DNA binding proteins and cell fate

Crisbel Marie Artuz

A thesis submitted for the degree of Doctor of Philosophy
(Biochemistry and Molecular Genetics)

School of Biotechnology and Biomolecular Sciences
The University of New South Wales

January 2015

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: ARTUZ

First name: CRISBEL MARIE Other name/s:

Abbreviation for degree as given in the University calendar: PhD

School: BABS (Biotechnology and Biomolecular Sciences) Faculty: SCIENCE

Title: DNA binding proteins and cell fate

Abstract 350 words maximum:

Sequence-specific DNA binding proteins, known as transcription factors, play a central role in the control of eukaryotic gene regulation. Understanding the mechanisms through which DNA binding domains recognise their target sequences will greatly improve our understanding of genetic diseases that result from mutations in DNA binding domains and gene promoters. Such information will also assist in the design of factors capable of artificially controlling gene expression.

The zinc finger motif, commonly present in tandem arrays of three or more fingers, is the most prevalent DNA recognition structure found in eukaryotic transcription factors. The first project in this thesis aimed to better understand how zinc finger domains bind DNA by examining the two-zinc finger motif of the transcriptional regulator and oncogene ZNF217. By performing a comprehensive mutagenesis analysis, we were able to identify the amino acid residues that are essential for DNA recognition. Our findings indicate that ZNF217 binds to its preferred consensus site by a novel mechanism, an understanding of which may lead to a better appreciation of diseases that result from dysregulation of ZNF217 oncogenic function, and ultimately to the design of novel therapeutic strategies.

In the second project, we examined the potential of DNA binding proteins to alter gene expression networks and hence cell fate, in the context of reprogramming fibroblasts towards the megakaryocytic lineage. Megakaryocytes are required for the production of platelets, which are essential for blood coagulation. Reduction in their numbers causes a life-threatening condition termed thrombocytopenia, which is currently treated by platelet transfusions. However, this treatment is restricted by the short storage life and limited supply of platelet concentrates. To investigate alternative approaches, we examined the potential of ectopic expression of combinations of transcription factors to direct fibroblasts towards the megakaryocyte lineage. We have discovered that over-expression of a combination of GATA1 or its mutant isoform, GATA1 short (GATA1s), FLI1 and TAL1 can drive phenotypic changes consistent with partial reprogramming of fibroblasts towards the megakaryocyte lineage, laying the foundation for follow up studies.

Declaration relating to disposition of project thesis/dissertation

I hereby grant to the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or in part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all property rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

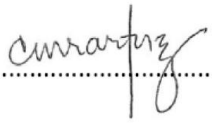
I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstracts International (this is applicable to doctoral theses only).

Signature	Witness	Date
	Thi Thanh Va	08 / 02 / 2015

The University recognises that there may be exceptional circumstances requiring restrictions on copying or conditions on use. Requests for restriction for a period of up to 2 years must be made in writing. Requests for a longer period of restriction may be considered in exceptional circumstances and require the approval of the Dean of Graduate Research.

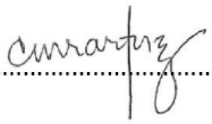
FOR OFFICE USE ONLY Date of completion of requirements for Award:

COPYRIGHT STATEMENT

Signed 

Date 08 / 02 / 2015

AUTHENTICITY STATEMENT

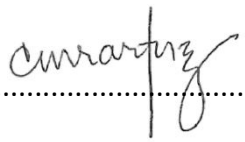
Signed 

Date 08 / 02 / 2015

Table of Contents

Originality Statement

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed 

Date 08 / 02 / 2015

Acknowledgements

I have been very fortunate to take a three and a half year-roller coaster ride of PhD experience with an amazing bunch of people, all of whom deserve to be mentioned as they made the ride enjoyable and unforgettable.

Firstly, I would like to express my heartfelt gratitude to my supervisor, Merlin Crossley, who has inspired me to continue my journey in research. Merlin's enthusiasm for Science has been enjoyable to witness and will remain to motivate me in my journey ahead. I am also extremely grateful to my co-supervisor, Richard Pearson, who took the time to go through the painful process of proof-reading this thesis. His patience and support has helped me throughout my PhD years.

A massive thank you goes to our past lab managers - Robert, Stella and Bev, for their huge efforts in chasing my orders and ensuring the lab was running smoothly. To the past and present Crossley lab members: Ali, Briony, Noe, Feyza, Thanh, Cassie, Vitri, Dimitri, Jon, Laura, Zara, Becky, Cath, Jasmine, Lu, Tanit, Alex and Gabbie, I want to give you all a big hug and thank you for making the lab, where I spent the greater majority of my days, feel like a home. All of you created a fun environment for me to work in. To Ali, Jon, Laura, Cath, Jasmine and especially Cassie, you guys made my lab experience bearable in the low times and I really appreciate your friendship and emotional support. I couldn't imagine going through those years without you! To Nirmani (Lutze-Mann lab), James (Brown lab), Jack, Tamar and Robin (Yang lab) thank you so much for your warm welcome when our lab came to UNSW. I will surely miss our endless laughter in the tissue culture room. I would also like to thank our collaborators from USYD, Joel Mackay and Marylene Vandevenne for their assistance in the ZNF217 project. I am very grateful for the financial support of an Australian Postgraduate Award (APA).

Most importantly, I would like to thank my family – Papa, Mama and Cris for their love and support throughout the years. I would have not been able to go through with this without your encouragement. I love you all very much.

Publications arising from this thesis

Journal articles

Vandevenne M, Jacques DA, Artuz CM, Nguyen CD, Kwan AH, Segal DJ, Matthews JM, Crossley M, Guss JM, Mackay JP (2013) New insights into DNA recognition by zinc fingers revealed by structural analysis of the oncoprotein ZNF217. *The Journal of Biological Chemistry* 288: 10616-10627.

Nunez N, Clifton MM, Funnell AP, Artuz CM, Hallal S, Quinlan KG, Font J, Vandevenne M, Setiyaputra S, Pearson RC, Mackay JP, Crossley M (2011) The multi-zinc finger protein ZNF217 contacts DNA through a two-finger domain. *The Journal of Biological Chemistry* 286: 38190-38201.

Conference abstracts

Artuz CM, Funnell APW, Pearson RCM, Gonda TJ, Crossley M. Directing the Differentiation of Fibroblasts towards the Megakaryocyte lineage. In: *ComBio 2013, Perth, Australia*. Poster presentation.

Artuz CM, Funnell APW, Pearson RCM, Gonda TJ, Crossley M. Directing the Differentiation of Fibroblasts towards the Megakaryocyte lineage. In: *The 34th Annual Lorne Genome Conference, Lorne, Australia*. Poster presentation.

Artuz CM, Funnell APW, Pearson RCM, Gonda TJ, Crossley M. Cellular reprogramming towards the megakaryocyte lineage. In: *ComBio 2012, Adelaide, Australia*. Poster presentation.

ABSTRACT

Sequence-specific DNA binding proteins, known as transcription factors, play a central role in the control of eukaryotic gene regulation. Understanding the mechanisms through which DNA binding domains recognise their target sequences will greatly improve our understanding of genetic diseases that result from mutations in DNA binding domains and gene promoters. Such information will also assist in the design of factors capable of artificially controlling gene expression.

The zinc finger motif, commonly present in tandem arrays of three or more fingers, is the most prevalent DNA recognition structure found in eukaryotic transcription factors. The first project in this thesis aimed to better understand how zinc finger domains bind DNA by examining the two-zinc finger motif of the transcriptional regulator and oncogene ZNF217. By performing a comprehensive mutagenesis analysis, we were able to identify the amino acid residues that are essential for DNA recognition. Our findings indicate that ZNF217 binds to its preferred consensus site by a novel mechanism, an understanding of which may lead to a better appreciation of diseases that result from dysregulation of ZNF217 oncogenic function, and ultimately to the design of novel therapeutic strategies.

In the second project, we examined the potential of DNA binding proteins to alter gene expression networks and hence cell fate, in the context of reprogramming fibroblasts towards the megakaryocytic lineage. Megakaryocytes are required for the production of platelets, which are essential for blood coagulation. Reduction in their numbers causes a life-threatening condition termed thrombocytopenia, which is

currently treated by platelet transfusions. However, this treatment is restricted by the short storage life and limited supply of platelet concentrates. To investigate alternative approaches, we examined the potential of ectopic expression of combinations of transcription factors to direct fibroblasts towards the megakaryocyte lineage. We have discovered that over-expression of a combination of GATA1 or its mutant isoform, GATA1 short (GATA1s), FLI1 and TAL1 can drive phenotypic changes consistent with partial reprogramming of fibroblasts towards the megakaryocyte lineage, laying the foundation for follow up studies.

Abbreviations

AChE	Acetylcholinesterase
APS	Ammonium persulphate
ARVEC	Arrayed RetroViral Expression Cloning
ATF	Artificial transcription factor
BFU-MK	Burst-forming unit megakaryocyte
C/EBP	CCAAT/enhancer binding protein
cDNA	complementary deoxyribonucleic acid
CFU-MK	Colony-forming unit
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
CRS	Core recognition sequence
CtBP	C-terminal Binding Protein
Cys (C)	Cysteine
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DTT	Dithiothreitol
EKLF	Erythroid Krüppel-like Factor
ELISA	Enzyme-linked immunosorbent assay
EMSA	Electrophoretic mobility shift assay
EPO	Erythropoietin
ESC	Embryonic stem cell
F6	Finger 6
F7	Finger 7
FDR	False discovery rate
FLI1	Friend leukaemia integration 1
GATA1	GATA-binding protein 1
GATA1s	GATA1 short
GATA2	GATA-binding protein 2
GFP	Green fluorescent protein
Glu (E)	Glutamic acid
Gly (G)	Glycine

GM-CSF	Granulocyte-macrophage colony-stimulating factor
GMP	Granulocyte/macrophage progenitor
Gp9	Glycoprotein 9
GST	Glutathione-S-Transferase
H3	Histone 3
H4	Histone 4
HAT	Histone acetyl transferase
HDAC	Histone deacetylases
HEPES	4-(2-Hydroxyethyl)-1-piperazin-ethan-sulfonsäure
His (H)	Histidine
HMT	Histone methyltransferase
HNF4	Hepatocyte nuclear factor 4
HSC	Haematopoietic stem cell
HSQC	Heteronuclear Single Quantum Coherence
IL	Interleukin
iPSC	induced pluripotent stem cell
IRES	Internal ribosome entry site
kDa	kiloDalton
KLF1	Krüppel-like Factor 1
KLF4	Krüppel-like Factor 4
LIF	Leukemia inhibitory factor
Lys (K)	Lysine
MEF	Murine embryonic fibroblast
MEP	Megakaryocyte/erythroid progenitor
MK	Megakaryocyte
MKLI	Megakaryocyte lineage induction
mRNA	messenger ribonucleic acid
NFE2	Nuclear factor, erythroid 2
OCT4	Octamer-binding transcription factor 4
Pax5	Paired box 5
Pax6	Paired box 6
PCR	Polymerase chain reaction
Pf4	Platelet factor 4

Pro (P)	Proline
PU.1	SFFV proviral integration 1 transcription factor
RNA	Ribonucleic acid
RNA pol II	RNA polymerase II
RT-PCR	Reverse transcriptase polymerase chain reaction
RUNX1	Runt-related transcription factor 1
SOX2	Sex-determining region Y (SRY)-box 2
SWI/SNF	Swith/sucrose non-fermentable
TAFs	TBP-associated factors
TAL1	T-cell acute lymphocytic leukaemia 1
TBP	TATA-binding protein
TF	Transcription factor
Thr (T)	Threonine
TPO	Thrombopoietin
TSA	Trichostatin A
ZNF217	Zinc finger 217

CHAPTER 1 - GENERAL INTRODUCTION

In this thesis, we explore two distinct projects linked by the theme of transcriptional regulation, a process controlled primarily by the actions of sequence-specific DNA-binding proteins known as transcription factors. The first project furthers our knowledge of the interaction between DNA-binding domains and their target sequences by studying a novel mechanism of DNA-binding by the zinc finger oncogene, ZNF217. Studying the mechanisms through which transcription factors bind DNA is essential for understanding genetic diseases that result from mutation of DNA-binding domains or binding sites. For example, the inherited bleeding disorder Haemophilia B, can result from mutations within the blood clotting factor IX promoter, which affect binding of transcription factors such as C/EBP, Androgen Receptor, HNF4, and ONECUT1/2, altering normal gene expression (Crossley & Brownlee, 1990; Crossley et al, 1992; Funnell et al, 2013; Reijnen et al, 1993; Reijnen et al, 1992). Furthermore, an understanding of DNA binding can also be applied to the construction of artificial transcription factors using in particular, zinc finger domains to recognise DNA sequences and modulate gene expression for both therapeutic and experimental purposes.

The ability of transcription factors to control gene expression is extended in the second project where we apply our understanding of how transcription factors direct developmental programs to the process of cellular reprogramming, focusing on the conversion of fibroblasts to megakaryocytes.

1.1 Transcriptional regulation of gene expression in eukaryotes

Classical gene expression is the process by which a gene's genomic DNA sequence is converted into messenger RNA (mRNA) by means of transcription, which in most cases is followed by translation of mRNA into protein. In eukaryotes, gene regulation is critical to generating and maintaining the numerous specialised cell types which, though genetically identical, are phenotypically diverse. To achieve this, all stages in the gene expression pathway from the initial transcription of DNA into RNA to the formation of the final protein product are regulated, with a particular emphasis in eukaryotic systems on control at the level of transcription (Proudfoot et al, 2002).

Eukaryotic DNA is assembled into a packaged structure called chromatin. The fundamental unit of chromatin, the nucleosome, is composed of an octamer of the four core histones (H3, H4, H2A, H2B) around which 147 base pairs (bp) of DNA are wrapped (Loidl, 2001). Due to this intricate packaging of DNA, regulation of eukaryotic gene expression relies on a complex interplay between *cis*-acting DNA elements, the basal transcriptional apparatus, chromatin modifying and remodelling complexes and most especially regulatory transcription factors. Transcription factors play a central role in transcriptional regulation.

1.1.1 Initiation of transcription by the basal transcriptional apparatus

In eukaryotes, transcription is initiated by a large, multi-protein complex known as the basal transcriptional apparatus that typically assembles at the core promoter regions of genes. This complex is composed of the 'general transcription factors' TFIIA, -B, -D, -E, -F and –H, together with RNA polymerase II, the enzyme

responsible for the transcription of protein-coding genes (Figure 1.1) (Roeder, 1996). Of these components, only TFIID can recognise and bind to the TATA element situated approximately 25 bp upstream of the transcriptional start site of many genes (Burley & Roeder, 1996; Nikolov & Burley, 1997). TFIID is a multi-subunit complex that contains the TATA-binding protein (TBP) and a set of as many as ten TBP-associated factors (TAFs). The TBP-DNA complex provides a platform to recruit RNA polymerase II and the remaining general transcription factors to initiate transcription. Notably, however, not all eukaryotic genes contain a TATA element, therefore, in non-TATA genes the basal transcriptional apparatus is assembled at other DNA elements, such as initiator element (Inr), downstream promoter element (DPE) and TFIIB recognition element (BRE) (Smale & Kadonaga, 2003).

In vitro, the recruitment of the basal transcription apparatus to the promoter regions is sufficient to initiate transcription. However, *in vivo* additional factors such as regulatory transcription factors and chromatin modifying and remodelling complexes are required to overcome the effects of chromatin structure that limits access to the basal transcriptional apparatus (Roeder, 2005; Thomas & Chiang, 2006).

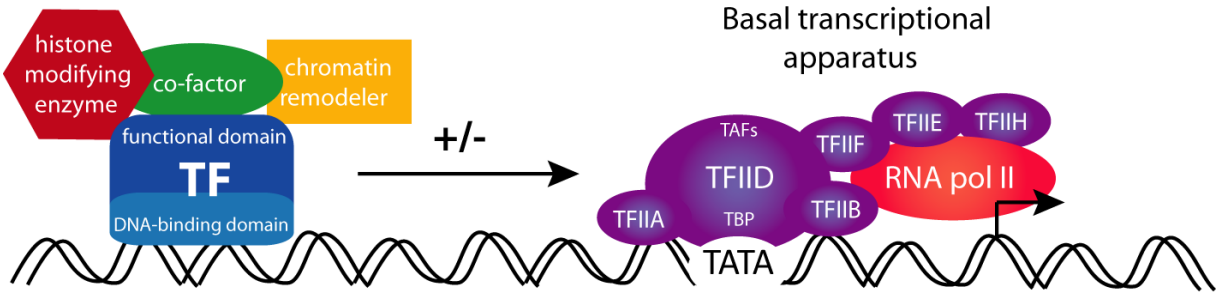


Figure 1.1. *Transcriptional regulation of gene expression in eukaryotes is primarily controlled by sequence-specific regulatory transcription factors and their co-factors, which can alter chromatin structure and influence assembly of the basal transcriptional apparatus.* The basal transcriptional apparatus is composed of the general transcription factors TFIIA, -B, -D, -E, -F and -H, and RNA polymerase II (RNA pol II). The TFIID complex which contains the TATA-binding protein (TBP) and several TBP-associated factors (TAFs) recognises the TATA element in the core promoter region of certain genes. The assembly of this apparatus is influenced by regulatory transcription factors (TF) that can bind to specific sequences of the regulatory elements of DNA through its DNA-binding domain and to co-factors through its functional domain. The co-factors then recruit histone-modifying enzymes and chromatin remodellers to alter chromatin structure. This transcription complex can either promote or impede the assembly of the basal transcriptional apparatus to regulate gene expression.

1.1.2 Transcriptional control by the regulatory transcription factors

In contrast to the general transcription factors, regulatory transcription factors are multi-domain proteins that feature a functional (activation/repression) domain and a DNA-binding domain. It is through the functional domain that regulatory transcription factors influence gene activity, by means of mediating protein-protein interactions with co-regulatory partner proteins (known as co-factors) (Figure 1.1). To promote transcription, activation domains interact with activating co-factors (co-activators), whereas repression domains recruit repressive co-factors (co-repressors) to impede transcription. Both transcription factors and their co-factors influence the

initiation of transcription through a number of mechanisms. Typically, transcriptional activators and their co-activators recruit and stabilise the basal transcriptional apparatus (Roeder, 2005). In contrast, transcriptional repressors and their co-repressors function by inhibiting the interactions between the basal transcription apparatus and activator complexes (Thiel et al, 2004). Other transcription factors act by bringing distal enhancer or silencer elements into closer proximity with the basal transcriptional machinery (Agalioti et al, 2000; Maniatis et al, 1998; Wieczorek et al, 2000), and there are many instances where transcription factors alter the chromatin structure (Kadonaga, 2004), providing an additional level of transcriptional regulation important for controlled gene expression.

The assembly of the basal transcriptional apparatus at the promoters of genes is influenced by the degree of inter-nucleosomal contact which establishes the chromatin structure. Chromatin structure can be classified into two categories. Heterochromatin consists of tightly packed nucleosomes (DNA and histone proteins) and is regarded as refractory to transcription, whereas euchromatin is a less compact structure that is accessible to gene transcription (Thiel et al, 2004). Given that chromatin structure affects gene transcription, transcription factors facilitate alteration of DNA packaging by recruiting enzymes responsible for covalent modifications to histone proteins that are associated with particular types of chromatin. For example, active euchromatin is associated with the acetylation of specific lysine residues in the N-termini of core histones, such as Lys-9 residue of H3 (H3K9) and Lys-5 residue of H4 (H4K5) (Strahl & Allis, 2000). As such transcriptional activators like the haematopoietic transcription factors GATA1 and EKLF/KLF1, can recruit histone acetyl transferases (HATs) to add acetyl groups to

these specific histone residues and activate transcription of erythroid genes (Boyes et al, 1998; Utley et al, 1998; Zhang & Bieker, 1998). In contrast, removal of acetyl groups, particularly on H3K9, by histone deacetylases (HDACs) (Ng & Bird, 2000) recruited by transcriptional repressors, results in increased chromatin condensation and is linked with inactive heterochromatin (Grewal & Moazed, 2003; Pearson et al, 2011; Strahl & Allis, 2000). Furthermore, open or closed chromatin structure is dependent on methylation of specific histone residues by histone methyltransferases (HMTs), also recruited by transcription factors. For example, methylation of H3K4, H3K36 and H3K79 are implicated in activation of transcription, while methylation of H3K9, H3K27 and H4K20 are associated with transcriptional repression (Kouzarides, 2007).

In addition to the chromatin/histone modifying enzymes mentioned, DNA-bound transcription factors also recruit chromatin remodelling complexes such as the ATP-dependent SWI/SNF complex (Cote et al, 1994) to promoters of genes to modify chromatin structure and hence regulate transcription. For example, the transcription factor C/EBPβ recruits SWI/SNF to activate transcription of myeloid genes (Kowenz-Leutz & Leutz, 1999). SWI/SNF can create euchromatic or heterochromatic states by changing nucleosome organisation that may result in nucleosome sliding, histone octamer transfer to another DNA molecule, dinucleosome formation and altered nucleosome structure as demonstrated by SWI/SNF remodelling activity *in vitro* (Flaus & Owen-Hughes, 2001; Narlikar et al, 2002).

One well characterised example that showcases transcriptional control by regulatory transcription factors is evident in the activation of the IFN-β gene by the transcription factors, NF-κB, IRF3/IRF7 and the ATF2/c-Jun heterodimer, following viral infection (Maniatis et al, 1998; Munshi et al, 1999). The transcription factors, together with the architectural HMGA1 protein, bind on the enhancer of the IFN-β gene and promote enhanceosome assembly (Thanos et al, 1993). The enhanceosome then recruits the GCN5 complex, a histone acetyltransferase, to acetylate the nucleosomes that flank the enhancer. Once the GCN5 complex leaves the promoter, the transcription factors then recruit the CBP-PolII holoenzyme complex, which further acetylates the nucleosomes. This then facilitates recruitment of the SWI/SNF complex by CBP, which in turn remodels the nucleosome/chromatin (Agalioti et al, 2000). As a result of this, the basal transcriptional apparatus is recruited to the core promoter and transcription of the IFN-β gene is then initiated (Agalioti et al, 2000).

While transcription factors can activate or repress genes using their functional domain as the site for interaction of co-factors, histone modifying enzymes and chromatin remodelers, the DNA- binding domain guides the regulatory transcription factors to the appropriate *cis*-regulatory elements of genomic DNA and dictates sequence-specific binding (Kadonaga, 2004). The specificity of DNA binding can be achieved through a number of different structural domains, and transcription factors are often categorised by the nature of their DNA recognition domains. DNA-binding domains include homeodomains (Scott et al, 1989), zinc fingers (Wolfe et al, 2000), basic leucine zippers (Amoutzias et al, 2007) and helix-loop-helices (Jones, 2004). Notably, the zinc finger is the most prevalent structural motif found for transcription factors in eukaryotes (Ding et al, 2009; Fu et al, 2009).

1.2 DNA binding by classical zinc finger transcription factors

Zinc fingers are the major class of sequence-specific DNA binding domains found in eukaryotes accounting for approximately 3% of genes in the human genome (Lu et al, 2003; Tupler et al, 2001). The most common configuration for this domain, termed the classical (C_2H_2) zinc finger, possesses two cysteine (Cys) and two histidine (His) residues that coordinate a zinc ion in a tetrahedral fashion that stabilises the structure (Dang et al, 2000) (Figure 1.2A). The spacing of the zinc-binding residues is highly conserved conforming to a consensus of Cys-X_{2-4}-Cys-X_{12}-His-X_{2-6}-His, where X is any amino acid (Simpson et al, 2003). Zinc finger motifs are usually composed of approximately 25-30 amino acid residues, and contain a two-stranded antiparallel β-strand at its N-terminal end and one α-helix at its C-terminal end, which is referred to as a ββα structure (Figure 1.2A) (Mackay & Crossley, 1998).

Zinc fingers have been extensively studied as sequence-specific DNA-binding modules. Typically, classical zinc finger proteins are found to bind DNA target sequences using tandem arrays of three or more zinc fingers. Reported three-dimensional structures of several such arrays bound to DNA, such as Zif268 (Pavletich & Pabo, 1991), Gli5 (Pavletich & Pabo, 1993), and TFIIIA (Foster et al, 1997), reveal a shared recognition mode for classical zinc finger-DNA interactions. The structures demonstrate that a single C_2H_2 zinc finger can interact specifically with three to four base pairs in the major groove of DNA, thus allowing sequence-specific recognition of nine or ten bases of DNA by three classical zinc finger motifs (Figures 1.2B and C). The nucleotide base contacts are mediated primarily via the

basic and hydrophobic side chains of the amino acid residues located on the N-terminal surface of the α-helix of each zinc finger module (Matthews & Sunde, 2002; Wolfe et al, 2000). Binding occurs through hydrogen-bond interactions at helical positions -1, +2, +3 and +6 (relative to the beginning of the α-helix) of each zinc finger to the DNA. Typically, the residue at position +6 of each zinc finger contacts the 5' base of the three nucleotides that are recognised by that finger, the residue at position +3 contacts the central base, and the residue at position -1 contacts the 3' base. These contacts are made on the same strand of DNA, the primary strand, whereas the residue at position +2 often contacts a flanking base on the other strand of DNA, which may be a part of the binding region of the next finger, creating some overlap in the recognition of bases (Figure 1.2B) (Pabo et al, 2001; Pavletich & Pabo, 1991). Using the 5' → 3' convention for the direction of a DNA strand and the N → C convention for the direction of a polypeptide strand, it can be described that the zinc finger peptide is antiparallel to the primary strand of DNA where most of the contacts are made (Figure 1.2C) (Pavletich & Pabo, 1991). Therefore, the overall arrangement has the first (N-terminal) finger at the 3' end of the binding site on the primary strand, the second finger binds near the centre and the third finger near the 5' end of the primary strand. It is notable that DNA sequence specificity is achieved through variations in the key contact amino acid residues that make up the binding sites of the finger domain. Additionally, residues flanking those helical positions that interact with DNA and those located in the β-strands may also make contacts with the phosphodiester backbone of DNA through non-specific electrostatic interactions and contribute to some extent to binding affinity.

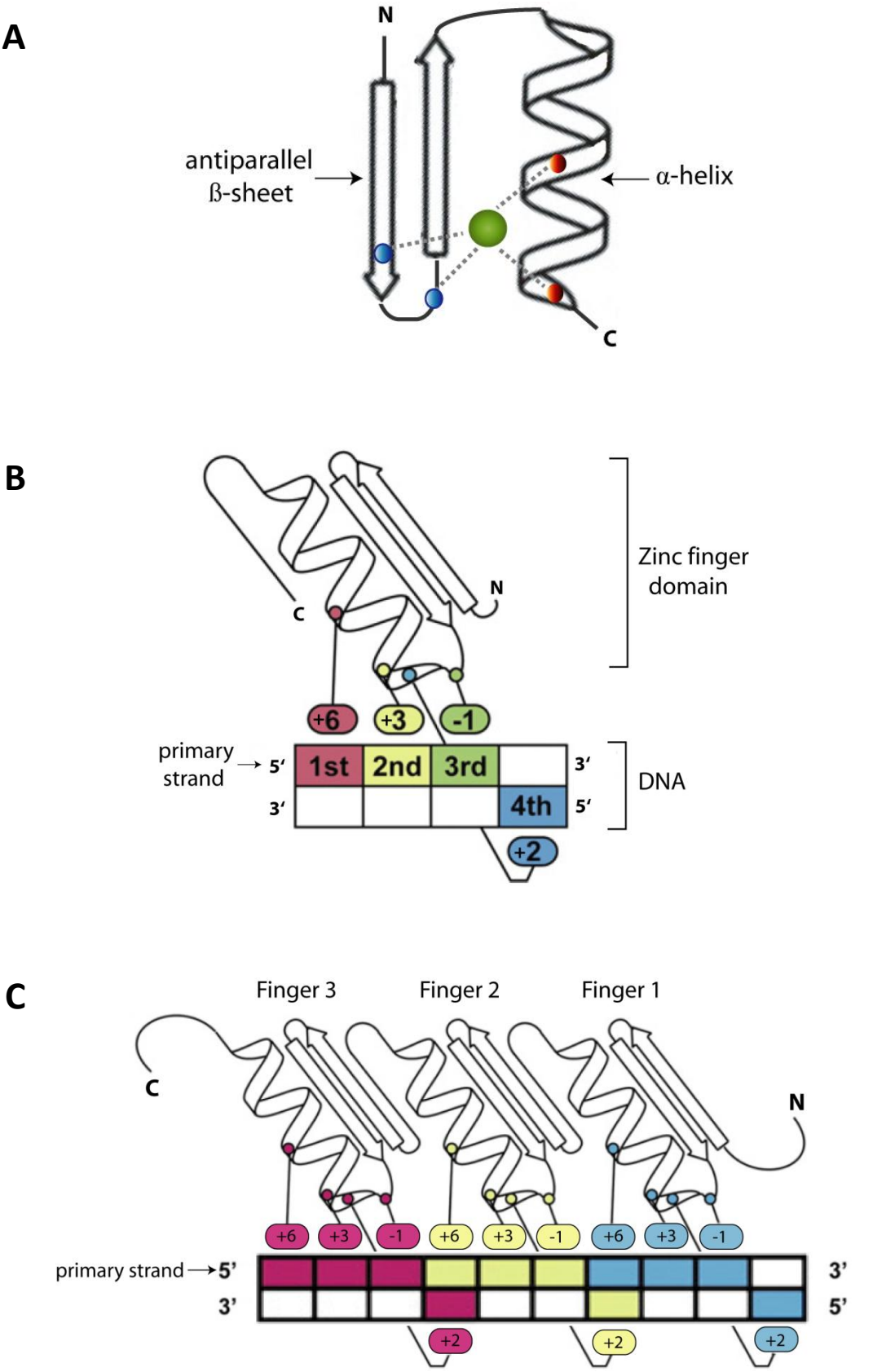


Figure 1.2. *Schematic representation of the classical C_2H_2 zinc finger motif and its interaction with double-stranded DNA.* **(A)** Schematic diagram of a C_2H_2 zinc finger showing

the two-stranded antiparallel β-sheet at the amino (N) terminal and the α-helix at the carboxyl (C) terminal (ββα structure). The pairs of Cys (blue circles) and His (red circles) residues coordinate the zinc ion (green circle) in a tetrahedral fashion. **(B)** Canonical interactions between the α-helical residues of a zinc finger domain and four specific base positions in DNA. Helical residues at positions -1, +2, +3 and +6 (relative to the start of the helix) that typically bind specific bases are indicated. **(C)** Model of how a 3-zinc finger protein interacts with DNA showing overlapping recognised bases and antiparallel binding of zinc fingers to DNA. The three zinc fingers mainly bind to the primary strand (non-coding strand of DNA), which are presented in the reverse direction to zinc fingers (Fingers 3 → 1 versus DNA 5' → 3'). Adapted from (Sera, 2009).

Furthermore, a highly conserved short amino acid linker sequence, Thr-Gly-Glu-Lys-Pro (TGEKP) separates adjacent zinc fingers in approximately half of known classical zinc fingers involved in DNA binding (Laity et al, 2000; Wolfe et al, 2000). This inter-domain linker is understood to stabilise the zinc finger-DNA complex by increasing the binding affinity and locking the conformation in place. When bound to DNA, each linker makes contact with the C-terminus of the α-helix of the preceding zinc finger via the Thr and Gly residues, termed C-capping, thereby enhancing the stability of the interaction (Foster et al, 1997; Laity et al, 2000; Wuttke et al, 1997). Moreover, the Lys residue of the TGEKP linker is able to make direct or water-mediated contact with the phosphate of the DNA backbone (Elrod-Erickson et al, 1996).

Because of the relatively conserved nature of zinc finger-DNA interactions, together with the stability and modularity of DNA binding, zinc fingers are now widely used as motifs in generating designer Artificial Transcription Factors (ATFs) to target a particular sequence of interest and modulate gene expression not only *in vitro* but

also *in vivo*. The modular nature of the zinc finger-DNA interaction makes it a relatively straight forward task to duplicate and modify designer zinc fingers to create proteins with novel DNA binding specificities (Klug, 2010; Papworth et al, 2006). Moreover, it is thought that binding of multiple fingers generally increases the specificity of binding by allowing recognition of longer sequences of DNA. As such, ATFs with 5- or 6- zinc fingers have been shown to specifically recognise genomic DNA sequences of greater than 16 bp (Bartsevich & Juliano, 2000; Beerli et al, 2000; Liu et al, 1997; Segal et al, 2006).

Although the conventional view has been that zinc finger proteins tend to bind DNA in clusters of three or more regularly and closely spaced finger modules to attain sequence-specific binding, there is a significant number of zinc finger proteins that do not have these three or more tandem-arranged zinc fingers such as the *Drosophila melanogaster* factors GAGA (Matharu et al, 2010) and Tramtrack (Fairall et al, 1993), having one and two zinc fingers respectively, raising the question of how they contact DNA. The zinc finger protein 217 (ZNF217), also binds DNA via a non-classical two-zinc finger domain, which is the focus of the first project of this thesis.

1.2.1 ZNF217, a multi-finger protein mediating DNA-binding via two zinc fingers

ZNF217 is a large multi-zinc finger protein that is composed of 1048 amino acid residues (Collins et al, 1998). Following its initial cloning, it was suggested that ZNF217 may function as a DNA-binding transcription factor due to the presence of eight C_2H_2 zinc fingers and a C-terminal proline-rich domain with potential to act as a functional domain (Figure 1.3) (Collins et al, 1998; Hanna-Rose & Hansen, 1996;

Mermod et al, 1989). These eight zinc fingers are irregularly spaced, with more than thirty amino acid residues separating adjacent fingers, with the exception of fingers 2-3 and fingers 6-7, which are separated by seven residues conforming to the spacing motif commonly found between adjoining zinc fingers involved in DNA-binding (Iuchi, 2001; Wolfe et al, 2000). Interestingly, only the sixth and seventh zinc fingers are separated by the canonical TGEKP linker, suggesting that of the eight zinc fingers of ZNF217, fingers 6-7 are the most likely to possess DNA-binding activity (Collins et al, 1998; Iuchi, 2001).

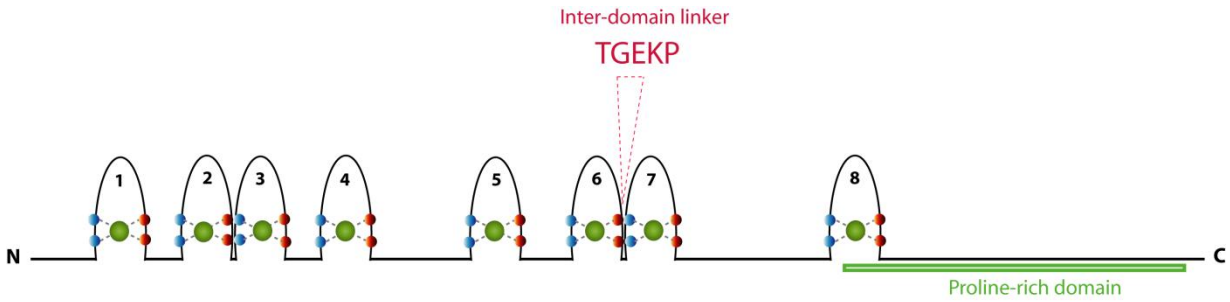


Figure 1.3. *Schematic of ZNF217*. Zinc fingers are shown as numbered arches with the Cys (blue) and His (red) residues coordinating the zinc ion (green) in the centre. The proline-rich domain and the canonical TGEKP linker located between zinc fingers 6 and 7 are indicated.

This hypothesis was first confirmed in a study, which proposed that the two zinc fingers recognise a 6-bp consensus sequence CAGAAY, where Y represents a C or T (Cowger et al, 2007). However, our laboratory has since shown this sequence to be a low affinity site and has provided an extended consensus to which stronger binding is observed (Nunez et al, 2011). The higher affinity recognition site consists of an 8-bp sequence, namely (T/A)(G/A)CAGAA(T/G/C). It is rare to have a paired zinc finger domain mediating sequence-specific contacts with an 8-bp recognition sequence as typically, three, rather than two, zinc fingers are required for recognition

of around 9 bases, and as such one aim of this thesis was to explore this novel mechanism of DNA binding.

ZNF217 is a putative oncogene. Its aberrant expression has been observed in numerous cancers, such as breast, ovarian, gastric and colon, and is correlated with poor prognosis (Li et al, 2014; Vendrell et al, 2012). Increased ZNF217 activity is associated with aggressive tumour development (Tanner et al, 1995) and it has been shown to immortalise cells such as human mammary epithelial cells (Nonet et al, 2001) and ovarian cells (Li et al, 2007), allowing them to overcome senescence. Patient survival is hence reduced in cases where ZNF217 is over-expressed (Ginestier et al, 2006; Ginzinger et al, 2000), although, as discussed in Chapter 3, less is known about the specific molecular mechanisms by which ZNF217 contributes to tumour progression.

In summary, studying the interaction between DNA and DNA-binding domains of transcription factors is an essential aspect of understanding gene regulation and is critical if one wishes to apply that knowledge to the manipulation of gene expression for human benefit. Following on from this, the next section of the introduction and the second topic of this thesis, discusses the emerging field of cellular reprogramming where transcription factors can be used to alter gene expression and cell fate, potentially in a therapeutic setting.

1.3 Cellular reprogramming by transcription factors

Transcription factors are regulators of gene expression central to the development and function of living cells. They play a key role in specifying normal development

and when over-expressed or ablated, certain transcription factors can induce cell fate changes in somatic cells resulting in cellular reprogramming. Cellular reprogramming, in a broad sense, is the process of stably converting one specialised cell type into a different cell type. The instructive role of transcription factors in lineage specification was first demonstrated in the 1980s when Harold Weintrub and colleagues demonstrated that adult somatic cells could be directly converted to another cell type by forced expression of the muscle-specific transcription factor, MyoD, which induced myotube formation in a fibroblast cell line (Davis et al, 1987). Subsequently, it was found that targeted expression of the *Pax6* gene in Drosophila induced the formation of ectopic eyes (Halder et al, 1995), and similarly, elimination of *Pax5* from mouse B lymphocytes led to their conversion to uncommitted haematopoietic progenitors (Cobaleda et al, 2007; Nutt et al, 1999). This powerful display of alteration of a cell's developmental program by individual transcription factors have refuted the old dogma that cell fate could only transition to progressively more differentiated states. However, it is notable that using single factors only works in a limited number of cases; nevertheless, these early conversions put forward an optimistic expectation that transcription factors could be used to direct lineage commitment.

1.3.1 Classes of transcription factor-mediated reprogramming

Cellular reprogramming by transcription factors can be categorised under two major classes. The first class involves the generation of induced pluripotent stem cells from differentiated somatic cells, which can subsequently be differentiated toward a new lineage. This contrasts with the second class, lineage reprogramming, in which

differentiated cells are converted directly to a different cell type without the induction of a pluripotent intermediary.

1.3.1.1 Induced pluripotent stem cell (iPSC) reprogramming

As the name suggests, induced pluripotent stem cell (iPSC) reprogramming involves the conversion of a somatic cell into a pluripotent stem cell that has the ability to give rise to all cells of the embryo (Figure 1.4A). In one of the most influential experiments in the past decade, Takahashi and Yamanaka paved the way for somatic cell reprogramming using transcription factors. Combinatorial expression of 24 candidate transcription factors, implicated in inducing pluripotency, was tested in mouse fibroblasts (Takahashi & Yamanaka, 2006). They demonstrated, by a process of elimination, that a set of four transcription factors, Oct4, Sox2, Klf4 and c-Myc (OSKM) were sufficient to revert somatic cells back to a pluripotent state, measured by the ability of the resulting cells to differentiate into all three embryonic germ layers: ectoderm, mesoderm and endoderm.

Since this monumental demonstration, the field of iPSC reprogramming has grown exponentially. In addition to mouse fibroblast cells, iPSCs have also been generated from human fibroblasts (Nakagawa et al, 2008; Park et al, 2008; Takahashi et al, 2007) and keratinocytes (Aasen et al, 2008; Maherali et al, 2008). These conversions have been successfully carried out using the same four factors and also different combinations. Moreover, many groups have generated iPSCs from cells of different tissues, including blood (Hanna et al, 2007; Loh et al, 2009), brain (Eminli et al, 2008; Kim et al, 2008), liver (Aoi et al, 2008), pancreas (Stadtfeld et al, 2008), stomach (Aoi et al, 2008), intestine and adrenals (Wernig et al, 2008).

The demonstration that certain transcription factors can reprogram towards pluripotency rekindled interest in the potential to directly convert mature cells to different cell types without reverting back or going through a pluripotent stage.

1.3.1.2 Lineage reprogramming

In contrast to iPSC reprogramming, lineage reprogramming involves cell type conversions that do not pass through or produce a pluripotent state (Figure 1.4). The resulting cells are therefore multipotent tissue stem cells, committed progenitors or terminally differentiated cells. Based on the starting cell and the resulting cell type, lineage reprogramming may be classified as transdifferentiation, transdetermination or dedifferentiation.

Transdifferentiation involves direct conversion between two differentiated and unrelated cell types (Figure 1.4B). There is no stem cell or progenitor cell involvement. The first example of this type of conversion was the aforementioned conversion of fibroblasts to muscle cells by the transcription factor MyoD (Davis et al, 1987). More recently several groups have explored this type of reprogramming ectopically expressing a variety of factors to alter cell fate. To date, fibroblasts have been converted to neurons (Caiazzo et al, 2011; Vierbuchen et al, 2010), hepatocytes (Huang et al, 2011; Sekiya & Suzuki, 2011), cardiomyocytes (Efe et al, 2011; Ieda et al, 2010) and macrophage-like cells (Feng et al, 2008) (Table 1.1).

Similarly, transdetermination is the reprogramming of a committed but not yet fully differentiated cell, such as a stem or progenitor cell, into a closely related cell type in

the same organ system (Figure 1.4C). Much of the research in this area has been performed in the haematopoietic system (Table 1.1) and includes the conversion of committed T-cell progenitors into functional macrophages and myeloid dendritic cells (Laiosa et al, 2006). The former was achieved by the overexpression of the transcription factor required for the formation of granulocyte-macrophage progenitors, C/EBPα, whereas the latter was generated by over-expressing the myeloid transcription factor PU.1. Additionally, committed B-cell progenitors can be reprogrammed to macrophages by enforced expression of both C/EBPα and PU.1 (Xie et al, 2004). Granulocyte-macrophage progenitors committed to the neutrophil/monocyte pathway can be converted to erythroid as well as basophilic and eosinophilic cells by the ectopic expression of the erythroid transcription factor, GATA1 (Heyworth et al, 2002). Furthermore, outside the haematopoietic system, *in vivo* transdetermination in the liver has been achieved by the endocrine transcription factor Ngn3, resulting in the generation of insulin-producing endocrine cells normally produced by the pancreas (Yechoor et al, 2009).

Finally, dedifferentiation is the process where a differentiated cell reverts to a less differentiated state. For example, a terminally differentiated cell can be converted into either a progenitor or a stem cell, and a progenitor back to a multipotent stem cell (Figure 1.4D). It is thought that deletion or knockdown of factors in a mature cell may be particularly important for dedifferentiation. This is because mature cells are believed to possess factors that are dedicated to maintaining the differentiated state. As an example, the ablation of *Pax5* in mature B lymphocytes, which is critical in establishing B cell identity, leads to the cell's conversion into uncommitted haematopoietic progenitors (Cobaleda et al, 2007; Nutt et al, 1999).

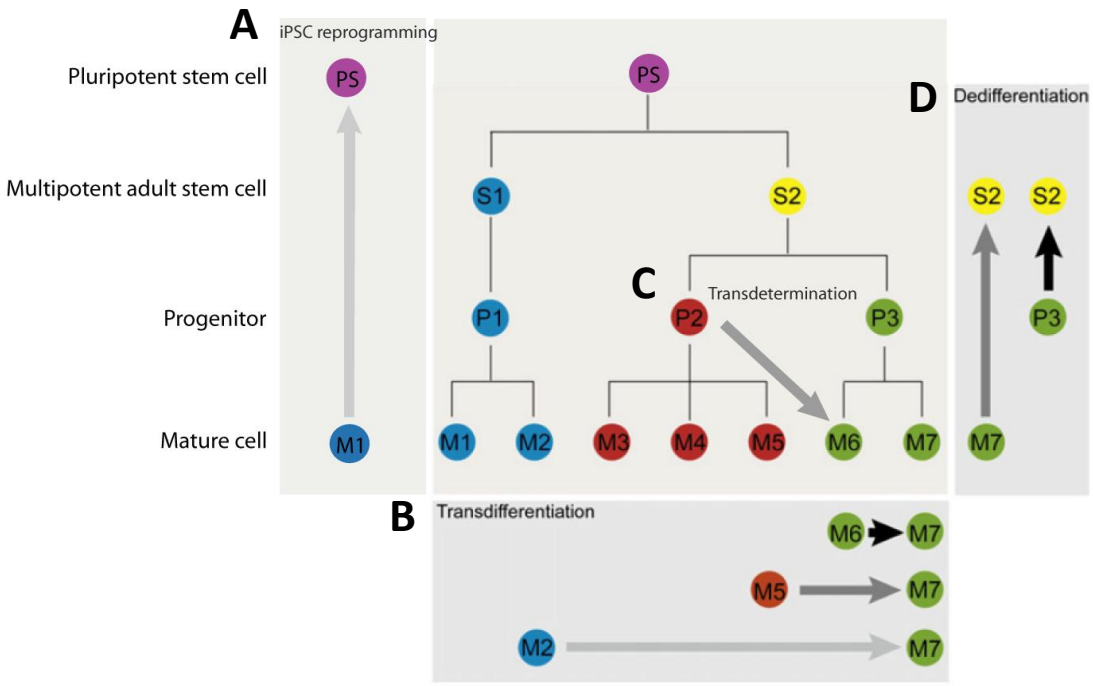


Figure 1.4. *Different routes of cell fate conversions that can be achieved by transcription factor-mediated reprogramming.* The central panel shows a hypothetical cell lineage where the pluripotent stem cell (PS) gives rise to multipotent tissue stem cells (S1, S2) which then produce progenitor cells (P1-P3), and finally different mature cell types (M1-M7). **(A)** In induced pluripotent stem cell (iPSC) reprogramming, a mature cell is converted to pluripotency. A notable example is the conversion of mouse fibroblasts to iPSCs by introduction of four transcription factors (Oct4, Sox2, Klf4 and c-Myc) (Takahashi & Yamanaka, 2006). **(B)** Transdifferentiation is the direct conversion of fate between two differentiated states. Examples include conversion of fibroblasts to neurons (Vierbuchen et al, 2010), hepatocytes (Huang et al, 2011) and cardiomyocytes (Ieda et al, 2010). **(C)** Transdetermination involves the conversion of a progenitor cell from its normal lineage into a closely related lineage. Examples include the conversion of committed mouse B and T cell progenitors into macrophages (Laiosa et al, 2006; Xie et al, 2004). **(D)** Dedifferentiation involves the conversion of a cell to a less differentiated state. This is exemplified in the loss of *Pax5* expression in mature B cells producing uncommitted haematopoietic progenitors (Cobaleda et al, 2007). Each arrow indicates how closely (dark arrow) or distantly (light arrow) related the cell types are. Adapted from (Zhou & Melton, 2008).

Table 1.1. *Transcription factor-mediated lineage reprogramming.* Examples are grouped according to the type of lineage reprogramming. Adapted from (Morris & Daley, 2013).

Initial cell type	Target cell type	Transcription factors	References
Transdifferentiation			
Mouse fibroblasts	Myocytes	MyoD	(Davis et al, 1987)
	Macrophages	C/EBPα/β, PU.1	(Feng et al, 2008)
	Cardiomyocytes	Gata4, Mef2c, Tbx5	(Ieda et al, 2010)
	Glutamatergic neurons	Ascl1, Brn2, Myt1l	(Vierbuchen et al, 2010)
	Dopaminergic neurons	Ascl1, Lmx1a, Nurr1	(Caiazzo et al, 2011)
	Motor neurons	Brn2, Ascl1, Myt1l, Lhx3, Hb9, Isl1, Ngn2	(Son et al, 2011)
	Hepatocytes	Gata4, Hnf1α, Foxa3	(Huang et al, 2011)
		$p19^{ARF}$ knockdown	
	Hepatocytes	Hnf4α, Foxa1/2/3	(Sekiya & Suzuki, 2011)
Human fibroblasts	Glutamatergic neurons	Ascl1, Brn2, Myt1l, NeuroD1	(Pang et al, 2011)
	Glutamatergic neurons GABAergic neurons	Ascl1, Myt1l, NeuroD2, miR-9/9, miR-124	(Yoo et al, 2011)
	Glutamatergic neurons	Brn2, Myt1l, miR-124	(Ambasudhan et al, 2011)
	Neurons	Ascl1, Brn2, Myt1l, Zic1, Olig2	(Qiang et al, 2011)
	Dopaminergic neurons	Ascl1, Brn2, Myt1l, Lmx1a, Foxa2	(Pfisterer et al, 2011)
	Dopaminergic neurons	Ascl1, Lmx1a, Nurr1	(Caiazzo et al, 2011)
	Motor neurons	Brn2, Ascl1, Myt1l, Lhx3, Hb9, Isl1, Ngn2	(Son et al, 2011)
Mouse cortical astrocytes	Glutamatergic neurons	Ngn2	(Heinrich et al, 2010)
	GABAergic neurons	Dlx2	(Berninger et al, 2007)
	GABAergic neurons	Ascl1, Dlx2	(Heinrich et al, 2010)
Mouse hepatocytes	Neurons	Ascl1, Brn2, Myt1l	(Marro et al, 2011)
Mouse pancreatic exocrine cells	Endocrine β-cells	Ngn3, Pdx1, MafA	(Zhou et al, 2008)
Transdetermination			
Mouse pre-B cells	Macrophages	C/EBPα/β	(Xie et al, 2004)
Mouse pre-T cells	Macrophages	C/EBPα/β	(Laiosa et al, 2006)
	Dendritic cells	PU.1	(Laiosa et al, 2006)
Mouse liver stem/progenitor cells	Pancreatic islet cells	Ngn3	(Yechoor et al, 2009)
Dedifferentiation			
Mouse B-cells	Haematopoietic progenitors	Pax5 knockout	(Cobaleda et al, 2007)
Mouse fibroblasts	Neuronal precursor cells	Brn2, Sox2, Foxg1	(Lujan et al, 2012)
Human fibroblasts	Haematopoietic progenitors	Oct4	(Szabo et al, 2010)

1.3.2 Transcription factor-mediated reprogramming: A solution to generating new cells for cell therapy?

From the previous section, it is apparent that ectopic expression of transcription factors can drive cellular reprogramming to redirect cells to a variety of alternative fates. This provides a novel technological platform upon which regenerative medicine can be built. The main goal of regenerative medicine is to produce new cells (cell therapy) that can repair or replace damaged cells, tissues or organs brought about by disease or injury, to restore or establish normal function (Mason & Dunnill, 2008).

The development of iPSC reprogramming, where the instructive factors are known and are capable of creating stem cells from a wide variety of cell types opened up new potential in this field. However, there are several issues surrounding the use of pluripotent stem cells for direct therapeutic applications. Firstly, transplanted iPS cells may give rise to potentially malignant teratomas *in vivo* (Miura et al, 2009; Takahashi & Yamanaka, 2006; Wernig et al, 2007). Furthermore, in cell therapy, iPSC reprogramming must be coupled with directed differentiation *in vitro* to produce lineage directed progenitor or mature cells. Thus, the methods involved in creating new cells from iPSCs are technically demanding and time consuming. Indeed, this second stage of directed differentiation is a critical rate limiting step in iPSC strategy with cell culture protocols largely remaining poorly defined. For example, definitive haematopoietic stem cells are yet to be successfully derived from pluripotent stem cells purely *in vitro* due to the difficulties of establishing culture conditions capable of maintaining and expanding such cells (Cherry & Daley, 2012).

Additionally, the efficiency with which cells can be reprogrammed is exceedingly low with a number of rate limiting stages to be overcome (Amabile & Meissner, 2009; Polo et al, 2010; Xu et al, 2013). Finally, challenges exist in fully purifying the resulting differentiated cells in culture from pluripotent cells, which as mentioned earlier, have the potential to form teratomas (Miura et al, 2009). Taken together, these limitations have prompted researchers to pursue alternative means to achieve cell fate conversions, focusing on direct lineage reprogramming.

Transcription factor-mediated lineage reprogramming provides a new avenue to directly convert abundant and accessible cells, such as fibroblasts, possibly from a patient, to any clinically relevant cell type. The major advantage of this approach over iPSC strategies is that it does not involve a pluripotent state and is hence suitable for direct clinical applications where mature and progenitor cells produced *in vitro* can be transplanted *in vivo* without the risk of forming tumours. This then raises the possibility of converting cells directly *in vivo* for *in situ* regeneration and repair, which would not be feasible in iPSC reprogramming. Indeed, this strategy has been successfully applied in treating a hyperglycaemic mouse model by generating insulin-producing endocrine β cells from pancreatic exocrine cells *in vivo* (Zhou et al, 2008).

There have been some notable successes in reprogramming cells within and toward the haematopoietic lineage (Table 1.1) such as the conversions of B and T lymphocytes to macrophages and dendritic cells (Laiosa et al, 2006; Xie et al, 2004), and the generation of haematopoietic progenitors from fibroblasts (Szabo et al, 2010). However, less work has been reported on directing cell fate towards the

generation of megakaryocytes, cells of considerable clinical importance, which when diseased or deficient result in the bleeding disorder thrombocytopenia.

1.3.3 Thrombocytopenia: a disorder seeking alternative treatments

Thrombocytopenia is a condition that is characterised by a decrease of platelets in blood. Platelets (also called thrombocytes) are small non-nucleated cell fragments derived from megakaryocytes (Italiano & Shivdasani, 2003). Platelets are largely involved in the process leading to the formation of blood clots (Davi & Patrono, 2007). Hence, a decrease in their number, which can either be caused by medical treatments such as chemotherapy or naturally occurring disorders, results in the inability of the blood to clot, leading to excessive bleeding. Morbidity and mortality from bleeding as a result of moderate to severe thrombocytopenia is a major problem that faces a wide range of patients. The magnitude of the problem can be gauged by considering the fact that approximately 1.5 million platelet transfusions are carried out each year to patients in the United States alone (Kaushansky, 2008; Wallace et al, 1995). This current treatment of platelet transfusions, unfortunately, is considered to be less than ideal (Kaushansky, 2008; Kruskall, 1997). Transfusions are expensive and the supply of platelet concentrates has been limited due to its short storage life. Additionally, they are associated with the risks of viral or bacterial infection.

With the limitations associated with current treatments for thrombocytopenia, our idea is to develop *in vitro* culture systems with the potential for a patient's own skin cells to be reprogrammed to produce a large number of megakaryocytes and hence platelets. This would address the issue of limited platelet supply and also avoid potential immune rejection, as the reprogrammed cells would genetically match the

recipient. Given recent advances in the field, lineage reprogramming by transcription factors is a feasible approach to this strategy. In considering this approach, it is first necessary to identify regulatory factors that are potential candidates for reprogramming towards the megakaryocyte lineage. To do this, it is useful to understand the pathways and factors involved in the differentiation of haematopoietic stem cells into mature megakaryocytes.

1.4 Megakaryocytes and platelets

Megakaryocytes (MKs), like all terminally differentiated haematopoietic cells, such as erythrocytes, macrophages, neutrophils, B- and T- cells originate from common haematopoietic stem cells (HSCs) that reside primarily in the bone marrow (Figure 1.5) (Ogawa, 1993). They are the largest (50-100 μm) and also one of the rarest cells in the bone marrow, and account for ~0.01% of nucleated bone marrow cells (Nakeff & Maat, 1974).

The derivation of MKs from HSCs termed megakaryopoiesis, involves successive lineage commitment steps (Figure 1.5). During haematopoiesis, HSCs give rise to the progenitors of two major lineages, the common lymphoid progenitor (CLP) (Kondo et al, 1997) and the common myeloid progenitor (CMP) (Akashi et al, 2000). The CLP then differentiates to lymphocytes such as the B- and T-cells, whereas the CMP gives rise to two progenitors: the granulocyte/macrophage progenitor (GMP) and the megakaryocyte/erythroid progenitor (MEP), from which MK progenitors are derived (Debili et al, 1996).

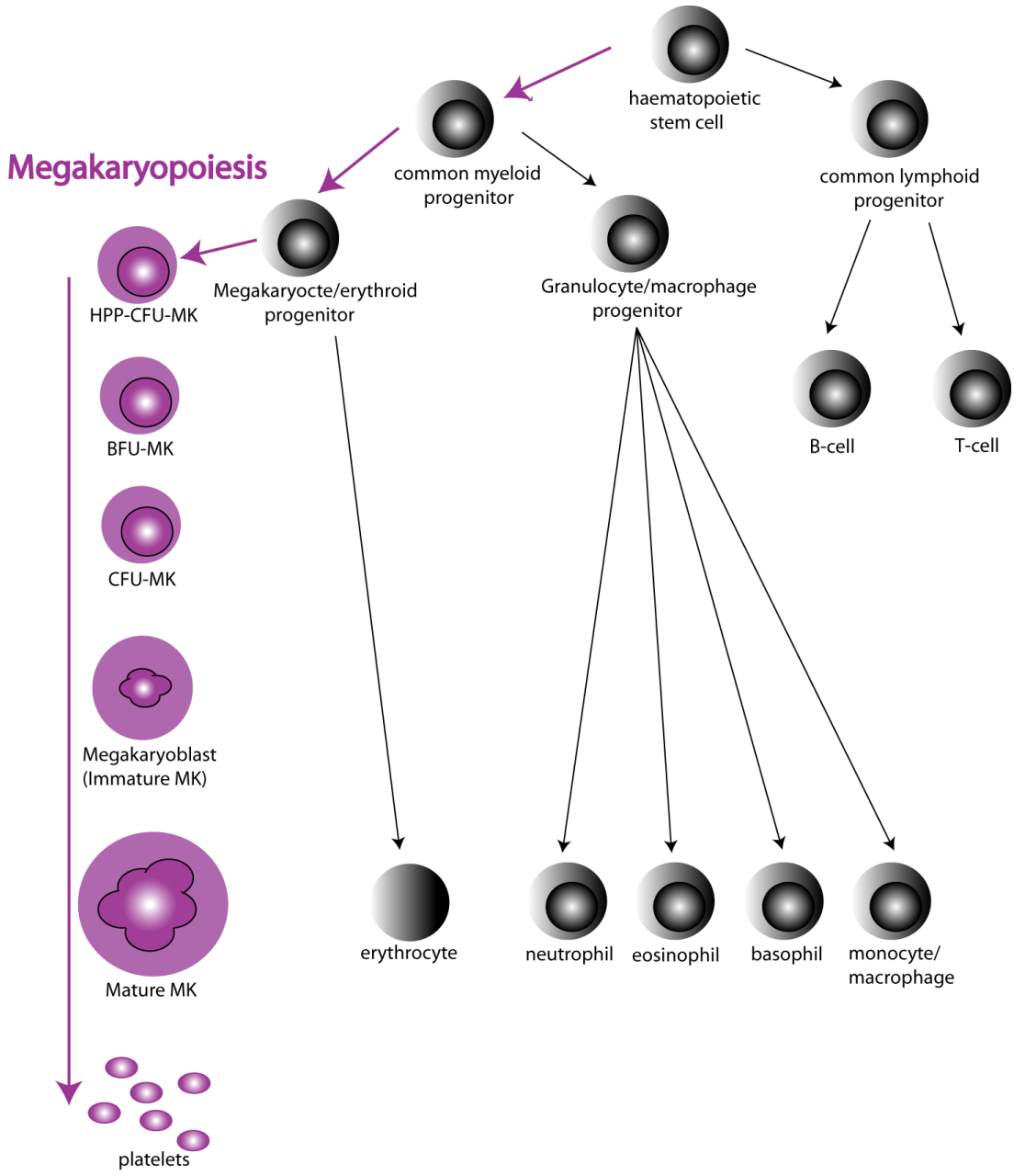


Figure 1.5. *Development of megakaryocytes and platelets from haematopoietic stem cells.* The haematopoietic stem cell has the ability to give rise to two progenitors: myeloid and lymphoid. These progenitors then follow distinct differentiation pathways to generate the mature myeloid and lymphoid cell types depicted. Megakaryopoiesis is one such pathway and is illustrated in more detail. Abbreviations of cell types are as follows: HPP-CFU-MK, high proliferative potential colony-forming unit megakaryocyte; BFU-MK, burst-forming unit megakaryocyte; CFU-MK, colony-forming unit megakaryocyte; MK, megakaryocyte.

The most primitive MK progenitors are the high proliferative potential-colony-forming unit-megakaryocyte (HPP-CFU-MK), followed by the burst-forming unit-megakaryocyte (BFU-MK). The latter produces the most differentiated MK progenitor cell, called colony-forming-unit-megakaryocyte (CFU-MK), which is capable of forming pure populations of MK and platelets *in vitro* and *in vivo*, and is characterised by a unique cell surface phenotype (Nakorn et al, 2003). CFU-MK then gives rise to diploid immature MKs called promegakaryoblasts, which then undergo a process of successive DNA replications without cytoplasmic and cell divisions, termed endomitosis to produce tetraploid immature MKs called megakaryoblasts. Further differentiation of megakaryoblasts results in even larger and more polyploid MKs that can contain up to 64 times the normal amount of DNA. During the transition from megakaryoblasts to mature MKs, the cells expand their cytoplasm, become full of platelet organelles and develop a highly complicated demarcation membrane system (Behnke, 1968; Breton-Gorius & Reyes, 1976; Deutsch & Tomer, 2006). The mature MKs, which are no longer capable of proliferation, now function as reservoirs for platelet formation. They shed their cytoplasm to produce proplatelets which allow the assembly and release of platelets into bone marrow sinusoids (Becker & De Bruyn, 1976; De Botton et al, 2002; Italiano et al, 1999). Platelet formation and release is a terminal process for the mature MKs that results in apoptosis and subsequent phagocytosis by macrophages (Patel et al, 2005; Radley & Haller, 1983; Zauli et al, 1997).

1.4.1 Regulation of megakaryocyte differentiation

MK development and subsequent platelet formation are coordinated by cytokines and transcription factors that act in concert at different stages to regulate these complex processes.

1.4.1.1 Cytokines involved in megakaryopoiesis

There are a number of cytokines that have been associated with the process of megakaryopoiesis. One of them, thrombopoietin (TPO) which is produced in the liver and marrow stroma (Kaushansky, 1995), is considered the major regulator that affects many aspects of growth and development of MKs from HSC precursors (de Sauvage et al, 1994; Lok et al, 1994). It has been shown to be responsible for stimulating the expression of characteristic MK cell surface CD antigens such as CD61, CD41 and CD42, which accompanies the maturation of MKs (Kaushansky et al, 1994). Moreover, it has also been demonstrated to induce endomitosis (Debili et al, 1995; Kaushansky et al, 1994) and to be the only cytokine responsible for the cytoplasmic reorganisation and formation of demarcation membranes, which allows for the release of platelets (Cramer et al, 1997).

In addition to TPO, cytokines such as interleukin (IL)-3, and granulocyte-macrophage colony-stimulating factor (GM-CSF), have also been shown to be important for normal megakaryopoiesis. Both cytokines work in the early stages of MK lineage development to stimulate MK colony formation, with IL-3 acting on bone marrow progenitor cells through the CFU-MK stage (Quesenberry et al, 1985; Robinson et al, 1987) and GM-CSF acting on the BFU-MK population (Aronica et al, 1995). Furthermore, additional cytokines including IL-6, IL-11, IL-12,

erythropoietin (EPO) and leukemia inhibitory factor (LIF) also contribute to MK development by stimulating proliferation of MK progenitors, and modulating MK maturation and platelet release, functioning in concert with TPO and acting synergistically with other cytokines (Gordon & Hoffman, 1992; Kaushansky & Drachman, 2002; Vainchenker et al, 1995).

1.4.1.2 Transcription factors involved in megakaryopoiesis

Several transcription factors and cytokines play essential roles in driving the differentiation of HSCs to MKs. In particular, their interplay is critical in deciding the fate of the bipotential MEP, where the choice between megakaryocytic and erythroid lineages depends on the expression of multiple transcription factors that contribute to lineage specification and differentiation. For the MEP to commit to the megakaryocytic lineage, it requires upregulation of MK-specific transcription factors, downregulation of erythroid-specific transcription factors and ongoing mediation of common erythro-megakaryocytic transcription factors. The factors that regulate the generation of the MK lineage will now be discussed.

The transcription factors TAL1/SCL and GATA2, which are characteristically expressed in multipotent HSCs and early haematopoietic progenitors, have been shown to be critically important in the earliest stages of embryonic haematopoiesis and/or vasculogenesis. Ablation of either *Tal1* or *Gata2* gene in mice causes embryonic lethality due to either a complete absence of blood or severe anaemia, respectively (Robb et al, 1995; Shivdasani et al, 1995a; Tsai et al, 1994; Visvader et al, 1998). Although the embryonic lethality of *Tal1* deletion has limited analyses of its role in haematopoiesis, conditional *Tal1*-knockout mice generated using inducible

Cre/LoxP technology has revealed an absolute requirement for TAL1 in megakaryopoiesis and erythropoiesis (Hall et al, 2003). Colony-formation assays performed on cells obtained from bone marrow and spleen of mice lacking TAL1 demonstrated the inability of the cells to generate early megakaryocytic and erythroid progenitors.

As differentiation occurs progressively down the haematopoietic lineage towards the MEP, GATA2 activates transcription of a closely-related transcription factor, GATA1, which in turn represses GATA2 expression (Grass et al, 2003; Ohneda & Yamamoto, 2002). Being family members, it is believed that GATA1 and GATA2 have both overlapping and unique roles in haematopoietic development (Fujiwara et al, 2004; Weiss & Orkin, 1995). It has been shown that GATA1 together with its co-factor, Friend of Gata1 (FOG1), promotes MK-erythroid differentiation by inhibiting the expression of the myeloid transcription factor PU.1 (Chou et al, 2009; Nerlov et al, 2000). Additionally, forced expression of this factor, results in the expression of erythro-megakaryocytic markers in murine myeloid cells (Visvader et al, 1992; Yamaguchi et al, 1998).

While GATA1 is primarily expressed in erythroid cells with targeted disruption of the gene in mice leading to death at embryonic day 10.5-11.5 due to severe anaemia (Fujiwara et al, 1996), it is also seen in MKs (Lemarchandel et al, 1993). Binding sites for this factor can be found in the enhancers of many MK-specific genes (Eisbacher et al, 2003). In mice, loss of *Gata1* selectively in MKs results in decreased platelet production, causing significant thrombocytopenia and increased numbers of abnormal MKs in the spleen and bone marrow (Shivdasani et al, 1997;

Vyas et al, 1999). The MKs derived from these mice show decreased polyploidisation, having small cytoplasm with disorganised internal membranes and lacking mature platelet granules, suggesting that *Gata1* expression is essential for the development of platelet organelles and cytoplasmic maturation during terminal differentiation.

The accumulation of immature MKs in the absence of GATA1 suggests a role in both terminal differentiation and negative regulation of proliferation. This is supported by the discovery of acquired somatic mutations associated with both transient myeloproliferative disorder and acute megakaryoblastic leukaemia in infants with Down syndrome (Greene et al, 2003; Wechsler et al, 2002). These mutations, which consist of various short deletions and insertions in the first coding exon of *GATA1* cause frame shifts and introduce premature stop codons. These prevent production of full-length GATA1 protein, but allow synthesis of a truncated mutant form, due to translation initiation from a downstream internal methionine. The mutant GATA1 isoform called GATA1 short (GATA1s) lacks part of the transcriptional activation domain but retains the ability to bind DNA and interact with the GATA co-factor FOG1 (Wechsler et al, 2002). GATA1s is believed to act as a dominant oncogene by specifically stimulating the proliferation of fetal MK progenitors (Li et al, 2005). The ability of this GATA1 variant to drive MK progenitor proliferation is significant when one is considering candidate factors for MK lineage reprogramming strategies.

As mentioned earlier, FOG1 is a co-factor of GATA1 with the two factors showing a similar expression pattern in erythroid and MK cells (Tsang et al, 1997). FOG1

interacts indirectly with target genes via association with GATA1 to regulate the early stages of MK differentiation. While targeted disruption of *Fog1* results in strikingly similar defects in erythroid development to those observed in *Gata1*-null mice, unlike *Gata1*-deficient mice, *Fog1*-deficient mice also exhibit a complete failure of megakaryopoiesis (Tsang et al, 1998). The severity of the MK defect in *Fog1*-null mice suggests that most or all critical GATA1-related activity during megakaryopoiesis requires interaction with FOG1. In support of this, the MKs of patients with X-linked thrombocytopenia and variable anaemia, who have GATA1 mutations that impair FOG1 binding, resemble murine *Gata1*-deficient cells (Mehaffey et al, 2001; Nichols et al, 2000; Yu et al, 2002).

Other transcription factors involved in megakaryopoiesis include FLI1 and RUNX1. Both factors have been reported to interact with GATA1 to synergistically activate MK genes (Elagib et al, 2003; Seth et al, 1993; Starck et al, 2003). Additionally, both were also implicated in MK lineage-determination, with FLI1 and RUNX1 over-expression driving haematopoietic cell lines to a megakaryocytic phenotype (Athanasiou et al, 1996; Niitsu et al, 1997). Moreover, FLI1 restricts MEPs to the MK lineage by enhancing GATA1 activity at megakaryocytic promoters while repressing the activity of other erythroid transcription factors at erythroid promoters (Starck et al, 2003). Disruption of either gene in mice is associated with the generation of abnormal, small, hypoploid MKs and associated thrombocytopenia (Hart et al, 2000; Ichikawa et al, 2004; Kawada et al, 2001; Putz et al, 2006; Spyropoulos et al, 2000).

Finally, NFE2 has been shown to be a critical transcription factor in late stage megakaryopoiesis. Although MKs lacking *NFE2* complete endomitosis and expand their cytoplasm, they develop a defective demarcation membrane system that is the source of proplatelet membranes essential for proplatelet formation (Schulze et al, 2006; Shivdasani et al, 1995b). In the absence of NFE2, MKs are unable to form and release platelets, with *Nfe2*-null mice exhibiting severe thrombocytopenia and an accumulation of MKs in the bone marrow (Lecine et al, 1998; Shivdasani & Orkin, 1995; Shivdasani et al, 1995b).

1.5 Aims of this thesis

Transcription factors are instrumental in the regulation of gene expression and a better understanding of how these factors contact and recognise DNA will improve our ability to understand and ultimately artificially control target gene expression. Typically, zinc finger proteins bind DNA through a tandem array of three or more fingers and this mechanism of DNA-binding is well characterised. However, a smaller subset of zinc finger proteins is able to interact with DNA via alternative means and in the first aim of the thesis, we investigate a novel mechanism of DNA-binding by the two-finger domain of the transcription factor ZNF217.

In the second aim of the thesis, the central role of transcription factors in controlling gene expression to direct developmental programs is explored. Specifically, we investigate the potential of candidate transcription factors to reprogram differentiated fibroblast cells toward the megakaryocyte lineage. With the prevalence of the platelet disorder thrombocytopenia and the lack of effective treatments, it is hoped that research in this area will help inform alternative therapeutic strategies.

CHAPTER 2 - MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals and Reagents

The chemicals and reagents that have been used in the experiments described in this thesis are listed below along with their suppliers. All chemicals and reagents used were of molecular biology grade unless alternatively specified.

- 4-(2-Hydroxyethyl)-1-piperazin-ethan-sulfonsäure (HEPES) buffer – Gibco-BRL Life Technologies
- acetic acid – Asia Pacific Specialty Chemicals, Seven Hills, NSW, Australia
- acetone – Ajax Laboratory Chemicals
- acetylthiocholiniodide – Sigma Chemical Company
- acrylamide (electrophoresis grade) – Sigma Chemical Company, St. Louis, MI, USA
- adenosine triphosphate (ATP) – Sigma Chemical Company
- adenosine 5'-[γ-^{32}P] triphosphate ([γ-^{32}P] ATP) – Perkin Elmer Life Sciences, Boston, MA, USA
- agar – Amyl Media, Dandenong, VIC, Australia
- agarose (DNA grade) – Progen Industries, Darra, QLD, Australia
- ammonium persulphate (APS) – Sigma Chemical Company
- ampicillin sodium salt – Progen Industries
- aprotinin – Sigma Chemical Company
- boric acid – Asia Pacific Specialty Chemicals
- bovine serum albumin – Sigma Chemical Company
- 3', 3", 5', 5"-tetrabromophenolsulfonephthalein (bromophenol blue) – Sigma Chemical Company
- casein peptone – Amyl Media
- chloroform – Biolab Scientific, Clayton, VIC, Australia
- Coomassie® brilliant blue R – Sigma Chemical Company

- copper sulphate – Sigma Chemical Company
- deoxynucleotide triphosphates (dNTPs) – Sigma Chemical Company
- D_2O – Sigma Chemical Company
- diethylpyrocarbonate (DEPC) – Sigma Chemical Company
- dimethylsulfoxide (DMSO) – Sigma Chemical Company
- dithiothreitol (DTT) – Sigma Chemical Company
- Dulbecco's Modified Eagle Medium (DMEM) (high glucose) – Gibco-BRL Life Technologies, Grand Island, NY, USA
- ethanol – Ajax Laboratory Chemicals
- ethidium bromide – Roche Molecular Biochemicals, Mannheim, Germany
- ethylenediaminetetraacetic acid (EDTA) disodium dehydrate – Ajax Laboratory Chemicals, Auburn, NSW, Australia
- ethylene glycol-bis[2-aminoethylether]-N,N,N',N'-tetraacetic acid (EGTA) – Sigma Chemical Company
- foetal calf serum (FCS) – Gibco-BRL Life Technologies
- formaldehyde – Sigma Chemical Company
- FuGENE® 6 transfection reagent – Roche Molecular Biochemicals
- GeneRuler™ DNA ladder mix – Progen Industries
- Geneticin® (G418) – Life Technologies
- glutathione-agarose beads – Sigma Chemical Company
- glutathione, reduced (GSH) – Roche Molecular Biochemicals
- glycerol - - Ajax Laboratory Chemicals
- glycine – Ajax Laboratory Chemicals
- Harris Hematoxylin Solution – Sigma Chemical Company
- hydrochloric acid (HCl) – Sigma Chemical Company
- hygromycin B – Life Technologies
- insulin – Sigma Chemical Company
- Iscove's Modified Dulbecco's Medium (IMDM) – Gibco-BRL Life Technologies, Grand Island, NY, USA
- isopropanol – Biolab Scientific, Northcote, New Zealand
- isopropyl-1-thio-β-D-galactopyranoside (IPTG) – Sigma Chemical Company

- JetStar™ 2.0 Plasmid Maxiprep kit – Genomed, Research Triangle Park, NC, USA
- leupeptin – Sigma Chemical Company
- lipoprotein, low density – Sigma Chemical Company
- β-mercaptoethanol – Sigma Chemical Company
- methanol – Ajax Laboratory Chemicals
- 3-[N-Morpholino]propanesulfonic acid (MOPS) – Life Technologies
- nonyl phenoxylpolyethoxylethanol (NP-40) (IGEPAL® CA-630) – Sigma Chemical Company
- penicillin, streptomycin and glutamine solution – Gibco-BRL Life Technologies
- phenol:chloroform:isoamyl alcohol (25:24:1) – Sigma Chemical Company
- phenylmethylsulphonylfluoride (PMSF) – Sigma Chemical Company
- phosphate buffered saline (PBS) tablets – Sigma Chemical Company
- polybrene – Sigma Chemical Company
- poly(dI-dC) – Amersham Pharmacia Biotech, Little Chalfont, Buckinghamshire, UK
- polyoxyethylenesorbitanmonolaurate (Tween™-20)– Sigma Chemical Company
- potassium chloride (KCl) – Sigma Chemical Company
- potassium hexacyanoferrate III – Sigma Chemical Company
- PureLink® HiPure Plasmid Filter Maxiprep Kit – Life Technologies
- puromycin dihydrochloride – Sigma Chemical Company
- Quick Spin G-25 columns for radiolabelled DNA purification – Roche Molecular Biochemicals
- Rainbow™ protein size standards – Amersham Pharmacia Biotech
- recombinant human thrombopoietin (TPO) – Stemcell Technologies and Abacus ALS, Australia
- recombinant mouse Interleukin-3 (IL-3) – Stemcell Technologies
- ribonucleotide triphosphates (rNTPs) – Finnzymes, Oy, Espoo, Finland
- RNase-Free DNase Set – Qiagen, Clifton Hill, VIC, Australia
- RNeasy mini kit – Qiagen
- skim milk powder – No Frills, Chullora, NSW, Australia

- sodium acetate – Ajax Laboratory Chemicals
- sodium azide – Asia Pacific Specialty Chemicals
- sodium chloride – Ajax Laboratory Chemicals
- sodium citrate – Asia Pacific Specialty Chemicals
- sodium dodecyl sulphate (lauryl sulphate sodium salt) (SDS) – Sigma Chemical Company
- sodium phosphate dibasic heptahydrate – Mallinckrodt AR
- SYBR® Green PCR master mix – Applied Biosystems, Foster City, CA, USA
- N,N,N',N'-tetramethylethylenediamine (TEMED) (electrophoresis grade) – Eastern Organic Chemicals, Rochester, NY, USA
- t-octylphenoxypolyethoxyethanol (Triton X-100) – Sigma Chemical Company
- transferrin (iron-saturated) (holo-transferrin) – Sigma Chemical Company
- TRI-REAGENT™ - Sigma Chemical Company
- tris-hydroxymethyl-methylamine (Tris) – Ajax Laboratory Chemicals
- trypan blue – Sigma Chemical Company
- yeast extract – Amyl media
- zinc sulphate – Ajax Laboratory Chemicals

2.1.2 Enzymes

- Antarctic phosphatase – New England Biolabs, MA, USA
- *Pfu* DNA polymerase (deoxynucleoside-triphosphate:DNA deoxynucleotidyl transferase, EC 2.7.7.7) – Stratagene, La Jolla, CA, USA
- Proteinase K – Astral Scientific, Gymea, NSW, Australia
- Pwo DNA polymerase (deoxynucleoside-triphosphate:DNA deoxynucleotidyl transferase, EC 2.7.7.7) – Roche Applied Science, Mannheim, Germany
- Ribonuclease A (RNase A) – Roche Molecular Biochemicals
- T4 DNA ligase – New England Biolabs, MA, USA
- T4 polynucleotide kinase (PNK) – New England Biolabs, MA, USA
- Type II restriction endocucleases (EC 3.1.21) – New England Biolabs, MA, USA

2.1.3 Antibodies

Antibodies used for Western blot and flow cytometry are listed below.

- Anti-GST monoclonal antibody – Sigma Chemical Company
- Anti-GATA1 rat monoclonal antibody, clone N6 – Santa Cruz Biotechnology, Santa Cruz, CA, USA
- Anti-Fli1 rabbit polyclonal antibody, clone C-19 – Santa Cruz Biotechnology, Santa Cruz, CA, USA
- Anti-β-actin mouse monoclonal antibody, clone AC-74 – Sigma Chemical Company
- ECL™ Anti-Rat IgG (NA 9350) – GE Life Sciences, Buckinghamshire, UK
- ECL™ Anti-Rabbit IgG (NA934V) – GE Life Sciences, Buckinghamshire, UK
- ECL™ Anti-Mouse IgG (NA931V) – GE Life Sciences, Buckinghmashire, UK
- PE Rat Anti-Mouse CD41, clone MWReg30 – BD Biosciences
- PE Rat IgG1, κ isotype control, clone R3-34 – BD Biosciences

2.1.4 Oligonucleotides

All oligonucleotides were synthesised by Sigma-Aldrich, Australia. A list of the names and sequences of the oligonucleotides used are contained in the Appendix.

2.1.5 Vectors and Plasmids

2.1.5.1 Vectors

Bacterial Expression Vector

- pGEX-4T-1 – GE Healthcare

Mammalian Retroviral Vectors

- pMSCVpuro – Clontech Laboratories, CA, USA
- pMSCVhyg – Clontech Laboratories, CA, USA
- pMSCVneo – Clontech laboratories, CA, USA
- pMXspuro – provided by Briony Jack

2.1.5.2 Gift plasmids

Bacterial Expression Plasmids

Plasmid	Provided by	Description
pGEX-4T-1-hZNF217 F6-7 (468-525)	Noelia Nunez	For bacterial over-expression of ZNF217 F6-7 region to be used in EMSA
pGEX-4T-1-hZNF217 F6-7 (468-525) C473A	Noelia Nunez	For bacterial over-expression of ZNF217 F6-7 C473A to be used in EMSA
pGEX-4T-1-hZNF217 F6-7 (468-525) Y485A	Noelia Nunez	For bacterial over-expression of ZNF217 F6-7 Y485A to be used in EMSA
pGEX-4T-1-hZNF217 F6-7 (468-525) L486A	Noelia Nunez	For bacterial over-expression of ZNF217 F6-7 L486A to be used in EMSA
pGEX-4T-1-hZNF217 F6-7 (468-525) N487A	Noelia Nunez	For bacterial over-expression of ZNF217 F6-7 N487A to be used in EMSA
pGEX-4T-1-hZNF217 F6-7 (468-525) H489A	Noelia Nunez	For bacterial over-expression of ZNF217 F6-7 H489A to be used in EMSA
pGEX-4T-1-hZNF217 F6-7 (468-525) L490A	Noelia Nunez	For bacterial over-expression of ZNF217 F6-7 L490A to be used in EMSA
pGEX-4T-1-hZNF217 F6-7 (468-525) T492A	Noelia Nunez	For bacterial over-expression of ZNF217 F6-7 T492A to be used in EMSA
pGEX-4T-1-hZNF217 F6-7 (468-525) T494A	Noelia Nunez	For bacterial over-expression of ZNF217 F6-7 T494A to be used in EMSA
pGEX-4T-1-hZNF217 F6-7 (468-525) C504A	Noelia Nunez	For bacterial over-expression of ZNF217 F6-7 C504A to be used in EMSA
pGEX-4T-1-hZNF217 F6-7 (468-525) Q510A	Noelia Nunez	For bacterial over-expression of ZNF217 F6-7 Q510A to be used in EMSA

Lentiviral Packaging Plasmids

- pMD.G (VSV-G) – provided by Prof. Tom Gonda
- pMDLg (Gag-pol) – provided by Prof. Tom Gonda
- pRSVrev – provided by Prof. Tom Gonda

Mammalian Retroviral and Lentiviral Plasmids

Plasmid	Provided by	Description
pMXspuro-mFog1	Briony Jack	A retroviral expression plasmid encoding murine Fog1.
pLV411	Prof. Tom Gonda	A lentiviral expression vector used as an empty vector control for over-expression studies. It contains a CMV promoter and an IRES-driven GFP.
pLV411-p45NFE2	Prof. Tom Gonda	A lentiviral expression plasmid encoding human p45 subunit of NFE2.
pLV411-MAFG	Prof. Tom Gonda	A lentiviral expression plasmid encoding human p18 subunit MAFG.
pLV411-MEF2C	Prof. Tom Gonda	A lentiviral expression plasmid encoding human MEF2C.
pLV411-FOG2	Prof. Tom Gonda	A lentiviral expression plasmid encoding human p45 subunit of FOG2
pLV411-ETS1	Prof. Tom Gonda	A lentiviral expression plasmid encoding human ETS1.
pLV411-GATA2	Prof. Tom Gonda	A lentiviral expression plasmid encoding human GATA2.
pLV411-OCT4	Prof. Tom Gonda	A lentiviral expression plasmid encoding human OCT4.
pLV411-FHL2	Prof. Tom Gonda	A lentiviral expression plasmid encoding human FHL2.
pLV411-RFX5	Prof. Tom Gonda	A lentiviral expression plasmid encoding human RFX5.
pLV411-MXD1	Prof. Tom Gonda	A lentiviral expression plasmid encoding human MXD1.
pLV411-E2F3	Prof. Tom Gonda	A lentiviral expression plasmid encoding human E2F3.
pLV411-c-Mpl	Prof. Tom Gonda	A lentiviral expression plasmid encoding human c-Mpl.

2.1.5.3 Plasmids

Bacterial Expression Plasmids

Plasmid	Oligos	Restriction Sites	Description
pGEX-4T-1-hZNF217 F6-7 (468-525) S474A	A3829/ A3830	EcoRI and XhoI	For bacterial over-expression of ZNF217 F6-7 mutants to be used in EMSA. Mutations were introduced by overlap PCR site-directed mutagenesis using the vector primers: A3928 and A3892, and the indicated internal primers containing the mutation.

pGEX-4T-1-hZNF217 F6-7 (468-525) Y475A	A3831/ A3832	EcoRI and XhoI	For bacterial over-expression of ZNF217 F6-7 mutants to be used in EMSA. Mutations were introduced by overlap PCR site-directed mutagenesis using the vector primers: A3928 and A3892, and the indicated internal primers containing the mutation.
pGEX-4T-1-hZNF217 F6-7 (468-525) G477A	A3833/ A3834	EcoRI and XhoI	
pGEX-4T-1-hZNF217 F6-7 (468-525) K478A	A3835/ A3836	EcoRI and XhoI	
pGEX-4T-1-hZNF217 F6-7 (468-525) F479A	A3837/ A3838	EcoRI and XhoI	
pGEX-4T-1-hZNF217 F6-7 (468-525) R481A	A3839/ A3840	EcoRI and XhoI	
pGEX-4T-1-hZNF217 F6-7 (468-525) S482A	A3841/ A3842	EcoRI and XhoI	
pGEX-4T-1-hZNF217 F6-7 (468-525) Y484A	A3845/ A3846	EcoRI and XhoI	
pGEX-4T-1-hZNF217 F6-7 (468-525) I488A	A3847/ A3848	EcoRI and XhoI	
pGEX-4T-1-hZNF217 F6-7 (468-525) R491A	A3849/ A3850	EcoRI and XhoI	
pGEX-4T-1-hZNF217 F6-7 (468-525) G495A	A3851/ A3852	EcoRI and XhoI	For bacterial over-expression of ZNF217 F6-7 mutants to be used in EMSA. Mutations were introduced by one-step site-directed mutagenesis using the indicated internal primers.
pGEX-4T-1-hZNF217 F6-7 (468-525) E496A	A3853/ A3854	EcoRI and XhoI	
pGEX-4T-1-hZNF217 F6-7 (468-525) K497A	A3855/ A3856	EcoRI and XhoI	For bacterial over-expression of ZNF217 F6-7 mutants to be used in EMSA. Mutations were introduced by overlap PCR site-directed mutagenesis using the vector primers: A3928 and A3892, and the indicated internal primers containing the mutation.
pGEX-4T-1-hZNF217 F6-7 (468-525) P498A	A3857/ A3858	EcoRI and XhoI	
pGEX-4T-1-hZNF217 F6-7 (468-525) Y499A	A3859/ A3860	EcoRI and XhoI	
pGEX-4T-1-hZNF217 F6-7 (468-525) K500A	A3861/ A3862	EcoRI and XhoI	
pGEX-4T-1-hZNF217 F6-7 (468-525) E502A	A3863/ A3864	EcoRI and XhoI	
pGEX-4T-1-hZNF217 F6-7 (468-525) F503A	A3865/ A3866	EcoRI and XhoI	

pGEX-4T-1-hZNF217 F6-7 (468-525) E505A	A3867/ A3868	EcoRI and XhoI	For bacterial over-expression of ZNF217 F6-7 mutants to be used in EMSA. Mutations were introduced by overlap PCR site-directed mutagenesis using the vector primers: A3928 and A3892, and the indicated internal primers containing the mutation.
pGEX-4T-1-hZNF217 F6-7 (468-525) Y506A	A3869/ A3870	EcoRI and XhoI	For bacterial over-expression of ZNF217 F6-7 mutants to be used in EMSA. Mutations were introduced by one-step site-directed mutagenesis using the indicated internal primers.
pGEX-4T-1-hZNF217 F6-7 (468-525) A507Q	A3871/ A3872	EcoRI and XhoI	For bacterial over-expression of ZNF217 F6-7 mutants to be used in EMSA. Mutations were introduced by overlap PCR site-directed mutagenesis using the vector primers: A3928 and A3892, and the indicated internal primers containing the mutation.
pGEX-4T-1-hZNF217 F6-7 (468-525) A509Q	A3873/ A3874	EcoRI and XhoI	
pGEX-4T-1-hZNF217 F6-7 (468-525) K511A	A3875/ A3876	EcoRI and XhoI	For bacterial over-expression of ZNF217 F6-7 mutants to be used in EMSA. Mutations were introduced by one-step site-directed mutagenesis using the indicated internal primers.
pGEX-4T-1-hZNF217 F6-7 (468-525) T512A	A3877/ A3878	EcoRI and XhoI	For bacterial over-expression of ZNF217 F6-7 mutants to be used in EMSA. Mutations were introduced by overlap PCR site-directed mutagenesis using the vector primers: A3928 and A3892, and the indicated internal primers containing the mutation.
pGEX-4T-1-hZNF217 F6-7 (468-525) S513A	A3879/ A3880	EcoRI and XhoI	For bacterial over-expression of ZNF217 F6-7 mutants to be used in EMSA. Mutations were introduced by one-step site-directed mutagenesis using the indicated internal primers.

pGEX-4T-1-hZNF217 F6-7 (468-525) Y516A	A3883/ A3884	EcoRI and XhoI	For bacterial over-expression of ZNF217 F6-7 mutants to be used in EMSA. Mutations were introduced by overlap PCR site-directed mutagenesis using the vector primers: A3928 and A3892, and the indicated internal primers containing the mutation.
pGEX-4T-1-hZNF217 F6-7 (468-525) L518A	A3885/ A3886	EcoRI and XhoI	
pGEX-4T-1-hZNF217 F6-7 (468-525) E519A	A3887/ A3888	EcoRI and XhoI	For bacterial over-expression of ZNF217 F6-7 mutants to be used in EMSA. Mutations were introduced by one-step site-directed mutagenesis using the indicated internal primers.
pGEX-4T-1-hZNF217 F6-7 (468-525) R520A	A3889/ A3890	EcoRI and XhoI	For bacterial over-expression of ZNF217 F6-7 mutants to be used in EMSA. Mutations were introduced by overlap PCR site-directed mutagenesis using the vector primers: A3928 and A3892, and the indicated internal primers containing the mutation.

Mammalian Retroviral and Lentiviral Plasmids

Plasmid	Oligos	Restriction Sites	Description
pMSCVhyg-Gata1	A3692/ A3693	BglII and XhoI	A retroviral vector encoding murine Gata1.
pMSCVhyg-Gata1s	A4288/ A3693	BglII and XhoI	A retroviral vector encoding murine Gata1 short (Gata1s).
pMSCVpuro-Fli1	A3690/ A3691	BglII and XhoI	A retroviral vector encoding murine Fli1.
pMSCVpuro-Ets1	A3702/ A3704	BglII	A retroviral vector encoding murine Ets1.
pMSCVneo-Tal1	A4291/ A4292	EcoRI and XhoI	A retroviral vector encoding murine Tal1.
pLV411-Runx1	A4213/ A4214	-	A lentiviral vector encoding murine Runx1.

pLV411-Meis1	A4209/ A4210	-	A lentiviral vector encoding murine Meis1.
pLV411-Fog1	A4207/ A4208	-	A lentiviral vector encoding murine Fog1.
pLV411-Gabpα	A4217/ A4218	-	A lentiviral vector encoding murine Gabpα.
pLV411-Tal1	A4211/ A4212	-	A lentiviral vector encoding murine Tal1.

2.1.6 Bacterial strains and culture media

The bacterial strain used for cloning, miniprep and maxiprep plasmid isolation was *Escherichia coli* (*E. coli*) DH5α (*supE44, ΔlacU169* [ϕ80*lac* ZΔM15], *hsdR17, recA1, endA1, gyrA96, thi-1, relA1*) (Bethesda Research Laboratories, Gaithersburg, MD, USA). The bacterial strain used for the over-expression of GST-fusion proteins was *E. coli* BL21 (genotype: B F-*dcm ompT hsd*S (r_{B-} m_{B-}) *gal*) (Stratagene).

Both bacterial strains were cultured in Luria-Bertani (LB) broth (or agar):

10 g/L casein peptone
5 g/L yeast extract
10 g/L sodium chloride
(15 g/L bacteriological agar)

LB was made up with MilliQ Water (MQW) and was sterilised by autoclaving. Filter-sterilised ampicillin (50 mg/mL in MQW) was added to cooled, autoclaved broth to a final concentration of 50 µg/mL. In the preparation of agar plates, antibiotics were added to agar containing media immediately prior to pouring. Broth and plates were stored at 4°C until use.

2.2 Methods

2.2.1 General methods

Routine molecular biological techniques were carried out as outlined in Sambrook *et al.* (Sambrook et al, 1989). Page references for each technique are indicated.

- restriction endonuclease digestion of DNA: 5.24-5.32
- agarose gel electrophoresis: 6.1-6.20
- polyacrylamide gel electrophoresis: 6.36-6.43, 6.45, 18.47-18.55
- agarose gel DNA purification: 6.22-6.23
- DNA ligation: 1.63-1.69
- transformation of competent bacterial cells: 1.74, 1.76, 1.86
- phenol/chloroform extraction of DNA: E.3-E.4
- ethanol precipitation of DNA/RNA: E.10-E.14
- mini-preparations of plasmid DNA: 1.21-1.31
- polymerase chain reaction (PCR): 14.1-14.4, 14.14-14.21
- nuclear extracts from cultured cells: 17.8-17.10
- Western blots: 18.60-18.61, 18.64-18.66, 18.69-18.74
- electrophoretic mobility shift assays (EMSAs): 17.13-17.17

2.2.2 Commercial services and kits

DNA sequencing reactions were conducted by the Australian Genome Research Facility (AGRF) Sydney or Brisbane nodes.

Techniques that involved the use of a commercial kit were carried out as advised in the manufacturers' protocols. A list of commercial kits used is displayed below.

- Chamber slides (Stemcell Technologies)
- DNA-*free*™ (Ambion, Austin, TX, USA) for DNase-treatment of RNA
- Immobilon Western HRP and AP Chemiluminescent HRP Substrate (ECL) (Merck Millipore, Billerica, MA, USA) for Western blot visualisation

- JetStar™ 2.0 Plasmid Maxiprep kit (Genomed, Research Triangle Park, NC, USA) for large-scale plasmid DNA purification
- PCR Cloning System with Gateway® Technology with pDONR™221 and One Shot® OmniMAX™ 2-T1^{R} Chemically Competent *E.coli* (Life Technologies) for the generation of Gateway ® entry clones from PCR products and donor vector
- LR Clonase™ II Enzyme Mix (Life Technologies) for the generation of expression plasmids from Gateway® entry clones and destination vector
- MegaCult®-C Collagen and medium without cytokines (Stemcell Technologies) for culturing cells in a collagen-based medium
- PCR Cloning System with Gateway® Technology (Life Technologies) for the generation of entry clones from Gateway® donor vector and PCR product
- PureLink® HiPure Plasmid Filter Maxiprep Kit (Life Technologies) for large-scale plasmid DNA purification
- Quantikine® ELISA Mouse CXCL4/PF4 Immunoassay kit (R&D Systems, Minneapolis, MN, USA) for PF4 ELISA
- RNeasy mini kit (Qiagen, Clifton Hill, VIC, Australia) for purification of RNA samples
- SuperScript® VILO™ cDNA Synthesis Kit (Life Technologies)
- Wizard® SV Gel and PCR Clean-Up System (Promega)

2.2.3 Equipment

- Branson Digital Sonifier® S-450D (Branson Sonic Power, Danbury, CO, USA) was used for lysis of bacterial cells
- Sephadex® G-25 Quick Spin™ columns (Roche Molecular Biochemicals) were used to purify radiolabeled EMSA probes
- Sturdier™ Vertical Slab Gel Unit (Hoefer Scientific Instruments) was used to run EMSA gels
- Typhoon™ FLA 9000 biomolecular imager (GE Healthcare Life Sciences) was used for the visualisation of radioactive EMSA gels
- ImageQuant™ (Version 3.3) software (Molecular Dynamics) was used to analyse the Typhoon™ FLA 9000-generated files

- Machine for 1D NMR - 600 MHz Bruker DRX-600 spectrometer
- XCell *SureLock*™ Mini-Cell (Life Technologies) was used for NuPAGE® Novex® SDS-PAGE
- XCell II™ Blot Module (Life Technologies) was used for Western blotting
- Biotrace™ nitrocellulose membranes (Pall gelman Sciences, Ann Arbor, MI, USA) were used for Western blotting
- Fujifilm Luminescent Image Analyzer LAS-3000 (Fujifilm Medical Systems, Stamford, CT, USA)
- Primer3™ Version 0.4.0 (Humana Press, Totowa, NJ, USA) was used to design real time RT-PCR primers
- Applied Biosystems® 7500 Fast Real-Time PCR System (Life Technologies) was used to perform real time RT-PCR runs
- Applied Biosystems® 7500 Fast Real-Time PCR System Software version 2.0.4 (Life Technologies) was used for real time RT-PCR data analysis
- BD LSRFortessa™ SORP cell analyser (BD Biosciences, San Jose, USA) was used for flow cytometry analysis
- BD FACS Jazz™ Cell Sorter (BD Biosciences, Cytopeia, USA) was used to sort for GFP positive cells
- FlowJo vX.0.6 (Tree Star, Inc., Ashland, OR, USA) was used for flow cytometry analysis
- Partek® Genomics Suite™ 6.6, Copyright© 2013 (Partek Inc., St. Louis, MO, USA) was used to analyse microarray data
- Olympus FSX100™ Fluorescence Microscope (Olympus) was used to perform phase contrast, GFP fluorescence and bright field imaging of cells
- Eppendorf Mastercycler® (Eppendorf AG, Barkhausenweg, Hamburg, Germany) thermal cycler was used for all PCRs
- Countess® Automated Cell Counter (Life Technologies) was used to perform cell counts
- Thermo Scientific NanoDrop™ 1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA) was used to quantify RNA and DNA

2.2.4 PCR, one-step and overlap PCR site-directed mutagenesis

Background on PCR and the protocol used for PCR amplification are seen in sections 14.1-14.4, 14.14-14.21 of 'Molecular Cloning. A Laboratory Manual' (Sambrook et al, 1989). PCRs were conducted with *Pfu* DNA polymerase, unless stated otherwise, with conditions and cycle parameters as recommended by the supplier.

One-step and overlap PCR site-directed mutagenesis were used to introduce site-directed mutations into cDNA sequences. Two complementary internal primers were synthesised to introduce each mutation. These internal primers contained the desired mutation but were complementary to the cDNA sequence for at least 11 nucleotides 5' and 3' of the site of the desired mutation. In the case of overlap PCR, two end primers were also designed which were complementary to at least 24 nucleotides of the vector sequence. These vector primers were at least 500 nucleotides 5' and 3' of the N- and C-termini of the cDNA nucleotide sequence. With respect to one-step PCR, a high fidelity DNA polymerase, Pwo, is used. The PCR products generated are complete, unmethylated plasmids with the introduced mutation. The residual template plasmid is then digested with *Dpn*I.

2.2.5 General cloning

PCR products were purified with Wizard® SV Gel and PCR Clean-Up System (Promega) and were subjected to appropriate restriction digests. Vectors were similarly digested and were subsequently treated with Antarctic phosphatise for 1 h to prevent self-ligation. Digested PCR products and vector were purified by agarose gel electrophoresis, excised, subsequently extracted from gel slabs with Wizard® SV

Gel and PCR Clean-Up System (Promega) and allowed to ligate overnight at 4°C or for 2 h at room temperature. Ligation products were then used to transform competent *E. coli*. Briefly, 100 µL competent cell suspension was added to each ligation mixture and incubated at 4°C for 10 min. Samples were then heat-shocked at 42°C for 2 min. 200 µL prewarmed LB broth was added and samples were incubated at 37°C for 20-30 min. Samples were then spread onto prewarmed LB-agar plates (containing the relevant antibiotic) which were incubated overnight at 37°C. Transformants were assayed for uptake of correct constructs by mini-preparation of plasmid DNA followed by checking digests, agarose gel electrophoresis and sequencing. For large-scale DNA preparation, the JetStar™ 2.0 Plasmid Maxiprep kit or PureLink® HiPure Filter Maxiprep kit was used.

2.2.6 Bacterial over-expression and purification of GST and GST-fusion proteins

E. coli BL21 cells were transformed with plasmids, and transformants were grown on LB supplemented with ampicillin (100 µg/mL) (LB + amp) for 18 h (overnight) at 37°C. Overnight culture (15 mL) was used to inoculate fresh LB + amp media (300 mL), which was incubated for a further 2-4 h at 37°C at 180 rpm or until the culture reached an absorbance of 0.6 at 600 nm. Protein over-expression was then induced by the addition of 0.1 mM IPTG and 0.01 mM $ZnSO_4$ and incubated overnight at 20°C with shaking at 180 rpm. Cells were harvested by centrifugation (15 min, 17,000 x g, 4°C).

Cell pellets were resuspended in 5-10 mL ice-cold lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5% (w/v) NP-40; supplemented prior to use with 1%

(v/v) Triton X-100, 1 mM DTT, 10 mM $MgCl_2$, 0.2 mM PMSF, 5 µg/mL leupeptin and 5 µg/mL aprotinin) and sonicated (8 x 15 s pulses at 25% intensity). Cell debris was removed by centrifugation (15 min, 12,000 x g, 4°C). The soluble fraction containing the desired protein was incubated with pre-swollen glutathione-agarose beads (800 µL of settled beads per L of culture) overnight at 4°C with gentle rotation, allowing GST binding with glutathione. The beads were then washed extensively with cold lysis buffer, and bound GST-fusion protein was eluted by incubation with reduced glutathione solution (20 mM reduced glutathione, 100 mM Tris pH 7.5, 120 mM NaCl) at 4°C for 30 min with gentle rotation. After elution, the protein was quantified by obtaining the absorbance at 280 nm. Eluted protein was snap frozen and stored at -80°C until required.

2.2.7 Polyacrylamide gel electrophoresis of GST and GST-fusion proteins

Purified GST and GST-fusion proteins (from pGEX-4T-1 constructs) were subjected to mini SDS/polyacrylamide gel electrophoresis (SDS/PAGE). Samples were boiled with an equal volume of SDS loading buffer (125 mM Tris pH 6.8, 4% SDS, 20% glycerol, 40 µg/mL bromophenol blue, [100 mM DTT, added just prior to use]) for 5 min at 95°C, spun down briefly and the whole volume was loaded on 10% NuPAGE® Novex® Bis-Tris gels. Rainbow™ protein standards were also loaded. Gels were subjected to electrophoresis in 1 x MOPS SDS running buffer (Life Technologies) for 55 min at 200 V. The gels were then stained with Coomassie® brilliant blue R overnight and destained.

2.2.8 Electrophoretic mobility shift assay (EMSA)

Single-stranded oligonucleotide was 5' end-labelled with [γ-^{32}P]-ATP using T4 polynucleotide kinase using provided 10 x kinase buffer. To generate double-stranded DNA probes, complementary oligonucleotides were annealed in TNE buffer (50 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA). Radiolabelled DNA was then purified by centrifugation (4,000 x g, 4 min) through Sephadex® G25 Quick Spin™ columns. Sequences of oligonucleotides used in the production of radiolabelled probes are displayed in the Appendix.

EMSA reactions were set up in a total volume of 30 µL, comprising 0.2 pmol of ^{32}P-labelled probe (final concentration 6 nM), 1 µg GST or recombinant GST-fusion protein and gel shift buffer (20 mM HEPES pH 7.9, 100 mM KCl, 0.2 mM EDTA, 0.2 mM EGTA, 0.1% NP-40, 5% glycerol), 0.5 mM PMSF, 0.5 mM DTT, 1 µg BSA, 0.5 µg poly(dI-dC) and 1 mM $ZnSO_4$. Reactions were incubated on ice for 20 min and then loaded onto a pre-cooled 6% native polyacrylamide gel (made with 40% acrylamide [19:1, acrylamide: bisacrylamide], 45 mM Tris, 45 mM boric acid, 1 mM EDTA, 0.1% APS and 0.1% TEMED). Gels were cast and run in a Sturdier™ vertical slab gel unit (Hoefer Scientific Instruments) and subjected to electrophoresis at 250 V for 1 h 45 min at 4°C in 0.5 x TBE buffer (45 mM Tris, 45 mM boric acid, 1mM EDTA). Gels were dried and scanned on a Typhoon™ FLA 9000 biomolecular imager (GE Healthcare Life Sciences) and the data analysed using ImageQuant™ (Version 3.3) software (Molecular Dynamics).

2.2.9 One dimensional (1D) NMR spectroscopy

Purified protein samples were concentrated with the assistance from Marylène Vandevenne using Centricon YM-3 devices (Santorius Stedim Biotech, Dandenong, VIC, Australia) with an appropriate molecular weight cut-off filter. Samples contained 1 mM DTT, 100 mM Tris pH 7.5, 120 mM NaCl and 50 mM reduced glutathione, to which 5% (v/v) D_2O and 2 µM 2,2- Dimethyl-2-silapentane-5-sulfonate pH 6.0 was added. 1D NMR spectroscopy and analysis were performed with assistance from collaborators, Joel Mackay and Marylène Vandevenne. Spectra of samples were recorded at 5°C on a 600 MHz Bruker DRX-600 spectrometer equipped with a 5 mm BBI probe head. Spectra were processed using TOPSPIN (Bruker) and analysed using Sparky (Goddard TD) running on Linux workstations. 1D NMR spectra were referenced to DSS at 0.00 ppm.

2.2.10 Mammalian cell culture

Murine embryonic fibroblasts (MEFs), Ecopack™ 2-293 retroviral packaging cells (Clontech) and HEK-293FT cells were cultured at 37°C and 5% CO_2 in a standard medium (HG DMEM, supplemented with 10% heat-inactivated FCS, 1% penicillin/streptomycin/glutamine solution). MEFs were maintained under selection in 2.5 µg/mL puromycin, 500 µg/mL hygromycin and/or 400 µg/mL Geneticin® (G418), where appropriate. Additionally, after transduction, MEFs were cultured in a standard medium supplemented with 50 ng/mL recombinant human TPO.

2.2.11 Generation of retroviral and lentiviral vectors

Retroviral vectors were created by inserting the coding sequence of murine *Gata1* and *Gata1s* into the multiple cloning site (MCS) of pMSCVhyg (Clontech

Laboratories, CA, USA), whereas the open reading frame (ORF) of murine *Fli1* and *Ets1* were inserted into the MCS of pMSCVpuro (Clontech). The full length cDNA of *Tal1* was inserted into the MCS of the retroviral pMSCVneo vector (Clontech). PCR and cloning were carried out as described in Sections 2.2.4 and 2.2.5.

Lentiviral vectors containing the coding sequence for murine *Runx1, Meis1, Fog1, Gabpα* and *Tal1* were created by the use of the Gateway® Technology with Clonase II kit which consists of the PCR Cloning System with Gateway® Technology with pDONR™221 and One Shot® OmniMAX™ 2-T1^{R} Chemically Competent *E.coli* kit, and LR Clonase™ II Enzyme Mix kit (Life Technologies). Primer design and cloning were performed according to the manufacturer's instructions.

2.2.12 Retroviral and lentiviral transductions

To generate retroviruses, Ecopack™ 2-293 cells were transfected with 5 µg of plasmid DNA (pMSCV plasmids) and to generate lentiviruses, HEK-293FT cells were transfected with 2.5 µg plasmid DNA (pLV411 plasmids) mixed with packaging plasmids pMD.G (VSV-G) (742 ng), pMDLg (Gag-pol) (1.1 µg) and pRSVrev (462 ng) in a total volume of 10 µL, using FuGENE6 transfection reagent (Roche) following the manufacturer's instructions. Twelve hours prior to infection, MEF cells were seeded at $1.5x10^5$ cells per 60 mm dish. At 48 hours after transfection, the media containing the virus particles was harvested from the Ecopack™ 2-293 or HEK-293FT cells, filtered (pore size 45 µm) and transferred to the MEF cells, where the media was removed prior to infection. The cells were incubated with 8 µg/mL Polybrene (Sigma) for 24 hours. Transduced cells were selected with the appropriate antibiotics: puromycin (2.5 µg/mL), hygromycin (500

µg/mL) and Geneticin® (400 µg/mL). Alternatively, to induce differentiation into MK lineage according to Matsubara's group, see Section 2.2.19 (Ono et al, 2012).

2.2.13 RNA extraction and cDNA synthesis

Cell samples were washed with PBS prior to RNA extraction. Total RNA was extracted with TRI REAGENT™ as per the supplier's protocol but with an additional centrifuge step at 12,000 g for 10 min at 4°C following homogenisation to reduce possible genomic DNA contamination. RNA was cleaned by the use of RNeasy mini kit (Qiagen) and was subsequently DNase-treated with DNA-*free*™ kit (Ambion) following the manufacturer's instructions for the 'rigorous' treatment. The purified RNA was used to synthesis cDNA by use of the SuperScript® VILO™ cDNA Synthesis Kit (Invitrogen) following the manufacturer's instructions. cDNA negative controls (-RT) were made using the addition of water in place of the reverse transcriptase.

2.2.14 Real time RT-PCR

Quantitative real time PCR reactions (final volume 20 µL) were set up in duplicate or triplicate with approximately 10 ng of cDNA, Power SYBR® Green PCR Master Mix (Life Technologies) and forward and reverse oligonucleotides primers (400 nM). Real time RT-PCR thermal cycling was performed using the Applied Biosystems 7500 Fast Real-Time PCR System (Life Technologies) using the machine's default settings. Results were analysed using the 7500 Software v2.0.4 (Life Technologies). Expression levels for genes of interest were normalised to *18S* rRNA levels. As negative controls, duplicate –RT and no template reactions were included.

2.2.15 Real time RT-PCR primers

Primer3™ Version 0.4.0 was used to design paired real time PCR primers. Primer pairs were designed to cross exon-exon junctions where possible to prevent amplification of any contaminating genomic DNA. Specificity of primers was verified by conducting genomic sequence searches using the Basic Local Alignment Search Tool (BLAST; http://blast.ncbi.nlm.nih.gov/Blast). A list of all real time RT-PCR primers can be found in the Appendix.

2.2.16 Western blotting

Western blotting of nuclear extracts was performed as described previously (Sambrook et al, 1989). Protein concentration was estimated by UV spectrophotometry at 280 nm using a NanoDrop (Thermo Fisher Scientific) and equivalent amounts of samples were loaded. Samples were run on NuPAGE® Novex 10% Bis-Tris Gels (Life Technologies) at 200 V for 55 min using XCell *SureLock* (Life Technologies) as per manufacturer's instructions. Proteins were transferred to nitrocellulose membrane using the XCell™ Blot Module (Life Technologies). Membranes were blocked in TBST (50 mM Tris pH 7.4, 150 mM NaCl, 0.05% Tween-20) with 3% (w/v) skim milk powder for 2 x 15 min, and were then washed with TBST for 3 x 15 min prior to primary antibody binding. To detect Gata1, the membrane was probed with primary GATA1 antibody clone N6 (1 in 200 dilution in 3% skim milk in TBST) overnight at 4°C on a shaker and was then washed with TBST for 5 x 15 min, followed by 1 hour incubation in a secondary anti-Rat IgG diluted 1 in 30,000 in TBST. For Fli1 detection, the membrane was incubated with primary Fli1 antibody clone C-19 (1 in 200 dilution in 3% skim milk in TBST) overnight at 4°C with gentle shaking. The membrane was then washed 5 x 15 min

with TBST and a secondary anti-Rabbit IgG diluted 1 in 30,000 in TBST was applied and incubated for 1 hour. After the secondary antibody incubation, the membranes were washed again in TBST 5 x 5 min. Antibodies were detected using Immobilon Western HRP and AP Chemiluminescent HRP Substrate (Millipore) and chemiluminescent bands were detected with Fujifilm Luminescent Image Analyzer LAS-3000. Rainbow™ protein size standards were loaded in each gel for size estimation. Membranes were stripped in 0.2M NaOH for 8 min and probed with β-actin antibody (1 in 30,000 in TBST) to serve as loading control.

2.2.17 Affymetrix Microarrays

RNA extraction was performed on MEF cells co-transduced with either empty pMSCVhyg and pMSCVpuro vectors or pMSCVhyg-Gata1 and pMSCVpuro-Fli1 four and seven months post-transduction as described in Section 2.2.13. RNA at 50 ng/uL (total of 500 ng for each sample) was provided to the Ramaciotti Centre (The UNSW Australia) for hybridization to Affymetrix GeneChip® Mouse Gene 1.0 ST arrays and scanning. Results were analysed using Partek® Genomics Suite™ 6.6, Copyright© 2013, Partek Inc., St. Louis, MO, USA.

2.2.18 Acetylcholinesterase detection assay

Cells already expressing GATA1 and FLI1 that have an additional factor introduced were grown in standard medium (HG DMEM, supplemented with 10% heat-inactivated FCS, 1% PSG solution) supplemented with 50 ng/mL recombinant human TPO, 2.5 µg/mL puromycin and 500 µg/mL hygromycin. Alternatively, bone marrow cells (obtained by flushing out the cells with 1 mL IMDM with 2% FCS from two femurs of a wildtype mouse using a 23G needle and 3mL syringe) and

MEF cells transduced with retro- and lentiviruses containing empty vectors, *Gata1*, *Fli1* and *Tal1* were cultured according to the manufacturer's instructions in a collagen-based medium on chamber slides using MegaCult®-C (Stemcell Technologies). Dehydrating, fixing and acetylcholinesterase staining of MEF cells were then performed as per the supplier's protocol.

2.2.19 Megakaryocyte differentiation of fibroblast

24 hours after MEF cells were transduced with combinations of retro- and lentiviruses containing various transcription factors, the cells were cultured in a standard medium (HG DMEM, supplemented with 10% heat-inactivated FCS, 1% PSG solution) for two days. To differentiate the cells into MK lineage, the confluent cells were trypsinised and detached from the tissue culture dishes and cultured in megakaryocyte lineage induction (MKLI) medium for eight days as described previously (Ono et al, 2012). The cells were then subjected to flow cytometry analysis as described in the next section.

2.2.20 Immunostaining for CD41 flow cytometric analysis

Expression of the megakaryocyte cell surface marker CD41 was analysed using the directly-labelled R-phycoerythrin (PE)-conjugated anti-mouse CD41 antibody (BD Biosciences). A single cell suspension was resuspended in FACS buffer (10 mM EDTA, 5% FCS, 0.05% NaN_3, PBS) at a concentration of 1 x 10^6 cells/ 100 µL. Antibody was added at its recommended concentration and the samples were incubated for 30 min in the dark at 4°C. The cells were washed with 5 times volume of FACS buffer and centrifuged at 500 g for 10 min at 4°C. Cells were then resuspended at 2 x 10^6 cells/mL for flow cytometric analysis. Cell exclusion dye

such as Topro-3 was used for live/dead cell discrimination. The data were collected using BD LSRFortessa™ SORP cell analyser (BD Biosciences), and analysed using FlowJo vX.0.6 (Tree Star).

CHAPTER 3 - A NOVEL DNA-BINDING MECHANISM BY THE MULTI-ZINC FINGER PROTEIN ZNF217

3.1 Introduction

ZNF217 is a multi-zinc finger protein that is aberrantly expressed in a number of cancers including breast, ovarian and colon (Peiro et al, 2002; Rooney et al, 2004; Yaswen & Stampfer, 2002), where its dysregulation has been implicated in tumour progression. The *ZNF217* gene is located on the q13.2 region of chromosome 20, which is amplified in a variety of cancer types showing aggressive tumour behaviour (Collins et al, 1998). The gene encodes a 1048 amino acid protein containing eight C_2H_2 zinc fingers (Collins et al, 1998). Of the eight zinc fingers, only the sixth and seventh fingers are separated by the canonical TGEKP linker, which is found in zinc finger proteins with DNA-binding activity (Iuchi, 2001; Laity et al, 2000; Wolfe et al, 2000). The additional presence of a C-terminal proline-rich sequence suggests a potential transcriptional regulatory domain consistent with the hypothesis that ZNF217 is a DNA-binding protein capable of modulating gene expression.

Until recently, a lack of experimental evidence confirming that ZNF217 can bind DNA had left researchers unsure of its biological roles and molecular mechanisms of action. The discovery by our group that ZNF217 interacts with the transcriptional co-repressor C-terminal Binding Protein (CtBP) (Quinlan et al, 2006) shed some light on how ZNF217 may function. As CtBP cannot directly bind DNA, it requires a DNA-binding transcription factor for recruitment to gene regulatory regions

(Chinnadurai, 2002; Turner & Crossley, 2001). The discovery that ZNF217 can bind CtBP was therefore strongly suggestive that ZNF217 may function to regulate gene expression by binding to DNA and recruiting gene regulatory cofactors. Furthermore, consistent with ZNF217's proposed role in cancer progression, CtBP is known to repress apoptotic, cell cycle and adhesion genes such as *Bax*, *p21* and *E-cadherin*, respectively, and can hence contribute to tumorigenesis and metastasis (Grooteclaes et al, 2003).

3.1.1 ZNF217 is a DNA-binding protein that recognises an eight base pair DNA consensus sequence via its zinc fingers 6 and 7

Despite being identified in 1998 (Collins et al, 1998), it was not until recently that the proposed DNA-binding activity of ZNF217 was confirmed. The Torchia laboratory and our group have independently shown that ZNF217 can directly bind DNA through zinc fingers six and seven (from here on referred to as F6-7) and can repress gene expression (Cowger et al, 2007; Nunez et al, 2011). The Torchia group reported that the two-finger domain recognises a 6 base pair (bp) consensus sequence of CAGAAY, where Y represents a C or T, whereas we observed that this site was bound by ZNF217 with low affinity prompting further investigations to search for higher affinity binding sites.

Electrophoretic mobility shift assay (EMSA) experiments performed in our laboratory revealed that ZNF217 F6-7 can bind strongly to an 8-bp Core Recognition Sequence (CRS) of TGCAGAAT (Nunez et al, 2011). The newly identified sequence revealed that the addition of two bases, TG, 5' to the previously described 6-bp CAGAAY core (Cowger et al, 2007) was able to convey high affinity binding

(Nunez et al, 2011). Further investigation into the extended CRS by EMSA revealed that all eight bases are important for DNA binding by ZNF217 and by individually mutating each base to every other possible base, a high affinity ZNF217 DNA recognition site of (T/A)(G/A)CAGAA(T/G/C) was determined (Nunez et al, 2011).

Importantly, the binding of two zinc fingers to an eight-base recognition sequence is somewhat unexpected given that zinc finger DNA binding domains more typically use three fingers to bind to sequences of nine bases (Foster et al, 1997; Pavletich & Pabo, 1991), with each finger making direct DNA contacts with three base pairs via the -1, +2, +3 and +6 amino acid positions of its α-helix (Choo & Klug, 1997).

3.1.2 Structural studies of ZNF217 F6-7 reveal residues implicated in DNA binding

^{15}N-HSQC NMR experiments have previously been carried out to determine potential amino acid residues of ZNF217 that might bind to the 8-bp CRS (Clifton, 2007; Nunez et al, 2011). Nine residues in ZNF217 F6-7 were identified that showed significant chemical shifts on binding to the CRS. These were residues Y485, L486, N487, H489, L490, T492 (on the α-helix of finger 6), T494 (on the canonical TGEKP linker between ZNF217 F6-7), and residues C504 and Q510 (on the β-hairpin and α-helix of finger 7 respectively).

However, as chemical shift changes in ^{15}N-HSQC may result not only from direct DNA contacts but also from local perturbations in structure arising from the binding event, further investigation was required to confirm key residues facilitating the interaction between ZNF217 and its consensus DNA-binding site.

3.2 Results – Identification of residues that are important for DNA recognition by mutagenesis scanning of ZNF217 F6-7

To determine if the residues in ZNF217 F6-7 identified by ^{15}N-HSQC NMR (Clifton, 2007; Nunez et al, 2011) are essential for DNA recognition, we assessed DNA binding by these residues using a mutagenesis strategy. To fully evaluate the F6-7 DNA binding domain, we extensively investigated the residues that make up both fingers, which include the nine previously identified candidate residues, by mutation to either alanine (A) or, where alanine was already present, to glutamine (Q).

Our hypothesis was that mutation of residues involved in making direct contact with DNA would disrupt the interaction between human ZNF217 F6-7 and the 8-bp CRS. To test this, we generated site-directed mutants of ZNF217 F6-7 and over-expressed them as GST-fusion proteins in bacteria (Figure 3.1). As a control, the region of the wildtype ZNF217 (amino acids 469-525) encompassing zinc fingers 6 and 7 was included (GST-ZNF217 F6-7) (Figure 3.1).

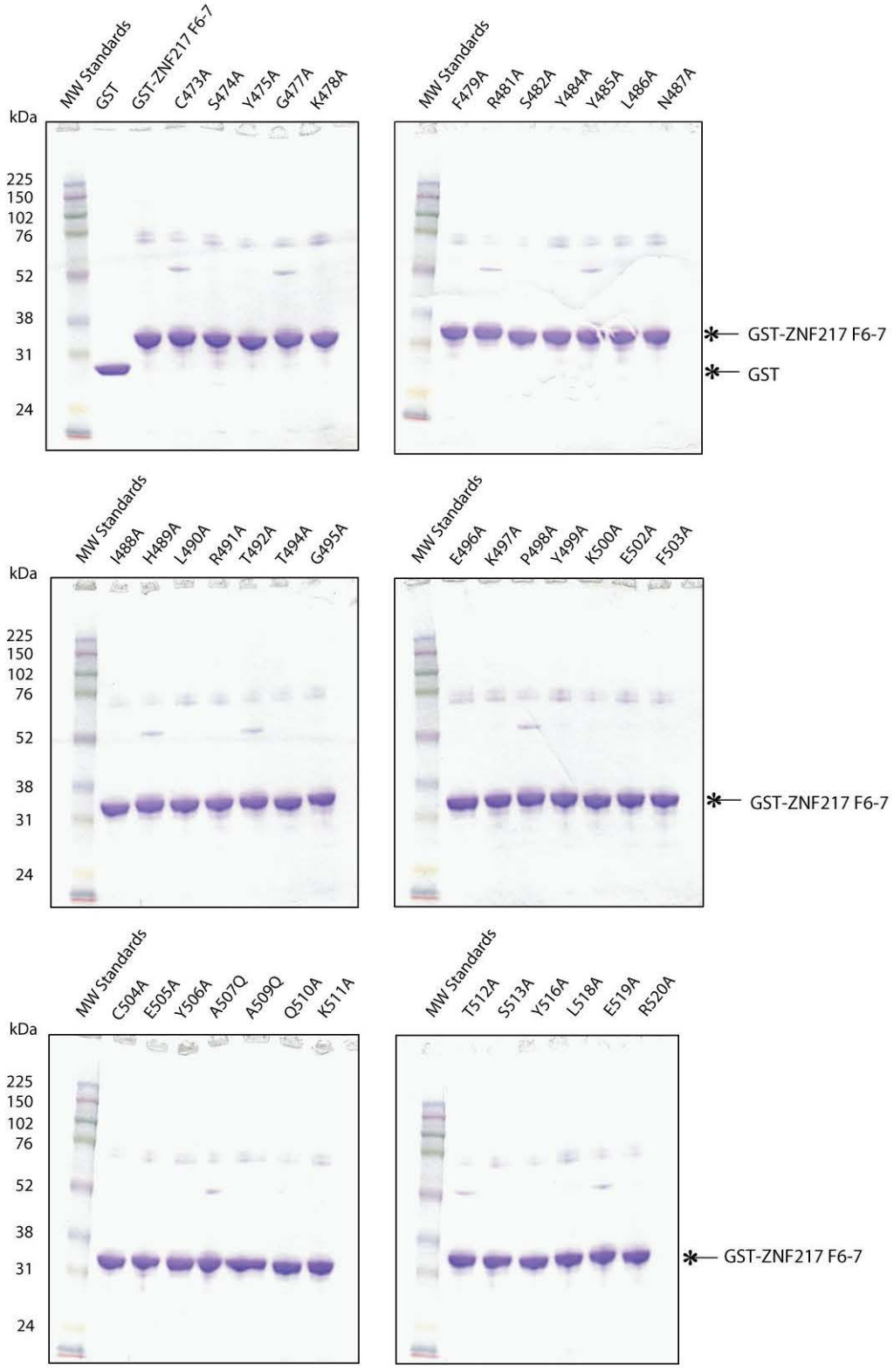


Figure 3.1. *Bacterial expression and purification of ZNF217 F6-7 and indicated mutants.* GST-fusion proteins were over-expressed in *E.coli* BL21 cells and purified by GSH-affinity. 5 µg of purified protein was run on 10% SDS-PAGE gels at 200V for 50 minutes and stained with Coomassie Blue for visualisation. GST alone is ~26 kDa, GST-ZNF217 F6-7 is ~33 kDa.

Having purified the various GST fusion proteins, we tested the ability of the ZNF217 F6-7 mutant proteins to bind the 8-bp CRS by EMSA (Figure 3.2).

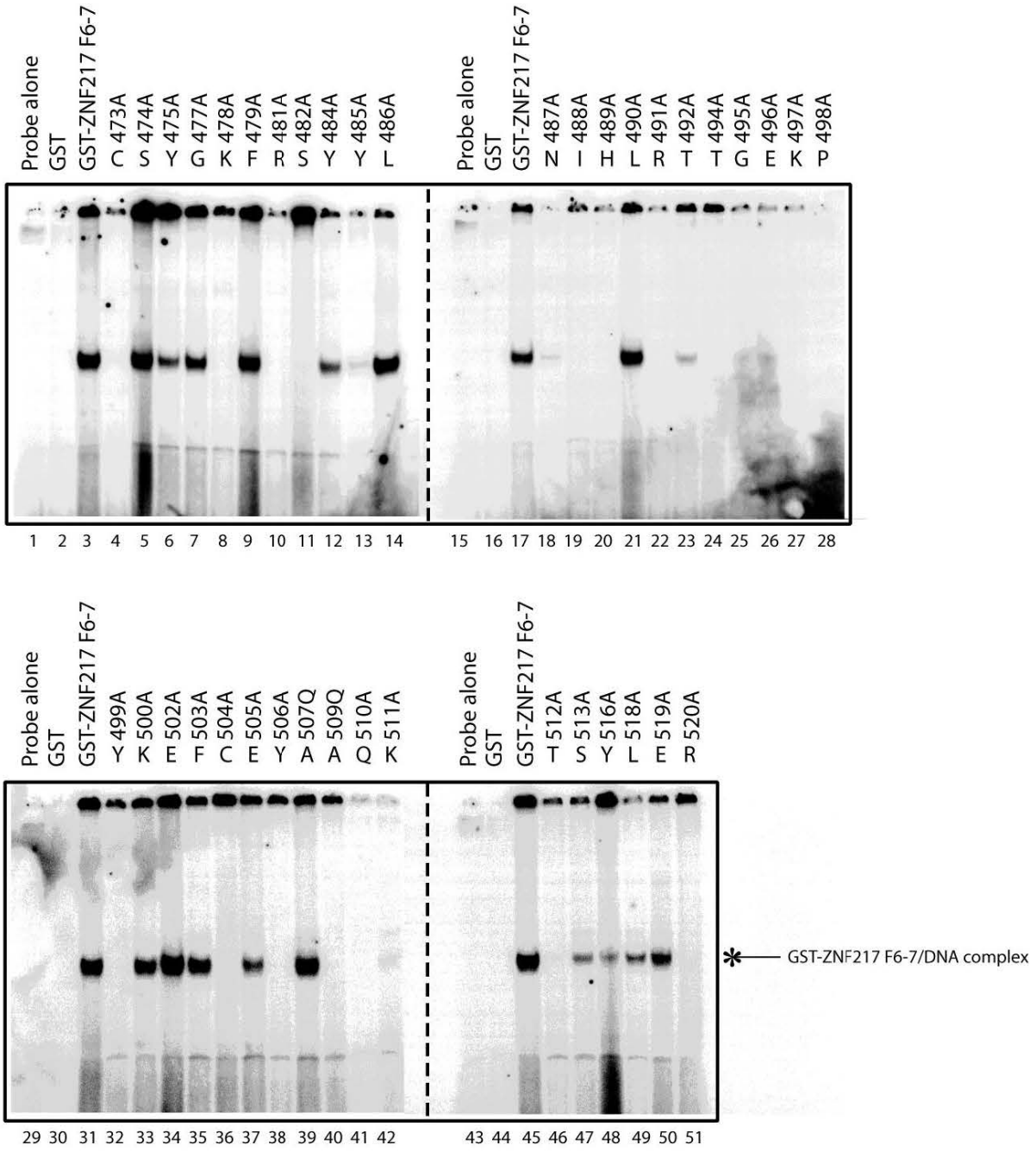


Figure 3.2. *Mutagenesis scan of the ZNF217 F6-7 DNA interaction.* Residues of zinc fingers 6 and 7 were mutated to alanine (or glutamine, where alanine was already present) and over-expressed in *E.coli* BL21 cells as GST fusion proteins. 1 µg of each of the GST fusion proteins was used in EMSA to detect binding to a radiolabelled probe containing the 8-bp CRS, TGCAGAAT. Panel below is a schematic showing the amino acid sequence of ZNF217 F6-7. The zinc chelating residues are in bold and denoted with asterisks (*), the TGEKP linker is in red, and positions -1, +2, +3 and +6 of zinc fingers 6 and 7 are indicated above.

To eliminate the possibility of mutant proteins not binding to the consensus site because of a misfolded protein structure, we assessed protein folding by one-dimensional (1D) ^{1}H NMR spectroscopy (Figure 3.3). In this technique, correctly folded proteins display a characteristic signature of hydrogen peaks, which is disrupted when mutations generate structural changes resulting in disordered regions (Kwan et al, 2011). The 1D NMR spectra of finger 6 (Figure 3.3A) and finger 7 (Figure 3.3B) mutants with reduced or no binding to the CRS, confirmed correct folding, except for Y506A, as demonstrated by the absence of peaks (Figure 3.3B) .

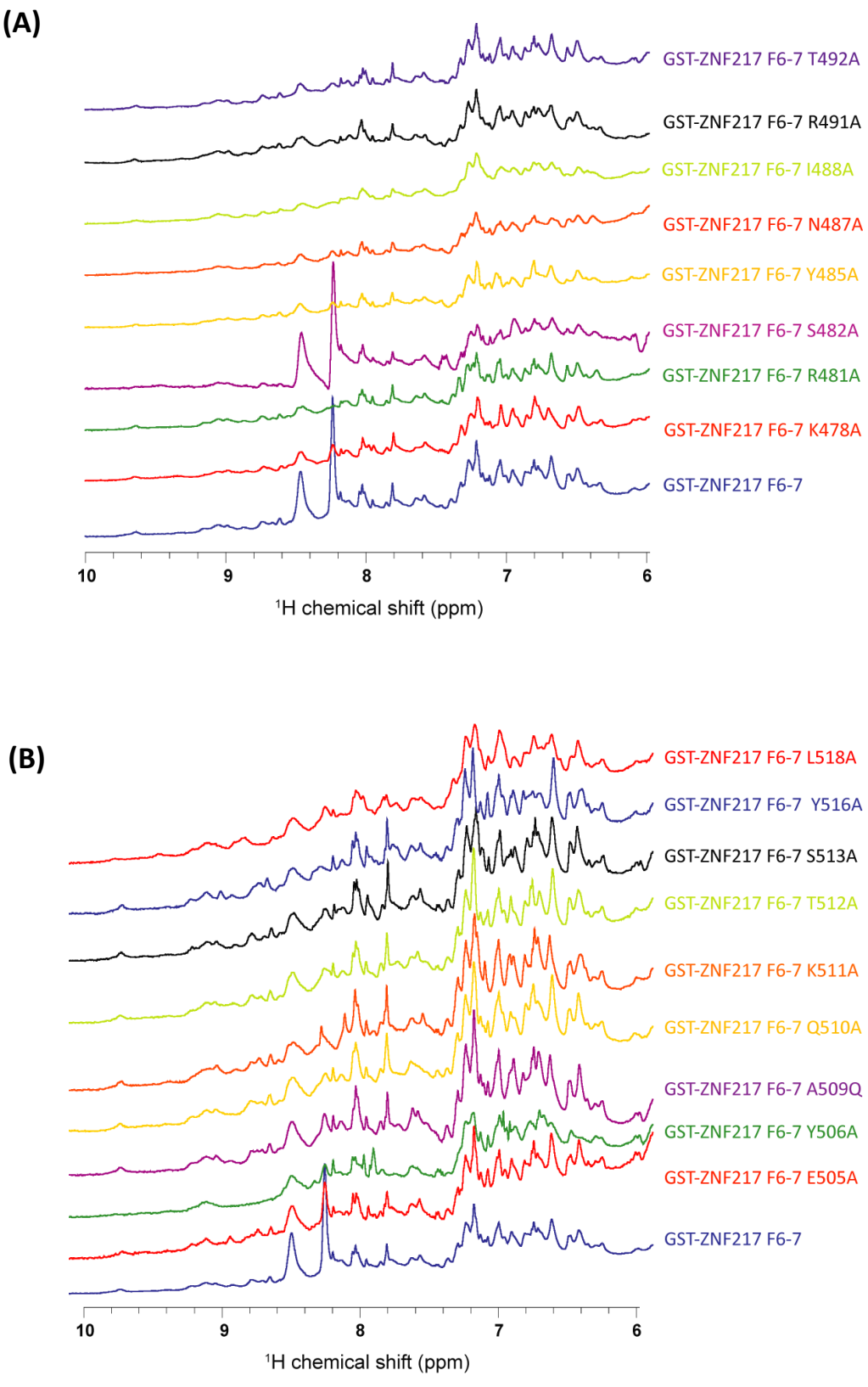


Figure 3.3. *One-dimensional ^{1}H NMR Spectra of GST-ZNF217 F6-7 and indicated mutants.* Shown are the 600 MHz 1D ^{1}H NMR spectra of the **(A)** Finger 6 and **(B)** Finger 7 mutants that were found to reduce or eliminate DNA binding as identified by EMSA (Figure 3.2). All

mutants folded correctly except for Y506A which showed absence of peaks particularly in the region of 6.4 to 7.4 ppm.

3.2.1 Mutating the canonical binding residues of classical zinc fingers compromises DNA binding of ZNF217 F6-7

In classical zinc finger proteins, the residues that typically make sequence-specific contacts with DNA are found at positions -1, +2, +3 and +6 (relative to the start of the α-helix) of each zinc finger (Choo & Klug, 1997; Foster et al, 1997; Pavletich & Pabo, 1991). Hence, it is predicted that mutating the residues at these positions will disrupt the interaction between ZNF217 F6-7 and its canonical DNA binding site. Indeed, by EMSA it was confirmed that all of the aforementioned helical positions in both fingers 6 and 7 of ZNF217 are critical for binding to this consensus (Figure 3.2). We mutated S482 (Figure 3.2, lane 11), Y485 (lane 13) and I488 (lane 19) at helical positions -1, +3 and +6 of finger 6 respectively, and also Q510 (lane 41) and T512 (lane 46) at positions -1 and +2 respectively of finger 7. We found that alanine substitution at each of these sites completely abolished DNA binding suggesting that these residues strongly contribute to DNA recognition at this site. Consistent with this, Y485 and Q510 have previously been identified as candidate DNA contact residues based on their chemical shifts in ^{15}N-HSQC experiments (Clifton, 2007; Nunez et al, 2011). In addition, mutation of residues Y484 (Figure 3.2, lane 12) at position +2 of finger 6, and S513 (lane 47) and Y516 (lane 48) at positions +3 and +6 respectively of finger 7 reduced but did not abolish DNA binding, implying that these too have roles in DNA specificity and affinity.

3.2.2 Interfering with the conserved TGEKP linker between ZNF217 F6-7 abolishes DNA binding

Disrupting the TGEKP linker between F6-7 seems to greatly influence the capability of ZNF217 to bind DNA. By mutating each linker residue (T494, G495, E496, K497 and P498) to alanine, the ability of ZNF217 F6-7 to recognise the 8-bp CRS is abolished (Figure 3.2, lanes 24-28). This result is consistent with previous reports of reduced binding affinity of the zinc finger protein TFIIIA as a result of mutation of the conserved TGEKP linker residues (Choo & Klug, 1993; Clemens et al, 1993; Clemens et al, 1994). The conserved linker is reported to have an important role in increasing the binding affinity of the zinc finger thus stabilising the protein-DNA complex (Laity et al, 2000; Pavletich & Pabo, 1991; Wuttke et al, 1997).

3.2.3 Disrupting the zinc-ligating residues of ZNF217 F6-7 abrogates DNA binding

As can be seen from Figure 3.2, mutation of the zinc-ligating residues at C473 and H489 of finger 6 (lanes 4 and 20 respectively) and at C504 of finger 7 (lane 36) to alanine abrogates *in vitro* DNA binding by ZNF217 F6-7. Previously our lab has shown by 1D NMR spectroscopy that these mutants do not fold correctly (Nunez, 2012). This is in agreement with the accepted role of zinc-ligating residues in determining and stabilising zinc finger structural integrity (Dang et al, 2000; Matthews & Sunde, 2002; Simpson et al, 2003) and it is hence unsurprising that mutating them results in a misfolded peptide unable to bind DNA.

3.2.4 Additional residues in ZNF217 F6-7 contribute to DNA binding

In addition to the zinc-ligating residues, TGEKP linker and canonical DNA binding positions, ten further residues were identified by alanine/glutamine substitution as having a role in the DNA-ZNF217 interaction. Eight of these residues appeared critical to binding with their mutation either mostly or completely eliminating the *in vitro* interaction of F6-7 with DNA. These key residues are K478 (Figure 3.2, lane 8), R481 (lane 10), N487 (lane 18), R491 (lane 22), T492 (lane 23), Y499 (lane 32), A509 (lane 40) and K511 (lane 42). A further two amino acids at Y475 (Figure 3.2, lane 6) and E505 (lane 37) also make some contribution to DNA binding, as their mutation reduced but did not abolish DNA binding. In agreement with 1D NMR data (Figure 3.3B) suggesting structural misfolding, mutant Y506A was unable to bind to the consensus DNA probe in EMSA (Figure 3.2, lane 38).

3.3 Discussion

The mechanisms through which zinc fingers recognise their target DNA sequences have been extensively studied (Pavletich & Pabo, 1991; Pavletich & Pabo, 1993; Wolfe et al, 2001; Wolfe et al, 2000; Wuttke et al, 1997). However, the vast majority of work has centred on classical tandem arrays of three or more zinc fingers separated by characteristic TGEKP linkers, where each finger makes contact with three base pairs of double stranded DNA. However, a number of DNA-binding zinc finger proteins that do not contain three or more zinc fingers of this type have also been discovered and investigated. These include the single GAGA zinc finger (Matharu et al, 2010) and the two-zinc finger protein Tramtrack (Fairall et al, 1993). This raises the question of how these more unusual factors contact DNA.

In this chapter, we have examined an eight-zinc finger protein ZNF217 that has previously been reported to bind eight base pairs of DNA via only two of its classical zinc fingers (Nunez et al, 2011). We have carried out a comprehensive mutagenesis scan of ZNF217 zinc fingers 6 and 7 to directly assess the contribution of each amino acid to DNA binding. Previous work using ^{15}N-HSQC NMR spectroscopy had implicated nine residues in DNA binding (Clifton, 2007; Nunez et al, 2011). Our mutagenesis experiments have demonstrated that five of these residues are indeed essential for DNA binding. These residues are Y485, N487, T492, T494 and Q510 (Figure 3.4, underlined and bold red). However, we also discovered that not all of the residues are involved in making direct contact with DNA. Two residues, L486 and L490, did not appear to be essential for binding, suggesting that the chemical shift changes observed for these residues might in fact be caused by conformational changes arising from the binding event. Moreover, the involvement of a further two residues in DNA binding, H489 and C504, is difficult to ascertain as both are responsible in coordinating the zinc ion essential for the formation of a stable zinc finger secondary structure (Dang et al, 2000; Mackay & Crossley, 1998; Matthews & Sunde, 2002; Simpson et al, 2003).

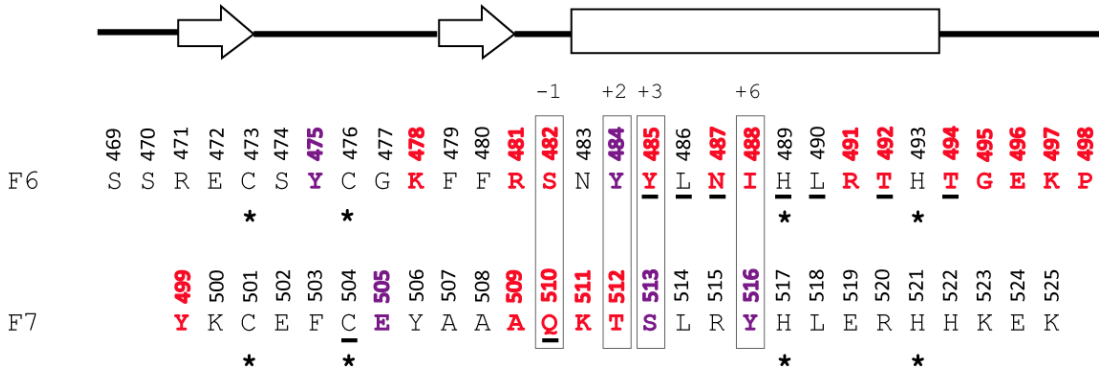


Figure 3.4. *Residues important for ZNF217 F6-7 DNA recognition*. Shown are the amino acid sequences of Finger 6 (F6) and Finger 7 (F7) of ZNF217. The residues that typically make sequence-specific contacts with DNA in classical zinc fingers are boxed and the numbering is based on the start of the α-helix. Residues that underwent significant chemical shift changes upon the addition of DNA to ZNF217 F6-7 as identified by ^{15}N-HSQC NMR experiments (Nunez et al, 2011) are underlined. The zinc-ligating residues are indicated with asterisks (*). Residues that were shown by site-directed mutagenesis to mostly or completely eliminate DNA binding are in bold and coloured red, whereas residues that reduced but did not abolish DNA binding are in bold and coloured purple.

In addition to residue T494, we found that all other residues of the TGEKP linker are important for DNA recognition (Figure 3.4, bold red). This conserved linker is found in approximately half of all C_2H_2 zinc finger proteins where it functions as a structural cap to the C-terminus of the α-helix of the preceding zinc finger thereby stabilising the DNA-protein complex (Laity et al, 2000). Typically, a hydrogen bond is formed between the amide of the glycine residue of the linker (G495 in ZNF217) and the carbonyl group of the third residue from the end of the helix, commonly an arginine or lysine (arginine, R491 in finger 6) (Pavletich & Pabo, 1991; Wuttke et al,

1997). It is probable that mutation of these residues resulted in the abrogation of DNA binding due to effects on zinc finger structure where C-capping cannot occur.

Among the residues located in the canonical DNA binding positions in the α-helix, Y485 at helical position +3 of finger 6, and Q510 at helical position -1 of finger 7, showed both chemical shift changes in NMR experiments and severe disruption to ZNF217-DNA binding in EMSA (Figure 3.4, bold red and underlined) (Nunez et al, 2011). Taken together, these results provided compelling evidence that the two residues are directly involved in making contact with the nucleotide bases of DNA.

Following the work described in this chapter, the crystal structure of ZNF217 F6-7 complexed to the DNA recognition sequence of TGCAGAAT, has been published (Protein Data Bank (PDB) codes: 4F2J and 4IS1), and has confirmed that Y485 and Q510 do indeed make base-specific contacts with DNA (Vandevenne et al, 2013). The crystal structure shows that the hydrophobic ring of Y485 interacts with the methyl group of a thymine (T) base, present in the DNA binding consensus site (Figure 3.5), via a methyl-π interaction. Although, this type of interaction occurs in a number of protein-nucleic acid complexes, it has not previously been reported for classical zinc finger domains. The structure also confirmed that residue Q510, forms a hydrogen bond with a cytosine (C) base of ZNF217's canonical binding site (Figure 3.5).

The solving of the crystal structure also revealed that in addition to direct DNA interaction, Y485 makes an electrostatic interaction with the sugar-phosphate backbone of DNA, together with residues Y484 and T492 (Vandevenne et al, 2013).

Mutation of Y484 was found to have a small effect on DNA binding in EMSA, while this study and ^{15}N-HSQC NMR indicate that T492 makes direct contact with DNA (Nunez et al, 2011). The crystal structure also indicated that residues T512 (+2 of finger 7 helix) and Y516 (+6 of finger 7 helix), identified as key residues in our mutagenesis studies, contact DNA. T512 hydrogen bonds to a guanine (G) base of the binding site, while Y516 forms a second methyl-π interaction with the methyl group of a thymine (T) base of the consensus site, along with a van der Waals contact with an adenine (A) base of the CRS (Figure 3.5) (Vandevenne et al, 2013). Of the non-canonical DNA-binding residues, mutation of R481 at the -2 position of the α-helix of finger 6 completely eliminated DNA binding by ZNF217. In agreement with this, the three-dimensional structure of the ZNF217-DNA complex confirmed that the guanidinium group of R481 makes a double hydrogen bond with the oxygen and nitrogen atoms of a guanine (G) base of the ZNF217 binding sequence (Figure 3.5) (Vandevenne et al, 2013).

We found that our previously identified DNA binding consensus sequence of (T/A)(G/A)CAGAA(T/G/C) established by mutagenesis analysis (Nunez et al, 2011) differs from the consensus sequence ATTCC(G/A)AC derived from ZNF217 chromatin immunoprecipitation (ChIP)-Chip data (Krig et al, 2007). This may be due to the fact that ChIP assays cannot differentiate between direct binding of transcription factors to DNA and indirect binding via association with other transcription factors bound to recognition sites (Farnham, 2009). As eukaryotic DNA is organised into nucleosomes which are further packaged into chromatin, a transcription factor's ability to gain access to its target genes is dependent on chromatin structure (Zhang et al, 2000). It has been reported that transcription factors

bind to differing subsets of target genes in different cell types as a result of chromatin accessibility (Kajimura et al, 2008). As a result, chromatin structure will influence and define each tissue-specific subset of ZNF217 target genes.

Defining the direct interaction between DNA and the zinc finger residues of ZNF217, and the crystal structure of the ZNF217-DNA complex, may well contribute to the generation of new artificial transcription factors (ATFs) capable of targeting nucleosomal DNA containing the ZNF217 consensus sequence, to manipulate gene expression. Given that ZNF217 is a putative oncogene, it would be beneficial to control the expression of its target genes.

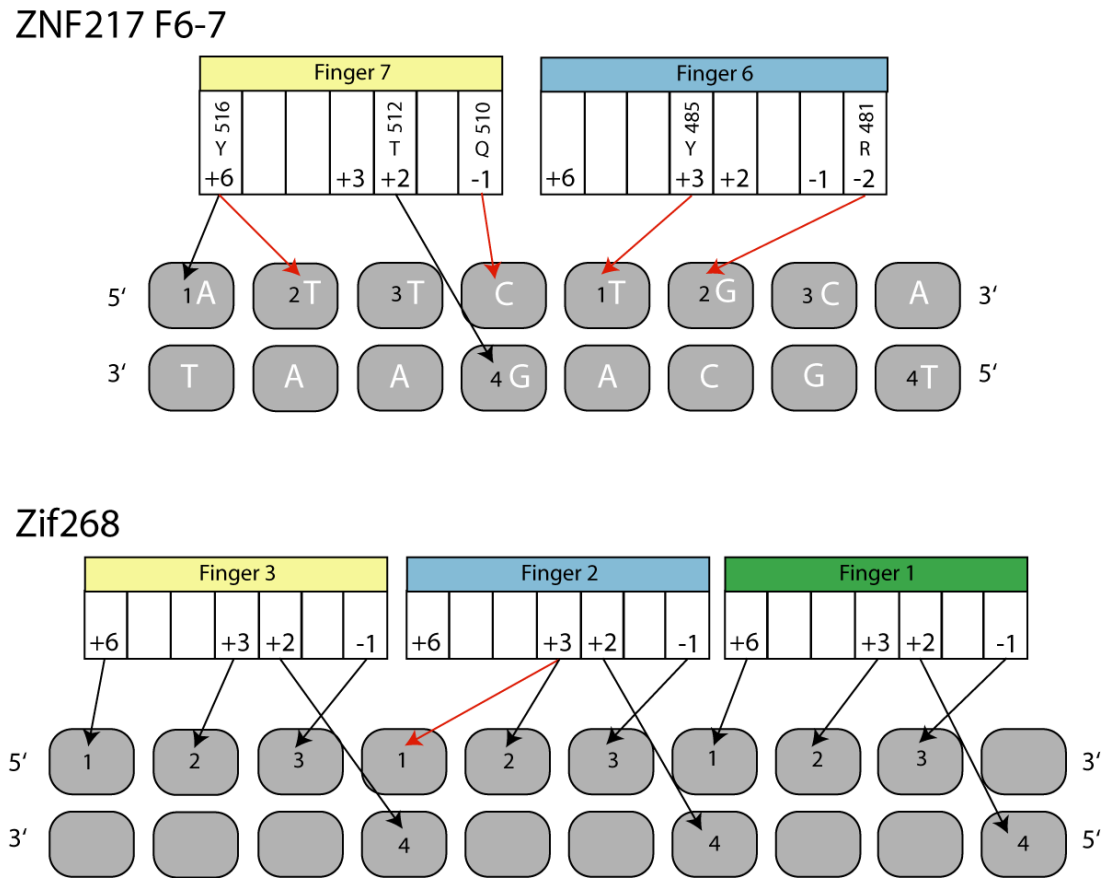


Figure 3.5. *A comparison of sequence–specific base contacts made by the classical zinc fingers of Zif268 and fingers 6 and 7 of ZNF217.* In the canonical interaction of classical zinc fingers, residues in the +2, +3 and +6 positions of the α-helix, together with the residue immediately preceding the helix (the -1 position), typically make base-specific contacts with DNA (represented by black arrows). Of the three or four bases typically recognised by a zinc finger, the residue at position +6 of the helix contacts the 5' base (1), the +3 residue contacts the central base (2) and the -1 residue contacts the 3' base (3) on the same non-coding strand of DNA (represented here as the top DNA strand). The residue at position +2 generally contacts a fourth base (4) on the other strand of DNA called the coding strand (bottom DNA strand). The interactions for ZNF217 F6-7 (Protein Data Bank (PDB) ID: 4IS1) (Vandevenne et al, 2013) and Zif268 (PDB ID: 1AAY) (Pavletich & Pabo, 1991) illustrated here were derived from the published crystal structures of the zinc finger-DNA complexes. Non-canonical interactions are represented in red. ZNF217 F6-7 interactions with its 8-bp consensus sequence, TGCAGAAT (represented as white letters), are depicted in more detail. The residues that make direct contact with DNA are indicated above their respective position in the helix.

In summary, our analysis of the molecular mechanism whereby ZNF217 contacts DNA has now been validated by the very recent crystal structure of the ZNF217-DNA complex. Overall, it was found that ZNF217 displays both similarities and differences from other zinc finger domains in the mechanism by which it recognises DNA (Vandevenne et al, 2013), which are highlighted by a comparison of the pattern of sequence-specific interactions (Vandevenne et al, 2013) with that reported for Zif268 (Elrod-Erickson et al, 1996; Pavletich & Pabo, 1991) (Figure 3.5). Fingers 1 and 3 of Zif268 follow the canonical pattern of binding with positions -1, +2, +3 and +6 of the recognition helix making contact with specific base positions in the DNA. Of the three bases typically recognised by a classical zinc finger, the residue at position +6 of the helix contacts the 5' base on one strand, the +3 residue contacts the central base, and the -1 residue contacts the 3' base. The residue at position +2 generally contacts the base pair immediately 3' to the 3-bp site. However, in the ZNF217 complex, only two canonical DNA contacts are observed via residues T512 (at position +2) and Y516 (at position +6) of the helix of finger 7 (Figure 3.5), indicating an unusual binding strategy noticeably different from that of canonical zinc finger-DNA interactions.

Currently, the only other structure of a two-zinc finger protein bound to DNA to be reported is that of Tramtrack (Fairall et al, 1993), where the manner that the second finger of Tramtrack contacts DNA resembles the "standard" interaction pattern for classical zinc fingers. However, ZNF217 offers a different mode of interaction that may possibly be used by other zinc finger proteins containing single or multiple two-finger units. Among such multi-zinc finger proteins are the mammalian transcriptional regulators BCL11A and ZNF219, which play important roles in

haematopoiesis and cell differentiation (Takigawa et al, 2010; Wu et al, 2013; Yu et al, 2012). They contain paired-zinc finger domains that are known to contact DNA and share over 50% sequence identity with fingers 6 and 7 of ZNF217 (Avram et al, 2002; Sakai et al, 2003). Additionally, other proteins such as ZNF536, which like ZNF217 bind the transcriptional co-factor CtBP, may also exhibit DNA-binding activity through double finger motifs (Qin et al, 2009).

CHAPTER 4 – DIRECTING FIBROBLASTS TOWARDS THE MEGAKARYOCYTE LINEAGE

4.1 Introduction

Megakaryocytes (MKs) are blood cells that produce platelets, which are essential for the process of haemostasis and blood clotting (Davi & Patrono, 2007; Varga-Szabo et al, 2008). Hence, a decline in platelet numbers or thrombocytopenia, which can result from medical treatments such as chemotherapy, stem cell transplants or naturally occurring diseases, leads to bleeding disorders (Haddad et al, 1999; Vadhan-Raj, 2009; Yamazaki et al, 2006). Current treatments for this condition include platelet transfusions, which are limited by a number of practical issues such as shortage of platelet supply and the possibility of immune reactions (Kaushansky, 2008; Stroncek & Rebulla, 2007). With the steady increase of platelet transfusions over the past decades due to increasing numbers of patients undergoing chemotherapy or stem cell transplantation, *in vitro* differentiation systems have been developed to generate platelet-producing MKs from haematopoietic stem cells (HSCs) (Matsunaga et al, 2006; Reems et al, 2010; Schipper et al, 2003) and embryonic stem cells (ESCs) (Fujimoto et al, 2003; Gaur et al, 2006; Takayama et al, 2008). However, the use of HSCs for generating MKs is restricted by the difficulty of obtaining these cells from bone marrow, peripheral blood or cord blood. Additionally, HSCs possess low capacity for *in vitro* expansion, and tend to yield insufficient MKs and platelets for clinical use (Reems et al, 2010). While ESCs, unlike HSCs, can proliferate *in vitro* to produce copious amount of cells (Thomson et

al, 1998), complicated experimental techniques are required to differentiate cells down the MK lineage (Fujimoto et al, 2003; Lu et al, 2011; Takayama et al, 2008). Thus, both ESCs and HSCs are problematic sources and simpler alternative experimental strategies are being sought.

To this end, fibroblast cells are of particular interest because they are abundant, easily obtained and readily grown in culture. Recent advances in the field of lineage conversion by transcription factors have shown that fibroblasts can be reprogrammed into neurons (Vierbuchen et al, 2010), cardiomyocytes (Ieda et al, 2010), hepatocytes (Huang et al, 2011) and macrophages (Feng et al, 2008). Motivated by this, we have explored the potential of transcription factors to directly reprogram differentiated fibroblasts towards the MK lineage as a novel alternative to the treatment of thrombocytopenia.

4.2 Up-regulation of MK-specific genes in fibroblast cells

To identify transcriptional regulators with the potential to reprogram toward the MK lineage, we began by testing over-expression of four factors, GATA1, FLI1, FOG1 and ETS1. For example, we chose *Ets1* since its expression increases during megakaryocytic differentiation (Terui et al, 2000), where it functions to bind and activate the promoters of MK-specific genes such as glycoprotein IIb *(Gp2b* or *CD41*) and platelet factor 4 (*Pf4*) (Lemarchandel et al, 1993; Minami et al, 1998). The other three factors are similarly implicated in megakaryocytic biology (Athanasiou et al, 1996; Shivdasani et al, 1997; Spyropoulos et al, 2000; Starck et al, 2003; Tsang et al, 1998; Vyas et al, 1999).

In this preliminary work, murine embryonic fibroblasts (MEFs) were retrovirally transduced with vectors expressing each of the four factors individually and also in pairwise combinations. Quantitative real time RT-PCR was then performed to confirm the over-expression of the four factors (Appendix, Figure A1) and also assess expression of one representative megakaryocyte/platelet-specific gene *Pf4* (Figure 4.1).

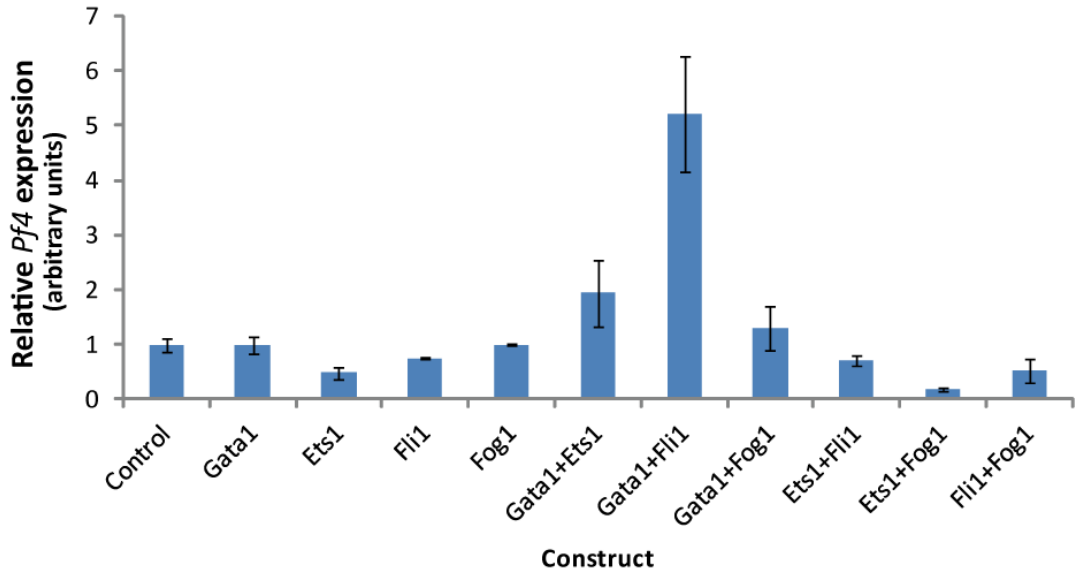


Figure 4.1. *Forced expression of MK transcriptional regulators in murine embryonic fibroblasts (MEFs). Pf4* mRNA was assessed by quantitative real time RT-PCR in MEFs following transduction with retroviral vectors containing the indicated MK lineage-specific transcription factors. Control sample represents MEFs transduced with vector alone. *Pf4* mRNA level was normalised to *18S* rRNA level and then normalised again to the control sample, which was set at 1. Error bars represent standard error of the mean (n=3).

The ectopic expression of individual candidate MK transcriptional regulators in MEFs appeared to have little effect on expression of the endogenous *Pf4* gene (Figure 4.1). However, when two transcription factors were co-transfected, certain paired combinations resulted in up-regulation of *Pf4* expression. A marginal increase was achieved by GATA1 and ETS1 and a more dramatic five-fold up-regulation was seen in cells over-expressing GATA1 and FLI1. While GATA1 and ETS1 have previously been shown to activate *Pf4* expression in MK cells (Minami et al, 1998) and GATA1 and FLI1 have been shown to interact and mediate synergistic activation of glycoprotein IX (*Gp9* or *CD42a*) and glycoprotein Ibα (*Gp1bα or*

CD42b) (Eisbacher et al, 2003) in human K562 blood cells, we were encouraged to see up-regulation of the *Pf4* marker gene in unrelated fibroblasts.

To further explore our preliminary observations, we next generated fibroblast cell lines stably over-expressing GATA1 and FLI1. MEFs were retrovirally co-transduced with pMSCVhyg-Gata1 and pMSCVpuro-Fli1 vectors and grown in the presence of puromycin and hygromycin to select for cells stably expressing both factors. Following antibiotic selection for two weeks, samples were collected from control cells transduced with empty vectors and from cells infected with retroviruses containing *Gata1* and *Fli1*. Real time RT-PCR (Figure 4.2A) and Western blotting (Figure 4.2B) were then performed to confirm forced expression of GATA1 and FLI1 in these cells. While endogenous GATA1 and FLI1 could not be detected in control MEFs by Western blot, we observed significantly higher levels of both proteins in transduced MEFs, as indicated by the presence of bands migrating at the expected mass (Figure 4.2B).

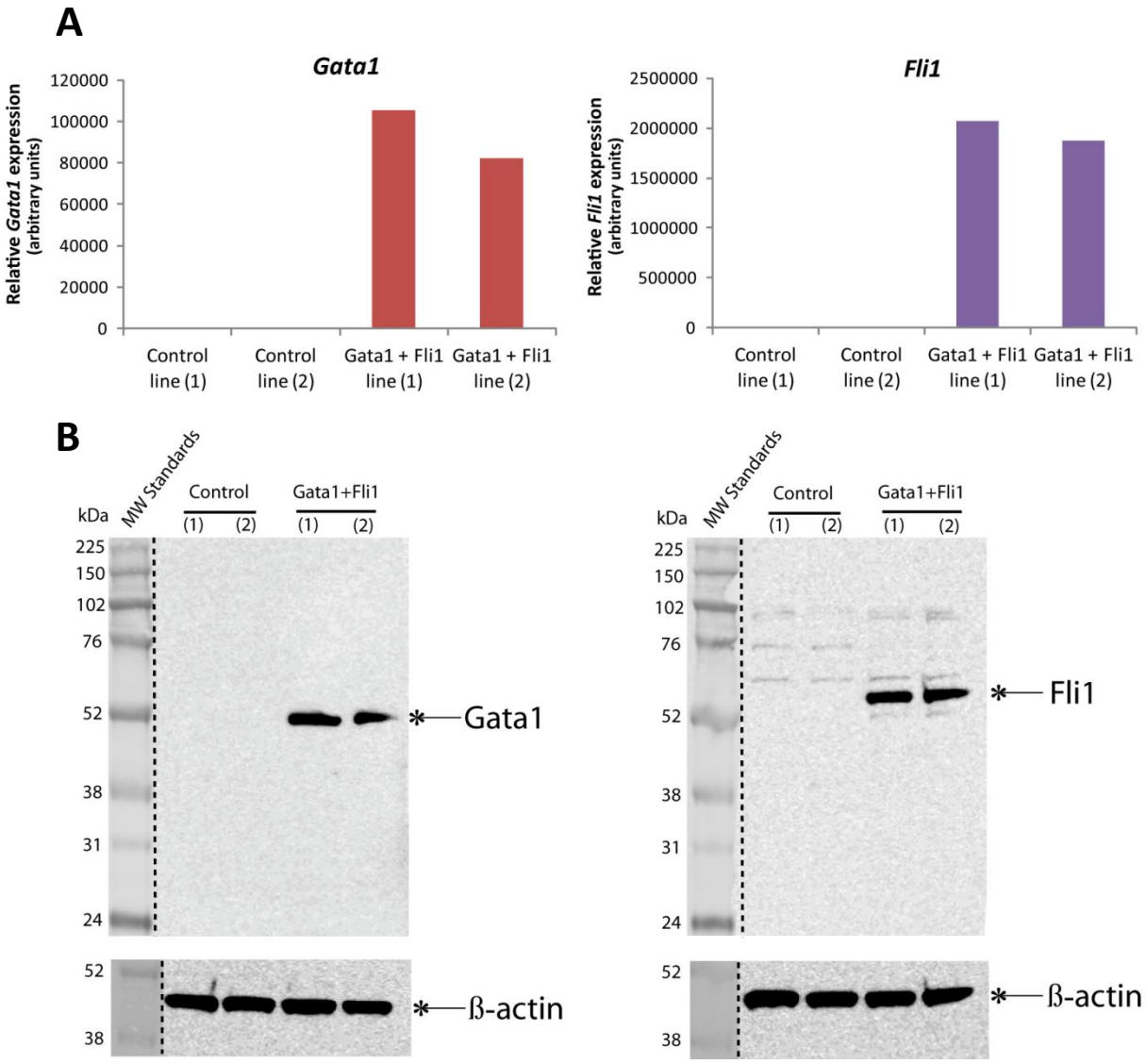


Figure 4.2. *Generation of fibroblast cell lines stably over-expressing GATA1 and FLI1*.

(A) MEFs retrovirally co-transduced with either empty pMSCVhyg and pMSCVpuro vectors (n=2), or pMSCVhyg-Gata1 and pMSCVpuro-Fli1 (n=2), were cultured in puromycin and hygromycin to select for cells expressing both factors. Total RNA was extracted from cells after two weeks of antibiotic selection and expression of *Gata1* and *Fli1* quantified by real time RT-PCR. *Gata1* and *Fli1* mRNA levels were normalised against *18S* rRNA levels and normalised again to the control sample that gave the lowest signal for the gene, which was set as 1. **(B)** Nuclear extracts from control and GATA1/FLI1 MEFs were analysed by Western blotting using antibodies specific for GATA1 (top left panel) and FLI1 (top right panel). Membranes were stripped and re-probed with a β-actin antibody (bottom panels) to confirm equal sample loading.

To investigate the effect of stable over-expression of GATA1 and FLI1 on the phenotype of the fibroblast cells, the transcript levels of key megakaryocytic markers including the chemokine PF4, and the MK cell surface receptors, CD61 and CD41, were also analysed. In cells stably over-expressing GATA1 and FLI1, we found a robust nine-fold up-regulation of *Pf4* mRNA compared to control cells (Figure 4.3) in agreement with our preliminary transient transfection experiments (Figure 4.1). Expression of *CD61* also increased, while *CD41* levels appeared unaffected by ectopic expression of both factors (Figure 4.3). By tracking gene expression levels of *Pf4* and *CD61* over time, we were able to determine that significant reprogramming of these MK loci is dependent upon prolonged over-expression of GATA1 and FLI1 (Figure 4.4).

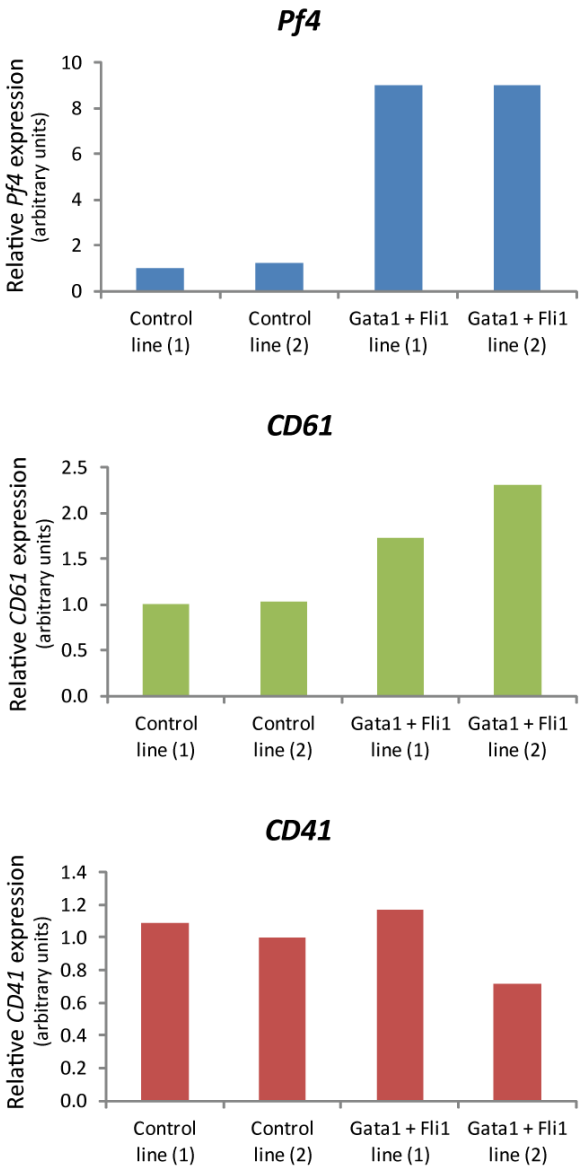


Figure 4.3. *Forced expression of GATA1 and FLI1 alters the expression of megakaryocytic markers.* Total RNA from fibroblast cell lines stably over-expressing GATA1 and FLI1 (n=2) were analysed by real time RT-PCR for the expression of key megakaryocytic markers *Pf4*, *CD61* and *CD41*. Control samples represent cell lines co-transduced with empty pMSCVhyg and pMSCVpuro vectors (n=2). Transcript levels for each gene were normalised against *18S* rRNA levels and then again normalised to the control sample that gave the lowest signal for the gene, which was set as 1.

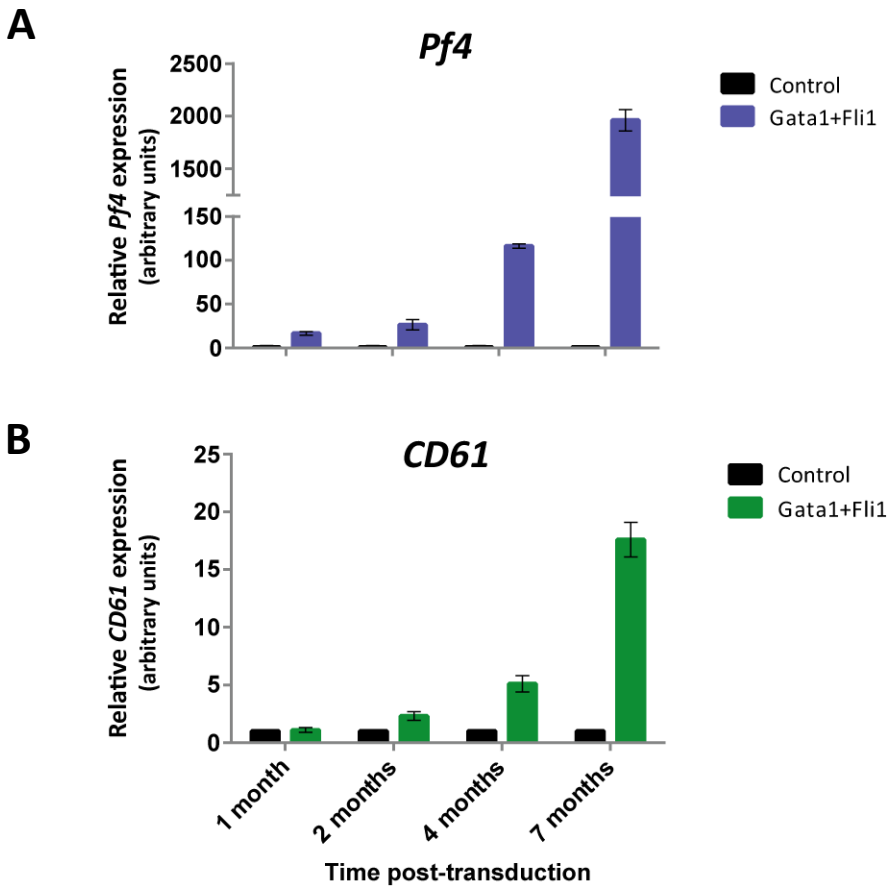


Figure 4.4. *Reprogramming of the Pf4 and CD61 loci over time.* Total RNA was extracted from MEFs stably expressing GATA1 and FLI1 at time points 1, 2, 4 and 7 months post-transduction. **(A)** *Pf4* and **(B)** *CD61* mRNA levels were analysed by real time RT-PCR. Control samples represent cell lines co-transduced with empty pMSCVhyg and pMSCVpuro vectors. Transcript levels were normalised against *18S* rRNA levels and then normalised again to the control sample, which was set as 1. Error bars represent standard error of the mean of two independent cell lines.

To further investigate the extent of MK differentiation in MEFs over-expressing GATA1 and FLI1, Affymetrix microarray analysis was performed to compare the global gene expression patterns with control cells transduced with empty vectors. Total RNA was prepared individually from two independent samples of Control and Gata1+Fli1 MEF cells, after four and seven months in culture. Following successful

quality control analysis, Affymetrix microarrays were performed on two separate occasions for the two different time points.

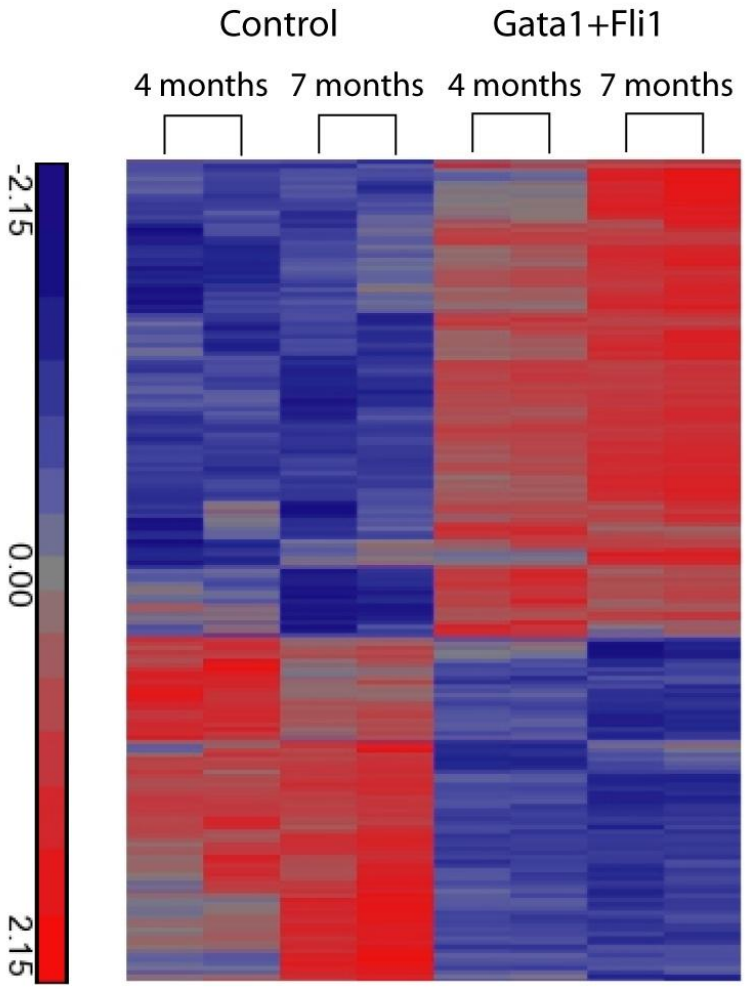


Figure 4.5. *Gene expression changes in fibroblasts over-expressing GATA1 and FLI1.* The heat map displays results for all differentially expressed genes whose average expression changed by at least 1.5 fold in GATA1/FLI1 cells at four and seven months time points compared to control, at a p-value with false discovery rate (FDR) <0.05. Up-regulated genes are shown in red and down-regulated in blue. This map was generated by Partek Genomics Suite™ 6.6.

In analysing the microarray data, GATA1 and FLI1 over-expressing cells were compared to control cells at each of their respective time points. Differentially

expressed genes were defined as having an average expression fold change of ≥ 1.5 at a p-value with false discovery rate (FDR) of less than 0.05. A heat map was then generated to visualise the gene expression profiles of control and Gata1+Fli1 samples at four and seven months post-generation of the stable cell lines (Figure 4.5). The heat map illustrates the significant differences between the gene expression patterns of control and Gata1+Fli1 expressing cells and highlights that these changes are more pronounced following sustained long term expression of these factors, supporting our previous observations (Figure 4.4).

Lists of genes with a fold up- or down-regulation of greater than 1.5 with a p-value of less than 0.05 in GATA1 and FLI1 expressing cells compared to control at four and seven months were generated. The genes were ranked in order of fold change and the top 20 genes for each list were selected. The tables are included in the appendix. Additionally, pathway analysis was performed using DAVID (the Database for Annotation, Visualization, and Integrated Discovery) bioinformatics resources (Dennis et al, 2003; Huang da et al, 2009). Lists of the functional groups to which genes were associated that changed by at least 1.5 fold up or down in GATA1/FLI1 cells at four and seven months time points compared to control with a p-value and false discovery rate (FDR) of less than 0.05, are shown in the appendix (Table 5 and Table 6).

From examining the lists of differentially expressed genes, some general comments can be made about the pattern of gene expression between GATA1/FLI1 expressing cells and control. Unexpectedly, the most up-regulated genes in GATA1/FLI1 expressing cells at four and seven month time points have a range of functions,

which include regulation of adipogenesis (*Pparg*), metabolic processes (*Mgat4a*, *Chst15*), cell signalling (*Usp18*, *Cxcl12*), maintenance of cell structure and cell adhesion (*Podxl*, *Gpr56*, *Dock4*, *Cytip*, *Nid1*). Similarly, the genes that are most down-regulated in GATA1/FLI1 expressing cells also have varying functions, although, as expected, a number of them are commonly expressed in fibroblasts. These include *Fbln5*, *Tagln*, *Fbn1* and *Cdh11*. Down-regulation of these genes indicates that the GATA1/FLI1 expressing cells are moving away from a fibroblast-specific gene program.

The remaining differentially regulated genes were also analysed and any genes which had previously been implicated in megakaryocyte differentiation are listed in Table 4.1 along with other MK genes of interest that did not show any change. Microarray data analysis revealed that after four and seven months of stable expression of GATA1 and FLI1, a number of MK-specific genes were up-regulated compared to control cells (Table 4.1). While *Gp5* and *CD41* expressions were unchanged in the GATA1/FLI1 cells, consistent with our previous real time RT-PCR analysis (Figure 4.3), *Gp9*, *CD61* and *F2rl2* all increased their expression by more than 1.5 fold at four months, with further up-regulation observed by seven months. Although the expression of genes such as *Pf4*, *F2r* and *Thbs1* were shown to have a fold change less than 1.5 fold at four months, *F2r* and *Thbs1* achieved a 1.7 fold up-regulation by seven months while *Pf4* increased by 4 fold at this time point. Similarly, even though *Ppbp* was seen to be down-regulated at an earlier time point, its expression increased by culturing the cells longer. In summary, many MK markers were up-regulated after seven months of stable expression of GATA1 and FLI1.

Table 4.1. *Affymetrix microarray analysis of megakaryocytic markers in fibroblasts over-expressing GATA1 and FLI1.* Samples were prepared from MEF cells transduced with either empty vectors or Gata1 and Fli1 at four months and seven months post-transduction. Microarray analysis was performed using Partek Genomics Suite 6.6 and resultant fold changes are shown. Genes that have a fold change up- or down-regulation of ≥ 1.50 in the Gata1+Fli1 cells compared to the empty vector control are considered differentially expressed.

Gene symbol	Gene description	Gata1+Fli1 compared to Control	
		Fold change (4 months)	Fold change (7 months)
Pf4	Platelet factor 4	1.34	4.27
Gp9	Glycoprotein 9 (platelet)	2.02	3.62
Itgb3 (CD61)	Integrin beta 3	1.53	2.31
F2rl2	Coagulation factor II (thrombin) receptor-like 2	1.67	1.73
F2r	Coagulation factor II (thrombin) receptor	1.43	1.72
Thbs1	Thrombospondin 1	1.29	1.65
Ppbp	Pro-platelet basic protein	-2.11	1.24
Gp5	Glycoprotein 5 (platelet)	-1.01	-1.06
Itga2b (CD41b)	Integrin alpha 2b	-1.16	-1.14

To confirm our microarray analysis, we performed real time RT-PCR validation of the array data for the samples collected at seven months (Figure 4.6). We examined expression of *Gata1*, *Fli1* and the three most up-regulated MK genes, *Pf4*, *Gp9* and *CD61*. *Gata1* and *Fli1* were found by microarray analysis to have fold changes of 28.9 and 41.6, respectively. However, real time RT-PCR showed fold changes in the tens and hundreds of thousands for these genes (Figure 4.6A and Figure 4.6B). Although there is an agreement with the pattern of expression observed using microarray and real time RT-PCR for all the five genes tested, the fold change values produced by the two techniques differ in magnitude. The up-regulation of MK genes in GATA1 and FLI1 cells was much less by microarray than by real time RT-PCR

analysis. This highlights that microarray analysis can under represent the extent of a change in expression due to signals from highly abundant transcripts becoming saturated, thereby demonstrating the importance of validation of any genes of interest identified from microarrays.

In addition to analysing *Pf4* gene expression by real time RT-PCR at the seven month time point, we also examined protein expression by performing enzyme-linked immunosorbent assay (ELISA) on cell culture supernatant (Figure 4.6D). In this assay, culture medium was collected after culturing cells in medium for four days. The medium was centrifuged to remove cell debris and the supernatant collected and applied to microplate wells pre-coated with monoclonal antibodies specific for mouse PF4. Any PF4 protein present in the supernatant is bound by the antibody and can be detected by the addition of an enzyme-linked PF4 antibody capable of driving a chromogenic reaction. Figure 4.6D shows that while no PF4 is detected in the media of control cells, the GATA1 and FLI1 cell lines are able to secrete high amounts of PF4 (~1 ng/mL), as indicated by the yellow-coloured wells.

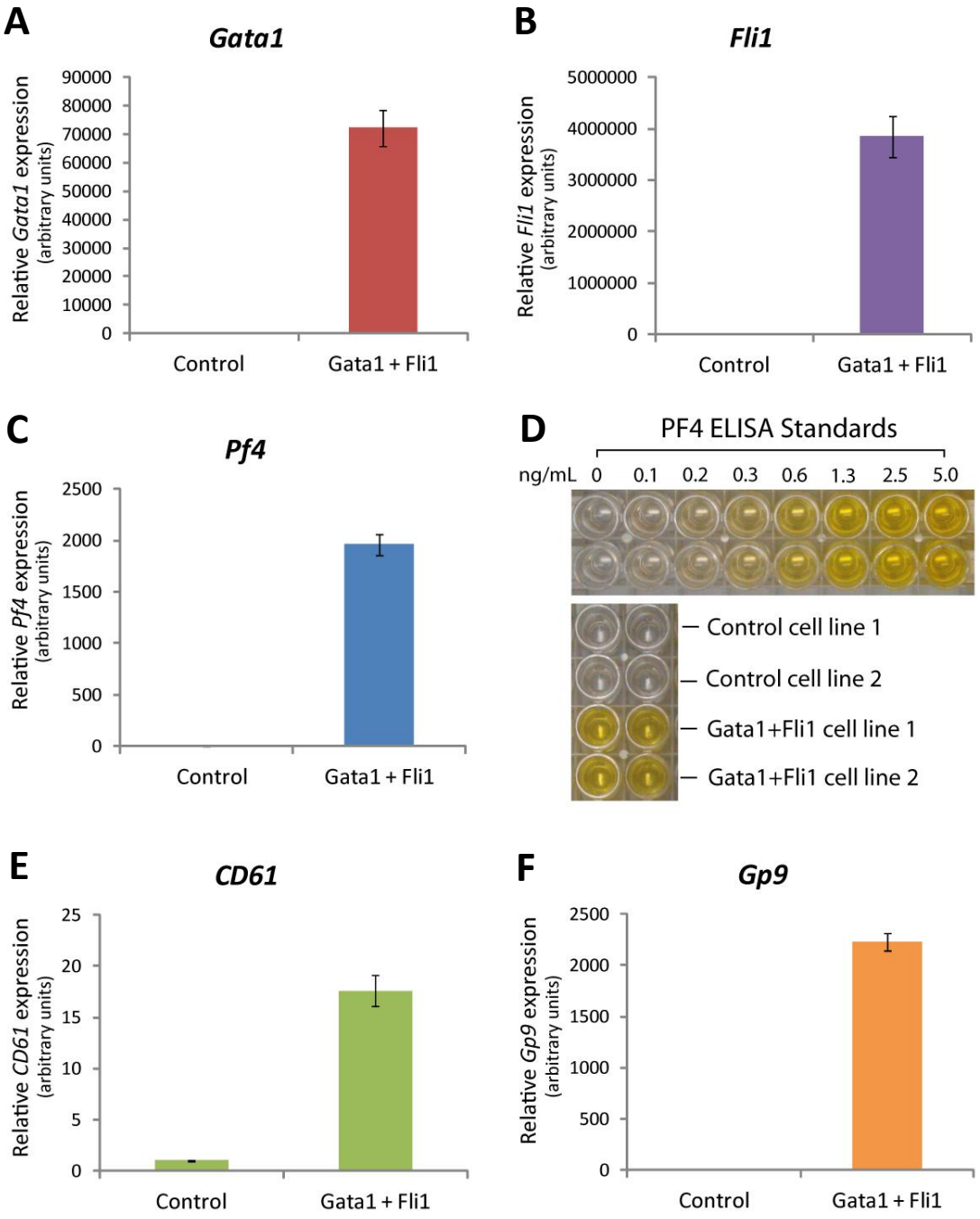


Figure 4.6. *Elevated megakaryocyte gene expression following ectopic expression of GATA1 and FLI1.* Real time RT-PCR validation of *Gata1* and *Fli1* over-expression, and various megakaryocyte-specific genes. Samples were collected from stable cell lines cultured for 7 months following transduction. Control samples represent cell lines co-transduced with empty pMSCVhyg and pMSCVpuro vectors. RNA signals were normalised against *18S* rRNA levels and normalised again to control sample, which was set as 1. Shown are expression levels of **(A)** *Gata1*, **(B)** *Fli1*, **(C)** *Pf4*, **(E)** *CD61* and **(F)** *Gp9*. Error bars represent standard error of the mean (n=2). **(D)** ELISA performed on cell culture media from control and

Gata1+Fli1 MEFs. It shows that the Gata1+Fli1 cell lines express PF4 protein as indicated by yellow-coloured wells. Concentrations of PF4 standards are indicated above the wells.

Although, forced expression of GATA1 and FLI1 in fibroblasts was sufficient for up-regulation of a number of MK-specific genes, we were unable to detect any morphological changes consistent with a MK phenotype, despite culturing cells for extended periods up to seven months (data not shown). These results suggested to us that additional or other factors would be required to drive a proper reprogramming process.

Gene expression is heavily influenced by chromatin structure (Grewal & Moazed, 2003; Kadonaga, 2004; Strahl & Allis, 2000; Thiel et al, 2004), with tightly packed heterochromatin restricting transcription factor access to regulatory regions. To address the possibility that reprogramming potential is limited by chromatin state, we tried culturing the stable cell lines in media containing trichostatin A (TSA), a potent histone deacetylase (HDAC) inhibitor, which remodels chromatin to an open structure by increasing histone acetylation (Toth et al, 2004; Yoshida et al, 1995; Yoshida et al, 1990). For example, TSA can revert mouse embryoid bodies to an undifferentiated state (Lee et al, 2004) and help promote cardiomyocyte differentiation of embryonic stem cells (Kawamura et al, 2005). Unfortunately, in our hands we found that sustained addition of TSA to the fibroblast culture medium resulted in high levels of cell death, even at the lowest recommended working concentrations (data not shown). Given this observation, we concentrated on the alternative strategy of assessing other candidate factors with the potential to boost the reprogramming of fibroblasts to MKs.

4.3 Screening for additional MK reprogramming factors

Encouraged by the partial reprogramming achieved by forced expression of GATA1 and FLI1, and the observation that typically at least three transcription factors are required to fully direct cell-specific gene programs capable of complete cell conversion (Huang et al, 2011; Ieda et al, 2010; Vierbuchen et al, 2010), we decided to test a number of further candidate factors. We continued to examine transcription factors and co-factors known to have roles in megakaryopoiesis, such as RUNX1 (Elagib et al, 2003; Ichikawa et al, 2004), NFE2 (a heterodimer composed of p45 and p18 MAF subunits) (Lecine et al, 1998; Onodera et al, 2000; Shivdasani et al, 1995b), MEF2C (Gekas et al, 2009), MEIS1 (Azcoitia et al, 2005; Hisa et al, 2004; Okada et al, 2003) and FOG1 (Tsang et al, 1998). These factors exhibit a clear knockout phenotype where deletion of the gene results in severe abnormality in megakaryopoiesis and associated thrombocytopenia. We also included the co-factor FOG2 (Jeanpierre et al, 2008; Tevosian et al, 1999) and transcription factors ETS1 (Lemarchandel et al, 1993; Minami et al, 1998) and GABPα (Pang et al, 2006) as they have been shown to interact with GATA1 and/or activate the promoters of MK-specific genes such as *Pf4*, *CD41* or *c-Mpl*. Our analysis also explored a number of additional megakaryocytic transcription factors, FHL2, RFX5, MXD1 and E2F3, which show increased expression in MK cells (Fuhrken et al, 2008).

We also considered candidate factors that are expressed in early haematopoietic and MK progenitors such as TAL1 (Robb et al, 1995; Shivdasani et al, 1995a) and GATA2 (Huang et al, 2009; Tsai et al, 1994), reasoning that such factors might have potency in initiating and maintaining a MK-specific gene expression program. We

also hoped that their inclusion might promote the generation of a pool of proliferating MK progenitor cells rather than a more limited number of terminally differentiated cells. In line with this hypothesis, we also examined the pluripotency factor OCT4, which has previously been reported to directly convert fibroblasts to haematopoietic progenitors capable of subsequent megakaryocytic differentiation (Szabo et al, 2010).

Finally, candidate genes were not limited to transcriptional activators and co-activators of MK differentiation. The *c-Mpl* gene which encodes the thrombopoietin (TPO) receptor (see Chapter 1.4.1) was also included as TPO is the major cytokine that regulates MK differentiation (de Sauvage et al, 1994; Kaushansky et al, 1994; Wendling et al, 1994).

In summary, we identified 17 candidate regulators with potential to enhance GATA1 and FLI1 reprogramming of fibroblasts toward the MK lineage. Each candidate factor was lentivirally transduced into cells that were already stably over-expressing GATA1 and FLI1. The pLV411 expression plasmid (Skalamera et al, 2012; Skalamera et al, 2011) used in these experiments placed the open reading frame of each candidate gene upstream of an internal ribosome entry site (IRES) followed by a green fluorescent protein (GFP) reporter, allowing us to assess and track expression by fluorescence.

To evaluate the effect of adding a third additional factor, we analysed expression of the MK markers *Pf4* and *CD61*, comparing transcript levels with cells expressing GATA1 and FLI1 alone. Initial real time RT-PCR analysis performed on cells

collected four days post-transduction of the additional gene indicated that none of the three-factor combinations expressed higher levels of *Pf4* and *CD61* compared to GATA1 and FLI1 alone (data not shown).

Since we previously found that sustained over-expression of GATA1 and FLI1 resulted in increased up-regulation of MK markers (Figure 4.4), we decided to culture cells expressing three factors for a longer period of time. Accordingly, transcript levels were determined for *Pf4* and *CD61* following a further two weeks in culture (Figure 4.7). Following prolonged over-expression, we found that a number of factors were able to modestly boost *Pf4* expression compared to cells expressing GATA1 and FLI1 alone, although this was associated with little effect on *CD61*. Additionally, the transcription factor TAL1, although appearing not to increase *Pf4*, was able to drive significant up-regulation of *CD61* expression, highlighting it for further investigation.

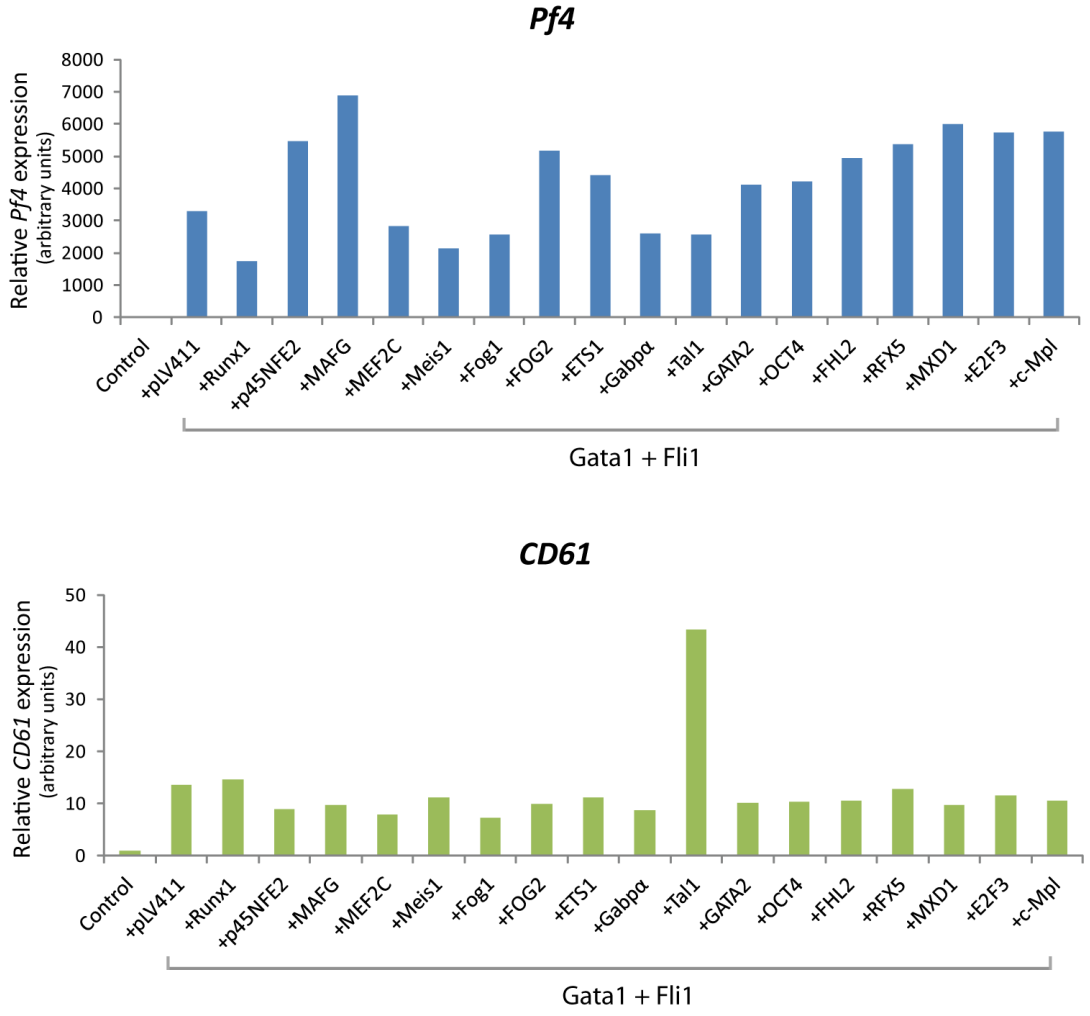


Figure 4.7. *Screening for a third factor that can promote MK differentiation in fibroblasts.* Candidate factors were transduced into cells expressing GATA1 and FLI1, and expression of *Pf4* and *CD61* assessed by real time RT-PCR. Control sample represents a cell line co-transduced with empty vectors (pMSCVhyg, pMSCVpuro and pLV411). Transcript levels were normalised against 18S rRNA levels and then normalised again to the control sample, which was set as 1. Each bar represents n=1.

As cells expressing a third factor also expressed GFP, we were able to use fluorescence microscopy to determine the number of cells that maintained over-expression of this additional factor for the duration of the study. This analysis revealed an initial transduction efficiency of 80-90%; however, after three weeks of culture we found that only 5-10% of the cell population continued to express the

additional factor (Figure 4.8). This observation suggested that analysis of the total cell population may be masking successful reprogramming, and that analysis of individual cells would be required for a more accurate assessment.

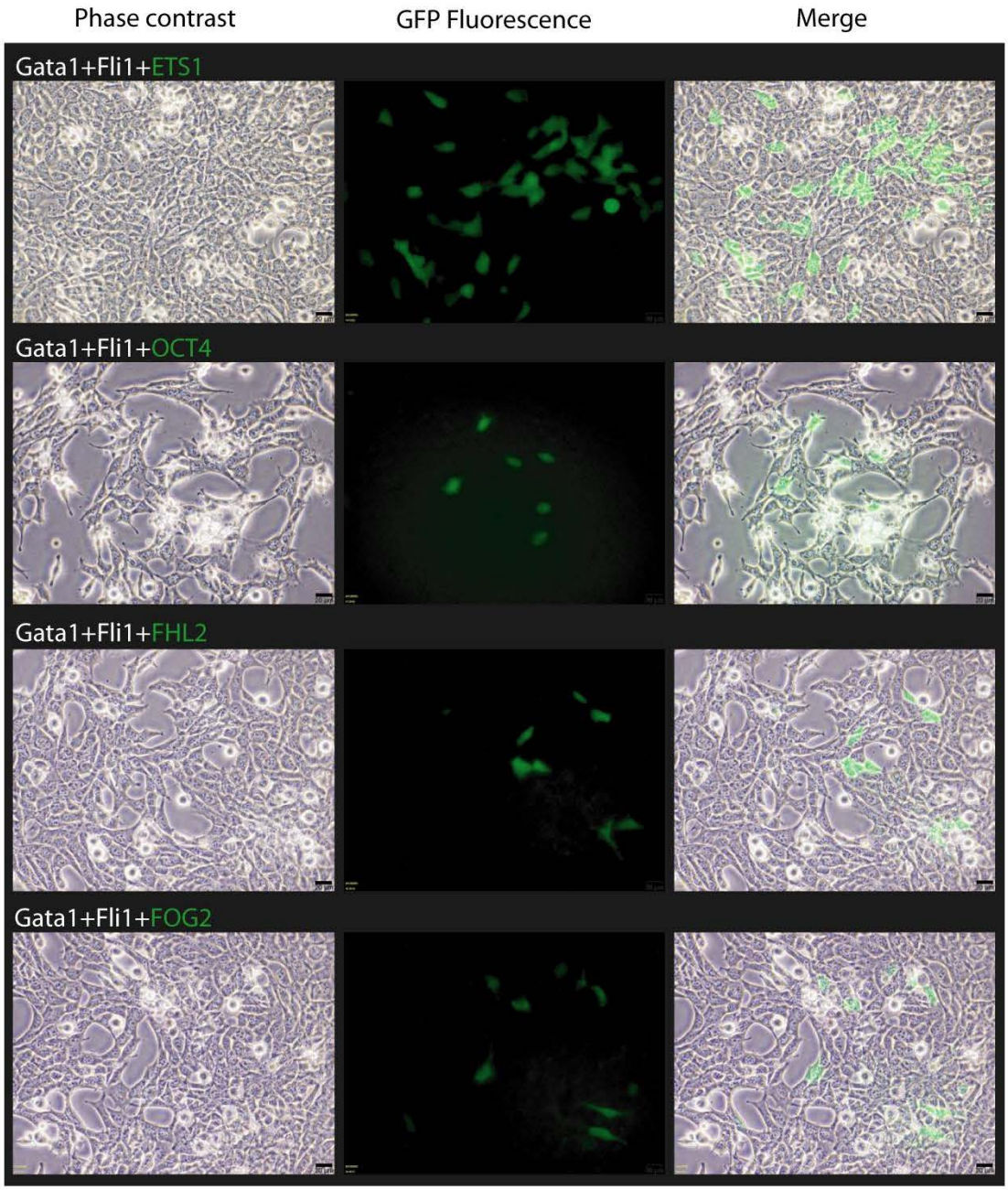
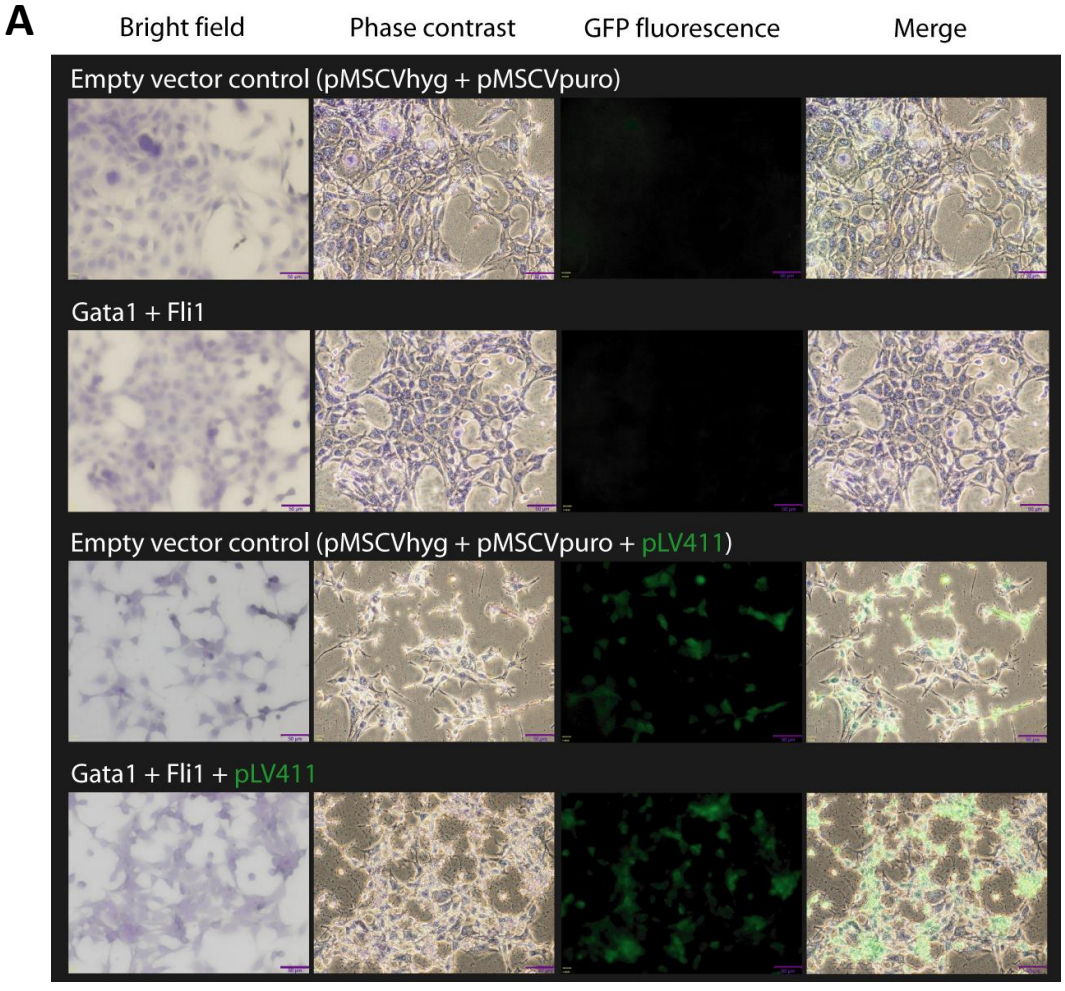


Figure 4.8. *Few cells continue to express an additional third factor three weeks after transduction.* MEF cells expressing GATA1 and FLI1 were transduced with lentiviral vectors

containing the additional factor and an IRES-driven GFP reporter, allowing expression to be monitored by GFP fluorescence. Scale bars represent 20 µm.

4.3.1 Expression of GATA1, FLI1 and TAL1 induces acetylcholinesterase activity in murine embryonic fibroblasts

Given the low efficiency of sustained expression of the third factor, we decided to assess reprogramming in single cells expressing GFP by an acetylcholinesterase (AChE) assay that is widely used to identify cells of the megakaryocytic lineage (Alford et al, 2010; Huang et al, 2009; Jackson, 1973; Matsumura-Takeda et al, 2007). In this assay, cells that express AChE, such as mouse MKs (Jackson, 1973), stain brown when treated with assay substrate whereas AChE-negative cells appear violet due to the prominence of the haematoxylin counterstain (Figure 4.9).



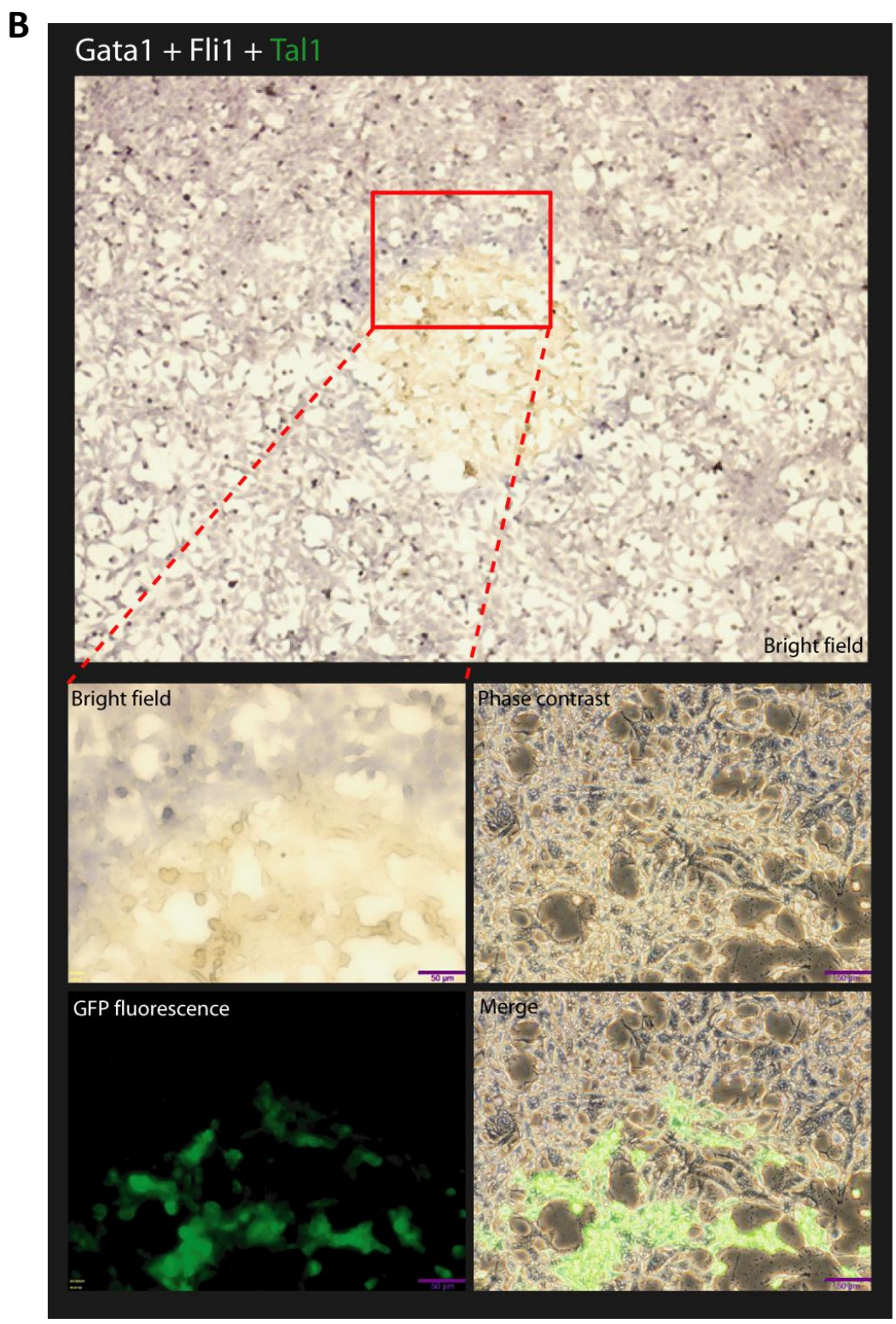


Figure 4.9. *Acetylcholinesterase (AChE) assay comparing MEFs over-expressing* ***(A)*** *GATA1/FLI1 and* ***(B)*** *GATA1/FLI1/TAL1.* In (B), MEFs already expressing GATA1 and FLI1 were transduced with *Tal1*, and cultured for three weeks prior to assaying for the presence of (AChE) activity. The boxed area in (B) is represented in more detail in the lower panels. AChE activity results in a brown stain, while AChE-negative cells appear violet due to the

haematoxylin counterstain. Cells were subjected to bright field microscopy to distinguish brown and violet staining, and GFP fluorescence microscopy to identify cells expressing the additional third factor. Scale bars represent 50 µm.

The absence of brown staining in cells that were transduced with empty vectors confirmed that MEFs do not express AChE (Figure 4.9A). We also determined that over-expression of GATA1 and FLI1 failed to induce AChE activity (Figure 4.9A). After careful observation of the seventeen additional factors, we found that co-expression of GATA1, FLI1 and TAL1 was capable of inducing AChE activity. The upper panel in Figure 4.9B depicts an area in cells transduced with these three factors where brown staining is evident. Detailed analysis of the boxed area revealed that these AChE-positive cells also express GFP, while violet cells negative for AChE lack GFP expression (Figure 4.9B, lower panel).

To confirm this observation, we cultured fibroblasts over-expressing GATA1, FLI1 and TAL1 in a collagen-based medium more typically used for optimal detection of MK progenitors in colony assays (Alford et al, 2010; Matsumura-Takeda et al, 2007). Bone marrow cells, which give rise to MK colonies, were included as a positive control in this assay. After eight days in culture, the fibroblasts and bone marrow cells were fixed and stained for AChE activity.

Again, we found that the combination of GATA1, FLI1 and TAL1 induced AChE activity in fibroblasts, producing brown-stained cells, while expression of GATA1 and FLI1 (denoted in Figure 4.10 as Gata1 + Fli1 + pLV411) did not produce any

brown-stained cells (Figure 4.10). The light brown colour of the AChE-positive cells in the GATA1+FLI1+TAL1 combination suggests that AChE activity in these cells is low compared to the dark red brown-stained cells in the bone marrow sample (Figure 4.10). This suggests again that although the addition of TAL1 further promotes the conversion of fibroblasts toward a MK phenotype, only partial reprogramming has been achieved.

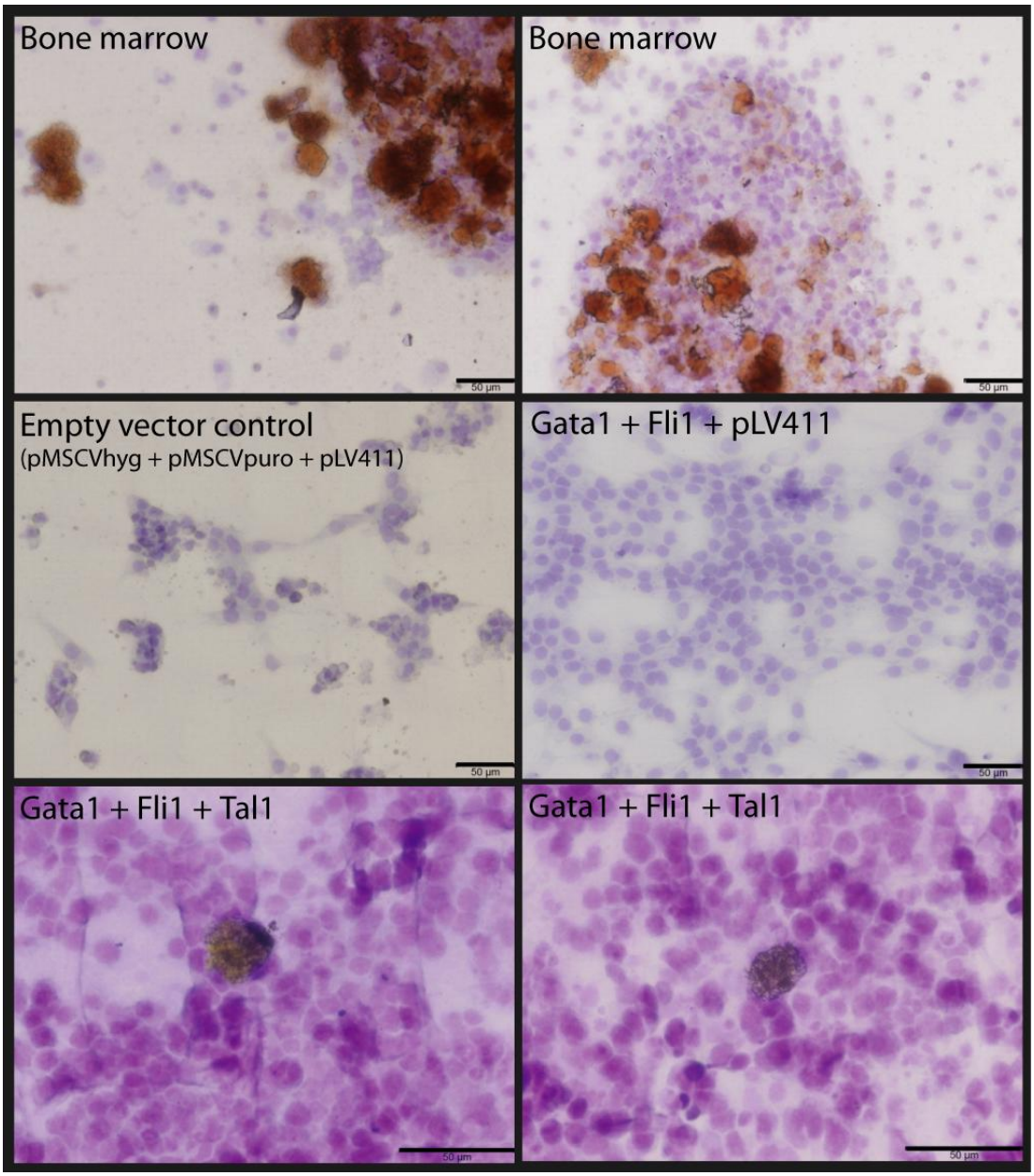


Figure 4.10. *Acetylcholinesterase (AChE) activity of fibroblasts and bone marrow cells cultured in collagen-based medium.* Bone marrow cells and mouse fibroblast transduced

with *Gata1*/*Fli1* and *Gata1*/*Fli1*/*Tal1* were cultured in a collagen-based medium optimised for detection of MK progenitors. After eight days in culture, cells were fixed and stained for AChE activity. AChE-positive cells appear brown while AChE-negative cells are violet. Scale bars represent 50 µm.

4.3.2 Forced expression of GATA1, FLI1 and TAL1 further enhances the expression of MK markers

To further address the low efficiency with which a third factor was expressed in the total cell population, we decided to use fluorescence-activated cell sorting (FACS) to purify GFP positive cells. Fibroblasts were transduced with *Gata1*, *Fli1* and *Tal1*, or empty vector controls, and cells expressing GFP were collected by FACS. RNA was then extracted from these cells and analysed by real time RT-PCR for *Pf4*, *CD61* and *Gp9* mRNA expression (Figure 4.11).

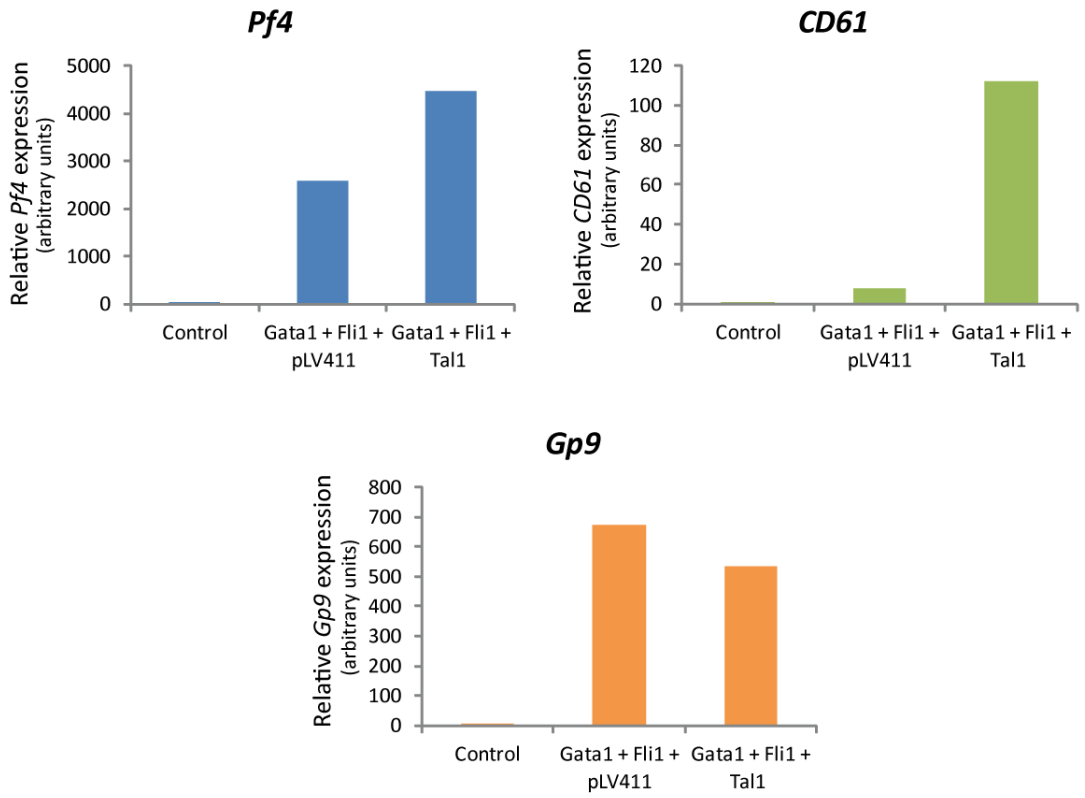


Figure 4.11. *Expression of MK markers is further up-regulated in Gata1+Fli1 cells in the presence of Tal1.* Control MEFs transduced with empty pMSCVhyg + pMSCVpuro + pLV411 vectors, and Gata1+Fli1 MEFs transduced with either empty pLV411 vector or pLV411-Tal1, were subjected to fluorescence-activated cell sorting (FACS) to select for GFP-positive cells. Samples were analysed by real time RT-PCR to determine *Pf4*, *CD61* and *Gp9* expression. mRNA levels were normalised against *18S* rRNA levels and again normalised to control sample, which was set as 1. Each bar represents n=1.

We found that addition of TAL1 to GATA1/FLI1 cells resulted in noticeably increased expression of *Pf4* and *CD61* transcripts, although *Gp9* levels remained similar to the doubly transduced cells (Figure 4.11), suggesting that TAL1 can further enhance the expression of MK markers. This observation, taken together with the induction of AChE activity in fibroblasts following forced expression of TAL1, GATA1 and FLI1, provides evidence that these three factors can achieve at least partial reprogramming of fibroblasts towards the megakaryocyte lineage.

4.4 A GATA1 isoform, GATA1 short, can substitute for GATA1 and perform better at reprogramming MK gene loci

To continue to refine our choice of reprogramming factors, we decided to investigate a GATA1 isoform, GATA1 short (GATA1s). This GATA1 variant lacks the N-terminal transcriptional activation domain, but retains the ability to bind DNA and interact with the GATA1 co-factor FOG1 (Wechsler et al, 2002). While GATA1 is important at certain stages of MK development, particularly in deciding the lineage choice of megakaryocyte-erythroid progenitors and also in terminal MK differentiation, it plays a negative role in MK progenitor proliferation (Shivdasani et al, 1997; Vyas et al, 1999). In contrast, GATA1s has been shown to drive proliferation of fetal MK progenitors (Li et al, 2005). We therefore reasoned that GATA1s may have greater potential in our aim of converting fibroblasts to proliferating MK cells.

In order to evaluate the potential of GATA1s in MK reprogramming, we began by comparing expression of MK markers in cells transduced with *Gata1s* and *Fli1* with cells expressing *Gata1* and *Fli1*. Real time RT-PCR analysis performed on cells three weeks post-transduction revealed that replacing GATA1 with GATA1s further increased the expression of *Pf4*, *CD61* and *Gp9* (Figure 4.12A), suggesting that the GATA1 mutant isoform is better able to direct MK-specific gene expression changes in fibroblasts. We also investigated whether GATA1s can synergise with FLI1 and TAL1 to further promote MK reprogramming by quantifying expression of MK marker genes in MEF cells stably expressing all three factors (Figure 4.12B). Real time RT-PCR revealed that forced expression of these three factors resulted in

further up-regulation of *CD61* and *Gp9*, compared to levels in cells expressing GATA1s and FLI1 alone. Taken together, these data suggest that GATA1s may be a preferred factor to GATA1 in selecting the combination of factors best able to direct MK reprogramming of fibroblasts.

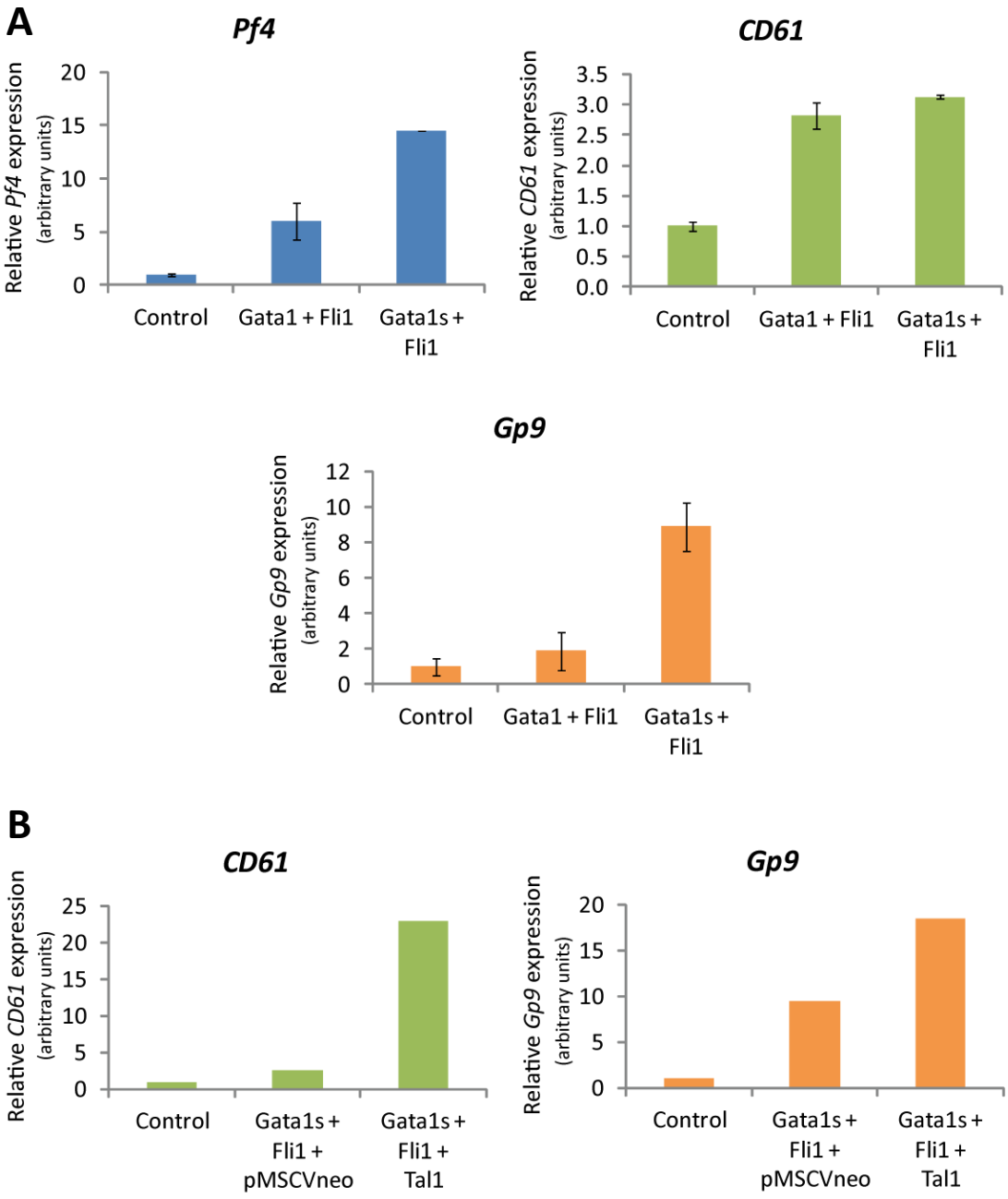


Figure 4.12. *Improved up-regulation of MK marker expression by the GATA1 mutant isoform, GATA1 short* (*GATA1s*). Expression of *Pf4*, *CD61* and *Gp9* was assessed by

quantitative real time RT-PCR in MEF cells three or four weeks post-transduction with **(A)** *Gata1s* and *Fli1*, and **(B)** *Gata1s, Fli1* and *Tal1*. Control cells in (A) were transduced with pMSCVhyg and pMSCVpuro vectors while control cells in (B) have an additional empty pMSCVneo vector. Transcript level for each gene has been normalised against *18S* rRNA and normalised again to control sample, which was set as 1. Error bars in (A) represent standard error of the mean (n= at least 2).

4.5 p45NFE2 and MAFG in MK differentiation of fibroblasts

During the course of our studies, Yumiko Matsubara and colleagues reported that a combination of three transcription factors: p45NFE2, MAFG and MAFK, can promote MK lineage redirection in mouse 3T3 and human dermal fibroblasts, as assessed by CD41 expression, DNA ploidy and cell morphology (Ono et al, 2012). NFE2 functions as a heterodimer consisting of a p45 subunit and a MAF protein (Andrews et al, 1993a; Andrews et al, 1993b; Igarashi et al, 1994; Toki et al, 1997), which is required for DNA-binding. NFE2 has been shown to be crucial in MK differentiation as mice lacking either of the subunits are defective in MK maturation and proplatelet formation (Lecine et al, 1998; Shivdasani et al, 1995b).

While Matsubara was able to report fibroblast conversion into MKs, the efficiency of reprogramming appeared low and this led us to investigate whether addition of our candidate factors, GATA1, GATA1s, FLI1 and TAL1, might enhance p45-MAF reprogramming.

We adopted the published cell culture method used to drive reprogramming of fibroblasts (Ono et al, 2012) and tested combinations of p45NFE2, MAFG, GATA1,

GATA1s, FLI1 and TAL1 in murine 3T3 fibroblasts. This protocol involves culturing cells in a megakaryocyte lineage induction (MKLI) medium, a serum-free medium that contains the cytokine TPO, which has previously been shown to drive MK reprogramming in some but not all cell lines (Matsubara et al, 2010), again suggesting that additional factors are required for efficient reprogramming.

The reprogramming potential of the various combinations of factors was assessed by observing cell morphology following transduction and culture in MKLI medium (Figure 4.13). After three days in culture, both untransduced and transduced 3T3s appeared as loosely aggregated spheres of floating cells (black arrows). This morphology, suggestive of the start of MK reprogramming, is similar to that previously reported for 3T3-L1 following three days culture in MKLI medium (Matsubara et al, 2010). After eight days, in two independent experiments, spheres of floating cells were still observed in both untransduced and transduced 3T3 cells, with larger spheres observed in some of the combinations (white triangles). This contradicted the previous report (Matsubara et al, 2010), which stated that untransduced 3T3 cells die after four days of culture in MKLI medium. Additionally, the increase in size of the floating spheres at day 8 is not consistent in the two experiments. In our hands, we found that a proportion of both untransduced and transduced 3T3 cells underwent similar morphological changes suggesting some degree of reprogramming that seemed to be independent of the factors (Figure 4.13).

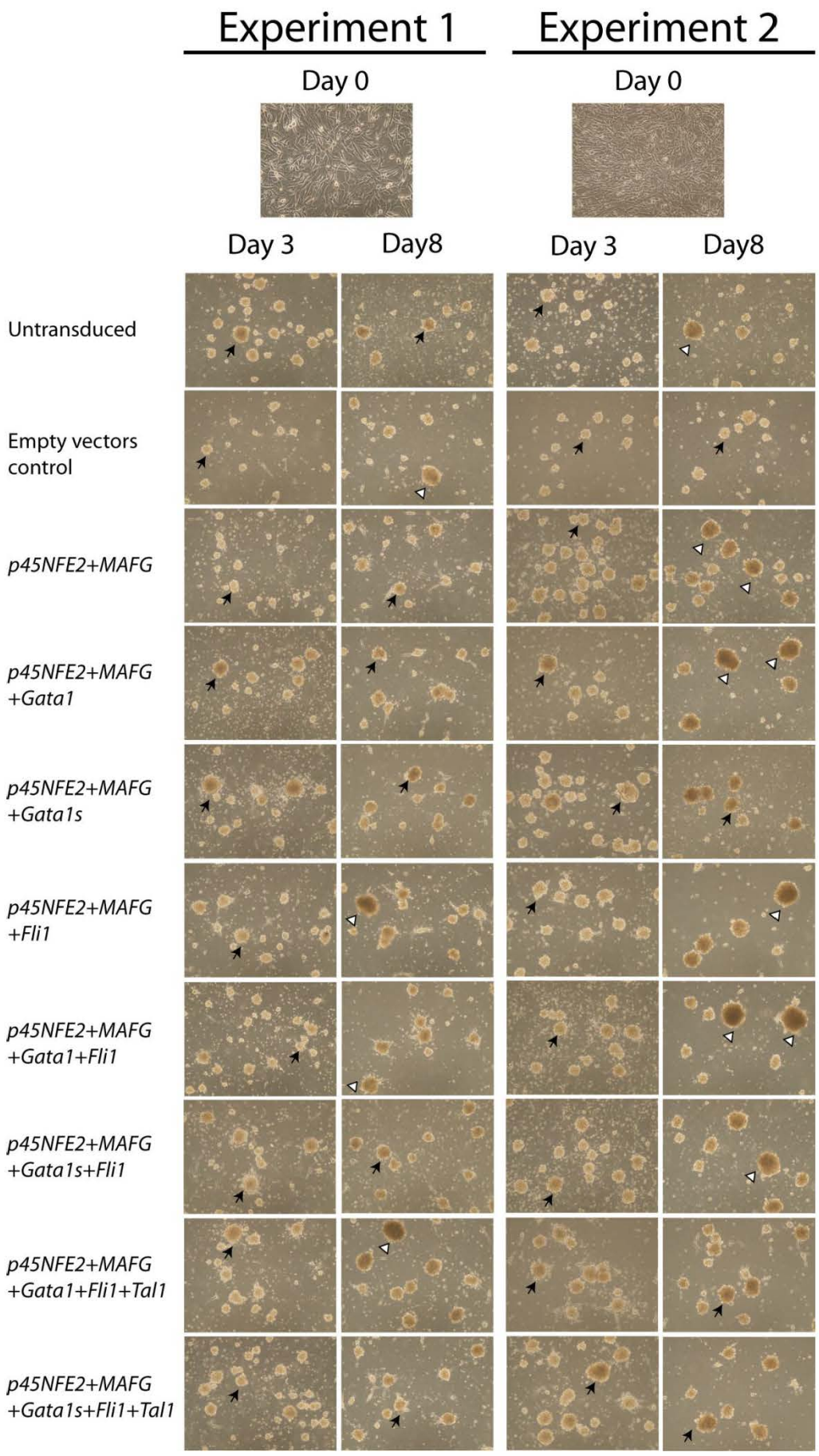
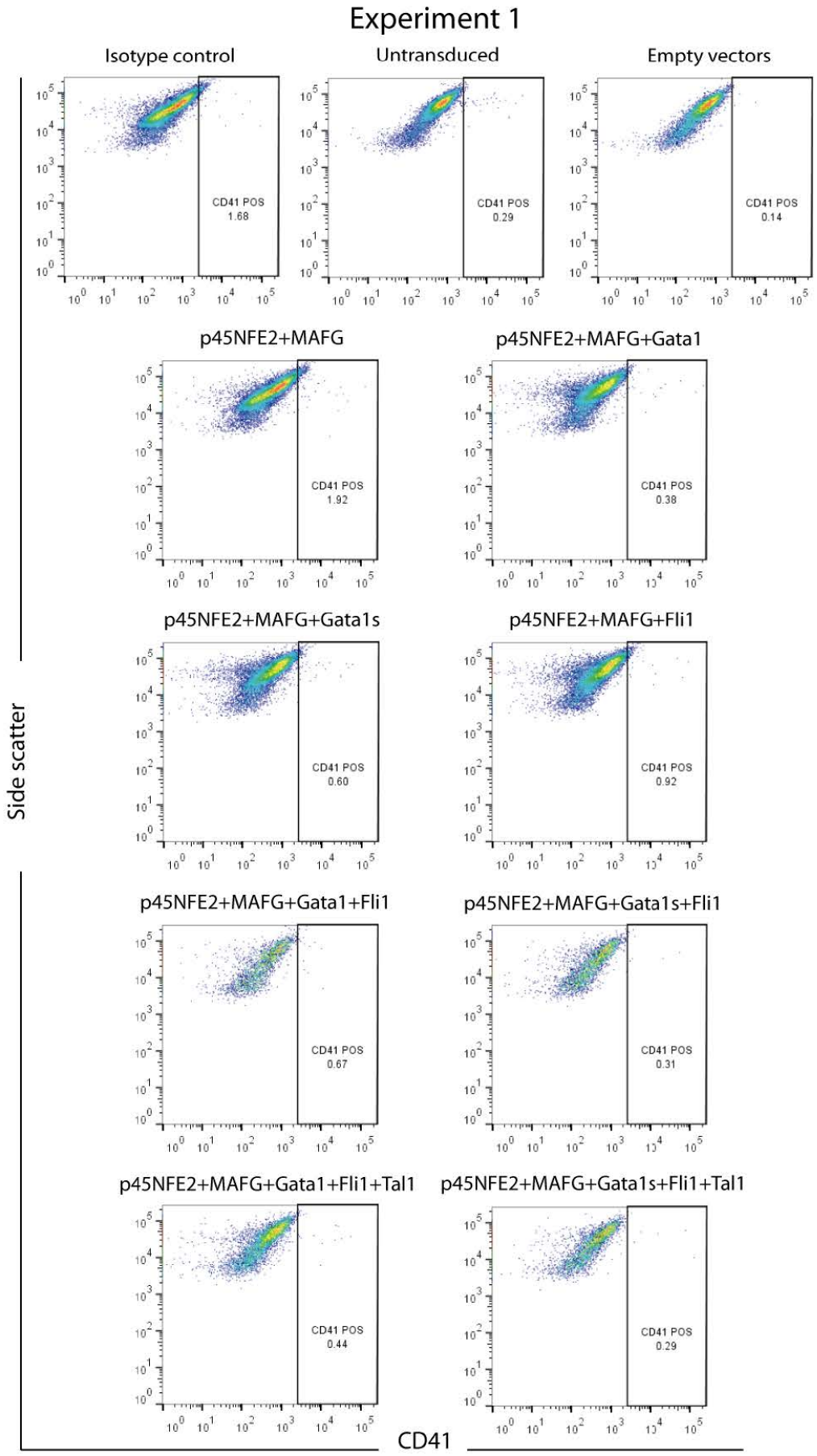

Experiment 1
Experiment 2
Day 0
Day 0
Day 3
Day8
Day 3
Day8
Untransduced
Empty vectors control
p45NFE2+MAFG
p45NFE2+MAFG +Gata1
p45NFE2+MAFG +Gata1s
p45NFE2+MAFG +Fli1
p45NFE2+MAFG +Gata1+Fli1
p45NFE2+MAFG +Gata1s+Fli1
p45NFE2+MAFG +Gata1+Fli1+Tal1
p45NFE2+MAFG +Gata1s+Fli1+Tal1

Figure 4.13. *Morphology of 3T3 fibroblasts cultured in megakaryocyte lineage induction (MKLI) medium.* 3T3 fibroblasts were transduced with combinations of *p45NFE2*, *MAFG*, *Gata1*, *Gata1s*, *Fli1* and *Tal1*, and cultured in MKLI medium for eight days. Shown are micrographs taken at Day 3 and Day 8 of culture for two independent experiments. Black arrows indicate loosely aggregated spheres of floating cells whereas white triangles show larger spheres seen at Day 8 in culture.

After eight days in culture, flow cytometry was used to examine cells for expression of the MK cell surface marker CD41 (Figure 4.14). In two independent experiments, we were unable to detect increased CD41 expression above background levels observed for control cells. In particular, in contrast to the published report (Ono et al, 2012), we found that 3T3 cells transduced with *p45NFE2* and *MAFG* failed to express CD41 at levels higher than control untransduced cells. Furthermore, we failed to observe any effects on CD41 levels upon forced expression of additional factors. Overall, our attempts to reproduce the published data using our reagents were unsuccessful and therefore the ease of using p45/MAFG as a general reprogramming strategy remains in question.


Experiment 1
Isotype control
CD41 POS
1.68
Untransduced
CD41 POS
0.29
Empty vectors
CD41 POS
0.14
p45NFE2+MAFG
CD41 POS
1.92
p45NFE2+MAFG+Gata1
CD41 POS
0.38
p45NFE2+MAFG+Gata1s
CD41 POS
0.60
p45NFE2+MAFG+Fli1
CD41 POS
0.92
p45NFE2+MAFG+Gata1+Fli1
CD41 POS
0.67
p45NFE2+MAFG+Gata1s+Fli1
CD41 POS
0.31
p45NFE2+MAFG+Gata1+Fli1+Tal1
CD41 POS
0.44
p45NFE2+MAFG+Gata1s+Fli1+Tal1
CD41 POS
0.29
Side scatter
CD41

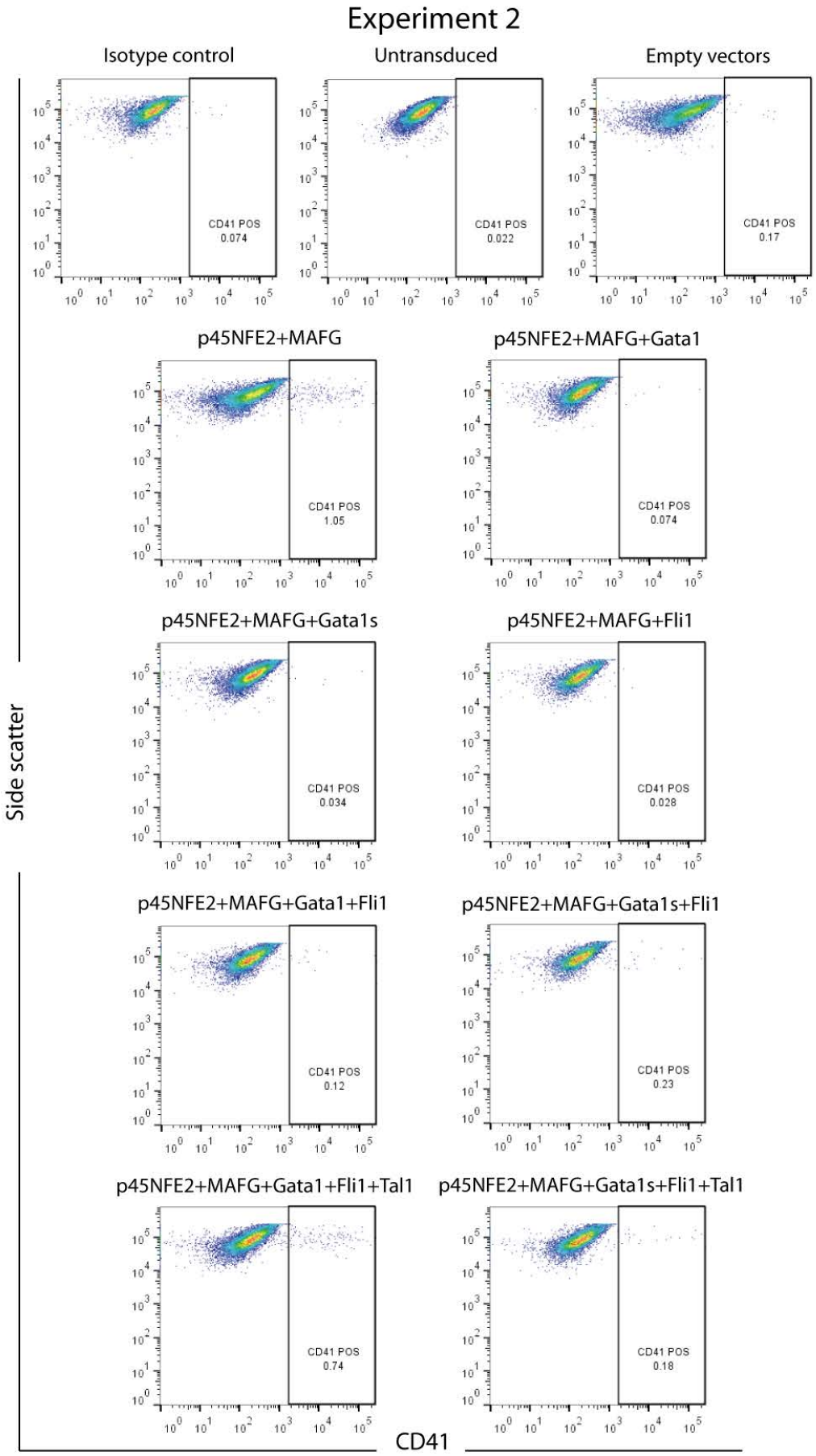

Experiment 2
Isotype control
CD41 POS
0.074
Untransduced
CD41 POS
0.022
Empty vectors
CD41 POS
0.17
p45NFE2+MAFG
CD41 POS
1.05
p45NFE2+MAFG+Gata1
CD41 POS
0.074
p45NFE2+MAFG+Gata1s
CD41 POS
0.034
p45NFE2+MAFG+Fli1
CD41 POS
0.028
p45NFE2+MAFG+Gata1+Fli1
CD41 POS
0.12
p45NFE2+MAFG+Gata1s+Fli1
CD41 POS
0.23
p45NFE2+MAFG+Gata1+Fli1+Tal1
CD41 POS
0.74
p45NFE2+MAFG+Gata1s+Fli1+Tal1
CD41 POS
0.18
Side scatter
CD41

Figure 4.14. *CD41 expression in 3T3 fibroblasts transduced with various combinations of reprogramming factors.* 3T3 cells were transduced with control empty vectors and combinations of plasmids expressing p45NFE2, MAFG, GATA1, GATA1s, FLI1 and TAL1, and cultured in megakaryocyte lineage induction medium. After eight days, cells were analysed by flow cytometry for expression of CD41, a megakaryocyte-specific cell surface receptor. CD41 expression is presented in the x-axis. The percentage of CD41-positive (CD41 POS) cells for each sample is indicated. Shown are data from two independent experiments.

4.6 Discussion

In this chapter, our aim was to determine a set of transcriptional regulators that can convert fibroblast cells towards the MK lineage. Previous studies have shown that combinatorial expression of multiple transcription factors and not just a single factor is required to convert fibroblasts to neuronal (Vierbuchen et al, 2010), hepatic (Huang et al, 2011) and cardiac (Fu et al, 2013; Ieda et al, 2010) lineages. We found that ectopic expression of transcriptional activators, GATA1 and FLI1, can induce MK gene expression changes in murine fibroblasts. Previously these factors had been demonstrated to activate expression of MK genes such as *Gp9* and *Gp1bα* in human K562 erythroleukaemia cells (Eisbacher et al, 2003). Consistent with this, we have now shown that forced expression of these factors can drive up-regulation of a number of megakaryocytic markers including *Gp9, Pf4* and *CD61*. Real time RT-PCR and microarray analyses on cells stably expressing GATA1 and FLI1 have shown that reprogramming of MK loci appears dependent upon prolonged over-expression of both factors, resulting in higher gene expression changes over time. This may be due to the gradual conversion of tightly condensed heterochromatic regulatory regions to a more open structure within the cell population, facilitated by the recruitment of histone modifying enzymes and/or chromatin remodelers (Blobel

et al, 1998; Utley et al, 1998) by the two transcription factors, although the very long term nature of the effect was somewhat unexpected.

In an attempt to further promote chromatin remodelling at MK loci, we supplemented the culture medium with the HDAC inhibitor TSA (Yoshida et al, 1995; Yoshida et al, 1990). TSA treatment can decrease the expression of histone deacetylases that cause chromatin condensation, and increase expression of histone acetylases resulting in open chromatin structure permissible to transcriptional regulation and hence gene activation (Dey & Evans, 2011; Kang & Roh, 2011). It has been shown that addition of TSA promoted cardiac differentiation of embryonic stem cells due to activation of the master regulator of cardiogenesis, GATA4 (Kawamura et al, 2005). However, we found that the presence of this drug in the medium led to high levels of cell death and previous studies have indicated that HDAC inhibitors can induce apoptosis (Schnichels et al, 2012; Toth et al, 2004). However, the use of other chromatin remodeler inhibitors is worth considering in future studies. DNA methyltransferase inhibitors such as 5-Azacytidine and 5-Aza-2′-deoxycitidine have been shown to reactivate silenced genes by inducing demethylation (Beltran et al, 2008; Christman, 2002; Huangfu et al, 2008a). Moreover, other HDAC inhibitors, which include suberoylanilide hydroxamic acid (SAHA) and valproic acid (VPA), were found to greatly improve reprogramming efficiency of fibroblasts to pluripotent stem cells, allowing conversion of cells with only two transcription factors instead of the original four factor combination (Huangfu et al, 2008a; Huangfu et al, 2008b).

Despite the up-regulation of some MK-specific genes, we found that the combination of GATA1 and FLI1 achieved only partial reprogramming, with expression at other MK loci remaining unchanged and a lack of morphological changes in cells. Moreover, the microarray studies clearly indicated that we were seeing numerous sets of genes changing rather than a systematic reprogramming towards the megakaryocyte lineage. Accordingly, we investigated the use of additional transcriptional regulators in further refining lineage redirection. By screening for factors that can aid in MK conversion of fibroblasts using expression analysis and a MK-specific acetylcholinesterase (AChE) assay, we identified the early haematopoietic transcription factor TAL1 as a third factor capable of synergising with the other two factors to influence MK differentiation. TAL1 was previously demonstrated to be a critical transcription factor in haematopoiesis, absolutely required for the generation of early megakaryocytic progenitor cells (Hall et al, 2003). Hence, its inclusion may aid in the production of a pool of MK progenitors that may subsequently differentiate into platelet-producing cells.

We also investigated the GATA1 truncated mutant isoform, GATA1 short (GATA1s) and found that in combination with FLI1 alone, and both FLI1 and TAL1, this factor is capable of driving expression of MK-specific genes to levels higher than those seen in combinations containing wildtype GATA1. Additionally, this short GATA1 variant seems to be a better candidate factor for MK reprogramming as it has an added advantage of driving MK progenitor proliferation (Li et al, 2005) rather than promoting terminal differentiation.

A recent report that p45NFE2 and MAF proteins appear capable of directing 3T3 fibroblast conversion to MKs (Ono et al, 2012) led us to investigate these factors in combination with the transcriptional regulators that we found to have reprogramming potential. However, in two independent experiments, we were unable to reproduce the results of this work using our reagents, as defined by an up-regulation of CD41 expression (Ono et al, 2012). We noted that the authors reported noticeable up-regulation of CD41 in the presence of p45NFE2 and MAFG, whereas we found levels remained at background. The reasons for this are currently unclear although there are a few minor differences between the two methodologies that we have considered. Firstly, in our analysis of p45/MAF factors, we focused on MAFG, the essential MAF protein for megakaryopoiesis, as demonstrated by *Mafg*-null mice, which exhibit impaired proplatelet formation and altered MK gene expression, in contrast to *Mafk*-null mice which appear normal (Kotkow & Orkin, 1996; Motohashi et al, 2000; Onodera et al, 2000). It is therefore formally possible that inclusion of MAFK in our reprogramming strategy might result in increased CD41 expression. We also noted that the reported reprogramming was driven by transfection of mouse *p45Nfe2* and *Mafg*, whereas we used human genes. However, these human and mouse proteins are highly homologous (Chan et al, 1993) with an amino acid (aa) identity of 89% (331/373 aa) and 99% (160/162 aa), respectively. Therefore, it is uncertain that this would explain our inability to reproduce the published reprogramming data. Nevertheless, in this thesis, we have identified combinations of three alternative transcription factors (GATA1/GATA1s, FLI1 and TAL1) that do appear capable of initiating and driving partial fibroblast to MK reprogramming.

In order to further drive the reprogramming process initiated by our set of transcription factors, it may be worthwhile to consider the addition of a combination of cytokines involved in megakaryopoiesis to the culture medium of transduced MEFs. In addition to regulation by transcription factors, cytokine signalling also plays an important role in megakaryocyte differentiation, as discussed in Chapter 1 Section 1.4.1. Cytokines such as IL-3, GM-CSF and especially TPO, can stimulate proliferation and development of MKs, and have been used in the generation of MK-like cells from human ESCs and 3T3-L1 cells (Matsubara et al, 2010; Pick et al, 2013). Moreover, TPO addition to the culture medium of HSCs has been reported to greatly enrich the production of MKs (Bruno et al, 2003; De Bruyn et al, 2005; Shim et al, 2004), whereas its absence in *TPO*-null mice significantly reduced MK numbers (Bunting et al, 1997). Although TPO was used in the medium of transduced 3T3 fibroblasts, these cells were cultured for a relatively short period of time and this may have limited the effect of aiding the reprogramming process. Longer term culture in the presence of TPO and other cytokine combinations may prove more effective in driving lineage conversion. A further alternative strategy might be to co-culture transduced MEFs with bone marrow stroma cell lines such as OP9. These cells provide a rich source of haematopoietic cytokines and are thus used as feeder cells for the differentiation of ESCs and iPSCs into haematopoietic cells and eventually into MKs (Fujimoto et al, 2003; Gaur et al, 2006; Niwa et al, 2009).

CHAPTER 5 – SUMMARY AND CONCLUSIONS

5.1 Summary

In this thesis, we explored two projects to better understand the basic science and potential applications of gene regulatory proteins. The first examined a novel mechanism of sequence-specific DNA binding used by the oncogenic transcription factor ZNF217 to interact with target genes. The second investigated how forced ectopic expression of transcription factors can direct lineage reprogramming, focusing on the conversion of fibroblasts to megakaryocytes.

We found that the eight-zinc finger protein ZNF217 binds DNA via only its sixth and seventh fingers, F6-7, in contrast to the more usual tandem array of three or more fingers. Using comprehensive alanine scanning mutagenesis on this two-finger domain, we were able to determine the amino acid residues essential for DNA recognition by ZNF217. We found that mutation of R481 and Y485 of finger 6, and Q510, T512 and Y516 of finger 7 severely affected DNA binding by ZNF217. In support of this, the recent solution of the crystal structure of ZNF217 bound to DNA has confirmed the importance of these amino acids, by demonstrating that they make direct contact with specific nucleotide bases within the DNA recognition sequence (Vandevenne et al, 2013). As expected, we also found that mutation of both zinc-ligating residues, responsible for the structural integrity of the finger domain, and TGEKP linker residues, involved in stabilising the protein-DNA interaction, abolished the ability of ZNF217 to bind DNA.

Overall, the pattern of binding of ZNF217 F6-7 was found to be somewhat unique relative to the canonical mode of zinc finger-DNA interaction. In a typical interaction, the helical residues at position -1, +2, +3 and +6 (relative to the start of the alpha helix) of each zinc finger make specific contact with three to four bases of DNA. The residue at helical position +6 binds the first base recognised by a zinc finger on one strand, the +3 residue contacts the second base, and the -1 residue binds the third base, while the helical residue at position +2 contacts the fourth base on the other strand of DNA. In a ZNF217-DNA complex, only two canonical interactions are observed through helical residues at position +2 (T512) and +6 (Y516) of finger 7, highlighting the unusual binding that is made by this zinc finger transcription factor.

Understanding ZNF217's DNA-binding mechanism provides insight into the probable binding mechanism adopted by other related two-finger binders such as BCL11A, ZNF219 and ZNF536, as discussed in more detail in Chapter 3. Additionally, determining the mode of interaction of ZNF217 with DNA will help inform the design and development of artificial factors capable of binding and modulating the activity of ZNF217 target genes. The engineering of such factors is an exciting and important therapeutic prospect, given that ZNF217 is an oncogene (Peiro et al, 2002; Quinlan et al, 2007; Rooney et al, 2004; Yaswen & Stampfer, 2002), which regulates genes implicated in cancer progression, such as plakophilin 2 and keratin 18 (Alam et al, 2011; Demirag et al, 2011; Krig et al, 2007). Indeed, previous studies have shown the potential of utilising artificial transcription factors (ATFs) to regulate the expression of clinically relevant genes (Gommans et al, 2005; Sera, 2009). Zinc finger-based ATFs developed to target the *VEGF-A* gene have

been shown to increase and decrease its expression in *in vitro* and *in vivo* models (Liu et al, 2001; Rebar et al, 2002; Snowden et al, 2003). Controlling *VEGF-A* expression has therapeutic benefits in stimulating neovascularisation to treat ischemia and aid in wound healing, while its repression has been shown to prevent angiogenesis associated with tumour growth, rheumatoid arthritis and diabetic retinopathy (Sera, 2009). It has been reported that an ATF that repressed *VEGF-A* in a human glioma xenograft mouse model resulted in enhanced survival rates of athymic mice (Kang et al, 2008). Additionally, zinc finger ATFs have also been reported to stimulate the expression of *Bax* (Falke et al, 2003) and *Maspin* (Beltran et al, 2007) genes, which have been considered as targets for cancer therapy. *Bax* is a pro-apoptotic gene that drives cancer cells into programmed cell death, while *Maspin* is a tumour suppressor gene. Recently, a synthetic zinc finger transcription factor made to bind stretches of CAG repeats in the *Huntingtin* (*HTT*) gene, found in polyglutamine disorders such as Huntington's disease, was shown to repress the mutant protein in human cell lines (Garriga-Canut et al, 2012). Moreover, when the ATF was delivered to the brains of a Huntington's disease mouse model, acute reduction in the expression of the mutant gene and also the protein was achieved, which lead to a delay in the onset of Huntington's disease symptoms (Garriga-Canut et al, 2012).

The second project examined the potential of transcription factors to direct cellular reprogramming of fibroblasts towards the megakaryocyte (MK) lineage. MKs produce blood cell fragments called platelets that are crucial to normal blood clotting. If platelet numbers decline, a life-threatening condition called thrombocytopenia develops, a major problem experienced by patients undergoing

chemotherapy and stem cell transplants. This study was motivated by a lack of effective treatments for this condition and the current reliance on less than ideal blood transfusions. Inspired by recent research showing that transcription factors can drive cellular reprogramming toward multiple lineages (Feng et al, 2008; Huang et al, 2011; Ieda et al, 2010; Vierbuchen et al, 2010), we explored the potential of the MK gene regulatory network to direct the conversion of fibroblasts with the ultimate aim of generating platelet-producing MK-like cells.

We showed that ectopic expression of the erythro-megakaryocytic transcription factor GATA1 and the MK-specific factor FLI1 resulted in gene expression changes in murine embryonic fibroblasts. MK-specific genes such as *Pf4*, *CD61* and *Gp9* were up-regulated in cells expressing both factors compared to controls suggesting that the two factors are able to regulate endogenous genes embedded in natural chromatin. This is an essential pre-requisite for the initiation of reprogramming. We then examined whether inclusion of additional transcriptional regulators could further promote and more properly direct reprogramming, as successful strategies have generally required the expression of at least three transcription factors (Huang et al, 2011; Ieda et al, 2010; Vierbuchen et al, 2010). Consequently, we demonstrated that addition of the early haematopoietic transcription factor TAL1 to cells expressing GATA1 and FLI1 resulted in further up-regulation of MK-specific gene expression and induction of MK-like acetylcholinesterase activity, suggesting further progression along the reprogramming pathway.

We also tested the truncated GATA1 mutant, GATA1 short (GATA1s), which in contrast to GATA1 positively regulates proliferation of MK progenitors. We

determined that in combination with FLI1 it performs better than wildtype GATA1 in driving MK gene expression. Additionally, when GATA1s was combined with both FLI1 and TAL1, further increase in *CD61* and *Gp9* expressions were achieved in comparison to GATA1s and FLI1 alone. Although, further investigation is needed to fully assess GATA1s' potential in MK reprogramming with FLI1 and TAL1 compared to combination with GATA1 such as testing for acetylcholinesterase activity. Nevertheless, forced expression of GATA1s has shown that it may have greater potential than GATA1 in converting fibroblasts into proliferating MK cells. Although it should be remembered that GATA1s also has oncogenic activity so it may never be possible to use it in therapeutic situations.

It has been reported that forced expression of p45NFE2/MAF in mouse and human fibroblasts can induce MK lineage conversion (Ono et al, 2012). However, despite a number of attempts, we were unable to repeat this study using our reagents and also found that addition of p45NFE2/MAFG to our combination of GATA1 or GATA1s, FLI1 and TAL1 did not result in any detectable CD41 expression in fibroblasts, the marker used in the study to assess MK reprogramming. Hence, the general application of p45NFE2/MAFG in MK reprogramming remains to be confirmed. Further analysis such as real time RT-PCR and microarray on cells over-expressing p45/MAF and in combination with other transcription factors are required to evaluate and determine the potential of these factors in MK-specific gene program.

5.2 Conclusions

The research undertaken in this thesis has provided important information on the mechanisms that allow gene regulatory proteins to bind DNA and has also investigated the potential of these factors to convert cell fate. In particular, in the first study, we identified the key amino acid residues of the DNA-binding domain of ZNF217 that are required for sequence-specific interaction with DNA. The results of this work have recently been validated by the crystal structure of ZNF217 bound to its core recognition sequence of TGCAGAAT, which confirmed these residues do indeed make direct contact with the consensus site. The discovery that ZNF217 binds DNA by a novel mechanism has expanded our knowledge of how zinc fingers interact with their target sequences. This has implications in our understanding of how gene regulatory proteins function in both health and disease, and will help inform the design of novel zinc finger-based artificial transcription factors for both experimental and therapeutic benefit.

The second study investigated the potential of DNA-binding transcription factors to direct cellular reprogramming by facilitating gene expression changes in terminally differentiated cells. We were interested in converting fibroblasts into platelet-producing megakaryocytes as an alternative approach to treating thrombocytopenia and associated bleeding disorders. We have shown that over-expression of the megakaryocytic transcriptional activators GATA1 or GATA1s, FLI1 and TAL1 in murine embryonic fibroblasts can up-regulate expression of megakaryocyte-specific genes including *Pf4, CD61* and *Gp9* and also induce acetylcholinesterase activity characteristic of megakaryocytes and their lineage progenitors, indicative of partial

reprogramming of fibroblasts towards the megakaryocyte lineage. To fully achieve complete and properly directed cell conversion, comprehensive screening for further factors may be helpful. Future strategies could make use of high-throughput library screening systems, such as the Arrayed RetroViral Expression Cloning (ARVEC) platform that is capable of highly automated genomic screening (Skalamera et al, 2011). Development of screening strategies using ARVEC should identify additional transcriptional regulators that can further promote the conversion of fibroblasts into MKs and also improve reprogramming efficiency and specificity.

REFERENCES

Aasen T, Raya A, Barrero MJ, Garreta E, Consiglio A, Gonzalez F, Vassena R, Bilic J, Pekarik V, Tiscornia G, Edel M, Boue S, Izpisua Belmonte JC (2008) Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. *Nature biotechnology* **26:** 1276-1284

Agalioti T, Lomvardas S, Parekh B, Yie J, Maniatis T, Thanos D (2000) Ordered recruitment of chromatin modifying and general transcription factors to the IFN-beta promoter. *Cell* **103:** 667-678

Akashi K, Traver D, Miyamoto T, Weissman IL (2000) A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404:** 193-197

Alam H, Gangadaran P, Bhate AV, Chaukar DA, Sawant SS, Tiwari R, Bobade J, Kannan S, D'Cruz A K, Kane S, Vaidya MM (2011) Loss of keratin 8 phosphorylation leads to increased tumor progression and correlates with clinico-pathological parameters of OSCC patients. *PloS one* **6:** e27767

Alford KA, Slender A, Vanes L, Li Z, Fisher EM, Nizetic D, Orkin SH, Roberts I, Tybulewicz VL (2010) Perturbed hematopoiesis in the Tc1 mouse model of Down syndrome. *Blood* **115:** 2928-2937

Amabile G, Meissner A (2009) Induced pluripotent stem cells: current progress and potential for regenerative medicine. *Trends in molecular medicine* **15:** 59-68

Ambasudhan R, Talantova M, Coleman R, Yuan X, Zhu S, Lipton SA, Ding S (2011) Direct reprogramming of adult human fibroblasts to functional neurons under defined conditions. *Cell stem cell* **9:** 113-118

Amoutzias GD, Veron AS, Weiner J, 3rd, Robinson-Rechavi M, Bornberg-Bauer E, Oliver SG, Robertson DL (2007) One billion years of bZIP transcription factor evolution: conservation and change in dimerization and DNA-binding site specificity. *Molecular biology and evolution* **24:** 827-835

Andrews NC, Erdjument-Bromage H, Davidson MB, Tempst P, Orkin SH (1993a) Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein. *Nature* **362:** 722-728

Andrews NC, Kotkow KJ, Ney PA, Erdjument-Bromage H, Tempst P, Orkin SH (1993b) The ubiquitous subunit of erythroid transcription factor NF-E2 is a small basic-leucine zipper protein related to the v-maf oncogene. *Proceedings of the National Academy of Sciences of the United States of America* **90:** 11488-11492

Aoi T, Yae K, Nakagawa M, Ichisaka T, Okita K, Takahashi K, Chiba T, Yamanaka S (2008) Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science* **321:** 699-702

Aronica SM, Mantel C, Gonin R, Marshall MS, Sarris A, Cooper S, Hague N, Zhang XF, Broxmeyer HE (1995) Interferon-inducible protein 10 and macrophage inflammatory protein-1 alpha inhibit growth factor stimulation of Raf-1 kinase activity and protein synthesis in a human growth factor-dependent hematopoietic cell line. *The Journal of biological chemistry* **270:** 21998-22007

Athanasiou M, Clausen PA, Mavrothalassitis GJ, Zhang XK, Watson DK, Blair DG (1996) Increased expression of the ETS-related transcription factor FLI-1/ERGB correlates with and can induce the megakaryocytic phenotype. *Cell growth & differentiation : the molecular biology journal of the American Association for Cancer Research* **7:** 1525-1534

Avram D, Fields A, Senawong T, Topark-Ngarm A, Leid M (2002) COUP-TF (chicken ovalbumin upstream promoter transcription factor)-interacting protein 1 (CTIP1) is a sequence-specific DNA binding protein. *The Biochemical journal* **368:** 555-563

Azcoitia V, Aracil M, Martinez AC, Torres M (2005) The homeodomain protein Meis1 is essential for definitive hematopoiesis and vascular patterning in the mouse embryo. *Developmental biology* **280:** 307-320

Bartsevich VV, Juliano RL (2000) Regulation of the MDR1 gene by transcriptional repressors selected using peptide combinatorial libraries. *Molecular pharmacology* **58:** 1-10

Becker RP, De Bruyn PP (1976) The transmural passage of blood cells into myeloid sinusoids and the entry of platelets into the sinusoidal circulation; a scanning electron microscopic investigation. *The American journal of anatomy* **145:** 183-205

Beerli RR, Dreier B, Barbas CF, 3rd (2000) Positive and negative regulation of endogenous genes by designed transcription factors. *Proceedings of the National Academy of Sciences of the United States of America* **97:** 1495-1500

Behnke O (1968) An electron microscope study of the megacaryocyte of the rat bone marrow. I. The development of the demarcation membrane system and the platelet surface coat. *Journal of ultrastructure research* **24:** 412-433

Beltran A, Parikh S, Liu Y, Cuevas BD, Johnson GL, Futscher BW, Blancafort P (2007) Re-activation of a dormant tumor suppressor gene maspin by designed transcription factors. *Oncogene* **26:** 2791-2798

Beltran AS, Sun X, Lizardi PM, Blancafort P (2008) Reprogramming epigenetic silencing: artificial transcription factors synergize with chromatin remodeling drugs to reactivate the tumor suppressor mammary serine protease inhibitor. *Molecular cancer therapeutics* **7:** 1080-1090

Berninger B, Costa MR, Koch U, Schroeder T, Sutor B, Grothe B, Gotz M (2007) Functional properties of neurons derived from in vitro reprogrammed postnatal astroglia. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **27:** 8654-8664

Blobel GA, Nakajima T, Eckner R, Montminy M, Orkin SH (1998) CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. *Proceedings of the National Academy of Sciences of the United States of America* **95:** 2061-2066

Boyes J, Byfield P, Nakatani Y, Ogryzko V (1998) Regulation of activity of the transcription factor GATA-1 by acetylation. *Nature* **396:** 594-598

Breton-Gorius J, Reyes F (1976) Ultrastructure of human bone marrow cell maturation. *International review of cytology* **46:** 251-321

Bruno S, Gunetti M, Gammaitoni L, Dane A, Cavalloni G, Sanavio F, Fagioli F, Aglietta M, Piacibello W (2003) In vitro and in vivo megakaryocyte differentiation of fresh and ex-vivo expanded cord blood cells: rapid and transient megakaryocyte reconstitution. *Haematologica* **88:** 379-387

Bunting S, Widmer R, Lipari T, Rangell L, Steinmetz H, Carver-Moore K, Moore MW, Keller GA, de Sauvage FJ (1997) Normal platelets and megakaryocytes are produced in vivo in the absence of thrombopoietin. *Blood* **90:** 3423-3429

Burley SK, Roeder RG (1996) Biochemistry and structural biology of transcription factor IID (TFIID). *Annual review of biochemistry* **65:** 769-799

Caiazzo M, Dell'Anno MT, Dvoretskova E, Lazarevic D, Taverna S, Leo D, Sotnikova TD, Menegon A, Roncaglia P, Colciago G, Russo G, Carninci P, Pezzoli G, Gainetdinov RR, Gustincich S, Dityatev A, Broccoli V (2011) Direct generation of functional dopaminergic neurons from mouse and human fibroblasts. *Nature* **476:** 224-227

Chan JY, Han XL, Kan YW (1993) Isolation of cDNA encoding the human NF-E2 protein. *Proceedings of the National Academy of Sciences of the United States of America* **90:** 11366-11370

Cherry AB, Daley GQ (2012) Reprogramming cellular identity for regenerative medicine. *Cell* **148:** 1110-1122

Chinnadurai G (2002) CtBP, an unconventional transcriptional corepressor in development and oncogenesis. *Molecular cell* **9:** 213-224

Choo Y, Klug A (1993) A role in DNA binding for the linker sequences of the first three zinc fingers of TFIIIA. *Nucleic acids research* **21:** 3341-3346

Choo Y, Klug A (1997) Physical basis of a protein-DNA recognition code. *Current opinion in structural biology* **7:** 117-125

Chou ST, Khandros E, Bailey LC, Nichols KE, Vakoc CR, Yao Y, Huang Z, Crispino JD, Hardison RC, Blobel GA, Weiss MJ (2009) Graded repression of PU.1/Sfpi1 gene transcription by GATA factors regulates hematopoietic cell fate. *Blood* **114:** 983-994

Christman JK (2002) 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene* **21:** 5483-5495

Clemens KR, Wolf V, McBryant SJ, Zhang P, Liao X, Wright PE, Gottesfeld JM (1993) Molecular basis for specific recognition of both RNA and DNA by a zinc finger protein. *Science* **260:** 530-533

Clemens KR, Zhang P, Liao X, McBryant SJ, Wright PE, Gottesfeld JM (1994) Relative contributions of the zinc fingers of transcription factor IIIA to the energetics of DNA binding. *Journal of molecular biology* **244:** 23-35

Clifton MM (2007) Chapter 7: Characterisation of ZNF217 fingers 6 & 7. PhD Thesis, Molecular and Microbial Biosciences, The University of Sydney,

Cobaleda C, Jochum W, Busslinger M (2007) Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. *Nature* **449:** 473-477

Collins C, Rommens JM, Kowbel D, Godfrey T, Tanner M, Hwang SI, Polikoff D, Nonet G, Cochran J, Myambo K, Jay KE, Froula J, Cloutier T, Kuo WL, Yaswen P, Dairkee S, Giovanola J, Hutchinson GB, Isola J, Kallioniemi OP, Palazzolo M, Martin C, Ericsson C, Pinkel D, Albertson D, Li WB, Gray JW (1998) Positional cloning of ZNF217 and NABC1: genes amplified at 20q13.2 and overexpressed in breast carcinoma. *Proceedings of the National Academy of Sciences of the United States of America* **95:** 8703-8708

Cote J, Quinn J, Workman JL, Peterson CL (1994) Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science* **265:** 53-60

Cowger JJ, Zhao Q, Isovic M, Torchia J (2007) Biochemical characterization of the zinc-finger protein 217 transcriptional repressor complex: identification of a ZNF217 consensus recognition sequence. *Oncogene* **26:** 3378-3386

Cramer EM, Norol F, Guichard J, Breton-Gorius J, Vainchenker W, Masse JM, Debili N (1997) Ultrastructure of platelet formation by human megakaryocytes cultured with the Mpl ligand. *Blood* **89:** 2336-2346

Crossley M, Brownlee GG (1990) Disruption of a C/EBP binding site in the factor IX promoter is associated with haemophilia B. *Nature* **345:** 444-446

Crossley M, Ludwig M, Stowell KM, De Vos P, Olek K, Brownlee GG (1992) Recovery from hemophilia B Leyden: an androgen-responsive element in the factor IX promoter. *Science* **257:** 377-379

Dang DT, Pevsner J, Yang VW (2000) The biology of the mammalian Kruppel-like family of transcription factors. *The international journal of biochemistry & cell biology* **32:** 1103-1121

Davi G, Patrono C (2007) Platelet activation and atherothrombosis. *The New England journal of medicine* **357:** 2482-2494

Davis RL, Weintraub H, Lassar AB (1987) Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51:** 987-1000

De Botton S, Sabri S, Daugas E, Zermati Y, Guidotti JE, Hermine O, Kroemer G, Vainchenker W, Debili N (2002) Platelet formation is the consequence of caspase activation within megakaryocytes. *Blood* **100:** 1310-1317

De Bruyn C, Delforge A, Martiat P, Bron D (2005) Ex vivo expansion of megakaryocyte progenitor cells: cord blood versus mobilized peripheral blood. *Stem cells and development* **14:** 415-424

de Sauvage FJ, Hass PE, Spencer SD, Malloy BE, Gurney AL, Spencer SA, Darbonne WC, Henzel WJ, Wong SC, Kuang WJ, et al. (1994) Stimulation of megakaryocytopoiesis and thrombopoiesis by the c-Mpl ligand. *Nature* **369:** 533-538

Debili N, Coulombel L, Croisille L, Katz A, Guichard J, Breton-Gorius J, Vainchenker W (1996) Characterization of a bipotent erythro-megakaryocytic progenitor in human bone marrow. *Blood* **88:** 1284-1296

Debili N, Wendling F, Katz A, Guichard J, Breton-Gorius J, Hunt P, Vainchenker W (1995) The Mpl-ligand or thrombopoietin or megakaryocyte growth and differentiative factor has both direct proliferative and differentiative activities on human megakaryocyte progenitors. *Blood* **86:** 2516-2525

Demirag GG, Sullu Y, Gurgenyatagi D, Okumus NO, Yucel I (2011) Expression of plakophilins (PKP1, PKP2, and PKP3) in gastric cancers. *Diagnostic pathology* **6:** 1

Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome biology* **4:** P3

Deutsch VR, Tomer A (2006) Megakaryocyte development and platelet production. *British journal of haematology* **134:** 453-466

Dey D, Evans GR (2011) Generation of Induced Pluripotent Stem (iPS) Cells by Nuclear Reprogramming. *Stem cells international* **2011:** 619583

Ding G, Lorenz P, Kreutzer M, Li Y, Thiesen HJ (2009) SysZNF: the C2H2 zinc finger gene database. *Nucleic acids research* **37:** D267-273

Efe JA, Hilcove S, Kim J, Zhou H, Ouyang K, Wang G, Chen J, Ding S (2011) Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy. *Nature cell biology* **13:** 215-222

Eisbacher M, Holmes ML, Newton A, Hogg PJ, Khachigian LM, Crossley M, Chong BH (2003) Protein-protein interaction between Fli-1 and GATA-1 mediates synergistic expression of megakaryocyte-specific genes through cooperative DNA binding. *Molecular and cellular biology* **23:** 3427-3441

Elagib KE, Racke FK, Mogass M, Khetawat R, Delehanty LL, Goldfarb AN (2003) RUNX1 and GATA-1 coexpression and cooperation in megakaryocytic differentiation. *Blood* **101:** 4333-4341

Elrod-Erickson M, Rould MA, Nekludova L, Pabo CO (1996) Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions. *Structure* **4:** 1171-1180

Eminli S, Utikal J, Arnold K, Jaenisch R, Hochedlinger K (2008) Reprogramming of neural progenitor cells into induced pluripotent stem cells in the absence of exogenous Sox2 expression. *Stem cells* **26:** 2467-2474

Fairall L, Schwabe JW, Chapman L, Finch JT, Rhodes D (1993) The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature* **366:** 483-487

Falke D, Fisher M, Ye D, Juliano RL (2003) Design of artificial transcription factors to selectively regulate the pro-apoptotic bax gene. *Nucleic acids research* **31:** e10

Farnham PJ (2009) Insights from genomic profiling of transcription factors. *Nature reviews Genetics* **10:** 605-616

Feng R, Desbordes SC, Xie H, Tillo ES, Pixley F, Stanley ER, Graf T (2008) PU.1 and C/EBPalpha/beta convert fibroblasts into macrophage-like cells. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 6057-6062

Flaus A, Owen-Hughes T (2001) Mechanisms for ATP-dependent chromatin remodelling. *Current opinion in genetics & development* **11:** 148-154

Foster MP, Wuttke DS, Radhakrishnan I, Case DA, Gottesfeld JM, Wright PE (1997) Domain packing and dynamics in the DNA complex of the N-terminal zinc fingers of TFIIIA. *Nature structural biology* **4:** 605-608

Fu F, Sander JD, Maeder M, Thibodeau-Beganny S, Joung JK, Dobbs D, Miller L, Voytas DF (2009) Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays. *Nucleic acids research* **37:** D279-283

Fu JD, Stone NR, Liu L, Spencer CI, Qian L, Hayashi Y, Delgado-Olguin P, Ding S, Bruneau BG, Srivastava D (2013) Direct Reprogramming of Human Fibroblasts toward a Cardiomyocyte-like State. *Stem cell reports* **1:** 235-247

Fuhrken PG, Chen C, Apostolidis PA, Wang M, Miller WM, Papoutsakis ET (2008) Gene Ontology-driven transcriptional analysis of CD34+ cell-initiated megakaryocytic cultures identifies new transcriptional regulators of megakaryopoiesis. *Physiological genomics* **33:** 159-169

Fujimoto TT, Kohata S, Suzuki H, Miyazaki H, Fujimura K (2003) Production of functional platelets by differentiated embryonic stem (ES) cells in vitro. *Blood* **102:** 4044-4051

Fujiwara Y, Browne CP, Cunniff K, Goff SC, Orkin SH (1996) Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proceedings of the National Academy of Sciences of the United States of America* **93:** 12355-12358

Fujiwara Y, Chang AN, Williams AM, Orkin SH (2004) Functional overlap of GATA-1 and GATA-2 in primitive hematopoietic development. *Blood* **103:** 583-585

Funnell AP, Wilson MD, Ballester B, Mak KS, Burdach J, Magan N, Pearson RC, Lemaigre FP, Stowell KM, Odom DT, Flicek P, Crossley M (2013) A CpG mutational hotspot in a ONECUT binding site accounts for the prevalent variant of hemophilia B Leyden. *American journal of human genetics* **92:** 460-467

Garriga-Canut M, Agustin-Pavon C, Herrmann F, Sanchez A, Dierssen M, Fillat C, Isalan M (2012) Synthetic zinc finger repressors reduce mutant huntingtin expression in the brain of R6/2 mice. *Proceedings of the National Academy of Sciences of the United States of America* **109:** E3136-3145

Gaur M, Kamata T, Wang S, Moran B, Shattil SJ, Leavitt AD (2006) Megakaryocytes derived from human embryonic stem cells: a genetically tractable system to study megakaryocytopoiesis and integrin function. *Journal of thrombosis and haemostasis : JTH* **4:** 436-442

Gekas C, Rhodes KE, Gereige LM, Helgadottir H, Ferrari R, Kurdistani SK, Montecino-Rodriguez E, Bassel-Duby R, Olson E, Krivtsov AV, Armstrong S, Orkin SH, Pellegrini M,

Mikkola HK (2009) Mef2C is a lineage-restricted target of Scl/Tal1 and regulates megakaryopoiesis and B-cell homeostasis. *Blood* **113:** 3461-3471

Ginestier C, Cervera N, Finetti P, Esteyries S, Esterni B, Adelaide J, Xerri L, Viens P, Jacquemier J, Charafe-Jauffret E, Chaffanet M, Birnbaum D, Bertucci F (2006) Prognosis and gene expression profiling of 20q13-amplified breast cancers. *Clinical cancer research : an official journal of the American Association for Cancer Research* **12:** 4533-4544

Ginzinger DG, Godfrey TE, Nigro J, Moore DH, 2nd, Suzuki S, Pallavicini MG, Gray JW, Jensen RH (2000) Measurement of DNA copy number at microsatellite loci using quantitative PCR analysis. *Cancer research* **60:** 5405-5409

Goddard TD, Kneller DG *SPARKY 3*, University of California, San Francisco.

Gommans WM, Haisma HJ, Rots MG (2005) Engineering zinc finger protein transcription factors: the therapeutic relevance of switching endogenous gene expression on or off at command. *Journal of molecular biology* **354:** 507-519

Gordon MS, Hoffman R (1992) Growth factors affecting human thrombocytopoiesis: potential agents for the treatment of thrombocytopenia. *Blood* **80:** 302-307

Grass JA, Boyer ME, Pal S, Wu J, Weiss MJ, Bresnick EH (2003) GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proceedings of the National Academy of Sciences of the United States of America* **100:** 8811-8816

Greene ME, Mundschau G, Wechsler J, McDevitt M, Gamis A, Karp J, Gurbuxani S, Arceci R, Crispino JD (2003) Mutations in GATA1 in both transient myeloproliferative disorder and acute megakaryoblastic leukemia of Down syndrome. *Blood cells, molecules & diseases* **31:** 351-356

Grewal SI, Moazed D (2003) Heterochromatin and epigenetic control of gene expression. *Science* **301:** 798-802

Grooteclaes M, Deveraux Q, Hildebrand J, Zhang Q, Goodman RH, Frisch SM (2003) C-terminal-binding protein corepresses epithelial and proapoptotic gene expression programs. *Proceedings of the National Academy of Sciences of the United States of America* **100:** 4568-4573

Haddad E, Cramer E, Riviere C, Rameau P, Louache F, Guichard J, Nelson DL, Fischer A, Vainchenker W, Debili N (1999) The thrombocytopenia of Wiskott Aldrich syndrome is not related to a defect in proplatelet formation. *Blood* **94:** 509-518

Halder G, Callaerts P, Gehring WJ (1995) Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science* **267:** 1788-1792

Hall MA, Curtis DJ, Metcalf D, Elefanty AG, Sourris K, Robb L, Gothert JR, Jane SM, Begley CG (2003) The critical regulator of embryonic hematopoiesis, SCL, is vital in the adult for megakaryopoiesis, erythropoiesis, and lineage choice in CFU-S12. *Proceedings of the National Academy of Sciences of the United States of America* **100:** 992-997

Hanna-Rose W, Hansen U (1996) Active repression mechanisms of eukaryotic transcription repressors. *Trends in genetics : TIG* **12:** 229-234

Hanna J, Wernig M, Markoulaki S, Sun CW, Meissner A, Cassady JP, Beard C, Brambrink T, Wu LC, Townes TM, Jaenisch R (2007) Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science* **318:** 1920-1923

Hart A, Melet F, Grossfeld P, Chien K, Jones C, Tunnacliffe A, Favier R, Bernstein A (2000) Fli-1 is required for murine vascular and megakaryocytic development and is hemizygously deleted in patients with thrombocytopenia. *Immunity* **13:** 167-177

Heinrich C, Blum R, Gascon S, Masserdotti G, Tripathi P, Sanchez R, Tiedt S, Schroeder T, Gotz M, Berninger B (2010) Directing astroglia from the cerebral cortex into subtype specific functional neurons. *PLoS biology* **8:** e1000373

Heyworth C, Pearson S, May G, Enver T (2002) Transcription factor-mediated lineage switching reveals plasticity in primary committed progenitor cells. *The EMBO journal* **21:** 3770-3781

Hisa T, Spence SE, Rachel RA, Fujita M, Nakamura T, Ward JM, Devor-Henneman DE, Saiki Y, Kutsuna H, Tessarollo L, Jenkins NA, Copeland NG (2004) Hematopoietic, angiogenic and eye defects in Meis1 mutant animals. *The EMBO journal* **23:** 450-459

Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4:** 44-57

Huang P, He Z, Ji S, Sun H, Xiang D, Liu C, Hu Y, Wang X, Hui L (2011) Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* **475:** 386-389

Huang Z, Dore LC, Li Z, Orkin SH, Feng G, Lin S, Crispino JD (2009) GATA-2 reinforces megakaryocyte development in the absence of GATA-1. *Molecular and cellular biology* **29:** 5168-5180

Huangfu D, Maehr R, Guo W, Eijkelenboom A, Snitow M, Chen AE, Melton DA (2008a) Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nature biotechnology* **26:** 795-797

Huangfu D, Osafune K, Maehr R, Guo W, Eijkelenboom A, Chen S, Muhlestein W, Melton DA (2008b) Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nature biotechnology* **26:** 1269-1275

Ichikawa M, Asai T, Saito T, Seo S, Yamazaki I, Yamagata T, Mitani K, Chiba S, Ogawa S, Kurokawa M, Hirai H (2004) AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. *Nature medicine* **10:** 299-304

Ieda M, Fu JD, Delgado-Olguin P, Vedantham V, Hayashi Y, Bruneau BG, Srivastava D (2010) Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142:** 375-386

Igarashi K, Kataoka K, Itoh K, Hayashi N, Nishizawa M, Yamamoto M (1994) Regulation of transcription by dimerization of erythroid factor NF-E2 p45 with small Maf proteins. *Nature* **367:** 568-572

Italiano JE, Jr., Lecine P, Shivdasani RA, Hartwig JH (1999) Blood platelets are assembled principally at the ends of proplatelet processes produced by differentiated megakaryocytes. *The Journal of cell biology* **147:** 1299-1312

Italiano JE, Jr., Shivdasani RA (2003) Megakaryocytes and beyond: the birth of platelets. *Journal of thrombosis and haemostasis : JTH* **1:** 1174-1182

Iuchi S (2001) Three classes of C2H2 zinc finger proteins. *Cellular and molecular life sciences : CMLS* **58:** 625-635

Jackson CW (1973) Cholinesterase as a possible marker for early cells of the megakaryocytic series. *Blood* **42:** 413-421

Jeanpierre S, Nicolini FE, Kaniewski B, Dumontet C, Rimokh R, Puisieux A, Maguer-Satta V (2008) BMP4 regulation of human megakaryocytic differentiation is involved in thrombopoietin signaling. *Blood* **112:** 3154-3163

Jones S (2004) An overview of the basic helix-loop-helix proteins. *Genome biology* **5:** 226

Kadonaga JT (2004) Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell* **116:** 247-257

Kajimura S, Seale P, Tomaru T, Erdjument-Bromage H, Cooper MP, Ruas JL, Chin S, Tempst P, Lazar MA, Spiegelman BM (2008) Regulation of the brown and white fat gene programs through a PRDM16/CtBP transcriptional complex. *Genes & development* **22:** 1397-1409

Kang H, Roh S (2011) Extended exposure to trichostatin A after activation alters the expression of genes important for early development in nuclear transfer murine embryos. *The Journal of veterinary medical science / the Japanese Society of Veterinary Science* **73:** 623-631

Kang YA, Shin HC, Yoo JY, Kim JH, Kim JS, Yun CO (2008) Novel cancer antiangiotherapy using the VEGF promoter-targeted artificial zinc-finger protein and oncolytic adenovirus. *Molecular therapy : the journal of the American Society of Gene Therapy* **16:** 1033-1040

Kaushansky K (1995) Thrombopoietin: the primary regulator of megakaryocyte and platelet production. *Thrombosis and haemostasis* **74:** 521-525

Kaushansky K (2008) Historical review: megakaryopoiesis and thrombopoiesis. *Blood* **111:** 981-986

Kaushansky K, Drachman JG (2002) The molecular and cellular biology of thrombopoietin: the primary regulator of platelet production. *Oncogene* **21:** 3359-3367

Kaushansky K, Lok S, Holly RD, Broudy VC, Lin N, Bailey MC, Forstrom JW, Buddle MM, Oort PJ, Hagen FS, et al. (1994) Promotion of megakaryocyte progenitor expansion and differentiation by the c-Mpl ligand thrombopoietin. *Nature* **369:** 568-571

Kawada H, Ito T, Pharr PN, Spyropoulos DD, Watson DK, Ogawa M (2001) Defective megakaryopoiesis and abnormal erythroid development in Fli-1 gene-targeted mice. *International journal of hematology* **73:** 463-468

Kawamura T, Ono K, Morimoto T, Wada H, Hirai M, Hidaka K, Morisaki T, Heike T, Nakahata T, Kita T, Hasegawa K (2005) Acetylation of GATA-4 is involved in the differentiation of embryonic stem cells into cardiac myocytes. *The Journal of biological chemistry* **280:** 19682-19688

Kim JB, Zaehres H, Wu G, Gentile L, Ko K, Sebastiano V, Arauzo-Bravo MJ, Ruau D, Han DW, Zenke M, Scholer HR (2008) Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. *Nature* **454:** 646-650

Klug A (2010) The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Quarterly reviews of biophysics* **43:** 1-21

Kondo M, Weissman IL, Akashi K (1997) Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91:** 661-672

Kotkow KJ, Orkin SH (1996) Complexity of the erythroid transcription factor NF-E2 as revealed by gene targeting of the mouse p18 NF-E2 locus. *Proceedings of the National Academy of Sciences of the United States of America* **93:** 3514-3518

Kouzarides T (2007) Chromatin modifications and their function. *Cell* **128:** 693-705

Kowenz-Leutz E, Leutz A (1999) A C/EBP beta isoform recruits the SWI/SNF complex to activate myeloid genes. *Molecular cell* **4:** 735-743

Krig SR, Jin VX, Bieda MC, O'Geen H, Yaswen P, Green R, Farnham PJ (2007) Identification of genes directly regulated by the oncogene ZNF217 using chromatin immunoprecipitation (ChIP)-chip assays. *The Journal of biological chemistry* **282:** 9703-9712

Kruskall MS (1997) The perils of platelet transfusions. *The New England journal of medicine* **337:** 1914-1915

Kwan AH, Mobli M, Gooley PR, King GF, Mackay JP (2011) Macromolecular NMR spectroscopy for the non-spectroscopist. *The FEBS journal* **278:** 687-703

Laiosa CV, Stadtfeld M, Xie H, de Andres-Aguayo L, Graf T (2006) Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors. *Immunity* **25:** 731-744

Laity JH, Dyson HJ, Wright PE (2000) DNA-induced alpha-helix capping in conserved linker sequences is a determinant of binding affinity in Cys(2)-His(2) zinc fingers. *Journal of molecular biology* **295:** 719-727

Lecine P, Villeval JL, Vyas P, Swencki B, Xu Y, Shivdasani RA (1998) Mice lacking transcription factor NF-E2 provide in vivo validation of the proplatelet model of thrombocytopoiesis and show a platelet production defect that is intrinsic to megakaryocytes. *Blood* **92:** 1608-1616

Lee JH, Hart SR, Skalnik DG (2004) Histone deacetylase activity is required for embryonic stem cell differentiation. *Genesis* **38:** 32-38

Lemarchandel V, Ghysdael J, Mignotte V, Rahuel C, Romeo PH (1993) GATA and Ets cis-acting sequences mediate megakaryocyte-specific expression. *Molecular and cellular biology* **13:** 668-676

Li J, Song L, Qiu Y, Yin A, Zhong M (2014) ZNF217 is associated with poor prognosis and enhances proliferation and metastasis in ovarian cancer. *International journal of clinical and experimental pathology* **7:** 3038-3047

Li P, Maines-Bandiera S, Kuo WL, Guan Y, Sun Y, Hills M, Huang G, Collins CC, Leung PC, Gray JW, Auersperg N (2007) Multiple roles of the candidate oncogene ZNF217 in ovarian epithelial neoplastic progression. *International journal of cancer Journal international du cancer* **120:** 1863-1873

Li Z, Godinho FJ, Klusmann JH, Garriga-Canut M, Yu C, Orkin SH (2005) Developmental stage-selective effect of somatically mutated leukemogenic transcription factor GATA1. *Nature genetics* **37:** 613-619

Liu PQ, Rebar EJ, Zhang L, Liu Q, Jamieson AC, Liang Y, Qi H, Li PX, Chen B, Mendel MC, Zhong X, Lee YL, Eisenberg SP, Spratt SK, Case CC, Wolffe AP (2001) Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. *The Journal of biological chemistry* **276:** 11323-11334

Liu Q, Segal DJ, Ghiara JB, Barbas CF, 3rd (1997) Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proceedings of the National Academy of Sciences of the United States of America* **94:** 5525-5530

Loh YH, Agarwal S, Park IH, Urbach A, Huo H, Heffner GC, Kim K, Miller JD, Ng K, Daley GQ (2009) Generation of induced pluripotent stem cells from human blood. *Blood* **113:** 5476-5479

Loidl P (2001) Introduction: assembly, remodeling and modification of chromatin. *Cellular and molecular life sciences : CMLS* **58:** 663-664

Lok S, Kaushansky K, Holly RD, Kuijper JL, Lofton-Day CE, Oort PJ, Grant FJ, Heipel MD, Burkhead SK, Kramer JM, et al. (1994) Cloning and expression of murine thrombopoietin cDNA and stimulation of platelet production in vivo. *Nature* **369:** 565-568

Lu D, Searles MA, Klug A (2003) Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. *Nature* **426:** 96-100

Lu SJ, Li F, Yin H, Feng Q, Kimbrel EA, Hahm E, Thon JN, Wang W, Italiano JE, Cho J, Lanza R (2011) Platelets generated from human embryonic stem cells are functional in vitro and in the microcirculation of living mice. *Cell research* **21:** 530-545

Lujan E, Chanda S, Ahlenius H, Sudhof TC, Wernig M (2012) Direct conversion of mouse fibroblasts to self-renewing, tripotent neural precursor cells. *Proceedings of the National Academy of Sciences of the United States of America* **109:** 2527-2532

Mackay JP, Crossley M (1998) Zinc fingers are sticking together. *Trends in biochemical sciences* **23:** 1-4

Maherali N, Ahfeldt T, Rigamonti A, Utikal J, Cowan C, Hochedlinger K (2008) A high-efficiency system for the generation and study of human induced pluripotent stem cells. *Cell stem cell* **3:** 340-345

Maniatis T, Falvo JV, Kim TH, Kim TK, Lin CH, Parekh BS, Wathelet MG (1998) Structure and function of the interferon-beta enhanceosome. *Cold Spring Harbor symposia on quantitative biology* **63:** 609-620

Marro S, Pang ZP, Yang N, Tsai MC, Qu K, Chang HY, Sudhof TC, Wernig M (2011) Direct lineage conversion of terminally differentiated hepatocytes to functional neurons. *Cell stem cell* **9:** 374-382

Mason C, Dunnill P (2008) A brief definition of regenerative medicine. *Regenerative medicine* **3:** 1-5

Matharu NK, Hussain T, Sankaranarayanan R, Mishra RK (2010) Vertebrate homologue of Drosophila GAGA factor. *Journal of molecular biology* **400:** 434-447

Matsubara Y, Suzuki H, Ikeda Y, Murata M (2010) Generation of megakaryocytes and platelets from preadipocyte cell line 3T3-L1, but not the parent cell line 3T3, in vitro. *Biochemical and biophysical research communications* **402:** 796-800

Matsumura-Takeda K, Sogo S, Isakari Y, Harada Y, Nishioka K, Kawakami T, Ono T, Taki T (2007) CD41+/CD45+ cells without acetylcholinesterase activity are immature and a major megakaryocytic population in murine bone marrow. *Stem cells* **25:** 862-870

Matsunaga T, Tanaka I, Kobune M, Kawano Y, Tanaka M, Kuribayashi K, Iyama S, Sato T, Sato Y, Takimoto R, Takayama T, Kato J, Ninomiya T, Hamada H, Niitsu Y (2006) Ex vivo large-scale generation of human platelets from cord blood CD34+ cells. *Stem cells* **24:** 2877-2887

Matthews JM, Sunde M (2002) Zinc fingers--folds for many occasions. *IUBMB life* **54:** 351-355

Mehaffey MG, Newton AL, Gandhi MJ, Crossley M, Drachman JG (2001) X-linked thrombocytopenia caused by a novel mutation of GATA-1. *Blood* **98:** 2681-2688

Mermod N, O'Neill EA, Kelly TJ, Tjian R (1989) The proline-rich transcriptional activator of CTF/NF-I is distinct from the replication and DNA binding domain. *Cell* **58:** 741-753

Minami T, Tachibana K, Imanishi T, Doi T (1998) Both Ets-1 and GATA-1 are essential for positive regulation of platelet factor 4 gene expression. *European journal of biochemistry / FEBS* **258:** 879-889

Miura K, Okada Y, Aoi T, Okada A, Takahashi K, Okita K, Nakagawa M, Koyanagi M, Tanabe K, Ohnuki M, Ogawa D, Ikeda E, Okano H, Yamanaka S (2009) Variation in the safety of induced pluripotent stem cell lines. *Nature biotechnology* **27:** 743-745

Morris SA, Daley GQ (2013) A blueprint for engineering cell fate: current technologies to reprogram cell identity. *Cell research* **23:** 33-48

Motohashi H, Katsuoka F, Shavit JA, Engel JD, Yamamoto M (2000) Positive or negative MARE-dependent transcriptional regulation is determined by the abundance of small Maf proteins. *Cell* **103:** 865-875

Munshi N, Yie Y, Merika M, Senger K, Lomvardas S, Agalioti T, Thanos D (1999) The IFN-beta enhancer: a paradigm for understanding activation and repression of inducible gene expression. *Cold Spring Harbor symposia on quantitative biology* **64:** 149-159

Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochiduki Y, Takizawa N, Yamanaka S (2008) Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nature biotechnology* **26:** 101-106

Nakeff A, Maat B (1974) Separation of megakaryocytes from mouse bone marrow by velocity sedimentation. *Blood* **43:** 591-595

Nakorn TN, Miyamoto T, Weissman IL (2003) Characterization of mouse clonogenic megakaryocyte progenitors. *Proceedings of the National Academy of Sciences of the United States of America* **100:** 205-210

Narlikar GJ, Fan HY, Kingston RE (2002) Cooperation between complexes that regulate chromatin structure and transcription. *Cell* **108:** 475-487

Nerlov C, Querfurth E, Kulessa H, Graf T (2000) GATA-1 interacts with the myeloid PU.1 transcription factor and represses PU.1-dependent transcription. *Blood* **95:** 2543-2551

Ng HH, Bird A (2000) Histone deacetylases: silencers for hire. *Trends in biochemical sciences* **25:** 121-126

Nichols KE, Crispino JD, Poncz M, White JG, Orkin SH, Maris JM, Weiss MJ (2000) Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nature genetics* **24:** 266-270

Niitsu N, Yamamoto-Yamaguchi Y, Miyoshi H, Shimizu K, Ohki M, Umeda M, Honma Y (1997) AML1a but not AML1b inhibits erythroid differentiation induced by sodium butyrate and enhances the megakaryocytic differentiation of K562 leukemia cells. *Cell growth & differentiation : the molecular biology journal of the American Association for Cancer Research* **8:** 319-326

Nikolov DB, Burley SK (1997) RNA polymerase II transcription initiation: a structural view. *Proceedings of the National Academy of Sciences of the United States of America* **94:** 15-22

Niwa A, Umeda K, Chang H, Saito M, Okita K, Takahashi K, Nakagawa M, Yamanaka S, Nakahata T, Heike T (2009) Orderly hematopoietic development of induced pluripotent stem cells via Flk-1(+) hemoangiogenic progenitors. *Journal of cellular physiology* **221:** 367-377

Nonet GH, Stampfer MR, Chin K, Gray JW, Collins CC, Yaswen P (2001) The ZNF217 gene amplified in breast cancers promotes immortalization of human mammary epithelial cells. *Cancer research* **61:** 1250-1254

Nunez N (2012) Examination of the molecular roles of the zinc finger transcription factors ZNF217 and Klf8. PhD Thesis, School of Molecular Biosciences, University of Sydney,

Nunez N, Clifton MM, Funnell AP, Artuz C, Hallal S, Quinlan KG, Font J, Vandevenne M, Setiyaputra S, Pearson RC, Mackay JP, Crossley M (2011) The multi-zinc finger protein ZNF217 contacts DNA through a two-finger domain. *The Journal of biological chemistry* **286:** 38190-38201

Nutt SL, Heavey B, Rolink AG, Busslinger M (1999) Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature* **401:** 556-562

Ogawa M (1993) Differentiation and proliferation of hematopoietic stem cells. *Blood* **81:** 2844-2853

Ohneda K, Yamamoto M (2002) Roles of hematopoietic transcription factors GATA-1 and GATA-2 in the development of red blood cell lineage. *Acta haematologica* **108:** 237-245

Okada Y, Nagai R, Sato T, Matsuura E, Minami T, Morita I, Doi T (2003) Homeodomain proteins MEIS1 and PBXs regulate the lineage-specific transcription of the platelet factor 4 gene. *Blood* **101:** 4748-4756

Ono Y, Wang Y, Suzuki H, Okamoto S, Ikeda Y, Murata M, Poncz M, Matsubara Y (2012) Induction of functional platelets from mouse and human fibroblasts by p45NF-E2/Maf. *Blood* **120:** 3812-3821

Onodera K, Shavit JA, Motohashi H, Yamamoto M, Engel JD (2000) Perinatal synthetic lethality and hematopoietic defects in compound mafG::mafK mutant mice. *The EMBO journal* **19:** 1335-1345

Pabo CO, Peisach E, Grant RA (2001) Design and selection of novel Cys2His2 zinc finger proteins. *Annual review of biochemistry* **70:** 313-340

Pang L, Xue HH, Szalai G, Wang X, Wang Y, Watson DK, Leonard WJ, Blobel GA, Poncz M (2006) Maturation stage-specific regulation of megakaryopoiesis by pointed-domain Ets proteins. *Blood* **108:** 2198-2206

Pang ZP, Yang N, Vierbuchen T, Ostermeier A, Fuentes DR, Yang TQ, Citri A, Sebastiano V, Marro S, Sudhof TC, Wernig M (2011) Induction of human neuronal cells by defined transcription factors. *Nature* **476:** 220-223

Papworth M, Kolasinska P, Minczuk M (2006) Designer zinc-finger proteins and their applications. *Gene* **366:** 27-38

Park IH, Zhao R, West JA, Yabuuchi A, Huo H, Ince TA, Lerou PH, Lensch MW, Daley GQ (2008) Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451:** 141-146

Patel SR, Hartwig JH, Italiano JE, Jr. (2005) The biogenesis of platelets from megakaryocyte proplatelets. *The Journal of clinical investigation* **115:** 3348-3354

Pavletich NP, Pabo CO (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science* **252:** 809-817

Pavletich NP, Pabo CO (1993) Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science* **261:** 1701-1707

Pearson RC, Funnell AP, Crossley M (2011) The mammalian zinc finger transcription factor Kruppel-like factor 3 (KLF3/BKLF). *IUBMB life* **63:** 86-93

Peiro G, Diebold J, Lohrs U (2002) CAS (cellular apoptosis susceptibility) gene expression in ovarian carcinoma: Correlation with 20q13.2 copy number and cyclin D1, p53, and Rb protein expression. *American journal of clinical pathology* **118:** 922-929

Pfisterer U, Kirkeby A, Torper O, Wood J, Nelander J, Dufour A, Bjorklund A, Lindvall O, Jakobsson J, Parmar M (2011) Direct conversion of human fibroblasts to dopaminergic neurons. *Proceedings of the National Academy of Sciences of the United States of America* **108:** 10343-10348

Pick M, Azzola L, Osborne E, Stanley EG, Elefanty AG (2013) Generation of megakaryocytic progenitors from human embryonic stem cells in a feeder- and serum-free medium. *PloS one* **8:** e55530

Polo JM, Liu S, Figueroa ME, Kulalert W, Eminli S, Tan KY, Apostolou E, Stadtfeld M, Li Y, Shioda T, Natesan S, Wagers AJ, Melnick A, Evans T, Hochedlinger K (2010) Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nature biotechnology* **28:** 848-855

Proudfoot NJ, Furger A, Dye MJ (2002) Integrating mRNA processing with transcription. *Cell* **108:** 501-512

Putz G, Rosner A, Nuesslein I, Schmitz N, Buchholz F (2006) AML1 deletion in adult mice causes splenomegaly and lymphomas. *Oncogene* **25:** 929-939

Qiang L, Fujita R, Yamashita T, Angulo S, Rhinn H, Rhee D, Doege C, Chau L, Aubry L, Vanti WB, Moreno H, Abeliovich A (2011) Directed conversion of Alzheimer's disease patient skin fibroblasts into functional neurons. *Cell* **146:** 359-371

Qin Z, Ren F, Xu X, Ren Y, Li H, Wang Y, Zhai Y, Chang Z (2009) ZNF536, a novel zinc finger protein specifically expressed in the brain, negatively regulates neuron differentiation by repressing retinoic acid-induced gene transcription. *Molecular and cellular biology* **29:** 3633-3643

Quesenberry PJ, Ihle JN, McGrath E (1985) The effect of interleukin 3 and GM-CSA-2 on megakaryocyte and myeloid clonal colony formation. *Blood* **65:** 214-217

Quinlan KG, Nardini M, Verger A, Francescato P, Yaswen P, Corda D, Bolognesi M, Crossley M (2006) Specific recognition of ZNF217 and other zinc finger proteins at a surface groove of C-terminal binding proteins. *Molecular and cellular biology* **26:** 8159-8172

Quinlan KG, Verger A, Yaswen P, Crossley M (2007) Amplification of zinc finger gene 217 (ZNF217) and cancer: when good fingers go bad. *Biochimica et biophysica acta* **1775:** 333-340

Radley JM, Haller CJ (1983) Fate of senescent megakaryocytes in the bone marrow. *British journal of haematology* **53:** 277-287

Rebar EJ, Huang Y, Hickey R, Nath AK, Meoli D, Nath S, Chen B, Xu L, Liang Y, Jamieson AC, Zhang L, Spratt SK, Case CC, Wolffe A, Giordano FJ (2002) Induction of

angiogenesis in a mouse model using engineered transcription factors. *Nature medicine* **8:** 1427-1432

Reems JA, Pineault N, Sun S (2010) In vitro megakaryocyte production and platelet biogenesis: state of the art. *Transfusion medicine reviews* **24:** 33-43

Reijnen MJ, Peerlinck K, Maasdam D, Bertina RM, Reitsma PH (1993) Hemophilia B Leyden: substitution of thymine for guanine at position -21 results in a disruption of a hepatocyte nuclear factor 4 binding site in the factor IX promoter. *Blood* **82:** 151-158

Reijnen MJ, Sladek FM, Bertina RM, Reitsma PH (1992) Disruption of a binding site for hepatocyte nuclear factor 4 results in hemophilia B Leyden. *Proceedings of the National Academy of Sciences of the United States of America* **89:** 6300-6303

Robb L, Lyons I, Li R, Hartley L, Kontgen F, Harvey RP, Metcalf D, Begley CG (1995) Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. *Proceedings of the National Academy of Sciences of the United States of America* **92:** 7075-7079

Robinson BE, McGrath HE, Quesenberry PJ (1987) Recombinant murine granulocyte macrophage colony-stimulating factor has megakaryocyte colony-stimulating activity and augments megakaryocyte colony stimulation by interleukin 3. *The Journal of clinical investigation* **79:** 1648-1652

Roeder RG (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends in biochemical sciences* **21:** 327-335

Roeder RG (2005) Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS letters* **579:** 909-915

Rooney PH, Boonsong A, McFadyen MC, McLeod HL, Cassidy J, Curran S, Murray GI (2004) The candidate oncogene ZNF217 is frequently amplified in colon cancer. *The Journal of pathology* **204:** 282-288

Sakai T, Hino K, Wada S, Maeda H (2003) Identification of the DNA binding specificity of the human ZNF219 protein and its function as a transcriptional repressor. *DNA research : an international journal for rapid publication of reports on genes and genomes* **10:** 155-165

Sambrook J, Fritsch EK, Maniatis T (1989) *Molecular Cloning. A Laboratory Manual.*, 2nd edn.: Cold Spring Harbour Laboratory Press.

Schipper LF, Brand A, Reniers N, Melief CJ, Willemze R, Fibbe WE (2003) Differential maturation of megakaryocyte progenitor cells from cord blood and mobilized peripheral blood. *Experimental hematology* **31:** 324-330

Schnichels S, Schultheiss M, Hofmann J, Szurman P, Bartz-Schmidt KU, Spitzer MS (2012) Trichostatin A induces cell death at the concentration recommended to differentiate the RGC-5 cell line. *Neurochemistry international* **60:** 581-591

Schulze H, Korpal M, Hurov J, Kim SW, Zhang J, Cantley LC, Graf T, Shivdasani RA (2006) Characterization of the megakaryocyte demarcation membrane system and its role in thrombopoiesis. *Blood* **107:** 3868-3875

Scott MP, Tamkun JW, Hartzell GW, 3rd (1989) The structure and function of the homeodomain. *Biochimica et biophysica acta* **989:** 25-48

Segal DJ, Crotty JW, Bhakta MS, Barbas CF, 3rd, Horton NC (2006) Structure of Aart, a designed six-finger zinc finger peptide, bound to DNA. *Journal of molecular biology* **363:** 405-421

Sekiya S, Suzuki A (2011) Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. *Nature* **475:** 390-393

Sera T (2009) Zinc-finger-based artificial transcription factors and their applications. *Advanced drug delivery reviews* **61:** 513-526

Seth A, Robinson L, Thompson DM, Watson DK, Papas TS (1993) Transactivation of GATA-1 promoter with ETS1, ETS2 and ERGB/Hu-FLI-1 proteins: stabilization of the ETS1 protein binding on GATA-1 promoter sequences by monoclonal antibody. *Oncogene* **8:** 1783-1790

Shim MH, Hoover A, Blake N, Drachman JG, Reems JA (2004) Gene expression profile of primary human CD34+CD38lo cells differentiating along the megakaryocyte lineage. *Experimental hematology* **32:** 638-648

Shivdasani RA, Fujiwara Y, McDevitt MA, Orkin SH (1997) A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *The EMBO journal* **16:** 3965-3973

Shivdasani RA, Mayer EL, Orkin SH (1995a) Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. *Nature* **373:** 432-434

Shivdasani RA, Orkin SH (1995) Erythropoiesis and globin gene expression in mice lacking the transcription factor NF-E2. *Proceedings of the National Academy of Sciences of the United States of America* **92:** 8690-8694

Shivdasani RA, Rosenblatt MF, Zucker-Franklin D, Jackson CW, Hunt P, Saris CJ, Orkin SH (1995b) Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. *Cell* **81:** 695-704

Simpson RJ, Cram ED, Czolij R, Matthews JM, Crossley M, Mackay JP (2003) CCHX zinc finger derivatives retain the ability to bind Zn(II) and mediate protein-DNA interactions. *The Journal of biological chemistry* **278:** 28011-28018

Skalamera D, Dahmer M, Purdon AS, Wilson BM, Ranall MV, Blumenthal A, Gabrielli B, Gonda TJ (2012) Generation of a genome scale lentiviral vector library for EF1alpha promoter-driven expression of human ORFs and identification of human genes affecting viral titer. *PloS one* **7:** e51733

Skalamera D, Ranall MV, Wilson BM, Leo P, Purdon AS, Hyde C, Nourbakhsh E, Grimmond SM, Barry SC, Gabrielli B, Gonda TJ (2011) A high-throughput platform for lentiviral overexpression screening of the human ORFeome. *PloS one* **6:** e20057

Smale ST, Kadonaga JT (2003) The RNA polymerase II core promoter. *Annual review of biochemistry* **72:** 449-479

Snowden AW, Zhang L, Urnov F, Dent C, Jouvenot Y, Zhong X, Rebar EJ, Jamieson AC, Zhang HS, Tan S, Case CC, Pabo CO, Wolffe AP, Gregory PD (2003) Repression of vascular endothelial growth factor A in glioblastoma cells using engineered zinc finger transcription factors. *Cancer research* **63:** 8968-8976

Son EY, Ichida JK, Wainger BJ, Toma JS, Rafuse VF, Woolf CJ, Eggan K (2011) Conversion of mouse and human fibroblasts into functional spinal motor neurons. *Cell stem cell* **9:** 205-218

Spyropoulos DD, Pharr PN, Lavenburg KR, Jackers P, Papas TS, Ogawa M, Watson DK (2000) Hemorrhage, impaired hematopoiesis, and lethality in mouse embryos carrying a targeted disruption of the Fli1 transcription factor. *Molecular and cellular biology* **20:** 5643-5652

Stadtfeld M, Brennand K, Hochedlinger K (2008) Reprogramming of pancreatic beta cells into induced pluripotent stem cells. *Current biology : CB* **18:** 890-894

Starck J, Cohet N, Gonnet C, Sarrazin S, Doubeikovskaia Z, Doubeikovski A, Verger A, Duterque-Coquillaud M, Morle F (2003) Functional cross-antagonism between transcription factors FLI-1 and EKLF. *Molecular and cellular biology* **23:** 1390-1402

Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* **403:** 41-45

Stroncek DF, Rebulla P (2007) Platelet transfusions. *Lancet* **370:** 427-438

Szabo E, Rampalli S, Risueno RM, Schnerch A, Mitchell R, Fiebig-Comyn A, Levadoux-Martin M, Bhatia M (2010) Direct conversion of human fibroblasts to multilineage blood progenitors. *Nature* **468:** 521-526

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131:** 861-872

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

Takayama N, Nishikii H, Usui J, Tsukui H, Sawaguchi A, Hiroyama T, Eto K, Nakauchi H (2008) Generation of functional platelets from human embryonic stem cells in vitro via ES-sacs, VEGF-promoted structures that concentrate hematopoietic progenitors. *Blood* **111:** 5298-5306

Takigawa Y, Hata K, Muramatsu S, Amano K, Ono K, Wakabayashi M, Matsuda A, Takada K, Nishimura R, Yoneda T (2010) The transcription factor Znf219 regulates chondrocyte differentiation by assembling a transcription factory with Sox9. *Journal of cell science* **123:** 3780-3788

Tanner MM, Tirkkonen M, Kallioniemi A, Holli K, Collins C, Kowbel D, Gray JW, Kallioniemi OP, Isola J (1995) Amplification of chromosomal region 20q13 in invasive breast cancer: prognostic implications. *Clinical cancer research : an official journal of the American Association for Cancer Research* **1:** 1455-1461

Terui K, Takahashi Y, Kitazawa J, Toki T, Yokoyama M, Ito E (2000) Expression of transcription factors during megakaryocytic differentiation of CD34+ cells from human cord blood induced by thrombopoietin. *The Tohoku journal of experimental medicine* **192:** 259-273

Tevosian SG, Deconinck AE, Cantor AB, Rieff HI, Fujiwara Y, Corfas G, Orkin SH (1999) FOG-2: A novel GATA-family cofactor related to multitype zinc-finger proteins Friend of GATA-1 and U-shaped. *Proceedings of the National Academy of Sciences of the United States of America* **96:** 950-955

Thanos D, Du W, Maniatis T (1993) The high mobility group protein HMG I(Y) is an essential structural component of a virus-inducible enhancer complex. *Cold Spring Harbor symposia on quantitative biology* **58:** 73-81

Thiel G, Lietz M, Hohl M (2004) How mammalian transcriptional repressors work. *European journal of biochemistry / FEBS* **271:** 2855-2862

Thomas MC, Chiang CM (2006) The general transcription machinery and general cofactors. *Critical reviews in biochemistry and molecular biology* **41:** 105-178

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM (1998) Embryonic stem cell lines derived from human blastocysts. *Science* **282:** 1145-1147

Toki T, Itoh J, Kitazawa J, Arai K, Hatakeyama K, Akasaka J, Igarashi K, Nomura N, Yokoyama M, Yamamoto M, Ito E (1997) Human small Maf proteins form heterodimers with CNC family transcription factors and recognize the NF-E2 motif. *Oncogene* **14:** 1901-1910

Toth KF, Knoch TA, Wachsmuth M, Frank-Stohr M, Stohr M, Bacher CP, Muller G, Rippe K (2004) Trichostatin A-induced histone acetylation causes decondensation of interphase chromatin. *Journal of cell science* **117:** 4277-4287

Tsai FY, Keller G, Kuo FC, Weiss M, Chen J, Rosenblatt M, Alt FW, Orkin SH (1994) An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature* **371:** 221-226

Tsang AP, Fujiwara Y, Hom DB, Orkin SH (1998) Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. *Genes & development* **12:** 1176-1188

Tsang AP, Visvader JE, Turner CA, Fujiwara Y, Yu C, Weiss MJ, Crossley M, Orkin SH (1997) FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell* **90:** 109-119

Tupler R, Perini G, Green MR (2001) Expressing the human genome. *Nature* **409:** 832-833

Turner J, Crossley M (2001) The CtBP family: enigmatic and enzymatic transcriptional co-repressors. *BioEssays : news and reviews in molecular, cellular and developmental biology* **23:** 683-690

Utley RT, Ikeda K, Grant PA, Cote J, Steger DJ, Eberharter A, John S, Workman JL (1998) Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. *Nature* **394:** 498-502

Vadhan-Raj S (2009) Management of chemotherapy-induced thrombocytopenia: current status of thrombopoietic agents. *Seminars in hematology* **46:** S26-32

Vainchenker W, Debili N, Mouthon MA, Wendling F (1995) Megakaryocytopoiesis: cellular aspects and regulation. *Critical reviews in oncology/hematology* **20:** 165-192

Vandevenne M, Jacques DA, Artuz C, Nguyen CD, Kwan AH, Segal DJ, Matthews JM, Crossley M, Guss JM, Mackay JP (2013) New insights into DNA recognition by zinc fingers revealed by structural analysis of the oncoprotein ZNF217. *The Journal of biological chemistry* **288:** 10616-10627

Varga-Szabo D, Pleines I, Nieswandt B (2008) Cell adhesion mechanisms in platelets. *Arteriosclerosis, thrombosis, and vascular biology* **28:** 403-412

Vendrell JA, Thollet A, Nguyen NT, Ghayad SE, Vinot S, Bieche I, Grisard E, Josserand V, Coll JL, Roux P, Corbo L, Treilleux I, Rimokh R, Cohen PA (2012) ZNF217 is a marker of poor prognosis in breast cancer that drives epithelial-mesenchymal transition and invasion. *Cancer research* **72:** 3593-3606

Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Sudhof TC, Wernig M (2010) Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463:** 1035-1041

Visvader JE, Elefanty AG, Strasser A, Adams JM (1992) GATA-1 but not SCL induces megakaryocytic differentiation in an early myeloid line. *The EMBO journal* **11:** 4557-4564

Visvader JE, Fujiwara Y, Orkin SH (1998) Unsuspected role for the T-cell leukemia protein SCL/tal-1 in vascular development. *Genes & development* **12:** 473-479

Vyas P, Ault K, Jackson CW, Orkin SH, Shivdasani RA (1999) Consequences of GATA-1 deficiency in megakaryocytes and platelets. *Blood* **93:** 2867-2875

Wallace EL, Churchill WH, Surgenor DM, An J, Cho G, McGurk S, Murphy L (1995) Collection and transfusion of blood and blood components in the United States, 1992. *Transfusion* **35:** 802-812

Wechsler J, Greene M, McDevitt MA, Anastasi J, Karp JE, Le Beau MM, Crispino JD (2002) Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. *Nature genetics* **32:** 148-152

Weiss MJ, Orkin SH (1995) GATA transcription factors: key regulators of hematopoiesis. *Experimental hematology* **23:** 99-107

Wendling F, Maraskovsky E, Debili N, Florindo C, Teepe M, Titeux M, Methia N, Breton-Gorius J, Cosman D, Vainchenker W (1994) cMpl ligand is a humoral regulator of megakaryocytopoiesis. *Nature* **369:** 571-574

Wernig M, Lengner CJ, Hanna J, Lodato MA, Steine E, Foreman R, Staerk J, Markoulaki S, Jaenisch R (2008) A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nature biotechnology* **26:** 916-924

Wernig M, Meissner A, Foreman R, Brambrink T, Ku M, Hochedlinger K, Bernstein BE, Jaenisch R (2007) In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448:** 318-324

Wieczorek E, Lin Z, Perkins EB, Law DJ, Merchant JL, Zehner ZE (2000) The zinc finger repressor, ZBP-89, binds to the silencer element of the human vimentin gene and complexes with the transcriptional activator, Sp1. *The Journal of biological chemistry* **275:** 12879-12888

Wolfe SA, Grant RA, Elrod-Erickson M, Pabo CO (2001) Beyond the "recognition code": structures of two Cys2His2 zinc finger/TATA box complexes. *Structure* **9:** 717-723

Wolfe SA, Nekludova L, Pabo CO (2000) DNA recognition by Cys2His2 zinc finger proteins. *Annual review of biophysics and biomolecular structure* **29:** 183-212

Wu X, Satpathy AT, Kc W, Liu P, Murphy TL, Murphy KM (2013) Bcl11a controls Flt3 expression in early hematopoietic progenitors and is required for pDC development in vivo. *PloS one* **8:** e64800

Wuttke DS, Foster MP, Case DA, Gottesfeld JM, Wright PE (1997) Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: determinants of affinity and sequence specificity. *Journal of molecular biology* **273:** 183-206

Xie H, Ye M, Feng R, Graf T (2004) Stepwise reprogramming of B cells into macrophages. *Cell* **117:** 663-676

Xu Y, Wei X, Wang M, Zhang R, Fu Y, Xing M, Hua Q, Xie X (2013) Proliferation rate of somatic cells affects reprogramming efficiency. *The Journal of biological chemistry* **288:** 9767-9778

Yamaguchi Y, Zon LI, Ackerman SJ, Yamamoto M, Suda T (1998) Forced GATA-1 expression in the murine myeloid cell line M1: induction of c-Mpl expression and megakaryocytic/erythroid differentiation. *Blood* **91:** 450-457

Yamazaki R, Kuwana M, Mori T, Okazaki Y, Kawakami Y, Ikeda Y, Okamoto S (2006) Prolonged thrombocytopenia after allogeneic hematopoietic stem cell transplantation: associations with impaired platelet production and increased platelet turnover. *Bone marrow transplantation* **38:** 377-384

Yaswen P, Stampfer MR (2002) Molecular changes accompanying senescence and immortalization of cultured human mammary epithelial cells. *The international journal of biochemistry & cell biology* **34:** 1382-1394

Yechoor V, Liu V, Espiritu C, Paul A, Oka K, Kojima H, Chan L (2009) Neurogenin3 is sufficient for transdetermination of hepatic progenitor cells into neo-islets in vivo but not transdifferentiation of hepatocytes. *Developmental cell* **16:** 358-373

Yoo AS, Sun AX, Li L, Shcheglovitov A, Portmann T, Li Y, Lee-Messer C, Dolmetsch RE, Tsien RW, Crabtree GR (2011) MicroRNA-mediated conversion of human fibroblasts to neurons. *Nature* **476:** 228-231

Yoshida M, Horinouchi S, Beppu T (1995) Trichostatin A and trapoxin: novel chemical probes for the role of histone acetylation in chromatin structure and function. *BioEssays : news and reviews in molecular, cellular and developmental biology* **17:** 423-430

Yoshida M, Kijima M, Akita M, Beppu T (1990) Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A. *The Journal of biological chemistry* **265:** 17174-17179

Yu C, Niakan KK, Matsushita M, Stamatoyannopoulos G, Orkin SH, Raskind WH (2002) X-linked thrombocytopenia with thalassemia from a mutation in the amino finger of GATA-1 affecting DNA binding rather than FOG-1 interaction. *Blood* **100:** 2040-2045

Yu Y, Wang J, Khaled W, Burke S, Li P, Chen X, Yang W, Jenkins NA, Copeland NG, Zhang S, Liu P (2012) Bcl11a is essential for lymphoid development and negatively regulates p53. *The Journal of experimental medicine* **209:** 2467-2483

Zauli G, Vitale M, Falcieri E, Gibellini D, Bassini A, Celeghini C, Columbaro M, Capitani S (1997) In vitro senescence and apoptotic cell death of human megakaryocytes. *Blood* **90:** 2234-2243

Zhang L, Spratt SK, Liu Q, Johnstone B, Qi H, Raschke EE, Jamieson AC, Rebar EJ, Wolffe AP, Case CC (2000) Synthetic zinc finger transcription factor action at an endogenous chromosomal site. Activation of the human erythropoietin gene. *The Journal of biological chemistry* **275:** 33850-33860

Zhang W, Bieker JJ (1998) Acetylation and modulation of erythroid Kruppel-like factor (EKLF) activity by interaction with histone acetyltransferases. *Proceedings of the National Academy of Sciences of the United States of America* **95:** 9855-9860

Zhou Q, Brown J, Kanarek A, Rajagopal J, Melton DA (2008) In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature* **455:** 627-632

Zhou Q, Melton DA (2008) Extreme makeover: converting one cell into another. *Cell stem cell* **3:** 382-388

APPENDIX

Oligonucleotides for EMSA

A3144	Torchia Flank 4-21 wt F	TCCATTGCAGAATTGTGG
A3145	Torchia Flank 4-21 wt R	CCACAATTCTGCAATGGA

Oligonucleotides for cloning and sequencing

A3829	Znf217 S474A Fwd	CAAGAGAGTGTGCTTATTGTGGAAA
A3830	Znf217 S474A Rev	TTTCCACAATAAGCACACTCTCTTG
A3831	Znf217 Y475A Fwd	GAGAGTGTAGTGCTTGTGGAAAGTT
A3832	Znf217 Y475A Rev	AACTTTCCACAAGCACTACACTCTC
A3833	Znf217 G477A Fwd	GTAGTTATTGTGCAAAGTTTTTCCG
A3834	Znf217 G477A Rev	CGGAAAAACTTTGCACAATAACTAC
A3835	Znf217 K478A Fwd	GTTATTGTGGAGCGTTTTTCCGTTC
A3836	Znf217 K478A Rev	GAACGGAAAAACGCTCCACAATAAC
A3837	Znf217 F479A Fwd	ATTGTGGAAAGGCTTTCCGTTCAAA
A3838	Znf217 F479A Rev	TTTGAACGGAAAGCCTTTCCACAAT
A3839	Znf217 R481A Fwd	GAAAGTTTTTCGCTTCAAATTATTA
A3840	Znf217 R481A Rev	TAATAATTTGAAGCGAAAAACTTTC
A3841	Znf217 S482A Fwd	AGTTTTTCCGTGCAAATTATTACCT
A3842	Znf217 S482A Rev	AGGTAATAATTTGCACGGAAAAACT
A3845	Znf217 Y484A Fwd	TCCGTTCAAATGCTTACCTCAATAT
A3846	Znf217 Y484A Rev	ATATTGAGGTAAGCATTTGAACGGA
A3847	Znf217 I488A Fwd	ATTACCTCAATGCTCATCTCAGAAC
A3848	Znf217 I488A Rev	GTTCTGAGATGAGCATTGAGGTAAT
A3849	Znf217 R491A Fwd	ATATTCATCTCGCAACGCATACAGG
A3850	Znf217 R491A Rev	CCTGTATGCGTTGCGAGATGAATATT
A3851	Znf217 G495A Fwd	GAACGCATACAGCTGAAAAACCATA
A3852	Znf217 G495A Rev	TATGGTTTTTCAGCTGTATGCGTTC
A3853	Znf217 E496A Fwd	CGCATACAGGTGCAAAACCATACAA
A3854	Znf217 E496A Rev	TTGTATGGTTTTGCACCTGTATGCG
A3855	Znf217 K497A Fwd	CATACAGGTGAAGCACCATACAAATGT
A3856	Znf217 K497A Rev	ACATTTGTATGGTGCTTCACCTGTATG
A3857	Znf217 P498A Fwd	CAGGTGAAAAAGCATACAAATGTGA

A3858	Znf217 P498A Rev	TCACATTTGTATGCTTTTTCACCTG
A3859	Znf217 Y499A Fwd	GTGAAAAACCAGCCAAATGTGAATT
A3860	Znf217 Y499A Rev	AATTCACATTTGGCTGGTTTTTCAC
A3861	Znf217 K500A Fwd	AAAAACCATACGCATGTGAATTTTG
A3862	Znf217 K500A Rev	CAAAATTCACATGCGTATGGTTTTT
A3863	Znf217 E502A Fwd	CATACAAATGTGCATTTTGTGAATA
A3864	Znf217 E502A Rev	TATTCACAAAATGCACATTTGTATG
A3865	Znf217 F503A Fwd	ACAAATGTGAAGCTTGTGAATATGC
A3866	Znf217 F503A Rev	GCATATTCACAAGCTTCACATTTGT
A3867	Znf217 E505A Fwd	GTGAATTTTGTGCATATGCTGCAGC
A3868	Znf217 E505A Rev	GCTGCAGCATATGCACAAAATTCAC
A3869	Znf217 Y506A Fwd	AATTTTGTGAAGCTGCTGCAGCCCA
A3870	Znf217 Y506A Rev	TGGGCTGCAGCAGCTTCACAAAATTC
A3871	Znf217 A507Q Fwd	TTTGTGAATATCAAGCAGCCCAGAA
A3872	Znf217 A507Q Rev	TTCTGGGCTGCTTGATATTCACAAA
A3873	Znf217 A509Q Fwd	ATATGCTGCACAACAGAAGACATC
A3874	Znf217 A509Q Rev	AGATGTCTTCTGTTGTGCAGCATATT
A3875	Znf217 K511A Fwd	CTGCAGCCCAGGCGACATCTCTGAG
A3876	Znf217 K511A Rev	CTCAGAGATGTCGCCTGGGCTGCAG
A3877	Znf217 T512A Fwd	CAGCCCAGAAGGCATCTCTGAGGTA
A3878	Znf217 T512A Rev	TACCTCAGAGATGCCTTCTGGGCTG
A3879	Znf217 S513A Fwd	CCCAGAAGACAGCTCTGAGGTATCA
A3880	Znf217 S513A Rev	TGATACCTCAGAGCTGTCTTCTGGG
A3883	Znf217 Y516A Fwd	CATCTCTGAGGGCTCACTTGGAGAG
A3884	Znf217 Y516A Rev	CTCTCCAAGTGAGCCCTCAGAGATG
A3885	Znf217 L518A Fwd	TGAGGTATCACGCGGAGAGACATCA
A3886	Znf217 L518A Rev	TGATGTCTCTCCGCGTGATACCTCA
A3887	Znf217 E519A Fwd	GGTATCACTTGGCGAGACATCACAA
A3888	Znf217 E519A Rev	TTGTGATGTCTCGCCAAGTGATACC
A3889	Znf217 R520A Fwd	ATCACTTGGAGGCACATCACAAGGA
A3890	Znf217 R520A Rev	TCCTTGTGATGTGCCTCCAAGTGAT
A3928	pGEX4T1 inta Fwd	AGTTTGAATTGGGTTTGGAGTTTC
A3892	pGex4T1 3internal Rev	TCTTCAGCATCTTTTACTTTCACC
A3930	pGEX4T1 seq Fwd	GGGCTGGCAAGCCACGTTTGGTG
A3931	pGEX4T1 seq Rev	CCGGGAGCTGCATGTGTCAGAGG
A3692	BglII Gata1 fwd	GAAGATCTGCCACCATGGATTTTCCTGGTCTAGG
A3693	XhoI Gata1 3p rev	CCGCTCGAGTCAAGAACTGAGTGGGGC

A4288	BglII Gata1s fwd	ATTAAGATCTATGGAGGGAATTCCTGGGGGCTCA
A3690	BglII Fli1 fwd	GAAGATCTGCCACCATGGACGGGACTATTAAGG
A3691	XhoI Fli1 3p rev	CCGCTCGAGCTAGTAGTAGCTGCCTAAGTGTGA
A3702	BglII Ets1 fwd	GAAGATCTGCCACCATGAAGGCGGCCGTCGAT
A3704	BglII Ets1 3p R	ATTAAGATCTCTAGTCAGCATCCGGCTTTACATC
A4291	EcoRI mTal1 Fwd	ATTAGAATTCGCCACCATGACGGAGCGGCCGCCGAGC
A4292	XhoI mTal1 Rev	ATTACTCGAGTCACCGGGGGCCAGCCCCATC
A3727	pMSCV 5 prime	CCCTTGAACCTCCTCGTTCGACC
A3728	pMSCV 3 prime	GAGACGTGCTACTTCCATTTGTC
A4264	pLV411 seqF	GTGTCGTGAGGAATTAGCTTG
A4265	pLV411 seqR	AGACGGCAATATGGTGGA
A4207	Fog1attB1F	
	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCGGAGACATGTCCAGGAGGAAACAGAGC	
A4208	Fog1 attB2R	
	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACTTTACGTGCTCGGCGGC	
A4209	Meis1 attB1F	
	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAGGCCGATGGCGCAAAGGTACGACGAC	
A4210	Meis1 attB2R	
	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACTGAGCGTGAATGTCCAT	
A4211	Tal1 attB1F	
	GGGGACAAGTTTGTACAAAAAAGCAGGCTGCCCCAGGATGACGGAGCGGCCGCCGAGC	
A4212	Tal1 attB2R	
	GGGGACCACTTTGTACAAGAAAGCTGGGTTTCACCGGGGGCCAGCCCCATC	
A4213	Runx1 attB1F	
	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAACGATGGCTTCAGACAGCATTTTT	
A4214	Runx1 attB2R	
	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGTAGGGCCGCCACACGGC	
A4217	Gabpα attB1F	
	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTTCAACCATGACTAAGAGAGAAGCAGAA	
A4218	Gabpα attB2R	
	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAAATCTCTTTGTCTGCCTG	

Oligonucleotides for real time RT-PCR

A1560	18SrRNA fwd	CACGGCCGGTACAGTGAAAC
A1561	18SrRNA rev	AGAGGAGCGAGCGACCAA
A3797	mGata1rtF	AGCATCAGCACTGGCCTACT
A3798	mGata1rtR	AGGCCCAGCTAGCATAAGGT
A3799	mFli1rtF	CAACCAGCCAGTGAGAGTCA
A3800	mFli1rtR	GCCCACCAGCTTGTTACATT
A3801	mEts1rtF	CCGAGCAGCAAAGAAATGAT
A3802	mEts1rtR	GACGTGGGTTTCTGTCCACT
A3803	mFog1rtF	CATGGCTAGTCCCTGGAGTG
A3804	mFog1rtR	GTCTGGATGCTCCCGTAGAA
A4244	mPf4 exon 2-3 rtF	GCGGTTCCCCAGCTCATAG
A4245	mPf4 exon 2-3 rtR	CCGGTCCAGGCAAATTTTC
A4285	GpIX rtF	TACCAGCCCACAAAAGGTGT
A4287	GpIX rt2R	GGGCAAGCCTGAGTATCTGT
A3805	mGp3artF (CD61)	TGGCTGTGAGTCCTGTGTGT
A3806	mGp3artR (CD61)	GCCTCACTGACTGGGAACTC
A3807	mGp2brtF (CD41)	AGCCACTTTGGCTTCTCAGT
A3808	mGp2brtR (CD41)	CACAGGAATACGGCTCCAGT

Quantitative PCR analysis results

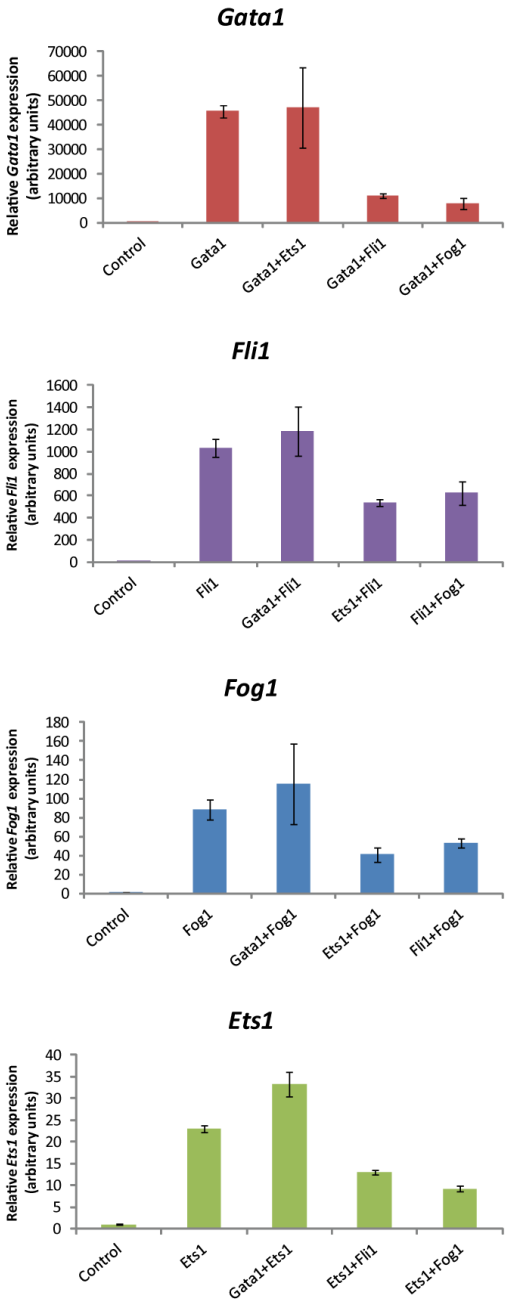


Figure A1. *Over-expression of MK transcriptional regulators in murine embryonic fibroblasts (MEFs). Gata1, Fli1, Fog1* and *Ets1* mRNA were assessed by real time RT-PCR in MEFs following transduction with retroviral vectors containing the indicated MK transcriptional activators. Control samples represent MEFs transduced with vector alone. Transcript levels for each gene were normalised against *18S* rRNA levels and then normalised again to the control sample, which was set at 1. Error bars represent standard error of the mean (n= at least 2).

Microarray analysis results

Table 1. *Genes up-regulated in Gata1+Fli1 cells four months post-transduction.* Genes were ranked by fold change compared to control cells and the top 20 genes were selected. Gene function or process involved information is derived from information in Entrez Gene (**http://www.ncbi.nlm.nih.gov/gene**).

Accession Number	Gene	Description	Gata1+Fli1 /Control (4 months)	Function or Process involved
NM_013563	Il2rg	interleukin 2 receptor, gamma chain	25.56	Cytokine receptor activity, positive regulation of B cell differentiation
NM_029758	Fam49a	family with sequence similarity 49, member A	19.16	Unknown
NM_013723	Podxl	podocalyxin-like	11.30	Negative regulation of cell adhesion
NM_009779	C3ar1	complement component 3a receptor 1	10.81	Complement cascade
NM_001195 084	Plscr2	phospholipid scramblase 2	10.11	Phospholipid translocation between a lipid bilayer
NM_172803	Dock4	dedicator of cytokinesis 4	10.06	Regulation of adherens junction between cells, cytoskeleton reorganisation
NM_009061	Rgs2	regulator of G-protein signaling 2	10.04	GTPase activator activity, inhibits signal transduction
NM_133871	Ifi44	interferon-induced protein 44	9.99	Formation of microtubular structure
NM_001033 141	Ecscr	endothelial cell surface expressed chemotaxis and apoptosis regulator	9.39	Chemotaxis regulation, angiogenesis
NM_011909	Usp18	ubiquitin specific peptidase 18	8.49	Immune response interferon signaling
NM_001198 894	Gpr56	G protein-coupled receptor 56	8.04	Positive regulation of cell adhesion
NM_172872	Kank4	KN motif and ankyrin repeat domains 4	7.70	Cytoskeleton formation, actin polymerisation regulation
NM_011854	Oasl2	2'-5' oligoadenylate synthetase-like 2	7.70	Interferon signaling
NM_008882	Plxna2	plexin A2	7.58	Axon guidance
NM_001127 330	Pparg	peroxisome proliferator activated receptor gamma	7.54	Adipogenesis, fat cell differentiation
NM_011852	Oas1g	2'-5' oligoadenylate synthetase 1G	7.40	Interferon signaling
NM_013468	Ankrd1	ankyrin repeat domain 1 (cardiac muscle)	6.77	Regulation of lipid metabolism, negative regulation of cardiac genes
NM_023386	Rtp4	receptor transporter protein 4	6.29	Unknown

NM_173870	Mgat4a	mannoside acetylglucosaminyltransferase 4, isoenzyme A	6.27	Transferase activity,involved in protein metabolism
NM_009721	Atp1b1	ATPase, Na+/K+ transporting, beta 1 polypeptide	6.06	ATP hydrolysis

Table 2. *Genes up-regulated in Gata1+Fli1 cells seven months post-transduction.* Genes were ranked by fold change compared to control cells and the top 20 genes were selected. Gene function or process involved information is derived from information in Entrez Gene (http://www.ncbi.nlm.nih.gov/gene).

Accession Number	Gene	Description	Gata1+Fli1 /Control (7 months)	Function or Process involved
NM_008730	Nptx1	neuronal pentraxin 1	31.49	Synapse regulation
NM_173870	Mgat4a	mannoside acetylglucosaminyltrans-ferase 4, isoenzyme A	27.04	Transferase activity,involved in protein metabolism
NM_029758	Fam49a	family with sequence similarity 49, member A	26.13	Unknown
NM_001198894	Gpr56	G protein-coupled receptor 56	16.50	Positive regulation of cell adhesion
NM_139200	Cytip	cytohesin 1 interacting protein	15.86	Regulation of cell adhesion in leukocytes
NM_175429	Kctd12b	potassium channel tetramerisation domain containing 12b	15.36	Unknown
NM_009721	Atp1b1	ATPase, Na+/K+ transporting, beta 1 polypeptide	14.31	ATP hydrolysis
NM_008509	Lpl	lipoprotein lipase	13.37	Adipogenesis
NM_001166456	Slc38a1	solute carrier family 38, member 1	13.07	Amino acid transport
NM_001033149	Ttc9	tetratricopeptide repeat domain 9	12.83	Unknown
NM_001033141	Ecscr	endothelial cell surface expressed chemotaxis and apoptosis regulator	12.66	Chemotaxis regulation, angiogenesis
NM_001077202	Hs6st2	heparan sulfate 6-O-sulfotransferase 2	12.46	Heparin biosynthesis
NM_010917	Nid1	nidogen 1	12.06	Maintenance of cell structure
NM_011171	Procr	protein C receptor, endothelial	11.38	Immune response
NM_001012477	Cxcl12	chemokine (C-X-C motif) ligand 12	11.22	Chemokine signalling in T lymphocytes and monocytes
NM_029000	Gvin1	GTPase, very large interferon inducible 1	11.12	Unknown

NM_013723	Podxl	podocalyxin-like	10.33	Negative regulation of cell adhesion
NM_176933	Dusp4	dual specificity phosphatase 4	10.30	Protein dephosphorylation, inactivation of MAPK activity
ENSMUST00000077472	Chst15	carbohydrate (N-acetylgalactosamine 4-sulfate 6-O) sulfotransferase 15	9.78	Sugar metabolism
NM_001127330	Pparg	peroxisome proliferator activated receptor gamma	9.76	Adipogenesis, fat cell differentiation

Table 3. *Genes down-regulated in Gata1+Fli1 cells four months post-transduction.* Genes were ranked by fold change compared to control cells and the top 20 genes were selected. Gene function or process involved information is derived from information in Entrez Gene (**http://www.ncbi.nlm.nih.gov/gene**).

Accession Number	Gene	Description	Gata1+Fli1 /Control (4 months)	Function or Process involved
NM_007472	Aqp1	aquaporin 1	-12.10	Water transmembrane transporter
NM_011812	Fbln5	fibulin 5	-10.01	Promotes elastic fibre assembly and stability in fibroblasts
NM_021361	Nova1	neuro-oncological ventral antigen 1	-9.90	Regulation of RNA splicing in neurons
NM_010207	Fgfr2	fibroblast growth factor receptor 2	-8.91	Regulation of cell growth and differentiation
NM_011526	Tagln	transgelin	-8.55	Actin-binding protein found in fibroblasts
NM_008846	Pip5k1b	phosphatidylinositol-4-phosphate 5-kinase, type 1 beta	-8.43	PI metabolism
NM_001242423	Fam105a	family with sequence similarity 105, member A	-8.42	Unknown
NM_011576	Tfpi	tissue factor pathway inhibitor	-8.01	Negative regulation of blood coagulation
NM_001111027	Runx1t1	runt-related transcription factor 1; translocated to, 1 (cyclin D-related)	-7.34	Regulation of adipogenesis
NM_001008424	Cdsn	corneodesmosin	-7.22	Promotion of cell adhesion
NM_008002	Fgf10	fibroblast growth factor 10	-7.05	Wound healing, cell proliferation and differentiation

NM_001136072	Meis2	Meis homeobox 2	-6.66	Regulates expression of *Pax6* which is the master regulator of eye development
NM_019634	Tspan7	tetraspanin 7	-6.44	Neuronal function
NM_010230	Fmn1	formin 1	-6.22	Cell structure organisation
ENSMUST00000103234	Fbn1	fibrillin 1	-6.15	Secreted by fibroblasts for the formation of elastic fibres
ENSMUST00000169713	Plce1	phospholipase C, epsilon 1	-6.09	Lipid metabolic process
NM_010636	Klf12	Kruppel-like factor 12	-5.93	Negative regulation of transcription
NM_011377	Sim2	single-minded homolog 2 (Drosophila)	-5.83	Regulation of transcription, neural development
NM_009866	Cdh11	cadherin 11	-5.59	Selectively expressed by fibroblasts, mediator of fibroblast inflammation
NM_011864	Papss2	3'-phosphoadenosine 5'-phosphosulfate synthase 2	-5.51	Carbohydrate metabolism

Table 4. *Genes down-regulated in Gata1+Fli1 cells seven months post-transduction.* Genes were ranked by fold change compared to control cells and the top 20 genes were selected. Gene function or process involved information is derived from information in Entrez Gene (**http://www.ncbi.nlm.nih.gov/gene**).

Accession Number	Gene	Description	Gata1+Fli1 /Control (7 months)	Function or Process involved
NM_010728	Lox	lysyl oxidase	-58.49	Maintenance of cell structure
NM_019634	Tspan7	tetraspanin 7	-38.89	Neuronal function
NM_007759	Crabp2	cellular retinoic acid binding protein II	-28.61	Retinol metabolism
NM_011812	Fbln5	fibulin 5	-24.03	Promotes elastic fibre assembly and stability in fibroblasts
NM_001111027	Runx1t1	runt-related transcription factor 1; translocated to, 1 (cyclin D-related)	-22.19	Regulation of adipogenesis
NM_007472	Aqp1	aquaporin 1	-22.11	Water transmembrane transporter
NM_001008424	Cdsn	corneodesmosin	-21.63	Promotion of cell adhesion
NM_145526	P2rx3	purinergic receptor P2X, ligand-gated ion channel, 3	-20.84	ATP binding
NM_009866	Cdh11	cadherin 11	-20.34	Selectively expressed by fibroblasts, mediator of fibroblast inflammation

NM_013496	Crabp1	cellular retinoic acid binding protein I	-19.92	Retinol metabolism
NM_011864	Papss2	3'-phosphoadenosine 5'-phosphosulfate synthase 2	-17.28	Carbohydrate metabolism
NM_153782	Fam20a	family with sequence similarity 20, member A	-16.17	Unknown
NM_028572	Vgll3	vestigial like 3 (Drosophila)	-16.11	Unknown
NM_172553	Alx1	ALX homeobox 1	-15.30	Regulation of transcription
NM_001077361	Fhl1	four and a half LIM domains 1	-13.97	Muscle development
NM_008862	Pkia	protein kinase inhibitor, alpha	-13.43	Negative regulation of protein kinase activity
NM_001136072	Meis2	Meis homeobox 2	-13.37	Regulates expression of *Pax6* which is the master regulator of eye development
NM_018857	Msln	mesothelin	-13.14	Marker for mesothelial cells
NM_018797	Plxnc1	plexin C1	-12.19	Axon guidance
NM_008002	Fgf10	fibroblast growth factor 10	-10.83	Wound healing, cell proliferation and differentiation

Table 5. *Pathways up-regulated in Gata1+Fli1 cells four and seven months post-transduction.* Pathways with a fold up-regulation of greater than 1.5, p-value and FDR of less than 0.05 generated from DAVID bioinformatics resources analysis are listed. GO = gene ontology, FDR = false discovery rate.

GO ID	GO name	No. of Genes	*P*-value	Fold enrichment	FDR
4 months post-transduction					
GO:0001944	vasculature development	29	7.96E-08	3.25	1.39E-04
GO:0006955	immune response	41	1.25E-07	2.52	2.18E-04
GO:0001568	blood vessel development	28	1.74E-07	3.21	3.02E-04
GO:0009615	response to virus	14	7.91E-07	5.68	1.38E-03
GO:0017124	SH3 domain binding	14	3.90E-06	4.96	5.75E-03
GO:0048514	blood vessel morphogenesis	22	8.01E-06	3.11	1.39E-02
GO:0001730	2'-5'-oligoadenylate synthetase activity	5	2.38E-05	23.02	3.51E-02
GO:0001525	angiogenesis	17	2.02E-05	3.58	3.52E-02
7 months post-transduction					
GO:0006955	immune response	60	4.96E-09	2.26	8.87E-06
GO:0035295	tube development	41	2.07E-08	2.67	3.70E-05
GO:0030695	GTPase regulator activity	47	1.11E-06	2.16	1.72E-03
GO:0060589	nucleoside-triphosphatase regulator activity	47	1.76E-06	2.12	2.71E-03
GO:0035239	tube morphogenesis	28	1.76E-06	2.82	3.15E-03
GO:0060429	epithelium development	37	2.27E-06	2.37	4.06E-03
GO:0042127	regulation of cell proliferation	59	3.55E-06	1.88	6.35E-03
GO:0005886	plasma membrane	205	5.81E-06	1.32	7.97E-03
GO:0007155	cell adhesion	60	5.37E-06	1.85	9.61E-03
GO:0022610	biological adhesion	60	5.37E-06	1.85	9.61E-03
GO:0001944	vasculature development	34	8.71E-06	2.33	1.56E-02
GO:0017124	SH3 domain binding	17	1.32E-05	3.61	2.03E-02
GO:0030334	regulation of cell migration	18	1.95E-05	3.36	3.49E-02
GO:0009725	response to hormone stimulus	25	1.97E-05	2.66	3.52E-02
GO:0007167	enzyme linked receptor protein signaling pathway	35	2.42E-05	2.19	4.32E-02

Table 6. *Pathways down-regulated in Gata1+Fli1 cells four and seven months post-transduction.* Pathways with a fold down-regulation of greater than 1.5, p-value and FDR of less than 0.05 generated from DAVID bioinformatics resources analysis are listed. GO = gene ontology, FDR = false discovery rate.

GO ID	GO name	No. of Genes	*P*-value	Fold enrichment	FDR
4 months post-transduction					
GO:0031012	extracellular matrix	41	1.74E-13	3.93	2.29E-10
GO:0005578	proteinaceous extracellular matrix	40	2.24E-13	3.99	2.94E-10
GO:0005576	extracellular region	109	2.30E-12	1.95	3.02E-09
GO:0019838	growth factor binding	18	1.90E-10	7.35	2.78E-07
GO:0044421	extracellular region part	60	7.41E-10	2.37	9.75E-07
GO:0001501	skeletal system development	33	1.45E-09	3.49	2.51E-06
GO:0007167	enzyme linked receptor protein signaling pathway	32	4.00E-09	3.43	6.92E-06
GO:0003700	transcription factor activity	59	6.22E-09	2.26	9.11E-06
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	26	7.83E-09	3.98	1.35E-05
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	36	1.70E-08	2.96	2.94E-05
GO:0030247	polysaccharide binding	20	6.00E-08	4.59	8.80E-05
GO:0001871	pattern binding	20	6.00E-08	4.59	8.80E-05
GO:0042127	regulation of cell proliferation	45	5.50E-08	2.46	9.52E-05
GO:0008201	heparin binding	16	1.06E-07	5.66	1.56E-04
GO:0048729	tissue morphogenesis	27	1.16E-07	3.37	2.00E-04
7 months post-transduction					
GO:0005578	proteinaceous extracellular matrix	41	5.26E-09	2.81	7.18E-06
GO:0031012	extracellular matrix	41	1.65E-08	2.70	2.25E-05
GO:0005576	extracellular region	129	4.55E-08	1.59	6.22E-05
GO:0044421	extracellular region part	70	2.58E-07	1.90	3.53E-04
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	25	2.39E-05	2.65	4.20E-02
GO:0007155	cell adhesion	51	2.61E-05	1.86	4.59E-02
GO:0022610	biological adhesion	51	2.61E-05	1.86	4.59E-02