MOLECULAR MECHANISMS INVOLVED IN THE INDUCTION OF PLURIPOTENCY

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr.rer.nat.)
submitted to the Department of Biology, Chemistry, Pharmacy
of the Freie Universität Berlin

by

Guifré Ruiz Acero
from Barcelona
March 2012

This Ph.D. was carried out at the Max Planck Institute for Molecular Genetics under the supervision of Prof. Hans Lehrach between December 2008 and March 2012

1st Reviewer: Prof. Dr. Hans Lehrach
Max Planck Institute for Molecular Genetics

2nd Reviewer: Prof. Dr. Rupert Mutzel
Free University Berlin

Date of defence: 03.07.2012

To my parents, in gratitude

Acknowledgements

I am indebted to many people for their long-lasting support and encouragement, which was invaluable for the successful completion of this work. It is a pleasure to have the opportunity to thank them, even though I am afraid that these words will not be enough to reflect the deep gratitude and great respect that I feel for many of them.

Firstly, I would like to thank Prof. Dr. Hans Lehrach, who initiated the thesis project, supported it scientifically and financially and gave me the unique opportunity to use the excellent research facilities in his department. I shall be always inspired by his wisdom and perseverance, and will pursue to attain his visionary leadership.

I have a particular debt of gratitude to Dr. Sylvia Krobitsch, who believed in this work from day one and shape my thinking and writing along these years. I am very grateful to her for supporting me with intelligence, skill, and dedication, throughout the experimentation, analysis, writing and rewriting.

I want to thank Dr. Christoph Wierling for introducing me to mathematical modelling and critically reviewing the work. I am also very thankful to Dr. Alexander Kühn, who helped me with the modelling and searched patiently for the best outcome of the simulations. I cannot forget the aid of Dr. Mireia Nogales with all statistical issues that came along this project.

My understanding and personal development have been largely influenced by friends and colleagues as Dr. Boris Greber, Dr. Justyna Jagodziska, Dr. Smita Sudheer, Dr. Thore Brink and Dr. Marc Jung, to which I will be always thankful. I would like to express my most sincere gratitute to Silke Wehrmeyer, Christian Kähler, Christian Linke, Dr. Linda Hallen, Dr. Franzi Welzel and Anja Nowka, who offered me a warm welcome in their group and generously shared their knowledge and expertises with me.

During my years at Max Planck Institute for Molecular Genetics I had contact to many extraordinary people, who I shall thank for their scientific discussion, advice and continuous support along my thesis, among them account Dr. Florian Mertes, Uta Marchfelder, Dr. Karin Habermann, Dr. Hannah Mueller, Dr. Hans-Jörg Warnatz, Dr. Hendrick Hache, Dr. Markus Ralser, Dr. Nils Rademacher, Dr. Tracie Pennimpede and Dr. Robert Querfurth.

I owe my parents and brothers the greatest debt of gratitude for their unconditional support that helped me to overcome the most difficult moments. I would like express my gratitude also to Maren, Jürgen and Anika Beese for their support.

Most of all I thank my wife and soul mate, Katharina Beese, for everything—her encouragement, critiques, tolerance, patience, empathy and love.

I would also like to thank the foundations and supporters of my work: Fundación La Caixa, Deutscher Akademischer Austausch Dienst, and the Max Planck Society.

Contents

1. Abstract

The landmark discovery that lineage-restricted cells can acquire a stable pluripotent state through the ectopic expression of a specific set of transcription factors has expanded the boundaries of regenerative medicine to a yet indefinite end. Induced pluripotent stem cells (iPS) have emerged as an invaluable tool for generating patient-specific therapies. Even though, many technical advances have been made to improve the methodology less is known about the molecular mechanisms involved in the process of reprogramming.

In order to investigate the molecular mechanisms that promote or hinder the induction of pluripotency I have studied the genome-wide transcriptional profile of human fibroblasts after over-expression of the reprogramming transcription factors (OCT4, SOX2, KLF4, and c-MYC) independently and in combination. The analysis of the transcriptome at the early stages of the reprogramming permitted the identification of the foremost pathways involved in the reprogramming of somatic cells. Moreover, these pathways were given functional meaning by assembling a mathematical model that helped to propose hypotheses about the molecular mechanisms driving the reprogramming process. These hypotheses have been validated in loss of function experiments by targeting genetically or pharmacologically the most relevant proteins involved, as p53 and TGFβ receptor, during the reprogramming process.

The data presented in this work demonstrate that the reprogramming of somatic cells begins in the pursuit of a rapid, embryonic-like, cell cycle. The synergistic effect of the reprogramming factors regulates the TGFβ and the p53 signalling pathway reducing their related anti-growth responses, barriers to reprogramming. Regulation of these pathways eliminates the G_1-S transition checkpoint promoting a cell cycle progression similar to embryonic stem cells. In addition, I probed that this regulation of the cell cycle has down-stream effects on DNA damage responses and apoptosis. Faster G_1-S transitions reduced the activity on replication-related repair mechanisms in favour of homologous recombination that became the most prominent repair mechanism in the cells.

2. Zusammenfassung

Die bahnbrechende Entdeckung, dass somatische Zellen einen stabilen pluripotenten Zustand durch die ektopische Expression von spezifischen Transkriptionsfaktoren annehmen können, hat die Möglichkeiten der regenerativen Medizin wesentlich erweitert. Induzierte pluripotente Stammzellen (iPS) haben sich zu einem außerordentlich wertvollen Instrument entwickelt, um patientenspezifische Therapien zu ermöglichen. Obwohl große technologische Fortschritte in der Verbesserung der Methodik erzielt wurden, sind die zugrundeliegenden molekularen Mechanismen im Prozess der Zellreprogrammierung weitgehend unbekannt.

Um die molekularen Mechanismen zu untersuchen, welche die Induktion der Pluripotenz vorantreiben oder verhindern, wurden in dieser Arbeit genomweit Transkriptionsprofile humaner Fibroblasten nach Überexpression der Reprogrammierungsfaktoren, OCT4, SOX2, KLF4 und c-MYC, einzeln und in Kombination untersucht. Die Analyse des Transkriptoms in der frühesten Phase der Reprogrammierung ermöglichte die Identifizierung von Signalwegen, die primär für die Reprogrammierung von somatischen Zellen verantwortlich sind. Diese Signalwege wurden in ein mathematisches Modell integriert, das zur funktionalen Analyse der identifizierten Komponenten und Signalwege genutzt wurde. Dies ermöglichte die Aufstellung von Hypothesen über die der Reprogrammierung zugrundeliegenden molekularen Mechanismen. Die aufgestellten Hypothesen wurden anhand der beiden wichtigsten involvierten Proteine, p53 und TGFβ-Rezeptoren, während des Prozesses der Reprogrammierung überprüft. Dabei wurden p53 genetisch und der TGFβ-Rezeptor pharmakologisch untersucht.

Die in dieser Arbeit erzielten Ergebnisse zeigen, dass die beginnende Reprogrammierung somatischer Zellen einen an den embryonalen Zellzyklus angelehnten Mechanismus verfolgt. Der synergistische Effekt der Reprogrammierungsfaktoren reguliert den TGFβ- und den p53-Signalweg und mindert damit die Anti-Wachstums-Reaktionen, die Hürden für die Reprogrammierung wären. Die Unterdrückung dieser Signalwege schaltet die G_1-S Phase Kontrollpunkte aus und ermöglicht so einen den embryonalen Stammzellen ähnlichen Zellzyklus. Schnellere G1-S Übergänge reduzieren die Aktivität des Replikations-Reparaturmechanismus und bevorzugen die homologe Rekombination, welche damit zum dominierenden Reparaturmechanismus in der Zelle wird.

3. Introduction

3.1. From embryonic stem cells to induced pluripotency

In mammalian development, differentiation has been considered to be unidirectional. During cellular differentiation, cells become increasingly specialized and further restricted in their developmental potential. Trans-differentiation and dedifferentiation are considered rare events *in vivo*. Nuclear-transfer experiments have shown that nuclei of somatic cells can be reprogrammed to an embryonic state though. Thus, nuclei retain a fully developmental potential that can be induced under certain circumstances (Hochedlinger and Jaenisch 2002).

In 1996, fifty years after the first somatic cell-nuclear transfer (SCNT) experiments Wilmut and colleagues reported the cloning of “Dolly” the sheep, the first mammal to be cloned from an adult cell (Campbell, McWhir et al. 1996). A year later, Thomson and colleagues reported the isolation and stable maintenance of human embryonic stem cells (hESCs) (Thomson, Itskovitz-Eldor et al. 1998). This immediately led to speculations that one could use a combination of both to derive hESCs from patients and apply them for therapeutic purposes (“therapeutic cloning”).

Cell fusion experiments between pluripotent cell lines and somatic cells yielded pluripotent tetraploid hybrid cells, suggesting that embryonic stem cells (ESCs) harbour factors with reprogramming activity (Miller and Ruddle 1976; Tada, Tada et al. 1997; Cowan, Atienza et al. 2005). After intense research on the genetic network responsible for pluripotency in ESCs, Yamanaka’s and Thomson’s team (Takahashi and Yamanaka 2006; Yu, Vodyanik et al. 2007) achieved to induce pluripotency in somatic cells through ectopic expression of a defined set of transcription factors present in hESCs.

The derivation of induced pluripotent stem (iPS) cells is of such great importance because of the ease and reproducibility of generating them. For the first time, it is a feasible way to generate sufficient numbers of patient-specific pluripotent stem cells. Such cells could be used for regenerative and therapeutic purposes, as demonstrated in mouse models of, for example, sickle cell anemia and Parkinson´s disease, respectively (Hanna, Wernig et al. 2007; Wernig, Zhao et al. 2008). In addition, the iPS-cell technology facilitates the generation of disease models that can be used in high-throughput screening and mechanistic studies (Dimos, Rodolfa et al. 2008; Park, Arora et al. 2008; Rubin 2008).

3.2. Regulatory network in embryonic stem cells

Only after enough insight into the regulatory network controlling the pluripotent state was gained, researchers could successfully undertake the reprogramming of somatic cells. Prior to such endeavour, where cell type memory shall be erased and a new developmental program established, the number and identity of the master regulators, as well as their mode of transcription and epigenetic regulation had to be known. All the knowledge acquired on ESCs was therefore key for successful reprogramming experiments.

Human ESCs are derived from the inner cell mass (ICM) of the developing blastocyst. ESCs are self-renewing pluripotent cells. These cells have the capacity to produce limitless daughter cells (self-renewal) and generate all kind of cells of the body (pluripotency). The molecular mechanisms underlying this developmental cell stage are considered fundamental to understand development.

ESCs have been subjected to systems-level and molecular approaches that allowed the identification of a small number of “core” transcription factors that can propagate and sustain an embryonic transcriptional program. Transcription factor complexes bound to specific DNA target sequences are regulators of gene expression programs. Their binding on the promoter of target genes activates or represses them, or alternatively maintains them in a “stand by” status that is resolved upon differentiation. Transcription factors can activate gene expression stimulating the activity of RNA polymerase II, recruiting transcriptional machinery to the regulatory sequences or engaging chromatin-modifying enzymes to alter the local DNA-histone structure, making it more accessible to the transcriptional machinery (Li, Carey et al. 2007; Fuda, Ardehali et al. 2009).

The identification of the master regulators of pluripotency started with the recognition of the unique expression pattern of Oct4 (Octamer-binding transcription factor 4) and Nanog, being almost exclusively expressed at high levels in ESCs during mice development (Chambers, Colby et al. 2003; Chambers and Smith 2004). Genetic experiments showed their essential role in maintaining a stable pluripotent state (Nichols, Zevnik et al. 1998; Niwa, Miyazaki et al. 2000; Chambers, Colby et al. 2003; Mitsui, Tokuzawa et al. 2003; Chambers and Smith 2004). OCT4 and NANOG were found to co-occupy many promoters together with SOX2 (SRY (sex-determining region Y)-box 2) in chromatin immunoprecipitation (ChIP) experiments performed in hESCs (Boyer, Lee et al. 2005). In fact a substantial fraction of the actively transcribed protein-coding genes and microRNAs (miRNAs) in mouse and human ESCs (mESCs and hESCs) depends on

the cooperative binding of Oct4, Sox2 and Nanog (Chen, Xu et al. 2008; Marson, Levine et al. 2008; Boyer, Lee et al. 2005).

Analysis of the promoters bound by the core factors showed that they activate expression of genes necessary to maintain ESCs state, while contributing to repress genes encoding lineage-specific transcription factors that could induce differentiation. Importantly, OCT4, SOX2, and NANOG activate their own transcription in a feed-forward loop in hESCs (Boyer, Lee et al. 2005). This regulatory motif generates the possibility to induce and maintain the pluripotent regulatory network (Bolouri and Davidson 2010), since once expressed they preserve their own transcription and that of their targets active.

The transcriptional regulatory network govern by *Oct4*, *Sox2* and *Nanog* has been linked to genes involved in transcriptional regulation, chromatin modifications and post-transcriptional regulation through non-coding RNAs and microRNAs (Ng and Surani 2011). Other genes, highly expressed in mouse ESC, act co-operatively with these factors to maintain this pluripotency-associated network as *Tcf3* (Cole, Johnstone et al. 2008), Smads (Ying, Nichols et al. 2003; Beattie, Lopez et al. 2005; James, Levine et al. 2005) or *Klf4* (Wu, Chen et al. 2006; Zhang, Tam et al. 2006; Orkin, Wang et al. 2008).

Another set of genes, including *c-Myc*, *Zfx*, *Rex1* and *Ronin*, are considered to complement this pluripotency-associated network in mESC controlling metabolic processes in the cell (Kim, Chu et al. 2008; Dejosez, Levine et al. 2010). As *c-Myc* (Cartwright, McLean et al. 2005), some other genes maintaining the pluripotency-related network are also found up-regulated in tumors such as *Stat3* (Matsuda, Nakamura et al. 1999), *E-Ras* (Takahashi, Bronson et al. 2003), *Klf4* (Li, McClintick et al. 2005) and *β-catenin* (Kielman, Rindapaa et al. 2002).

3.2.1. Core transcriptional machinery

The most important transcriptional regulatory control in mouse and human ESCs appears to come from a small number of transcription factors, namely Oct4, Sox2, and Nanog that act together with other transcription factors, as Klf4 or c-Myc, to maintain the pluripotent state (Greber and Schöler 2008). Moreover, during the last years the role of microRNAs in the control of the pluripotent state has been acquiring more importance and they are now considered as fine-tuners of the expression of key genes for ESC identity (Marson, Levine et al. 2008). Below are summarized the most important features of the factors that have been relevant in the initial experiments of reprogramming.

3.2.1.1. *Oct4*

Octamer-binding transcription factor 4, also known as Oct3, Oct4 and Pou5f1, is expressed in unfertilized oocytes, ESCs and primordial germ cells (Schöler, Hatzopoulos et al. 1989). Its expression is essential for the development of the inner cell mass *in vivo*, the derivation of ESCs and the maintenance of the pluripotent state (Nichols, Zevnik et al. 1998). *Oct4* expression is tightly controlled in mouse ESCs. A less than 2-fold increase in expression causes differentiation into primitive endoderm and mesoderm, whereas repression of *Oct4* induces loss of pluripotency and differentiation into trophectoderm (Niwa, Miyazaki et al. 2000).

Oct4 activity in the germ line is required for the viability of mouse germ cells (Kehler, Tolkunova et al. 2004). Abnormal activity of this transcription factor is a well known oncogenic driver in tumors of germ cell origin (Gidekel, Pizov et al. 2003; Looijenga, Stoop et al. 2003). Besides this, a rapid expansion of progenitor cells and an increased metastatic potential in intestine or skin cells after ectopic activation of *Oct4* has been reported, indicating that it can also act as a powerful oncogene in mouse somatic cells (Hochedlinger, Yamada et al. 2005).

3.2.1.2. *Sox2*

SRY (sex-determining region Y)-box 2, known as Sox2, is a transcription factor involved in the self-renewal of mouse and human ESCs. It has an important role in maintaining mouse and human ESCs pluripotency and heterodimerizes in a complex with Oct4 (Yuan, Corbi et al. 1995). In hESCs, SOX2 shares many target genes with OCT4 and NANOG (Boyer, Lee et al. 2005). In addition to hESCs, *SOX2* is also expressed in the extraembryonic ectoderm, trophoblast stem (TS) cells and the developing central nervous system (neural stem cells) (Avilion, Nicolis et al. 2003). In these cell lineages, *SOX2* expression is restricted to cells with stem-cell characteristics supporting their self-renewal capability and it is no longer expressed in cells with less developmental potential (Avilion, Nicolis et al. 2003). Interestingly, forced expression of *Oct4* can compensate for loss of *Sox2* in mouse ESCs (Masui, Nakatake et al. 2007). In direct reprogramming of mouse embryonic fibroblasts has been shown that Sox2 can be replaced by Sox1, Sox3 and, to a lesser extent, Sox15 or Sox18 (Nakagawa, Koyanagi et al. 2008).

Sox proteins share a highly conserved DNA binding domain referred to as High Mobility Group (HMG). *Sox* genes functions are essential for central neural development. In particular Sox2, which is expressed in mouse and human Neural Stem Cells (NSC), is

required to maintain NSC properties and its depletion causes abnormal differentiation in the eye and the brain of the embryo and the adult (Pevny and Nicolis 2010).

3.2.1.3. *Klf4*

Krüppel-like factor 4 (Klf4) is a transcription factor expressed in a variety of tissues, including the epithelium of the intestine, the kidney and the skin of mice (Segre, Bauer et al. 1999). Depending on the target gene and interaction partner, Klf4 can both activate and repress transcription (Rowland and Bernards 2006). A growing body of evidence suggests that KLF4 can function both as an oncoprotein and tumor suppressor in human cells (Zhao, Hisamuddin et al. 2004). Constitutive expression of *KLF4* suppresses cell proliferation by blocking G_1-S progression in the cell cycle (Zhao, Hisamuddin et al. 2004). In human colorectal carcinoma, KLF4 appears to be down-regulated, being its promoter hypermethylated (Zhao, Hisamuddin et al. 2004). Recently, it has been demonstrated that the forced over-expression of *Klf4* in mESCs inhibits differentiation in erythroid progenitors (Li, McClintick et. al 2005). Its role in the reprogramming process is also not fully understood, but it can be replaced with other Klf family members (Klf2 and Klf5) (Nakagawa, Koyanagi et al. 2008), or the unrelated factors Nanog and Lin28 (Yu, Vodyanik et al. 2007) in the reprogramming of mouse fibroblasts.

3.2.1.4. *c-Myc*

c-MYC is a pleiotropic transcription factor that has been linked in humans to several cellular functions, including cell-cycle regulation, proliferation, growth, differentiation and metabolism (Schmidt 1999). Knock-out c-Myc mice embryos are severely growth retarded and die before embryonic day (E) 10.5, because of severe hematopoietic and placental defects (Trumpp, Refaeli et al. 2001). This factor tends to be highly expressed in the majority of rapidly proliferating cells and is generally low or absent during quiescence (Schmidt 1999). c-MYC is also important for self-renewing of human progenitor cells and during their differentiation, particularly in interactions between stem cells and the local microenvironment (Masui, Nakatake et al. 2007). A large number of binding sites have been reported throughout the human genome and c-MYC appears to be involved in recruiting chromatin-remodelling activities to promoters (Knoepfler, Zhang et al. 2006). The role of *c-Myc* in reprogramming is not clear yet. It is dispensable for the generation of iPS cells in mouse and human (Huangfu, Osafune et al. 2008; Nakagawa, Koyanagi et al. 2008; Wernig, Meissner et al. 2008) but the efficiency of reprogramming decreases dramatically. It can also be replaced by other

family members, such as n-Myc and l-Myc, to reprogram mouse somatic cells to an ESC-like status (Nakagawa, Koyanagi et al. 2008).

3.2.2. Transcriptional regulation

In order to understand why OCT4, SOX2, KLF4, and c-MYC and not other transcription factors are able to change the gene expression program of a cell and its progeny, it is fundamental to understand how these factors control transcription and remodel chromatin structure. The next two sections provide a synthesis of key concepts that explain the driving force towards pluripotency of these reprogramming factors.

Recent studies brought insight into how these transcription factors (TFs) influence expression of their target genes (Kagey, Newman et al. 2010). TFs can regulate transcription at different steps and by different mechanisms (Fuda, Ardehali et al. 2009). The initiation of transcription of a gene requires the assembly of general transcription factors (GTFs) in the core promoter that form a preinitiation complex. This multiprotein complex recruits other auxiliary proteins—co-activators—that enable RNA polymerase II (Pol II) to bind to the transcription start site (TSS) and start the synthesis of a short RNA molecule (approx. 35 bp). Therewith, Pol II escapes the core promoter and pauses. Elongation of the transcript takes place just after phosphorylation of Pol II at the carboxy-terminal domain (CTD) has occurred. This phosphorylation is done by the kinase complex positive transcription elongation factor b (p-TEFb).

Oct4, Sox2, and Nanog have been shown to interact with the mediator, a co-activator, in many active genes in the core regulatory circuitry of mouse ESCs (Kagey, Newman et al. 2010). The mediator associates with the cohesion-loading factor Nipbl and brings together distal regulatory regions and the core promoter. This enhances the binding of Pol II to core promoters and thereby transcription initiation (Kagey, Newman et al. 2010).

On the other side, c-Myc is found in mESCs bound to E-box sequences at the core promoter, where recruits p-TEFb releasing RNA polymerase II from stopping (Rahl, Lin et al. 2010). This type of transcriptional regulation increases the number of complete transcripts. This regulatory mechanism has been postulated to underlie c-Myc up-regulation in cancer cells, and could explain the increase in efficiency observed when c-Myc is co-expressed to induce pluripotency in mouse and human somatic cells (Jaenisch and Young 2008; Rahl, Lin et al. 2010).

A large proportion of the actively transcribed genes in human ESCs are bound and regulated by both the core transcription factors and c-MYC (Boyer, Lee et al. 2005; Knoepfler, Zhang et al. 2006). Simultaneous binding of OCT4, SOX2, and NANOG to the

promoters of selected genes induce transcription by recruiting Pol II, while c-MYC regulates its release from the pause region enabling fully transcription (Young 2011).

3.2.3. Epigenetic regulation

Each cell type is thought to have a characteristic epigenetic pattern that correlates with its differentiation potential (Waddington 1957). During differentiation processes transcription factors and co-factors induce the changes in chromatin structure proper of the specific cell type. Changes defining the identity of a cell will be kept after duplication until new differentiation signals induce novel changes in chromatin structure. Master regulators as the reprogramming factors interact frequently with nucleosome modifying or mobilizing enzymes and induce the rearrangement of local chromatin structure to maintain a developmental state in the cells. Therefore the understanding of the reprogramming process has inherently associated the discovery of the mechanisms that re-set chromatin structures to an embryonic state (Hochendlinger and Plath 2009).

In mouse and human ESCs among the most studied chromatin modifiers are the Polycomb group (PcG) chromatin regulators and SetDB1, which have been implicated in repression of these lineage-specific regulatory genes (Azuara, Perry et al. 2006; Bernstein, Mikkelsen et al. 2006; Boyer, Plath et al. 2006; Bilodeau, Kagey et al. 2009; Yeap, Hayashi et al. 2009). Oct4 can bind sumoylated SetDB1, which can trimethylate histone 3 lysine 9 (H3K9) and propagate a transcriptionally repressive mark on promoters of many of these genes (Bilodeau, Kagey et al. 2009; Yeap, Hayashi et al. 2009). PcG complexes can associate with histone H3K9me3 and further contribute to repression (Azuara, Perry et al. 2006; Bernstein, Mikkelsen et al. 2006; Boyer, Plath et al. 2006).

In addition PcG protein complexes catalyze the trimethylation of histone 3 lysine 27 (H3K27me3). They participate in the silencing of genes encoding key regulators of development that remain in ESCs in a state that is “poised” for activation during differentiation (Bernstein, Mikkelsen et al. 2006; Bracken, Dietrich et al. 2006; Endoh, Endo et al. 2008). PcG proteins are thought to inhibit transcription by restraining poised RNA polymerase molecules (Stock, Giadrossi et al. 2007; Zhou, Zhu et al. 2008).

Multiple repressive mechanisms are involved in the silencing of lineage-specific regulators. Nevertheless, ESCs have significantly more promoters than differentiated cell types with a repressive mark, histone 3 lysine 9 trimethylation (H3K9me3) and an active mark like histone 3 lysine 4 trimethylation (H3K4me3) at the promoters of developmental regulators (Bernstein, Mikkelsen et al. 2006). The trimethylation of histone 3 lysine 4 at the promoters of genes is catalyzed by the TrxG (Tritorax Group) proteins and

antagonizes partially the functions of PcG proteins. This combination of histone modifications has been called the 'bivalent mark', because upon differentiation it resolves to an active (H3K4me3) or an inactive (H3K9me3) mark depending on the cell type they commit to be. It is thought to maintain developmental regulators ready to be expressed in response to differentiation cues (Young 2011).

DNA methylation is essential for mammalian development (Jones and Takai 2001). Complete silencing of Oct4 and Nanog genes is required for proper differentiation of mouse and human ESCs and this is attained, at least partially, by promoter methylation (Feldman, Gerson et al. 2006). Mouse and human ESCs express five DNA methyl-transferases (Dnmt1, 2, 3a, 3b and 3l) and 60-80% of all CpG dinucleotides are methylated (Meissner 2010). While Dnmt-deficient mouse ESCs can be maintained in culture, they are markedly deficient in differentiation (Jackson, Krassowska et al. 2004).

3.3. Induced pluripotency with defined factors

The reprogramming of the somatic cells by nuclear transfer into an oocyte or by fusion with ES cells implied that the unfertilized eggs and ES cells contain factors that can confer pluripotency to somatic cells (Hochedlinger and Jaenisch 2003). Prior experiments had shown that the over-expression or deletion of transcription factors can induce cell fate changes. For instance it was known that the over-expression of the transcription factor Myod in mouse fibroblasts converted them into myogenic cells (Davis, Weintraub et al. 1987), or that the deletion of *Pax5* from mouse B cells induced their dedifferentiation into hematopoietic progenitors (Nutt, Heavey et al. 1999).

Thus, while research on ESCs was uncovering the fundamental transcriptional regulation of pluripotency, different research groups started to over-express in somatic cells combinations of predominantly ES cell-specific proteins. In 2006 Kazutoshi Takahashi and Shynia Yamanaka achieved to induce pluripotency in adult mouse fibroblasts by ectopic expression of Oct4, Sox2, Klf4, and c-Myc after an initial screening of 24 candidates (Takahashi and Yamanaka 2006). The same combination of factors could also reprogram human cells (Takahashi, Tanabe et al. 2007). A second team identified a partially overlapping combination of factors—Oct4, Sox2, Nanog, and Lin28 (Yu, Vodyanik et al. 2007).

The reprogrammed cells were found to be morphologically similar to mouse and human ESCs forming compact colonies. Comparison of global transcriptional data showed high resemblance to stem cells. Epigenetic landmarks like the unmethylation of Oct4 and Nanog promoters recall those observed in mESCs and hESCs. The induced pluripotency was tested *in vivo* and *in vitro*. Following injection into blastocysts, iPS cells

contributed to mouse embryonic development. Cells were also able to differentiate into various cell types, such as neurons, cardiomyocytes, pancreatic and hepatic cells. These results show that the iPS cells are morphologically, molecularly and functionally similar to ESCs (Takahashi and Yamanaka 2006; Takahashi, Tanabe et al. 2007).

3.3.1. Technical considerations in the induction of pluripotency

Researchers tempting to reprogram somatic cells with a set of defined factors had to take in consideration some issues like donor cell type or DNA-delivery methods when designing their initial experiment. Choices made in those issues bent initially the results towards success or failure. Previous attempts to reprogram with SCNT or cell fusion as well as the experiments meant to transform somatic cells were used as reference. After the initial experiments cell-type and DNA-delivery method were quickly extended.

On behalf of the cell type, fibroblasts were a fine cell-type candidate as they had been successfully reprogrammed by nuclear transfers in mouse (Hochedlinger and Jaenisch 2003) and by cell fusion in both mouse and human (Tada, Takahama et al. 2001; Cowan, Atienza et al. 2005). They were also used as “feeder cells” in ES routine culture for optimal cell growth, making them compatible with the ES cell culture conditions. Another important consideration was the reporter system integrated in the genome of the donor cells. A stem cell-specific promoter driving antibiotic resistance was used to identify faithful reprogramming (Takahashi and Yamanaka 2006).

One year after the reprogramming of human fibroblasts cells (Takahashi, Tanabe et al. 2007) many other cell types were successfully reprogrammed, among them: stomach cells, liver cells, neural progenitor cell, lymphocytes, B-cells, keratinocytes, human cord blood cells, human amniotic cells, human peripheral blood cells, human platelets, human astrocytes, and cells from human adipose tissues (Aasen, Raya et al. 2008; Aoi, Yae et al. 2008; Hanna, Markoulaki et al. 2008; Kim, Zaehres et al. 2008; Stadtfeld, Brennand et al. 2008; Utikal, Maherali et al. 2009).

The choice of virus-mediated gene delivery method was surely an advantage in the initial attempts to generate iPS cells. The protein identity, combination, amount or expression time necessary for the process was at that point unknown. Safe and efficient methods had already been developed and long-term screening of dozens of genes at the same time could be performed.

3.3.1.1. Viral delivery system

Takahashi and Yamanaka (Takahashi, Tanabe et al. 2007) used a Moloney Murine Leukemia Virus (MMLV)-based system, a retrovirus, to over-express their candidates, while Thomson's team performed their experiments with lentiviruses (Yu, Vodyanik et al. 2007). Lentiviruses infect both dividing and non-dividing cells and can be used with the Vesicular Stomatitis Virus glycoprotein (VSVg), which is highly effective transducing a wide variety of cells.

Retroviruses are very efficient methods to introduce genetic material into the cells. These systems cause the integration of foreign sequences into the host genome (Pfeifer, Lim et al. 2010). Even though the integrated sequences are efficiently silenced in mouse and human ESCs (Cherry, Biniszkiewicz et al. 2000) they can by themselves reactivate upon differentiation. If this occurs, exogenous Klf4 and c-Myc may give rise to tumors in cells derived from iPS cells. Therefore, integrative methods are not desirable for cell therapies. The use of different vectors to deliver each transcription factor (Oct4, Sox2, Klf4, c-Myc) results in a high number of genomic integrations ranging from 6-10 integrations in most iPS cell lines (Maherali and Hochedlinger 2008).

DoxyCycline-inducible lentiviruses allow the temporal control of factor expression and they had been used to generate the so-called secondary iPS (Hockemeyer, Soldner et al. 2008). The establishment of iPS with that sort of vector and subsequent differentiation to fibroblasts made it possible to generate cell lines containing inducible factors. Reactivation of the integrated transgenes in the differentiated cells produced "secondary" iPS cells (Maherali and Hochedlinger 2008). The reprogramming efficiency for the secondary iPS cells was 100-fold higher than that found in the primary iPS cells. Throughout the use of the inducible lentiviruses, researchers showed also that expression of the factors is necessary until the late stage of iPS (Mikkelsen, Hanna et al. 2008; Li, Liang et al. 2010).

3.3.1.2. Non-integrative approaches

Induced pluripotent cells could also be generated with adenoviruses that show extremely low rates of integration (Stadtfeld, Nagaya et al. 2008; Zhou and Freed 2009). Successful reprogramming with this system answered an open question at that time because it was not known if integration events were essential for *in vitro* reprogramming. The success of the non-integrating vector with transient gene expression to generate iPS cells provides an opportunity to potentially develop a non-viral delivery strategy, which is safe, cost-effective and easier to manufacture and manipulate.

Sequences encoded in plasmids transfected by means of lipid or cationic polymers are episomally expressed and exhibit short duration of gene expression. Nevertheless, reprogramming of mouse somatic cells was achieved with multiple transfections of two different plasmids. A polycistronic vector containing Oct4, Sox2 and Klf4 was simultaneously transfected with a plasmid expressing c-Myc (Okita, Nakagawa et al. 2008). Although plasmid incorporation into the host genome was seen using the multiple transfection methods, no integration was detected using the transient transfection protocol (Okita, Nakagawa et al. 2008). A single plasmid containing all the four reprogramming factors has also been used to generate iPS (Belmonte, Gonzalez et al. 2009). In this latter report nucleofection transfection technology, which delivers DNA directly to the nucleus, was used to enhance gene expression.

Alternatively, episomal plasmid vectors are able to replicate themselves autonomously as extrachromosomal elements. These vectors therefore exhibit prolonged expression of the coding sequences in target cells and can be stably maintained in transfected cells via drug selection. In the reprogramming field one of those systems, derived from the Epstein-Barr virus, has been used for reprogramming of human somatic cells (Yu, Hu et al. 2009). In the study, three combinations of oriP/EBNA1, all of which include OCT4, SOX2, NANOG, LIN-28, c-MYC, KLF4, and SV40LT (SV40 large T antigen), successfully generated the first human iPS cells, which have not undergone genomic manipulation. Even though, the reprogramming efficiency of this approach was extremely low (3-6 colonies per 10^6 input cells).

Other non-integrative approaches have shown comparable induction efficiencies to the retroviral methods. For instance, the piggyback (PB) transposon is capable of excising itself without leaving any exogenous DNA in the cell genome (Elick, Lobo et al. 1997). With the expression of transposase in the programmed cell lines, precise excision of the integrated sequences can be achieved (Kaji, Norrby et al. 2009; Woltjen, Michael et al. 2009). Furthermore, iPS cells have been generated from fibroblasts using a PB method to deliver a polycistronic construct, where the reprogramming factors were linked with a 2A peptide linker and cloned between PB 5' and 3' terminal repeats (Woltjen, Michael et al. 2009).

Purified mouse recombinant reprogramming factor proteins (Oct4, Sox2, Klf4, and c-Myc) have been used to generate protein-induced pluripotent stem cells (piPS cells). The different factors were fused to poly-arginine peptide tags that facilitate passage of the proteins through the plasma membrane upon addition to the culture media. After four rounds of protein and succeeding culture for 30-35 days in the presence of histone deacytelase inhibitor and valproic acid (VPA) a handful of iPS cells

colonies from the initial 5 x 10^4 mouse embryonic fibroblasts (MEF) could be isolated and further propagated (Kim, Kim et al. 2009).

3.3.2. Pathways involved in reprogramming

After the initial experiments some of the signalling pathways involved in the reprogramming process were identified and targeted genetically and/or chemically and probed to increase and accelerate reprogramming (Banito, Rashid et al. 2009; Hong, Takahashi et al. 2009; Kawamura, Suzuki et al. 2009; Li, Collado et al. 2009; Marion, Strati et al. 2009; Utikal, Polo et al. 2009). Because cells that will be successfully reprogrammed cannot be tagged and monitored and the method is so inefficient, a great deal of the molecular events occurring after the over-expression of the factors is yet unknown. Nevertheless, the importance of the p53 signalling pathway or certain clusters of miRNAs has already been highlighted. Furthermore, certain processes such as the cell cycle or the acquisition of epithelial characteristics are expected to happen, since they are main differences between fibroblasts and ESCs, and have been shown to determine the success of the reprogramming process.

Several reports demonstrated that the down-regulation of tumor suppressor components such as *p53*, *p21*, *p16*, and *p19* in mouse fibroblasts enhances the efficiency and kinetics by which iPS cells are generated (Banito, Rashid et al. 2009; Hong, Takahashi et al. 2009; Kawamura, Suzuki et al. 2009; Li, Collado et al. 2009; Marion, Strati et al. 2009; Utikal, Polo et al. 2009). Consistent with this idea, *INK4/ARF* locus is up-regulated with age and reprogramming is less efficient with older cells (Li, Collado et al. 2009). So far, the loss of these proteins has been linked to higher proliferation rates or induction of immortalization (Utikal, Polo et al. 2009).

The first days after oocyte fertilization are characterized by rapid duplication of the cells and progressive decrease of cell size (Murray and Hunt 1993). These features are a consequence of shorter growth phases before mitosis occurs. Human ESCs show the same cell division pattern that is required to keep their high proliferation rate and pluripotency (Becker, Ghule et al. 2006). Initially, reprogramming dependence on cell cycle was analyzed using different cell types, which have specific requirements on cell division components as well as diverse proliferation rates. The proliferation rate of the donor cells correlated with the efficiency of reprogramming and it was therefore believed to be an important aspect in the process.

The presence of LIN-28 in the combination of reprogramming factors suggested early a possible involvement of microRNAs in the process of reprogramming. There is a known stem cell specific miRNA signature (Houbaviy, Murray et al. 2003; Suh, Lee et al.

2004) that is involved in the regulation of the cell cycle in stem cells (Hatfield, Shcherbata et al. 2005; Wang, Medvid et al. 2007; Qi, Yu et al. 2009). Different families of miRNA have been shown to regulate transcriptionally—directly or indirectly—core transcription factors of pluripotency. Furthermore, co-expression of some of them improved the reprogramming efficiency of mouse and human fibroblasts (Subramanyam, Belair et al. 2010). Recently, it has been published the reprogramming of mouse fibroblasts by miRNAs over-expression, without co-expression of transcription factors (Anokye-Danso, Trivedi et al. 2011; Miyoshi, Ishii et al. 2011).

ESCs have some epithelial features that underlie some of their functional and morphological characteristics. ESCs grow as clusters of cells that maintain an extensive range of cell-cell contacts, due to specialized membrane structures, such as tight junctions, adherent junctions, desmosomes and gap junctions. Under normal conditions epithelial cells are motile but their relocation is restricted within the cluster of cells. Instead, fibroblasts of mesenchymal origin, as the one used in reprogramming experiments, grow in culture with few focally contacts to the neighbouring cells and tend to be highly motile (Li, Liang et al. 2010; Samavarchi-Tehrani, Golipour et al. 2010). Therefore, it can be very well expected that the transition from fibroblasts to stem cells recall at least partially a mesenchymal-epithelial transition (Samavarchi-Tehrani, Golipour et al. 2010). This conversion is known from embryonic development and the TGFβ (Transforming Growth Factor ß) signalling pathway is one of the main players on it (Xu, Lamouille et al. 2009).

3.4. Mathematical modelling

In order to understand the reprogramming process the temporal and spatial expression of the genes activated by the reprogramming factors is essential. The reprogramming factors and their targets have to modulate the most elemental processes of the cells as DNA transcription, RNA processing and transport, RNA translation and degradation, posttranslational modification and degradation of proteins. The regulation of these processes is achieved through networks of interactions between DNA, RNA, proteins and small molecules that give rise to genetic regulatory networks. These genetic regulatory networks lie beneath cell behaviour, but involve so many components connected through intermingled positive and negative feedback loops that go beyond intuitive analysis.

To make quantitative predictions about the behaviour of the cells that depend on such an array of interactive macromolecules acting co-ordinately, it is necessary to apply mathematical and computational tools that can be able to show spatial and temporal

organization. The study of genetic regulatory networks has received a major boost from the development of experimental techniques like cDNA miroarrays and oligonucleotide chips, which allow the rapid measurement of gene expression levels in a massively parallel way (Brown and Botstein 1999; Lockhart and Winzeler 2000). Many efforts during the last years have been made to develop computational tools to deduce network structure from high-throughput experimental data and model the dynamic behaviour of the regulatory motifs found (Liang, Fuhrman et al. 1998; Chen, He et al. 1999; D'Haeseleer, Wen et al. 1999; Friedman, Linial et al. 2000).

There is currently a huge research effort laid on the development of modelling approaches able to describe the molecular dynamics of cellular systems. The starting point is the attribution of mathematical equations to the different interactions between macromolecules and the physico-chemical phenomena that undergo in a cellular context. The type of equations to be used depends on the biological questions under consideration (Hasty, McMillen et al. 2001). The simplest modelling approach uses Boolean functions to describe the transcriptional state of a certain gene that can be then active or inactive (ON/OFF). Its status depends on the state of other genes that act as activators of repressors. The collection of genes interacting to each other in this way is called Boolean Network (Chen, He et al. 1999). This kind of model can already describe qualitatively the activation and repression of gene transcription and can recall features of biological systems as global complex behaviour, self-organization, stability or periodicity (Somogyi and Greller 2001).

Models can describe also quantitatively regulatory information such as activation/repression or also simulate more complex temporal behaviours of their components if a set of ordinary differential equations (ODE) describes the regulatory interactions between the components. These differential equations are derived from kinetic rate laws (Chen, He et al. 1999), and make possible their mathematical manipulation. Particularly with ordinary differential equations a very detailed description of a (dynamical) gene regulation system is possible. It has been a useful technique to describe biological systems, especially regulatory loops (Blüthgen, Legewie et al. 2009) or crosstalk between different signalling pathways (Borisov, Aksamitiene et al. 2009).

In the context ESCs kinetic models and ordinary differential equations have been already used to study the regulatory interactions between the core transcription factors, Oct4, Sox2, and Nanog and their transcriptional activation of down-stream targets in mice (Chickarmane, Troein et al. 2006; Chickarmane and Peterson 2008). These studies support that ESCs can maintain its pluripotent state through the autoregulatory loop of the core transcription factors independently of the signalling that induced it in the first instance, yet were able to be down-regulated in response to differentiation cues. In

addition, they explored different transcriptional outcomes on the target genes depending on the binding affinities of the factors to their target promoters (Chickarmane, Troein et al. 2006).

A dynamical model for lineage determination, specifically differentiation to *trophectoderm* and *endoderm* has been also developed (Chickarmane and Peterson 2008). The model can successfully describe lineage commitment and recall partially the transcriptional dynamics observed in stem cells after knock-down or over-expression of Oct3/4 (Niwa, Miyazaki et al. 2000).

3.5. Aims of the project

The reprogramming of somatic cells into pluripotent cells by over-expression of four transcription factors (OCT4, SOX2, KLF4, and c-MYC) brought a lot of attention since its first publication in 2006. Since then, much effort has been placed on the elucidation of the molecular mechanisms underlying the induction of differentiation potential in committed cells, as well as the roadblocks that have to be overcome. The method has been hindered by its low efficiency, though.

In this context, the aim of my Ph. D. project is to address the molecular mechanisms that contribute or hinder the induction of pluripotency and to understand the different roles of the transcription factors in this process. Since the reprogramming process is expected to be a complex process involving many different genes genome-wide information will be required. The starting point can be then the analysis of the transcriptome at the early stages of the reprogramming. The data acquired can then render a concrete genetic regulatory network that brings together a set of transcriptional events that explain phenotypic differences.

Even though less of the existing genetic complexity in the cell participates in the process of reprogramming, the amount of participants and their intricate relationships demand specialized tools to represent them and assess their functional significance. In order to rationalize the systems-level changes of the regulatory network a mathematical model can be assembled based on the transcriptional and biochemical regulatory information available for the relevant participants.

Such a mathematical model can mimick the cellular processes involved in the induction of pluripotency, and here particularly, it can describe some of the roadblocks of the process. The model can greatly facilitate further investigation on the molecular mechanisms that underlie reprogramming enabling working hypothesis to be tested *in silico*. The resulting *in silico* simulations should assist the experimental work meant to probe the molecular mechanisms behind the reprogramming of somatic cells (Figure 1).

The analysis of the molecular mechanisms that underlie reprogramming can yield strategies to improve the methodology and solve key questions regarding cell fate determination and disease biology.

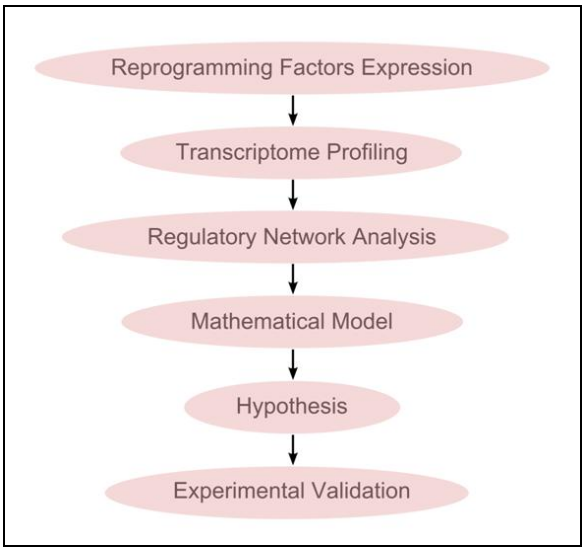

Figure 1. Outline of the project. This scheme shows the milestones of the project according to the described aims.

4. Material and Methods

4.1. Material

4.1.1. Reagents and kits

Table 1. Reagents and kits. Below is a relation of the substances used. On the left column can be found the name of the substance and on the right column the name of the company by which it was distributed.

Chemicals	Company
4',6-diamidino-2-phenylindole (DAPI)	Invitrogen
6x Loading Buffer	Fermentas
Acetic Acid	Merck
Agarose	Invitrogen
Bovine Serum Albumin (BSA)	Sigma
Bromophenol Blue	Sigma
Dimethylsulfoxide (DMSO)	Sigma
Dispase	Invitrogen
Dithiothreitol	Sigma
dNTP	Pharmacia
Dulbecco´s Modified Eagle Medium (DMEM)	Invitrogen
Ethanol	Merck
Ethidium Bromide	Sigma
Ethylenediaminetetraacetate (EDTA)	Sigma
Fetal Bovine Serum (FBS)	Biochrom AG
Fluoromount G	Southern Biotech
Gelatine from bovine skin, Type B	Sigma
GeneRuler 1 kb DNA ladder	Fermentas
Glycerol	Merck
Glycine	Merck
HCl	Merck
Knockout Dulbecco´s Modified Eagle Medium (KO-DMEM)	Invitrogen
Knockout Serum Replacement	Invitrogen
L-glutamine	Invitrogen
Linear Amplification Kit	Ambion
Lipofectamine	Invitrogen

Chemicals	Company
Matrigel	Becton Dickinson
Mitomycin C	Roche
M-MLV Reverse Transcriptase Buffer	Promega
mRNA – H9	Dr. Greber, MPI-MB
mRNA – SW420 and SW680	Dr. Schweiger, MPI-MG
Non-essential Amino Acids	Invitrogen
NucleoSpin Plasmid Purification Kit	Macherey-Nagel
Oligo(dT)15-Primer (0,5 ug/uL)	Promega
Formaldehyde	Sigma
Penicillin-Streptomycin	Invitrogen
Phosphate-Buffered Saline (PBS)	Invitrogen
Protease Inhibitor (Complete PI Cocktail Tablets)	Roche
Puromycin	Roche
RNase-free DNase set	Qiagen
Rneasy Mini Kit	Qiagen
SB431542	Sigma
Sodium Chloride	Merck
Sodium Dodecyl Sulfate	Merck
ß-Mercaptoethanol	Sigma
SYBR Green Master Mix	Fermentas
TEMED	Roth
Tris(hydroxymethyl)aminomethane	Merck
Triton X-100	Merck
Trypsin	Invitrogen
Tween 20	Roth

4.1.2. Buffers and solutions

Table 2. Buffers and solutions. Description of the different buffers and solutions that were used. On the left column is specified the name of the buffer or solution. On the right column are the amounts and components necessary to make them up.

Buffer/solution	Composition
1x Tris Buffered Saline Buffer (TBS)	8 g Sodium Chloride 20 mL 1 M Tris-HCl pH 7.6 ad 1000 mL dH_2O
1x Tris Buffered Saline/Tween Buffer (TBST)	1000 mL 1x TBS 1 mL Tween 20
4% Formaldehyde	12.5 mL of 16% Formaldehyde 37.5 PBS
50X Tris/Acetate/EDTA (TAE) Buffer	242 g Tris Base 57.1 mL Acetic Acid 100 mL of 500 mM EDTA (pH 8.0)
Dispase Solution	5 g Dispase 10 mL Knockout DMEM
Mitomycin C	10 mg Mitomycin C 10 mL PBS
Puromycin	10 mg Puromycin 10 mL dH_2O

4.1.3. Equipment

Table 3. Materials. This table presents the list of material employed. On the left column are registered the name of the different products that are sold by the companies recorded on the right column.

Material and machines	Company
12-well Cell Culture Plates	Techno Plastic Products
6-well Cell Culture Plates	Techno Plastic Products
ABI-PRISM 7900HT Sequence Detection System	Applied Biosystems
Acrodisc Syringe Filters, 0.45 um	Pall Corporation
Agarose Gel Chamber	MPI-MG Werkstatt
Analytical Balance AT 250	Mettler
Cell culture dish, 60 mm, ultra low attachment	Techno Plastic Products
Cell Culture Freezing Box	Nalgene
Cell Culture Incubator Innova CO-170	New Brunswick Scientific
Centrifuge Avanti J-25	Beckman Coulter
Cover Slips, 12/18 mm diameter	Roth
Cryovials, 2 mL	Corning

Material and machines	Company
Filter Units. 0.2 um pore size	Nalgene
Freezer –20 °C	Bosch
Freezer –80 °C	Forma Scientific
Gel Documentation System	Alpha Innotech
Heatblock Compact 5350	Eppendorf
Illumina Human-8 BeadChips	Illumina
Illumina BeadStation 500 platform	Illumina
LSM520 Meta Microscope	Zeiss
Magnetic Stirrer MR3001	Heidolph
Microscope Slide	Roth
Nano Drop ND-1000 UV-Spectrometer	Nano Drop Technologies
Optical 96/384-well Reaction Plate	Applied Biosystems
Optical Adhesive Covers	Applied Biosystems
Parafilm	American National Can Company
PCR-Machine PTC-100 and PTC-225	MJ-Research
Petri Dishes	Greiner
Pipettes	Eppendorf
Polypropylene Tubes	Corning
Precision Balance PM3000	Mettler
Scalpel	Braun
Sterile Glass Beads	MPI-MG Werkstatt
Sterile Needles BD Microlance	Beckson Dickinson
Sterile Pipettes	Sarstedt
T50 Cell Culture Flasks	Techno Plastic Products
T75 Cell Culture Flasks	Techno Plastic Products
Tabletop Centrifuge 5224, 5810R, 5417C	Eppendorf
UV-Transilluminator UVT-28M	Herolab
Vortex-Genie 2 Machine	Scientific Industries
Waterbath	Memmert

4.1.4. Cell lines

Table 4. Cell lines. The different experiments presented in this work were done on the cell lines listed below. On the left column is written the name of cell line. The company that provide the cell line and their catalogue number is detailed in the right column.

Cell Lines	Company/Catalog Number
Human Newborn Foreskin Fibroblasts (HFF-1)	ATCC/SCRC-1041
Human Embryonic Kidney Cells (HEK293T)	ATCC/CRL-1573
Mouse Embryonic Fibroblasts (CF1)	MPG MOLGEN, Mouse Facility

4.1.5. Cell culture media

Table 5. Cell culture media. Description of the media used to grow and freeze the cells. On the left column is specified the purpose of the media. On the right column are the amounts and components necessary to make them up.

Cell Culture Media	Composition
Fibroblast/HEK freezing media	50% of FBS 40% of DMEM 10% of DMSO
Fibroblast/HEK culture media	450 mL of DMEM 50 mL of FBS 5 mL of 200 mM L-glutamine 5 mL of Non-Essential Amino Acids 5 mL of Penicillin-Streptomycin
Induced pluripontent stem cell freezing media	50% of Knockout Serum Replacement 40% of Knockout DMEM 10% of DMSO
Induced pluripotent stem cell media (Unconditioned Media: UM)	400 mL of Knockout DMEM 100 mL of Knockout Serum Replacement 5 mL of 200 mM L-glutamine 5 mL of Penicillin-streptomycin 5 mL of Non-Essential Amino Acids 35 µL of 0.14 M ß-Mercaptoethanol

4.1.6. Bacterial strains

Table 6. Bacterial strains. On the following list are related the different *Escherichia coli* strains that were used to propagate plasmids. On the left column can be found the name of the strain and the company that distribute it. On the right column are detailed genotypic specifications.

Bacteria Strain (Company)	Genotype
XL1-Blue (Stratagene)	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1* *lac* [F´ *proAB lacI*q*Z_M15* Tn*10* (Tetr)]
XL10-Gold (Stratagene)	*Tetr _(mcrA)183 _(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F´ proAB lacIqZ_M15 Tn10 (Tetr) Amy Camr]*
DH-5-α (Invitrogen)	*F- φ80lacZ_M15 _(lacZYA-argF) U169 recA1 endA1 hsdR17 (rk-, mk+) phoA supE44 λ- thi-1 gyrA96 relA1*

4.1.7. Enzymes

Table 7. Enzymes. The name of the enzymes that were used has been listed on the left column. The company that distribute them can be found in the right column.

Enzymes	Company
M-MLV Reverse Transcriptase	Promega
*Not*I	New England Biolabs
*Eco*RI	New England Biolabs

4.1.8. Primary antibodies

Table 8. Primary antibodies. The following table relates the primary antibodies that were employed. The target gene to which the antibodies bind is specified on the most left column and the animal source in which they were produced is specified in the column aside. Under the columns labels "WB" and "IF" can be found the dilution used in Western Blot and immunofluorescence experiments respectively. The company that distribute each antibody and their catalogue number is detailed on the right column.

Target	Source	IF	Company/Catalog Number
Human c-MYC	Rabbit	1:200	Santa Cruz Biotechnology; SC-7645
Human KLF4	Rabbit	1:200	Santa Cruz Biotechnology; SC-20691
Human SOX2	Goat	1:200	Santa Cruz Biotechnology; SC-17320
Human OCT3/4	Goat	1:200	Santa Cruz Biotechnology; SC-8629

4.1.9. Secondary antibodies

Table 9. Secondary antibodies. This table lists the secondary antibodies used in immunofluorescence experiments. The source of the primary antibody to which the antibodies bind is specified on the left column. The second column indicates the kind of tag the secondary antibody is harbouring. The next column on the right shows the animal source in which the antibody was produced. The following column indicates the dilution at which the antibodies were applied. The company that distribute each antibody and their catalogue number is detailed on the right column.

Target	Tag	Source	Dilution	Company/Catalog Number
Anti-Goat	Alexa Fluor 594	Chicken	1:200	Invitrogen; A21468
Anti-Rabbit	Alexa Fluor 546	Goat	1:500	Invitrogen; A11010

4.1.10. Oligonucleotides

Primer3 (Rozen and Skaletsky 2000) was used to design primers for real-time PCR. The DNA sequence of interest was introduced. An amplicon maximal size of 100 base pairs and optimal melting temperature (Tm) of 60 °C were set before the software was let to pick the primers. The primer pair retrieved by the software was checked for specificity against an assembly of the human genome (NCBI36/hg18) available at http://genome.ucsc.edu/cgi-bin/hgPcr?org=Human&db=hg18&hgsid=88954760.

Table 10. Oligonucleotides. In the following table are listed the oligonucleotides used to carry out Real-Time-PCRs. On the left column are listed the target genes. The middle column indicates the direction of the primer and the right column shows the sequence.

Target	Primer	Sequence (5' → 3')
ACTIN	fwd	GTTGTCGACGACGAGCG
	rev	GCACAGAGCCTCGCCTT
AIF	fwd	TGTTGAGTTGGCCAAGACTG
	rev	CGTGCTTGTAGCTCTGCATT
BAD	fwd	CAGTCACCAGCAGGAGCA
	rev	GTAGGAGCTGTGGCGACTC
BAX	fwd	GTGGCAGCTGACATGTTTTC
	rev	TTGCTGGCAAAGTAGAAAAGG
BCL2	fwd	GGTGGAGGAGCTCTTCAGG
	rev	ACGCTCTCCACACACATGAC
BCL2XL	fwd	CATGGCAGCAGTAAAGCAAG
	rev	AGCTGGGATGTCAGGTCACT

Target	Primer	Sequence (5' → 3')
BIRC5	fwd	TCTGCTTCAAGGAGCTGGA
	rev	TCTCCGCAGTTTCCTCAAAT
BRCA1	fwd	GGAACCCCTTACCTGGAATC
	rev	ACTCTGGGGCTCTGTCTTCA
BRCA2	fwd	AGGCTTCAAAAAGCACTCCA
	rev	TTGTGCGAAAGGGTACACAG
CASP3	fwd	TCGCTTCCATGTATGATCTTTG
	rev	CTGCCTCTTCCCCCATTCT
CASP7	fwd	GGATCGCATGGTGACATTTT
	rev	GGTTGAGGATTCAGCAAATGA
CASP9	fwd	CTGCATTTCCCCTCAAACTC
	rev	AGGTTCTCAGACCGGAAACA
CDK1	fwd	CAACTCCATAGGTACCTTCTCCA
	rev	GCGGAATAATAAGCCGGGAT
CDK2	fwd	AATTCATGGATGCCTCTGCT
	rev	TGGAGCAGCTGGAACAGATA
CDK4	fwd	CAGATGGAGCTTGTTCAGGAG
	rev	AGGGGACTTAAACGCCACTT
CDK6	fwd	CAGATGGAGCTTGTTCAGGAG
	rev	AGGGGACTTAAACGCCACTT
CMYC	fwd	ACGTCTCCACACATCAGCAC
	rev	GTCCAACTTGACCCTCTTGG
Cyclin A1	fwd	CTAGAGCAGGGGGACAGAGA
	rev	GTGTGCCGGTGTCTACTTCA
Cyclin A2	fwd	GCACCCCTTAAGGATCTTCC
	rev	CGCAGGCTGTTTACTGTTTG
Cyclin D1	fwd	TGGTGAACAAGCTCAAGTGG
	rev	CTGGCATTTTGGAGAGGAAG

Target	Primer	Sequence (5' → 3')
Cyclin D2	fwd	GCTGGAGGTCTGTGAGGAAC
	rev	CCAAGAAACGGTCCAGGTAA
Cyclin D3	fwd	GACAGGCCTTGGTCAAAAAG
	rev	CGGGTACATGGCAAAGGTAT
Cyclin E1	fwd	CAGATGGAGCTTGTTCAGGAG
	rev	AGGGGACTTAAACGCCACTT
DAXX	fwd	GTGGCCATAGGGGATCAAAT
	rev	TGCGAGGTTCTGAGAATTGC
GAPDH	fwd	AATGAAGGGGTCATTGATGG
	rev	AAGGTGAAGGTCGGAGTCAA
GADD45α	fwd	CAGAAGACCGAAAGGATGGA
	rev	AGTGATCGTGCGCTGACTC
GFP	fwd	CTGAGGGCTACATCCAGGAG
	rev	GTATCGCCCTCGAACTTCAC
JNK	fwd	GCCAGACCGAAGTCAAGAAT
	rev	CAAGCACCTTCATTCTGCTG
KLF4	fwd	ACCCACACAGGTGAGAAACC
	rev	ACGGTAGTGCCTGGTCAGTT
MDM2	fwd	GAGCAGGCAAATGTGCAATA
	rev	TTGTTCCGAAGCTGGAATCT
MEKK1	fwd	TTTGGATGGTCAACAGGACA
	rev	ACTGTGCACTCAGGGGAACT
NBN	fwd	TTTGACTGGCGTTGAGTACG
	rev	ATGATTTCGGCTGATCGACT
OCT4	fwd	GTGGAGGAAGCTGACAACAA
	rev	ATTCTCCAGGTTGCCTCTCA
p16	fwd	GACATCCCCGATTGAAAGAA
	rev	TTTACGGTAGTGGGGGAAGG

Target	Primer	Sequence (5' → 3')
p19	fwd	GCTGCAGGTCATGATGTTTG
	rev	CTGCCAGATGGATTGGAAGT
p21	fwd	GCGACTGTGATGCGCTAAT
	rev	GTGTCTCGGTGACAAAGTCG
p53	fwd	TGGCTCTGACTGTACCACCA
	rev	CCAGTGTGATGATGGTGAGG
RAD1	fwd	GCCAGGGACTTTAACTGCAC
	rev	TTGCAGACTGTCACCACTCC
TAK1	fwd	ACAGCAGAGTCGGTGGAGAT
	rev	CTTTGGCGCAAATCCTGAG
TGFβR2	fwd	TAACCTGCTGCCTGTGTGAC
	rev	CCCACTGTTAGCCAGGTCAT
SMAD3	fwd	CGCAGGTTCTCCAAACCTAT
	rev	GGCTCGCAGTAGGTAACTGG
SMAD4	fwd	ATACAGCACCCCAGCTCTGT
	rev	GTGGAAGCCACAGGAATGTT
SMURF2	fwd	AGCGAGACCTGGTTCAGAAA
	rev	TCTCTTCCCTGGAAACCTCA
SOX2	fwd	GTATCAGGAGTTGTCAAGGCAGAG
	rev	TCCTAGTCTTAAAGAGGCAGCAAAC

4.1.11. Plasmids

4.1.11.1. pMXs-hOCT3/4, pMXs-hSOX2, pMXs-hKLF4, pMXs-hc-MYC

The reprogramming factors were cloned independently in a 4600 bp long pMXs backbone vector. This vector contains long tandem repeats (LTR) derived from Moloney-Murine Leukemia Virus and is suited for expression in mammalian cells. These vectors are used in combination with the pUMVC3-gag-pol vector and the pCMV-VSV-G vector in order to produce functional virions. The plasmid contains a gene conferring resistance to Ampicillin that allows its purification from *E. coli* bacteria at high copy number. A more detailed description of the procedure to manufacture the vector backbone is explained elsewhere (Kitamura, Koshino et al. 2003). This vector was used in the pioneer work of (Takahashi, Tanabe et al. 2007).

The human sequence for OCT3/4, 1100 bp long, has been cloned in the pMXs vector (Figure 2) using *EcoR*I restriction sites. This vector was purchased at Addgene under the catalogue number 17217 and is named pMXs-hOCT3/4. The human sequence for SOX2 is 960 bp long and was cloned in using *EcoR*I and *Not*I restriction sites. The vector containing human SOX2 was purchased at Addgene under the catalogue number 17218 and is named pMXs-hSOX2. Human KLF4 has been cloned in a pMXs vector using attB sites. This latter insert is 1420 bp long. This vector was purchased at Addgene under the catalogue number 17219 and is named pMXs-hKLF4. The pMXs-hc-MYC contains an insert 1320 bp long encoding human c-MYC. The pMXs-hc-MYC vector was purchased as the others at Addgene under the catalogue number 17220.

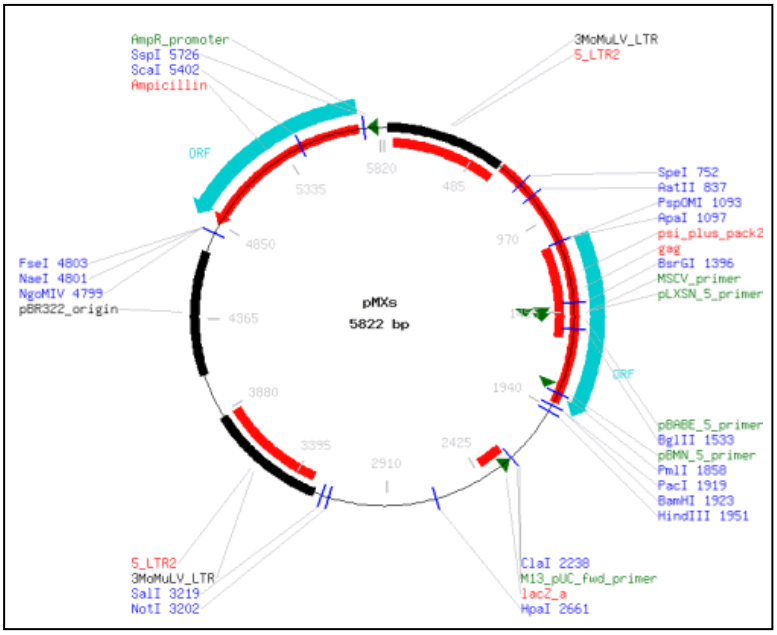

Figure 2. pMXs backbone vector. Map as it is found in Addgene database. The reprogramming factors, OCT4, SOX2, KLF4, and c-MYC have been cloned after the viral packaging signal and the gag protein ("psi_plus_pack2" and "gag" in the map). Long tandem repeats are labelled as "5_LTR2" in red letters. The Ampicillin resistance gene and its promoter are located between bp 4850 and 5820. Black color is used to depict structural motifs. Light blue delimits open reading frames (ORFs). Dark blue writing indicates restrictions points. Green color specifies motifs such as promoters or primer binding sites. Red color is used for labelling selected features.

4.1.11.2. pLENTI-GFP

This construct was kindly gifted by Dr. Nils Rademacher. The pLenti6/VB5-D-TOPO vector (Figure 3) was used as backbone, in which GFP was cloned. This vector is suited for the lentivirus-mediated over-expression of proteins in mammalian cells. The vector is 6963 bp long and contains a gene conferring resistance to Ampicillin and a Blasticidin resistance gene.

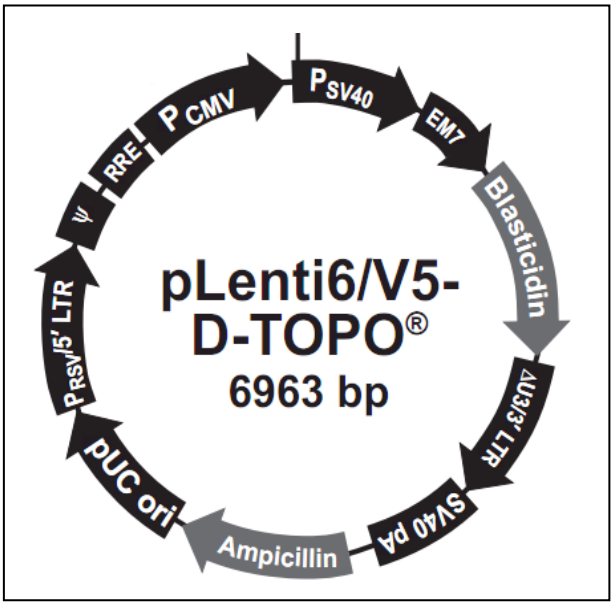

Figure 3. pLenti6/VB5-D-TOPO backbone vector. This map can be found at Invitrogen website. A CMV promoter labelled as "P_{CMV}" precedes the multiple cloning site, where GFP was cloned. Antibiotic resistance genes are highlighted in gray. The long tandem repeats have been labelled as "5' LTR" and "3' LTR".

4.1.11.3. pUMVC3-gag-pol

The vector backbone is 9251 bp (Figure 4) and contains the *gag* and *pol* genes from Moloney-Murine Leukemia Virus and is suited for expression in mammalian cells. This vector is required together with a pMXs, pMKO or pLKO vector and the p-CMV-VSV-G vector for virion packaging. A more detailed description of the procedure to manufacture the vector backbone is explained elsewhere (Stewart, Dykxhoorn et al. 2003). The plasmid contains a Kanamycin-resistance gene that allows its purification from *E. coli* bacteria at low copy number. This vector was purchased at Addgene under the catalogue number 8449.

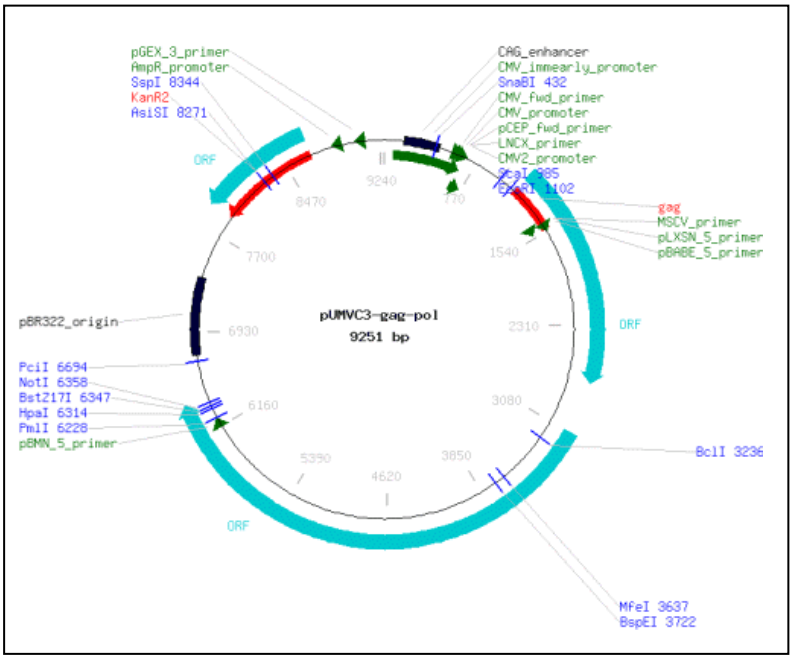

Figure 4. pUMVC3-gag-pol vector. Map as it is found in Addgene database. The Kanamycin resistance gene is labelled as "KanR2" in red. The sequence encoding *gag* and *pol* are located between 1117-2733 bp and 3061-6420 bp respectively. Black color is used to depict structural motifs. Light blue delimits open reading frames (ORFs). Dark blue writing indicates restrictions points. Green color specifies motifs such as promoters or primer binding sites. Red color is used for labelling selected features.

4.1.11.4. pCMV-VSV-G

The vector backbone is 6363 bp and contains the envelope glycoprotein from Vesicular Stomatitis Virus and is suited for expression in mammalian cells. This vector is required together with a pMXs, pMKO or pLKO vector and the pUMVC3-gag-pol vector for virion packaging. A more detailed description of the procedure to manufacture the vector backbone is explained elsewhere (Stewart, Dykxhoorn et al. 2003). The plasmid contains a Kanamycin-resistance gene that allows its purification from *E. coli* bacteria at low copy number. This vector was purchased at Addgene under the catalogue number 8454.

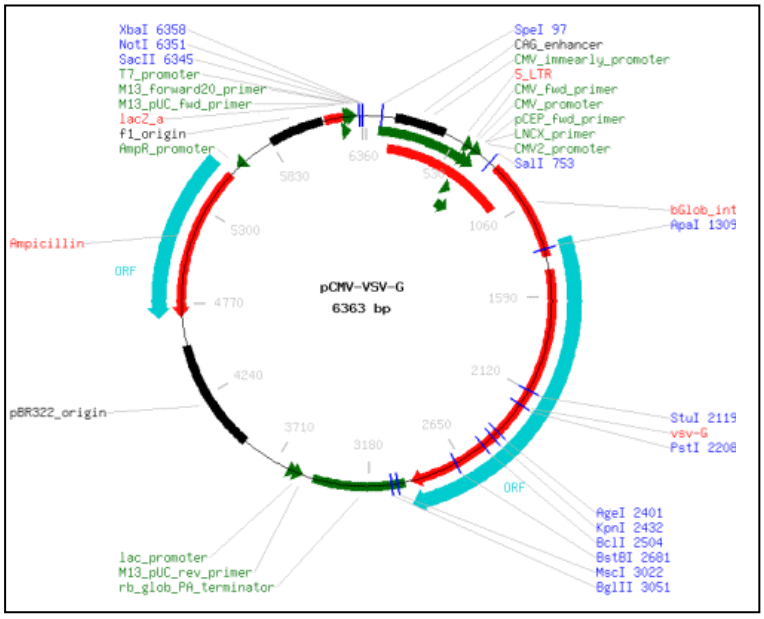

Figure 5. pCMV-VSV-G vector. Map as it is found in Addgene database. The ampicillin resistance gene and the virion envelope are depicted as “ORF” in light blue and labelled as “Ampicillin” and “vsv-G” respectively. Black color is used to depict structural motifs. Dark blue writing indicates restrictions points. Green color specifies motifs such as promoters or primer binding sites. Red color is used for labelling selected features.

4.1.11.5. pMKO.1 puro p53-shRNA

The vector backbone is pMKO.1 puro shRNA and is 6700 bp long (Figure 6). A 58 bp long short hairpin RNA sequence targeting p53 was cloned using *AgeI* and *EcoRI* restriction sites. This vector has a retroviral origin and is suited for transducing mammalian cells. After transduction positive clones can be selected by addition of puromycin. The plasmid contains an Ampicillin-resistance gene that allows its purification from *E. coli* bacteria at high copy number. This vector was purchased at Addgene under the catalogue number 10672.

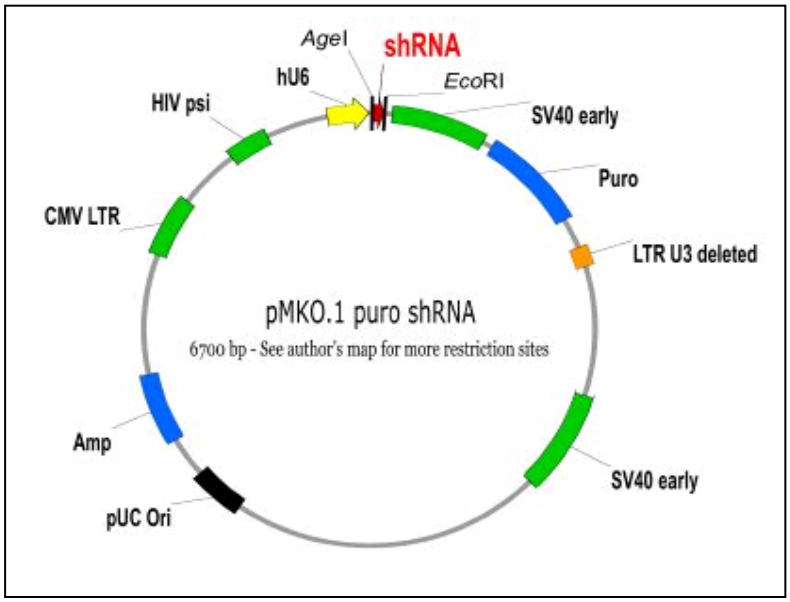

Figure 6. pMKO.1 puro shRNA vector. Map as it is found in Addgene database. The cloning site, "shRNA", is flanked by *Age*I and *Eco*RI restriction sites. The transcription of the shRNA is driven by the human RNA polymerase III promoter ("hU6") and is illustrated as yellow arrow. Antibiotic resistance genes are highlighted in blue boxes. The viral packaging signal from human immodeficiency virus is labelled as "HIV psi". The pUC origin of replication is depicted in a black box. Long tandem repeats are labelled as "LTR".

4.1.11.6. pLKO.1 puro scramble-shRNA

The vector backbone is pLKO.1 puro shRNA and is 7032 bp long (Figure 7). A 60 bp long short hairpin RNA scramble sequence was cloned using *Age*I and *EcoR*I restriction sites. This vector has a retroviral origin and is suited for transducing mammalian cells. After transduction positive clones can be selected by addition of puromycin. The plasmid contains an Ampicillin-resistance gene that allows its purification from *E. coli* bacteria at high copy number. This vector was purchased at Addgene under the catalogue number 1864.

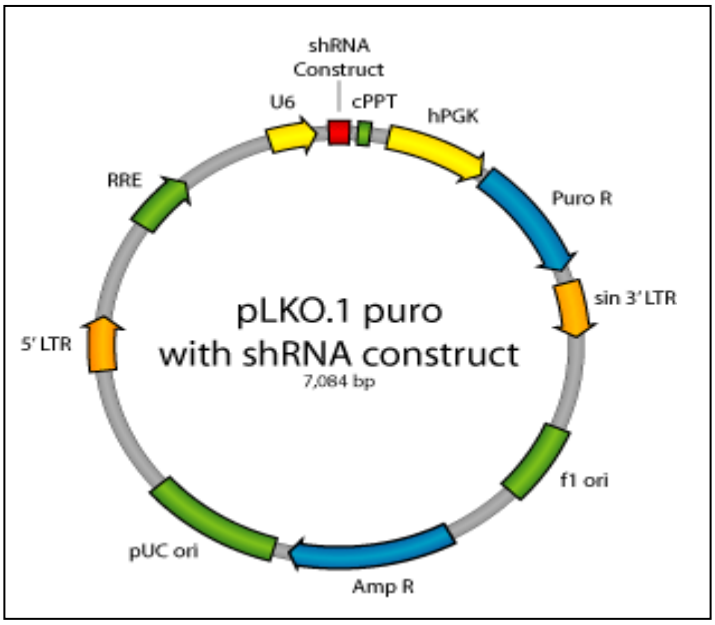

Figure 7. pLKO.1 puro shRNA vector. Map as it is found in Addgene database. In blue boxes are depicted the ampicillin ("Amp") and the puromycin ("Puro") resistance gene. The shRNA construct is cloned between the U6 promoter (yellow arrow) and the central polypurine tract (cPPT). "RRE" contains the *Rev* response element and it is illustrated as green arrow. The long tandem repeats have been represented as dark yellow arrows.

4.2. Methods

4.2.1. Thawing and maintaining fibroblasts

A vial of frozen fibroblasts was removed from the liquid nitrogen tank and thawed in a warm water bath. When just a small crystal of ice remained then vial was sterilized with 70% (v/v) ethanol. Afterwards the cell suspension was pipetted dropwise into 10 mL of warm fibroblast culture media to reduce osmotic shock. To remove the DMSO present in fibroblast freezing media (see section Table 5), cells were pelleted by centrifugation at 300 g for 5 min. Then the supernatant was removed, cells were resuspended in fibroblast media (see section Table 5) and plated in filtered flasks. Once cells have reached 90% confluency, they were washed with PBS and incubated in warmed trypsin solution. After one to three min the same volume of trypsin solution was added to the wells or flasks and the mixture was collected and centrifuged at low speed (300 g) for 5 min. Cells were diluted in a 1:3 or 1:4 ratio and seeded into new wells/flasks.

Fibroblasts, kidney cells and iPS-like cells were routinely incubated at 37 °C, 5% CO_2 and 95% humidity in a New Brunswick Scientific incubator.

4.2.2. Cryopreservation of fibroblasts

Culture media was removed and cells were washed with PBS and trypsinized for 5 min at 37 °C. Trypsin was inactivated by adding 10 mL of fibroblasts culture media (Table 5). Then cells were centrifuged at low speed (300 g for 5 min). The supernatant was then removed and the cells resuspended in fibroblast freezing media (Table 5). Cells were aliquoted in cryovials and placed inside a pre-cooled freezing container in a -80 °C freezer. After 24 h cryovials were transferred to a liquid nitrogen tank for long-term storage.

4.2.3. Preparation of iPS-like culture plates

The plates used for reprogramming experiments or to grow iPS-like cells were previously coated with Matrigel. First Matrigel solution was thawed overnight at 4°C to avoid gel formation. The day after cold Knockout DMEM was added to the 10 mL of Matrigel and mixed. Bottle containing Matrigel and falcons were kept at 4 °C and 1 mL aliquots were pipetted into each tube and stored at -20 °C. Before plating the cells, one aliquot was diluted in 14 mL of cold Knockout DMEM to coat 3 x 6-well plates. Plates were then sealed off with parafilm and kept overnight at 4 °C. The Matrigel solution was removed immediately before plating fibroblasts cells.

Typically 250000 inactivated mouse or human fibroblasts (see section 4.2.4) per well of 6-well plate were seeded and grown till attached (next day morning). Then

fibroblasts media was replaced with unconditioned media (UM) supplemented with 4 ng/ml of FGF2 (fibroblast growth factor 2). After 24 h iPS-like cells were plated.

4.2.4. Inactivation of fibroblasts (Feeders preparation)

Flasks were coated with 0.2% gelatine and incubated at room temperature for at least 2 h. After this time, flasks were washed with PBS and fibroblasts feeders were plated at the desired density. Fresh fibroblasts growth media was then added and flasks were incubated until next day. Cells were then inactivated using 10 µg/ml of mitomycin C that was added to the media and incubated for 2 h at 37°C. Mitomycin C containing media was then removed and cells were washed twice with PBS. Cells were then split using trypsin solution and seeded at desired density onto gelatine-coated plates with warm media (see section 4.2.3).

4.2.5. Conditioned Media preparation

To prepare conditioned media (CM) it is necessary to plate the inactivated cells at high density, 75000 cells/cm^2. After attachment on the plates, the flasks were washed with PBS and media was changed to unconditioned media (Table 5) containing 4 ng/mL FGF2. Usually 30 mL of media was used for a T150 flask. Cells were then incubated for 24 h and the media was collected and kept at 4 °C. New media containing FGF2 was then added to the flasks. Media collection was done during six consecutive days. After that cells were split and frozen to further use as feeders. All collected media was centrifuged for 5 min at 300 g and filtered through a 0.2 um filter unit. Media was aliquoted in 45 mL Falcon tubes and stored at -20 °C.

4.2.6. Passage of human iPS-like cells on Matrigel

Matrigel-coated plates were prepared as previously described. New conditioned media was added and the plate was kept in the incubator until cells were transferred. iPS-like colonies were typically grown until 75-90% confluency and then cut manually using a injection needle. Then cells were placed into the incubator for 10-15 min. After that, media was removed and cells were washed with PBS. Cells are then incubated with 1 mL of dispase (Table 2) until the edges of the colonies detached. Cells were gently washed with 2 mL of unconditioned media (Table 5). Fresh unconditioned media (UM) was then added and the plate was placed back on the incubator for another 10-15 min. After cell clumps were gently remove with a cell scraper and transferred into a 15 mL falcon tube for brief centrifugation. Cells were then resuspended in UM (supplemented

with FGF2) and seeded in Matrigel-coated plates with inactivated mouse or human fibroblasts (see section 4.2.3) in a split ratio of 1:3.

Culture methods and conditions were adapted from hESCs (Greber, Lehrach et al. 2008).

4.2.7. Transfection of packaging cells

Human embryonic kidney cells (HEK293) were grown as described previously (see section 4.2.1) until they reached 90-95% confluency. The amounts of plasmids listed in the table below (Table 11) were mixed in 0.5 mL of DMEM. Lipofection reagent was mixed in a separate tube with 0.5 mL of DMEM, and both solutions were incubated for 5 min at room temperature. Solutions were then mixed and incubated for 20 min at room temperature. Media was then changed by fresh fibroblasts culture media without antibiotics and the mixed solution was added dropwise. Cells were then placed back in the incubator. During the next three days after transfection the media was changed for antibiotic-free media in 24 h intervals. Supernatants were centrifuged at 300 g for 5 min and filtered through a 0.45 µm filter. Before transduction of donor cells, the virion-containing supernatants were incubated in the water bath for 10 min to equilibrate the temperature to 37 °C.

Table 11. Transfection mixture. In the left column can be found the plasmid or the reagent used and the amount of them is related in the column on the right. These amounts are necessary to package a single reprogramming factor or GFP.

Plasmid	Amount (µg)
pMXs or pLENTI-GFP	9
GAG-POL	8,1
VSV-G	0,9
Reagent	**Amount (µL)**
Lipofectamine	45

4.2.8. Infection of fibroblasts

The day before transduction fibroblasts were seeded at 200000 cells pro well in a 6-well plate. The next day the media was replaced by virion-containing supernatants, typically 2-3 mL of each factor (OCT4, SOX2, KLF4, and/or c-MYC) or GFP per well, and cells were incubated for 12-16 h. After that time, supernatants were replaced by new aliquots. After two infection rounds media was changed for standard fibroblast growth media and cells were incubated for 12-16 h. Next, media was changed for CM, in which 4 ng/mL of FGF2 had been added. Cells were incubated in this media for 48 h. Cells

were then split and mRNA was harvested, or seeded in a fresh Matrigel-coated plate containing feeders (see section 4.2.3) in case that the experiment was meant to obtain colonies. To obtain colonies the cells were grown in CM containing 4 ng/mL of FGF2 during three to four weeks changing the media every two days.

4.2.9. Derivation of p53-knock-down cell lines

Fibroblasts were grown as described previously (see section 4.2.1) until they reached 90-95% confluency and transfected as described previously (see section 4.2.7) with the amounts of DNA and transfection reagent related in Table 12. The day after transfection 1 µg/mL puromycin was added to the media. After fourteen days cells were split onto new plates and antibiotic pressure was reduced to 0.25 µg/mL. $2x10^5$ fibroblasts cells were then transduced simultaneously with OCT4, SOX2, KLF4, or with OCT4, SOX2, KLF4, and c-MYC as described in section 4.2.8. Four days after transduction cells were washed twice with PBS and pelleted (see section 4.2.8). mRNA was then harvested as described below (see section 4.4.1).

Table 12. Transfection mixture. In the left column can be found the plasmid or the reagent used and the amount of them is related in the column on the right. The amounts here displayed are necessary to package one of both short-haipin containing plasmids.

Plasmid	Amount (µg)
pMKO-shp53 or pLKO-shscramble	4,5
GAG-POL	4,05
VSV-G	0,45
Reagent	**Amount (µL)**
Lipofectamine	27

4.2.10. Inhibition of TGFβ signalling pathway receptors

KLF4 over-expression in fibroblasts was carried out as described (see section 4.2.8). SB431542, a selective inhibitor of a subset of activin receptor-like kinase receptors, specifically ALK4, ALK5 (TGFβR1) and ALK7, was then added to media in a final concentration of 2 µM. Four days after transduction, cells were harvested and pelleted as described in section 4.2.1. Messenger RNA was isolated as described in section 4.4.1.

4.3. Protein-Based Methods

4.3.1. Immunocytochemistry

Cells were washed with TBST and fixed with 4% Formaldehyde (Table 2) for 10 min at 37 °C. After washing twice with TBST, cells were permeabilized with 1% Triton X-100 for 10 min at room temperature. Cells were washed twice with TBST and blocked in TBST containing 5% BSA (Table 2) for 45 min at room temperature with gentle rocking. Afterwards cells were incubated for 1 h with primary antibodies diluted to working concentration in TBST containing 1% BSA. Then cells were washed three times with TBST containing 0.1% BSA and incubated with secondary antibodies diluted in TBST (1% BSA) for 1 h in the dark with gentle rocking. After two washing step with TBST (0.1% BSA), DAPI solution (Table 2) diluted in TBST was added to the wells and plates were left 10 to 20 min in the dark. It followed two washing steps with TBST and fluorescence was analyzed under the microscope (Table 3).

4.4. RNA/DNA protocols

4.4.1. RNA isolation

Total RNA was isolated using the Rneasy Mini Kit from Qiagen according to the instructions given by the manufacturer. Briefly, cells were lysed and homogenized with a guanidine-containing buffer. Ethanol was added and the mixture was loaded into a spin column. Samples were centrifuged ($1x10^4$ g for 1 min) and washed. DNase I treatment was then performed on column using the RNase-free DNase set from Qiagen (Table 1). Membrane was washed twice with an ethanol-containing buffer and the mRNA eluted with RNase-free water.

4.4.2. Reverse transcription PCR

Reverse transcription was then carried out as follows: 500 ng of mRNA and 1 µg/µL oligo-dT were incubated for 3 min at 70 °C and cooled on ice for 2 min. Then the primed mRNA was supplemented with a master mix made off the components listed in Table 13. After 1 h incubation at 42 °C the reaction was stopped at 65 °C for 10 min.

Table 13. Reverse Transcription Components. In the left column can be found the plasmid or the reagent used and the amount of them is related in the column on the right.

Component	Amount (µL)
5x M-MLV Reverse Transcriptase Buffer	5
dNTP (25 mM)	0.5
M-MLV Reverse Transcriptase (200 U/uL)	0.1
dH2O	9.4

4.4.3. RNA quantification and quality control

Due to the instability of RNA and the difficulty of obtaining ribonuclease-free samples, integrity and purity checks are essential steps before any RNA-dependent application. For these reasons, before samples were hybridized onto DNA microarrays absorbance ratios were measured of each sample. In Table 14 are listed the concentrations and absorbance ratios of the samples that were hybridized onto Illumina 8-Sample BeadChips. The same quantification and quality control (see section 4.4.6) was followed in the knock-down of p53 (see section 5.4.3.1), the inhibition of TGFβ signalling (see section 5.4.3.2), and the comparison of different lines detailed in section 5.4.5.1.

Table 14. Sample quantification and quality control. The different samples on study were subjected to quality control: F1 to F3 are non-transduced fibroblasts, G1 to G3 are GFP-transduced fibroblasts, O1 to O3 are OCT4-transduced fibroblasts, S1 to S3 are SOX2-transduced fibroblasts, K1 to K3 are Klf4-transduced fibroblasts, M1 to M3 are c-MYC-transduced fibroblasts. The simultaneous over-expression of OCT4, SOX2, and KLF4 has been abbreviated as 3T1 to 3T3, and the combination of the OCT4, SOX2, KLF4, and c-MYC as 4T1 to 4T3. The ratio between the absorbance at 260 nm and 280 nm (260/280) and, 260 nm and 230 nm (260/230) are also shown.

Lab ID	ID	Conc (ng/µl)	260/280	260/230	chip barcode/field		quality control
GR-10-01	F1	320,56	2,21	1,48	4311471020	A	passed
GR-10-02	F2	315,65	2,21	1,42	4311471020	B	passed
GR-10-03	F3	209,23	2,22	1,07	4311471020	C	passed
GR-10-04	G1	254,97	2,22	1,37	4311471020	D	passed
GR-10-05	G2	332,05	2,19	1,72	4311471020	E	passed
GR-10-06	G3	258,84	2,2	1,6	4311471020	F	passed
GR-10-07	O1	299,37	2,21	1,17	4311471020	G	passed
GR-10-08	O2	259,61	2,22	1,3	4311471020	H	passed
GR-10-09	O3	278,48	2,22	1,35	4311471051	A	passed
GR-10-10	S1	256,61	2,04	0,75	4311471051	B	passed
GR-10-11	S2	277,6	2,2	1,29	4311471051	C	passed
GR-10-12	S3	212,99	2,22	1,17	4311471051	D	passed
GR-10-13	K1	325,47	2,18	1,2	4311471051	E	passed
GR-10-14	K2	302,85	2,21	1,51	4311471051	F	passed
GR-10-15	K3	284,32	2,19	1,82	4311471051	G	passed
GR-10-16	M1	396,24	2,19	1,88	4311471051	H	passed
GR-10-17	M2	345,87	2,2	1,34	4311471063	A	passed
GR-10-18	M3	197,85	2,21	1,15	4311471063	B	passed
GR-10-19	3T1	224,51	2,17	1,58	4311471063	C	passed
GR-10-20	3T2	436,24	2,18	1,6	4311471063	D	passed
GR-10-21	3T3	421,19	2,17	1,65	4311471063	E	passed
GR-10-22	4T1	266,46	2,24	1,35	4311471063	F	passed
GR-10-23	4T2	407,45	2,19	1,28	4311471063	G	passed
GR-10-24	4T3	389,89	2,19	1,6	4311471063	H	passed

Samples received	Hybridization	Chiptype
2010 03 16	2010 03 22	HumanRef-8 v3

4.4.4. Quantitative real-time PCR

Real-time PCR was carried out on the Applied Biosystems 7900 instrument in 96-well optical reaction plates. In Table 15 is detailed the protocol that was run in the ABI instrument (Table 3). The final reaction volume of 20 µL consisted of 10 µL of SYBR Green PCR Master Mix (Applied Biosystems), 2.5 µM of each primer (3 µL) and 7 µL of cDNA (1:8 dilution). Calculation of the relative mRNA levels and their standard deviations were carried out using the comparative Ct method (ABI instruction manual) using the mRNA levels of β-Actin and GAPDH genes as controls for normalization. Primers are

listed in section 4.1.10. At least three biological and three technical replicates were assayed for each target/condition.

Table 15. Real-time protocol. This table shows a detailed description of the parameters used to run real-time PCRs. The number of cycles is listed in the first column, the temperature is given in Celsius degrees in the second column (°C) and the time lasting each cycle is also specified (min: '; seconds: ").

Cycles	T (°C)	Time
1x	50	2'
1x	95	10'
40x	95	15"
	60	1'
1x	95	15"
1x	60	15"
1x	95	15"

If a transcript could not be detected a Ct value equal the total amount of amplification cycles (Ct = 40) was used to make possible further calculations. Very often unspecific amplification occurs at low amounts of starting target sequence. Furthermore, in RT-PCRs to calculate accurately the relative amount between different targets it is necessary to ensure that the amplification of the target sequence is amplified with the same efficiency independently of the initial amount of cDNA. Therefore, to test the efficiency of the amplification of each primer pair at different input concentrations, a series of RT-PCR reactions with a standard dilution of raw positive control samples were run. In addition, the specificity of the amplification in the melting curves resolved by the analysis program SDS 2.1 were checked and the annealing temperature of the amplicon was displayed and compared to that of a positive control. Since the annealing temperature of an amplicon depends on the GC content, identity and order of the sequence, this analysis allows the identification non-specific amplification.

4.4.5. Plasmid isolation

Plasmid isolation was carried out with the NucleoSpin Plasmid Kit from Macherey-Nagel according to the manufacturer's instructions. Briefly, the pelleted bacteria were resuspended and plasmid DNA was liberated in the solution by SDS/alkaline lysis buffer. After neutralizing the resulting lysates, they were loaded into a NucleoSpin column. After centrifugation ($1x10^4$ g for 1 min) the membrane was washed twice and pure plasmid DNA was eluted with a 5 mM Tris/HCl buffer with pH 8.5.

4.4.6. RNA and DNA quantification

The quantity and quality of RNA and plasmid DNA was examined using the NanoDrop (Table 3). 1-2 µL of sample was applied to the NanoDrop and the absorbance ratios at 260/280 (nm) and 260/230 (nm) were measured. The software calculated the concentration based on Beer-Lambert's law.

4.4.7. Agarose gel electrophoresis

RNA and DNA quality control was determined using agarose gel electrophoresis and ethidium bromide staining. 0.8-1.5% gels were obtained by mixing agarose (Table 1) and 1x TAE buffer (Table 2). Ethidium bromide was added directly to the gel before solidifying. The length of the amplicons was specified according to the GeneRuler 1 kb DNA ladder (Table 1). Samples were loaded together with 6x loading buffer (Table 1) and gels were run in an electrophoresis chamber at 60 V for 30-60 min. Nucleic acids were visualized with UV light using a gel documentation system (Table 3).

4.4.8. Illumina 8-Sample BeadChip hybridisation

Biotin-labelled cRNA was produced by means of a linear amplification kit (Table 1) using 400 ng of quality-checked total RNA as input. Chip hybridisations, washing, Cy3-streptavidin staining and scanning were performed on an Illumina BeadStation 500 platform (Table 3) using reagents and following protocols supplied by the manufacturer. cRNA samples were hybridised on Illumina human-8 v3 BeadChips.

4.5. Data analysis

4.5.1. Initial analysis of the microarray data

All basic expression data analysis was carried out using the manufacturer's software BeadStudio 3.0 (Illumina). Raw data were background-subtracted and normalised using the "rank invariant" algorithm. Normalized data were then filtered for significant expression on the basis of negative control beads. Biological replicates were averaged. Selection for differentially expressed genes was performed on the basis of arbitrary thresholds for fold changes plus statistical significance according to an Illumina custom model (Kuhn, Baker et al. 2004).

Signals with highly confident statistical values were considered and labelled as expressed tags (detection P-value <0.01). Differentially expressed genes were further listed based on the signal intensity ratio between the factor in study and non-transduced fibroblasts, considering genes: up-regulated (signal intensity ratio (factor/mock) >1.5), down-regulated (signal intensity ratio (factor/mock) <0.67), undetected to detected

(detection P-value >0.05 to <0.01), and detected to undetected (detection P-value <0.01 to >0.05).

4.5.2. Pathway analysis

The Illumina Gene ID was used to input the genes found differentially expressed in DAVID (Huang, Lempicki et al. 2009). They were further filtered according to Gene Ontology terms or mapped to KEGG pathways. The confidence threshold was set in P-value <0.05, and the background for databases to *Homo sapiens*.

Heatmaps of each relevant pathway were made by collecting the official gene names from KEGG pathway database (www.genome.jp/kegg/pathway.html) and relating them to their gene-expression ratio and confidence value (P-value). Gene lists were then filtered for P-value <0.01. Each of these lists (e.g. apoptosis, cell cycle, etc.) was then analysed with J-Express (Dysvik and Jonassen 2001) using a hierarchical clustering algorithm for a classification of the genes and samples into groups of distinct patterns of gene expression. A complete linkage clustering was conducted based on a euclidean distance function. The heatmaps referring to microarray data (Figure 15 and 16) were performed by Dr. Hendrik Hache.

4.5.3. Meta-analysis of public datasets

Datasets were selected from the Gene Omnibus Expression database (http://www.ncbi.nlm.nih.gov/geo/) out of differential gene expression studies. For Human ESCs were selected two replicates of the cell line H1 (GSM515699 and GSM515700) and two replicates of the cell line H9 (GSM515697 and GSM515698) that were hybridized in Illumina microarrays. For colon cancer cells were chosen three replicates of the cell line SW-420 (GSM21712, GSM21713 and GSM21714) and three replicates of the cell line SW-680 (GSM21715, GSM21716 and GSM21718) that were hybridized in Affymetrix microarrays.

A central pre-requisiste of any meta-analysis approach is the consolidation of the different ID types coming from different versions or platforms of microarrays. The Ensembl database (version 58; http://www.ensembl.org/index.html) was used as the backbone annotation for all studies. IDs were mapped to human Ensembl gene IDs. In total, information on 10177 ENSEMBL genes was mapped.

Affymetrix gene chip annotations were adapted from the latest genome annotation (version 12). Affymetrix data were normalized with GC RMA. Illumina gene chip annotations were adapted from the latest genome annotation (version 3). Illumina

chips were not renormalized and information by gene was displayed using the mean for the associated probes. This meta-analysis was completed by Dr. Mireia Nogales.

4.6. Modelling

The most relevant pathways were introduced in PyBioS (Wierling, Herwig, et. al., 2007; Klipp, Liebermeister, et al. 2009), a software for the development and simulation of mathematical models. Pathway information was retrieved from public databases: KEGG PATHWAY DATABASE (Kanehisa and Goto 2000) and REACTOME (Joshi-Tope, Stein et al. 2005), or were introduced manually based literature data. Each component of the pathways under study was created pair wise as active/inactive species. Interaction between the components was then added and modelled as kinetic reaction. Inactive components were considered to have a basal synthesis and degradation rate, while a degradation rate was enclosed to each active component. The kinetic reactions were considered irreversible, except for complex formation reactions, which are described as reversible reaction with a forward and reverse reaction. The kinetic reactions were simplified to mass action kinetics. All kinetic reactions used in the model are listed in Table 16.

The expression data obtained from the microarray were used to set the protein synthesis ratios. All other parameters were randomized. A Monte Carlo approach was used to solve the set of differential equations representing the kinetic rate laws (Wierling, Kühn et al. 2012). This is achieved by consecutive rounds of random parameterization and calculation of steady state concentrations. Kinetic parameters are randomized with values from a log-normal distribution with mean $\mu = 2.5$ and standard deviation $\sigma = 0.5$. Random parameterization and calculation of the steady state was repeated 100 times for the control and perturbation experiment. A simulation interval of 20000 arbitrary units was used to ensure that the steady state had been reached. This iterative process yields a distribution of steady state results for each component of the model. The geometric mean of the distribution for each component is then used to analyze the differences between control and perturbation experiment. The log2-ratios of relevant model components are then plotted into a heat map (Figure 33 and Figure 37). The simulations presented in this work were carried out by Dr. Alexander Kühn. Dr. Christoph Wierling coordinated and supervised the efforts concerning modelling.

Table 16. Kinetic rate laws. During the assembly of the model different types of reactions were considered and listed in this table. Its biochemical representation has been added for species S1 and S2. The mathematical expression is shown in the column labelled as “Kinetic rate law”, where k is the kinetic parameter that has been sampled within the Monte Carlo approach for each individual reaction. Brackets indicate concentrations of species as G (gene transcript), TF (transcription factor), E (enzyme), or I (inhibitor). k_D represents the dissotiation constant of a complex.

Reaction	Biochemical representation	Kinetic rate law
Synthesis	→ S1	v = k
- with gene	→ S1	v = k * [G]
Transactivation	→ S1	v = k * [G] * [TF]
- with inhibitor	→ S1	v = k * [G] * [TF] / ((1 / k_D) * [I])
Product formation	S1 → S2	v = k * [S1]
- with enzyme	S1 → S2	v = k * [S1] * [E]
- with enzyme & inhibitor	S1 → S2	v = k * [S1] * [E] / ((1 / k_D) * [I])
Complex formation	S1 + S2 → S1:S2	v = k * [S1] * [S2] – k * k_D * [S1:S2]
Degradation	S2 →	v = k * [S2]

5. Results

The derivation of induced pluripotent stem (iPS) cells is of such great importance because they could be used for therapeutic purposes and enables the generation of disease models. The combined over-expression of OCT4, SOX2, KLF4 and c-MYC in somatic cells induces an evolutionarily conserved pluripotency network (Takahashi and Yamanaka 2006) similar to the one found in ESCs. However the process is as yet poorly understood at the molecular level, therefore I aimed to clarify the genetic regulatory network behind the reprogramming process and the foremost molecular mechanisms.

As outlined above (Figure 1), in order to study the molecular mechanisms behind the induction of pluripotency, the reprogramming experiment was reproduced (section 5.1) and the transcriptome of the cells analyzed four days after infection (section 5.2). The most relevant regulatory events were validated (section 5.3), and brought together into a regulatory network by means of a mathematical model (section 5.4.1). The model of the regulatory network enabled to test *in silico* hypotheses about the molecular mechanisms active or inactive in the cells (section 5.4.2). Finally, these hypotheses were validated experimentally (section 5.4.3).

5.1. Infection of human fibroblasts

Thus, first of all human fibroblasts were transduced with different transcription factors independently (OCT4, SOX2, KLF4, and c-MYC), as well as the combinations that yield iPS cells (OCT4/SOX2/KLF4 and OCT4/SOX2/KLF4/c-MYC). Cells were then grown in hESCs standard conditions for 48 h (see section 4.2.6). Besides these transcription factors, GFP was transduced independently in the cells to identify non-specific events in further analysis. As outline in Figure 8, mRNA was harvested four days after transduction from three different biological replicates of non-transduced fibroblasts, GFP-, OCT4-, SOX2-, KLF4- and c-MYC-transduced fibroblasts, and two combinations of the factors over-expressed simultaneously: OCT4/SOX2/KLF4 and OCT4/SOX2/KLF4/c-MYC (see section 4.4.1).

In order to verify that the protocol used was able to reprogram the donor cells, one replicate of fibroblasts transduced with the different combinations of factors was split 48 h after transduction. Cells were then seeded onto fresh Matrigel-coated plates containing feeder cells (see section 4.2.3), mimicking the conditions in which hESCs grow (see section 4.2.6). From this time point on conditioned media (CM; see section 4.2.5), necessary for iPS growth, was used instead of fibroblast media (see section

4.2.1). CM was freshly added every two days. The first colonies were visible during the third week for fibroblasts over-expressing the four reprogramming factors and during the fourth week in fibroblasts over-expressing the three factor combination. After the fifth week colonies showing the typical morphology of hESCs, well-defined edges and high nucleus-to-cytoplasm rate, were manually picked and further passaged onto fresh plates, grown for a week in hESCs conditions and frozen.

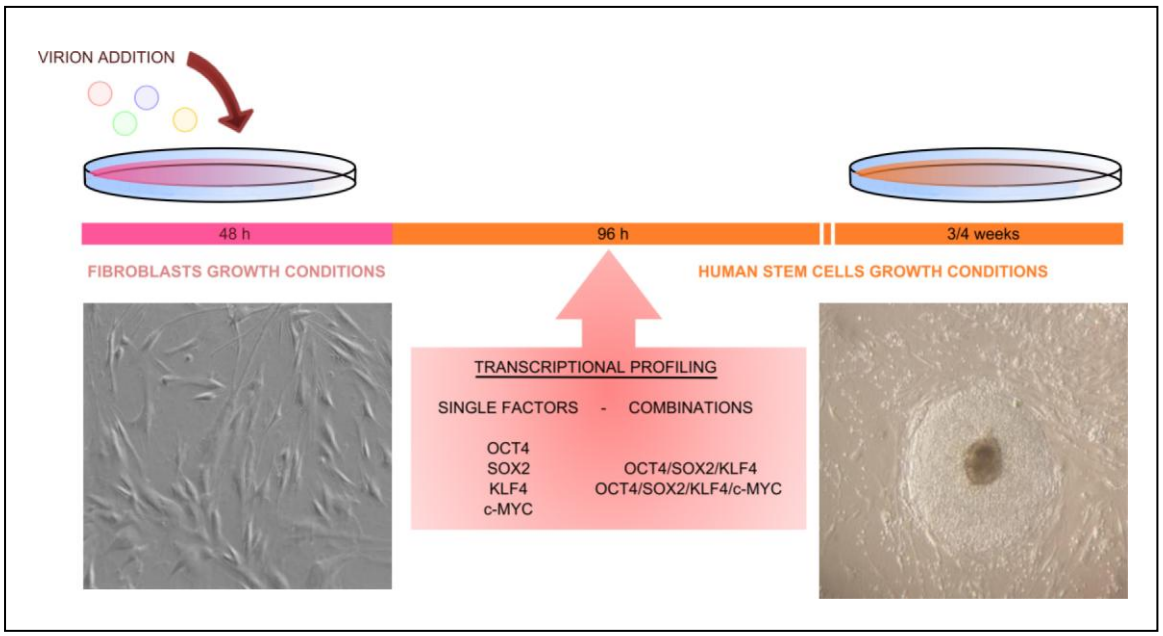

Figure 8. Outline of the experimental design. Human fibroblasts are transduced with each of the reprogramming factors OCT4, SOX2, KLF4, and c-MYC and combinations of them that achieve induced pluripotent cells: OCT4, SOX2, and KLF4 (3TF) and OCT4, SOX2, KLF4, and c-MYC (4TF). Transduced fibroblasts are plated and grown on human embryonic stem cell conditions thereafter. Four days after seeding, mRNA is harvested and hybridized onto microarrays. The picture of the left corresponds to donor fibroblasts, while picture on the right shows a colony of iPS-like cells that had been passaged onto a fresh plate after the reprogramming experiment.

5.1.1. Confirmation of virus-mediated protein over-expression

In order to confirm that the over-expression of the factors with the employed system was working, the presence of the transduced transcripts was tested by quantitative real time-PCR (qRT-PCR) in each sample. In Figure 9 and Figure 10 are shown the results of the qRT-PCR performed on cDNA synthesized from each of the replicates listed in Table 14 are shown. Each sample was tested for transcripts of an internal control, namely housekeeping genes *β-Actin* or *GAPDH*, and the transcript of the protein that had been transduced. The internal control was then used to normalize the Ct values. The relative amount of transcript-of-interest found in each sample was then compared to that found in non-transduced fibroblasts (mock). ESCs mRNA (H9) was used as positive control for OCT4, SOX2, KLF4, and c-MYC expression.

As shown in Figure 9, there were significant fold-changes for GFP, OCT4 and SOX2 (15- to 20-fold change (log_2 scale)), since non-transduced fibroblasts show no expression of these genes. On the other hand KLF4 and c-MYC transcripts could be detected in non-transduced fibroblasts, therefore yielding more moderate fold changes (4- to 8-fold change (log_2 scale)). The transcripts of KLF4 and c-MYC were also detected in the H9 cell line used as positive control, even though transcript levels of these two proteins are similar in H9 than in fibroblasts.

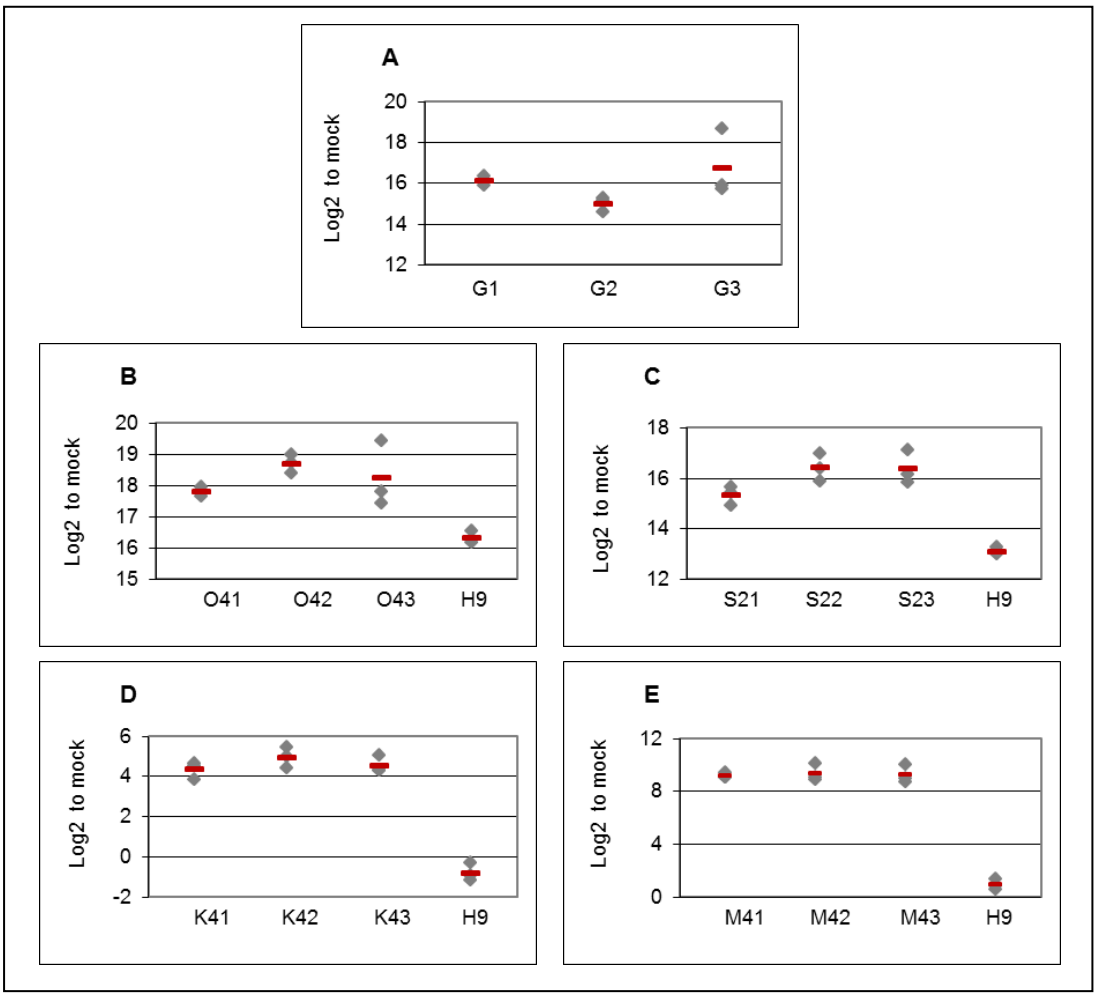

Figure 9. Real-time confirmations of virus-mediated protein expression in fibroblasts. Plots show the amounts of OCT4, SOX2, KLF4, and c-MYC transcripts relative to non-transduced cells in the three biological replicates in which the reprogramming factors were over-expressed. The specificity and efficiency of the primers were tested using a positive control (H9 or GFP-transduced fibroblasts). Once the primers had been validated, qRT-PCR for each sample was tested for transcripts of an internal control (β-Actin and GAPDH; see section 4.1.10) along with that of protein that had been transduced (GFP, OCT4, SOX2, KLF4, or c-MYC). The internal control was then used to normalize the Ct values. The relative amount of transcript found in each sample was then compared to that found in non-transduced fibroblasts (mock). The Y-axis indicates the logarithmic fold change and the X-axis the abbreviation for each replicate. **A**: G1, G2, G3 are three biological replicates of fibroblasts transduced with GFP. **B**: O41, O42, O43 are three replicates of OCT4 over-expressing cells. **C**: S21, S22, S23 are the replicates of SOX2. **D**: K41, K42, K43 are abbreviations for the three KLF4 replicates. **E**: MY1, MY2, MY3 are three replicates of fibroblasts over-expressing c-MYC. Three technical replicates were averaged for mock and H9. The red bar indicates the average of the three replicates.

The cDNA from fibroblasts transduced with the different combination of factors, OCT4, SOX2, and KLF4 (3TF) and OCT4, SOX2, KLF4, and c-MYC (4TF), was also checked for the expression of the different factors. In Figure 10 are shown the amounts of OCT4, SOX2, KLF4, and c-MYC transcript levels for the three biological replicates in which the reprogramming factors were over-expressed simultaneously. Higher transcript levels were found for OCT4 and SOX2 than for KLF4 and c-MYC relative to non-transduced fibroblasts, as it was the case in single transfections. Interestingly, the relative amount of transcript levels for KLF4 and c-MYC were lower (2- to 5-fold change (log 2)) when the factors were over-expressed simultaneously with OCT4 and SOX2.

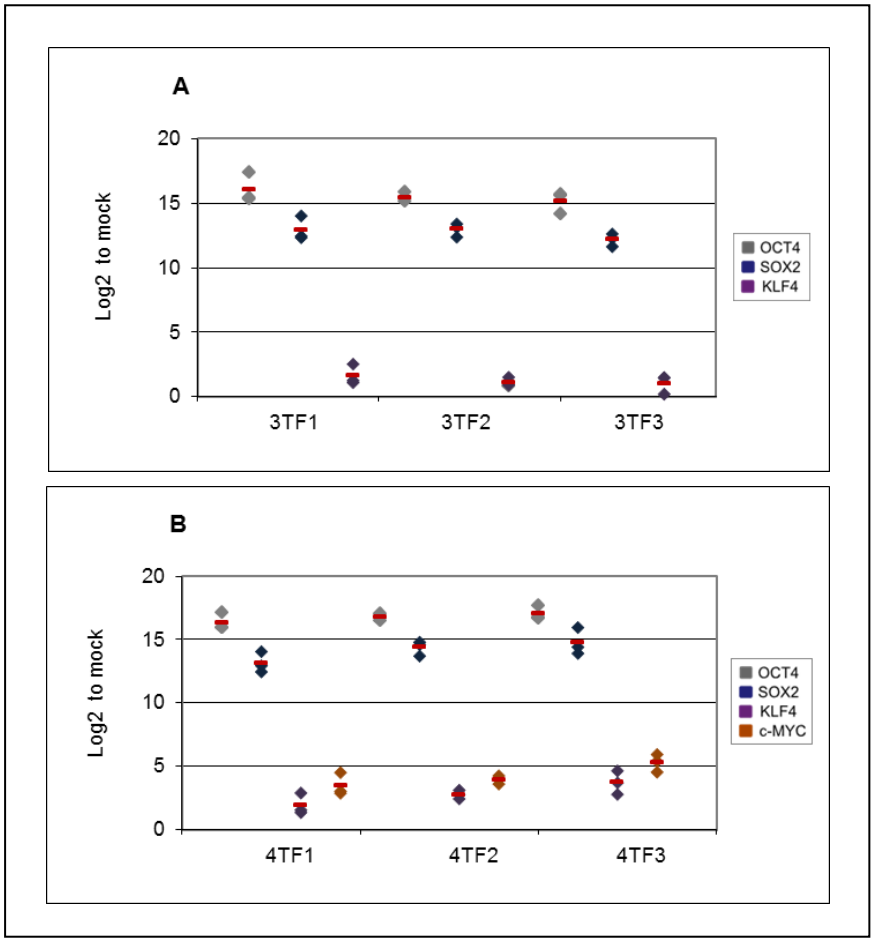

Figure 10. Real-time confirmations of virus-mediated protein expression in fibroblasts. Plots show the level of transcripts from the different proteins when they were over-expressed simultaneously. The relative amount of each transcript relative to non-transduced (mock) cells is shown on the Y-axis (logarithmic fold-changes). The X-axis lists the abbreviations used for each sample. **A**: 3TF1, 3TF2, 3TF3 are abbreviations of three biological replicates over-expressing OCT4, SOX2, and KLF4. **B**: 4TF1, 4TF2 and 4TF3 are biological replicates from fibroblasts over-expressing OCT4, SOX2, KLF4, and c-MYC.

If it can be regarded that infections occur independently of each other, and do not depend on the amount of infectious particles added, as has been shown by other research groups (Papapetrou, Tomishima et al. 2009), then the lower levels of KLF4 and c-MYC could suggest that OCT4 and/or SOX2 regulate directly or indirectly the expression of KLF4 and c-MYC.

Overall, the presence of transcripts from each of the reprogramming factors was confirmed in all samples. In order to quantify the number of cells that were expressing the different transcription factors transduced, immunocytochemistry was performed (see section 4.3.1). To this end antibodies targeting OCT4, SOX2, KLF4, and c-MYC were used (see sections 4.1.8 and 4.1.9). Immunocytochemistry was performed on fifty-thousand cells pooled together from three biological replicates where the four factors were simultaneously over-expressed. The number of positive cells in a total of one hundred cells was counted (Figure 11). Non-transduced fibroblasts were used to test for unspecific targeting of the antibodies. The efficiency of infection was estimated to be in 75% of total cells. It has been shown that infections occur at the same rate in cells independently of the use of one or more factors (Papapetrou, Tomishima et al. 2009). It can then be estimated that around 55% of the cells were over-expressing simultaneously two factors, approximately 40% three factors and 30% all four factors.

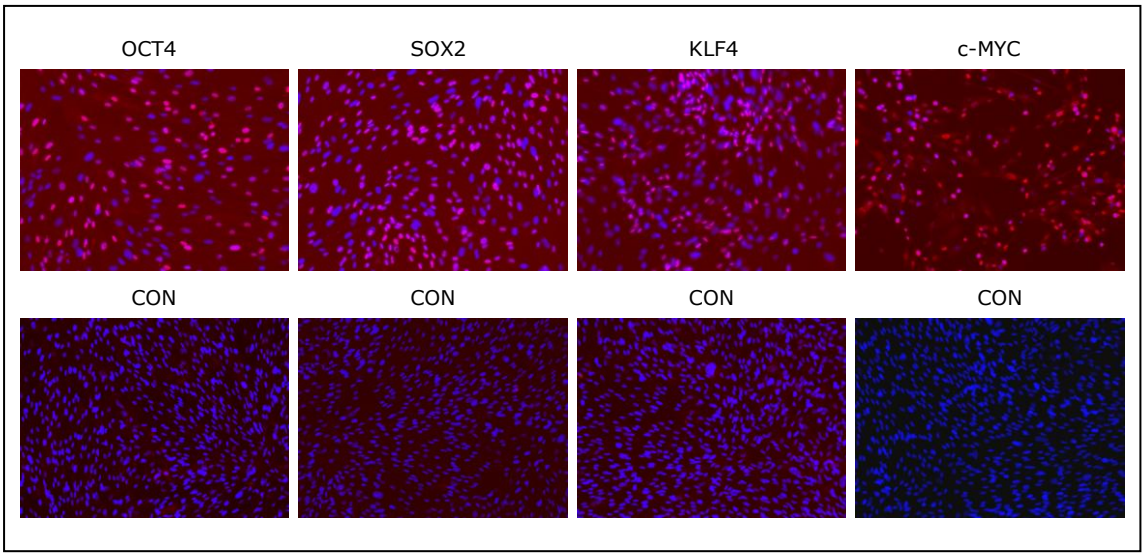

Figure 11. Infection efficiency. Immunocytochemistry detection of the transcription factors transduced in fibroblasts cells 96 h after transduction. Non-transduced cells (CON) were also treated with the different antibodies to check that unspecific staining was not occurring. DAPI was used to stain the nucleus of the plated cells and thus calculate the efficiency of the transduction. Images show the overlap between the DAPI (blue) and the factor specific staining (red) after over-expression.

5.1.2. Expression analysis of the microarray data

Once it was confirmed that 75% of fibroblasts were successfully transduced with single factors, cRNA was hybridized to DNA-microarrays (Illumina human-8 v3 BeadChips; see section 4.4.8). The different samples hybridized are related in Table 14. Briefly, three biological replicates of the following samples were submitted to genome-wide transcriptional analysis: non-transduced fibroblasts, GFP-transduced fibroblasts, OCT4-transduced fibroblasts, SOX2-transduced fibroblasts, KLF4-transduced fibroblasts, c-MYC-transduced fibroblasts and samples over-expressing simultaneously of OCT4, SOX2, and KLF4 (abbreviated as 3TF or 3T), and samples over-expressing simultaneously OCT4, SOX2, KLF4, and c-MYC (abbreviated as 4TF or 4T).

To analyze all of the microarray datasets globally, biological replicates were averaged, and a cluster analysis was performed. An overall view of similarity between samples can be gained by clustering samples using the correlation values. Samples that have the most similar expression profiles are then clustered together. The results of the cluster analysis are presented in Figure 12. The different samples cluster into two main groups, with most of the single transduced samples in one, and the combinations of factors and c-MYC in the other one.

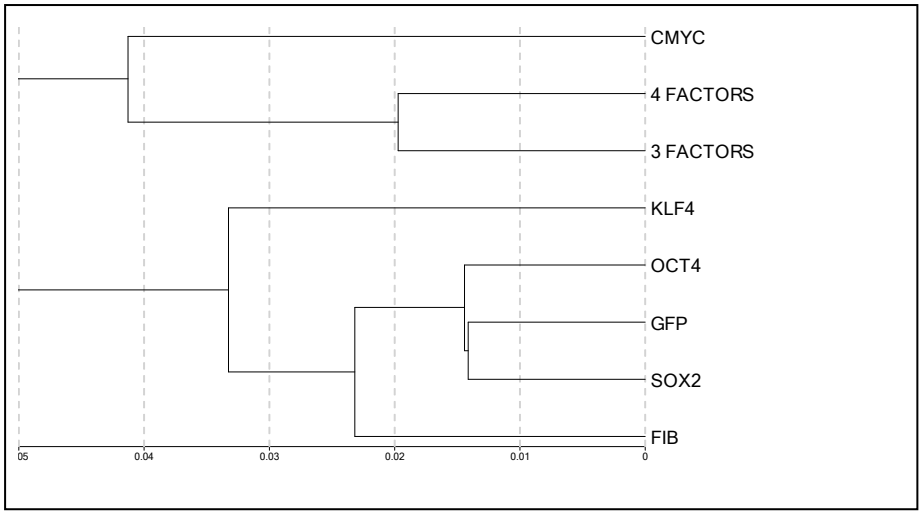

Figure 12. Cluster analysis. Three different biological replicates have been averaged and all genes classified as expressed (detection P-value <0.01) have been considered for this analysis. Control fibroblasts (FIB) were used as reference. This cluster analysis has been made with BeadStudio 3.0 (Illumina) and its implemented method for hierarchical clustering based on cubic spline normalization of the raw data and Pearson-correlation as distance. The scale shows 1 minus the correlation coefficient between the signal intensities.

OCT4, SOX2, and KLF4 clustered together with the GFP control. The differences between samples are lower than the differences between any of them as compared to

non-transduced fibroblasts (labelled as “FIB” in Figure 12). Amongst the single factors, KLF4 showed the closest profile to the other group of samples. On the other hand, the profiles of the combinations of factors are very similar to each other and share large similarity with c-MYC. Moreover, this cluster analysis suggests an important role for c-MYC and KLF4 during the first four days after infection since they induced a gene expression profile most similar to that of the combination of factors.

The most meaningful information from microarray experiments is gained through the analysis of genes differentially expressed between the control and the treated sample. In order to start this sort of analysis, biological replicates were averaged and the different datasets were sorted based upon P-value scores, taking into account only highly reliable values (P-value <0.01) for further examination. Hence, between 8000 and 9200 genes were considered detected after filtering the data (“Expressed” in Table 17). The fold-change between the intensity values for a certain gene in each sample against the signal intensity of the same gene in non-transduced fibroblasts was then calculated. Genes were classified as up-regulated (“Up” column in Table 17) if the ratio between treated sample over control was higher than 1.5, and alternatively, genes were classified as down-regulated if the ratio between treated sample over control was lower than 0.67 (“Down” column in Table 17).

Because this ratio classification criteria relies upon confident values on both, sample and control, genes that might be just expressed, or silenced, after over-expression—giving confident values merely for sample or control—would not be considered. Thus, an additional listing based upon significance values was included for further analysis. Genes that were not detected in mock (P-value >0.01), but were detected after over-expression (P-value >0.01), or conversely, were detected in mock but were not detected after over-expression were also listed (columns labeled as “U to D” and “D to U” in Table 17, where D is an abbreviation for detected, and U is an abbreviation for undetected). Hence, differentially expressed genes (“Differentially Expressed” column with light blue background in Table 17) comprise genes up-regulated, down-regulated, or genes that changed its detection status after over-expression.

Nevertheless, in order to discriminate between genes differentially expressed due to the methodology employed, and those specifically regulated by the reprogramming factors, fibroblasts were also transduced with GFP. Thus, to highlight the particular effect of each of the reprogramming factors and their combinations, genes found differentially expressed after GFP over-expression were subtracted from each of the datasets and alternative lists were created (bold numbers in Table 17).

A general analysis of genes differentially expressed can then be started at the over-expression of single factors, where it is notable that c-MYC caused the largest change in the transcriptome, altering the expression of more than 5000 genes, more than the amount altered by the combination of factors. Furthermore, it had an overall activating effect on the genes, as judge by the larger amount of genes up-regulated than down-regulated. On the other hand, KLF4 shows many more genes down-regulated than up-regulated. This effect was also apparent when KLF4 was over-expressed in combination with the other factors. Among single transcription factors, OCT4 showed the lowest number of differentially expressed genes, being just 417 more than in GFP. Instead, following 3TF and 4TF over-expression 3480 and 4749 genes respectively showed changes in expression. Between 3TF and 4TF the largest difference is found in the number of up-regulated genes, which increased notably upon c-MYC addition.

Table 17. General analysis of differentially expressed genes. All basic expression data analysis was carried out using BeadStudio 3.0 (Illumina). Raw data were background-subtracted and normalized using the "rank invariant" algorithm. Normalized data were then filtered for significant expression on the basis of negative control beads. Signals with highly confident statistical values were considered and labelled as expressed (detection P-value <0.01). Differentially regulated genes were further filtered between up-regulated (signal intensity ratio (factor/mock) >1.5; columns labelled as “Up”), down-regulated (signal intensity ratio (factor/mock) <0.67; columns labelled as “Down”), genes that were not expressed in mock but were found expressed after over-expression (detection P-value mock >0.01 and detection P-value after over-expression <0.01; columns labelled “U to D” as abbreviation for undetected to detected) and genes that were expressed in mock but not expressed after over-expression (detection P-value mock <0.01 and detection P-value >0.01); columns labelled “D to U” as abbreviation for detected to undetected). In the columns labelled with “(-GFP)” have been subtracted the genes that were also found up/down-regulated or changed their detection status following GFP-over-expression. Genes up-regulated, down-regulated or genes that changed its detection status after over-expression were added and listed as differentially expressed (light blue background column). 3TF is the abbreviation for OCT4, SOX2, and KLF4 simultaneous over-expression and 4TF is the abbreviation for OCT4, SOX2, KLF4, and c-MYC simultaneous over-expression.

	Expressed	Up	Up (-GFP)	Down	Down (-GFP)	U to D	U to D (-GFP)	D to U	D to U (-GFP)	Differentially Expressed
MOCK	8417									
GFP	8531	607		412		428		314		1761
OCT4	8540	771	**417**	632	**374**	450	**210**	327	**144**	2180
SOX2	8911	896	**551**	711	**414**	712	**410**	218	**84**	2537
KLF4	8551	814	**508**	1103	**801**	540	**306**	406	**220**	2863
c-MYC	8704	2904	**2347**	940	**688**	734	**448**	447	**262**	5025
3TF	8914	999	**565**	1374	**1027**	802	**492**	305	**150**	3480
4TF	9144	1637	**1185**	1659	**1335**	1090	**780**	363	**209**	4749

Common and specific regulation among the different factors and GFP can be initially studied by grouping genes shared between the different datasets. Thus, genes listed as up- and down-regulated, as well as genes that changed their detection

threshold (underlined numbers in Table 17) were used to construct the Venn diagrams shown in Figure 13. Most of the genes differentially expressed in fibroblasts transduced with GFP are shared with one or more of the factors, suggesting a large transcriptional background due to the methodology. Hence, OCT4, SOX2, and KLF4 have a number of genes that are specifically regulated by each of them (see Figure 13A). SOX2 has the highest specific regulation for up-regulated genes (553 genes of 1608 total genes: 38%), while KLF4 shows the largest number of down-regulated genes that are not found in GFP, OCT4 or SOX2 (608 genes of 1509 total genes: 40%; see Figure 13C).

The larger number of differentially expressed genes specific to one factor was found in the c-MYC dataset, where 1456 (from a total of 3638: 40%) genes achieved the detection threshold or were up-regulated (see Figure 13B). Even though, some of the activating effects that c-MYC created in the cells after over-expression were recapitulated when it was over-expressed in combination—c-MYC and 4TF shared exclusively 726 genes—, the fact that as much as 40% of genes regulated by c-MYC are not differentially expressed after 4TF over-expression suggests that c-MYC transcriptional activation is regulated to a large extent by the other factors.

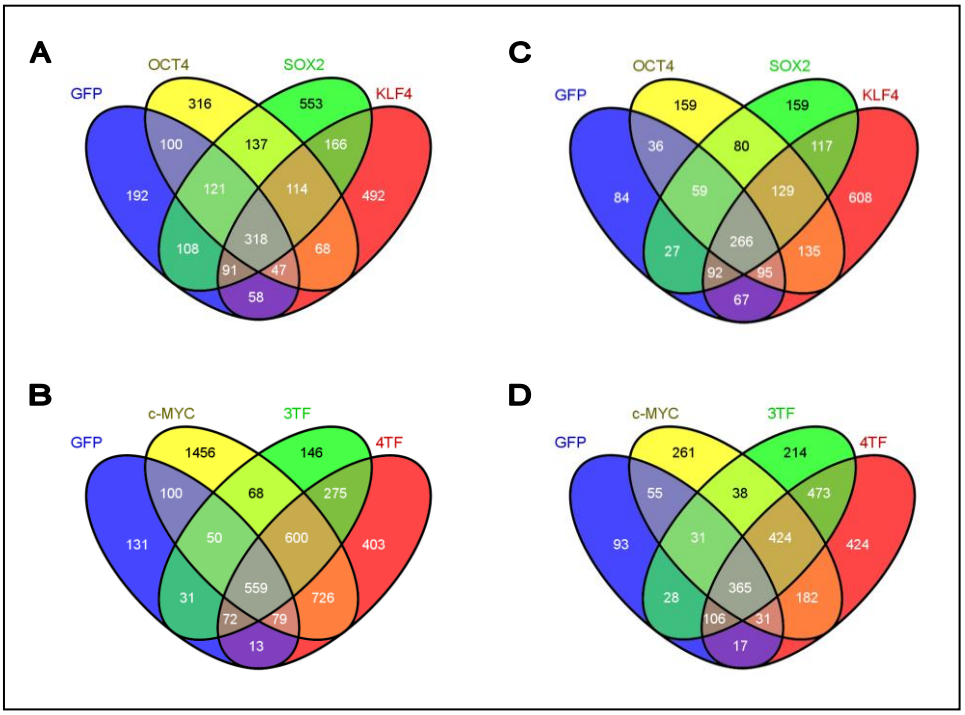

Figure 13. Common and specific regulation. Genes listed in Table 17 as differentially expressed were grouped into two sets: **A**, **B**: up-regulated together with undetected to detected genes, and **C**, **D**: down-regulated together with detected to undetected genes (right). These lists of genes were created for each sample (GFP, OCT4, SOX2, KLF4, c-MYC, 3TF (OCT4, SOX2, and KLF4) and 4TF (OCT4, SOX2, KLF4, and c-MYC)) and compared to the others using VENNY (Oliveros J.C. 2007; http://bioinfogp.cnb.csic.es/tools/venny/index.html).

It is striking that the overall genes that increased their expression after 3TF over-expression, independently if they were shared or not with GFP and/or c-MYC, also increased also their expression after 4TF over-expression (see Figure 13B). Albeit the gene network up-regulated after 3TF includes a collective background shared by GFP, c-MYC and 4TF (559 genes), common regulatory features between c-MYC and 3TF that are conserved in 4TF (600 genes), and genes uniquely regulated in 3TF and 4TF, a total of 1506 genes (600+559+275+72), which represents more than 83% of the gene network up-regulated after 3TF over-expression (1801 genes) is conserved following 4TF over-expression.

Those 1506 genes that 4TF shares with 3TF make up around 55% (from a total of 2727) of the genes that increased their expression following 4TF over-expression, although. The remaining genes were shared just with c-MYC (726 genes: 27%), or found exclusively in the 4TF dataset (403 genes: 15%). Hence, the combination of factors could be achieving also regulatory effects on the gene network that go beyond the regulatory features observed in single factors.

Amongst the down-regulated or undetected genes similar distribution patterns were found (see Figure 13D). A large amount of genes in the 3TF dataset is shared with 4TF (1368 of a total of 1679 genes decreasing their expression in 3TF: 81%), being found in 4TF a remarkable number of specific genes (424 genes from a total of 2022: 21%).

Thus, this analysis supports that virus-induced over-expression has a strong effect on transcription. Besides this, the regulatory network established after 3TF over-expression conditions c-MYC activating effects. Even though differentially expressed genes in 4TF can be explained principally by 3TF and c-MYC gene regulation, the simultaneous combination of the factors may give rise to regulatory features not achieved by any of the factors alone.

5.2. Pathway analysis of the microarray data

High-throughput experiments, including microarray expression experiments, tend to contain false positives. This makes it difficult to distinguish between genes truly involved in the biological process under study and those altered inherently by the technique. In addition, it is difficult to understand the functional significance of differentially expressed genes in multi-gene datasets and pin point relevant relationships between the many candidate genes that such experiments can yield. Therefore, researchers tend to subgroup genes into pathways depending on the discrete cellular function in which they participate. Pathway analysis within large datasets requires

statistical quantification of the assembly to extract with confidence true positive genes for use in further experimental analysis.

Hence, genes listed previously as differentially regulated—in the case of the reprogramming factor/s specifically regulated (bold numbers in Table 17)—were grouped in up-regulated genes, which included genes considered undetected in non-transduced fibroblasts but detected after over-expression ("U to D" in Table 17), and down-regulated genes, including genes considered detected in non-transduced fibroblasts but detected after over-expression ("D to U" in Table 17), were introduced in the software DAVID (Huang, Lempicki et al. 2009) in order to classify those pathways over-represented in the different groups of genes.

These analyses were limited to those pathways concerning signal transduction pathways and biological processes, since adaptive changes in an established transcriptional regulatory network are set principally by signal transduction pathways. Adding biological processes to the analyses enables assessing the functional relevance of the changes in the transcriptional regulatory network. The KEGG pathway database was used for these analyses, because it contains a complete and comprehensive set of pathways. The confidence threshold for the pathway analysis was set on P-value <0.05. In the analysis of c-MYC and 4TF datasets metabolic pathways were included, because the number of specific (GFP-background subtracted) differentially expressed genes classified in metabolic pathways was close to a third of the total considered. The results of these analyses are summarized in Figure 14 and the raw data, including average signal intensities, P-values and ratios for each gene are presented in the appendix.

5.2.1. Pathways commonly regulated

As mentioned before (see section 5.1.2), changes were found in the transcriptome profiles independently of the factor expressed. The combination of signalling pathways and processes differentially expressed indicates that retroviral mediated over-expression caused an activation of transcripts related to the p53 signalling pathway (see "GFP" in Figure 14). Furthermore, in most of the datasets this was accompanied by up-regulated death-ligand related apoptotic signalling. Another pathway was enriched, the phosphatidylinositol signalling system, which may influence the p53 signalling pathway and apoptosis (Majerus, Zou et al. 2008). Even though, this system can play also important roles in many different cellular functions as cell growth, proliferation, differentiation, motility, survival and intracellular trafficking (Yuan and Cantley 2008). After transduction cells also showed changes in the expression of cell cycle proteins, more prominently in those involved in DNA replication. In addition, genes

involved in replication-related mechanisms such as mismatch repair were also found to be up-regulated.

While the pathways above were found over-represented among the genes with higher expression following transduction, just genes of the Hedgehog signalling pathway were significantly enriched among the genes that decreased their transcription. This pathway in combination with the TGFβ and the Wnt signalling pathways are essential in development and controls cell fate and morphogenesis (Capdevila and Izpisúa Belmonte 1999; Ingham and McMahon 2001).

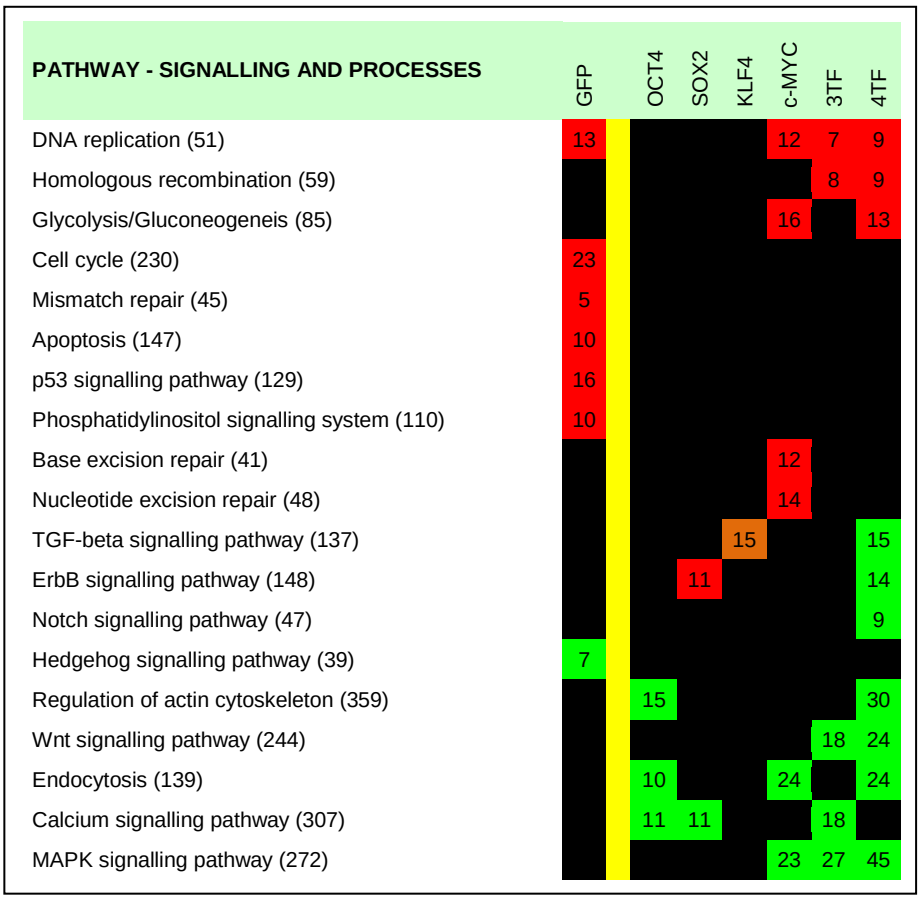

Figure 14. Pathway analysis. List of pathways that were found to be differentially expressed (up-regulated: red; down-regulated: green; orange: up- and down-regulated). The analysis was done using the KEGG pathway database and filtering for signal transduction pathways and cellular processes. The number of genes that are considered in each pathway is indicated after the name of the pathway. The first column shows GFP, which is used as a subtractive control for the other groups of samples. In this way, specific regulation by the factor/s should be emphasized. The number of genes found differentially expressed in each dataset and for each pathway is shown in the columns. The confidence threshold was set on P-value <0.05. Signal intensities, P-values and ratios for each gene differentially expressed that belong to the pathways highlighted here are listed in the appendix.

5.2.2. Pathways differentially regulated after KLF4 transduction

Because the expression profile of the combination of factors shows more resemblance to the KLF4 and c-MYC profiles (see Figure 12), I started first to analyse the changes produced by those factors. Additionally, most changes in expression that were observed in GFP, OCT4, and SOX2 were found amplified in the KLF4 dataset. For these reasons a detailed analysis of this dataset is prioritized here.

As pointed out, genes involved in apoptosis changed their transcription in most datasets. Among them, the transcriptional up-regulation of death ligands (e.g. *FAS*) or intrinsic apoptotic signals (e.g. *ENDOG*) were widely observed. A weak down-stream response is common between GFP, OCT4, SOX2, and KLF4, regarding many genes directly involved in the degradation of cellular structures with no or low change in expression (e.g. *CASP6*, *DFFA*). Nevertheless, KLF4 showed specific effects on the apoptotic cascade, since it is observed down-regulation of key genes in the process (e.g. *CASP3*). In addition, this factor could also be promoting survival by inhibiting *TP53* (Figure 15B).

KLF4 over-expression showed specific effects on transcription, particularly in the TGFβ pathway that was found over-represented in the list of genes with higher expression, as well as in the list of genes with lower expression after transduction (see Figure 14). The TGFβ signalling pathway has crucial roles in embryonic development by controlling tissue morphogenesis, stem cell proliferation, apoptosis, differentiation and migration (Hogan, Blessing et al. 1994).

The regulation of the TGFβ signalling pathway was remarkably altered (see Figure 15A). Most of the differences such as the deregulation of genes encoding secreted proteins (e.g. *BMP2/4/5*, *TGFB2*, *NODAL*, GDF5), or the up-regulation of inhibitors of the pathway (*SMURF1/2*) could be observed in all datasets. Strikingly, key members of the cascade (e.g. *TGFR2*, *SMAD4*) were found down-regulated, while effectors of the pathway like ID transcription factors (e.g. *ID1/2/3*) or the cell cycle inhibitor, *CDKN2B* also named *p15*, were strongly up-regulated.

Overall these data show that KLF4 over-expression can have important consequences on cell survival and proliferation. Importantly, apoptotic markers were down-regulated and inhibitors of the cell cycle were found up-regulated, suggesting that this factor may be driving cells into senescence.

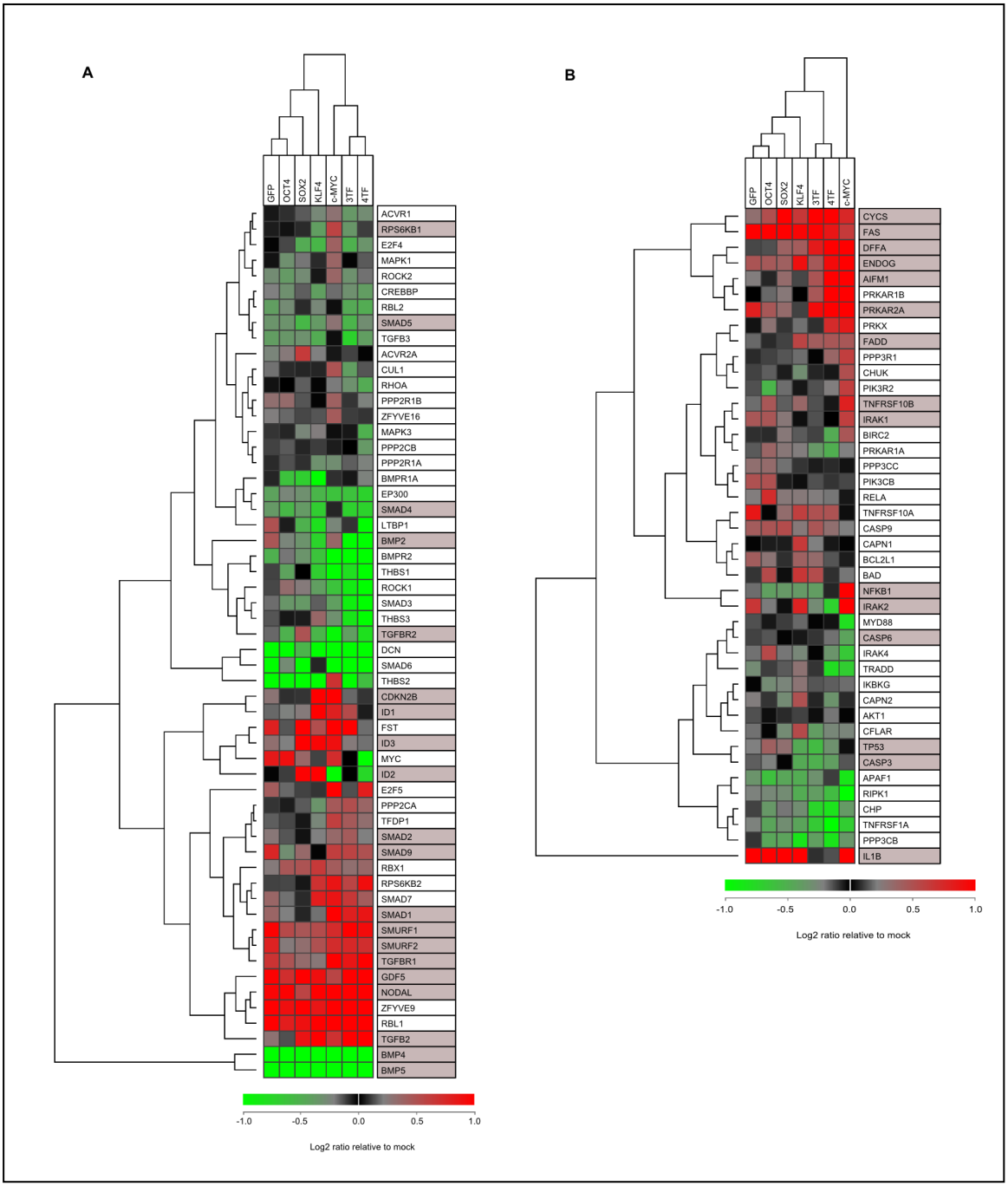

Figure 15. Expression data from DNA microarrays – TGFβ pathway and apoptosis. In columns are indicated the factor/s over-expressed (3TF is OCT4, SOX2, and KLF4, and 4TF is OCT4, SOX2, KLF4, and c-MYC). In rows are listed genes. Red represents up-regulation (log2 ratio between factor expressed versus non-transduced fibroblasts >1.5); Green represents down-regulation (ratio between factor expressed versus non-transduced fibroblasts <0.67). Just statically significant values (detection P-value <0.01) were used. This plot was carried out with J-Express (see section 4.5.2) using a Euclidean distance function and a complete linkage algorithm. Samples and genes were clustered base on their relative amount of expression. Three different biological replicates have been averaged. **A**: TGFβ pathway; **B**: Apoptosis. Genes mentioned in the text are displayed on a grey background.

5.2.3. Pathways differentially regulated after c-MYC transduction

The over-expression of c-MYC had wide-spread effects on cell transcription (see Table 17). Processes such as DNA replication or repair systems (e.g. mismatch repair, base excision repair and homologous recombination) were classified among the pathways over-represented with high confidence values. Once again, apoptotic cascades show very different patterns in transduced cells than in non-transduced fibroblasts. As well, this transcription factor induced many changes in metabolic and energy production pathways that reflected in higher protein synthesis rates (see Table 18).

Increased levels of c-MYC induced chances in cell cycle regulation (Figure 16B). Even though GFP-expressing cells showed already changes in phosphatases that regulate the cell cycle at different points (e.g. CDC45L, CDC25C), Cyclins (e.g. CCNA2, CCND2, CCNB2) and members of the Minichromosome Maintainance Complex (e.g. MCM2/4), c-MYC over-expression increased the transcription of more genes participating in the progression between the G_1 and S phases of the cell cycle (e.g. *CDK4/6*, *CCND1/3, E2F3/5*). It was therefore not surprising to find many different subunits of the DNA replication machinery up-regulated such as components of DNA polymerases, subunits of the helicase and replication factors (e.g. *POLA1*, *POLE*, *POLE3*, *MCM3/6* and *RFC1*). Yet some inhibitors of the cell cycle were also up-regulated over threshold levels (e.g. *CDKN1A*). The ratios between the average signal intensity after c-MYC over-expression and the control (fibroblasts) for each gene mentioned here have been listed in the appendix (Table 24).

After over-expression of this factor cell death could readily be observed in the growth plates. Consequently many more genes of the apoptotic pathway changed their expression. Extrinsic (e.g. *FAS*, *DFFA*) and intrinsic (e.g. *AIFM1*, *ENDOG*) signalling components were up-regulated. Death receptors and adaptor proteins of the extrinsic pathway were also up-regulated (e.g. *TNFRSF10B*, *FADD*). On the other hand, pro-survival genes (e.g. *PRKAR2A*, *IL1B*, *NF-κB1*) were found up-regulated (Figure 15B).

The results from c-MYC over-expression showed some similarities with KLF4 over-expression in the TGFβ signalling pathway. Deregulation of genes encoding secreted proteins (e.g. *BMP2/4/5*, *TGFB2*, *NODAL*, *GDF5*), and up-regulation of some transcription factors (e.g. *ID1/3*) were common regulatory events. c-MYC showed however an extended activation of the pathway, which included up-regulation of genes coding for receptors (e.g. *TGFR1*), adaptor proteins (e.g. *ZFYVE16*), signal transduction kinases (e.g. *RPS6KB1*) and transcription factors of the BMP branch (e.g. *SMAD1/5*).

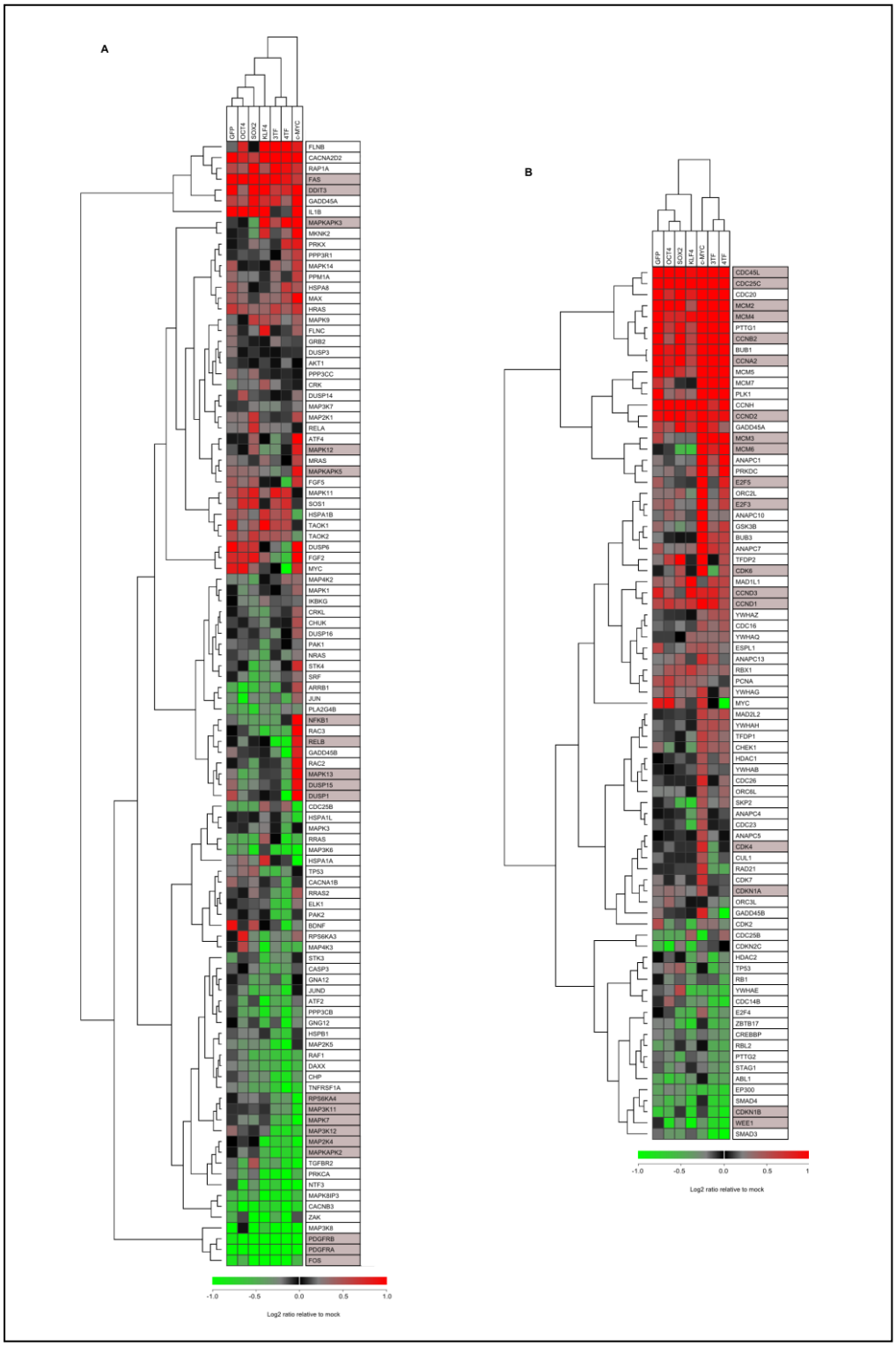

Figure 16. Expression data from DNA microarrays – MAP kinase signalling pathway and cell cycle. In columns are indicated the different datasets, while in rows are listed genes. Red represents up-regulation (log2 ratio between factor expressed versus non-transduced fibroblasts >1.5); Green represents down-regulation (ratio between factor expressed versus non-transduced fibroblasts <0.67). Just statically significant values (detection P-value <0.01) were used. This plot was carried out with J-Express (see section 4.5.2) using a Euclidean distance function and a complete linkage algorithm. Samples and genes were clustered base on their relative amount of expression. Three different biological replicates have been averaged. On panel **A** are shown the members of the MAP kinase signalling pathway; on panel **B** are cell cycle related genes. Genes mentioned in the text are displayed on a grey background.

The MAP kinase pathway, transducer of the mitogenic stimulus of the growth factors contained in the media, was listed amongst the pathways over-represented in the set of genes that decreased their expression following transduction. There were regulatory features of the pathway that were found in all the datasets, such as the down-regulation PDGF receptors and the FOS protein, or the up-regulation of pro-apoptotic proteins as DDIT3 or FAS. More commonly it was observed a decrease in the expression of members that was often shared between c-MYC and the combination of factors, and occasionally also KLF4, for instance the down-regulation of ERK5 branch (e.g. *MAPK7*), the reduced levels of ATF4 activators (e.g. *MAPK12*, *MAPKAPK2/5*, *and RPS6KA4*) or the decrease in JNK activators (e.g. *MAP3K11/12*, *MAP2K4*). Specific regulatory features of c-MYC over-expression were related to stress response proteins that were found up-regulated (e.g. *NF-κB1*, *RELB, DUSP1/15, and MAPK13*).

As mentioned before metabolic activities after c-MYC over-expression are fundamentally different from those found in non-transduced fibroblasts. Glycolysis and oxidative phosphorylation appeared among the metabolic pathways most significantly changed. Higher rates of transcription on components mediating glucose consumption correlated with the up-regulation of enzymes involved in pyruvate metabolism and higher expression of biosynthetic precursors.

Even though, some pyruvate should be converted to lactate, since lactate dehydrogenase (*LDHA*) expression was highly enhanced, many genes involved in the synthesis of cysteine, methionine, alanine, aspartate and glutamate were found up-regulated. These amino acids may serve as substrates for the biosynthesis of proteins and nucleic acids, or generate oxidizable substrates in order to produce ATP in the mitochondria through oxidative phosphorylation. With regard to oxidative phosphorylation is worthwhile to mention the extensive up-regulation of components of the electron transport chain. In addition purine and pyrimidine metabolism are also found among the metabolic pathways differentially expressed.

The metabolic flux necessary for the synthesis of lipids is also partially represented in Table 18. Fatty acid metabolism and steroid biosynthesis, precursors of cholesterol, are necessarily activated in order to produce lipid membranes for daughter cells. Genes coding for lipogenic enzymes such as ATP citrate lyase (*ACL*) and fatty acid synthase (*FASN*) were found to be up-regulated. These key enzymes for fatty acid synthesis were intensively transcribed after c-MYC over-expression.

Table 18. Metabolic pathways differentially expressed in c-MYC transduced cells. This plot shows the different metabolic pathways that were found to be over-represented, when the genes listed for c-MYC as up-regulated in Table 17 were analysed. The first column shows the pathway analysed. The second column indicates the number of genes of each pathway found up-regulated. The last column is indicated the P -value. Pathways with a light grey background have P-value >0.05 and are considered therefore to be less significant. In parentheses are indicated the number of genes considered in each pathway.

Pathway	Count	P-value
Oxidative phosphorylation (207)	31	8,0E-10
Purine metabolism (245)	27	6,5E-06
Pyrimidine metabolism (154)	20	1,1E-05
Cysteine and methionine metabolism (65)	8	6,1E-03
Pyruvate metabolism (75)	8	1,5E-02
Valine, leucine and isoleucine degradation (61)	8	2,4E-02
Glycolysis / Gluconeogenesis (85)	9	4,3E-02
Fatty acid metabolism (49)	7	4,7E-02
Alanine, aspartate and glutamate metabolism (56)	6	5,1E-02
Biosynthesis of unsaturated fatty acids (29)	5	5,4E-02
Amino sugar and nucleotide sugar metabolism (111)	7	6,9E-02
One carbon pool by folate (27)	4	8,6E-02
Fatty acid elongation in mitochondria (19)	3	9,5E-02
Steroid biosynthesis (28)	4	9,9E-02

The results presented here support a multi-layered role for c-MYC. This transcription factor enhanced the transcription of many cell cycle components, especially those involved in the regulation of G_1 to S phase. In addition, it was able to increase the expression of enzymes essential for energy production and nucleotide biosynthesis necessary for the duplication of the genome. On the other hand, stress-related proteins and apoptotic signalling were clearly up-regulated.

5.2.4. Pathways differentially regulated after simultaneous over-expression of OCT4, SOX2, and KLF4

Surprisingly the pathways found differentially regulated after simultaneous over-expression of OCT4, SOX2, and KLF4 (3TF) recapitulated just partially the profiles found in the single factors (see Table 17). While apoptotic signalling recalled the profile observed following GFP over-expression, the key differences found after KLF4 over-expression were also present. Down-regulation of Ca^{2+}-related signalling was here noted as it was in OCT4 and SOX2 datasets. Simultaneously, regulation on the MAP kinase pathway and DNA replication resembled that found in cells with high levels of c-MYC. Yet a mechanism meant to repair DNA double-strand breaks, homologous recombination, and the Wnt signalling pathway appeared for the first time among the

pathways over-represented. Besides this, the TGFβ showed a distinctive profile not observed before after the over-expression of any of other factors.

Among the pathways that were recapitulated from the single transductions is apoptosis, which showed more intermediates of the extrinsic pathway (e.g. *AIFM1*, *DFFA*) than in GFP control cells, but included the inhibitory effect on *TP53* and *CASP3* observed in KLF4 over-expression (Figure 15B). Another pathway found differentially regulated can influence apoptosis, since high concentrations of intracellular Ca^{2+} can also trigger apoptosis (Berridge, Bootman et al. 2003). The Ca^{2+} signalling pathway, which controls the intake, storage and secretion of Ca^{2+}, can alter in addition the transcriptional regulation of the MAP kinase and it is dependent on the phosphatidylinositol signalling pathways (Berridge, Bootman et al. 2003).

DNA replication and repair mechanisms appeared differentially regulated (Figure 14). The cell cycle profile recalled the one found in c-MYC -over-expressing cells, with a large number, although not particularly strong, of up-regulated components of the cell cycle (Figure 16B). CDK inhibitors no longer showed large increases in expression or were down-regulated (e.g. *CDKN1A*, *CDKN1B* and *WEE1*). Replication-related repair mechanisms were no longer up-regulated as in GFP or c-MYC datasets; instead, homologous recombination components were up-regulated.

The Wnt signalling pathway, which controls basic developmental processes such as proliferation and differentiation of progenitor cells, or changes cell motility and adhesion features in more differentiated cell types (Capdevila and Izpisúa Belmonte 1999), showed many changes at the transcriptional level. The Wnt signalling pathway displayed a strong remodelling of genes coding secreted components and receptors (e.g. *WNT5A, FZD2*), as well as down-stream effectors transcription factors (e.g. *LEF1;* see in the appendix Table 25). It should be noticed that some of the targets activated by the Wnt signalling pathway depend on the binding of the complex form by SMAD3 and SMAD4 to β-catenin and its co-activators (see section 6.2).

One of the most surprising observations is the pattern observed in the TGFβ pathway because it showed remarkable differences to that of the single factors that made up the combination over-expressed. In fact, the profile in this pathway is more closely related to fibroblasts over-expressing c-MYC, as it was the cell cycle profile (see cluster on top of the Figure 15A). Because the TGFβ pathway can act up-stream of the cell cycle, and could explain at some point the differences found in cell cycle progression, more attention was paid to the regulatory events around the TGFβ pathway.

The combination of factors achieves a stronger up-regulation of the receptor inhibitors, *SMURF1/2*, than single factors, induced additionally the expression of inhibitory SMADs (e.g. *SMAD9*) (Figure 15A), and some transcription (*SMAD1* and

SMAD2). But the most striking differences come from the down-stream targets that were found highly expressed in the KLF4 dataset (e.g. IDs, *CDKN2B*), but practically did not change their expression in the 3TF dataset. Thus, while the expression of the different components of the pathway does not change drastically, the functional outcome of the pathway does.

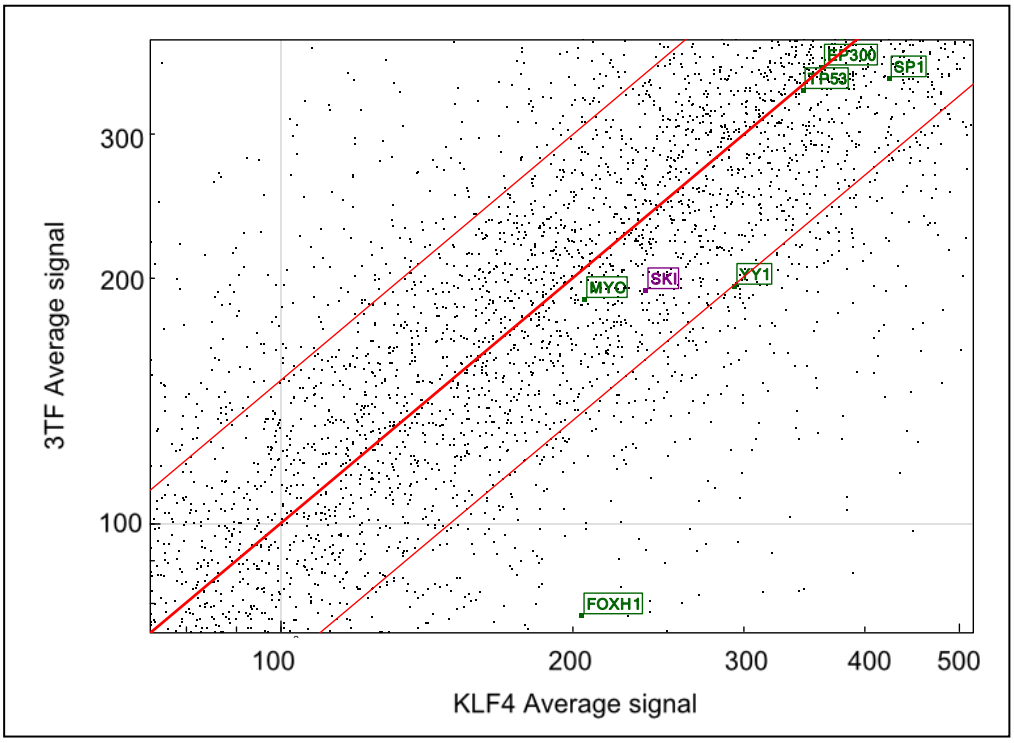

Figure 17. SMAD2/3 interacting partners in KLF4 and 3TF. The transcription factors that bind to promoters together with SMAD2/3 are labelled in green and inhibitors of SMADs in purple. The average intensity of the signals obtained in the microarray from three biological replicates of fibroblasts over-expressing KLF4 (X-axis) or the combination of 3 transcription factors (OCT4, SOX2, and KLF4; Y-axis) have been plotted against each other using BeadStudio. Red lines indicate up-regulation (ratio between differential score of 3TF versus KLF4 >1.5; top line), no change in regulation (ratio between differential score of 3TF versus KLF4 =1; middle line) and down-regulation (ratio between differential score of 3TF versus KLF4 <0.67; bottom line).

Expression of the cell cycle inhibitor, CDKN2B, is a known mechanism triggered by the TGFβ pathway to stop the cell cycle. SMADs act in combination of other transcription factors in order to activate CDKN2B transcription (Massague et al., 2000). Therefore, it is reasonable to think that changes in co-activators or co-repressors are responsible for the target shift found in the pathway. To find SMADs co-activators and co-repressors that are actually bound to promoters I used publicly available protein-protein interaction data. The STRING database (http://string-db.org/) has a large compendium of such datasets, but in addition combines genomic information, co-

expression data, and low- and high-scale experiments to retrieve known and predicted protein-protein interactions.

Thus, I analysed the changes in expression of the interacting partners retrieved by STRING. Figure 17 shows a comparison between the datasets for KLF4 and the three factor combination, where some of the SMAD-interacting partners have been highlighted. The SMAD-regulated target selection is known to be a very complicated process that requires specific partner transcription factors (Massague and Wotton 2000). Some of the most prominent DNA-binding partners in transcriptional activation processes, for example *FOXH1* (Chen, Rubock et al. 1996) is strongly down-regulated. The YY1 transcription factor, a protein that is involved in repressing and activating a diverse number of promoters (Gordon, Akopyan et al. 2006) is also down-regulated. Other DNA-binding partners of the SMADs, such as TCFs transcription factors belonging to the Wnt signalling pathway, were also down-regulated. Hence, the different landscape of transcription factors and co-binders in the nucleus could very well explain the different functional outcomes found in the TGFβ pathway.

5.2.5. Pathways differentially regulated after simultaneous over-expression of OCT4, SOX2, KLF4, and c-MYC

In human fibroblasts it is enough to over-express OCT4, SOX2, and KLF4 to reprogram the cells, but the efficiency at which this transformation happens is approximately 100-fold higher when c-MYC is added (Takahashi, Tanabe et al. 2007). This dataset should be able to highlight some essential aspects of the process and was therefore included in the analysis of the genetic regulatory network of the reprogramming process.

The regulatory signalling observed for 3TF was here reproduced practically to completion; even though, addition of c-MYC to that combination of factors set up many of the expression patterns found after the over-expression of c-MYC alone (see section 5.2.3). Intracellular processes regulated by c-MYC, such as DNA replication and cellular metabolism are significantly changed (Figure 14). The Wnt, TGFβ and MAP kinase signalling pathways that were found to be extensively remodelled after the over-expression of 3TF, were also regulated in the combination of four transcription factors. Likewise observed after 3TF over-expression was homologous recombination that appeared to be up-regulated again. Besides this, pathways important in development such as the Notch and ErbB were regulated.

Regarding DNA replication, the activation of cell cycle components induced by c-MYC, was also apparent here, when it was over-expressed in combination with the other factors. In addition, a down-regulation of CDK inhibitors (e.g. *CDKN1B* and *WEE1*) was

detected, as was observed after over-expression of 3TF. Again, activation of the replication machinery of the cell, as in c-MYC and 3TF datasets was noted, and correlated with an increase in the expression of proteins from the homologous recombination repair system (see Table 17 and Figure 16B).

The over-expression of four reprogramming factors brought about remodelling of the Wnt and TGFβ pathways very similar to that of the three factor combination (see section 5.2.4). As mentioned above, the regulation of the TGFβ signalling pathway was remarkably altered in most datasets (see Figure 15A). The deregulation of genes coding secreted proteins and receptors (e.g. *BMP2/4/5*, *TGFβ2*, *NODAL*, *GDF5, TGFR1/2*) was a general phenomenon that was accompanied by the down-regulation of key intracellular members of the cascade (e.g. *SMAD4*) and the up-regulation of inhibitors (e.g. *SMURF1/2*). Similarly the Wnt pathway showed a strong down-regulation of ligands and receptors (e.g. *WNT3*, *WNT5A, FZD2/6*), lower expression of transcription factors (e.g. *LEF1*), and reduced transcription of active constituents of the β-catenin degradation complex (e.g. CSNK1A1, CSNK1E; see in the appendix Table 26).

The MAP kinase pathway was extensively down-regulated. A profile already observed in the 3TF dataset, but readily enhanced in the 4TF dataset. Components of all the branches of the pathway, including stress-related kinases, counted lower expression values (e.g. *RRAS*, *RAF1*, *MAPK3/7/13*, *and JUND*). Just a common background to all datasets related to apoptosis remains with higher expression values (see section 5.2.1).

Genes from the apoptotic pathway were transcriptionally activated (e.g. *FAS*, *CYCS*), as in all datasets analysed so far, even though the combination of the four transcription factors showed a wider activation of the intrinsic apoptotic pathway (e.g. *DFFA*, *ENDOG*), as c-MYC did (Figure 15B). Interestingly, some of the effects of c-MYC on the apoptotic cascade were outweighed in the four factor combination, especially visible is the absence of regulation on the NF-κB pathway (e.g. *IL1B*, *IRAK1/2*, *NF-κB1*; Figure 15B).

As happened after the over-expression of c-MYC (Table 18), the transcription of many metabolic pathways were rearranged (Table 19). The addition of c-MYC to OCT4, SOX2, and KLF4, induced the expression of enzymes involved in nucleotide biosynthesis metabolism, including 5'-monophosphate dehydrogenase, serine hydroxymethyltransferase, adenosine kinase, adenylate kinase 2 and phosphoribosyl pyrophosphate amidotransferase. Furthermore, enhanced expression of glycolytic enzymes correlated with an increase in the expression of proteins involved in down-stream processes, such as pyruvate metabolism and oxidative phosphorylation. Some of the pyruvate produced during glucose oxidation might be used to synthesize several

non-essential amino acids, since their metabolic enzymes were found to be up-regulated.

Table 19. Metabolic pathways differentially expressed in 4TF transduced cells. This plot shows the different metabolic pathways that were found to be differently regulated, when the genes listed for 4TF as up-regulated in Table 17 were analysed. The first column shows the pathway analysed, the second column indicates the number of genes of each pathway found to be up-regulated, and the last column indicates the P-value. Pathways with a light grey background have P-value >0.05 and are considered therefore to be less reliable.

Pathway	Count	P-value
Purine metabolism (245)	35	7,3E-10
Pyrimidine metabolism (145)	26	3,7E-09
Cysteine and methionine metabolism (65)	10	4,1E-04
Lysine degradation (60)	10	2,9E-03
Pyruvate metabolism (75)	8	2,0E-02
One carbon pool by folate (27)	5	2,2E-02
Glycolysis / Gluconeogenesis (85)	10	2,3E-02
Valine, leucine and isoleucine degradation (61)	8	3,2E-02
Oxidative phosphorylation (207)	16	3,8E-02
Alanine, aspartate and glutamate metabolism (56)	6	6,2E-02
Glycine, serine and threonine metabolism (79)	6	6,2E-02
Propanoate metabolism (66)	6	7,0E-02
Ubiquinone and other terpenoid-quinone biosynthesis (41)	3	8,2E-02
Amino sugar and nucleotide sugar metabolism (111)	7	8,5E-02
Pentose phosphate pathway (52)	5	9,4E-02

The transcriptional profile of the cells following the over-expression of all four factors can be essentially explained by the combination of effects of the three factor combination and the metabolic changes induced by c-MYC. Hence, cells show high competence to replicate their DNA, by activating components that directly trigger progression from G_1 to S phase, but also the anabolic metabolism required for the synthesis of proteins, nucleic acids and lipids. In addition, cells did show less apoptotic signalling than expected and an increase availability of homologous recombination components.

5.3. Validation of microarray data by qRT-PCR

Microarray technology is based on short oligonucleotides attached to a glass surface, to which cDNA harbouring fluorescent tags is able to hybridize in a sequence-specific manner, thus allowing their relative quantification. Data acquired in such a way has an intrinsic level of noise due to target labelling, non-specific binding, and microarray processing among others, increasing the number of false positives. It is therefore

necessary to confirm the results that may contain biological significance by an alternative method.

To this end, I used qRT-PCR on cDNA synthesized out of the mRNA that was originally harvested to produce the samples hybridized onto microarrays (see section 4.4.3). Using Primer3 software primers were designed (see section 4.1.10) to amplify targets of the different pathways or processes that were classified among the significantly changed (Figure 14). The next subsections present the results, and are grouped according to signalling pathway or process.

5.3.1. p53 signalling pathway regulation after transduction

The first gene that I validated is *TP53* (the gene encoding for p53 protein), since its up-regulation has important consequences in cell cycle, apoptosis and DNA damage responses that were among the processes significantly changed in most datasets (Figure 14). p53 is tightly regulated by different proteins that modify it post-translationally and control its function (see section 5.4.1.1 for more details and references). MDM2 has ubiquitin ligase activity on the C-terminus of p53 and targets it for degradation. This activity of MDM2 is reduced however, if p19 is bound to MDM2, resulting in the overall stabilization of p53. Therefore, *MDM2* and *p19* expression were checked at the same time as *TP53* expression.

Up-regulation of *TP53* was found in all the samples that were hybridized onto microarrays. This can be partially attributed to the methodology, since even fibroblasts over-expressing GFP showed a 4.5-fold change (log_2 scale) in transcript levels than non-transduced fibroblasts (mock in Figure 18). Nonetheless, the levels of *TP53* in cells over-expressing c-MYC, as well as the combinations of factors (3TF and 4TF) were much more strongly induced, achieving a 6- to 8-fold change (log_2 scale). Conversely, KLF4-over-expressing fibroblasts displayed lower expression of *TP53* than fibroblasts over-expressing GFP.

Transcriptional activation of *MDM2* could also be detected in all samples, but it was most prominent after c-MYC over-expression, where an increase of around 2-fold change (log_2 scale) over the GFP control was observed. MDM2 can be transcriptionally activated by p53 binding to its promoter, creating, in this way a negative feedback loop (see section 5.4.1.1). In such a case, p53 is functional only transiently. This loop can be interfered by p19, which inhibits MDM2, and as a consequence stabilizes p53. This regulatory mechanism may be taking place, when OCT4, SOX2, and KLF4 (3TF) are simultaneously expressed, since the expression of *p19* was approximately twice than in GFP-expressing fibroblasts.

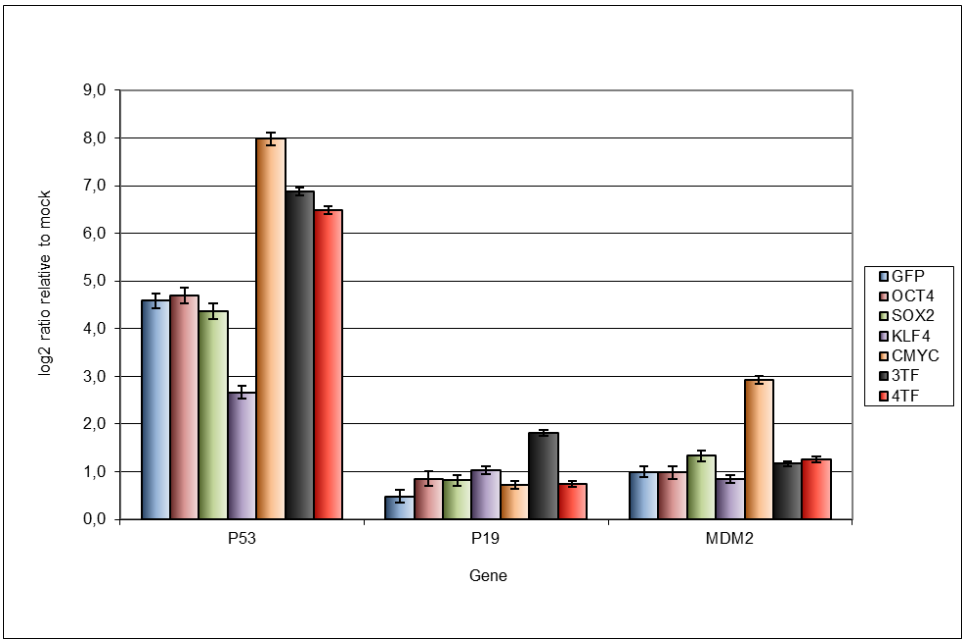

Figure 18. p53 regulation. Up-regulation of transcripts belonging to the p53 regulatory loop as measured by qRT-PCR. The amount of each transcript relative to non-transduced cells is on the Y-axis (logarithmic fold change). Along the X-axis are the gene targets under analysis. The legend shows the samples that have been analysed. All three biological replicates hybridized onto the microarrays have been analysed. The amounts shown are the average of three different technical replicates run in each of the three biological replicates. The error bars indicate the standard deviation of the average of all experimental replicates. Abbreviations: 3TF: OCT4, SOX2, and KLF4; 4TF: OCT4, SOX2, KLF4, and c-MYC.

Overall, a general up-regulation of *TP53* expression was detected in all samples, as expected from the analysis of commonly regulated pathways (see section 5.2.1). The most remarkable changes were found in cells over-expressing c-MYC and the combinations of factors (3TF and 4TF), where *TP53* was strongly induced. Notably, the over-expression of KLF4 reduced the levels of *TP53* below control levels.

5.3.2. Expression changes in apoptotic signalling

The microarray data indicated regulatory events in apoptosis common to all datasets. An over-expression of proteins can induce abnormal cellular responses in the cells and trigger apoptosis. The activation of pro-apoptotic proteins, or the reduction of survival proteins that hold back the otherwise unavoidable degradation of cellular structures, can reduce the efficiency of reprogramming.

Therefore, qRT-PCR validation of some of the anti-apoptotic (*BCL2*, *BCLXL*) and pro-apoptotic markers (*BAD*, *BAX*) that modulates the permeability of the mitochondrial membrane, and which were found to be deregulated in the microarray data, was performed. Other components of the intrinsic pathway (*AIF*) that can modulate the

release of pro-apoptotic proteins from the mitochondria were also checked, as well as, some of the inhibitors of apoptosis (*BIRC5*). The actual proteolytic cleavage effectors, namely caspases (*CASP3*, *CASP7*), and an apoptosome component (*CASP9*) that participates in the activation of the caspases were additionally tested (see section 5.4.1.2 for more details and references about the apoptotic cascade).

Pro-apoptotic markers such as *BAD* and *BAX* roughly doubled their expression in the GFP control, as well as in OCT4, SOX2, and c-MYC (Figure 19). Strikingly, KLF4, 3TF, and 4TF samples showed reduced levels of transcripts for both proteins. Microarray results for other pro-apoptotic markers could be confirmed, such as the up-regulation of *AIF* that increased its expression around 3- to 5-fold change (log_2 scale) in all samples analysed. Nonetheless, 3TF samples showed levels below the GFP control, while c-MYC-containing samples had higher expression levels. This peptide can induce the release of cytochrome c from the mitochondria, triggering the activation of the caspase–cascade (see section 5.4.1.2).

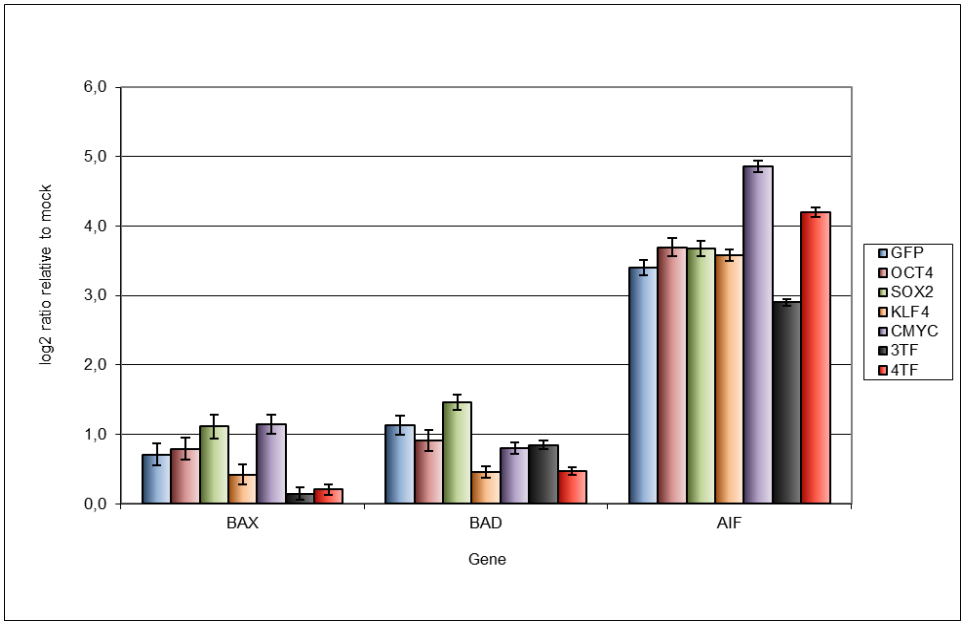

Figure 19. Pro-apoptotic genes. Regulation of transcripts belonging to the apoptotic pathway as measured by qRT-PCR. The relative amount of each transcript relative to non-transduced cells is scaled along Y-axis (logarithmic fold change). Along the X-axis are the gene targets under analysis. The legend shows the samples that have been analysed. All three biological replicates hybridized onto the microarrays were analysed. The error bars indicate the standard deviation of the average of all experimental replicates. Abbreviations: 3TF: OCT4, SOX2, and KLF4. 4TF: OCT4, SOX2, KLF4, and c-MYC.

Interestingly, down-stream caspases such as *CASP3*, *CASP7* and *CASP9* were not commonly up-regulated, showing more often similar or lower expression levels than GFP-expressing cells. Only in fibroblasts over-expressing c-MYC (Figure 20) their

transcription increased. The expression of these caspases recapitulates partially the results found for *BAD* and *BAX,* where KLF4 and the combinations of factors showed the lowest levels of expression.

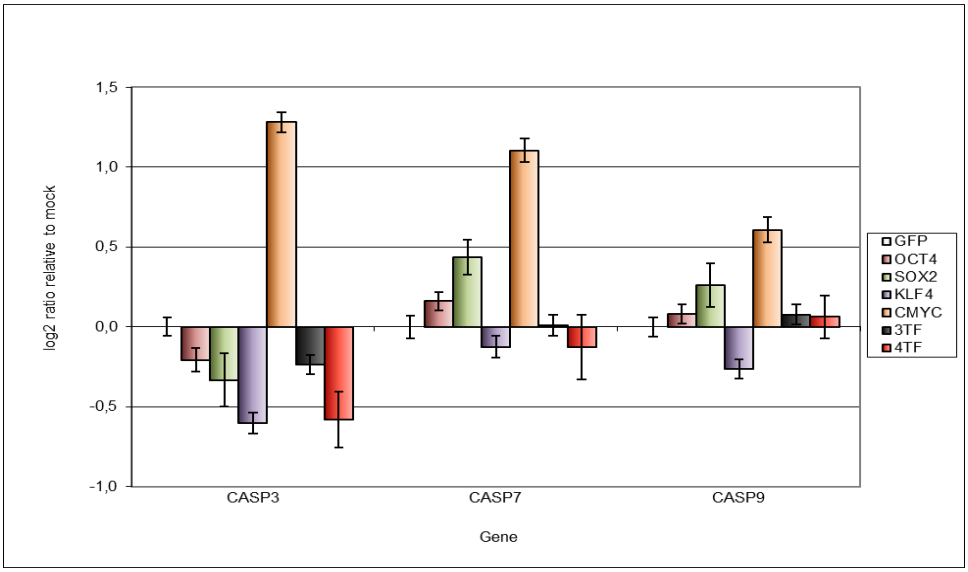

Figure 20. Caspases. Regulation of transcripts belonging to the caspase cascade as measured by qRT-PCR. The relative amount of each transcript relative to non-transduced cells is scaled along Y-axis (logarithmic fold change). Along the X-axis are the gene targets under analysis. All three biological replicates hybridized onto the microarrays were analysed. The amounts shown are the average of three different technical replicates run in each biological replicate. The error bars indicate the standard deviation of the average of all experimental replicates. Abbreviations: 3TF: OCT4, SOX2, and KLF4; 4TF: OCT4, SOX2, KLF4, and c-MYC.

Inhibitors of apoptosis such as *BIRC5* (IAP family of proteins) or pro-survival proteins can stop the apoptotic cascade (see section 5.4.1.2). *BIRC5* was found to be up-regulated in all samples including GFP (Figure 21), where achieved a 1.5-fold change (log_2 scale). It is notable the strong up-regulation found in c-MYC, 3TF, and 4TF samples with 3- to 4-fold change (log_2 scale) in expression. Transcripts for pro-survival proteins such as *BCL2* (Figure 21) that can counteract the effect of *BAD* or *BAX* showed expression levels 1- to 2-fold (log_2 scale) higher than GFP after over-expression of SOX2, c-MYC and 4TF. On the other hand, in OCT4 and 4TF samples were detected just half of the *BCLXL* transcripts, another survival protein, that were detected in the GFP control.

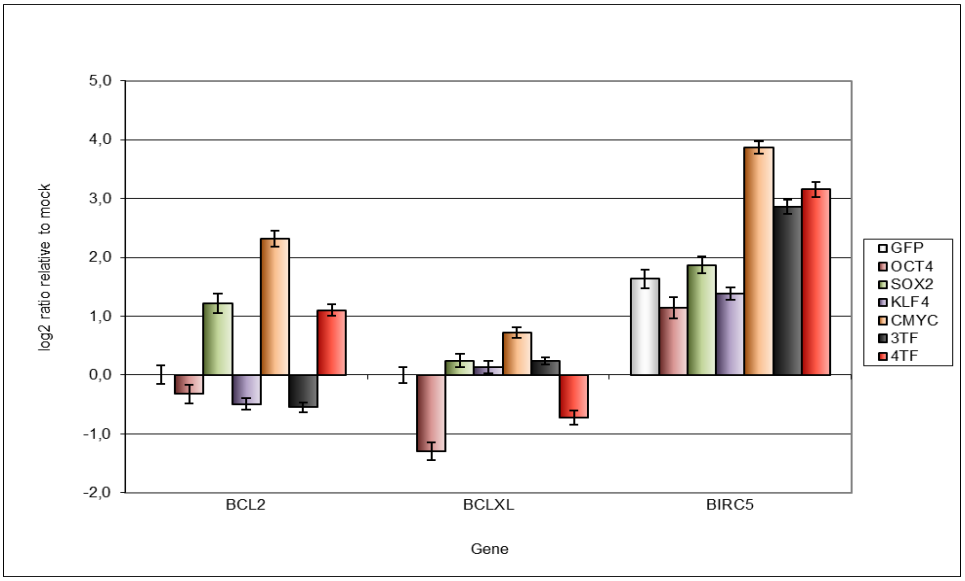

Figure 21. Pro-survival genes. Regulation of pro-survival genes as measured by qRT-PCR. The relative amount of each transcript relative to non-transduced cells is scaled along Y-axis (logarithmic fold change). Along the X-axis are the gene targets under analysis. The legend shows the samples that have been analysed. The amounts shown are the average of three different technical replicates run in each of the three biological replicates. The error bars indicate the standard deviation of the average of all experimental replicates. Abbreviations: 3TF: OCT4, SOX2, and KLF4; 4TF: OCT4, SOX2, KLF4, and c-MYC.

This data confirmed the overall change in apoptosis suggested by the pathway analysis. The microarray showed changes in most samples, including the GFP control, in the intrinsic pathway but no large activation of down-stream cascades, which is also observed here (see section 5.2.1). The over-expression of c-MYC showed up-regulation of pro- and anti-apoptotic proteins. Conversely, KLF4 and the combination of factors showed consistently the lowest levels of pro-apoptotic markers, especially remarkable in 4TF, since c-MYC is expected to up-regulate them. Furthermore, the combination of factors OCT4, SOX2, and KLF4, showed a higher increase in *BIRC5* expression than any of the factors individually. Also the four factor combination, OCT4, SOX2, KLF4, and c-MYC, showed a large increase in *BIRC5* expression, but in addition up-regulation of *BCL2* as observed in SOX2 or c-MYC over-expressing cells.

5.3.3. Expression changes in cell cycle components

The larger amount of differentially expressed genes in a dataset that belonged to a single process relate to cell cycle. Furthermore, genes involved in the replication of DNA (G_1-S phase), such as Cyclins and CDKs, increased their expression after GFP, c-MYC, 3TF and 4TF over-expression. Thus, transcription of genes involved in the G_1 and S phase of the cell cycle were analysed by qRT-PCR (Figure 22).

In the cluster analysis shown in Figure 12B the GFP-, OCT4-, SOX2- and KLF4-expressing cells clustered together, while c-MYC-over-expressing cells clustered closer to the cells that expressed the combination of factors (top dendogram). While it is expected for c-MYC, it is not an obvious effect of the combined expression of the reprogramming factors. Therefore, I paid special attention to the combination of factors and validated the expression of cell cycle components on these samples.

As observed in the microarray data (Figure 16B) C*DK4* and *CDK6* and their interaction partners *Cyclin D1*, *D2* and *D3*, which drive the first growth phase after mitosis, are up-regulated in comparison to non-transduced fibroblasts. Also *CDK2*, *Cyclin E1* and *Cyclin A2* that prompt the replication of DNA (synthesis phase) were found to be up-regulated (for more details and references on cell cycle see section 5.4.1.3).

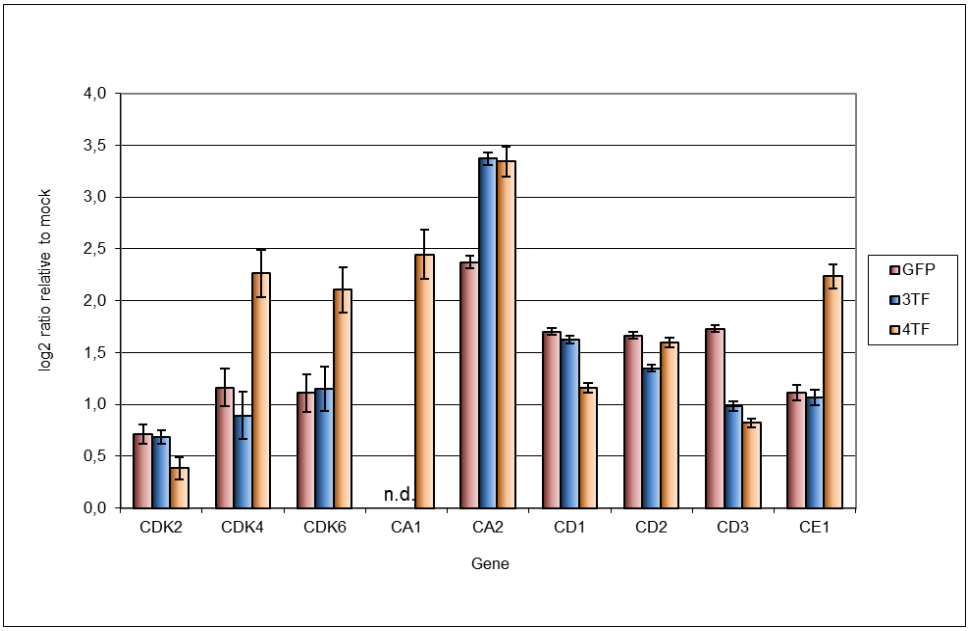

Figure 22. Cell cycle regulation. Regulation of transcripts belonging to the cell cycle as measured by qRT-PCR. The relative amount of each transcript relative to non-transduced cells is scaled along Y-axis (logarithmic fold change). Along the X-axis are the gene targets under analysis. The legend shows the samples that have been analysed. All three biological replicates hybridized onto the microarrays were analysed. The amounts shown are the average of three different technical replicates run in each biological replicate. The error bars indicate the standard deviation of the average of all experimental replicates. Abbreviations: 3TF: OCT4, SOX2, and KLF4; 4TF: OCT4, SOX2, KLF4, and c-MYC; CA1/2: *Cyclin A1/2*; CD1/2/3: *Cyclin D1/2/3*; CE1: *Cyclin E1*.

The levels of transcription for *CDK2/4/6*, *Cyclin D1/2* (CD1 and CD2 respectively in Figure 22) were similar between control cells (GFP) and 3TF, ranging a 0.5- to 1.5-fold change (log_2 scale). Hence, the three factor combination shows a close profile to GFP-expressing fibroblasts, being the most remarkable difference the higher up-regulation of *Cyclin A2* (CA2 in Figure 22*),* which triggers DNA replication as well as co-operates in

driving the cells through S-phase to the following growth phase (G_2) (see section 5.4.1.3). The four factors combination shows the effects expected on the cell cycle after c-MYC over-expression more prominently. *Cyclins* and *CDKs* involved in driving G_1-S, such as *CDK4/6*, *Cyclin A2* and *Cyclin E1* (this latter is abbreviated as CE1 in Figure 22), are transcriptionally up-regulated above GFP control levels, often doubling the amount of transcripts and reaching 2- to 3.5-fold change (log_2 scale). Also the germ-cell specific *Cyclin A1* (CA1 in Figure 22) was drastically augmented.

Cell cycle is a tightly regulated process and somatic cell proliferation is usually restrained under certain circumstances (e.g. DNA damage) by cell cycle inhibitors that induce senescence (see section 5.4.1.3). Since *p16* and *p21* (also named *CDKN2A* and *CDKN1A*) were listed as up-regulated genes in the initial analysis of the microarray data, their transcripts were also monitored by qRT-PCR.

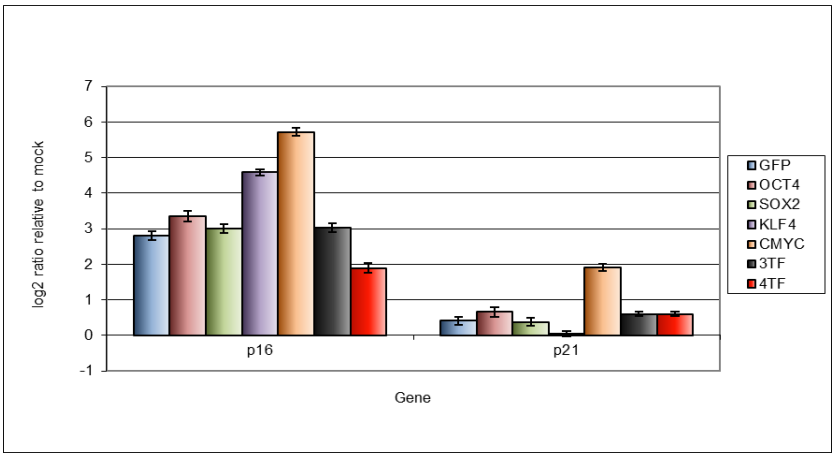

Figure 23. Cell cycle inhibitors. Transcriptional regulation of proteins that inhibit the cell cycle measured by qRT-PCR. The relative amount of each transcript relative to non-transduced cells is scaled along Y-axis (logarithmic fold change). Along the X-axis are the gene targets under analysis. The legend shows the samples that have been analysed. All three biological replicates hybridized onto the microarrays were analysed. The amounts shown are the average of three different technical replicates run in each biological replicate. The error bars indicate the standard deviation of the average of all experimental replicates. Abbreviations: 3TF: OCT4, SOX2, and KLF4; 4TF: OCT4, SOX2, KLF4, and c-MYC.

Indeed, up-regulation of the *p16* transcript was observed in all tested samples (Figure 23), even though in most cases achieved similar levels than the GFP control, being 3-fold (log_2 scale) higher than non-transduced fibroblasts. Notably, c-MYC and KLF4 samples showed the highest levels of *p16* transcript, while the combination of four factors had the lowest levels. Also, *p21* showed up-regulation in fibroblasts expressing c-MYC (Figure 23), were the relative amount of transcript to non-transduced fibroblasts

was 2-fold (log_2 scale) higher, while GFP-expressing cells did not achieved a 0.5-fold change (log_2 scale).

This data supports the up-regulation of G_1-S phase proteins after the over-expression of GFP and the three factor combination, further boosted by c-MYC, when it was added to the reprogramming cocktail. On the other hand, at least some of the cells among the transduced fibroblasts expressed cell cycle inhibitors and the most striking differences to non-transduced fibroblasts were found in KLF4 and c-MYC samples (see sections 5.2.2 and 5.2.3).

5.3.4. TGFβ signalling pathway regulation after transduction

The TGFβ pathway expression profile showed key differences between the different datasets (see section 5.2.4). This signalling pathway has an important role in both differentiation signalling and antigrowth cascades that can induce senescence and apoptosis in cells (for more details and references see section 5.4.1.4). Consequently, the next validation efforts concentrated on the expression profile of selected genes belonging to the TGFβ signalling pathway.

Among the relevant components of the pathway are the TGFβ receptors that participate in the transduction of extra- or intracellular cues, phosphorylating and activating SMAD2/3 and/or kinase cascades of the MAP kinase signalling pathway (see section 5.4.1.4). There are also known inhibitors of the receptors as SMAD-specific E3 ubiquitin-protein ligase 1 and 2 (SMURF1/2). The microarray datasets showed large changes in the expression of *TGFβR2*, *SMAD3* and *SMURF2*, depending on the sample considered, those genes were selected for qRT-PCR validation.

Thus, GFP over-expression caused already a down-regulation of the *TGFβR2,* which was recapitulated and enhanced in OCT4- and KLF4-expressing fibroblasts (Figure 24). Instead, SOX2, c-MYC and the combination of factors showed levels comparable to non-transduced fibroblasts. The expression of the TGFβ receptor inhibitor *SMURF2* was up-regulated in most of the samples, even though SOX2, KLF4, and 3TF showed levels comparable to the GFP control. Only c-MYC samples showed larger up-regulations than GFP samples, being *SMURF2* expression around 2-fold (log_2 scale) higher. Only in OCT4 and 4TF samples decrease the expression of these inhibitors in comparison to the GFP control.

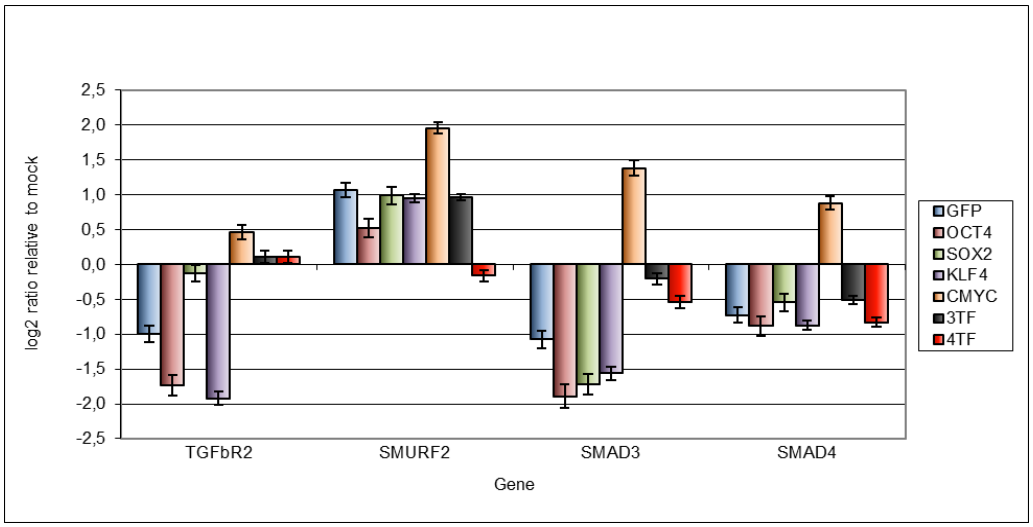

Figure 24. TGFβ signalling pathway. Transcriptional regulation of proteins that inhibit the cell cycle measured by qRT-PCR. The relative amount of each transcript relative to non-transduced cells is scaled along Y-axis (logarithmic fold change). Along the X-axis are the gene targets under analysis. The legend shows the samples that have been analysed. All three biological replicates hybridized onto the microarrays were analysed. The amounts shown are the average of three different technical replicates run in each biological replicate. The error bars indicate the standard deviation of the average of all experimental replicates. Embryonic stem cell mRNA (H9) was used as a positive control for the primers. Abbreviations: 3TF: OCT4, SOX2, and KLF4; 4TF: OCT4, SOX2, KLF4, and c-MYC.

Intracellular components like *SMAD3* and *SMAD4* were consistently down-regulated in most of the samples including the GFP control (Figure 24). *SMAD3* showed stronger down-regulations, ranging between 1- to 2-fold (log_2 scale) less than non-transduced cells. Only after over-expression of c-MYC higher level of transcript of these transcription factors were observed, being *SMAD3* expression for instance 2-fold (log_2 scale) higher than in GFP-expressing cells. Just the combination of factors did not show remarkable changes in *SMAD3* expression.

Hence, the TGFβ signalling pathway was indeed differently regulated and showed diverse expression patterns depending on the factor that was over-expressed, but most remarkably the combination of factors did show a different regulation of the pathway, not observed in any of the factors making up the combinations. While OCT4 and KLF4 showed a consisted down-regulation of genes participating in signal transduction, cells over-expressing c-MYC showed higher transcriptional activity for all genes within the pathway, contrasting with the combination of factors that displayed expression levels similar to non-transduced fibroblasts or the GFP control. Thus, these results agree with the analysis presented above (see section 5.2.4), where main differences in the expression profile of this pathway between any of the single factors and the combination of factors was noted.

5.3.5. DNA damage response – Homologous recombination

Many of the processes by which cells counteract DNA damages—mismatch repair, nucleotide excision repair, base excision repair, or homologous recombination (HR)—were found among the processes differentially expressed (Figure 14). The analysis of the homologous repair system was prioritized here, since its regulation was more specific by the simultaneous over-expression of the transcription factors and those combinations are the ones that achieved the reprogramming of the cells.

HR is activated by somatic cells to repair DNA lesions that can compromise genomic integrity and fidelity (see section 5.4.1.5 for details and references). The activation of homologous recombination can have important implications for the cell, since it can trigger p53-dependent and -independent antigrowth mechanisms. In order to repair breaks in the DNA, proteins are expressed that are assembled surrounding the DNA lesion, like BRCA1/2 and NBN. Simultaneously, proteins such as RAD1 or GADD45α are synthesized to stop cell cycle progression and allow time for repair (more details in section 5.4.1.5).

In Figure 25 results of the qRT-PCR reactions that were carried out to check the expression of the above mentioned components of the HR repair system have been plotted. The GFP control showed a general up-regulation of homologous recombination components, found also in single factor over-expressions, where fold-changes were most of the time slightly reduced. Only c-MYC over-expressing cells displayed a higher up-regulation in all genes analysed. The combinations of factors had the closest profile to c-MYC samples, with a comparable activation of well-known markers like *BRCA1* and *BRCA2*. For instance, *BRCA2* expression was 2-fold ($\log_2$ scale) higher than in non-transduced fibroblasts, while achieved a 4-fold ($\log_2$ scale) change in c-MYC samples. Similar fold changes were found in all samples for *GADD45α*, varying slightly depending on the identity of the factor(s) expressed.

Remarkably, *Nibrin* (*NBN*) was not up-regulated in GFP- and SOX2-expressing fibroblasts, but down-regulated in the other samples considered, except for c-MYC cells. Alternatively, *RAD1* doubled its transcription in control cells (GFP), which was recapitulated after single factor over-expression, contrasting with 3TF and 4TF over-expression, where no difference to non-transduced fibroblasts was found.

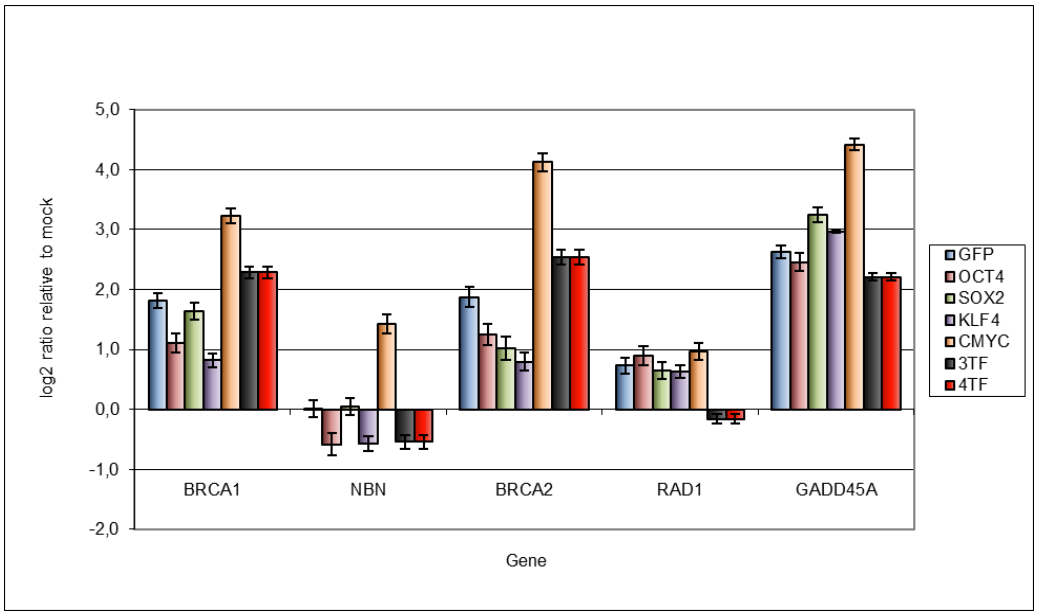

Figure 25. Homologous recombination. Transcriptional regulation of proteins that participate in the repair of DNA breaks measured by qRT-PCR. The relative amount of each transcript relative to non-transduced cells is scaled along Y-axis (logarithmic fold change). Along the X-axis are the gene targets under analysis. The legend shows the samples that have been analysed. All three biological replicates hybridized onto the microarrays were analysed. The amounts shown are the average of three different technical replicates run in each biological replicate. The error bars indicate the standard deviation of the average of all experimental replicates. Embryonic stem cell mRNA (H9) was used as a positive control for the primers. Abbreviations: 3TF: OCT4, SOX2, and KLF4; 4TF: OCT4, SOX2, KLF4, and c-MYC.

Thus, homologous recombination components were indeed regulated. While not all the components showed strong regulation, well known markers did (*BRCA1/2*), being above GFP-expressing cells for c-MYC and the combination of factors, and below control cells (GFP) for OCT4, SOX2 and KLF4.

5.3.6. Comparison between qRT-PCR and microarray results

As pointed out above (see section 5.3) microarray experiments have a certain level of noise that produces false positives. In order to estimate the overall reproducibility of the microarray data the results obtained in the microarray and the qRT-PCR have been plotted together. In Figure 26 can be observed the results of the qRT-PCR presented in the last sections represented in the x-axis against the ratios obtained in the microarray experiment that have been laid down in the y-axis. In both cases the expression of the gene-of-interest is normalized against non-transduced fibroblasts and the ratio represent it as logarithmic fold-change. The ratio thresholds used for the analysis of the microarray data in order to discriminate up- and down-regulation has been also depicted in the plot as red lines. The points in the plot are depicted as dots if the microarray data was considered high confident (P-value <0.01).

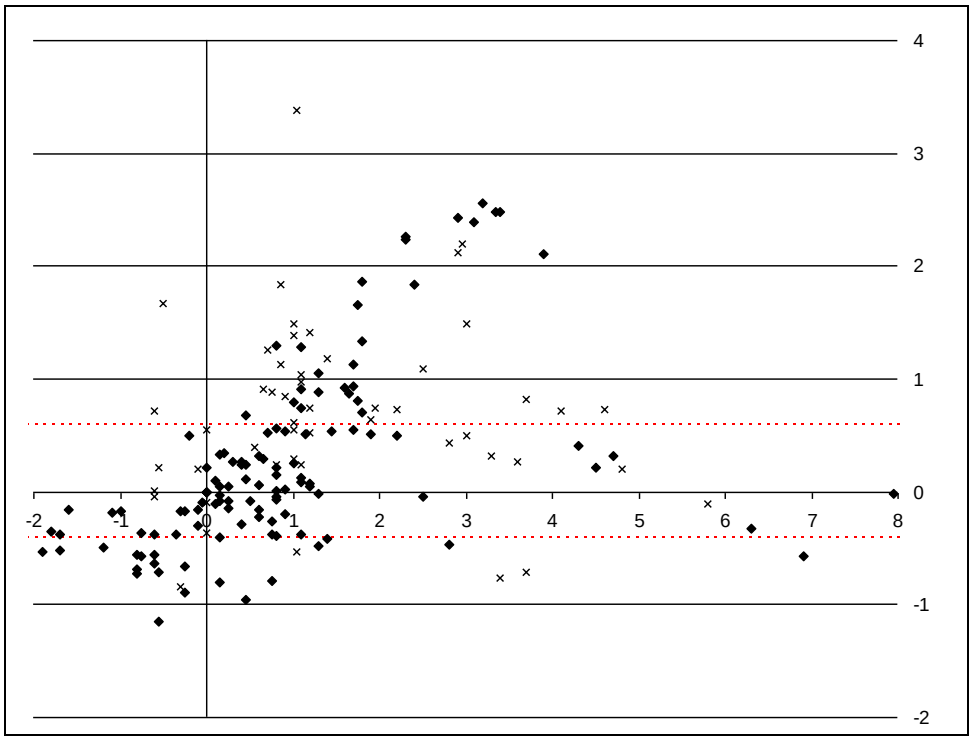

Figure 26. Comparison between RT-PCR and microarray. The results obtained with qRT-PCR (X-axis) have been plotted against the results obtained with the microarray (Y-axis). The red lines indicate the threshold used to filter the microarray data; thus, genes considered up-regulated were those found to have a signal intensity ratio (factor/mock) >1.5 (upper red line), and those listed as down-regulated had a signal intensity ratio (factor/mock) <0.67. Microarray signals with high confident values (P-value <0.01) are depicted as dots, otherwise as crosses (P-value >0.01).

The relative changes were consistently lower in the microarray than in RT-PCR. Nevertheless, in more than 72% of the cases the qRT-PCR correlates positively with the results of the microarray. If thresholds and confident values used to filter the microarray data are considered 82% of the results correlate positively (dots out of the interval defined by red lines in Figure 26). The thresholds used to filter the microarray data discriminate then most of the signals that could not be validated by qRT-PCR. Thus, the lists of differentially expressed genes obtained with those filtering criterias have high reliability. In fact the pathway analysis was based on these lists, and each of the pathways validated was indeed differentially regulated.

5.4. Modelling

Even though, pathway analysis can highlight cellular processes and signalling pathways that have been affected after certain experiments, it is not always enough to determine their functional relevance. Up-regulation or down-regulation of one or more genes does not let predict the implications for a certain cellular function, unless they are

integrated in the biochemical processes taking place in the cells. In addition, different signalling pathways are very often acting co-ordinately to regulate the same processes, which makes still more complicated to understand how the cell will respond to many simultaneous changes. In the last years, a great effort has been made to create bioinformatic tools able to face the complexity of simultaneous changes in complex regulatory networks and describe their spatial and temporal behaviour (see section 3.4) that can be now helpful to understand complex regulatory events.

Thus, the next step was to assemble a mathematical model of regulatory signalling based on the results of the pathway analysis previously generated. The aim is to understand the different molecular mechanisms that trigger these responses and be able to rationalize their interdependence. The model should be able to describe the functional meaning of the different transcriptional changes found and should help to propose hypothesis about the molecular mechanisms operative in the cells after the over-expression of the reprogramming factors.

5.4.1. Assembly of the model

A scheme of the most important features contained in the model is shown in Figure 27. The model was started combining the MAP kinase signalling pathway and the cell cycle, which establishes a regulatory framework for proliferation (Yang, Sharrocks et al. 2003). The model has extrinsic signalling that activates the different pathways (e.g. growth factors). The outcome of the simulation will yield qualitative information about the accumulation of down-stream effectors of each pathway that have been linked to symbolic tags (e.g. mitosis, apoptosis) in order to display generally the activity of each pathway. Thus, the model is started by growth factors that activate the MAP kinase signalling pathway, which promotes cell cycle progression. The cell cycle is divided in the four different stages corresponding to the G_1-phase that ends up with synthesis of S-phase proteins, S-phase when the actual replication of DNA occurs, a second growth phase, namely G_2, and mitosis.

Thereafter, pathways linked before to anti-growth mechanisms were included, such as the p53 pathway (Vogelstein, Lane et al. 2000), the TGFβ signalling pathway (Massague and Wotton 2000) or DNA damage response (Ciccia and Elledge 2010). The TGFβ signalling pathway is induced after replication stress and can stop the cell cycle by virtue of cell cycle inhibitors, as well as participate in the activation of p53 (Massague and Wotton 2000; Wagner and Nebreda 2009). On the other hand, the DNA damage checkpoint preceding mitosis can initiate signal transduction cascades among others the p53 signalling pathway that prompt cells to stop proliferating or to die (Harper and

Elledge 2007; Nagata 1999). The p53 signalling pathway is therefore a common node of different surveillance mechanisms in the signalling network considered here.

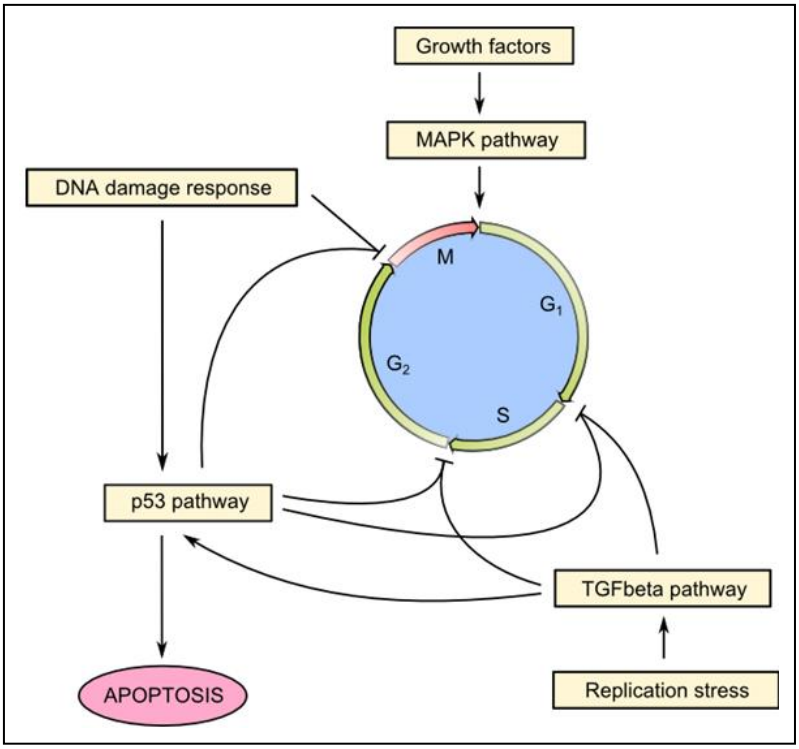

Figure 27. Kinetic model. Overview of the different features added in the model. Growth factors and the MAPK signalling pathway trigger the cell cycle that has been represented as a circle divided in four different phases. The cell cycle checkpoints are linked to different pathways through lines ending with perpendicular segments. Arrows indicate activation events.

The regulatory information available for these pathways in curated databases (e.g.: KEGG pathways; Reactome) and pertinent literature were introduced in the simulation tool PyBios (Wierling, Herwig et al. 2007; Klipp, Liebermeister et al. 2009) to end up with a model of 164 components and 304 reactions. As kinetic parameters remain unknown, we applied a generalized approach for the construction of the transcriptional kinetics (Kühn, Wierling et al. 2009). PyBios can then solve the set of ordinary differential equations (ODE) and calculate the behaviour of the system using a Monte Carlo approach (see section 4.6). In this way, up- and down-regulation of the different components of the network in response to a perturbation experiments can be simulated with randomized parameters.

5.4.1.1. p53 signalling pathway

The transcription factor p53 is involved in development, differentiation and response to cellular stress. It is activated to diverse cellular stresses and regulates target

genes that induce cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism (Vogelstein, Lane 2000). In Figure 28 is represented the regulatory signalling concerning p53, while its down-stream targets appear in the apoptosis and cell cycle sections (see section 5.4.1.2 and 5.4.1.3). The cascade of events starts upon DNA damage ataxia telangiectasia mutated protein (ATM) and/or DNA-dependent protein kinase (DNA-PK) phosphorylate N-terminal serines or threonines stabilizing active p53 ("p53_active" in Figure 28) (Savic, Yin et al. 2009). Other known activators of p53 are p38 and JNK, which are stress-related MAP kinases that can be activated by TGF receptors (Wagner and Nebreda 2009), have been also included in the model as activators of p53.

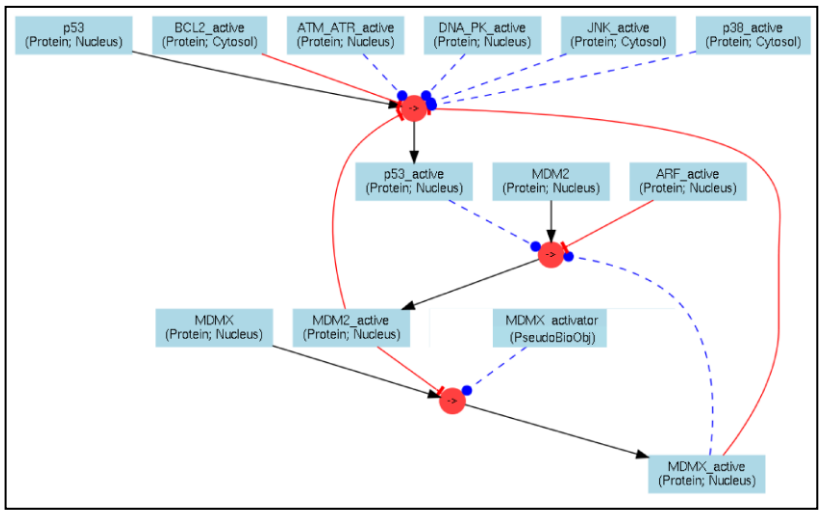

Figure 28. Scheme of p53 regulation. Illustration of the p53 signalling pathway as it has been implemented in PyBios. The blue boxes represent components of the model, while the red dots represent reactions. In the blue boxes is written the name of the component, the type of molecule it is and its location. Components added for convenience is labelled as pseudobioobject ("PseudoBioObj"). Components are linked to reactions by lines that are black if they are reactants of the reaction, blue if they are acting as activators and red if they are acting as repressors in the reaction. Basal synthesis and degradation reactions for each of the components have not been represented to maintain the scheme simple.

Accumulation of p53 in the mitochondria is known to be a key event in the induction of p53-dependent apoptosis (Mihara, Erster et al. 2003). The association of BCL2 to p53 in the membrane of the mitochondria prevents cell death (Mihara, Erster et al. 2003), and therefore BCL2 has been implemented in the model as negative regulator of p53 that mimicks this proto-oncogenic property of BCL2.

The levels of functional p53 are tightly regulated in the cells. Ubiquitination of the protein by MDM2 or binding of MDMX to its transcriptional activation domain are known to unstabilize p53 that is targeted to the proteosome if other compensatory modifications

are not present (Toledo and Wahl 2006). Even though the nature of their inhibitory effect on p53 is different, both interactions have been considered in the model as inhibition of p53 activation. Besides this, p53 activates transcriptionally *MDM2* ("MDM2" to "MDM2_active") forming a negative feedback loop that has been added in the model (Vogelstein, Lane et al. 2000). In addition, MDMX can interact with MDM2, avoiding its degradation and has been implemented as activator in MDM2 activation (Toledo and Wahl 2006). Unfortunately, MDMX mode of activation has not been characterized and it was necessary to introduce a figurative activator ("MDMX_activator"). Finally, oncogene over-expression is known to induce the transcription of *ARF* (alternative reading frame) that inhibits MDM2, resulting in p53 stabilization (Vogelstein, Lane et al. 2000).

Activated p53 can induce the expression of proteins that block the cell cycle, like p21, the 14-3-3 adaptor proteins or GADD45α. p21 achieves to bind and inactivate many different complexes important at different stages of the cell cycle (CDK4/6-Cyclin D and CDK2-Cyclin A), while 14-3-3 and GADD45α can inhibit the binding of key components (e.g. Cdc25C; CDK1), impairing their function (Pietenpol and Stewart 2002; Rosemary Siafakas and Richardson 2009).

Different mechanisms are simultaneously triggered by p53 in order to induce apoptosis (Slee, O'Connor et al. 2004). p53 can activate the transcription of secreted proteins such as death ligands (e.g. FAS ligand), and simultaneously repress the promoter of *BCL2*, a survival gene. p53 activates pro-apoptotic factors such as *BAX*, *BAK*, *NOXA* or *PUMA*, which facilitate the release of cytochrom c and apoptosis-inducing factors from the mitochondria (Slee, O'Connor et al. 2004). All these interactions have been included in the apoptosis section below (see section 5.4.1.2).

5.4.1.2. Apoptosis

Cell death can occur after signalling molecules, like cytokines or hormones, bind to receptors, or after certain stresses as nutrient deprivation or radiation, leading to the activation of extrinsic and/or intrinsic signalling pathways (Yuan, Lipinski et al. 2003). The extrinsic apoptotic cascade is initiated in the model by interaction of an extracellular death ligand such as FASL (labelled as "FasL_ECR" in Figure 28) with a receptor, to which binds an adaptor protein (FADD) forming a complex that is able to anchor and activate precursors of initiator caspases such as CASP8 and CASP10 (Nagata 1999). Once activated by restriction these cysteine proteases can induce the translocation of Bid into the mitochondria, which is displayed in the model as an activation of Bid resulting in "Bid_active". This leads, as the accumulation of BAX or BAK proteins, to the permeabilization of the mitochondrial membrane and the release of cytochrome c from

the mitochondria (“CytC_Mit”) into the cytosol (“CytC_Mit”) (Samali, Zhivotovsky et al. 1999).

If cytochrome c is released from mitochondria, it binds to apoptotic protease activating factor-1 (APAF-1), which can then interact and activate pro-CASPASE-9 to create a protein complex known as an apoptosome (“APAF1_CASP9” in Figure 29). The apoptosome activates by proteolytic cleavage effector caspases (e.g. CASP3/6/7) (Samali, Zhivotovsky et al. 1999). These will begin with the degradation of endogenous protein structures (e.g. ACTIN, LAMIN) and induction of the DNA fragmentation by endonucleases (degradation of inhibitor of caspase-activated DNase (ICAD)) (Thornberry and Lazebnik 1998). The activation of CASP3 (“CASP3_active”) is irreversible and therefore has been linked in the model to a representative reaction (“Survival” to “Apoptosis”) that summarizes the activation of this cascade.

Other mechanisms are known in which effector caspases are activated by CASP8 and CASP10 independently of mitochondrial pathways and end up in the activation of DFF45, the degradation of the genome, and finally cell death (Omata, Suzuki et al. 2008). To distinguish these molecular events the activation of DFF45 has been connected to a symbolic reaction with the outcome of “Degradation_ON”. Thus, if “Degradation_ON” increases and “Apoptosis” does not cell death is occurring independently of the integrity of the mitochondria.

In addition, activation of death receptors can also trigger pro-survival signalling through the NF-kB pathway (Kucharczak, Simmons et al. 2003). The NF-kB pathway plays a critical role in infectious processes and coordinates immune responses (Ghosh and Karin 2002). Between the genes regulated by NF-kB are included BCL2, BCLXL, and members of the IAP family of proteins. The latter proteins are characterized by the presence of a homologous domain named the baculoviral IAP repeat (BIR) domain that can directly bind to and inhibit caspase function (Deveraux, Stennicke et al. 1999). In the model are represented by “IAP” that once activated (“IAP_active”) inhibits all effector caspases, acting as a pro-survival protein.

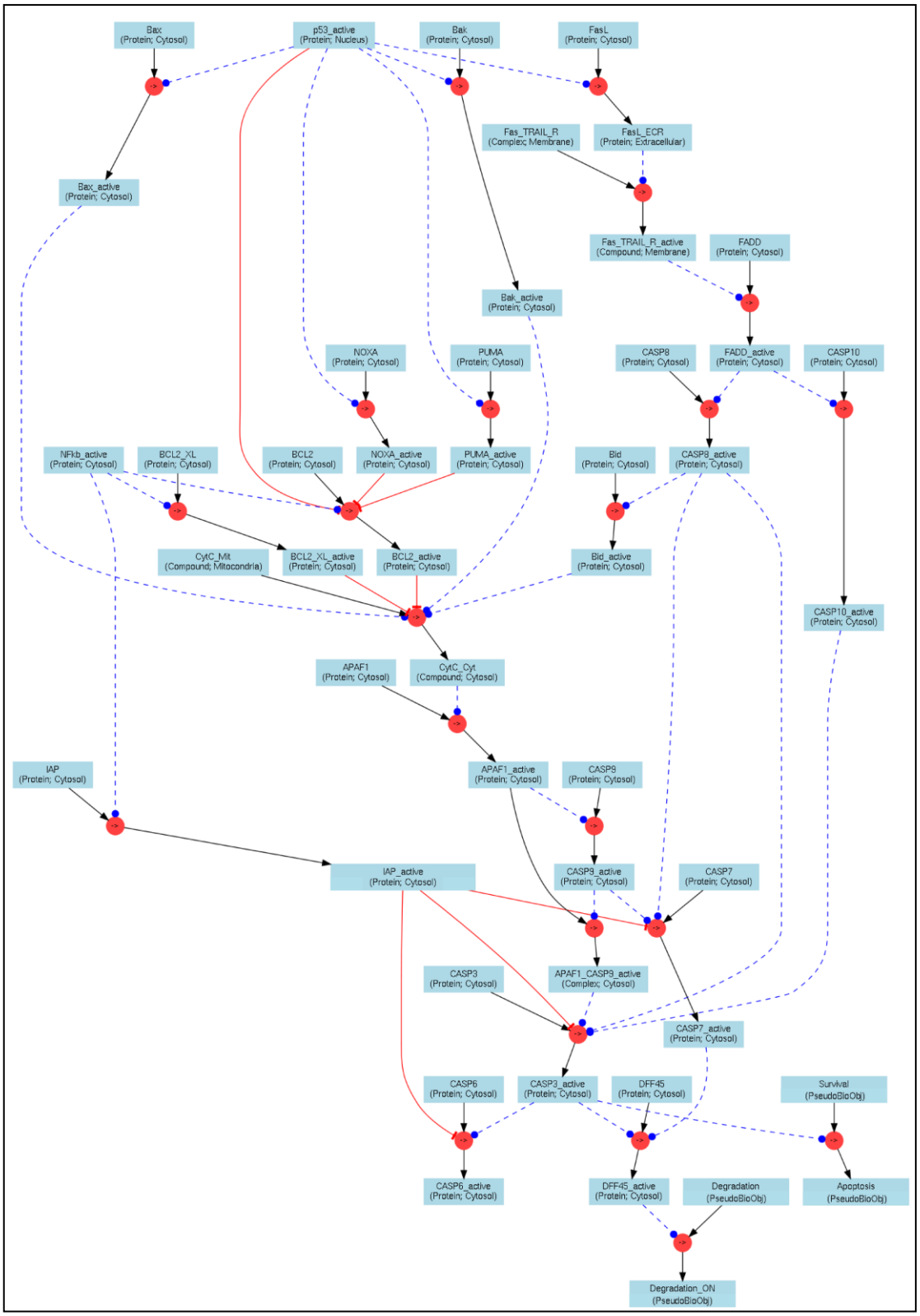

Figure 29. Scheme of apoptosis. Illustration of the apoptotic signalling as it has been implemented in PyBios. The blue boxes represent components of the model, while the red dots represent reactions. In the blue boxes is written the name of the component, the type of molecule it is and its location. Components added for convenience is labelled as pseudobioobject ("PseudoBioObj").Components are linked to reactions by lines that are black if they are reactants of the reaction, blue if they are acting as activators and red if they are acting as repressors in the reaction. Basal synthesis and degradation reactions for each of the components have not been represented to maintain the scheme simple.

Other important survival mechanisms rely on the control of mitochondrial function by regulating the permeability of the membrane that may release apoptotic inducing proteins into the cytoplasm such as cytochrome c. The BCL2 family of proteins regulates to a large extent this process (Donovan and Cotter 2004). Accumulation of inhibitors of pore formation such as BCL2 and BCLXL in the membrane counteracts the action of pro-apoptotic proteins (e.g. BAX, BAK) that otherwise would form homodimers and a channel through the membrane (Gross, Jockel et al. 1998).

As mentioned earlier p53 has many transcriptional targets that induce apoptosis. Among the best well-studied are FAS ligand, BAX, BAK, PUMA, and NOXA (Slee, O'Connor et al. 2004) that have been implemented in the model. While Fas ligand would trigger receptor-mediated cell death (Nagata 1999), BAX, BAK, PUMA and NOXA localize to the mitochondria and promote cytochrome c release (Slee, O'Connor et al. 2004).

5.4.1.3. Cell cycle

As it has been implemented in the model the cell cycle starts with the extracellular mitogenic signals, here generalized as "Growth_Factor", that are sensed by means of different receptors in the cell and initiate a signal transduction cascade (Figure 30). This signal transduction, the MAP kinase signalling pathway have been simplified here to a single event ("MAPK_pathway_OFF/ON") since it was not the focus of this study. This pathway retrieves the information into the nucleus by means of transcription factors, such as RAS and c-MYC, which trigger the transcription of D-type Cyclins (D1, D2 and D3), and other components necessary to promote cell cycle like phosphatases (e.g. CDC25A) (Murray and Hunt 1993). D-type Cyclins will then bind and activate CDK4 or CDK6 (G_1 phase) (Matsuoka, Yamaguchi et al. 1994). The model shows the complex form by CDK4/6 with Cyclins type D (labelled as "C_D_CDK4_6" in the model) that once activated displace the so-called pocket proteins, such as RB, p107, and p130, from the promoter of S-phase proteins, where E2F proteins can then activate their transcription (Cobrinik 2005). The inhibitory effect of the pocket proteins have been represented in the model as an "Rb_active" inhibitory reaction on the activation of E2F proteins.

E-type (E1 and E2) and A-type Cyclins (A1 and A2) will be then expressed and bind CDK2 ("C_E_CDK2" and "C_A_CDK2" in Figure 30) and further phosphorylate pocket proteins to complete their inactivation (Cobrinik 2005). CDK2-Cyclin E/A complexes drive transcription of proteins necessary for DNA synthesis like the Origin Recognition Complex (ORC) and the Minichromosome Maintainance Complex (MCM) (Murray and Hunt 1993). These events have been implemented in the model as a single reaction ("DNA_biosinthesis_OFF/ON") that represents S phase transition.

After DNA replication CDK2-Cyclin A2 (Cyclin A1 in germ cells) complex drives the transition from S phase to mitosis, a period known as the G2 phase (Vermeulen, Van Bockstaele et al. 2003). Finally, CDK1 is activated by A-type Cyclins, facilitating the formation of the CDK1-Cyclin B complexes responsible for driving cells to mitosis (Vermeulen, Van Bockstaele et al. 2003). Activation of this complex has been considered here as the key molecular event leading to initiation of mitosis, and for this reason has been linked to the output reaction “Mitosis_ON/OFF”.

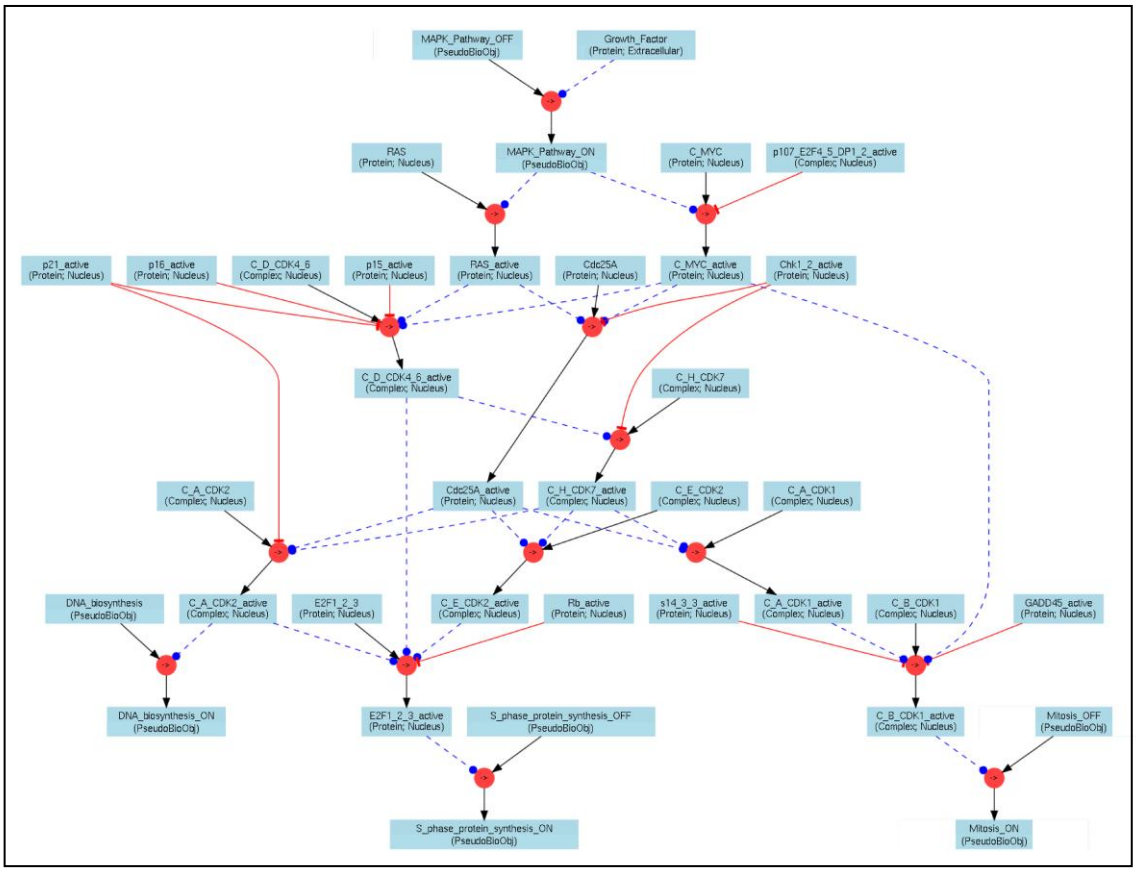

Figure 30. Cell cycle interaction map. Illustration of the cell cycle as it has been implemented in PyBios. The blue boxes represent components of the model, while the red dots represent reactions. In the blue boxes is written the name of the component, the type of molecule it is and its location. Components added for convenience is labelled as pseudobioobject (“PseudoBioObj”). Components are linked to reactions by lines that are black if they are reactants of the reaction, blue if they are acting as activators and red if they are acting as repressors in the reaction. Basal synthesis and degradation reactions for each of the components have not been represented to maintain the scheme simple.

All CDKs are activated by the phosphorylation carried out by CDK7-Cyclin H complex (“C_H_CDK7” in the model) that is a constituent of the transcription factor IIH (TFIIH) complex (Martinez, Afshar et al. 1997). The activation of the CDK7-Cyclin H complex has been ascribed to the complex CDK4/6-Cyclin D in order to model the sequential activation of CDK-Cyclin complexes, even though no biological evidence has been found so far to support it. Other activation events on CDKs are known to be

necessary, those required the activity of phosphatases (e.g. Cdc25A/B/C) and have also been included in the model, though as one, “Cdc25A”, since they are very similar structurally and functionally (Murray and Hunt 1993).

Two families of inhibitors regulate CDK activity: the INK4, including p15 and p16, and the Cip/Kip family, composed of p21 and p27. These small proteins can inhibit cell cycle progression by competitive binding to CDKs, Cyclin or CDK-Cyclin complexes, impairing their activation (Cobrinik 2005). Therefore these proteins appear in the model as inhibitors of CDK4/6-Cyclin D activation. Additionally, p21 is known to inhibit CDK2-Cyclin A activation, which has also been included in the model. Cell cycle inhibitors can be activated by different signalling pathway, as the TGFβ and p53 signalling pathways, and processes as replication stress or DNA damage (Cobrinik 2005).

DNA damage related mechanisms can also stop the cell cycle by activating proteins such as 14-3-3 adaptor proteins and GADD45α that inhibit the binding of CDK1 to Cyclin B (“C_B_CDK1” in the model) (Harper and Elledge 2007). Alternatively, checkpoint proteins can be activated (e.g. Chk1/2), which induce inhibitory phosphorylations on CDKs (Harper and Elledge 2007). The action of these proteins (“Chk_1_2_active”) has been implemented in the model attributing them an inhibitory role in the activation of “CDC25A” and “C_H_CDK7”, necessary activators of CDKs (Martinez, Afshar et al. 1997).

5.4.1.4. TGFβ signalling pathway

There are more than 40 peptides belonging to the TGFβ family of cytokines (Massague 2000). In the model signal transduction is initiated either by TGFβ as prototype of the family and/or by a component tagged as “Replication_Stress” (see Figure 31). The latter has been added because independently of extracellular cytokines the TGFβ pathway can be activated as response to DNA replication malfunction (Cobrinik 2005). A cytokine such as TGFβ starts a signalling cascade bringing together TGFβ receptor type I and type II (“TGFbR1” and “TGFbR2” respectively in the model). Receptor type II activates then type I receptors by phosphorylating the cytoplasmatic domain of the protein, implemented in the model as “TGFbR1_2_active”. These receptors transduce then extra- and intracellular information phosphorylating different cytosolic targets, from which SMADs and stress-related MAP kinases were considered the most relevant for this work (Massague and Wotton 2000).

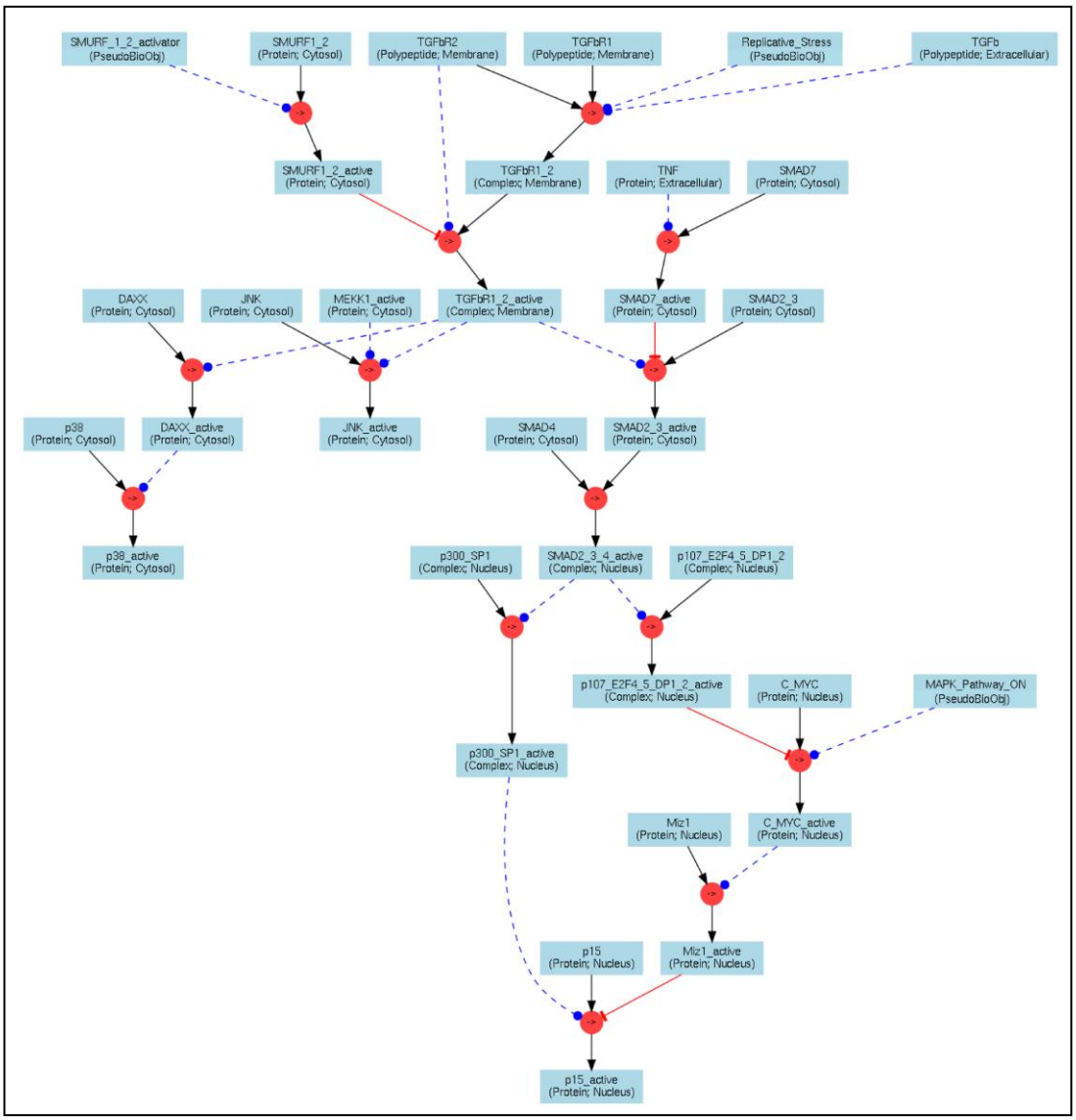

Figure 31. Scheme of the TGFβ signalling pathway. Illustration of the model as it has been implemented in PyBios. The blue boxes represent components of the model, while the red dots represent reactions. In the blue boxes is written the name of the component, the type of molecule it is and its location. Components added for convenience is labelled as pseudobioobject ("PseudoBioObj"). Components are linked to reactions by lines that are black if they are reactants of the reaction, blue if they are acting as activators and red if they are acting as repressors in the reaction. Basal synthesis and degradation reactions for each of the components have not been represented to maintain the scheme simple.

Different receptor-regulated SMADs (R-SMADs: SMAD1/2/3/5/8) are anchored and activated depending on the receptor pairing (Massague 2000). Because the anti-growth properties of this pathway have been linked to SMAD2/3, the model just contains the activation of those transcription factors ("SMAD2/3" to "SMAD2/3_active"). Activated R-SMADs can translocate to the nucleus after forming complexes with SMAD4, where together with other co-activators or co-repressors regulate a wide variety of target genes. The model contains two of the well-characterized cytostatic interaction: the up-regulation of p15 and the down-regulation of c-MYC (Siegel and Massague 2003). In order to

regulate the expression of *p15*, the SMADs form an activating complex together with p300 and SP1 on the promoter of the gene. This regulatory event is implemented in the model through a positive regulator, “p300_SP1_active”, which acts on the activation of p15. The down-regulation of c-MYC is mediated by a complex containing p107, E2F4/5, and DP1/2, and for that reason appears an inhibitor labelled as “p107_E2F4_5_DP1_2_active” on c-MYC activation. Thus, “SMAD2_3_active” regulate positively both complexes recapitulating the effects on cell cycle upon TGFβ activation (Siegel and Massague 2003).

SMAD6 and SMAD7 are inhibitors of the TGFβ pathway. They compete for receptor and SMAD4 binding with R-SMADs and target the receptor for degradation (Massague 2000). These inhibitory SMADs are represented in the model by SMAD7 that has been shown to be activated by TNFα (“TNF” in the model) (Nakao, Okimura 2002). Other genes are also involved in the inhibition of this pathway, such as secreted peptides that bind to TGFβ ligands holding back the activation of the receptors, or oncogenes like SKI or SKIL that interact with R-SMADs repressing their ability to activate target genes, but the DNA microarray data did not show large changes in their expression (Massague and Wotton 2000; Bonnon and Atanasoski 2011). Instead, intracellular ubiquitin ligases such as SMURF1/2 (“SMURF1_2” in Figure 31) that tag receptors to degradation were found to be differently regulated in most of the datasets and have been included in the model as inhibitors of TGFβ receptors activation. Unfortunately, no information was found regarding the activation of these ubiquitin ligases and an imaginary activator had to be included in the model to mimick their up-regulation in those datasets were was found positively regulated.

TGFβ receptors can also initiate other signalling transducing cascades independently of R-SMADs. Moreover, anti-growth signals can also be driven through non-SMAD pathways such as the p38 and Jun N-terminal kinase (JNK) (Wagner and Nebreda 2009; Mu, Gudey et al. 2011). These activations are independent of the receptor-kinase activity (Mu, Sundar et al. 2011). The activation of p38 takes place after a cascade of phosphorylating events (Wagner and Nebreda 2009) that have been summarized in the model with the activation of one intermediate such as DAXX. Other signalling pathways are also known to activate JNK (Ip and Davis 1998), and therefore have been represented in the model by MEKK1 that should serve as indicator of alternative regulators of JNK. As mentioned earlier (see section 5.4.1.1) activation of p38 and/or JNK, can end up in the stabilization of p53.

5.4.1.5. DNA damage response – Homologous recombination

Homologous recombination (HR) is essential for the repair of DNA double-strand breaks (DSBs) and the accurate maintenance of the genetic material of the cell (Ouyang, Woo et al. 2008). Since the components of this pathway were over-represented among the genes differentially expressed (see Figure 14), the most important events necessary to repair a DSB have been included in the model (Figure 32).

After DNA damage response, ATM or ATR are quickly phosphorylated and accumulates at double-strand breaks, which has been represented in the model as the activation of "ATM/ATR" to "ATM/ATR_active". ATM phosphorylates numerous substrates as DNA damage response, including histone H2AX, MDC1, p53, 14-3-3 adaptor proteins, GADD45α, and CHK1/2 (Savic, Yin et al. 2009). Details on the activation of p53 and the down-stream signalling of 14-3-3 adaptor proteins, GADD45α, and CHK1/2, which induce G_2-M cell cycle arrest, can be found in section 5.4.1.1.

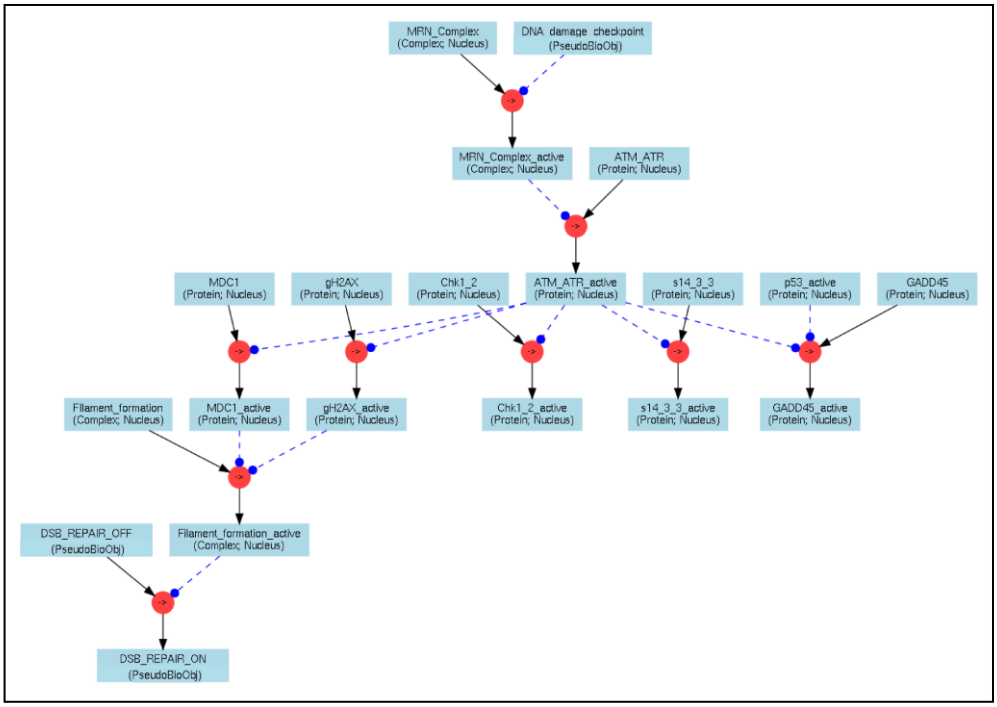

Figure 32. Scheme of homologous recombination. Illustration of the cascade of events leading to DNA-break repair as it has been implemented in PyBios. The blue boxes represent components of the model, while the red dots represent reactions. In the blue boxes is written the name of the component, the type of molecule it is and its location. Components added for convenience is labelled as pseudobioobject ("PseudoBioObj"). Components are linked to reactions by lines that are black if they are reactants of the reaction, blue if they are acting as activators and red if they are acting as repressors in the reaction. Basal synthesis and degradation reactions for each of the components have not been represented to maintain the scheme simple.

Even though, the activation of ATM/ATR is a key event on HR repair, the recognition of DNA double-strand breaks is mediated by the MRN complex composed by RAD50, NBN (also named NBS1) and MRE11, that binds at DNA break sites (Hopfner, Putnam et al. 2002). This step has been modelled as the activation of the MRN complex ("MRN_complex" to "MRN_complex_active"). It is then, when the MRN complex recruits and activates the catalytic function of the ATM protein kinase through direct interaction of ATM and NBN (Lee and Paull 2004). MDC1 and γH2AX bind adjacent to ATM that phosphorylates them, which has been included in the model as independent activation steps in order to use them as markers in the simulation experiments.

ATM-mediated phosphorylation events enable the formation of complexes compose partly by BRCA1/BARD1 and BRCA2/DSS1, which will induce the loading of RAD51 (Yang, Li et al. 2005), after 5′-to-3′ exonuclease activity of MRE11 excises nucleotides at the end of the DNA break, converting the edges at the break point into ssDNA, to which RPA (replication protein A) can bind (Trujillo, Yuan et al. 1998). The formation of BRCA1/2-containing complexes at the edges of the DNA strand break allows the formation of a filament, composed of multimers of the RAD51 recombinase that bound to ssDNA. These latter steps have been summarized in the model with a reaction that results in the accumulation of "Filament_formation_active".

The RAD51 nucleoprotein aligns the broken DNA strands to the complementary sequence in the sister chromatid forming a D-loop intermediate (Holliday junction). At this point the polymerase δ elongates the DNA chain using the intact DNA as a template (Maloisel, Fabre et al. 2008). The Holliday junction is thought to be resolved by enzyme complexes such as BLM in complex with topoisomerase IIIα or MUS81–EME1 complex (Hartlerode and Scully 2009). These last steps have not been implemented in the model, instead a symbolic reaction reflecting the activity on this pathway has been added ("DSB_REPAIR_OFF" to "DSB_REPAIR_ON").

5.4.2. Simulations - Reprogramming

5.4.2.1 Over-expression of the reprogramming factors

After assembling the model I introduced the expression data from the microarray in the model as the rate at which the active components will be produced. This modelling strategy assumes that the production of active species with functional meaning is just dependent on the mRNA levels found and directly proportional to them. The geometric mean of the distribution of results obtained after 100 Monte Carlo simulations (see section 4.6) will represent here the steady state concentration of the different components of the model. The final result will then predict the concentration of the

component once the steady state is reached given the initial state defined by the expression data obtained with the microarray. The heatmaps showed in the next figures (Figure 33 and Figure 37) highlight the differences in the steady state reached by cells over-expressing the different proteins and non-transduced fibroblasts.

Thus, in Figure 33 can be found the results of the simulation of the over-expression of GFP, OCT4, SOX2, KLF4, c-MYC, 3TF (OCT4, SOX2, and KLF4) and 4TF (OCT4, SOX2, KLF4, and c-MYC). The outcomes capture most of the expectations that the results previously described would generate. As observed in the cluster analysis (Figure 12) and explained in section 5.2.1, GFP, OCT4, and SOX2 samples share many similarities in their expression profile (e.g. p53 signalling, apoptosis, cell cycle) and therefore reached here a comparable steady state. Some of the patterns are also partially recalled in the other samples.

Among the common pathways most significantly changed (see section 5.2.1) was the p53 signalling pathway. Due to the higher levels of p53-down-stream targets the model predicts the induction of apoptotic signalling through death ligands (see "FAS-TRAIL" in Figure 33) and induce apoptosis (see "APOPTOSIS" in Figure 33). The up-regulation of intrinsic apoptotic signalling was confirmed by qRT-PCR (see "AIF" in Figure 19), and seem to be effectively responsible for cell death (see "DNA degradation" in Figure 33). Though, apoptotic signalling is counteracted by pro-survival proteins (e.g. BCL or BCLXL) and inhibitors of apoptosis (e.g. IAP) and here constrain apoptosis to a certain extent, the simulations on the current model stress that they do not abolish completely the activation of the apoptotic cascade. The similarities found in the pathway analysis between KLF4 and the 3TF combination in apoptosis regulation (see 5.2.4) are reflected here in the predicted steady states, since these samples showed the lowest levels of apoptotic signalling (see "APOPTOSIS" in Figure 33). This is principally due to the down-regulation of p53, which reduces the activation of down-stream targets in the steady state. The simulations also mirrored the partial neutralization in 4TF of the death-inducing signalling observed in c-MYC (see section 5.2.5).

Remarkably, the levels of p53 are not significantly changed in most of the samples. The model suggests that the up-regulation of its inhibitor *MDM2* is responsible in some cases (GFP, OCT4, SOX2, and c-MYC) to keep the level of active p53 at the levels found in non-transduced fibroblasts. The levels of p53 are partially responsible for the activation of *p16* and *p21* observed in single transductions. The down-regulation of these cell-cycle inhibitors explained partially the higher levels of mitotic activity found in samples were the factors were over-expressed in combination.

The up-regulation of cell cycle machinery in GFP, 3TF and 4TF that were verified by qRT-PCR (Figure 22) translate in higher rates of Cyclin-CDK complexes (see Figure

33). These changes are especially prominent in c-MYC-containing samples and 3TF over-expression (see “CYCLIN D-CKD4/6” in Figure 33).

They showed also a strong up-regulation of G_1-S phase transition complexes as Cyclin E/A-CDK2 that was not observed for OCT4, SOX2, and KLF4. All samples shared in some measure an up-regulation of cell cycle dynamics in comparison to non-transduced fibroblasts that seems to be effective in functional terms (see section 5.2.1). Nevertheless, OCT4, SOX2, and KLF4, seem to slow down the cell cycle since it would be expected a similar increase of cell cycle components concentrations as in control samples (“GFP”) and it is not the case.

The validation of the microarray data suggested that homologous recombination was active in the cells after over-expression of the proteins (Figure 25). Activity of this pathway is summarized in the model by a symbolic tag, “DSB repair”, which shows an increase in activity in most of the samples, especially following c-MYC, 3TF and 4TF over-expression as it happened to be based on the qRT-PCR (see section 5.3.5).

A general down-regulation of components of the TGFβ pathway is predicted and was expected based on the qRT-PCR results found (Figure 24). GFP- and SOX2-expressing fibroblasts retain more similar levels to non-transduced fibroblasts. The up-regulation of inhibitors of the pathway (e.g. “SMURF1/2” and “SMAD7”) was not fully recapitulated in this model. Also, the activation of *p15* after c-MYC and KLF4 over-expression (see 5.2.2 and 5.2.3) would be expected to remain active in the cells in the steady state, but was not successfully mimicked by this model (see section 6.5).

5.4.2.2. Inhibition of senescence and apoptosis

From the previous simulations it is reasonable to think that apoptosis and senescence might be barriers for reprogramming. Apoptosis is induced in this network by p53 that can also up-regulate cell cycle inhibitors, like *p16* or *p21*, causing senescence. Senescence can be also established by the TGFβ signalling pathway through p15 and DNA damage responses by activation of checkpoint proteins as CHK1/2. Additionally both of these pathways can at the same time activate p53.

Alternatively, if some of the cells are successfully reprogrammed it is because they can overcome senescence and apoptosis. The model can be used to simulate some of the mechanisms that the simultaneous over-expression of the reprogramming factors triggered in the cells and help them to surmount these anti-growth mechanisms. Simulating the knock-down of proteins can be helpful to identify down-stream effects of the proteins on the steady state. Furthermore, those effects can be compared with the simulation of the over-expression of factors, and if the same patterns of down-stream regulations are found, hypotheses can be postulated regarding the participation of the

perturbed protein on the establishment of the steady state found in the over-expression of reprogramming factors.

Thus, I simulated the inhibition of the different pathways that may stop cell proliferation or induce cell-death and study how the steady state changes. In Figure 33 can be observed the outcome of p53 inhibition, the impairment of DNA damage responses—exemplified here with the knock-down of ATM and DNA-PK—, as well as the inhibition of the TGFβ signalling in 3TF and 4TF over-expression. As expected the inhibition of p53 or ATM/DNA-PK has main effects on the apoptotic cascade that is strongly down-regulated or inactive. Both pathways have also a remarkable effect in the steady state regulation of cell cycle inhibitors like p21 or GADD45α. The inhibition of ATM/DNA-PK as up-stream inducer of the DNA damage response results additionally in the inactivation of homologous recombination ("DSB repair" in Figure 33). The inhibition of TGFβ pathway at the receptor level (most left columns in Figure 33) would cause the complete shutdown of the pathway leading to SMAD2/3 inactivation, which would cause down-regulation of *p15*. Additionally p38 or JNK are no longer activated by TGFβ receptors, which causes a decrease in active p53. The knock-down of other components of the TGFβ receptor as SMAD4 would rather avoid the activation of the gene coding for p15, but not the post-translational activation of non-SMAD signalling such as the MAP kinase proteins, p38 and JNK.

These simulations allow searching for similarities between the virtual knock-downs and the profiles observed in the over-expression of the combination of factors and postulate some hypothesis. First, as observed in the knock-down of p53, the decrease in apoptotic signalling observed in 3TF and 4TF over-expression can be very well due to the down-regulation of p53. Since most of the down-stream effects caused by the *in silico* knock-down of p53 are readily observed in the simulations corresponding to the combination of factors (see columns "3TF *p53 ko*" and "4TF *p53 ko*" in Figure 33). It is especially clear, when 4TF and c-MYC over-expression are compared, given that most of the components of the model differentially regulated between both correspond to down-stream targets of p53.

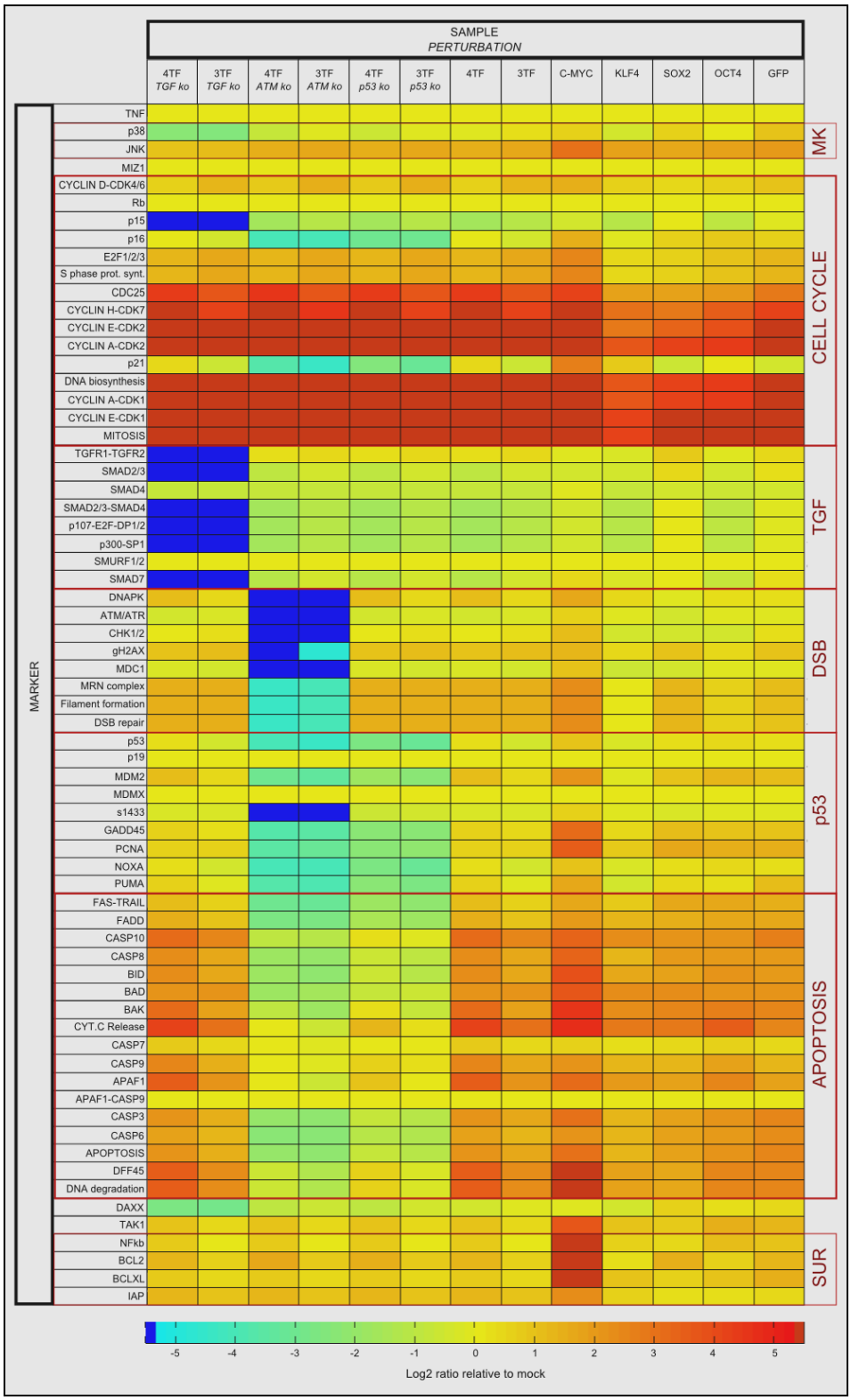

Figure 33. Simulation of the over-expression of the reprogramming factors. The components of the model are ordered in rows. The columns show the different simulation results, when the expression data for each experiment, from left to right GFP, OCT4, SOX2, KLF4, c-MYC, 3TF, and 4TF, was introduced as parameter for each of the components listed rows. The following columns show the simulation of different perturbation experiments (indicated in cursive letters) on the 3TF or 4TF datasets. The boxes were coloured based on the geometric mean of the Monte Carlo simulation following the scale showed at the bottom. Abbreviations: MK: MAP kinase signalling pathway, DSB: homologous recombination, and SUR: survival proteins.

On the other hand, up-stream regulators of p53 as ATM/ATR might not be distorted, since the DNA damage response cascade is still active in the cells, and its inactivation would impair double-strand break repair (see columns “3TF *ATM ko*” and “4TF *ATM ko*”). Decrease on p53 activity has functional meaning for the cells not just in apoptosis, but also in cell cycle regulation that becomes faster as G_1-S phase loses restrictive activities (e.g. p16 and p21).

The down-regulation of the TGFβ signalling pathway is set in these simulation experiments at the level of receptor/s (columns “3TF *TGF ko*” and “4TF *TGF ko*” in Figure 33), which cause a strong decrease in the TGFβ signalling, but also on MAP kinases, p38 and JNK, and the cell cycle inhibitor p15. The same components are down-regulated in 3TF and 4TF simulations in comparison to GFP control. Therefore, it is reasonable to postulate that the combination of factors could be impairing TGFβ signalling at the level of the receptors. Furthermore, it is worthwhile to test experimentally this hypothesis because it would explain the presence of *TP53* transcripts but not its apoptotic down-stream effects, since a post-translational activator as p38 is no longer active.

The lowest levels of TGFβ and p53 signalling are observed in KLF4 samples. This protein is present in combinations, 3TF and 4TF, and the simulations reinforce the idea that KLF4 anti-apoptotic effects on the cells are present in the combination of the factors. While this would be an important contribution to facilitate the reprogramming, the other factors seem to be absolutely necessary to induce faster cell cycle rates that in KLF4 are hindered by cell cycle inhibitors (see section 5.2.2).

5.4.3. Validation of *in silico* predictions

5.4.3.1 p53 knock-down during the reprogramming process

The model assembled and the simulations performed point out to p53 as outcoming node of different anti-proliferative mechanisms triggered by the TGFβ signalling pathway and transduced by MAP kinases. As such, its effects on cell cycle are especially adverse for further growth and the completion of the reprogramming process. The simulations allowed the inhibition of the different pathways activating p53 and predicted an improvement in the growth rates of the cells in a p53-knock-down scenario. I validated the postulated hypothesis and tested if knocking-down p53 would improve the proliferative capacity of the cells and decrease at the same time the apoptotic signalling induce by c-MYC.

Thus, the over-expression of 3TF (OCT4, SOX2, and KLF4) and 4TF (OCT4, SOX2, KLF4, and c-MYC), as was originally done to obtained the microarray data (see section 4.2.8), were reproduced in fibroblast lines expressing a short hairpin targeting

TP53 mRNA for degradation (see section 4.2.9). A cell expressing a short hairpin of a scramble mRNA was used as control (labelled as “SC” after the sample name in Figure 34 and 35). Four days after transduction mRNA was harvested and cDNA was produced (see section 4.4.2) to carry out qRT-PCR reactions shown Figure 34 and 35.

I checked then the levels of *p53* transcript as well as some of its direct targets, like p21 or p16 that are directly involved in the establishment of senescence. The cell lines created showed reduced levels of *p53*, ranging from a 6- to a 3-fold change (log_2 scale). *p16* and *p21* showed consistent down-regulations, being *p16* expression 3-fold (log_2 scale) lower in 3TF and 4TF in cell lines expressing shRNAs targeting *p53* (Figure 34).

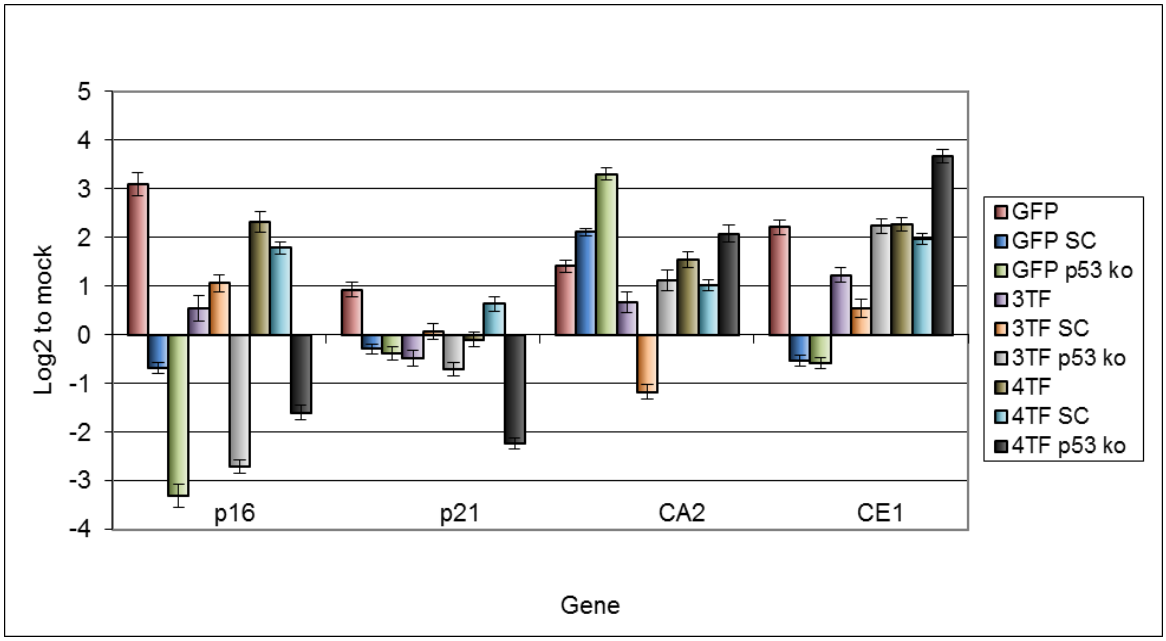

Figure 34. p53 knock-down. Transcriptional regulation of *TP53* and cell cycle components by qRT-PCR. The relative amount to non-transduced fibroblasts of each transcript is scaled in the Y-axis that represents logarithmic fold changes. In the X-axis are the gene targets under analysis. The legend shows the samples that have been analysed. 3TF is the abbreviation used for fibroblasts over-expressing OCT4, SOX2, and KLF4. 4TF is the abbreviation used for fibroblasts over-expressing OCT4, SOX2, KLF4, and c-MYC. “SC” after the sample name indicates that is a cell line expressing a scramble shRNA. If the cell lines are expressing a p53 shRNA is indicated after the sample. The amounts shown are the averaged of three different technical replicates. The error bars indicate the standard deviation of the average of all experimental replicates.

To verify that this reduction in the cell cycle inhibitors was increasing the proliferation potential of the cells I tested transcript levels of Cyclin A2 and Cyclin E1. These Cyclins are active in the synthesis phase of the cell cycle and their up-regulation after p53 knock-down can be expected (see section 5.4.1.1). And indeed, higher rates of expression were found for both of them. While *Cyclin E1* transcript levels showed larger

changes than *Cyclin A2*, their transcription was at least doubled in cells over-expressing the combination of factors that had at the same time *TP53* transcriptionally impaired

On the other hand, the simulations predicted that the regulation of *TP53* after the over-expression of the reprogramming factors were responsible for the low expression of pro-apoptotic proteins. Thus, I added to the analysis *BAX* and *CASP3* and checked their levels of expression in p53 knock-down cell lines. Figure 35 shows that the expression of *BAX* following the over-expression of the combination of factors is independent of p53 knock-down. Instead in GFP-expressing fibroblasts a 3-fold change (log_2 scale) decrease could be detected. *CASP3* showed low transcript levels in samples over-expressing GFP or any of the combination of factors. In p53 knock-down cells achieved levels comparable to non-transduced fibroblasts following 3TF over-expression, and again were independent of *TP53* transcript levels following 4TF over-expression.

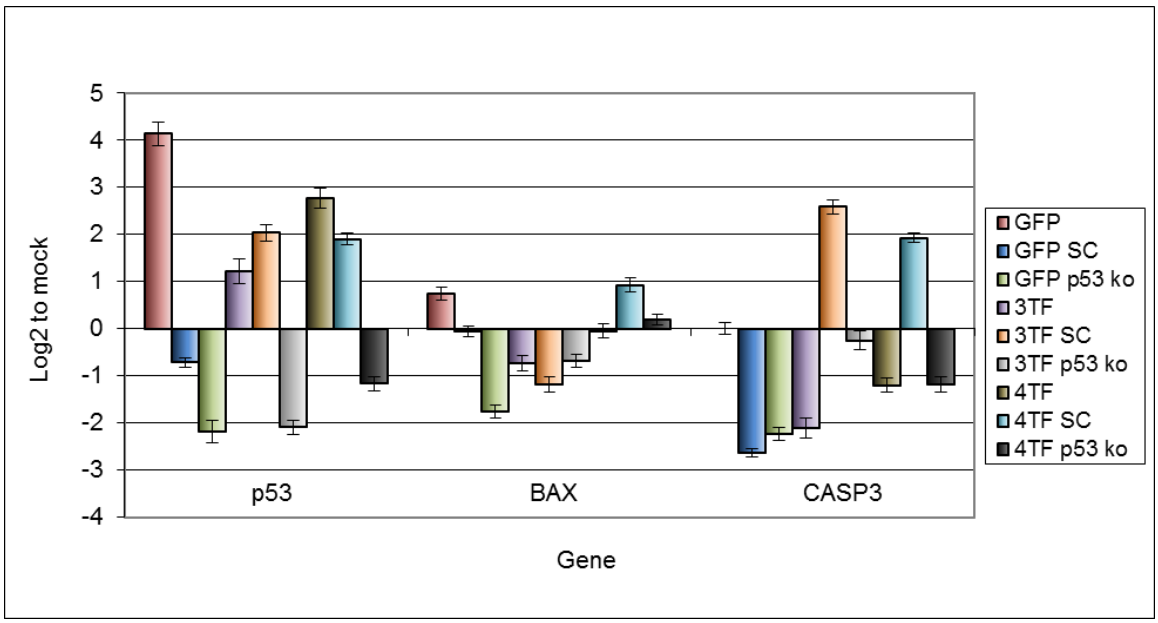

Figure 35. p53 knock-down. Transcriptional regulation of *TP53* and down-stream targets of the apoptotic pathway measured by qRT-PCR. The relative amount to non-transduced cells of each transcript is scaled in the Y-axis that represents logarithmic fold changes. In the X-axis are the gene targets under analysis. The legend shows the samples that have been analysed. 3TF is the abbreviation used for fibroblasts over-expressing OCT4, SOX2, and KLF4. 4TF is the abbreviation used for fibroblasts over-expressing OCT4, SOX2, KLF4, and c-MYC. "SC" after the sample name indicates that is a cell line expressing a scramble shRNA. If the cell lines are expressing a p53 shRNA is indicated after the sample. The amounts shown are the averaged of three different technical replicates. The error bars indicate the standard deviation of the average of all experimental replicates.

Altogether, these results correlate with the predicted effects of p53 knock-down, inducing the expression of G1-S phase components of the cell cycle and down-regulating cell cycle inhibitors (see section 5.4.2.2). Interestingly, following simultaneous over-expression of the reprogramming factors, apoptotic signalling is not anymore

depending on p53. Even though the simulations were useful to understand that p53 is responsible for the apoptotic signalling following the over-expression of GFP, these validations suggest that is not the level of p53 transcription the key event to avoid its death-inducing effects.

5.4.3.2. Inhibition of TGFβ signalling during the reprogramming process

One of the most interesting issues that came out during the analysis of the microarray data is the absence of down-stream apoptotic signalling in the combination of four factors, since c-MYC clearly induced high levels. As shown in the above simulations, apoptotic signalling depends to a large extent on p53 (see section 5.4.2.2), and its activation can be a major mechanistic issue to understand cell survival after over-expression of the combination of factors, since its transcriptional levels cannot explain the absence of apoptotic signalling in 4TF samples.

The over-expression of single factors as OCT4 and KLF4 caused a strong down-regulation of the *TGFβR2* that correlated with the down-regulation of TGFβ signalling (see section 5.3.4). The model suggests that the inhibition of TGFβ signalling could have strong effects on the genetic network on study. The down-regulation of components of the TGFβ signalling pathway can act in a direct manner on the expression of the cell cycle inhibitor *p15*, but in addition can affect indirectly stress-related kinases that may not be activated. Reduced levels of activators of p53 would explain the lower apoptotic activity of p53, even when it is actively expressed. An expressed but inactive p53 protein would explain the apoptotic profile observed following 4TF over-expression, and could be contributing factor to cell cycle progression.

To analyse the impact that TGFβ signalling has on the network, I used an inhibitor of TGFβ/Activin receptors after over-expression c-MYC in order to mimick the TGFβ regulation observed after the over-expression of the reprogramming factors. I set up the over-expression of c-MYC as previously described (see section 5.1) and added in the media a small molecule that inhibits TGFβ receptor 1—but not TGFβ receptor 2—(see section 4.2.10) to analyse its effects on *TP53* regulation. Four days after transduction mRNA was harvested and cDNA was produced (see section 4.4.1) to carry out the qRT-PCR reactions shown in Figure 36. To test if the same molecular mechanism could be active on the combination of factors, results were compared with the three and four factor over-expression.

The addition of the inhibitor, which impairs the activation of the TGFβ receptors, had no remarkable effect on the transcription of most the genes analysed (Figure 36). Transcript levels of *TGFR1* were down-regulated, while *TGFR2* levels remained practically constant. GFP-expressing fibroblasts showed a low up-regulation of *TP53*,

p16, *p21* and *JNK* expression, after addition of the inhibitor, and *p38* was slightly down-regulated.

The most remarkable difference in expression was found in c-MYC-expressing fibroblasts. After impairment of TGFβ signalling *TP53* was modestly down-regulated, but notably the expression levels of *p38* were halved, displaying levels comparable to the combination of factors. At the same time cell cycle inhibitors, *p16* and *p21* were found slightly down-regulated. Other p53 activators such as JNK were not affected upon inhibitor addition.

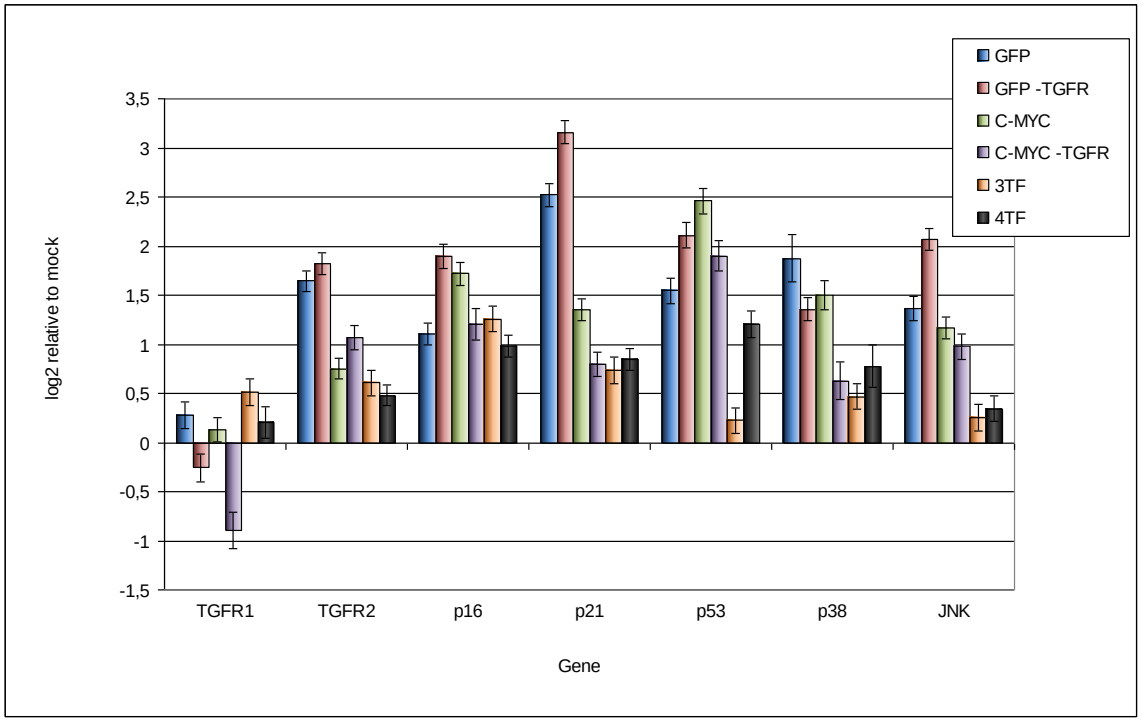

Figure 36. Inhibition of TGFβ receptors. Transcriptional regulation of the TGFβ signalling pathway and its down-stream targets measured by real-time PCR. The relative amount of each transcript relative to non-transduced fibroblasts is scaled in the Y-axis that represents logarithmic fold changes. In the X-axis are depicted the genes under analysis. The legend shows the samples that have been analysed. 3TF is the abbreviation used for fibroblasts over-expressing OCT4, SOX2, and KLF4. 4TF is the abbreviation used for fibroblasts over-expressing OCT4, SOX2, KLF4, and c-MYC. Samples were TGF receptors were inhibited is indicated as "-TGFR" after the name of the sample. The amounts shown are the averaged of three different technical replicates. The error bars indicate the standard deviation of the average of all experimental replicates. Embryonic stem cell mRNA (H9) was used as a positive control for the primers.

In overall, these results show that the down-regulation of TGFβ/Activin receptors can indeed contribute to the activation of stress-related kinases that activate p53, as predicted by the simulation experiments (5.4.2.2). With these results, it was confirmed that the levels of these kinases after the over-expression of the reprogramming factors are low, even when c-MYC is simultaneously over-expressed. Therefore it remains

feasible that the regulatory events induced by the reprogramming factors on the TGFβ pathway are partially responsible for p53 and cell cycle inhibitors regulation.

5.4.4. Evasion of senescence and apoptosis—Cancer cell lines

Since the modelling approach was successful to postulate molecular mechanisms to overcome cell cycle control, it is tested at this point if it can used also to compare molecular mechanisms. If possible it would help then to answer questions such if the mechanisms that contribute to overcome cell cycle control in reprogramming are found in other instances. In fact, the molecular mechanisms that help to overcome senescence and apoptosis in the process of reprogramming shall be expected in other cells. Evasion of senescence and apoptosis belong to the hallmarks of cancer (Hanahan, Weinberg 2011). Cells that progress to anomalous proliferation rates have to avoid the cellular systems that try to guarantee controlled proliferation.

The simulations above suggest that the anti-proliferative features of TGFβ signalling are reduced following over-expression of the combination of factors. In addition, regulation of p53 signalling may contribute to decrease the levels of senescence and apoptosis, and pro-survival signalling may add resistance to otherwise apoptotic cues. These very same traits have been well characterized in the development of colon cancer (Markowitz and Bertagnolli 2009).

For these reasons, I decided to compare the over-expression of the four reprogramming factors with two colon cancer cell lines SW420 and SW680. I added to the comparison a human embryonic stem cell line (H9) to study if the molecular mechanisms found might just be a normal operative behaviour of the cells associated to pluripotency. To this end I retrieved publicly available microarray data from these cells and normalized them (see section 4.5.3), in order to compare them. The model can be then used to study the functional differences between the different cell lines. Moreover, substituting the expression values of different set of components (e.g. signalling pathway) from one set to another it is possible to pin point similarities or differences between the molecular mechanisms that underlie the behaviour of the different cell lines.

In Figure 37 from left to right are represented the simulation results for 4TF over-expression (labelled as TF4) compared to the results of two cancer cells lines, as SW480 and SW620, and a human embryonic cell line (H9). As explained in section 5.4.2 the expression data is introduced in the model as kinetic parameter in order to simulate the steady state that cells would achieve with such an initial expression profile. The heatmap in Figure 37 shows the log2 ratio of the geometric mean of the most relevant components after 100 Monte Carlo simulations. The different experiments or cell lines are compared to non-transduced fibroblast that is the control dataset.

The results show that all the cell lines achieve a highly active cell cycle (see “S phase prot. synt.”, “DNA biosynthesis”, and “MITOSIS” in Figure 37). They have also in common higher synthesis rates of DNA damage response components (see “DSB repair”). In a cellular context of DNA damage can be also expected high rates of cell death and indeed most components of the apoptotic cascade (see “APOPTOSIS”) show levels far above the control (non-transduced fibroblasts).

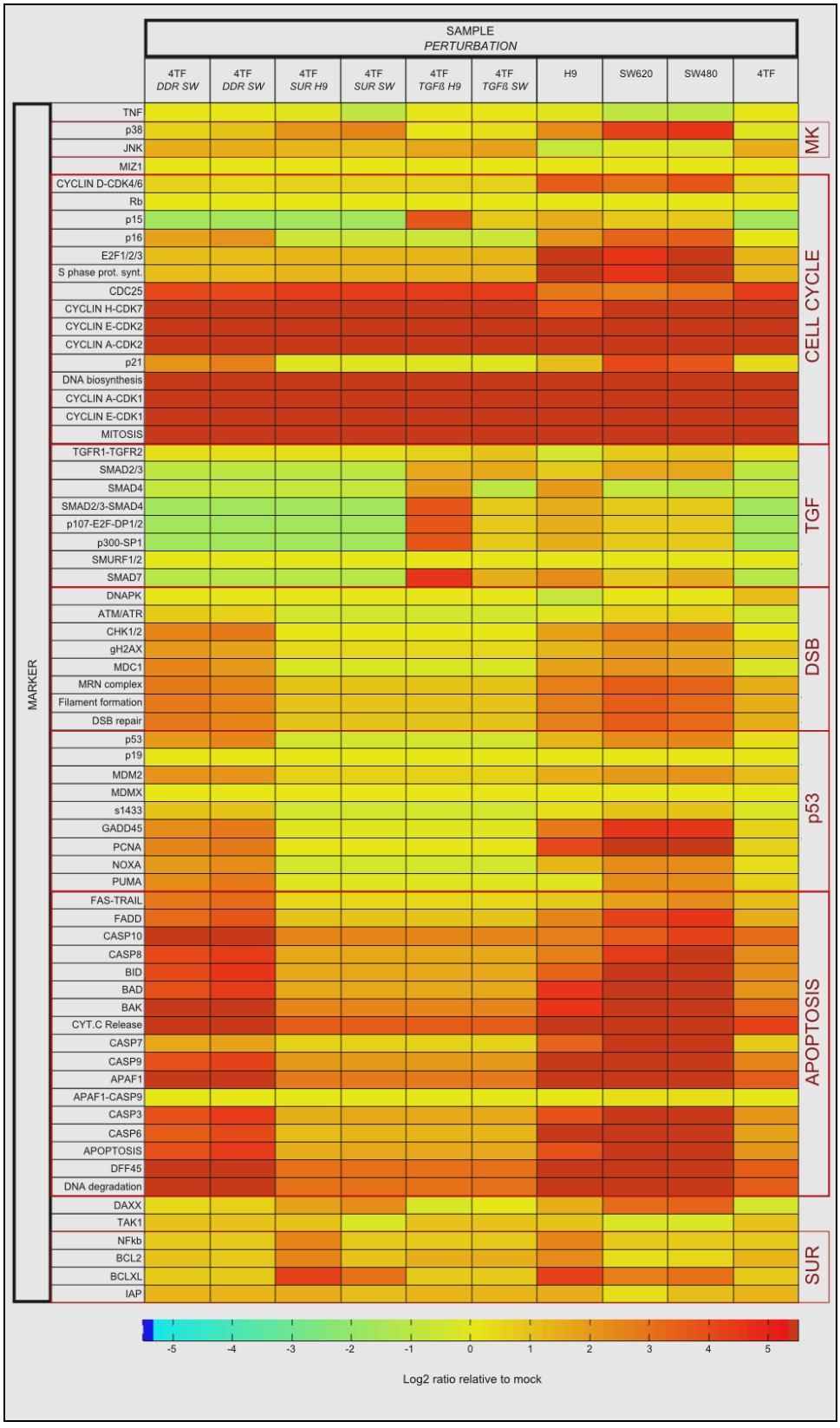

Figure 37. Over-expression of the reprogramming factors vs. high proliferative cells. The components of the model (active species) are ordered in rows. The columns show the different simulation results, when the expression data for each component was introduced as parameter in the model. From left to right 4TF over-expression, SW420 (colon cancer cell line), SW680 (colon cancer cell line), H9 (human embryonic cell line). The following rows show the simulation of 4TF over-expression, when the initial values for the TGFβ pathway, survival proteins or DNA damage response

components were interchanged for those found in a cancer cell line (SW420) or human embryonic cells (H9). Thus, 4TF *TGFβ SW* and 4 TF *TGFβ H9* show the simulation, when the values of TGFβ signalling pathway of SW420 and H9 were introduced in the 4TF dataset. 4TF *SUR SW* and 4TF *SUR H9* represent the outcome of the simulation, when the 4TF dataset has the expression values for survival proteins found in SW420 and H9, respectively. 4TF *DDR SW* and 4TF *DDR H9* show the result of interchanging the expression values found in the DNA damage response system in SW420 and H9 for those in the over-expression of 4TF. The boxes were coloured based on the geometric mean of the Monte Carlo simulation following the scale showed at the bottom. Abbreviations: MK: MAP kinase signalling pathway, DSB: homologous recombination, and SUR: survival proteins

Strikingly, the patterns for the TGFβ signalling pathway among the different datasets are very different. While the four transcription factors show a general down-regulation of the pathway, colon cancer cells are predicted to have active TGFβ signalling, even though low *SMAD4* expression (see "SMAD4" in Figure 37). On the other hand, human embryonic stem cells have high levels of expression in most of the TGFβ signalling pathway components; except at the level of receptors (see "TGFR1-TGFR2" in Figure 37). Increase activity at the level of receptors correlates with the levels of p38 and p53 signalling activity.

The TGFβ pathway profile found in SW420 can be substituted in the 4TF dataset in order to analyse the functional consequences that such a profile (e.g. low SMAD4 activity) would have in the steady state induce by 4TF. In the columns labelled as "4TF *TGFβ SW"* and "4TF *TGFβ H9"* can be observed the effects of the expression levels found in SW420 and H9, when substituted in the expression data of 4TF over-expression. In "4TF *TGFβ SW"* can be observed that active TGFβ signalling has an immediate effect on p15 and p38, even if SMAD4 is sparse. Instead, low activity in up-stream TGFβ signalling (e.g. TGFβ receptors) is important to control p38 activity.

Cell growth depends on the ratio between apoptosis and cell survival, and therefore the mechanisms that increase the chances of a cell to survive stress conditions can easily unbalance the outcome. To study the survival mechanism acting on the cells after the simultaneous over-expression of OCT4, SOX2, KLF4, and c-MYC, the expression values of survival proteins such as BCL2, BCLXL and BIRC5 ("IAP" in the Figure 37) have been substituted in the 4TF dataset for those found in SW420 and H9 ("4TF *SUR SW"* and "4TF *SUR H9"* in Figure 37, respectively). The results suggest that cells are using different strategies to overcome apoptosis. While cancer cells keep high levels in one anti-apoptotic protein (see BCLXL), cells over-expressing 4TF or stem cells show a wide range of active pro-survival proteins (see "BCL2", "BCLXL" and "IAP"). This last strategy can be extremely effective, as shown in "4TF *SUR H9"*, where the levels of apoptosis could be reduced to basal levels.

Apoptosis depends to a great extent on the DNA damage response systems of the cell that may induce p53 stabilization and trigger programmed cell death. Human ESCs and colon cancer cell lines display high levels of proteins involved in DNA repair and maintenance (see “DSB repair” in Figure 37). Substituting the profile show by cancer or stem cells in the 4TF over-expression dataset can illustrate the effects of the DNA repair system pathway on the network. The results of such simulations, using the expression levels of SW420 and H9, can be found in the columns labelled as “TF4 *DDR SW”* and “4TF *DDR H9”*, respectively. Higher activities on the pathway brought higher levels of CDK_inhibitors (see “p16” and “p21”), and higher apoptotic signalling (see “p53” and “CASP3”).

In overall, these simulations suggest that these cell lines have common features with the 4TF over-expression such as fast cell cycle, very active homologous recombination repair, and pronounced apoptotic signalling. Differences to the 4TF over-expression and among the cell lines were found in the TGFβ pathway, since this pathway was predicted to be active in H9 and both colon cancer cell lines conversely to 4TF. Nevertheless, specific down-regulation of certain components such as TGF receptors for H9, and *SMAD4* for colon cancer cell lines was also predicted.

5.4.4.1. Comparison of cancer cell lines and transduced cells by qRT-PCR

In order to compare the transcript levels of some of the key proteins of the model that were relevant for the previous analysis (see section 5.4.4), mRNA from two cancer cells lines, as SW480 and SW620, and a human embryonic cell line (H9) were used to synthesize cDNA and carry out qRT-PCR reactions (see section 4.4.2 and 4.4.3). A breast cancer cell line (MCF7) was added to the analysis to determine if the profiles observed are generally found in oncogenic processes, or are specific for colon cancer.

Figure 38 shows the results obtained on the analysis of *TP53*, a DNA damage response marker as BRCA1, a cell cycle kinase like CDK1 and key components of the TGFβ pathway such as TGFβR2 and SMAD4. Since the molecular mechanisms in study seemed to depend on KLF4 and c-MYC (see section 5.1.2), I started analysing if the cell lines expressed these proto-oncogenes. High expression of c-MYC was found only in the colon cancer cell lines. Instead, KLF4 was highly expressed in 4TF and ESCs samples.

All cell lines, as well as 4TF over-expression showed high levels of *BRCA1*. Besides this, the down-regulation of the TGFβ pathway is a common trait of the cancer cell lines analysed. *TGFβR2* and *SMAD4* expression were found down-regulated in SW480, SW620 and MCF7. All these cell lines showed low expression of *TP53*, and the highest expression of the proliferation marker *CDK1*.

As expected from the simulations above (Figure 37) hESCs showed a 4-log2-fold down-regulation in the expression of TGFβ receptor 2, even though *SMAD4* expression level remained comparable to non-transduced fibroblasts. Only in 4TF samples the levels of these TGFβ components were similar to non-transduced fibroblasts. In H9 as well as in 4TF the expression of p53 remain being 3-log2-fold higher than in non-transduced fibroblasts.

In overall, these results correlate with the simulations in the predicted high rates of proliferation and DNA damage responses of the cells in study. In addition, these results support the predictions based on the simulation experiments, which suggested that regulation of the TGFβ signalling pathway is an important feature of highly proliferative cells. Furthermore, the model was able to predict differences in the regulation of the pathway, and pin point the most relevant regulatory events.

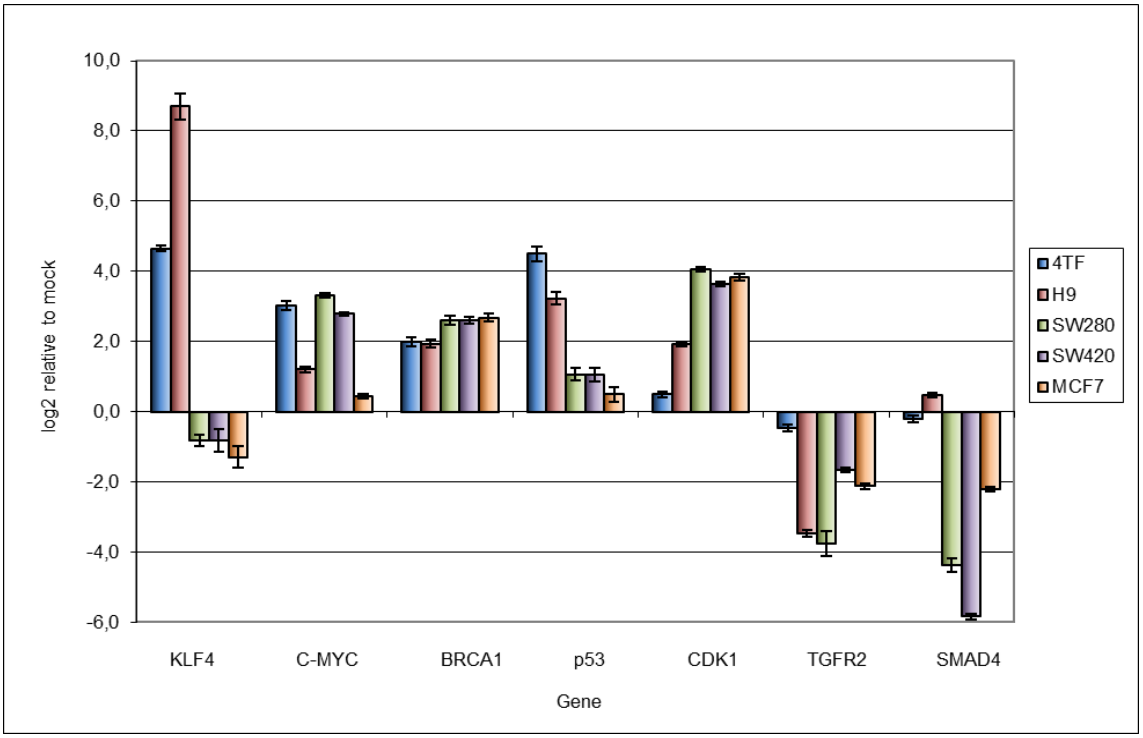

Figure 38. Comparison of the reprogramming process with cancer cell lines. Transcriptional regulation of the TGFβ signalling pathway and its down-stream targets measured by qRT-PCR. The relative amount of each transcript relative to non-transduced cells is scaled in the Y-axis that represents logarithmic fold changes. In the X-axis are the gene targets under analysis. The legend shows the samples that have been analysed. 4TF is the abbreviation used for fibroblasts over-expressing OCT4, SOX2, KLF4, and c-MYC. All three biological replicates hybridized onto the microarrays have been analysed. The amounts shown are the averaged of three different technical replicates run in each biological replicate. The error bars indicate the standard deviation of the average of all experimental replicates. Embryonic stem cell mRNA (H9) was used as a positive control for the primers.

6. Discussion

Somatic cells can be reprogrammed to iPS cells by the delivery of a few pluripotency-related transcriptional factors. Since the original description of iPS cells in Shinya Yamanaka´s landmark report (Takahashi and Yamanaka 2006), studies of transcription factor induced reprogramming to the iPS cell state have mainly branched into two fields of research.

First, no longer hindered by the technical and ethical limitation associated with somatic cell nuclear transfer (SCNT) and cell fusion, transcription factor mediated reprogramming provides a new avenue to investigate basic questions of cellular plasticity and pluripotency. Secondly, the iPS cell technology enables the derivation of patient- and disease-specific pluripotent stem cell lines, which has opened the door to disease modelling, drug discovery and cell replacement strategies.

The detailed events occurring between the time of exogenous expression of the reprogramming factors and the establishment of the iPS cell state remain largely uncharacterised. This is primarily due to the low efficiency and slow kinetics of the process and the fact that cells that will successfully complete the reprogramming process cannot be pre-selected.

To understand the molecular mechanisms underlying the induction of pluripotency it is necessary to determine the different elements involved and how they interplay to control cellular plasticity. Through the analysis of the transcriptome at the early stages of the reprogramming were identified the most important pathways involved in the reprogramming of somatic cells. The modelling approach undertaken here uncovered the functional meaning of those pathways acting co-ordinately. Below are discussed the results found, as well as, the most relevant technical hurdles faced up.

6.1. Virus-mediated over-expression

The comparison between the datasets obtained from the over-expression of the transcription factors (OCT4, SOX2, KLF4, and c-MYC) independently or in combination highlighted the roles of the different factors and the synergistic effects once they can act together. Nevertheless, the over-expression of GFP as a control for the methodology generated a considerable background that had to be considered in the analysis of the acquired data to interpret it effectively.

Even though, over-expression of proteins using virus-derived systems is a powerful tool for studying genetics, because of its high efficiency and the wide host

range of cells that can be infected, its side effects on the cells should not be underestimated. After infection stress responses are active in at least some of the cells and, those can modulate other cells behaviour by secretion of ligands into the medium. The viral envelope used (gp120; ENV) is known to trigger p53 signalling and other components of the system, like the reverse transcriptase or the regulatory protein Tat, interact and modulate the activity of p53 and cell cycle-related proteins (CDK4, CDK1) (Perfettini, Castedo et al. 2005). This is a plausible cause of the observed background after GFP over-expression (Figure 14).

Besides retroviral transduction, several methods have been developed for generating human iPS cells as DNA-transposition-based systems, transient plasmid delivery and integration/plasmid-free systems. Some of these methods are known to induce p53 responses and, to date lacks a systematic study to evaluate to which extent p53 signalling pathway activation is inherent to the reprogramming process, or is a technical side effect. The comparison of transcriptional data using alternative methods to viral-mediated over-expression can render interesting conclusions not solely about the implications of p53 in the process. They would also help to highlight the leading molecular mechanisms that otherwise remain intermingled with bystander processes.

Another approach able to reduce the method-related background, and would facilitate the investigation of the most relevant molecular mechanisms comes with the use of inducible vectors that in mouse allowed the creation of secondary iPS (see section 3.3.1.1). These secondary iPS cells allow the induction of the factors in a cell population that contain *a priori* a successful combination of the factors integrated in the genome. Viral components would not be added into the media and therefore it would be an appropriate system to avoid health risks and technical artefacts. The version of these vectors suitable for the infection of human cells was also built and made available on public sources, and even though a report was published (Hockemeyer *et al*. 2008), these experiments have not been reproduced yet.

6.2. Factor-specific regulation

The analysis of OCT4, SOX2, KLF4, and c-MYC induced transcriptional changes provided significant insight into the action of the reprogramming factors. While OCT4, SOX2, and KLF4 regulated pathways with a crucial role in development (see Figure 14), c-MYC changed the expression of more than half of the genes considered expressed, altering cell cycle and cellular metabolism drastically (see section 5.2.3). Since c-MYC and KLF4 had the closest profile in the cluster analysis to the combination of factors, I

expected these two factors to explain more of the regulatory signalling in the combination of factors (see Figure 12).

After over-expression of KLF4 two cellular processes, namely cell cycle and apoptosis, were principally affected. Inhibitors of cell cycle were found up-regulated, whereas apoptotic signalling was down-regulated. Hence, ectopic expression of this factor in human fibroblasts contributes to their survival, yet holding back their duplication in accordance with previously published results (Zhao, Hisamuddin et al. 2004). TGFβ and Wnt signalling pathways can be considered plausible candidates playing a role in these regulatory events, since they were found differentially regulated. Accurate prediction, at the level of single genes, of the regulatory outcome of the simultaneous regulation of these pathways is rather difficult at the moment, since the final effectors of these pathways SMAD- and TCF-transcription factors are known to bind different targets depending on the presence of other transcription factors and co-factors (Massague et al., 2000; Riese et. al., 1997).

KLF4 caused specific effects on the apoptotic cascade. Especially relevant is the down-regulation of *TP53* and its down-stream factors. These results are in agreement with published results, where KLF4 was found to bind to *TP53* promoter and repress its transcription (Rowland, Bernards et al. 2005). This regulation can have a large impact on cell survival and it is actually the foundation for the oncogenic properties of this factor (Zhao, Hisamuddin et al. 2004).

On the other hand c-MYC over-expressing fibroblasts acquired a high proliferation phenotype accompanied of obvious cell death. The broad transcriptional regulation of this factor, nonetheless well expected (Rahl, Lin et al. 2010), makes difficult to know which of the different pathways may facilitate reprogramming. Cell cycle, metabolic changes or DNA damage repair were between the most relevant processes affected. c-MYC is known to regulate the transcription of cell cycle components involved in the regulation of G_1 to S phase transition (Obaya, Mateyak et al. 1999).

Besides this, over-expression of this factor was able to increase the expression of enzymes essential for energy production, protein synthesis and nucleotide synthesis, without which cells cannot undertake the duplication of the genome. The large changes in biosynthetic pathways have been observed after oncogene over-expression or tumor-suppressor depletion (DeBerardinis, Lum et al. 2008; Jones and Thompson 2009). Glycolysis is then favoured and relatively little pyruvate is dispatched to the mitochondria that requires oxygen to process it. This favours the accumulation of glycolytic intermediates that can be used for nucleoside and amino acid synthesis. This energetic metabolism allows cells to adapt to hypoxia conditions and therefore is found in many cancer cells (DeBerardinis, Lum et al. 2008).

Highly proliferative cells have usually very active DNA damage response mechanisms that can repair, or alternatively, trigger senescence or apoptosis upon genome replication or segregation errors (Bartkova, Horejsi et al. 2005; Gorgoulis, Vassiliou et al. 2005). The over-expression of oncogenes like c-MYC is known to bring in a hyper-proliferation state that induces senescence in the cells by virtue of *p16* and *p21* activation (Di Micco, Fumagalli et al. 2006). c-MYC over-expression mimicks a long term pro-mitotic signalling that is known to stabilize p53, which may trigger apoptosis (Prendergast 1999). Nevertheless, p53-independent apoptotic mechanisms have been also characterized (Sakamuro, Eviner et al. 1995). It is therefore not surprising that the pathway analysis summarized in Figure 14 include repair mechanisms and the apoptotic cascade.

6.3. Combined expression: OCT4, SOX2, and KLF4

Some of the transcriptional regulation observed after simultaneous over-expression of OCT4, SOX2, and KLF4 recalled the profiles found after KLF4 over-expression as was the case for apoptotic signalling, yet some processes as DNA replication and certain repair mechanisms, namely homologous recombination, recapitulated the over-expression of c-MYC. Strikingly, synergistic effects of these factors achieved to restore a cell cycle that was otherwise compromised by KLF4-induced senescence. The increased of replication components, which achieved levels close to fast proliferative cells, could be due to an inactive G_1-S checkpoint and the disappearance of its anti-growth effects. This would explain the absence of replication-related repair mechanisms such as mismatch repair, nucleotide excision repair, or base excision repair, which were observed otherwise following c-MYC or GFP over-expression (see Figure 14). Checkpoint deficiency is one distinctive feature of ESCs in opposition to differentiated cell types and it is suspected to be critical to maintain pluripotency (Boheler 2009).

Two main signalling pathways has been implicated in stopping the cell cycle upon replication stress, the p53 and the TGFβ signalling pathway (Hartwell and Kastan 1994; Montgomery, Goggins et al. 2001). These two pathways are linked by TGFβ receptors that can activate stress-related MAP kinases, which can induce p53 stabilization (Mu, Gudey et al. 2011). In fact the analysis of both signalling pathways after 3TF over-expression showed remarkable features (see section 5.2.4). Strikingly, *TP53* expression was up-regulated (see section 5.3.1), while components of the apoptotic cascade that can be largely dependent on p53, were down-regulated in comparison to control cells (see section 5.3.2). While the expression profile of pro-apoptotic markers recalled KLF4-

over-expressing fibroblasts, the expression of *TP53* achieved the levels found in c-MYC samples (see section 5.3.1). On the other hand the over-expression of 3TF showed a different profile in the expression of TGFβ pathway components and their transcriptional co-activators that did not recall the profile of any of the factors when they were expressed independently (see section 5.2.4).

Because p53 stability and function depend on its post-translational modifications (Savic, Yin et al. 2009), it is reasonable to postulate that the synergistic effects of the factors change the expression of pathway components and/or their co-activators avoiding subsequently activation of proteins involved in anti-growth mechanisms including post-translational activators of p53. Since the last effectors of the TGFβ pathway, the SMADs, can activate together with other transcription factors cell cycle inhibitors (Massague and Wotton 2000) and the TGFβ receptors can activate stress-related MAP kinases able to stabilize p53, a plausible mechanism by which the reprogramming factors avoid both anti-proliferative effects is the down-regulation of the receptor/s.

The impairment of *TP53* transcription did cause a reduction in the levels of cell cycle inhibitors, but did not affect the expression of apoptotic markers. Thus, in the context of reprogramming *TP53* participates in the induction of *p16* or *p21* and the down-regulation of G_1-S phase Cyclins. In contrast, *TP53* does not promote any longer cell death after the over-expression of 3TF, since apoptotic markers remained unaffected after p53 knock-down. Hence, cells might be impaired to duplicate, but are protected from p53-dependent apoptosis, which recapitulates KLF4 over-expression. These regulatory effects on the apoptotic signalling observed after 3TF and KLF4 over-expression deserves further research, since they could promote oncogenic properties in iPS cells and its derivatives.

Other researches worked before on the role of *TP53* in the reprogramming process, linking it to the efficiency of the process (Zhao, Yin et al. 2008). Different approaches have used to impair *TP53* function during the reprogramming process. Short-interfering RNA (Hong, Takahashi et al. 2009) or *TP53* null fibroblasts (Li, Collado et al. 2009; Kawamura, Suzuki et al. 2009) have been used to abrogate *TP53* activity during the generation of iPS cells, and to demonstrate that the *TP53* pathway is a roadblock of the process. In the report of Hong et al. the role of different down-stream targets of p53 was also analysed. They could demonstrate that it is *p21* the down-stream factor that limits at the most the generation of iPS cells. Conversely, the up-regulation of MDM2, a p53 inhibitor, enhances the efficiency of reprogramming (Hong, Takahashi et al. 2009). Other researchers backed up their results with similar findings (Kawamura, Suzuki et al. 2009). It is important to notice that permanent suppression of p53 causes a

higher rate of genomic instability in the iPS cells generated, constraining their use in further applications (Hong, Takahashi et al. 2009).

The inhibition of the TGFβ receptors undertaken in this work was meant to mimick the regulatory effect of the combination of factors on p53. The inhibition of the receptors did cause a transcriptional reduction of one of the MAP stress-related kinases, p38. Other MAP stress-related kinases such as JNK were not affected. Thus, inhibition of TGFβ/Activin receptors can indeed contribute to the activation of stress-related kinases that activate p53. The regulatory events induced by the reprogramming factors on the TGFβ pathway shall be at least partially responsible for the lower activity of p53 in those cells. Actually, TGFβ receptors over-expression has been reported to decrease the efficiency of reprogramming (Li, Liang et al. 2010). In order to mimick completely in c-MYC samples the regulation of *TP53* signalling observed following the over-expression of KLF4 or 3TF might be necessary to perturb simultaneously the Wnt pathway, since both pathways were simultaneously regulated after KLF4 or 3TF over-expression (see Figure 14).

Besides TGFβ signalling mediated activation of *TP53*, other mechanisms could be also contributing to suppress p53-induced death, such as changes in other regulators that activate post-transcriptionally p53, repression of co-factors required for p53 binding on the promoters of pro-apoptotic genes, or inhibition of effector proteins involved in apoptosis. In this regard, elevated expression of the p53 inhibitor MDM2 overcomes the tumour suppression function of p53 that is associated to its ability to induce apoptosis (Freedman, Wu et al. 1999). Moreover, since p53 requires co-factors such as ASPP1/2, JMY, p63 or p73 to enhance its ability to activate pro-apoptotic genes (Samuels-Lev, O'Connor et al. 2001; Shikama, Lee et al. 1999; Flores, Tsai et al. 2002), down-regulation of these co-factors can be very well expected to decrease p53-mediated apoptosis (Slee, O'Connor et al. 2004). Alternatively, down-regulation of effector proteins such as APAF-1 or CASP3 can confer cells resistance to apoptosis (Soengas, Capodieci et al. 2001). Similarly, up-regulation of pro-survival genes of the BCL2 family of proteins can impair the apoptotic cascade and contribute to hinder p53-induced apoptosis (Schmitt, Fridman et al. 2002).

However, the TGFβ pathway has a major role in the maintenance of the pluripotent state in hESCs, although. Human ESCs have been initially cultivated with mouse fibroblasts acting as feeders (Thomson, Itskovitz-Eldor et al. 1998), or in conditioned media by those cells (Xu, Inokuma et al. 2001). The cytokines secreted in the media sustain undifferentiated stem cells and have been extensively investigated. It has been shown that the activation of the TGFβ/Activin/Nodal pathway and its down-stream effectors, SMAD2/3, can under defined conditions, sustain the growth of human

embryonic pluripotent cells without the need of feeders or conditioned media (Xu, Rosler et al. 2005). Either bFGF (Greber, Lehrach et al. 2008) or Activin A (Beattie, Lopez et al. 2005) have been successfully used to keep high levels of phosphorylated SMAD2/3 in hESCs (Greber, Lehrach et al. 2008). Because phosphorylated SMAD2/3 are absolutely necessary for the maintenance of the pluripotent state, cells that will be successfully reprogrammed have to find the means to activate SMAD2/3 without the intervention of TGFβ receptors, for example through Activin or FGF receptors.

6.4. Combined expression: OCT4, SOX2, KLF4, and c-MYC

The addition of c-MYC increases drastically the efficiency and the kinetics at which fibroblasts convert into pluripotent cells (Takahashi and Yamanaka 2006; Takahashi, Tanabe et al. 2007). Consequently the genetic regulatory network induced by the four-factor combination shall highlight essential aspects of the reprogramming process. The transcriptional profile of the cells after the combination of four factors can be essentially explained with the combination of effects of the three factor combination and the metabolic changes induced by c-MYC. Hence, cells show high competence to replicate their DNA as the three factor combination, but with a larger activation of components that directly trigger progression from G_1 to S phase. Additionally, they acquired the anabolic metabolism needed for the synthesis of proteins, nucleic acids and lipids. Furthermore, cells were more resistance to certain apoptotic cues and showed a larger activation of homologous recombination components.

The genetic regulatory network established by 4TF was prominently characterized by cell cycle regulation. It is to remark although, that microarrays like highly curated databases are often based in well-established knowledge. This means that the retrieved information can be very well bias depending on the extent of knowledge that biological sciences have on a particular field and cell cycle is among the first processes studied in cellular biology.

However to acquire stem cell-like properties somatic cells must undergo through remodelling of the mechanisms controlling cell cycle, to which c-MYC is a major contributor (Obaya, Mateyak et al. 1999). Human ESCs cycle is shorter and with just brief growth phases, controlled mainly by S-phase and M-phase initiating complexes (Becker, Ghule et al. 2006). Cells in the G_1 phase are particularly responsive to intracellular and extracellular signals based on which they decide their fate. Therefore, hESCs not just accelerate their division rate by minimizing G_1 phase duration, but can also avoid differentiation signalling that are active during the early G_1 phase. Accordingly, the analysis of cell cycle related genes after 4TF over-expression showed an up-

regulation of G_1-S transition Cyclins and replication-related genes as it does in human ESCs, where c-MYC participates in its regulation. These results agree with the initial observation that highly proliferative cells are more readily to be reprogrammed, which has been the topic of some publications (Ho, Chronis et al. 2011; Ruiz, Panopoulos et al. 2010).

The addition of c-MYC to the reprogramming cocktail can achieve a germ-cell like program on the cell cycle (see section 5.3.3). Importantly, this leads in somatic cells to p53-mediated senescence or apoptosis, but in combination with OCT4, SOX2 and KLF4 the outcome was much reduced than expected. This correlates with results published for ESCs, where expression of CDK inhibitors as p21 is barely detectable (Becker, Ghule et al. 2006). In fact, pharmacological activation of p53 that led to the up-regulation of p21 was shown to increase the number of cells in G_1 phase and the levels of differentiation (Maimets, Neganova et al. 2008). It has previously been demonstrated that apoptotic cues originated by cell cycle checkpoint proteins are absence in hESCs, which would explain the levels of karyotype abnormalities observed in those cells (Mantel, Guo et al. 2007).

Notably, the catabolic processes induced after c-MYC over-expression were also found, when it was expressed with the other factors. Cells can then fulfil the higher metabolic demands that faster duplicating cells have. The metabolic changes observed shifts the ATP production toward glycolysis instead of oxidative phosphorylation (Saretzki, Armstrong et al. 2004), which reduces the amount of reactive oxygen species, by-product of oxidative phosphorylation that are the main source of endogenous DNA damage. To this regard, hESCs are sensitive to the level of reactive oxygen species that may induce differentiation and base their ATP production principally in glycolysis (Ito, Hirao et al. 2006).

High proliferative rates bring about necessarily DNA damage induced apoptosis and senescence in healthy physiological conditions. Double strand break is considered to be very toxic for the cells and even oven one single break can be lethal (Rich, Allen et al. 2000). Different DSB repair mechanisms can repair double strand breaks, mainly non-homologous end joining and homologous recombination. The choice between these pathways relies between other factors on cell type. The data presented here shows that following over-expression of the reprogramming factors, either 3TF as 4TF, cells has a strong preference for the homologous recombination pathway.

Homologous recombination uses as template the replicated chromatid and it is therefore happening after S phase (Haber 2000; Rothkamm, Kruger et al. 2003), while non-homologous end joining happens in G_1 and early S phase and is an error-prone process leading more often to loss of genetic information (Lees-Miller and Meek 2003;

Lieber, Ma et al. 2003). Human ESCs are much less tolerant to inaccurate repair than differentiated cell types and spend most of the cell cycle in S phase, where a sister chromatid is present. It is therefore believed that homologous recombination is the DSB system repair of choice in hESCs. After over-expression of the four factors and with the abolition of G_1-S checkpoint, cells show also preference for homologous recombination repair. This could also be consequence of the regulatory events around p53, since it has been demonstrated that inactivation of p53 results in much higher rates of homologous recombination events (Mekeel, Tang et al. 1997).

In overall, the reprogramming of somatic cells begins in the pursuit of an embryonic-like cell cycle. The synergistic effect of the reprogramming factors (OCT4, SOX2, KLF4, and c-MYC) regulates the TGFβ and the p53 signalling pathway reducing their related anti-growth responses, barriers to reprogramming (see Figure 39). These regulatory events eliminate the G_1-S transition checkpoint promoting a cell cycle progression similar to ESCs. Cell cycle progression is further supported by reorganization of the metabolic activity that shunts metabolites into biosynthetic pathways. This regulation of the cell cycle has down-stream effects on DNA damage responses and apoptosis. Faster G_1-S transitions reduced the activity on replication-related repair mechanisms in favour of homologous recombination that becomes the most prominent repair mechanism in the cells. Furthermore, cells show thereafter no apoptotic signalling derived from the G_1-S checkpoint transition.

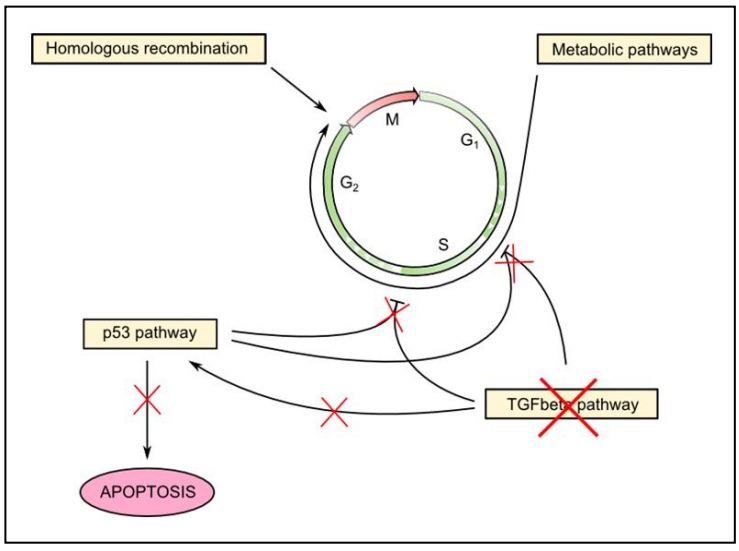

Figure 39. Induced regulatory network. Overview of the most remarkable features induced by the reprogramming factors on the genetic regulatory network of human fibroblasts.

6.5. Mathematical modelling

In order to interpret the functional relevance of the different nodes and edges that form a genetic regulatory network, it is necessary to find a mathematical representation to any given interaction between the components. While the kinetic laws presented are not able to model complex reactions found in the cell, they capture the intuitive ideas about transcriptional activation and repression, protein synthesis and degradation, phosphorylation and dephosphorylation. Large networks bring into play many kinetic parameters that are very often unknown, although. The modelling approach undertaken here can surmount this problem and predict the outcome of complex genetic regulatory networks. Furthermore, multifaceted perturbation experiments can be performed and complicated findings or hypothesis can be recapitulated.

A mathematical model can be fed with the expression levels of the different components and can be used to understand the underlying genetic regulatory network in the context of this particular situation, for example a certain cell type, developmental state, disease or inter-species differences. This is therefore a modelling approach especially suited to characterize the similarities and differences between two or more different states and underline the functional consequences of the differences.

Nevertheless, caution should be taken when interpreting the results, since models are always simple representations of more complicated events. Furthermore in this sort of biological models there is in addition a large amount of unknown factors that do not allow estimating faithfully the outcome of simulated experiments. While large models have to wait for further knowledge (e.g. kinetic rates) to predict accurately cell behaviour, they can be used effectively to analyse to which extend a network can explain the phenotypes observed, or how the diferent components of the network influence them. In other words, different pathways or processes might be thought to result in a certain phenotype, and a mathematical model can assess if those pathway are enough to explain the phenotype, or other pathway and processes need to be considered. What's more a mathematical model can also serve to determine the relevance of each of the components to explain the phenotype.

Here particularly the layout of the genetic regulatory network helped to understand the essence of the roadblocks to reprogramming. The model was useful to devise experiments in order to surmount those barriers, or alternatively, to demonstrate active mechanisms triggered by the reprogramming factors that accomplish the same goal. Hence, simulation of *TP53* inactivation suggested that this tumor-suppressor can be a cornerstone for reprogramming. Its inhibition (or inactivation) was predicted to increase the survival and proliferation capabilities of the cells, which correlated with the

phenotype observed after over-expression of the reprogramming factors. The activation of *TP53* can be induced by the DNA damage response systems of the cells and/or by non-SMAD TGFβ signalling. The model suggested that the down-regulation of the TGFβ receptor/s can be a powerful mechanism to reduce the levels of active p53 and indeed the over-expression of the reprogramming factors achieved reduced levels of active p53 by regulating TGFβ signalling (see section 5.4.3.2), enabling in this way the reprogramming process.

Thus, the genetic regulatory network established by OCT4, SOX2, KLF4 and c-MYC drives the molecular structure, function and behaviour of the donor cell. The reprogramming factors induce, establish and regulate the interactions between the different components of the cells, including miRNAs and chromatin regulating enzymes, and mathematical models can be an excellent tool to understand their complex interactions and predict complex behavioural outcomes.

Nonetheless model assembly and prediction potential can be troublesome. One regulatory motif is extremely relevant to maintain induced pluripotency that is the positive feed forward loop of the core transcriptional machinery of stem cells (OCT4, SOX2). This regulatory motif is such a key element, because maintain the cells in the pluripotent state without the need of exogenous factors. Otherwise, pluripotency should be maintained by the exogenous factors that induced it, and they would hinder afterwards their differentiation to other cell types. The modelling of such a regulatory motif could not be implemented in the current model because no steady state is achieved. To complete it, is necessary to know how are regulated the total amounts of OCT4 and SOX2 in the cell.

The model shall require another regulatory loop to explain *p15* activation (see section 5.2.2). The TGFβ signalling pathway is known to participate in *p15* activation through binding of SMAD2/3 on its promoter. Motifs of other transcription factors can be also found in its promoter and are expected to regulate its transcription (e.g. Wnt pathway), even though I could not found experimental data in this regard. In the model presented here a stable activation of *p15* requires the activity of TGFβ pathway. Since the TGFβ signalling pathway was found down-regulated the regulatory outcome based on the regulatory information available was the down-regulation of *p15*.

There are different alternatives to explain the up-regulation of *p15* in the absence of TGFβ signalling. Additional players on its regulation shall be in most cases included, since p15 is not a transcription factor that could regulate positively its own transcription. Thus, a different signalling pathway might be responsible for the activation and stable expression of this CDK-inhibitor. Alternatively, the SMAD2/3 could be inducing p15 and be thereafter down-regulated. In this case, a different transcription factor/s would be then

stabilizing the expression of *p15*. In any case additionally regulators are necessary to recapitulate *p15* stable expression.

6.6 Reprogramming and cancer

The reprogramming process is to some extent the transformation of slow proliferative cells to fast Cycling cells. As such, cellular barriers to an uncontrolled proliferation will be triggered and the outcome senescence or apoptosis. In order to overcome these barriers proper genetic information processing, metabolic needs and cell communication requirements shall be met. Evasion of the regulatory systems is possible but leads to pathology in multicellular organisms (Hanahan, Weinberg 2011).

Some of the molecular mechanisms found the reprogramming process, recall colon rectal cancers, in which successive mutations acquire on proliferative (e.g. *ERAS*, *RAF*) and apoptosis related genes (e.g. *TP53*, *BAX*), and inactivation of the TGFβ pathway (e.g. *SMAD4*, *TGFβR2*) and the uncontrolled proliferation of cells (Guda, Natale et al. 2009). In colon rectal cancers there is a low expression of *KLF4*, but other members of this family of transcription factors are highly expressed: *KLF5* or *KLF11*.

c-MYC plays a critical role in cell proliferation and tumorogenesis in those cells and its down-regulation by transforming growth factor-ß signalling is necessary for TGFβ to inhibit cell proliferation. KL5, on the other hand, acts as pro-proliferative transcription factor that reverses function to become an anti-proliferative TGFβ co-factor upon TGFβ stimulation. These two transcription factors are thought to be main drivers of colon cancer cells (Guo and Wang 2009) and do play a major role in the gene regulatory network in early stages of reprogramming.

The comparison of the simulations among cell lines brought insides about the common and specific molecular mechanisms that they share (see section 5.4.4). Cells along the reprogramming process achieve a pronounced deregulation of the TGFβ pathway altering the expression of co-activators and co-repressors (see section 5.2.4), avoiding so directly and indirectly up-regulation of CDK-inhibitors and stress-related kinases, while colon cancer restrains the anti-growth properties of this pathway by down-regulating key components of the pathway (e.g. *SMAD4* and *TGFR2* in section 5.4.4.1). As a matter of fact, TGF-β type II receptor is inactivated by mutation in most human gastrointestinal cancers with microsatellite instability and *SMAD4* wild type function is lost in nearly half of all pancreatic carcinomas (Montgomery, Goggins et al. 2001). On the other hand, hESCs grow thanks to an active TGFβ pathway, even though signalling induced by TGFβR2 is also minimized suggesting that this is necessary for high

proliferation rates. SMAD2/3 activation is then necessarily induced by other receptors of the family (e.g. Activin receptors).

Survival mechanisms showed also interesting differences between the cell lines and the cells in which reprogramming factors were over-expressed. The simulations results suggest that the up-regulation of different survival proteins is a common mechanism between hESCs and 4TF cells. Again, colon cancer stem cells counteract apoptotic signalling with the up-regulation of a single but key pro-survival protein (see section 5.4.4). Even though, both mechanisms remain as powerful strategies to surmount apoptosis, a multifaceted up-regulation seems to offer better protection because different survival mechanisms shall be activated depending on the nature of the apoptotic cues.

Signalling pathways independent of *TP53* can autonomously stop the cell cycle or trigger cellular death. The genetic regulatory network at the early stages of reprogramming is largely affected by those pathways, and can be very well expected to hinder cell cycle progression due to anomalies in DNA duplication and segregation. Human ESCs and colon cancer cells showed also high activity in those pathways, supporting the need of efficient repair mechanisms.

These data should raise concerns about the use of these cells for regenerative therapies. Cells to be reprogrammed are under strong selective pressure and chromosomal abnormalities, CNVs or mutations in genes involved in the process will fix preferentially in the genome. These kinds of events have been observed before for hESCs in continuous cell culture growth (Baker, Harrisson et al. 2007). Thus, if the levels of oncogenes as *c-MYC* or tumor-suppressors as *TP53* are directly related with the efficiency of the process, genetic defects shall accumulate faster on those genes during or after the reprogramming process. Cells with these kinds of genetic defects will make them more prompt to form cancers and unsafe for cell therapy.

Nonetheless, it is clear that reprogramming and oncogenic transformation share similarities. Thus, unravelling of the molecular mechanisms behind the induction of pluripotency can greatly help the cancer field. Cancer hallmarks as unlimited proliferation, evasion of apoptosis or metastasis can come at some point from embryonic features that differentiated types acquired due to genetic or genomic defects. The reprogramming experiment offers therefore a platform of inestimable value to investigate the acquisition of such oncogenic properties. Moreover, because different cells can be reprogrammed, the analysis can be extended to many different cell types that can help to understand oncogenic transformation in different tissues.

6.7 Reprogramming and other diseases

As pointed out above regulation of cell cycle and apoptosis are pivotal issues in the reprogramming process. Anomalies in these processes are a known feature of other diseases such as Alzheimer's, Parkinson's, and Huntington's disease (Takuma, Yan et al. 2005; Thomas and Beal 2007; Sawa, Tomoda et al. 2003). Some molecular features observed in the reprogramming of somatic cells could also help to understand these diseases.

The over-expression of reprogramming factors abolished a cell cycle checkpoint, to some extent by regulating p53 activators such as p38. This same mechanism that controls fibroblasts proliferation could be active in neurons that do not divide in healthy conditions. Actually, cell cycle re-entry is a common phenotype of different neurodegenerative diseases (Krantic, Mechawar et al. 2005). Thus, the lost or deregulation of cell cycle checkpoint components, such as p53, could very well explain the proliferative irregularities observed in these neuropathologies. As a matter of fact, p38 has been shown to associate with the tau protein that is abnormally phosphorylated in Alzheimer disease (Zhu, Rottkamp et al. 2000). If this interaction impairs p38 regulation upon p53, the same mechanism that is driving fibroblasts reprogramming, could be driving neuronal cell death.

Besides this, other pathological features of these diseases can benefit from further knowledge in the molecular mechanisms involved in the reprogramming of somatic cells. For instance, p53 is known to have deleterious effects on Huntington's disease (Ryan, Zeitlin et al. 2006; Feng, Jin et al. 2006). Its deletion has been shown to have protective effects in mouse models of Huntington disease (Bae, Xu et al. 2005). In this case, further knowledge on how the reprogramming factors regulate p53-mediated apoptosis could serve to devise strategies to avoid undesirable death of neurons in Huntington's disease.

Furthermore, mitochondrial dysfunction is known to be involved in the pathogenic cell death common to these diseases (Facecchia, Fochesato et al. 2011). The impairment of the mitochondrial electron transport chain is a known cause of mitochondrial dysfunction, and its associated oxidative stress (Orsucci, Filosto et al. 2009). The regulatory network established after over-expression of the reprogramming factors show a strong regulation of respiratory chain proteins. Molecular details of how mitochondria components are regulated can just help to devise strategies to diminish or eliminate the detrimental effects observed neurodegeneration.

6.8 Novelty and impact of the work

In overall, this work sets for the first time a framework to understand the reprogramming process from a systems biology point of view. Considering the induction of pluripotency as the outcome of many different cues that shall be coherent with the known cellular biology. Up to now, cell cycle was known to be a central process on the reprogramming but no other researcher have been able to relate to the loss G_1-S transition checkpoint, and pin point in detail the most relevant molecular events. Also DNA damage responses were thought to play a role following over-expression of the reprogramming factors, but until now the functional meaning of the changes observed had not been explained. Furthermore, this work opens a new line of investigation regarding the regulation of apoptotic pathways involved in the reprogramming process, from which disease biology can take great profit. Besides this, it is the first time in the reprogramming literature that metabolic reprogramming has been mentioned and studied.

In addition it presents a general mathematical model for cell cycle progression that can be used to analyse similarities and differences between different cell types, disease processes or developmental stages. The approach that I started allows to analyse further published data from different sources and to undertake new kind of analysis. The vast amount of transcriptional profiling done during the last decade can be compared, even being in different platforms or carried out by different research groups. Thanks to the simulation strategy similarities can be found between molecular mechanisms or functional features between different datasets that are otherwise very difficult to interpret. Once those similarities have been found small or large scale experiments done in one case study can be simulated in the other and its impact estimated. This kind of analysis will help to optimize bench work, speeding up biologically relevant results and reduce research costs.

7. Abbreviations

3TF/3T	OCT4, SOX2, and KLF4
4TF/4T	OCT4, SOX2, KLF4, and c-MYC
BMP	Bone Morphogenic Protein
BSA	bovine serum albumin
cDNA	Complementary Deoxyribonucleic Acid
ChIP	Chromatin Immunoprecipitation
CM	Conditioned Media
c-Myc	v-myc Myelocytomatosis Viral Oncogene Homolog
Ct	Threshold Cycle
CTD	Carboxy-terminal Domain
dH2O	Distilled Water
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
Dnmt	*De novo* methyl transferases
E. coli	*Escherichia coli*
EBNA1	Epstein-Barr Nuclear Antigen-1
ESC/s	Embryonic Stem Cell/s
FBS	Fetal Bovine Serum
FGF2	Fibroblast Growth Factor 2
GFP	Green Fluorescence Protein
GTFs	General Transcription Factors
H3K27me3	Histone 3 Lysine 27 Trimethylation
H3K4me3	Histone 3 Lysine 4 Trimethylation
H3K9	Histone 3 Lysine 9
H3K9me3	Histone 3 Lysine 9 Trimethylation
hESC/s	Human Embryonic Stem Cell/s
HEK	Human Embryonic Kidney
HMG	High Mobility Group
ICM	Inner Cell Mass
IF	Immunofluorescence
iPS	Induced Pluripotent Stem
Klf4	Krüppel-like Factor 4
LTR	Long Tandem Repeats

MEF	Mouse Embryonic Fibroblast
mESC/s	Mouse Embryonic Stem Cell/s
miRNA/s	MicroRNA/s
MMLV	Moloney Murine Leukemia Virus
NSC	Neural Stem Cells
Oct4	Octamer-binding Transcription Factor 4
ODE	Ordinary Differential Equations
oriP	Latent Origin DNA Replication
ORF	Open Reading Frame
PB	Piggyback
PBS	Phosphate Buffered Saline
PcG	Polycomb Group
PCR	Polymerase Chain Reaction
piPS	Protein-induced Pluripotent Stem
Pol II	Polymerase II
p-TEFb	Positive Transcription Elongation Factor b
qRT-PCR	Quantitative Real Time-Polymerase Chain Reaction
RNA	Ribonucleic Acid
RT-PCR	Real Time-Polymerase Chain Reaction
SCNT	Somatic Cell-nuclear Transfer
Sox2	SRY (sex-determining region Y)-box 2
ß-ACT	ß-Actin
TBS	Tris Buffered Saline
TBST	Tris Buffered Saline Tween
TFs	Transcription Factors
TGFβ	Transforming Growth Factor ß
TrxG	Tritorax Group
TS	Trophoblast Stem
TSS	Transcription Start Site
TuJ1	Smooth Muscle Actin
UM	Unconditioned Media
VPA	Valproic Acid
VSVg	Vesicular Stomatitis Virus glycoprotein

8. Semantics

x	times
%	percent
°C	degrees Celsius
µg	micrograms
µL	microlitre
µM	micromolar
g	grams / gravity
hr/hrs	hours
M	mol
mg	milligram
min/’	min
ml	millilitre
mm	millimetre
mM	millimolar
ng	nanogram
nm	nanometre
s/”	second
U	unit
V	volt
vs.	versus
bp	base pair

9. References

Aasen, T., A. Raya, et al. (2008). "Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes." Nat Biotechnol **26**(11): 1276-1284.

Anokye-Danso, F., C. M. Trivedi, et al. (2011). "Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency." Cell Stem Cell **8**(4): 376-388.

Aoi, T., K. Yae, et al. (2008). "Generation of pluripotent stem cells from adult mouse liver and stomach cells." Science **321**(5889): 699-702.

Avilion, A. A., S. K. Nicolis, et al. (2003). "Multipotent cell lineages in early mouse development depend on SOX2 function." Genes Dev **17**(1): 126-140.

Azuara, V., P. Perry, et al. (2006). "Chromatin signatures of pluripotent cell lines." Nat Cell Biol **8**(5): 532-538.

Banito, A., S. T. Rashid, et al. (2009). "Senescence impairs successful reprogramming to pluripotent stem cells." Genes Dev **23**(18): 2134-2139.

Baker D.E., N.J. Harrison, et al. (2007). "Adaptation to culture of human embryonic stem cells and oncogenesis in vivo." Nature Biotechnology **25**(2):207-15.

Bartkova, J., Z. Horejsi, et al. (2005). "DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis." Nature **434**(7035): 864-870.

Beattie, G. M., A. D. Lopez, et al. (2005). "Activin A maintains pluripotency of human embryonic stem cells in the absence of feeder layers." Stem Cells **23**(4): 489-495.

Becker, K., P. Ghule, et al. (2006). "Unique cell cycle kinetics of human embryonic stem cells." Journal of Bone and Mineral Research **21**: S390-S390.

Becker, K. A., P. N. Ghule, et al. (2006). "Self-renewal of human embryonic stem cells is supported by a shortened G1 cell cycle phase." J Cell Physiol **209**(3): 883-893.

Belmonte, J. C. I., F. Gonzalez, et al. (2009). "Generation of mouse-induced pluripotent stem cells by transient expression of a single nonviral polycistronic vector." Proceedings of the National Academy of Sciences of the United States of America **106**(22): 8918-8922.

Bernstein, B. E., T. S. Mikkelsen, et al. (2006). "A bivalent chromatin structure marks key developmental genes in embryonic stem cells." Cell **125**(2): 315-326.

Berridge M.J., M.D. Bootman, et al. (2003). “Calcium signalling: dynamics, homeostasis and remodelling” Nat Rev Mol Cell Biol **4**:517-29.

Bilodeau, S., M. H. Kagey, et al. (2009). "SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state." Genes Dev **23**(21): 2484-2489.

Blüthgen, N., S. Legewie, et al. (2009). "A systems biological approach suggests that transcriptional feedback regulation by dual-specificity phosphatase 6 shapes extracellular signal-related kinase activity in RAS-transformed fibroblasts." FEBS J **276**(4): 1024-1035.

Boheler, K. R. (2009). "Stem cell pluripotency: a cellular trait that depends on transcription factors, chromatin state and a checkpoint deficient cell cycle." J Cell Physiol **221**(1): 10-17.

Bolouri, H. and E. H. Davidson (2010). "The gene regulatory network basis of the "community effect," and analysis of a sea urchin embryo example." Dev Biol **340**(2): 170-178.

Bonnon, C. and S. Atanasoski (2011). "c-Ski in health and disease." Cell Tissue Res.

Borisov, N., E. Aksamitiene, et al. (2009). "Systems-level interactions between insulin-EGF networks amplify mitogenic signaling." Mol Syst Biol **5**: 256.

Boyer, L. A., T. I. Lee, et al. (2005). "Core transcriptional regulatory circuitry in human embryonic stem cells." Cell **122**(6): 947-956.

Boyer, L. A., K. Plath, et al. (2006). "Polycomb complexes repress developmental regulators in murine embryonic stem cells." Nature **441**(7091): 349-353.

Bracken, A. P., N. Dietrich, et al. (2006). "Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions." Genes Dev **20**(9): 1123-1136.

Brown, P. O. and D. Botstein (1999). "Exploring the new world of the genome with DNA microarrays." Nat Genet **21**(1 Suppl): 33-37.

Campbell, K. H., J. McWhir, et al. (1996). "Sheep cloned by nuclear transfer from a cultured cell line." Nature **380**(6569): 64-66.

Capdevila, J. and J.C. Izpisúa Belmonte (1999) "Extracellular modulation of the Hedgehog, Wnt and TGF-beta signalling pathways during embryonic development." Curr Opin Genet Dev **9**(4): 427-33.

Cartwright, P., C. McLean, et al. (2005). "LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism." Development **132**(5): 885-896.

Chambers, I., D. Colby, et al. (2003). "Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells." Cell **113**(5): 643-655.

Chambers, I. and A. Smith (2004). "Self-renewal of teratocarcinoma and embryonic stem cells." Oncogene **23**(43): 7150-7160.

Chen, T., H. L. He, et al. (1999). "Modeling gene expression with differential equations." Pac Symp Biocomput: 29-40.

Chen, X., M. J. Rubock, et al. (1996). "A transcriptional partner for MAD proteins in TGF-beta signalling." Nature **383**(6602): 691-696.

Chen, X., H. Xu, et al. (2008). "Integration of external signaling pathways with the core transcriptional network in embryonic stem cells." Cell **133**(6): 1106-1117.

Cherry, S. R., D. Biniszkiewicz, et al. (2000). "Retroviral expression in embryonic stem cells and hematopoietic stem cells." Mol Cell Biol **20**(20): 7419-7426.

Chickarmane, V. and C. Peterson (2008). "A computational model for understanding stem cell, trophectoderm and endoderm lineage determination." PLoS One **3**(10): e3478.

Chickarmane, V., C. Troein, et al. (2006). "Transcriptional dynamics of the embryonic stem cell switch." PLoS Comput Biol **2**(9): e123.

Ciccia, A. and S. J. Elledge (2010). "The DNA damage response: making it safe to play with knives." Mol Cell **40**(2): 179-204.

Cobrinik, D. (2005). "Pocket proteins and cell cycle control." Oncogene **24**(17): 2796-2809.

Cole, M. F., S. E. Johnstone, et al. (2008). "Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells." Genes Dev **22**(6): 746-755.

Cowan, C. A., J. Atienza, et al. (2005). "Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells." Science **309**(5739): 1369-1373.

D'Haeseleer, P., X. Wen, et al. (1999). "Linear modeling of mRNA expression levels during CNS development and injury." Pac Symp Biocomput: 41-52.

Davis, R. L., H. Weintraub, et al. (1987). "Expression of a single transfected cDNA converts fibroblasts to myoblasts." Cell **51**(6): 987-1000.

DeBerardinis, R. J., J. J. Lum, et al. (2008). "The biology of cancer: metabolic reprogramming fuels cell growth and proliferation." Cell Metab **7**(1): 11-20.

Dejosez, M., S. S. Levine, et al. (2010). "Ronin/Hcf-1 binds to a hyperconserved enhancer element and regulates genes involved in the growth of embryonic stem cells." Genes Dev **24**(14): 1479-1484.

Deveraux, Q. L., H. R. Stennicke, et al. (1999). "Endogenous inhibitors of caspases." J Clin Immunol **19**(6): 388-398.

Di Micco, R., M. Fumagalli, et al. (2006). "Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication." Nature **444**(7119): 638-642.

Dimos, J. T., K. T. Rodolfa, et al. (2008). "Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons." Science **321**(5893): 1218-1221.

Donovan, M. and T. G. Cotter (2004). "Control of mitochondrial integrity by Bcl-2 family members and caspase-independent cell death." Biochimica Et Biophysica Acta **1644**(2-3): 133-147.

Dysvik, B. and I. Jonassen (2001). "J-Express: exploring gene expression data using Java." Bioinformatics **17**(4): 369-370.

Elick, T. A., N. Lobo, et al. (1997). "Analysis of the cis-acting DNA elements required for piggyBac transposable element excision." Mol Gen Genet **255**(6): 605-610.

Endoh, M., T. A. Endo, et al. (2008). "Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity." Development **135**(8): 1513-1524.

Facecchia, K., L. A. Fochesato, et al. (2011). "Oxidative toxicity in neurodegenerative diseases: role of mitochondrial dysfunction and therapeutic strategies." J Toxicol 2011:683728.

Feldman, N., A. Gerson, et al. (2006). "G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis." Nat Cell Biol **8**(2): 188-194.

Flores, E. R., K. Y. Tsai, et al. (2002). "p63 and p73 are required for p53-dependent apoptosis in response to DNA damage." Nature **416**(6880): 560-564.

Freedman, D. A., L. Wu, et al. (1999). "Functions of the MDM2 oncoprotein." Cellular and Molecular Life Sciences **55**(1): 96-107.

Friedman, N., M. Linial, et al. (2000). "Using Bayesian networks to analyze expression data." J Comput Biol **7**(3-4): 601-620.

Fuda, N. J., M. B. Ardehali, et al. (2009). "Defining mechanisms that regulate RNA polymerase II transcription in vivo." Nature **461**(7261): 186-192.

Ghosh, S. and M. Karin (2002). "Missing pieces in the NF-kappaB puzzle." Cell **109**: S81-96.

Gidekel, S., G. Pizov, et al. (2003). "Oct-3/4 is a dose-dependent oncogenic fate determinant." Cancer Cell **4**(5): 361-370.

Gordon, S., G. Akopyan, et al. (2006). "Transcription factor YY1: structure, function therapeutic implications in cancer biology." Oncogene **25**(8): 1125-1142.

Gorgoulis, V. G., L. V. Vassiliou, et al. (2005). "Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions." Nature **434**(7035): 907-913.

Greber, B., H. Lehrach, et al. (2008). "Control of early fate decisions in human ES cells by distinct states of TGFbeta pathway activity." Stem Cells Dev **17**(6): 1065-1077.

Greber, B., H. Schöler (2008). "A breakthrough in stem cell research? Reprogramming somatic cells into pluripotent stem cells" Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz **51**(9): 1005-13.

Gross, A., J. Jockel, et al. (1998). "Enforced dimerization of BAX results in its translocation, mitochondrial dysfunction and apoptosis." EMBO J **17**(14): 3878-3885.

Guda, K., L. Natale, et al. (2009). "Infrequent detection of germline allele-specific expression of TGFBR1 in lymphoblasts and tissues of colon cancer patients." Cancer Res **69**(12): 4959-4961.

Guo, X. and X. F. Wang (2009). "Signaling cross-talk between TGF-beta/BMP and other pathways." Cell Research **19**(1): 71-88.

Haber, J. E. (2000). "Partners and pathwaysrepairing a double-strand break." Trends in Genetics **16**(6): 259-264.

Hanna, J., S. Markoulaki, et al. (2008). "Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency." Cell **133**(2): 250-264.

Hanna, J., M. Wernig, et al. (2007). "Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin." Science **318**(5858): 1920-1923.

Hanahan, D., R.A. Weinberg (2011). "Hallmarks of cancer: the next generation." Cell **144**(5):646-74.

Harper, J. W. and S. J. Elledge (2007). "The DNA damage response: ten years after." Mol Cell **28**(5): 739-745.

Hartlerode, A., R. Scully (2009). "Mechanisms of double-strand break repair in somatic mammalian cells." Biochem J **423**(2): 157-168.

Hartwell, L. H. and M. B. Kastan (1994). "Cell cycle control and cancer." Science **266**(5192): 1821-1828.

Hasty, J., D. McMillen, et al. (2001). "Computational studies of gene regulatory networks: in numero molecular biology." Nat Rev Genet **2**(4): 268-279.

Hatfield, S. D., H. R. Shcherbata, et al. (2005). "Stem cell division is regulated by the microRNA pathway." Nature **435**(7044): 974-978.

Ho, R., C. Chronis, K.Plath (2011). "Mechanistic insights into reprogramming to induced pluripotency." Journal of Cell Physiology **226**(4): 868-78.

Hochedlinger, K. and R. Jaenisch (2002). "Nuclear transplantation: lessons from frogs and mice." Current Opinion in Cell Biology **14**(6): 741-748.

Hochedlinger, K. and R. Jaenisch (2003). "Nuclear transplantation, embryonic stem cells, and the potential for cell therapy." N Engl J Med **349**(3): 275-286.

Hochedlinger, K. and K. Plath (2009). "Epigenetic reprogramming and induced pluripotency." Development **136**(4): 509-23.

Hochedlinger, K., Y. Yamada, et al. (2005). "Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues." Cell **121**(3): 465-477.

Hockemeyer, D., F. Soldner, et al. (2008). "A drug-inducible system for direct reprogramming of human somatic cells to pluripotency." Cell Stem Cell **3**(3): 346-353.

Hogan, B.L., M. Blessing ,et. al. (1994). " Growth factors in development: the role of TGF-beta related polypeptide signalling molecules in embryogenesis. Development. 53-60.

Hong, H., K. Takahashi, et al. (2009). "Suppression of induced pluripotent stem cell generation by the p53-p21 pathway." Nature **460**(7259): 1132-1135.

Hopfner, K. P., C. D. Putnam, et al. (2002). "DNA double-strand break repair from head to tail." Curr Opin Struct Biol **12**(1): 115-122.

Houbaviy, H. B., M. F. Murray, et al. (2003). "Embryonic stem cell-specific MicroRNAs." Developmental Cell **5**(2): 351-358.

Huang D.W., B.T. Sherman, R. A. Lempicki (2009). "Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources." Nature Protoc **4**(1):44-57.

Huangfu, D., K. Osafune, et al. (2008). "Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2." Nat Biotechnol **26**(11): 1269-1275.

Ingham, P.W., and A.P. McMahon (2001). "Hedgehog signaling in animal development: paradigms and principles." Genes Dev **15**(23): 3059-87

Ip, Y.T., R.J. Davis (1998). "Signal transduction by the c-Jun N-terminal kinase (JNK)--from imflammation to development." Curr Opin Cell Biol **10**(2): 205-19.

Ito, K., A. Hirao, et al. (2006). "Reactive oxygen species act through p38 MAPK to limit the lifespan of hematopoietic stem cells." Nat Med **12**(4): 446-451.

Jackson, M., A. Krassowska, et al. (2004). "Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells." Mol Cell Biol **24**(20): 8862-8871.

Jaenisch, R. and R. Young (2008). "Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming." Cell **132**(4): 567-582.

James, D., A. J. Levine, et al. (2005). "TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells." Development **132**(6): 1273-1282.

Jones, P. A. and D. Takai (2001). "The role of DNA methylation in mammalian epigenetics" Science **293**(5532): 1068-1070.

Jones, R. G. and C. B. Thompson (2009). "Tumor suppressors and cell metabolism: a recipe for cancer growth." Genes Dev **23**(5): 537-548.

Joshi-Tope G., L. Stein et al. (2005) "Reactome: a knowledgebase of biological pathways." Nucleic Acids Res **33**:D428-32.

Kagey, M. H., J. J. Newman, et al. (2010). "Mediator and cohesin connect gene expression and chromatin architecture." Nature **467**(7314): 430-435.

Kaji, K., K. Norrby, et al. (2009). "Virus-free induction of pluripotency and subsequent excision of reprogramming factors." Nature **458**(7239): 771-775.

Kanehisa, M. and S. Goto (2000). "KEGG: Kyoto Encyclopedia of Genes and Genomes." Nucleic Acids Res 28, 27-30.

Kawamura, T., J. Suzuki, et al. (2009). "Linking the p53 tumour suppressor pathway to somatic cell reprogramming." Nature **460**(7259): 1140-1144.

Kehler, J., E. Tolkunova, et al. (2004). "Oct4 is required for primordial germ cell survival." EMBO Rep **5**(11): 1078-1083.

Kielman, M. F., M. Rindapaa, et al. (2002). "Apc modulates embryonic stem-cell differentiation by controlling the dosage of beta-catenin signaling." Nat Genet **32**(4): 594-605.

Kim, D., C. H. Kim, et al. (2009). "Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins." Cell Stem Cell **4**(6): 472-476.

Kim, J., J. Chu, et al. (2008). "An extended transcriptional network for pluripotency of embryonic stem cells." Cell **132**(6): 1049-1061.

Kim, J. B., H. Zaehres, et al. (2008). "Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors." Nature **454**(7204): 646-650.

Kitamura, T., Y. Koshino, et al. (2003). "Retrovirus-mediated gene transfer and expression cloning: powerful tools in functional genomics." Exp Hematol **31**(11): 1007-1014.

Klipp, E., W. Liebermeister, C., et al. (2009) "Systems Biology - A Textbook." Wiley-VCH, Weinheim.

Knoepfler, P. S., X. Y. Zhang, et al. (2006). "Myc influences global chromatin structure." EMBO J **25**(12): 2723-2734.

Krantic, S., N. Mechawar et al. (2005). “Molecular basis of programmed cell death involved in neurodegeneration”. Trends Neurosci **28**(12): 670-6.

Kucharczak, J., M. J. Simmons, et al. (2003). "To be, or not to be: NF-kappaB is the answer--role of Rel/NF-kappaB in the regulation of apoptosis." Oncogene **22**(56): 8961-8982.

Kuhn, C., C. Wierling, et al. (2009). "Monte Carlo analysis of an ODE Model of the Sea Urchin Endomesoderm Network." BMC Syst Biol **3**: 83.

Kuhn, K., S. C. Baker, et al. (2004). "A novel, high-performance random array platform for quantitative gene expression profiling." Genome Research **14**(11): 2347-2356.

Lee, J. H. and T. T. Paull (2004). "Direct activation of the ATM protein kinase by the Mre11/Rad50/Nbs1 complex." Science **304**(5667): 93-96.

Lees-Miller, S. P. and K. Meek (2003). "Repair of DNA double strand breaks by non-homologous end joining." Biochimie **85**(11): 1161-1173.

Li, B., M. Carey, et al. (2007). "The role of chromatin during transcription." Cell **128**(4): 707-719.

Li, H., M. Collado, et al. (2009). "The Ink4/Arf locus is a barrier for iPS cell reprogramming." Nature **460**(7259): 1136-1139.

Li, R., J. Liang, et al. (2010). "A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts." Cell Stem Cell **7**(1): 51-63.

Li, Y., J. McClintick, et al. (2005). "Murine embryonic stem cell differentiation is promoted by SOCS-3 and inhibited by the zinc finger transcription factor Klf4." Blood **105**(2): 635-637.

Liang, S., S. Fuhrman, et al. (1998). "Reveal, a general reverse engineering algorithm for inference of genetic network architectures." Pac Symp Biocomput: 18-29.

Lieber, M. R., Y. Ma, et al. (2003). "Mechanism and regulation of human non-homologous DNA end-joining." Nat Rev Mol Cell Biol **4**(9): 712-720.

Lockhart, D. J. and E. A. Winzeler (2000). "Genomics, gene expression and DNA arrays." Nature **405**(6788): 827-836.

Looijenga, L. H., H. Stoop, et al. (2003). "POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors." Cancer Res **63**(9): 2244-2250.

Maherali, N. and K. Hochedlinger (2008). "Induced Pluripotency of Mouse and Human Somatic Cells." Control and Regulation of Stem Cells **73**: 157-162.

Maimets, T., I. Neganova, et al. (2008). "Activation of p53 by nutlin leads to rapid differentiation of human embryonic stem cells." Oncogene **27**(40): 5277-5287.

Majerus, P. W., J. Zou, et al. (2008). "The role of inositol signaling in the control of apoptosis." Adv Enzyme Regul **48**: 10-7.

Maloisel, L., F. Fabre, et al. (2008). "DNA polymerase delta is preferentially recruited during homologous recombination to promote heteroduplex DNA extension." Mol Cell Biol **28**(4): 1373-1382.

Mantel, C., Y. Guo, et al. (2007). "Checkpoint-apoptosis uncoupling in human and mouse embryonic stem cells: a source of karyotpic instability." Blood **109**(10): 4518-4527.

Marion, R. M., K. Strati, et al. (2009). "A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity." Nature **460**(7259): 1149-1153.

Markowitz, S. D. and M. M. Bertagnolli (2009). "Molecular origins of cancer: Molecular basis of colorectal cancer." N Engl J Med **361**(25): 2449-2460.

Marson, A., S. S. Levine, et al. (2008). "Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells." Cell **134**(3): 521-533.

Martinez, A.M., M. Afshar, et al. (1997). " Dual phosphorylation of the T-loop in cdk7: its role in controlling Cyclin H binding and CAK activity." EMBO **16**: 343-54.

Massague, J. and D. Wotton (2000). "Transcriptional control by the TGF-beta/Smad signaling system." EMBO J **19**(8): 1745-1754.

Massague, J. (2000). " How cells read TGF-beta signals." Nat Rev Mol Cell Biol **1**:169-78.

Masui, S., Y. Nakatake, et al. (2007). "Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells." Nat Cell Biol **9**(6): 625-635.

Matsuda, T., T. Nakamura, et al. (1999). "STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells." EMBO J **18**(15): 4261-4269.

Matsuoka, S., M. Yamaguchi, et al. (1994). "D-type Cyclin-binding regions of proliferating cell nuclear antigen." J Biol Chem **269**(15): 11030-11036.

Meissner, A. (2010). "Epigenetic modifications in pluripotent and differentiated cells." Nat Biotechnol **28**(10): 1079-1088.

Mekeel, K.L., W. Tang et al. (1997). "Inactivation of p53 results in high rates of homologous recombination." Oncogene **14**(15): 1847-57.

Mihara, M., S. Erster, et al. (1993). "p53 has a direct apoptogenic role in the mitochondria." Mol Cell **11**(3): 577-90.

Mikkelsen, T. S., J. Hanna, et al. (2008). "Dissecting direct reprogramming through integrative genomic analysis." Nature **454**(7200): 49-55.

Miller, R. A. and F. H. Ruddle (1976). "Pluripotent teratocarcinoma-thymus somatic cell hybrids." Cell **9**(1): 45-55.

Mitsui, K., Y. Tokuzawa, et al. (2003). "The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells." Cell **113**(5): 631-642.

Miyoshi, N., H. Ishii, et al. (2011). "Reprogramming of mouse and human cells to pluripotency using mature microRNAs." Cell Stem Cell **8**(6): 633-638.

Montgomery, E., M. Goggins, et al. (2001). "Nuclear localization of Dpc4 (Madh4, Smad4) in colorectal carcinomas and relation to mismatch repair/transforming growth factor-beta receptor defects." Am J Pathol **158**(2): 537-542.

Mu, Y., S. K. Gudey, et al. (2011). "Non-Smad signaling pathways." Cell Tissue Res.

Mu, Y., R. Sundar, et al. (2011). "TRAF6 ubiquitinates TGFbeta type I receptor to promote its cleavage and nuclear translocation in cancer." Nat Commun **2**: 330.

Murray, A. W. and T. Hunt (1993). The cell cycle : an introduction. New York, W.H. Freeman.

Nagata, S. (1999). "Fas ligand-induced apoptosis." Annu Rev Genet **33**: 29-55.

Nakagawa, M., M. Koyanagi, et al. (2008). "Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts." Nat Biotechnol **26**(1): 101-106.

Nakao, A., K. Okumura et al. (2002). "Smad7: a new key player in TGF-β-associated disease." Trends in Mol Med **8**(8): 361-363.

Ng, H. H. and M. A. Surani (2011). "The transcriptional and signalling networks of pluripotency." Nat Cell Biol **13**(5): 490-496.

Nichols, J., B. Zevnik, et al. (1998). "Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4." Cell **95**(3): 379-391.

Niwa, H., J. Miyazaki, et al. (2000). "Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells." Nat Genet **24**(4): 372-376.

Nutt, S. L., B. Heavey, et al. (1999). "Commitment to the B-lymphoid lineage depends on the transcription factor Pax5." Nature **401**(6753): 556-562.

Obaya, A. J., M. K. Mateyak, et al. (1999). "Mysterious liaisons: the relationship between c-Myc and the cell cycle." Oncogene **18**(19): 2934-2941.

Okita, K., M. Nakagawa, et al. (2008). "Generation of mouse induced pluripotent stem cells without viral vectors." Science **322**(5903): 949-953.

Omata, K., R. Suzuki, et al. (2008). "Identification and characterization of the human inhibitor of caspase-activated DNase gene promoter." Apoptosis 13(7): 929-37.

Orkin, S. H., J. Wang, et al. (2008). "The transcriptional network controlling pluripotency in ES cells." Cold Spring Harb Symp Quant Biol **73**: 195-202.

Orsucci., D., M. Filosto, et al., (2009). "Electron transfer mediators and other metabolites and cofactors in the treatment of mitochondrial dysfunction." Nutr Rev **67**(8): 427-38.

Ouyang, K. J., L. L. Woo, et al. (2008). "Homologous recombination and maintenance of genome integrity: cancer and aging through the prism of human RecQ helicases." Mech Ageing Dev **129**(7-8): 425-440.

Papapetrou, E. P., M. J. Tomishima, et al. (2009). "Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction and differentiation." Proc Natl Acad Sci U S A **106**(31): 12759-12764.

Park, I. H., N. Arora, et al. (2008). "Disease-specific induced pluripotent stem cells." Cell **134**(5): 877-886.

Patterson, G. I. and R. W. Padgett (2000). "TGF beta-related pathways. Roles in Caenorhabditis elegans development." Trends in Genetics **16**(1): 27-33.

Perfettini, J. L., M. Castedo, et al. (2005). "Essential role of p53 phosphorylation by p38 MAPK in apoptosis induction by the HIV-1 envelope." Journal of Experimental Medicine **201**(2): 279-289.

Pevny, L. H. and S. K. Nicolis (2010). "Sox2 roles in neural stem cells." Int J Biochem Cell Biol **42**(3): 421-424.

Pfeifer, A., T. Lim, et al. (2010). "Lentivirus Transgenesis." Methods in Enzymology, Vol 477: Guide to Techniques in Mouse Development, Part B: Mouse Molecular Genetics, Second Edition **477**: 3-15.

Pietenpol, J. A. and Z. A. Stewart (2002). "Cell cycle checkpoint signaling: cell cycle arrest versus apoptosis." Toxicology **181-182**: 475-481.

Prendergast, G. C. (1999). "Mechanisms of apoptosis by c-Myc." Oncogene **18**(19): 2967-2987.

Qi, J., J. Y. Yu, et al. (2009). "microRNAs regulate human embryonic stem cell division." Cell Cycle **8**(22): 3729-3741.

Rahl, P. B., C. Y. Lin, et al. (2010). "c-Myc regulates transcriptional pause release." Cell **141**(3): 432-445.

Rich, T., R. L. Allen, et al. (2000). "Defying death after DNA damage." Nature **407**(6805): 777-783.

Rosemary Siafakas, A. and D. R. Richardson (2009). "Growth arrest and DNA damage-45 alpha (GADD45alpha)." Int J Biochem Cell Biol **41**(5): 986-989.

Rothkamm, K., I. Kruger, et al. (2003). "Pathways of DNA double-strand break repair during the mammalian cell cycle." Mol Cell Biol **23**(16): 5706-5715.

Rowland, B. D. and R. Bernards (2006). "Re-evaluating cell-cycle regulation by E2Fs." Cell **127**(5): 871-874.

Rowland, B.D., R.A. Bernards, et al. (2005). "The KLF4 tumour suppressor is a transcriptional repressor of p53 that acts as a context-dependent oncogene." Nature cell biology, **7** (11): 1074-1082.

Rozen, S., and H. J. Skaletsky (2000). "Primer3 on the WWW for general users and for biologist programmers." Methods Mol Biol 132:365:86

Rubin, L. L. (2008). "Stem cells and drug discovery: the beginning of a new era?" Cell **132**(4): 549-552.

Ruiz S., A.D. Panopoulos, et al. (2010). "A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity." Current Biology **21**(1):45-52.

Sakamuro, D., V. Eviner, et al. (1995). "c-Myc induces apoptosis in epithelial cells by both p53-dependent and p53-independent mechanisms." Oncogene **11**(11): 2411-2418.

Samali, A., B. Zhivotovsky, et al. (1999). "Apoptosis: cell death defined by caspase activation." Cell Death Differ **6**(6): 495-496.

Samavarchi-Tehrani, P., A. Golipour, et al. (2010). "Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming." Cell Stem Cell **7**(1): 64-77.

Samuels-Lev, Y., D. J. O'Connor, et al. (2001). "ASPP proteins specifically stimulate the apoptotic function of p53." Mol Cell **8**(4): 781-794.

Saretzki, G., L. Armstrong, et al. (2004). "Stress defense in murine embryonic stem cells is superior to that of various differentiated murine cells." Stem Cells **22**(6): 962-971.

Savic, V., B. Yin, et al. (2009). "Formation of dynamic gamma-H2AX domains along broken DNA strands is distinctly regulated by ATM and MDC1 and dependent upon H2AX densities in chromatin." Mol Cell **34**(3): 298-310.

Sawa, A., T. Tomoda, B.I. Bae (2003). "Mechanisms of neuronal cell death in Huntington's disease." Cytogenet Genome Res **100**: 287-95.

Shikama, N., C. W. Lee, et al. (1999). "A novel cofactor for p300 that regulates the p53 response." Mol Cell **4**(3): 365-376.

Schmidt, E. V. (1999). "The role of c-myc in cellular growth control." Oncogene **18**(19): 2988-2996.

Schmitt, C. A., J. S. Fridman, et al. (2002). "Dissecting p53 tumor suppressor functions in vivo." Cancer Cell **1**(3): 289-298.

Schöler, H. R., A. K. Hatzopoulos, et al. (1989). "A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor." EMBO J **8**(9): 2543-2550.

Segre, J. A., C. Bauer, et al. (1999). "Klf4 is a transcription factor required for establishing the barrier function of the skin." Nat Genet **22**(4): 356-360.

Siegel, P.M. and J. Massague (2003). "Cytostatic and apoptotic actions of TGF-beta in homeostasis and cancer." Nat Rev Cancer **3**:807-21

Slee, E. A., D. J. O'Connor, et al. (2004). "To die or not to die: how does p53 decide?" Oncogene **23**, 2809-2818.

Soengas, M. S., P. Capodieci, et al. (2001). "Inactivation of the apoptosis effector Apaf-1 in malignant melanoma." Nature **409**(6817): 207-211.

Somogyi, R. and L. D. Greller (2001). "The dynamics of molecular networks: applications to therapeutic discovery." Drug Discov Today **6**(24): 1267-1277.

Stadtfeld, M., K. Brennand, et al. (2008). "Reprogramming of pancreatic beta cells into induced pluripotent stem cells." Curr Biol **18**(12): 890-894.

Stadtfeld, M., M. Nagaya, et al. (2008). "Induced pluripotent stem cells generated without viral integration." Science **322**(5903): 945-949.

Stewart, S. A., D. M. Dykxhoorn, et al. (2003). "Lentivirus-delivered stable gene silencing by RNAi in primary cells." RNA **9**(4): 493-501.

Stock, J. K., S. Giadrossi, et al. (2007). "Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells." Nat Cell Biol **9**(12): 1428-1435.

Subramanyam, D., C. D. Belair, et al. (2010). "PML-RAR{alpha} and Dnmt3a1 cooperate in vivo to promote acute promyelocytic leukemia." Cancer Res **70**(21): 8792-8801.

Suh, M. R., Y. Lee, et al. (2004). "Human embryonic stem cells express a unique set of microRNAs." Dev Biol **270**(2): 488-498.

Tada, M., T. Tada, et al. (1997). "Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells." EMBO J **16**(21): 6510-6520.

Tada, M., Y. Takahama, et al. (2001). "Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells." Curr Biol **11**(19): 1553-1558.

Takahashi, C., R. T. Bronson, et al. (2003). "Rb and N-ras function together to control differentiation in the mouse." Mol Cell Biol **23**(15): 5256-5268.

Takahashi, K., K. Tanabe, et al. (2007). "Induction of pluripotent stem cells from adult human fibroblasts by defined factors." Cell **131**(5): 861-872.

Takahashi, K. and S. Yamanaka (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." Cell **126**(4): 663-676.

Takuma, K., S.S. Yan, et al. (2005). "Mitochondrial dysfunction, endoplasmic reticulum stress, and apoptosis in Alzheimer's disease." J Pharmacol Sci **97**: 312-6.

Thomas, B., M. Beal (2007). "Parkinson's disease." Human Molecular Genetics **16**(2): 183-9.

Thomson, J. A., J. Itskovitz-Eldor, et al. (1998). "Embryonic stem cell lines derived from human blastocysts." Science **282**(5391): 1145-1147.

Thornberry, N. A. and Y. Lazebnik (1998). "Caspases: enemies within." Science **281**(5381): 1312-1316.

Toledo, F. and G. M. Wahl (2006). "Regulating the p53 pathway: in vitro hypotheses, in vivo veritas." Nat Rev Cancer **6**(12): 909-923.

Trujillo, K. M., S. S. Yuan, et al. (1998). "Nuclease activities in a complex of human recombination and DNA repair factors Rad50, Mre11, and p95." J Biol Chem **273**(34): 21447-21450.

Trumpp, A., Y. Refaeli, et al. (2001). "c-Myc regulates mammalian body size by controlling cell number but not cell size." Nature **414**(6865): 768-773.

Utikal, J., N. Maherali, et al. (2009). "Sox2 is dispensable for the reprogramming of melanocytes and melanoma cells into induced pluripotent stem cells." Journal of Cell Science **122**(Pt 19): 3502-3510.

Utikal, J., J. M. Polo, et al. (2009). "Immortalization eliminates a roadblock during cellular reprogramming into iPS cells." Nature **460**(7259): 1145-1148.

Vermeulen, K., D. R. Van Bockstaele, et al. (2003). "The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer." Cell Proliferation **36**(3): 131-149.

Vogelstein, B., D. Lane, et al. (2000). "Surfing the p53 network." Nature **408**(6810): 307-310.

Waddington, C. H. (1957). The strategy of the genes : a discussion of some aspects of theoretical biology. London, Allen & Unwin.

Wagner, E. F. and A. R. Nebreda (2009). "Signal integration by JNK and p38 MAPK pathways in cancer development." Nat Rev Cancer **9**(8): 537-549.

Wang, Y., R. Medvid, et al. (2007). "DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal." Nat Genet **39**(3): 380-385.

Wernig, M., A. Meissner, et al. (2008). "c-Myc is dispensable for direct reprogramming of mouse fibroblasts." Cell Stem Cell **2**(1): 10-12.

Wernig, M., J. P. Zhao, et al. (2008). "Neurons derived from reprogrammed fibroblasts functionally integrate into the fetal brain and improve symptoms of rats with Parkinson's disease." Proc Natl Acad Sci U S A **105**(15): 5856-5861.

Wierling, C., R. Herwig, et al. (2007). "Resources, standards and tools for systems biology." Brief Funct Genomic Proteomic **6**(3): 240-251.

Wierling, C., A. Kühn A, et al. (2012). "Prediction in the face of uncertainty: a Monte Carlo-based approach for systems biology of cancer treatment" Mutation Research (in press).

Woltjen, K., I. P. Michael, et al. (2009). "piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells." Nature **458**(7239): 766-770.

Wu, Q., X. Chen, et al. (2006). "Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells." J Biol Chem **281**(34): 24090-24094.

Xu, C., M. S. Inokuma, et al. (2001). "Feeder-free growth of undifferentiated human embryonic stem cells." Nat Biotechnol **19**(10): 971-974.

Xu, C., E. Rosler, et al. (2005). "Basic fibroblast growth factor supports undifferentiated human embryonic stem cell growth without conditioned medium." Stem Cells **23**(3): 315-323.

Xu, J., S. Lamouille, et al. (2009). "TGF-beta-induced epithelial to mesenchymal transition." Cell Research **19**(2): 156-172.

Yang, H., Q. Li, et al. (2005). "The BRCA2 homologue Brh2 nucleates RAD51 filament formation at a dsDNA-ssDNA junction." Nature **433**(7026): 653-657.

Yang, S. H., A. D. Sharrocks, et al. (2003). "Transcriptional regulation by the MAP kinase signaling cascades." Gene **320**: 3-21.

Yeap, L. S., K. Hayashi, et al. (2009). "ERG-associated protein with SET domain (ESET)-Oct4 interaction regulates pluripotency and represses the trophectoderm lineage." Epigenetics Chromatin **2**(1): 12.

Ying, Q. L., J. Nichols, et al. (2003). "BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3." Cell **115**(3): 281-292.

Young, R. A. (2011). "Control of the embryonic stem cell state." Cell **144**(6): 940-954. Review

Yu, J., K. Hu, et al. (2009). "Human induced pluripotent stem cells free of vector and transgene sequences." Science **324**(5928): 797-801.

Yu, J., M. A. Vodyanik, et al. (2007). "Induced pluripotent stem cell lines derived from human somatic cells." Science **318**(5858): 1917-1920.

Yuan, H., N. Corbi, et al. (1995). "Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3." Genes Dev **9**(21): 2635-2645.

Yuan, J., M. Lipinski, et al. (2003). "Diversity in the mechanisms of neuronal cell death." Neuron **40**(2): 401-413.

Yuan, T.L. and L.C. Cantley (2008). " PI3K pathway alterations in cancer: variations on a theme." Oncogene **27**: 5497-5510.

Zhang, J., W. L. Tam, et al. (2006). "Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1." Nat Cell Biol **8**(10): 1114-1123.

Zhao, W., I. M. Hisamuddin, et al. (2004). "Identification of Kruppel-like factor 4 as a potential tumor suppressor gene in colorectal cancer." Oncogene **23**(2): 395-402.

Zhao, Y., X. Yin, et al. (2008). "Two supporting factors greatly improve the efficiency of human iPSC generation." Cell Stem Cell **3**: 475-479.

Zhou, W. and C. R. Freed (2009). "Adenoviral gene delivery can reprogram human fibroblasts to induced pluripotent stem cells." Stem Cells **27**(11): 2667-2674.

Zhou, W., P. Zhu, et al. (2008). "Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation." Mol Cell **29**(1): 69-80.

Zhu, X., C. Rottkamp et al. (2000). " Activation of p38 kinase links tau phosphorylation, oxidative stress, and cell cycle-related events in Alzheimer Disease." J Neuropathol Exp Neurol **59**(10): 880-8.

10. Appendix

Table 20. Pathways enriched among differentially expressed genes following GFP over-expression. These table shows the raw data obtained from BeadStudio 3.0 (Illumina) from genes differentially expressed that belong to pathways highlighted in Table 14. The first column relates the gene name. In the remaining columns have been listed the detection P-value ("Det P-value") and the average signal intensity ("AVG_Signal") found in fibroblasts and GFP datasets. The last column relates the ratio of average signal intensity relative to non-transduced fibroblasts ("Fibroblasts") for each gene and sample. Samples refer to the average of three different biological replicates.

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	GFP Det P-value	GFP AVG_Signal	Ratio GFP/FIB
DNA replication					
FEN1	0,00E+00	539,13	0,00E+00	1287,01	2,39
LIG1	4,51E-03	128,63	0,00E+00	257,15	2,00
MCM2	3,01E-03	143,52	0,00E+00	354,41	2,47
MCM4	1,02E-05	253,90	0,00E+00	597,20	2,35
MCM5	0,00E+00	218,25	0,00E+00	460,88	2,11
MCM7	3,11E-05	705,47	2,32E-06	1134,49	1,61
POLA2	0,00E+00	267,79	0,00E+00	627,32	2,34
POLE2	4,51E-03	93,19	0,00E+00	265,74	2,85
PRIM1	4,51E-03	104,88	0,00E+00	239,52	2,28
RFC2	1,63E-03	37,70	5,98E-04	78,50	2,08
RFC3	1,84E-02	28,84	8,85E-04	98,43	3,41
RFC4	0,00E+00	470,52	0,00E+00	744,83	1,58
RNASEH2A	4,51E-03	102,82	0,00E+00	283,29	2,76
Cell cycle					
ANAPC2	3,80E-03	39,54	2,61E-03	60,53	1,53
BUB1	9,02E-03	89,35	0,00E+00	310,83	3,48
CCNA2	0,00E+00	302,60	0,00E+00	1082,55	3,58
CCNB1	3,76E-02	28,05	4,51E-03	132,02	4,71
CCNB2	0,00E+00	435,70	0,00E+00	1120,35	2,57
CCND2	0,00E+00	3944,83	0,00E+00	7549,73	1,91
CCND3	0,00E+00	5316,95	0,00E+00	9289,73	1,75
CCNH	4,51E-03	98,37	0,00E+00	264,49	2,69
CDC20	0,00E+00	617,43	0,00E+00	2011,17	3,26
CDC25A	3,42E-02	18,64	1,99E-03	94,91	5,09
CDC25C	5,87E-03	23,19	7,83E-06	95,68	4,13
CDC7	3,01E-02	36,99	9,02E-03	97,79	2,64
E2F2	6,62E-02	20,10	4,51E-03	128,48	6,39
GADD45A	0,00E+00	1818,01	0,00E+00	2764,76	1,52
MCM2	3,01E-03	143,52	0,00E+00	354,41	2,47
MCM4	1,02E-05	253,90	0,00E+00	597,20	2,35
MCM5	0,00E+00	218,25	0,00E+00	460,88	2,11
MCM7	3,11E-05	705,47	2,32E-06	1134,49	1,61
MYC	1,50E-03	184,28	0,00E+00	332,77	1,81

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	GFP Det P-value	GFP AVG_Signal	Ratio GFP/FIB
PLK1	4,51E-03	97,82	1,50E-03	216,16	2,21
PTTG1	0,00E+00	1168,96	0,00E+00	2552,42	2,18
SMAD2	1,18E-02	70,38	3,67E-03	84,82	1,21
TTK	1,05E-02	84,95	0,00E+00	288,44	3,40
p53 signalling pathway					
BID	1,66E-02	422,41	2,15E-03	434,81	1,03
CCNB1	3,76E-02	28,05	4,51E-03	132,02	4,71
CCNB2	0,00E+00	435,70	0,00E+00	1120,35	2,57
CCND2	0,00E+00	3944,83	0,00E+00	7549,73	1,91
CCND3	0,00E+00	5316,95	0,00E+00	9289,73	1,75
FAS	1,31E-03	47,67	2,54E-04	104,04	2,18
GADD45A	0,00E+00	1818,01	0,00E+00	2764,76	1,52
GTSE1	1,95E-02	49,36	4,51E-03	161,49	3,27
IGFBP3	1,63E-04	100,94	1,02E-05	223,58	2,21
PERP	3,01E-03	142,82	1,50E-03	220,68	1,55
PMAIP1	0,00E+00	251,91	0,00E+00	443,19	1,76
RCHY1	1,77E-02	56,82	1,55E-03	81,71	1,44
RRM2	1,35E-02	75,42	0,00E+00	319,98	4,24
RRM2B	1,50E-02	62,31	9,02E-03	104,87	1,68
SERPINE1	0,00E+00	3314,43	0,00E+00	5918,00	1,79
SESN2	1,65E-02	56,22	9,02E-03	113,28	2,01
Phosphatidylinositol signalling system					
DGKB	4,08E-02	7,80	2,88E-03	35,82	4,59
DGKH	1,42E-02	17,00	5,70E-03	41,21	2,42
INPP4B	1,50E-02	61,89	4,51E-03	160,55	2,59
IPPK	2,11E-02	44,91	9,02E-03	101,67	2,26
ITPKA	2,41E-02	42,15	7,52E-03	119,02	2,82
ITPR1	1,35E-02	72,40	4,51E-03	134,92	1,86
PIK3C3	1,35E-02	74,13	9,02E-03	116,72	1,57
PIP5K1A	1,35E-02	70,38	9,02E-03	102,55	1,46
PLCB1	1,77E-02	68,30	2,54E-04	148,91	2,18
PLCB4	2,61E-03	33,60	1,55E-03	63,93	1,90
Mismatch repair					
LIG1	4,51E-03	128,63	0,00E+00	257,15	2,00
PMS2	2,22E-04	74,99	1,91E-04	124,56	1,66
RFC2	1,63E-03	37,70	5,98E-04	78,50	2,08
RFC3	1,84E-02	28,84	8,85E-04	98,43	3,41
RFC4	0,00E+00	470,52	0,00E+00	744,83	1,58
Apoptosis					
AKT3	1,31E-02	22,73	7,36E-03	43,30	1,91
BID	1,66E-02	422,41	2,15E-03	434,81	1,03
FAS	1,31E-03	47,67	2,54E-04	104,04	2,18
IL1A	3,61E-02	31,70	9,02E-03	104,71	3,30
IL1B	0,00E+00	311,05	0,00E+00	1078,05	3,47

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	GFP Det P-value	GFP AVG_Signal	Ratio GFP/FIB
IRAK2	0,00E+00	601,64	0,00E+00	954,86	1,59
IRAK3	1,05E-02	85,34	4,51E-03	134,70	1,58
PRKAR2A	1,50E-03	177,59	0,00E+00	312,27	1,76
TNFRSF10A	1,20E-02	77,10	4,51E-03	151,68	1,97
TNFRSF10D	0,00E+00	262,09	0,00E+00	474,64	1,81
PPAR signalling pathway					
ACOX2	0,00E+00	475,54	0,00E+00	298,68	0,63
CHKB	1,16E-03	922,90	5,38E-03	570,95	0,62
CYP27A1	0,00E+00	1400,55	0,00E+00	716,92	0,51
EHHADH	1,50E-03	190,08	4,51E-03	126,16	0,66
PCK2	7,14E-04	245,71	1,09E-02	314,68	1,28
PLTP	0,00E+00	2712,54	0,00E+00	1503,40	0,55
RXRA	0,00E+00	2412,73	0,00E+00	1589,86	0,66
SCD	0,00E+00	14280,18	0,00E+00	8483,19	0,59
SLC27A1	0,00E+00	1242,98	0,00E+00	795,60	0,64
Hedgehog signalling pathway					
BMP4	6,27E-05	1729,04	3,14E-03	480,30	0,28
BMP5	1,50E-03	173,97	3,16E-02	52,35	0,30
GLI3	3,01E-03	148,01	1,20E-02	90,03	0,61
PRKACA	6,88E-03	27,17	2,61E-02	44,25	1,63
PRKACB	6,58E-03	45,69	1,94E-02	53,60	1,17
SMO	0,00E+00	333,35	1,50E-03	202,55	0,61
WNT2	0,00E+00	515,94	9,02E-03	109,26	0,21

Table 21. Pathways enriched among differentially expressed genes following OCT4 over-expression. These table shows the raw data obtained from BeadStudio 3.0 (Illumina) from genes differentially expressed that belong to pathways highlighted in Table 14. The first column relates the gene name. In the remaining columns have been listed the detection P-value ("Det P-value") and the average signal intensity ("AVG_Signal") found in fibroblasts and OCT4 datasets. The last column relates the ratio of average signal intensity relative to non-transduced fibroblasts ("Fibroblasts") for each gene and sample. Samples refer to the average of three different biological replicates.

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	OCT4 Det P-value	OCT4 AVG_Signal	Ratio OCT4/FIB
Regulation of actin cytoskeleton					
ACTN1	0,00E+00	9534,15	0,00E+00	5883,10	0,62
BDKRB1	0,00E+00	3267,96	0,00E+00	2068,52	0,63
BDKRB2	0,00E+00	742,42	0,00E+00	481,45	0,65
FGD1	0,00E+00	724,51	0,00E+00	454,87	0,63
FGFR3	5,23E-03	31,29	1,39E-02	29,98	0,96
FN1	5,33E-04	264,45	2,69E-04	169,77	0,64
ITGA1	0,00E+00	720,05	0,00E+00	416,38	0,58
ITGA10	0,00E+00	255,11	2,26E-02	57,52	0,23
ITGA11	0,00E+00	1631,33	0,00E+00	912,56	0,56
ITGAV	0,00E+00	5066,86	0,00E+00	3332,79	0,66
MYLK	0,00E+00	2597,43	1,13E-06	1417,52	0,55

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	OCT4 Det P-value	OCT4 AVG_Signal	Ratio OCT4/FIB
PDGFC	0,00E+00	891,12	0,00E+00	585,29	0,66
PIK3CD	1,50E-03	170,58	7,52E-03	104,70	0,61
SSH2	0,00E+00	754,94	0,00E+00	460,88	0,61
VAV2	4,51E-03	100,75	1,35E-02	85,64917	0,85
Calcium signalling pathway					
ADCY2	4,51E-03	92,85	1,35E-02	84,99	0,92
ADCY7	4,51E-03	95,73	1,20E-02	88,45	0,92
ADCY9	0,00E+00	1643,05	0,00E+00	932,57	0,57
ATP2A2	0,00E+00	5091,92	0,00E+00	3097,83	0,61
BDKRB1	0,00E+00	3267,96	0,00E+00	2068,52	0,63
BDKRB2	0,00E+00	742,42	0,00E+00	481,45	0,65
CACNA1C	1,50E-03	183,77	1,65E-02	73,67	0,40
MYLK	0,00E+00	2597,43	1,13E-06	1417,52	0,55
OXTR	0,00E+00	216,96	1,65E-02	78,17	0,36
P2RX4	0,00E+00	794,48	0,00E+00	514,39	0,65
TRPC1	0,00E+00	618,95	0,00E+00	397,52	0,64
Endocytosis					
ARRB1	1,02E-05	151,78	3,66E-04	85,79	0,57
CBLB	0,00E+00	705,77	0,00E+00	450,03	0,64
CLTB	6,70E-03	1595,00	2,28E-02	1631,02	1,02
DNM1	0,00E+00	223,26	1,35E-02	84,96	0,38
EHD4	0,00E+00	2144,88	0,00E+00	1421,89	0,66
FGFR3	5,23E-03	31,29	1,39E-02	29,98	0,96
GRK5	0,00E+00	2032,10	0,00E+00	1346,14	0,66
PLD1	0,00E+00	299,39	1,50E-03	187,91	0,63
RAB11FIP1	5,17E-03	559,05	3,31E-03	287,31	0,51
SMAP1	1,02E-05	633,41	9,16E-05	398,35	0,63

Table 22. Pathways enriched among differentially expressed genes following SOX2 over-expression. These table shows the raw data obtained from BeadStudio 3.0 (Illumina) from genes differentially expressed that belong to pathways highlighted in Table 14. The first column relates the gene name. In the remaining columns have been listed the detection P-value ("Det P-value") and the average signal intensity ("AVG_Signal") found in fibroblasts and SOX2 datasets. The last column relates the ratio of average signal intensity relative to non-transduced fibroblasts ("Fibroblasts") for each gene and sample. Samples refer to the average of three different biological replicates.

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	SOX2 Det P-value	SOX2 AVG_Signal	Ratio SOX2/FIB
T cell receptor signalling pathway					
AKT2	2,41E-02	41,29	4,51E-03	123,56	2,99
CBL	1,50E-03	193,41	0,00E+00	307,63	1,59
CBLB	0,00E+00	705,77	0,00E+00	1121,12	1,59
CD247	2,55E-01	2,58	1,81E-03	39,18	15,18
MAP2K1	0,00E+00	1428,72	0,00E+00	2282,69	1,60
MAPK11	7,64E-04	99,93	1,63E-04	179,47	1,80
MAPK9	0,00E+00	807,46	0,00E+00	1245,52	1,54
NFAT5	4,57E-02	63,83	3,09E-03	97,43	1,53

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	SOX2 Det P-value	SOX2 AVG_Signal	Ratio SOX2/FIB
NFATC3	1,12E-02	46,31	3,20E-03	72,12	1,56
PAK7	4,21E-02	10,41	9,78E-03	26,11	2,51
PIK3CA	1,65E-02	54,80	6,02E-03	96,27	1,76
PPP3CA	1,20E-02	77,32	6,02E-03	106,85	1,38
RELA	0,00E+00	301,99	0,00E+00	499,28	1,65
SOS1	4,51E-03	105,83	1,50E-03	183,32	1,73
ErbB signalling pathway					
AKT2	2,41E-02	41,29	4,51E-03	123,56	2,99
CAMK2B	2,33E-02	8,92	5,27E-03	21,46	2,41
CAMK2D	1,93E-02	129,79	4,20E-03	168,29	1,30
CBL	1,50E-03	193,41	0,00E+00	307,63	1,59
CBLB	0,00E+00	705,77	0,00E+00	1121,12	1,59
EREG	2,41E-02	41,72	9,02E-03	85,36	2,05
MAP2K1	0,00E+00	1428,72	0,00E+00	2282,69	1,60
MAPK9	0,00E+00	807,46	0,00E+00	1245,52	1,54
PAK7	4,21E-02	10,41	9,78E-03	26,11	2,51
PIK3CA	1,65E-02	54,80	6,02E-03	96,27	1,76
SOS1	4,51E-03	105,83	1,50E-03	183,32	1,73
Calcium signalling pathway					
ADCY2	4,51E-03	92,85	1,35E-02	79,08	0,85
AGTR1	0,00E+00	241,99	4,52E-06	154,32	0,64
ATP2A2	0,00E+00	5091,92	0,00E+00	3030,37	0,60
BDKRB1	0,00E+00	3267,96	0,00E+00	1266,80	0,39
BDKRB2	0,00E+00	742,42	0,00E+00	262,71	0,35
CACNA1C	1,50E-03	183,77	1,20E-02	81,77	0,44
EDNRA	0,00E+00	517,04	0,00E+00	315,41	0,61
ITPR3	0,00E+00	1372,47	0,00E+00	775,35	0,56
P2RX4	0,00E+00	794,48	0,00E+00	465,35	0,59
PLCD4	3,01E-03	145,27	1,80E-02	62,45	0,43
PTGER1	0,00E+00	456,62	0,00E+00	302,08	0,66

Table 23. Pathways enriched among differentially expressed genes following KLF4 over-expression. These table shows the raw data obtained from BeadStudio 3.0 (Illumina) from genes differentially expressed that belong to pathways highlighted in Table 14. The first column relates the gene name. In the remaining columns have been listed the detection P-value ("Det P-value") and the average signal intensity ("AVG_Signal") found in fibroblasts and KLF4 datasets. The last column relates the ratio of average signal intensity relative to non-transduced fibroblasts ("Fibroblasts") for each gene and sample. Samples refer to the average of three different biological replicates.

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	KLF4 Det P-value	KLF4 AVG_Signal	Ratio KLF4/FIB
TGF-beta signalling pathway					
ACVRL1	1,50E-03	177,80	1,65E-02	60,93	0,34
BMP2	0,00E+00	2426,98	0,00E+00	1465,33	0,60
BMP6	1,14E-01	13,19	1,50E-03	199,24	15,11
ID1	1,63E-03	345,05	1,13E-06	2087,09	6,05
ID2	0,00E+00	503,18	0,00E+00	1060,26	2,11

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	KLF4 Det P-value	KLF4 AVG_Signal	Ratio KLF4/FIB
ID3	0,00E+00	2408,36	0,00E+00	4721,28	1,96
LTBP1	1,72E-03	161,00	1,06E-02	91,65	0,57
RBL1	2,48E-02	17,03	5,23E-03	39,42	2,31
RBX1	0,00E+00	2416,95	0,00E+00	3791,31	1,57
RPS6KB2	0,00E+00	2171,63	0,00E+00	3519,97	1,62
SMAD4	0,00E+00	2185,40	0,00E+00	1358,67	0,62
SMURF1	7,60E-03	28,14	1,39E-02	40,66	1,44
TGFB2	2,41E-02	43,64	3,01E-03	134,92	3,09
THBS1	0,00E+00	23037,12	0,00E+00	14213,93	0,62
ZFYVE16	4,51E-03	92,41	1,20E-02	83,68	0,91

Table 24. Pathways enriched among differentially expressed genes following c-MYC over-expression. These table shows the raw data obtained from BeadStudio 3.0 (Illumina) from genes differentially expressed that belong to pathways highlighted in Table 14. The first column relates the gene name. In the remaining columns have been listed the detection P-value ("Det P-value") and the average signal intensity ("AVG_Signal") found in fibroblasts and c-MYC datasets. The last column relates the ratio of average signal intensity relative to non-transduced fibroblasts ("Fibroblasts") for each gene and sample. Samples refer to the average of three different biological replicates.

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	c-MYC Det P-value	c-MYC AVG_Signal	Ratio c-MYC/FIB
Nucleotide excision repair					
CDK7	0,00E+00	1145,15	0,00E+00	1858,81	1,62
CUL4A	7,24E-03	266,49	2,83E-03	481,00	1,80
DDB1	4,52E-06	7116,19	1,13E-06	11030,50	1,55
ERCC2	0,00E+00	292,16	0,00E+00	521,18	1,78
GTF2H3	0,00E+00	477,40	0,00E+00	1343,23	2,81
GTF2H4	0,00E+00	777,20	0,00E+00	1225,41	1,58
POLD1	3,01E-03	150,63	0,00E+00	338,32	2,25
POLE	4,51E-03	91,68	1,50E-03	147,59	1,61
POLE3	0,00E+00	2568,61	0,00E+00	9310,45	3,62
POLE4	0,00E+00	4415,10	0,00E+00	7149,94	1,62
RAD23B	0,00E+00	1263,20	0,00E+00	2015,82	1,60
RFC1	0,00E+00	519,01	0,00E+00	792,69	1,53
RPA2	0,00E+00	1143,56	0,00E+00	2472,44	2,16
RPA3	0,00E+00	1302,98	0,00E+00	3848,94	2,95
Base excision repair					
APEX1	0,00E+00	5357,74	0,00E+00	11657,11	2,18
MPG	1,08E-03	38,82	1,38E-05	89,13	2,30
MUTYH	2,30E-04	146,05	6,03E-06	337,23	2,31
NEIL2	0,00E+00	233,54	0,00E+00	379,84	1,63
NTHL1	0,00E+00	633,02	0,00E+00	1449,89	2,29
PARP1	0,00E+00	3101,70	0,00E+00	6197,41	2,00
PARP2	1,02E-05	178,06	0,00E+00	355,91	2,00
POLB	0,00E+00	1315,15	0,00E+00	1995,17	1,52
POLD1	3,01E-03	150,63	0,00E+00	338,32	2,25
POLE	4,51E-03	91,68	1,50E-03	147,59	1,61

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	c-MYC Det P-value	c-MYC AVG_Signal	Ratio c-MYC/FIB
POLE3	0,00E+00	2568,61	0,00E+00	9310,45	3,62
POLE4	0,00E+00	4415,10	0,00E+00	7149,94	1,62
DNA replication					
MCM3	0,00E+00	1764,81	0,00E+00	4672,12	2,65
MCM6	0,00E+00	1580,33	0,00E+00	4684,09	2,96
POLA1	9,02E-03	89,20	0,00E+00	248,20	2,78
POLD1	3,01E-03	150,63	0,00E+00	338,32	2,25
POLE	4,51E-03	91,68	1,50E-03	147,59	1,61
POLE3	0,00E+00	2568,61	0,00E+00	9310,45	3,62
POLE4	0,00E+00	4415,10	0,00E+00	7149,94	1,62
RFC1	0,00E+00	519,01	0,00E+00	792,69	1,53
RNASEH1	0,00E+00	1756,14	0,00E+00	3031,21	1,73
RPA2	0,00E+00	1143,56	0,00E+00	2472,44	2,16
RPA3	0,00E+00	1302,98	0,00E+00	3848,94	2,95
SSBP1	0,00E+00	4178,10	0,00E+00	8932,55	2,14
Glycolysis / Gluconeogenesis					
ALDH1A3	0,00E+00	544,22	0,00E+00	1851,82	3,40
ALDH1B1	3,01E-03	155,84	0,00E+00	461,61	2,96
ALDOA	6,08E-04	5164,34	5,76E-04	8165,80	1,58
DLAT	0,00E+00	583,81	0,00E+00	1322,66	2,27
DLD	0,00E+00	521,97	0,00E+00	842,03	1,61
ENO1	0,00E+00	18661,41	0,00E+00	46831,72	2,51
GPI	0,00E+00	1743,43	0,00E+00	4043,50	2,32
LDHA	0,00E+00	11707,23	0,00E+00	30691,38	2,62
LDHB	0,00E+00	2280,24	0,00E+00	5201,63	2,28
PDHA1	0,00E+00	830,65	0,00E+00	1760,02	2,12
PDHB	0,00E+00	8243,50	0,00E+00	14413,74	1,75
PFKM	0,00E+00	1038,82	0,00E+00	3200,63	3,08
PGAM1	0,00E+00	3410,17	0,00E+00	6604,54	1,94
PGK1	0,00E+00	930,57	0,00E+00	1646,34	1,77
PGM2	0,00E+00	288,65	0,00E+00	826,16	2,86
TPI1	0,00E+00	14020,27	0,00E+00	26516,65	1,89
Endocytosis					
AP2A1	4,71E-04	188,45	9,62E-03	113,85	0,60
AP2A2	0,00E+00	313,57	1,50E-03	185,76	0,59
CLTB	6,70E-03	1595,00	1,63E-02	2183,77	1,37
EHD2	0,00E+00	741,50	1,50E-03	156,01	0,21
EHD3	0,00E+00	226,96	1,65E-02	72,38	0,32
FGFR3	5,23E-03	31,29	4,17E-02	38,47	1,23
GRK4	2,59E-04	69,99	2,64E-02	99,23	1,42
GRK5	0,00E+00	2032,10	0,00E+00	1188,99	0,59
HLA-F	0,00E+00	461,50	0,00E+00	260,21	0,56
HSPA1A	0,00E+00	848,69	0,00E+00	292,82	0,35
HSPA1L	1,50E-03	164,37	9,02E-03	96,92	0,59

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	c-MYC Det P-value	c-MYC AVG_Signal	Ratio c-MYC/FIB
HSPA2	0,00E+00	824,57	1,65E-02	65,47	0,08
IQSEC2	4,51E-03	125,25	1,05E-02	95,23	0,76
KIT	4,51E-03	93,22	1,50E-02	78,26	0,84
LDLRAP1	3,01E-03	152,59	1,65E-02	64,77	0,42
PARD6G	4,52E-06	390,97	4,07E-05	172,17	0,44
PIP5K1C	0,00E+00	1638,82	0,00E+00	694,84	0,42
PLD1	0,00E+00	299,39	1,50E-03	175,49	0,59
PLD2	0,00E+00	281,10	1,50E-03	153,69	0,55
RAB11B	0,00E+00	247,38	6,02E-03	116,81	0,47
RAB22A	0,00E+00	2054,95	0,00E+00	1150,55	0,56
RAB5B	0,00E+00	6537,17	0,00E+00	4024,48	0,62
SMURF1	7,60E-03	28,14	3,70E-02	46,96	1,67
TGFBR2	6,51E-04	2306,33	1,72E-03	1184,80	0,51
MAPK signalling pathway					
CACNA1C	1,50E-03	183,77	1,50E-02	80,70	0,44
CDC25B	0,00E+00	3938,76	0,00E+00	2173,79	0,55
FGF7	4,51E-03	100,28	1,35E-02	98,02	0,98
FGFR3	5,23E-03	31,29	4,17E-02	38,47	1,23
FLNA	0,00E+00	223,64	1,65E-02	78,21	0,35
HSPA1A	0,00E+00	848,69	0,00E+00	292,82	0,35
HSPA1L	1,50E-03	164,37	9,02E-03	96,92	0,59
HSPA2	0,00E+00	824,57	1,65E-02	65,47	0,08
IL1R1	1,50E-03	189,72	7,52E-02	50,01	0,26
MAP2K4	0,00E+00	954,35	0,00E+00	498,06	0,52
MAP3K11	0,00E+00	517,86	0,00E+00	284,81	0,55
MAP3K12	1,50E-03	188,57	6,02E-03	120,90	0,64
MAP3K6	0,00E+00	780,46	0,00E+00	405,53	0,52
MAPK7	9,95E-05	263,09	3,73E-04	106,29	0,40
MAPK8IP3	1,09E-03	904,83	4,49E-03	605,54	0,67
MAPKAPK2	2,22E-04	240,67	2,54E-04	140,83	0,59
NR4A1	9,07E-03	24,97	1,49E-01	32,89	1,32
NTF3	0,00E+00	645,10	0,00E+00	263,11	0,41
PTPN5	5,70E-03	22,80	3,19E-02	40,24	1,76
RPS6KA4	0,00E+00	463,65	0,00E+00	239,67	0,52
STMN1	5,59E-03	388,34	6,75E-02	482,39	1,24
TGFBR2	6,51E-04	2306,33	1,72E-03	1184,80	0,51
TNFRSF1A	0,00E+00	5957,80	0,00E+00	3698,28	0,62

Table 25. Pathways enriched among differentially expressed genes following silmutaneous over-expression of OCT4, SOX2 and KLF4. These table shows the raw data obtained from BeadStudio 3.0 (Illumina) from genes differentially expressed that belong to pathways highlighted in Table 14. The first column relates the gene name. In the remaining columns have been listed the detection P-value ("Det P-value") and the average signal intensity ("AVG_Signal") found in fibroblasts and 3TF (abbreviation for OCT4, SOX2 and KLF4) datasets. The last column relates the ratio of average signal intensity relative to non-transduced fibroblasts ("Fibroblasts") for each gene and sample. Samples refer to the average of three different biological replicates.

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	3TF Det P-value	3TF AVG_Signal	Ratio 3TF/FIB
Homologous recombination					
BLM	7,22E-02	18,24	1,50E-03	133,63	7,33
EME1	3,76E-02	29,85	6,02E-03	93,46	3,13
MRE11A	3,91E-05	59,84	2,32E-06	103,92	1,74
NBN	1,32E-02	24,74	4,49E-03	56,19	2,27
RAD51C	4,96E-02	331,82	9,18E-03	540,36	1,63
RPA2	0,00E+00	1143,56	0,00E+00	1719,73	1,50
RPA3	0,00E+00	1302,98	0,00E+00	2937,59	2,25
TOP3A	1,50E-03	212,80	0,00E+00	347,04	1,63
DNA replication					
MCM3	0,00E+00	1764,81	0,00E+00	3541,36	2,01
MCM6	0,00E+00	1580,33	0,00E+00	2751,42	1,74
POLA1	9,02E-03	89,20	1,50E-03	149,76	1,68
POLE	4,51E-03	91,68	1,50E-03	180,66	1,97
POLE3	0,00E+00	2568,61	0,00E+00	5035,30	1,96
RPA2	0,00E+00	1143,56	0,00E+00	1719,73	1,50
RPA3	0,00E+00	1302,98	0,00E+00	2937,59	2,25
Wnt signalling pathway					
CHP	0,00E+00	795,08	0,00E+00	474,73	0,60
CTBP1	0,00E+00	728,62	0,00E+00	431,45	0,59
DAAM2	3,01E-03	139,27	6,02E-03	89,15	0,64
DVL2	0,00E+00	912,19	0,00E+00	494,08	0,54
DVL3	0,00E+00	1105,43	0,00E+00	643,20	0,58
EP300	0,00E+00	503,02	0,00E+00	328,15	0,65
FRAT2	0,00E+00	478,49	0,00E+00	305,58	0,64
FZD2	0,00E+00	3249,84	0,00E+00	1802,60	0,55
LEF1	0,00E+00	306,13	3,01E-03	106,20	0,35
NKD2	0,00E+00	575,29	1,38E-01	43,18	0,08
PPP2R5A	0,00E+00	1099,45	0,00E+00	615,10	0,56
PRICKLE2	0,00E+00	1281,25	0,00E+00	741,02	0,58
PRKCA	0,00E+00	1549,11	0,00E+00	811,29	0,52
ROCK1	0,00E+00	242,58	1,50E-03	131,69	0,54
SMAD3	0,00E+00	4188,80	0,00E+00	2248,44	0,54
SMAD4	0,00E+00	2185,40	0,00E+00	1338,90	0,61
TBL1XR1	0,00E+00	258,50	1,50E-03	130,30	0,50
WNT5A	0,00E+00	3827,64	0,00E+00	1920,79	0,50
MAPK signalling pathway					
CACNA1C	1,50E-03	183,77	1,65E-02	69,59	0,38

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	3TF Det P-value	3TF AVG_Signal	Ratio 3TF/FIB
CHP	0,00E+00	795,08	0,00E+00	474,73	0,60
ECSIT	4,51E-03	94,97	1,65E-02	73,37	0,77
ELK1	0,00E+00	1720,02	0,00E+00	1126,57	0,65
FGF7	4,51E-03	100,28	4,51E-02	54,26	0,54
FLNA	0,00E+00	223,64	6,02E-03	92,55	0,41
HSPA2	0,00E+00	824,57	1,20E-02	75,51	0,09
IL1R1	1,50E-03	189,72	4,36E-02	54,30	0,29
MAP2K4	0,00E+00	954,35	0,00E+00	585,82	0,61
MAP2K5	1,02E-05	286,51	5,54E-05	164,38	0,57
MAP3K12	1,50E-03	188,57	3,01E-03	111,41	0,59
MAP3K14	4,51E-03	109,68	1,05E-02	78,43	0,72
MAP3K6	0,00E+00	780,46	0,00E+00	471,64	0,60
MAP4K4	1,72E-03	1149,23	3,55E-03	720,05	0,63
MAPK8IP3	1,09E-03	904,83	1,16E-03	493,53	0,55
MAPKAPK2	2,22E-04	240,67	4,52E-04	138,51	0,58
NFKB2	9,79E-04	111,71	4,81E-03	68,90	0,62
NR4A1	9,07E-03	24,97	1,27E-01	37,93	1,52
NTF3	0,00E+00	645,10	0,00E+00	410,43	0,64
PAK2	4,52E-06	510,95	1,13E-06	327,68	0,64
PLA2G6	2,94E-03	34,95	4,94E-02	48,69	1,39
PRKCA	0,00E+00	1549,11	0,00E+00	811,29	0,52
PTPN5	5,70E-03	22,80	3,75E-02	43,99	1,93
RAF1	0,00E+00	1816,87	0,00E+00	1173,20	0,65
RELB	0,00E+00	242,52	1,50E-03	131,27	0,54
STMN1	5,59E-03	388,34	2,63E-02	736,05	1,90
TNFRSF1A	0,00E+00	5957,80	0,00E+00	3755,80	0,63
Calcium signalling pathway					
ADCY9	0,00E+00	1643,05	0,00E+00	464,89	0,28
AGTR1	0,00E+00	241,99	1,02E-05	129,08	0,53
ATP2A2	0,00E+00	5091,92	0,00E+00	2429,73	0,48
BDKRB1	0,00E+00	3267,96	0,00E+00	1162,36	0,36
BDKRB2	0,00E+00	742,42	3,01E-03	116,43	0,16
CACNA1C	1,50E-03	183,77	1,65E-02	69,59	0,38
CHP	0,00E+00	795,08	0,00E+00	474,73	0,60
EDNRA	0,00E+00	517,04	0,00E+00	301,71	0,58
HRH1	4,51E-03	104,38	1,65E-02	71,83	0,69
HTR2A	0,00E+00	948,29	0,00E+00	499,30	0,53
HTR2B	0,00E+00	302,78	2,41E-02	60,52	0,20
MYLK	0,00E+00	2597,43	1,13E-06	924,63	0,36
P2RX4	0,00E+00	794,48	0,00E+00	353,74	0,45
PLCD1	0,00E+00	525,31	0,00E+00	331,99	0,63
PLCD3	0,00E+00	315,94	0,00E+00	207,80	0,66
PLCG1	7,24E-05	524,48	4,52E-04	240,40	0,46
PRKCA	0,00E+00	1549,11	0,00E+00	811,29	0,52

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	3TF Det P-value	3TF AVG_Signal	Ratio 3TF/FIB
TRPC1	0,00E+00	618,95	0,00E+00	387,25	0,63

Table 26. Pathways enriched among differentially expressed genes following silmutaneous over-expression of OCT4, SOX2, KLF4 and c-MYC. These table shows the raw data obtained from BeadStudio 3.0 (Illumina) from genes differentially expressed that belong to pathways highlighted in Table 14. The first column relates the gene name. In the remaining columns have been listed the detection P-value ("Det P-value") and the average signal intensity ("AVG_Signal") found in fibroblasts and 4TF (abbreviation for OCT4, SOX2, KLF4 and c-MYC) datasets. The last column relates the ratio of average signal intensity relative to non-transduced fibroblasts ("Fibroblasts") for each gene and sample. Samples refer to the average of three different biological replicates.

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	4TF Det P-value	4TF AVG_Signal	Ratio 4TF/FIB
Homologous recombination					
BLM	7,22E-02	18,24	1,50E-03	153,07	8,39
EME1	3,76E-02	29,85	7,52E-03	80,80	2,71
MRE11A	3,91E-05	59,84	6,90E-06	103,76	1,73
POLD1	3,01E-03	150,63	0,00E+00	312,77	2,08
RAD51C	4,96E-02	331,82	6,20E-03	888,48	2,68
RPA1	0,00E+00	2053,80	0,00E+00	3086,57	1,50
RPA2	0,00E+00	1143,56	0,00E+00	2236,25	1,96
RPA3	0,00E+00	1302,98	0,00E+00	2898,74	2,22
TOP3A	1,50E-03	212,80	0,00E+00	480,24	2,26
Glycolysis / Gluconeogenesis					
ALDH1B1	3,01E-03	155,84	0,00E+00	383,87	2,46
ALDH3A1	7,64E-01	-8,63	0,00E+00	1161,55	-134,62
ALDH3A2	1,85E-02	103,87	9,16E-05	226,75	2,18
DLAT	0,00E+00	583,81	0,00E+00	927,43	1,59
ENO1	0,00E+00	18661,41	0,00E+00	39222,54	2,10
FBP1	4,51E-01	-1,95	0,00E+00	310,92	-159,66
GPI	0,00E+00	1743,43	0,00E+00	3603,50	2,07
LDHA	0,00E+00	11707,23	0,00E+00	22899,59	1,96
PFKM	0,00E+00	1038,82	0,00E+00	1700,39	1,64
PGAM1	0,00E+00	3410,17	0,00E+00	5456,78	1,60
PGAM2	1,67E-01	7,28	7,52E-03	80,19	11,02
PGM2	0,00E+00	288,65	0,00E+00	434,99	1,51
TPI1	0,00E+00	14020,27	0,00E+00	22088,18	1,58
DNA replication					
MCM3	0,00E+00	1764,81	0,00E+00	5443,87	3,08
MCM6	0,00E+00	1580,33	0,00E+00	4360,85	2,76
POLA1	9,02E-03	89,20	0,00E+00	192,81	2,16
POLD1	3,01E-03	150,63	0,00E+00	312,77	2,08
POLE	4,51E-03	91,68	0,00E+00	222,91	2,43
POLE3	0,00E+00	2568,61	0,00E+00	7893,73	3,07
RPA1	0,00E+00	2053,80	0,00E+00	3086,57	1,50
RPA2	0,00E+00	1143,56	0,00E+00	2236,25	1,96
RPA3	0,00E+00	1302,98	0,00E+00	2898,74	2,22

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	4TF Det P-value	4TF AVG_Signal	Ratio 4TF/FIB
MAPK signalling pathway					
ATF2	0,00E+00	315,70	0,00E+00	210,02	0,67
BDNF	6,54E-04	404,51	9,68E-04	229,83	0,57
CACNA1C	1,50E-03	183,77	5,86E-02	50,96	0,28
CHP	0,00E+00	795,08	0,00E+00	451,74	0,57
DAXX	0,00E+00	878,47	0,00E+00	496,24	0,56
DUSP1	0,00E+00	3922,86	0,00E+00	1962,78	0,50
ELK1	0,00E+00	1720,02	0,00E+00	1068,00	0,62
FGF2	0,00E+00	264,41	1,50E-03	170,51	0,64
FGF5	2,83E-05	132,48	7,24E-05	85,38	0,64
FGF7	4,51E-03	100,28	1,43E-01	43,04	0,43
FGFR1	7,07E-04	87,64	3,91E-02	44,39	0,51
FLNA	0,00E+00	223,64	6,02E-03	94,70	0,42
GADD45B	1,50E-03	210,49	4,51E-03	107,70	0,51
GNA12	0,00E+00	1392,53	0,00E+00	923,08	0,66
GNG12	0,00E+00	1536,67	0,00E+00	817,83	0,53
HSPA2	0,00E+00	824,57	1,35E-02	67,11	0,08
HSPB1	0,00E+00	10971,31	0,00E+00	6870,50	0,63
IL1R1	1,50E-03	189,72	1,49E-01	42,19	0,22
JUND	0,00E+00	21877,40	0,00E+00	12295,09	0,56
MAP2K4	0,00E+00	954,35	0,00E+00	603,89	0,63
MAP2K5	1,02E-05	286,51	9,16E-05	145,96	0,51
MAP3K12	1,50E-03	188,57	4,51E-03	107,45	0,57
MAP3K4	1,16E-03	540,09	2,03E-02	529,86	0,98
MAP3K6	0,00E+00	780,46	0,00E+00	439,38	0,56
MAP4K4	1,72E-03	1149,23	7,06E-03	522,69	0,45
MAPK7	9,95E-05	263,09	2,23E-05	171,19	0,65
MAPK8IP3	1,09E-03	904,83	2,71E-03	484,35	0,54
MAPKAPK2	2,22E-04	240,67	4,52E-04	146,26	0,61
MYC	1,50E-03	184,28	6,02E-03	91,02	0,49
NFKB2	9,79E-04	111,71	1,70E-02	80,49	0,72
NR4A1	9,07E-03	24,97	4,51E-01	35,05	1,40
NTF3	0,00E+00	645,10	0,00E+00	208,28	0,32
PAK2	4,52E-06	510,95	1,13E-06	338,09	0,66
PLA2G6	2,94E-03	34,95	1,66E-02	50,84	1,45
PPP3CB	0,00E+00	3100,89	0,00E+00	1739,50	0,56
PRKCA	0,00E+00	1549,11	0,00E+00	543,66	0,35
PTPN5	5,70E-03	22,80	1,19E-01	42,49	1,86
RAF1	0,00E+00	1816,87	0,00E+00	1102,15	0,61
RELB	0,00E+00	242,52	1,50E-03	130,59	0,54
RRAS	0,00E+00	7520,62	0,00E+00	4262,34	0,57
RRAS2	0,00E+00	607,64	0,00E+00	405,42	0,67
STMN1	5,59E-03	388,34	1,87E-02	550,58	1,42
TGFBR2	6,51E-04	2306,33	1,13E-04	1319,63	0,57

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	4TF Det P-value	4TF AVG_Signal	Ratio 4TF/FIB
TNFRSF1A	0,00E+00	5957,80	0,00E+00	2883,15	0,48
ZAK	0,00E+00	4659,97	0,00E+00	1706,05	0,37
Wnt signalling pathway					
CHP	0,00E+00	795,08	0,00E+00	451,74	0,57
CSNK1A1	1,12E-05	636,28	3,36E-05	409,89	0,64
CSNK1E	5,50E-04	2806,86	1,78E-04	1840,17	0,66
CTBP1	0,00E+00	728,62	0,00E+00	407,06	0,56
DAAM2	3,01E-03	139,27	7,52E-03	80,46	0,58
DVL2	0,00E+00	912,19	0,00E+00	459,02	0,50
DVL3	0,00E+00	1105,43	0,00E+00	630,30	0,57
EP300	0,00E+00	503,02	0,00E+00	306,35	0,61
FZD2	0,00E+00	3249,84	0,00E+00	848,77	0,26
FZD6	3,01E-03	148,59	6,02E-03	94,93	0,64
LEF1	0,00E+00	306,13	4,51E-03	114,87	0,38
MYC	1,50E-03	184,28	6,02E-03	91,02	0,49
NKD2	0,00E+00	575,29	2,03E-01	39,95	0,07
PPP2R5A	0,00E+00	1099,45	0,00E+00	541,70	0,49
PPP3CB	0,00E+00	3100,89	0,00E+00	1739,50	0,56
PRICKLE2	0,00E+00	1281,25	0,00E+00	724,72	0,57
PRKCA	0,00E+00	1549,11	0,00E+00	543,66	0,35
ROCK1	0,00E+00	242,58	6,02E-03	100,65	0,41
SFRP1	0,00E+00	12751,53	0,00E+00	7791,30	0,61
SMAD3	0,00E+00	4188,80	0,00E+00	1885,84	0,45
SMAD4	0,00E+00	2185,40	0,00E+00	1326,36	0,61
TBL1X	0,00E+00	1956,63	0,00E+00	1141,83	0,58
TBL1XR1	0,00E+00	258,50	1,50E-03	122,85	0,48
WNT5A	0,00E+00	3827,64	0,00E+00	1370,92	0,36
Regulation of actin cytoskeleton					
BCAR1	0,00E+00	1422,77	0,00E+00	642,74	0,45
BDKRB1	0,00E+00	3267,96	0,00E+00	480,26	0,15
BDKRB2	0,00E+00	742,42	1,35E-02	67,60	0,09
FGD1	0,00E+00	724,51	0,00E+00	390,56	0,54
FGF2	0,00E+00	264,41	1,50E-03	170,51	0,64
FGF5	2,83E-05	132,48	7,24E-05	85,38	0,64
FGF7	4,51E-03	100,28	1,43E-01	43,04	0,43
FGFR1	7,07E-04	87,64	3,91E-02	44,39	0,51
FN1	5,33E-04	264,45	2,19E-02	113,87	0,43
GNA12	0,00E+00	1392,53	0,00E+00	923,08	0,66
GNG12	0,00E+00	1536,67	0,00E+00	817,83	0,53
IQGAP1	0,00E+00	533,29	0,00E+00	284,04	0,53
ITGA1	0,00E+00	720,05	0,00E+00	199,86	0,28
ITGA11	0,00E+00	1631,33	1,13E-06	223,52	0,14
ITGA2	0,00E+00	1026,28	0,00E+00	597,33	0,58
ITGA3	1,16E-03	1159,36	7,42E-03	745,16	0,64
ITGAV	0,00E+00	5066,86	0,00E+00	3237,23	0,64
MYH9	0,00E+00	24058,16	0,00E+00	12286,39	0,51
MYL9	0,00E+00	1368,62	1,13E-06	451,86	0,33
MYLK	0,00E+00	2597,43	1,02E-05	399,35	0,15

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	4TF Det P-value	4TF AVG_Signal	Ratio 4TF/FIB
NCKAP1	0,00E+00	1603,12	0,00E+00	1006,18	0,63
PAK2	4,52E-06	510,95	1,13E-06	338,09	0,66
PDGFC	0,00E+00	891,12	0,00E+00	510,14	0,57
PIK3CD	1,50E-03	170,58	4,51E-03	110,51	0,65
PPP1CB	0,00E+00	1726,48	0,00E+00	908,60	0,53
RAF1	0,00E+00	1816,87	0,00E+00	1102,15	0,61
ROCK1	0,00E+00	242,58	6,02E-03	100,65	0,41
RRAS	0,00E+00	7520,62	0,00E+00	4262,34	0,57
RRAS2	0,00E+00	607,64	0,00E+00	405,42	0,67
SSH2	0,00E+00	754,94	0,00E+00	423,04	0,56
TGF-beta signalling pathway					
ACVRL1	1,50E-03	177,80	8,27E-02	47,52	0,27
BMP2	0,00E+00	2426,98	0,00E+00	539,61	0,22
BMP4	6,27E-05	1729,04	3,24E-01	111,06	0,06
BMPR2	0,00E+00	1098,65	0,00E+00	403,74	0,37
EP300	0,00E+00	503,02	0,00E+00	306,35	0,61
ID2	0,00E+00	503,18	0,00E+00	298,89	0,59
LTBP1	1,72E-03	161,00	5,47E-02	64,72	0,40
MYC	1,50E-03	184,28	6,02E-03	91,02	0,49
ROCK1	0,00E+00	242,58	6,02E-03	100,65	0,41
SMAD3	0,00E+00	4188,80	0,00E+00	1885,84	0,45
SMAD4	0,00E+00	2185,40	0,00E+00	1326,36	0,61
SP1	0,00E+00	529,71	0,00E+00	313,88	0,59
TGFBR2	6,51E-04	2306,33	1,13E-04	1319,63	0,57
THBS1	0,00E+00	23037,12	0,00E+00	6864,03	0,30
THBS3	0,00E+00	221,63	6,02E-03	101,88	0,46
ErbB signalling pathway					
CBL	1,50E-03	193,41	6,02E-03	100,78	0,52
ELK1	0,00E+00	1720,02	0,00E+00	1068,00	0,62
MAP2K4	0,00E+00	954,35	0,00E+00	603,89	0,63
MYC	1,50E-03	184,28	6,02E-03	91,02	0,49
NCK1	0,00E+00	1962,67	0,00E+00	1166,72	0,59
PAK2	4,52E-06	510,95	1,13E-06	338,09	0,66
PIK3CD	1,50E-03	170,58	4,51E-03	110,51	0,65
PLCG1	7,24E-05	524,48	5,54E-05	299,15	0,57
PRKCA	0,00E+00	1549,11	0,00E+00	543,66	0,35
RAF1	0,00E+00	1816,87	0,00E+00	1102,15	0,61
SHC1	0,00E+00	8481,03	5,67E-10	5579,88	0,66
SHC3	1,50E-03	176,96	7,52E-03	80,03	0,45
STAT5A	0,00E+00	444,36	1,05E-02	75,31	0,17
STAT5B	0,00E+00	215,23	1,50E-03	130,64	0,61
Endocytosis					
AP2A1	4,71E-04	188,45	7,17E-03	106,72	0,57
CBL	1,50E-03	193,41	6,02E-03	100,78	0,52
CHMP1B	0,00E+00	3656,42	0,00E+00	2405,04	0,66
CLTB	6,70E-03	1595,00	2,28E-02	2422,24	1,52
DAB2	0,00E+00	2840,46	0,00E+00	1860,88	0,66
DNM1	0,00E+00	223,26	2,11E-02	62,36	0,28
EHD2	0,00E+00	741,50	1,50E-03	163,31	0,22
FAM125A	0,00E+00	2072,66	0,00E+00	1171,78	0,57
GRK5	0,00E+00	2032,10	0,00E+00	630,59	0,31
HLA-E	0,00E+00	637,62	0,00E+00	255,72	0,40

GENE SYMBOL	Fibroblasts Det P-value	Fibroblasts AVG_Signal	4TF Det P-value	4TF AVG_Signal	Ratio 4TF/FIB
HLA-F	0,00E+00	461,50	0,00E+00	246,88	0,53
HSPA2	0,00E+00	824,57	1,35E-02	67,11	0,08
PARD6G	4,52E-06	390,97	7,24E-05	119,19	0,30
PLD2	0,00E+00	281,10	4,51E-03	118,34	0,42
PRKCI	1,50E-03	184,31	4,51E-03	108,69	0,59
RAB11B	0,00E+00	247,38	6,02E-03	89,86	0,36
RAB11FIP3	0,00E+00	2271,36	0,00E+00	1303,55	0,57
RAB22A	0,00E+00	2054,95	0,00E+00	1259,06	0,61
RAB4A	0,00E+00	553,76	0,00E+00	365,88	0,66
STAM2	0,00E+00	273,34	1,50E-03	130,68	0,48
TGFBR2	6,51E-04	2306,33	1,13E-04	1319,63	0,57
VPS28	0,00E+00	2329,62	0,00E+00	1241,41	0,53
VPS37D	0,00E+00	1269,86	0,00E+00	849,17	0,67
VPS4B	0,00E+00	2690,52	0,00E+00	1707,10	0,63
Notch signalling pathway					
CTBP1	0,00E+00	728,62	0,00E+00	407,06	0,56
DTX3	0,00E+00	602,63	0,00E+00	206,00	0,34
DVL2	0,00E+00	912,19	0,00E+00	459,02	0,50
DVL3	0,00E+00	1105,43	0,00E+00	630,30	0,57
EP300	0,00E+00	503,02	0,00E+00	306,35	0,61
LFNG	2,24E-03	112,42	2,13E-02	128,53	1,14
NUMB	0,00E+00	3919,27	0,00E+00	2371,60	0,61
RFNG	1,53E-08	1487,00	3,63E-08	993,53	0,67
SNW1	1,50E-03	178,84	4,51E-03	117,10	0,65