Forward programming of human pluripotent stem cells to a megakaryocyte-erythrocyte bi-potent progenitor population: an *in vitro* system for the production of platelets and red blood cells for transfusion medicine.

Amanda Louise Dalby

University of Cambridge

St Catharine's College

This dissertation is submitted for the degree of Doctor of Philosophy

June 2017

Summary

Forward programming of human pluripotent stem cells to a megakaryocyte-erythrocyte bi-potent progenitor population: an *in vitro* system for the production of platelets and red blood cells for transfusion medicine.

Amanda Louise Dalby.

There exists a need to produce platelets *in vitro* for use in transfusion medicine, due to increased platelet demands and short shelf life. Our lab uses human induced pluripotent stem cells (iPSCs), as an attractive alternative supply, as iPSCs can be cultured indefinitely and differentiate into almost any cell type. Using a technique called forward programming, we over express three key haematological transcription factors (TFs), pushing iPSCs towards the megakaryocyte lineage, to produce mature megakaryocytes, the platelet precursor cell type.

A major limitation of the forward programming technique is a reliance of lentiviral transduction to overexpress the three TFs, which leads to a number of issues including heterogeneity and high experimental costs. To overcome this, I have developed an inducible iPSC line by inserting the forward programming TFs into a genomic safe harbour, using genome editing techniques. TF expression is strictly controlled, with the TFs expressed only after chemical induction. Inducing forward programming is an efficient method for producing mature megakaryocytes and these cells maintain higher purity in long-term cultures, when compared to cells produced by the lentiviral method.

Removing the requirement of lentiviral transduction is a major advancement, making forward programming more amenable to scaling-up, thus moving this technology closer towards our goal of producing *in vitro* platelets for use in transfusion medicine. I have also shown that forward programming generates a bi-potent progenitor population, from which erythroblasts can be generated, by altering only media conditions. As for megakaryocyte cultures, inducing forward programming improves the purity of erythroblasts produced, compared to the lentiviral method.

I have developed single cell progenitor assays combined with index sorting of different cell surface markers, to allow retrospective analysis of cells which successfully generate colonies. The aim of this work is to better characterise the progenitor cells produced by forward programming, to allow further study of this cell type. Single cell RNA-seq of megakaryocytes revealed heterogeneity in long-term cultures and also identified novel candidate surface markers that may help to further characterise the progenitor cell population.

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the relevant Degree Committee.

Layout

- List of Content
- List of Abbreviations
- Abstract
- Introduction
- Main Materials and Methods
- Results Chapters:
    - Chapter 1. Erythroblast Forward Programming
    - Chapter 2. Testing Different Polycistronic Cassettes
    - Chapter 3. Inducible Forward Programming
    - Chapter 4. Understanding the Heterogeneity of Long-Term Megakaryocyte Cultures
- Discussion and Future Perspectives
- Acknowledgements
- Bibliography

List of Diagrams

List of Tables

List of Vector Maps

List of Figures

Chapter 1. Erythroblast Forward Programming

Table of Abbreviations

3G	Third generation
AAVS1	Adeno-associated virus integration site 1
Ab	Antibody
ADP	Adenosine diphosphate
BCL-2	B-cell lymphoma 2
BEL-A	Bristol erythroid line adult (immortalised cell line)
BFU-E	Burst-forming unit erythroid
bHLH	Basic helix-loop-helix
BM	Bone marrow
BSA	Bovine serum albumin
Ca^{2+}	Calcium ion
CaR	Calcium receptor
CFU	Colony-forming unit
CFU-E	Colony-forming unit erythroid
CHIP-PCR	Chromatin immunoprecipitation with quantitative polymerase chain reaction
CHIP-seq	Chromatin immunoprecipitation with massively parallel DNA sequencing
c-KIT	Stem cell factor receptor (cytokine receptor-tyrosine kinase)
CLOUD-HSPCs	Continuum of low-primed undifferentiated haematopoietic stem and progenitor cells
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
cMPL	Thrombopoietin receptor
CMV	Cytomegalovirus
c-MYC	V-Myc Avian Myelocytomatosis Viral Oncogene Homolog
CRISPR/Cas9	Clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins
crRNA	CRISPR-RNA
DAG	1,2-diacylglycerol
δ-granule	Dense granule
DNA	Deoxyribonucleic acid
Dox	Doxycycline
DSB	Double strand break
dTomato	Dimeric Tomato
E2A	Equine rhinitis A virus
EB medium	Erythroblast medium
eGFP	Enhanced green fluorescent protein
EPO	Erythropoietin
EPOR	Erythropoietin receptor
ERG1	ERG, ETS transcription factor 1
Ery	Erythrocyte
Ery-FoP	Erythroblast forward programming
ESC	Embryonic stem cell
Ext	Extension
FA	Formaldehyde
FACs	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
FGF2	Fibroblast growth factor 2

FGF-4	Fibroblast growth factor 4
FLI1	Friend leukaemia integration transcription factor 1
FLT-3	FMS-like tyrosine kinase 3
FoP	Forward programming
FoP-Erys	Erythroblasts produced by forward programming
FoP-MKs	Megakaryocytes produced by forward programming
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GATA1	GATA binding factor 1
GATA2	GATA binding factor 2
G-CSF	Granulocyte colony-stimulating factor
GMP	Granulocyte-Macrophage progenitor
GMP	Good manufacturing practice
GPIbα	Glycoprotein Ibα
GPIIB	Glycoprotein IIB
GPVI	Glycoprotein VI
h(ESC/iPSC)	Human (ESC/iPSC)
Hb	Haemoglobin
HbA	Adult haemoglobin
HbF	Foetal haemoglobin
HDR	Homology directed repair
HEK293T	Human embryonic kidney 293 cells
HEL cell line	Human erythroleukaemia cell line
HGF	Hepatocyte growth factor
HLA	Human leukocyte antigen
HMBS	Hydroxymethylbilane synthase
HR	Homologous recombination
HSC	Haematopoietic stem cell
HSPCs	Haematopoietic stem and progenitor cells
HTC116	Human colorectal carcinoma cell line
I	Insulin
IL-	Interleukin- (eg. IL-3)
imMKCL	Immortalised megakaryocyte progenitor cell line
IMS	Invaginated membrane system
indels	Insertions or deletions
iPSC	Induced pluripotent stem cell
IRES	Internal ribosome entry site
ITS	Insulin-transferrin-selenium
Kb	Kilobase
KLF1	Krüppel-like factor 1
KLF4	Krüppel-like factor 4
KOSR	KnockOut serum replacement
LAA	L-Asorbic acid 2-phosphate sesquimagnesium salt hydrate
LDB1	LIM domain binding 1
LIF	Leukaemia inhibitory factor
LMO2	LIM domain only 2
lncRNA	Long noncoding RNA
LSSmOrange	Large Stokes shift monomeric orange
LT-HSC	Long term haematopoietic stem cell
LY1	Ly1 antibody reactive
MARS-seq	Massively parallel single-cell RNA-sequencing

MEF	Mouse embryonic fibroblast
MEF2c	Myocyte-specific enhancer 2c
MEP	Megakaryocyte-Erythroid progenitor
miRNA	microRNA
MK	Megakaryocyte
MK-FoP	Megakaryocyte forward programming
MMP	Multipotent progenitor
NHEJ	Non-homologous end joining
NHS	National health service
NK	Natural killer
NPV	Negative predictive value
Oct4	Octamer-binding transcription factor 4
ON	Over night
OPTi-OX	Optimised induced overexpression
ORF	Open reading frame
P2A	Porcine teschovirus-1
PAM	Protospacer adjacent motif
PBX1	Pre-B cell leukaemia transcription factor 1
PenStrep	Penicillin and Streptomycin
PF4	Platelet factor 4
PPV	Positive predictive value
PSC	Pluripotent stem cell
PtdSer	Phosphatidylserine
Puro	Puromycin
qPCR	Quantitative polymerase chain reaction
RBC	Red blood cell
RhD	Rhesus D
RIPA	Radioimmunoprecipitation assay
rLV	Recombinant lentivirus
RNA	Ribonucleic acid
RT	Room temperature
RT-qPCR	Real time quantitative polymerase chain reaction
rTTA	Reverse tetracycline trans-activator
RUNX1	Runt related transcription factor 1
SCF	Stem cell factor
scRNA-seq	Single-cell RNA-sequencing
SDF-1	Stromal-derived factor 1
SOX2	Sex determining region Y
T	Transferrin
T2A	Thoseaasigna virus
TAL1/SCL	Stem cell leukaemia
TALEN	Transcription activator-like effector nucleases
T-ALL	T-cell acute lymphoblastic leukaemia
TCR	T-cell receptor
Tet	Tetracycline
TF	Transcription factor
TG	Transgene
TPO	Thrombopoietin
tracrRNA	Trans-activating crRNA
TRE	Tetracycline responsive element

VEGF	Vascular endothelial growth factor
vWF	Von Willebrand Factor
WHO	World health organisation
WT	Wild type
ZFN	Zinc finger nuclease

Abstract

Forward Programming to Produce Megakaryocytes and Platelets

There exists a need for the production of platelets *in vitro*, to overcome several issues surrounding current platelet collection including; donor-dependency, cost, biosafety and increasing demands. Our lab uses a technique called 'Forward Programming' (FoP), whereby three key haematological transcription factors (TFs) are overexpressed in human pluripotent stem cells (PSCs), in order to produce mature megakaryocytes (MKs), as an *in vitro* alternative to providing platelets. The mature MKs produced are higher in number and purity compared to existing methods of MK production from PSC that use a directed differentiation approach. FoP-MKs are capable of producing functional platelets and have been shown to activate and take part in thrombus formation both *in vitro* and *in vivo*. Our protocol uses a minimal number of cytokines, follows good medical practice (GMP) guidelines and is currently being employed by a commercial company (Platelet Biogenesis, of the USA) to produce a high number of *in vitro* MKs, with the goal of producing platelets for use in transfusion medicine.

Forward Programming Produces Bi-Potent Progenitor Cells

During my PhD I have shown that FoP, using the same three TFs (*GATA1*, *TAL1* and *FLI1*), generates a bi-potent progenitor cell population, which can produce MKs as well as immature red blood cells, erythroblasts. I have shown that FoP produces a high number of erythroblasts when culture conditions are altered to favour red cell development by the exchange of thrombopoietin (TPO) used in MK culture conditions, for erythropoietin (EPO). The maturation of these cells is currently limited and reflects similar maturation levels found by other researchers producing red cells from PSCs, with poor enucleation and expression of embryonic and foetal, but not adult, globins. The bi-potent progenitor cell population produced mimics the Megakaryocyte-Erythroid Progenitor (MEP) cell found *in vivo*. MEPs are difficult to study *in vitro*, as they cannot be easily isolated or cultured. Therefore, the discovery that FoP can generate this cell type has important implications for understanding the basic biology of this cell type and what governs the cell-fate making process, and also offers new perspectives for translational medicine.

Removing Lentiviral Requirement by Producing an Inducible Stem Cell Line for Forward Programming

During my PhD I have worked on overcoming a major limitation of the forward programming

technique, which is a reliance of lentiviral vector transduction to overexpress the three TFs. The use of lentiviruses is costly, results in heterogeneous cultures and presents regulatory issues, potentially limiting the use of the resulting cell product in humans. The main aim of my work was to remove the need for lentiviral transduction to effect forward programming. To do this, I first demonstrated that the three TFs inserted into the same lentiviral vector backbone resulting in a polycistronic cassette, allowed forward programming to occur. I produced a number of polycistronic cassettes, with the three TFs in different configurations, and found that when *FLI1* was placed in the centre of the three TFs an erythroblast bias was observed. A vector containing *FLI1* in the final gene position however, gave promising results for both MK and erythroblast production, high-lighting that the polycistronic vector was also able to produce a bi-potent progenitor population. This vector was subsequently chosen to target to a genomic safe harbour, using genome editing techniques, to generate an inducible PSC line.

The resulting inducible PSC cell line stably expresses the three TFs directly from its genome after chemical induction with doxycycline, utilising the TET-ON inducible system. I have shown that TF expression is strictly controlled, meaning the inducible line can be kept in culture long term, as for normal PSCs. An advantage of inducible FoP over the lentiviral method is that it results in minimal cell death at the start of programming and overall efficiency of programming is higher. Data from multiple clones show that inducible FoP is a highly efficient method for producing mature MKs. Over long-term culture, of up to 100 days, the purity of the MKs produced is higher (>90%) when the inducible cell line is used, compared to cells produced by the lentiviral method, which decrease in purity overtime (to approximately 60%).

Using the inducible PSC line, FoP in EPO produces a highly pure red cell culture (98%) in 20 days, including a percentage of red blood cells expressing maturity markers, such as Rhesus D, and showing enucleation. These cells are indicative of maturation to the final reticulocyte red cell stage. Similar to what I observed for MK cultures, purity is also improved in induced cells cultured in EPO, compared to those which are forward programmed using the lentiviral method (98% versus approximately 50%).

Removing the requirement of lentiviral transduction in FoP reduces the complexity and the biosafety risk of the method. It also increases efficiency and improves cell outcome, as well as greatly reduces the cost. This work demonstrates a major advancement in forward programming that will make it more amenable to scaling-up, which will help to make the jump from bench to bedside more attainable in the near future, moving this technology closer towards our goal of producing *in vitro* platelets for use in transfusion medicine.

Characterising the Megakaryocyte Progenitor Population

Finally, I have developed single cell progenitor assays to try and identify the progenitor population in FoP-MK cultures at early (day 9-13) and late (day 40) time-points. Early-time point progenitors appear to be characterised by CD41 and CD235 co-expression. Long-term progenitors appear to be characterised by low CD42 expression and high CD235 expression.

Understanding Megakaryocyte Heterogeneity in Long-Term Cultures

I performed single cell RNA-seq (scRNA-seq) on long-term FoP-MKs, which expressed different Rainbow vector combinations, therefore, expressed different combinations of transgenes (TGs). The results show that despite the absence or presence of different TG markers, all cells showed equal expression of total (endogenous plus exogenous) *GATA1*, *TAL1* and *FLI1*. Within the cells sequenced, four populations of cells were discovered; mature MKs, MKs, MK progenitors and cells which were dying and potentially represent mature MKs that are poised to give rise to platelets. scRNA-seq has provided a new list of novel candidate genes which can help in future to isolate and study the progenitor cells identified.

Introduction

1. Clinical Overview of Transfusion Medicine

1.1 The History of Blood Transfusion

The transfusion of blood refers to the administration of whole blood or blood components. The first human-to-human transfusion was performed in 1818, by the English obstetrician James Blundell, in order to treat a patient experiencing major bleeding (Blundell J, 1818). Blundell described the use of a syringe for this transfusion and made note of avoiding air intake into the veins, as well as the incompatibility of heterologous donors. It the year 1900 the ABO blood system was classified by Landsteiner, for which he was later awarded a Nobel Prize, after observing that red cells from certain patients would agglutinate with the serum of others (Landsteiner K, 1900). Currently there are 36 recognised red blood group systems, representing more than 300 different antigens, as listed by the International Society of Blood Transfusion. As such, matching blood products to patients is imperative to avoid the risk of severe and potentially fatal transfusion reactions, with all transfusion being antigen matched for the two most common antigens, ABO and Rhesus D (RhD).

1.2 Blood Donation

Blood donations are made by two methods; either as a whole blood donation which can be processed subsequently into separate platelet, red blood cell and plasma components, or alternatively as a single compartment donation, using apheresis to return the remaining compartments back to the donor. In industrialised countries blood component therapy replaced whole blood therapy in the middle of the twentieth century, thanks to a large number of inventions including refrigeration, anticoagulant and preservative solutions, plastic blood bags and infectious disease testing (Arya RC, 2011). As such, blood component therapy is currently considered the gold standard treatment for the majority of transfused patients. The most important consideration for any blood product intended for transfusion is that it remains of high-quality and will deliver the intended benefits to a patient. This is a complicated task, considering that some blood components are stored for considerable lengths of time before use, and with each blood component requiring specific storage conditions. For example, red cells require cold temperature storage, just above freezing, while platelets require room temperature storage and continuous agitation to prevent aggregation (Arya RC, 2011). Additionally, some blood components undergo additional steps after collection, such as irradiation, leukodepletion or cryopreservation, which adds further complexity to their storage and usage (Acker JP, 2016).

1.3 Regional Procedures and Dependence on Transfusions

Differences between regional health service regulations means that there is no global standard procedure for the storage of blood cell components. For example, in the United States of America (USA) standard red cells units can be stored for up to 42 days while in the United Kingdom (UK) they are stored for just 35 days. In the USA platelets are stored for 5 days, while in the UK this can be extended to 7 days if bacterial screening has been implemented (Arya RC, 2011; Harris AM, 2012). Fifty to eighty percent, depending on the country, of administered transfusions relate to very few clinical situations. These are; severe haemorrhage related to pregnancy or childbirth, trauma- usually as a result of road traffic accidents and severe anaemia in the young- often a consequence of malaria (Ala F, 2012).

While most wealthy countries have an adequate supply of blood, the use of which is largely pre-planned and predictable, this is not the case for certain areas of the world. In wealthy countries blood donors and donated blood are rigorously screened, meaning the frequency of disease transmission from transfusions is very low, but certain regions lack the infrastructure or resources to implement such screening measures (Leparc GF, 2015). In sub-Saharan Africa there are two major issues surrounding transfusions: unsafe blood and blood shortages. These issues can often lead to serious health consequences, such as the transmission of life threatening conditions including hepatitis and HIV, and can lead to fatalities. The lack of blood for emergency situations in sub-Saharan Africa is the leading cause of maternal death due to postpartum haemorrhage. Blood shortages can be the result of; non-affordable blood, a lack of donors, unwillingness of relatives to donate and inadequate supplies or transport (Bates I, 2008).

1.4 Use of Donated Blood

In low-income countries, the majority (67%) of blood transfusions are administered to children below the age of 5, whereas in high-income countries the majority (79%) are used in the over 60s. In high-income countries transfusions are used as supportive care for patients undergoing cardiovascular or transplant surgery, after major trauma, or as therapy for patients with cancer and haematological malignancies (WHO, 2016). The World Health Organisation (WHO) recommends whole blood use in severe malarial cases of anaemia and overall whole blood use remains high in low-income countries (85%) (WHO, 2015). However, the majority of transfusions in high-income countries deliver blood components, not whole blood, due to improved diagnostic and treatment options, as a result of better developed healthcare.

Red cell units are administered with the main aim of improving oxygen delivery to organs and are primarily used to treat anaemia. Anaemia can develop; as the result of severe blood loss (during surgery or as a result of trauma), in situations affecting the normal production of red cells (such as during chemotherapy treatment), as a consequence of defective haemoglobin (such as in thalassaemia or sickle cell anaemia), or by the increased destruction of red blood cells (during malaria infections) (Acker JP, 2016).

Platelet concentrates are administered for treating patients with acute haemorrhage and for preventing bleeding in patients with a severe reduction in their platelet count, known as thrombocytopenia, which is a common complication arising from haematological malignancies or their treatment (Stanworth SJ, 2015). Platelet concentrates are kept at room temperature, increasing the risk of bacterial infection, and their the haemostatic and metabolic function are known are deteriorate overtime, in a process called platelet storage lesion, owing to their short shelf life (Védy D, 2008; Thon JN, 2008). Due to the short shelf life of platelets it can be difficult to keep up with the demand for their use, especially over holiday periods when typically fewer donors are available, and also makes administration difficult in remote areas.

1.5 Limitations of Transfusion Medicine

The major limitation of transfusion medicine is maintaining an adequate and safe supply of blood or blood components from donors. Many countries rely solely on voluntary non-remunerated donors, based on the assumption that donors may lie about their health if donation is accompanied by a monetary reward, which can increase the prevalence of HIV infected blood (Ala F, 2012). There is good evidence for avoiding unnecessary red cell transfusions, to minimize the risk of transfusion-transmitted infections or adverse reactions, which has accounted for a 3-4% annual decrease in the demand for red blood cell transfusions in the UK, a trend which is also reflected globally (Carson JL, 2016; NHSBT, 2016). Conversely, the demand for platelet transfusions in recent years has risen due to an ageing population, an increase in the incidence and prevalence of patients suffering from haematological malignancies, and changes in the treatment of such malignancies (Estcourt LJ, 2014).

An increase in platelet transfusion demand, alongside the short shelf life and issues surrounding their biosafety, means there exists a clinical need to find an alternative supply of platelets for use in transfusion medicine. The pursuit of generating red blood cells *in vitro* is also being heavily investigated, since some global regions still have higher demands than can be met by supply. Therefore, deriving blood cells from alternative sources, such as embryonic or induced pluripotent stem cells, could offer a number of benefits. Deriving blood cells *in vitro* could offer the opportunity

to produce universal cells that are pathogen free and abundantly available. There are a number of labs in the world pursuing different avenues to try and achieve these goals (discussed later), including our own. In order to understand how researchers are attempting to produce blood cells *in vitro*, it is important to first understand how this process occurs *in vivo*.

2. Haematopoietic Overview

2.1 Blood Cell Development

Haematopoiesis describes the process of blood cell production. More than 100 billion mature blood cells are produced per day in human adults. In haematopoiesis, haematopoietic stem cells (HSCs) differentiate into the entire repertoire of blood cells found in the vascular system, generating a huge diversity of cell types with broad-ranging functions; from carrying oxygen, to thrombus formation, to fighting infection. HSCs normally reside in the bone marrow of adults and have been extensively studied, especially in mouse models of transplantation, due to the fact that it has been possible to purify these cells. It has been shown that a single HSC, transplanted into a lethally irradiated mouse, is capable of reconstituting the entire repertoire of cells of the blood system (Osawa, 1996).

'The Haematopoietic Tree' (***Diagram 1***) shows the main cell types produced during haematopoiesis, with HSCs sitting at the top of this hierarchy. HSCs have multipotent potential, while each branch of the haematopoietic tree below it represents a cell which is more restricted in its differentiation potential. The range of cells produced can be separated into two main lineages; the lymphoid and myeloid lineages. The earliest differentiation event for HSCs results in a multipotent progenitor cell (MMP), a cell which maintains multi-potent potential but has lost the ability to self-renew. MMPs further differentiate into the oligo-potent common lymphoid progenitor (CLP), or, common myeloid progenitor (CMP) (Reya T, 2001). Mature lymphoid cells include T lymphocytes, B lymphocytes and natural killer (NK) cells. Mature myeloid cells include megakaryocytes, erythrocytes, monocytes, macrophages, dendritic cells, eosinophils, basophils and neutrophils.

Recently the hierarchical view of haematopoiesis has been challenged by several research groups, following the advent of single-cell profiling. Using eleven surface markers to index sort and perform single-cell RNA-seq alongside functional characterisation of cells, it has been proposed that HSCs and their immediate progenitors represent a continuum of low-primed undifferentiated haematopoietic stem and progenitor cells, so called 'CLOUD-HSPCs', from which uni-lineage-restricted cells arise directly (Velten L, 2017). The model of differentiation proposed is incompatible with the hierarchical model, where HSCs pass through distinct progenitor stages, becoming more lineage restricted with every branch-point crossed. Instead, it has provided evidence that CLOUD-HSPCs gradually acquire

lineage priming at the transcript level in multiple directions, with a bias towards certain cell-fates. In this study, HSCs were found to express genes related to the lymphoid/myeloid and megakaryocyte/erythroid lineages, suggesting this is one of the earliest transcriptional priming events occurring in primitive HSCs, which was confirmed by functional data.

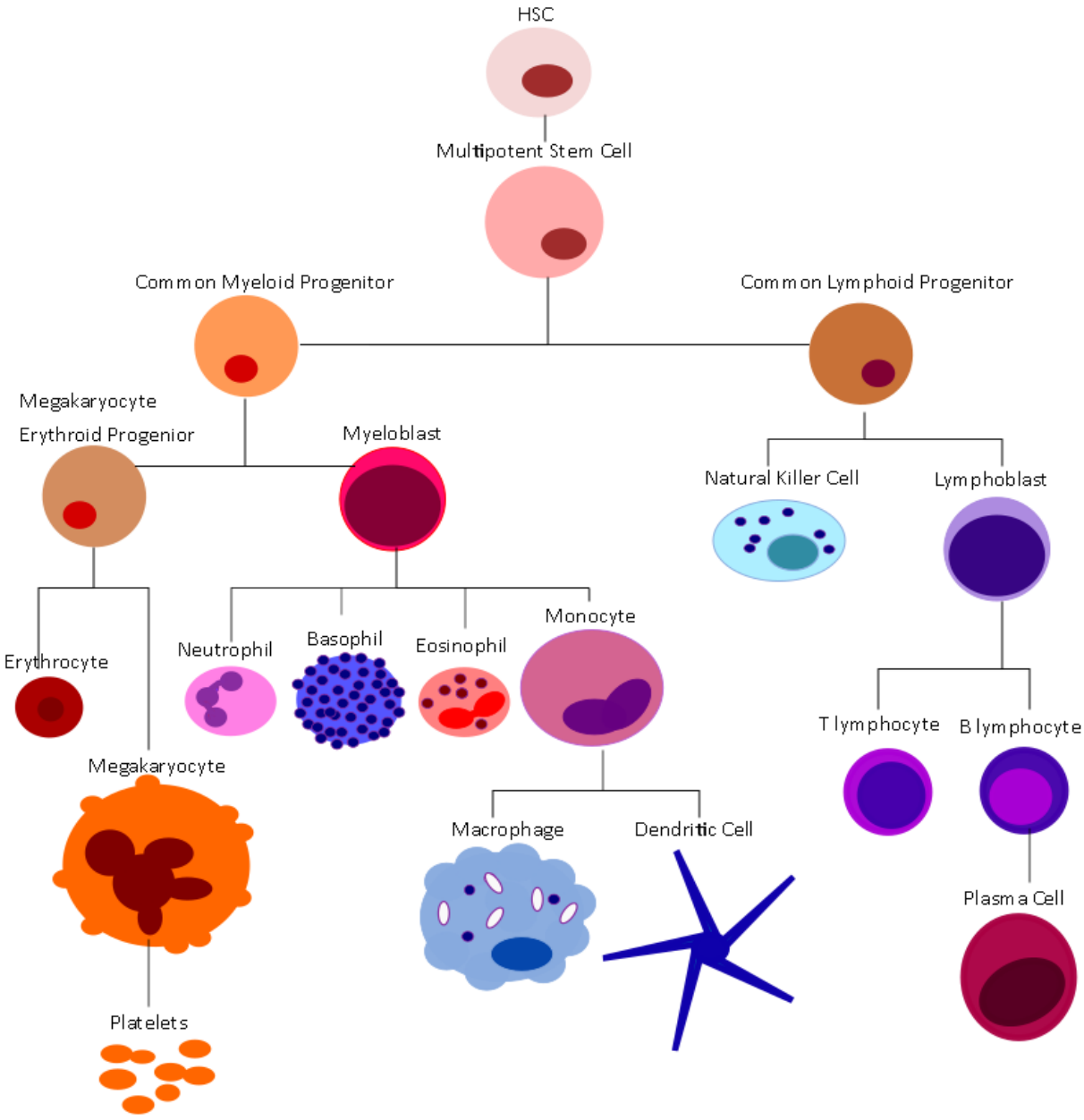


Diagram 1- The Haematopoietic Tree. Haematopoietic stem cells (HSCs) sit at the top of the hierarchy and are capable of producing all types of haematopoietic cells found in the body.

2.2 Lineage Commitment in Haematopoietic Progenitors

The way in which a progenitor cell can give rise to more than one differentiated cell type has been an area of intense study, as scientists are keen to understand the biological mechanisms driving fate-decision choices. It is well documented that transcription factors (TFs) play a major determining role in deciding a progenitor cells fate. For example the TFs *GATA1* and *PU.1* act in opposing lineage specific and indeed lineage-specifying ways. *GATA1* and *PU.1* switch on separate networks of effector genes to instruct a common myeloid progenitor (CMP) to differentiate into either a megakaryocyte-erythroid progenitor (MEP) or granulocyte-macrophage progenitor (GMP), respectively (Ferreira R, 2005; Rosmarin AG, 2005).

Two main mechanisms were proposed to try and describe events leading to the decision-making process, and the resulting gene regulatory network changes, in progenitor cells. The first is described as the 'intrinsic' or stochastic model of cell fate control. In this model, cell-fate is thought of as a pre-existing program, adopted by a progenitor cell in a random fashion. This model considers external factors, such as cytokines, to lack fate determining capacity. Instead, they simply act as growth or survival signals in a cell which is already lineage committed. The second model is described as the 'extrinsic' or deterministic model, where external signals are considered to be responsible for altering the genetic program of a cell. Acting to either switch on or switch off the appropriate set of genes governing a differentiation state, thus, initiating the signal transduction cascades responsible for determining a differentiated cell fate (Enver T, 1998).

It has been shown that these two mechanisms actually work together to decide the cell-fate decision of progenitors (***Diagram 2***). Mathematical modelling of the well-established interactions of *GATA1* and *PU.1* describes CMPs as being in a 'metastable' cell state, between the two cell fates it can differentiate into. It was modelled that to leave this metastable state two processes had to occur. First, the cell had to leave its progenitor state. The way in which MEPs or GMPs did this was determined by distinct signals, resulting in a change of their global transcriptomes. This global change occurred in a very similar fashion whether a CMP differentiated into a MEP or GMP. This has been described as a common destabilisation phase, occurring over a time period of 24-48 hours. The second process that occurs is an abrupt change in the transcriptome, after which point cells cannot be forced to change their differentiation trajectory by changing media signals, indicating that the point of commitment occurs between 24-48 hours of exiting the progenitor state. The near symmetrical bifurcation model that occurs at the start of the destabilisation phase coincides with experimental evidence of the observed stochasticity of fate determination (Enver T, 1998). However, cytokines can skew the bias of this bifurcation process, by initiating small intrinsic changes

of gene expression which contribute to the larger transcriptome changes later acquired, since cell death is low in media conditions which promote either the MEP or GMP lineage. Thus, both the 'intrinsic' and 'extrinsic' mechanisms control the cell-fate decision making in progenitors (Huang S, 2007).

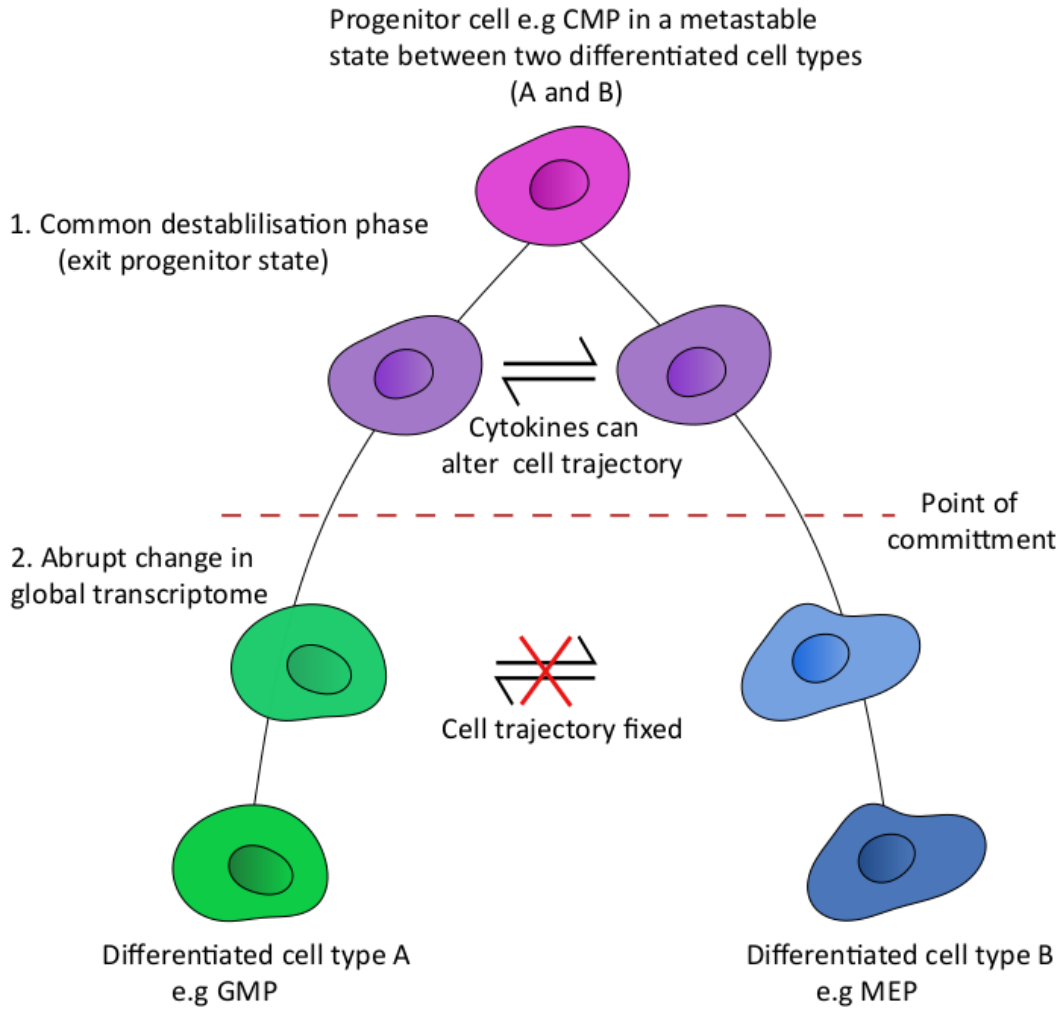


Diagram 2- The two stages of progenitor cell fate decision. Progenitor cells exit their progenitor state, towards a more differentiated cell type. This step, known as the common destabilisation phase, occurs similarly in cells following different cell fate trajectories. At this stage, cytokines are able to re-direct the cell fate trajectory. The second step involves abrupt global transcriptome changes, after which point cells are committed to their final cell fate.

2.3 Megakaryopoiesis

A common bi-potent megakaryocyte erythroid progenitor (MEP) cell has been identified at the last branch point before the differentiation of erythrocytes and megakaryocytes (MKs) (Klimchenko, 2009). MKs are very large polynucleated cells that reside in the bone marrow in small numbers. They are the precursor cells responsible for producing platelets, with over 10^{11} platelets being produced on average per day (Bluteau, 2009). Megakaryopoiesis describes the formation of MK cells from HSCs. Megakaryopoiesis and platelet production can be described in four main biological steps;

1. Haematopoietic stem cell (HSC) maintenance in the bone marrow (BM).
2. HSC differentiation of megakaryocyte progenitor cells.
3. Megakaryocyte production, maturation and endomitosis.
4. Platelet assembly and release.

2.3.1 HSC Maintenance in the Bone Marrow

Adult stem cells are present in self-renewing tissues such as the skin, intestinal epithelium and haematopoietic system. Typically a stem cell niche is an environment in which stem cells reside in order to self-renew, proliferate, and produce a sub-set of progenitor cells. These can then differentiate and mature, producing a large number of specialised cell types, enabling tissue repair and maintenance. In 1978, a stem cell niche for HSCs in the BM was first hypothesised as an environment that tightly controls stem cells entering the cell-cycle, to prevent them from exhaustion and DNA damage (Schofield, 1978). The BM is made up of different types of niche environments that play an important role in HSC maintenance and differentiation, by providing a complex mixture of cellular, physical, chemical and structural cues. It has long been known that a small number of HSCs and progenitors can also be found in the peripheral blood (Goodman, 1962), and that these cells are capable of homing back to the bone marrow (Wright, 2001).

Single cell tracking of HSCs and progenitors has found that the most multipotent cells (HSCs) home to the endosteal region of the BM, found close to the bone surface, while more committed haematopoietic cells are found localized in the more central bone marrow region (Nilsson, 2001). Due to the observed preference in environment of HSCs, many investigators have tried to identify the cell types found in the endosteal niche, to try and elucidate the exact factors which are essential for stem cell maintenance. While the full picture still remains elusive, progress has been made in discovering a number of cell types and factors that play a role in this tightly regulated process.

A large number of different cell types are important for maintaining stem cells in a niche environment. In the BM, mesenchymal stem cell derived cells, such as osteoblasts, play an important

role in HSC maintenance. *In vitro*, osteoblasts have been shown to support the expansion of human $CD34^+$ haematopoietic progenitor derived cells, through their secretion of hepatocyte growth factor (HGF) and granulocyte colony-stimulating factor (G-CSF) (Taichman, 2001). Not only important for HSC maintenance, osteoblasts have also been implemented in maintaining quiescence in HSCs in the adult BM, through a number of signalling pathways. Angiotensin-1 produced by osteoblasts activates the receptor tyrosine kinase Tie2, found on the surface of HSCs. When bound to Tie2, angiotensin-1 causes HSCs to tightly adhere to osteoblasts in the niche, resulting in enhanced HSC survival and quiescence (Arai, 2004). It has also been demonstrated that thrombopoietin, produced by osteoblasts, binds to its receptor, cMPL, on the surface of long-term HSCs (LT-HSCs), resulting in cell contact with the osteoblasts and quiescence (Qian, 2007). Activated osteoblasts, found at sites of bone-remodelling, have also been demonstrated to negatively regulated HSC pool size, thus reduce expansion by producing osteopontin, an acidic matrix glycoprotein which binds to CD44 and α4 integrin receptors on the HSC surface (Stier, 2005).

Another cell type, osteoclasts, help to maintain the HSC pool in the endosteal niche. Osteoclasts regulate a process known as bone resportion, where bone tissue is broken down to assist in repair, remodelling and maintenance of bones. This process results in secreted calcium, which also helps to maintain blood calcium levels. During development, HSCs translocate to the BM for adult haematopoiesis, from the fetal level. It is thought that the calcium rich environment found in the endosteal niche is an important signal for HSCs, detected by the calcium sensing receptor (CaR) on their cell surface, that helps HSCs home to the BM (Adams, 2006).

Nestin positive mesenchymal stem cells (MSCs) have also been shown to be important niche cells, that highly express HSC maintenance genes and downregulate these genes when stimulated. MSCs closely associated with HSCs physically, and if deleted result in a significant reduction in the number of HSCs found in the BM (Méndez-Ferrer, 2010). MSCs can be split into two populations. Of interest, one sub-population is comprised of very rare cells (~0.002% of bone marrow) that highly express nestin and are found only along small arterioles within the endosteal niche, known as periarteriolar cells (Kunisaki, 2013). Depleting periarteriolar cells alters the location of HSCs away from arterioles, and is associated with HSCs phenotypically switching from a quiescent to non-quiescent state. This leads to a statistically reduced frequency of HSCs in both the BM and the spleen, indicating that periarteriolar cells help to maintain the pool of HSCs. As evidenced by the large body of work gone into trying to identify the main components of HSC maintenance, the BM niche responsible is very complex and populated by a large number of cell types which are important to this.

2.3.2 HSC Differentiation of Megakaryocyte Progenitor Cells

As previously described, HSCs give rise to multipotent progenitor cells, which in turn give rise to the megakaryocyte erythroid progenitor (MEP) cell. Historically, studies on HSCs and their progenitors have relied on fluorescence-activated cell sorting (FACS), using monoclonal antibodies to stain a specific panel of surface receptors. This has been done with limited success due to many antibodies staining more than one cell type. At best, typically only 50% of HSCs isolated in this way are active in transplantation assays (Kent DG, 2009). However, in more recent years, thanks to the development of new sequencing and transcriptomics techniques, as well as the scaling down of these technologies to the single-cell level, the classical hierarchical view of haematopoiesis is changing.

A new era of single-cell biology is emerging, with investigators no longer satisfied at looking at cells at the whole population level, because this can lead to heterogeneity when looking at gene expression, for example. Multiple cell types may be present in a 'purified' population and different cell cycle or transcriptional states of a specific population of cells can lead to heterogeneity. When looking at gene expression change in a population, it is hard to decipher whether the change is happening at the level of a few individual cells, or in a much larger sub-set of the population. Thus, information can be lost when looking at cells at a whole population level. There is also an issue with rare cell types, such as HSCs. These can be difficult to isolate and investigate for conventional population studies that often require large cell numbers. A major advancement in redefining the HSC phenotype was achieved using a combination of single-cell RNA-seq, to interrogate gene expression, and functional transplantation assays. A sub-set of HSCs was identified and enriched for, using a molecular dataset that linked long-term self-renewal and functional stem cell activity. This improved the purification of true HSCs to 67%, compared to the 50% previously achievable (Wilson, 2015).

The distribution and transcriptional states of myeloid progenitor cells has been reported with conflicting messages historically. A study using index sorting set to settle this, by measuring common surface markers of CMPs, and their two distinct progeny cell types; MEPs and GMPs (Paul F, 2015). Massively parallel single-cell RNA-seq (MARS-seq) and clustering analysis was performed on almost 3000 single-cells, allowing very rare cell types to be sequenced with a cell frequency as little as 0.2% (5–6 cells). From differential expression pattern analysis of almost 3500 genes, 19 transcriptionally distinct populations of cells were identified. Some of these sub-populations formed groups which expressed marker genes corresponding to expected cell phenotypes (erythrocyte, monocyte, and neutrophil progenitors). However, some had overlapping gene expression distribution, not in line with the classical hierarchical model of haematopoiesis. In fact, in contrast to the classical model, their study established that all myeloid progenitors are transcriptionally primed towards one of

seven (or more) different fates; MKs, erythrocytes, monocytes, neutrophils, eosinophils, basophils or dendritic cells. This diversity of primed cell types may represent earlier commitment in development than previously believed, or a more plastic regulation of different states.

Focusing on the MEP lineage, 8 sub-populations were index sorted based on MEP cell surface markers. Only one of the 8 sub-populations represented MK progenitors and showed expression of well-established MK markers, CD41 and platelet factor 4 (Pf4), and MK TFs, pre-B cell leukaemia transcription factor 1, friend leukaemia integration 1 transcription factor and myocyte-specific enhancer 2c (*Pbx1*, *Fli1*, *Mef2c*). The other 7 sub-populations were separated along a gradient of erythrocyte transcription, from an early progenitor expression profile, with progressive erythrocyte functional gene activation to a mature profile that highly expressed haemoglobin. The authors suggest that this represents a developmental continuum of differentiation. The least mature of these erythrocyte sub-populations, with the most progenitor potential, along with the MK progenitor sub-population shared some major transcriptional regulators such as GATA binding factor 2 (GATA2), not shared with the monocyte or granulocyte lineages.

In functional competitive assays, $CD41^+$ MK progenitors gave rise to a large proportion of erythrocytes, despite having a transcriptional association distinct to the MK, not the erythrocyte, lineage. $CD41^+$ progenitors also display an epigenetic profile partially compatible with HSCs, determined by chromatin analysis. This highlights that although transcriptional state is clearly an important factor in dictating cell fate outcome, MK progenitors are not completely lineage restricted. The observation of this plastic, rather than strictly controlled, progenitor cell fate may offer advantages to an organism when it encounters haematological stress. This study demonstrates how single-cell analysis is challenging the classical idea of haematopoiesis. It presents strong evidence for the transcriptional state of a cell determining its functional commitment, however, highlight that this process can be fluid and cell fate can be changed. Despite single-cell analysis not identifying a distinct MEP population, rather a couple of progenitor sub-populations that are already transcriptionally primed towards either the MK or erythrocyte lineage, this study identifies that the two cell lineages are closely connected developmentally.

2.3.3 Megakaryocyte Production, Maturation and Endomitosis

MKs account for only 1% of all myeloid cells, thus are a very rare cell type, making them difficult to study (Ogawa M, 1993). They are found primarily in the bone marrow, but also in the lungs and peripheral blood. A recent study in mice has suggested that the lungs, previously thought to play only a minor role in megakaryopoiesis, may in fact be a major site of platelet production (Lefrançais

E, 2017). Once HSC differentiate to increase the number of MK progenitors, these cells differentiate further into immature MKs with limited self-renewal capacity that must undergo maturation. The first step of maturation involves a MK-specific type of cell division, called endomitosis.

During endomitosis MKs undergo multiple rounds of cell division, replicating their DNA content as expected, but without cellular division. This is due to a failure of cytokinesis (cytoplasmic) division, often coupled with failure of karyokinesis (nuclear) division (Lordier L, 2008). Multiple rounds of endomitosis occur, resulting in MKs with many times the normal compliment of chromosomes (46), leading to a polylobulated nucleus. A single MK can accumulate a DNA content of up to 128n through endomitosis (Zimmet J, 2000). After this, cytoplasmic enlargement occurs, by a process called cytoplasmic maturation. This, along with endomitosis, is considered a fundamental property specifically acquired by the MK lineage to facilitate a high quantity of platelet production, by increasing mRNA and protein content (Reems JA, 2010; Zimmet J, 2000).

During cytoplasmic maturation, platelet-specific granules form, along with a network of specialised membranes that forms the invaginated membrane system (IMS, also known as the demarcation system- DMS), which adds to the MK cell volume (Yamane A, 2008). The IMS, a marker of MK maturation, is found throughout the cytoplasm as a complex of cisternae and tubules. For some time this system was thought to function by separating MK cytoplasm into small areas where platelets were assembled before release (Yamada E, 1957). Since, it has been shown that the IMS complex is continuous with the plasma membrane, and it is now thought to exist as a reservoir of membranes utilised during proplatelet formation (Schulze H, 2006). A protein often found in the plasma membrane skeleton of cells, spectrin, is important for stabilising the IMS, as it provides structural support by forming a 2D lattice. A study which disrupted the assembly of spectrin tetramers in mouse MKs resulted in mature MKs with an underdeveloped IMS, that were unable to produce proplatelets (Patel-Hett S, 2011). This suggests that spectrin is intimately involved in the assembly or stabilisation of the IMS, which is critical for proplatelet formation.

Other markers of MK maturation include the development of a dense tubular system, found to be the site of platelet prostaglandin biosynthesis, important for vasoconstriction and platelet aggregation (Gerrard JM, 1976). Another important structure found in mature MKs is the open canalicular system, a surface-connected channelled system important for granule release (Patel SR, 2005). Proteins specifically associated with platelets are found on the surface of mature MKs, such as the Von Willebrand Factor (vWF) and fibrinogen receptors. Other proteins, including vWF, are packaged into secretory granules and loaded into α-granules derived from Golgi complexes. Other proteins, including fibrinogen, are added to platelet-specific granules through endocytosis or

pinocytosis of plasma (Heijnen HF, 1998). Mitochondria and dense granules, also derived from Golgi complexes, are assembled during MK maturation. The resulting MK cell is terminally differentiated and equipped with all the necessary apparatus required to perform the major undertaking of platelet biogenesis.

2.3.4 Platelet Assembly and Release

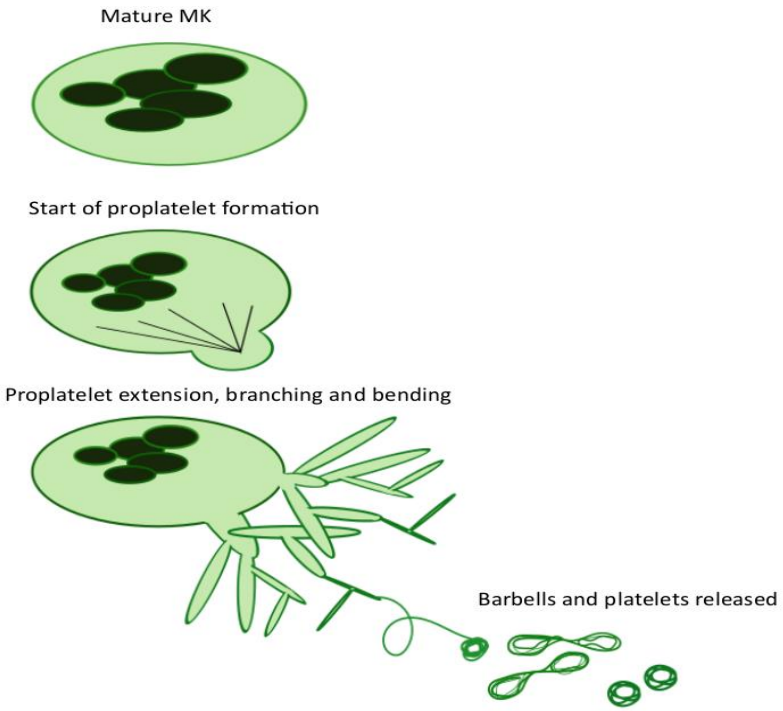


Diagram 3- Platelet formation. A mature megakaryocyte produces a single pseudopodia, marking the initiation of proplatelet formation. This extends branches and bends, forming long proplatelet extensions. Nascent platelets are found at the end of proplatelet extensions and are released, most commonly, as barbells.

A single mature MK is capable of releasing 1000-2000 platelets by undergoing massive cytoskeletal reorganisation. Platelets are very small, 2-3µm across, enucleated fragments with a distinctive disc shape that are formed from the cytoplasmic content of a mature MK. The process of platelet assembly has been studied for over a century (***Diagram 3***). It was first proposed in 1906 that platelets detach from MK pseudopods, initially introducing the idea that platelets were formed by MK extensions (Wright J, 1906). Currently, the most widely accepted theory of platelet formation is based on the proplatelet theory, which was first described in 1976. This built on the earlier observations and described psudeopod-like extensions forming from the mature MK that broke apart to produce individual platelets (Becker RP, 1976). The term proplatelet was introduced to

describe these long cytoplasmic processes, originating from the MK cell body. At the time, it was still believed that the IMS marked out specific regions inside the mature MK, where platelets were formed. Nowadays, a revised version of this theory is generally accepted, known as the 'flow model'. The flow model describes the IMS as a membrane reserve that undergoes evagination during proplatelet formation, enveloping putative platelets (Radley JM, 1982).

In vitro time-lapse microscopy capturing the process in real time has greatly improved our understanding of exactly how proplatelet formation occurs (Italiano JE Jr, 1999). Proplatelets are distinguished by numerous platelet-sized swellings, joined together by cytoplasmic bridges, giving them a 'beads-on-a-string' appearance. The mature MK organises cytoplasm into long extensions, 100-500μm in length, forming the 'string'. The 'beads' found along the proplatelet string are sites where the microtubule bundles diverge and reconvene, creating thickened areas within the proplatelet shaft. The branching process often starts at a single site of the mature MK where one or more broad pseudopodia form. This process has been visualised *in vitro* and over a period of 4-10 hours the processes lengthen, finally forming proplatelets with a diameter of 2-4μm. Microtubules, made up of αβ-tubulin dimers, are essential structural components driving proplatelet formation. When cultured MKs were treated with nocodazole, a drug which inhibits microtubule assembly, proplatelets were no longer able to form (Italiano JE Jr, 1999).

Prior to proplatelet formation, these microtubules form a condensed mass, just below the plasma membrane of the first pseudopodia that forms, before aligning into bundles which fill the cortex of the first process, marking the start of proplatelet formation. Each proplatelet extension contains bundles of microtubules in their core, which form tear-drop loops of 3-5μm at their distal end, referred to as bulbous tips. It is only at these tips that microtubule coiling occurs, a signature property of circulating platelets whereby a single long microtubule coils 8-12 times around the cortex of a platelet, which maintains the platelets hallmark disc shape (White JG, 1967). This indicates that the bulbous tip is the site of platelet assembly. The proplatelet extensions can bend and from these bends new branches form, amplifying the number of extensions and increasing the number of platelets produced by a single MK. At a bend, the proplatelet shaft folds back on itself, generating a new loop, which eventually elongates and becomes a new bulbous tip.

Tubulin microtubules also function to transport organelles from the cell body to the site of *de novo* platelet production in the bulbous tip. The movement of organelles can be bidirectional, due to the bipolar organisation of microtubules within the proplatelet shaft, however, the organelles appear to be become trapped at the tip. Organelle movement occurs due to two mechanisms; organelles

travelling along microtubules and microtubules moving relative to one another to indirectly transport organelles (Richardson JL, 2005).

Individual proplatelets *in vitro* are released from the residual cell body by a rapid retraction event. The processes of extension and retraction occur in continuous cycles, revealing that proplatelets are dynamic and unstable structures. The most common shaped proplatelet to be released after retraction is a barbell, where two platelet-like particles are joined by a single, thin cytoplasmic bridge. More recently an intermediate stage between proplatelet and barbell production has been described, termed the preplatelet (Thon JN, 2010). Preplatelets are 3-10μm and are able to convert between their 'giant' platelet state and the barbell state that has been described previously. Curiously, despite being believed to be the last step before individual platelet release, the barbell structures still undergo granule content exchange. Therefore, it has been proposed that the interconversion between preplatelets and barbell structures may represent a novel mechanism for distributing and sorting granule content in platelets before their final release.

In vivo the release of platelets remains difficult to study due to the rarity of MKs, although proplatelets are known to extend into the vascular sinusoids of the BM, where they enter the bloodstream and are released under shear stress. The process of retraction seen *in vitro* is believed to mimic what occurs *in vivo*, where the barbell structures are believed to break down further in the blood stream to generate singular platelets. *In vivo* platelet generation has been possible in mice using multiphoton intravital microscopy of the BM, which confirmed the concept that blood-flow induced shear stress is a biophysical determinant of platelet production (Junt T, 2007). While a lot of invaluable data and observations have been collected due to improved techniques for culturing MKs *in vitro*, our understanding of the final steps of platelet release remains incomplete.

2.4 Platelet Role: Thrombus Formation

Once released into the bloodstream, platelets make up one of the most abundant vascular components (***Diagram 4***). They are equipped with a huge array of granular contents that allow them to function in many ways in the vasculature. Platelet-specific alpha (α) granules contain a high number of proteins which are released only once a platelet has become activated. The rich granular content of platelets enables them to perform a multitude of functions including; maintaining normal haemostasis, forming thrombi at sites of vascular injury, promoting new vessel production and playing a role in inflammatory processes.

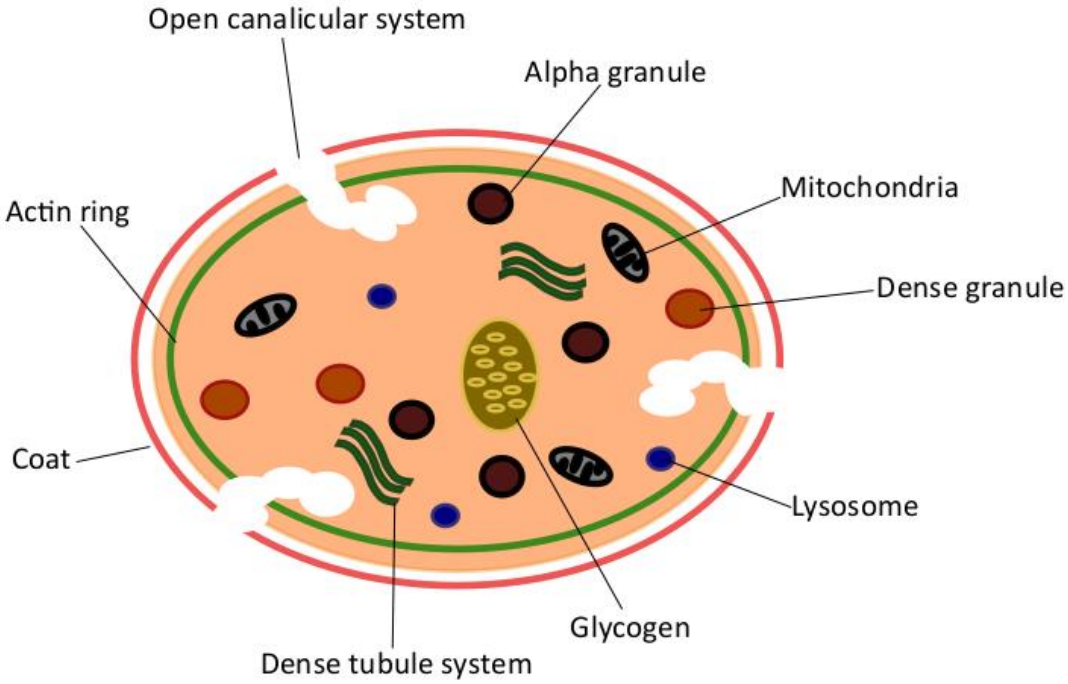


Diagram 4- Platelet structure. Platelets contain a large number of organelles, including alpha and dense granules.

During the short 10 day lifetime of a platelet it is unlikely to come into substantial contact with the endothelial wall lining the blood vessels. This is not the case, however, at sites of blood vessel injury, where the sub-endothelial extracellular matrix becomes exposed to circulating blood components. Platelets adhere to such sites of injury and subsequently become activated in order to limit blood loss and promote vessel healing. Upon extracellular matrix exposure, adhesion molecules such as collagen, von Willebrand factor (vWF), thrombospondin, laminin and fibronectin are exposed, all of which function as ligands to multiple platelet-surface receptors (Broos K, 2011). A number of tightly controlled processes are vital for stabilising platelet adhesion and promoting thrombus formation after initial platelet binding, these are; tethering, rolling, activation and firm adhesion. Since larger vessels, such as veins and arteries, are under a lower shear rate of blood flow than smaller vessels in the microvasculature, the methods for how these processes occur are different, depending on the site of adhesion.

At sites of low shear rates, collagen, fibronectin and laminin binding plays the primary role, while in high shear rates the binding of vWF to the platelet receptor glycoprotein Ibα (GPIbα) is vital to recruit fast moving platelets. GPIbα binds vWF that is either exposed by the damaged endothelial extracellular matrix or immobilised on exposed collagen. vWF can also self-associate, thus increasing

the amount of exposed vWF in the vicinity of vascular damage. Once immobilised, vWF captures platelets via an interaction with GPIbα. Mediated through fast association and disassociation bond rates, this allows platelets to tether to the region of injury by providing a high resistance to the tensile stress caused by the high shear rates in the veins and arteries (Savage B, 1996). This tethering is insufficient to produce stable binding but instead it acts to decelerate platelets, allowing them to roll in the direction of the blood flow. This rolling process puts non-activated platelets into close and continuous contact with the exposed extracellular matrix, crucially allowing further platelet-matrix interactions to stabilise the adhesion (***Diagram 5***).

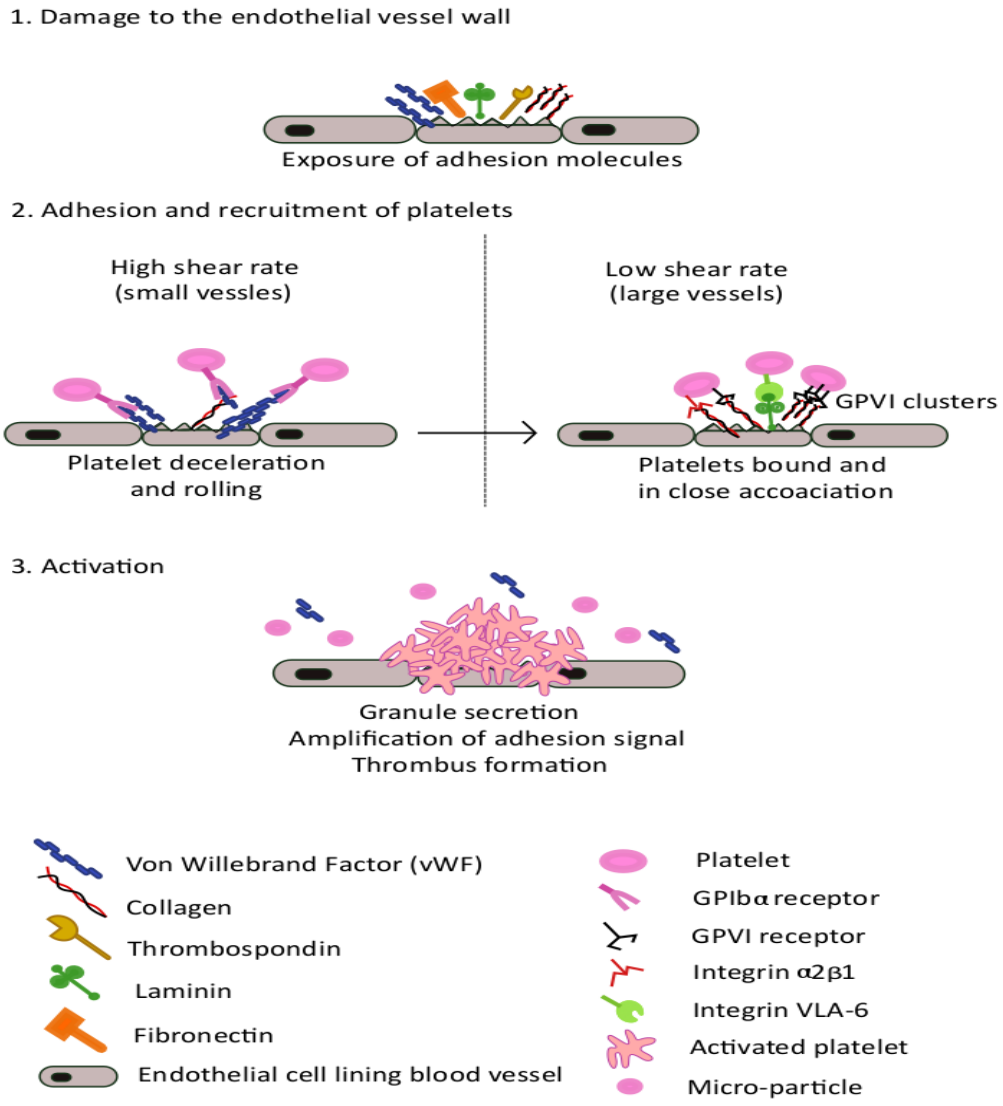


Diagram 5- Thrombus formation. Damaged vessel walls expose adhesion molecules, which bind to the surface receptors of platelets. In high shear environments, platelets are first decelerated through vWF binding, causing them to roll over the damaged surface. This leads to further platelet recruitment through multiple receptor-ligand interactions, causing platelets to adhere to the site of injury. Platelets become activated, releasing their granule content and micro-particles. These release even more adhesion molecules, add to the pro-coagulant rich surface and amplify adhesion signals, forming a thrombus.

Collagen type I and II are particularly potent adhesion molecules exposed by the endothelial extracellular matrix. They bind to the platelet-specific receptor, glycoprotein VI (GPVI), found constitutively on resting platelets. Upon collagen binding, dimers of GPVI form clusters, increasing the receptors avidity for collagen and also increasing the proximity of GPVI-associated signalling molecules, thought to be crucial for initiating and maintaining the downstream activation signal

(Poulter NS, 2017). A second important collagen receptor required for the binding of platelets, is the integrin α2β1 receptor. The platelet α2β1 receptor switches from a low-affinity conformation, through an intermediate, to a high-affinity conformation upon collagen binding (Van de Walle GR, 2005). Another important integrin receptor on the platelet surface, integrin αIIbβ3, binds fibrinogen. This binding has a slow bond formation rate with low resistance to tensile stress, so occurs in low shear areas, or as a downstream event to the GPIbα-vWF initiated tethering and rolling found in high shear locations (Savage B, 1996).

Platelet activation occurs once adhesion has been established, via a number of downstream signal transduction pathways. One way this is achieved is through the activation of a tyrosine-kinase cascade, which ultimately leads to an increase in cytosolic calcium ion (Ca^{2+}) and 1,2-diacylglycerol (DAG) concentrations (Broos K, 2011). This activation results in the platelet changing shape, secreting its granule content and ultimately leads to aggregation. Increased cytosolic Ca^{2+} levels trigger exposure of phosphatidylserine (PtdSer) on the surface of the platelet. PtdSer binds Ca^{2+} and provides a pro-coagulant surface for other factors involved in the coagulation cascade to bind, such as vitamin-K dependent clotting factors. Micro-particles are also released, which present PtdSer and tissue factor, which also increase blood coagulation (Morel O, 2008).

Once activated, platelets release the content of their granules to the local environment in order to amplify the signal transduction of platelet activation. By performing mass spectrometry to probe their proteome, the most abundant granule type, the α-granule, has been shown to contain over 280 different proteins (Maynard DM, 2007). While many of the proteins identified have unknown functions, others have been implemented in inflammation, cell-matrix interactions, wound healing, promoting angiogenesis and even includes novel proteins that could in future be used for detecting Alzheimer's disease from other forms of dementia. Unsurprisingly, such a large proteomic repertoire gives platelets the ability to mediate an impressively wide array of responses to activation, some of which remain poorly understood, as it is remains difficult to establish the exact effects of specific proteins.

Several important pro-thrombotic factors, including vWF and fibrinogen, are released by α-granules. These increase the cross-linking of platelets through their receptor binding, increasing platelet-platelet binding and aiding the recruitment of even more platelets to the site of thrombus formation. This is known as an outside-in signalling cascade, as the initial external stimulus (for example immobilised vWF on the collagen surface), leads to the secondary secretion of the same factor from within the platelet α-granule, and ultimately amplification of the aggregation potential. Alpha-granules, as well as the open canalicular system, also contain reserve supplies of a number of

receptors normally found on the membrane of a resting platelet (Suzuki H, 2003). These additional receptor copies become redistributed to the surface membrane or to released micro-particles, thus, further amplifying the adhesion potential and platelet response.

Dense (δ) granules contain much fewer proteins than α-granules. They have a number of membrane-anchored proteins but also contain Ca^{2+} ions and metabolites such as histamine, involved in inflammation. The main role of δ-granules in amplification of the platelet response signal is through their release of the nucleotide adenosine diphosphate (ADP), which plays a major role in platelet activation. ADP binds to two G-protein coupled receptors, which results in the activation of two separate signalling pathways. The result of one of these activation pathways is to change the shape of the platelet, which occurs before aggregation. This increases the surface area exposed by the platelet, which helps to promote binding to other cell and matrix components (Kahner BN, 2006). Together, these processes lead to an increase in the number of activated platelets at the site of injury, leading to aggregation and thrombus formation, preventing bleeding.

2.5 Thrombopoietin- The Master Cytokine of Megakaryopoiesis

The term thrombopoietin (TPO) was first introduced to describe a cytokine that was essential for the production of platelets in 1958, long before TPO had been cloned and purified (Kelemen E, 1958). However, it was not until the TPO receptor, cMPL (also known as the myeloproliferative leukemia protein, CD110 or MPL), was first cloned in 1992 that the quest to discover TPO really gained traction (Vigon I, 1992). cMPL was identified as a haematopoietic cytokine receptor by the presence of four spatially conserved cysteine residues, alongside a juxtamembrane pentapeptide sequence which is shared by other receptors of this family, including receptors for erythropoietin (EPO), granulocyte colony-stimulating factor, interleukin (IL)-3, IL-5, IL-7, IL-9, and IL-11, amongst others (Cosman D, 1993). Initially cMPL was considered to be an orphan receptor, but due to the cell line from which it was cloned, the bi-potent megakaryocyte erythroid progenitor human erythroleukaemia (HEL) cell line (Long MW, 1990), scientists began to postulate that TPO may be the cMPL ligand.

In 1994, after years of trying to identify TPO, three groups almost simultaneously reported the cloning of TPO cDNA from a number of species including canine, murine and human (Bartley TD, 1994; De Sauvage FJ, 1994; Lok S, 1994). The primary site of TPO production was identified as the liver, but other organs including the kidney were shown to also express TPO. Injections of the recombinant protein into mice increased the number of circulating platelets by fourfold, in a seven day time-period (Lok S, 1994). It was found that the 353 amino acid protein had a novel two-domain

structure. The N-terminal domain, sharing sequence homology with erythropoietin (EPO), was able to bind to cMPL (Bartley TD, 1994). It was later discovered that despite sharing N-terminal homology, EPO and TPO do not compete for binding to their respective receptors (Broudy VC, 1997). The C-terminal domain of TPO is unique, having not been identified in any other known protein and has been shown to function in two ways. First, it prolongs the half-life of the hormone in circulation, as it contains multiple sites of modified N- and O- linked carbohydrates (Harker LA, 1996). Secondly, it contains a glycan domain, responsible for increasing secretion of TPO and which additionally functions as an inter-molecular chaperone, likely within the endoplasmic reticulum, to prevent degradation of the hormone (Linden HM, 2002).

Shortly after the identification of TPO and production of the recombinant protein, *in vitro* cultures of MKs improved greatly. MKs expanded from human CD34+ cells isolated from peripheral blood were shown to be able to generate proplatelets in culture in the presence of TPO (Choi ES, 1995). This was the first demonstration of MK expansion, growth and full maturation to produce platelets *in vitro*. TPO was also shown to be essential for platelet production in animal models. Mice deficient for either TPO or its receptor, show severe thrombocytopenia (low platelet count). They also show significantly reduced cell counts from other haematological lineages, including erythroid, granulocyte-macrophage and multipotent progenitors in the BM, spleen and peripheral blood (Carver-Moore K, 1996). In TPO-deficient mice, the administration of recombinant TPO increased the number of circulating platelets and MK progenitors, as well as the increasing the number of erythroid, myeloid and mixed progenitors in both the BM and spleen. This study showed that not only do TPO and c-MPL play an important role in MK development, but they also act much earlier on in haematopoiesis, at the stem cell or early progenitor cell level.

Further evidence supported this idea when it was found that deleting TPO or cMPL resulted in the reduction of not only megakaryocytes and platelets, but also CMPs and HSCs (Kimura, 1998). The TPO receptor, cMPL, was later found to be highly enriched on the surface of HSCs, even more so than on CMP or MEPs (Terskikh, 2003). This was surprising, since TPO signalling is most closely associated with megakaryocytic fate decisions. However, a role for TPO as an early-acting cytokine in the induction of HSC proliferation was demonstrated in a study investigating the role of TPO, stem cell factor (SCF) and IL-3 on HSCs (Domen, 2000). Single plated mouse HSCs were prevented from undergoing growth factor deprivation–induced apoptosis, by overexpressing the protein B-cell lymphoma 2 (BCL-2), to block apoptosis. BCL-2 over-expressing HSCs, and wild type (WT) HSCs, were then cultured in different cytokines, in order to elucidate if those that induced a growth response could be separated from those which prevented apoptosis. Transgenic cells were not protected from

apoptosis in serum-free conditions, in contrast to when they were cultured in serum. However, they did survive and proliferate rapidly in the presence of SCF, whereas WT cells died. This suggests that both BCL-2 and SCF are required to prevent apoptosis, but signal through different pathways, in HSCs. SCF was found to provide a strong signal for proliferation, with the loss of HSC self-renewal and associated differentiation to all progenitor lineages. TPO stimulated the production of large mature MKs from both mutant and WT HSCs but the plating efficiency was higher in mutant cells, which shows that TPO alone does not prevent WT HSCs from undergoing apoptosis. IL-3 was shown to block both pathways of apoptosis, and stimulated proliferation and the rapid appearance of myeloid cells of various lineages (Domen, 2000). This demonstrated that TPO is a crucial key regulator of MK development and platelet production, but that it also plays an important role in HSC maintenance. Thus, TPO can be described as a pan-haematopoietic cytokine.

One of the major regulation mechanisms for TPO is via receptor-mediated uptake and destruction, a mechanism first described for another haematopoietic growth factor, macrophage colony stimulating factor (Bartocci A, 1987). An auto-regulatory feedback loop is generated by platelets, as cMPL on the platelet surface removes TPO from circulation and it is degraded following binding (***Diagram 6***). If the number of platelets rises, the more TPO is removed from system, reducing MK production of platelets. Conversely, when there are less platelets in circulation TPO levels rise, thus, driving MKs to produce more platelets. However, there are a number of exceptions to this simple receptor-mediated uptake and destruction mechanism. For example, not all cMPL expressing cells contribute to this feedback loop. Endothelial cells also express cMPL but have been found not to contribute to blood levels of TPO (Geddis AE, 2006). The inflammatory marker IL-6 has also been implicated in increasing TPO levels by increasing TPO transcription in the liver, leading to inflammation-induced thrombocytosis (Kaser A, 2001).

Ageing platelets also play a role in TPO production. As platelets age in circulation, they become disialylated and bind to the hepatic Ashwell-Morell receptor (AMR). Binding of the AMR induces hepatic TPO gene transcription and translation, resulting in platelet production regulation (Grozovsky R, 2015).

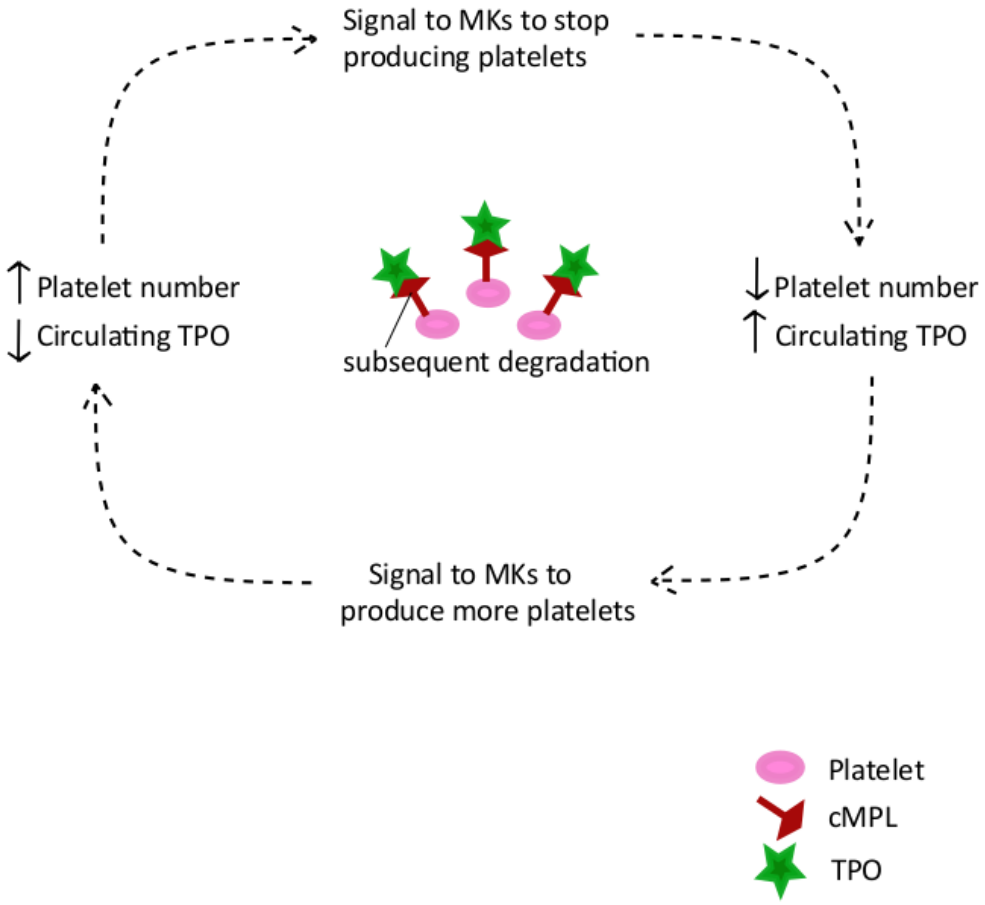


Diagram 6- Thrombopoietin regulatory feedback loop. TPO in the bloodstream binds cMPL found on the surface of platelets, which is subsequently degraded. If platelet levels rise, more TPO is bound, resulting in a decrease in circulating TPO levels. Reduced TPO levels signals to MKs to produce fewer platelets. When fewer platelets are detected, the circulating levels of TPO rise, signalling to MKs to increase their production of platelets.

2.6 Thrombopoietin-independent Megakaryopoiesis

TPO or cMPL null mice are severely thrombocytopenic, however, their platelet counts are not completely eliminated and remain at around 10% that of a WT mouse. A number of studies have therefore tried to determine which other cytokines may be playing a role in maintaining this residual platelet production. cMPL null mice have been crossed with several other mouse strains, lacking different cytokine receptors. From these studies, IL-3, 6, 11 and leukemia inhibitory factor (LIF) have all been shown not to play a role in platelet production (Chen Q, 1998; Gainsford T, 2000). Two MK-active chemokines have been shown to mediate interactions of BM endothelial cells to promote platelet production in a TPO-independent manner. Stromal-derived factor 1 (SDF-1) together with fibroblast growth factor 4 (FGF-4) can restore platelet production in both TPO and cMPL null mice

(Avecilla ST, 2004). These two chemokines mediate MK progenitor localisation to the vascular niche in the bone marrow, promoting MK survival and maturation, as well as platelet release. Through this, it has been suggested that their major role is in allowing progenitors to relocate to a permissive microenvironment.

2.7 Stem Cell Factor in Haematopoiesis and Megakaryopoiesis

Mouse experiments in 1990 revealed two important factors required for HSC and niche maintenance; stem cell factor (SCF, also known as kit-ligand or steel factor) (Zsebo, 1990) and its receptor, cytokine receptor-tyrosine kinase (c-Kit, also known as CD117) (Reith, 1990). These factors were discovered after the mutations of the two loci where these genes are encoded, chromosome 10 and 5 respectively in mice, were investigated. The mutated loci resulted in animals with a similar phenotype that included anaemia and mast cell deficiency. Haematopoiesis could not be corrected in mice with a SCF mutation, even after WT or mutant c-Kit BM cells had been transplanted, whereas haematopoiesis was fully restored in c-Kit mutant mice after similar transplantations were performed. These experiments demonstrated that the c-Kit receptor was essential for HSC function, as haematopoiesis was restored, while the SCF ligand was an essential environmental element for haematopoiesis but was not contained in BM cells, as transplants failed to restore HSC function.

SCF has also been shown to play a role in MK maturation. When added to *in vitro* cultures of CD34+ cells derived from cord blood, SCF significantly increased the level of polyploidisation seen in CD41+ MKs, as well as reduced TPO-induced apoptosis (Kie JH, 2002). Testing the effects on MK growth of a number of different cytokines; Flt3 ligand, IL-3, IL-6, IL-11, LIF, EPO, granulocyte colony stimulating factor (G-CSF) and SCF, only SCF was found to reduce the apoptotic fraction of MKs cells, particularly in late-maturation stage MKs. This suggests that SCF delays TPO-induced apoptosis in mature MKs. MK-lineage specific TPO-induced apoptosis has been previously described in *in vitro* MK cultures, showing that although TPO is important for the proliferation of MK progenitors, alone it is not capable of sustaining their long-term expansion (Ryu KH, 2001). Thus a combination of TPO and SCF is beneficial for *in vitro* MK proliferation and maturation.

2.8 Other Cytokines Controlling Megakaryopoiesis

Applying a multi-step statistical model to quantify the individual and interactive effects of 13 different cytokines on megakarypoiesis *in vitro* allowed the optimal concentrations and combinations of cytokines to obtain MK expansion, maturation and platelet production, to be obtained (Cortin V, 2005). In combination with TPO, the cytokines SCF, IL-6 and IL-9 significantly increased the number of mature MKs in culture. TPO in combination with just Flt3 ligand was found

to only increase the expansion of MK progenitor cells and had no effect on MK maturation. Cytokines which had a negative effect included EPO and IL-8, which both inhibited MK maturation. Optimising the concentrations of TPO, SCF, IL-6 and IL-9, resulted in a MK culture of high purity (90%), which maximised MK expansion and maturation. Increasing the concentration of SCF only in this cocktail further improved both MK expansion and maturation, although this had a negative effect on cell purity, reducing it by almost half.

2.9 Erythropoiesis

As previously described, the megakaryocyte-erythroid progenitor (MEP) cell is bi-potent and can give rise to cells from either the MK or erythrocyte lineage. Erythropoiesis describes the stages of maturation, from erythroblast progenitor to mature erythrocyte (red blood cell). During development, the initial stages of haematopoiesis occur in the yolk sac, where primitive erythrocytes are produced, vital for the survival of the embryo (Tavassoli M, 1991). This is followed by definitive waves of erythropoiesis in the foetal liver and postnatal bone marrow, occurring in three distinct stages. The first stage of definitive erythropoiesis gives rise to the earliest recognisable erythroid-specific progenitor, the burst-forming unit erythroid progenitor (BFU-E). BFU-E progenitors give rise to colony forming unit erythroid (CFU-E) progenitors. The second stage describes progression of nucleated precursors, with the earliest identified as the proerythroblast.

The proerythroblast undergoes 3-4 rounds of cell division as it progressively differentiates towards the reticulocyte, with each cell division being distinguished by a decrease in cell size, enhanced chromatin condensation, increased haemoglobinisation and by changes to the cell membrane organisation (***Diagram 7***). Morphologically distinct populations follow the proerythroblast, which are; basophilic, polychromatic and orthochromatic erythroblasts (Chen K, 2009). In the third stage, orthochromatic erythroblasts generate mature reticulocytes by expelling their nuclear content. A loss of intermediate filaments occurs during the cell divisions from proerythroblast to orthochromatic erythroblast, resulting in a free moving nucleus which is relocates to the cell periphery, occupying an acentric position, before enucleation.

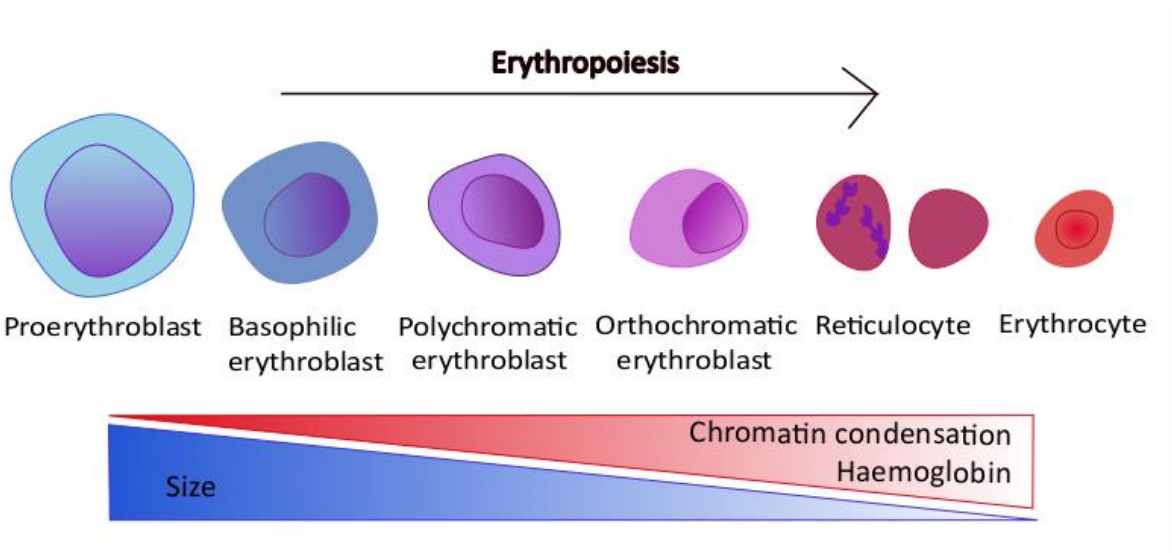


Diagram 7- Erythropoiesis. Proerythroblasts undergo 3-4 rounds of division, producing cells which decrease in size, contain more condensed chromatin and increased levels of haemoglobin. The final stage of differentiation produces the enucleated erythrocyte found in circulation.

The process of definitive erythropoiesis occurs in locations known as erythroblastic islands, specialised microenvironments, found in the BM. In the centre of an erythroblastic island, a macrophage known as a nurse cell, with a unique immunophenotypic signature, is responsible for nurturing erythropoietic cells throughout maturation. The nurse cell extends cytoplasmic protrusions, forming a ring around the erythroblasts and enabling nurse-erythroblast cell interactions, responsible for anchoring cells and maintaining erythroblastic island integrity. Nurse cells are thought to aid erythropoiesis in multiple ways. They have been attributed with providing nutrients required for erythropoiesis, as well as proliferative and survival signals. Nurse cells are also responsible for phagocytosing the expelled nuclei of mature reticulocytes. There is also evidence for nurse cells providing signals for enucleation, a role which has yet to be clearly defined. Erythroblasts are capable of proliferation, maturation and enucleation *in vitro*, but these processes are typically inefficient and very few cells complete the final stage of enucleation. Enucleation has been found to be greatly improved when erythrocytes are co-cultured with macrophages. It is thought that the nurse-erythroblast cell interactions allow cross-talk which controls regulatory feedback mechanisms within the erythoblastic island, that can trigger intracellular signalling pathways to regulate gene expression (Manwaani D, 2008).

Reticulocytes produced straight after nuclear expulsions are immature, with a different cell and cytoskeletal membrane to mature reticulocytes. Immature reticulocytes are found in the BM in healthy individuals but can be released into circulation in times of erythropietic stress. Immature

reticulocytes are motile, multilobular and stain intensely for RNA, while more mature cells are non-motile (Chasis JA, 1989). Mature reticulocytes circulate in the bloodstream for about a day before maturing into erythrocytes (Bessman JD, 1990). Erythrocytes are specialised cells responsible for transporting oxygen around the body and are the most abundant enucleated cell type found in the circulation. The human rate of reticulocyte production is 2 million per second. Erythrocytes are between 6-8μm wide, with a distinct biconcave disc shape, which aids to increase their surface area and therefore their capacity for binding oxygen and also to allow them to travel through narrow capillaries.

2.10 Haemoglobin Expression During Development

The most important component of erythrocytes, and the one which gives them their distinctive red colour, is the haemoglobin tetramer. Haemoglobin is composed of two α-like and two β-like globin peptide chains as well as the haem moieties required for carrying oxygen. A number of different β-like globin molecules are produced during distinct stages of development, due to the fact that the human β-globin locus of chromosome 11 is developmentally regulated. Two distinct switching events occur in expression of the β-like globin genes in humans. The fist β-like globin gene to be expressed is the embryonic ε-globin, expressed during primitive haematopoiesis in the yolk sac. The first globin switching event coincides with the switch from primitive haematopoiesis to the first stage of definitive erythropoiesis in the foetal liver, from expression of ε-globin to foetal globin to γ-globin (McGrath K, 2008). α and γ-globin chains make up the foetal haemaoglobin (HbF), which is the major haemoglobin found throughout gestation. The second globin switch occurs shortly after birth, with a switch from γ-globin to β-globin expression, a transcriptional switch which occurs at the level of erythroid progenitors, resulting in adult haemoglobin (HbA) (Sankaran VG, 2013).

2.11 Erythropoietin- The Master Cytokine of Erythropoiesis

CFU-E progenitors, generated during the first wave of definitive erythropoiesis, are dependent on the cytokine erythropoietin (EPO) and its receptor, EPOR, for proliferation and maturation. Studies on mice which lack functional EPO, or EPOR, die at around embryonic day 13, due to a block in definitive erythropoiesis at the CFU-E progenitor stage. These studies show that both EPO and EPOR are essential for the *in vivo* survival of CFU-E progenitors in the foetal liver and BM. However, a lack of either EPO or EPOR protein does not block lineage commitment occurring before this stage, in primitive erythropoiesis, as BFU-E progenitors were unaffected (Wu H, 1995). In the adult, EPO has been shown to slow down the breakdown of DNA, preventing apoptosis occurring in progenitors, enabling them to differentiate (Koury MJ, 1990).

EPO expression by the kidneys is tightly controlled during development. At the yolk sac stage, where primitive erythroid progenitors are abundant, the expression of EPOR but not EPO can be detected in mice. Primitive progenitors are not EPO dependent and so are able to mature in the yolk sac. EPO expression is first detected in the embryo when foetal circulation occurs, at the first stage of definitive erythropoiesis. Since primitive and definitive erythroid progenitors have been observed together in the yolk sac, it is proposed that the lack of EPO stimulation on EPO-dependent definitive progenitors allows definitive erythroid progenitors to exist which are poised for maturation once EPO is expressed. EPO expression marks a major control point for the second wave of erythropoiesis, showing EPO is a critical cytokine to signal and coordinate erythrocyte developmental pathways (Lee R, 2001).

3. Transcriptional Regulation of Megakaryopoiesis and Erythropoiesis

3.1 Transcriptional Regulation Overview

Since the genetic content of every cell within an individual's body is identical, multi-layered mechanisms exist that allow differentiated cells to be produced which fulfil distinct, specific requirements. The differences between distinct differentiated cell types, such as neurones and cardiomyocytes, or developmentally more closely associated cell types, such as megakaryocytes and erythrocytes, is due to their expression of different genomic and proteomic profiles. Most cells will differ at the level of gene expression patterns, but those that do not have different protein expression levels. Thus, the mechanisms controlling cell fate differ at the transcriptional (gene), translational (mRNA) and protein structure levels.

Further to this, genetic expression can be controlled at the epigenetic level, whereby heritable changes to the structure and function of genes occur without altering the genetic code (Jaenisch R, 2003). Small, non-coding RNA molecules such as microRNA (miRNA) alter protein translation, by targeting certain mRNA molecules for degradation or translation repression. A large repertoire of over 530 miRNAs has been identified in human platelets. The most abundant family of let-7 miRNAs in platelets has been shown to play a role in cell differentiation (Plé H, 2012). Lineage specific long noncoding RNAs (lncRNAs) regulate gene expression at both the transcriptional and post-transcriptional level. Over 1100 lncRNAs have been identified in erythroblasts, megakaryocytes and MEPs and some have been shown to play a role in the terminal maturation of these cells. Many of these lncRNAs identified are regulated by important TFs for these lineages, such as GATA1 and TAL1 (Vikram R. Paralkar, 2014).

3.2 GATA-binding Transcription Factors: GATA1 and GATA2

The GATA family of TFs is made up of proteins which have high affinity binding for the consensus sequence T/A (GATA) A/G (Merika M, 1993). GATA proteins have two independent transactivation domains, at the N- and C- terminals, as well as 2 zinc finger DNA binding domains (Kaneko H, 2012). There are 6 GATA TFs in total, involved in proliferation, differentiation and cell survival. Of these, *GATA1, GATA2* and *GATA3* are known as haematopoietic GATA factors, and have distinct but often overlapping patterns of expression. *GATA1* is expressed in erythroid and MK cells, as well as in mast cells, eosinophils and haematopoietic progenitor cells. *GATA2* is also expressed in MKs, haematopoietic progenitors, mast cells and a wider variety of other cells and tissues. While the third haematopoietic GATA factor, *GATA3*, is highly expressed in T-lymphoid cells (Merika M, 1993).

Targeted gene mutations of the *GATA1* gene in mice showed that this factor is essential for erythroid development, as erythroid differentiation was blocked and these mice failed to give rise to mature reticulocytes (Pevny L, 1991). *GATA1* is required for activating a number of erythroid genes involved in maturation, as well as for activating globin genes. This study demonstrated that despite similar binding motifs, the family of GATA-binding proteins cannot compensate for an absence of *GATA1*. *GATA2*-null mice have severe anemia and, as for *GATA1*-null mice, embryonic lethality. *GATA2* gene dosage experiments on *GATA2*-null mice revealed that *GATA2* plays two functionally distinct roles at the HSC level. The first is in the production and expansion of HSCs in the embryo during primitive haematopoiesis. And the second is in the proliferation of HSCs in the BM of adults during definitive haematopoiesis (Ling KW, 2004).

GATA1 and GATA2 recognize similar DNA motifs and act sequentially during haematopoiesis in orchestrating a broad program of gene activation and repression. *GATA2* is most highly expressed during early blood cell development, and its expression in erythropoiesis declines as cells become more differentiated, while the opposite is seen for *GATA1* expression. GATA1 actually replaces GATA2 at a number of its chromatin occupancy sites, in a process which is known as '*GATA* switching' (Weiss MJ, 1994). Genome wide chromatin occupancy analysis has revealed that GATA switching occurs at greater than one third of sites occupied by *GATA1*, in cells of both the MK and erythroid lineages, with *GATA1* and *GATA2* acting generally in opposing ways on their switch target genes (Doré LC, 2012). The main difference found between the two lineages was in their expression of *GATA2*. *GATA2* expression was greatly reduced in erythroid cells compared to MK cells, despite the GATA switching events occurring at the GATA2 locus controlling its expression being the same in the two cell types. Thus, the difference in *GATA2* expression is likely due to the differential recruitment of different co-factors to these sites, or different kinetics of GATA2 displacement. The

same study also found that *GATA1* binds many of the same genes in MKs and erythroid cells, but that this occurs at different binding sites, allowing for differential control of the same genes between the lineages. As for differential *GATA2* expression, these differences are most likely due to co-factors that associate with *GATA1*. MK-associated TFs containing an ETS domain were found more frequently associated with *GATA1* in MK cells compared to erythroid cells.

3.3 ETS-binding Transcription Factors: ETS-1 and FLI1

ETS factors encompass a family of more than 40 members, which are characterized by an 85 amino-acid region of homology known as the ETS domain. The ETS domain mediates binding to a core DNA motif, known as the ETS recognition element. The ETS recognition element has the following consensus sequence; 5′-GGA(A/T)-3′ (Karim FD, 1990). In the genome wide *GATA1/2* chromatin occupancy site study mentioned previously, *ETS-1* in particular was found by ChiP-PCR to co-occupy GATA2 sites at critical MK genes (Doré LC, 2012). *ETS-1* expression is downregulated by the miRNA, MIRN155, which itself is highly expressed in haematopoietic progenitors but decreases sharply during MK differentiation. When MIRN155 is overexpressed, MK proliferation and differentiation is impaired (Romania P, 2008). Therefore, *ETS-1* has been implicated in the regulatory axis of the MEP lineage fate choice, as the reduction in MIRN155 and increase in *ETS-1* are required for MK proliferation and differentiation.

ETS-1 and *GATA1* have been shown to bind to the promoter of the human Glycoprotein IIB (*GPIIB*) gene, which is only expressed in MKs. Their trans-acting activity was shown to act additively on the GPIIB gene promoter, in co-transfection assays in HeLa cells. This GATA-ETS association was also shown at the human GPIIB enhancer and at the rat platelet factor 4 promoter (Lemarchandel V, 1993). These results indicated that GATA and ETS cis-acting sequences are important for determining the expression of MK-specific genes.

FLI1 is another member of the ETS family of transcriptional activators and is also an important regulator of differentiation in MKs. *FLI1* was originally discovered as a gene that was commonly activated as a result of pro-viral insertion of the Friend leukemia virus in mice (Ben-David Y, 1990). *FLI1*, along with *ETS-1*, is located at the chromosome region of 11q23.1. Patients identified carrying a deletion of this chromosomal region, suffer from thrombocytopenia (Breton-Gorius J, 1995). Both the *FLI1* and *ETS-1* gene products bind to and trans-activate the *cMPL* and gene promoter, along with *GATA1* (Deveaux S, 1996) . The effects of these three TFs binding at the *cMPL* promoter are additive, and led to the proposal that the presence of GATA and ETS binding sites, closely spaced, is a hallmark of MK regulatory regions (Deveaux S, 1996).

FLI1 has also been identified as playing a major role at the top of a genetic hierarchy of blood and endothelial cell development (Liu F, 2008). During embryogenesis, the haemangioblast precursor cell gives rise to both endothelial and blood cells. Using an antisense morpholino to the *FLI1* start codon in Xenopus and zebrafish embryos resulted in a substantial reduction of haemangioblasts. Loss of *FLI1* led to reduced expression of haemangioblast genes, as well as genes later expressed in erythroid, myeloid and endothelial cells. Therefore, *FLI1* is required to form haemangioblast cells in the mesoderm germ layer in the early embryo, and in turn, for producing endothelial and blood cells.

3.4 TAL1

The *TAL1/SCL* (stem cell leukaemia) gene, was first identified due to the translocation t(1;14)(p34;q11), which generates a fusion gene with *TAL1* and the T-cell receptor (*TCR*) gene. This translocation is responsible for 3% of T-cell acute lymphoblastic leukaemia (T-ALL) cases, and a further 23% of T-ALL cases are the result of a constitutively active *TAL1* gene (Aifantis I, 2008). *TAL1* has been described as a master regulator of haematopoiesis, after being identified as an important gene product in multipotent haematopoietic stem cells, as well as being an essential protein for the differentiation of almost all cells in the haematopoietic lineages, with the exception of T cells (Green T, 1996).

TAL1 belongs to the basic helix-loop-helix (bHLH) family of TFs. The bHLH domain is capable of protein-dimerisation as well as DNA binding. TAL1 proteins do not have intrinsic DNA binding activity, due to its bHLH domain being unable to form homodimers, therefore both its DNA binding and TF properties rely on its interaction with other proteins. TAL1 is capable of dimerising with certain other bHLH proteins and will readily bind any of the ubiquitously expressed class A bHLH TFs, such as the products of the *E2A* gene, E12 and E47 (Hsu HL, 1994). Once TAL1 has formed a heterodimer, it is able to recognise and bind to DNA in a sequence-specific manner. These heterodimers preferentially recognise a subset of the E-box element motifs, with the consensus sequence AACAGATGGT, a cis-acting element found in a variety of eukaryotic transcription enhancers. In this consensus sequence AACAG is the recognition site for the class A bHLH polypeptide, while ATGGT is the recognition site for the TAL1 polypeptide (Hsu HL, 1994).

A second study by Hsu showed how TAL1 binding can regulate a single target gene in a positive or negative manner, depending on the cellular environment. The study found that E47-E47 homodimers were more potent activators of an artificial reporter gene, than TAL1-E47 heterodimers. In the presence of TAL1, E47 would more readily form heterodimers, than

homodimers, and that these TAL1-E47 heterodimers were more stable. E47-E47 homodimers would readily dissociate in the presence of Id peptides, a family of HLH proteins that regulate the transcriptional activities of bHLH proteins (Benezra R, 1990). Thus, in an environment which permits E47-E47 homodimer formation, TAL1 can act to repress a gene by recruiting E47 to form TAL1-E47 heterodimers, which has lower transcriptional activity. Conversely, TAL1 acts as an activator of gene expression in the presence of Id proteins, as TAL1-E47 are more resistant to Id directed dissociation, compared to E47-E47, despite being a less potent activator. Id protein levels have been observed to fluctuate during differentiation in response to growth factors, with Id1 and Id2 protein levels seen to rapidly decrease in erythroid cells as they are induced to differentiate (Benezra R, 1990).

TAL1 has been shown to be essential for haematopoiesis and it is thought its role in blood formation has been conserved throughout vertebrate evolution, as *TAL1* homologues are expressed in the mesodermal precursors to both endothelial and blood cells in mice, zebrafish and Xenopus (Elefanty AG, 1997). Mice embryos lacking *TAL1* display retarded growth and die between embryonic stages E9.5-10.5, with absolute anaemia. By assessing the contribution of *TAL1*-/- embryonic stem cells (ESCs) to haematopoietic cells in chimeric adult mice, Elefanty *et al* showed that expression of the TAL1 protein is localised to the neural, vascular and haematopoietic tissues. They have also shown that *TAL1*-null ESCs fail to express haematopoietic-restricted genes, such as *GATA1*, when differentiated *in vitro*. However, genes responsible for ventral mesoderm formation, such as *BRACHYURY*, as well as genes co-expressed in both haematopoietic and endothelial lineages, such as *GATA2*, are expressed at the same levels as in *TAL1* heterozygous and wild type ESCs. This study gives further evidence to support the fact that *TAL1* is essential for putative hemangioblasts to commit to the haematopoietic lineage, but that its expression is not an essential requirement of endothelial differentiation (Elefanty AG, 1997).

A *TAL1* stem cell enhancer has been identified by the Göttgens *et al*, 19 kilobases (Kb) from the *TAL1* promoter, termed +19. The +19 enhancer contains three regions of conserved critical TF binding sites; region 1 containing ETS (GGAT) and Myb (TAACAG) binding sites, region 2 containing an ETS (GGAA) binding site and region 3 containing a GATA binding site. These sites were shown to be on the same surface of the DNA strand, allowing a multi-protein complex of ETS and GATA factors to assemble and activate the enhancer. *FLI1, GATA2* and *ELF1* form this complex, to initiate the expression of *TAL1* and establish a transcriptional programme for both HSC and blood cell development. This 'enhanceosome' acts to integrate lineage commitment signals, and it is likely that its composition alters during development. The authors propose that different GATA and ETS family

members might bind to it in response to distinct signals, or in different cell types, allowing multi-lineage control (Göttgens B, 2002).

TAL1 has been shown to interact with growth factors during adult haematopoiesis. For example the TAL1 protein, through interactions with vascular endothelial growth factor (VEGF) protein, functions to suppress apoptosis at the onset of haematopoiesis. In particular, overexpression of *TAL1* can rescue the near absence of primitive erythroid precursors in VEGF perturbed mouse yolk sacs, by re-establishing the survival of erythroid cells and increasing expression of erythroid specific genes (Martin R, 2004). Also, the *TAL1* gene occupies the regulatory sequences of the SCF receptor in primary hematopoietic progenitors. It has been found that TAL1 enhances the sensitivity of these progenitors to SCF, reducing apoptosis. It also support the survival of MEPs, by acting downstream of the SCF receptor, c-KIT, and establishing a positive feedback loop in multipotent and MEP cells (Lacombe J, 2013).

3.5 Transcription Factor complexes in Haematopoiesis

A heptad of key TFs expressed in haematopoietic stem and progenitor cells (HSPC), binds to and primes an extensive set of MK specific genes, alongside a smaller number of erythroid-specific genes, during the early stages of haematopoiesis (Wilson NK, 2010). This heptad of TFs includes; *TAL1, FLI1, RUNX1, GATA2, LYL1, LMO2* and *ERG1* (***Diagram 8***). A combination of global transcriptome profiling and chromatin immunoprecipitation with massively parallel DNA sequencing (ChIP-seq) found that the principal TFs in this complex- the ETS and GATA factors plus *TAL1*, remain bound to their *cis*-regulatory modules even after a gene undergoes further transcriptional activation upon lineage commitment (Pimkin M, 2014). This study also identified that the majority of genes bound by this core complex of TFs undergo GATA switching. This suggests that GATA switching regulates the change from low levels of expression in HSPCs to high expression levels after lineage commitment, which was found to be particularly true of MK genes. It also provided evidence that the GATA switch is associated preferentially with genes undergoing transcriptional activation, supported by the fact that many genes primed in HSPCs which are activated in MKs, are silenced in erythrocytes.

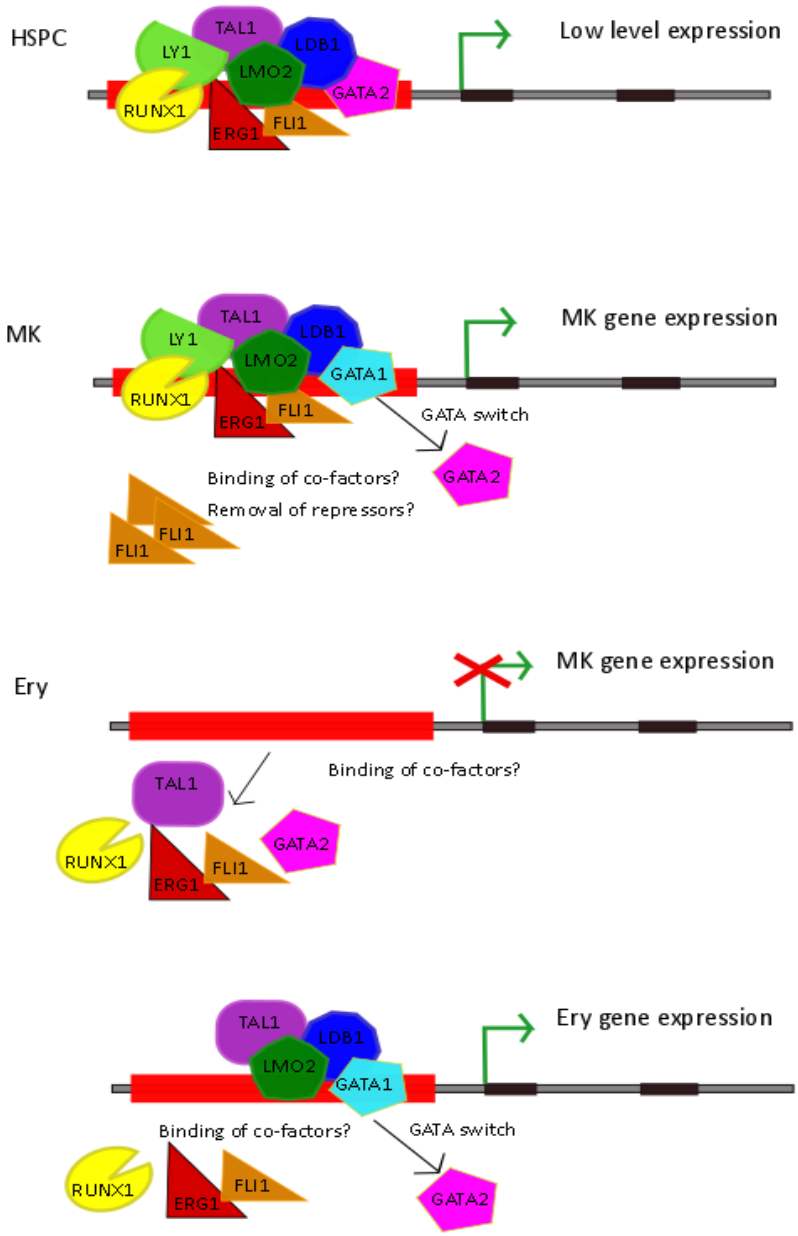


Diagram 8- Potential transcription factor binding mechanisms in developmental gene regulation during megakaryopoiesis and erythropoiesis. A heptad of TFs (TAL1, FLI1, RUNX1, GATA2, LYL1, LMO2* and *ERG1), plus LDB1 which mediates binding of several of these TF proteins, primes a number of MK-associated genes, resulting in low levels of expression in haematopoietic stem and progenitor cells (HSPCs). In MKs, further transcriptional induction of MK genes is mediated by the GATA switch from GATA2 to GATA1, plus additional mechanisms such as increased levels of FLI1, binding of further co-activators and removal of repressors. In erythroblasts, terminal MK genes are silenced by the departure of MK-specific activators, such as the GATA proteins and ETS factors. Erythrocyte gene expression is achieved by GATA switching, along with the binding of erythrocyte co-activators and loss of FLI1.

Genome-wide analysis of the binding of GATA1, GATA2, RUNX1, FLI1 and TAL1 in primary human MKs was investigated through ChIP-seq. The simultaneous binding of all five factors was the most common binding pattern observed, occurring near to known haematopoietic regulators. These five TFs are controlled by a positive auto-regulatory feedback loop and contain extensive cross-regulatory links, forming a densely connected core network that largely acts as a positive regulator of MK gene expression (***Diagram 9***) (Tijssen MR, 2011). This study also highlighted that a large number of MK genes are bound by GATA1 and FLI1 plus additional factors, most likely forming larger, multi-protein complexes in order to elicit a positive regulatory effect. Of these additional factors, RUNX1 was highly represented and is known to bind to both GATA1 and FLI1 proteins. Expression levels of *RUNX1* remain high during MK differentiation but become diminished in erythroid differentiation. These results suggest a role for RUNX1, plus additional factors, to work in combination with FLI1 and GATA1 in activating MK genes, but that its absence inhibits the expression of these genes which acts to promote erythroid differentiation.

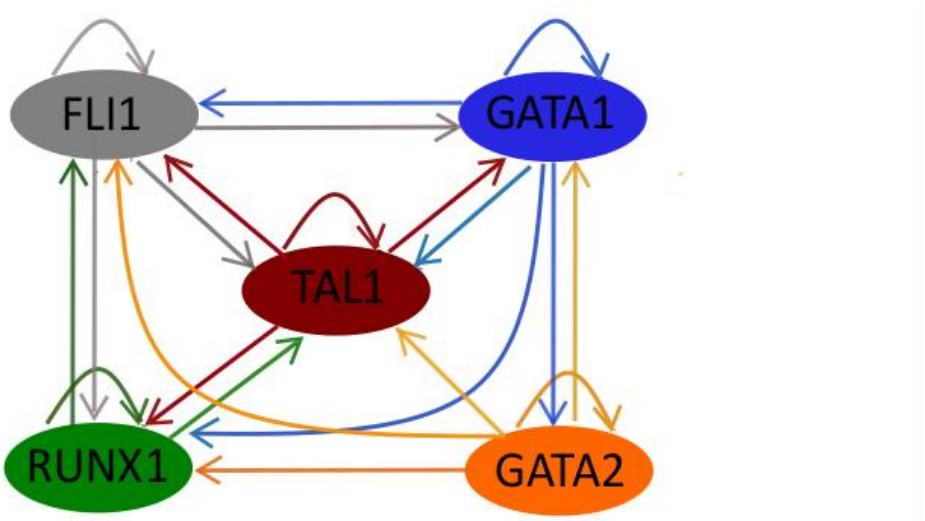


Diagram 9- Auto-regulatory and cross-regulatory links between the transcription factors FLI1, GATA1, TAL1, RUNX1 and GATA2.

4. Stem Cell Biology

4.1 Stem Cell Definitions

Cells which fulfil the following two requirements; self-renewal (the ability to replicate into identical daughter cells without differentiating) and maintain pluri/multi-potency (the ability to generate multiple cell types), are classified as stem cells. The concept of such a cell type was first described in 1963, through the pioneering *in vivo* work on blood system regeneration. Bone marrow (BM) cells

from immunogenic donor mice were transplanted into recipient mice and ten days later it was found that cellular colonies had formed in the recipient's spleen. Analysis of these colonies showed that they had been generated from a small sub-set of the donor BM cells, and that these cells possessed unusual cellular qualities enabling them to both self-renew and produce multiple myeloerythroid cells (Becker, 1963). The stem cell and regenerative medicine fields have since identified multiple types of stem cells, responsible for different roles depending on the level of potency they retain. ***Table 1*** describes different types of stems cells, along with examples of each.

Type of Stem Cell	**Can generate:**	**Example of cell:**
Toti-potent	All embryonic and extraembryonic tissues	Zygote
Pluri-potent	All embryonic tissues	Embryonic stem cell (ESC), induced pluripotent stem cell (iPSC)
Multi-potent	All lineages of a tissue/organ	Haematopoietic stem cell (HSC), neural stem cell (NSC)
Oligo-potent	A limited number of lineages	Common myeloid progenitor (CMP), common lymphoid progenitor (CLP)
Bi-potent	Two different lineages	Megakaryocyte erythroid progenitor (MEP)
Uni-potent/Progenitor cell	One lineage	Macrophage progenitor

Table 1. List of stem cell definitions.

4.2 Embryonic Stem Cells

The term 'embryonic stem cell' (ESC) was first coined in 1981. Cells isolated from the inner cell mass of late blastocysts of mice were cultured in conditioned media before being injected into recipient mice. They were shown to be capable of producing teratocarcinomas, a hallmark quality of pluripotent stem cells (Martin GR, 1981). ESCs were also shown to be able to survive in culture indefinitely, while maintaining their pluripotent potential (Evans MJ, 1981). These discoveries marked a turning point in the investigation of mammalian development and expanded the field of experimental genetics. In 1991, a series of experiments were reported in which ES cells were transduced with promoter traps, using a promoter-less novel reporter which encoded both β-galactosidase and neomycin phosphotransferase for selection. These ES cells led to the derivation of transgenic mice with mutated target genes, whose expression could be visualised by staining with X-gal. 24 heterozygous mice were inter-crossed, identifying 9 strains in which homozygosity resulted in embryonic lethality. This approach offered a unique feature of being able to study genes that were inaccessible by classical genetic approaches, as it allowed genes which were not essential for development to be studied by combining the ability to select for insertions within genes, with the ability to track their activity during development (Friedrich G, 1991).

4.3 Induced Pluripotent Stem Cells

ESCs, despite offering huge promise in areas such as disease-modelling and drug efficacy screening, remain surrounded by ethical controversy. This is especially true for human ESCs, as their isolation involves the destruction of an embryo. In 2006, these ethics-related issues were alleviated, with the generation of pluripotent stem cells directly from mouse embryonic and adult fibroblasts (Takahashi K, 2006). This was the first demonstration that fully differentiated, mature, somatic cells could be reprogrammed to an embryonic-like state, by overexpressing a minimal number of defined TFs through retrovirus-mediated transfection. The four factors used were; Oct3/4, Sox2, c-Myc and Klf4. These factors are often referred to as the 'Yamanaka' factors when describing the derivation of induced pluripotent stem cells (iPSCs). iPSCs exhibit similar morphological and growth properties as ESCs, express ESC marker genes, and also give rise to tumours containing tissues from all three germ layers when injected into nude mice. In 2007, human iPSCs were first derived using the same method (Takahashi K, 2007).

The fields of translational and regenerative medicine greatly expanded after the discovery of iPSCs as it offered the opportunity to easily derive patient-specific iPSC lines for disease modelling and to investigate the development of difficult to isolate cell types, such as cardiac tissue. There are now multiple ways of producing human iPSCs from somatic cells, as the focus of work over the past decade has been to try and improve reprogramming efficiency, which was very low (approximately 0.01-0.02%) for the retrovirus-mediated method described previously (Takahashi K, 2007). There have also been advancements to eliminate the tumorigenic potential associated with the use of the oncogenes c-Myc and Klf4 to derive iPSCs. Different protocols differ in their choice of somatic cell type, reprogramming factors used, delivery methods of these factors and culture conditions. Different methods have their own advantages and disadvantages. For example, adenovirus delivery has the advantage of being transgene and vector-free, resulting in no genomic DNA integration but has the disadvantage of being slow and inefficient. Other non-integrative methods for generating iPSCs include the use of RNA Sendai virus and episomes (Ono M, 2012, Yu J, 2009). Whereas, lentiviral or retroviral delivery have the advantages of being very efficient and stable, but the disadvantage that transgenes (TGs) are integrated into genomic DNA and residual expression of TGs is a possibility (González F, 2011). Therefore, when considering iPSC derivation, the downstream application of the cells should also be considered.

4.4 Pluripotent Stems Cells: A Potential Source of Blood Cells

Most *ex vivo* human megakaryopoiesis relies on the donation of cord blood or peripheral blood apheresis samples, in order to isolate HSCs for culture and differentiation. A huge limitation with this technique is the limited number of HSCs that can be isolated, compounded by the fact that primary haematopoietic stem cells cannot be expanded and maintained for long periods in culture. Due to the difficulty in both obtaining and maintaining HSC cultures *in vitro*, pluripotent stems cells (PSCs- either ESCs or iPSCs) offer an attractive alternative cell source, since they can be cultured indefinitely, with many characterised lines readily available for use.

Due to the fact that platelets and red blood cells are enucleated, they will survive γ-radiation treatment. This would allow these cell types to be irradiated prior to clinical use, in order to eradicate any remaining undifferentiated cells with oncogenic potential. This would generate a pure and safe population of cells for transfusion. Other advantages, such as being able to more easily match blood cells to recipients based on human leukocyte antigen (HLA) class I expression, could be achieved by the careful selection of very few starting iPSC lines of selected HLA haplotypes. These considerations could reduce the number of patients suffering from platelet refractoriness, where an immune reaction to the transfused platelets eliminates them, often the result of repeat transfusions (Jia *et al*, 2014). Or they could provide support to patients who are already alloimmunised, such as multiparous women.

These advantages mean PSCs are currently the source of much excitement in the field of transfusion medicine, as they offer a promising avenue for the production of blood cells *in vitro*. The most commonly used method to derive mature cell types from PSCs is through directed differentiation. In this technique, cytokines and growth factors are used to supplement media, directing PSCs through the different developmental stages that they would undergo during normal development *in vivo*. Methods exist for differentiating PSCs into almost all cell types, including megakaryocytes and reticulocytes.

4.5 Generating Megakaryocytes *in vitro*

The first study to show that MK differentiation from hESCs was possible demonstrated a yield of 0.1-0.4 MKs per input ESC (Gaur M, 2006). The second report of hESC-derived MKs, which promoted the emergence of sac-like structures termed embryonic stem cell-derived sacs (ES-sacs) by culturing with vascular endothelial growth factor (VEGF), increased MK yield to 2-5 per input ESC (Takayama N, 2008). Despite this improvement in MK yield, neither protocol was feasible to give rise to the

number of cells required for clinical use and had other disadvantages, namely the reliance on xenogeneic cell co-culture, as well as the use of animal serum.

In 2011, hESC-derived platelets were first shown to be able to function *in vivo* by contributing to thrombus formation in a laser injury mouse model (Lu SJ, 2011). This study brought two advantages compared to previous ones; the MKs derived were in feeder and serum-free conditions (although these were subsequently plated onto animal feeder cells to maximise platelet production) and they showed increased MK yield, with a range of 18-118 MKs produced per hESC, across three lines tested. More recently, immortalised MK progenitor cell lines (imMKCLs) have been reported, through the overexpression of BMI1 and BCL-XI, to supress senescence and apoptosis respectively, and the controlled expression of c-MYC, to promote proliferation (Nakamura S, 2014). This approach enables cells to expand for up to 5 months of culture, which can produce platelet-sized particles expressing MK markers CD41 and CD42, by halting the expression of the three immortalising genes. However, this study approximated the requirement of 2.5E+10 imMKCLs for the production of a single platelet unit for transfusion (1.0E+11 platelets), which would require considerable large-scale production.

4.6 Forward Programming to Generate Megakaryocytes

Our lab developed a novel method of differentiating PSCs, termed 'Forward Programming' (Moreau T, 2016). Unlike the majority of directed differentiation protocols, forward programming of PSCs into MKs is feeder-free, chemically defined, performed in a reduced cytokine setting and adheres to strict good manufacturing practice (GMP) guidelines. This method produces an increased number of mature MKs, up to 26 times more, compared to directed differentiation methods, in just 20 days. Furthermore, the forward programming strategy generates a MK biased progenitor, capable of sustaining long term, pure MK cultures over several months, resulting in 2.0E+05 mature MKs per input PSC. The mature MKs produced by forward programming release functional platelets, allowing the prospective collection of several platelet transfusion units from as few as 1 million input PSCs.

Forward programming was designed with a long-term goal of being scaled up to allow for the production of MKs *ex vivo* for use in transfusion medicine, hence, the advantage of working in feeder and serum free conditions. The MKs produced can also be cryopreserved at any stage throughout culture. Due to this, the massive mature MK expansion achieved and the high purity of cells, forward programming has a major advantage over most existing directed differentiation protocols, as it is closer to being a clinically compatible method for this goal.

Forward programming was developed after identifying three key haematopoietic transcription factors (TFs) that, when ectopically expressed in PSCs, push them to 'jump' developmentally from iPSC to MK cell. Forward programming uses the three TFs; *TAL1*, *GATA1* and *FLI1*, which have been identified as master regulators of haematopoiesis and are capable of generating MKs directly from PSCs. In silico predictions, based on ChIP-seq data, suggests that these three TFs potentially activate 51% of the genes involved in megakaryopoiesis (Tijssen MR, 2011). Once these genes have been switched on, they potentially in turn switch on a further 45% of the genes responsible for megakaryopoiesis, based on interaction databases, resulting in a total of 98% of the genes involved being activated (Gieger C, 2011). Without any one of the three TFs, *TAL1*, *GATA1* and *FLI1*, MK production from iPSCs is completely abolished in forward programming.

4.7 Methods of Targeting Human Pluripotent Stem Cells

Owing to their self-renewal and pluripotent cell properties, genome editing in PSCs is often desirable, for example in order to correct patient-specific mutations, considering all other cell types have limited culture periods meaning any gene modifications in them cannot always be studied extensively before a culture dies. Also, since PSCs are amenable to cryopreservation, modified cell lines can be banked and used for future purposes. Below is a description of the most widely used gene-editing techniques utilised in PSCs.

4.7.1 Zinc Finger Nucleases (ZFNs)

The first report to describe the successful electroporation of hESCs to insert vector DNA by homologous recombination (HR) was published in 2007. The protocol required a high cell volume (1.0E+07) and resulted in a low yield of colonies surviving selection (20-500), as extensive mortality is associated with electroporating hESCs (Costa, 2007). The difficulty in culturing such large volumes of stem cells at the time, as well as the high mortality following electroporation, led to few reports of successful HR, highlighting that this approach was difficult in such cells and that methods to achieve it were lacking.

The discovery and use of zinc-finger nucleases (ZFNs) was reported around the same time, which propelled work in targeting both hESCs and iPSCs. ZFNs consist of a FokI nuclease domain fused to a DNA recognition domain. The DNA recognition domain is made up of C2H2 zinc-finger motifs, engineered to recognise a specific sequence of DNA that can be tailored to target almost any genetic locus of choice. Once the target DNA is bound by two ZFN fusion proteins, the FokI nuclease domains dimerize, become activated and cut the DNA. This results in a double-strand break, which can be repaired through error-prone non-homologous end joining (NHEJ), or homology-directed repair

(HDR). If homologous vector DNA is present, HDR enables the incorporation of exogenous DNA (Lombardo, 2007).

The Jaenisch lab demonstrated the use of ZFNs for developing a highly efficient and robust expression system, targeting five distinct genetic loci of three different genes in hESCs and iPSCs (Hockemeyer, 2009). One of the gene loci studied, the AAVS1 locus on chromosome 19, encodes the ubiquitously expressed *PPP1R12C* gene. This is a well characterized genomic site, shown previously to allow the long-term, stable expression of TGs in a number of cell types, including hESCs (Smith, 2008). The ZFNs used had short homology arms (~500 bases) and were used to incorporate two different donor vectors into the first intron of *PPP1R12C*. One of these vectors coded for a splice acceptor, 2A sequence and a Puromycin resistance gene. It relied on the endogenous promotor to drive expression of Puro, to allow antibitoic selection of targeted clones. The second vector tested an exogenous PGK promotor to drive puromycin resistance. Each vector was incorporated with ~50% efficiency in hESCs and iPSCs, with successful targeting to one or both alleles, without off-target integrations. Importantly, targeted cells remained pluripotent and maintained a normal karyotype.

The same study also used the AAVS1 locus to demonstrate the over-expression of an inducible vector; containing a minimal CMV promoter, the tetracycline operator minimal promotor and eGFP. Targeting the vector in the same and reverse orientation to the PPP1R12C gene resulted in similar targeting efficencies (40-47%). Clones were transduced with an rTTA rLV and showed doxycycline-inducible GFP expression, with loss of expression after doxycyline removal. This work elegantly showed that the integration of exogenous genes could be achieved at specified sites of the genome. The expression of the targeted TGs could be driven using both endogenous and exogenous promotors and induced using the TET-ON system, in hESCs or iPSCs.

4.7.2 Transcription Activator-Like Effector Nucleases (TALENs)

Shortly after the discovery of ZFNs, another genome editing tool was discovered; transcription activator-like effector nucleases (TALENs). Similar to ZFNs, TALENs consist of a non-specific FokI nuclease domain, fused to a customizable DNA-binding domain containing highly conserved repetitive units, derived from transcription activator-like effectors (TALEs) (Joung, 2010). TALEs are secreted proteins of the *Xanthomonas* bacteria that alter gene transcription in the host cells of plants that it infects (Boch, 2010). TALENs can be designed easily and rapidly, using a simple protein-DNA code, to regulate modular TALE domains to specific target DNA sequences. A follow up study by the Jaenisch lab showed that TALENs could genetically modify hESCs and iPSCs with a similar efficiency to ZFNs (Hockemeyer, 2011). TALE repeats recognise a single base-pair, unlike ZFNs which

are slightly more restricted due to a triplet DNA recognition motif, allowing greater flexibility during their design. This, along with the ease of design over ZFNs, quickly propelled TALENs to become the most popular method at the time for achieving genome editing (Gaj, 2013).

4.7.3 Clustered regularly interspaced short palindromic repeats (CRISPR/Cas9)

More recently, the use of CRISPR/Cas9 genome editing tools have propelled the targeting of PSCs, as well as many other cell types. Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins with helicase and nuclease activity were first demonstrated to mediate adaptive immunity in infection experiments of the lactic acid bacterium *Streptococcus thermophilus* with lytic phages (Barrangou R, 2007). Based on these finding, researchers were able to determine the exact sequences required for Cas9 recruitment to enable DNA cleavage, and combine these into a single chimeric RNA-guide, directing the Cas9 nuclease type II enzyme, to exploit their use in mammalian cells for genome editing (Jinek M, 2012). Through the engineering of several chimeric guides they found they could rationally design chimeric RNA guides to target any DNA sequence of interest, with limited design constraints. The single constraint found was that the target DNA sequence must contain an adjacent GG dinucleotide, known as a protospacer adjacent motif (PAM), which is essential for Cas9 binding and cleavage of the target sequence.

The designed chimeric RNA fused together 20 nucleotides from the 3’ end of a CRISPR-RNA (crRNA), for target recognition, followed by a hairpin structure which allows interaction with the 5’ end of trans-activating crRNA (tracrRNA). This guide RNA recruits the Cas9 protein to the target DNA, where two distinct endonuclease domains perform DNA cleavage. A HNH nuclease domain cleaves the complimentary strand to the guide RNA, while a RuvC-like domain cleaves the non-complimentary strand, forming a double strand (DS) break in the target DNA. Due to the introduction of DS breaks through CRISPR-Cas9, two different DNA repair pathways can be exploited (Takata M, 1998). Gene knock outs can be generated through the error-prone non-homologous end joining (NHEJ) pathway, if a repair template is not provided. NHEJ creates small insertions or deletions (indels), which often leads to frameshifts or premature stop codons in the DNA sequence. Alternatively, the DS break can be used to insert DNA, by homologous recombination (HR), by providing a homologous repair template.

Due to the ease of design and minimal restraints on the target DNA sequence, the CRISPR-Cas9 system has quickly been established as an efficient, versatile and programmable tool for genome editing. Despite being a relatively new technology, there are many examples to date of CRISPR/Cas9 being used to generate modified PSCs, demonstrating their ease of use and design.

Aims and Hypotheses

1) Evidence based on clonogenic assays performed on day 9 forward programmed cells, resulting in both MK and CFU-E colonies, suggests that the MK forward programming (FoP) protocol produces a bi-potent progenitor cell population. The first aim of my PhD is to show whether this is the case, and whether these cells are capable of maturing into both MKs and erythroblasts. It will be important to fully characterise the erythroblasts produced to see if they share a similar phenotype to erythrocytes produced *in vivo*, as has been done for FoP-MKs previously (Moreau et al, 2016).
2) The existing FoP protocol requires lentiviral transduction to achieve overexpression of the three transcription factors (GATA1, TAL1 and FLI1) in iPSCs. This step leads to a number of issues (discussed in further detail in Chapter 2), namely; poor reproducibility and limitations for future large scale production of *in vitro* platelets. Therefore, the second aim of my PhD is to show whether it is possible to FoP with a single lentivirus, containing all three TFs in tandem, by generating a polycistronic cassette. A polycistronic cassette will ensure that all three TFs have been overexpressed and we hypothesise this will result in a more homogenous population of cells that will FoP more efficiently, improve MK and erythroblast yield and importantly, lead to better reproducibility.
3) If the polycistronic cassette enables forward programming, this will be a good proof of principle that an equal copy number of the three TFs can produce MKs and erythroblasts. This will allow us to move onto the third aim of my PhD, to produce an inducible, stable iPSC line for forward programming. The aim would be to insert the polycistronic cassette into a genomic safe harbour of iPSCs, under the control of an inducible system, to allow controllable expression of the three TFs. In doing so, I aim to achieve more reproducible results and provide a more scalable forward programming platform for moving towards the production of platelets for clinical transfusion purposes.
4) Long-term forward programmed MK cultures can be maintained for up to 100 days, after which the cultures crash. We hypothesise this is due to an exhaustion of MK-progenitors in the culture, which have yet to be well defined. The final aim of my PhD is to look at the whole transcriptome of long-term cells, to try and identify whether multiple populations are indeed present and to try and identify a progenitor signature that will enable future isolation for further study.

Main Materials and Methods

Tissue culture

All tissue culture (TC) is performed in a biosafety containment level 2 (CL2) TC laboratory, with appropriate safety measures taken including personal protective equipment (PPE), TC handling in microsafety cabinet Class 2 (MSCII), biohazardous waste specific disposal route and dedicated laboratory space. Care must be taken to keep all equipment is sterile to avoid contamination of samples and maintain highest TC standards while avoiding use of antibiotics in all cultures. Standard TC-treated polystyrene culture plates are used for adherent and suspension cultures (Corning; 6, 12, 24-well plates, T25, T75 and T150 ventilated flasks). All TC is carried in normoxic conditions in incubators set at 37°C, 5% CO2. All culture media are kept at 4°C, brought to room temperature before use and have a shelf-life of a month after opening. The cytokine stock solutions are kept as frozen aliquots at -20°C (BMP4, FGF2, SCF) or -80°C (TPO) and for up to five days at 4°C after thawing (except TPO which is maintained at -20°C after the first thaw); the cytokines are never freeze-thawed more than twice.

Human Cell lines

Induced PSC lines used: **Bob** (ID: A1ATD1), **BobC** (ID: A1ATD1-c), **BBNX** (BRC iPSC Cambridge Facility) were all derived from skin fibroblasts with the monocistronic iPS reprogramming kit (Vectalys), consisting of four retroviral vectors encoding: OCT4, SOX2, KLF4, v-MYC. Passages: 20-55. The Bob line was derived from a patient with an alpha1 anti-trypsin gene mutant, which has been corrected in the BobC line (Yusa et al, 2011). **Qolg** (ID: HPSI1113i-qolg_3, Wellcome Trust Sanger Institute, HipSCI project) and **S4** (ID: S4 -SF5, BRC iPSC Cambridge Facility) were both derived from skin fibroblasts, with Sendai reprogramming kit CytoTune 1 (Thermo Fisher Scientific) with virus encoding: OCT4, SOX2, KLF4, v-MYC. Passages: 31-47. **FFDK** (ID: FSFE11a, Wellcome Trust Sanger Institute, Cellular Generation and Phenotyping (cGAP)) was derived from human skin fibroblasts using episomal vectors encoding: OCT4, SOX2, KLF4, v-MYC. Passages: 25-60.

Embryonic PSC line used: **H9**, derived from totipotent cells of the inner cell mass of a human embryo (Thomson et al. 1998); Passages: 30-42 (WiCell, hPSC Reg ID: WAe009-A).

No karyotyping has taken place for any of the above lines in the Ghevaert lab during the time taken to produce this thesis.

Other cell lines used: HEK293T cells (human embryonic kidney, ATCC CRL-11268), HCT116 cells (human colon carcinoma, ATCC CCL-247*).*

Media

PSC lines: AE6 media is supplemented with Zebrafish fibroblast growth factor 2 (FGF2) 15ng/ml (R&D) and Human Activin A 15ng/ml (Cambridge Stem Cell Institute) for culturing PSC lines. AE6 media (in house version of commercial Essential-6): DMEM/F12 (Life Technologies, with L-Glutamine 2.5mM, HEPES 15mM, D-Glucose 3.2g/L, Phenol red 0.02mM), supplemented by 0.054% NaHCO3 (Thermo Fisher Scientific), 64mg/L L-Ascorbic Acid (Thermo Fisher Scientific), 20mg/L Insulin, 11mg/L Transferrin and 0.0134mg/L Selenium (ITS, Life Technologies).

HEK293T cells: DMEM+Glutamax (Life Technologies) supplemented with 10% Fetal Bovine Serum (FBS, Fisher Scientific) and 1% Penicillin and Streptomycin (PenStrep, Life Technologies).

HCT116 cells: McCoy's 5A (Life Technologies) supplemented with 10% FBS and 1% PenStrep.

Passaging Cell lines

Cells are split when they reach approximately 70-80% confluency. When required, cells were counted after staining with Trypan Blue solution (0.4%, Thermo Fisher Scientific), using disposable C-Chip haemocytometers (Neubauer).

PSC lines: PBS (Sigma Aldrich) + EDTA (1mM) is applied to PBS washed wells and left to incubate 5 min, RT. PBS/EDTA solution is aspirated, 1ml media added, then the well is scraped using the end of a 5ml serological pipette (Sigma Aldrich). Clumps of cells are collected in AE6 before re-suspending with a P1000 tip before plating at the required dilution (usually 1:30 for routine maintenance). For the clean-up of background differentiation, 1mg/ml Collagenase Type IV and Dispase II (both Life Technologies) are added to PBS washed wells, before being incubated for 30 min-1 hour at 37^{o}C, 5% CO_2. Wells are then washed to collect clumps, further broken by gentle pipetting, and added to a 15ml falcon. 10ml media is added and clumps are left to settle to the bottom for 10 min. This wash step is repeated twice, before cells are re-suspended in AE6 and plated. When required, PSCs are split as single cells using TrypLE (described below) and spun at 100g, 5 mins. Single cell PSCs require the addition of Rho-associated protein kinase (ROCK) Inhibitor Y-27632 (Stemcell Technologies) for 24 hours at 37^{o}C, 5% CO_2.

HEK239T and HCT116 cells:

To split cells into single cells, TrypLE (Life Technologies) is added to washed cells (1x PBS) and incubated, 37^{o}C, 5% CO_2, 5min. At least 3 times volume of media is added to TrypLE, before collecting cells. Cells are spun at 300g, 5mins. Care is taken when feeding these cells lines as they detach easily from TC plastic, so media is added to the top of the flask.

Cryopreserving and Thawing cells

PSC lines: Cells are cryopreserved in 90% (vol:vol) KnockOut Serum Replacement (KOSR, Life Technologies), with 10% (vol:vol) Dimethyl Sulfoxide (DMSO, Sigma Aldrich), at a concentration of approximately 1-2E+06 cells/ml.

HEK293T and HCT116 cells: Cells are cryopreserved in 90% Fetal Bovine Serum (FBS, Life Technologies), with 10% DMSO, at a concentration of approximately 5E+06cells/ml.

Forward Programmed MKs and Erythroblasts: Cells are cryopreserved in Ivscove's Modified Dulbecco's Medium (IMDM, Life Technologies) + 20% (vol:vol) FBS and 5% (vol:vol) DMSO at a concentration of approximately 1-2E+06 cells/ml.

Cells are stored in 1.5ml cryotubes (Nunc) then stored at -80^{o}C, inside a My Frosty Freezing container (Thermo Fisher) to control freezing to -1^{o}C/min, before being transferred for long-term storage at -150^{o}C. Cells are thawed at 37°C and then slowly adding 1ml of basal medium, dropwise, while agitating cells. Cells are then quenched in 10ml basal medium before being centrifuging at 300g, 5 min. Cell pellets are then re-suspending in appropriate media before plating. PSCs recovery after thawing can be improved by the addition of ROCK Inhibitor Y-27632 (Stemcell Technologies) for 24 hours at 37^{o}C, 5% CO_2.

Recombinant Lentiviral Vector Production

This protocol describes the production of amphotrope (VSV-G pseudotyped particles) recombinant HIV-1 lentiviral particles using transient co-transfection into HEK 293T cells. The method has been optimized for a 2nd generation production system using three plasmids (vector, helper and envelope) developed in Didier Trono lab (pWPT-GFP, psPAX2 and pMD2.G respectively - Addgene references: 12255, 12260, 12259). The transfection method used is cationic polymer based. Resulting lentiviral particles are self-inactivating (SIN vectors) due to deletion of the U3 region in the 3'LTR of the vector plasmid. Transgene expression is under the control of the internal EF1α promoter.

(Day 0) 18x10^{6} 293T cells per T150/flask are seeded and left to reattach for 24 hours, 37^{o}C, 5% CO_2. All subsequent stages of this protocol must be performed in a dedicated biosafety level 2 (CL2) viral laboratory including reinforced safety measures and decontamination procedures. (Day 1) Per T150: In a sterile 15ml tube, add 1.5ml RPMI media (basal media, without serum), plus 25µg vector plasmid pWPT (Addgene #12255), 6µg HIV-1 envelope plasmid, pMD2.G (Addgene #12259), 14µg HIV-1 helper plasmid, psPAX2 (Addgene #12260) and 1.5µl Turbofect Transfection Reagent (Fermentas)/µg total DNA. Mix and incubate 30 mins, RT. (Note for larger vector plasmids (such as

Polycistronic vectors) 30µg DNA was used). Add transfection solution dropwise to cells, to ensure even distribution of DNA-polymer complexes and incubate 24 hours, 37°C, 5% CO_2. (Day 2) Wash cells 1x PBS and add 25ml fresh media, incubate 24-48 hours, 37°C, 5% CO_2. (Day 3/4) Filter supernatant containing viral particles, through a 0.45µm minisart filter (Sartorius) with a 50ml syringe, into a 50ml tube. Add DNaseI 5U/ml (Roche) and 1mM $MgCl_2$, and incubate 30' at RT. Add 8ml Lenti-X concentrator solution (Takara/Clontech) mix well, incubate 30mins, 4°C. Centrifuge in an aerosol-tight container 1500g, 45mins, 4°C. Aspirate supernatant keeping viral pellet and re-centrifuge for a further 5mins and aspirate remaining supernatant. Re-suspend pellet in 250µl AE6 media and aliquot into small volumes into 1ml conical cryotubes (Nunc) and keep at -80°C.

Lentiviral Vector Titration

The principle relies on parallel quantification of the copy number of the HIV-1 specific RRE sequence and the HMBS endogene after transduction of the HCT116 cell line (colon carcinoma, ATCC CCL-247). Cells are transduced in parallel by LV batches to be titered and by a control vector ubiquitously expressing a reporter gene that can be detected by flow cytometry (e.g. a GFP expressing vector). Three days after transduction, genomic DNA is purified; in addition, fluorescence is analysed by flow cytometry for GFP transduced control cells. The viral RRE sequence versus HMBS endogene copy number ratio is then measured by QPCR for each sample using a plasmid standard curve. From GFP transduced control sample data, a standard curve correlating the RRE/HMBS ratio to TU values (deduced from flow cytometry data) is obtained. Eventually, TU present in the initial LV volume used for transduction can be deduced from the RRE/HMBS value of test samples and LV titre calculated (TU/mL).

(Day 0) Seed $1x10^5$ HCT116 cells per well of a 24 well plate. Seed enough wells to transduce 2 wells per virus produced to titre, plus a non-transduced well, plus 4 wells for transducing with a GFP virus of a known titre. Incubate ON, 37°C, 5% CO_2. All subsequent steps of this protocol, until gDNA processing, must be performed in a CAT2 viral laboratory. (Day 1) Add Polybrene (PB) to media to get final concentration of 10ug/ml per well. Thaw lentiviral batches on ice and add 2µl and 1µl of the test virus to 2 wells. Dilute GFP virus (known titre >5E+8TU/mL) 1/100 and add 125µl, 25 µl, 5 µl and 1µl to 4 wells. Gently shake and incubate 24 hours, 37°C, 5% CO_2. (Day 2) Wash wells 1x PBS and add 1ml fresh media/well. Incubate 48 hours, 37°C, 5% CO_2. (Day 4) Collect GFP-transduced well and any other well containing GFP in the construct, after 1x wash with PBS, with 100µl TrypLE, incubate 5mins, 37°C, 5% CO_2. Add 300µl media to collect cells take a one quater for flow analysis by putting directly into 500µl PBS + 1% formaldehyde (FA, Sigma). Add large volume of PBS and spin 300g, 5mins, RT. Resuspend pellet in 150µl cell lysis buffer (Wizard SV genomic DNA purification system,

Promega). All other wells, perform cell lysis directly in well after 1x PBS wash. These samples are then processed as per the instruction for gDNA extraction with Wizard SV genomic DNA purification system. GFP samples are analysed by flow cytometry to get an accurate titre for standard curve. qPCR is performed with the gDNA samples, to amplify the lentiviral provirus (RRE primers) and the human endogene HMBS. Serial dilutions of a lentiviral pWPT plasmid encompassing the HMBS sequence are used to provide standard curve points and absolute quantification of RRE/HMBS copy number. From flow cytometry data the following equation allows you to calculate the average titre of the GFP: (10E+5 cells) x (% GFP +ve) / (volume lentivirus in ml) = Titre (TU/ml). Based on this you can calculate the absolute number of TU added to cells on day 1 eg if the calculated titre is $4.5x10^8$ TU/ml, then 4500 TU were added to cells in 0.01uL sample. From qPCR data calculate the RRE/HMBS copy number ratio and draw a standard curve using the GFP control sample TU values (Y-axis) and the RRE/HMBS values (X-axis). Using the GFP standard curve it is possible to calculate the mean functional titre for each lentiviral batch tested, based on the RRE/HMBS ratio obtained by qPCR from the 2 transduction points.

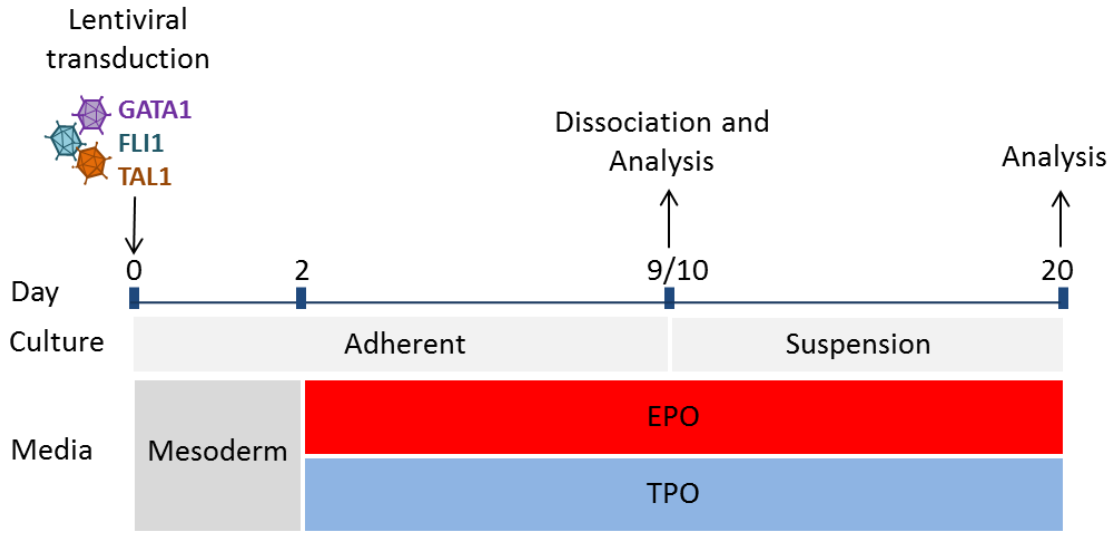


Diagram 2.1- The Main Steps of the Forward Programming Protocol. Briefly: iPSCs are seeded 24 hours before lentiviral transduction at day 0, at the same time mesoderm media is added to cells. After 2 days in mesoderm media, cells are then grown in either TPO (for MK FoP) or EPO (for erythroblast FoP). At day 9/10, cells are dissociated and analysis performed. This is routinely flow cytometry, and occasionally by cytospin, CFU assays, qPCR. After further culture in TPO or EPO medium, cells are analysed at day 20, routinely by flow cytometry and occasionally by cytospin, qPCR, Western blot. Cultures can be maintained further than 20 days.

MK Forward Programming

(Day -1) ~1x10^5 cells are collected using PBS+EDTA method and reseeded as small clumps onto a vitronectin coated well of a 12-well plate, in AE6 media +FGF2 +Activin A (as above) and allowed to reattach for 24 hours. Single cells can also be seeded, following the same protocol as described above for PSCs. The single cell method is stated in specific chapters when used. All subsequent steps of this protocol must be performed in a CL2 viral laboratory. (Day 0) Cells are transduced with the appropriate volume of recombinant lentivirus to get the desired multiplicity of infection (MOI). Conventional forward programming requires GATA1, TAL1 and FLI1 lentiviruses, thawed on ice, all used at an MOI 20 (unless otherwise stated). All lentiviruses used were either produced in-house (by method described above), or produced commercially (Vectalys). Lentivirus mix is added to 0.5ml mesoderm-inducing FLyB medium: AE6 + FGF2 20ng/ml + BMP4 10ng/ml (R&D) + PI3K Inhibitor LY-294002 10µM (Sigma)/ well of a 12-well plate. For transduction of multiple wells a master mix is prepared. At this stage protamine sulphate (PS) 10µg/ml (Sigma) is also added to the FLyB-virus mix. Cells and transduction mix are left to incubate for 24 hours. (Day 1) Cells are washed 1x PBS before 0.5ml fresh FLyB media is added and incubated for 24 hours. (Day 2) Medium is changed to 0.5ml MK-1 media: CellGro SCGM (CellGenix) + Human TPO 100ng/ml (CellGenix) + Human SCF 25ng/ml (Life Technologies), for MK differentiation (unless otherwise stated in specific chapter methods). Cells can be left for 48-72 hours before fresh medium is added. The first medium addition requires 0.5ml MK media with 2x concentration of cytokines. Subsequent medium changes (50% exchange) involve careful removal of half the medium by titling the plate and collecting only medium, not cells, before adding half the volume fresh medium (2x cytokines). (Day 9/10) The supernatant of each well is collected into a 15ml falcon tube and the well rinsed 1x PBS (0.5ml), before being pooled with the supernatant. To collect the adherent cell fraction 300µl TrypLE is added to wells and incubated for 10 mins at 37^{o}C, 5% CO_2. The TrypLE-cell mix is added to the corresponding falcon tube and quenched with 10 ml PBS before being centrifuged 300g, 5min, RT. Cell pellet is then re-suspended in 0.2ml MK-2 media (TPO 20ng/ml, SCF 25ng/ml). 5%-10% cells are used for flow cytometry analysis to monitor MEP and MK markers (see below). Remaining cells are re-plated onto TC 6 well plates in a total of 2ml MK-2 media. At this stage it is safe to remove cells from the CL2 viral laboratory to be subsequently handled in a standard CL2 TC laboratory. Medium is refreshed every 48-72 hours, (50% exchange). (Day 20/21) Cells are collected and stained again to monitor MK maturation by flow cytometry. Cells can be re-plated and maintained long-term by refreshing the MK-2 media every 3 days (50% exchange) and checking MK purity every 7-10 days. Cell density can be adjusted once an accurate cell count has been obtained by flow cytometry. Optimal cell density is 2E+5-1.5E+06 cells/ml. Cells can be split 1:5 when cell density exceeds 1.5E+06 cells/ml. Forward programmed MKs

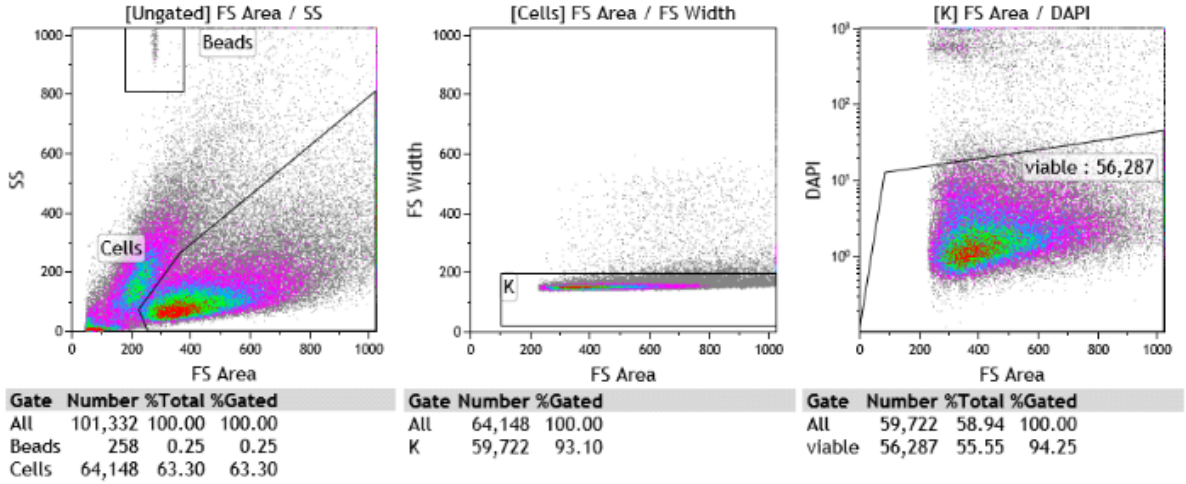
(FoP-MKs) can be frozen in IMDM 20% FBS 5% DMSO ideally at 0.5-1E+6 cells per vial. Upon thawing, these cells can be placed back in MK-2 media (TPO 20ng/ml, SCF 25ng/ml), refreshing every 2-3 days.

Erythroblast Forward Programming

Performed as for MK forward programming, with TPO containing MK medium substituted for EPO containing Ery media; CellGro SCGM (CellGenix) + recombinant human EPO 2U/ml (R&D) + SCF 50ng/ml (Life Technologies) + Insulin 10µ/ml (Roche) + Transferrin 30µg/ml (Roche) + IL-3 10ng/ml (R&D).

Flow Cytometry

Flow cytometry was performed on a Gallios Flow Cytometer (Beckman Coulter) and analysed using the Kaluza analysis 1.5a software. Background fluorescence was set against matched isotype control antibodies and compensation matrix defined using single-color stained cells. Flow count fluorospheres and DAPI were used to determine viable cell count in samples.



Gate	Number	%Total	%Gated
All	101,332	100.00	100.00
Beads	258	0.25	0.25
Cells	64,148	63.30	63.30

Gate	Number	%Gated
All	64,148	100.00
K	59,722	93.10

Gate	Number	%Total	%Gated
All	59,722	58.94	100.00
viable	56,287	55.55	94.25

An example of the gating strategy used for all flow cytometry analysis presented throughout this thesis. Cells are initially gated, removing debris and fluorospheres from analysis. Next, cells are gating on singlets, to remove doublets from analysis. Finally, DAPI negative viable cells are gated, to remove dead cells from downstream analysis. The resulting viable, single cell population are then analysed for routine FoP markers (found in **Table 2**).

Immunostaining for flow cytometry analysis is routinely done in 20uL culture medium or flow buffer (DPBS 0.5% BSA 2mM EDTA) in an Eppendorf tube, including separate unstained and single colour stained controls, prepared from pooled samples. Routine day 9/10 MKFOP analysis requires CD41a/CD235a/CD42a staining. Abs and suppliers are listed in **Table 2**. An Ab master mix is prepared

before Abs are added to cells (1:10-200 dilution depending on Ab and determined after titration; **Table 2**), mixed and incubated 20-30mins, RT, in the dark. 1ml Flow buffer is added to Ab-cell mix and spun at 300g, 5mins, RT. Cell pellet is re-suspended in 0.5ml flow buffer + DAPI, flow count fluorospheres (1000/100µl, Beckman Coulter) + 0.5% FA (Sigma). Flow count fluorospheres require mixing and incubation at 37^{o}C for 30 mins prior to use. Day 20+ cells, now MKs, must be centrifuged at 120g, 8min (acceleration and brake 3). Routine day 20 analysis is as for day 9/10 (CD41a/CD235a/CD42a or CD42b. At this stage cells do not need to be fixed, so are re-suspended in flow buffer.

Colony Forming Unit (CFU) Assays

To determine the progenitor amount in FOP cultures, we used clonogenic assays in semi-solid medium. MethoCult methylcellulose with recombinant cytokines for human cells (#4435, StemCell Technologies), for assessing both MK and Erythroid potential, or MethoCult methylcellulose without cytokines for human cells (#4230, StemCell Technologies) supplemented with TPO 100ng/mL and SCF 50ng/mL for assessing MK potential only, were used for CFU assays. 15,000 cells in 300µL IMDM are added to 3ml methylcellulose (MC) medium and vortexed well before being incubated for 10 mins to remove air bubbles. Using a P1000, 600ul is pipetted into a 35mm round dish (non-treated, Corning #430588), starting in the middle and spiralling outwards towards the sides. Another 600ul is added to the same dish, before gently swirling to make sure all the way to the edge of the plate is coated. This is then repeated with a second 35mm dish to duplicate the same condition resulting in a duplicate assay with approximately 5000 cells/replicate dish. A 3rd 35mm dish has dH_2O added, and the lid removed to keep CFU dish fully humidified. All 3 dishes are placed into a 10cm^2 Petri dish and placed in an incubator. CFU assays should be disturbed as little as possible, and not removed from the incubator for 14 days, before counting colonies. Colony types counted include mixed, CFU-E, MK and MK progenitor colonies.

Cell Morphology Analysis

Cells were spun onto a glass slide using CytoSep single funnel (Simport), at 400g for 5 mins before being methanol fixed and stained using the Rapid Romanowsky staining kit (TCS Biosciences Ltd), following manufacturer instructions.

Mammalian Cell RNA extraction and cDNA synthesis

Cells were collected and processed with the Qiagen RNeasy mini kit for total RNA extraction as per the manufacturer guidelines, including the optional DNAse I treatment (NEB). RNA concentrations were obtained using the Qubit Fluorometer (Thermo Fisher) using the Qubit RNA HS Assay kit

(Thermo Fisher), following manufacturer guidelines. cDNA synthesis was performed using Maxima first strand cDNA synthesis kit for RT-qPCR (Thermo Scientific) following manufacturer guidelines, including minus reverse transcriptase and no template controls. Equal starting RNA concentrations were used for cDNA synthesis, usually 50ng. RNA is kept at -80°C and cDNA -20°C for long-term storage.

Recombinant DNA Cloning

Cloning strategies are described in specific chapter methods sections. Plasmids were grown ON in LB broth + appropriate antibiotic (in text) or on plates made from LB Agar + appropriate antibiotic. All restriction enzymes (RE) used were supplied from New England Biolabs (NEB) and were used with optimal buffer supplied. RE digests were routinely performed at 37^{o}C for 1 hour. 1-2% TAE or TBE gels were used to analyse DNA fragments (TAE gels were used when the product needed to be gel purified). SafeView Nucleic acid stain (NBS Biologicals Ltd) or SYBR safe DNA gel stain (Invitrogen) was added to visualise DNA or for DNA extraction. 6x Orange G loading buffer (Sigma) was added to DNA samples before loading into a gel. Gel purification was done following the manufacturer instructions for the QIAquick gel extraction kit (Qiagen). Dephosphorylation was performed with either Antarctic phosphatase (NEB), or using calf intestinal alkaline phosphatase (NEB), as per manufacturer instructions. Klenow (NEB) was also used for modifying DNA ends, following manufacturer instructions. PCR products intended for cloning were purified using the QIAquick PCR purification kit (Qiagen), following manufacturer instructions. Ligations were performed with T4 DNA ligase (NEB or Takara). Chemically competent cells used for transformations include One shot DBH10 cells (Invitrogen), One shot Stbl3 (Life Technologies) and Mix & Go cells (Zymo Research), following manufacturer guidelines.

PCRs performed to generate fragments for Gibson cloning, were performed using Q5 Hot start high-fidelity DNA polymerase (NEB) and the Gibson Assembly master mix (NEB) was used for Gibson cloning, following manufacturer's instructions. Gateway reactions were performed with Gateway LR Clonase II Plus Reaction mix (Invitrogen), following manufacturer instructions. All primers, used for sequencing or PCR, were supplied by Sigma.

Plasmid DNA preparation

DNA was purified using both QIAGEN QIAprep spin miniprep kit or plasmid Midi kit for larger concentrations, after plasmids were grown up ON in LB broth + appropriate antibiotic. DNA concentrations were obtained using a NanoDrop 2000 UV-Vis spectrophotometer (Thermo Fisher) or

the Qubit Fluorometer (Thermo Fisher) using the Qubit DNA Broad Range Assay kit (Thermo Fisher), following manufacturer guidelines.

Analytical and Preparative PCR

PCRs were performed with sequence specific primers (Sigma Aldrich) described in text, **Table 4-5**, using Phusion high fidelity DNA polymerase (NEB) following manufacturer guidelines.

Real Time Quantitative PCR

Quantitative PCR (qPCR) was performed using BrilliantII SYBR green master mix (Agilent) in 0.2ml non-skirted low profile 96-well PCR plates (Thermo Scientific), using optically clear flat 8 cap strips (Thermo scientific). Plates were run on the Stratagene Mx3000P machine (Agilent), with the thermal profile "normal 2 step" (95°C/10", 60°C/30"; 40 cycles) including a terminal melting curve. Data was analysed using the software MxPro- Mx3000P. Duplicate wells were used for all samples. Minus reverse transcriptase negative controls and no template controls were included in qPCR samples to double check no contaminating gDNA remained after cDNA synthesis or that reagents used were contaminated. Relative gene expression was calculated by the 2-ΔCt method using HMBS as housekeeping gene for normalization. qPCR primer pairs (**Table 3**) designed to amplify only cDNA, to detect all known isoforms, and to have no reported off-target matches searching the human NCBI RefSeq database were tested within 80-120% PCR efficiencies with single dissociation curves. We used UTR targeting (absent from transgenes) to monitor endogene expression while transgene specific primer pairs used a common reverse primer specific to the viral vector.

Statistical Analysis

Two-tailed *t*-tests have been performed on all results presented where appropriate, with *P* values of 0.05 or lower indicated to show statistically significant results. N values, along with details of technical and biological replicates, are indicated in figure legends.

Materials and Methods Tables

Table 2. Antibodies used for Flow Cytometry and Western Blot

Antibody	Fluorochrome	Assay concentration	Catalogue Number	Manufacturer
CD14	FITC	1:10 dilution	345784	BD Pharmingen
CD36	FITC	1:10 dilution	561820	BD Pharmingen
CD36	Pe	1:10 dilution	555455	BD Pharmingen
CD41a	APC	1:10 dilution	559777	BD Pharmingen
CD41a	APC-H7	1:100 dilution	561422	BD Pharmingen
CD42a	FITC	1:10 dilution	558818	BD Pharmingen
CD42b	APC	1:20 diultion	551061	BD Pharmingen
CD66c	PE	1:10 dilution	551478	BD Pharmingen
CD71	PE	1:10 dilution	561938	BD Pharmingen
CD71	APC-H7	1:100 dilution	563671	BD Pharmingen
CD235a	PE	1:200 dilution	555570	BD Pharmingen
CD235a	PE-Cy7	1:100 dilution	563666	BD Pharmingen
CD235a	APC	1:10 dilution	551775	BD Pharmingen
BAND3	PE	1:200 dilution	9439PE (BRIC 6-Pe)	IBGRL Research Products
Calcein-AM		1:20000 dilution	17783	Life Technologies

For Western blot			
Gene name	**Probe name**	**Assay concentration**	**Manufacturer**
α-globin	sc-514378	1:2000 dilution	Santa Cruz
β -globin	sc-21757	1:2000 dilution	Santa Cruz
γ-globin	sc-21756	1:2000 dilution	Santa Cruz
ξ-globin	ab156041	1:200 dilution	Abcam
β-actin	A5441	1:15000 dilution	Sigma Aldrich

Table 3. Primers used for RT-qPCR (all human sequences)

Taqman Probes		
Primer name	**Gene name**	**Probe name (Life Technologies)**
GAPDH	GAPDH	Hs99999905_m1
α-globin	HBA1+HBA2	Hs00361191_g1
β -globin	HBB	Hs00758889
γ-globin	HBG1+HBG2	Hs00361131_g1
ξ-globin	HBE1	Hs00362215_g1
ζ-globin	HBZ	Hs00923579-m1
RT-qPCR Primers		
Gene name	**Forward primer**	**Reverse primer**
GAPDH	AAGGTGAAGGTCGGAGTCAAC	GGGGTCATTGATGGCAACAATA
HMBS	ATTACCCCGGGAGACTGAAC	GGCTGTTGCTTGGACTTCTC
MDH1	GGGTGTCCTGGACGGTGTCCT	CCCTTCTTGGCATGGAGCCCAC
FLI1 endogenous	GGGCTCGGCTGCAGACTTGG	AGATGGGCTGCCGCTCCGTA
GATA1 endogenous	TTGCCACATCCCCAAGGCGG	GGGGGAGGGGCTCTGAGGTC
TAL1 endogenous	AGCAAAGACCCGGGTGTGCATC	CCTCTAGCTGGGGGTCACTGCG
Polycistronic transgene	CCAGACACAGAGTGCCTACC	AGGCAGTTCAGCTGTCACA
WPRE (pWPT Re)	Use transgene forward primer	GCAGCGTATCCACATAGCGTAAAAGG
FLI1 transgene	CCCGCCATCCTAACACCCACG	Use WPRE
GATA1 transgene	GGTGGCTCCGCTCAGCTCAT	Use WPRE
TAL1 transgene	(in pTRIP) AGGCGGTGGACTTGAACTTT	TCTAGCCAGGCACAATCAGC
RT-qPCR primers for lentiviral vector titration		
HMBS	ATTACCCCGGGAGACTGAAC	GGCTGTTGCTTGGACTTCTC
RRE	TTTGTTCCTTGGGTTCTTGG	GATGCCCCAGACTGTGAGTT

Results

Chapter 1

Erythroblast Forward Programming

Introduction

Forward programming (FoP) was established in our lab for the purpose of generating platelets *in vitro*. During the development of this technique, Dr Moreau observed that day 10 forward programmed cells had the ability to generate both MK and CFU-E colonies when cultured in enriched methylcellulose CFU assays, used to assess progenitor potential of haematopoietic cells. This led to the hypothesis that FoP must generate bi-potent progenitor cells, like the megakaryocyte-erythroid progenitors (MEPs) found *in vivo*, in order for these two colony types to arise. This observation therefore raised the question of whether FoP could potentially be used as a method for producing erythroblasts and reticulocytes (mature, enucleated red cells).

Other methods exist for producing erythrocytes *in vitro*, however, these have a number of drawbacks. Firstly, some rely on using CD34+ cells derived from cord or peripheral blood, a non-renewable source of cells (Olivier EN, 2006 and Giarratana MC, 2005). Those which utilise ESCs or iPSCs as the starting material, rely on directed differentiation approaches, which are often time consuming and involve complex cytokine cocktails at various stages of culture (Olivier EN, 2016 and Dias J, 2011). Additionally, they often require co-culture with mouse feeder cells, or serum use, which make these protocols unsuitable for producing clinical grade cells for human use. Finally, many ESC/iPSC protocols have a poor erythrocyte cell number outcome, are very poor at producing mature enucleated erythrocytes and fail to complete haemoglobin switching from an embryonic or foetal phenotype to adult (Dorn I, 2015). Thus, major drawbacks exist for current *in vitro* methods of deriving erythrocytes.

A good manufacturing-compatible (GMP) protocol is still required, which additionally should be amenable to scaling-up when needed, in order to produce the large quantities of cells required for transfusion. Currently, few methods exist which would fulfil these requirements. One recently published protocol which shows promise, produces a highly pure population of erythrocytes (approximately 95% purity by day 31 of culture) in both feeder and serum-free conditions (Olivier EN, 2016). This protocol also supports the large expansion of cells, with a single iPSC giving rise to 50,000-200,000 erythroid cells by day 31 of culture. However, this protocol requires many cytokines and a number of small molecules throughout various stages of culture, which would add substantial cost to large-scale manufacture. Despite achieving approximately 10% enucleation, the authors themselves discuss the need to attain more efficient enucleation in order to obtain a viable method for transfusion of cells into man, as transfusing nucleated cells into a patient has associated risks. The cells produced by their protocol express foetal alpha and gamma globins predominantly, with a small amount of adult beta globin expression, which the authors suggest could be improved with

additional stimuli. Encouragingly, since a predominantly foetal, not embryonic, phenotype is observed, similar to that of cord blood, these cells would be compatible with clinical benefits post-transfusion.

More recently, an alternative approach has been published, whereby an immortalised human erythroid line (BEL-A) was established by utilising a Tet-inducible HPV16-E6/E7 expression system in CD34+ cells, first developed in 2013 (Kurita R, 2013). The CD34+ cells were matured for several days before the addition of doxycycline to switch on the viral oncogenes. Immortalised cells were expanded for 100 days and show a pro- to early basophilic erythroblast phenotype. BEL-A cells are then transferred to a primary erythroid culture medium (containing doxycycline) for 6 days, then a tertiary medium (excluding doxycycline) to promote maturation towards orthochromatic normoblasts and reticulocytes (Trakarnsanga K, 2017). This approach negates two main issues with other existing *in vitro* methods of erythrocyte production. Firstly, this immortalised line can reproducibly generate large quantities of progenitor cells. Secondly, once differentiated, these cells display a mature adult phenotype, producing adult beta-globin and approximately 30% of cells enucleate. Additionally, BEL-A cells show a similar survival rate as donor RBCs in mice *in vivo*, and show further maturation after 24 hours of transfusion. However, drawbacks of this method include the expensive expansion media, StemSpan SFM (StemCell Technologies Inc, £349/500ml), used to maintain the immortalised BEL-A line, which could hinder cost-effective large scale culture, highlighting a requirement for advancements in the culturing very large cell quantities.

Forward Programming: An Alternative Method to Generate Erythroblasts *in vitro*?

As previously shown, to enable forward programming iPSCs must be transduced with all 3 exogenous TFs used, in order to follow the FoP developmental pathway (Moreau T, 2016). Thus, to generate bi-potent progenitors requires *GATA1*, *TAL1* and *FLI1*. However, it was not known whether the presence of the *FLI1* TG in these cells, and the resulting *FLI1* endogenous gene expression which it initiates, would inhibit differentiation along the erythrocyte lineage in liquid culture, since *FLI1* is a known inhibitor of this pathway (Athanasiou M, 2000). Since the transduction of the 3TFs in iPSCs is stochastic, yet we see a highly pure population of MKs in a short time span, we hypothesised that a thrombopoietin (TPO) containing promoting medium is highly selective for only the cells that contain the 'optimum' levels of TGs for generating MKs. Therefore, we reasoned that if we tried FoP in erythroblast (EB) media conditions containing erythropoietin (EPO), this may select for cells which had received lower levels of *FLI1* TG, and contain a more 'optimal' mix of TGs to generate erythroblasts.

In order to explore whether FoP had the potential to generate erythroblasts in liquid culture, the normal FoP protocol was adjusted, using an EB medium containing EPO instead of TPO. This resulted in an increase in the number of CD41-/CD235+ cells generated, compared to that obtained using MK medium. This was the first evidence to suggest that FoP does generate a truly bi-potent progenitor cell population and that FoP could be used to generate erythroblasts in culture *in vitro*. We wanted to know whether we could optimise the number of erythroblasts produced, and characterize these cells, to see if FoP would be an efficient means of generating erythroblasts *in vitro*. This would offer an exciting avenue of research for studying both erythropoiesis *in vitro*, but that could also lead to the generation of erythrocytes for use in transfusion medicine. It could also facilitate the work of disease modelling, in particular by providing a more reliable source of erythroid cells for researchers working on malarial infection. For example, genome editing in mESCs differentiated into erythrocytes has revealed that the glycohphorin C receptor, found on the surface membrane of erythrocytes, is crucial for the invasion of the parasite *Plasmodium berghei* (Yiangou L, 2016). This would not have been possible in erythrocytes, as their enucleation makes genetic approaches impossible. Crucially, it would have a major advantage over existing methods, as it could also provide an *in vitro* method for studying the cell-fate decision of the bi-potent MEP cell, which is currently not well understood and difficult to study.

Chapter Overview

The following chapter describes work which shows for the first time that FoP reproducibly generates a functional, bi-potent progenitor cell population which optimally produces erythroblasts in EPO-containing medium, or MKs in TPO-containing medium, in three different iPSC lines. Medium conditions to produce erythroblasts from iPSCs have been optimised for FoP to produce erythroblasts. Characterisation of these erythroblasts show that FoP, like existing directed differentiation methods, does not produce adult-like cells, with embryonic and foetal globins but no adult beta globin detected. Cell expansion was good, with approximately 140 erythroblasts produced per starting iPSC, in just 20 days of culture. This shows an alternative approach for studying the development of erythroblasts, as well as bi-potent progenitor cells *in vitro*.

Materials and Methods

Experiments described in this chapter used techniques, such as qPCR (for endogenous and TG expression only) and flow cytometry, described in the main Materials and Methods chapter. Only chapter-specific differences are described below.

Testing Different Initial Forward Programming Media Conditions

The experiment described in **Fig 1.2**- **Fig 1.5** of this chapter was performed to try and establish the best media conditions for generating erythroblasts by FoP. The iPSC line Bob, cultured in CDM12 supplemented with FGF2 (R&D) and Activin-A (Cambridge Stem Cell Research), both 15ng/ml, was seeded in clumps at a density of approximately 10E+05 cells/well of a Gel-MEF pre-coated 12 well plate, into 24 wells. Transduction with *GATA1*, *TAL1* and *FLI1* rLV was performed as described in the main Materials and Methods chapter, and mesoderm induction for day 0 and 1 followed as described. At day 2, different cytokine conditions were added to duplicate wells, in CellGro SCGM serum free media (CellGenix). The following table describes the conditions and concentrations of cytokines used until day 20 of programming. Media was changed every 2-3 days, by removing approximately half of the old media and replacing with 2x concentrated media.

Media conditions day 2-8	
TPO and SCF	TPO 100ng/ml, SCF 50ng/ml
EPO and SCF	EPO 2U/ml, SCF 50ng/ml
TPO only	TPO 100ng/ml
EPO only	EPO 2U/ml
SCF only	SCF 50ng/ml
No cytokines	-
EB media	EPO 2U/ml, SCF 50ng/ml, Insulin 10µl/ml, Transferrin 30µg/ml, IL-3 10ng/ml
Media conditions day 8-20	
MK media	TPO 100ng/ml, IL-1β 10ng/ml
EB media	EPO 2U/ml, SCF 50ng/ml, Insulin 10µl/ml, Transferrin 30µg/ml, IL-3 10ng/ml

IL-1β (Miltenyi), other cytokine suppliers described in Main Materials and Methods section.

Cells were dissociated at day 8 and 20 using 0.5ml Collagenase IV and Dispase II (both 1mg/ml, Life Technologies). Flow cytometry was performed on the 9 colour CyAn flow cytometer (Beckman-Coulter) and analysed using the Summit 4.3 software.

Colony Forming Unit Progenitor Assays

The main protocol for performing CFU assays is described in the Main Materials and Methods section. All experiments described in this chapter were performed in MethoCult methylcellulose with recombinant cytokines for human cells (#4434, StemCell Technologies), for assessing both MK

and CFU-E potential. Colony counts presented in **Fig 1.6** onwards are the result of blinded counts obtained by 2 or more people.

TaqMan qPCR for Haemoglobin Expression

cDNA was produced following methods described in the main Materials and Methods chapter. For the initial media test experiment (described above), haemoglobin expression was detected using TaqMan probes. TaqMan probes for β-actin, GAPDH, Haemoglobin alpha, beta, gamma epsilon and zeta were used, and sequences are listed in **Table 3** (Main Materials and Methods). TaqMan Gene expression Master Mix (Thermo Fischer Scientific) was used, following manufacturers protocol to prepare qPCR mix. Real time thermos-cycling was performed on the Stratagene Mx3000P (Agilent Technologies), data was acquired and analysed using MxPro qPCR software (Agilent). Fluorescence data was collected for ROX (reference dye, filter position 2) and FAM (filter position 4), with the following thermal profile setup; Segment 1: 1 cycle, 2 min at 50^{o}C, Segment 2: 1 cycle, 10 min at 95^{o}C, Segment 3: 40 cycles, 15 sec at 95^{o}C followed by 1 min at 60^{o}C.

EPO and TPO Experiments

The remaining experiments described in this chapter follow the forward programming protocol described in the main Materials and Methods chapter.

iPSC lines used

The iPSC lines Bob and BobC were used for the majority of experiments described in this chapter. These 2 lines have an almost identical genomic sequence, with the BobC line having been derived from Bob, with a patient mutation corrected (described in Main Materials and Methods). These lines perform similarly in FoP and were both used due to the experiments being performed some time apart, after which time culture of Bob was no longer routine in the lab. Bob was used for the experiment describing different initial cytokines tested from day 2-8 (**Fig 1.2- 1.5**). BobC was used for the experiment to described the non-transduced outcome of FoP (**Fig 1.1**), and the subsequent experiments performed in EPO and TPO experiments, from day 2-20 of FoP (**Fig 1.6-1.9**). FFDK was used as a second line to test whether the results shown in BobC would be replicated, as this line is not similar to Bob or BobC.

DRAQ5 enucleation assay

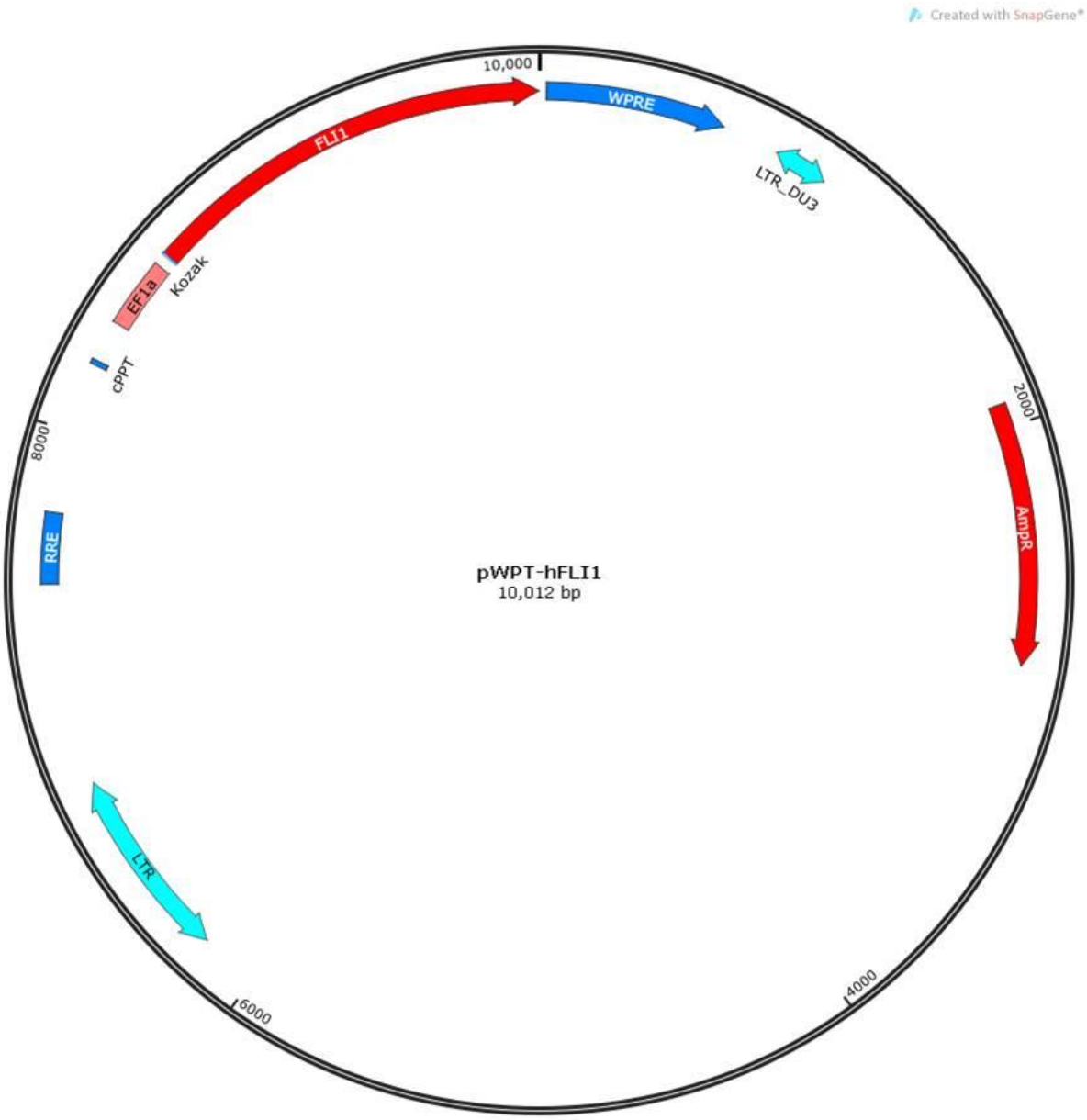
50µl cell suspension was stained with DRAQ5 1:400 (Biostatus cat # DR50050, 5mM), CD235-Pe 1:200 and made up to 100µl before incubating at 4^{o}C for 5 min. DAPI 1:1000 is added before incubating at 4^{o}C for 10 min. 400µl PBS is added directly to cell mixture and analysed immediately

by flow cytometry. DAPI negative cells are gated on CD235 and DRAQ5. DRAQ5 stains DNA of live cells, therefore cells which are DRAQ5 negative and CD235 positive are enucleated red cells.

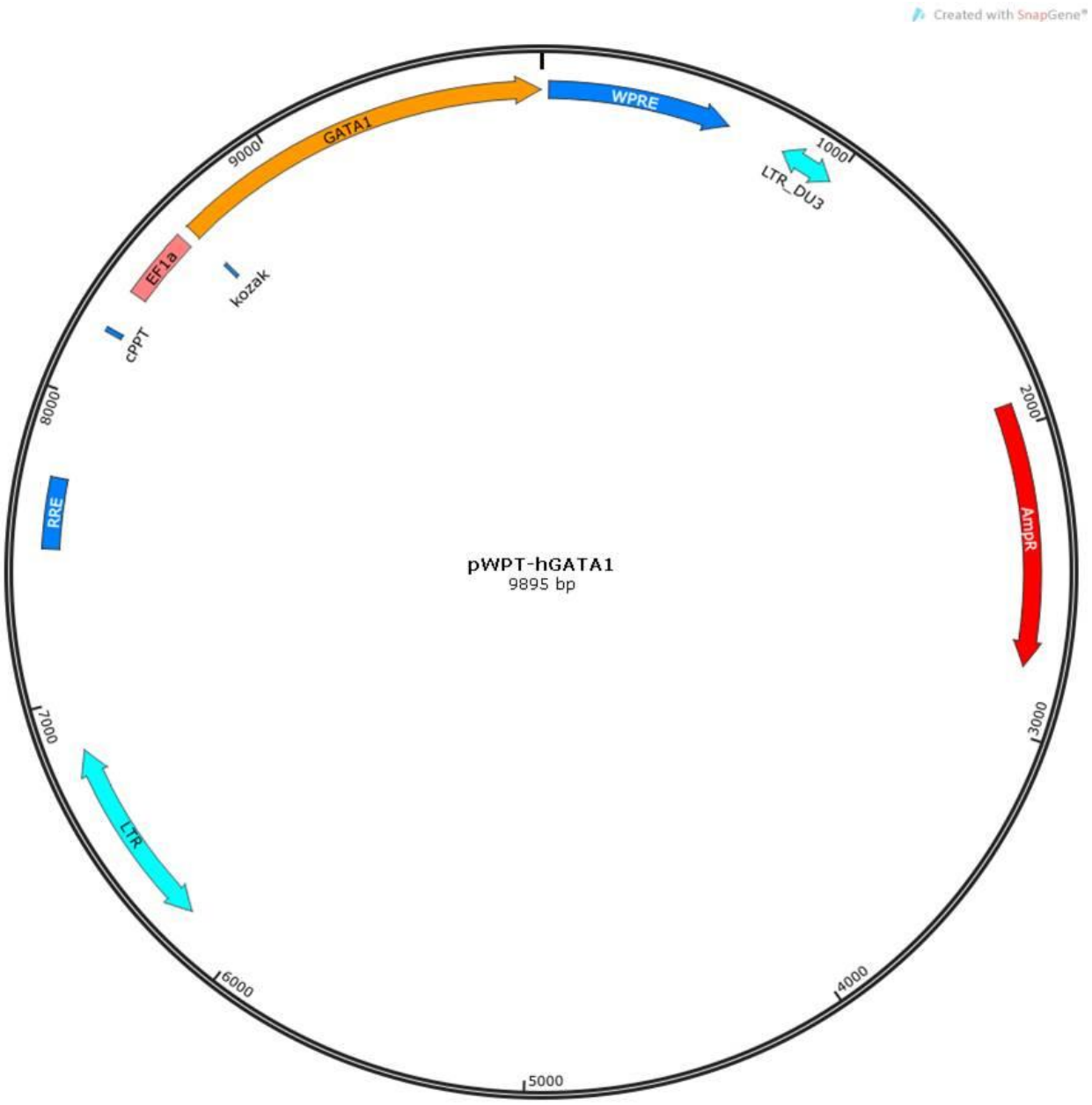
Western blot

Cells were lysed in Radioimmunoprecipitation assay (RIPA) buffer containing proteinase inhibitors (complete, Roche) and sonicated for 5 mins (cycles of 30 secs on/30 secs off) at 4^{o}C, before alpha, beta, gamma and zeta-globin and β-actin characterisation, using antibodies described in **Table 2** (Main Materials and Methods).

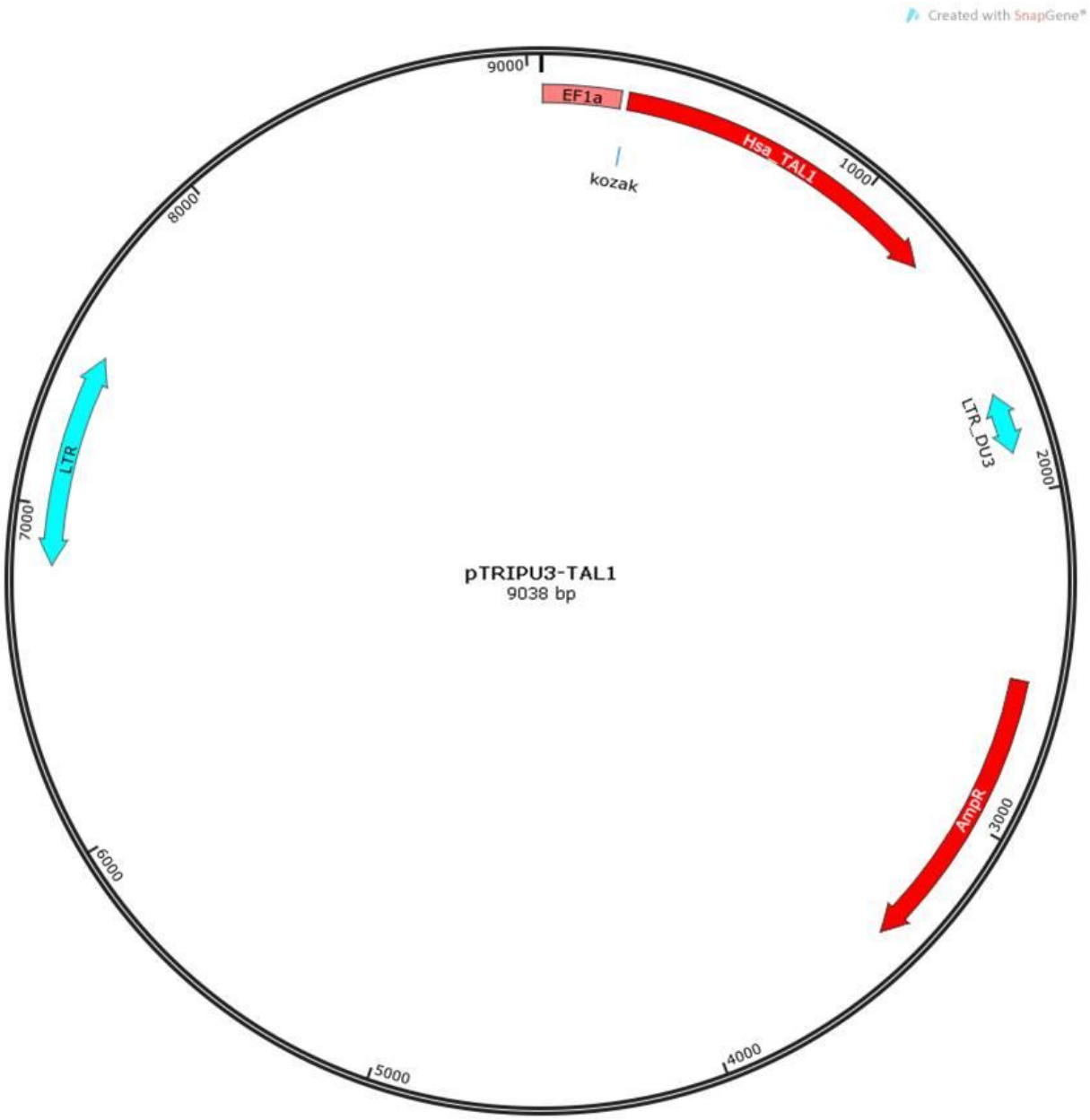
Chapter 1: Vector Maps



Map 1.1 pWPT-FLI1



Map 1.2 pWPT-GATA1



Map 1.3 pTRIP-TAL1

Results

In the results section, the following cell type definitions are used; CD41+/CD235+ cells are bi-potent progenitors, CD41-/CD235+ cells are erythroblasts and CD41+/CD235- cells are MKs. CD41a antibodies detect the Integrin-αIIb surface membrane protein, found on megakaryocytes and platelets. CD41 forms part of the glycoprotein IIb-IIIa complex, representing the most abundant surface protein found on platelets (Phillips DR, 1988). CD235a antibodies detect Glycophorin A, a major erythrocyte membrane sialoglycoprotein (Dahr W, 1987). The co-expression of CD41a and CD235a has been demonstrated on bi-potent progenitor cells differentiated from human ESCs previously (Klimchenko O, 2009). CD71, Band3 and Rhesus D (RhD) antibodies have been used to further characterise the erythroblasts produced. CD71 antibodies detect the transferrin receptor, expressed at high levels on the surface of erythroblasts during all stages of development, apart from in mature erythrocytes where it is absent (Pan BT, 1983). Band3 antibodies detect the membrane transport protein, Band 3, which is synthesised in increasing quantities throughout all stages of erythroblast differentiation, including terminal differentiation (Hanspal M, 1993). Rhesus D antibodies detect Rhesus D, found on the surface of erythroblasts and erythrocytes and is a protein used for determining blood group classifications (Mollison PL, 1993).

We have previously shown that mature MKS are produced in TPO containing media, after lentiviral transduction in iPSCs with the three TFs *GATA1*, *TAL1* and *FLI1* (Moreau et al, 2016). For the context of the work presented in this thesis, it is important to demonstrate that non-transduced (NT) iPS cells do not acquire any markers associated with FoP, when cultured in the same media conditions. **Fig 1.1** shows NT cells grown in either MK medium (TPO) or erythroblast medium (EPO), described in the Materials and Methods, at day 9 and 20. In TPO and EPO, NT cells do not express CD41a or CD235a, two markers expected to be expressed by day 9 of efficient FoP (**Fig 1.1A** and **B**). In EPO, 59% of cells express the transferrin receptor CD71 but do not express the erythroid differentiation marker, CD36. Neither do these cells express the myeloid markers CD14 or CD66c (**Fig 1.1B**). Apart from some cells expressing CD71, which is most likely the result of being cultured in the presence of the cytokine transferrin, NT cells show no evidence of having differentiated into haematopoietic cells. Furthermore, by day 20 of FoP, the viability of cells in both TPO and EPO is reduced to lower than 8% (**Fig 1.1C**), showing that NT cells do not survive in these media conditions for the duration of FoP experiments. It is due to these results, which have been replicated in other FoP experiments containing NT controls (data not shown), that we concluded NT cells do not forward program, and thus are not included as a standard control for experiments shown throughout this thesis.

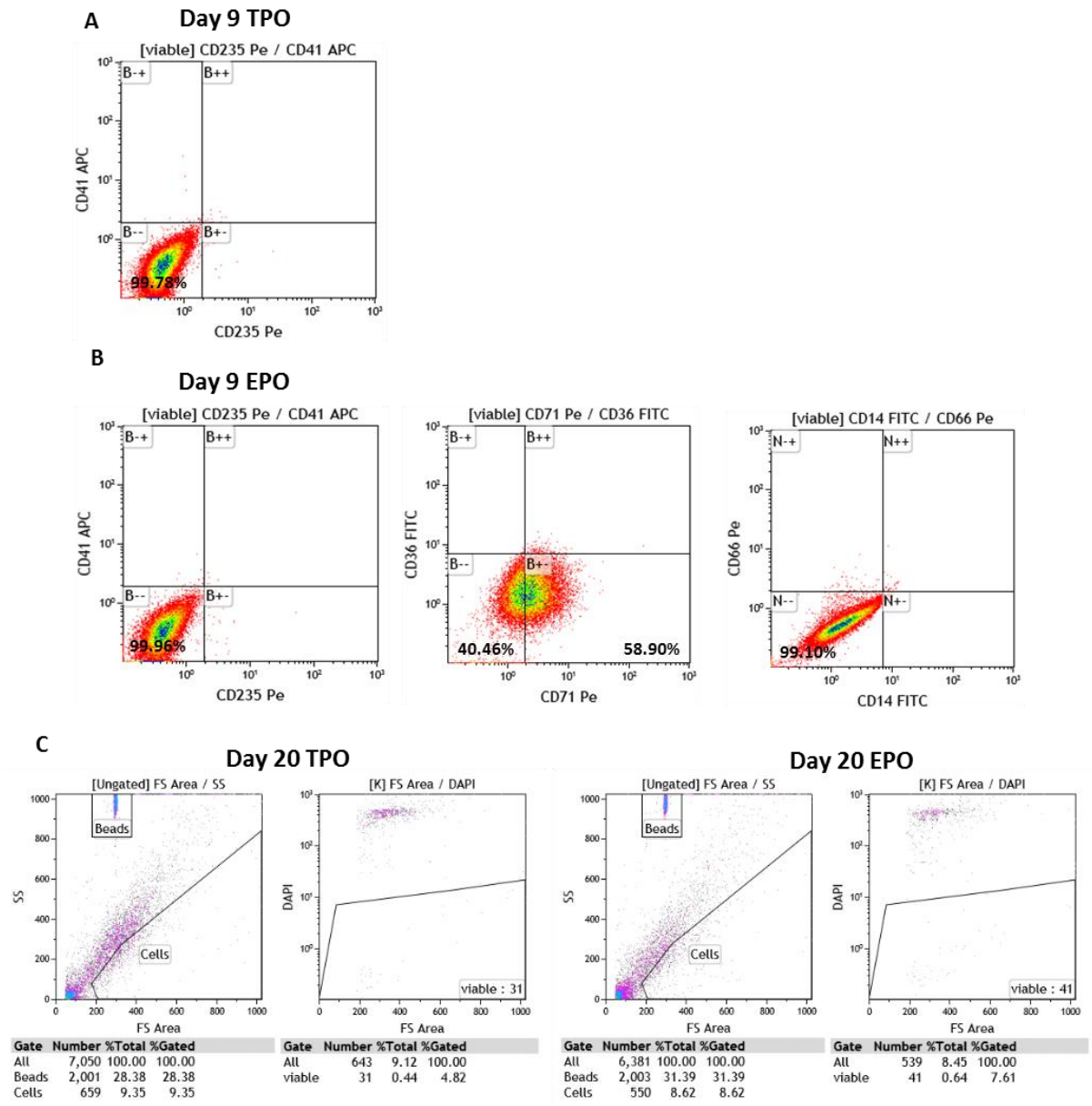


Fig 1.1 The outcome of MK and Ery-FoP on non-transduced BobC cells. The iPSC line BobC was forward programmed, without the addition of lentivirus at day 0. All other steps of the FoP protocol were performed as normal. After 2 days in mesoderm medium cells were grown in either MK medium (TPO), or erythroblast medium (EPO), for the remaining 18 days of culture. Cells were dissociated at day 9 and routine flow cytometry analysis performed. **A**) Flow cytometry dot plot of day 9 cells cultured in TPO, stained for the markers CD41a and CD235a. **B**) Flow cytometry dot plot of day 9 cells cultured in EPO, stained for the markers; CD41a and CD235a (left), CD71 and CD36 (middle), CD14 and CD66c (right). **C**) Flow cytometry plots to show viability of cells at day 20 of FoP, grown in TPO (left) and EPO (right). N=1.

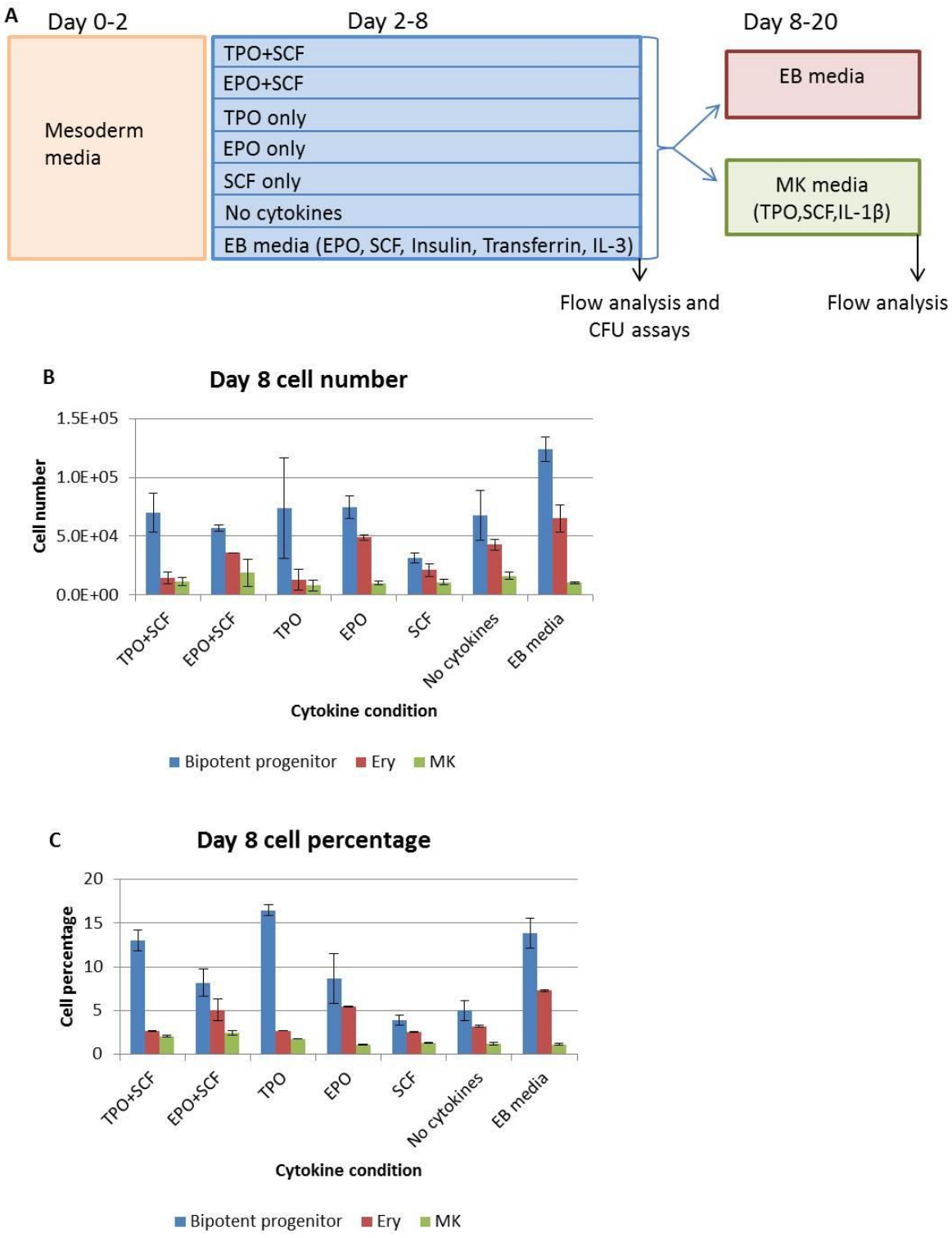


Fig 1.2 Comparing the day 8 outcome of FoP in different cytokine conditions. Bob iPSCs were transduced at day 0 with the three TFs: GATA1, TAL1 and FLI1 (MOI 20 each).FoP was performed with a 2 day mesoderm induction, before cells were cultured in different cytokine conditions until day 8. **A**) Schematic of experiment performed, detailing different media conditions tested. **B**) Day 8 average cell number from each different cytokine condition tested. **C**) Average day 8 cell

percentages of forward programmed cells. Graphs show bi-potent progenitors (CD235+/CD41+), Erythroblasts (CD235+/CD41-) and MKs (CD235-/CD41+), N=2 (technical replicates), error bars= data range.

In order to test whether liquid cultures of FoP produce bi-potent progenitors that could generate erythroblast cells, and to try and find the optimal condition for these cells, an experiment was performed which tested 7 different cytokine settings between day 2 and 8 of culture. **Fig 1.2A** shows a schematic of the experiment performed, and the day 8 cell numbers and percentage of FoP cells produced by each condition tested. At day 8, the condition which produced the highest number of bi-potent progenitors and erythroblasts was the EB media condition, which contained a combination of cytokines and growth factors clearly important for haematopoiesis. The remaining conditions produced similar cell numbers, with SCF only producing the fewest cells overall, even less than the no cytokine condition (**Fig 1.2B**). The highest purity of erythroblasts also came from the EB media condition, with overall FoP percentages being fairly low for all conditions at the stage, below 25% (**Fig 1.2C**).

CFU assays are used to evaluate the progenitor potential of haematopoietic cells. In order to assess if particular cytokine conditions lead to an enrichment of certain types of progenitors, CFU assays were performed in enriched methylcellulose on day 8 cells (**Fig 1.3**). The total number of colonies counted varied widely, with SCF only producing the fewest, 91.5 colonies on average, and TPO only producing the highest, 1029 on average. The number of CFU-E colonies was highest, while mixed and MK colony numbers were fairly similar for all conditions (**Fig 1.3A**). Flow cytometry performed on CFU colonies show a difference in distribution of CD235 and CD41 (**Fig 1.3B**). Both TPO containing conditions (TPO only and TPO+SCF) and the SCF only condition showed a greater percentage of CD41+/CD235- cells (71%, 53% and 48% respectively), than any other population. EPO+SCF and EB media showed the most even distribution of cells between the 3 populations with approximately 30% of each, while EPO only and the no cytokines condition showed the highest percentage of erythroblasts (38% and 39% respectively).

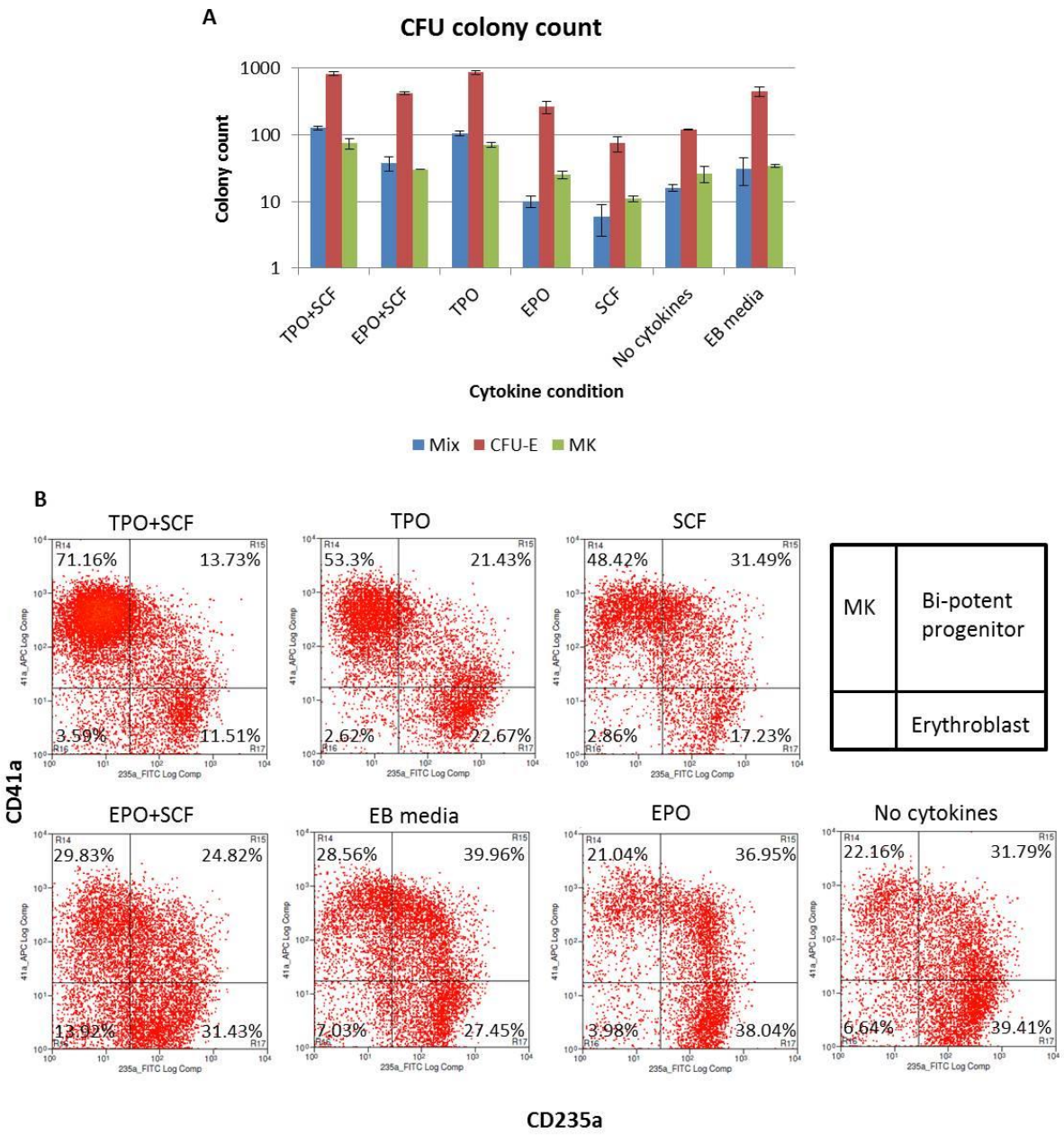


Fig 1.3 Comparing the CFU outcome of FoP in different cytokine conditions. Day 8 forward programmed Bob cells were seeded into clonogenic assays after culturing in different cytokine conditions (day 2-8, described in **Fig 1.2A**). Colonies were counted 14 days after cells were seeded. **A**) The number of mixed, CFU-E and MK colonies counted from CFU assays. **B**) Representative dot plots from CFU assays for each cytokine condition tested, showing the distribution of CD41 and CD235 cells. N=2 (technical replicates performed for CFU assays), error bars= data range.

At day 8, cells were split and maintained in either EB media, or MK media for the remainder of the experiment, before being analysed again at day 20. **Fig 1.4** shows the day 20 results for all conditions. Total cell number was highest in the condition that had been maintained in EB media

since day 2 to day 20 of FoP, and this condition produced the highest number of erythroblasts. Cell number was low in the remaining conditions, with the no cytokines condition producing the fewest cells (**Fig 1.4A**). The day 20 percentages show that despite erythroblast percentage being low previously at day 8 (approximately just 5% of each condition's cell population) EB media increased the erythroblast proportion to approximately 50% of the cell population for all conditions (**Fig 1.4B**). For cells maintained in MK media after day 8, the condition which was previously in EB media also generated the highest total cell number and highest MK cell number (**Fig 1.4C**). For all conditions, the percentage of MKs had increased from less than 2% at day 8, to above 40% by day 20 (**Fig 1.4D**).

Representative dot plots of CD41 and CD235 expression show the difference in cell populations generated at day 20, for cells which were initially maintained in EB media until day 8. There is a shift in cells towards the erythroblast population when maintained in EB media for 20 days, while a shift towards an MK population is observed when cells are switched to MK media after day 8, with an almost complete loss of erythroblast cells (**Fig 1.4E**). qPCR to detect expression of endogenous and TG FoP TFs was performed on these cells (shown in Fig 1.3E), as this condition at day 20 produced the highest number of erythroblasts in EB media, and the highest number of MKs in MK media. In the EB media condition, *FLI1* expression was the lowest for all FoP genes, with an expression value of 0.08 for endogenous and 0.8 for the transgene, relative to the housekeeping gene *MDH1*. G*ATA1* endogenous and transgene expression were highest for the three TFs in EB media. Overall expression of all three TFs was higher in the cells cultured in MK media after day 8. The highest expression was of *FLI1* transgene, with an expression value of 7.14, relative to the housekeeping gene *MDH1* (**Fig 1.4F**).

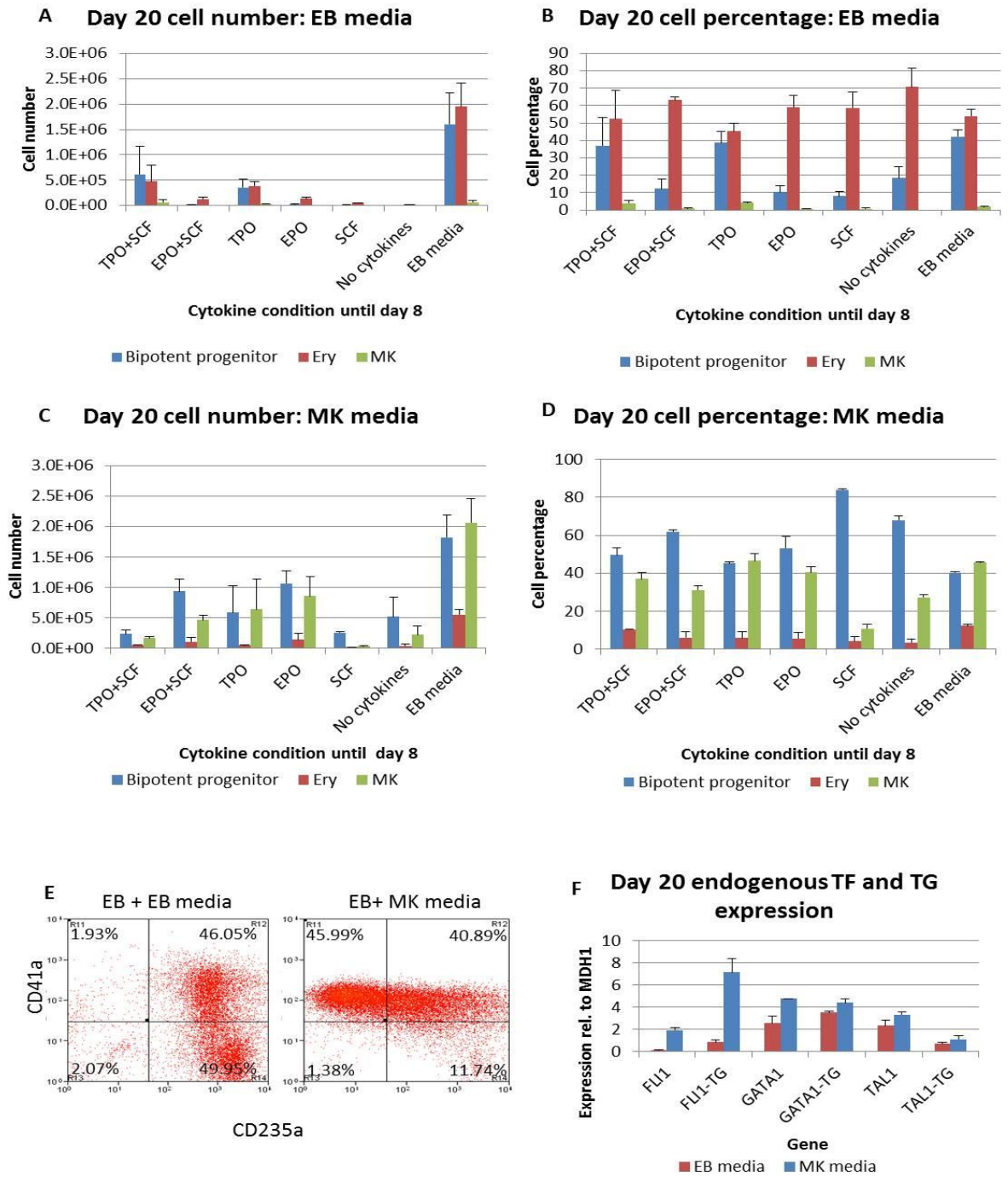


Fig 1.4 Comparing the day 20 outcome of FoP in different cytokine conditions. Bob cells cultured in different initial cytokine conditions (days 2-8) were switched to EB (EPO, IL-3, transferrin, insulin + SCF in CellGro) or MK (TPO + SCF in CellGro) media at day 8, until day 20. **A**) Day 20 cell number, after switching to EB media at day 8. **B**) Day 20 cell percentage after switching to EB media at day 8. **C**) Day 20 cell number, after switching to MK media at day 8. **D**) Day 20 cell percentage, after switching to MK media at day 8. **E**) Representative dot plots from the conditions cultured in EB media initially (day 2-8), then in either EB or MK media until day 20. **F**) qPCR data to show relative quantification (RQ) values of *FLI1*, *GATA1* and *TAL1* endogenous and transgene expression, relative

to *MDH1*, from the same conditions shown in **E**. Graphs show bi-potent progenitors (CD235+/CD41+), Erythroblasts (CD235+/CD41-) and MKs (CD235-/CD41+), N=2 (technical replicates), error bars= data range.

Fig 1.5 shows the results from further characterisation of the erythroblasts produced from the EB+EB condition at day 20. Almost no gamma or beta haemoglobin expression was detected, while alpha, epsilon and zeta haemoglobins were expressed (**Fig 1.5A**). Rhesus D expression was detected only in the CD41-/CD235+ erythroblast population of cells produced by FoP (**Fig 1.5B**). A cytospin of these cells shows a large number of small cells, with the characteristic morphology of proerythroblasts and polychromatophilic erythroblasts, as well a small number of cells with eccentric condensed nuclei, characteristic of a late normoblast (arrowed), a later stage of reticulocyte differentiation. Additionally, the cells generated in EB media produce a red pellet when centrifuged (**Fig 1.5C**). This experiment established that the EB media was best for supporting and promoting erythroblast generation, and will be referred to as the EPO condition, and the MK media shall be referred to as the TPO condition, for all further experiments.

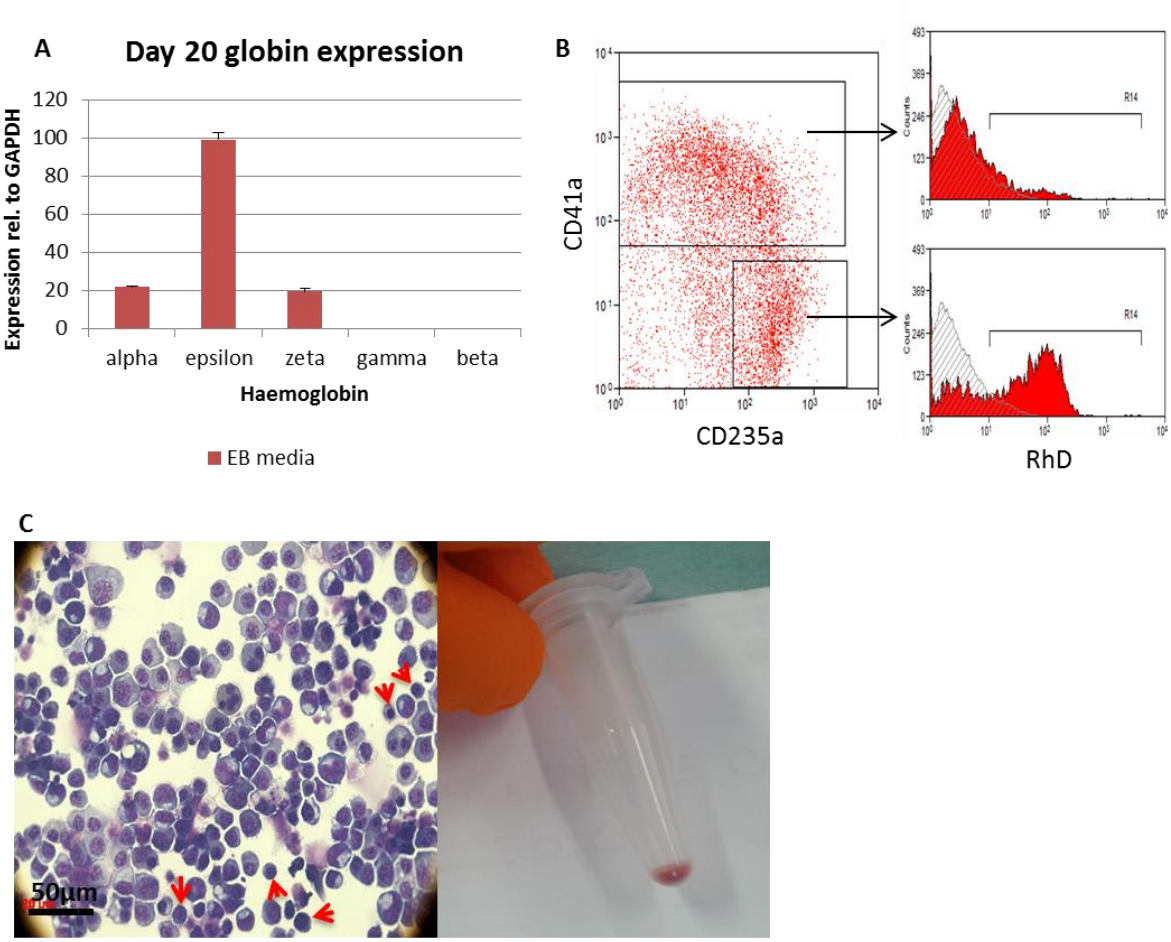


Fig 1.5 Day 20 Characterisation of CD235+ cells. Day 20 CD235+ Bob cells cultured in EB (EPO, IL-3, transferrin, insulin + SCF in CellGro) medium (day 2-20) are shown. **A**) qPCR data to show haemoglobin expression at day 20, relative to the housekeeping gene *GAPDH*. **B**) Representative dot plot CD41 and CD235 expression in day 20 forward programmed cells. Right upper panel: Rhesus D (RhD) expression in CD41+ MK and CD41+/CD235+ bi-potent progenitor cells. Right lower panel: RhD expression in CD41-/CD235+ erythroblast cells. **C**) (Left) Cytospin showing day 20 cells cultured in EB media only (day 2-20). Small cells with a large nucleus are indicative of a more mature erythrocyte (arrowed). Scale bar= 50µm. (Right) Red cell pellet produced when these cells are centrifuged. N=1.

Based on the observations that FoP can generate erythroblasts as well as MKs, we wanted to show that this strategy could be reproduced in different cell lines. **Fig 1.6** shows the day 9 results of 5 experiments, performed in either EPO or TPO media, with the starting iPSC line BobC. **Fig 1.6A** shows a schematic of how the experiment was performed. At day 9, the bi-potent progenitor cell is produced in the highest quantity in both conditions. Interestingly, a bias towards the erythroblast lineage is seen, with a higher erythroblast cell number generated in EPO compared to TPO (**Fig 1.6B**). The total cell number produced is similar, as is the total percentage of FoP cells produced in both conditions, with comparable variance (**Fig 1.5C**). Representative dot plots show how the populations

of FoP cells differ between the two conditions, with a shift towards CD41+^{low}/CD235+ and CD41-/CD235+ cells in EPO, and a shift towards CD41+/CD235+^{low} and CD41+/CD235- cells in TPO (**Fig 1.5D**).

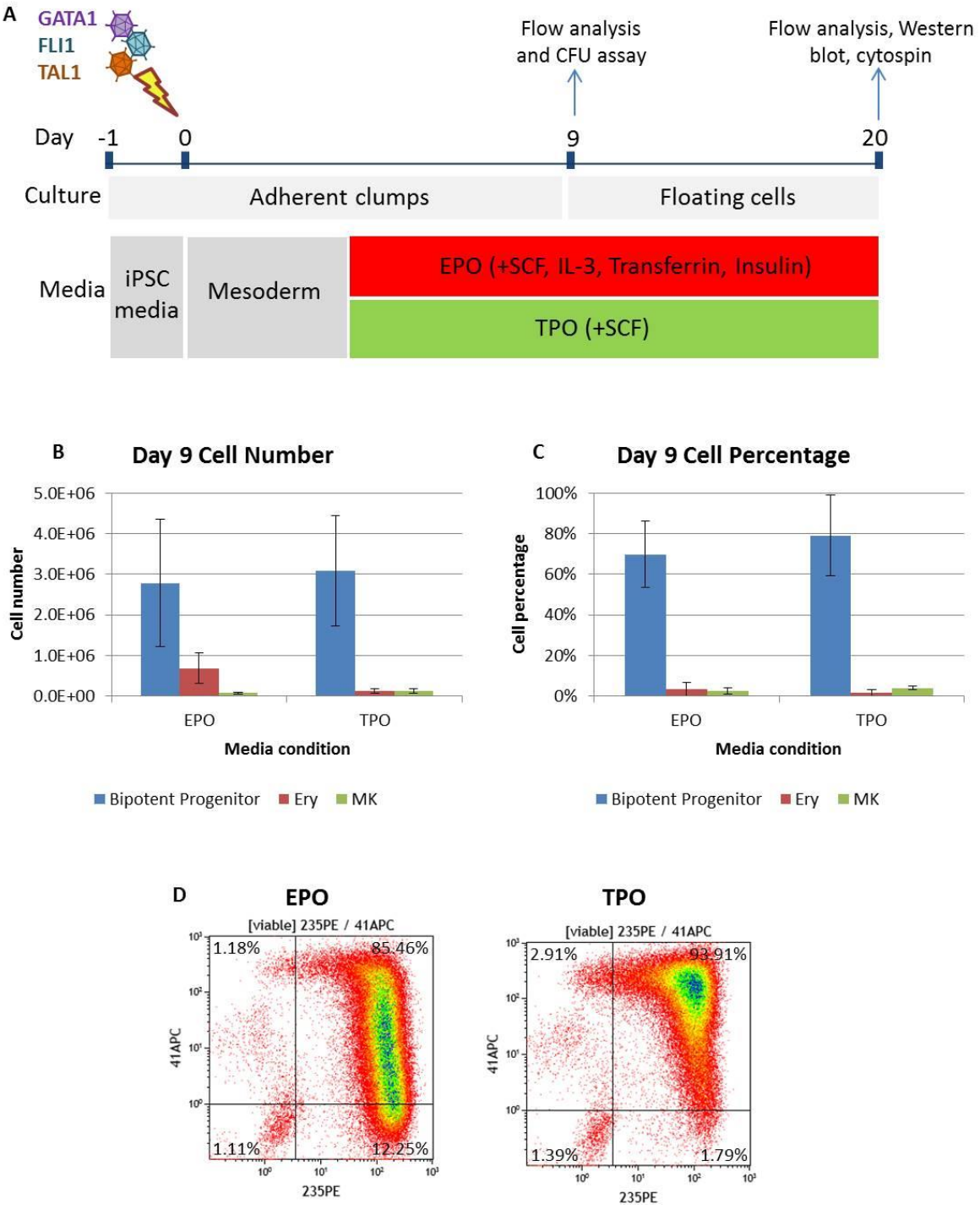


Fig 1.6 Day 9 outcome of FoP in EPO and TPO. BobC iPSCs were transduced at day 0 with GATA1, TAL1 and FLI1 lentivirus (MOI 20/ TF). Cells were maintained for 2 days in mesoderm media before being cultured in TPO medium (containing TPO and SCF to promote MK differentiation), or EPO

medium (containing EPO, IL-3, Insulin, Transferrin and SCF to promote erythroid differentiation) for the remainder of the experiment. **A**) Schematic of experiment performed, showing medium conditions tested and details of analysis performed. **B**) EPO and TPO day 9 cell numbers showing FoP populations only. **C**) EPO and TPO day 9 cell percentages of FoP populations. **D**) Representative dot plots to show the day 9 distributions of CD41 and CD235 expression in cells cultured in EPO and TPO. Graphs show bi-potent progenitors (CD235+/CD41+), Erythroblasts (CD235+/CD41-) and MKs (CD235-/CD41+), N=5 (biological replicates), error bars= standard deviation.

In order to characterise the cells generated by our strategy further, day 9 cells were seeded into enriched methylcellulose to perform CFU assays. **Fig 1.7** shows how the resulting CFU colonies were classified, and the colony outcome for this experiment. Four types of colonies were counted; mixed, CFU-E, MK and MK progenitor colonies. Mixed colonies are the largest in size and comprise of a mixture of both MK and erythroblast cells. CFU-E colonies are the smallest in size and are comprised of erythroblasts. MK colonies, comprising of MK cells, were divided into MK and MK progenitor colonies, based on their size. MK progenitor colonies are visibly larger and have a higher cell density, while MK colonies are smaller, with slightly more spread out and larger cells, representative of more mature MKs. Representative light microscope images are shown for each colony type (**Fig 1.7A**). Representative dot plots for CD41 and CD235 are shown for all colony types. The average percentages of the different FoP cell types from four individual colonies is shown (flow cytometry performed by Dr Moreau and Dr Jose Ballester-Beltran). This shows that mixed colonies are truly a mixture of CD41 single positive and CD235 single positive cells, thus, contains both MKs and erythroblasts. CFU-E colonies are mostly made up of CD235+ single erythroblast cells, with very few CD41+ MK cells. MK and MK progenitor colonies are made up of predominantly CD41 single positive MK cells, with fewer cells co-expressing CD235 and very few CD235 single positive cells. No other colony types were found.

The CFU colony number outcome of cells seeded at day 9 shows that a higher proportion of CFU-E colonies were generated from cells which had been cultured in EPO, while a higher number of MK and MK progenitor colonies were generated from cells in TPO (**Fig 1.7B**). The total colony number counted overall was similar for both conditions.

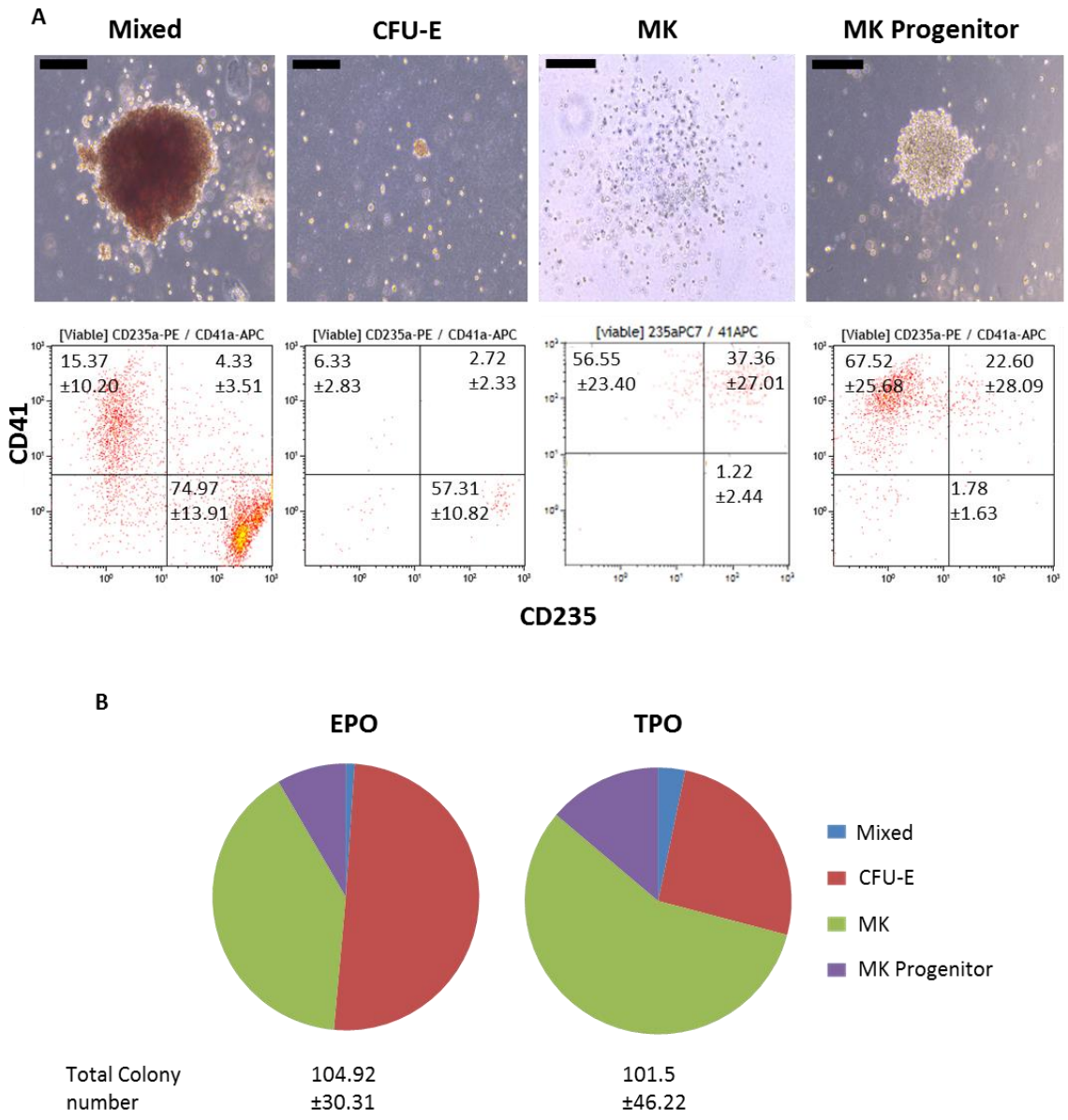


Fig 1.7 Progenitor potential of FoP in EPO and TPO. CFU clonogenic assays were performed on day 9 BobC FoP cells that were cultured (day 2-9) in EPO or TPO conditions. Colonies were counted 14 days after cells were seeded. All counts were conducted blind, with at least 2 people performing counts. **A**) Morphology of mixed, CFU-E, MK and MK progenitor colonies and representative flow plots to show CD41 and CD235 expression in all colony types. Flow cytometry was performed on individual colonies and average percentages shown from at least 5 individual colonies, ± standard deviation. **B**) CFU colony counts and distribution for EPO and TPO conditions. Scale bars= 200µm. N=4 (Day 9 cells seeded in CFU assays were produced by 4 biological replicate experiments. Each CFU assay was performed in duplicate (technical replicates). Technical replicate results were averaged, before overall average and ± standard deviation worked out from the average of 4 biological replicates).

The day 20 erythroblast and MK cell outcome is shown in **Fig 1.8**. By day 20 the EPO condition sustained a higher number of erythroblasts, while TPO sustained a higher number of MKs (**Fig 1.8A** and **Fig 1.8B**). The percentage of bi-potent progenitors remained similar at day 20 for both media conditions, with the largest difference between the erythroblast and MK percentages, which followed the same pattern as cell number (**Fig 1.8C**). Day 20 flow cytometry analyses show the bias in cell populations depending on culture conditions, towards erythroblasts in EPO and MKs in TPO. Cytospins provide evidence that EPO produces small cells, some of which have eccentric condensed nuclei, characteristic of a late normoblast, as well as proerythroblast and polychromatophilic erythroblasts, and evidence of an enucleated erythrocyte. TPO produces larger cells, morphologically similar to megakaryoblasts, with a number of polynucleated cells (arrowed), characteristic of more mature MKs. Cells pellets show that EPO produces red cells, while TPO produces predominantly white cells (**Fig 1.8D**).

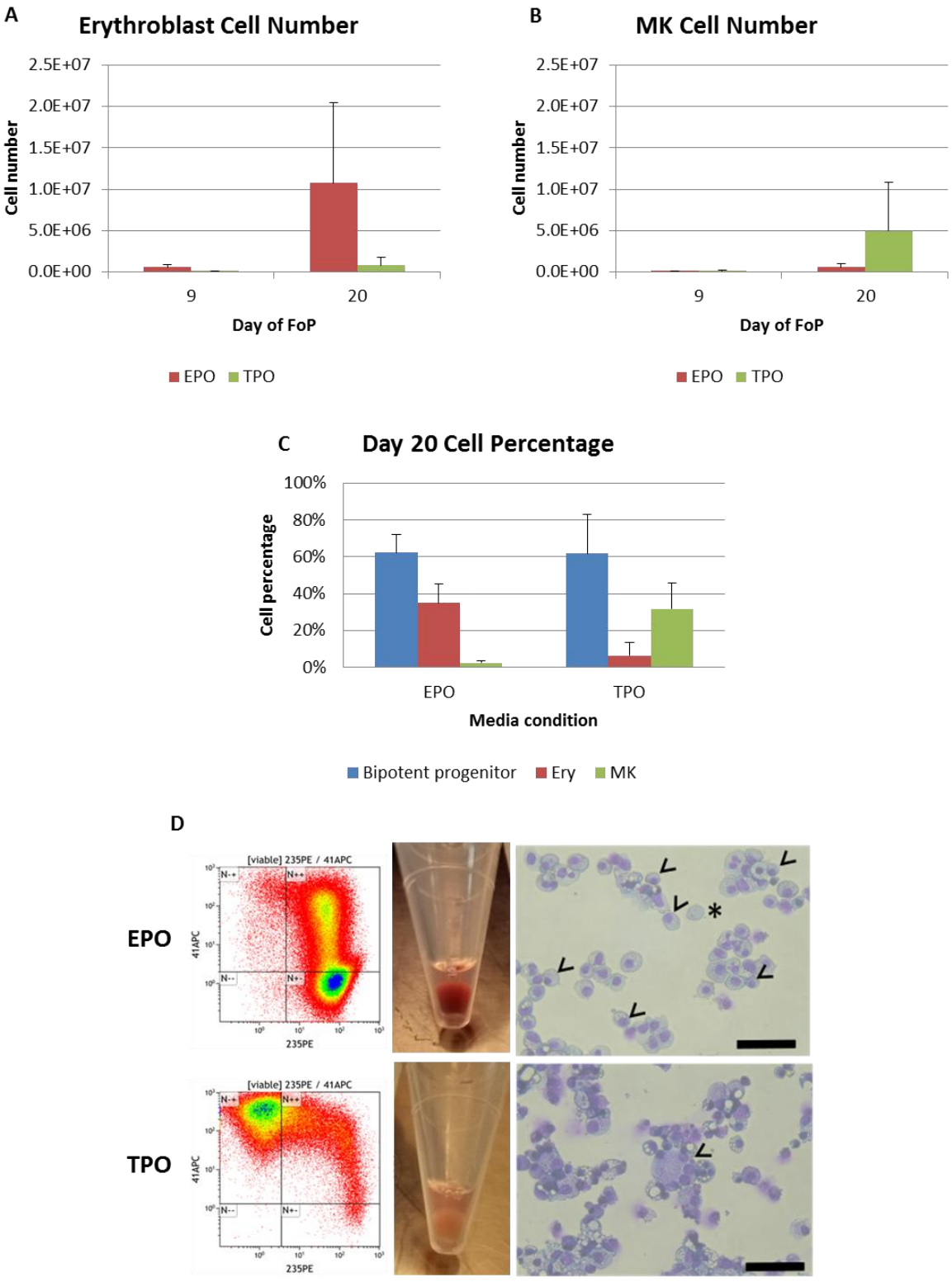


Fig 1.8 Day 20 outcome of FoP in EPO and TPO. The experiment described for FoP of BobC cells (**Fig 1.5**) was continued until day 20, when further analysis was performed. **A**) Day 9 and 20 erythroblast (CD41-/CD235+) cell number. **B**) Day 9 and 20 MK (CD41+/CD235-) cell number. **C**) Day 20 percentage of bi-potent progenitors (CD41+/CD235+), erythroblasts and MKs (as previously described). **D**) Left panel: Representative dot plots for CD41 and CD235 staining. Middle panel: Cell

pellets were red in EPO and white in TPO. Right panel: Cytospins show normoblasts (arrowed) and enucleated erythrocyte (starred) in EPO (top), and polynucleated MK (arrowed) in TPO (bottom). Scale bar = 50µm. N=4 (biological replicates), error bars= standard deviation.

Characterisation of CD235+ erythroblast cells from the EPO condition shows the majority of these cell stain positively for other erythroid markers, BAND3 and CD71 (**Fig 1.9A**). Western blot for haemoglobins show that these cells produce epsilon, alpha and gamma embryonic globins, but that the adult beta globin is not expressed. Cord blood derived erythroblasts, which have an adult phenotype, as shown for comparison (**Fig 1.9B**, Western blot data produced by Dr Ballester-Beltran).

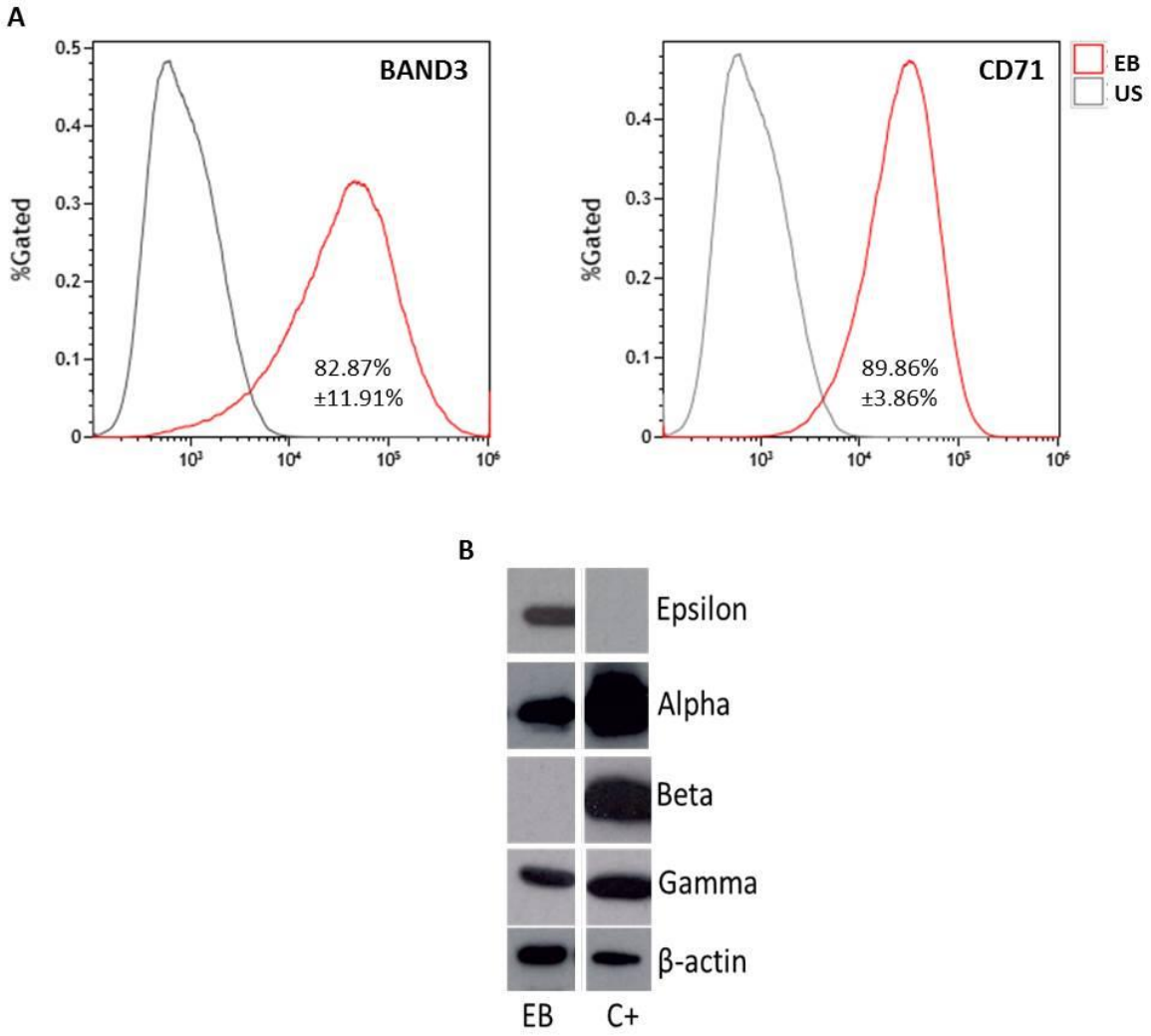


Fig 1.9 Day 20 characterisation of CD235+ cells grown in EPO. Data from day 20 CD235+ BobC cells forward programmed in EPO media (EPO, SCF, IL-3, insulin and transferrin in CellGro) from day 2-20, is shown. **A**) Band3 and CD71 expression in EPO grown cells, compared to unstained controls (N<3 biological replicates, ± standard deviation). **B**) Western blot of epsilon, alpha, beta and gamma haemoglobins, with the positive control β-actin. EB= EPO media, US= unstained, C+= day 16 cord blood derived erythroblasts.

A second cell line, FFDK, was used to perform the same experiments to see if these results would be reproduced in a different iPSC line. **Fig 1.10** shows the day 9 results for FFDK. A higher proportion of MK cells were seen at this stage, compared to BobC, but the same pattern was observed, that EPO produces a higher number of erythroblasts than TPO (**Fig 1.10A**). Day 9 cell percentages show a higher proportion of erythroblasts in EPO and MKs in TPO at this stage, which was not seen in BobC (**Fig 1.10B**). Day 20 cell numbers show that only EPO supports erythroblast proliferation (**Fig 1.10C**). While EPO supported MK proliferation in FFDK, it is not as great by day 20 as in TPO (**Fig 1.10D**). Cell number overall is lower than was seen for BobC. Representative dot plots at day 9 show that cells are more committed to either the erythroblast or MK lineages, compared to BobC at the same stage. Day 20 dot plots show distinct populations for the two media conditions, with very few bi-potent progenitors at this stage, unlike in BobC (**Fig 1.10E**). Further characterisation of the CD235+ erythroblast population from EPO shows that the majority of these cells stain positive for both BAND3 and CD71 (**Fig 1.10F**).

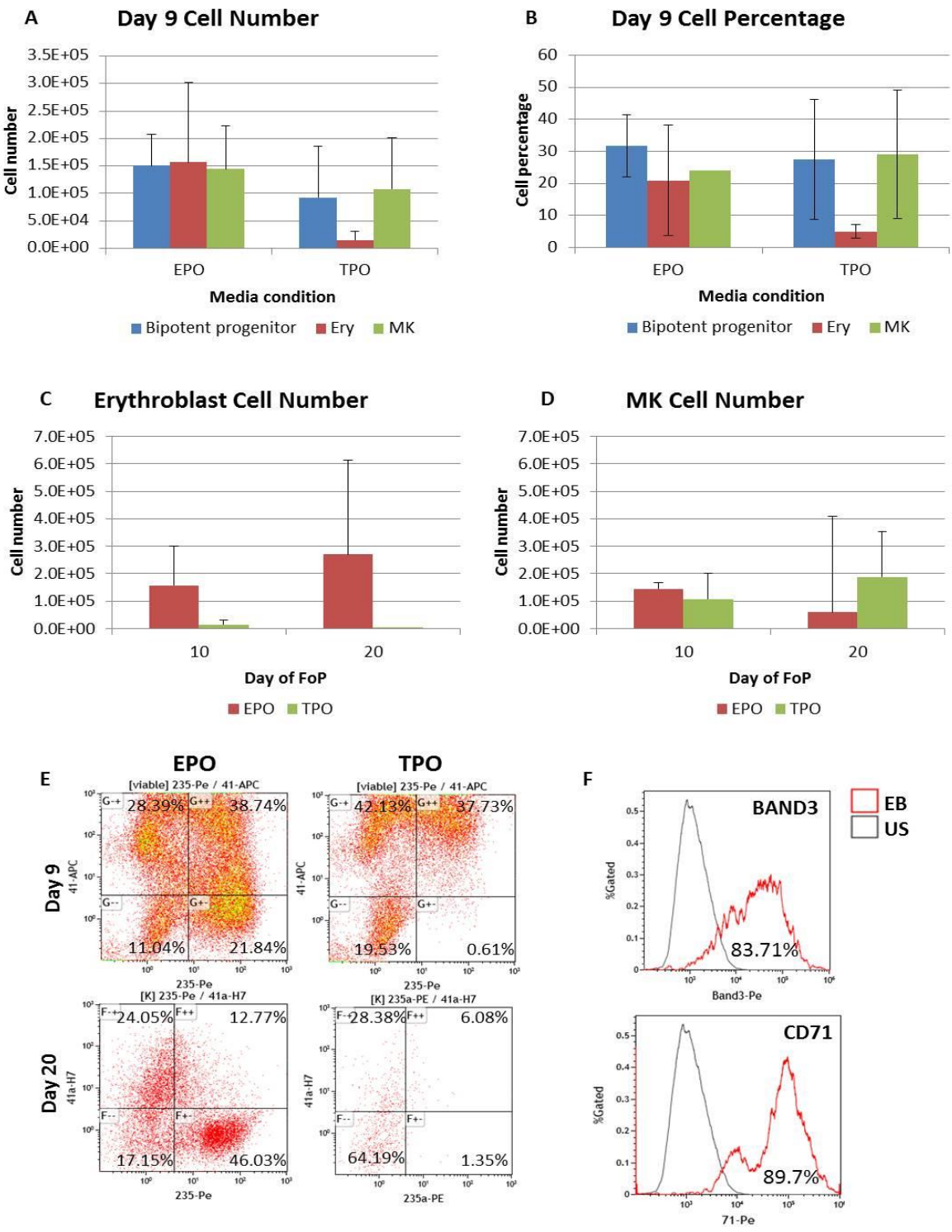


Fig 1.10 Results of FoP in TPO and EPO in a second iPSC line. FFDK, was transduced with GATA1, TAL1 and FLI1 (MOI 20/ TF) and forward programmed in EPO or TPO conditions from day 2-20. Data shown is from cells collected at days 9 and 20 of FoP. **A**) Day 9 cell number in EPO and TPO. **B**) Day 9 cell percentage. **C**) Erythroblast cell number in EPO and TPO at days 9 and 20. **D**) MK cell number in EPO and TPO at days 9 and 20. **E**) Representative dot plots from EPO and TPO cells, showing the distribution of CD41 and CD235 cells, at day 9 (top) and 20 (bottom). **F**) BAND3 and CD71 expression

in CD235+ cells, from the EPO condition only, compared to unstained control cells. Graphs show bi-potent progenitors (CD235+/CD41+), Erythroblasts (CD235+/CD41-) and MKs (CD235-/CD41+), US= unstained, N=3 (biological replicates), error bars= standard deviation.

Discussion

This chapter describes, for the first time, the *in vitro* generation and bifurcation of bi-potent progenitor cells produced by forward programming of human iPSCs. These progenitors are capable of generating cells of both the erythroid and MK lineages, showing a bias for lineage commitment in different cytokine conditions as early as day 8 of programming. Erythroblasts produced by FoP are CD235 positive, express other common markers of erythroid differentiation, and produce embryonic and fetal globins.

Forward Programming Allows Bifurcation into Megakaryocytes and Erythroblasts

The first experiment described where multiple cytokine conditions were tested between day 2 and 8 of FoP in Bob cells, resulted in relatively similar day 8 cell number outcomes (**Fig 1.1B**). Day 20 bi-potent progenitor, erythroblast and MK cell numbers were highest when cells were maintained initially (days 2-8) in EB medium, containing EPO, SCF, IL-3, transferrin and insulin (**Fig1.3A** and **Fig1.3C**). Surprisingly, the TPO+SCF condition, which is currently used as routine MK-FoP media, did not produce many MK cells by day 20. This suggests that the initial addition of IL-3, transferrin and insulin throughout days 2-8 , are responsible for driving a higher rate of cell proliferation, and/or provide the most protection from apoptosis in liquid culture, compared to the conditions with fewer cytokines. Transferrin is often used in growth media as it regulates the transport and uptake of iron into cells, an essential element for cell growth and metabolism (Zhang D, 2012). Insulin has been shown to be particularly important in confluent cell cultures, as it encourages the up-take and efficient use of nutrients, increasing cell size and metabolic activity (Griffiths, 1970). IL-3 has been shown to reduce the length of the G1 cell-cycle phase, resulting in rapid cell proliferation and has long been associated with supporting the proliferation and development of haematopoietic precursors in culture (Vander Heiden MG, 2001; Donahue RE, 1988).

The difference between endogenous and TG expression of the three FoP TFs on the whole population of cells present at day 20 in EB+EB and EB+MK is striking (**Fig 1.3F**). After 20 days, a large number of erythroblasts and very few MK cells were present in the EB+EB condition (**Fig 1.3A**), coinciding with the very low expression levels detected for total *FLI1* (endogenous and TG). *FLI1* expression has previously been shown to be downregulated during EPO-induced erythroblast differentiation (Tamir A, 1999). Conversely, high expression levels were detected for total *FLI1* in cells from the EB+MK condition, for which a high number of MKs and low number of erythroblasts were present at day 20 (**Fig 1.3C**). FLI1 has also been shown to be highly expressed in MK progenitors, and further induced during MK differentiation (Okada Y, 2011). Total *GATA1* and *TAL1*

expression was lower in EB+EB cells, but expressed at comparative ratios as in cells from the EB+MK condition. This data suggests that *FLI1* TG is likely to be the key TF used in FoP, responsible for determining an early cell fate decision that is responsible for directing cells towards either MK or erythroblast lineage commitment. This data highlights the importance of the overexpression of exogenous TFs in FoP for switching on their endogenous gene counterparts and demonstrates how their expression, along with the expression of downstream genes which they control, can direct cell fate.

The colony number outcome of a clonogenic assay performed on day 8 cells showed a difference in the progenitor potential for the different cytokine conditions tested (**Fig 1.2A**). The number of colonies produced was lowest in the condition with the lowest percentage of bi-potent progenitors present at day 8, in the SCF only condition (**Fig 1.1C**). The highest colony number was in the TPO only condition, which had the highest progenitor percentage at day 8, suggesting the day 8 progenitor outcome is a good indicator for colony potential. This result also indicates that the CD41+/CD235+ population contains the bi-potent progenitors produced during FoP, as colony number did not correlate to the percentage of either erythroblasts or MKs at day 8. The conditions TPO+SCF and TPO only, produced the highest numbers of all three colony types recorded (mixed, CFU-E and MK), suggesting that TPO alone may play a major role in maintaining or supporting the bi-potent progenitor population during the initial stages of FoP. TPO has been long been known to act early in haematopoiesis, regulating haematopoietic stem cell expansion, as well as the development and differentiation of MKs (Solar GP, 1998).

When the bulk CFU colonies were analysed for CD41 and CD235 expression, an obvious bias towards the MK population was observed for colonies arising from both TPO containing conditions, while the EPO containing conditions showed a shift towards the erythroblast population (**Fig 1.2B**). This data corroborates previous work demonstrating murine MEPs respond to TPO by inducing MK differentiation and to EPO by producing erythrocytes (Ng AP, 2012). This result suggests that FoP progenitor cells cultured in TPO-containing medium until day 8 are already primed for MK differentiation while culture in the presence of EPO primes them for erythroblast differentiation. Interestingly, the condition with the greatest shift towards the CD235 single positive population was the no cytokines condition, suggesting that in the absence of differentiation stimulus, bi-potent progenitors preferentially differentiate along the erythrocyte lineage. This suggests that erythropoiesis may represent a ‘default’ cell fate decision for FoP cells.

The day 20 outcome was unexpectedly poor for MKs in the initial TPO+SCF condition (**Fig 1.3C**), our standard MK culture condition. This may be due to poor transduction efficiency, which may explain

low cell number overall at day 8. A master mix containing the three FoP lentiviral vectors was added to all conditions, and the higher cell number resulting from the initial EB condition may be due to the additional cytokines contained in this condition, such as increased cell survival mentioned previously. It could be possible that the batch of TPO used for this experiment was sub-optimal. However, since MK production was observed in the EB+MK condition at day 20, this seems less likely.

Overall, this experiment provided evidence that FoP does generate bi-potent progenitors, as only CFU-E, MK or mixed colonies are observed in CFU assays of FoP cells cultured in a number of cytokine conditions. It also suggests that initial media conditions can affect the lineage cells will preferentially differentiate along, with TPO supporting MK-FoP and EPO supporting Ery-FoP. It demonstrates that the cytokines included in EB media best support the proliferation and differentiation of erythroblasts, and this is now our standard medium for differentiating FoP-erythroblasts. Despite a potential problem with transduction efficiency, the data generated from this experiment shows that FoP is a very selective protocol. The conditions tested indicate that cytokines are important for providing a permissive environment for the survival of cells which have received the optimal mix of TGs. TG expression, particularly of *FLI1*, enforces cell identity in these cells.

Bifurcation Replicated in Different iPSC Lines

The next series of experiments performed confirmed in additional iPSC lines, BobC and FFDK, that EPO media (EPO, SCF, IL-3, transferrin and insulin) best supports the generation of erythroblasts, and TPO media (TPO and SCF) best supports the generation of MKs produced by FoP. In the cell line BobC, the day 9 cell number and percentages was very similar in EPO and TPO conditions (**Fig 1.5B** and **Fig 1.5C**). Overall, cell percentages and numbers were higher in this experiment, compared to the first (**Fig 1.1**), showing improved FoP efficiency. This is likely due to performing all experiments shown in **Fig 1.5** with commercially produced rLVs, which were produced at higher titres than the ones produced in the lab and used for the previous experiment.

In clonogenic assays, the day 9 cells produced gave rise to a higher number of CFU-E colonies when cultured in EPO and a higher number of MK and MK progenitor colonies when cultured in TPO (**Fig 1.6B**). What cannot be determined, based on the data presented here, is whether the increase in CFU-E colonies from EPO cells is due to there being a higher number of erythroblast cells present at day 8 in EPO, or whether this is due to a bias at the progenitor population for erythroblast differentiation. We believe that the progenitor cells reside within the CD41+/CD235+ population, as this population is most abundant at early time-points and reduces over time, as does clonogenic

potential (data not shown). Additionally, the vast majority of cells at day 9 were CD41+/CD235+, with few lineage restricted cells overall (**Fig 1.5**), suggesting this population contains the bi-potent progenitor cells. However, to formally prove this, the CD41+/CD235+ population of cells should be sorted, from EPO and TPO, to repeat CFU assays. If the results are the same as those presented in **Fig 1.6B**, this would show that some progenitors are already biased to one lineage over the other by day 9 of FoP. This seems most likely, since the day 9 CD41+/CD235+ population looks different in EPO and TPO by flow cytometry (**Fig 1.5D**), with the EPO cells shifted towards a CD41 low population, while in TPO the population remains CD41 high. In this case, it is likely that the progenitor population contains a mixture of both bi-potent and more committed progenitor cells, which would explain the results seen in **Fig 1.6B**. This proposed work is currently on-going, in order to formally show where the bi-potent progenitor population resides.

Characterisation of Erythroblasts

The day 20 cell number of erythroblasts was highest in EPO, with very few MKs, while the reverse trend was observed for TPO (**Fig 1.7A** and **Fig 1.7B**). At day 20, the appearance of normoblasts and a red cell pellet in EPO showed erythrocyte differentiation, while the appearance of poly-nucleated megakaryocytes and a white pellet show MK differentiation in TPO (**Fig 1.7D**). The MKs generated by the MK-FoP have been well characterised and the functionality of the platelets derived from them demonstrated (Moreau T, 2016). We therefore concentrated on the characterisation of the erythroblasts produced by Ery-FoP in EPO. Analysis of expression of other erythroblast markers, such as BAND3 and CD71, as well as haemoglobin protein expression (**Fig 1.8**) confirmed the nature of the cells obtained with our protocol.

CD71 expression is detected in erythroblasts during all stages of differentiation before it finally disappears in terminally differentiated erythrocytes (Pan BT, 1983). Since a high proportion of the CD235+ cells produced at day 20 were also positive for CD71 this confirmed that Ery-FoP produces erythroblasts and indicated that very few, or no, mature erythrocytes were produced. The erythroblasts produced by FoP are not adult in phenotype, as they lack the adult beta-globin. No enucleated reticulocytes were observed by cytospin and flow cytometry using the DRAQ5 nuclear stain confirmed this, with single figure percentages found to be enucleated (data not shown). The phenotype of Ery-FoP cells, lacking adult globin, is a similar feature to other PSC derived erythroblasts (Lapillonne H, 2010 and Olivier EN, 2016). Importantly, these findings were corroborated in an additional iPSC line, FFDK (**Fig 1.9**). Few existing protocols (mentioned in the Introduction of this chapter) are able to produce enucleated cells, and most retain a embryonic or foetal phenotype, so the cells produced by Ery-FoP are quite comparable to other erythroblasts

produced *in vitro*. However, like MK-FoP, Ery-FoP uses minimal cytokines in comparison to other protocols, is GMP-compatible and produces a relatively high number of cells in a short period of time, which are all beneficial features of this protocol.

These results show that FoP generates a bi-potent progenitor cell population which is capable of producing an enriched population of erythroblasts in EPO, or MKs in TPO, in just 20 days. This represents an exciting future avenue of research, using FoP to further explore the basic biology of the bi-potent progenitor and to elucidate the mechanisms controlling cell-fate decision in these cells *in vitro*. The number of erythroblasts produced per starting PSC is not as high as others have reported (Olivier EN, 2016), but is good despite Ery-FoP using minimal cytokines, compared to existing protocols. Due to poor enucleation and lack of adult-phenotype switching, the Ery-FoP protocol presented here requires further optimisation, in order to represent a viable alternative source of erythrocytes for use in transfusion medicine. For example, steps to improve cell yield could be achieved by trying different media conditions to those tested here. Maturation could be encouraged by transducing cells with additional TFs, such as Krüppel-like factor 1 (*KLF1)*, which is known for its role in terminal erythroid differentiaiton and importantly, for activating expression of the β-globin gene (Tallack MR, 2012). Adult swithcing of globins is not an absolute requirement for transfusable erythroid cells and in certain cases may be undesierable. The reactivation of foetal-globin in patients with sickle cell anaemia and β thalassaemia, where there is an inherited deficiency in β-globin production, is being explored as a disease management approach, as patients with unusually high levels of foetal-globin present with milder illness (Breda L, 2016). However, overall maturation would need to be encouraged to produce enucleated erythrocytes that would be more suitable for use in clinical transfusion.

Results

Chapter 2

Testing Different Polycistronic Cassettes

Introduction

We have previously shown that lentiviral Forward Programming (FoP) efficiently produces a highly pure population of mature MKs and generates a large numbers of MKs per starting iPSC cell (Moreau, 2016). The protocol is performed in a chemically defined serum-free, minimal-cytokine setting and the expanding long-term cultures can be cryopreserved to generate cell bank of FoP-MKs. Thanks to the ease and efficiency of FoP over existing MK differentiation methods, our protocol is now being used by a growing number of researchers in Cambridge, as well as internationally, to study the biology of megakaryocytes and platelets. As the ultimate aim of our work is to produce platelets *in vitro* for transfusion medicine, it is important that the protocol can practically transition from the bench-top, to the large-scale clinical manufacture of iPSC-derived platelets.

The first consideration of scaling-up FoP is cost: if large-scale production cannot be done inexpensively then the resulting platelets could be too expensive for an organisation such as the NHS to administer, and would not be a competitive option over the current supply of donor platelets (at a cost of £193 per transfusion unit in 2017). Since the FoP protocol currently relies on just a handful of cytokines kept at low concentrations over the duration of culture, the requirement of expensive growth factors has already been minimised. One of the most expensive steps in the current FoP arises from the need to lentivirally transduce iPSCs with three different transcription factors (TFs). Technically recombinant lentiviral vector (rLV) production is a costly and lengthy procedure, further complicated by batch variations that introduce experimental discrepancies. Of the three genes overexpressed to drive FoP, FLI1 rLV have been consistently difficult to produce with high titres, despite efforts to improve production internally and by specialised platforms (Vectalys). One proposed hypothesis to explain this is that FLI1 RNA conformation may impair rLV packaging or the downstream processes of transduction (reverse transcription, nucleus import, integration), resulting in low titres.

The second consideration of scaling-up is reproducibility: the protocol needs to be able to robustly and routinely produce high volumes of mature MKs. Currently FoP variability occurs most due to inter-line variability between iPSC lines, but even in core lines routinely used to FoP, intra-line variability is an issue. The lentiviral transduction step is the most likely cause of most intra-line variability, due to different transduction efficiencies between experiments, the requirement of all three TGs to enter the same cell and random vector integration leading to unique polyclonal patterns, but also due to batch variability as mentioned previously.

A third consideration is ensuring all FoP reagents are good manufacturing practice (GMP) grade for use in a clinical setting. We would need to source GMP grade cytokines, as well as rLVs. While GMP

grade rLVs can be produced, (Ausubel, 2012), their use as theraputic agents in the clinic is still viewed with caution, notably after a series of serious side effects in gene therapy trials using related retroviral vectors (Thrasher AJ, 2006). Indeed, due to random insertions biased to intragenic genomic regions of host cells, these vectors are assocaited with a risk of insertional mutagenesis including activation of proto-oncogenes or inactivation of tumour suppressors (Amado, 1999). While this is theoretically not relevant in the context of *in vitro* production of short-lived anucleate platelets, the exposure to rLVs during the FoP process would still require the implementation of specific safety controls, especially since the end product is destined to general transfusion purpose.

Reducing Lentiviral Requirement: Generating a Polycistronic Forward Programming Vector

As outlined, one of the major drawbacks of the current FoP technology, which is likely to cause the most issues when the time comes for large-scale production, is the reliance of rLVs. It is therefore of utmost importance to minimise or eliminate the rLV component of the FoP protocol. This chapter describes the generation of a polycistronic vector, containing a single EF1α promoter and the coding sequences of eGFP, *GATA1*, *TAL1* and *FLI1* in tandem, from which a single rLV was produced to test in FoP. The polycistronic vector minimises rLV requirement of FoP, thus also reducing cost, and provides a more homogenous cell population at the start of FoP by ensuring cells receive all three of the necessary TFs.

To achieve transcription of a single mRNA transcript and translation of separate proteins from each gene, two options were considered; either internal ribosome entry sites (IRES) or 2A peptides could be used to separate ORFs. IRES sequences are ~450 nucleotides in length, which is much longer than 2A sequences, typically between 54-66 nucleotides. Genes downstream of an IRES sequence have been associated with low levels of gene expression, usually between 20-50% than that of the gene upstream of the IRES sequence (Mizuguchi, 2000). Due to rLV vector capacity being limited, and the association of non-stoichiometric gene expression of sequential genes using IRES, the 2A peptide option was chosen.

For the 2A system to work, stop codons of each ORF, apart from the final one, are removed. Self-cleaving 2A oligopeptides contain a highly conserved c-terminal motif: D(V/I)EXNPGP. When protein synthesis occurs along an ORF that contains a 2A sequence, translation is paused by the ribosome at the glycine (G) and proline (P) codons of the 2A motif. This results in the nascent protein chain, up to and including the glycine residue, being released. This process is known as 'ribosome skipping' and results in two separate proteins being produced (Doronina, 2008). Consideration of the gene order when using 2A sequences is important, as the resulting proteins will have either a 2A peptide

attached to the C-terminus (for genes upstream of a 2A sequence), or a proline attached to the N-terminus (for genes downstream of a 2A sequence) (Hu, 2014), and cleavage efficency may be affected by the N-terminal protein following the 2A sequence (Szymczak, 2004).

The use of different 2A variants within a polycistronic cassette can reduce the risk of intramolecular homologous recombination (HR). Three of the 2A sequences most commonly used are;

- E2A: equine rhinitis A virus
- P2A: porcine teschovirus-1
- T2A: *Thoseaasigna* virus

In 2004, the resulting 2A-peptide linked proteins, encoded by a 2A polycistronic vector containing four genes, were shown to be stable in a number of *in vitro* cell lines, including an early embryonic stem cell line, and in mouse experiments *in vivo*. Expression of the four genes was shown to be stoichiometric (Szymczak, 2004). In 2009, the Jaenisch lab demonstrated that a polycistronic vector using the 2A sequences P2A, T2A, and E2A to separate the ORFs of Oct4, Sox2, Klf4, and c-Myc respectively, could be used to successfully reprogram mouse embryonic and somatic cells, as well as human somatic cells, into iPSCs. This work also demonstrated that three 2A peptides can be used to separate up to four genes, in a single lentiviral vector. Interestingly, it was shown that a single vector was sufficient for reprogramming to occur in mouse embryonic fibroblasts (MEFs). The vector used was tetracycline-inducible, with the four reprogramming genes under the control of the tetracycline operator minimal promoter. A constitutive FUW lentiral vector was co-transduced, to provide the reverse tetracycline responsive trans-activator (rTTA).

The work by Jaenisch and colleagues resulted in sub-optimal reprogramming of iPSCs compared to using separate vectors. Since transduction using separate vectors permits integration of different numbers of rLV for each factor, reprogramming may require a specific number of integrations, perhaps favouring higher expression of one factor over another. While this indeed might also be the case for MK-FoP, we were encouraged by qPCR data from the lab that suggested TG expression (and presumably, therefore, transduction of the three FoP TFs) of forward programmed cells was relatively equal. The Jaenisch lab later showed improved efficiency of reprogramming with a polycistronic vector, by testing multiple 2A sequences they found very subtle differences in vector design could have an impact on protein expression, as they described non-stoichiometric expression of their reprogramming factors (Carey BW, 2011). This highlights the importance of vector design, and a potential issue with non-stoichiometric expression, when trying to overexpress multiple proteins with a polycistronic system.

Chapter Overview

This chapter will describe the generation of a polycistronic vector (pWPT-GATA1-FLI1-TAL1) and how initial cell experimental results led to the hypothesis that gene order of this vector was impeding MK development, but beneficial for erythroblast development. After generating new polycistronic vectors with altered gene order, further experimental results added to the evidence that the original vector was not favourable for MK-FoP. The new vectors allowed the bifurcation of progenitors, generating both mature MKs and erythroblasts. This body of work shows that the expression of the three FoP TFs from a polycistronic cassette allows efficient MK and erythroblast production in permissive conditions, and is important evidence that an inducible cell line would be capable of forward programming.

Materials and Methods

Generating the Polycistronic expression Cassettes

The Polycistronic Cassettes (PCs) were assembled and cloned into the recombinant lentiviral backbone pWPT in place of the original eGFP sequence (***Map 2.1***) using standard molecular biology methods as described in the general method section. All PCRs for DNA cloning fragment synthesis were performed using the Phusion Taq high-fidelity DNA polymerase according to manufacturer instructions (NEB, cat # M0530S). All restrictions enzymes come from NEB and have been used in optimal digestion buffers. Plasmids were dephosphorylated with Antarctic phosphatase according to manufacturer instructions (NEB, cat # M0289S). Gibson Assembly (NEB, cat # E5510S) was performed according to manufacturer instructions. DNA fragment and plasmid purifications were achieved using Qiagen plasmid Mini or Midi kits (cat # 12123 and 12143 respectively) following standard protocols. Plasmid constructs were cloned and amplified by heat transformation of the bacteria strains DH5α, Stbl3 (ThermoFisher Scientific, cat # 18265017 and C737303 respectively), XL-10 Gold (Agilent Technologies, cat # 200314) and Mix & Go (Zymo Research, cat # T3007). All final plasmid constructs have been verified by Sanger Sequencing (Source Bioscience), of modified functional regions using primers listed in **Table 2.1**.

PC-Entry vector: pWPT-Entry2A

We first generated a bespoke pWPT-Entry2A vector including the P2A and T2A sequences separated by unique cloning sites (MluI, SpeI, SalI) in order to allow a versatile and simple directed cloning of up to three individual coding sequences in a polycistronic cassette under the control of the ubiquitous EF1a promoter. The pWPT-eGFP plasmid (***Map 2.1***) was first co-digested with MluI and SalI to remove the eGFP coding sequence. The Entry2A fragment was generated by a joining PCR of the two oligonucleotides EntryP2A_Fo and EntryT2A_Re using the two primers MluIP2A_Fo and SalIT2A_Re. Oligonucleotides (2µM) and primers (0.2µM) were combined with in a 50µl PCR reaction and thermocycled following recommended conditions, with an annealing temperature of 60^{o}C, with 35 cycles. The resulting Entry2A fragment was double digested (MluI/SalI) and ligated to the opened pWPT vector to obtain the pWPT-Entry2A plasmid.

PC vector: pWPT-GATA1-FLI1-TAL1co

The GATA1 coding sequence (NM_002049.3) was first cloned into the pWPT-Entry2A. The GATA1 sequence was obtained by PCR using the available pWPT-GATA1 plasmid as a template and the primers 2AEntry-GATA1 Fo and 2AEntry-GATA1 Re including MluI restriction sites. A Kozak sequence

was also included in the primer 2AEntry-GATA1 Fo before the GATA1 initiation codon for efficient translation and the stop codon omitted in the primer 2AEntry-GATA1 Re to avoid translation termination before the 2A peptide. The pWPT-Entry2A vector was linearized by MluI digestion, dephosphorylated to avoid recircularization and ligated to the MluI digested GATA1 PCR fragment.

The FLI1 sequence (NM_002017.4) was PCR amplified from the available pWPT-FLI1 plasmid using the primers 2AEntry-FLI11 Fo and 2AEntry-FLI11 Re including SpeI restriction sites. The FLI1 coding sequence was subsequently cloned into the resulting pWPT-GATA1-2A plasmid using a similar strategy and the SpeI restriction enzyme.

The unbalanced GC-rich TAL1 coding sequence was difficult to PCR amplify, it was therefore synthesised (GeneArt) to include the addition of SalI restriction sites at both ends, a 3' stop codon, and codon optimised to correct GC-content distribution. The codon optimised TAL1 (TAL1co) was then cloned into the pWPT-GATA1-2A-FLI1-2A vector after SalI digestion. The resulting polycistronic vector pWPT-GATA1-FLI1-TAL1co (***Map 2.2***) was sequenced verified.

PC1 vector: pWPT-eGFP-GATA1-FLI1-TAL1co

We decided it would be beneficial to insert the fluorescent marker gene eGFP into the vector, in order to easily identify cells transduced with the polycistronic cassette in FoP. The GFP coding sequence was PCR amplified from the pWPT-GFP vector (***Map 2.1***) using the primers MluI_eGFP_Fo and MluI_eGFP(mut.stop)_Re including MluI restriction sites. A Kozak sequence was also included in the primer MluI_eGFP_Fo before the eGFP initiation codon. The primer MluI_eGFP(mut.stop)_Re omitted a stop codon and added a unique NotI restriction site. The resulting eGFP sequence was subsequently cloned into the pWPT-GATA1-FLI1-TAL1co plasmid using a similar strategy and the MluI restriction enzyme, resulting in eGFP replacing the GATA1 fragment.

An E2A sequence was generated by annealing the two oligonucleotides E2A_GATA1_Fo and E2A_GATA1_Re including NotI restriction sites. 100µM of each oligonucleotide was added to 5µl NEB buffer 2.1, plus 35µl dH_2O, to a final volume of 50µl. A touchdown PCR was performed using the thermocycling conditions: 95°C 10min, 95°C 3min (-1°C/cycle) x70 cycles, 4°C hold. A joining PCR (previously described) using the primers E2A_GATA1_Fo3 and E2A_GATA1_Re2 joined the E2A oligonucleotide and GATA1 coding sequence. The E2A-GATA1 fragment was cloned into the pWPT-GFP-FLI1-TAL1co vector using the NotI restriction enzyme, resulting in the pWPT-GFP-GATA1-FLI1-TAL1co vector (***Map 2.3***), which was sequence verified and will be subsequently referred to as PC1.

PC2: pWPT-eGFP-GATA1-TAL1co-FLI1

Initial results for MK-FoP were poor with PC1, and a erythroblast bias was observed, which we hypothesised may be due to the positioning of FLI1 in the vector. Therefore, we decided to generate two more polycistronic vectors; PC2-with FLI1 moved to the final gene position, and PC3- with a codon optimised FLI1 sequence. FLI1 codon optimisation was done to increase the amount of FLI1 protein produced, by replacing rare codons with ones that have tRNA more abundantly available in human transduced cells, to try and improve FoP.

PC1 was double digested with SpeI and SalI to remove the FLI1,T2A and TAL1co fragments, resulting in a pWPT-GFP-GATA1 vector. A new TAL1co fragment was PCR amplified from the synthetoc TAL1CO plasmid (GeneArt- used to generate PC vector) using the primers SpeI_AsiSI_Tal1CO_Fo and SalI_AsiSI_Tal1CO_Re, removing the stop codon at the 3' end of TAL1co and including AsiSI (5' and 3'), SpeI (5') and SalI (3') sites. The resulting mutated stop TAL1co PCR product was digested (SalI/SpeI) and ligated to the open pWPT-GFP-GATA1 vector, generating a pWPT-GFP-GATA1-TAL1co vector.

A new T2A fragment was generated by annealing the two oligonucleotides T2A_FLI1oligo_Fo and T2A_FLI1oligo_Re, as previously described. A FLI1 fragment was amplified from the pWPT-FLI1 vector using the primers T2A_FLI1_Fo and SalI_FLI1STOP_Re, including a stop codon and SalI site (3'). The T2A and FLI1 PCR products were joined in a joining PCR using the primers SalI_T2A_Fo and SalI_FLI1STOP_Re, as previously described. The resulting T2A-FLI1 PCR product was used as a template for a final PCR to generate a fragment for Gibson assembly, using the primers G_TCO_FLI1_Fo and G_TCO_FLI1_Re, before a Gibson assembly was performed with the SalI linearized pWPT-GFP-GATA1-TAL1co vector. The resulting vector, pWPT-GFP-GATA1-TAL1co-FLI1 (***Map 2.4***), will be subsequently referred to as PC2.

PC3: pWPT-eGFP-GATA1-TAL1co-FLI1co

A FLI1co fragment was generated using the synthetic pWPT-FLI1co vector (GeneArt) and the primers BsiWI-FCO-Fo and G_TCO_Fli1CO_Re. The gel purified PCR product was amplified using Gibson primers Gibson_FCO_Fo2 and G_TCO_Fli1CO_Re, before a Gibson assembly was performed with the SalI linearized pWPT-GFP-GATA1-TAL1co vector. The resulting vector, pWPT-GFP-GATA1-TAL1co-FLI1co(***Map 2.5***), will be subsequently referred to as PC3.

pWPT-FLI1co

In order to compare MK production using three individual rLV vectors with PC3 vector, TAL1co and FLI1co single vectors were also generated. pWPT-TAL1co was already available (generated by Dr Amanda Evans). pWPT-FLI1 was double digested (MluI/SalI) to remove the FLI1 ORF. The T2A-FLI1co Gibson PCR product used for PC3 cloning, was amplified with the primers MluI_K_FLI1CO_Fo and FLI1COadd_Re inserting a Kozak sequence and MluI site (5'). The PCR product was MluI/SalI digested and liagated with the linearized pWPT vector, resulting in the pWPT-FLI1co vector (***Map 2.6***).

Flow Cytometry

Since the PC vectors contain GFP TG, they were stained without a FITC-conjugated antibody. Cells were therefore stained with CD41a-APC-H7 (1:100), CD42b-APC (1:20) and CD235a-PE (1:200). Erythroblast characterisation on was performed with the following antibody combination; CD36-PE (1:10), CD71-APC-H7 (1:100), CD235a-APC (1:200).

PSC lines used

This Chapter shows experimental results of forward programming primarily with the iPSC BobC line. The majority of results were then verified in a second iPSC line, FFDK, which behaves differently in FoP to BobC, to check whether results were iPSC line dependent. Testing of the initial PC vector (**Map 2.2**) was performed on the iPSC line S4 as a second line, which, like FFDK, behaves differently in FoP to BobC. All three lines are routinely used in the lab for FoP but due to experiments taking place over a long time period, to do with re-cloning of the PC vector, S4 was not available for all experiments shown here.

Forward Programming

All experiments presented in this Chapter follow the standard forward programming protocol for MK or erythroblast culture, as described in the Main Materials and Methods Chapter. All transductions were performed at an MOI of 20, unless otherwise stated in the text.

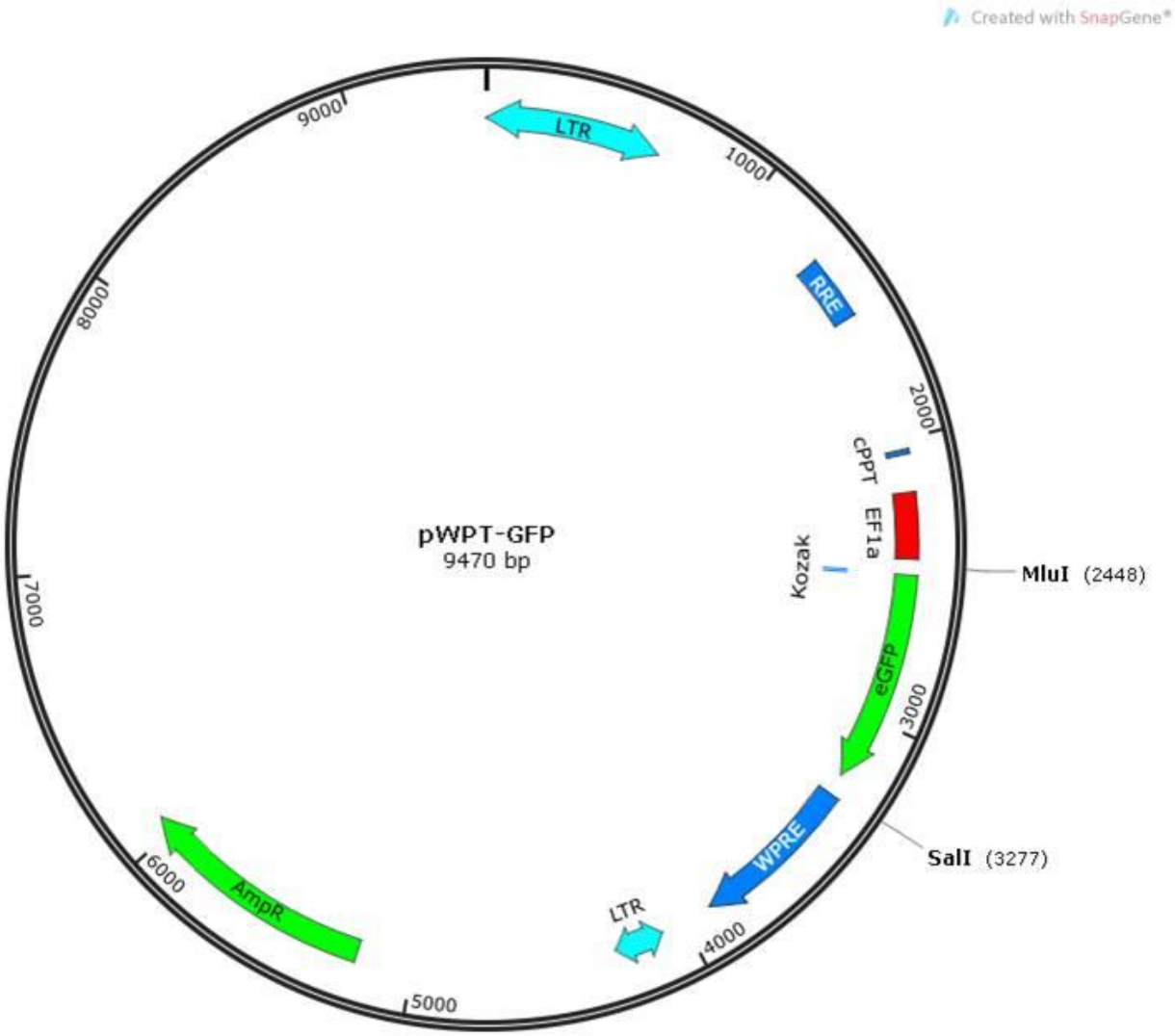
Materials and Methods Tables

Table 2.1. Primers and oligonucleotides used for generating and sequencing PC/1/2/3 Vectors

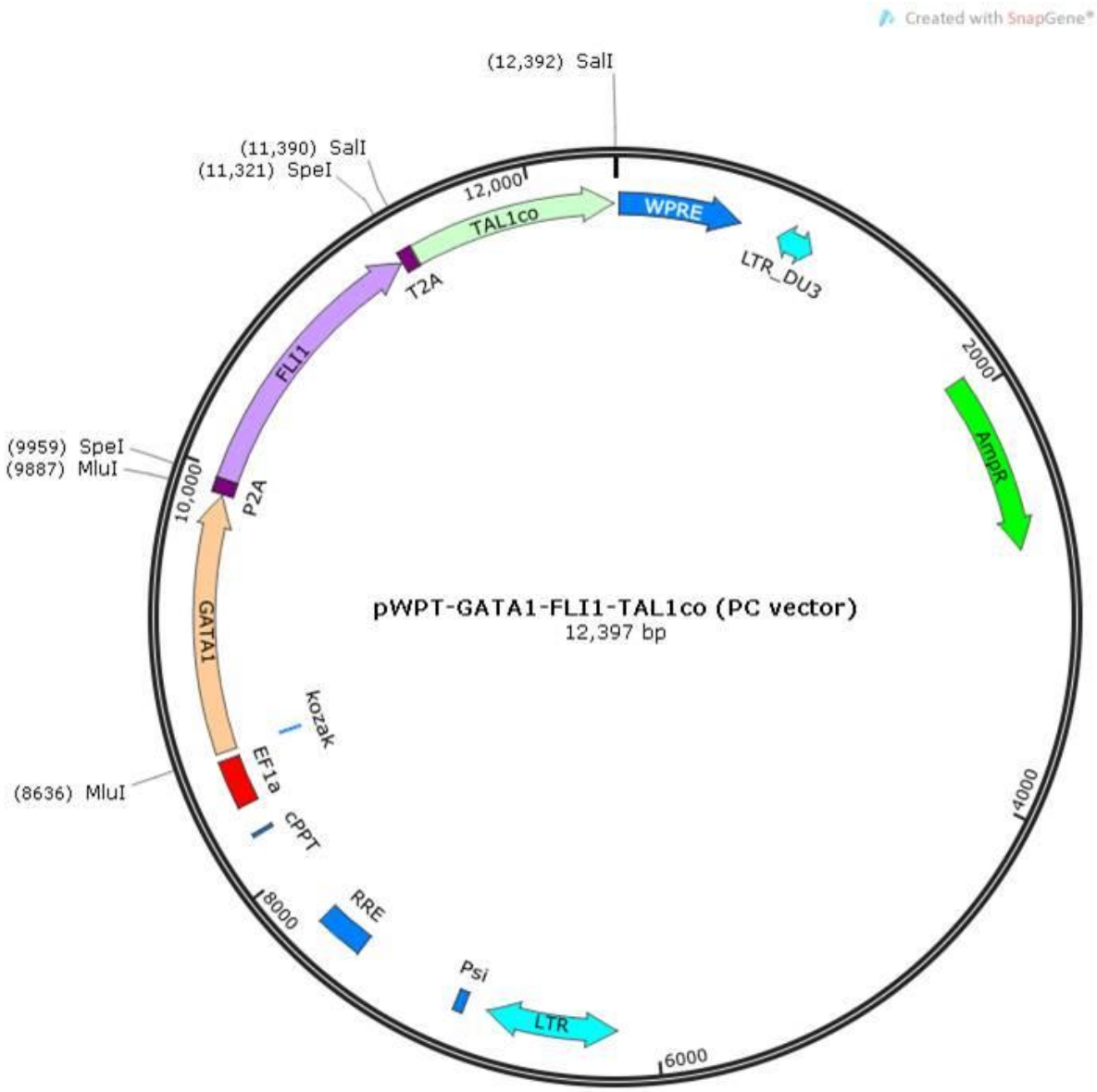
Oligonucleotide/Primer name	**Sequence 5'-3'**
EntryP2A_Fo	GCCACGCGTGGAAGCGGAGCTACTAACTTCAGCCTGCTGAAGCA GGCTGGAGACGTGGAGGAGAACCCTGGACCTACTAGTGGA
EntryT2A_Re	GGCGTCGACAGGTCCAGGATTCTCCTCGACGTCACCGCATGTTAG CAGACTTCCTCTGCCCTCTCCGCTTCCACTAGTAGGTC
Mlu1P2A_Fo	GCCACGCGTGGAAGCGGA
SalIT2A_Re	GGCGTCGACAGGTCCAGG
2AEntry-GATA1 Fo	GCCACGCGTGCCACCATGGAG
2AEntry-GATA1 Re	GCCACGCGTTGAGCTGAGCGGAGCCAC
2AEntry-FLI1 Fo	GCCACTAGTATGGACGGGACTATTAAG
2AEntry-FLI1 Re	GCCACTAGTGTAGTAGCTGCCTAAGTG
WPT_Seqi_Fo	TCTTTTTCGCAACGGGTTT
WPT_Seqi_Re	CCACATAGCGTAAAAGGAGCA
GATA1_TG_Fo	GGTGGCTCCGCTCAGCTCAT
FLI1_TG_Fo	CCCGCCATCCTAACACCCACG
FLI1_Seqi_Fo1	TCAGTCAGAAGAGGAGCTTGG
FLI1_Seqi_Re1	GGGCCGTTGCTCTGTATTCT
GATA1_Seqi_Fo2	TGGTGGCTTTATGGTGGTG
TAL1_TG_Fo	CCTCCATCCTGCCATGCT
MluI_eGFP_Fo	GCCACGCGTTGATCAGCCACCATGGTGAGCAAGGGC
MluI_eGFP(mut.stop)_Re	GCCACGCGTGCGGCCGCCCCGGGGTAGCTACTAGCTAGT
E2A_GATA1_Fo	TGCGGCCGCTGGAAGCGGACAGTGTACTAATTATGCTCTCT TGAAATTGGCTGGAGATGTTGAGAGCAACCCTGGACCTTTC GAAATGGAGTTCCCT
E2A_GATA1_Re	AGGGAACTCCATTTCGAAAGGTCCAGGGTTGCTCTCAACATC TCCAGCCAATTTCAAGAGAGCATAATTAGTACACTGTCCGCTT CCAGCGGCCGCA
E2A_GATA1_Fo3	TATTCCGTTGCGGCCGCTGGAAGCG
E2A_GATA1_Re2	ATTGAGCCGCGGCCGCCTGAGCTGAGCG
Gibson Primers for PC2/PC3	
TCO_Fli1_Fo	GGACCTAGAGCGATCGCGTTAAAGTCGACTTGCG
TCO_Fli1_Re	GATTATCGGAATTCCCTCGAGGATTGTCGACTCAGTAG
BsiWI-FCO-Fo	
Gibson_FCO_Fo2	GACCTAGAGCGATCGCGTTGTCGACTTGCGATCGCGGAA
G_TCO_Fli1CO_Re	GATTATCGGAATTCCCTCGAGGAGTCGACTCAGTAGTA
Additional primers used for sequencing	
TAL1CO_Seqi_Re	AGCCGCAGGATCTCGTTCTT
GATA1_seq_1Re	GTGGGAGAAAAGAAGGTACTGG
GATA1_seq_2Re	ATTCCCGCTACCGCTG
FLI1_seq_1Re	ATTTGCTAACGCTGCAGTCC
FLI1_seq_1Fo	GGAGTATGACCACATGAATGG

FLI1_seq_2Re	GCCGCATCACAATACTGG
FLI1_seq_2Fo	GGTGAACTTTGTCCCTCC
TAL1CO_seq_1Re	AGGCAGTTCAGCTGTCACA
TAL1CO_seq_1Fo	CCAGACACAGAGTGCCTACC
TAL1CO_seq_2Fo	GCAAGAACGAGATCCTGC
eGFP_Seqi_Re	TAGCTACTAGCTAGTCGAGA
E2A_GATA1_Fo	GCCTGCGGCCGCTGGAAGCG
E2A_GATA1_Re	GCCGCGGCCGCCTGAGCTGAGCG
E2A_GATA1_Fo2	CCTGGACCTTTCGAAATGGAGTTCCCTGGC

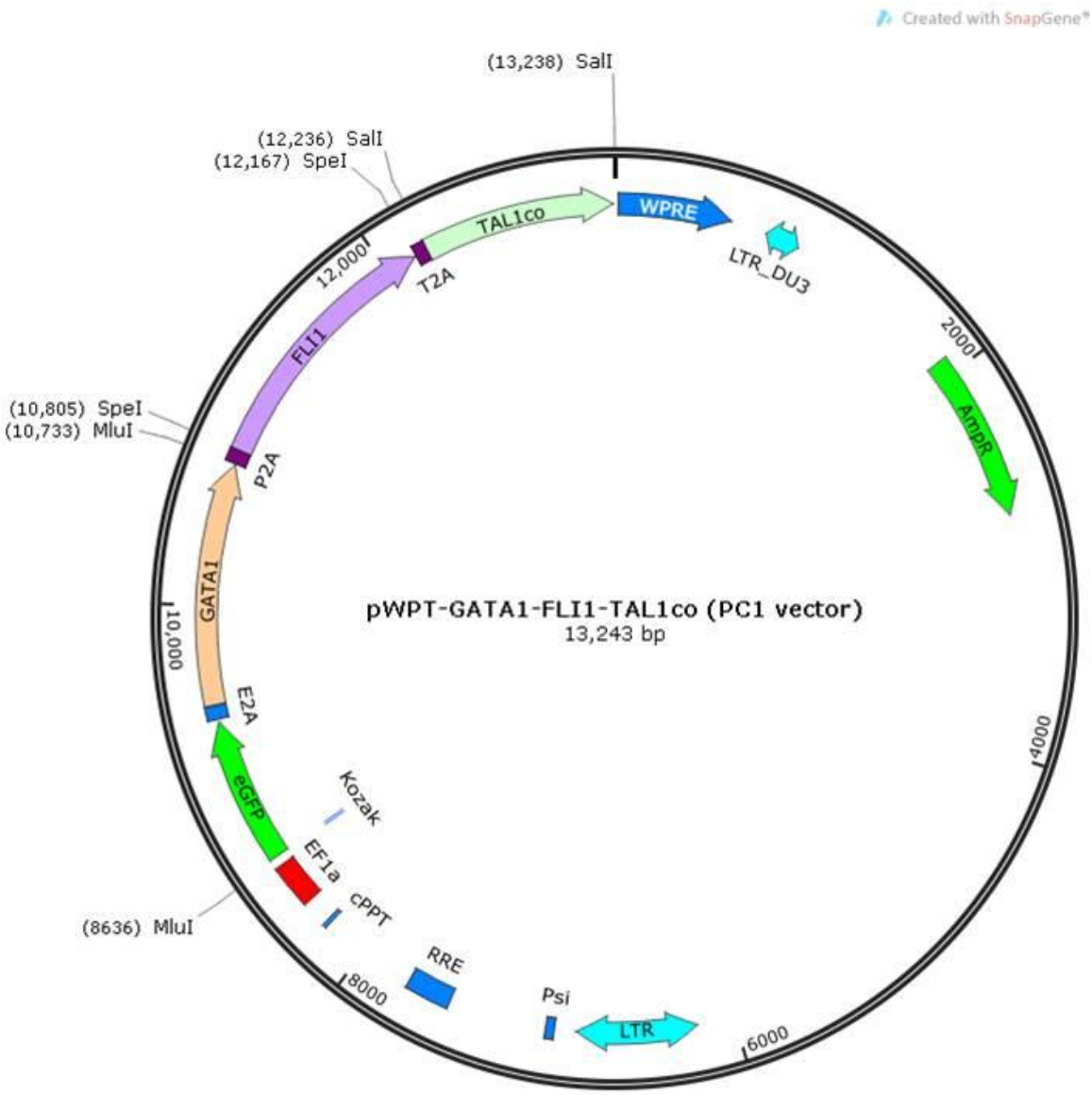
Chapter 2: Vector Maps



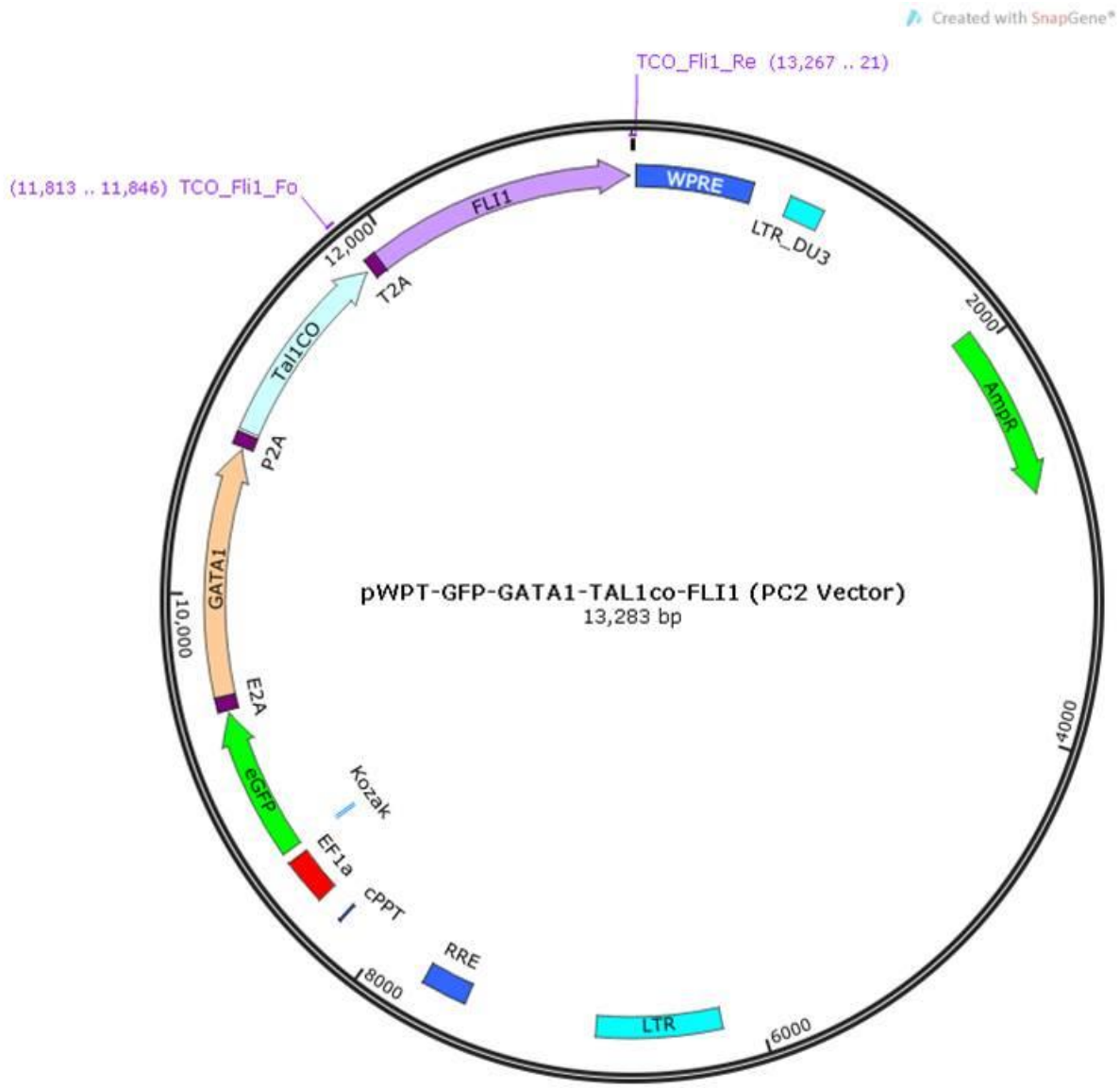
Map 2.1 pWPT-GFP starting vector for Polycistronic cloning strategy. Restriction enzyme sites used for cloning are shown.



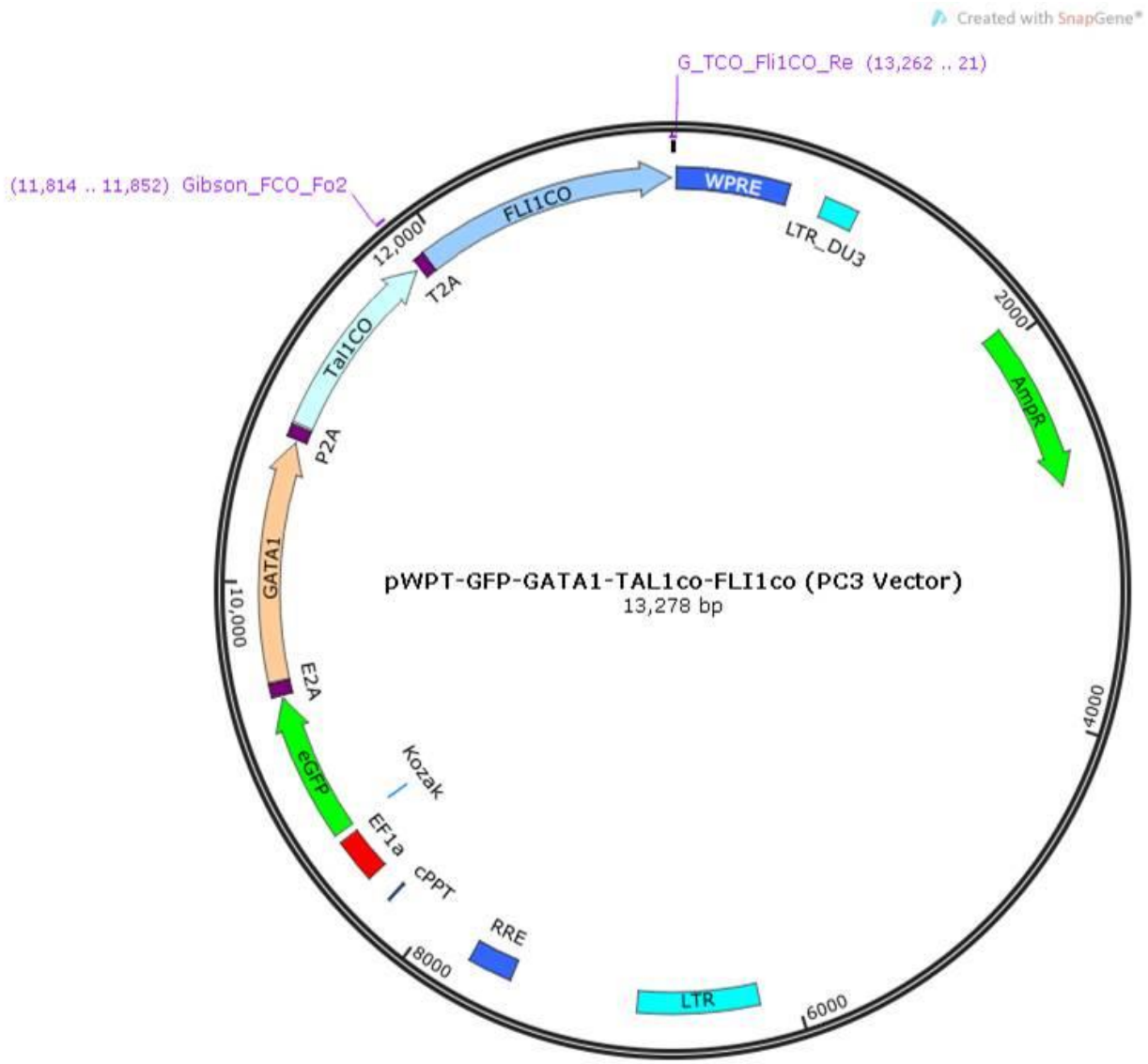
Map 2.2 PC vector: pWPT-GATA1-FLI1-TAL1Cco. Restriction enzyme sites used for cloning steps are shown.



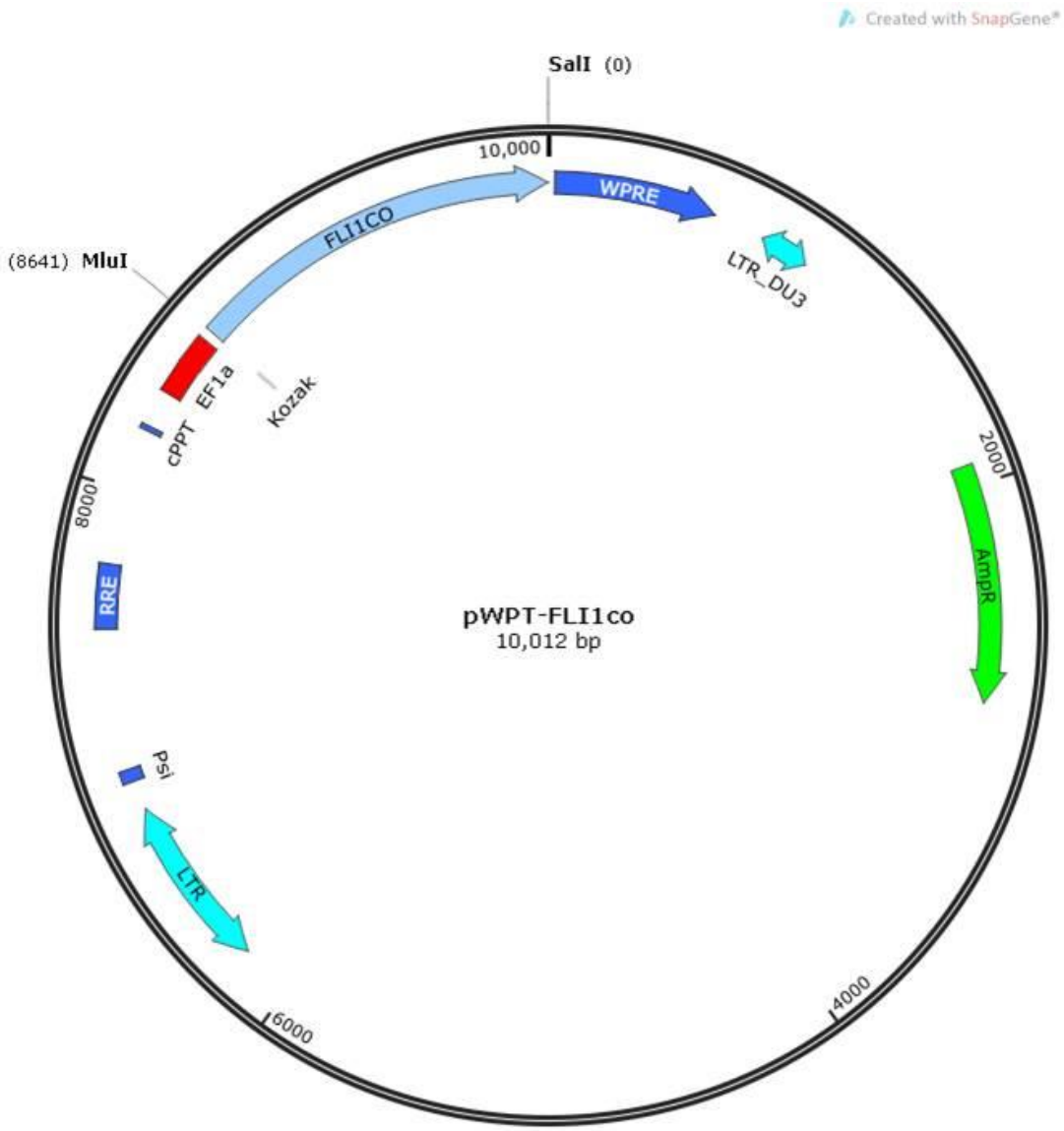
Map 2.3 PC1 Vector: pWPT-GFP-GATA1-FLI1-TAL1co. Restriction enzyme sites used for cloning steps are shown.



Map 2.4 PC2 vector: pWPT-GFP-GATA1-TAL1co-FLI1. Gibson primers used to produce final T2A-FLI1 fragment are shown.



Map 2.5 PC3 vector: pWPT-GFP-GATA1-TAL1co-FLI1co. Gibson primers used to produce final T2A-FLI1co fragment are shown.



Map 2.6 pWPT-FLI1co. Restriction enzyme sites used for cloning steps are shown.

Results

In the results section the following cell type definitions are used; CD41+/CD235+ cells are bi-potent progenitors, CD41-/CD235+ cells are erythroblasts, CD41+/CD235- cells at day 9-10 are MKs, and CD41+/CD42+ cells from day 20 onwards are mature MKs. CD42 antibodies detect glycoprotein Ib (GPIb), expressed on the surface of mature MKs and corresponds to a late step of MK differentiation and is also a component of the GPIb-V-IX complex on platelets (Chang Y, 2007).

Testing the pWPT-GATA1-FLI1-TAL1 Vector in TPO: Low MK and High Erythroblast Potential

The pWPT-GATA1-FLI1-TAL1co (PC) recombinant lentivirus (rLV) was first used to transduce two iPSC lines, BobC and S4, with increasing multiplicity of infection (MOI: 20, 50 and 100), in an initial experiment to test whether stoichiometric expression of the three TFs would enable MK-FoP to occur in TPO. **Fig 2.1** shows the results for BobC. The number and percentage of cells at day 10 increased with MOI, with the MOI 100 condition resulting in the highest number and percentage of bi-potent progenitors and erythroblasts, from 1.0E+05 iPSCs seeded (**Fig 2.1A/B**). The MOI 20 and 50 conditions showed little difference in progenitor and MK number, while erythroblast cell number was higher in MOI 20. Unexpectedly, the highest proportion of cells produced was erythroblasts.

By day 20 the same pattern was followed, with the MOI 100 condition giving rise to the highest number of progenitor, erythroblast and MK cells (**Fig 2.1C**). A drop in cell number, compared to day 10, was observed for all cell types in all conditions tested. In particular, the expected MK growth and maturation was not observed. Day 20 shows an increase in both progenitor and erythroblast percentages for all conditions tested while MK percentage failed to increase, remaining below 3% (**Fig 2.1D**). Interestingly, the erythroblast purity was positively correlated with MOI while the bi-potent progenitor purity decreased with MOI.

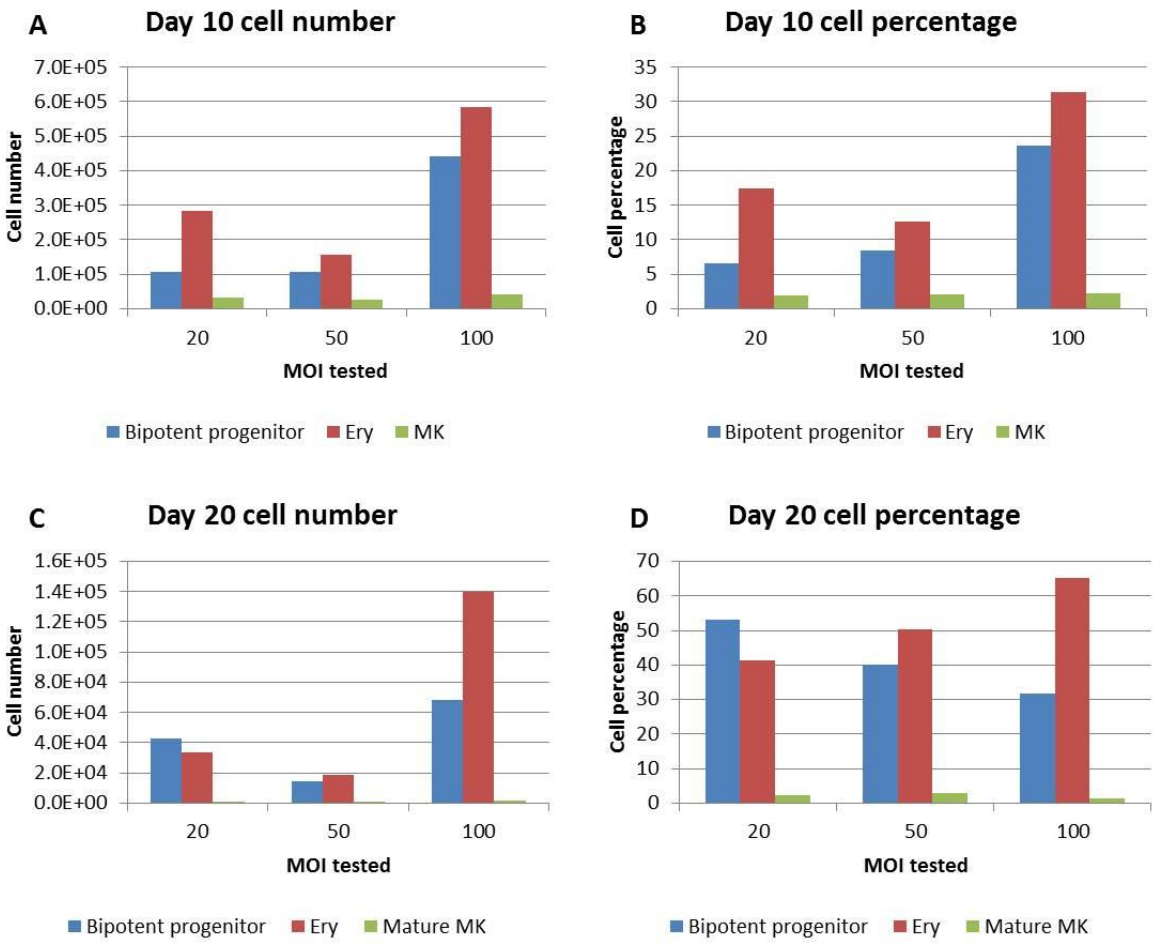


Fig 2.1 Testing the PC vector with increasing MOI in BobC MK-FoP. The PC vector (pWPT-GATA1-FLI1-TAL1, **Map 2.2**) was transduced with an increasing MOI (20, 50 and 100) in BobC cells. After 2 days of mesoderm induction, MK medium (TPO and SCF in CellGro) was used throughout the next 18 days of culture. **A**) Day 10 cell number for CD41+/CD235+ bi-potent progenitors, CD41-/CD235+ erythroblasts and CD41+/CD235- MKs. **B**) Day 10 cell percentages of populations present in **A**. **C**) Day 20 cell number for bi-potent progenitors and erythroblasts (as before) and CD41+/CD42+ mature MKs. **D**) Day 20 cell percentages of populations present in **C**. N=1.

Fig 2.2 shows results for the second cell line tested, S4. This line reproduced the same pattern of results observed for BobC, with the highest cell numbers and percentages observed in the MOI 100 condition at day 10 (**Fig 2.2A/B**).This line also failed to produce mature MKs by day 20 and showed the same trend of erythroblast purity positively correlating with MOI, while bi-potent progenitor purity showed the reverse trend (**Fig 2.2C/D**).

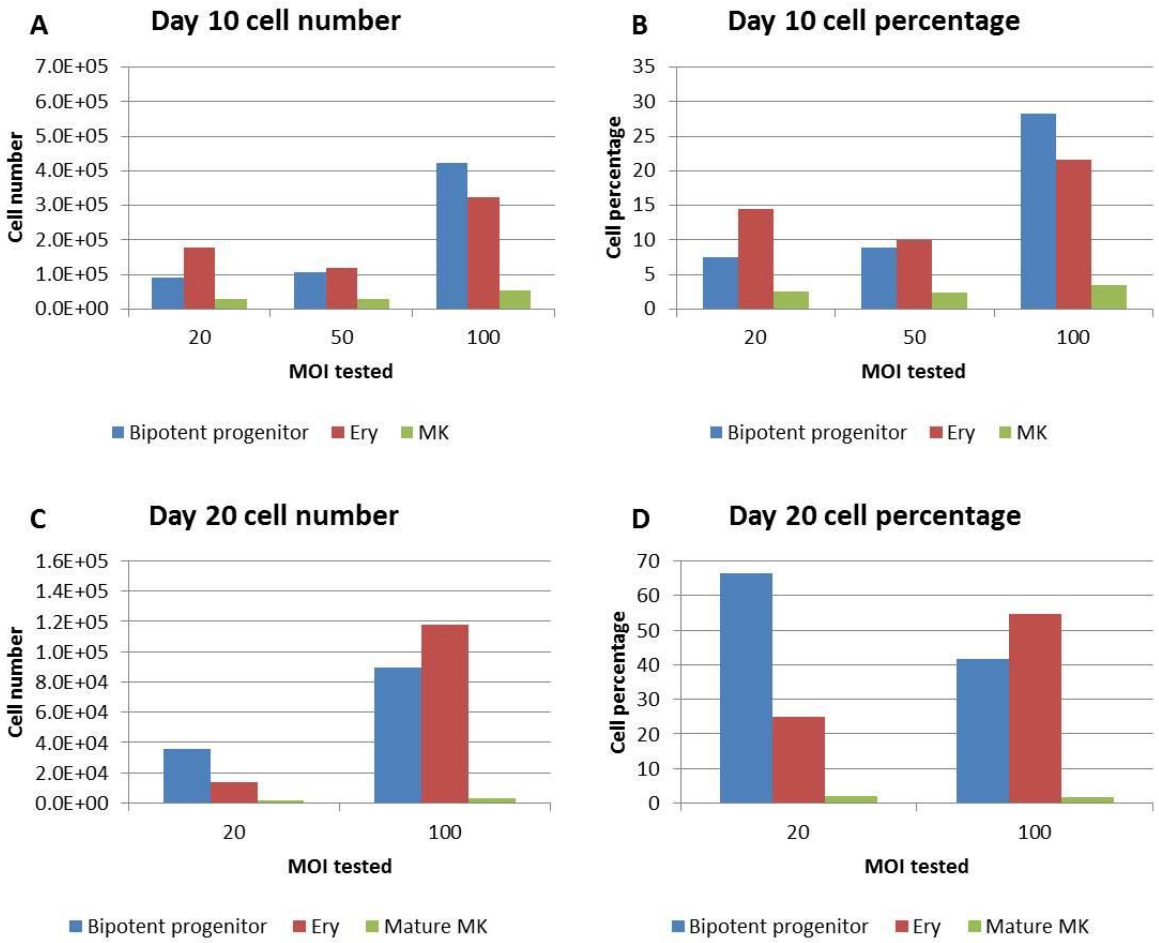


Fig 2.2 Testing the PC vector with increasing MOI in S4 MK-FoP. The PC vector (pWPT-GATA1-FLI1-TAL1, **Map 2.2**) was transduced with an increasing MOI (20, 50 and 100) in S4 cells. After 2 days of mesoderm induction, MK medium (TPO and SCF in CellGro) was used throughout the next 18 days of culture. **A**) Day 10 cell number for CD41+/CD235+ bi-potent progenitors, CD41-/CD235+ erythroblasts and CD41+/CD235- MKs. **B**) Day 10 cell percentages of populations present in **A**. **C**) Day 20 cell number for bi-potent progenitors and erythroblasts (as before) and CD41+/CD42+ mature MKs. **D**) Day 20 cell percentages of populations present in **C**. N=1. Day 20 data is missing for MOI 50 due to contamination of culture.

Overall, these results show that the PC vector is capable of forward programming, seen by the emergence of progenitors, erythroblasts and MKs at day 10. Low MK cell number and percentage at day 20 show that this vector failed to generate MKs, which was replicated in two iPSC lines known to FoP successfully using separate rLVs. However, the PC vector did unexpectedly produce a high number and percentage of erythroblasts, despite the MK-promoting TPO culture conditions used. We hypothesised that TF gene order in the PC vector may be responsible for the observed results. This incited us to revise the design of the PC cassette, notably by modifying the location and composition of the *FLI1* coding sequence known to be a key driver of MK-FOP.

At this stage, we also added GFP as a marker gene to the PC vector sequence, generating pWPT-eGFP-GATA1-FLI1-TAL1co (PC1). To revise this vector, *FLI1* was moved to the last gene position, generating pWPT-eGFP-GATA1 -TAL1co- FLI1 (PC2). For the final vector, *FLI1* was codon optimised pWPT-eGFP-GATA1 -TAL1co- FLI1co (PC3). We hypothesised that PC1 might be best for erythroblast production, and PC3 best for MK production. Since all the PC vectors contained a codon optimised *TAL1* (TAL1co) sequence, and the PC3 vector only contained a FLI1co sequence, we also generated rLVs for the single codon optimised genes, pWPT-TAL1co and pWPT-FLI1co. This was done to test the proper controls for each of the PC vectors, as described in **Table 2.2.**

3 rLV mix:	Abbreviation used: (Fig 2.3 and Fig 2.4)	Represents:
pWPT-GATA1, pWPT-FLI1, pGEM-TAL1	3TF	Classical 3TF FoP condition
pWPT-GATA1, pWPT-FLI1, **pWPT-TAL1co**	TCO	PC1 and PC2 control
pWPT-GATA1, **pWPT-FLI1co**, **pWPT-TAL1co**	TCO/FCO	PC3 control

Table 2.2 Control conditions for Polycistronic Cassettes.

No Clear Effect of Codon Optimisation on MK Outcome

Another reason for testing the codon optimised single gene rLVs is due to the fact that codon optimisation introduces many point mutations, which could theoretically introduce novel and unwanted properties and may result in toxicity or decreased efficiency (Schambach A, 2013). Therefore, the classical 3TF control, alongside the 3 rLV controls for PC1/2 and PC3 were tested in two iPSC lines in TPO, to see if codon optimisation of *TAL1* and *FLI1* had any effect on MK-FoP. **Fig 2.3** shows the results in BobC. Day 9 cell number shows a similar outcome for the 3TF and TCO conditions, while TCO/FCO failed to produce many cells (**Fig 2.3A**). A high percentage of bi-potent progenitors was produced in the 3TF and TCO conditions (**Fig 2.3B**). By day 20, the highest number of cell produced for the 3TF and TCO conditions was erythroblasts, while the TCO/FCO condition produced the highest number of mature MKs (**Fig 2.3C**). The percentage of bi-potent progenitors

and mature MKs increased the most in the TCO/FCO condition, while the erythroblast percentage increased most in the 3TF and TCO conditions (**Fig 2.3D**).

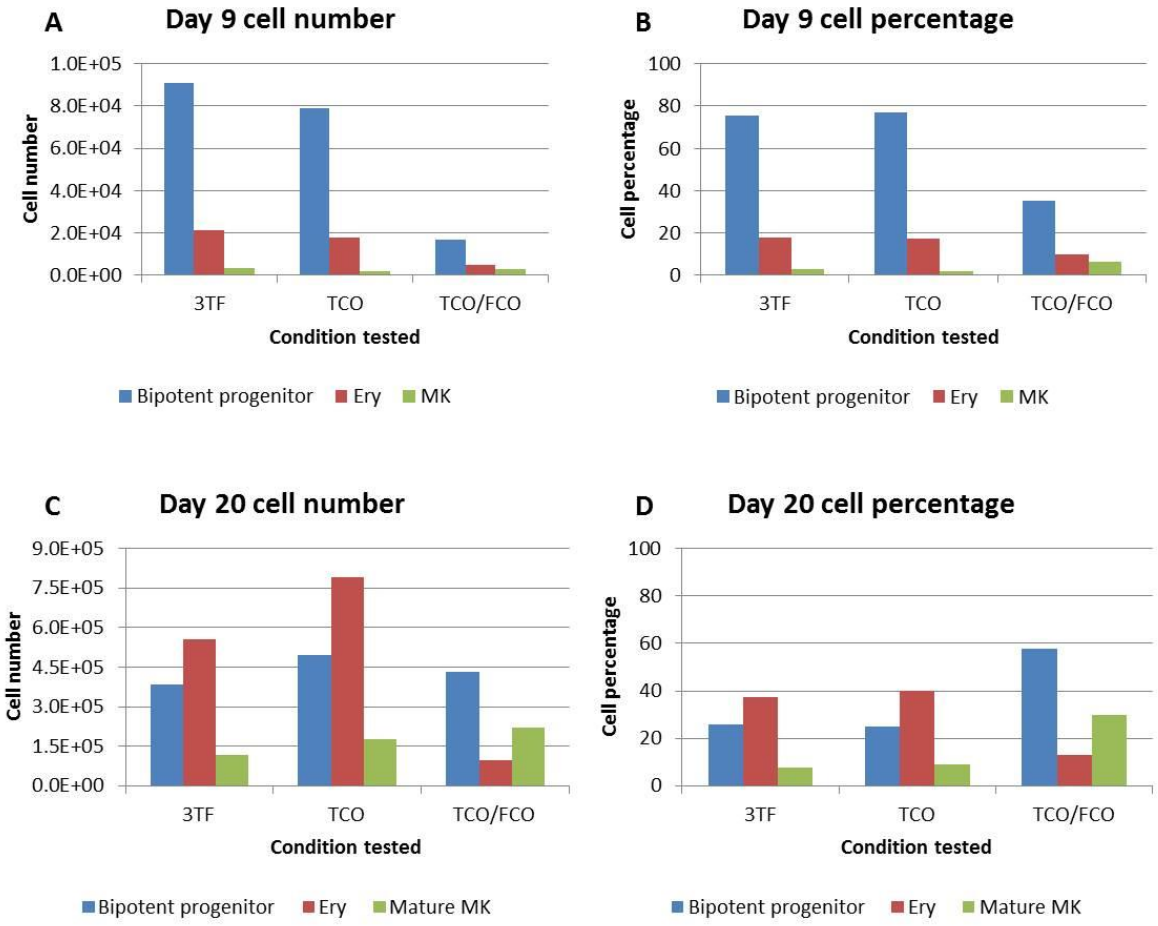


Fig 2.3 Testing transduction of single vector controls for forward programming in BobC MK-FoP. The iPSC line BobC was transduced with the following vector combinations; 1) Three TFs (3TF) = single lentiviral vectors used for GATA1, FLI1 and TAL1. 2) Codon optimised TAL1 (TCO) = single lentiviral vectors used for GATA1, FLI1 and TAL1co. 3) Codon optimised TAL1 and FLI1 (TCO/FCO) = single lentiviral vectors used for GATA1, TAL1co and FLI1co. All conditions were maintained in MK medium (TPO and SCF in CellGro) after 2 days in mesoderm medium. **A**) Day 9 cell number for CD41+/CD235+ bi-potent progenitors, CD41-/CD235+ erythroblasts and CD41+/CD235- MKs. **B**) Day 9 cell percentages of populations present in **A**. **C**) Day 20 cell number for bi-potent progenitors and erythroblasts (as before) and CD41+/CD42+ mature MKs. **D**) Day 20 cell percentages of populations present in **C**. N=1.

FFDK data is shown in **Fig 2.4**. The day 9 cell numbers were lower compared to BobC, with the 3TF and TCO/FCO conditions resulting in similar cell numbers and the TCO condition producing very few cells (**Fig 2.4A**). Cell percentages were similar for all conditions, with a higher percentage of MK cells at this stage than BobC (**Fig 2.4B**). Day 20 outcome was also different to that observed in BobC, with the TCO/FCO condition producing the lowest number of mature MKs (**Fig 2.4C**), despite similar cell percentages for all conditions (**Fig 2.4D**).

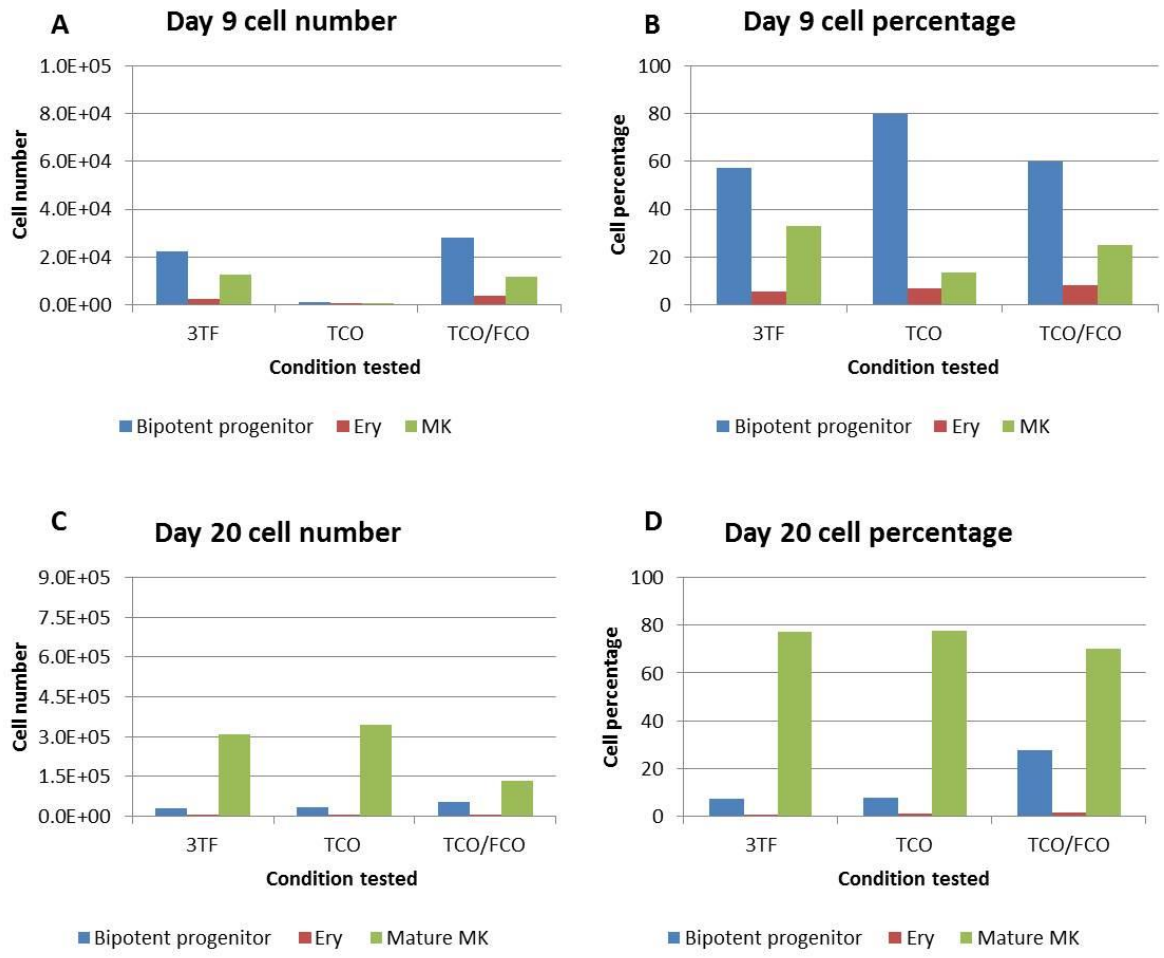


Fig 2.4 Testing transduction of single vector controls for forward programming in FFDK MK-FoP. The iPSC line FFDK was transduced with the following vector combinations; 1) Three TFs (3TF) = single lentiviral vectors used for GATA1, FLI1 and TAL1. 2) Codon optimised TAL1 (TCO) = single lentiviral vectors used for GATA1, FLI1 and TAL1co. 3) Codon optimised TAL1 and FLI1 (TCO/FCO) = single lentiviral vectors used for GATA1, TAL1co and FLI1co. All conditions were maintained in MK medium (TPO and SCF in CellGro) after 2 days in mesoderm medium. **A**) Day 9 cell number for CD41+/CD235+ bi-potent progenitors, CD41-/CD235+ erythroblasts and CD41+/CD235- MKs. **B**) Day 9 cell

percentages of populations present in **A**. **C**) Day 20 cell number for bi-potent progenitors and erythroblasts (as before) and CD41+/CD42+ mature MKs. **D**) Day 20 cell percentages of populations present in **C**. N=1.

Overall, it was difficult to know from this experiment whether codon optimisation was beneficial for MK-FoP. In both lines the TAL1co sequence resulted in a small increase in mature MK cell number at day 20. In BobC, FLI1co increased the mature MK outcome, however, the opposite was observed for FFDK. In light of this, we still wanted to test the new PC vectors, but did not include the TCO and TCO/FCO controls in further experiments. All experiments shown after this include the 3TF control as a measure that FoP worked.

Testing PC1/2/3 in TPO: PC1 has Lowest MK Potential but all PC Vectors Generate Long-term MK Cultures

FoP experiments were performed in BobC and FFDK with the 3TF control, PC1, PC2 and PC3 vectors in TPO and EPO. This was to test our hypothesis that gene order within the PC vector might lead to more efficient MK production by PC2 and PC3, and more efficient erythroblast production by PC1. **Fig 2.5** shows the average results for three experiments for BobC in TPO. Day 9 cell number was slightly higher in PC2 and PC3 for all cell types, despite cell percentages being similar for all conditions (**Fig 2.5A/B**). Day 20-22 mature MK cell number is shown separately for three experiments, as the outcome of each varied greatly. Overall, the highest number of mature MKs was generated by the PC2 and PC3 conditions, while the PC1 and 3TF conditions generated very few mature MKs (**Fig 2.5C**). Percentages at day 22 was generally higher for PC2 and PC3, but remained below 23% for two experiments, which is a bit lower than expected at this time point (**Fig 2.5D**). For experiment 1, MKs were maintained until day 68, showing all PC vectors were capable of producing long-term MK cultures (**Fig 2.5E**).

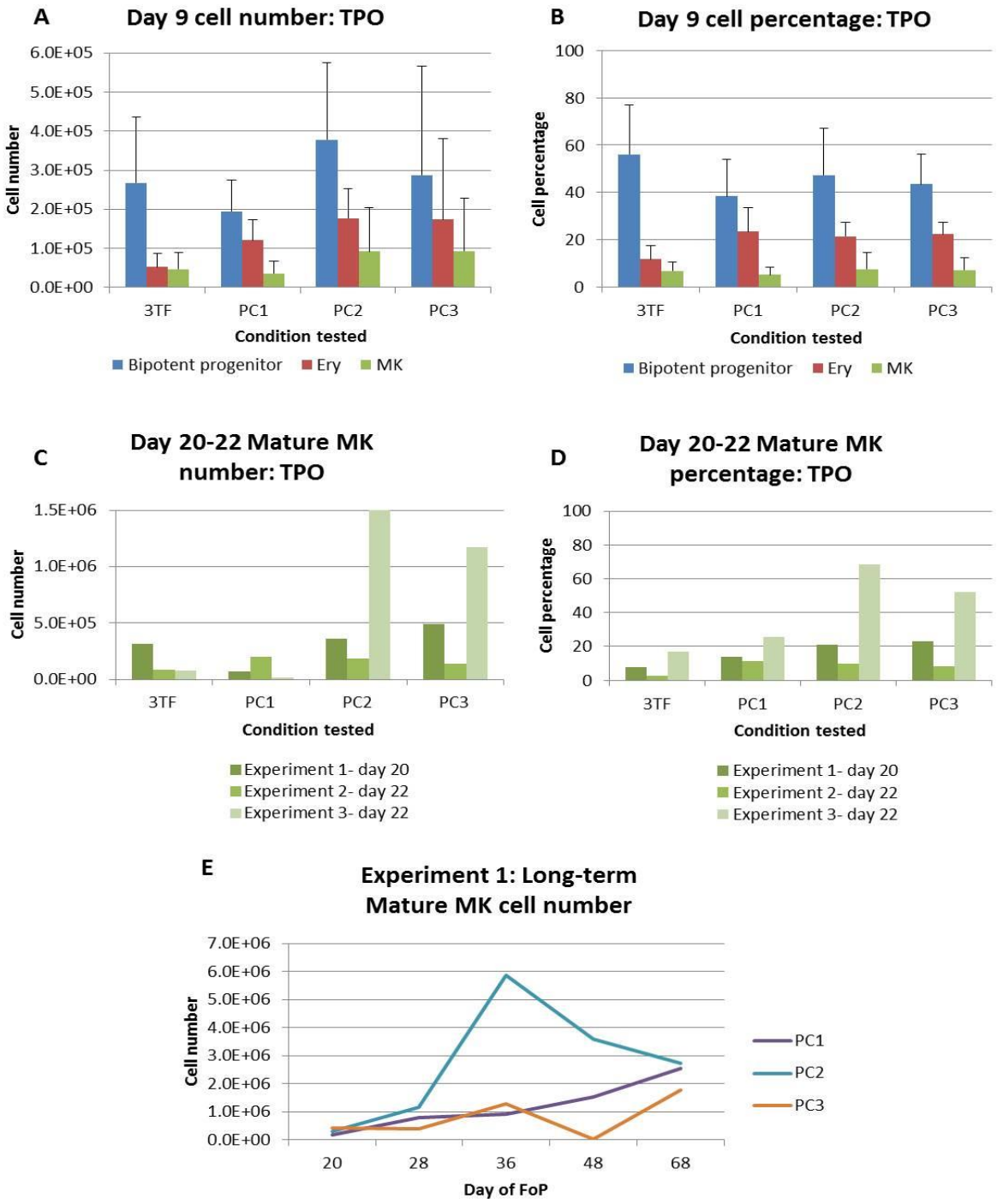


Fig 2.5 Testing the PC1, PC2 and PC3 vectors in BobC MK-FoP. 3TF (GATA1, FLI1 and TAL1 single lentiviral vector control), PC1, PC2 and PC3 vectors were used to transduce BobC cells. After 2 days of mesoderm induction, MK medium (TPO + SCF in CellGro) was used to culture cells for the remainder of the experiment. Cells were harvested at day 9, 20 or 22, 28, 36, 48 and 68 of culture. **A**) Day 9 cell number for CD41+/CD235+ bi-potent progenitors, CD41-/CD235+ erythroblasts and CD41+/CD235- MKs. **B**) Day 9 cell percentages of populations present in **A**. **C**) Day 20-22 cell number for CD41+/CD42+ mature MKs. **D**) Day 20-22 cell percentages of mature MKs. **E**) Experiment 1

mature MK number for the three PC vectors, maintained for 68 days of culture. N=3, biological replicates (data collected from experiments on day 20 for n=1 and day 22 for n=2), N=1 for time points after day 20, error bars=standard deviation.

Fig 2.6 shows the average results for FFDK in TPO. Day 9 cell numbers were highest in the 3TF condition, being lower for all three PC vectors. Overall cell number was lower than those observed for BobC, despite similar cell percentages, which were highest for the PC1 vector (**Fig 2.7A/B**). The 3TF condition at this stage produced significantly more MKs versus PC1 and PC3 (*P*= 0.006 and 0.004 respectively). By day 20-22 the number of mature MKs had not proliferated as much as in BobC, with the outcome of each experiment being different. The 3TF condition produced the highest mature MK cell number in experiment 1, while the 3TF, PC2 and PC3 conditions were all similar for experiment 2, and PC3 performed best in experiment 3. PC1 performed worst overall (**Fig 2.6C**). At this stage there was no statistical significance in the number of MKs produced by the different conditions. The day 22 percentage of mature MKs was consistently lowest for PC1, with the remaining conditions performing similarly and generating a higher percentage than observed in BobC at the same time-point (**Fig 2.6D**).

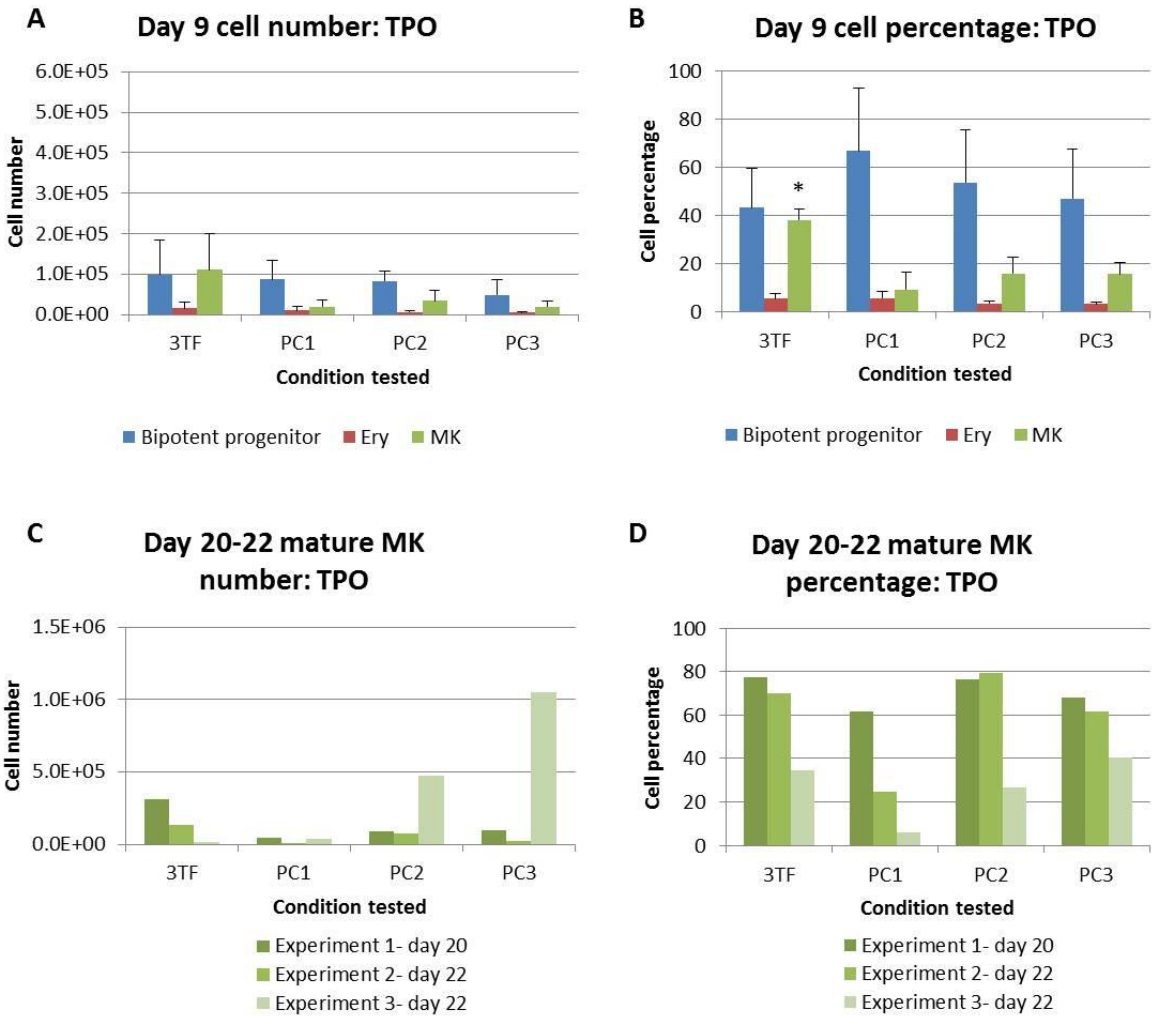


Fig 2.6 Testing the PC1, PC2 and PC3 vectors in FFDK MK-FoP. 3TF (GATA1, FLI1 and TAL1 single lentiviral vector control), PC1, PC2 and PC3 vectors were used to transduce FFDK cells. After 2 days of mesoderm induction, MK medium (TPO + SCF in CellGro) was used to culture cells for the remainder of the experiment. Cells were harvested at day 9 and 20 or 22. **A**) Day 9 cell number for CD41+/CD235+ bi-potent progenitors, CD41-/CD235+ erythroblasts and CD41+/CD235- MKs. **B**) Day 9 cell percentages of populations present in **A**. (P<0.01 by two tail t-test versus PC1 and PC3). **C**) Day 20-22 cell number for CD41+/CD42+ mature MKs. **D**) Day 20-22 cell percentages of mature MKs. N=3, biological replicates (data collected from experiments on day 20 for n=1 and day 22 for n=2), error bars=standard deviation.

PC1/2/3 in EPO: All Vectors Generate Erythroblasts

Fig 2.7 shows the average results for BobC in EPO of two experiments. The day 9 number of cells was highest in PC2 and PC3, with PC1 and the 3TF control giving similarly lower cell numbers (**Fig 2.7A**). Compared to BobC in TPO, the number of erythroblasts was increased, with almost three times as many in PC2 and twice as many in PC3. PC2 produced significantly more erythroblasts compared to the 3TF condition (P=0.04). The percentage of bi-potent progenitors was reduced in EPO, compared to TPO, while the inverse was seen for erythroblast percentage (**Fig 2.7B**). The day 22 erythroblast cell number is shown separately for the two experiments due to large variation. The outcome of experiment 1 was a high erythroblast number in all conditions, highest in PC3. Experiment 2 resulted in low overall erythroblast outcome, which was highest in the 3TF control (**Fig 2.7C**). No statistical significance was observed for the number of erythroblasts produced between the different conditions tested at this stage. The purity of erythroblasts was similar between experiments, and was high for all PC vectors, around 70-80%, and lowest in the 3TF condition, around 40% (**Fig 2.7D**). Globin expression was checked by qPCR, on cells from experiment 1, as described in **Fig 2.7**. Compared to the 3TF, all PC vectors expressed higher quantities of mRNA for all globins tested, apart from beta-globin which was not expressed in any condition. With epsilon and zeta expression, the erythroblasts produced had an embryonic phenotype (**Fig 2.7E**).

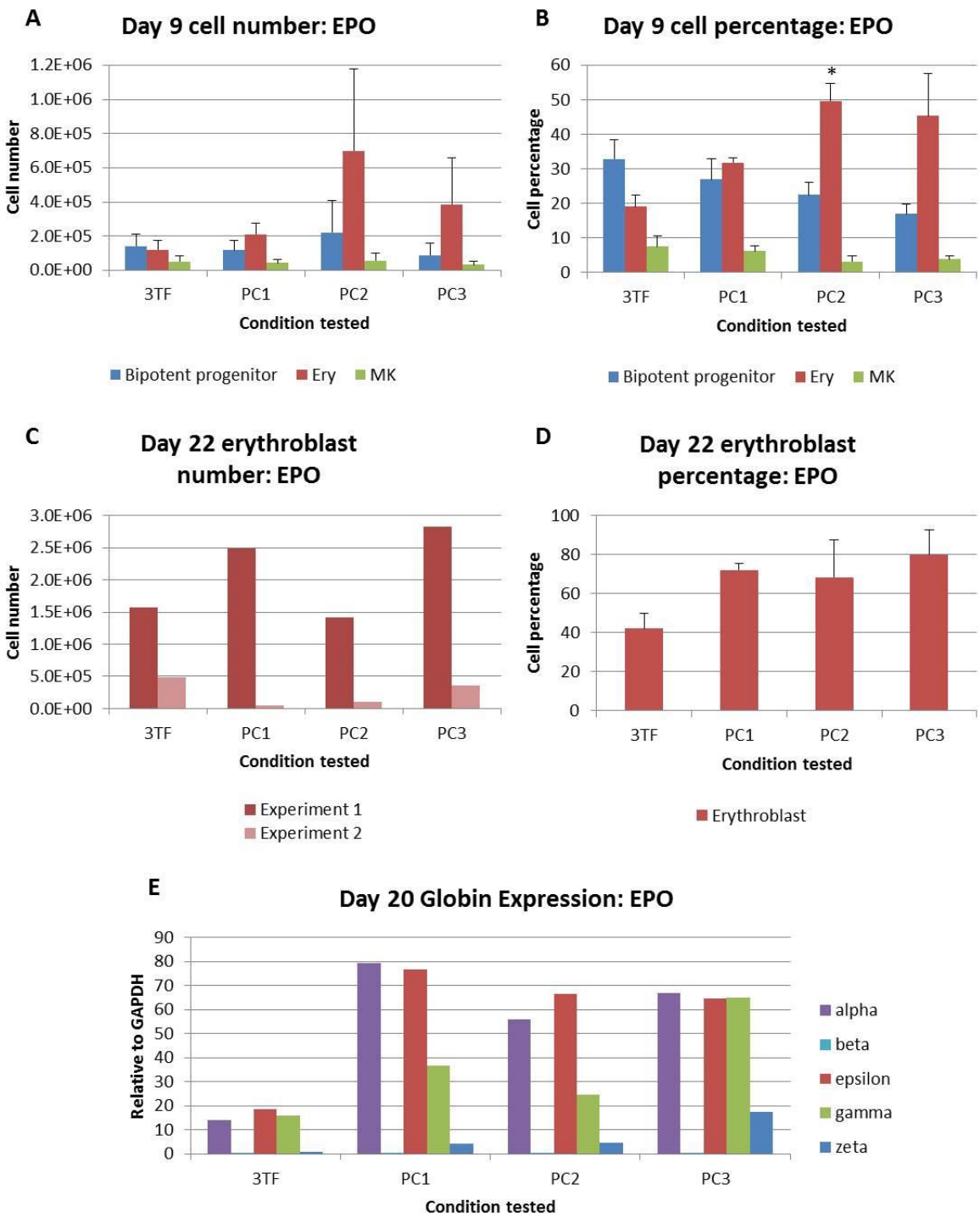


Fig 2.7 Testing the PC1, PC2 and PC3 vectors in BobC Ery-FoP. 3TF (GATA1, FLI1 and TAL1 single lentiviral vector control), PC1, PC2 and PC3 vectors were used to transduce BobC cells. After 2 days of mesoderm induction, erythroblast medium (EPO, IL-3, transferrin, insulin + SCF in CellGro) was used to culture cells for the remainder of the experiment. Cells were harvested at days 9 and 22. **A**) Day 9 cell number for CD41+/CD235+ bi-potent progenitors, CD41-/CD235+ erythroblasts and CD41+/CD235- MKs. **B**) Day 9 cell percentages of populations present in **A**. (*P*=0.04 by two tail *t*-test

versus 3TF). **C**) Day 22 cell number for erythroblasts. **D**) Day 22 cell percentages for erythroblasts. **E**) Gene expression measured by qPCR for globin genes at day 20, values shown are the relative quantification, relative to *GAPDH*. N=2 (biological replicates), error bars=range, N=1 for qPCR data and was performed on cells from experiment 1 (qPCR data generated by Dr Jose Ballester Beltran).

Fig 2.8 shows the average results for FFDK in EPO. As in TPO, FFDK produced fewer cells compared to BobC, generating noticeably less erythroblasts. The 3TF gave the highest number of cells overall, with an even distribution of cells between the three cell populations. Only PC1 produced a higher number of erythroblasts than the 3TF control (**Fig 2.8A**). Cell percentages were fairly similar for all conditions at this time point (**Fig 2.8B**). By day 22, PC2 produced the highest number of erythroblast cells in experiment 1, showing lower proliferation than in BobC. Experiment 2 generated the highest number of erythroblasts in PC1, showing much better proliferation than in BobC (**Fig 2.8C**). Erythroblast purity was high for the PC vectors, between 70-83%, similar to the outcome in BobC (**Fig 2.8D**). The expression levels of endogenous *GATA1*, *TAL1*, *FLI1* and polycistronic transgene (PC TG) were measured by qPCR for PC1 cells, at three time points (days 10, 20 and 32). Expression values are shown alongside those from erythrocytes cultured from peripheral blood derived CD34+ cells (PB cells), at day 16. *GATA1* and *TAL1* expression are similar throughout culture for PC1 cells, and are similar to PB cells. *FLI1* expression decreases from day 10 in PC1 cells, resulting in similarly low expression levels at day 32 as PB cells. PC TG expression is low throughout for PC1, with a slight increase at day 32 (**Fig 2.8E**).

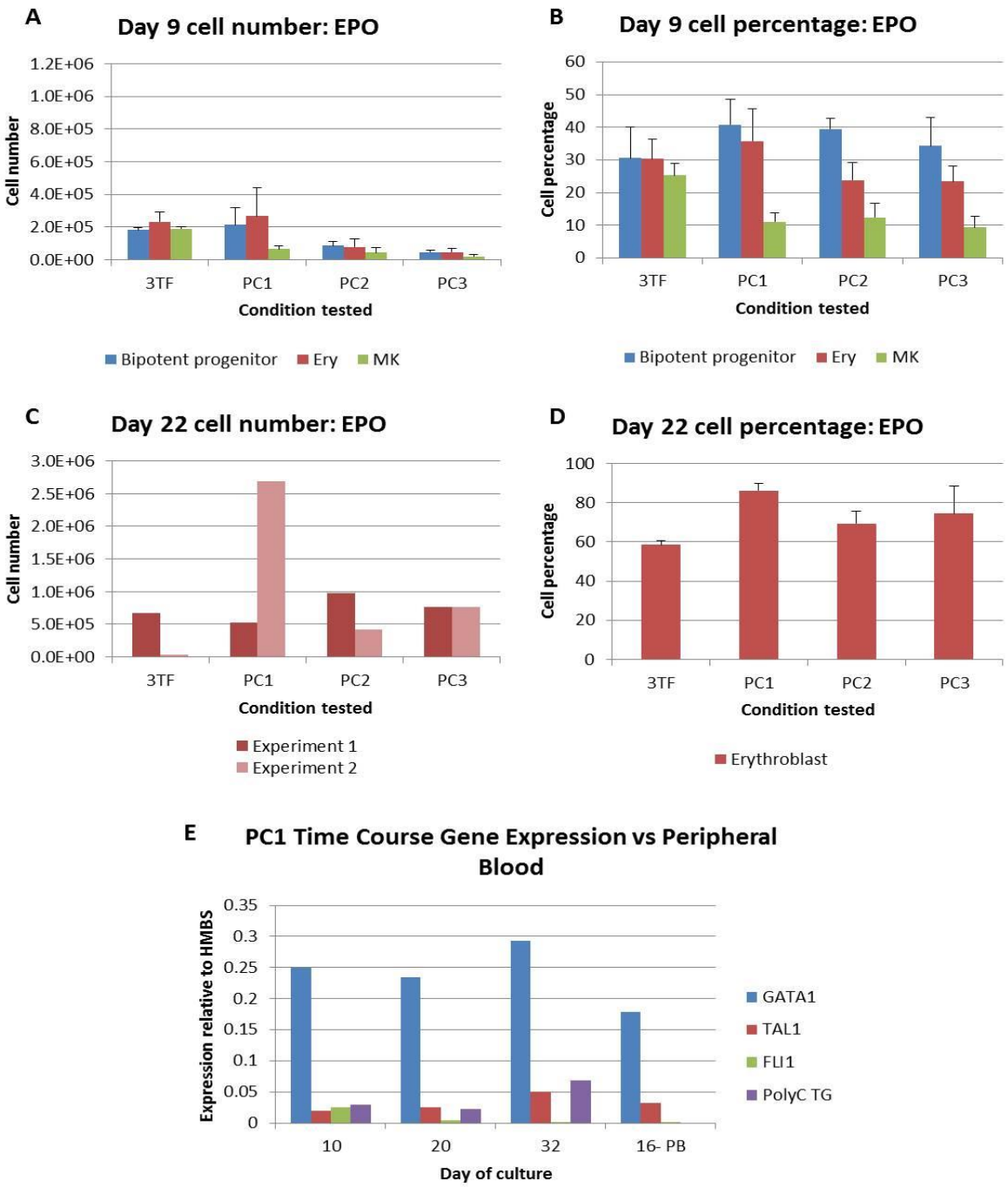


Fig 2.8 Testing the PC1, PC2 and PC3 vectors in FFDK Ery-FoP. 3TF (GATA1, FLI1 and TAL1 single lentiviral vector control), PC1, PC2 and PC3 vectors were used to transduce FFDK cells. After 2 days of mesoderm induction, erythroblast medium (EPO, IL-3, transferrin, insulin + SCF in CellGro) was used to culture cells for the remainder of the experiment. Cells were harvested at days 9 and 22. **A**) Day 9 cell number for CD41+/CD235+ bi-potent progenitors, CD41-/CD235+ erythroblasts and CD41+/CD235- MKs. **B**) Day 9 cell percentages of populations present in **A**. **C**) Day 22 cell number for erythroblasts. **D**) Day 22 cell percentages of erythroblasts. **E**) Gene expression of endogenous *GATA1*, *TAL1*, *FLI1* and polycistronic transgene (PC TG) was measured at 3 different time points for

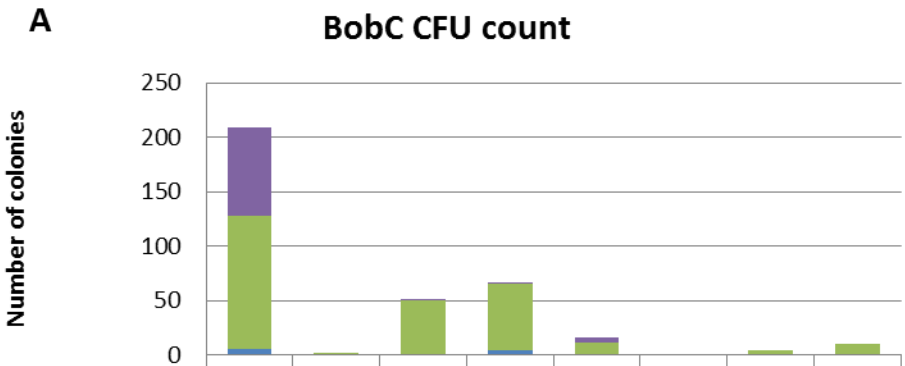
cells transduced with PC1. Relative quantification values are shown compared to erythrocytes cultured from peripheral blood CD34+ cells. N=2 (biological replicates), error bars=range. qPCR data n=1 is from a separate experiment to those shown in the rest of the figure, expression values relative to *HMBS*.

Overall, comparing TPO and EPO, these experiments show that media conditions direct cell fate to preferentially produce MKs in TPO and erythroblasts in EPO when using the PC vectors. All PC vectors produced mature MKs in TPO, with PC2 and PC3 showing a higher efficiency for MK production compared to PC1, in both iPSC lines tested. All PC vectors produced erythroblasts in EPO, with the highest erythroblast cell numbers in BobC generated by PC3, and in FFDK by PC1 and PC2. Erythroblasts produced show higher globin expression than the 3TF control, although the phenotype remains embryonic. Compared to PB derived erythrocytes, PC1 shows similar endogenous TF expression levels.

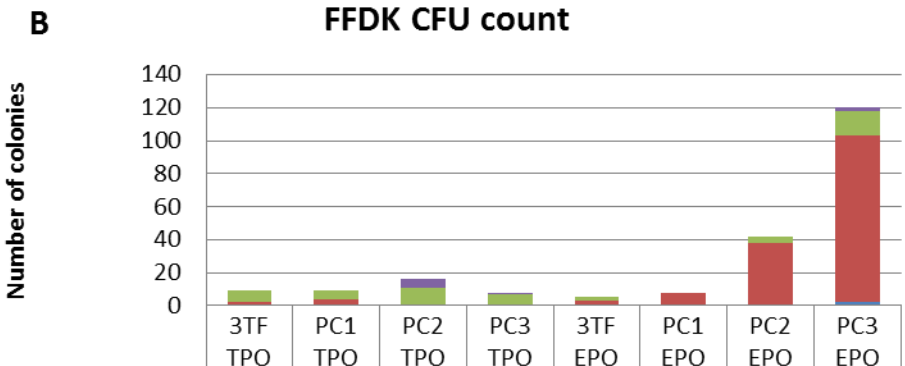
PC1/2/3 in Colony Assays: Higher MK potential in TPO

Day 9 cells from one experiment (performed in both lines in TPO and EPO at the same time), were used to perform CFU assays, to assess progenitor potential for each vector. **Fig 2.9** shows the number of colonies counted for the following colony types; mixed, CFU-E, MK and MK progenitor. In BobC, PC3 had the highest mixed, MK and MK progenitor potential of the PC vectors, with PC1 giving rise to the fewest. For the EPO conditions, the mixed, MK and MK progenitor potential was reduced for all vectors. The CFU-E counts are not shown for BobC due to counting inaccuracies resulting in unreliable data for this colony type (**Fig 2.9A**).

For FFDK, the total number of colonies counted was lower than for BobC, reflecting the trend in cell number seen in the liquid culture of this line. The MK potential was highest in the TPO PC2 condition and lowest for PC1. The number of CFU-E colonies was low in all FFDK TPO conditions. In EPO, the mixed, MK and MK progenitor potential was lower than for TPO in all conditions, apart from in PC3 which showed an increase in all colony types. CFU-E potential was increased in EPO, with PC3 giving rise to the highest number of CFU-E colonies (**Fig 2.9B**).



	3TF TPO	PC1 TPO	PC2 TPO	PC3 TPO	3TF EPO	PC1 EPO	PC2 EPO	PC3 EPO
MK progenitor	81	0	1	1	5	0	0	0
MK	122	2	50	61	12	1	5	11
Mixed	6	0	1	5	0	0	0	0



	3TF TPO	PC1 TPO	PC2 TPO	PC3 TPO	3TF EPO	PC1 EPO	PC2 EPO	PC3 EPO
MK progenitor	0	0	5	1	0	0	0	2
MK	7	5	11	7	2	0	4	15
CFU-E	1	4	0	0	3	8	38	101
Mixed	1	0	0	0	0	0	0	2

Fig 2.9 Progenitor potential of PC1, PC2 and PC3 transduced cells in MK and Ery-FoP. 3TF (GATA1, FLI1 and TAL1 single lentiviral vector control), PC1, PC2 and PC3 vectors were used to transduce BobC and FFDK cells. After 2 days mesoderm induction cells were cultured in MK (TPO) and erythroblast (EPO) media. At day 9 of FoP cells were seeded into clonogenic assays to assess progenitor potential and left for 14 days, before counting colonies. Colonies counted were scored as mixed, CFU-E, MK or MK progenitor colonies. **A**) BobC CFU colony count. **B**) FFDK CFU colony count. CFU-E missing from BobC data due to unreliable data collection. N=1.

Overall, the results of the CFU assays show in both lines that mixed, MK and MK progenitor potential is higher when cells are cultured in TPO for the first 9 days of FoP, with the exception of PC3 for

FFDK. PC1 showed the lowest MK potential overall in both lines. In FFDK the EPO culture conditions increased CFU-E potential.

Transgene Expression is Lower in PC-Transduced Cells Cultured in EPO

The level of transgene expression of cells cultured in different conditions was measured using the GFP marker of the PC vectors and gene expression by qPCR, shown in **Fig 2.10**. In day 9 cells, the 3TF showed a low level of background GFP signal. For all PC vectors, the percentage of GFP expressing cells was higher in TPO compared to EPO, seen for both lines tested (**Fig 2.10A**). In BobC PC3-transduced cells, the distribution of GFP positive cells is similar for TPO and EPO conditions, with the majority of GFP cells contributed by the MK population, and the least GFP positive cells found in the erythroblast population (**Fig 2.10B**). Looking at gene expression by qPCR, endogenous *GATA1* and *TAL1* are similar for TPO and EPO conditions. The largest difference observed for endogenous genes is in *FLI1* expression levels, which are lower in EPO cells. PC TG expression is also lower in EPO cells. This trend was observed in both cell lines tested (**Fig 2.10C**).

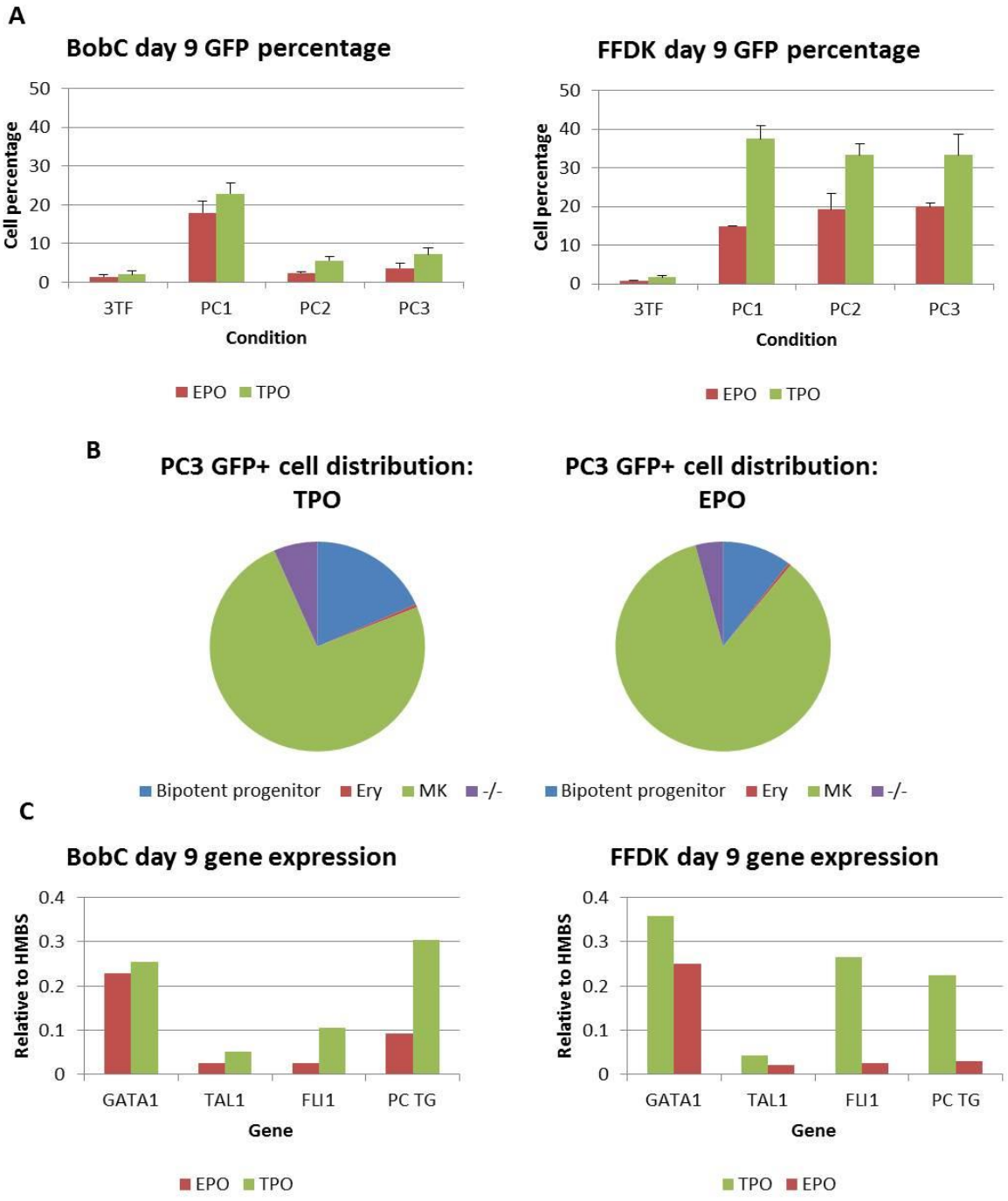


Fig 2.10 EPO media selects for cells with lower TG expression by day 9 of culture. 3TF (GATA1, FLI1 and TAL1 single lentiviral vector control), PC1, PC2 and PC3 vectors were used to transduce BobC and FFDK cells. After 2 days mesoderm induction cells were cultured in MK (TPO) and erythroblast (EPO) media. Cells were analysed at day 9 of culture. **A**) Percentage of GFP positive cells cultured in TPO and EPO, in BobC and FFDK. **B**) GFP positive cell distribution in BobC PC3 transduced cells for CD41+/CD235+ bi-potent progenitors, CD41-/CD235+ erythroblasts, CD41+/CD235- MKs and CD41-/CD235- double negative cells (-/-). **C**) Day 9 endogenous *GATA1*, *TAL1*, *FLI1* and polycistronic transgene (PC TG) expression in TPO and EPO cells from BobC and FFDK, transduced with PC vector

at day 0. Relative quantification values are shown, relative to the housekeeping gene *HMBS*. TPO: n=3 (biological replicates), error bars= standard deviation, EPO: n=2 (biological replicates), error bars= range. n=1 for **B** and **C**.

Discussion

The aim of the work presented in this chapter was to generate a single polycistronic rLV vector for use in FoP to minimise the reliance of three separate rLVs in order to; reduce procedural costs, reduce stochasticity at the start of FoP, reduce variability and generate a more homogenous population of cells. Testing the first version of the PC vector (pWPT-GATA1-FLI1-TAL1co) highlighted a bias towards the erythroblast lineage. We hypothesised that this was due to a potentially altered function of the FLI1-2A protein produced. In order to overcome this, gene order within the vector was rearranged with the generation of PC2 (pWPT-eGFP-GATA1-TAL1co-FLI1), and the *FLI1* sequence codon optimised to try and improve MK-FoP potential, generating PC3 (pWPT-eGFP-GATA1-TAL1co-FLI1co). The three PC vectors were compared to test whether any showed better potential for MK-FoP or Ery-FoP, in TPO and EPO conditions respectively. The results from two iPSC lines suggest a better MK-FoP potential with PC2 and PC3, and lowest MK-FoP potential with PC1, while all vectors show the potential for Ery-FoP. Importantly, this chapter shows the expression of the 3TFs from a polycistronic vector efficiently forward programmes iPSCs towards both the MK and erythroblast lineages, seen by >50% cells expressing the markers CD235 and CD41 a day 9, providing good evidence that an inducible iPSC line could work.

Polycistronic Vector Forward Programmes iPSCs but Evidence of a Erythroblast Bias

Initial FoP experiments performed with the PC rLV resulted in unusually high erythroblast and limited MK production, with similar results in two iPSC lines tested, BobC and S4 (**Fig 2.1** and **Fig 2.2**). These results led us to consider whether gene order in the poylcistronic vector could be affecting gene function. Of the three FoP TFs, we know from previous work that *FLI1* is the most essential for generating MKs. When the *FLI1* TG is not well transduced in standard FoP with 3 rLVs, the erythroblast developmental pathway is favoured, as *FLI1* normally inhibits this pathway (Athanasiou M, 2000). Considering the placement of the *FLI1* sequence in the PC vector, we questioned whether the 2A fragment fused to the C-terminus, resulting in a FLI1-2A protein, might be affecting FLI1 functionality, which could explain why the PC vector showed an erythroblast bias.

The C-terminus of FLI1 has been shown to be important for DNA-binding properties. For example, the fusion oncogene, *EWS/FLI1*, generated by chromosomal translocation, results in the N-terminus of the Ewing sarcoma (*EWS*) gene fused to the C-terminal DNA-binding domain of *FLI1* (Ano, 2004). The resulting *EWS/FLI1* fusion gene is an aberrant TF with potent transforming properties, responsible for over 95% of cases of Ewing sarcoma, a paediatric tumour of neuroectodermal origin (Ohno T, 1993). Looking at the PC vector in more detail, the T2A sequence downstram of *FLI1* leads

to an additional 22 amino acids being joined to the C-terminal end of the FLI1 protein sequence. We reasoned that the additional amino acids could be interfering with the normal DNA binding properties of the FLI1 C-terminal end and thus affect its functionality as a TF.

Previously, I had generated a *FLI1-ERT2* vector, where the C-terminal end of *FLI1* was fused to the eostrogen receptor that binds tamoxifen (*ERT2*), to establish whether controlling expression of *FLI1* during FoP could lead to improved erythroblast producion. However, after testing this vector several times, it failed to generate cells by FoP, suggesting that this fusion protein was not functional (data not shown). This observation added evidence to the idea that fusing another sequence to the C-terminal end of *FLI1* could impair function. While a large protein, such as ERT2, appeared to have rendered *FLI1* completely inactive due to the complete abolition of FoP, the much shorter 2A sequence must have had a more subtle effect, as FoP was still occuring and this would not be possible if *FLI1* was totally non-functional. Finally, we concluded that the most likely explanation for the efficiency of MK-FoP being reduced and Ery-FoP being increased with the PC vector, was due to the 2A sequenced fused to the C terminus of *FLI1* reducing or limiting its TF ability. In order to try and overcome this, we redesigned the PC vector, altering gene order to put *FLI1* at the final gene position. At the same time, we generated a third PC vector, with a codon optimised *FLI1* (GeneArt) in the final gene position, to try and improve MK-FoP efficicency as much as possible.

Codon Optimisation does not Impede MK Production

Testing the single gene vectors pWPT-TAL1co and pWPT-FLI1co in combinations to mimick the correct controls for the PC vectors showed that codon optimisation of these genes does not impair MK production. In BobC the combination of TAL1co and FLI1co showed the greatest increase in MK production compared to the 3TF classical FoP combination of vectors. While in FFDK the TAL1co sequence alone improved MK production. These conditions do not take into account the functional effect of 2A fusion proteins produced by the polycistronic cassette, however, this data was encouraging that the new PC3 vector may help to improve MK-FoP, in line with our hypothesis.

PC Vectors Generate MKs in Long-term BobC Cultures

Once generated the three PC vectors were tested alongside the normal 3TF FoP control in TPO and EPO. We hypothesised that in TPO, due to gene order, PC1 would have the lowest MK potential, PC2 an intermediate potential and PC3 the highest MK potential.The results of this experiment show a MK potential in all vectors in both iPSC lines tested, with lower overall MK production in FFDK (**Fig 2.5C** and **Fig 2.6C**). In line with our predicition, PC1 showed the lowest MK potential, consistently generating very few mature MKs. Both PC2 and PC3 resulted in high MK numbers, consistently

higher than the 3TF control in BobC. This shows that moving *FLI1* to the final gene position of the PC vector, removing the 2A sequence at the C-terminal end, was beneficial for MK-FoP. Importantly, the PC vectors established long-term MK cultures in BobC (**Fig 2.5E**). This suggests that FoP with the PC vectors is occuring in a similar manner to normal 3TF FoP, generating a repopulating pool of progenitor cells. In EPO, we hyptohesised that PC1 would be the best PC vector to produce erythroblasts, due to the potentially reduced function of the FLI1-2A protein produced. However, this was not the case. In BobC, PC3 performed best by producing the highest number of erythroblasts of any of the PC vectors, while in FFDK PC1 and PC2 produced the highest number in two experiments (**Fig 2.7C** and **Fig 2.8C**).

These experiments highlight an issue of working with iPSCs, that the results of different lines can vary considerably. Normally FoP in FFDK is quicker, with the appearance of mature MKs earlier than in BobC. Indeed this experiment showed a higher percentage of mature MKs at day 20 in FFDK (**Fig 2.5D** and **Fig2.6D**) but these cells failed to proliferate. Another issue raised is the variability of day 20-22 outcome from the same iPSC line. Vector batch can explain some of the observed variance, as experiment 1 was performed with lab-produced rLVs, while experiment 2 and 3 were performed with vectors produced commercially. However, this does not explain all the variation seen, as experiments 2 and 3 showed a large variation too. These differences could be due to different transduction efficicencies of the 2 experiments, which would also help to explain why the standard deviation of cell numbers at day 9 of culture was large, particularly for BobC. This may be due to differences in the cells at the time of transduction, for example if cells were in different cell cycle stages, this may effect how well they are able to be transduced. Unfortunately GFP data was not collected from earlier timepoints, which may have helped to identify experiments where transduction was good. As seen in **Fig 2.10A** the GFP percentage of cells was already low at day 9 of FoP, suggesting silencing of the TGs before this timepoint.

Poor transduction efficiency with the PC vectors may be due to their large size, >13kb, which is much larger than the size of the three single gene vectors used in 3TF FoP and may also explain why it was difficult to produce high titre virus. We did not spend time optimising the lentiviral production for the PC vectors, and a previous report describes that virus production is heavily decreased for large (13.2kb) packaged inserts (Al Yacoub N, 2007). The large vectors result in large mRNA transcripts, which themselves might be unstable and prone to degradation. The codon optimisation step performed on *TAL1* in all vectors, and *FLI1* in PC3, should result in increased mRNA stability, export and translation of these genes by removing cryptic splice sites, unwanted secondary structures,

instability motifs and improving codon usage for *Homo sapien* eukaryotic cells (Schambach A, 2013). This may to some degree explain why PC3 was most efficient in the majority of experiments.

Interestingly, in BobC, the experiment that produced few MKs in TPO but produced a high number of erythroblasts in EPO, is the result of cells transduced at the same time with the exact same rLV master mix. Conversely, the experiment which gave the opposite result, a high MK number in TPO and low erythroblast number in EPO, was also the result of cells transduced at the same time with the same rLV master mix. This suggests that MK outcome in TPO and erythroblast outcome in EPO might be closely linked and could relate to the transduction efficiency of each experiment. It is likely that a high MK and low erythroblast outcome is the result of low transduction efficiency, and the opposite result is due to a high trandscution efficiency.

CD41 Expressing Cells are Highly GFP Positive

With the addition of the marker eGFP in the PC vectors, we can assess the expression of the transgenes (TGs) in cells at day 9. In both lines tested, the GFP signal was lower for all vectors in cells cultured in EPO, compared to those in TPO, which was corroborated by TG expression data (**Fig 2.10A** and **Fig 2.10C**). It was also possible to establish the distribution of GFP in the different cell populations produced by FoP. CD41- cell populations (CD41-/CD235- cells and erythroblasts) have a low GFP signal, while CD41+ cells (bi-potent progenitors and MKs) have a high GFP signal (**Fig 2.10B**). This shows TG expression is highest specifically in CD41+ cells. EPO cells express less TG and lower levels of endogenous *FLI1* and preferetially differentiate along the erythroblast lineage (**Fig 2.10C**). FoP appears to be a highly selective protocol, only allowing cells which receive the optimal combination of all three TGs at the time of transduction to survive in culture. This data suggests highly transduced cells will preferentially differentiate along the MK lineage, or are specifically selected for in TPO at an early time-point. In EPO, the situation is reversed, with EPO selecting for poorly transduced cells or cells which have silenced the FoP TGs.

The percentage of GFP positive cells, in TPO and EPO for BobC, showed a decrease in percentage between day 9 and 22 for all PC conditions (data not shown). This suggests that overtime TGs are being silenced. Integration of rLV vectors favours gene expression hotspots, which leads often to insertions being surpressed in human ES cells at the time of transduction, before TG expression can occur (Xia X, 2007). Where integrations into transcriptionally active sites are not surpressed, they can be subjected to silencing by histone mdifications. Of relevance, the EF1α promoter has been shown to be subjected to DNA methylation at CpG islands, silencing TG expression *in vitro* (Zhang F,

2010). This is not likely to be a problem in FoP, since we know the endogenous genes are switched on early in FoP by their TG counterparts (**Fig 2.10C**).

PC Vectors Enable Bifurcation of Progenitors

By testing the PC vectors in TPO and EPO, the results seen for the 3TF condition (Results Chapter 1) were replicated in 2 lines. It was unknown whether the expression of the three factors from a polycsitronic vector would alter the progression of FoP, by not allowing cells to select the optimal combination of the three factors, which is what we expect happens when using the three separate TF vectors. However, the expression of all three factors from the same vector does allow the generation of a true bi-potent progenitor, which can differentiate into either MKs or erythroblasts. The PC1 vector shows the lowest MK potential in TPO and supports our hypothesis that FLI1-2A is less functional than normal FLI1.

As described in the introduction, vector design can play an important role in the level of gene expression in a polycistronic vector and is not neccesarily always stochiometric (Carey BW, 2011). We did not check that the expression of the three FoP TGs from the polycistronic vector was stochiometric by western blot. While we could potentially test multiple 2A sequences to find ones which deliver the optimal level of each gene, which is unknown at this stage, this work was primarily to investigate whether expression from a single vector would allow FoP to occur, in order to provide evidence for whether an inducible iPSC line would work. This work showed that reducing the rLV requirement can allow FoP to occur efficiently, as most experiments gave rise to higher cell numbers with PC vectors compared to the 3TF control; however, it is likely that we could optimise this system further.

Results

Chapter 3

Inducible Forward Programming

Introduction

Having demonstrated that the expression of exogenous *GATA1*, *TAL1* and *FLI1* by the polycistronic vectors efficiently generates MKs and RBCs, the next goal of my work was to further improve FoP technology by eliminating the use of rLVs completely. As discussed previously, the rLV aspect of the current FoP protocol introduces experimental variability due to rLV batch variation, as well as incurring large costs. Generating a stable, inducible iPSC line to FoP would remove these limitations. The practical aspect of having a line which can be maintained as any other iPSC line, allowing it to be cryopreserved, thawed and expanded with relative ease, but which can also be induced to FoP when required, will be of great benefit to researchers. An inducible line would also enable greater reproducibility between FoP experiments and allow greater control over TG expression timing.

A number of issues arise when targeting human pluripotent stem cells (PSCs), including silencing of transgenes in randomly inserted sites, thought to be mediated through methylation-dependent and independent mechanisms of integrated donor DNA (Smith JR, 2008). Silencing in PSCs, although common and a potentially significant problem for generating a stable PSC line to FoP, can be avoided by testing multiple promoters, as Andrews and colleagues demonstrated (Liew CG, 2007). They transfected four different promoters (UbiquitinC, Rosa26, CMV and CAGG) to drive eGFP expression, which demonstrated an unpredictable variation in transgene expression among different hESC lines tested. Only the CAGG promoter, linked to a polyoma virus mutant enhancer (PyF101) and IRES, achieved stable transgene expression in the absence of selection for more than 120 passages. They found that the CMV promoter was unable to generate stably transfected clones. This work demonstrates the difficulty faced in trying to generate a stably transfected stem cell line.

Another issue described is variegated transgene expression in targeted iPSCs. Pederson and colleagues described both silencing and variegation issues, when only 50% of eGFP targeted hESCs in their study showed expression, which was lost once cells were removed from selection upon differentiation (Smith JR, 2008). However, by targeting eGFP to the Adeno-Associated Virus Integration Site 1 (AAVS1) locus, using the adeno-associated virus type 2 (AAV2), the group showed reduced variegation of expression, reduced silencing tendencies for extended periods of time after selection withdrawal, and higher eGFP mRNA levels, compared to randomly integrated control cells. The AAVS1 locus, the preferred site for AAV integration, located on chromosome 19 is associated with an open-chromatin state due to the presence of a DNase I-hypersensitive site and contains native insulating sequences, protecting the site from silencing (Ogata T, 2003). Thus, the AAVS1 is now regarded as a genomic safe harbour for the integration of exogenous DNA (Sadelain M, 2011).

For this reason, we chose the AAVS1 locus as the target site for integrating our vector for producing an inducible iPSC line.

The second consideration for this work was selecting the targeting method to perform gene editing in iPSCs. In the past, achieving efficient stable transfection of human ESCs has been difficult. Electroporation techniques originally developed for mouse ES cells, which have high levels of homologous recombination (HR), had to be adapted in order to work in human ESCs. Also, screening rare targeting events was difficult, due to the cloning from single human ES cells being much less efficient than their murine counterparts (Zwaka TP, 2003). As discussed in the Introduction Chapter, several strategies now exist that have overcome these issues and have been demonstrated successfully in iPSCs; ZFNs, TALENs and CRISP/Cas9. However, several issues are still associated with targeting iPSCs including high cell mortality, due to high DNA damage sensitivity, and low transfection rates, due to low nuclease activity (Liu JC, 2014 and Hendriks W, 2016). Nuclease activity is usually determined in an easy to transfect cell line, such as HEK293T, prior to iPSC transfection to investigate efficiency of creating double strand breaks (DSBs) at a particular loci of interest (Hendriks W, 2016). However, a nuclease with high cleavage efficiency in HEK293T cells may not always introduce DSBs in iPSCs, which may be due to target locus accessibility and DNA-damage repair mechanisms differing between an immortalized cell line and iPSC line, or due to the target sequence itself (Chari R, 2015 and Liu JC, 2014).

Targeting Strategy

The strategy chosen to generate a stable, inducible iPSC to use in FoP, was to target the polycistronic vector to the AAVS1 genomic safe harbour of iPSCs, and induce TG expression using the inducible TET-ON third generation (3G) system. For the targeting of iPSCs we decided to use TALENs initially, as these had been shown to work at a similar efficiency to ZFNs, and had been well documented in their transfection of iPSCs (Hockemeyer, 2011). The AAVS1 genomic safe harbour was chosen, due to the targeting efficiency previously described and the success that had been shown using this locus to overexpress TGs in iPSCs. Similarly, the TET-ON system had also been demonstrated to work well in vectors targeted to the AAVS1 locus, so we used the most recent version available, the 3G system (Clontech). To avoid rTTA rLV transduction, we decided to incorporate the rTTA component into the same targeting vector. To perform the initial proof-of-principle experiments, the fluorescent marker H2BVenus was inserted. Initially we used TALENs to target the AAVS1 locus but for reasons described in the following results sections, we also tested ZFNs (optimized inducible overexpression Strategy- OPTi-OX) in this locus and separated the rTTA and TRE components.

Chapter Overview

This chapter will describe the work of generating the targeting vectors and overcoming several issues to do with targeting the AAVS1 locus, in collaboration with Dr Rosen, Dr Pance and Dr Carobbio (all Wellcome Trust Sanger institute, Cambridge). Finally, the targeting strategy originally devised was changed in favour of the OPTi-OX system (Bertero A, 2016 and Pawlowski M, 2017) developed in Dr Kotter's lab, in collaboration with Dr Pawlowski, Dr Bertero, Dr Ortmann, Professor Vallier and Dr Kotter (all Laboratory for Regenerative Medicine, Cambridge). The OPTi-OX system provided evidence that inducible FoP is achievable and efficient in producing MKs and erythroblasts.

Materials and Methods

Generation of Targeting Vectors

The vectors to engineer the inducible cell line were built in collaboration with Dr Barry Rosen, Dr Alena Pance, Dr Stefania Carobbio and Yalem Bekele (all Wellcome Trust Sanger Institute, Cambridge). We used the starting L1L2 vector (***Map 3.1***) and standard molecular biology methods as described in the general method section to generate all constructs. All restrictions enzymes come from NEB and have been used in optimal digestion buffers. DNA fragment and plasmid purifications through the cloning steps were achieved using Qiagen plasmid Mini, Midi or Plasmid *Plus* 96 Miniprep Kits (cat # 12123, 12143 and 16181 respectively), following standard protocols. Gibson assembly cloning kits (NEB, cat # E5510S) were performed as per manufacturer instructions. Gateway cloning was performed using Gateway LR Clonase Enzyme mix (ThermoFisher Scientific, cat # 11791019) as per manufacturer instructions. Plasmid constructs were cloned and amplified by heat transformation of the Stbl2 bacteria strains DH10B or DH5α (ThermoFisher Scientific, cat # 12331013 and 12297016 respectively). All final plasmid constructs have been verified by Sanger Sequencing of modified functional regions using primers listed in **Table 3.1**. In silico vector maps were generated using SnapGene software.

Vector 1 (all in one): AAVS1-rTTA-TRE-H2BVenus

This 'all-in-one' vector was produced to target the AAVS1 locus of human iPSCs, to insert the rTTA component of the TET-ON system under the control of the endogenous AAVS1 promoter and enable the TRE promoter to drive a fluorescent Venus marker gene in the presence of doxycycline.

The starting L1L2 vector (***Map 3.1***) for this project contained L1 and L2 Gateway sequences, to allow insertion of AAVS1 homology arms. It also contained a PGK promoter to drive puromycin (Puro) resistance, to allow selection of correctly targeted clones. This vector was double digested (RsrII/PacI) to replace the H2BVenus ORF, with a synthetic rTTA3G fragment. The rTTA fragment (minus an ATG start codon) was synthesised (GeneArt), flanked by RsrII and PacI sites, digested gel extracted and ligated with the open L1L2 vector, to produce the L1L2-rTTA vector.

The polyadenylation (polyA) signal, SV40pA (***Map 3.1***), was replaced with a stronger transcription termination signal, the synthetic 3xpolyA sequence (GeneArt).

A synthetic TRE-Sv40pA fragment (GeneArt), flanked by AscI sites was digested and ligated with the AscI digested L1L2-rTTA open vector, resulting in the L1L2-rTTA-TRE vector.

A unique restriction site (AccI), was used to insert the H2BVenus ORF, under the control of the TRE promoter. H2BVenus (***Map 3.1***) was amplified using primers with complementary ends to the AccI digested vector and inserted into the linearized L1L2-rTTA-TRE vector by Gibson assembly, resulting in the L1L2-rTTA-TRE-H2BVenus vector (***Map 3.2***).

A two-step Gateway reaction was performed on the L1L2-rTTA-TRE-H2BVenus vector and an intermediate R1R2 vector, containing AAVS1 homology arms. During the Gateway reaction, L1+L2 sites exchange material with R1+R2 sites, to produce the final vector for targeting iPSCs. The Gateway reaction was performed using 60ng of the L1L2-rTTA-TRE-H2BVenus vector and 100ng of an AAVS1- R1R2 intermediate vector (***Diagram 3.1***). The resulting AAVS1-rTTA-TRE-H2BVenus vector (***Map 3.3***) will be subsequently referred to as Vector 1.

Mouse (all-in-one) Vector: Rosa26-KrTTA-TRE-H2BVenus

A mouse vector, containing the same targeting components as vector 1 was also generated, to target mouse ES cells in the Rosa26 locus. For this mouse all-in-one vector:

1. A Kozak sequence was added before the rTTA ORF of L1L2-rTTA-TRE-H2BVenus (***Map 3.2***), generating a L1L2-KrTTA-TRE-H2BVenus vector.
2. Homology arms of the mouse Rosa26 locus added by Gateway reaction, using the intermediate R1R2_mROSA26 vector.

A three-step Gateway reaction was performed to generate the targeting mouse vector from the L1L2-KrTTA-TRE-H2BVenus. The first stage incorporates the mouse Rosa26 homology arms with a R1R2 intermediate vector. The second stage incorporates a spectinomycin ORF and diphtheria toxin under a PGK promoter, with a L3L4 vector (***Diagram 3.2***). TALEN targeting in human stem cells requires circular vectors, whereas linear DNA is used for electroporating mouse ES cells. Therefore, if the construct does not integrate into the correct locus, the diphtheria toxin is expressed and will kill cells in which integration has not occurred. 60ng of both the L1L2-KrTTA-TRE-H2BVenus and L3L4 plasmids, along with 200ng of the mRosa26-R1R2 intermediate vector were used to for perform the Gateway reaction, generating the final targeting vector Rosa26-KrTTA-TRE-H2BVenus (***Map 3.4***).

Vector 2 (all-in-one): AAVS1-EF1α-rTTA-TRE-H2BVenus

Cloning of the AAVS1-EF1α-rTTA-TRE-H2bVenus vector (***Map 3.5***) was performed by Steffani Carobbio, following similar cloning techniques to those previously described and will be subsequently referred to as Vector 2.

Vectors 3, 4 and 5: Separated rTTA and TRE Components and an All-in-one Vector

Alongside the all-in-one vector, containing both the TET components (rTTA and TRE), two more vectors were generated to separate these major components. This would enable the rTTA and TRE separate vectors to be targeted together, one to each allele of the AAVS1 locus or to different alleles if needed, should the all-in-one strategy not work. I constructed the vectors by performing Gateway reactions to insert the AAVS1 homology arms, as previously described. The resulting vectors were:

- AAVS1-Ef1α-rTTA (***Map 3.6***) will be subsequently referred to as Vector 3.
- AAVS1-KBlsc-TRE-H2BVenus (***Map 3.7***): containing the replaced Kozak-ATG-Blasticidin ORF, will be subsequently referred to as Vector 4.
- AAVS1-Ef1α-rTTA-KBlsc-TRE-H2BVenus (***Map 3.8***): all-in-one vector, all components of Vector 3 and 4 combined, will be subsequently referred to as Vector 5.

Optimised Inducible Expression System Cloning Strategy

The second part of the inducible cell line work was performed in collaboration with Dr Matthias Pawlowski, Dr Alessandro Bertero, Dr Daniel Ortmann, Dr Ludovic Vallier and Dr Mark Kotter (all Laboratory of Regenerative Medicine, Cambridge). They kindly provided the reagents for ROSA26 targeting (CRISPR/Cas9n + guideRNA vectors, plus the donor vector Rosa26-CAG-rTTA, containing a CAG promoter to drive rTTA expression, ***Map 3.9***-Vector 6) and AAVS1 targeting (ZFN vectors, and donor vector AAVS1-TRE-GFP, ***Map 3.10***-Vector 7) (Bertero A, 2016).

Generating Forward Programming Targeting Vector: AAVS1-TRE-PC3

iPSCs previously targeted with the Rosa26-CAG-rTTA vector (***Map 3.9***-Vector 6) were targeting with the AAVS1-TRE-GFP vector (***Map 3.10***-Vector 7). Targeted clones resulted in sustained expression of GFP, so Vector 7 was used to produce the final inducible vector to target the three FoP TFs to the AAVS1 locus. The GFP ORF from Vector 7 was excised with NcoI/EcoRI (***Map 3.10***). The same digest of the PC3 vector (***Map 2.5***) excised the eGFP_GATA1_TAL1CO_FLI1CO fragment, which was inserted into the open Vector 7. The resulting AAVS1-TRE-PC3 vector (***Map 3.11***) was used for targeting Rosa26-rTTA homozygous iPSC lines and will be subsequently referred to as Vector 8.

Performing Cell Experiments

Testing the Mouse Vector

Mouse embryonic stem cells of the JM8.N4 cell line, derived from the mouse strain C57BL/6N (Pettitt, 2009), were electroporated with the linearized (AsiSI) mRosa26_all-in-one vector. 1 million

cells in suspension were electroporated with 1-2µg donor DNA using the Amaxa 4D (Lonza). Following electroporation, positive cells were selected with 0.8ug/ml puromycin for two weeks. Venus expression was induced by addition of 500ng/ml doxycycline. This work was performed by Dr Alena Pance.

Electroporating Human iPSCs

iPSC cells were washed twice with PBS and treated with TrypLE for 5 min, at 37^0C, 5% CO_2. Excess TrypLE was aspirated carefully by keeping the plate flat, so as not to disturb colonies. Basal media was added and cells re-suspended three times to give a single-cell suspension. Cells were stained with Trypan Blue and counted on a haemocytometer. 0.5E+06-1E+06 cells were added to a 15ml falcon tube, topped up to 10ml with wash media before centrifuging at 300g, for 5 min. Cells were re-suspended in DNA mix (18µl Solution 1, 82µl Solution 2 (both Human Stem Cell Nucleofector Kit 1, Lonza), 2µg vector DNA, 4µg/TALEN, ZFN or CRISPR/Cas9n + gRNA vector DNA). The DNA + cell mix was transferred to a cuvette (Nucleofector Kit 1) and electroporated using the B-016 cycle on the Amaxa Nucleofector II (Lonza). Following electroporation, 500µl iPSC media (AE6+FGF+Activin), supplemented with ROCK Inhibitor (1000x) for the first 24 hours, was added to cells. A Pasteur pipette (Nucleofector Kit 1) was used to split cells equally into a 6 well plate, pre-treated with vitronectin. For lines targeted with both Vector 6 (Rosa26-CAG-rTTA) and Vector 8 (AAVS1-TRE-PC3) vectors, Vector 6 was targeted initially, selected and genotyped, before homozygous clones were targeted with Vector 9.

An experiment was also performed with the Amaxa 4D-Nucleofector X Unit using the P3 Primary Cell 4D-Nucleofector X Kit L (Lonza), following the same protocol, using machine setting CA 137.

iPSC Lines Used

The lines BobC, S4 and Qolg, lines which FoP well, were used to test initial vectors (Vectors 1-5) before changing targeting strategy. At the time of changing strategy, the lines H9_Rosa26_rTTA, BBNX_Rosa26_rTTA and Bob_Rosa26_rTTA were shared with us by Dr Pawlowski, which had already been targeted in the Rosa26 locus with Vector 6 (Rosa26-CAG-rTTA, **Map 3.9**) and genotyped. After this, the lines FFDK and BobC were used to test new vectors, as these lines are known to FoP well and S4 was no longer being routinely cultured in the lab at the time.

Transduction with rTTA Lentivirus in Vector 1-targeted iPSCs

Polyclonal wells of selected iPSCs were transduced with M2rTTA rLV virus (MOI 20) in stem cell media supplemented with 10µg/ml protamine sulphate, following normal transduction procedure

(previously described). Cells were washed after 24 hours with PBS and doxycycline added (1µg/ml) to induce Venus expression.

293T Cell Transfection with Vectors 3, 4 and 5

0.5E+06 293T cells were seeded/well of a 6 well plate and incubated for 24 hours at 37^0C, 5% CO_2. 3µl Turbofect and 2µg DNA were added to 100µl 293T media, briefly vortexed then incubated for 30mins at RT, before being added to cells. After 24 hours, cells received fresh media, supplemented with 1µg/ml doxycycline. After 48 hours fluorescence was checked under a fluorescent microscope.

Selection of Correctly Targeted iPSCs

Kill curves for each selection agent were performed on each iPSC line prior to targeting. After splitting iPSCs into a 6 well plate, antibiotic selection across a range of concentrations was performed. Initial concentrations to test were chosen based on literature reviews of the same selective agent in a similar cell type, where possible. Cells were maintained in different concentrations for 7 days and the lowest concentration to result in complete cell death by day 7 was chosen as the concentration to use on targeted cells.

After electroporation, cells were grown in iPSC media until a confluency of ~75%. Selection was maintained until clones were large enough to split. Different antibiotics were used during different stages of the cloning strategy, which are summarised in **Table 3.2**.

Splitting Selected Clones

Several methods of splitting individual colonies were tried, the following method was found to be the most effective with the highest rate of clone survival. 1mg/ml Collagenase Type IV and Dispase II (both Life Technologies) were added to cells and incubated at 37^0C, 5% CO_2 checking at regular 10min intervals, until the edges of a colony appeared to be loose under the microscope. Then, in the incubator, the colony was gently lifted with a P200 and aspirated into a 1.5ml Eppendorf. Wash media was added to fill the Eppendorf and left for 10min, to allow the colony to settle. The colony was washed twice and re-suspended in 1ml stem cell media. Each colony was split into 2 wells of a 12 or 24 well plate (pre-coated with vitronectin). Clones were expanded, before collecting one well for gDNA and subsequent PCR genotyping. The second well was maintained in culture or cryopreserved, until genotyping results were obtained.

PCR Genotyping

gDNA was collected in 120µl lysis buffer (Wizard SV Genomic DNA Purification System, Promega) and purified following the manufacturer's protocol. Where possible, genotyping PCRs were performed in the following order, to minimise the number of PCRs required:

1) Genomic locus PCR: Clones which were negative for the WT allele were selected for genotyping fully, while any clones with the WT allele still present (not targeted, or heterozygous for the insert) were discarded.
2) Off-targets PCR: At this stage, any clones positive for an off-target integration were also discarded. This was tested for by using a primer within the donor vector sequence, and one outside of the sequence to be targeted. Any positive bands for this PCR suggest the targeting has not occurred correctly by integrating part of the sequence which should not be present in the genome. This is a quick way of detecting off-targets but will not detect all off-targets, which can only be accurately discovered by sequencing the genome of targeted cells. Sequencing was not performed on any of the targeted clones presented in this Chpater.
3) 5' and 3' integration PCRs: Any clones that did not produce the correct band for 5' and 3' integration were also discarded.

The chosen clones therefore had correct insertion at the 5' and 3' site of integration, were homozygous for the insert, and had no off-targets detected. These were then expanded in culture to cryopreserve larger stocks, before inducible FoP was performed. For clones targeted with Vector 5 the 5' and 3' genotyping PCRs were optimised, the primers used are summarised in **Table 3.3**. Genotyping for Rosa26 and AAVS1 inserts were optimised for clones targeted with iOX Vector 6 (Rosa26-CAG-rTTA) and Vector 9 (AAVS1-TRE-PC3), the primers used are described in **Table 3.4**. All PCRs were performed with LongAmp Taq polymerase (New England BioLabs). The conditions are described in **Table 3.5**, and thermocycling conditions in **Table 3.6**.

Inducing Expression of Transgenes

Venus and GFP expression was induced for a minimum of 48 hours using 1µg/ml doxycycline (Doxycycline hyclate, Sigma-Aldrich) before fluorescence was detected.

Inducible Forward Programming

Inducible forward programming was achieved by the addition of doxycycline (1ug/ml, unless otherwise stated), every 2 days. The normal forward programming protocol was followed, without the use of lentiviral vectors and the associated wash steps. iPSCs were split into small clumps of

approximately 1xE5 cells per well, or single cells using TrypLE (previously described), and allowed to recover for 24 hours. To initiate forward programming, doxycycline and mesoderm media were added and left for 2 days, before being replaced with MK/RBC media, plus doxycycline. A high TPO concentration was used (100ng/ml), until day 8-10 of the protocol, when cells were split using TrypLE. After splitting a low TPO concentration was maintained (20ng/ml). Fresh media and doxycycline was added every 2-3 days. Flow cytometry was performed periodically, as previously described.

Flow Cytometry

Since the inducible lines contain GFP TG, they were stained without a FITC-conjugated antibody. As for PC-transduced cells, inducible cells were stained with CD41a-APC-H7 (1:100), CD42b-APC (1:20) and CD235a-PE (1:200). Erythroblast characterisation on inducible cells was performed with the following antibody combination; CD36-PE (1:10), CD71-APC-H7 (1:100), CD235a-APC (1:200). (All antibodies supplied by BD-Pharmingen).

Platelet-like-particle Assay

MKs and platelet-like-particles (PLPs) were collected directly from the culture dish, after gentle swirling of the culture and homogonisation with a P1000. 100µl of cells was added to $1/9^{th}$ volume of Anticoagulant citrate-dextrose (ACD, Sigma) and 100µl of ACD-cell mixture used per flow tube. Calcein-AM (1:20,000), CD41-APC-H7 (1:200) and CD42a-APC (1:100) were used to stain cells, mixed by gently flicking the tube. Samples were incubated at 37^0C for 20 mins in an Eppendorf ThermoMixer Temperature Control Device (Eppendorf cat#:5382000023). Finally, 1ml RT PBE and 10,000 count beads were added per tube. Flow analysis was performed as previously described.

Materials and Methods Tables

Table 3.1. Primers used for generating and sequencing Vectors 1-5

Oligo/Primer name	Sequence 5'-3'
Used for producing Gibson PCR product (Vector 1)	
TET_H2bV_Fo	TACCACTTCCTACCCTCGTAAAGTCTGACTAGGAGGCCACCATG
TET_H2bV_R	TGCAGCTTGAACTGAACTCCTGCAGCTCTTCTTACTTGTACAGCTC
Used for sequence verifying Vectors 1,2,3,4,5 and mouse Rosa26 vector	
R1R	CGTGGTATCGTTATGCGCCT
R2R	TCTATAGTCGCAGTAGGCGG
R3	GCGGATAACAATTTCACACAGGA
R4	TGTAAAACGACGGCCAGT
PGKR	CCATTGCTCAGCGGTGCTGTCC
H2bR	CCTTAGTCACCGCCTTCTTG
Fchk2F	GTATCTGCAACCTCAAGCTAGC
PGKPuropA F	TGTGGTTTCCAAATGTGTCAG
Fchk2R	GCTAGCTTGAGGTTGCAGATAC
PuroF2	CCATGACCGAGTACAAGCCCACG
L1L2 FchkF	CGGATTTGAACGTTGTGAAG
PGKPuro 3248R	ATAGCAGCTTTGCTCCTTCGCTTTC

Table 3.2 Antibiotic concentration used on different iPSC lines.

Cell line tested	Puromycin (µg/ml)	Neomycin (µg/ml)	Blasticidin (µg/ml)
Bob	1	-	5
BobC	1	250	5
FFDK	0.5	200	2
S4	1	-	5
Qolg	-	-	2
BBNX_rTTA	0.5	-	3
Bob_rTTA	-	-	3
H9_rTTA	-	-	2.5

'-' Indicates selective agent not tested in that line.

Puromycin (Puromycin dihydrochloride from *Streptomyces alboniger*, Sigma-Aldrich), Neomycin (G418 disulphate salt solution, Sigma-Aldrich), Blasticidin (Blasticidin S hydrochloride, Fisher Scientific).

Table 3.3 Vector 5 targeted iPSCs, genotyping primers for 5' and 3' integration in AAVS1.

AAVS1		Wild-type	Correct Insert	Annealing temp
5'-end integration				
R1R	TCCGCGTATTGCTATGGTGC	none	2883 bp	58
p'2	GCCACGAAAACAGATCCAGG			2' 30" ext
3'-end integration				
R2R	TCTATAGTCGCAGTAGGCGG	none	1137 bp	57
d'3	CCTACTCTCTTCCGCATTGG			1' ext

Table 3.4 Vector 6 and Vector 9 targeted iPSCs, genotyping primers for integration in AAVS1 and Rosa26.

AAVS1 Genomic Locus		Wild-type	Correct Insert	Annealing temp
F_ genome	CTGTTTCCCCTTCCCAGGCAGGTCC	1692 bp	None- too large	65
R_genome	TGCAGGGGAACGGGGCTCAGTCTGA			1'30" ext
5'-end integration				
F_(genome)	CTGTTTCCCCTTCCCAGGCAGGTCC	none	991 bp	65
R_(puro)	TCGTCGCGGGTGGCGAGGCGCACCG			1' ext
3'-end integration				
Fli1CO_seqi_2Fo	GCAGCAGACTGGCCAACCCT	none	1700 bp	64
R_(genome)	TGCAGGGGAACGGGGCTCAGTCTGA			1' 30'' ext
Off-target integration			Off targets	
OptTal1_TG_Fo	ACCGCCAGATCTCTGCATC	none	2748 bp	57
AVS1 off-target_R	ATGCTTCCGGCTCGTATGTT			2' 45'' ext
ROSA26 Genomic Locus		Wild-type	Correct Insert	Annealing temp
F_ROSA	GAGAAGAGGCTGTGCTTCGG	2186 bp	None- too GC rich	63
R_ROSA	ACAGTACAAGCCAGTAATGGAG			3' ext
3'-end integration				
F_ROSA_rTTA	GAAACTCGCTCAAAAGCTGGG	none	1895 bp	54
R_ROSA	ACAGTACAAGCCAGTAATGGAG			1'50" ext
5'-end integration				
F_ROSA	GAGAAGAGGCTGTGCTTCGG	none	1274 bp	60
R_(ROSA vector)	AAGACCGCGAAGAGTTTGTCC			1'20" ext
Off-target integration			Off targets	
F_ROSA_rTTA	GAAACTCGCTCAAAAGCTGGG	none	1794 bp	55
ROSA off-target_R	TGACCATGATTACGCCAAGC			1'45" ext

Table 3.5 Genotyping PCR conditions used on targeted iPSCs.

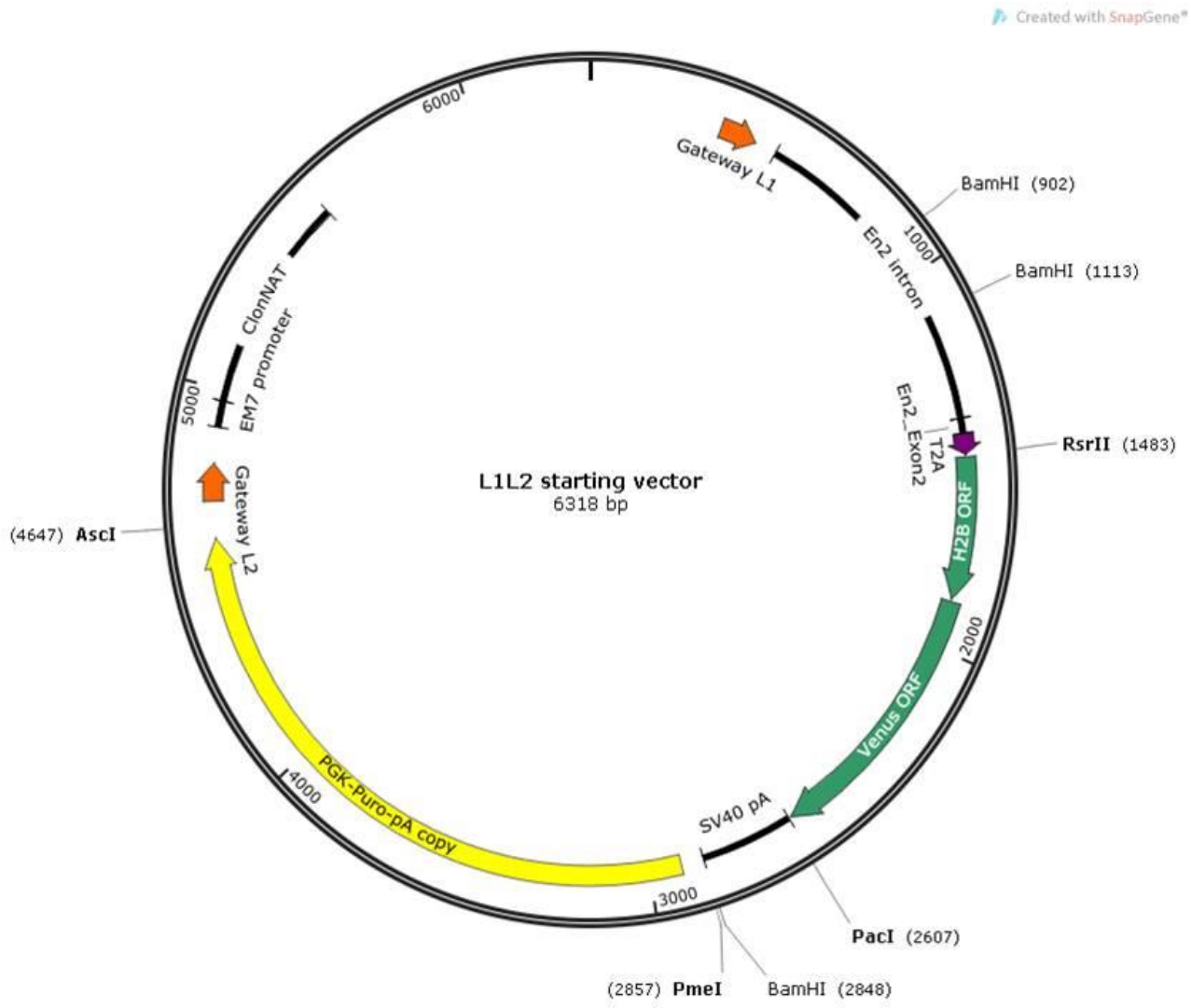
Component	**All genotyping PCRs 25 µl reaction**	**Rosa26 Genomic locus only reaction**	**Final Concentration**
5X LongAmp *Taq* Reaction Buffer	5 µl	2µl	1X
10 mM dNTPs	0.75 µl	-	300 µM
10µM dNTPs	-	0.3µl	300 µM
10 µM Forward Primer	1 µl	-	0.4 µM (0.05–1 µM)
5 µM Forward Primer	-	1µl	0.4 µM (0.05–1 µM)
10 µM Reverse Primer	1 µl	-	0.4 µM (0.05–1 µM)
5 µM Reverse Primer	-	1µl	0.4 µM (0.05–1 µM)
Template DNA	variable	variable	100-1000 ng
LongAmp *Taq* DNA Polymerase	1 µl	0.4µl	5 units/50 µl PCR
DMSO	-	0.2µl	
Nuclease-free water	to 25 µl	up to 10µl	

Table 3.6 Cycling conditions used for genotyping PCRs on a thermocycler (Thermo Fischer Scientific).

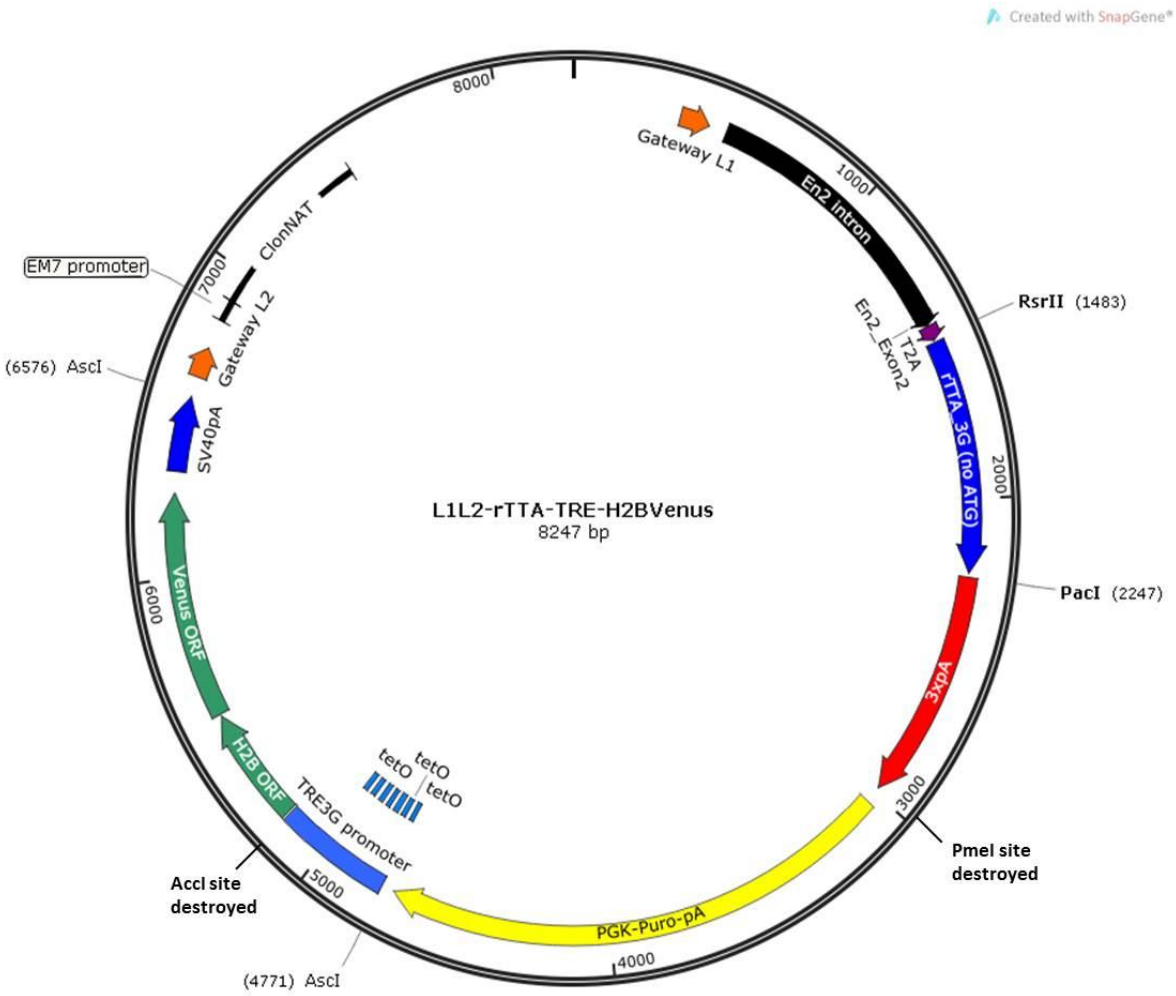
Vector 5 targeted cells			
Step	Temp (^{o}C)	Time	Notes
Initial denaturation	94	4'	
5 cycles	94	30"	-1^{o}C/cycle
	variable	30"	
	65	50"/kb	
30 cycles	94	30"	
	variable	30"	
	65	50"/kb	
Final extension	65	10'	
Hold	4		

Rosa26 Genomic locus PCR				All other PCRs		
Step	Temp (^{o}C)	Time		Temp (^{o}C)	Time	
Initial denaturation	94	5'		94	30"	
Cycles	94	15"	25x	94	30"	30x
	63	30"		variable	60"	
	65	3'		65	50s/kb	
Final extension	65	5'		65	10'	
Hold	4			4		

Chapter 3: Vector Maps



Map 3.1 L1L2 starting vector for cloning project. Restriction enzyme sites used for cloning steps are shown.



Cloning steps to generate Vector 1 from L1L2 starting vector (Map 3.1)

Digest:	Removed:	Inserted:	Notes:
RsrII/PacI	H2BVenus ORF	rTTA (no ATG)	
PacI/PmeI	SV40pA	3xpA	PmeI site destroyed
AscI	-	TRE-SV40pA	
AccI	-	H2BVenus ORF	AccI site destroyed

Map 3.2 L1L2-rTTA-TRE-H2bVenus vector (final vector before Gateway reaction).

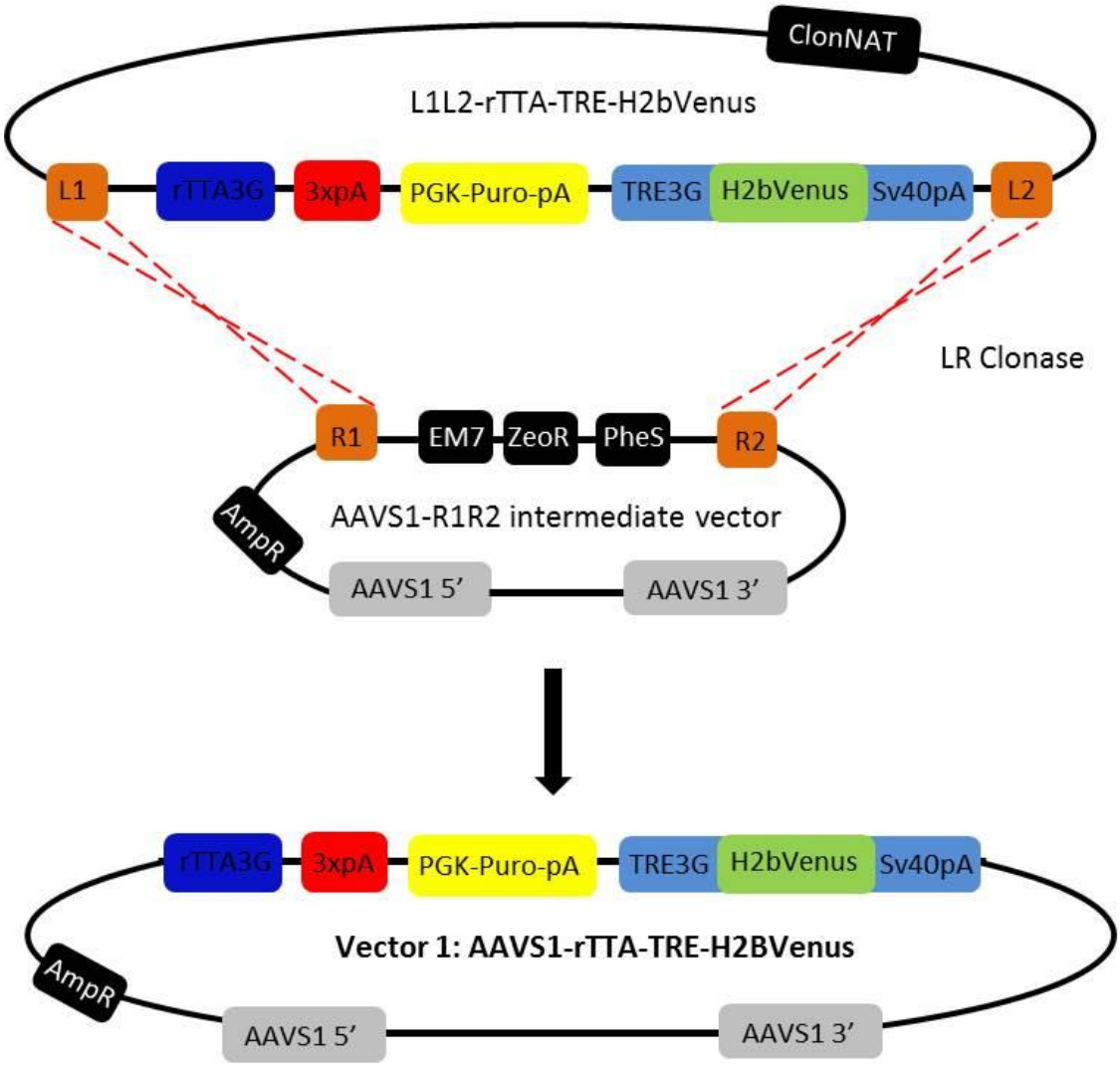
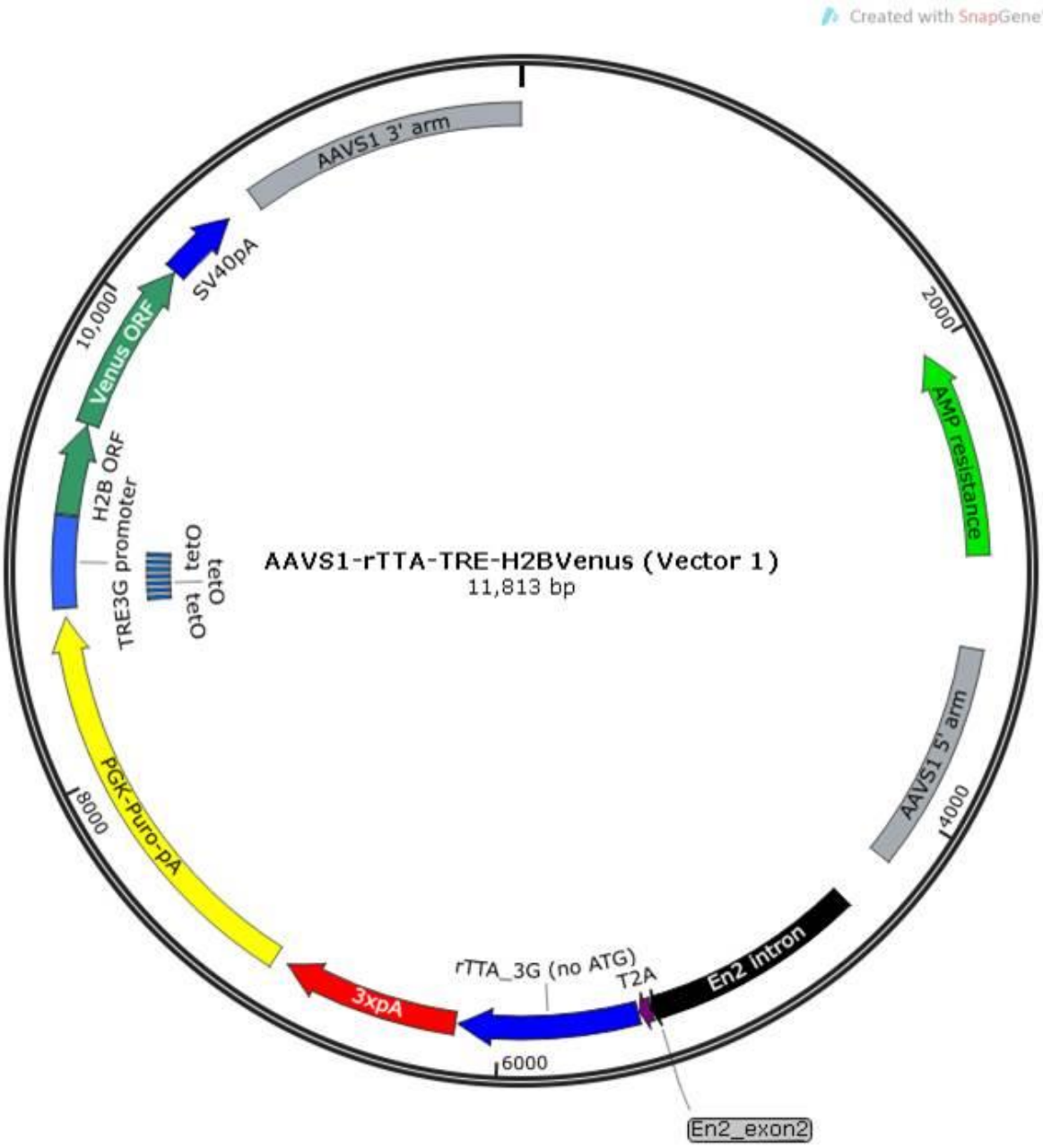


Diagram 3.1 2-way Gateway to generate final Vector 1 using L1L2-rTTA-TRE-H2BVenus and AAVS1-R1R2 intermediate vectors.



Map 3.3 Vector 1: AAVS1-rTTA-TRE-H2BVenus final targeting vector.

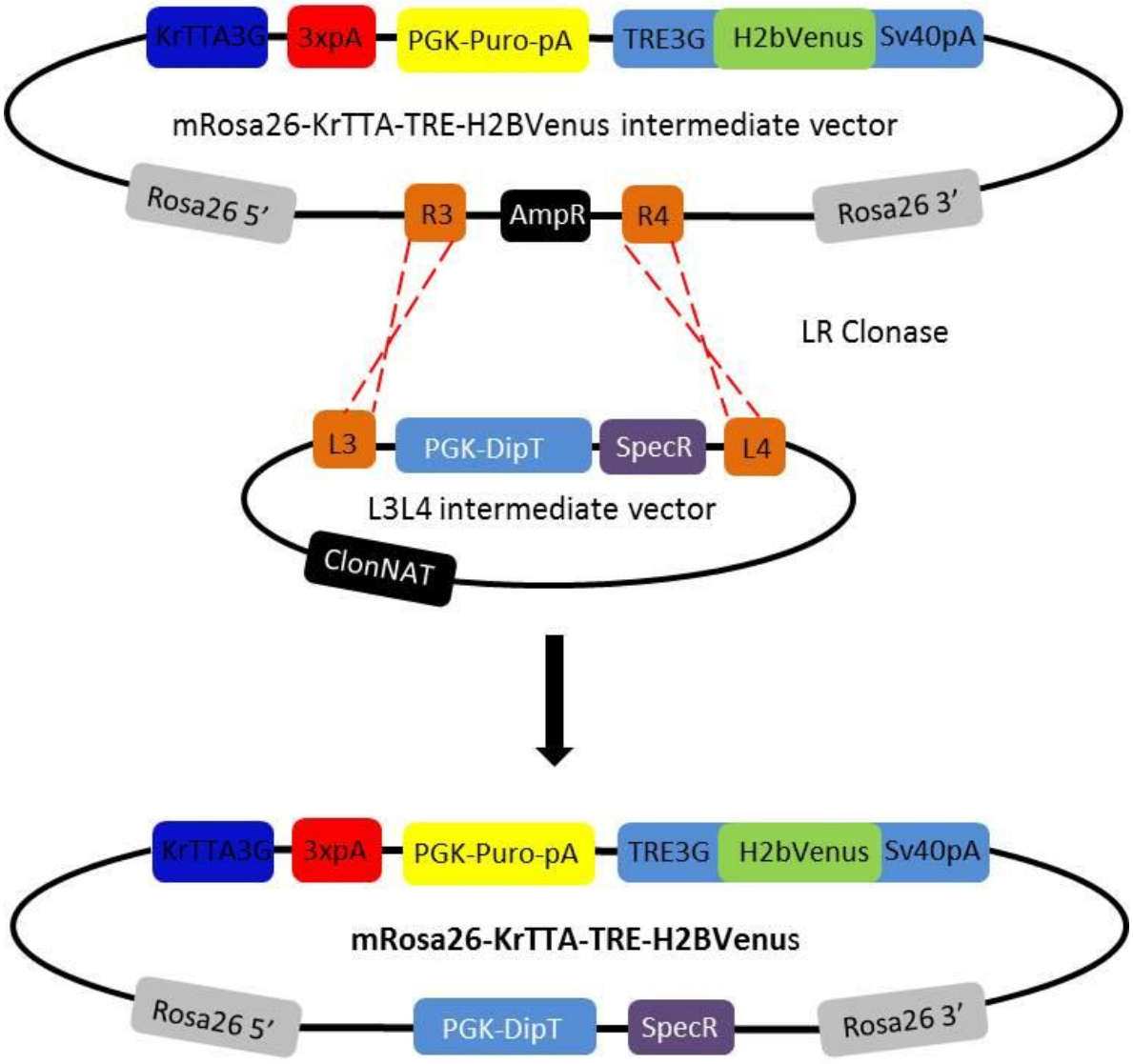
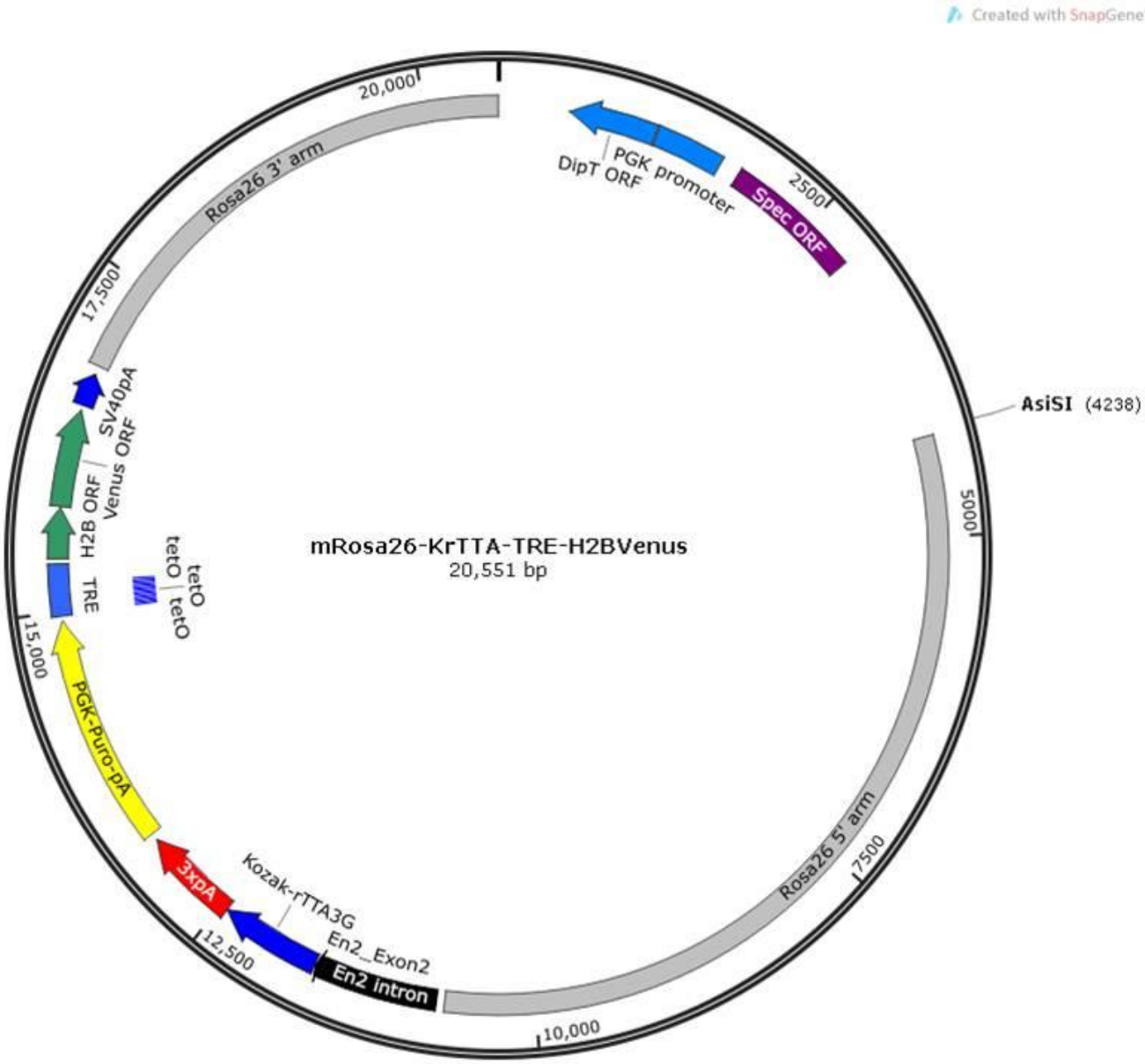
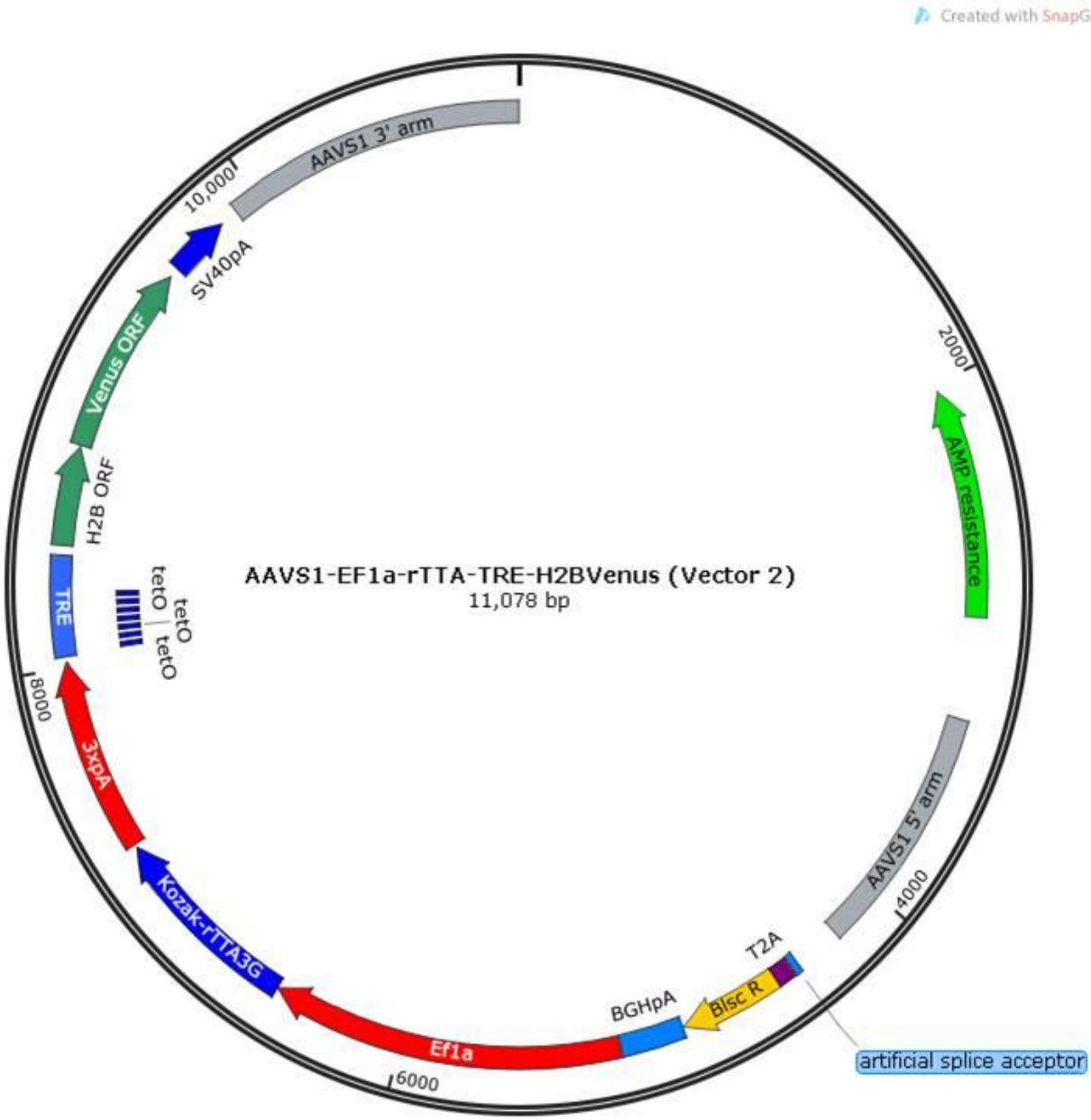


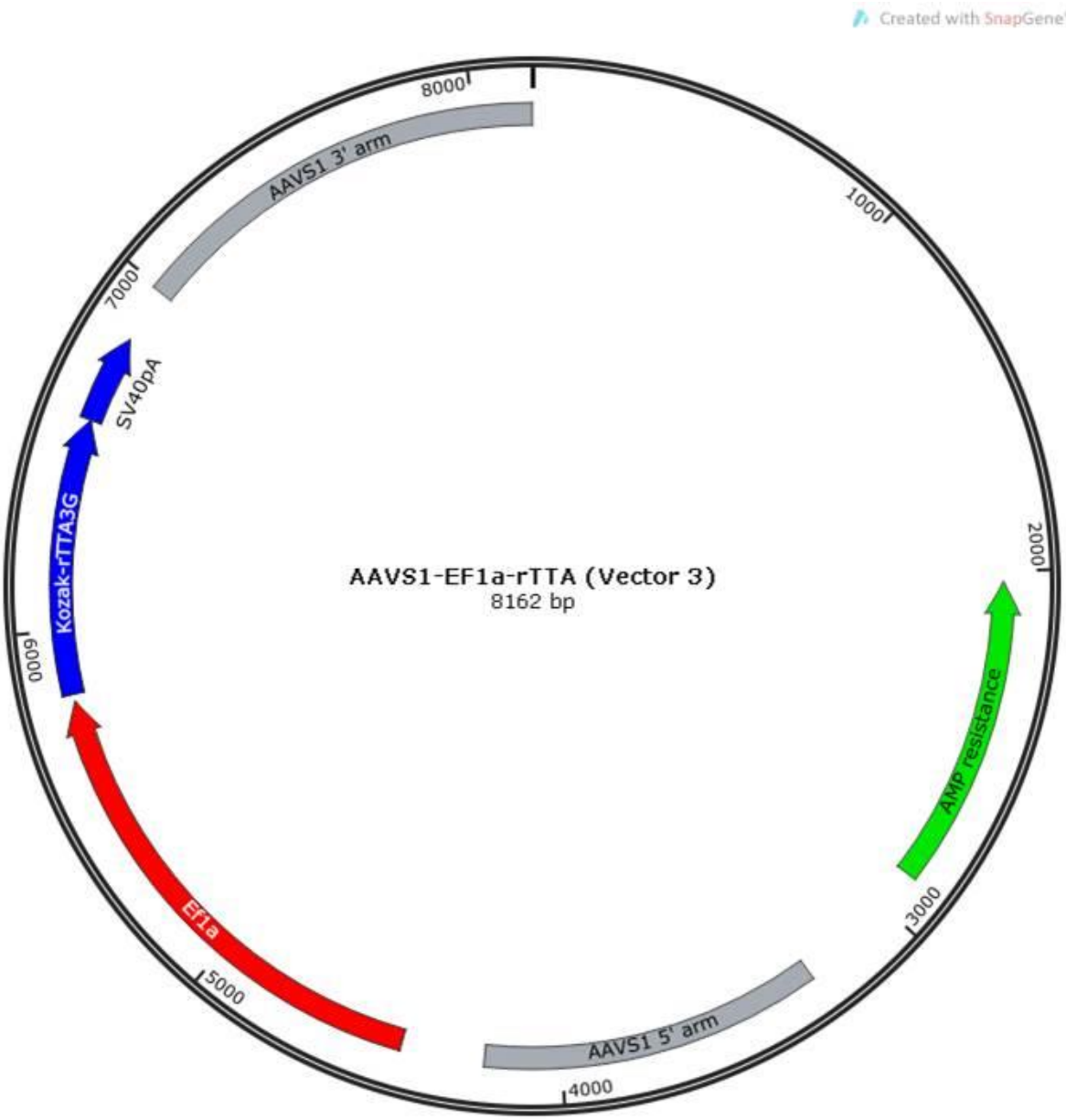
Diagram 3.2 The second stage of the 3-way Gateway Reaction to generate Rosa26 mouse targeting vector. The first stage of the gateway reaction is the same as that shown in Diagram 3.2 (not depicted) but with the L1L2-KrTTA-TRE-H2BVenus vector generated for mouse cell work.



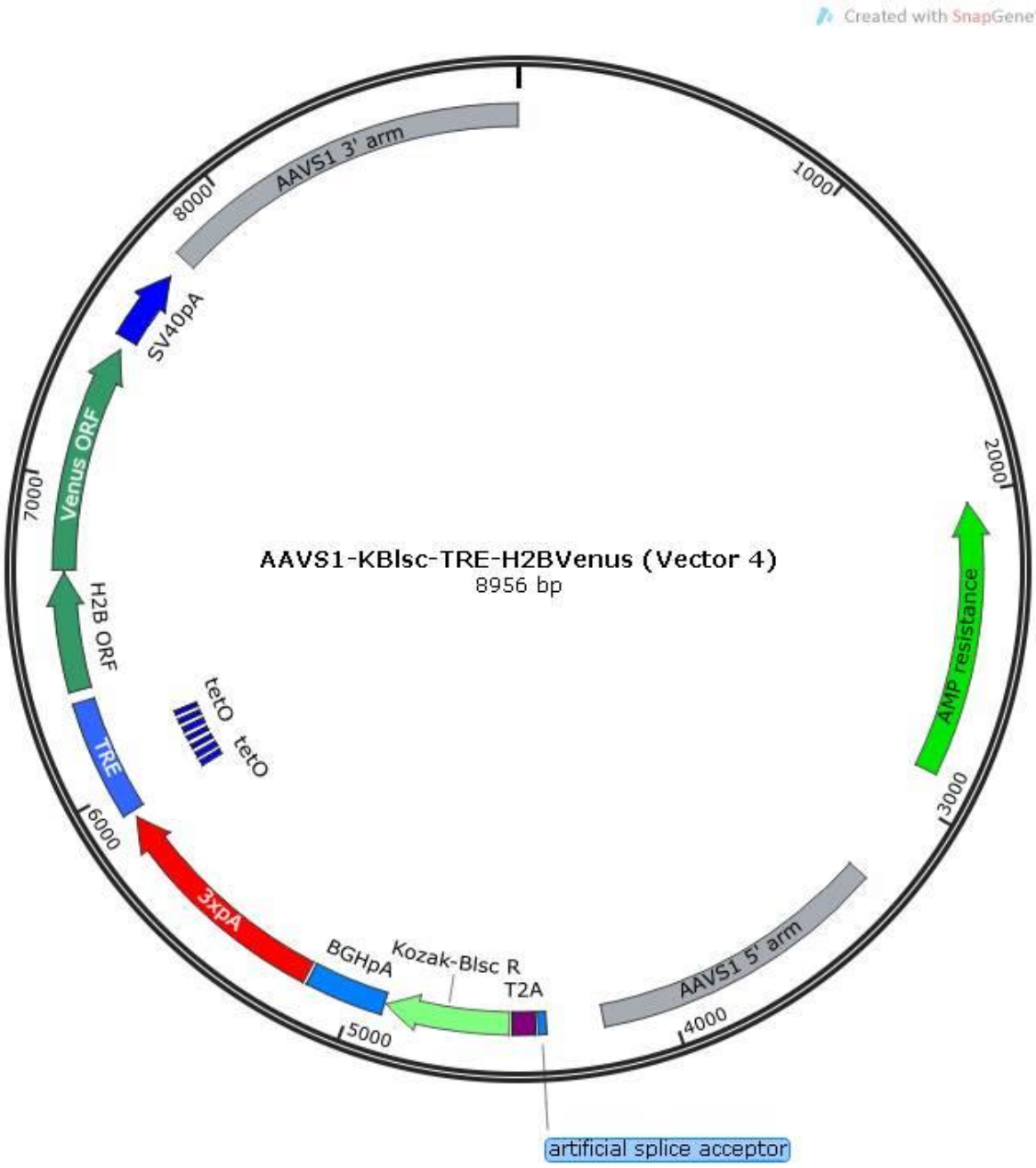
Map 3.4 Final mouse targeting vector Rosa26-KrTTA-TRE-H2BVenus. AsiSI restriction site used to linearize vector prior to mouse ES cell testing.



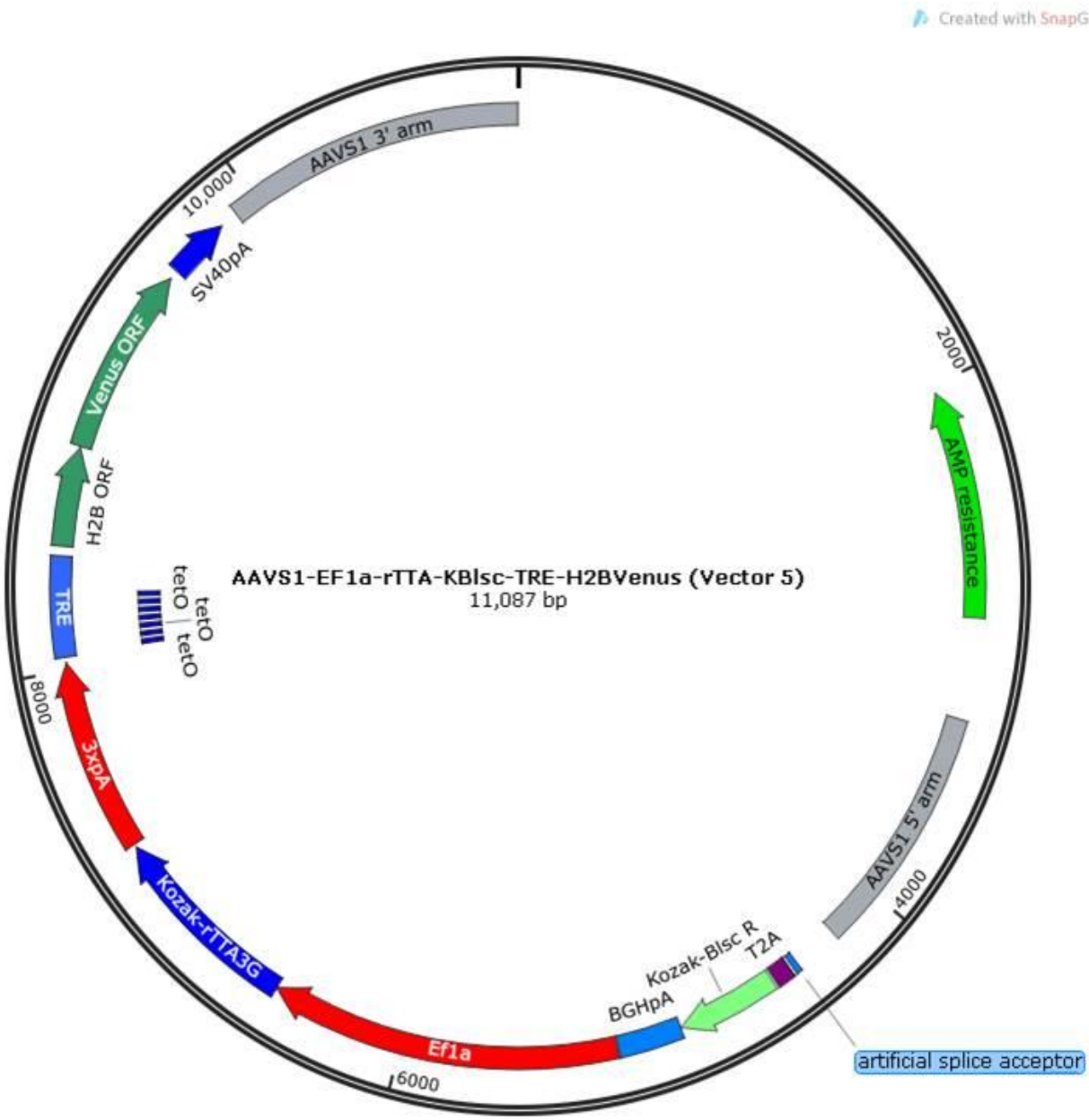
Map 3.5 Vector 2: AAVS1-Ef1α-rTTA-TRE-H2bVenus vector.



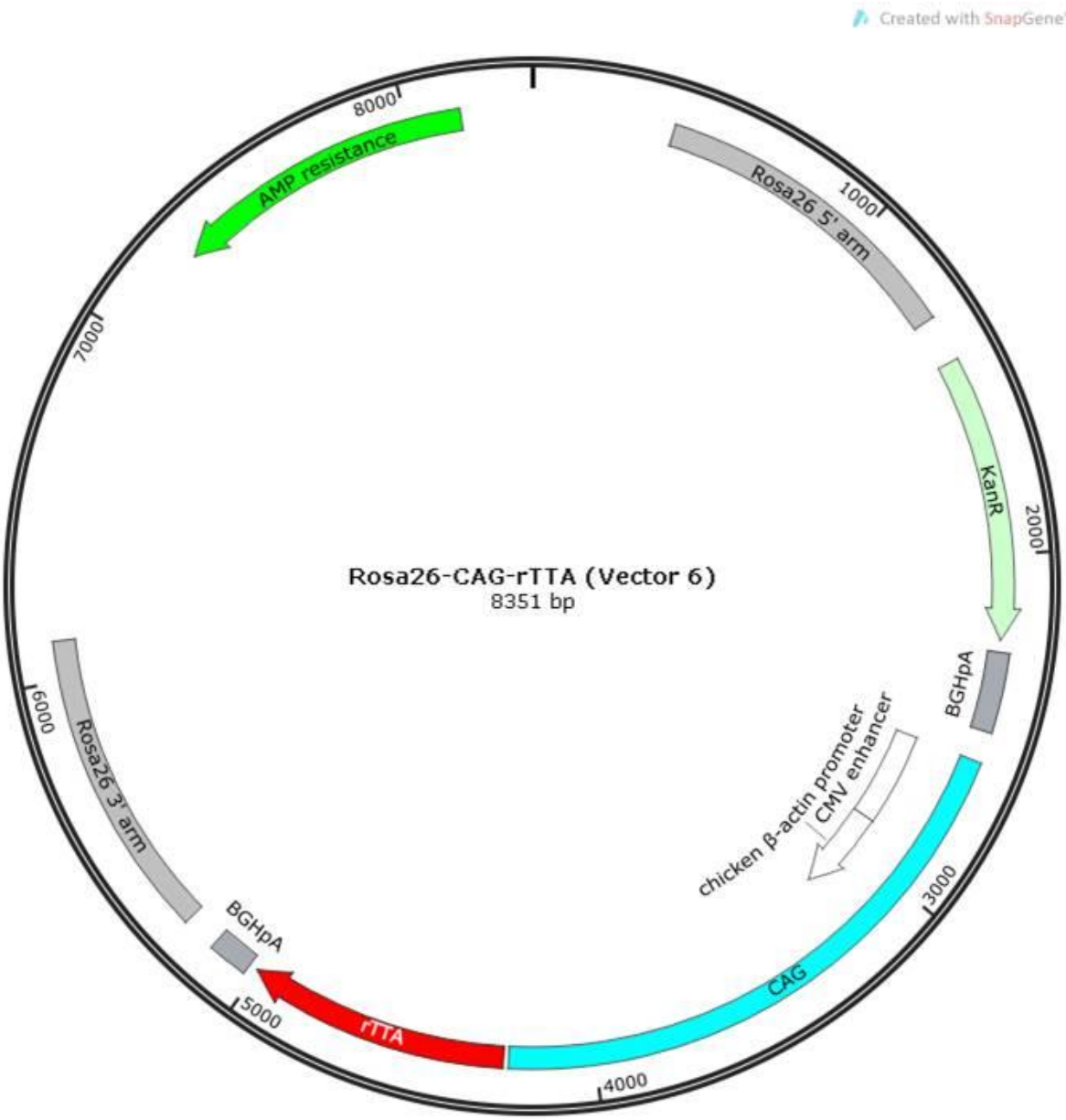
Map 3.6 Vector 3: AAVS1-Ef1α-rTTA.



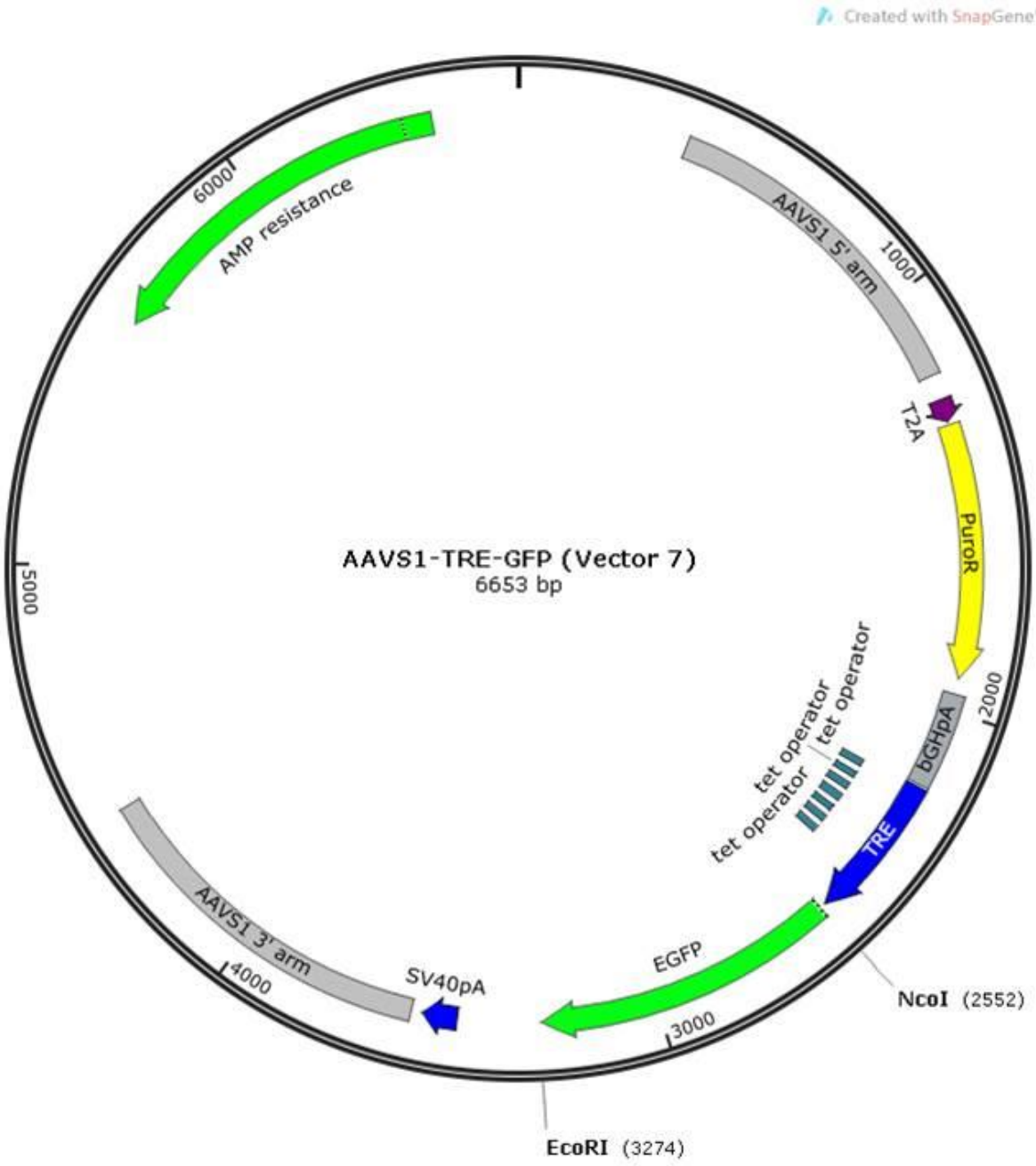
Map 3.7 Vector 4: AAVS1-KBlsc-TRE-H2BVenus.



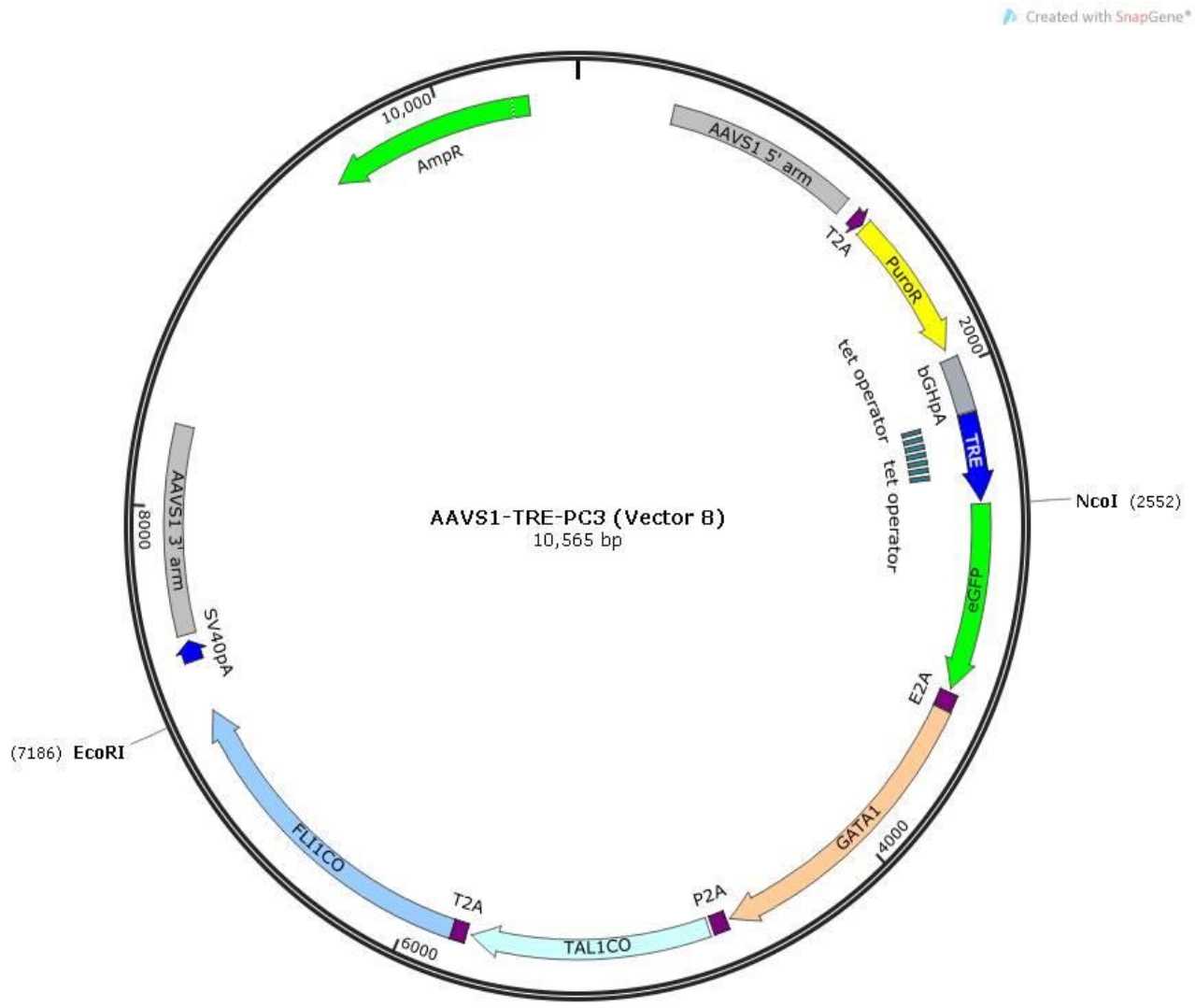
Map 3.8 Vector 5: AAVS1-Ef1α-rTTA-KBlsc-TRE-H2BVenus.



Map 3.9 Vector 6 Rosa26-CAG-rTTA OptiX vector targeted to Rosa26.



Map 3.10 Vector 7 AAVS1-TRE-GFP OptiX responder vector targeted to AAVS1 locus of rTTA-Rosa26 lines. Restriction enzyme sites used to remove eGFP fragment, to replace with PC3 fragment, are shown.



Map 3.11 Vector 8 AAVS-TRE-PC3 final targeting vector used to generate inducible iPSC lines. Restriction enzyme sites used to insert PC3 fragment are shown.

Results

Table 3.7 summarises the vectors generated and tested that will be discussed in this results section, highlighting the main components and differences in targeting strategy for each.

Vector Name	Promoter controlling rTTA expression	Promoter + Selection gene	TRE + H2BVenus	Targeted locus	Targeting method	Vector Map
Mouse Vector All-in-one	Mouse Rosa26	PGK, Puromycin	Yes	Mouse Rosa26	Homologous Recombination	**3.4**
Vector 1 All-in-one	AAVS1	PGK, Puromycin	Yes	AAVS1	TALENs	**3.3**
Vector 2 All-in-one	EF1α	AAVS1, Blasticidin	Yes	AAVS1	TALENs	**3.5**
Vector 3 rTTA only	EF1α	None	No	AAVS1	TALENs	**3.6**
Vector 4 TRE-H2BVenus only	No rTTA	AAVS1, Blasticidin	Yes	AAVS1	TALENs	**3.7**
Vector 5 All-in-one	EF1α	AAVS1, Blasticidin	Yes	AAVS1	TALENs	**3.8**
OPTi-OX (optimized inducible overexpression) Strategy						
Vector 6 rTTA only	CAG	Rosa26, Neomycin	No	Rosa26	ZFNs	**3.9**
Vector 7 TRE-GFP only	No rTTA	AAVS1, Puromycin	Yes	AAVS1	ZFNs	**3.10**
Vector 8 PC3 only	No rTTA	AAVS1, Puromycin	No	AAVS1	ZFNs	**3.11**

Table 3.7 Summary of inducible vectors tested and targeting strategy used.

In the results section the following cell type definitions are used; CD41+/CD235+ cells are bi-potent progenitors, CD41-/CD235+ cells are erythroblasts, CD41+/CD235- cells at day 8-9 are MKs, and CD41+/CD42+ cells at day 19 onwards are mature MKs.

Targeting the Mouse Vector to Mouse ES Cells: Venus Fluorescence Detected

Mouse ES cells were electroporated with the linearised mouse Rosa26 targeting vector, performed by Dr Alena Pance (Wellcome Trust Sanger Institute, Cambridge). Low level fluorescence was observed in some cells. Although the fluorescence was patchy this provided important experimental evidence that the construct and TET components were functional in mouse ES cells (data not shown).

Targeting Vector 1 to Human iPSCs: No Venus Fluorescence Detected

Following the promising results from the mouse cells, human Vector 1, was tested in two iPSC lines. The results of two experiments are summarised in **Table 3.8**.

Condition	Total number of clones after selection	Fluorescence after induction
S4 control (-TALENs)	None	-
BobC control (-TALENs)	None	-
S4	9	None
BobC	7	None

Table 3.8 Summary of testing Vector 1.

None of the selected clones in either iPSC line were fluorescent after doxycycline induction. Reagent issues were ruled out, as fresh stocks of both puromycin and doxycycline were used. Results from the mouse vector targeted to mouse ES cells suggested that the TET components were functional, meaning that the reason for lack of fluorescence in this case might be human iPSC-specific, rather than a problem with the vector.

rTTA Not Expressed Under AAVS1 Endogenous Promoter

In order to understand the lack of fluorescence in the transfected cells, we first verified that the AAVS1 promoter was active and that there was expression of the rTTA gene. Clonal wells of puromycin resistant colonies were collected, and expression levels of rTTA were assessed by qPCR. This analysis revealed extremely low levels of rTTA with the highest expression value being 0.006 relative to *HMBS* (**Fig. 3.1A**). The average Ct value for rTTA was 9 cycles higher than for *HMBS* (34.6 compared to 25.5). This finding led to the conclusion that the targeted AAVS1 promoter was not strong enough to drive expression of sufficient levels of rTTA. These findings are consistent with those of Matthias Pawlowski (a PhD student at the time in Dr Mark Kotter's lab at the Laboratory for Regenerative Medicine), who had also been working on targeting the AAVS1 locus with an inducible TET3G vector.

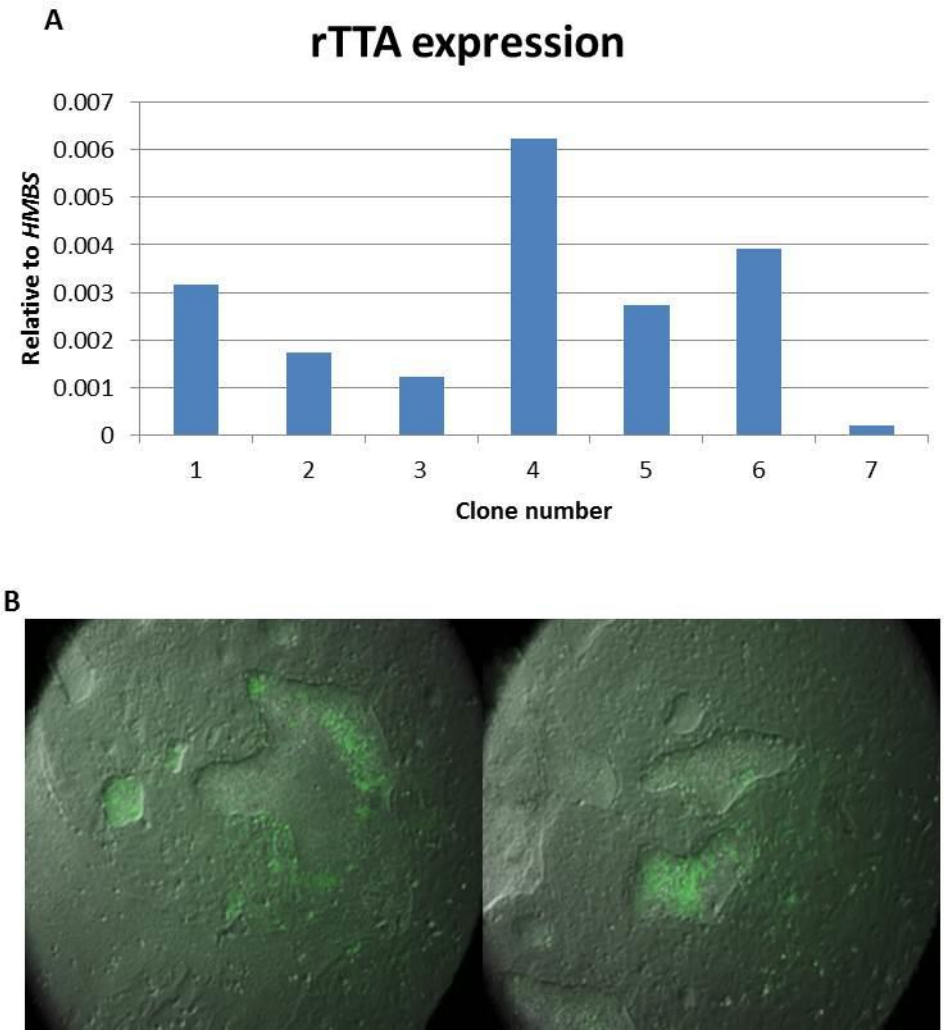


Fig. 3.1 Vector 1 targeted iPSCs do not express rTTA. Data presented here is from BobC iPSC clones, after targeting with Vector 1 and puromycin selection. **A**) qPCR data to show relative quantification values of rTTA expression relative to HMBS of 7 targeted clones. **B**) Light and fluorescent images overlayed of targeted iPSCs show Venus expression in only a few cells after rTTA rLV transduction and induction with doxycycline. N=1.

To further demonstrate that the lack of rTTA expression was the reason for not observing Venus fluorescence, and to verify that the remaining components of the vector were functional, targeted cells were transduced with rTTA lentivirus. The lentivirus used was a 2nd generation rTTA, not identical in sequence to the rTTA used in Vector 1, but believed to be compatible with the 3rd generation TRE. After doxycycline induction, the level of fluorescence observed was not as high as expected and not observed in all cells (**Fig. 3.1B**). There may be a number of reasons for this; poor transduction efficiency, the polyclonal nature of the wells, or rTTA2G-TET3G compatibility. Importantly, however, this experiment demonstrated that the remaining components of the construct (the TRE promoter and H2BVenus reporter gene) were functional in iPSCs. Additionally, it

was observed that upon removal of doxycycline Venus fluorescence was lost, showing the system was doxycycline responsive.

Ef1α Promoter: Resolving rTTA Expression

After rTTA was found not to be expressed under the AAVS1 endogenous promoter, the decision was made to put rTTA under the expression of the EF1α promoter, shown previously to be a strong exogenous promoter in many cell types (Qin, 2010). At this stage a second major change was also made to Vector 1. PGK has previously been reported to be a bi-directional promoter (Johnson, 1990), thus to avoid unwanted transcriptional interference, the PGK_Puro component was replaced with a Blasticidin resistance ORF. Blasticidin resistance expression would be put under the control of the endogenous AAVS1 promoter, replacing the rTTA component. The blasticidin ORF vector was kindly shared by Professor Rudolph Jaenisch (The Whitehead Institute and Department of Biology at M.I.T, Cambridge MA). The replacement vector, Vector 2, therefore contained an EF1α promoter to drive rTTA expression, a blasticidin selection gene driven by AAVS1 endogenous promoter, as well as the marker H2B-Venus.

An immediate issue with selection was observed when targeting iPSCs with Vector 2. After two experiments, in two iPSC lines (BobC and S4), virtually nothing survived in the test conditions (+TALENs). The maximum number of selected colonies obtained was 4, which were not checked for fluorescence. TALEN targeting should have an efficiency of approximately 50%, therefore the lack of cell survival lead to doubts about the correct expression of the blasticidin resistance gene. Steps taken to clone the blasticidin resistance ORF were thoroughly re-checked and concluded that a Kozak sequence, removed during cloning steps of Vector 2, may be necessary for its expression.

The Kozak sequence was restored, and two new vectors to separate the TET components were generated, giving the possibility to target them to separate alleles, or even different loci, if necessary should the all-in-one system fail. Thus an rTTA only component (Vector 3), a TRE only component (Vector 4) and a new all-in-one containing both rTTA and TRE components (Vector 5), were generated. After restoring the Kozak sequence, Vectors 3, 4 and 5 were initially tested in 293T cells to quickly assess if cloning of the EF1α promoter had been successful in restoring rTTA transcription. The results of 2 experiments, one to test the separate components (Vectors 3+4) and one to test the all-in-one vector (Vector 5) have been combined in **Fig.3.2**. No Venus expression was observed from non-transduced cells (condition A-data not shown), as expected. Venus fluorescence was seen in many cells transduced with the separate and all-in-one vectors (conditions B and D) at low DNA concentration (2ng/ml). Higher fluorescence levels were observed in more cells receiving the higher

DNA concentration (conditions C and E). This experiment indicated that rTTA was now being transcribed under the EF1α promoter and that the TRE_H2bVenus responder element was functional, as inducible expression of Venus had been observed.

Condition	Vector	DNA conc. (µg)
A	Non-transduced	0
B	Vectors 3+4	2/ vector
C	Vectors 3+4	4/ vector
D	Vector 5	2/ vector
E	Vector 5	4/ vector

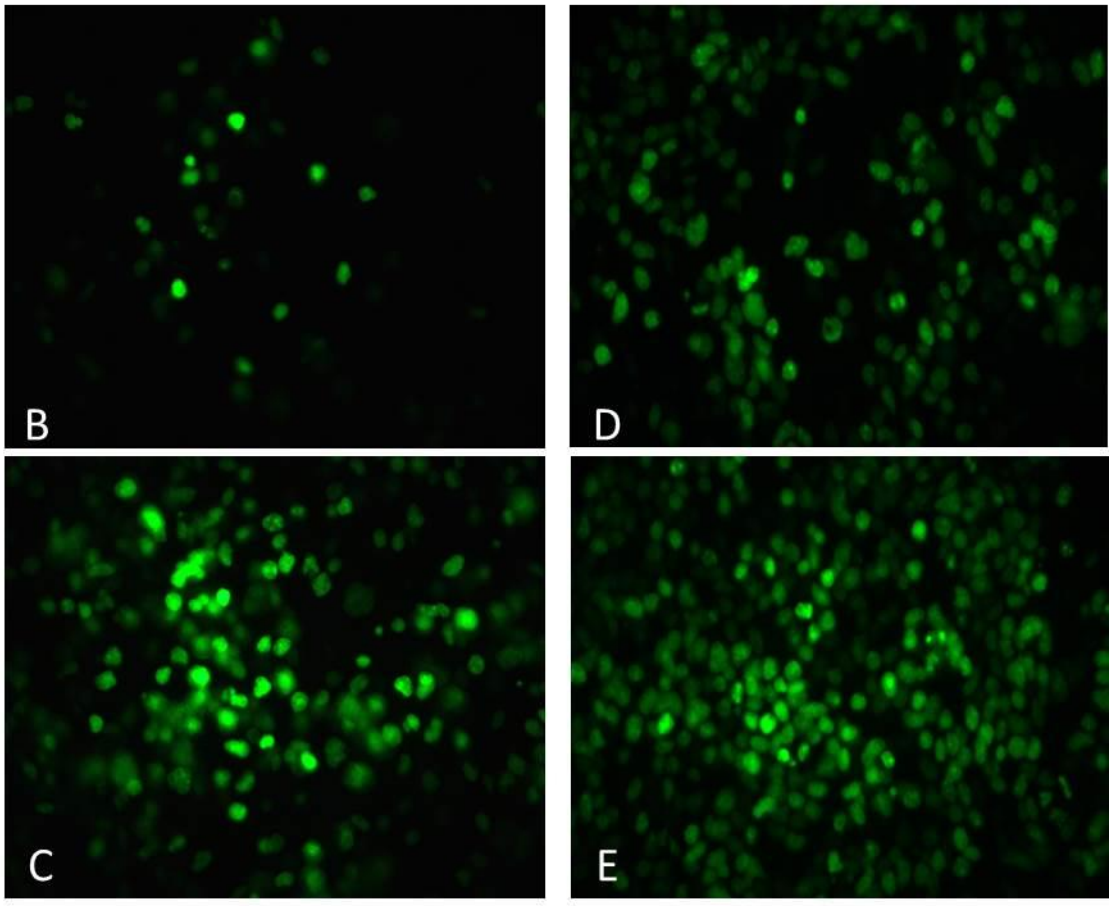


Fig. 3.2 rTTA expression is restored under the EF1α promoter. 293T cells were transduced with the Vectors 3+4 in combination, or Vector 5 only, before induction with doxycycline. The conditions tested are shown, with relative fluorescent microscope images. Venus fluorescence was detected in **B** and **C**, transfected with Vectors 3+4 (separate TET components), and **D** and **E**, transfected with Vector 5 (all-in-one vector). N=1.

Vectors 3, 4 and 5 in iPSCs: Venus Fluorescence Observed despite Low Targeting Efficiency

Having determined that rTTA was now being expressed under the EF1α promoter, iPSCs were electroporated to test that expression of the blasticidin resistance gene had been restored by replacing the Kozak sequence. The results of 4 separate experiments are summarised in **Table 3**.9.

Condition	Number of experiments performed	Number of clones after selection	Fluorescence after induction
Vectors 3+4 control (-TALENs)	1	None	-
Vectors 3+4	1	BobC=2, S4=3	None
Vector 5 control (-TALENs)	3	BobC=6, S4=0, Qolg=5	1 clone for BobC and Qolg
Vector 5	3	BobC=8, S4=0, Qolg=8	2 BobC clones

Table 3.9 Summary of testing Vectors 4-6.

From the separate vectors (Vectors 3+4), a total of 5 clones survived selection but no fluorescence was observed after induction. Genotyping PCRs were performed on Vector 5 targeted and selected clones, to check correct integration. A large number of primers and conditions were tested and although most clones were unsuccessful, one BobC clone had correct PCR products for both 3' and 5' integration in the AAVS1 locus. Subsequent sequencing of the PCR products for this clone sequenced 10 bases of the AAVS1 locus outside of the 3' homology arm of the vector and 14 bases outside of the 5' homology arm, suggesting that this clone was correctly targeted. Fluorescence of this clone, and another which genotyping failed in, was heterogeneous, with approximately 50% of cells visibly green. After 8 days of induction, doxycycline was removed to passage cells. When it was re-added, the cells were no longer fluorescent, suggesting the inserted vector had been silenced in targeted clones. An alternative explanation could be that non-GFP cells carrying genetic abnormalities had a growth advantage and were selected for over this time. An additional observation in this experiment was the number of colonies surviving selection was still sub-optimal, compared to what was expected if both the TALEN targeting and the blasticidin selection were functioning efficiently. For control cells to be surviving and fluorescent, at a similar efficiency to targeted cells, suggested that the TALEN targeting was inefficient- as control wells would represent the efficiency of HR in iPSCs.

The Optimized Inducible Overexpression System

In view of the problems with potential silencing and low colony numbers experienced with our own vectors, we opted to try the system developed by Matthias Pawlowski, termed the OPTi-OX system (Pawlowski M, 2017 and Bertero A, 2016, *patent 1619876.4*). This system uses TET-ON 3G and

successfully achieved induction of different genes at different time-points in a number of iPSC lines. The major differences between the OPTi-OX system and the one I had developed, was that the rTTA and TRE components are targeted separately, to two different genomic safe harbours, using CRISPR/Cas9 nickase (Cas9n) or zinc finger nucleases (ZFNs), respectively. A vector containing a CAG promoter to drive the expression of rTTA is targeted to the human Rosa26 locus and a TRE-driven responder vector is targeted to the AAVS1 locus. The CAG promoter drives a more robust and higher rTTA expression compared to the endogenous Rosa26 promoter or exogenous EF1α promoter (Pawlowski M, 2017). A number of stem cell lines were available, that had been genotyped and confirmed positive for the homozygous insert of the rTTA vector. These lines were shared with us, in order to target my responder vector containing TRE_H2bVenus (Vector 4) to the AAVS1 locus. Work from this point was undertaken with the advice and guidance of Matthias Pawlowski, as well as Alessandro Bertero (a PhD student at the time in Ludovic Vallier's lab) and Daniel Ortmann (a postdoc in the Vallier lab), who had helped establish the OPTi-OX system.

Three iPSC lines shared by Dr Pawlowski, which had been targeted and genotyped for a homozygous rTTA integration in the Rosa26 locus (H9, BBNX and Bob rTTA_Rosa26 lines), were electroporated with Vector 4. The results of this experiment are described in **Table 3.10**.

rTTA_Rosa26 line tested:	Number of clones after selection	Fluorescence after induction
H9	0	-
Bob	0	-
BBNX	3	1 fluorescent colony

Table 3.10 Testing Vector 4 in rTTA_Rosa26 iPSC lines.

Both the H9 and Bob lines resulted in no colonies after blasticidin selection, while BBNX gave rise to 3 colonies. After doxycycline induction, only one colony was fluorescent, with approximately 90% of cells displaying fluorescence. After a further 10 days in doxycycline, this clone had almost no fluorescent cells visible. This experiment led further support to the hypothesis that silencing was occurring to a component in our inducible system. Or alternatively highlighted that outgrowth of a genetically abnormal clone was occurring. The OPTi-OX system had been robustly characterised by Dr Pawlowski and silencing of the rTTA component of these lines seemed unlikely. This suggested that the TRE responder element was potentially being subjected to silencing in the AAVS1 locus. The low colony survival numbers again highlighted a targeting or selection issue.

A second experiment was performed using the Amaxa 4D rather than the Amaxa II machine. Due to concerns regarding degradation of large vectors, >10kb, by the former, this machine had not be tested previously for the larger Vectors 1 and 2. Alongside targeting Vector 5 (all-in-one), a GFP

responder (Vector 7) and the associated AAVS1 ZFN's were tested. Two rTTA_Rosa26 lines were tested, Bob and BBNX. The results of this experiment are summarised in **Table 3.11**.

rTTA_Rosa26 line tested:	Number of clones after selection	Fluorescence after induction
Bob + Vector 5 + TALENs	0	-
BBNX + Vector 5 + TALENs	4	2 fluorescent colonies
Bob + Vector 7 + ZFNs	8	All fluorescent
BBNX + Vector 7 + ZFNs	10	All fluorescent

Table 3.11 Summary of testing Vectors 5 and 7 in rTTA_Rosa26 iPSC lines.

After selection, only 4 colonies survived after targeting with Vector 5 in the BBNX line, while none survived in the Bob line. The number of colonies surviving after targeting with Vector 7 was higher, with 8 in Bob and 10 in BBNX. After induction with doxycycline these clones showed a striking difference between the TALEN and ZFNs conditions. Firstly, only 2 of the TALEN clones were fluorescent, whereas all 18 ZFN clones were. Secondly, the 2 TALEN colonies lost fluorescence within 4 days of doxycycline induction, whereas the ZFN clones remained very homogenously fluorescent for 25 days+ of induction.

A Vector 7 targeted BBNX clone was split into 2 wells, after 8 days of induction. One well continued to have fresh doxycycline added daily (+Dox, green), and the second had doxycycline removed at the time the cells were split (-Dox, blue). **Fig 3.3** shows the percentage of GFP positive cells, gated on DAPI negative viable single cells, at different time points. A negative control, iPSCs from the same clone which never received doxycycline, is also shown (No Dox, red). After 3 days without doxycycline, the expression of GFP remained as high for the –Dox condition, as the +Dox condition (**Fig 3.3A**). After 7 days, GFP expression in the –Dox condition had fallen to the same level as the negative control. Fluorescence remained high for the +Dox condition (**Fig 3.3B**). After 14 days, the expression of GFP follows the same pattern as was observed after 7 days for the No Dox, +Dox and –Dox conditions. After the re-addition of doxycycline to the –Dox condition for 3 days (purple), GFP expression increased to the same level as the +Dox condition (**Fig 3.3C**). The number of cells in the +Dox condition dramatically dropped compared to the NT and – Dox condition. Viability was just 2.2% for the +Dox condition after 7 days, compared to 66% and 50% for the –Dox and NT conditions respectively. This might be due to toxicity of sustained high levels of GFP in cells (Liu et al. 1999), or to Dox toxicity over time.

Fig. 3.3D shows representative fluorescent microscope images of loss and gain of GFP expression, confirming inducibility of GFP using the OPTi-OX system in iPSCs. There was no evidence of silencing of GFP, as the expression remained strong and homogenous in the clones (40x, +Dox condition).

Based on this experiment, with a higher targeting efficiency seen for the ZFNs and GFP responder vector, compared to the TALENs and Venus responder, we decided to switch to the OPTi-OX system to engineer a forward programming iPS cell line.

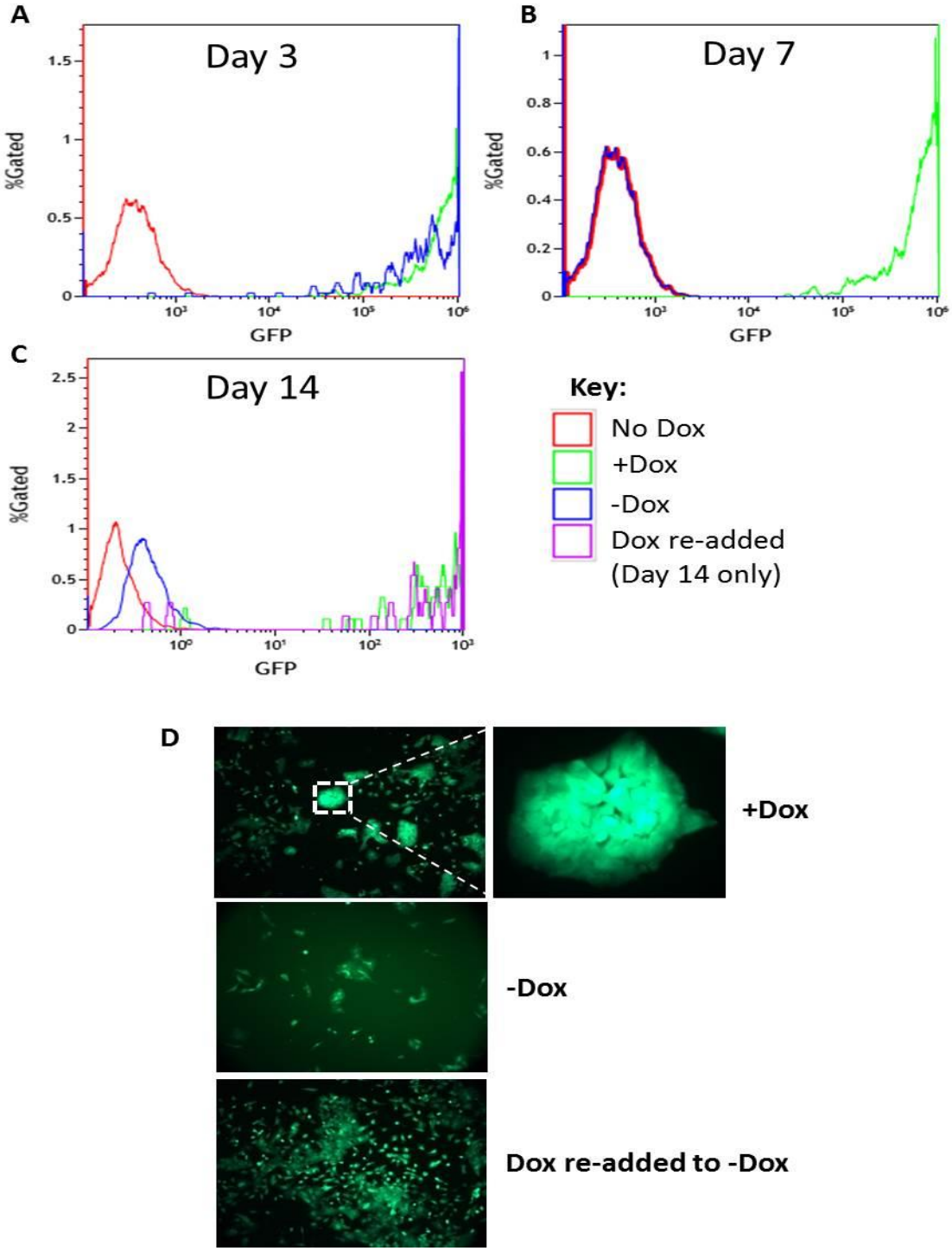


Fig. 3.3 GFP expression of Vector 7 targeted BBNX iPSCs, after doxycycline induction. The percentage of viable cells expressing GFP is shown for different time points. **A**) After 3 days +/- Dox. **B**) After 7 days –Dox. **C**) After 14 days of initial +/- Dox. Targeted iPSCs which never received doxycycline= No

Dox (red), Doxycycline removed= -Dox (blue), Doxycycline maintained= +Dox (green), Doxycycline re-added to –Dox condition 11 days after doxycycline was removed= Dox re-added (purple). **D**) GFP fluorescence images observed in +Dox condition, -Dox after 6 days, and Dox re-added for 4 days. All images 10x optical zoom, 40x zoom for in-lay. N=1.

Testing rTTA iPSC Lines in Lentiviral MK Forward Programming: BBNX Produces MKs

To test if the GFP targeted rTTA lines available (Bob and BBNX) could produce MKs, a lentiviral FoP experiment was performed. Additionally, to test if the system was immune to silencing, GFP expression was measured by flow cytometry in the presence of doxycycline at multiple time points. **Fig. 3.4** shows the results of FoP in these two rTTA lines. The MFI of GFP remained high throughout the experiment for both lines in the +doxycycline condition (**Fig. 3.4A**). No GFP was expressed in the –doxycycline conditions. In the BBNX_GFP line the number of CD41+ MK cells expanded in both the + and –doxycycline conditions, while the Bob_GFP line failed to produce a high number of MKs (**Fig. 3.4B**). In the induced BBNX_GFP line the percentage of GFP+/CD41+ cells increased form 0% at day 2, to 81.5% at day 6 (**Fig. 3.4C**). Together, these results show that the Bob_GFP line does not FoP to produce MKs, seen by the absence of CD41+ cells, while the BBNX_GFP line does. GFP expression in induced cells is immune to silencing during FoP, seen by high GFP MFI and GFP+ cell percentage maintained throughout the experiment. Another important observation is that doxycycline addition to FoP media did not appear to impede MK production, seen by similar MK cell numbers produced in + and –Dox conditions. Based on this experiment, the BBNX_rTTA_Rosa26 line was targeted with the PC3 construct into the AAVS1 locus, resulting in the line BBNX_rTTA_Rosa26_PC3_AAVS1 (BBNX_PC3).

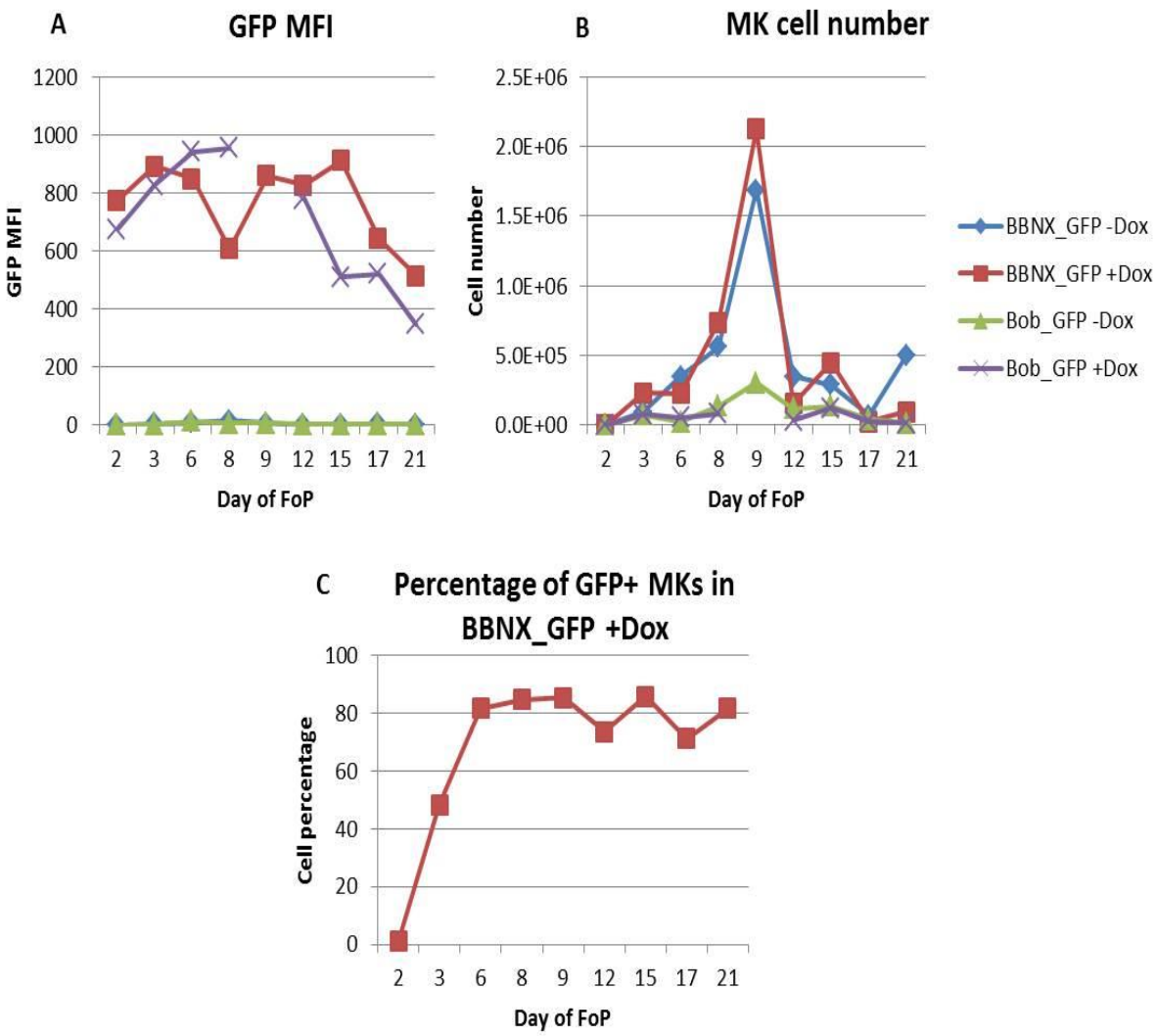


Fig 3.4 BBNX_GFP produces MKs, while Bob_GFP does not. After GFP targeting to the AAVS1 locus in two Rosa26_rTTA iPSC lines (BBNX and Bob), forward programming was performed on the lines in MK medium (TPO + SCF in CellGRo), to asses if the two lines could still programme after two rounds of genome targeting. The lines were forward programmed in the presence of absence of doxycycline (Dox) throughout to test for Dox toxicity. **A**) GFP mean fluorescent intensity (MFI) at multiple time-points for both lines, forward programmed with and without 1µg/ml Dox. **B**) MK (CD41+/CD235-) cell number at the same time-points of FoP. **C**) Percentage of GFP+/41+ cells in BBNX_GFP. *Day 9 data missing for Bob_GFP due to mechanical fault when acquiring data. N=1. The key for all graphs is the same.

BBNX-PC3 line: Induced Forward Programming Fails to Produce MKs After Early Silencing Event Identified

The BBNX_PC3 iPSC line was maintained for several months, undergoing regular passaging. To test whether the OPTi-OX system was robust, with no leakiness (no TG expression in the absence of doxycycline), two pluripotency markers (TRA160 and SSEA4) were used to stain BBNX_PC3 iPSCs. The line stained 99.7% double positive for these markers, showing it maintained pluripotency in normal iPSC maintenance media (**Fig. 3.5A**). This line was induced to express the inserted TGs using doxycycline in order to achieve forward programming, termed inducible FoP (iFoP). Six concentrations of doxcycyline were tested (0, 0.1, 0.5, 1, 5 and 10µg/ml) to test if doxycycline dose would affect iFoP efficiency and to try and establish the optimal concentration for use. Starting from 2.5E+04 cells at day 0, the highest cell expansion was observed at day 5, for all conditions, except the 10µg/ml (**Fig. 3.5B**). By day 10, all conditions had decreased in cell number. The total percentage of CD41+ cells was highest at day 5 for most +Dox conditions, before decreasing in all conditions by day 9 (**Fig. 3.5C**). The percentage of single positive GFP cells ranged from 51-44% at day 2 for all + doxycycline conditions, while the background GFP percentage was just 0.67% in the 0µg/ml condition. At day 9, the percentage of GFP cells was reduced for all conditions (**Fig. 3.5D**).

These results suggest failure to induce FoP in BBNX_PC3. In all but one of the doxycycline conditions cell number was lower than in the negative control (0µg/ml), showing a lack of cell proliferation. The percentage of CD41+ cells was extremely low in all induced conditions and reduced over time. GFP percentage reduced dramatically between days 2-5 for all induced conditions. Together, this suggested an early silencing event of the FoP TGs, resulting in a failure to switch on the endogenous TFs, which prevented FoP.

To test the hypothesis of an early silencing event, qPCR was performed on cells at day 2, 4 and 7, for doxycycline concentrations of 0, 0.5 and 1µg/ml. rTTA and PC3 TG expression were measured to check for levels of mRNA transcripts relative to the house keeping gene *HMBS*. rTTA mRNA, which should be ubiquitously expressed from the Rosa26 locus, was present in all conditions at all time points tested, as expected (**Fig. 3.5E**). PC3 TG expression was absent in the 0µg/ml condition, at all time points, as expected. In the 0.5 and 1µg/ml conditions, PC3 TG expression was detected at day 2, with expression values of 11.8 and 9 relative to *HMBS*, respectively. The expression was reduced by day 4, and further still at day 7 resulting in final expression values of 0.03 relative to *HMBS* (**Fig. 3.5F**). This result supported the hypothesis that the BBNX_PC3 inducible line had experienced an early silencing event after doxycycline induction, preventing inducible FoP from occurring successfully.

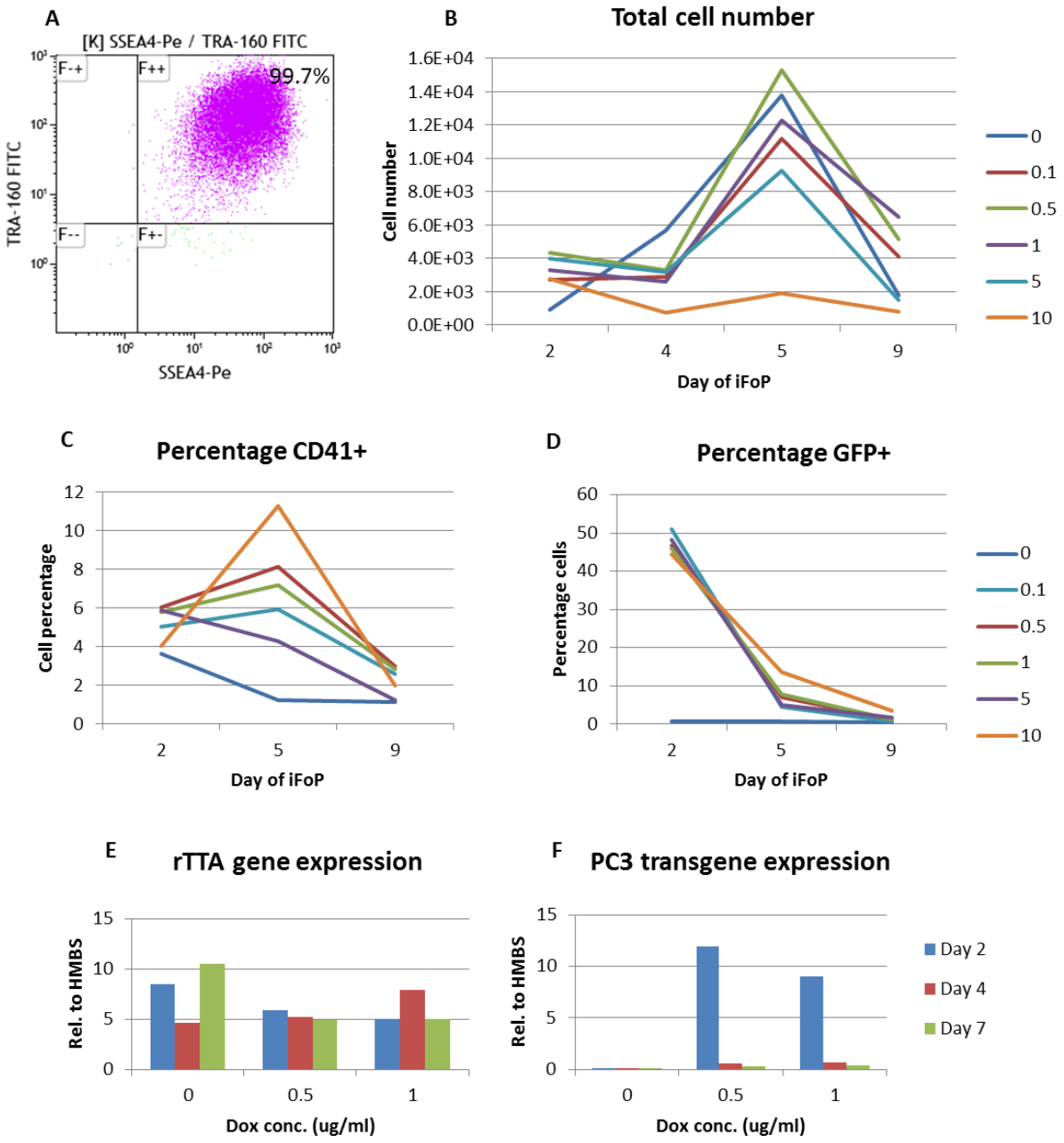


Fig 3.5 BBNX_PC3 iPSC line is stable but does not forward programme after doxycycline induction. The BBNX_PC3 inducible iPSC line was forward programmed by induction with different concentrations of doxycycline (Dox), in MK media (TPO + SCF in CellGro). 2.5E+04 cells were seeded 24 hours prior to Dox induction. At different early time-points cells were collected for flow cytometry analysis and RNA taken for qPCR. **A**) BBNX_PC3 iPSCs, maintained in iPSC media for several weeks, stain double positive for two pluripotency markers, SSEA4 and TRA-160. **B**) Total cell number shown for a range of doxycycline concentrations on days 2, 4, 5 and 9 of iFoP. **C**) Total percentage of CD41+ cells at days 2, 5 and 9. **D**) Total percentage of GFP+ cells at days 2, 5 and 9. **E**) qPCR data shows rTTA mRNA expression at days 2, 4 and 7 from cells induced with 0, 0.5 and 1µg/ml

Dox. **F**) qPCR data shows PC3 transgene mRNA expression at days 2, 4 and 7 from cells induced with 0, 0.5 and 1µg/ml Dox. qPCR data shows relative quantification (RQ) values, relative to housekeeping gene *HMBS*. N=1.

The maximum GFP expression at day 2 had been lower than expected, at 50% (**Fig. 3.5D**), which lead to concerns that the line tested was not clonal. To test this, the line was sub-cloned and an iFoP experiment performed on 11 sub-clones, along with the original clone as a control. **Fig. 3.6** shows the results for this experiment. The total cell number at day 13 for the original clone +Doxycycline was 9.8E+0.3, and the average total for the 11 sub-clones was similar, at 8.6E+0.3 (**Fig. 3.6A**). The percentage of CD41+ and GFP+ cells was similar for the induced original clone, and for the sub-clones (**Fig. 3.6B**). Due to similar cell numbers and percentages between the original and sub-cloned lines, the evidence did not suggest that the original line was polyclonal. Repeat iFoP experiments in BBNX_PC3 (n=7, data not shown) all resulted in failure to induce FoP.

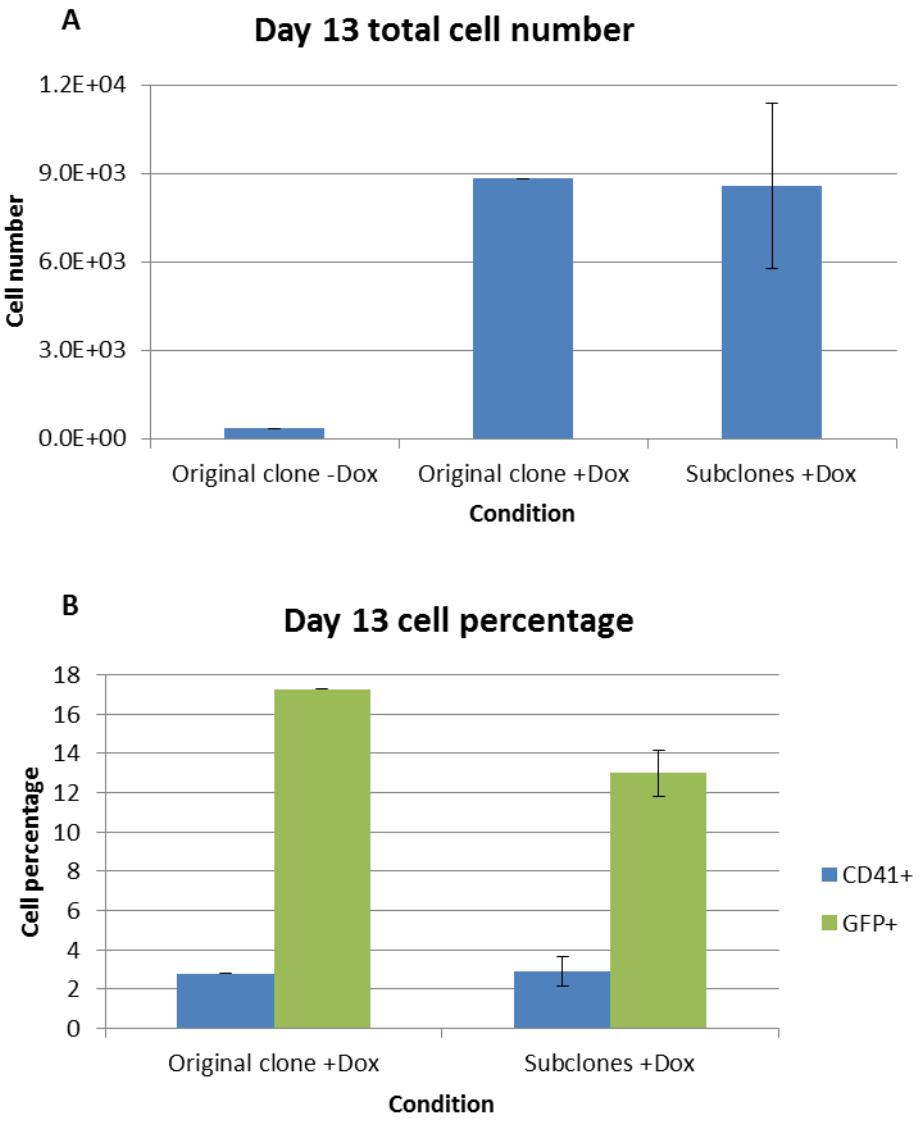


Fig 3.6 Inducible forward programming on 11 sub-clones of BBNX-PC3 did not reveal the original clone to be polyclonal. The BBNX_PC3 line was sub-cloned, with 11 sub-cones grown up from single cells. Inducible FoP was performed in the original line, plus the sub-clones, with 1µg/ml doxycycline (Dox), in MK media (TPO + SCF in CellGro). Flow cytometry was performed at day 13 of iFoP. **A**) Day 13 total cell number of the control clone (original clone from which the 11 sub-clones were established) in the absence or presence of doxycycline (-Dox, +Dox respectively), alongside the average cell number of 11 sub-clones in the presence of doxycycline (+Dox subclones). **B**) Day 13 CD41+ and GFP+ cell percentages of +Dox conditions only. There were <400 viable cells in –Dox condition, therefore flow cytometry analysis on so few cells is not accurate and percentages are misleading. Error bars represent standard error of the 11 sub-clones.

Next, two in-house lines, used routinely to FoP (BobC and FFDK), were generated harbouring homozygous inserts for rTTA in Rosa26 and PC3 in AAVS1. They were also forward programmed by inducing TG expression, to try and understand whether the silencing event observed in BBNX_PC3 was an iPSC line-specific issue or a TG-specific issue, since GFP had not been silenced in BBNX_GFP when forward programmed with lentivirus (**Fig 3.4**).

Testing Induced Forward Programming in Different iPSC Lines: BobC-PC3 Produces Mature MKs

Fig. 3.7 shows the iFoP average of 2 individual clones from the BobC_rTTA_Rosa26_PC3_AAVS1 line (BobC_PC3) across a doxycycline concentration curve (0.2-1µg/ml). Representative dot plots of flow cytometry analysis performed at day 9, show that in the absence of Dox this line does not acquire expression of CD41a or CD235a, two markers of FoP. Due to the number of cells which did acquire expression of either of these markers (0.01% of the totoal population) no further data is shown for the –Dox condition, as flow cytometry analysis is not accurate on so few cells and to show percentages based on this would be misleading. At day 9 the average bi-potent progenitor and MK cell number increased with increasing doxycycline concentration. The number of bi-potent progenitors produced was significantly higher in 1µg/ml compared to 0.2µg/ml (*P*=0.03). Erythroblasts showed an inverse trend, decreasing in number as doxycycline concentration increased (**Fig. 3.7B**). Cell percentages follow the same pattern, with higher bi-potent and MK cell percentages in high doxycycline, and high erythroblast percentage in low doxycycline (**Fig. 3.7C**).

At day 9 the 1µg/ml condition was split into two wells, with one maintained in doxycycline and one removed from treatment. At day 20 the mature MK cell number remained low for all conditions, with a maximum of 7E+05 MKs in 1µg/ml, but by day 29 mature MK cell number had increased. The lowest average number, 7.7E+06, was seen for the lowest doxycycline concentration. The highest cell number, 2.6E+07, was seen for the 0.8µg/ml condition, but the standard deviation for this is large. The condition removed from doxycycline had the second lowest cell number, with 7.8E+06 MKs (**Fig. 3.7D**). Day 20 percentages of mature MKs were above 40%, and had increased to over 80% in all conditions by day 29, showing good maturation and an increase in mature MK cell number (**Fig. 3.7E**).

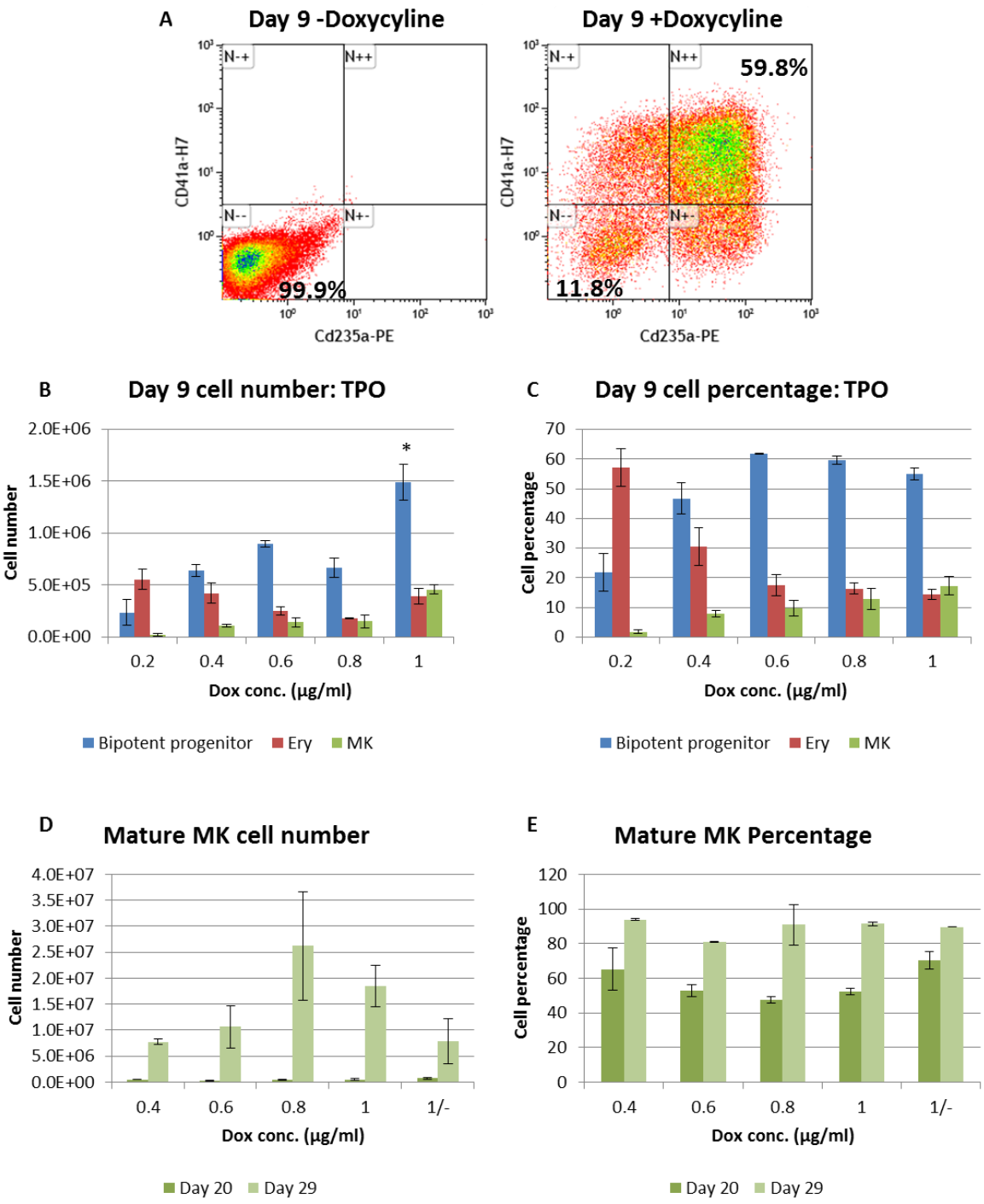


Fig 3.7 BobC_PC3 efficiently produces mature MKs by inducible forward programming. 2 clones of the iPSC line BobC_PC3 were forward programmed in MK medium (TPO + SCF in CellGro), with the addition of doxycycline (Dox) every 2 days at varying concentrations. **A**) Day 9 representative dot plots show the expression of CD41a and CD235a of the line forward programmed with no Dox (-Doxycycline) and with 1µg/ml Dox (+Doxycycline). **B**) Day 9 cell number of bi-potent progenitors (CD235+/CD41+) , Erythroblasts (CD235+/CD41-) and MKs (CD235-/CD41+), in increasing doxycycline concentrations. (*P*=0.03 by two tail *t*-test versus 0.2µg/ml). **C**) Day 9 cell percentages. **D**) Number of

mature MKs (CD41+/CD42+) at days 20 and 29 of iFoP. **E**) Percentages of mature MKs. '1/-' represents the condition in which doxycycline was removed at day 9. Error bars represent range of data, n=2 (biological replicates).

Fig. 3.8 shows qPCR analysis of cells from day 9 and 29, for endogenous *FLI1*, *GATA1* and *TAL1* and PC3 TG, with their expression shown relative to the house keeping gene *MDH1*. At day 9 the lowest doxycycline condition expressed no *FLI1* and had the highest *GATA1* and *TAL1* expression levels. The remaining conditions all had similar levels of endogenous expression (**Fig. 3.8A**). By day 29, when the percentage of mature MKs was similar for all conditions, the expression level of all three endogenous genes was also similar, with *FLI1* showing the largest increase since day 9 (**Fig. 3.8B**). PC3 TG expression was low at both time-points, with expression values ranging from 0.06-0.25 relative to *MDH1*, and showed a higher level of variance at day 9. At day 29, expression followed a dose response to doxycycline concentration. No TG expression was detected in the –doxycycline condition at day 29, showing that expression of the TGs is robustly ceased after doxycycline is removed from the system (**Fig. 3.8C**).

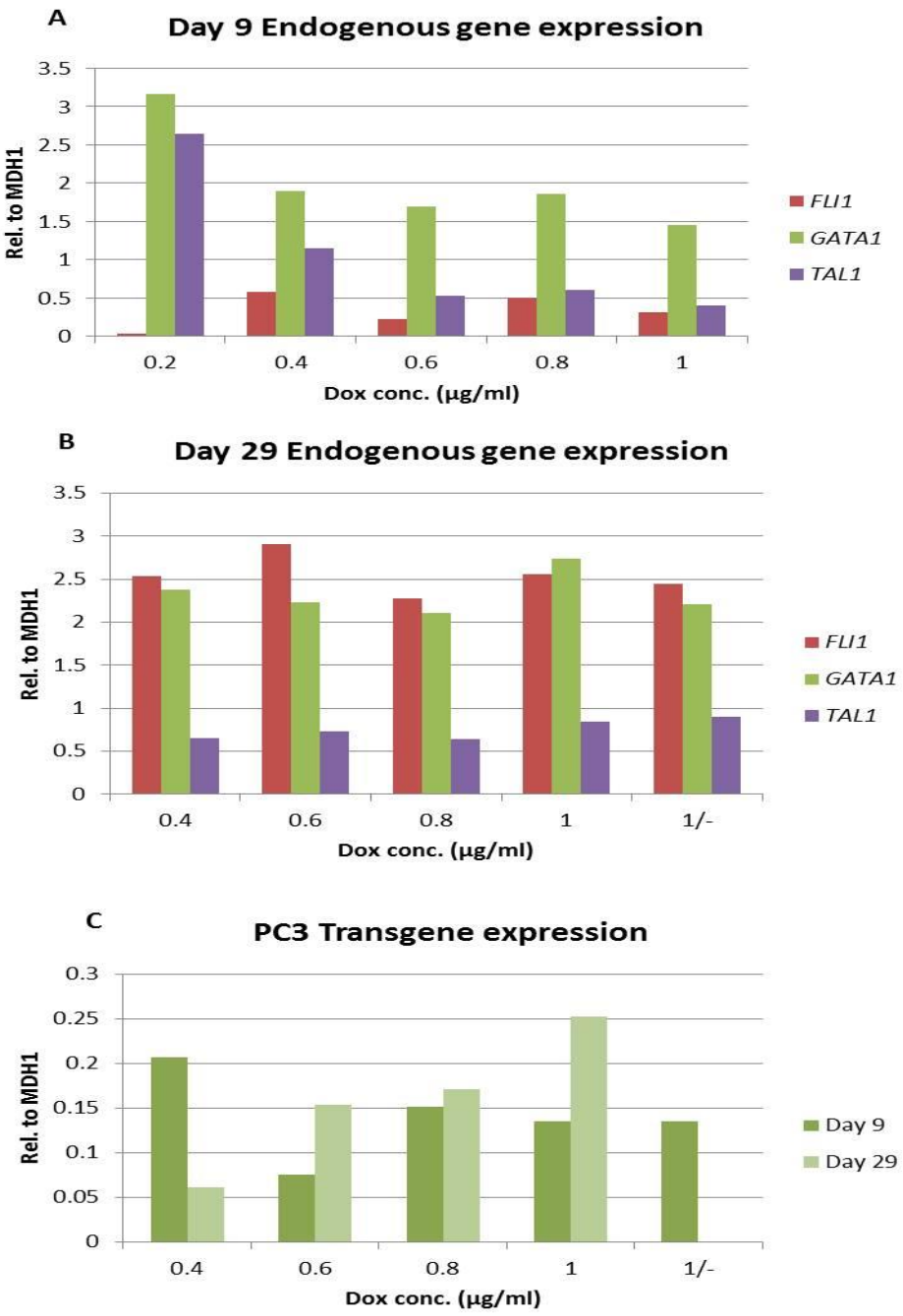


Fig 3.8 BobC_PC3 gene expression during inducible forward programming. qPCR was performed on cells collected from one clone of BobC-PC3 of the inducible MK-FoP experiment, described in **Fig 3.7**. **A**) qPCR data from day 9 to show endogenous *FLI1*, *GATA1* and *TAL1* expression in increasing doxycycline concentrations. **B**) qPCR data from day 29 to show endogenous gene expression. **C**) qPCR data of PC3 TG expression at days 9 and 29. '1/-' represents the condition in which doxycycline was removed at day 9. All values shown are the relative quantification, expressed relative to the housekeeping gene *MDH1*. 0.2µg/ml data not shown at day 29 due to technical error. No dox control

condition was not included for qPCR analysis, due to these cells not having forward programmed and their culture discontinued at day 9.

Induced Forward Programming Efficiently Produces MKs in TPO and Erythroblasts in EPO

To test whether iFoP produced a true bi-potent progenitor and could give rise to erythroblasts, another iFoP experiment was performed across a doxycycline concentration curve in both TPO and EPO. Different iPSC seeding methods were also tested to see whether seeding as clumps or single cells would improve iFoP efficiency. Graphs show the results from just the single cell seeding method, results from the clump seeding method were very similar, showing the same trends, but resulted in fewer cells overall. **Fig. 3.9** shows the day 8 results for TPO and EPO. Similar results to the previous experiment were observed in TPO, with the lowest concentration (0.2µg/ml) of doxycycline resulting in the highest number of erythroblasts, and higher concentrations resulting in a higher number of bi-potent progenitor and MK cells (**Fig 3.9A**). Day 8 cell percentages show a similar pattern, with higher concentrations showing improved iFoP efficiency in TPO (**Fig. 3.9B**).

In EPO, the 0.2µg/ml condition again resulted in the highest erythroblast cell number, with the number of erythroblasts almost double in EPO compared to TPO. In all doxycycline concentrations erythroblasts contributed highest to the overall cell number (**Fig. 3.9C**). The highest percentage of erythroblasts was in the 0.2µg/ml condition with more than 70% of cells (**Fig. 3.9D**). All conditions were split at day 8, into + and –doxycycline, as the previous experiment showed doxycycline was no longer essential after cells had been forward programmed, we wanted to test whether this was the case for cells in EPO as well as TPO.

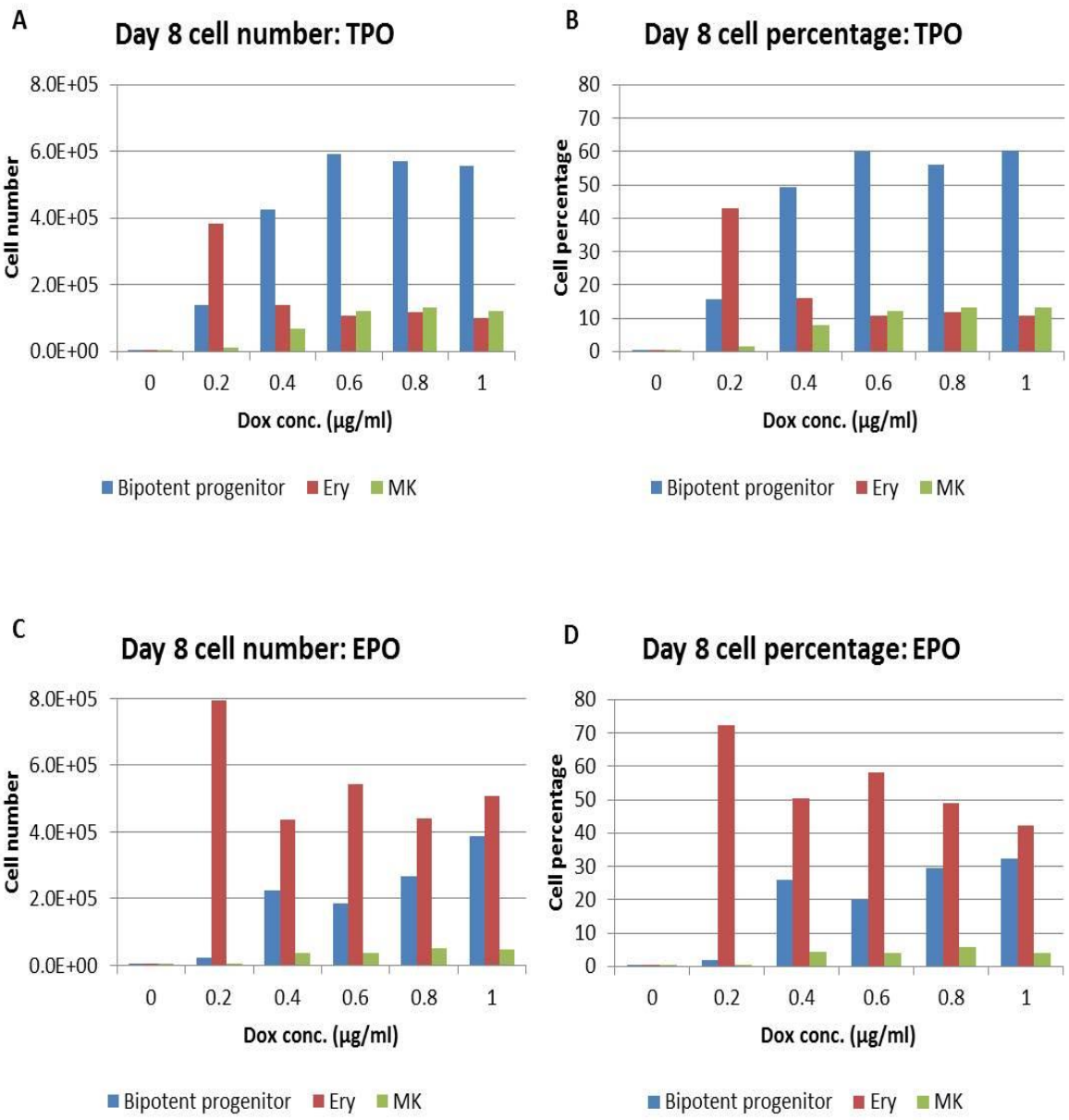


Fig 3.9 Inducible forward programming produces different cell outcomes when BobC_PC3 is programmed in TPO or EPO. BobC_PC3 was forward programmed in MK medium (TPO + SCF in CellGro) or erythroblast medium (EPO, IL-3, insulin, transferrin + SCF in CellGro) from day 2 onwards of FoP, after 2 days of mesoderm induction. Both medium conditions were performed in an increasing concentration of doxycycline (Dox), added every 2 days to medium. Flow cytometry was performed at day 8. **A**) Day 8 cell numbers show bi-potent progenitors (CD235+/CD41+), Erythroblasts (CD235+/CD41-) and MKs (CD235-/CD41+) in TPO. **B**) Day 8 cell percentages in TPO. **C**) Day 8 cell numbers in EPO. **D**) Day 8 cell percentages in EPO. N=1.

Fig. 3.10 shows the results for mature MKs in TPO and erythroblasts in EPO at day 19. At day 19, the number of mature MKs was highest in the highest doxycycline concentration, in both the + and –Dox conditions (**Fig. 3.10A**). The percentage of mature MKs was similar for all conditions (**Fig. 3.10B**). In EPO, the number of erythroblasts was highest in 0.4µg/ml, with cell number higher in the –Dox condition (**Fig. 3.10C**). Erythroblast percentage was very high in all conditions tested (**Fig. 3.10D**).

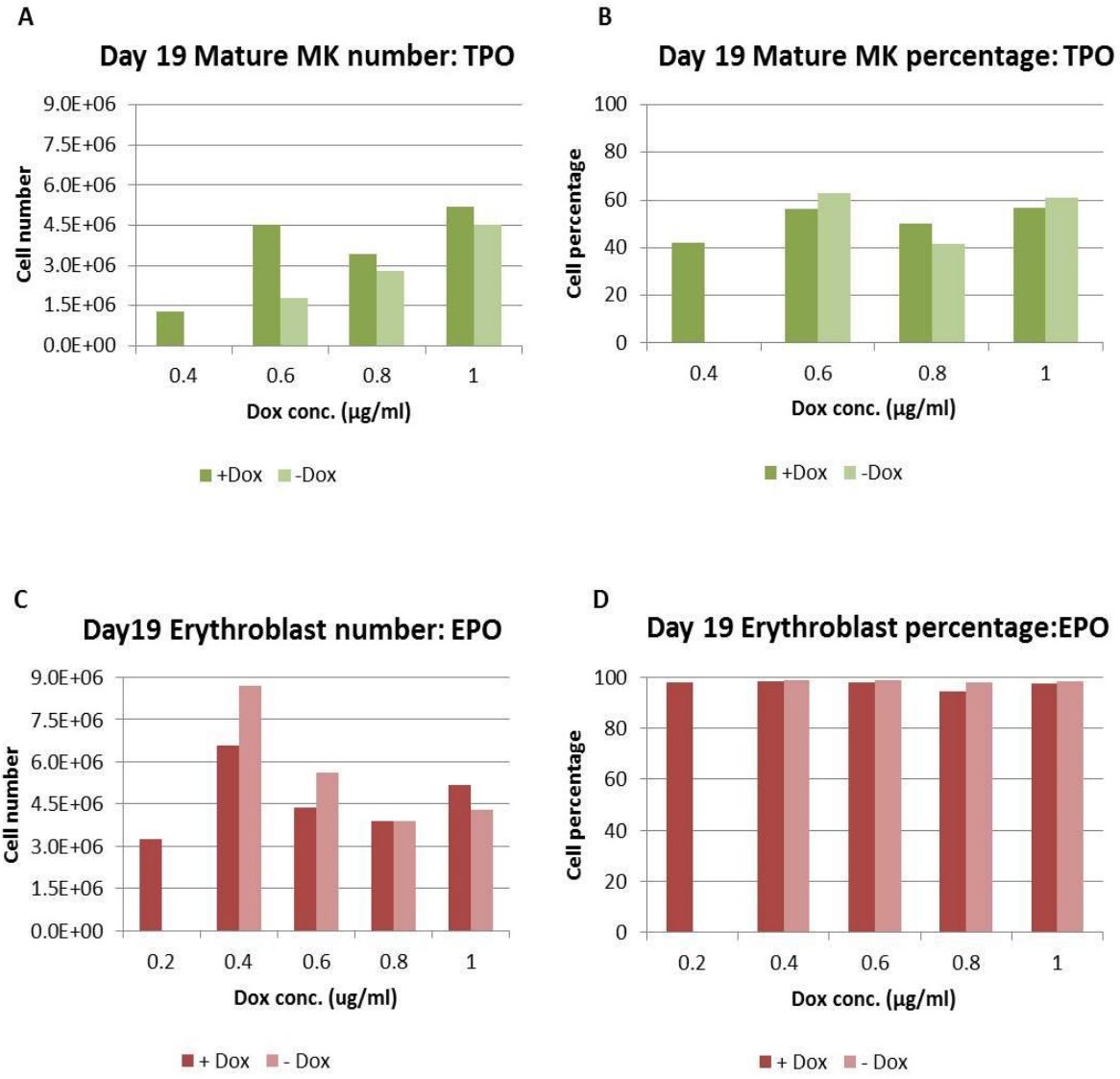


Fig 3.10 Removing doxycycline at day 8 does not impede megakaryocyte maturation in TPO or erythroblast production in EPO. The experiment described in **Fig 3.9** was continued until day 19, when flow cytometry analysis was repeated. At day 8 all conditions were split into 2 new wells, one which continued to received Dox addition every 2 days (+Dox), and one which no longer received Dox (-Dox). **A**) Number of mature MKs in TPO (CD41+/CD42+) with doxycycline maintained or removed at day 8 (+Dox and –Dox respectively). **B**) Percentage of mature MKs in TPO. Erythroblast cell number in EPO (CD235+/CD41-) with doxycycline maintained or removed at day 8 (+Dox and –

Dox respectively) **C**) Erythroblast cell number in EPO (CD235+/CD41-) with doxycycline maintained or removed at day 8 (+Dox and –Dox respectively) **D**) Percentage of erythroblasts in EPO. N=1. (no data for 0.4 –Dox TPO or 0.2-Dox EPO, due to contamination).

Induced MKs Maintain High Purity and Cell Number Long Term

The 1µg/ml Dox condition was split at day 8, to + or – Dox and these conditions were maintained long term for approximately 100 days in culture, with flow cytometry performed periodically throughout this time-period. **Fig. 3.11** shows the results of this long-term experiment. The + and -doxycycline conditions behaved extremely similarly in terms of cell expansion. Following an increase until day 30, a small drop in cell number was observed, followed by a second period of cell expansion (**Fig. 3.11A**). At around day 65 the -doxycycline condition had expanded more than the +doxycycline condition. A maximum expansion of 7508 and 1586 times respectively was seen at day 80. After day 80 both conditions started to decrease in cell number and the cultures crashed not long after day 100. The percentage purity of CD41+/CD42+ cells was above 90% for all time-points after day 19 for both conditions, until day 90 when it dropped just below 90% (**Fig. 3.11B**).

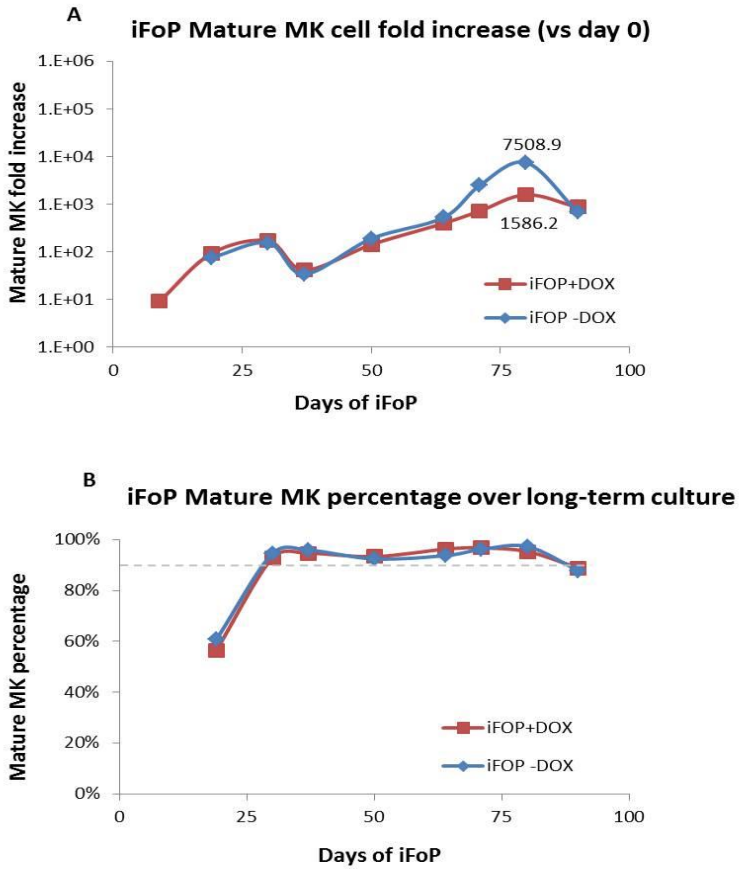


Fig 3.11 Inducible forward programming of BobC_PC3 produces long-term expanding megakaryocyte cultures of high purity in TPO. The 1µg/ml TPO condition (described in **Fig 3.9**) was split at day 8 to continue Dox treatment (+Doc) or stop Dox treatment (-Dox). These two conditions were then maintained in MK medium (+ or – Dox) until day 90 of FoP. **A**) Mature MK (CD41+/CD42+) cell fold increase relative to day 0 iPSC cell input shown over 90 days of culture of the 1µg/ml doxycycline condition, with sustained or removed doxycycline after day 8 (iFoP +Dox and iFoP -Dox respectively). Highest cell fold indicated by values at day 80. **B**) MK purity (CD41+/CD42+ percentage) over long-term iFoP culture.

Platelet-like particle assays were performed on the long-term cultures of iFoP MKs, at day 80 by Dr Annett Müller. **Fig. 3.12B** shows the results compared to donor platelets and MKs produced by lentivral FoP. While donor platelets produce 99% CD42+/Calcein AM+ particles, the levels are lower for the iFoP (+ and – Dox)and lentiviral FoP MKs but they are comparable to each other, with the iFoP derived MKs producing a less abundant double negative population. This shows that iFoP MKs are capable of producing platelet-like particles, in a similar manner as lentiviral FoP MKs.

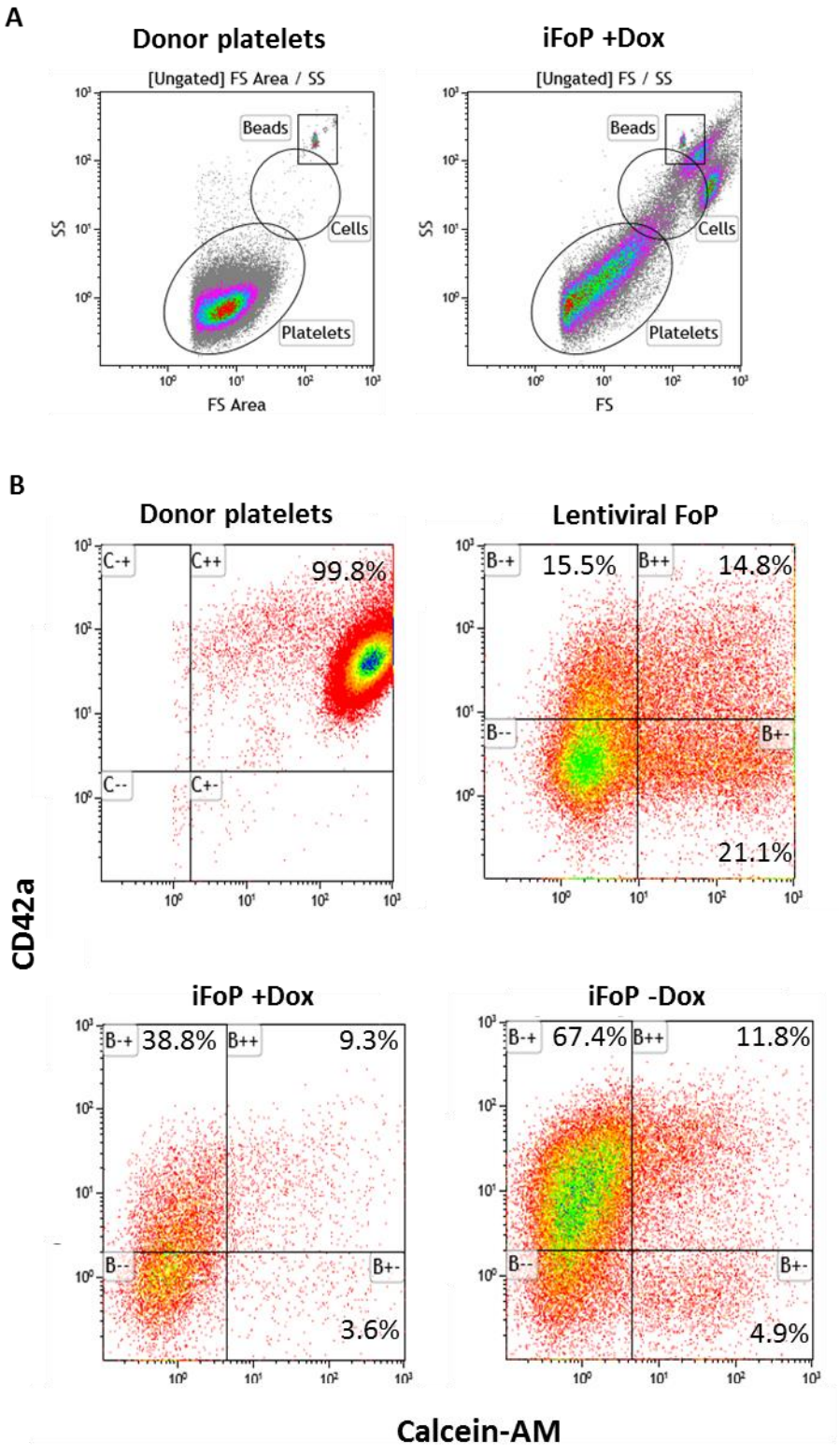


Fig 3.12 Long term inducible forward programmed megakaryocytes capable of producing platelet-like particles in a similar manner to lentiviral forward programmed megakaryocytes. **A**) Representative dot plots show cell gating strategy used, with platelet size determined by cells present in a donor platelet bag (left) alongside long-term (day 80) iFoP+Dox cells (right). **B**)

Representative flow cytometry CD42/Calcein-AM dot plots show double positive platelet-like particle cell populations for donor platelets, lentiviral FoP-MKs and iFoP-MKs (+ and – doxycycline after day 8). Gated on single, DAPI-ve, platelet sized cells. Data generated by Dr Annett Muller.

Characterising Maturity of Induced Erythroblasts: Cells Mature and Capable of Enucleating

To further characterise the maturity of erythroblasts produced by iFoP at day 30 and 37, two maturation markers, CD71 and CD36, were used to stain cells, in addition to CD235. Only data from the 0.4µg/ml conditions is shown, as this condition had been deemed the best for producing erythroblasts due to high cell numbers. Staining double positive for CD71 and CD36 indicates a less mature phenotype, while the loss of both markers indicates the phenotype of an erythrocyte (Mao B, 2016). At day 30, the percentage of double positive CD71/CD36 cells was low for + and -Dox (**Fig. 3.13A**). 22% of cells in +Dox stained single positive for CD71, which most likely reflects the pro-erythroblast or polychromatophilic erythroblast, which stain $CD71^{high}/CD36^{low}$ in normal blood cells. A higher percentage of cells, 48%, had this phenotype in the –Dox condition. The highest percentage of mature cells, phenotypically the same as a mature erythrocyte (CD71-/CD36-), is 76% in +doxycycline compared to 46% in the –doxycycline condition. Cytospin images (**Fig 3.13 B/C**) show enucleated erythrocytes at day 29 and 35 of culture, confirming the maturation status observed by flow cytometry.

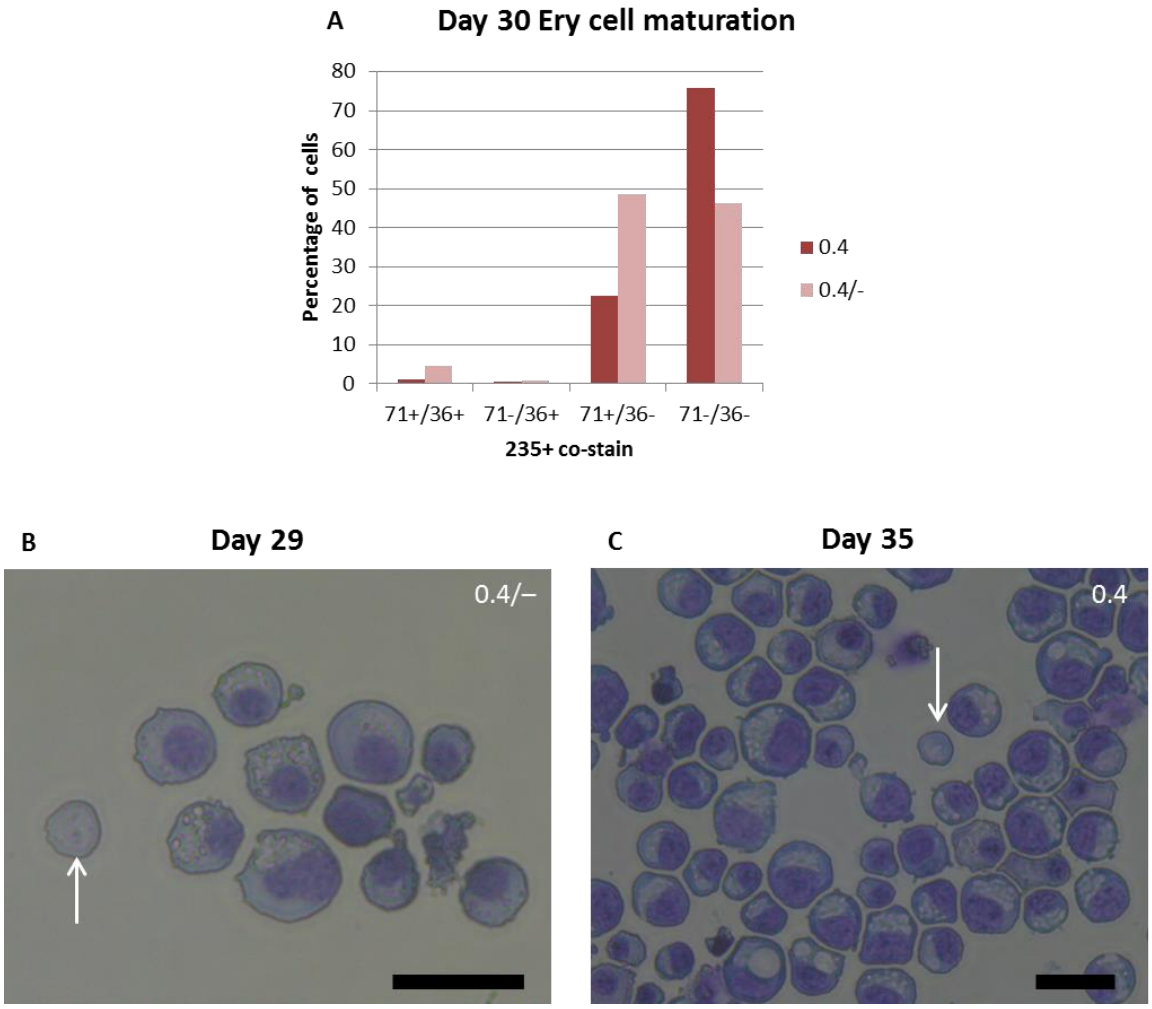


Fig 3.13 Erythroblasts produced by inducible forward programming are capable of maturing and enucleating in EPO medium. The 0.4µg/ml doxycycline was split at day 8 and cultured in EPO with Dox (0.4) or without Dox (0.4/-) after this time point. **A**) Day 30 erythroblast maturation status of CD235+ cells, determined by staining with CD71 and CD36, shows least mature phenotype (left) to most mature phenotype (right). **B + C**) Cytospin images to show enucleated erythrocytes (arrowed) in EPO at day 29 and 35. Scale bar= 20 µm. N=1.

Testing Induced Forward Programming in Different iPSC Lines: FFDK-PC3 Fails to Produce a High Volume of MKs

The second in-house line generated, FFDK_rTTA_Rosa26_PC3_AAVS1 (FFDK_PC3), was tested across a concentration curve of doxycycline. **Fig. 3.14** shows the results of two experiments. At day 9 the condition with the highest cell number was the negative control, 0µg/ml doxycycline. As the concentration of doxycycline increased, the number of cells decreased, in an inverse manner (**Fig.**

3.14A). The percentage of CD41+ cells was low in 0μg/ml and 0.1μg/ml doxycycline, less than 2% for all cell types. At 0.2μg/ml and above CD41+ cells were present and the percentage increased with increasing doxycycline concentrations. A maximum percentage of bi-potent progenitors was observed in 0.4μg/ml, at 67.8%. In 0.6μg/ml the highest percentage of erythroblasts and MKs were observed, 6.7% and 21.6% respectively (**Fig. 3.14B**). Representative dot plots show 0.1% of cells stained CD235+/CD41+ in 0μg/ml doxycycline, compared to 16.7% in 0.2μg/ml (**Fig. 3.14C/D**).

The FFDK_PC3 line appeared to have FoP potential in high concentrations of doxycycline, shown by the appearance of CD41+ cells, but the cell number remained extremely low at day 9, without recovering over an extended culture period, showing an inability to proliferate. To test whether doxycycline toxicity was the reason for lower cell number in higher concentrations, a kill curve was performed on the original untargeted FFDK iPSC line. The kill curve showed no toxic effect on FFDK, even at the highest doxycycline concentration, 1μg/ml. Next, minor adjustments were made to the iFoP protocol, to reduce the frequency of doxycycline addition, in order to see if this would improve cell outcome. However, the altered iFoP protocol was found to be less efficient than the previous experiment (data not shown). Finally, to test whether the line had acquired any off-target effect, undetected by genotyping, that had rendered the FFDK_PC3 line incapable of forward programming, we tested two clones in a lentiviral FoP experiment, to ensure TG overexpression. From these experiments very few viable cells remained at day 9 and cell percentages were low (n=3, performed by Dr Annett Müller, data not shown). Due to this, we reasoned that the targeted FFDK_PC3 line tested is not suitable for iFoP and no longer suitable for lentiviral FoP, most likely due to undesired off-target integration of the rTTa or PC3 vector, or due to the survival of a mutated clone offering a selection/growth advantage due to low efficiency of electroporation.

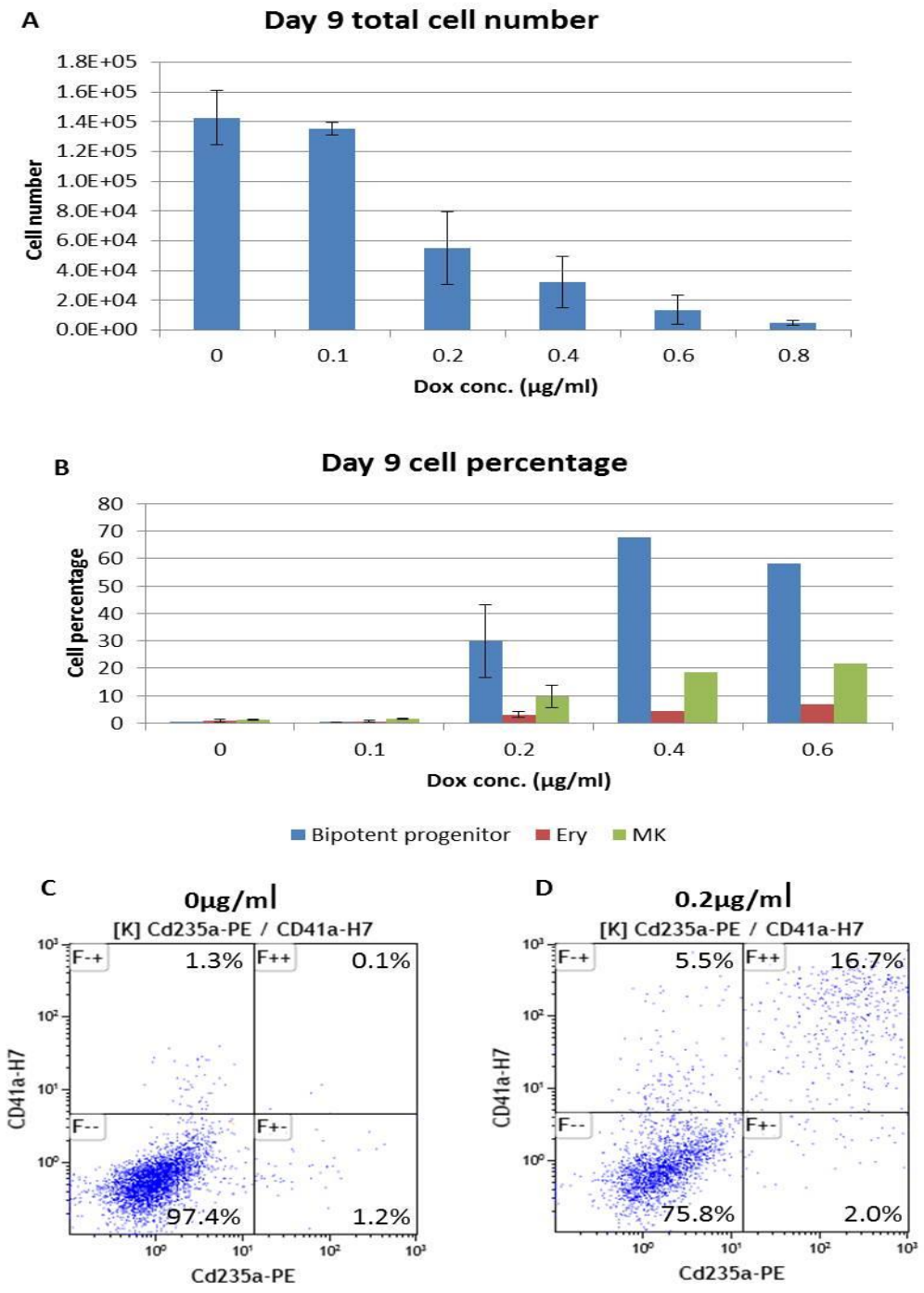


Fig 3.14 FFDK_PC3 line shows inducible forward programming potential in increasing doxycycline concentrations but generates very few cells. **A**) Day 9 total cell number of a FFDK_PC3 clone in increasing doxycycline concentrations. **B**) Day 9 cell percentages show iFoP populations of bi-potent progenitors (CD235+/CD41+), Erythroblasts (CD235+/CD41-) and MKs (CD235-/CD41+) in increasing doxycycline concentrations. **C**) Representative flow cytometry dot-plot from 0µg/ml condition showing CD235/CD41 co-staining. **D**) Representative flow cytometry dot-plot from 0.2µg/ml condition. Error bars represent range of data for n=2. When too few (<1000) viable cells were available to analyse, error bars are absent (n=1) or the condition is not shown.

Discussion

This Results Chapter has described the work of designing and testing various vectors for targeting the TET-ON components to the AAVS1 and Rosa26 genomic safe harbours of human iPSCs. Several vectors initially tested were found to be either non-functional or ultimately were found to be silenced. To overcome the issue of silencing, the two TET components (rTTA and TRE) were separately targeted to each of the previously mentioned genomic safe harbours, using the OPTi-OX system. Once inserted into three iPSC lines, induced FoP (iFoP) was performed. Only one of these lines, BobC_PC3, was able to iFoP efficiently, highlighting iPSC inter-line variability. BobC_PC3, when induced to FoP in TPO maintained a highly pure population of mature MKs in long-term culture (**Fig 3.11**). In EPO, a highly pure population of erythroblasts was produced, with evidence of maturation and enucleation (**Fig 3.13**). Initial doxycycline concentration and iPSC seeding method altered cell number outcome, and has been optimised for TPO and EPO culture conditions. It was found that doxycycline was not required after 8 days of iFoP, as cells by this time have switched on endogenous gene transcription and are already forward programmed (**Fig. 3.8**).

rTTA Not Expressed Under AAVS1 Promoter

After initial experiments with Vector 1 resulted in low level, patchy Venus expression as a result of very low or no rTTA expression (**Fig 3.1**), we hypothesised that the AAVS1 endogenous promoter targeted was not strong enough to drive sufficient rTTA expression. However, results from Vector 2 suggested an inefficient splice acceptor was being used in the vector, which was identical to the splice acceptor that had been used in Vector 1 to splice the endogenous promoter and rTTA gene fragment. Thus, the low rTTA expression observed was most likely due to the ineffective splice acceptor in Vector 1. Interestingly, Matthias Pawlowski independently identified insufficient rTTA expression under the same promoter, suggesting that this endogenous promoter is not strong enough to drive sufficient rTTA expression, in line with our original hypothesis.

Venus Expression Silenced in iPSCs, using All-in-One Targeting Vector Strategy

Despite issues with targeting and/or blasticidin selection, which resulted in sub-optimal numbers of clones, the cells that did show expression of Venus with Vector 5 resulted in silencing of expression. Cells which survived selection were likely to be targeted correctly, so when Venus expression was checked and many cells were found to be non-fluorescent, it is likely that these cells had already silenced the inserted DNA. The patchiness of Venus fluorescence can be explained, in part, due to the fact that the H2B targets this protein to the histone of cells, thus cells in different stages of the cell cycle will show differences in fluorescence (Kanda T, 1998). Due to the fact that the inserts are at

the same locus, the difference in fluorescence could also be partly explained by observing a mixture of heterozygous and homozygous inserted clones. Nevertheless, the reduction in the number of fluorescent cells within a single clone overtime represent a silencing event which cannot be explained by fluorescent variability, since each clone was grown from a single iPSC. Due to the observed silencing of Venus occurring randomly in cells from a single clone, it is likely that the silencing events occurred in a cell-independent manner during cell replication.

Interestingly, Matthias Pawlowski independently found an all-in-one vector of the OPTi-OX system, targeted to the AAVS1 locus, was inefficient for driving TG expression in iPSCs, as this too showed evidence of being silenced. This suggests that there is an issue in iPSCs when trying to express rTTA and its target, TRE, from the AAVS1 locus. One thought was that this may be due to the size of the targeted insert. However, our collaborator Stefania Carobbio used the same AAVS1 TALENs to insert a larger insert of 15kb into iPSCs successfully, without observing silencing. These observations resulted in the conclusion that the TET components targeted together were leading to a silencing event occurring in iPSCs at the AAVS1 locus.

Optimised Inducible Overexpression System Allows Robust TG Expression During Forward Programming

Due to strict gene regulation in iPSCs, one of the biggest concerns with generating an inducible iPSC line was that TG expression could be switched off or silenced in either the iPSC state, as seen with the previous all-in-one strategy, or during forward programming. To test this in the OPTi-OX system, the BBNX_GFP and Bob_GFP lines were forward programmed with rLVs, with the addition of doxycycline. Importantly, GFP expression was observed throughout and was not silenced, showing the OPTi-OX system does allow TG expression over the course of forward programming (**Fig 3.4**), crucial for the inducible line with FoP TFs to work. There was a slight reduction in GFP MFI overtime, more so for Bob_GFP, but this line failed to forward programme efficiently. The BBNX_GFP line, which did produce mature MKs, showed varying MFIs throughout the experiment but this remained high throughout. From this we were confident that TGs inserted into the AAVS1 locus would be protected from silencing in both the iPSC state and during FoP.

Next, it was important to establish that the system was in no way 'leaky', i.e that there was no expression of the inserted TGs at any time when doxycycline was absent. To check this, the BBNX_PC3 line was cultured for several months as per routine iPSC maintenance. A pluripotency check showed that the line had maintained pluripotency (**Fig 3.5**) and cell morphology was normal and comparable to the original, untargeted BBNX line. Crucially, unwanted spontaneous

differentiation (determined by cell morphology change) was not observed in any targeted iPSC lines generated with the OPTi-OX system. This is important for showing that this targeting strategy generates a stable iPSC line that can be cultured indefinitely and retains the main features of pluripotency. This is not only important for providing an easy to use tool for researchers, as the line can be frozen, thawed and maintained, while retaining the ability to FoP when required, but this is also a crucial requirement for the scaling up of FoP technology.

Inducible Forward Programming is iPSC-line Dependent

The first iPSC line described that was targeted with the OPTi-OX system, to induce expression of PC3, was BBNX_PC3. This line was unable to iFoP, as a result of a loss of insert or PC3-specific silencing event occurring before day 4 (**Fig 3.5**). This suggests that TG overexpression is required past this stage (day 4), in order to allow time for endogenous genes to be switched on, to drive FoP. The loss of insert or silencing event must have been due to the nature of the PC3 insert specifically, as the BBNX_GFP line had not silenced GFP during FoP (**Fig 3.4**). This may have been triggered due to the larger insert size of the PC3 vector (10.5kb), compared to the GFP only vector (6.6kb). One possible explanation is that the larger PC3 insert brought about epigenetic changes at the AAVS1 locus during differentiation that GFP alone did not, such as methylation or chromatin remodelling. This in turn may prevent the transcriptional machinery from being able to access the promoter at this locus and result in a failure to initiate transcription.

Another targeted line tested, FFDK_PC3, was also unable to iFoP. Interestingly, this line showed evidence of increased forward programming potential, compared to BBNX_PC3, as seen by the appearance of FoP cells by day 9 (**Fig 3.14**). This led us to conclude that the PC3 insert had not been silenced in this line. However, this line showed poor viability, in three clones, in high doxycycline concentrations. Initially, the poor viability was considered to be a doxycycline toxicity issue, however, the untargeted FFDK line revealed no evidence of toxicity. This led us to question whether something may have occurred in the line during the targeting steps, which had rendered the line incapable of producing MKs. This indeed appeared to be the case, as two clones from the FFDK_PC3 line failed to produce enough MKs to be analysed by flow cytometry when the 3TFs were overexpressed using the rLV FoP method. An unwanted off-target event may have occurred resulting in poor FoP efficiency, which highlights inter-line variability of iPSCs and should be an important consideration for researchers wishing to implement such an inducible system.

It is essential that we are able to verify the findings shown in this thesis, in other inducible lines in the future. It is also important for us to show whether the three TFs within the polycistronic

cassette, used to generate the inducible line, have stoichiometric expression and that the proteins produced are cleaved as expected. If this is not the case, steps could be taken to modify the polycistronic construct, which may be beneficial when testing new inducible PSC lines.

Inducible Forward Programming Efficiently Produces MKs and Erythroblasts in BobC

The only targeted line, of the three tested, to successfully produced MKs and erythroblasts by iFoP, was BobC_PC3. Multiple clones from this line were induced, showing iFoP is highly reproducible and efficient (**Fig 3.7** and **Fig 3.9**). At early time points of iFoP, the overall percentage of cells which have acquired CD41, CD235, or both, was higher than at the same time point by rLV FoP. In the highest doxycycline condition tested, the average percentage at day 9 was 86.5% (**Fig. 3.7B**, n=2), compared to rLV experiments where the average percentage was 63.1% (Chapter 2, **Fig. 2.6B**, n=3). The increased efficiency in iFoP is most likely explained by the reduced stochasticity at the start of experiments, as all cells contain the three TFs, compared to those transduced with rLVs. While the efficiency of iFoP is higher, it is not 100%, which could be due to varying levels of rTTA or TG expression in individual cells, or that a concentration of doxycycline higher than 1µg/ml might be required to achieve 100% efficiency, as a dose response to doxycycline concentration was observed (**Fig3.7** and **Fig 3.9**). Higher doxycycline concentrations were only tested in the BBNX_PC3 line, which eventually did not iFoP, so the concentration could potentially be further optimised for the BobC_PC3 line.

Another advantage of iFoP over the rLV FoP technique is that cell death is greatly reduced at the start of experiments. We know that in rLV FoP any cell not receiving all three TFs will not FoP, and so this accounts for some of the cell death seen. Another reason for large cell death is due to cells being transduced with too many rLV particles, or due to rLV insertions into regions of the genome which are not conducive for cell division.

The fold increase in mature MKs produced is approximately doubled at day 25 for iFoP experiments, with around 1.0E+02 MKs per starting iPSC (**Fig 3.11A**), compared to the same time point of rLV experiments, with just 1.0E+01 MKs per starting iPSC (Moreau T, 2016). Over the long-term iFoP culture, the total number of MKs produced was not as high as we have seen previously from rLV experiments. However, the long-term experiment has not been repeated yet, and variance in cell number could be mostly due to difference in how the cells were handled between different people over such a long time period. An alternative explanation is that iFoP cells may not continue to proliferate as well over time, and this may be due to the fact that a fixed ratio of 1:1:1 for the TFs is already set in these cells. It could be that the rLV setting allows for an altered ratio of TFs in certain

cells, which may favour a more proliferative state. So far we do not have enough data from the iFoP line to know whether this is the case, although seeing as the effect is seen much later than day 9, when we know FoP has already occurred and the inducible line has shown that TGs are no longer required after this point, it seems less likely.

Interesting, the efficiency of iFoP at day 9 did not have any effect on the percentage of cells present at day 19/20. It did not matter what the initial doxycycline concentration was, because by day 19/20 and 29 the percentages of mature MKs in TPO and erythroblasts in EPO were extremely similar, with the only exception being the 0.2µg/ml condition in TPO (**Fig 3.7D**, **3.10A** and **3.10C**). Even more interesting was the finding that doxycycline removal at day 8/9 of culture equally did not have any effect on the percentage of cell maturation. This shows that by day 8/9, after endogenous *FLI1*, *GATA1* and *TAL1* have been switched on, FoP has already occurred and the resulting gene networks switched on by the endogenous genes are more important for driving maturation and proliferation.

What is striking from the inducible experiments performed on BobC_PC3, is the high level of purity seen in cultures that were maintained long-term (**Fig. 3.11**). For mature MKs, the purity was maintained above 90% for almost 60 days in culture, while in a rLV setting the purity has been shown to peak at around 90% at day 30, but reduced to 65% by day 90, as CD42 expression is lost over time (Moreau T, 2016). In the EPO setting, the purity of erythroblasts at day 19 was approximately 98%. This is greatly increased compared to the rLV setting, where at a similar time point a high percentage of bi-potent progenitors is maintained and erythroblast purity is approximately 35% (Chapter 1, **Fig 1.7C**). Higher purity was also observed in the conditions which had doxycycline removed after day 9, so suggests that it is more the result of endogenous gene expression after TG induction has ceased. A good way to check if the differences observed in purity is due to differences in endogenous gene expression, would be to perform qPCR on both iFoP and rLV FoP cells to check expression levels at multiple time points. At the total population level, removing TG expression in all cells at the same time could help in the maturation steps by making the population more homogenous in iFoP, whereas the process is most likely staggered in the population of transduced cells, due to the heterogeneity of TG expression.

In the lowest concentration of doxycycline tested there was a detrimental effect on cell number outcome by day 19, with the fewest mature MKs in TPO (**Fig 3.10A**), and the fewest erythroblasts in EPO (**Fig 3.10C**). 0.2µg/ml produced the lowest percentage of bi-potent progenitors at day 8, showing that efficient FoP requires a high percentage of bi-potent progenitors in order to generate high cell numbers. In TPO, removing doxycycline at day 9 had little effect on the cell number outcome of conditions which had received high doxycycline initially (0.8 and 1 µg/ml), compared to

in low doxycycline conditions (0.4 and 0.6µg/ml), where removal of doxycycline had a detrimental effect on cell number, despite a similar number of a FoP cells at day 8 between all conditions. The same effect was not observed for erythroblast numbers in EPO. This suggests that doxycycline concentration may affect the rate of proliferation of MK cells but not erythroblasts, with high doxycycline required initially to maintain higher levels of proliferation. At the TG level, this means a higher level of TG expression initially may be beneficial for MK cell number overall.

It is reassuring that the inducible line followed the same cell number dynamics that we would expect from an rLV experiment, with an expansion phase followed by a crash at day 90. This is good evidence that the same mode of differentiation is being taken by iFoP cells as transduced cells, and verifies that we have not generated an immortalised differentiated cell line, which was not the aim of this work. Also of importance, is that the iFoP MKs produced are capable of producing live platelet-like particles, to a similar extent as rLV FoP MKs (**Fig 3.13**). iFoP MKs produce a higher percentage of CD42+ particles than rLV FoP MKs, especially true of the –Dox long term iFoP condition. Further characterisation of the MKs and platelet-like particles produced by iFoP is required, but the results so far are encouraging, as iFoP MKs appear to look and behave the same as rLV MKs, which we know are capable of producing functional platelets (Moreau T, 2016).

Most surprisingly, since the presence of FLI1 TG is certain in all cells, is how well erythroblasts were produced using iFoP, and the presence of fully mature, enucleated erythrocytes. Starting with 0.4µg/ml doxycycline and removing it at day 9, was found to produce the highest expansion of erythroblasts by day 19 (**Fig3.10C**). This preliminary work is extremely promising, and is the starting point for being able to optimise conditions, for example using different cytokine combinations or concentrations to help further mature the erythrocytes produced. It also could allow for the TFs to be switched off at an even earlier time-point, removing *FLI1* TG earlier, which may also help to promote maturation.

It is encouraging that results so far support iFoP as an attractive alternative to rLV FoP, as it is a highly reproducible method for generating an increased purity of mature cells, for both MKs and erythroblasts. The production of a stable iPSC line, which has been genetically manipulated to FoP, should facilitate taking this technique from bench to bedside more easily, to achieve the *in vitro* production of mature blood cells for use in transfusion medicine. However, the work presented here highlights how difficult it is to predict what will happen when working with and modifying iPSCs, as each line showed huge variation. Further validation, by karyotyping and genomic sequencing are required, to show that the targeting strategies used did not result in any unwanted off-target effects in the inducible line BobC that did iFoP well.

Results

Chapter 4

Understanding the Heterogeneity of Long-Term Megakaryocyte Cultures

Introduction

Single cell RNA-sequencing (scRNA-seq) is a powerful tool, allowing high-throughput and high-resolution transcriptomic analysis of individual cells within a population. This technology has increased our knowledge of cellular diversity within populations which outwardly appear to be homogenous. Importantly, this technique also allows information to be gathered on very rare cell types. The major advantage of scRNA-seq is that it enables gene expression dynamics to be studied of low level transcripts, which would otherwise be masked in population-averaged bulk analysis (Wills QF, 2013). It is has increased in popularity over recent years due to protocols becoming more robust and sequencing more economical to perform.

An example of how scRNA-seq has been used to dissect direct cellular reprogramming pathways and developmental programs was demonstrated in mouse embryonic fibroblasts (MEFs) reprogrammed towards induced neuronal (iN) cells (Treutlein B, 2016). In this study, high-resolution scRNA-seq data enabled the reconstruction of the programming path taken by MEFs towards iN cells at multiple time-points of reprogramming, after induction of the three transcription factors (TFs); *Ascl1*, *Brn2* and *Myt1l* (Vierbuchen T, 2010). The authors wanted to identify whether a heterogeneous cellular response occurred in MEFs during reprogramming, while also trying to identify mechanisms that might prevent this process from being successful. An important issue regarding the reprogramming of any cell type is whether pre-determined mechanisms, such as epigenetic regulation, prevents a proportion of cells from undergoing reprogramming, or, whether a proportion of cells do not reprogram simply due to inefficient reprogramming procedures.

The findings of that study suggest that all MEFs were capable of reprogramming successfully, as the majority of cells had silenced MEF-associated genes after 48 hours post initiation of the reprogramming protocol. Although this does not rule out epigenetic variation between MEFs, the authors suggest that MEF heterogeneity is unlikely to contribute significantly to reprogramming efficiency. However, they found that silencing of one of the TFs used, *Ascl1*, and divergence from the neuronal differentiation pathway towards an alternative myogenic fate, were both significant factors contributing to the failure of reprogramming. This suggests that intermediate steps in the pathway from MEF to iN cell are unstable, some of which may be governed by epigenetic factors. This study also identified that reprogrammed cells transition through a neuronal precursor cell (NPC) stage, expressing a number of associated NPC genes, between MEF and iN cell stages. This finding was in contrast to what was believed to happen during direct somatic lineage reprogramming, where intermediate progenitor cells states were not expected to arise, unlike during pluripotent stem cell (PSC) directed differentiation (Li XJ, 2005).

A recent single cell study on haematopoietic stem cells (HSCs) revealed a powerful approach to identify the molecular regulators of HSC function by combining scRNA-seq and single cell functional assays, linked by flow cytometric index sorting (Wilson NK, 2015). Studying gene regulation in HSCs has previously been complicated due to there often being contaminating non-HSC cell types present in isolated populations and additionally, due to the identification of functionally different sub-types of HSCs (Challen GA, 2010). One of the caveats of functional assays, often performed on HSCs to assess progenitor potential, is that the transcriptome of the cell cannot be studied once it has differentiated, and as such the status of HSC can only be given retrospectively. However, through the bioinformatic integration of datasets generated by both transcriptome and functional analysis, key molecules associated with durable long-term self-renewal were identified in HSCs, giving insights into the molecular mechanisms controlling these properties and enabling the purity of HSC sorting to be increased.

The forward programming (FoP) approach, pertaining to this chapter, relies on the overexpression of three TFs in PSCs, directing them towards mature megakaryocytes (MKs). FoP initially generates a heterogeneous population of cells when transducing PSCs with recombinant lentivirus (rLV), of which only the cells receiving a mix of TFs are selected for during the protocol, eventually resulting in a highly pure population of MKs. The FoP protocol has been shown to produce a bi-potent progenitor cell population (Chapter 1). We know from clonogenic assays that approximately 1% of cells at day 9/10 of FoP give rise to progenitors and this progenitor potential decreases overtime, by 10 fold between day 21 and day 100 (Moreau T, 2016). MK-FoP cells can be kept in culture for approximately 100 days, after which the culture number declines dramatically, without recovering. This, we believe, signifies the end of the life-span of the bi-potent, or MK-progenitor cell population in the culture, responsible for maintaining such long cultures in the first place. Long-term cultures (above 30 days) contain a mixture of immature and mature MKs, as well as progenitor cells, yet these cell populations remain undistinguishable currently by our methods for cell analysis using flow cytometry. Previous attempts to find other surface markers which could be used to distinguish progenitor cells in the heterogeneous population generated by forward programming have failed. Dr Moreau has sorted MK-FoP cells based on KIT and CD34 expression, two markers associated with HSCs (Matsuoka S, 2001), and found that neither marker is expressed by the majority of cells which give rise to MK or MK progenitor colonies.

We believe that scRNA-seq could help us to answers some of the fundamental questions surrounding FoP. This technique could help us to elucidate the correct ratio, or optimal mix, of TFs required for successful MK-FoP, and improve efficiency of generating MKs from PSCs. It could help us

to identify whether a very small sub-set of cells with an optimal TF mix are responsible for generating long-term cultures, and whether these are heterogeneous. It could also give clues as to the biological mechanisms controlling the pathway from PSC to MK. We have published data to show that cells transitioning from PSC to MK by FoP may at least in part recapitulate the key steps of haematopoiesis in the embryo (Moreau T, 2016). Mesoderm commitment is highly beneficial to MK-FoP outcome, consistent with normal blood cell development. Also, the expression of haemogenic endothelium markers, such as FLK1, CD34 and VE-Cadherin are detected early, from day 2 of FoP. Early expression of CD41 suggests early MK-commitment in MK-FoP, which may be a unique feature of this protocol. Additionally, scRNA-seq could help us to determine the molecular profile of the progenitor cells within MK-FoP cultures, helping us to purify and study these cells. To be able to determine the differences between progenitor and differentiated cell states would enable us to study the development of both cell types in more detail and potentially help us to increase MK yield from cultures which would benefit the clinical aim of this work.

The precise molecular mechanisms governing MK-FoP remain unknown as we have not been able to study these questions at the resolution required to obtain informative data, especially since rLV FoP is uncontrollable and untraceable using current cellular markers. The first step to being able to perform scRNA-seq on MK-FoP cells was to address the issue of transgene traceability in rLV-transduced cells. In order to achieve traceability, we needed to generate rLV vectors with marker genes, for which a different fluorescent protein was cloned into each of the single TF viral backbones. The fluorescent proteins eGFP, dTomato and LSSmOrange were chosen and are followed by a 2A sequence in the vectors containing GATA1, TAL1 and FLI1 respectively, to achieve a single mRNA transcript (as described in Chapter 2). The resulting vectors are collectively known as the 'Rainbow' vectors.

Once the Rainbow vectors were produced, we performed a number of MK-FoP experiments to determine whether MKs produced in long-term cultures were a mixture of cells (containing several combinations of rainbow colours) or a homogenous population of cells which had received primarily all three TFs (and therefore contain all three rainbow colours). **Diagram 4.1** shows the changes in Rainbow cell dynamics, of the 8 possible combinations, found in MK-FoP cells of a long-term experiment performed by Dr Moreau. At the start of FoP cells were predominantly positive for all three rainbow colours, indicating cells which received all three FoP TGs. However, long-term cultures did not contain a single population of rainbow cells and instead, over time, the number of possible combinations of colours increased and the total 8 different populations were observed. We believed these different populations arise from cells which have undergone silencing of different rLVs, most

likely as a consequence of insertion into regions of the genome that are not permissive for cell division.

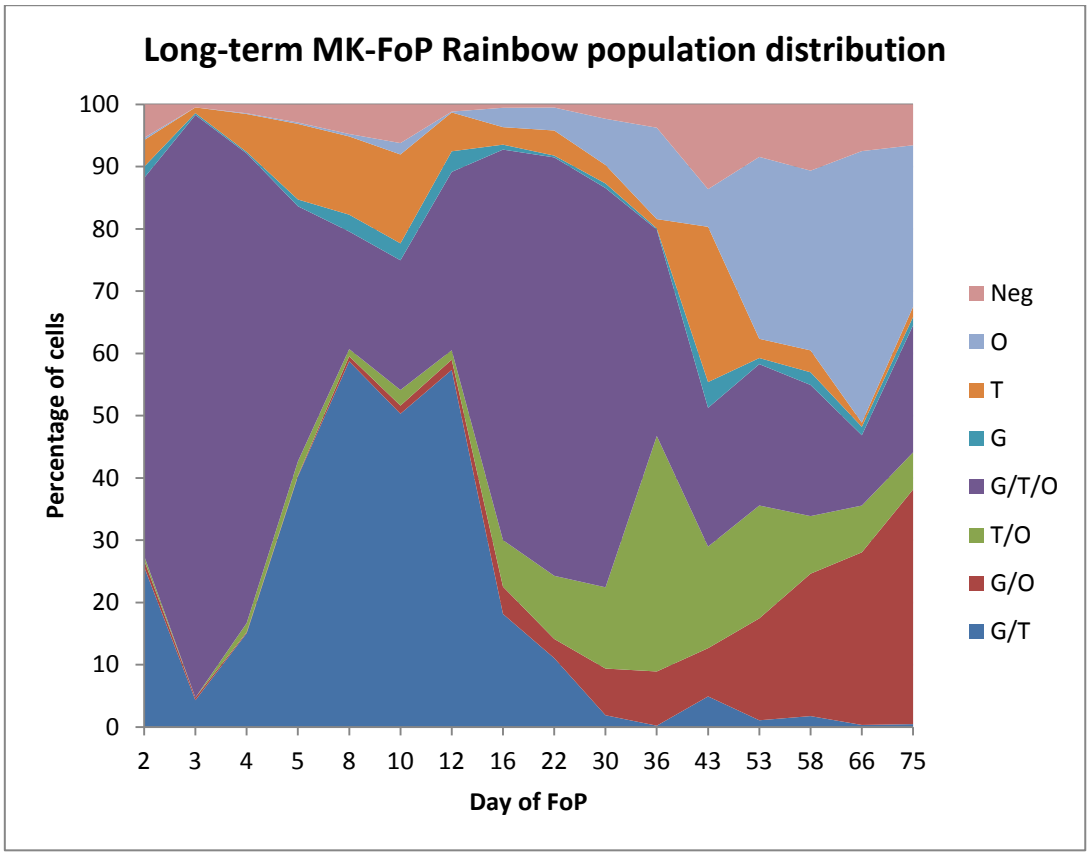


Diagram 4.1 *Long-term Megakaryocyte Rainbow Population Distributions. The percentage of cells for the 8 combinations of rainbow colours found in a long-term Rainbow transduced MK-FoP experiment in BobC cells. O= LSSmOrange_FLI1+, G= GFP-GATA1+, T= dTomato-TAL1+, Neg= no fluorescence detected.*

Chapter Overview

The following chapter describes results where the focus was to gain a better understanding of the heterogeneity of cells in long-term MK-FoP cultures. It describes the establishment of single cell progenitor assays, designed to try and identify the progenitor cell population. It also describes the results of a scRNA-seq experiment performed on the 8 populations found in long-term rainbow MK-FoP cells at day 40. Day 40 was chosen, based on the experimental data shown in Diagram 4.1, as this time-point showed a good range of different Rainbow populations and would allow us to address whether these different populations contained different cell types.

Materials and Methods

Rainbow Vectors

The Rainbow Vectors describe the three single rLVs used for normal forward programming, with the addition of a unique fluorescent reporter protein to each. The fluorescent markers eGFP, dTomato and LSSmOrange were chosen due to their excitation and emission wave lengths being far apart, to easily detect by flow cytometry, while still being able to stain cells with the common markers used during FoP (CD235, CD41 and CD42). Each vector was generated from the pWPT backbone (Addgene 12255) with the coding sequence of the FoP TFs GATA1, TAL1 and FLI1 (Refseq NM_002049.3, NM_003189.5 and NM_002017.4 respectively) downstream of the fluorescent reporter and E2A sequence. The three rainbow vectors were generated by Annett Muller and Niloufar Hojatoleslami using standard molecular cloning strategies:

- pWPT-GFP-GATA1 (**Map 4.1**)
- pWPT-dTomato-TAL1 (**Map 4.2**)
- pWPT-LSSmOrange-FLI1 (**Map 4.3**)

A mixture of all three Rainbow rLVs was produced commercially by Vectalys to high titre and purity for experiments described in this chapter.

iPSC Lines ued

For setting up single cell assays, the iPSC lines Qolg and BobC were used, as these lines were being routinely used for FoP in the lab at the time and both FoP well. After the single cell protocol was established, only BobC was used for Rainbow and sorting experiments.

Single Cell Progenitor assays

I optimised the flow cytometry sorting of single cells into 96 well plates in order to perform clonogenic assays. Prior to cell sorting, 96 round-bottom well plates (Corning) were prepared by aliquoting 100µl MK media (TPO 20ng/ml, SCF 25ng/ml) plus 1% Penicillin and Streptomycin (PenStrep, Life Technologies) into all wells. Serum free methylcellulose-based medium for human cells (Methocult H4236, Stemcell Technologies), plus TPO 20ng/ml, SCF 25ng/ml and 1% Penicillin and Streptomycin can also be used, by carefully pipetting 100µl per well. Cells were prepared by staining with CD41a APC-H7 (1:100), CD235a Pe-Cy7 (1:100) and CD42b APC (1:20) (all BD-Pharmingen) for 20 min at RT in the dark. Cells were washed with 1ml PBS and spun at 300g, for 5 min before re-suspending in 0.5ml PBE.

Cells were sorted on the BD AriaIII Cell Sorter, a 16 fluorescent channel cell sorter, at the Cambridge NIHR BRC Cell Phenotyping Hub. The megakaryocytic cell line CHRF was transduced with single reporter gene rLVs and used to set up gating and compensation for florescent reporters on the AriaIII.

The following configuration was used for detecting stained Rainbow transduced cells.

Excitation laser (nm)	Filter	Detects
633	660/20	APC
633	780/60	APC-H7
561	780/60	Pe-Cy7
561	610/20	dTomato
488	530/30	eGFP
405	610/20	LSSmOrange

After single cell sorting into 96 well plates, remaining cells were sorted into bulk populations, into 5ml Polypropylene tubes (Sigma-Aldrich) in PBS + 10% FBS. Bulk sorted populations were used to perform CFU assays (as described in the main Materials and Methods section), to check for the clonogenic potential of the cell population.

Single Cell RNA-seq

Several sorts were performed on the AriaIII, to ensure single cells were being sorted into each well of a 96 well plate and each well checked under the microscope until satisfactory sorting had been achieved and the machine settings recorded for the single cell RNA-seq sort. Single cell RNA-seq (scRNA-seq) was performed with the help and guidance of Sonia Nestorowa and Winnie Lau (both Göttgens group, Department of Haematology). The protocol provided is adapted from the Smartseq2 protocol (Picelli S, 2014). All steps of the scRNA-seq protocol were performed in a UV-sterilised hood with laminar flow and all surfaces kept free of RNase and DNA. The same hood was used for all steps except the cDNA amplification step. All centrifugation was performed at 4^{o}C for pre-amplification steps.

1. **Single Cell Lysis**

Cell lysis buffer was prepared by adding 1µl SUPERase-In RNase Inhibitor (20U/L, Thermo Fisher Scientific) to 19µl 0.2% (vol/vol) Triton X-100 solution (Sigma Aldrich, cat # T9284). 2.3µl was aliquoted per well of a 96 V-bottom well plate, keeping on ice. Cells were sorted in the lowest possible volume, as described above, and index data collected for CD235, CD41, CD42, eGFP, dTomato and LSSmOrange. The plate was sealed with an adhesive lid, vortex briefly and spun down

at 300g for 1 min at 4°C. At this stage the plates were frozen at -80°C, and can be stored for up to six months. An additional 2 test plates were also prepared, with just 1 column of cells being sorted, with no index data collect, for use in downstream steps to check the number of PCR cycles required which differs depending on the amount of input RNA. Single cells were sorted into 7 of the 8 wells in each test plate and 30 cells were sorted into the final well. Test plates were taken up to the stage of cDNA library quality check. One was used to test 19 PCR cycles, and one to test 21 cycles in the PCR amplification step. For the cells sorted in this experiment, 18 cycles was found to be sufficient and it is best to use as few PCR cycles as possible, as not a lot of cDNA is required for library preparation.

2. Reverse Transcription

The annealing mixture was prepared as follows:

	1 x test column (8 single cells)	**1 x 96 well plate**
ERCC 20x (*)	1 µl	10 µl
Oligo-dT 100 µM	1 µl	10 µl
dNTP 10 mM	10 µl	100 µl
dH_2O	8 µl	80 µl
	20 µl	**200 µl**

*ERCC dilution is cell type/batch specific. For this experiment a dilution of 1:300,000 was used. ERCC RNA Spike-In Mix (Invitrogen, cat# 4456740), OligodT30VN dissolved in TE buffer (Sigma, with HPLC purification), dNTP mix (10mM, Thermo Fisher, cat# 10319879).

2µl annealing mix was aliquoted per well by lightly touching the edge of the well and was followed by centrifugation at 700g, 1 min. Plates were incubated at 72°C for 3 min in a GS4 multi-block thermocycler (G-Storm) before being placed immediately on ice and re-centrifuged at 700g, 1 min. This step ensures the oligo-dT primer is hybridised to the poly(A) tail of the mRNA molecule.

The reverse transcription mix was prepared as follows:

	Per well	**1 x 96 well plate**
Superscript II RT (200 U/ µl)	0.5 µl	50 µl
RNase inhibitor (20 U/ µl)	0.25 µl	25 µl
5 x Superscript II first strand buffer	2 µl	200 µl
100 mM DTT	0.5 µl	50 µl
5 M Betaine	2 µl	200 µl
1 M $MgCl_2$	0.06 µl	6 µl
Oligo TSO (100 µM)	0.1 µl	10 µl
dH_2O	0.29 µl	29 µl
	5.7 µl	**570 µl**

Superscript II Reverse Transcriptase (200U/ml, Thermo Fisher, cat # 18064-014), Superscript II first strand buffer (5x, Invitrogen, cat # 18064-014), Dithiothreitol (DTT, 1M, Invitrogen, cat # 15508013),

Betaine (BioUltra ≥99.0%, Sigma-Aldrich, cat # BO300), Magnesium chloride (1M, Thermo Fisher, cat # AM9530G, oligo TSO (dissolved in TE buffer, Exiqon).

5.6µl reverse transcription mix was added per well, before centrifugation at 700g, 1 min. Plates were then thermocycled following the conditions described in **Table 4.2**. Centrifugation was then repeated.

3. PCR Pre-Amplification

The PCR mix was prepared as follows:

	1 well	**1 x 96 well plate**
KAPA HiFi Hotstart ReadyMix (2x)	12.5 µl	1250 µl
IS PCR primer (10 µM)	0.25 µl	25 µl
dH_2O	2.25 µl	225 µl
	15 µl	**1500 µl**

KAPA HiFi Hotstart ReadyMix (2x, KAPA Biosystems, cat # KK2601), IS PCR primer (dissolved in TE buffer, Sigma Aldrich, HPLC purified).

15µl PCR mix was added per well and the plate was centrifuged at 700g, 1 min. PCR conditions are described in **Table 4.3.** At this stage PCR products can be stored at -20^{o}C or -80^{o}C for over 6 months.

4. PCR Purification

The following steps were performed manually for test plates or using the Biomek FxP Laboratory automation workstation (Beckman Coulter) for sample plates. Agencourt AMPure XP beads (Beckman Coulter, cat # A63881) were equilibrated to room temperature 15 mins prior to starting and were vortexed briefly. 16.25µl Ampure XP beads were added to each well, resulting in a 1:06 sample:bead ratio. Sample and beads were homogenised by pipetting up and down 10 times. Samples were transferred to a 96 well plate compatible for use with a magnet stand. Samples were incubated for 8 mins at RT to allow DNA to bind to the beads. The plate was then placed onto the magnet stand and incubated for 5 min, to allow beads to collect at one corner of the well and leave the solution clear. The clear supernatant was removed carefully without disturbing the beads. Beads were then washed with 200µl freshly prepared 80% ethanol (vol.vol), incubated for 30 secs before the ethanol was removed. The wash step was repeated and samples allowed to dry completely for 5 min at RT, until cracks appear on the surface of the beads. 22µl of EB solution (10mM Tris-Cl, pH 8.5, Qiagen, cat # 19086) was added to resuspend beads by pipetting 10 times. The plate was removed from the magnet and incubated for 2 min at RT. The plate was replaced on the magnet and

incubated for 2 min, until the solution was clear and beads had collected at one side. 20µl of the supernatant was collected, without disturbing the beads, and transferred to a fresh 96 well plate.

5. Quality check of cDNA Library

The size distribution was checked for 11 samples per plate using an Agilent High-Sensitivity DNA Chip (Agilent, cat # 5067-4626), run on the Agilent Bioanalyzer System.

6. cDNA Library Quantification

Library quantification was performed using the Scientific Quant-iT^{TM} PicoGreen double stranded DNA assay kit (Thermo Fisher, cat # P7589), following manufacturers protocol.

7. Library Preparation

Tagmentation was carried out using the Nextera XT DNA sample preparation kit (Illumina, cat # FC-131-1096), following the Fluidigm tagmentation protocol. The Nextera XTDNA library prep kit is optimised for 1ng total input DNA. The Fluidigm protocol recommends an optimal concentration of 0.1-0.3ng/µl cDNA per single cell. The majority of samples were diluted between 0.1-0.15ng/µl in EB buffer. The NT buffer was thawed at RT and vortexed to resuspend precipitates. Tagment DNA buffer was also thawed at RT, inverted to mix and briefly spun in a microcentrifuge. Sample plates were kept on ice. The following reagents were added to a 1.5ml Eppendorf tube to make up the Pre-Mix:

Reagent	Volume per sample (µl)	X 96 well plate (10 % overage)	4 x 96 well plates
Tagmentation DNA buffer	2.5	264	1056
Amplification Tagment Mix	1.25	132	528
Sample	1.25		
Total	**5.0**		

The Pre-Mix was vortexed briefly and equal volumes aliquoted into each tube of a 8-tube strip. 3.75µl of Pre-Mix was added to each well of a new, non-skirted 96 well plate ('Library prep' plate), using an 8-channel pipette. 1.25µl of sample was added quickly at RT to the 'Library prep' plate, before being sealed and centrifuged at 2000rpm, for 30 secs to remove bubbles. The 'Library prep' plate was then added to the thermocycler for 10 mins at 55^{o}C, then held at 10^{o}C. NT buffer was aliquoted into each tube of an 8-tube strip, before 1.25µl was added quickly to each well in order to neutralise the tagmented samples. The plate was sealed and centrifuged at 2000rpm, 1 min. The Nextera PCR Master Mix (NPM) was aliquoted to each well of a 8-tube strip, before 3.75µl was added to each well. The Nextera XT 96-Index kit (384 samples, Illumina, cat # FC-131-1002) was used

to index each well. 1.25µl of Index Primer 1 (N701-N712) was added to the corresponding well of each row of the 'Library prep' plate using an 8-channel pipette. 1.25µl of Index Primer 2 (S517 and S502-S508) was added to the corresponding well of each column of the 'Library prep' plate using a 12-channel pipette, as described in **Table 4.4**. Fresh Index Adapter Replacement Caps (Illumina, cat # 15026762) were used for Index primers after each use. It is essential to double check Index order, if you wish to integrate metadata later. Plates were sealed with adhesive film and centrifuged at 2000rpm, 1 min. Plates were placed into the thermocycler to perform the following PCR amplification as described in **Table 4.5**. At this stage amplified products can be stored at -20^{o}C for long-term storage.

8. **Library Pooling and Clean-up**

Agilent AMPure XP beads were warmed to RT and vortexed briefly. Samples were pooled into a 1.5ml Eppendorf/ 96 well plate, then divided equally into a total of 3 Eppendorfs. Three bead volumes were tested, 0.6, 0.7 and 0.8% of the total pooled volume, in order to choose the best sample based on the final Bioanalyser results, for quantification and sequencing. Beads and pooled samples were mixed well by pipetting up and down 5 times, before a 5 min incubation at RT. Tubes were place on a magnetic stand for 2 mins and supernatant removed carefully without disturbing the beads. Beads were washed with freshly prepared 70% ethanol and incubated for 30 secs before removing supernatant. Wash steps were repeated, before allowing beads to air dry completely for 10-25 mins. Samples were removed from the magnet, eluted in 50µl EB solution, vortexed and incubated for 2 mins at RT. Samples were replaced on the magnet and allowed to stand for 2 mins, before the entire supernatant was removed to a fresh 1.5ml Eppendorf.

9. **Quality Check the Amplified Pooled Library**

The library size distribution was checked on the Agilent Bioanalyser. Pooled libraries were then quantified using the KAPA qPCR Quantification kit (KAPA Biosystems, cat # KK4824), following manufacturer protocol using a black, flat-bottomed, medium binding Fluotrac 384-well Microplate (Greiner Bio-One, cat # 781076).

10. **Sequencing**

Single end (50bp) sequencing was performed on one lane of the HiSeq-4000 (Illumina) per 96 cells at the CRUK Cambridge Institute Genomics Core facility.

scRNA-seq Computational Analysis

Computational analysis was performed by James Baye (Cambridge Stem Cell Institute), under the supervision of The Göttgens lab (Department of Haematology, University of Cambridge). All analysis

was performed in R 3.4.0. Raw sequencing reads (fastq files) were aligned with the R package *GSNAP* (Wu TD, 2010) to the human Ensembl reference genome GrCh38.81, with the addition of the 92 ERCC spike-in genes and the three fluorescent reporters eGFP, LSSmOrange and dTomato expressed exogenously in FoP-MKs. Read counts for each gene were then determined with HTSEQ-Count31 from the mapping files. The R package *scater* was used to perform rigorous pre-processing, quality control, normalization and visualization of scRNA-seq data (McCarthy DJ, 2017). Quality check consisted of discarding all cells with less than 500,000 read counts and outlier cells showing a high proportion of mitochondrial RNA reads or spike-in reads. All non-captured genes (read count equal to zero for all cells) were removed from further analyses. A total of 192 cells were sequenced (24 for each of the 8 Rainbow populations). 81 cells were removed from downstream analysis, 30 cells from the GFP+ plate and 51 cells from the GFP- plate. Read counts were then normalised to counts per million (CPM). Unsupervised clustering of cells based on their whole RNA expression was performed with the R package *SC3* (Kiselev VY, 2017). Two to five cluster groups were attempted with three and four clusters showing the best consensus. Differential gene expression analysis based on the negative binomial distribution were conducted with the R package *DESeq2,* with a maximum false discovery rate (FDR) of 1% and gene expression fold change greater than 2 (Love MI, 2014). Gene ontology results were obtained from the Gene Ontology Consortium database (Blake JA, 2015), filtering by biological processes with p-values inferior to 0.01.

Materials and Methods Tables

Table 4.1 Oligonucleotides used for Single cell RNA-seq library preparation

Oligo/Primer name	Sequence 5'-3'
TSO	AAGCAGTGGTATCAACGCAGAGTACATrGrG+G
Oligo-dT30VN	AAGCAGTGGTATCAACGCAGAGTAC(T30)VN
IS PCR oligo	AAGCAGTGGTATCAACGCAGAGT

Table 4.2 Thermocycling conditions for reverse transcription

Cycle	**Temperature (°C)**	**Time**	**Purpose**
1	42	90 mins	RT and template switching
10	50	2 mins	Unfolding of RNA secondary structures
	42	2 mins	Completion/continuation of RT and template switching
1	70	15 mins	Enzyme inactivation
-	4	Hold	Safe storage

Table 4.3 PCR Pre-amplification step PCR conditions.

Cycles	**Temperature (°C)**	**Time**	**Purpose**
1	98	3 min	Denature
21	98	20 secs	Denature
	67	15 secs	Anneal
	72	6 mins	Extend
1	72	5 mins	Extend
-	4	hold	-

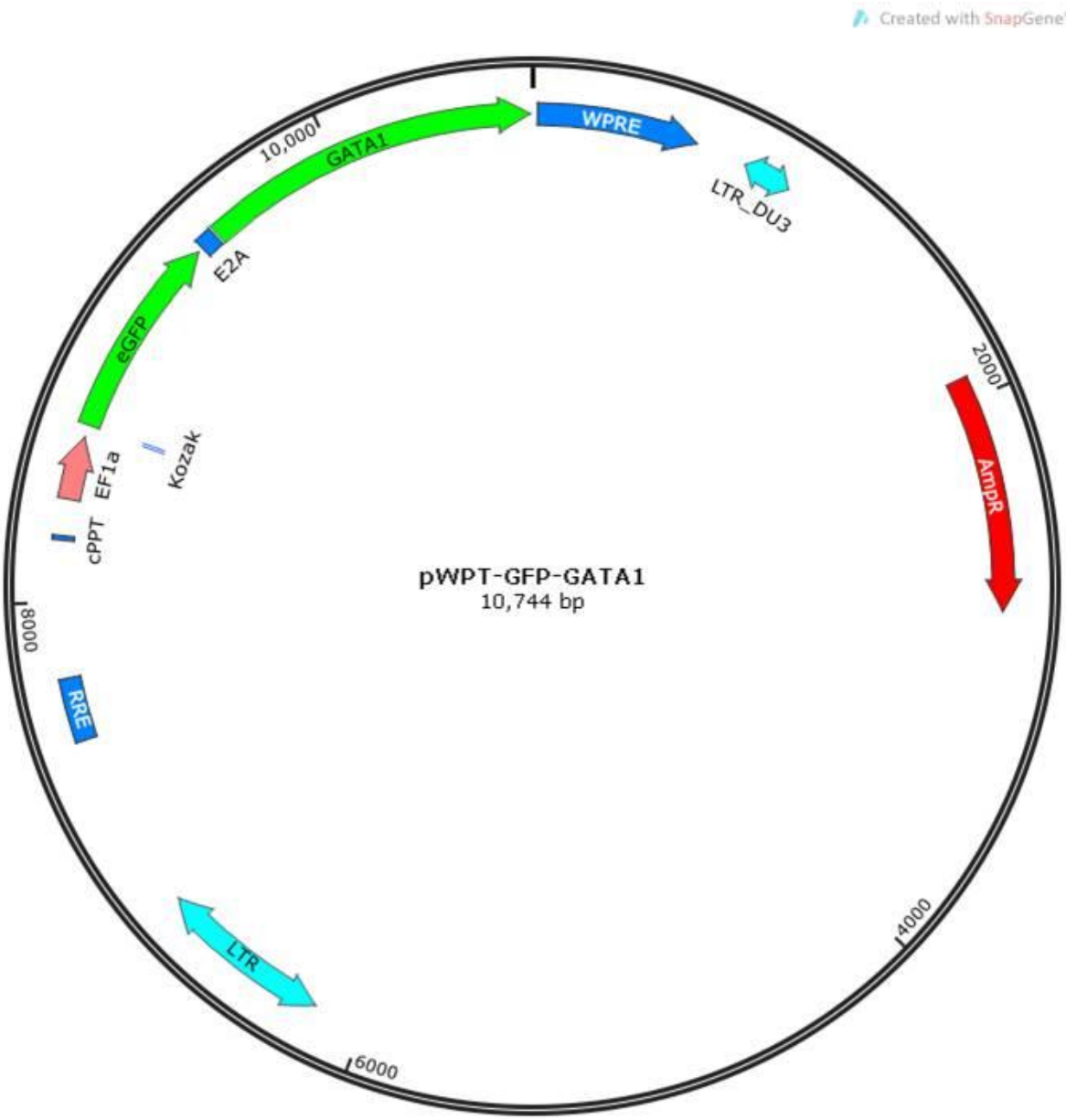
Table 4.4 Library Prep Plate Layout

		N701	N702	N703	N704	N705	N706	N707	N708	N709	N710	N711	N712
		1	2	3	4	5	6	7	8	9	10	11	12
S517	A	S517/ N701	S517/ N702	S517/ N703	S517/ N704	S517/ N705	S517/ N706	S517/ N707	S517/ N708	S517/ N709	S517/ N710	S517/ N711	S517/ N712
S502	B	S502/ N701	S502/ N702	S502/ N703	S502/ N704	S502/ N705	S502/ N706	S502/ N707	S502/ N708	S502/ N709	S502/ N710	S502/ N711	S502/ N712
S503	C	S503/ N701	S503/ N702	S503/ N703	S503/ N704	S503/ N705	S503/ N706	S503/ N707	S503/ N708	S503/ N709	S503/ N710	S503/ N711	S503/ N712
S504	D	S504/ N701	S504/ N702	S504/ N703	S504/ N704	S504/ N705	S504/ N706	S504/ N707	S504/ N708	S504/ N709	S504/ N710	S504/ N711	S504/ N712
S505	E	S505/ N701	S505/ N702	S505/ N703	S505/ N704	S505/ N705	S505/ N706	S505/ N707	S505/ N708	S505/ N709	S505/ N710	S505/ N711	S505/ N712
S506	F	S506/ N701	S506/ N702	S506/ N703	S506/ N704	S506/ N705	S506/ N706	S506/ N707	S506/ N708	S506/ N709	S506/ N710	S506/ N711	S506/ N712
S507	G	S507/ N701	S507/ N702	S507/ N703	S507/ N704	S507/ N705	S507/ N706	S507/ N707	S507/ N708	S507/ N709	S507/ N710	S507/ N711	S507/ N712
S508	H	S508/ N701	S508/ N702	S508/ N703	S508/ N704	S508/ N705	S508/ N706	S508/ N707	S508/ N708	S508/ N709	S508/ N710	S508/ N711	S508/ N712

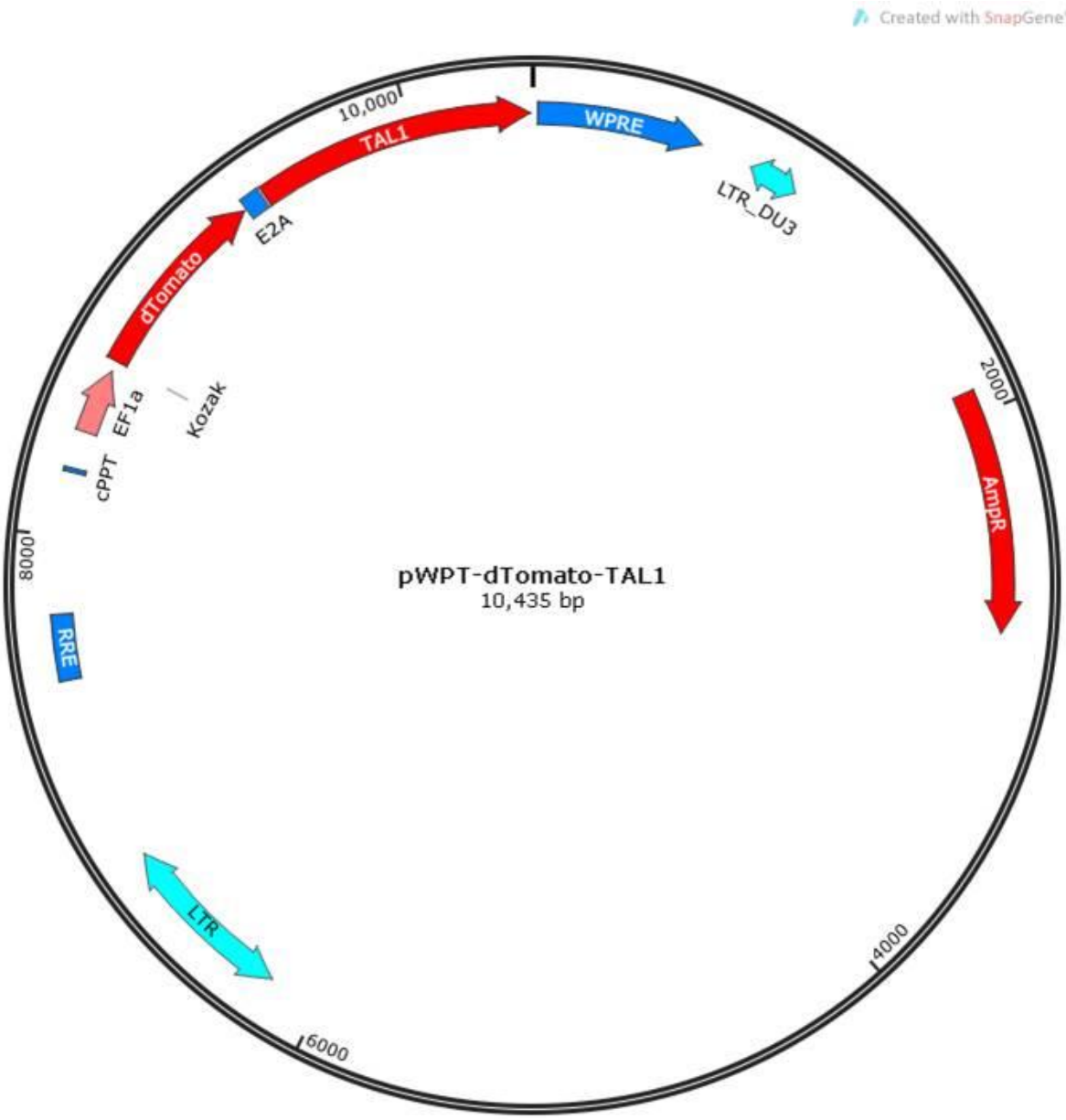
Table 4.5 PCR conditions for Library Prep plate

Temperature (°C)	**Time**	**Cycles**
72	3 mins	1
95	30 secs	1
95	10 secs	12
55	30 secs	
72	60 secs	
72	5 mins	1
10	hold	-

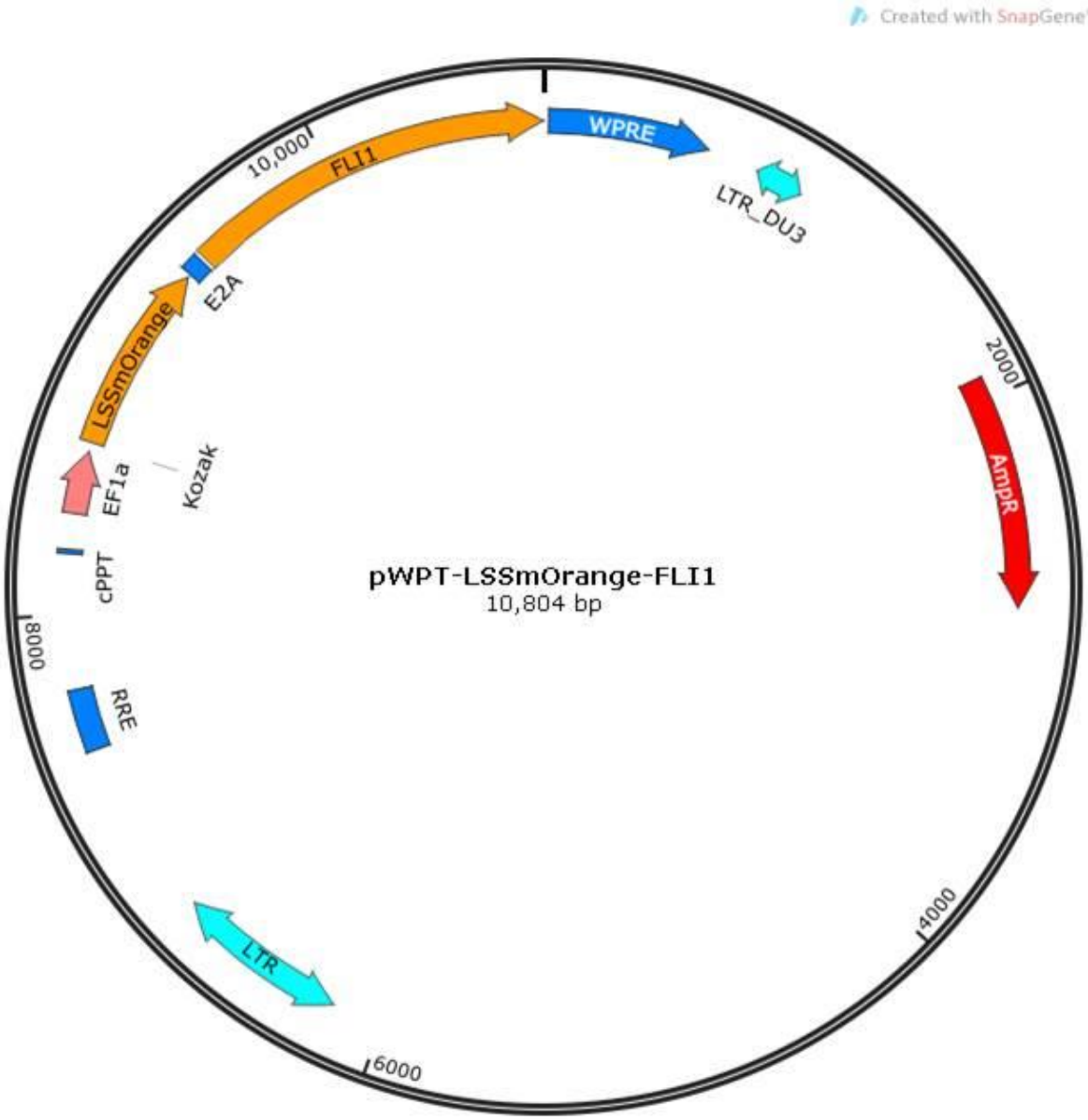
Chapter 4: Vector Maps



Map 4.1 pWPT-eGFP-GATA1.



Map 4.2 pWPT-dTomato-TAL1.



Map 4.3 pWPT-LSSmOrange-FLI1.

Results

In the results section the following cell type definitions are used; CD41+/CD235+ cells are bi-potent progenitors, CD41+/CD235- cells at day 10-13 are MKs, and CD41+/CD42+ cells are mature MKs. This results section describes only cells produced by MK-FoP, using TPO containing medium.

Establishing a Protocol for Single Cell Progenitor Assays

Results from a preliminary single cell sort, performed on day 13 FoP Qolg cells, transduced with the 3TFs recombinant lentivirus (rLVs, as previously described), are shown in **Fig 4.1**. Cells were sorted into two discreet populations, CD41+/CD235- and CD41+/CD235+ (**Fig 4.1A**). CFU progenitor assays were performed on the bulk sort of these two populations, as well as on unsorted cells to ensure the process of sorting did not adversely affect progenitor potential. Cells were seeded into non-enriched methylcellulose supplemented with 100ng/ml TPO and 25ng/ml SCF, in the absence or presence of Penicillin-Streptomycin (PenStrep, P/S) when cell numbers allowed, to test whether P/S adversely affects progenitor potential. MK progenitor colonies and MK colonies were counted, with the average colony number per cell seeded shown in **Fig 4.1B**. The CD41+/CD235+ population had higher potential for both colony types, compared to the CD41+/CD235- population. The progenitor potential was slightly increased in the unsorted population. The addition of P/S had no effect on the CD41+/CD235+ population and little effect on the unsorted population. Due to this and a number of contaminations after using the shared sorting facility, all further experiments were performed with the addition of P/S.

Sixty individual cells from the CD41+/CD235+ and CD41+/CD235- population were sorted for single cell progenitor assays into liquid (liq) medium and non-enriched methylcellulose (MC) (both supplemented with TPO and SCF, as previously described). Both conditions were tested in the presence or absence of P/S. **Fig 4.1C/D** and **E** show the results from sorting the CD41+/CD235+ population only. The single cell seeding of CD41+/CD235- cells resulted in colonies after 14 days of culture. Colony outcome was better in liquid medium, compared to methylcellulose (**Fig 4.1C**). A slight decrease in colony number and colony size was observed when single cells were cultured in the presence of P/S, however, to reduce the risk of contamination this condition was chosen for all further experiments performed on single cells. MK progenitor colonies and MK colonies produced from single cells in liquid medium (**Fig 4.1D**) were similar in appearance to those observed in bulk methylcellulose cultures (Chapter 1, **Fig 1.6A**). No MK progenitor colonies were observed from single cells seeded into methylcellulose. The MFI of CD41 and CD235, recorded by index sorting, for all single cells seeded in all conditions is shown in **Fig 4.1E**. This experiment showed that sorted single

cells could give rise to progenitors and all data presented hereafter come from experiments performed in liquid culture medium, with the addition of P/S.

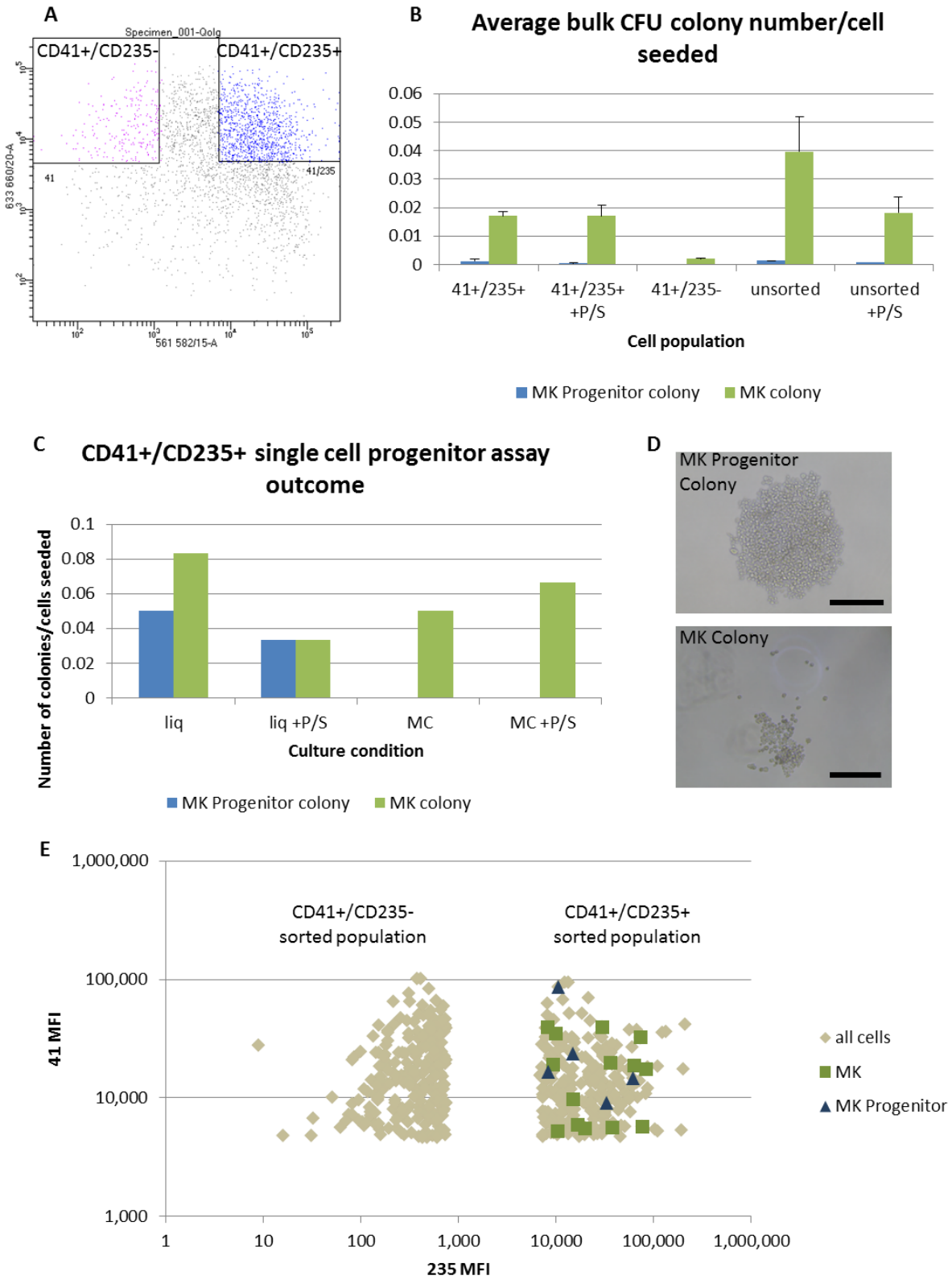


Fig 4.1 Determining the megakaryocyte progenitor cell signature. Day 13 forward programmed Qolg cells, transduced with 3TF lentivirus, were sorted by FACS based on cell surface expression of CD41 and CD235. On the day of sorting, cells were either cultured in a colony forming assay (CFU) or in

liquid media (both containing TPO and SCF to promote MK growth) and left for 14 days for colonies to grow before counting. **A**) Sorting strategy shows discreet CD41+/CD235- and CD41+/CD235+ cell populations, which were sorted.**B**) The average number of MK progenitor colonies and MK colonies produced/cell seeded in CFU assays from bulk sorted populations, and unsorted cells, in the presence or absence of PenStrep (P/S). **C**) The average number of colonies produced/cell seeded in single cell progenitor assays, from the CD41+/CD235+ population only. Both liquid (liq) and methylcellulose (MC) media conditions, in the presence or absence of P/S, were tested. **D**) An MK progenitor colony and MK colony formed from single cells in liquid medium. **E**) Mean fluorescence intensity (MFI) of CD41 and CD235, recorded by index sorting single cells seeded in progenitor assays. Cells which produced no colony (grey), MK colonies (green) and MK progenitor colonies (blue) are shown. MFI threshold values: 4500 for CD41, 7000 for CD235. Error bars = range. Scale bars= 50µm. N=1.

We hypothesised that the CD41+/CD235+ population contained the bi-potent and MK progenitor cells produced by forward programming (FoP). In order to test this, bulk sorting was performed on all 4 populations produced by FoP; CD41-/CD235-, CD41+/CD235-, CD41+/CD235+ and CD41-/CD235+. BobC cells transduced with the Rainbow mix of rLVs were sorted at day 9. The progenitor assay results of both the bulk and single cell sorting of these populations is shown in **Fig 4.2**. This sort did not divide the populations into discreet groups, which increases the risk of contaminating cell types being sorted (**Fig 4.2A**). The average colony number per cell seeded from the bulk sort shows that the majority of colonies were produced by cells originating in the CD41+/CD235+ population (**Fig 4.2B**), which gave higher colony numbers than the unsorted population.

Three hundred single cells were sorted from the CD41+/CD235- and CD41+/CD235+ populations. Five colonies were recorded in total after 14 days of culture. The colony types are shown, along with the index data (positive or negative for each recorded marker) for each originating single cell (**Fig 4.2C**). All originating cells were positive for CD41, CD42, dTomato and GFP. Three cells were negative for CD235 only and one cell was negative for LSSmOrange only. Flow cytometry was performed on the resulting colonies. All colonies contained a high percentage of CD41 positive cells and CD41 was gated on to show the percentage of CD235, CD42, dTomato, LSSmOrange and GFP expressing cells (**Fig 4.2D** and **Fig 4.2E**). A high percentage of cells stained triple positive for CD41, CD42 and CD235 in all colonies (>69%). A smaller percentage of cells stained for CD41 and CD42 only in all colonies, indicating mature MKs. All colonies contained cells expressing at least one of the three fluorescent proteins to varying levels, for example between 18-89% for GFP. This data shows that not all CD41 expressing cells were positive for all three fluorescent proteins.

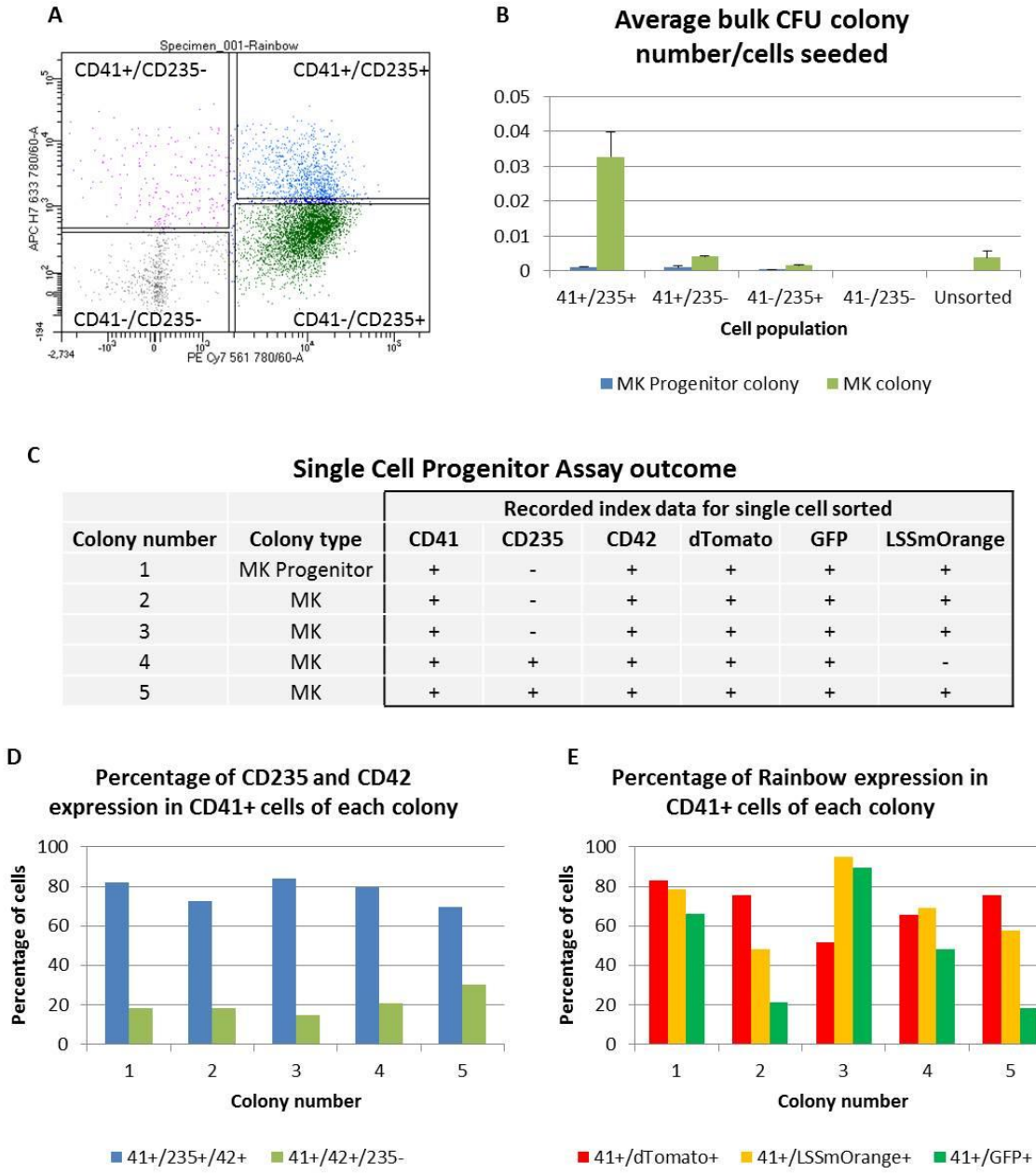


Fig 4.2 Sorting strategy of Rainbow transduced BobC cells. Forward programmed BobC cells were sorted on day 9 of FoP, after transduction with the Rainbow-FoP mix of lentivirus, based on the surface expression of CD41 and CD235. After sorting, cells were cultured for 14 days in colony forming assays (CFU) or in liquid culture to promote MK growth, before colonies were counted. **A**) Cells were sorted based on CD41 and CD235 expression into 4 populations; CD41-/CD235-, CD41+/CD235-, CD41+/CD235+ and CD41-/CD235+. **B**) Average MK progenitor colony and MK colony number/cell seeded from the bulk sort of the 4 populations, plus unsorted cells, in CFU assays. **C**) The number and type of colonies produced by progenitor assays on single cells. The recorded index data (+ or - =positive or negative) for CD42, CD235, CD41 and the three fluorescent markers of the single cell sorted which gave rise to each colony is shown. **D**) The 5 colonies derived from single cells

were analysed by flow cytometry. The percentage of CD41 positive cells co-expressing CD42 and/or CD235 are shown. **E**) The percentage of CD41 positive cells with co-expression of dTomato_TAL1, LSSmOrange_FLI1 and GFP_GATA1 are shown. Error bars= range. N=1.

Interrogating the Progenitor Potential of a Long-Term Culture

Next, a long-term Rainbow MK-FoP experiment, performed in the BobC cell line, was sorted at day 40 for both single cell progenitor assays and single cell (sc) RNA-seq. The remaining results for this chapter relate to this particular sort. The results of the single cell progenitor assays are shown in **Fig 4.3** and **Fig 4.4**. The percentage of the 8 possible Rainbow populations, over the 41 day culture period, shows that at day 9 the population of cells was more varied than at day 22 or 41 (**Fig 4.3A**). This was in contrast to what we had observed previously (**Diagram 4.1**). At day 9 the most common population was the GFP_GATA1+/dTomato_TAL1+ combination, while from day 22 this population was reduced and the dTomato_TAL1+/LSSmOrange_FLI1+ combination and triple positive combination were predominant. At day 44 the triple positive combination had increased further, becoming the most commonly observed. The percentage of cell types present at day 41 are shown by representative dot plots (**Fig 4.3B**). The percentage of mature MKs was high (77%). CD235 expression was higher than expected at this stage of programing, with a small percentage of cells (15%) staining triple positive for CD41/CD235/CD42. The majority of mature MKs however did not co-express CD235.

For the single cell progenitor assay, cells were sorted based only on CD235 expression which we had determined to be a probable marker of colony forming progenitors. CD235+ and CD235- cells were sorted and index data acquired for each of the three fluorescent TGs, CD42 and CD235. The outcome for the single cell progenitor assay shows that both populations gave rise to both MK progenitor colonies and MK colonies and that there was a higher frequency recorded for both colony types in the CD235+ population (**Fig 4.3C**).

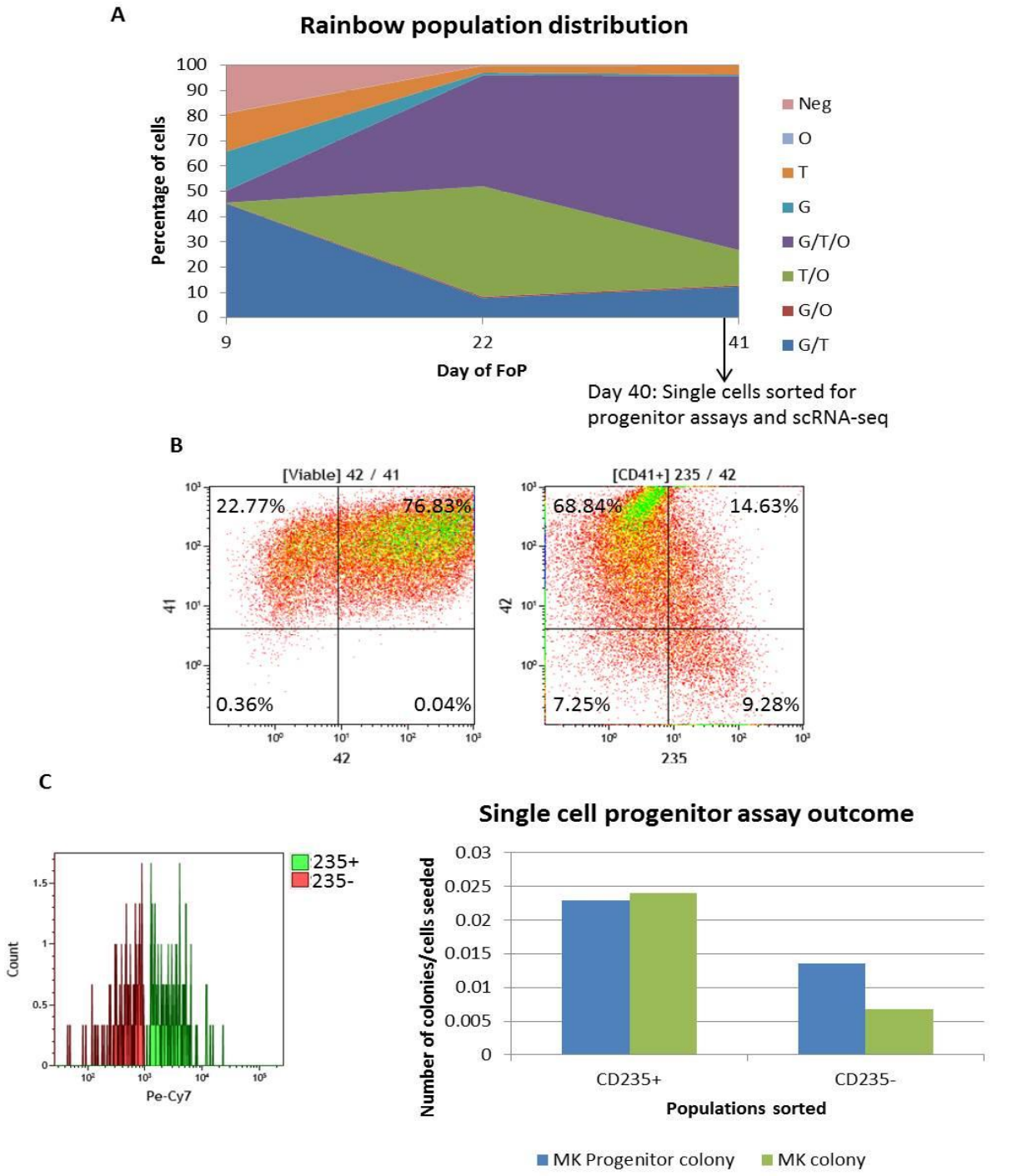


Fig 4.3 Sorting long-term Rainbow megakaryocytes. Rainbow transduced BobC cells were cultured for 41 days in MK media and sorted on day 40 for single cell progenitor assays and single cell RNA-seq. Progenitor assays were performed in liquid media containing TPO and SCF and colonies counted after 14 days of growth. **A**) The percentage of different Rainbow populations present in viable cells during a 41 day MK-FoP experiment. **B**) Representative dot plots show CD41/CD42 and CD41+/CD235/CD42 expression at day 41. **C**) Rainbow cells were sorted as single cells in a progenitor assay based on CD235 expression at day 40. An overlay of the CD235 signal from a plate sorted with

CD235+ cells and CD235- cells (left) and the number of resulting MK progenitor colonies and MK colonies per cell seeded (right). O= LSSmOrange-FLI1, T= dTomato-TAL1, G= GFP-GATA1, Neg= Negative for fluorescent markers. N=1.

The MFI of CD42, CD235 and the fluorescent proteins for each cell, acquired by index sorting, is shown in **Fig 4.4**. Due to a technical error CD41 expression was not recorded properly at the time of the sort and subsequently is not shown. The average mean fluorescence intensity (MFI) of CD42 for cells that gave rise to MK and MK progenitor colonies was lower than for cells which did not give rise to any colonies (**Fig 4.4A**). The average MFI of CD235 was higher in cells which gave rise to MK and MK progenitor colonies, compared to cells which gave no colonies (**Fig 4.4B**).

The average MFI of dTomato_TAL1 was slightly higher in cells that gave rise to MK colonies, compared to those that did not (**Fig 4.4C**). In contrast, for cells giving rise to MK progenitor colonies, the average MFI was lower than for cells which did not produce colonies. The average GFP_GATA1 MFI was higher in cells giving rise to both types of colonies compared to those that did not (**Fig 4.4D**). LSSmOrange_FLI1 average MFI was similar for all three categories (**Fig 4.4E**).

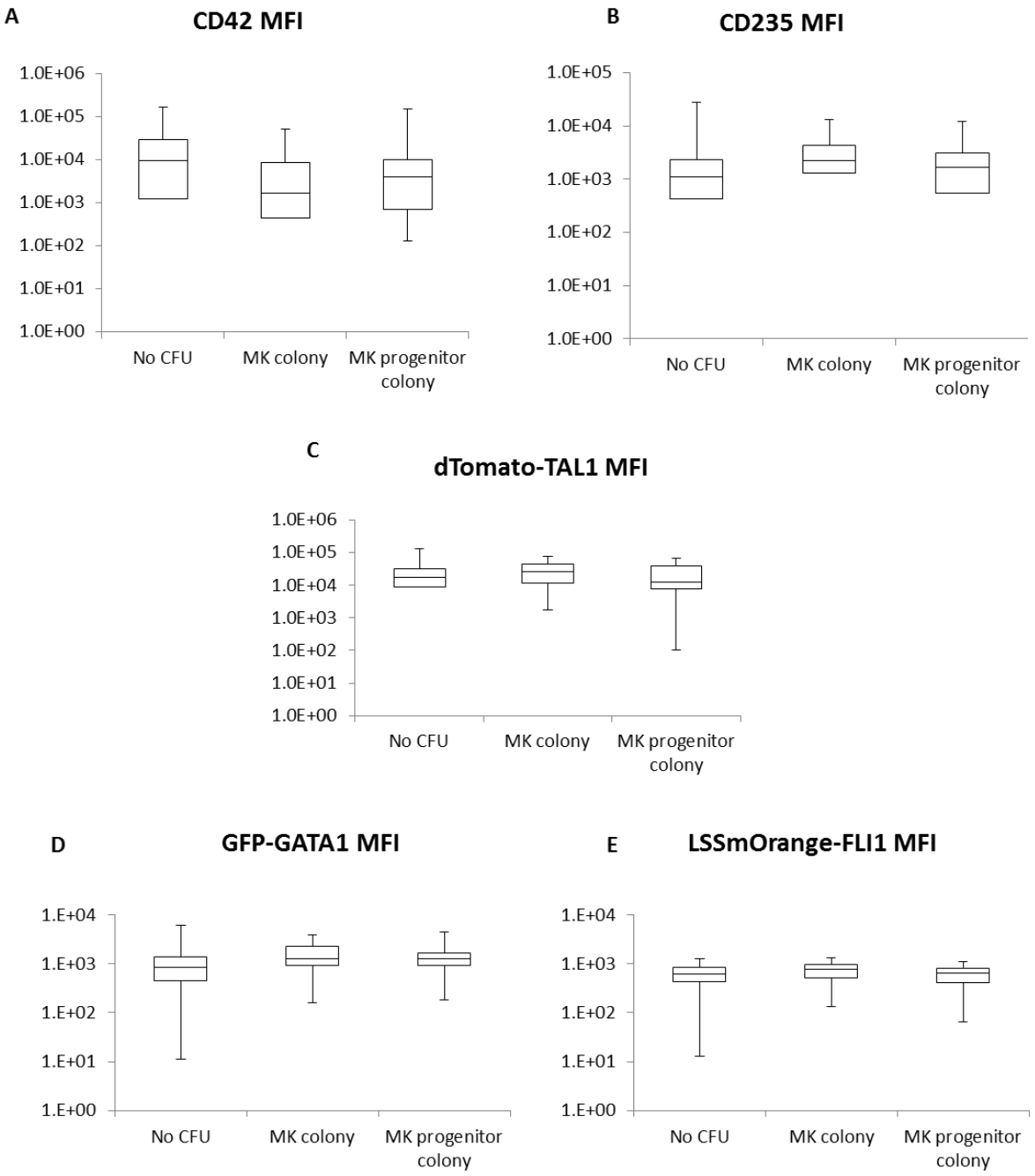


Fig 4.4 The indexed mean fluorescent intensity (MFI) of CD42, CD235 and fluorescent transgenes of single cells sorted for progenitor assays. BobC Rainbow MK-FoP day 40 cells were sorted (as previously shown). Box and whisker plots show the MFI recorded for **A**) CD42, **B**) CD235, **C**) dTomato-TAL1, **D**) GFP-GATA1 and **E**) LssmOrange-FLI1 of each single cell sorted, resulting in either no CFU colonies, MK colonies or MK progenitor colonies. N=1.

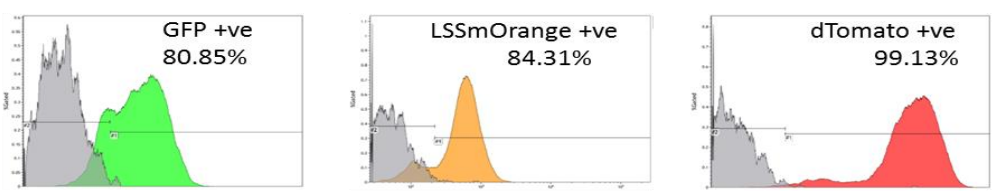
Single Cell RNA-seq Data

The remaining results shown in this chapter relate to the scRNA-seq data, analysed primarily by James Baye, with help and guidance from members of the Göttgens group. Two 96 well plates were used to sort the 8 different possible combinations of TGs that could be found in a day 40 culture of MK-FoP cells (>99% CD41+ MK committed cells, 76% CD41+/CD42+ mature MKs) transduced with the Rainbow rLV mixture at day 0. Cells were stained as previously described for single cell progenitor assays, with additional DAPI staining (final concentration 0.625µg/ml). For this experiment 24 cells were sorted per Rainbow population based on the MFI for each marker gene being above an unstained control threshold level (**Fig 4.5A**), resulting in a total of 192 cells being sorted. After aligning cDNA reads to the human reference genome (with the addition of the TGs GFP, dTomato and LSSmOrange), the next step of analysis involved quality control (QC), in order to ensure the data interrogated further downstream was reliable, did not include technical artefacts and was of high quality. QC steps eliminated 81 cells in total, with 111 cells retained for downstream analysis (**Fig 4.5B**).

Transgene and endogenous expression of GATA1, FLI1 and TAL1 in Rainbow Populations

The fluorescent transgene expression profile is shown for the 8 populations sorted in **Fig 4.5C**. The positive predictive value (PPV) for GFP was high, 0.98, with the majority of cells sorted based on GFP protein expression by flow cytometry also expressing GFP at the transcript level. However, the negative predictive value (NPV) was low, 0.47, and a high number of cells expressed GFP at the transcript level that had been sorted as GFP negative cells. The PPV for dTomato was not as high as for GFP, 0.66, with some cells that had been sorted based on positive dTomato protein expression showing no expression of dTomato at the transcript level. The NPV for this marker was higher, 0.91, showing good positive selection. LSSmOrange PPV was reasonably high, 0.83, with good positive selection for most cells. However, LSSmOrange NPV was low, 0.57, showing negative selection for this marker was also not robust for all cells.

A



B

Population	G/O/T	G/O	G/T	G	O/T	O	T	neg
GFP-GATA1	+	+	+	+	-	-	-	-
LSSmOrange-FLI1	+	+	-	-	+	+	-	-
dTomato-TAL1	+	-	+	-	+	-	+	-
Number of cells passed QC	17	17	16	16	16	14	9	6
Percentage of cells passed QC	71%	71%	67%	67%	67%	58%	38%	25%

C

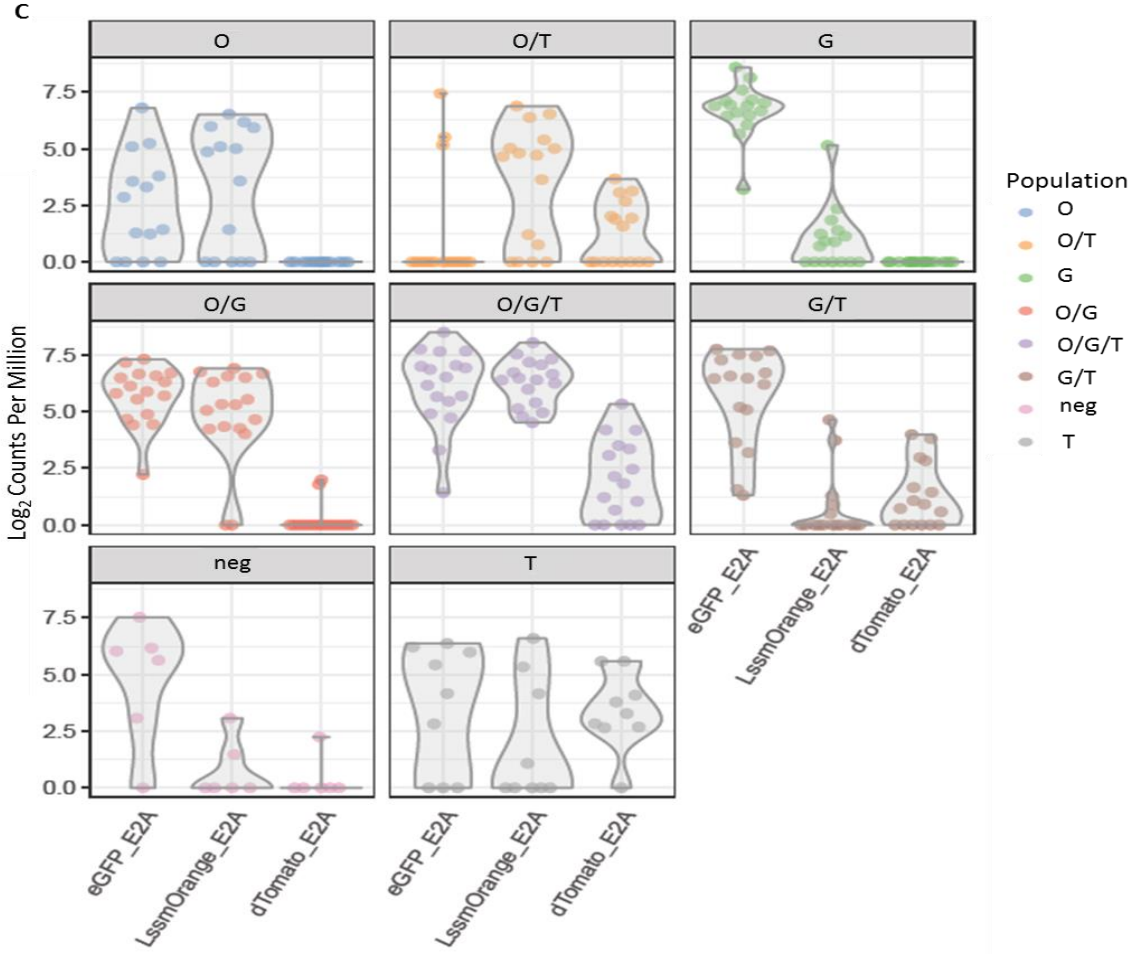


Fig 4.5 Quality control of the 8 different Rainbow populations sorted for scRNA-seq. **A**) Gating used to sort single cells by FACS for each fluorescent transgene marker. Cells with a mean fluorescent intensity (MFI) value higher than unstained negative control cells were selected as positive for each marker. 24 cells for each of the 8 different Rainbow populations were sorted for scRNA-seq. **B**) The resulting cell numbers for each population passing QC checks is shown. **C**) Fluorescent gene expression levels determined by scRNA-seq for each population sorted. O= LSSmOrange-FLI1, T= dTomato-TAL1, G= GFP-GATA1, neg= negative for all fluorescent markers.

The total expression of endogenous and transgenic GATA1, FLI1 and TAL1 in the different sorted populations is shown in **Fig 4.6**. All three TFs are expressed at similar levels, with no statistical difference found between the 8 populations.

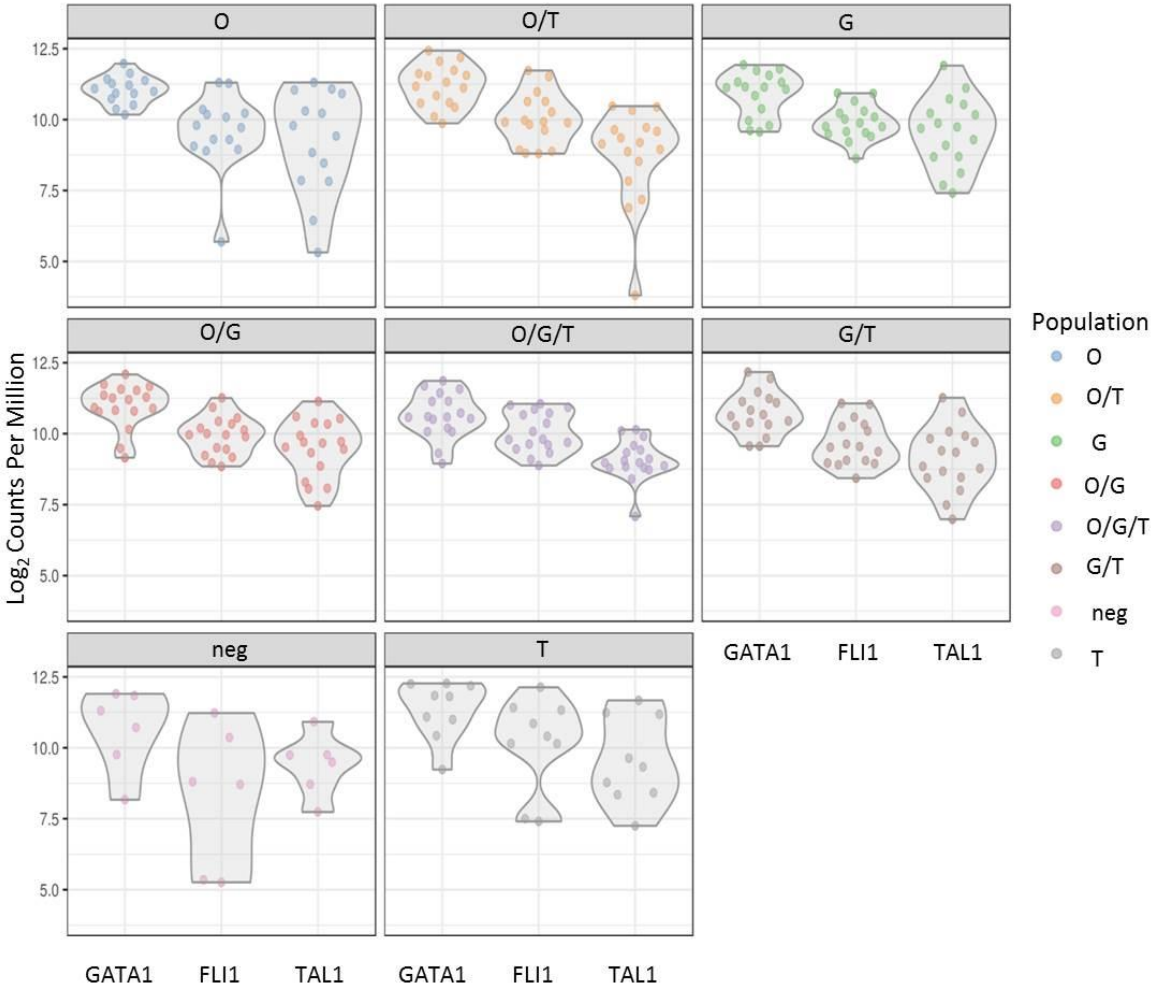


Fig 4.6 Endogenous GATA1, FLI1 and TAL1 gene expression levels determined by scRNA-seq for each population sorted. O= LSSmOrange-FLI1, T= dTomato-TAL1, G= GFP-GATA1.

Different Cell Populations Revealed by Unsupervised Clustering

Unsupervised clustering was performed on all cells based on their whole RNA expression patterns of 20,296 genes captured by scRNA-seq. Clustering showed the highest stability when cells were grouped into 3 or 4 subgroups (**Fig 4.7**).

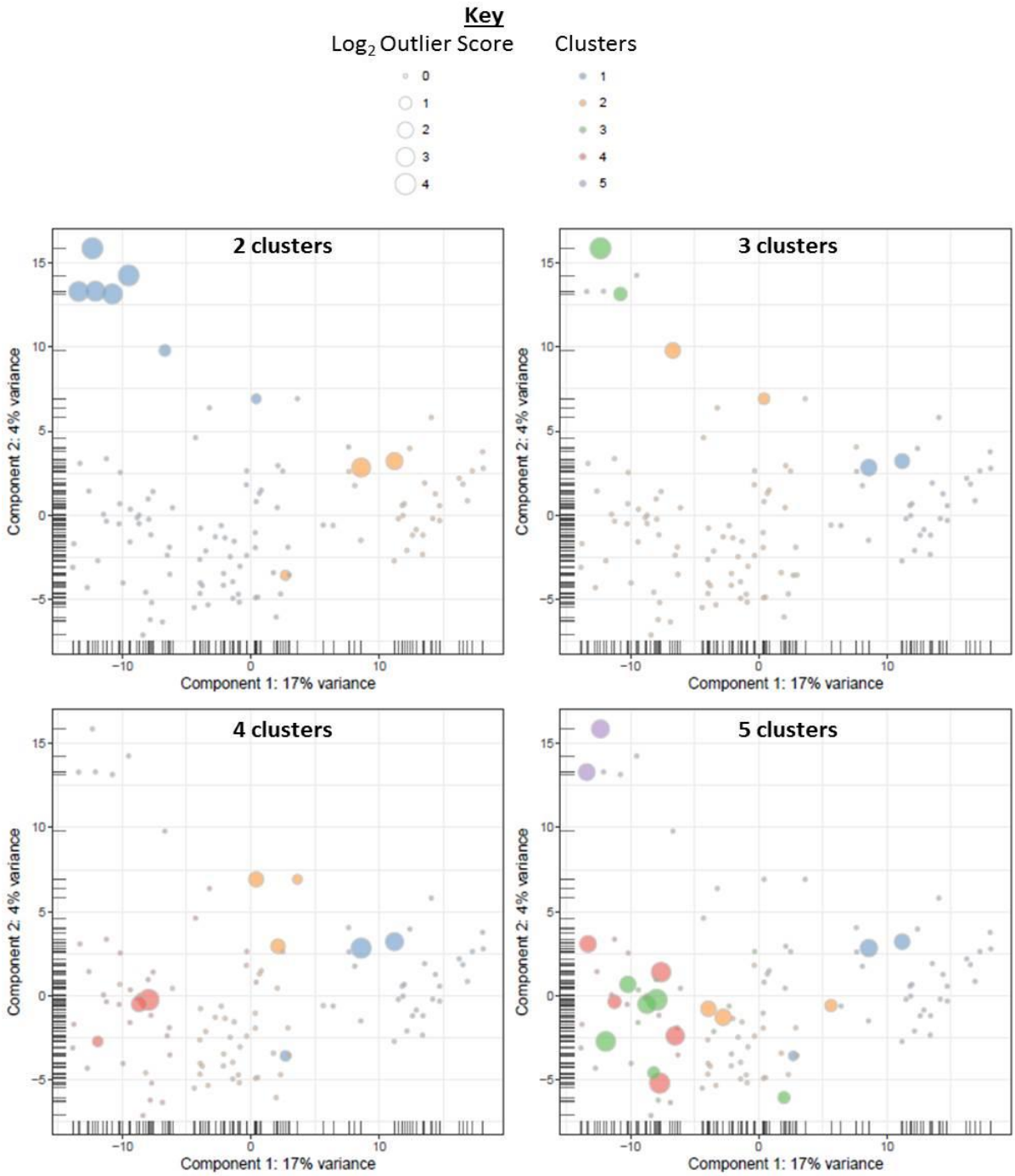


Fig 4.7 Unsupervised clustering of the scRNA-seq data into 2, 3, 4 or 5 groups. Clustering shows the greatest stability when 3 or 4 sub-groups are produced. The larger the circle, the less those data points fit with the cluster they are in, represented by colour.

Gene ontology analysis was performed on the upregulated genes of each cluster, in the three cluster formation (**Fig 4.8**). The largest cluster observed (orange), with 430 upregulated genes, revealed that these cells upregulate genes related to platelet biology such as those involved in platelet activation, blood coagulation, haemostasis and response to wounding. This cluster likely represents more mature megakaryocytes. A smaller cluster identified (blue), with 1020 upregulated genes, revealed that this subgroup upregulates genes related to cell replication such as those involved in cell division, DNA and nuclear division and the mitotic cell cycle. This cluster therefore likely represents proliferative MK progenitors. A third, smaller group of cells was identified (green), with 410 upregulated genes, revealed that this group upregulate genes related to cell death such as those related to cell cycle arrest, apoptotic signalling and G1 DNA damage response. This cluster, therefore, likely represents dying cells but not dead cells, which were excluded in QC steps.

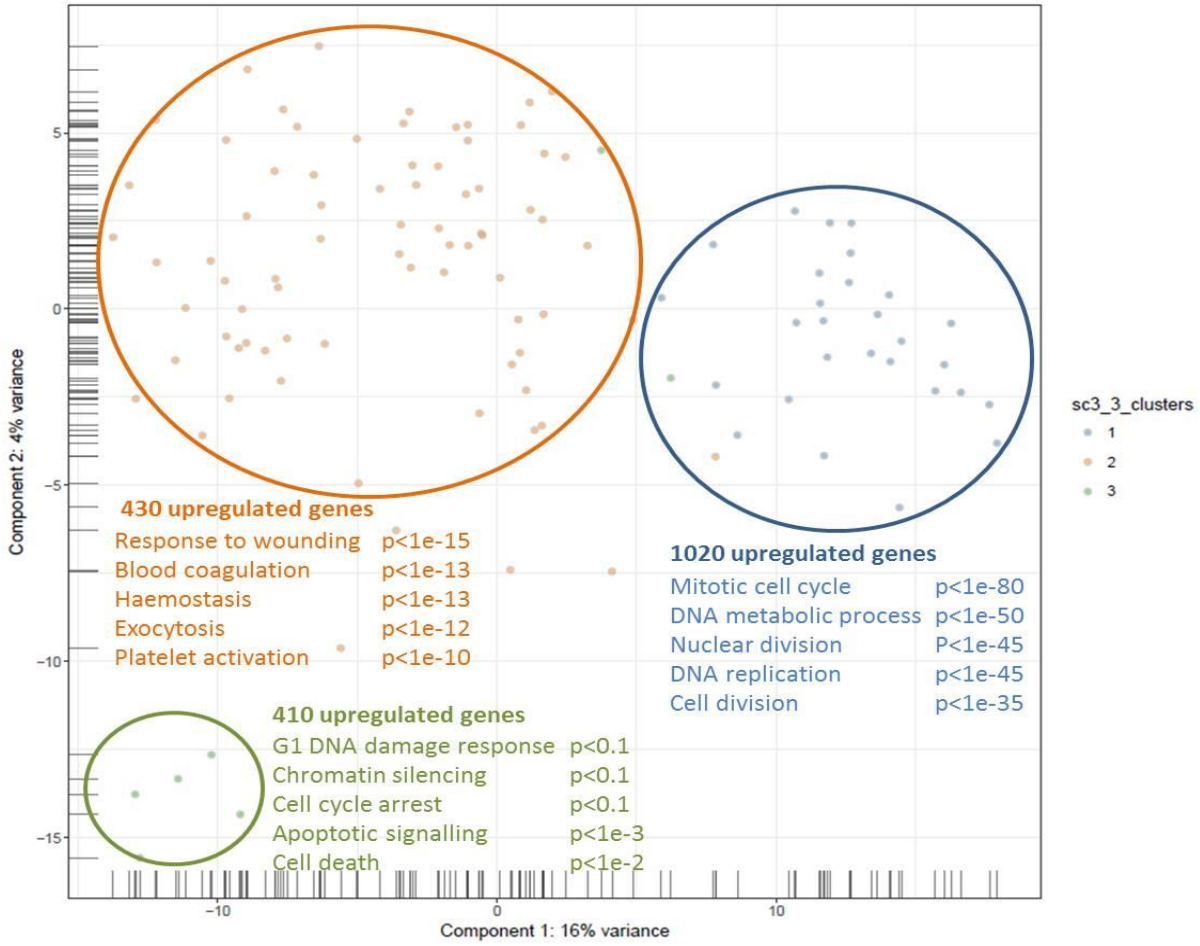


Fig 4.8 Unsupervised clustering based on single-cell whole RNA expression identifies three cell subtypes with different phenotypes identified through gene ontology. Unsupervised clustering segregates cells into three sub-groups, visualised by separate spatial location of cells on a PCA plot, illustrating high-level differences in RNA expression between single cells. Gene ontology on upregulated genes in each sub-group identities different cell phenotypes.

The largest sub-group identified by the three cluster formation (orange, **Fig 4.8**) was further divided into two sub-groups (red and yellow) when analysed by a four cluster formation (**Fig 4.9A**). The BluePrint consortium have previously published lists of upregulated genes in haematopoietic stem cells (HSC) and all downstream progenitors that give rise to, and including, MK and erythroblasts (Chen L, 2014). Gene expression levels of those upregulated in megakaryocyte-erythroid progenitors (MEPs), MKs and erythroblasts (EB) from the BluePrint report were compared with the four populations depicted in **Fig 4.9A**. The heat map shows that the MK progenitor (blue) group has strong expression of MK genes, and stronger expression of EB genes compared to the other sub-groups (**Fig 4.9B**). The far left MK sub-group (red), shows strong upregulation of MK gene expression strong downregulations of EB and MEP gene expression, meaning this cluster most likely represents a more mature MK phenotype compared to the intermediate MK cluster (yellow). The cells group most likely to contain dead cells shows strong downregulation of all gene types.

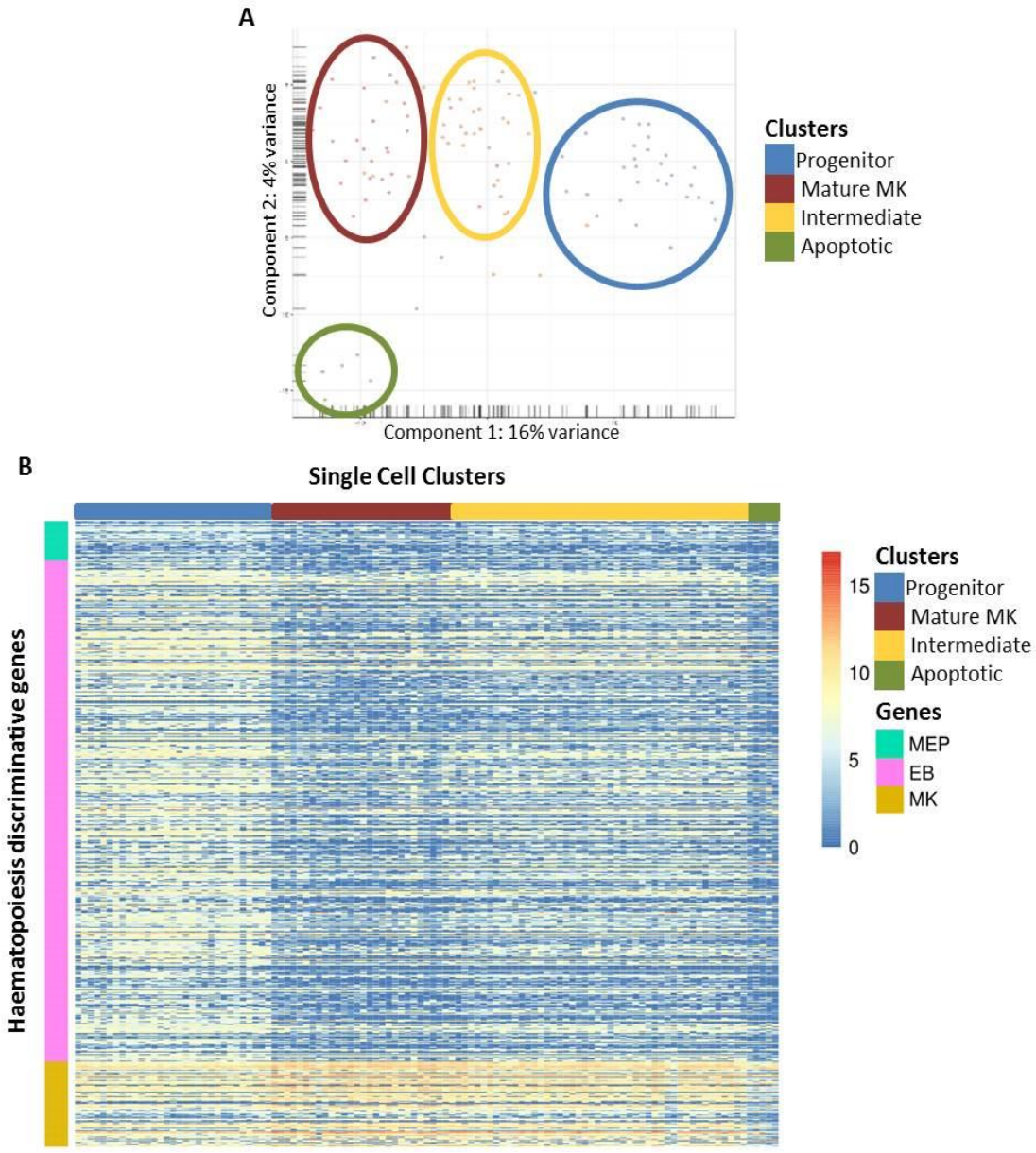


Fig 4.9 Evidence of MK-FoP cell clusters in distinct stages of differentiation. **A**) Unsupervised clustering into 4 sub-groups further divides the MK cell group into two subgroups. **B**) Heat map shows gene expression patterns of the 4 sub-groups of cells identified in long-term MK-FoP, with genes known to be upregulated in MEP, EB or MK cells. The progenitor cluster identified in MK-FoP cells shows higher expression of EB genes than other groups, while intermediate and mature MK cells show upregulation of MK genes specifically.

MK Gene Expression Varies in Different Populations

Gene expression of the three surface markers most frequently used to identify forward programmed cells, as well as genes related to MK maturation, are shown for the 4 cell populations in **Fig 4.10**. CD235 is not expressed in the majority of cells, but is most frequently expressed in progenitor cells, with a number of intermediate MKs showing low expression (**Fig 4.10A**). CD41 is expressed in all cells, with the lowest expression predominantly in the apoptotic cell cluster. CD41 expression is increased in progenitors and further increased in intermediate MKs and the highest expression is seen most frequently in mature MKs. CD42 expression follows a similar expression pattern, with some apoptotic cells expressing high levels of CD42. Two genes with known expression patterns in MKs and their progenitors were also assessed, shown in the riverplot (generated from Blueprint epigenome data available online) (**Fig 4.10B**). vWF expression was seen in only a small number of cells, belonging to all cell types observed. STMN1 expression was highest in the progenitor population, decreasing in the immature and mature MKs.

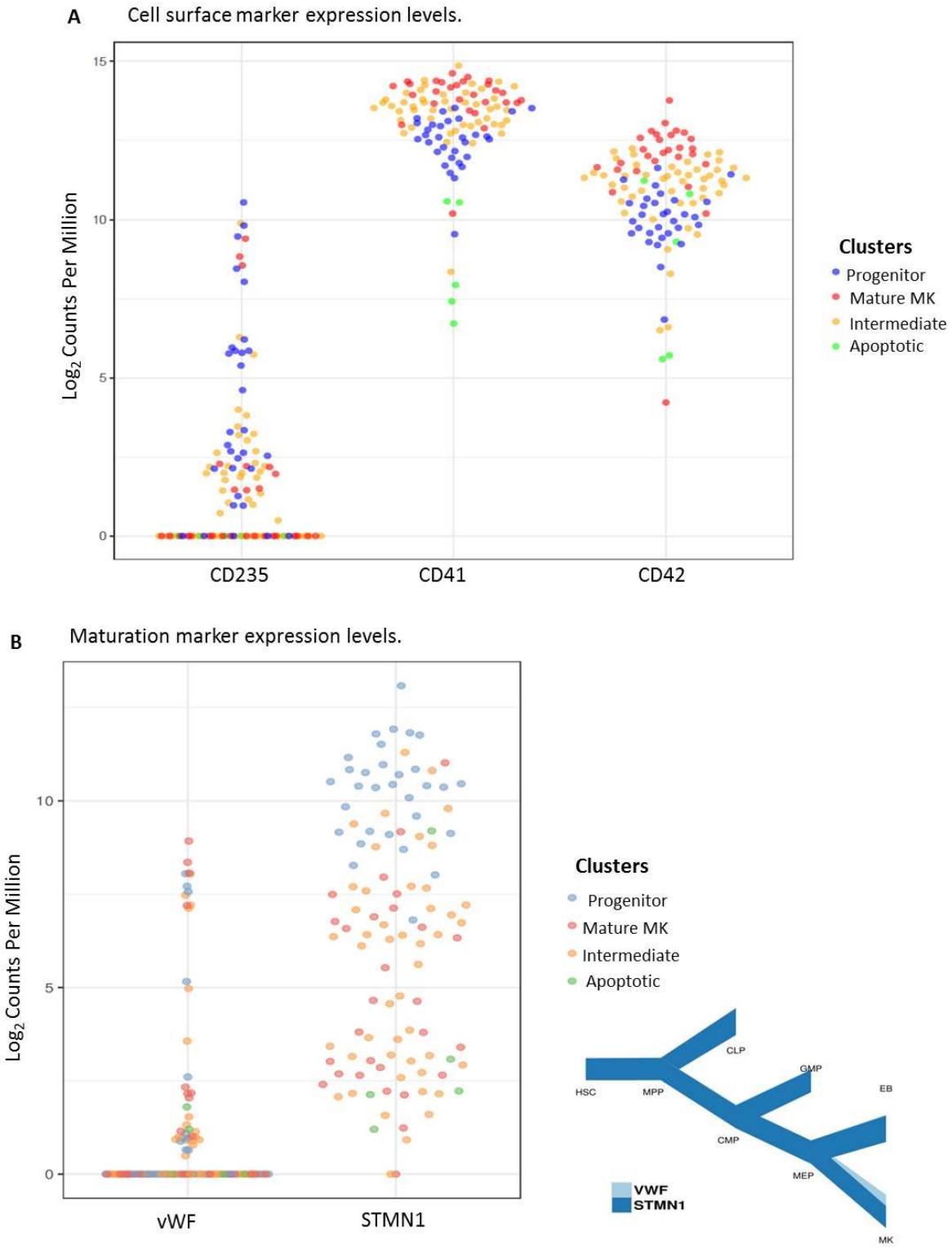


Fig 4.10 Expression of cell surface markers and maturation markers in the four populations identified by unsupervised clustering. A) CD235, CD41 and CD42 cell surface marker expression. B) vWF and STMN1 (Stathmin) expression, with a riverplot to show expression of these genes in haematopoietic cells (from Blueprint epigenome data).

Identifying Novel Candidate Genes to Interrogate MK Progenitors

Novel candidate surface marker genes were identified as being upregulated (Gavin Wright, Wellcome Trust Sanger Institute), in only the progenitor sub-group identified by unsupervised clustering (**Fig 4.9A**). From this list a number of genes which are of interest to our group have been selected and are shown in **Table 4.6**. The genes could offer better discrimination of the different cell populations identified by scRNA-seq in long term MK-FoP cultures and could potentially be used as better markers than CD41, CD42 and CD235, to sort progenitors from intermediate and mature MKs. Better isolation methods would allow us to carry out future downstream analysis to characterise the different cell phenotypes, by performing pro-platelet formation assays and colony assays. From this list, MICB has already been validated as a good candidate surface marker by Dr Moyra Lawrence (Ghevaert lab), who has shown MICB can be detected by flow cytometry on a small population of MK-FoP cells and that sorting these cells based on MICB expression leads to higher colony numbers in CFU assays (data not shown, N=1).

Gene	Membrane protein classification
CD24	GPI anchor
KCNS3	Multi pass membrane protein (6TM)
IL18R1	Single pass membrane protein
ADAM15	Single pass type I membrane protein
CD34	Single pass type I membrane protein
CD4	Single pass type I membrane protein
CD44	Single pass type I membrane protein
CD83	Single pass type I membrane protein
IFNGR2	Single pass type I membrane protein
MICB	Single pass type I membrane protein

Table 4.6 List of interesting novel plasma membrane genes for identifying MK progenitors. The genes listed are genes which are upregulated in MK progenitors and that also localise to the surface membrane.

Discussion

This chapter has described the development of single cell progenitor assays, performed on sorted cells in a 96 well plate format. The reason for developing such an assay was to enable us to perform index sorting, alongside obtaining a functional read-out of sorted cells, with the hope of identifying a surface marker signal that could identify progenitor cells in the heterogeneous MK-FoP culture. This is because classical progenitor assays do not allow retrospective analysis of cells. Single cell progenitor sorting assays were performed with the routine surface markers (CD41, CD42 and CD235) in order to establish the correct conditions for the assay and because we have previously hypothesised that the bi-potent progenitor cell belongs to the CD41+/CD235+ population (Chapter 1) and have been unsuccessful in identifying better surface markers so far. In parallel we developed the Rainbow vectors, with the aim of later adding these markers to the index sorting panel in order to establish a better signature for progenitor cells based on their level of transgene (TG) expression.

The Rainbow vectors exposed the heterogeneity of cells within long-term MK-FoP experiments in terms of transgene expression, which was the starting point for trying to better understand these cultures using single cell RNA-seq. Day 40 FoP-MKs were sorted based on their Rainbow expression and scRNA-seq data showed that different levels of transgene expression does not affect the total expression of *GATA1*, *TAL1* and *FLI1*. Therefore, total expression of the three FoP TFs is homogenous in long-term MKs. Cells clustered into four different groups, identified as; mature MKs, MKs, progenitors and dying cells. Based on these cell clusters, novel surface markers were found to be overexpressed in progenitor cells, which could allow for more efficient isolation of these cells in future for further study.

Early MK-Progenitors Appear to be CD41/CD235 Double Positive

Early time-point (day 9-13) bulk sorts were performed alongside single cell sorts to ensure that sorting cells did not reduce their progenitor potential and that the cultures used were for sorting were capable of giving rise to colonies in progenitor assays. Bulk sorting different populations for CFU assays resulted in no progenitor potential seen from the CD41-/CD235- population of cells, as expected, as this population represents cells which have failed to forward program (**Fig 4.2B**). The highest colony numbers per cell seeded were consistently produced by the CD41+/CD235+ population (**Fig 4.1B** and **Fig 4.2B**). A small number of colonies were produced in the CD41+/CD235- and CD41-/CD235+ bulk sorted populations, however, due to gating strategy used to sort cells in this experiment (**Fig 4.2A**) it is possible contaminating cells were sorted in either of these populations from the double positive population.

Data from the single cell progenitor assays support the findings that progenitor cells are capable of producing MK progenitor colonies and MK colonies from single cells after FACS sorting. As for the bulk sorted populations, single cells from the CD41+/CD235+ population showed the highest progenitor potential (**Fig 4.1C**). Strikingly, no colonies were generated from CD41+/CD235- sorted single cells. Combined with data from the bulk sorts, this suggested that the CD41+/CD235+ population contains MK progenitor cells produced by FoP, in line with our hypothesis. These experiments were performed in TPO only hence this data cannot confirm the presence of a bi-potent progenitor capable of giving rise also to CFU-E colonies.

The day 9 Rainbow transduced cells used to perform progenitor assays in **Fig 2**, contained a high proportion of cells expressing CD235 only, which reflects a poor MK-FoP experiment. The cells expressed quite low levels of the fluorescent TGs, with LSSmOrange-FLI1 expression being the lowest, which might explain the high frequency of CD235+ cells. As mentioned in the Methods section for this chapter, the Rainbow rLV mix used (produced by a commercial company) has an unknown ratio of the three TGs. This experiment, as well as others performed with the Rainbow mix that show a higher percentage of CD235+ cells than expected, leads us to believe that the MOI of LSSmOrange-FLI1 is lowest of the three TGs in this mix. The single cell progenitor assay performed on these cells produced five colonies (**Fig 4.2C**). Three of the five colonies were produced from CD235 negative cells, which could reflect the existence of a MK progenitor population, which does not express CD235. However, the MFI of CD235 for these cells was high (>1000) and the threshold MFI used to gate positive cells was not much higher (1895). Also the colonies produced by these single cells expressed a high percentage of CD235 (72-84%, **Fig 4.2D**), so the single cells sorted most likely were low CD235 expressing cells. This shows that the gating strategy could have been improved by making the gates more discreet for this experiment, as mentioned above for the bulk sorting. Interestingly, despite colony 1 being identified as an MK progenitor colony and the remaining colonies identified as MK colonies, the flow data collected shows these colonies to be very similar to each other in terms of their CD41, CD235 and CD42 expression. All colonies produced contained cells expressing at least one of the three fluorescent TGs (**Fig 4.2E**). Colony 4, originating from a cell with did not express LSSmOrange-FLI1 at the time of sorting, contained >60% CD41/LSSmOrange expressing cells. This suggests that either the cells in this colony were able to upregulate LSSmOrange-FLI1 expression, or the cell was very lowly expressing LSSmOrange at the time of sorting.

Rainbow Transduced Cells Produce a unique CD41/CD235/CD42 Triple Positive Population

All colonies produced by the single cell progenitor assays in the experiment described in **Fig 4.2** expressed a high percentage of cells which were CD41/CD235/CD42 triple positive (>70%), which is a population that is not normally observed in 3TF FoP. The long term Rainbow MK-FoP culture used for scRNA-seq and large-scale single cell progenitor assays, also contained a percentage of triple positive cells at day 41 (**Fig 4.3B**), which was lower (15%) at this later time-point. This suggests that the Rainbow vectors, with their unknown mix of TGs, uniquely generates a population of cells which contains both mature MK markers and erythroblast markers early in FoP, which is lost over time in TPO culture conditions. It is important to note that at day 41 the Rainbow culture contained primarily mature MKs which were CD41/CD42 double positive (69%), as expected from MK-FoP experiments.

Characterising Long-term MK-Progenitors

A long-term Rainbow MK-FoP experiment, performed by Dr Moreau, tested different MOIs and monitored Rainbow populations over time. This experiment (**Diagram 4.1**) showed that the Rainbow rLV mix used at an MOI of 40 could maintain mature MKs long-term. It also established that an MOI of 40 produced a large range of Rainbow populations by day 40, starting from a predominantly triple positive (GFP+/dTomato+/LSSmOrange+) population at day 9. Based on this preliminary experiment, I performed another long-term MK-FoP experiment, using the Rainbow mix at an MOI of 40, in order to perform single cell progenitor assays and scRNA-seq on a long-term culture, at day 40. This experiment, used for cell sorting, did not follow the same dynamic of Rainbow populations as previously seen, with the majority of populations being quite rare at day 40 and the culture mainly consisting of triple positive cells at this stage (**Fig 4.3A**). CD41 index data is not shown for the experiment described in **Fig 4.3** and **Fig 4.4**, due to technical issues.

High CD235 and low CD42 Expression Found for MK-Progenitors

The single cell sorting of CD235+ and CD235- cells for progenitor assays suggest that at day 40 the MK progenitor population is no longer defined by expression of CD235, as both populations gave rise to colonies (**Fig 4.3C**), however, the number of colonies was higher in the CD235 positive population. Since fewer than 10% of CD235+ cells did not express CD41 at day 41 (**Fig 4.3B**), it is likely that the majority of CD235+ cells which gave rise to colonies were also positive for CD41. Hence, the MK-biased progenitor population identified in previous experiments at early time points (day 9-13), appears to still exist in long-term cultures. Surprisingly, the progenitor potential was 4% (0.04 colonies/cells seeded, Fig **4.3C**) for the CD235+ population, which is higher than expected at this

time-point. This could be reflected by the fact that CD235 expression was higher in this Rainbow experiment, even at day 41, which is not usually seen in 3TF experiments and suggests that this culture contains higher numbers of progenitors than a 3TF experiment would at the same time-point.

Index data collected at the time of single cell sorting for progenitor assays help to narrow down the MK-biased progenitor cell signature in the long-term culture. Cells which gave rise to MK and MK progenitor colonies had a lower average CD42 MFI and higher average CD235 MFI than cells which did not give rise to colonies (**Fig 4.4A** and **Fig 4.4B**). This gives a more complete signature for a MK colony forming progenitor than previously known, of $CD42^{low}/CD235^{high}$. Since the CD235 MFI was higher on average for both colony types, this might suggest that the cells from the CD235 negative population that gave rise to colonies were actually expressing CD235, and might highlight an issue with contaminating cells being sorted by the sorting strategy used, which did not separate the CD235+ and CD235- populations very discreetly, as seen by the overlay of the CD235 expression from a plate of cells sorted from each population (**Fig 4.3C**).

Single Cell RNA-Seq Data Reveals All Rainbow Populations Express Similar Endogenous Levels of *GATA1*, *FLI1* and *TAL1*

The gating strategy used to sort single cells based on fluorescent protein expression, into the 8 different Rainbow populations, shows that a high percentage of cells selected by FACS for each marker were positive, compared to a negative control (**Fig 4.5A**). However, scRNA-seq expression data for the three fluorescent TGs shows that the sorting strategy used did not successfully capture the desired population in a few circumstances (**Fig 4.5C**). This highlights the fact that fluorescence at the protein level does not necessarily correlate to the transcript level. This may be due to a fluorescent marker being expressed at the mRNA level, which is subsequently degraded and not translated, therefore is not present at the protein level, or a fluorescent marker protein being present at too low a level to be detected by FACS and incorrectly categorised in the negative population. Incorrect negatively selected cells show relatively low mRNA transcript levels (**Fig 4.5C**). Interestingly, the majority of single dTomato positive and triple negative cells were discarded (15 and 18 cells respectively), from further analysis during the QC steps (**Fig 4.5B**). This raises the question whether these cells failed specifically due to the lack of GFP-GATA1 and LSSmOrange-FLI1, or all three TGs, which may represent predominantly dead cells, or whether this is unrelated to these populations performing poorly in the QC step.

The total expression levels of *GATA1*, *FLI1* and *TAL1* (endogenous and transgenic) shows no significant statistical difference between the different populations sorted (**Fig 4.6**). This shows that despite certain TGs not being expressed or expressed at low levels in certain populations, as seen for the fluorescent proteins, the endogenous gene expression level must be increased, in order for this equal expression to be observed. Unfortunately the scRNA-seq performed cannot distinguish between endogenous and transgenic expression of the three TFs. One cannot simply remove the expression value of the relevant fluorescent protein, as the sequencing performed has a 5' bias, so will over-represent the fluorescent protein expression levels relative to the transgenic TF, due to the ordering of all vectors being fluorescent gene-2A-TF. However, this data indicates that despite observed heterogeneity in TG expression, the total expression of the three TFs was the same in the different populations, which determines MK identity.

Gene Ontology Terms Define Three Distinct Cellular Phenotypes

Unsupervised clustering revealed three or four groups of cells with the highest stability (**Fig 4.7**). In the three-group clustering, gene ontology was performed which identified one cell group with platelet-biased functions, one with mitotic functions and one with increased cell death related functions (**Fig 4.**). It is therefore most likely that these groups reflect mature MKs, MK-progenitors and dead cells respectively, based on the gene ontology results. In the four-group clustering the platelet-biased group was sub-divided but gene ontology failed to discriminate many differentially expressed genes in these two groups. This is expected, since gene ontology is a tool best suited to reporting high-level variance between cells. Nevertheless it is likely that these two groups represent more mature MKs, the furthest group away from the MK-progenitor group, and intermediate MK cells (**Fig 4.9A**).

Evidence of Cell Clusters in Distinct Stages of Differentiation

By comparing the gene expression levels of genes known to be upregulated in MEPs, MKs and erythroblasts (EB) (Chen L, 2014), in the 4 populations identified in the long-term MK-FoP culture, distinct cell phenotypes were observed (**Fig 4.9**). The progenitor group showed upregulation of both MK and EB genes, consistent with the idea that this cell type reflects a bi-potent progenitor group generated by FoP. Of the two MK cell groups identified, one showed stronger upregulation of MK genes and downregulation of EB and MEP genes, compared to the other, consistent with the idea that the one group represents more mature MKs, while the other represents less mature MKs.

Based on the four groups identified (progenitor, immature MK, mature MK and apoptotic cells) the expression of cell surface markers and maturation markers corroborates these groups further (**Fig**

4.10). CD235, a red cell specific marker, was not expressed by the majority of cells and found most commonly in cells believed to represent bi-potent progenitors. CD41 and CD42 were most highly expressed in mature MKs. Stathmin (STMN1), an essential protein for the regulation of microtubule cytoskeletons, is downregulated during MK maturation but is known to be highly expressed in MEPs (Iancu-Rubin C, 2011). This gene followed the expected expression pattern in MK-FoP cells, with the highest expression seen for progenitor cells. vWF expression, which we have previously observed by flow cytometry to not be expressed in the majority of MK-FoP cells, showed sporadic expression in a number of cell types identified, and in very few cells overall. This reflects that MK-FoP cells do not express the same maturation markers as MKs from primary stem cells, which may have implications for their therapeutic use.

Novel Cell Surface Markers Identified could Improve MK-Progenitor Discrimination and Isolation

Finally, the identification of a number of novel cell surface markers could be important for enabling progenitor cell isolation (**Table 4.6**), improving sorting of these cells and helping us to further refine the progenitor cell surface identity for future progenitor assays and whole genome expression analysis. While CD41 and CD42 show some discrimination between the cell groups identified by unsupervised clustering of the scRNA-seq data, the expression for all cell types was high, revealing that these markers are not ideal for discriminating easily between progenitors and MKs. It should be noted that high RNA expression may not translate to high protein levels and that these markers would need to be experimentally validated and checked for external cell surface expression.

Overall the data presented here shows strong evidence that long term MK-FoP cultures consist of progenitors and MK cells in different stages of differentiation. The apoptotic cell group identified, interestingly, clusters closely to the mature MKs, raising the question- do these cells represent very mature MKs which are poised to produce platelets? There has long been an association of apoptosis with platelet production, since primary cultured MKs gave rise to the observation that peak platelet production *in vitro* corresponds with the onset of apoptosis in MKs (Zauli G, 1997). Therefore, this small group could represent platelet producing cells, and may be represented by such low numbers in this experiment due to these cells undergoing cell death, making them most likely to fail QC steps and be excluded from analysis. It would be interesting to perform scRNA-seq at multiple time points during FoP, to see if the dynamics of the cell groups identified in this analysis remain stable overtime and particularly if progenitor cells are able to self-maintain and/or become mature MKs. It would also be interesting to see whether the apoptotic cell group increases in number if cells are pushed towards platelet formation *in vitro* before sorting.

Discussion

and

Future Perspectives

The first aim of my PhD was to show whether FoP could generate bi-potent progenitors, capable of maturing into both MKs and erythroblasts. The work presented in Chapter 1 shows for the first time that the FoP technique provides a means of producing platelets and erythroblasts, both of which are clinically relevant cell types for use in transfusion medicine. Many of the issues surrounding platelet transfusions outlined in the Introduction of this thesis are also relevant for red blood cell transfusions, such as biosafety and donor dependency. For this reason, showing that the FoP protocol can reliably generate both cell types from hiPSCs, while also being GMP compatible and amenable to large-scale manufacture, is a huge advance for the field of transfusion medicine.

As for existing directed differentiation protocols, the platelets and erythroblasts generated through FoP are embryonic in phenotype, which could lead to a number of clinically relevant issues. Such as, will embryonic platelets and erythroblasts function in the same way as adult cells? Importantly, will they offer therapeutic benefit, to an adult recipient? Functional data from our lab has shown FoP produces platelets that are capable of thrombus formation both *in vitro* and *in vivo*, in a mouse laser-induced vascular injury model (Moreau et al, 2016). This shows promise that despite not being adult in phenotype FoP produced platelets will still function in the same way as donor platelets, suggesting this will not be an obstacle for their use in human trials.

Many published protocols to generate RBCs from PSCs, produce cells which lack the classical features of mature erythrocytes, including enucleation and adult beta globin expression, representing a lack of definitive haematopoiesis occurring in these cells. Critically, *in vitro* produced RBCs which lack an adult phenotype may not be able to function in the same way as RBCs produced *in vivo*. As reticulocytes mature *in vivo* they reduce in size, become biconcave in shape and acquire stability and deformability, allowing mature erythrocytes to pass through small arterioles easily (Waugh et al, 2001). One issue is that *in vitro* produced RBCs are much larger than their *in vivo* counterparts and thus may cause blockages in arterioles if they are unable to deform correctly. Secondly, without enucleation and adult haemoglobin, they will not match the oxygen carrying capacity of donor cells, so offer little advantage.

One potential way to improve maturation of FoP erythroblasts could be to over-expressing additional TFs. For example, Krüppel-like factor 1 (*KLF1*) has recently been shown to enhance both differentiation and maturation of red blood cells from PSCs, resulting in improved enucleation, when over-expressed at the stage of haematopoietic progenitors, during differentiation (Yang CT, 2017). However, the phenotype of these cells remains embryonic, with little β-globin expression. The lack of maturation seen in *in vitro* RBCs, may be due to the nature of the starting cells used- pluripotent stem cells, which may contain epigenetic features that need to be removed for full maturation to

occur during differentiation. Ashley Toye and Jan Frayne's groups in Bristol (both part of the NovoSang consortium) have made progress in generating mature erythrocytes, but from immortalised adult erythroblasts not PSCs. The cells produced show the hallmark features of erythrocytes and will be used in the near future for conducting the first in man trials (Trakarnsanga et al, 2017). The fact that these groups have stepped away from trying to generate RBCs from hPSCs, in favour of starting cells which are already committed to the haematopoietic lineage, also adds weight to the idea that PSCs may not be able to produce definitive haematopoietic cells in the minimal cytokine, 2D conditions such as we and others are currently using.

While other cell types may prove to be more useful to generate mature RBCs in vitro, deriving platelets from PSCs still holds a lot of promise. PSC-derived platelets could offer a huge number of therapeutic advantages over using donor platelets. In order to generate changes in the platelets produced, PSCs can be 'customised' more easily using techniques such as CRISPR-Cas9, than other cell types such as CD34+ cells, which cannot be cultured for long periods of time. The advantage of stating with PSCs is that after any genetic changes made, the new PSC line can be cryopreserved and cultured indefinitely. One example of customising PSCs has already taken place in our lab. PSCs have been genetically altered to stop expression of the β2 microglobulin (*B2M*) gene. The B2M protein is responsible for cell surface expression of MHC class I molecules in most cell types. Annett Müller has deleted the *B2M* gene from iPSCs and shown that these cells FoP in exactly the same manner as their parental line (unpublished). Mouse experiments will soon be underway to assess whether these HLA-null cells can evade an immune response and thus be used as a source of universal platelets.

A PSC source of universal platelets would alleviate the pressures of finding matched platelets for patients with rare blood types, or for those who have become alloimunised. A 2012 study calculated that 150 homozygous HLA-types volunteers could provide iPSC lines that would offer a match for 93% of the UK's population (Taylor et al, 2012). However, the establishment of such a large number of clinical grade iPSC lines would be a huge undertaking, with huge associated costs. To put this into perspective, the UK Stem Cell Bank currently has only 3 clinical grade stem cell lines available for use, one of which has 2 confirmed mutations in the gene *TP53*. Since only approximately 40% of all the PSC lines tested in our lab have been capable of producing mature MKs by FoP, even if the establishment of 150 lines was feasible, the reality is that many of these lines would not FoP.

Another example of how *in vitro* derived platelets could offer benefits over donor platelets has to do with the fact that platelets are activated only at sites of vascular injury, where they in turn release their granule content. There is a potential to use platelets to deliver drugs to specific regions of the

body, at local sites of injury. Alternatively, by genetically editing the starting PSCs used for FoP, platelets could be derived which overexpress certain proteins in their alpha granules, such as pro-thrombotic proteins or pro-angiogenic proteins, which could be administered to specific patients groups depending on their therapeutic requirements. This has been functionally demonstrated by David Wilcox and colleagues, who transduced HSCs in canines with a lentiviral vector that encoded for a modified version of the Factor VIII gene, downstream of the MK specific CD41 promoter and an α-granule targeting domain of VWF (Du et al, 2013). They found this strategy to alleviate haemophilia A symptoms *in vivo* and thus, we are adapting this strategy to target other proteins to the α-granules of FoP platelets.

Unlike many cell types, currently of interest to produce from PSCs for clinical use, RBCs and platelets offer a unique advantage, in that both these cell types are anucleate. This means any cells to be transfused into patients can be irradiated, removing any potentially oncogenic progenitors that may remain in cultures. However, a disadvantage of platelets is that they require careful handling to collect and store, so as not to become activated and clot. This makes collecting platelets from FoP-MK cultures difficult and may reduce the numbers of platelets collected. As with donor platelets the storage of such cells would mean they could not be kept for long time periods. Mature erythrocytes produced by the NovoSang consortium immortalised line, has highlighted issues with collection for these cells. The low purity of mature cells in their cultures means they need to be purified, which causes a large loss of cells and an overall reduction in the number of erythrocytes obtained (Trakarnsanga et al, 2017).

Since its publication in 2016, the FoP technique has gained widespread popularity with many researchers around the world investigating the developmental biology of MKs and platelet formation *in vitro*. Compared to existing directed differentiation (DD) protocols, FoP offers advantages in both cell purity and cell number outcome (Moreau et al, 2016). As useful as this technique is for studying megakaryopoiesis *in vitro*, the main aim of the Ghevaert lab is to produce platelets for clinical use. When I started my PhD, there were 2 major bottle necks in the FoP protocol, which presented huge drawbacks to being able to scale up this technology and offer *in vitro* derived platelets.

The first bottle neck was that we were unable to produce large enough numbers of MKs per starting iPSC, despite producing more MKs compared to DD methods, with a technique that was not very amenable to large scale manufacture. The second and third chapters of my thesis have shown the process of working towards a protocol which is a more a viable option for scaling up MK-FoP, using an inducible PSC line. Inducible MK-FoP offers a huge advantage over lentiviral FoP, by greatly

reducing the scale-up costs that will be involved, due to the elimination of lentiviral particles. Now, we also know that FoP-MKs can be cultured long term, with continued proliferation up to around 100 days, allowing more MKs to be produced per PSC seeded. This also allows banks of MK progenitors to be cryopreserved, which once thawed, quickly (in less than 7 days) generate platelets. Our understanding of long-term FoP-MK cultures is improving, thanks to work at Loughborough University by PhD student Elizabeth Cheesman, and we have been able to maximise MK cell number by minimising cell handling and following stricter cell feeding regimes. However, despite there being a great number of labs around the world aiming to produce cell types for use in a clinical setting, there still remains a lack of progress in the large-scale manufacture of cells that would meet this aim.

The second bottleneck is that the MKs produced *in vitro*, are poor producers of platelets, producing on average just 1 platelet per MK, compared to approximately 2000 platelets per MK *in vivo*. While other groups working towards the same goal have published studies that show higher platelet production per *in vitro* MK, there have been doubts cast over how accurately these 'platelet' events have been characterised, as many fail to discriminate between live and dead cells. Members of the Ghevaert lab have worked on developing protocols which accurately determine the number and quality of the platelet-like particles produced from FoP-MKs. This second bottleneck is most likely not going to improve through protocol adjustments, until we move FoP-MKs from a 2D static environment. That is why members of our group are working on developing a bioreactor, alongside a 3D collagen scaffold to mimic the bone marrow niche in which *in vivo* MKs normally reside.

Since *in vivo* MKs take both biochemical and physical cues from their niche environment, we believe that mimicking these cues *in vitro* will offer the best chance of improving platelet yield. We have already seen that simply seeding FoP-MKs onto a 3D scaffold increases the number of platelet-like particles produced, compared to static tissue culture conditions (unpublished). A number of transmembrane proteins have been identified, that can be mobilised onto the 3D scaffolds, which increase proplatelet formation of FoP-MKs (Maria Colzani and Holly Foster, unpublished). A bioreactor has also been developed in our lab, whereby the 3D scaffold sits between two chambers exposed to different flow rates, to mimic the shear stress MKs are exposed to *in vivo* from blood flow. Individually, all three of these components (3D scaffold, transmembrane proteins and bioreactor) increase platelet yield from FoP-MKs, but the idea is to combine all of these approaches, which will hopefully have a cumulative effect on increasing platelet production *in vitro*.

In collaboration with a company (Platelet BioGenesis, borne from Joseph Italiano and Jonathan Thon's labs at Harvard University in the USA), inducible and lentiviral FoP-MKs are also being tested in a second bioreactor developed independently of our lab. Platelet BioGenesis no longer rely on

their own DD method of generating PSC-derived MKs, due to low cell numbers, and has bought exclusive rights to the license for MK-FoP in order to try and generate *in vitro* platelets using their bioreactor. The lab of Alessandra Balduini, (Pavia University, Italy and Boston University, USA) also collaborates with our group to use FoP-MKs in their own version of a bioreactor, using silk-based vascular tubes, to study the biology of platelet formation and release, with the hope that increasing our knowledge about these processes will help to improve the yield *in vitro*. Over the past few years many of our once competitors are now collaborators, and this joining of forces can only be beneficial overall to achieving the ultimate goal of PSC derived platelets being used in clinical transfusion.

Bibliography

Acker JP, Marks DC, Sheffield WP. “Quality Assessment of Established and Emerging Blood Components for Transfusion.” *Journal of Blood Transfusion.* (2016): e4860284.

Adams GB, Chabner KT, Alley IR, Olson DP, Szczepiorkowski ZM, Poznansky MC, Kos CH, Pollak MR, Brown EM, Scadden DT. “Stem cell engraftment at the endosteal niche is specified by the calcium-sensing receptor.” *Nature* 439.7076 (2006): 599-603.

Aifantis I, Raetz E, Buonamici S. “Molecular pathogenesis of T-cell leukaemia and lymphoma.” *Nature Reviews Immunology* 8.5 (2008): 380-90.

Al Yacoub N, Romanowska M, Haritonova N, Foerster J. “Optimized production and concentration of lentiviral vectors containing large inserts.” *The Journal of Gene Medicine* 9.7 (2007): 579-84.

Ala F, Allain JP, Bates I, Boukef K, Boulton F, Brandful J, Dax EM, El Ekiaby M, Farrugia A, Gorlin J, Hassall O, Lee H, Loua A, Maitland K, Mbanya D, Mukhtar Z, Murphy W, Opare-Sem O, Owusu-Ofori S, Reesink H, Roberts D, Torres O, Totoe G et al. “External financial aid to blood transfusion services in sub-Saharan Africa: a need for reflection.” *PLoS Medicine* 9.9 (2012): e1001309.

Amado RG, Chen IS. “Lentiviral vectors; the promise of gene therapy within reach?” *Science* 285.5428 (1999): 674-6.

Ano S, Pereira R, Pironin M, Lesault I, Milley C, Lebigot I, Quang CT, Ghysdael J. “Erythroblast transformation by FLI-1 depends upon its specific DNA binding and transcriptional activation properties.” *The Journal of Biological Chemistry* 279.4 (2004): 2993-3002.

Arai F, Hirao A, Ohmura M, Sato H, Matsuoka S, Takubo K, Ito K, Koh GY, Suda T. “Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche.” *Cell* 118.2 (2004): 149-61.

Arya RC, Wander GS, Gupta P. “Blood component therapy: Which, when and how much.” *Journal of Anaesthesiology Clinical Pharmacology.* 27.2 (2011): 278-84.

Athanasiou M, Mavrothalassitis G, Sun-Hoffman L, Blair DG. “FLI-1 is a suppressor of erythroid differentiation in human hematopoietic cells.” *Leukemia* 14.3 (2000): 439-45.

Ausubel LJ, Hall C, Sharma A, Shakeley R, Lopez P, Quezada V, Couture S, Laderman K, McMahon R, Huang P, Hsu D, Couture L. “Production of CGMP-Grade Lentiviral Vectors.” *BioProcess International.* 10.2 (2012): 32-43.

Avecilla ST, Hattori K, Heissig B, Tejada R, Liao F, Shido K, Jin DK, Dias S, Zhang F, Hartman TE, Hackett NR, Crystal RG, Witte L, Hicklin DJ, Bohlen P, Eaton D, Lyden D, de Sauvage F, Rafii S. "Chemokine-mediated interaction of hematopoietic progenitors with the bone marrow vascular niche is required for thrombopoiesis." *Nature Medicine* 10.1 (2004): 64-71.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. "CRISPR provides acquired resistance against viruses in prokaryotes." *Science.* 315 (2007): 1709-12.

Bartley TD, Bogenberger J, Hunt P, Li YS, Lu HS, Martin F, Chang MS, Samal B, Nichol JL, Swift S, et al. "Identification and cloning of a megakaryocyte growth and development factor that is a ligand for the cytokine receptor Mpl." *Cell* 77.7 (1994): 1117-24.

Bartocci A, Mastrogiannis DS, Migliorati G, Stockert RJ, Wolkoff AW, Stanley ER. "Macrophages specifically regulate the concentration of their own growth factor in the circulation." *Proceedings of the National Academy of Sciences of the United States of America.* 84.17 (1987): 6179-83.

Bates I, Chapotera GK, McKew S, van den Broek N. "Maternal mortality in sub-Saharan Africa: the contribution of ineffective blood transfusion services." *BJOG: an international journal of obstetrics and gynaecology.* 115.11 (2008): 1331-9.

Becker AJ, McCulloch CE, Till JE. "Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells." *Nature* 197 (1963): 452–454.

Becker RP, De Bruyn PP. "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* 145.2 (1976): 183-205.

Ben-David Y, Giddens EB, Bernstein A. "Identification and mapping of a common proviral integration site Fli-1 in erythroleukemia cells induced by Friend murine leukemia virus." *Proceedings of the National Academy of Sciences of the United States of America.* 87.4 (1990): 1332-6.

Benezra R, Davis RL, Lockshon D, Turner DL, Weintraub H. "The protein Id: a negative regulator of helix-loop-helix DNA binding proteins." *Cell* 61.1 (1990): 49-59.

Bertero A, Pawlowski M, Ortmann D, Snijders K, Yiangou L, Cardoso de Brito M, Brown S, Bernard WG, Cooper JD, Giacomelli E, Gambardella L, Hannan NR, Iyer D, Sampaziotis F, Serrano F, Zonneveld MC, Sinha S, Kotter M, Vallier L. "Optimized inducible shRNA and CRISPR/Cas9

platforms for in vitro studies of human development using hPSCs." *Development* 143.23 (2016): 4405-4418.

Bessman JD. "Clinical Methods: The History, Physical, and Laboratory Examinations. 3rd edition." (1990): Chapter 156: Reticulocytes.

Blake JA, Christie KR, Dolan ME, Drabkin HJ, Hill DP, Ni L, Sitnikov D, Burgess S, Buza T, Gresham C, McCarthy F, Pillai L, Wang H, Carbon S, Dietze H, Lewis SE, Mungall CJ, Munoz-Torres MC, Feuermann M, Gaudet P, Basu S, Chisholm RL, Dodson RJ et al. "Gene ontology consortium: Going forward." *Nucleic Acids Research.* 43 (2015): D1049-56.

Blundell J. "Experiments on the Transfusion of Blood by the Syringe." *Medico-Chirurgical Transactions* 9.1 (1818): 56-92.

Bluteau D, Lordier L, Di Stefano A, Chang Y. " Regulation of megakaryocyte maturation and platelet formation." *The Journal of Thrombosis and Haemostasis* 7 (2009): 227–234.

Boch J, Bonas U. "Xanthomonas AvrBs3 family-type III effectors: discovery and function." *Annual Review of Phytopathology* 48 (2010): 419-36.

Breda L, Motta I, Lourenco S, Gemmo C, Deng W, Rupon JW, Abdulmalik OY, Manwani D, Blobel GA, Rivella S. "Forced chromatin looping raises fetal hemoglobin in adult sickle cells to higher levels than pharmacologic inducers." *Blood* 128.8 (2016): 1139-43.

Breton-Gorius J, Favier R, Guichard J, Cherif D, Berger R, Debili N, Vainchenker W, Douay L. "A new congenital dysmegakaryopoietic thrombocytopenia (Paris-Trousseau) associated with giant platelet alpha-granules and chromosome 11 deletion at 11q23." *Blood* 85.7 (1995): 1805-14.

Broos K, Feys HB, De Meyer SF, Vanhoorelbeke K, Deckmyn H. "Platelets at work in primary hemostasis." *Blood Reviews* 25.4 (2011): 155-67.

Broudy VC, Lin NL, Sabath DF, Papayannopoulou T, Kaushansky K. "Human platelets display high-affinity receptors for thrombopoietin." *Blood* 89.6 (1997): 1896-904.

Carey BW, Markoulaki S, Hanna JH, Faddah DA, Buganim Y, Kim J, Ganz K, Steine EJ, Cassady JP, Creyghton MP, Welstead GG, Gao Q, Jaenisch R. "Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells." *Cell Stem Cell* 9.6 (2011): 588-98.

Carson JL, Stanworth SJ, Roubinian N, Fergusson DA, Triulzi D, Doree C, Hebert PC. “Transfusion thresholds and other strategies for guiding allogeneic red blood cell transfusion.” *The Cochrane Database of Systematic Reviews.* 10 (2016).

Carver-Moore K, Broxmeyer HE, Luoh SM, Cooper S, Peng J, Burstein SA, Moore MW, de Sauvage FJ. “Low levels of erythroid and myeloid progenitors in thrombopoietin and c-mpl deficient mice.” *Blood* 88.3 (1996): 803-8.

Challen GA, Boles NC, Chambers SM, Goodell MA. “Distinct Hematopoietic Stem Cell Subtypes Are Differentially Regulated by TGFβ1.” *Cell Stem Cell* 6.3 (2010): 265-78.

Chang Y, Bluteau D, Debili N, Vainchenker W. “From hematopoietic stem cells to platelets.” *Journal of Thrombosis and Haemostasis* 5 (2007): 318-27.

Chari R, Mali P, Moosburner M,Church GM. “Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach.” *Nature Methods* 12 (2015): 823–826.

Chasis JA, Prenant M, Leung A, Mohandas N. “Membrane assembly and remodeling during reticulocyte maturation.” *Blood* 74.3 (1989): 1112-20.

Chen K, Liu J, Heck S, Chasis JA, An X, Mohandas N. “Resolving the distinct stages in erythroid differentiation based on dynamic changes in membrane protein expression during erythropoiesis.” *Proceedings of the National Academy of Sciences of the United States of America.* 106.41 (2009): 17413-8.

Chen L, Kostadima M, Martens JHA, Canu G, Garcia SP, Turro E, Downes K, Macaulay IC, Bielczyk-Maczynska E, Coe S, Farrow S, Poudel P, Burden F et al. “Transcriptional diversity during lineage commitment of human blood progenitors.” *Science* 345.6204 (2014): 1251033.

Chen Q, Solar G, Eaton DL, de Sauvage FJ. “IL-3 does not contribute to platelet production in c-Mpl-deficient mice.” *Stem Cells* 16 (1998): 31-6.

Choi ES, Nichol JL, Hokom MM, Hornkohl AC, Hunt P. “Platelets generated in vitro from proplatelet-displaying human megakaryocytes are functional.” *Blood* 85.2 (1995): 402-13.

Cortin V, Garnier A, Pineault N, Lemieux R, Boyer L, Proulx C. “Efficient in vitro megakaryocyte maturation using cytokine cocktails optimized by statistical experimental design.” *Experimental Hematology* 33.10 (2005): 1182-91.

Cosman, D. “The Hematopoietin Receptor Superfamily.” *Cytokine* 5 (1993): 95-106.

Costa M, Dottori M, Sourris K, Jamshidi P, Hatzistavrou T, Davis R, Azzola L, Jackson S, Lim SM, Pera M, Elefanty AG, Stanley EG. “A method for genetic modification of human embryonic stem cells using electroporation.” *Nature Protocols* 2.4 (2007): 792-6.

Dahr W, Beyreuther K, Moulds JJ. “Structural analysis of the major human erythrocyte membrane sialoglycoprotein from Miltenberger class VII cells.” *European Journal of Biochemistry.* 166.1 (1987): 27-30.

De Sauvage FJ, Hass PE, Spencer SD, Malloy BE, Gurney AL, Spencer SA, Darbonne WC, Henzel WJ, Wong SC, Kuang WJ, et al. “Stimulation of megakaryocytopoiesis and thrombopoiesis by the c-Mpl ligand.” *Nature* 369.6481 (1994): 533-8.

Deveaux S, Filipe A, Lemarchandel V, Ghysdael J, Roméo PH, Mignotte V. “Analysis of the thrombopoietin receptor (MPL) promoter implicates GATA and Ets proteins in the coregulation of megakaryocyte-specific genes.” *Blood* 87.11 (1996): 4678-85.

Dias J, Gumenyuk M, Kang H, Vodyanik M, Yu J, Thomson JA, Slukvin II. “Generation of red blood cells from human induced pluripotent stem cells.” *Stem Cells and Development* 20.9 (2011): 1639-47.

Domen J, Weissman IL. “Hematopoietic Stem Cells Need Two Signals to Prevent Apoptosis; Bcl-2 Can Provide One of These, Kitl/C-KIT Signaling the Other.” *The Journal of Experimental Medicine* 192.12 (2000): 1707–1718.

Donahue RE, Seehra J, Metzger M, Lefebvre D, Rock B, Carbone S, Nathan DG, Garnick M, Sehgal PK, Laston D, et al. “Human IL-3 and GM-CSF act synergistically in stimulating hematopoiesis in primates.” *Science* 421.4874 (1988): 1820-3.

Doré LC, Chlon TM, Brown CD, White KP, Crispino JD. “Chromatin occupancy analysis reveals genome-wide GATA factor switching during hematopoiesis.” *Blood* 119.16 (2012): 3724-33.

Dorn I, Klich K, Arauzo-Bravo MJ, Radstaak M, Santourlidis S, Ghanjati F, Radke TF, Psathaki OE, Hargus G, Kramer J, Einhaus M, Kim JB, Kögler G, Wernet P, Schöler HR, Schlenke P, Zaehres H. “Erythroid differentiation of human induced pluripotent stem cells is independent of donor cell type of origin.” *Haematologica* 100.1 (2015): 32-41.

Doronina VA, Wu C, de Felipe P, Sachs MS, Ryan MD, Brown JD. “Site-specific release of nascent chains from ribosomes at a sense codon.” *Molecular Cell Biology* 28.13 (2008): 4227-39.

Du LM, Nurden P, Nurden AT, Nichols TC, Bellinger DA, Jensen ES, Haberichter SL, Merricks E, Raymer RA, Fang J, Koukouritaki SB, Jcobi PM, Hawkins TB, Cornetta K, Shi Q, Wilcox DA. "Platelet-targeted gene therapy with human factor VIII establishes haemostasis in dogs with haemophilia A." *Nature Commuinications* 4 (2013): 2773.

Elefanty AG, Robb L, Birner R, Begley CG. "Hematopoietic-specific genes are not induced during in vitro differentiation of scl-null embryonic stem cells." *Blood* 90.4 (1997): 435-47.

Engler C, Marillonnet S. "Golden Gate Cloning." *Methods in Molecular Biology* 1116 (2004): 119-31.

Enver T, Heyworth CM, Dexter TM. "Do stem cells play dice?" *Blood* 92.2 (1998): 348-51.

Estcourt LJ, . "Why has demand for platelet components increased? A review." *Transfusion Medicine* 24.5 (2014): 260-8.

Evans MJ, Kaufman MH. "Establishment in culture of pluripotential cells from mouse embryos." *Nature* 292.5819 (1981): 154-6.

Ferreira R, Ohneda K, Yamamoto M, Philipsen S. "GATA1 function, a paradigm for transcription factors in hematopoiesis." *Molecular Cell Biology* 25.4 (2005): 1215-27.

Friedrich G, Soriano P. "Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice." *Genes & Development.* 5.9 (1991): 1513-23.

Gainsford T, Nandurkar H, Metcalf D, Robb L, Begley CG, Alexander WS. "The residual megakaryocyte and platelet production in c-mpl-deficient mice is not dependent on the actions of interleukin-6, interleukin-11, or leukemia inhibitory factor." *Blood* 95.2 (2000): 528-34.

Gaj T, Gersbach CA, Barbas CF. "ZFN, TALEN and CRISPR/Cas-based methods for genome engineering." *Trends in Biotechnology* 31.7 (2013): 397–405.

Gaur M, Kamata T, Wang S, Moran B, Shattil SJ, Leavitt AD. "Megakaryocytes derived from human embryonic stem cells: a genetically tractable system to study megakaryocytopoiesis and integrin function." *Journal of Thrombosis and Haemostasis.* 4.2 (2006): 436-42.

Geddis AE, Fox NE, Kaushansky K. "The Mpl receptor expressed on endothelial cells does not contribute significantly to the regulation of circulating thrombopoietin levels." *Experimental Hematology* 34.1 (2006): 82-6.

Gerrard JM, White JG, Rao GH, Townsend D. “Localization of platelet prostaglandin production in the platelet dense tubular system.” *American Journal of Pathology* 83.2 (1976): 283-98.

Giarratana MC, Kobari L, Lapillonne H, Chalmers D, Kiger L, Cynober T, Marden MC, Wajcman H, Douay L. “Ex vivo generation of fully mature human red blood cells from hematopoietic stem cells.” *Nature Biotechnology* 23.1 (2005): 69-74.

Gieger C, Radhakrishnan A, Cvejic A, Tang W, Porcu E, Pistis G, Serbanovic-Canic J, Elling U, Goodall AH, Labrune Y, Lopez LM, Mägi R, Meacham S, Okada Y, Pirastu N, Sorice R, Teumer A, Voss K, Zhang W, Ramirez-Solis R, Bis JC, Ellinghaus D et al. “New gene functions in megakaryopoiesis and platelet formation.” *Nature.* 7376.480 (2011): 201-8.

González F, Boué S, Izpisúa Belmonte JC. “Methods for making induced pluripotent stem cells: reprogramming à la carte.” *Nature Reviews. Genetics.* 12.4 (2011): 231-42.

Goodman JW, Hodgson GS. “Evidence for stem cells in the peripheral blood of mice.” *Blood* 19 (1962): 702–714.

Göttgens B, Nastos A, Kinston S, Piltz S, Delabesse EC, Stanley M, Sanchez MJ, Ciau-Uitz A, Patient R, Green AR. “Establishing the transcriptional programme for blood: the SCL stem cell enhancer is regulated by a multiprotein complex containing Ets and GATA factors.” *The EMBO Journal* 21.12 (2002): 3039-50.

Green, T. “Haematopoiesis. Master regulator unmasked.” *Nature* 383.6601 (1996): 577.

Griffiths, JB. “The Effect of Insulin on the Growth and Metabolism of the Human Diploid Cell, Wi-38.” *Journal of Cell Science* 7 (1970): 575-585.

Grozovsky R, Begonja AJ, Liu K, Visner G, Hartwig JH, Falet H, Hoffmeister KM. “The Ashwell-Morell receptor regulates hepatic thrombopoietin production via JAK2-STAT3 signaling.” *Nature Medicine* 21.1 (2015): 47-54.

Hanspal M, Prchal JT, Palek J. “Biogenesis of erythrocyte membrane skeleton in health and disease.” *Stem Cells.* 11.1 (1993): 8-12.

Harker LA, Marzec UM, Hunt P, Kelly AB, Tomer A, Cheung E, Hanson SR, Stead RB. “Dose-response effects of pegylated human megakaryocyte growth and development factor on platelet production and function in nonhuman primates.” *Blood* 88.2 (1996): 511-21.

Harris AM, Atterbury CLJ, Chaffe B, Elliott C, Hawkins T, Hennem SJ, Howell C, Jones J, Murray S, New HV, Norfolk D, Pirie L, Russell J, Taylor C. “Guideline on the Administration of Blood Components.” *British Committee for Standards in Haematology* (2012).

Heijnen HF, Debili N, Vainchencker W, Breton-Gorius J, Geuze HJ, Sixma JJ. “Multivesicular bodies are an intermediate stage in the formation of platelet alpha-granules.” *Blood* 91.7 (1998): 2313-25.

Hendriks W, Warren C, Cowan C. “Genome Editing in Human Pluripotent Stem Cells: Approaches, Pitfalls, and Solutions.” *Cell Stem Cell* 18.1 (2016): 53-65.

Hockemeyer D, Soldner F, Beard C, Gao Q, Mitalipova M, DeKelver RC, Katibah GE, Amora R, Boydston EA, Zeitler B, Meng X, Miller JC, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Jaenisch R. “Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases.” *Nature Biotechnology* 27.9 (2009): 851-7.

Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, Cost GJ, Zhang L, Santiago Y, Miller JC, Zeitler B, Cherone JM, Meng X, Hinkley SJ, Rebar EJ, Gregory PD, Urnov FD, Jaenisch R. “Genetic engineering of human pluripotent cells using TALE nucleases.” *Nature Biotechnology* 29.8 (2011): 731-4.

Hsu HL, Huang L, Tsan JT, Funk W, Wright WE, Hu JS, Kingston RE, Baer R. “Preferred sequences for DNA recognition by the TAL1 helix-loop-helix proteins.” *Molecular Cell Biology* 14.2 (1994): 1256-65.

Huang S, Guo YP, May G, Enver T. “Bifurcation dynamics in lineage-commitment in bipotent progenitor cells.” *Developmental Biology* 305.2 (2007): 695-713.

Iancu-Rubin C, Gajzer D, Tripodi J, Najfeld V, Gordon RE, Hoffman R, Atweh GF. “Down-regulation of stathmin expression is required for megakaryocyte maturation and platelet production.” *Blood* 117.17 (2011): 4580-9.

Italiano JE Jr, Lecine P, Shivdasani RA, Hartwig JH. “Blood platelets are assembled principally at the ends of proplatelet processes produced by differentiated megakaryocytes.” *The Journal of Cell Biology* 147.13 (1999): 1299-312.

J., Wright. “The origin and nature of blood platelets.” *Boston Medical and Surgical Journal* 154 (1906): 643-45.

Jaenisch R, Bird R. “Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals.” *Nature Genetics* 33 (2003): 245-54.

JD, Joung JK and Sander. “TALENs: a widely applicable technology for targeted genome editing.” *Nature Reviews. Molecular Cell Biology* 14.1 (2010): 49-55.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. “A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity.” *Science.* 17.6096 (2012): 816-21.

Johnson P, Friedmann T. “Limited bidirectional activity of two housekeeping gene promoters: human HPRT and PGK.” *Gene* 88.2 (1990): 207-13.

Junt T, Schulze H, Chen Z, Massberg S, Goerge T, Krueger A, Wagner DD, Graf T, Italiano JE Jr, Shivdasani RA, von Andrian UH. “Dynamic visualization of thrombopoiesis within bone marrow.” *Science* 317.5845 (2007): 1767-70.

K, Hu. “Vectorology and Factor Delivery in Induced Pluripotent Stem Cell Reprogramming.” *Stem Cells and Development* 23.12 (2014): 1301–1315.

Kahner BN, Shankar H, Murugappan S, Prasad GL, Kunapuli SP. “Nucleotide receptor signaling in platelets.” *Journal of Thrombosis and Haemostasis.* 4 (2006): 2317-26.

Kanda T, Sullivan KF, Wahl GM. “Histone–GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells.” *Current Biology* 8.7 (1998): 377-85.

Kaneko H, Kobayashi E, Yamamoto M, Shimizu R. “N- and C-terminal transactivation domains of GATA1 protein coordinate hematopoietic program.” *The Journal of Biological Chemistry* 287.25 (2012): 21439-49.

Karim FD, Urness LD, Thummel CS, Klemsz MJ, McKercher SR, Celada A, Van Beveren C, Maki RA, Gunther CV, Nye JA, et al. “The ETS-domain: a new DNA-binding motif that recognizes a purine-rich core DNA sequence.” *Genes & Development* 4.9 (1990): 1451-3.

Kaser A, Brandacher G, Steurer W, Kaser S, Offner FA, Zoller H, Theurl I, Widder W, Molnar C, Ludwiczek O, Atkins MB, Mier JW, Tilg H. “Interleukin-6 stimulates thrombopoiesis through thrombopoietin: role in inflammatory thrombocytosis.” *Blood* 98.9 (2001): 2720-5.

Kelemen E, Cserhati I, Tanos B. “Demonstration and some properties of human thrombopoietin in thrombocythaemic sera.” *Haematologica* 20.6 (1958): 350-5.

Kent DG, Copley MR, Benz C, Wöhrer S, Dykstra BJ, Ma E, Cheyne J, Zhao Y, Bowie MB, Zhao Y, Gasparetto M, Delaney A, Smith C, Marra M, Eaves CJ. “Prospective isolation and molecular characterization of hematopoietic stem cells with durable self-renewal potential.” *Blood* 113.25 (2009): 6342-50.

Kie JH, Yang WI, Lee MK, Kwon TJ, Min YH, Kim HO, Ahn HS, Im SA, Kim HL, Park HY, Ryu KH, Chung WS, Shin MH, Jung YJ, Woo SY, Park HK, Seoh JY. “Decrease in apoptosis and increase in polyploidization of megakaryocytes by stem cell factor during ex vivo expansion of human cord blood CD34+ cells using thrombopoietin.” *Stem Cells* 20.1 (2002): 73-9.

Kilpinen H, Goncalves A, Leha A, Afzal V, Alasoo K, Ashford S, Bala S, Bensaddek D, Casale FP, Culley OJ, Danecek P, Faulconbridge A, Harrison PW, Kathuria A, McCarthy D, McCarthy SA, Meleckyte R, Memari Y, Moens N, Soares F, Mann A, Streeter I et al. “Common genetic variation drives molecular heterogeneity in human iPSCs.” *Nature* 546.7658 (2017): 370-5.

Kimura S, Roberts AW, Metcalf D, Alexander WS. “Hematopoietic stem cell deficiencies in mice lacking c-Mpl, the receptor for thrombopoietin.” *Proceedings of the National Academy of Sciences* 95.3 (1998): 1195-200.

Kiselev VY, Kirschner K, Schaub MT, Andrews T, Yiu A, Chandra T, Natarajan KN, Reik W, Barahona M, Green AR, Hemberg M. “SC3: consensus clustering of single-cell RNA-seq data.” *Nature Methods* 14.5 (2017): 483-486.

Klimchenko O, Mori M, Distefano A, Langlois T, Larbret F, Lecluse Y, Feraud O, Vainchenker W, Norol F, Debili N. “A common bipotent progenitor generates the erythroid and megakaryocyte lineages in embryonic stem cell-derived primitive hematopoiesis.” *Blood* 114.8 (2009): 1506-17.

Klimchenko O, Mori M, Distefano A, Langlois T, Larbret F, Lecluse Y, Feraud O, Vainchenker W, Norol F, Debili N. “A common bipotent progenitor generates the erythroid and megakaryocyte lineages in embryonic stem cell-derived primitive hematopoiesis.” *Blood.* 114.8 (2009): 1506-17.

Koury MJ, Bondurant MC. “Erythropoietin retards DNA breakdown and prevents programmed death in erythroid progenitor cells.” *Science* 248.4953 (1990): 378-81.

Kunisaki Y, Bruns I, Scheiermann C, Ahmed J, Pinho S, Zhang D, Mizoguchi T, Wei Q, Lucas D, Ito K, Mar JC, Bergman A, Frenette PS. “Arteriolar niches maintain haematopoietic stem cell quiescence.” *Nature* 502.7473 (2013): 637-43.

Kurita R, Suda N, Sudo K, Miharada K, Hiroyama T, Miyoshi H, Tani K, Nakamura Y. “Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells.” *PLoS One.* 8.3 (2013): e59890.

Lacombe J, Krosl G, Tremblay M, Gerby B, Martin R, Aplan PD, Lemieux S, Hoang T. “Genetic interaction between Kit and Scl.” *Blood* 122.7 (2013): 1150-61.

Landsteiner K, . “Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe. .” *Zentbl. Bakt. Orig* 27 (1900): 357-362995-8.

Lapillonne H, Kobari L, Mazurier C, Tropel P, Giarratana MC, Zanella-Cleon I, Kiger L, Wattenhofer-Donzé M, Puccio H, Hebert N, Francina A, Andreu G, Viville S, Douay L. “Red blood cell generation from human induced pluripotent stem cells: perspectives for transfusion medicine.” *Haematologica* 95.10 (2010): 1651-9.

Lee R, Kertesz N, Joseph SB, Jegalian A, Wu H. “Erythropoietin (Epo) and EpoR expression and 2 waves of erythropoiesis.” *Blood* 98.5 (2001): 1408-15.

Lefrançais E, Ortiz-Muñoz G, Caudrillier A, Mallavia B, Liu F, Sayah DM, Thornton EE, Headley MB, David T, Coughlin SR, Krummel MF, Leavitt AD, Passegué E, Looney MR. “The lung is a site of platelet biogenesis and a reservoir for haematopoietic progenitors.” *Nature* 6.544 (2017): 105-9.

Lemarchandel V, Ghysdael J, Mignotte V, Rahuel C, Roméo PH. “GATA and Ets cis-acting sequences mediate megakaryocyte-specific expression.” *Molecular and Cell Biology* 13.1 (1993): 668-76.

Leparc GF. “Safety of the blood supply.” *Cancer Control* 22.1 (2015): 7-15.

Li XJ, Du ZW, Zarnowska ED, Pankratz M, Hansen LO, Pearce RA, Zhang SC. “Specification of motoneurons from human embryonic stem cells.” *Nature Biotechnology* 23.2 (2005): 215-21.

Liew CG, Draper JS, Walsh J, Moore H, Andrews PW. “Transient and stable transgene expression in human embryonic stem cells.” *Stem Cells* 25.6 (2007): 1521-8.

Linden HM, Kaushansky K. “The glycan domain of thrombopoietin (TPO) acts in trans to enhance secretion of the hormone and other cytokines.” *The Journal of Biological Chemistry* 277.38 (2002): 35240-7.

Ling KW, Ottersbach K, van Hamburg JP, Oziemlak A, Tsai FY, Orkin SH, Ploemacher R, Hendriks RW, Dzierzak E. “GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells.” *The Journal of Experimental Medicine* 7 (2004): 871-82.

Liu F, Walmsley M, Rodaway A, Patient R. “Fli1 acts at the top of the transcriptional network driving blood and endothelial development.” *Current Biology* 18.16 (2008): 1234-40.

Liu JC, Lerou PH, Lahav G. “Stem cells: balancing resistance and sensitivity to DNA damage.” *Trends in Cell Biology* 24 (2014): 268–274.

Lok S, Kaushansky K, Holly RD, Kuijper JL, Lofton-Day CE, Oort PJ, Grant FJ, Heipel MD, Burkhead SK, Kramer JM, et al. “Cloning and expression of murine thrombopoietin cDNA and stimulation of platelet production in vivo.” *Nature* 369.6481 (1994): 565-8.

Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Ando D, Urnov FD, Galli C, Gregory PD, Holmes MC, Naldini L. “Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery.” *Nature Biotechnology* 25.11 (2007): 1298-306.

Long MW, Heffner CH, Williams JL, Peters C, Prochownik EV. “Regulation of megakaryocyte phenotype in human erythroleukemia cells.” *Journal of Clinical Investigation* 85.4 (1990): 1072-84.

Lordier L, Jalil A, Aurade F, Larbret F, Larghero J, Debili N, Vainchenker W, Chang Y. “Megakaryocyte endomitosis is a failure of late cytokinesis related to defects in the contractile ring and Rho/Rock signaling.” *Blood* 112.8 (2008): 3164-74.

Love MI, Huber W, Anders S. “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.” *Genome Biology* 15.12 (2014): 550.

Lu SJ, Li F, Yin H, Feng Q, Kimbrel EA, Hahm E, Thon JN, Wang W, Italiano JE, Cho J, Lanza R. “Platelets generated from human embryonic stem cells are functional in vitro and in the microcirculation of living mice.” *Cell Research* 21.3 (2011): 530-45.

Manwaani D, Biecker J. “The Erythroblastic Island.” *Current Topics in Developmental Biology* 82 (2008): 23-53.

Mao B, Huang S, Lu X, Sun W, Zhou Y, Pan X, Yu J, Lai M, Chen B, Zhou Q, Mao S, Bian G, Zhou J, Nakahata T, Ma F. “Early Development of Definitive Erythroblasts from Human Pluripotent

Stem Cells Defined by Expression of Glycophorin A/CD235a, CD34, and CD36." *Stem Cell Reports* 7.5 (2016): 869-883.

Martin GR. "Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells." *Proceedings of the National Academy of Sciences of the United States of America.* 78.12 (1981): 7634-8.

Martin R, Lahlil R, Damert A, Miquerol L, Nagy A, Keller G, Hoang T. "SCL interacts with VEGF to suppress apoptosis at the onset of hematopoiesis." *Development* 131.3 (2004): 693-702.

Matsuoka S, Ebihara Y, Xu M, Ishii T, Sugiyama D, Yoshino H, Ueda T, Manabe A, Tanaka R, Ikeda Y, Nakahata T, Tsuji K. "CD34 expression on long-term repopulating hematopoietic stem cells changes during developmental stages." *Blood* 97.2 (2001): 419-25.

Maynard DM, Heijnen HF, Horne MK, White JG, Gahl WA. "Proteomic analysis of platelet alpha-granules using mass spectrometry." *Journal of Thrombosis and Haemostasis.* 5.9 (2007): 1945-55.

McCarthy DJ, Campbell KR, Lun AT, Wills QF. "Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R." *Bioinformatics* 33.8 (2017): 1179-1186.

McGrath K, Palis J. "Ontogeny of erythropoiesis in the mammalian embryo." *Current Topics in Developmental Biology* 82 (2008): 1-22.

Méndez-Ferrer S, Michurina TV, Ferraro F, Mazloom AR, Macarthur BD, Lira SA, Scadden DT, Ma'ayan A, Enikolopov GN, Frenette PS. "Mesenchymal and haematopoietic stem cells form a unique bone marrow niche." *Nature* 466.7308 (2010): 829-34.

Merika M, Orkin SH. "DNA-binding specificity of GATA family transcription factors." *Molecular Cell Biology* 13.7 (1993): 3999-4010.

Mizuguchi H, Xu Z, Ishii-Watabe A, Uchida E, Hayakawa T. "IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector." *Molecular Therapy* 1.4 (2000): 376-82.

Moreau T, Evans AL, Vasquez L, Tijssen MR, Yan Y, Trotter MW, Howard D, Colzani M, Arumugam M, Wu WH, Dalby A, Lampela R, Bouet G, Hobbs CM, Pask DC, Payne H, Ponomaryov T, Brill A, Soranzo N, Ouwehand WH, Pedersen RA, Ghevaert C. "Large scale production of megakaryocytes from human pluripotent stem cells by chemically defined forward programming." *Nature Communications* 7.7 (2016): 11208.

Mollison PL, Engelfriet CP, Conteras M. *The Rh blood group system.* 9th. Oxford: Black well Scientific Publication, 1993.

Morel O, Morel N, Freyssinet JM, Toti F. "Platelet microparticles and vascular cells interactions: a checkpoint between the haemostatic and thrombotic responses." *Platelets* 19.1 (2008): 9-23.

Nakamura S, Takayama N, Hirata S, Seo H, Endo H, Ochi K, Fujita K, Koike T, Harimoto K, Dohda T, Watanabe A, Okita K, Takahashi N, Sawaguchi A, Yamanaka S, Nakauchi H, Nishimura S, Eto K. "Expandable megakaryocyte cell lines enable clinically applicable generation of platelets from human induced pluripotent stem cells." *Cell Stem Cell.* 14.4 (2014): 535-48.

Ng AP, Kauppi M, Metcalf D, Di Rago L, Hyland CD, Alexander WS. "Characterization of thrombopoietin (TPO)-responsive progenitor cells in adult mouse bone marrow with in vivo megakaryocyte and erythroid potential." *Proc Natl Acad Sci USA* 109.7 (2012): 2364-9.

NHSBT. "NHS Blood and Transplant leads global campaign for blood donors of the future." *News Release* (2016).

Nilsson SK, Johnston HM, Coverdale JA. "Spatial localization of transplanted hemopoietic stem cells: inferences for the localization of stem cell niches." *Blood* 97 (2001): 2293-2299.

Ogata T, Kozuka T, Kanda T. "Identification of an insulator in AAVS1, a preferred region for integration of adeno-associated virus DNA." *Journal of Virology* 77.16 (2003): 9000-7.

Ogawa M. "Differentiation and proliferation of hematopoietic stem cells." *Blood* 81.11 (1993): 2844-53.

Ohno T, Rao VN, Reddy ES. "EWS/Fli-1 chimeric protein is a transcriptional activator." *Cancer Research* 53.24 (1993): 5859-63.

Okada Y, Nobori H, Shimizu M, Watanabe M, Yonekura M, Nakai T, Kamikawa Y, Wakimura A, Funahashi N, Naruse H, Watanabe A, Yamasaki D, Fukada S, Yasui K, Matsumoto K, Sato T, Kitajima K, Nakano T, Aird WC, Doi T. "Multiple ETS family proteins regulate PF4 gene expression by binding to the same ETS binding site." *PLoS One* 6.9 (2011): e24837.

Olivier EN, Marenah L, McCahill A, Condie A, Cowan S, Mountford JC. "High-Efficiency Serum-Free Feeder-Free Erythroid Differentiation of Human Pluripotent Stem Cells Using Small Molecules." *Stem Cells Translational Medicine* 5.10 (2016): 1394-1405.

Olivier EN, Qiu C, Velho M, Hirsch RE, Bouhassira EE. “Large-scale production of embryonic red blood cells from human embryonic stem cells.” *Experimental Hematology* 34.12 (2006): 1635-42.

Ono M, Hamada Y, Horiuchi Y, Matsuo-Takasaki M, Imoto Y, Satomi K, Arinami T, Hasegawa M, Fujioka T, Nakamura Y, Noguchi E. “Generation of Induced Pluripotent Stem Cells from Human Nasal Epithelial Cells Using a Sendai Virus Vector.” *PLoS One.* 7.8 (2012): e42855.

Osawa M, Hanada K, Hamada H, Nakauchi H. “Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell.” *Science* 273.5272 (1996): 242-245.

Pan BT, Johnstone RM. “Fate of the transferrin receptor during maturation of sheep reticulocytes in vitro: selective externalization of the receptor.” *Cell.* 33.3 (1983): 967-78.

Paralkar VP, Mishra T, Luan J, Yao Y, Kossenkov AV, Anderson SM, Dunain M, Pimkin M, Gore M, Sun D, Konuthula N, Raj A, An X, Mohandas N, Bodine DM, Hardison RC, Weiss M. “Lineage and species-specific long noncoding RNAs during erythro-megakaryocytic development.” *Blood* 123.12 (2014): 1927-37.

Patel SR, Hartwig JH, Italiano JE. “The biogenesis of platelets from megakaryocyte proplatelets.” *The Journal of Clinical Investigation* 115.12 (2005): 3348-54.

Patel-Hett S, Wang H, Begonja AJ, Thon JN, Alden EC, Wandersee NJ, An X, Mohandas N, Hartwig JH, Italiano JE Jr. “The spectrin-based membrane skeleton stabilizes mouse megakaryocyte membrane systems and is essential for proplatelet and platelet formation.” *Blood* 118.6 (2011): 1641-52.

Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, Winter D, Lara-Astiaso D, Gury M, Weiner A, David E, Cohen N, Lauridsen FK, Haas S, Schlitzer A, Mildner A, Ginhoux F, Jung S, Trumpp A, Porse BT, Tanay A, Amit I. “Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors.” *Cell.* 163.7 (2015): 1663-77.

Pawlowski M, Ortmann D, Bertero A, Tavares JM, Pedersen RA, Vallier L, Kotter MR. “Inducible and Deterministic Forward Programming of Human Pluripotent Stem Cells into Neurons, Skeletal Myocytes, and Oligodendrocytes.” *Stem Cell Reports* (2017): S2213-6711(17)30083-8. doi: 10.1016/j.stemcr.2017.02.016. [Epub ahead of print].

Pettitt SJ, Liang Q, Rairdan XY, Moran JL, Prosser HM, Beier DR, Lloyd KC, Bradley A, Skarnes WC. “Agouti C57BL/6N embryonic stem cells for mouse genetic resources.” *Nature Methods* 6.7 (2009): 493-5.

Pevny L, Simon MC, Robertson E, Klein WH, Tsai SF, D'Agati V, Orkin SH, Costantini F. “Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1.” *Nature* 349.6306 (1991): 257-60.

Phillips DR, Charo IF, Parise LV, Fitzgerald LA. “The platelet membrane glycoprotein IIb-IIIa complex.” *Blood.* 71.4 (1988): 831-3.

Picelli S, Faridani OR, Björklund AK, Winberg G, Sagasser S, Sandberg R. “Full-length RNA-seq from single cells using Smart-seq2.” *Nature Protocols* 9.1 (2014): 171-81.

Pimkin M, Kossenkov AV, Mishra T, Morrissey CS, Wu W, Keller CA, Blobel GA, Lee D, Beer MA, Hardison RC, Weiss MJ. “Divergent functions of hematopoietic transcription factors in lineage priming and differentiation during erythro-megakaryopoiesis.” *Genome Research.* 24.12 (2014): 1932-44.

Plé H, Landry P, Benham A, Coarfa C, Gunaratne PH, Provost P. “The Repertoire and Features of Human Platelet microRNAs.” *PLoS One* 7.12 (2012): e50746.

Poulter NS, Pollitt AY, Owen DM, Gardiner EE, Andrews RK, Shimizu H, Ishikawa D, Bihan D, Farndale RW, Moroi M, Watson SP, Jung SM. “Clustering of glycoprotein VI (GPVI) dimers upon adhesion to collagen as a mechanism to regulate GPVI signaling in platelets.” *Journal of Thrombosis and Haemostasis.* 15.3 (2017): 549-564.

Qian H, Buza-Vidas N, Hyland CD, et al. “Critical role of thrombopoietin in maintaining adult quiescent hematopoietic stem cells.” *Cell Stem Cell* 1.6 (2007): 671–684.

Qin JY, Zhang L, Clift KL, Hulur I, Xiang AP, Ren BZ, Lahn BT. “Systematic comparison of constitutive promoters and the doxycycline-inducible promoter.” *PLoS One* 10611 (2010).

Radley JM, Haller CJ. “The demarcation membrane system of the megakaryocyte: a misnomer?” *Blood* 60.1 (1982): 213-9.

Reems JA, Pineault N, Sun S. “In Vitro Megakaryocyte Production and Platelet Biogenesis:State of the Art.” *Transfusion Medicine Reviews* 24.1 (2010): 33-43.

Reith AD, Rottapel R, Giddens E, Brady C, Forrester L, Bernstein A. “W mutant mice with mild or severe developmental defects contain distinct point mutations in the kinase domain of the c-kit receptor.” *Genes & Development* 3 (1990): 390-400.

Reya T, Morrison SJ, Clarke MF, Weissman IL. “Stem cells, cancer, and cancer stem cells.” *Nature* 414.6859 (2001): 105-11.

Richardson JL, Shivdasani RA, Boers C, Hartwig JH, Italiano JE Jr. “Mechanisms of organelle transport and capture along proplatelets during platelet production.” *Blood* 106.13 (2005): 4066-75.

Romania P, Lulli V, Pelosi E, Biffoni M, Peschle C, Marziali G. “MicroRNA 155 modulates megakaryopoiesis at progenitor and precursor level by targeting Ets-1 and Meis1 transcription factors.” *British Journal of Haematology* 143.4 (2008): 570-80.

Rosmarin AG, Yang Z, Resendes KK. “Transcriptional regulation in myelopoiesis: Hematopoietic fate choice, myeloid differentiation, and leukemogenesis.” *Experimental Hematology* 33.2 (2005): 131-43.

Ryu KH, Chun S, Carbonierre S, Im SA, Kim HL, Shin MH, Shin HY, Ahn HS, Woo SY, Seoh JY, Fraser JK. “Apoptosis and megakaryocytic differentiation during ex vivo expansion of human cord blood CD34+ cells using thrombopoietin.” *British Journal of Haematology* 113.2 (2001): 470-8.

Sadelain M, Papapetrou EP, Bushman FD. “Safe harbours for the integration of new DNA in the human genome.” *Nature Reviews Cancer* 12.1 (2011): 51-8.

Sankaran VG, Orkin SH. “The Switch from Fetal to Adult Hemoglobin.” *Cold Spring Harbor Perspectives in Medicine.* 3.1 (2013): a011643.

Savage B, Saldívar E, Ruggeri ZM. “Initiation of platelet adhesion by arrest onto fibrinogen or translocation on von Willebrand factor.” *Cell* 84.2 (1996): 289=97.

Schambach A, Zychlinski D, Ehrnstroem B, Baum C. “Biosafety Features of Lentiviral Vectors.” *Human Gene Therapy* 24.2 (2013): 132-42.

Schofield R. “The relationship between the spleen colony-forming cell and the haemopoietic stem cell.” *Blood Cells* 4.1 (1978): 7-25.

Schulze H, Korpal M, Hurov J, Kim SW, Zhang J, Cantley LC, Graf T, Shivdasani RA. “Characterization of the megakaryocyte demarcation membrane system and its role in thrombopoiesis.” *Blood* 107.10 (2006): 3868-75.

Smith JR, Maguire S, Davis LA, Alexander M, Yang F, Chandran S, ffrench-Constant C, Pedersen RA. "Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration." *Stem Cells* 26.2 (2008): 496-504.

Smith JR, Maguire S, Davis LA, Alexander M, Yang F, Chandran S, Ffrench-Constant C, Pedersen RA. "Robust, Persistent Transgene Expression in Human Embryonic Stem Cells Is Achieved with AAVS1-Targeted Integration." *Stem Cells* 26.2 (2008): 496-504.

Solar GP, Kerr WG, Zeigler FC, Hess D, Donahue C, de Sauvage FJ, Eaton DL. "Role of c-mpl in early hematopoiesis." *Blood* 92.1 (1998): 4-10.

Sommer CA, Stadtfeld M, Murphy GJ, Hochedlinger K, Kotton DN, Mostoslavsky G. "Induced pluripotent stem cell generation using a single lentiviral stem cell cassette." *Stem Cells* 27.3 (2009): 543-9.

Sommer CA, Stadtfeld M, Murphy GJ, Hochedlinger K, Kotton DN, Mostoslavsky G. (n.d.).

Srinivas S, Watanabe T, Lin CS, William CM, Tanabe Y, Jessell TM, Costantini F. "Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus." *BMC Developmental Biology* 1 (2001): 4.

Stanworth SJ, Hudson CL, Estcourt LJ, Johnson RJ, Wood EM. "Risk of bleeding and use of platelet transfusions in patients with hematologic malignancies: recurrent event analysis." *Haematologica.* 100.6 (2015): 740-7.

Stier S, Ko Y, Forkert R, Lutz C, Neuhaus T, Grünewald E, Cheng T, Dombkowski D, Calvi LM, Rittling SR, Scadden DT. "Osteopontin is a hematopoietic stem cell niche component that negatively regulates stem cell pool size." *The Journal of Experimental Medicine* 201.11 (2005): 1781-91.

Suzuki H, Murasaki K, Kodama K, Takayama H. "Intracellular localization of glycoprotein VI in human platelets and its surface expression upon activation." *British Journal of Haematology.* 121.6 (2003): 904-12.

Szymczak AL, Workman CJ, Wang Y, Vignali KM, Dilioglou S, Vanin EF, Vignali DA. "Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector." *Nature Biotechnology* 22.5 (2004): 589-94.

Taichman R, Reilly M, Verma R, Ehrenman K, Emerson S. " Hepatocyte growth factor is secreted by osteoblasts and cooperatively permits the survival of haematopoietic progenitors." *British Journal of Haematology* 112 (2001): 438–448.

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. “Induction of pluripotent stem cells from adult human fibroblasts by defined factors.” *Cell.* 131 (2007): 861-72.

Takahashi K, Yamanaka S. “Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.” *Cell* 126.4 (2006): 663-76.

Takata M, Sasaki MS, Sonoda E, Morrison C, Hashimoto M, Utsumi H, Yamaguchi-Iwai Y, Shinohara A, Takeda S. “Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells.” *The EMBO Journal.* 17.18 (1998): 5497-508.

Takayama N, Nishikii H, Usui J, Tsukui H, Sawaguchi A, Hiroyama T, Eto K, Nakauchi H. “Generation of functional platelets from human embryonic stem cells in vitro via ES-sacs, VEGF-promoted structures that concentrate hematopoietic progenitors.” *Blood.* 111.11 (2008): 5298-306.

Tallack MR, Magor GW, Dartigues B, Sun L, Huang S, Fittock JM, Fry SV, Glazov EA, Bailey TL, Perkins AC. “Novel roles for KLF1 in erythropoiesis revealed by mRNA-seq.” *Genome Research* 22.12 (2012): 2385-98.

Tamir A, Howard J, Higgins RR, Li YJ, Berger L, Zacksenhaus E, Reis M, Ben-David Y. “ Fli-1, an Ets-related transcription factor, regulates erythropoietin-induced erythroid proliferation and differentiation: evidence for direct transcriptional repression of the Rb gene during differentiation.” *Molecular and Cellular Biology* 19.6 (1999): 4452-64.

Tavassoli M. “Embryonic and fetal hemopoiesis: An overview.” *Blood Cells* 18 (1991): 269-81.

Taylor CJ, Peacock S, Chaudhry AN, Bradley JA, Bolton EM. "Generating an iPSC bank for HLA-matched tissue transplantation based on known donor and recipient HLA types." *Cell Stem Cell* 11.2 (2012): 147-52.

Terskikh AV, Miyamoto T, Chang C, Diatchenko L, Weissman IL. “Gene expression analysis of purified hematopoietic stem cells and committed progenitors.” *Blood* 102.1 (2003): 94-101.

Thon JN, Montalvo A, Patel-Hett S, Devine MT, Richardson JL, Ehrlicher A, Larson MK, Hoffmeister K, Hartwig JH, Italiano JE Jr. “Cytoskeletal mechanics of proplatelet maturation and platelet release.” *The Journal of Cell Biology* 191.4 (2010): 861-74.

Thon JN, Schubert P, Devine DV. “Platelet Storage Lesion: a new understanding from a proteomic perspective.” *Transfusion Medicine.* 22.4 (2008): 268-79.

Thrasher AJ, Gaspar HB, Baum C, Modlich U, Schambach A, Candotti F, Otsu M, Sorrentino B, Scobie L, Cameron E, Blyth K, Neil J, Abina SH, Cavazzana-Calvo M, Fischer A. “Gene therapy: X-SCID transgene leukaemogenicity.” *Nature* 443.7109 (2006): E5-6.

Tijssen MR, Cvejic A, Joshi A, Hannah RL, Ferreira R, Forrai A, Bellissimo DC, Oram SH, Smethurst PA, Wilson NK, Wang X, Ottersbach K, Stemple DL, Green AR, Ouwehand WH, Göttgens B. “Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators.” *Developmental Cell.* 20.5 (2011): 597-609.

Trakarnsanga K, Griffiths RE, Wilson MC, Blair A, Satchwell TJ, Meinders M, Cogan N, Kupzig S, Kurita R, Nakamura Y, Toye AM, Anstee DJ, Frayne J. “An immortalized adult human erythroid line facilitates sustainable and scalable generation of functional red cells.” *Nature Communications* 14.8 (2017): 14750.

Treutlein B, Lee QY, Camp JG, Mall M, Koh W, Shariati SA, Sim S, Neff NF, Skotheim JM, Wernig M, Quake SR. “Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq.” *Nature* 534.7607 (2016): 391-5.

Van de Walle GR, Vanhoorelbeke K, Majer Z, Illyés E, Baert J, Pareyn I, Deckmyn H. “Two functional active conformations of the integrin alpha-2-beta-1, depending on activation condition and cell type.” *The Journal of Biological Chemistry.* 280.44 (2005): 36873-82.

Vander Heiden MG, Plas DR, Rathmell JC, Fox CJ, Harris MH, Thompson CB. “Growth Factors Can Influence Cell Growth and Survival through Effects on Glucose Metabolism.” *Molecular and Cellular Biology* 21.17 (2001): 5899-912.

Védy D, Robert D, Gasparini D, Canellini G, Waldvogel S, Tissot JD. “Bacterial contamination of platelet concentrates: pathogen detection and inactivation methods.” *Hematology Reviews* 1.1 (2008): e5.

Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, Hirche C, Lutz C, Buss EC, Nowak D, Boch T, Hofmann WK, Ho AD, Huber W, Trumpp A, Essers MA, Steinmetz LM. “Human haematopoietic stem cell lineage commitment is a continuous process.” *Nature Cell Biology* 19.4 (2017): 271-81.

Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Südhof TC, Wernig M. “Direct conversion of fibroblasts to functional neurons by defined factors.” *Nature* 463.7284 (2010): 1035-41.

Vigon I, Mornon JP, Cocault L, Mitjavila MT, Tambourin P, Gisselbrecht S, Souyri M. “Molecular cloning and characterization of MPL, the human homolog of the v-mpl oncogene: identification of a member of the hematopoietic growth factor receptor superfamily.” *Proceedings of the National Academy of Sciences of the United States of America* 89 (1992): 5640-44.

Waugh RE, Mantalaris A, Bauserman RG, Hwang WC, Wu, JH. "Membrane instability in late-stage erythropoiesis." *Blood* 96.6 (2001): 1869-75.

Weiss MJ, Keller G, Orkin SH. “Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells.” *Genes & Development* 8.10 (1994): 1184-97.

White JG, Krivit W. “An ultrastructural basis for the shape changes induced in platelets by chilling.” *Blood* 30.5 (1967): 625-35.

WHO. “Global Status Report on Blood Safety and Availability 2016.” (2016).

WHO. “Guidelines for the treatment of malaria. 3rd edition.” (2015).

Wills QF, Livak KJ, Tipping AJ, Enver T, Goldson AJ, Sexton DW, Holmes C. “Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments.” *Nature Biotechnology* 31.8 (2013): 748-52.

Wilson NK, Foster SD, Wang X, Knezevic K, Schütte J, Kaimakis P, Chilarska PM, Kinston S, Ouwehand WH, Dzierzak E, Pimanda JE, de Bruijn MF, Göttgens B. “Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators.” *Cell Stem Cell.* 7.4 (2010): 532-44.

Wilson NK, Kent DG, Buettner F, Shehata M, Macaulay IC, Calero-Nieto FJ, Sánchez Castillo M, Oedekoven CA, Diamanti E, Schulte R, Ponting CP, Voet T, Caldas C, Stingl J, Green AR, Theis FJ, Göttgens B. “Combined single-cell functional and gene expression analysis resolves heterogeneity within stem cell populations.” *Cell Stem Cell* 16.6 (2015): 712-24.

Wilson NK, Kent DG, Buettner F, Shehata M, Macaulay IC, Calero-Nieto FJ, Sánchez Castillo M, Oedekoven CA, Diamanti E, Schulte R, Ponting CP, Voet T, Caldas C, Stingl J, Green AR, Theis FJ, Göttgens B. “Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations.” *Cell Stem Cell* 16.6 (2015): 712-24.

Wright DE, Wagers AJ, Gulati AP, Johnson FL, Weissman IL. “Physiological migration of hematopoietic stem and progenitor cells.” *Science* 294 (2001): 1933–1936.

Wu H, Liu X, Jaenisch R, Lodish HF. “Generation of committed erythroid BFU-E and CFU-E progenitors does not require erythropoietin or the erythropoietin receptor.” *Cell* 83.1 (1995): 59-67.

Wu TD, Nacu S. “Fast and SNP-tolerant detection of complex variants and splicing in short reads.” *Bioinformatics* 26 (2010): 873-81.

Xia X, Zhang Y, Zieth CR, Zhang SC. “Transgenes Delivered by Lentiviral Vector Are Suppressed in Human Embryonic Stem Cells in a Promoter-Dependent Manner.” *Stem Cells and Development* 16.1 (2007): 167-76.

Yamada E. “The fine structure of the megakaryocyte in the mouse spleen.” *Acta Anatomica* 29.3 (1957): 267-90.

Yamane A, Nakamura T, Suzuki H, Ito M, Ohnishi Y, Ikeda Y, Miyakawa Y. “Interferon-2b–induced thrombocytopenia is caused by inhibition of platelet production but not proliferation and endomitosis in human megakaryocytes.” *Blood* 112.3 (2008): 542-50.

Yang CT, Ma R, Axton RA, Jackson M, Taylor AH, Fidanza A, Marenah L, Frayne J, Mountford JC, Forrester LM. “Activation of KLF1 Enhances the Differentiation and Maturation of Red Blood Cells from Human Pluripotent Stem Cells.” *Stem Cells* 35.4 (2017): 886-897.

Yiangou L, Montandon R, Modrzynska K, Rosen B, Bushell W, Hale C, Billker O, Rayner JC, Pance A. “A Stem Cell Strategy Identifies Glycophorin C as a Major Erythrocyte Receptor for the Rodent Malaria Parasite Plasmodium berghei.” *PLoS One.* 11.6 (2016): e0158238.

Yu J, Hu K, Smuga-Otto K, Tian S, Stewart R, Slukvin II, Thomson JA. “Human Induced Pluripotent Stem Cells Free of Vector and Transgene Sequences.” *Science* 324.5928 (2009): 797-801.

Zauli G, Vitale M, Falcieri E, Gibellini D, Bassini A, Celeghini C, Columbaro M, Capitani S. “ In vitro senescence and apoptotic cell death of human megakaryocytes.” *Blood* 90.6 (1997): 2234-43.

Zhang D, Lee HF, Pettit SC, Zaro JL, Huang N, Shen WC. “Characterization of transferrin receptor-mediated endocytosis and cellular iron delivery of recombinant human serum transferrin from rice (Oryza sativa L.).” *BMC Biotechnology* 30.12 (2012): 92.

Zhang F, Frost AR, Blundell MP, Bales O, Antoniou MN, Thrasher AJ. “A ubiquitous chromatin opening element (UCOE) confers resistance to DNA methylation-mediated silencing of lentiviral vectors.” *Molecular Therapy* 18.9 (2010): 1640-9.

Zimmet J, Ravid K. “Polyploidy: occurrence in nature, mechanisms, and significance for the megakaryocyte-platelet system.” *Experimental Hematology* 28.1 (2000): 3-16.

Zsebo KM, Williams DA, Geissler EN, Broudy VC, Martin FH, Atkins HL, Hsu RY, Birkett NC, Okino KH, Murdock DC, et al. “Stem cell factor is encoded at the Sl locus of the mouse and is the ligand for the c-kit tyrosine kinase receptor.” *Nature* 63.1 (1990): 213-24.

Zwaka TP, Thomson JA. “Homologous recombination in human embryonic stem cells.” *Nature Biotechnology* 21 (2003): 319-21.

Acknowledgements

I will start by thanking my wonderful supervisors and mentors, Cedric and Thomas. They have both given me so much of their time over the last few years, for which I am extremely grateful and they have shaped me into the scientist I am today. Writing my thesis has been the biggest challenge for me during my PhD- (even more so than getting the inducible PSC line to work!), I struggled to find a concise and clear writing style. I am so thankful for their patience, advice and help. I have learnt an awful lot from them and know that the advice and knowledge they have shared will help me in my future career. More than that, they have helped me produce a thesis I feel very proud of.

I would like to thank all the members of the Ghevaert lab, past and present, for all their encouragement, help, knowledge and support over the last 4 years. I feel lucky to have worked with such wonderful people during my PhD and will treasure the memories of my time in Cambridge because of them. The work that I have achieved would not have been possible without the help and guidance of my lab mates, who are all so hard working and generous. Thank you all so much. Special thanks to the PSC sub-group of the lab; Thomas, Amanda, Annett and Nilly, who have worked closely with me and helped me out on so many occasions. Thomas, you have been the best teacher, so enthusiastic and encouraging. I cannot thank you enough for all that you have done and shared. Amanda, you have always been so supportive and provided a listening ear when things have been tough and I've needed to rant. Annett, thank you for always making me see what really matters and saying things straight, your perspective on things always makes me feel happier. Thanks to Wing, who showed me my first PCR reaction, and forgot to add polymerase! You were so generous with your time when I was new to the lab and knew so little. To other members of the lab; Holly, Momal, Cav, Dan, Moyra, Amy, Nina, Maria, Meera, Guénaëlle, Harriet and Catherine, thank you for the wonderful company, I am lucky to have had you as colleagues.

Thank you to Alena, who has been a wonderful mentor. You have always looked out for me and encouraged me and got me through some difficult times. You taught me not to give up and I'm so glad I didn't, it has been so worth it in the end. Thank you for reading parts of my thesis and for your comments. Thank you to Marloes, Pepe, Chiara and Nicola. My work became entwined with yours by accident but it has led to a great collaboration on a very interesting story that has followed me since my Masters rotation. Thank you all for your hard work in helping to put together the erythroblast manuscript and making up some of the figures for my first chapter. Thank you to James, who achieved so much in a short rotation project and skilfully performed the analysis for my scRNA-seq experiment. Thank you for being so forthcoming with your time and knowledge, my final chapter would have been a lot harder to write without it!

Away from the lab I am lucky to have the support of so many wonderful people. I would like to thank my boyfriend, Matt. You have been my rock, helping me through the bad times and being the best company to enjoy the good times with. Thank you for always making me smile, for always teaching me new things and for supporting me no matter what. Thanks for coining the phrase “Think of the Slink”, it’s gotten me through a lot! You mean the world to me. Thank you Slinky, for always having a waggy tail, we could all learn a lot from you. Thank you to my wonderful, supportive family, and Matt’s. Thank you for having the belief in me that the work I do will make a difference (despite not really understanding it!) Mum, I don’t know what I’d do without you, I cannot put it into words. Simply, thank you for everything. Thank you to my closest friends, Ellie, Paneet and Fiona. Despite not being able to see each other as often as we’d like (we’re busy pretending to be grown-ups), whenever we’re together it feels like no time has passed at all since we were in our blue blazers at LHS or dancing the night away at Vice.

I never thought I’d be able to write a book, but here it is. Thank you to everyone that helped along the way. Amanda.