, A. G., Aronow, B. J., O'Donnell, K. A. and Dang, C. V. (2003). *An integrated database of genes responsive to the Myc oncogenic transcription factor: identification of direct genomic targets.* Genome Biol, **4**(10): R69.

Zeng, H., Datta, K., Neid, M., Li, J., Parangi, S. and Mukhopadhyay, D. (2003). *Requirement of different signaling pathways mediated by insulin-like growth factor-I receptor for proliferation, invasion, and VPF/VEGF expression in a pancreatic carcinoma cell line.* Biochem Biophys Res Commun, **302**(1): 46-55.

Zeng, Y., Forbes, K. C., Wu, Z., Moreno, S., Piwnica-Worms, H. and Enoch, T. (1998). *Replication checkpoint requires phosphorylation of the phosphatase Cdc25 by Cds1 or Chk1.* Nature, **395**(6701): 507-10.

Zha, H., Aime-Sempe, C., Sato, T. and Reed, J. C. (1996a). *Proapoptotic protein Bax heterodimerizes with Bcl-2 and homodimerizes with Bax via a novel domain (BH3) distinct from BH1 and BH2.* J Biol Chem, **271**(13): 7440-4.

Zha, J., Harada, H., Yang, E., Jockel, J. and Korsmeyer, S. J. (1996b). *Serine phosphorylation of death agonist BAD in response to survival factor results in binding to 14-3-3 not BCL-X(L).* Cell, **87**(4): 619-28.

Zhang, D., Bai, Y., Ge, Q., Qiao, Y., Wang, Y., Chen, Z. and Lu, Z. (2006). *Microarray-based molecular margin methylation pattern analysis in colorectal carcinoma.* Anal Biochem, **355**(1): 117-24.

Zhang, Y., Xiong, Y. and Yarbrough, W. G. (1998). *ARF promotes MDM2 degradation and stabilizes p53: ARF-INK4a locus deletion impairs both the Rb and p53 tumor suppression pathways.* Cell, **92**(6): 725-34.

Zhao, H. and Piwnica-Worms, H. (2001). *ATR-mediated checkpoint pathways regulate phosphorylation and activation of human Chk1.* Mol Cell Biol, **21**(13): 4129-39.

Zhao, R., Gish, K., Murphy, M., Yin, Y., Notterman, D., Hoffman, W. H., Tom, E., Mack, D. H. and Levine, A. J. (2000). *Analysis of p53-regulated gene expression patterns using oligonucleotide arrays.* Genes Dev, **14**(8): 981-93.

Zhou, Y. P., Pena, J. C., Roe, M. W., Mittal, A., Levisetti, M., Baldwin, A. C., Pugh, W., Ostrega, D., Ahmed, N., Bindokas, V. P., Philipson, L. H., Hanahan, D., Thompson, C. B. and Polonsky, K. S. (2000). *Overexpression of Bcl-x(L) in beta-cells prevents cell death but impairs mitochondrial signal for insulin secretion.* Am J Physiol Endocrinol Metab, **278**(2): E340-51.

Zhu, X., Orci, L., Carroll, R., Norrbom, C., Ravazzola, M. and Steiner, D. F. (2002). *Severe block in processing of proinsulin to insulin accompanied by elevation of des-64,65 proinsulin intermediates in islets of mice lacking prohormone convertase 1/3.* Proc Natl Acad Sci U S A, **99**(16): 10299-304.

Zindy, F., Eischen, C. M., Randle, D. H., Kamijo, T., Cleveland, J. L., Sherr, C. J. and Roussel, M. F. (1998). *Myc signaling via the ARF tumor suppressor regulates p53-dependent apoptosis and immortalization.* Genes Dev, **12**(15): 2424-33.

Zou, L., Cortez, D. and Elledge, S. J. (2002). *Regulation of ATR substrate selection by Rad17-dependent loading of Rad9 complexes onto chromatin.* Genes Dev, **16**(2): 198-208.

Appendix A: Gene abbreviations

Gene	Protein	Name
akt1-3	Akt1-3	v-akt murine thymoma viral oncogene homolog 1-3
apaf1	Apaf1	Apoptotic peptidase activating factor 1
atm	Atm	Ataxia telangiectasia mutated
atr	Atr	Ataxia telangiectasia and Rad3 related
bad	Bad	Bcl-associated death promoter
bak	Bak	BCL2-antagonist/killer
bax	Bax	Bcl2-associated X protein
bbc3	Puma/Bbc3	Bcl-2 binding component 3
bcl2	Bcl2	B-cell leukemia/lymphoma
bcl_W	Bcl_W	Bcl2-like 2
bcl_{XL}	Bcl_{XL}	Bcl2-like 1, extra large
bid	Bid	BH3 interacting domain death agonist
bim	Bim	Bcl2-like 11 (apoptosis facilitator)
birc5	Birc5/Survivin	Baculoviral IAP repeat-containing 5 (survivin)
brca1	Brca1	Breast cancer 1, early onset
cak	Cak	Cyclin activating kinase
casp1-14	Caspase 1-14	Caspase, apoptosis-related cysteine peptidase 1-14
cbp	Cbp	CREB binding protein
ccna	Cyclin A	Cyclin A
ccnb	Cyclin B	Cyclin B
ccnd	Cyclin D	Cyclin D
ccne	Cyclin E	Cyclin E
cdc25a	Cdc25a	Cell division cycle 25a

Gene	Protein	Name
cdc2a	Cdk1	Cyclin dependent kinase 1
cdk2-11	Cdk2-11	Cyclin dependent kinase 2-11
cdkn1a	$p21^{Cip1/Waf1}$	Cyclin dependent kinase inhibitor 1a
cdkn1b	$p27^{Kip1}$	Cyclin dependent kinase inhibitor 1b
cdkn2a	$p19^{Arf}$	Cyclin dependent kinase inhibitor 2a
cdkn2b	$p15^{Ink4b}$	Cyclin dependent kinase inhibitor 2b
cdkn2c	$p18^{Ink4c}$	Cyclin dependent kinase inhibitor 2c
cdt1	Cdt1	Chromatin licensing and DNA replication factor 1
cflar	Flip	Flice inhibitory protein
c-fos	c-Fos	FBJ osteosarcoma oncogene
chk1, 2	Chk1, 2	Checkpoint kinase 1, 2
cks	Cks	CDC28 protein kinase
c-myc	c-Myc	Myelocytomatosis, cellular
creg1	Creg1	Cellular repressor of E1A-stimulated genes 1
cst7	Cst7	cystatin F (leukocystatin)
csta	Csta	Cystatin A (stefin A)
cul1	Cul1	Cullin 1
cycs	Cyto c	Cytochrome c, somatic
e2f1-8	E2f1-8	E2F transcription factor 1-8
eif1-5	Eif1-5	Eukaryotic translation initiation factor 1-5
endog	Endog	Endonuclease g
fadd	Fadd	Fas-associated death domain
fas	Fas/CD95/Apo1	Fas (TNF receptor superfamily member 6)
flice	Flice/Caspase 8	Caspase 8
gadd45g	Gadd45g	Growth arrest and DNA-damage-inducible, gamma
gck	Gck	Glucokinase

Gene	Protein	Name
gcn5	Gcn5	General control of amino-acid synthesis 5
glp1	Glp1	Glucagon-like peptide 1
h2-aa	H2-Aa	Histocompatibility 2, class II antigen A, alpha
h2afx	H2afx	H2A histone family, member X
h2-d1	H2-D1	Histocompatibility 2, D region locus 1
h2-l	H2-L	Histocompatibility 2, D region
hb9	Hb9/Mnx1	Motor neuron and pancreas homeobox 1
hdac1, 2	Hdac1, 2	Histone deacetylase 1, 2
hus1	Hus1	HUS1 checkpoint homolog (S. pombe)
igf1	Igf1	Insulin-like growth factor 1
igf1bp1-7	Igfbp1-7	Insulin-like growth factor binding protein 1-7
igf1r	Igf1r	Insulin-like growth factor 1 receptor
il1r1	Il1r1	Interleukin-1 receptor 1
il6ra	Il6ra	Interleukin-6 receptor a
ins	Insulin	Insulin
inv	Inv	Involucrin
itga1-11	Itga1-11	Integrin, alpha 1-11
itgb1-8	Itgb1-8	Integrin, beta 1-8
ki67	Ki67	Antigen identified by monoclonal antibody Ki-67
kl	Klotho	Klotho
klk1-27	Klk1-27	Kallikrein 1-27
l-myc	l-Myc	Myelocytomatosis, lung
mad	Mad	MAX dimerization protein 1
max	Max	MYC associated factor X
mcm2-7	Mcm2-7	Minichromosome maintenance deficient 2-7
mdm2	Mdm2	Transformed 3T3 cell double minute 2, p53 binding protein

Gene	Protein	Name
miz1	Miz1	Zinc finger and BTB domain containing 17
nkx2.2	Nkx2.2	NK2 transcription factor related, locus 2
nkx6.1	Nkx6.1	NK6 transcription factor related, locus 1
n-myc	n-Myc	Myelocytomatosis, neuronal
pax4	Pax4	Paired box 4
pcnt	Pcnt	Pericentrin (kendrin)
pcsk1-3	Pcsk1-3	Proprotein convertase subtilisin/kexin type 1
pdx1	Pdx1/Ipf1	Pancreatic and duodenal homeobox 1
pgf	Pgf	Placental growth factor
pi3k	Pi3k	Phosphoinositide 3-kinase
plec1	Plec1	Plectin 1
pmaip1	Noxa/Pmaip1	Phorbol-12-myristate-13-acetate-induced protein 1
rad1-17	Rad1-17	Replication activation domain 1-17
rb1	Rb	Retinoblastoma 1
skp2	Skp2	S-phase kinase-associated protein 2
slc2a2	Glut2	Solute carrier family 2 (facilitated Glucose transporter), member 2
smac	Smac/Diablo	Second mitochondria-derived activator of caspases
smc1	Smc1	Structural maintenance of chromosomes 1
sp1, 3	Sp1, 3	Trans-acting transcription factor 1, 3
tgm2	Tgm2	Transglutaminase
tnf	Tnf	Tumour necrosis factor (TNF superfamily, member 2)
tnfr1, 2	Tnfr1, 2	Tumour necrosis factor receptor
tradd	Tradd	TNF receptor-associated death domain
trail	Trail	TNF-related apoptosis inducing ligand

Gene	Protein	Name
trp53	p53	Transformation related protein p53
trrap	Trrap	Ttransformation/transcription domain-associated protein
v-myc	v-Myc	Myelocytomatosis, viral

Appendix B: Gene lists

Supplementary Table 1: Cell cycle genes showing significant increase in expression within 8 hours of MycER^{TAM} activation in the pancreatic β-cells. p-value derived from analysis of the '4OHT' term in the *Envisage* model. Flags represent significant effects detected at specific time points ('*' = t-test p-value ≤ 0.05, '' = t-test p-value ≤ 0.01).**

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
akap8	BB037566	2.72 **	1.93 *	0.9	1.48 **	7.99E-06
akap8	BG069776	2.53 **	1.95 *	1.11	1.96 **	0.004869
anapc1	AV113524	2.73 **	0.43 **	4.82 **	2.34 **	0.001374
bop1	BM213936	2.75 **	2.86 **	2.96 **	3.1 **	7.74E-05
brms1l	AK003055	0.99	2.45 **	1.71 **	1.08	0.000107
bub1b	AU045529	1.54 *	3.86 **	21.18 **	2.47 **	0.000959
ccna2	X75483	0.91 **	5.41 *	14.48 **	3.11 **	0.010868
ccnb1-rs1	NM_007629	1.07	3.37 **	11.07 **	3.49 **	0.002143
ccnd1	NM_007631	2.02 **	2.25 *	0.84 *	1.93	0.001627
ccnd1	NM_007631	1.92 **	3.93 *	1.19 *	3.21	0.014837
ccnd1	NM_007631	3.59 **	4.3 *	1.68 *	3.41	0.003873
ccnd2	NM_009829	2.3 *	2.02	1.3	2.29 *	0.026191
ccnd2	AK007904	2.05 *	0.67	1.99	2.52 *	0.014806
ccne1	NM_007633	1.8 **	3.35 **	8.21 **	1.82 **	0.001283
ccne1	BB293079	1.82 **	7.54 **	1.86 **	2.13 **	2.73E-05
ccne2	AF091432	3.04 **	8.16 **	7.78 **	8.86 **	0.000871
cd40	BB220422	3.08	1.36	2.64 *	0.94	1.74E-05
cdc14b	AK013228	2.17 **	0.34 **	0.63	1.46	0.009572
cdc25a	C76119	2.73 **	2.77 *	2.16 **	3.81 **	0.00184
cdc25a	C76119	2.39 **	2.24 *	1.83 **	2.05 **	0.0064
cdc2a	NM_007659	1.06	4.54 **	14.85 **	3.45 **	0.004186
cdc37	AK013255	2.13 **	2.36 **	1.17 **	1.82 **	0.000284
cdc6	NM_011799	3.02 *	10.59 **	15.69 **	8.81 **	0.004407
cdc7	AB018574	5.79 **	4.32 **	4.27 **	5.19 **	0.000715
cdca5	NM_026410	0.74	15.39 **	25.14 **	10.75 **	0.000357
cdca5	NM_026410	0.83	2.93 **	18.54 **	3.03 **	0.003913
cdca7	AK011289	2.91 **	1.8 **	1.18 *	1.58 *	0.000223
cdca7l	BC006933	3.05 **	4.18 **	3.34 **	5.58 **	0.003352
cdkn1a	AK007630	1.56 **	2.12 **	1.62 **	1.49	0.005006
cdkn2a	NM_009877	2.54 **	2 **	1.49 **	2.7 **	0.000264
cdt1	AF477481	5 **	16.63 **	37.16 **	6.04 **	2.36E-06
cdt1	AF477481	3.62 **	5.24 **	13.12 **	8.31 **	0.002799

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
cenpm	NM_025639	1.06	3.52 **	10.38 **	1.76	0.015638
cep55	AK004655	1.21	2.63 **	10.02 **	3.18 **	0.001102
chek1	BB298208	1.15	3.84 *	2.76 **	2.05 **	0.004048
chek1	C85740	2.05	5.34 *	6.01 **	3.86 **	0.004993
chek2	NM_016681	1.22	2.17 **	3.81 **	2.29 **	0.019763
cks1b	NM_016904	1.5 *	3.21 **	4.22 **	2.77 **	3.47E-06
cks1b	NM_016904	1.3 *	3.55 **	7.9 **	4.26 **	0.000506
cks2	NM_025415	0.91	3.26 **	13.58 **	4.86 **	0.004375
cks2	NM_025415	0.96	2.86 **	12.67 **	2.33 **	0.011831
clspn	BG067086	1.37	3.64 **	10.13 **	1.53	0.035206
cnbp	BM237919	2.21 **	1.41	0.85	2.22 **	0.036888
cse1l	NM_023565	1.62 **	2.3 **	3.23 **	1.81 **	4.37E-05
cul5	BB702110	2.68	0.92 *	1.65	0.98	0.016909
cwf19l1	BB749215	1.7 *	2.28 **	0.63	0.96	0.006289
dbf4	NM_013726	3.49 **	5.48 **	10.84 **	4.82 **	3.29E-05
dhcr7	NM_007856	2.37 **	2.63 **	1.21 *	1.6 **	0.001596
dis3	BM232345	2.61 **	2.37 **	2.39 **	3.92 **	0.000507
dnajc2	BG067003	2.3 **	2.1 **	2.44 **	1.85 **	8.56E-06
dnajc2	BG067003	2.49 **	1.09 **	2.47 **	1.87 **	0.005595
dock4	BG068753	4.26 **	2.31 **	1.14	1.66 **	0.000871
dock5	AK004325	2.63 **	2.08 **	1.71 **	1.98 **	0.000205
e2f3	BQ176318	2.54 **	1.55	1.11	1.61 *	0.00021
e2f5	BC003220	2.31 *	0.73 *	1.3	2.21 *	0.029519
eef1e1	NM_025380	3.15 **	2.33 **	7.85 **	8.45 **	8.73E-05
frk	BB787292	2.23 **	1.18	0.94	0.59 **	0.015192
gmnn	NM_020567	1.57 *	6.01 **	9.71 **	5.67 **	5.62E-05
h2afx	NM_010436	1.16	2.42 **	5.22 **	1.4 **	0.001174
hells	NM_008234	2.59	14.95 **	20.22 **	10.82 **	0.00272
hells	AK021390	1.27	6.08 **	5.18 **	3.95 **	0.049171
hspa8	AK004608	3.32 **	3.85 **	1.06	1.15	0.000406
igf1	NM_010512	2.21	0.64	1.05	1.09 **	0.00325
jag2	AV264681	3.32 **	2.62	2.14	4.69 **	0.000236
jag2	AV264681	2.57 **	1.41	1.01	2.49 **	0.001409
mcm2	BB699415	1.66	3.08	4.09 **	2.93 **	0.00025
mcm3	C80350	4.05	3.93	13.89 **	9.87 **	0.002198
mcm3	BI658327	1.22	3.36	2.43 **	1.96 **	0.000242
mcm4	BC013094	1.69 **	3.01 **	4.47 **	2.7 **	0.024174
mcm5	AI324988	4.51 **	20.38 **	6.84 **	10.15 **	3.43E-05
mcm6	NM_008567	1.65 **	5.93 **	6.1 **	5.14 **	0.020149
mcm6	BB099487	2.38 **	10.85 **	12.34 **	5.08 **	0.00041
mcm7	NM_008568	1.64 **	2.64 **	4.61 **	2.47 **	0.001371
mcm7	BB464359	2.63 **	4.84 **	5.56 **	4.17 **	0.00015

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
mcm7	BB407228	2 **	4.95 **	3.71 **	3.06 **	0.000883
mki67	X82786	0.67	2.2 *	25.01 **	1.44 *	0.031045
mphosph9	BG067775	3.2 **	1.64	1.9	2.05 *	0.029834
ncaph	BB725358	0.5	4.14 **	33.38 **	6.24 **	0.003466
nup62	NM_053074	1.87	2.71	4.56 **	2.57 **	8.77E-06
nup62	AW240611	2.17	2.41	2.93 **	2.7 **	3.73E-05
odc1	S64539	1.79 **	2.34 **	1.69 **	1.88 **	2.12E-05
orc2l	BB830976	2.15	1.91 **	1.64 *	2.38 **	4.54E-06
orc2l	BB830976	0.9	3.94 **	1.89 *	2.54 **	0.005752
orc4l	BB620704	1.02	2.19 **	0.74 *	1.5 **	0.012472
orc6l	NM_019716	2.17 **	1.82 **	1.75 **	2.61 **	5.97E-05
p42pop	AF364868	1.51 *	3.02 **	1.45 *	1.92 **	0.001944
pa2g4	AA672939	1.95 **	4.83 **	1.32 **	1.59	4.15E-05
pa2g4	BM232515	2.24 **	1.35 **	1.7 **	3.14	0.004893
pa2g4	AI152156	2.55 **	1.85 **	0.98 **	2.27	0.000752
pa2g4	BM232515	2.1 **	5.55 **	2.78 **	2.66	0.000473
pa2g4	AA672939	2.14 **	2.25 **	0.46 **	1.13	0.000106
pcna	BC010343	0.96	3.32 **	20.99 **	4.87 **	0.005464
pola1	NM_008892	1.16	2.85 *	9.23 **	1.53	0.029454
ppargc1b	NM_133249	5.02 **	1.06	0.98	1.74 *	0.034594
ppp1r8	BC025479	2.48 **	0.68 **	2.25	2.11 *	0.048537
rad1	NM_011232	2.08 **	2.87 **	2.2 *	1.76	0.00377
rad50	NM_009012	1.68 *	3.48 **	1.6	4.18 **	0.021127
rcc2	AV122997	2.32 *	2.25 *	1.82 **	8.48 **	0.000133
rif1	AK018316	4.01	1.62	1.63	1.94	0.000908
rif1	AK018316	2.28	1.26	1.67	2.51	0.021757
ripk2	NM_138952	2.22 **	0.92	1.1	2.52 **	0.007429
ruvbl1	NM_019685	2.5 **	2.18 *	3.73 **	7.33 **	0.000116
sgol1	NM_028232	1.17	3.01	17.62 **	2.76	0.000468
skp2	AV259620	2.98 **	2.57 **	5.48 **	1.74 *	0.035912
skp2	BB055741	2.43 **	3.03 **	8.04 **	3.97 *	0.017959
skp2	BB784099	2.52 **	1.6 **	4.17 **	2.21 *	0.046724
skp2	AV259620	3.81 **	3.7 **	10.41 **	1.66 *	0.003072
smpd3	BF456582	1.3	2.07 **	1.69 **	3.34 **	0.044087
spc24	BF577722	0.81	2.77 **	14.54 **	1.45 *	8.74E-05
stmn1	BC010581	0.65 **	2.98	7.54 **	1.7	0.011405
tfdp1	BG075396	3.14 **	1.86 **	3.48 **	3.2 **	7.99E-06
tfrc	BB810450	4.29 **	0.78	0.73	1.85 *	0.009638
tnfsf5ip1	NM_134138	2.2 **	2	2.47 **	5.52 **	0.005229
tubb5	BG064086	1.11	3.14 **	2.31 **	2.16 **	0.003102
txnl4	AW552577	2.17 **	1.56 **	1.83 **	3.46 **	0.000117
uhrf1	BB702754	1.78	10.09 **	9.25 **	4.2 **	0.00156

Supplementary Table 2: Cell cycle genes showing significant decrease in expression within 8 hours of MycER^{TAM} activation in the pancreatic β-cells. p-value derived from analysis of the '4OHT' term in the *Envisage* model. Flags represent significant effects detected at specific time points ('*' = t-test p-value ≤ 0.05, '' = t-test p-value ≤ 0.01).**

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
abca7	NM_013850	0.62 **	0.36 **	0.79	0.49 **	0.001295
agt	AK018763	1.02	0.27 **	1.75 **	2.23 **	0.004482
ai844718	AI844718	0.99	0.31 **	1.35 *	0.42 **	0.010287
anapc1	AV113524	2.73 **	0.43 **	4.82 **	2.34 **	0.001374
appl1	BG073343	1.08	0.38 **	0.44 **	0.93	0.010668
arx	BB322201	0.58 **	0.63 **	0.4 **	0.57 **	0.000444
asah2	NM_018830	0.47 **	0.25 **	0.55 **	0.7 *	4.62E-07
bcl2l2	BB485989	1.29	0.41 **	0.53 **	0.53 **	0.012412
bin1	U60884	1.09	0.56 **	2.23 **	1.3 **	0.00724
bin1	BG293813	1.04	0.34 **	0.38 **	0.53 **	0.009659
btg1	L16846	0.47 **	0.42 **	1.09	0.64 **	0.002209
btg1	AW322026	0.62 **	0.4 **	0.99	0.74 **	0.002357
btg2	NM_007570	0.31 **	0.43 **	1.25 *	0.78	0.015696
btg2	NM_007570	0.45 **	0.34 **	1.71 *	1.11	0.025367
ccng2	U95826	0.38 **	0.42 **	1.81	0.66 *	0.000465
ccni	NM_017367	0.59 *	0.55 **	0.77	0.91 *	0.000169
ccnl2	AK008585	0.93	0.47 **	1.07 *	0.57	0.037282
cd9	NM_007657	0.44 **	0.5 *	2.43 *	2.02 *	0.035585
cdc14a	BB479310	0.55 *	0.5 **	0.8	0.87	0.038201
cdc14b	AK013228	2.17 **	0.34 **	0.63	1.46	0.009572
cdc23	BB492440	0.94	0.43 **	1.01	1.57 **	0.011765
cdkn1b	NM_009875	1.23	0.29 **	0.76 *	0.96 **	0.010159
cdkn2b	AF059567	0.47 **	0.32 **	0.76	0.45 **	0.001199
cdkn2c	BC027026	0.32 **	1.15	7.05 **	0.56 *	0.024661
cfl1	NM_007687	1.45	0.49 **	1.24	2.14 **	0.006256
cgrrf1	AV305616	0.67 **	0.42 *	1.71 **	0.6 **	0.022894
cgrrf1	AK004156	0.52 **	0.74 *	1.89 **	1.98 **	0.030366
col8a1	AV292255	0.44 **	0.47 **	0.98	4.08 **	0.009906
ctnna1	NM_009818	0.49 **	0.62 **	1.51 **	0.92	0.000473
d5ertd40e	C77487	0.59 **	0.59 **	1.25	0.45 **	0.013948
egfr	AV369812	0.76 *	0.47 **	0.97	1	0.000479
elk3	BC005686	0.63 *	0.46 **	0.62 *	0.72	0.035167
erbb3	BF140685	0.51 **	0.54 **	0.73 *	1.54	0.014831
esr1	NM_007956	0.62	0.25	1.8 **	1.49	0.010188
fgf1	AI649186	0.53 **	0.74 *	0.75 *	0.94	0.002184
flcn	BC025820	0.45 **	0.57 **	2	0.56 **	0.007496

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
frk	BB787292	0.63 **	0.35	0.45	0.97 **	0.003432
gadd45b	AK010420	0.73 *	0.33 **	2.11 **	0.48 **	0.001694
gadd45g	AK007410	0.58 *	0.4 **	2.08 **	0.63 **	1.30E-05
gng2	BB522409	0.49 **	1.08 *	0.9 **	0.78	0.009745
golt1a	BC024448	0.48 **	0.43 **	1.57 *	0.86	0.014803
gsk3b	BB831420	0.92	0.44 **	0.7 **	0.57 **	0.002762
htatip2	AF061972	0.84	0.41 **	0.78 **	0.72 **	0.016576
igf2	NM_010514	0.5 **	0.79	1.01	0.93	0.014592
il1r1	NM_008362	0.53 **	0.43 **	1.7 *	0.6 *	0.002892
il6ra	X53802	0.36 **	0.45 *	1	0.47 **	0.004703
itga6	BM935811	0.68 *	0.37 **	0.81	0.71	0.016939
kitl	BB815530	0.76 *	0.38 **	0.33 **	0.55 **	0.005882
lmo1	NM_057173	0.37 **	1.18	2.63 **	0.75 *	0.016203
lpp	BB089138	1.05	0.3 **	1.18 *	0.85 **	0.012568
lpp	BM236111	1.16	0.52 **	0.73 *	0.37 **	0.002906
mapk12	BC021640	0.75 *	0.52 **	0.69 *	1.17	0.021212
mapk7	NM_011841	1.02	0.41 **	1.11 *	1.02	0.010869
mitf	BB763517	0.46 *	0.43 **	0.98	0.93	0.01716
mta3	NM_054082	0.68 *	0.31 **	1.29 *	1.01	0.000871
mtag2	NM_016664	0.67 **	0.42 **	1.31	0.66 **	0.001173
ncaph	BB725358	0.5	4.14 **	33.38 **	6.24 **	0.003466
nek1	BB418199	1.16 *	0.49	0.47	0.69 **	0.000186
nfatc1	NM_016791	0.91	0.4	1.06	1.28	0.002721
nfkbia	BB096843	0.64 **	0.46 **	1.41	0.85 **	0.040282
nfu1	BC018355	0.9	0.41 **	3.89 **	5.25 **	0.000155
nr2c2	AV162817	1.11	0.48 **	1.21	1.11	0.026226
numb	U70674	0.7 **	0.4 **	1	0.5 **	0.00042
pard3	AW543460	0.93	0.55 *	0.66	0.87	0.011453
pard3	BE199556	0.53	0.54 *	1.37	0.62	0.000169
pard6a	NM_019695	0.69 **	0.38 **	1.67 **	0.53 **	0.004508
pbx1	L27453	0.95	0.43	0.93	1.05	0.037642
pbxip1	AV220340	0.42 **	0.6 **	0.41 **	0.61 **	0.004918
pdx1	AK020261	0.76 **	0.22 **	0.76 *	0.47 **	0.008211
pdx1	AK020261	0.69 **	0.38 **	0.71 *	0.74 **	0.041404
pgf	NM_008827	0.51 **	0.39 **	1.2	0.5 **	0.031444
pkd1	NM_013630	0.86	0.38 **	1.34	0.86	0.019938
ppargc1a	BB745167	0.51 **	1.08	0.36 **	0.32 **	0.00229
ppp1cc	BG071790	1.54	0.72 **	2.71 **	2.55 **	0.025228
ppp1r8	BC025479	2.48 **	0.68 **	2.25	2.11 *	0.048537
prkar1b	NM_008923	0.63 **	0.35 **	2.26 **	1.12 **	0.047205
prkar1b	BB274009	0.7 **	0.41 **	1.08 **	0.64 **	0.018337
prkar1b	BB283894	0.85 **	0.39 **	2.18 **	0.73 **	0.000177

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
rasgrf1	AF169826	0.69	0.49 **	1.23 **	0.92	0.0075
rasip1	AK003910	1.39	0.43 **	1.1	0.29 **	0.008586
rassf4	AV291679	0.4 **	0.26 *	0.83 *	0.54 **	0.003493
rb1cc1	BE570980	0.97	0.45 **	0.57	0.94	0.037941
reck	NM_016678	0.39 **	0.54 **	0.93	0.82	0.000164
rgs2	AF215668	1.51	0.32	2.55 **	1.02 *	0.022341
rnf6	BI738010	0.52 *	0.65 **	1.08	0.5 **	0.027362
rnf6	BI738010	0.77 *	0.46 **	1.15	0.46 **	0.040255
s100a6	NM_011313	0.5 **	0.94	1.45 *	1.62 **	0.0117
sash1	BI658899	0.54 **	0.99	1.03	0.97	0.00026
sesn3	NM_030261	0.77 *	0.51 **	1.19	0.73 **	0.015285
siah2	AA414485	0.52 **	0.76 **	1.32 **	0.87	0.011982
sipa1l1	BI153574	0.94	0.51 **	0.91	0.54	0.026869
smarca2	AK011935	0.82	0.48 **	0.79 *	0.8 *	0.001044
smarca2	BM230202	0.94	0.44 **	0.58 *	0.65 *	0.010874
stat5a	U36502	0.57 **	0.46 **	0.83	1	0.003117
tnfsf13	NM_023517	0.33 **	0.95	2.49	1.88	0.01273
trp53bp2	BB814564	0.85	0.43 **	1.6 *	1.1	0.037677
tspan5	AK015705	0.76 *	0.38 **	0.49 **	1.24	0.04554
tubb1	AW493179	0.41 **	0.96	1.25	0.8 *	0.000493
uchl1	NM_011670	0.47 **	0.34 **	1.75	3.78 **	0.030195
unknown	BF471533	0.56	0.4	0.94	1.17	0.019641
unknown	BM238838	1.95	0.32	0.35	0.38	0.001903

Supplementary Table 3: Cell cycle genes showing significant increase in expression within 8 hours of MycER^{TAM} activation in the skin suprabasal keratinocytes. p-value derived from analysis of the '4OHT' term in the *Envisage* model. Flags represent significant effects detected at specific time points ('*' = t-test p-value ≤ 0.05, '' = t-test p-value ≤ 0.01).**

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
akt1	NM_009652	0.79	3.33 **	1.07	2.72 **	0.003781
akt2	NM_007434	1.24	2.3 **	1.1	2.37 **	0.007853
atf5	AF375476	0.8	3.03 **	1.07	3.45 **	0.007942
b230120h23rik	BB561086	2.54 **	1.49 *	1.69 *	0.6 *	0.001607
bin1	U60884	1.28	3.12	1.15	3.1	0.00724
bop1	BM213936	0.77	2.37 **	1.38	2.51 **	7.74E-05
brms1	NM_134155	1.79	3.68 **	1.6	1.05	0.036994
camk2d	AF059029	0.72 **	2.07 **	0.97 **	1.22 **	0.009523
ccnd2	AK007904	1.63	3	1.03	2.4	0.014806
ccnd3	NM_007632	0.84 **	2.64 **	0.85 **	3.86	0.000512
cd34	NM_133654	0.98	2.76 **	1.04	3.37 **	0.006083
cd3g	M58149	1.55	2.22 **	1	0.8	0.044598
cd9	NM_007657	1.27	2.17 *	0.69	1.09	0.035585
cdc34	BI794243	1.22	2.21 **	1.52	1.94 *	0.037859
cdc37	AK013255	1.05	2.58 *	1.42	1.49	0.000426
cdk4	NM_009870	1.83	12.73 **	0.88	1.93	0.049585
cdk4	NM_009870	1.5	11.77 **	1.09	1.89	0.026067
cdk4	NM_009870	1.64	11.53 **	0.91	1.63	0.021676
cdk7	U11822	0.91	2.24 **	1.05	1.57 *	0.002113
cetn3	BC002162	0.69	2.22 **	0.99	1.33	0.01078
cfb	NM_008198	0.88	2.73 **	0.54 **	3.96 **	0.029819
cfl1	NM_007687	1.01	2.48 **	1.25	2.36 **	0.006256
cgref1	BC023116	0.67 **	2.26 **	0.82 *	2.59 **	0.033148
ciao1	AK004129	1.33	3.7 **	1.17	1.53	0.010593
cks1b	NM_016904	0.78 *	2.04 **	0.82	1.34	0.000506
cops5	NM_013715	1.03	2.01 **	0.93	1.71 **	0.008324
csf1	BM233698	1.67 **	3.02 **	1.79 **	3.89 **	0.000568
csf1r	AK004947	1.19	2.79 **	0.55 **	2.67 **	0.003082
ctnnb1	BI134907	1.34	2.36	1.09	1.67	0.034574
eef1e1	NM_025380	1.49	3.48 **	1.65 *	1.58 **	8.73E-05
erh	BB071632	0.65 **	3.66 **	0.91	2.17 **	4.37E-05
evi2a	NM_010161	0.98	3.68 **	0.64 *	3.18 **	0.049158
fcgr1	AF143181	0.62	5.88	1.32	8.27 **	0.000362
fes	BG867327	1.11	2.55 **	0.75 **	3.18 **	0.000477
gadd45g	AK007410	4.44 **	5.44 **	3.27 **	2.87 **	1.30E-05
gadd45gip1	BE368753	0.47 **	3.5 **	0.55 **	3.85 **	0.034335

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
hras1	NM_008284	1.19	2.87 **	0.8	1.42	0.013043
hras1	BC011083	1.06	2.2 **	0.86	1.36	0.016689
hspa8	BC006722	0.93	2.38	0.93	2.2	0.021573
hspa8	BC006722	1.13	2.48	0.94	2.6	0.009973
ifrd2	BB540964	1.06	2.23 **	1.59	3.3 **	0.00325
igf1	NM_010512	1.09 *	2.5 **	1.61	2.35 **	0.000129
igf1	AF440694	1.8 *	2.31 **	1.13	3.17 **	0.037463
igh-6	AI326478	0.14 **	3.06	0.32 **	3.26 **	0.001176
itgb2	NM_008404	1.3	3.05 **	0.62	4.39 **	0.017104
mis12	BC026790	0.83	2.06 **	1.08	1.86 **	0.003241
ndel1	BC021434	2.39 *	0.27 **	0.46 **	1.04	0.00203
nek6	BB528391	0.85	2.11 **	1.3	2.25 **	0.000155
nfu1	BC018355	1.14	2.29 **	0.89	1.56	0.005464
pcna	BC010343	1.42	3.23 **	0.86	3.18 **	0.031444
pgf	NM_008827	1.1	2.33 **	2.23 **	8.19 **	0.009284
polr3d	BC016102	1.47 *	4.21 **	1.52	2.63 **	0.00229
ppargc1a	BB745167	2.3 **	0.84	0.78	0.74	0.003812
pstpip1	U87814	0.98	2.73 **	1.53	1.98 **	0.010424
ptgds2	NM_019455	0.93	2.96 **	0.93	1.09	0.00377
rad1	NM_011232	0.75	3.33 **	0.79	2.22 *	2.13E-05
ran	AV090150	0.86	2.17 **	1.35	1.75 **	0.019522
rassf2	AK018504	1.03	3.77 **	1.01	1.5 **	0.003309
rogdi	BC006914	0.72	2.16 **	0.96	1.31	0.014577
rp23-143a14.5	NM_027136	0.71	2.22 **	0.98	1.87 **	0.000116
ruvbl1	NM_019685	0.97	2.21 **	1.57	2.16 **	0.005751
sept11	AV229846	2.16 **	1.02	1.46 **	0.9	0.005006
sept9	NM_017380	0.75	2.3 **	0.94	2.1 **	0.003502
sfpi1	NM_011355	0.77	2.53 **	0.91	4.66 **	0.002031
skp1a	AV347477	0.85 *	2.14 **	0.93	1.03	0.007487
slfn2	NM_011408	1.34	2.75 **	0.94	2.88 **	0.001677
tial1	NM_009383	0.76 **	2.33 **	0.91	2.83 **	0.01273
tnfsf13	NM_023517	1.03	2.13 *	0.35 **	1.47	0.001952
tnfsf5ip1	BC016606	1.1	3.17	1.06	1.99	0.000653
triap1	AK007514	1.02	3.6 **	0.92	2.11 **	0.034875
u2af1	NM_024187	0.98	2.16 **	1.2	1.52	0.002046
ywhab	NM_018753	0.84	2.23 **	1.09	2.42 **	0.026275

Supplementary Table 4: Cell cycle genes showing significant decrease in expression within 8 hours of MycER^{TAM} activation in the skin suprabasal keratinocytes. p-value derived from analysis of the '4OHT' term in the *Envisage* model. Flags represent significant effects detected at specific time points ('*' = t-test p-value ≤ 0.05, '' = t-test p-value ≤ 0.01).**

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
4732435n03rik	AV371987	0.77	0.42 **	0.5 **	0.81	0.000556
ahr	BE989096	1.22	0.47 **	0.92	0.51 **	0.031511
ai467657	AA419994	0.91	0.31 **	1.04	0.46 **	3.03E-05
asah2	NM_018830	0.65 **	0.48 **	0.9	0.89	4.62E-07
bmp2	AV239587	0.44 **	0.37 **	0.88 *	0.41 **	7.86E-05
bmp4	NM_007554	0.41 **	0.76	0.58 **	0.67 *	0.000697
btc	NM_007568	0.34	0.27 **	0.65 **	0.47 **	0.000705
btc	AV231340	0.78	0.33 **	0.68 **	0.57 **	0.006474
bub1b	AU045529	0.48 **	0.53 *	1.11	0.9	0.000959
camk2d	NM_023813	0.9 **	0.44 **	0.81 **	0.6 **	0.004559
camk2d	AV337193	0.56 **	0.47 **	0.64 **	0.35 **	0.000282
ccnb1-rs1	NM_007629	0.69	0.44 *	1.36	0.66	0.002143
cdc14b	AK013228	0.84	0.32 **	0.84	0.64 *	0.009572
cdc25a	C76119	1.19	0.44	1.08	0.81	0.00184
cdc73	BM935271	0.97	0.39 **	1.06	0.52 **	0.00153
cdkn1b	NM_009875	0.2	0.51 **	0.5	1.06 *	0.010159
chek1	C85740	0.83	0.43	1.45	1.15	3.29E-05
crip3	AF367970	0.48 **	0.87 *	0.61 **	0.91	0.001596
d5ertd40e	C77487	0.93	0.47 **	1.24	1.14	0.001553
dbf4	NM_013726	0.41 **	1.39 *	1.19	1.6 **	6.87E-05
dhcr7	NM_007856	0.48 **	0.49 **	0.99	1.41 *	0.027391
ednra	AW558570	0.94	0.43 **	0.84 *	0.46 **	0.000479
efnb1	NM_010110	0.39 **	0.51 **	1.11	1.09	0.014831
egf	NM_010113	1.64 **	0.47 **	0.93	0.91	2.13E-05
egfr	AV369812	1.15	0.4 **	1.37 *	0.49 **	0.010188
erbb3	BF140685	0.89	0.48 **	0.74	0.45 **	0.000618
ereg	NM_007950	0.16 **	0.27 **	0.63 **	0.93	0.003432
esr1	NM_007956	0.44	0.86	0.58	1.03	0.034335
figf	NM_010216	0.76 **	0.47 **	1	1.63 **	8.54E-05
frk	BB787292	0.45	0.98	0.51	0.7	0.004749
gadd45gip1	BE368753	0.47 **	3.5 **	0.55 **	3.85 **	0.000139
gas1	BB550400	1.16	0.4 **	0.99	0.42 **	0.00272
gspt1	AW537663	0.5 **	0.94	1.23	1.23	0.037463
h2-ea	U13648	1.15	0.3 **	0.59 **	1.66 **	0.000194
hells	NM_008234	0.36	0.53	1.38	1.33	0.002514
igh-6	AI326478	0.14 **	3.06	0.32 **	3.26 **	0.010831

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
il1a	BC003727	0.46 **	0.93	0.44 **	0.32 **	0.005882
jag1	AV359819	0.71 **	0.47 **	0.92	0.52 **	0.035102
kitl	BB815530	1.11	0.42 **	1.26	0.7	0.003406
kitl	BB815530	1.07	0.29 **	1.17	1.31	0.002906
loh11cr2a	BC004727	0.87	0.43 **	1.65 **	0.8 *	0.00041
lpp	BB557975	1.32	0.42 **	1.28	0.58 **	0.019714
lpp	BM236111	1.17	0.5 **	0.82	0.64 **	0.003241
mcm6	BB099487	0.51 **	0.59	1.21	1.16	0.024372
myc	BC006728	0.31 **	0.7 *	0.59 **	1.27	0.011453
ndel1	BC021434	2.39 *	0.27 **	0.46 **	1.04	0.007197
nipbl	BF661272	0.99	0.47 **	1.19	0.78 *	0.004969
pard3	AW543460	1.09 *	0.49 **	1.08 *	0.73	0.01133
pard3	BG063922	0.78 *	0.45 **	0.7 *	1.19	0.022582
pard6b	BE953582	0.52	0.33 **	0.96	0.91 **	0.005339
pard6b	BE953582	0.67	0.37 **	1.19	0.5 **	0.000177
pmaip1	NM_021451	1.43 *	0.44 **	0.82 *	0.9	0.009509
ppp1r13b	BG064715	0.7	0.52	0.78	1.19	0.038674
prkar1b	BB283894	0.29 **	0.58 **	0.47 **	1.11	0.043241
prkar2b	BB216074	0.59 **	1.16	0.94	1.5 *	0.002571
prox1	BE994433	1.18	0.5 **	1.08	0.91	0.026183
ptprv	NM_007955	0.49 **	1.36	0.97	2.1 **	0.000852
pycard	BG084230	0.49 **	1.34	0.83	1.44 *	0.000908
racgap1	NM_012025	0.49	0.86	0.99	1.76	0.021757
rb1cc1	BE570980	1.11	0.48	1.03	0.66	0.006484
rif1	AK018316	0.6	0.49	1.7	0.74	0.007547
rif1	AK018316	1.37	0.49	1.5	0.67	0.003072
runx2	D14636	0.5 **	0.58 **	0.81 *	0.84 *	0.027828
scin	NM_009132	1.12	0.49 **	1.16	0.62 **	0.006421
skp2	AV259620	0.53	0.49 *	1.02	1.11	3.71E-06
smarcb1	NM_011418	0.49	1.17	0.88	2.16	0.001903
sphk2	AK016616	0.41 **	0.61 **	0.93	0.69 *	0.014006
tgfa	M92420	0.83	0.32 **	0.78	0.43 **	0.006056
Unknown	BM238838	1.26	0.45	0.67	1.13	0.004993
Unknown	BB071777	0.89	0.43	0.82	0.53	0.009326
Unknown	AU042527	1.78	0.45	1.08	0.93	0.013948
vegfc	BB089170	0.79 **	0.39	0.85	0.79	0.016314
zbtb16	Z47205	1.95	0.52	0.89	0.51 *	0.031087
zwint	BC013559	0.66 **	0.48 **	1.25	0.75 **	0.026178

Supplementary Table 5: Apoptosis genes showing significant increase in expression within 8 hours of MycER^{TAM} activation in the pancreatic β-cells. p-value derived from analysis of the '4OHT' term in the *Envisage* model. Flags represent significant effects detected at specific time points ('*' = t-test p-value ≤ 0.05, '' = t-test p-value ≤ 0.01).**

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
bcl2	BI664467	1.32	2.5 **	0.99	1.29	0.007792
birc6	BG071331	2.41 **	0.59 *	0.94	0.92	0.044489
bnip1	BG073508	2.6 **	1.1	3.59 **	4.69 **	0.000214
camk1d	BG071931	2.91 **	2.31 **	0.76	1.29	0.010131
cd40	BB220422	3.08	1.36	2.64 *	0.94	1.74E-05
cdc2a	NM_007659	1.06	4.54 **	14.85 **	3.45 **	0.004186
cdkn1a	AK007630	1.56 **	2.12 **	1.62 **	1.49	0.005006
cdkn2a	NM_009877	2.54 **	2 **	1.49 **	2.7 **	0.000264
cse1l	NM_023565	1.62 **	2.3 **	3.23 **	1.81 **	4.37E-05
cugbp2	BB644164	1.45	2.38 *	1.19 **	1.06	0.010569
cycs	NM_007808	1.15 *	2.04 **	1.64 **	2.06 **	0.001638
eef1e1	NM_025380	3.15 **	2.33 **	7.85 **	8.45 **	8.73E-05
endog	NM_007931	2.66 **	3.12 **	2.27 **	2.07 **	7.61E-06
fas	BG976607	6.57	6.48	3.57	14.03	0.000139
fas	BG976607	4	7.72	2.13	4.96	0.006321
fas	BG976607	9.9	13.59	9.74	11.38	0.001117
fas	BG976607	2.74	1.87	2.58	2.58	0.002214
fastkd1	BE957020	2.54 **	1.16	1.87 **	1.61 *	0.011133
hells	NM_008234	2.59	14.95 **	20.22 **	10.82 **	0.00272
hells	AK021390	1.27	6.08 **	5.18 **	3.95 **	0.049171
htra2	AW323050	1.49	2.25	1.16 **	1.75 **	0.016942
igf1	NM_010512	2.21	0.64	1.05	1.09 **	0.00325
jag2	AV264681	3.32 **	2.62	2.14	4.69 **	0.000236
jag2	AV264681	2.57 **	1.41	1.01	2.49 **	0.001409
litaf	AV360881	3.17 **	0.81	1.78 **	1.7 **	0.003786
nup62	NM_053074	1.87	2.71	4.56 **	2.57 **	8.77E-06
nup62	AW240611	2.17	2.41	2.93 **	2.7 **	3.73E-05
parp1	BB767586	2.48 **	2.26 **	0.97	1.95 **	8.08E-05
pdcd11	AK003899	3.61 **	1.75 *	2.09 **	3.42 **	0.000226
prkar2a	AK004336	2.55 **	2.79 **	1.41 *	2.71 **	0.001012
prkar2a	AK004336	2.33 **	2.59 **	1.14 *	3.22 **	2.16E-05
ptrh2	BB178232	2.23 **	2.39 *	0.61	1.62	0.023708
ripk2	NM_138952	2.22 **	0.92	1.1	2.52 **	0.007429
siva1	AF033112	1.26	2.27 **	3.4 **	1.53 **	0.020671
tbrg4	BB139935	2.09	1.74	1.16 **	2.26 **	7.12E-05
tfdp1	BG075396	3.14 **	1.86 **	3.48 **	3.2 **	7.99E-06

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
thoc1	BC024951	2.15 **	0.91 **	1.4 *	1.58 **	thoc1
thoc1	BC024951	2.13 **	0.97 **	1.64 *	1.19 **	thoc1
thoc1	BG066490	2.18 **	3.27 **	1.33 *	2.4 **	thoc1
tnfsf5ip1	NM_134138	2.2 **	2	2.47 **	5.52 **	tnfsf5ip1
usp14	AW107924	1.94 **	2.33 **	0.91 **	1.04	usp14

Supplementary Table 6: Apoptosis genes showing significant decrease in expression within 8 hours of MycER^{TAM} activation in the pancreatic β-cells. p-value derived from analysis of the '4OHT' term in the *Envisage* model. Flags represent significant effects detected at specific time points ('*' = t-test p-value ≤ 0.05, '' = t-test p-value ≤ 0.01).**

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
aatk	NM_007377	0.64 **	0.4 **	0.93	1.16	0.000473
agt	AK018763	1.02	0.27 **	1.75 **	2.23 **	0.004482
apaf1	AK018076	0.5	1.14	1.42	1.32	0.007714
asah2	NM_018830	0.47 **	0.25 **	0.55 **	0.7 *	4.62E-07
bcl2l2	BB485989	1.29	0.41 **	0.53 **	0.53 **	0.012412
bnip3	NM_009760	0.75	0.48 **	1.55 **	0.67 **	0.014857
btg1	L16846	0.47 **	0.42 **	1.09	0.64 **	0.002209
btg1	AW322026	0.62 **	0.4 **	0.99	0.74 **	0.002357
btg2	NM_007570	0.31 **	0.43 **	1.25 *	0.78	0.015696
btg2	NM_007570	0.45 **	0.34 **	1.71 *	1.11	0.025367
card6	BB766747	0.44 **	0.55 **	0.73 *	0.51 **	4.66E-05
ctnna1	NM_009818	0.49 **	0.62 **	1.51 **	0.92	0.000473
dap	BC024876	0.77 *	0.52 **	1.19	0.71 *	0.004095
dlg5	BC021314	0.72 *	0.45 **	0.71 **	0.98	0.003389
efhc1	AK006489	0.57 **	0.41 **	1.09	0.62 **	0.001059
elmo3	AI481208	1.27	0.29 **	0.79	0.48 **	0.021212
eya1	BB760085	0.81 *	0.46 **	1.51 **	1.14	0.017138
gadd45b	AK010420	0.73 *	0.33 **	2.11 **	0.48 **	0.001694
gadd45g	AK007410	0.58 *	0.4 **	2.08 **	0.63 **	1.30E-05
gsk3b	BB831420	0.92	0.44 **	0.7 **	0.57 **	0.002762
htatip2	AF061972	0.84	0.41 **	0.78 **	0.72 **	0.016576
ikbkg	BB821318	1.72 **	0.49 **	0.71 *	3 **	0.004462
il1r1	NM_008362	0.53 **	0.43 **	1.7 *	0.6 *	0.002892
inhba	NM_008380	0.58 **	0.42 **	0.98	0.64 **	0.013148
irf6	NM_016851	0.71 *	0.44 **	1.19	0.81	0.019855
kitl	BB815530	0.76 *	0.38 **	0.33 **	0.55 **	0.005882
mapk10	BB453775	0.46 **	0.42 **	1.23 **	0.87 **	4.07E-06
mapk8ip1	BB546463	0.75 **	0.43 **	0.7	0.5 **	0.015573
mapk8ip2	AW536912	0.44 **	0.94	1.8 *	1.67 *	0.005901
mitf	BB763517	0.46 *	0.43 **	0.98	0.93	0.01716
nfkbia	BB096843	0.64 **	0.46 **	1.41	0.85 **	0.040282
nod1	BB138330	1.03	0.31 **	1.25	1.25	0.023552
nuak2	AK004737	0.5 **	1.13	1.97 **	0.7 **	0.024528
peg3	AB003040	0.45 **	0.25 **	0.79 **	0.55 **	0.000397
phlda1	NM_009344	0.51 **	1.06	1.57 *	0.7 *	0.030067
pik3ca	AI528567	0.83 *	0.48 **	1.53 **	0.92	0.002475

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
prkar1b	NM_008923	0.63 **	0.35 **	2.26 **	1.12 **	0.047205
rtn4	BE988775	0.67 **	0.61 **	0.57 *	0.66 **	0.001425
rtn4	AK003859	0.57 **	0.44 **	1.28 *	0.59 **	0.00012
serinc3	BM244064	0.65 *	0.5	1.23	0.82	0.011082
sgk	NM_011361	1.01	0.53 **	1.27	1.09	0.019309
sh3kbp1	BB326929	0.74 *	0.48 **	0.47 **	1.03	0.011982
siah2	AA414485	0.52 **	0.76 **	1.32 **	0.87	0.035316
sqstm1	BM232298	0.46 **	0.3 **	3.04 **	0.59 *	0.011586
stambp	AA289490	0.74 **	0.34 **	0.56 **	0.74 **	0.003117
stat5a	U36502	0.57 **	0.46 **	0.83	1	0.003252
tnfrsf22	BB366863	0.36 **	0.76 *	1.88 *	0.91	0.01273
tnfsf13	NM_023517	0.33 **	0.95	2.49	1.88	0.029572
traf6	AV244412	0.8 *	0.32 **	1.09	1.14	0.037677

Supplementary Table 7: Activation of MycER^{TAM} in pancreatic β-cells resulted in significant change in expression of genes involved in DNA damage response. p-value derived from analysis of the '4OHT' term in the *Envisage* model. Flags represent significant effects detected at specific time points ('*' = t-test p-value ≤ 0.05, '' = t-test p-value ≤ 0.01).**

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
atr	AF236887	2.31	4.04 **	2.57	2.87	4.44E-05
cdc2a	NM_007659	1.06	4.54 **	14.85 **	3.45 **	0.004186
chek1	C85740	2.05	5.34 *	6.01 **	3.86 **	0.0040477
chek1	BB298208	1.15	3.84 *	2.76 **	2.05 **	0.0049931
chek1	NM_007691	1.52	2.19 *	3.12 **	1.65 **	0.0058294
h2afx	NM_010436	1.16	2.42 **	5.22 **	1.4 **	0.0011743
hus1	AF076845	1.45	1.1	2.35 **	1.3	0.0146337
hus1	NM_008316	0.93	2.13	2.07 **	1.67	0.0185243
nbn	NM_013752	1.65 *	1.02	2.4 **	1.83 **	0.0116503
ptprv	NM_007955	1.01	0.95	1.06	0.95	0.043241
rad1	NM_011232	2.08 **	2.87 **	2.2 *	1.76	0.00377
rad51	NM_011234	1.07	6.97 **	30.92 **	2.39 **	0.000283
triap1	AK007514	1.76 **	1.12	2.03 **	1.76 *	0.0006526

Supplementary Table 8: Apoptosis genes showing significant increase in expression within 8 hours of MycER^{TAM} activation in the suprabasal keratinocytes. p-value derived from analysis of the '4OHT' term in the *Envisage* model. Flags represent significant effects detected at specific time points ('*' = t-test p-value ≤ 0.05, '' = t-test p-value ≤ 0.01).**

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
akt1	NM_009652	0.79	3.33 **	1.07	2.72 **	0.003781
akt2	NM_007434	1.24	2.3 **	1.1	2.37 **	0.007853
atf5	AF375476	0.8	3.03 **	1.07	3.45 **	0.007942
b230120h23rik	BB561086	2.54 **	1.49 *	1.69 *	0.6 *	0.001607
bag1	NM_009736	0.72	2.33 **	1.11	2.01 **	0.003798
bnip1	BG073508	0.56 **	2 **	0.84	2.44 **	0.000214
c1qtnf6	AK012868	0.59 *	3.55 **	0.57 **	6.15 **	0.018492
cd209b	AF374471	1.14	2.09 **	0.32 **	1.2 **	0.001609
cd3g	M58149	1.55	2.22 **	1	0.8	0.044598
ciapin1	NM_134141	1.17	2.72 **	1.14	2.49 **	0.032304
dnaja3	AK004575	1.2	2.62 **	1.33	2.47 **	0.007569
eef1e1	NM_025380	1.49	3.48 **	1.65 *	1.58 **	8.73E-05
eif5a	BF384094	1.5	2.56 **	1.38	1.71	7.99E-06
fas	BG976607	0.85	5.35	0.86 **	0.49	0.001117
fcer1g	NM_010185	0.91	3.71 **	0.84	4.54 **	0.001439
gadd45g	AK007410	4.44 **	5.44 **	3.27 **	2.87 **	1.30E-05
hras1	NM_008284	1.19	2.87 **	0.8	1.42	0.013043
hras1	BC011083	1.06	2.2 **	0.86	1.36	0.005963
hspa5	AJ002387	1.12	2.74 **	0.91	2.28 **	0.001325
igf1	NM_010512	1.09 *	2.5 **	1.61	2.35 **	0.00325
igf1	AF440694	1.8 *	2.31 **	1.13	3.17 **	0.000129
jmjd6	AK017622	1.08	2.31 **	1.09	2.42 **	0.000231
lsp1	NM_019391	1.08	3.67 **	0.66 **	1.92 **	0.008392
nek6	BB528391	0.85	2.11 **	1.3	2.25 **	0.00203
pdcd2l	AK003339	0.76	2.84 **	0.88 **	2.09 **	0.028861
pdcd2l	AK003339	0.96	2.14 **	0.74 **	1.62 **	0.005131
pigt	AK019717	1	2.08 **	0.93 *	1.47	0.002608
ptrh2	BC026947	0.98	2.26	0.92	1.35 *	0.035495
siva1	NM_013929	1.04	2.18 **	0.91	1.52	0.003626
siva1	AF033112	1.02	2.4 **	1.28	1.38	0.006496
sqstm1	BM232298	2.13 *	2.3 **	0.93	0.91	0.035316
tial1	NM_009383	0.76 **	2.33 **	0.91	2.83 **	0.001677
tlr1	AF316985	2.2 **	3.7 **	1.87 **	5.01 **	0.000375
tnfrsf12a	NM_013749	1.14	3.21 **	1.44 **	1.18	0.013343
tnfrsf12a	NM_013749	1.22	2.9 **	1.4 **	1.14	0.010767
tnfrsf4	NM_011659	1.25	2.24 **	1.94 **	1.04	0.005782

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
tnfsf13	NM_023517	1.03	2.13 *	0.35 **	1.47	0.01273
tnfsf5ip1	BC016606	1.1	3.17	1.06	1.99	0.001952
tnfsf9	NM_009404	0.95	2.24 **	1.04	0.81	0.0447
tradd	BB749262	0.8	3.51 **	0.78	1.45	0.000653
triap1	AK007514	1.02	3.6 **	0.92	2.11 **	0.010491

Supplementary Table 9: Apoptosis genes showing significant decrease in expression within 8 hours of MycER^{TAM} activation in the suprabasal keratinocytes. p-value derived from analysis of the '4OHT' term in the *Envisage* model. Flags represent significant effects detected at specific time points ('*' = t-test p-value ≤ 0.05, '' = t-test p-value ≤ 0.01).**

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
ahr	BE989096	1.22	0.47 **	0.92	0.51 **	0.031511
ai467657	AA419994	0.91	0.31 **	1.04	0.46 **	3.03E-05
asah2	NM_018830	0.65 **	0.48 **	0.9	0.89	4.62E-07
bcl2l11	BB667581	1.96 **	0.43 **	0.96	0.79 **	0.036131
birc4	BF134200	0.51 **	0.34 **	1.12	0.87	0.005291
bmf	BB212341	0.62 **	0.29 **	0.5 **	0.64 **	5.31E-05
c1qtnf7	BB039211	0.62 **	0.28 **	0.69 **	0.51 **	5.42E-06
elmo3	AI481208	0.5 **	0.89	0.67 *	0.9	0.021212
gas1	BB550400	1.16	0.4 **	0.99	0.42 **	8.54E-05
glo1	BC024663	1	0.35 **	1.3	1.13	0.03944
glo1	BC024663	1.13	0.32 **	1.3	0.79	0.032921
hells	NM_008234	0.36	0.53	1.38	1.33	0.00272
il1a	BC003727	0.46 **	0.93	0.44 **	0.32 **	0.000194
il1rap	BE285634	0.82 *	0.43	0.92 **	0.44	0.009618
kitl	BB815530	1.11	0.42 **	1.26	0.7	0.010831
kitl	BB815530	1.07	0.29 **	1.17	1.31	0.005882
myc	BC006728	0.31 **	0.7 *	0.59 **	1.27	0.019714
pmaip1	NM_021451	1.43 *	0.44 **	0.82 *	0.9	0.022582
ppp1r13b	BG064715	0.7	0.52	0.78	1.19	0.005339
prkar2b	BB216074	0.59 **	1.16	0.94	1.5 *	0.009509
ptprv	NM_007955	0.49 **	1.36	0.97	2.1 **	0.043241
pycard	BG084230	0.49 **	1.34	0.83	1.44 *	0.002571
rffl	AW123157	0.83	0.44 **	0.99	0.61 **	0.034411
scin	NM_009132	1.12	0.49 **	1.16	0.62 **	0.007547
serinc3	BM239368	0.99	0.46 **	0.8 *	1.04	0.017428
sgpl1	NM_009163	0.81 **	0.45 **	0.75 **	1.36 **	0.000419
sphk2	AK016616	0.41 **	0.61 **	0.93	0.69 *	0.006421
tm2d1	AF353993	0.49 **	1.02	0.88	1.79 **	0.023572
tns4	BB142697	0.67 *	0.44 **	0.92	0.93	0.005475
trib3	BB508622	1	0.51 **	0.71	0.63 *	0.001954
zbtb16	Z47205	1.95	0.52	0.89	0.51 *	0.031087

Supplementary Table 10: Mature islet β-cell genes up-regulated within 8 hours following activation of MycER^{TAM}. p-value derived from analysis of the '4OHT' term in the *Envisage* model. Flags represent significant effects detected at specific time points ('*' = t-test p-value ≤ 0.05, '' = t-test p-value ≤ 0.01).**

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
creg1	BC027426	2.09 **	2.25 **	1.52 **	2.31 **	0.001706
h2-aa	AV086906	7.66	1.52 **	0.74 *	4.26	0.006055
h2-d1	M34962	4.55	2.25	1.59	5.37 *	0.002308
h2-l	M86502	4.93	1.99	1.94	5.96 **	0.005144
h2-l	M69068	4.72	1.91	1.9	7.18 **	0.001647
pcnt	NM_008787	2.63	4.27	4.98 **	4.78	5.43E-06
pcsk2	BB357975	2.41 **	0.56 **	0.78 **	1.09 **	0.01453

Supplementary Table 11: Mature islet β-cell genes down-regulated within 8 hours following activation of MycER^{TAM}. p-value derived from analysis of the '4OHT' term in the *Envisage* model. Flags represent significant effects detected at specific time points ('*' = t-test p-value ≤ 0.05, '' = t-test p-value ≤ 0.01).**

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
1110003e01rik	NM_133697	0.64 **	0.35 **	1.82	0.6 **	0.000882
1810015c04rik	NM_025459	0.75	0.36 **	1.02	0.46 **	0.032155
acot11	NM_025590	0.76	0.32 **	1.04	0.82	0.000444
arfgap3	BG067878	0.5 **	0.52 **	1.13	0.45 **	0.014405
c3	K02782	0.9	0.52 *	2.85 *	2.19 **	0.006855
ddc	AF071068	0.63 **	0.49 **	1.64 *	1.08	0.010593
defb1	BC024380	0.33 **	0.41 **	1.02	0.23 **	5.13E-05
defb1	BC024380	0.29 **	0.48 **	0.91	0.2 **	4.09E-05
fkbp1b	NM_016863	0.87	0.48 **	1.21	0.67 *	0.001985
gramd3	AV259880	0.71 **	0.36 **	0.64 **	0.72 **	0.000283
h2-d1	NM_010380	0.54	0.41	2.18	1.12 *	0.004778
il1r1	NM_008362	0.53 **	0.43 **	1.7 *	0.6 *	0.002892
il6ra	X53802	0.36 **	0.45 *	1	0.47 **	0.004703
kl	BQ175355	0.84	0.37 **	0.98	0.89	0.021298
lamp2	BB390704	0.98	0.49 **	1.28	0.87	0.01545
mapk10	BB453775	0.46 **	0.42 **	1.23 **	0.87 **	4.07E-06
mapk10	L35236	0.47 **	0.23 **	1.64 **	1.77 **	0.007442
ndrg4	AV006122	0.43 **	0.4 **	0.85	0.77	0.009491
ndrg4	AI837704	0.31 **	0.35 **	1.24	0.93	0.046516
nupr1	NM_019738	0.6 *	0.37 **	1.31	0.51 *	0.00678
nupr1	NM_019738	0.61 *	0.53 **	1.29	0.67 *	0.03423
papss2	BF786072	0.21 **	0.19 **	1.39	0.22 **	1.54E-05
papss2	BF786072	0.32 **	0.16 **	0.69	0.3 **	0.000236
papss2	BF780807	0.68 **	0.13 **	0.97	0.38 **	0.000156
pcsk2	AI839700	1 **	0.48 **	0.74 **	0.96 **	0.037893
pcsk2	NM_008792	0.57 **	0.39 **	1.99 **	2.68 **	0.041463
pftk1	AI327038	0.43	1.16	1.36 **	0.81	0.023166
pgcp	BB468025	0.72	0.52 **	1.12	0.88	0.02741
pip5k1b	NM_008846	0.29 *	0.83 **	2.88 **	1.07 *	0.008443
pip5k1b	NM_008846	0.63 *	0.44 **	1.88 **	1.74 *	0.009431
psmb9	NM_013585	0.47 **	0.23 **	2.31 **	0.65 *	0.000561
ptprn	NM_008985	0.48 **	0.48 **	0.86	1.98 *	0.008885
rgs2	AF215668	1.51	0.32	2.55 **	1.02 *	0.022341
slc2a2	NM_031197	0.89 *	0.42 **	0.97	0.77 **	0.002479
snap25	BE952593	0.56 **	0.45 **	1.71 **	1.11	0.012387

		Fold change from control				
Gene Symbol	**GenBank**	**4 hours**	**8 hours**	**16 hours**	**32 hours**	**4OHT p-value**
sqrdl	AF174535	0.51 **	0.4 **	1.62 *	0.76	0.003079
tapbpl	BC017613	0.54 **	0.41 **	1.93 **	0.84	0.001013
tspyl4	BC017540	0.59 *	0.45 **	1.11	0.91	0.00143