# Characterisation of MESP1 binding partners and their use in the production of cardiac progenitor cells

Kendall, Sarah

*Awarding institution:*
King's College London

# Characterisation of MESP1 binding partners and their use in the production of cardiac progenitor cells

Sarah Leonorah Kendall

Thesis submitted for the degree of Doctor of Philosophy

Randall Division of Cell and Molecular Biophysics

King's College London

October 2020

# 1 Acknowledgements

First and foremost, I would like to thank my supervisor Dr Fiona Wardle, for all of her guidance, understanding and stem cell commiserations on this journey. Your advice has been very much valued. I would also like to thank my second supervisor, Professor Georgina Ellison–Hughes, for stepping in and allowing me to steal her lab equipment and tissue culture room when ours was being shaken by building works.

A huge thank you to my thesis committee: Dr Baljinder Mankoo, Dr Elisabeth Ehler and Dr Attila Csikasz-Nagy for your direction and humour during this time, and support during the rougher times.

In addition, I would like to thank the British Heart Foundation for funding this PhD and the yearly conferences that have helped to support my scientific journey.

A delayed thank you to my first ever lab supervisor Dr Emily Mercer. It was because of you that I caught the science bug, and this all began.

I would also like to thank my fantastic lab mates, past and present, and the wider office. The laughs, summer Pimm's on the grass and sneaking off for coffee, helped make all the late lab nights and long weekends bearable.

A big thank you to all my non-lab friends for putting up with me complaining about strange scientific concepts, talking absolute rubbish for four years and providing plenty of laughs.

Mike, thank you for not ever complaining about how many times you had to help me move to a new house, build furniture and the lethal cocktails. To amazing aunt and uncle, AG and UN, who have supported me throughout my time at university (and before!). Thank you for dealing with the annoying phone calls and coming to visit often, it was very appreciated. A further thanks to Drewe and Felicity for helping to boost my spirits throughout my time in academia. A big thank you to the rest of my family, of whom there are too many to mention, for learning to stop asking when I will finally graduate and what I want to do when I finish. You are all amazing.

Adam, thank you for keeping me sane, for dealing with me discussing science every 5 minutes, disappearing to lab constantly and your unwavering support. Couldn't have done this without you.

And last, but certainly not least, to my amazing Mum. Thank you for bringing me up to be strong, independent and far too inquisitive, which has come in handy for this PhD! For your endless help and lots of laughs, this one is for you!

# 2 Abstract

The transcription factor, *Mesoderm posterior 1* (*Mesp1*), is essential for mesoderm formation in the early embryo. Overexpression in an embryonic stem cell (ESC) model leads to cardiac differentiation. However dependent upon conditions, *Mesp1* overexpression can also induce skeletal muscle and haematopoietic lineages. The mechanism by which *Mesp1* induces these different phenotypes is unknown but could be due to different interacting partners. Identifying if such interactions play a role in the differentiation of stem cells into cardiac progenitor cells (CPCs), as well as if we can use these factors to reprogram somatic cells into CPCs is the pivotal question to be answered in order to provide a novel CPC production pathway. The ability to produce a pure CPC population would allow us to make a breakthrough in their use in regenerative medicine, opening up new avenues of treatment for heart failure patients.

To investigate this, we identified MESP1 and its binding partners through mass spectrometry of mouse embryonic stem cells (mESC) isolated at mesoderm stage of cardiac differentiation. Our analysis has identified 7 putative binding partners that have similar expression patterns to *Mesp1* in mESCs undergoing cardiac differentiation. Furthermore, we have verified that 2 of these putative binding partners, Eomesodermin and Methyl- CpG-binding domain protein 3, interact with MESP1 when co-expressed in human embryonic kidney cells (HEK). Furthermore, using literature-based evidence, we have identified 1 other potential binding partner, WD Repeat Domain 5. Utilising mESCs, we have identified that these binding partners have an effect on cardiac differentiation, and other binding partner expression patterns, that previously observed in *Mesp1* knockout mESCs. We further ascertained the effects of these binding partners in the reprogramming of cells. Overexpression in 3T3 fibroblasts has identified morphological and genetic changes, including inducing the expression of cardiac genes.

# 3 Table of Contents

# 4 Table of Figures

# 5 Table of Tables

# 6 List of abbreviations

ACEi – Angiotensin II converting enzyme inhibitors

Amot - Angiomotin

ARA – Angiotensin II receptor antagonists

bHLH – basic helix loop helix

BMC – bone marrow cell

BM-MESC – bone marrow mouse embryonic stem cell

BMP4 – bone morphogenic protein 4

ChIP – chromatin immunoprecipitation

COPD- chronic obstructive pulmonary disease

CPC – cardiopoietic cells

CRE - cAMP- responsive element

Creb1 – cAMP- responsive element binding protein 1

CVD – cardiovascular disease

EB – embryoid body

EMT – epithelial to mesenchymal transition

Eomes - Eomesodermin

EPI-ESC – epiblast embryonic stem cells

Etv2 – ETS variant 2

GHMT - Gata4, Hand2, Mef2c, Tbx5

GMT – Gata4, Mef2c, Tbx5

HEK – human embryonic kidney cell

hESC – human embryonic stem cell

hFGFb – human fibroblast growth factor b

hFGF10 – human fibroblast growth factor 10

iPSC – induced pluripotent stem cell

Linc – large intergenic non-coding RNA

Mass spec – mass spectrometry

MCP – multipotent cardiac progenitor

MEF – mouse embryonic fibroblasts

mESC – mouse embryonic stem cell

Myh6 – myosin heavy chain 6

Myl7 – myosin light chain 7

NT-ESC – nuclear transfer embryonic stem cell

OD – optical density

OKSM – Oct4, Klf2, Sox2, c-Myc

PS - primitive streak

RIME - rapid immunoprecipitation mass spectrometry

siRNA – small interfering RNA

SNP – single nucleotide polymorphisms

SRF – serum response factor

T - Brachyury

VEGF- vascular endothelial growth factor

Wdr5 – WD repeat domain 5

# 1 Chapter 1

## 1.1 Introduction

### 1.1.1 Heart disease is a global threat to life

Cardiovascular disease (CVD) is the most prevalent non-communicable disease worldwide, leading to approximately 18 million deaths in 2017 [1]. CVD is thought to affect as many as 40.5% of the US population [2] and 20% of the UK population [3].

An increasing global population and change in demographics, with an upwards shift in the proportion of the population over the age of 65, underlies a predicted 10% increase in the prevalence of CVD by 2030 [2]. Due to the increased prevalence, the cost of treating CVD is expected to increase by 200% in the next 20 years [4].

CVD is an umbrella term, covering a range of diseases that affect the heart and circulatory system. These can be broadly subdivided into; those that affect cardiogenesis and its morphology, which affect roughly 1 in every 150 live births a year [5], and adult cardiomyopathies. Whilst adult cardiomyopathies can arise from the inheritance of a single gene variant, most are assumed to develop as a result of a combination of detrimental factors, including life style choices or the development of a single nucleotide polymorphisms [6–9]. Whilst all cardiomyopathies affect a single organ, the heart, they each have different aetiologies underlying their disease pathology.

### 1.1.2 The different forms of heart disease

The predominant form of cardiovascular disease is coronary heart disease (CHD), otherwise referred to as coronary artery disease [5]. This is the result of the accumulation of atherosclerotic plaques, made up of a collection of lipid molecules, leukocytes and inflammatory mediators, within the endothelial lining of medium arteries [10]. This narrowing of the arteries results in the formation of two types of lesions: stenotic and non-stenotic. The development of a stenotic lesion results in remodelling of the artery, with a small lipid core and a thick fibrous cap, that protrudes into the lumen of the artery, restricting blood flow. This typically presents with symptoms of chest pain, clinically identified as angina pectoris, with a reduction in oxygenated blood flow to the heart [11]. The more common form of arterial lesion is non-stenotic lesions, universally known for their role in myocardial infarctions. These are lipid rich, with a thinner fibrous cap, that are capable of undergoing thrombosis [11]. Arteries containing these lesions often compensate through

enlargement of the vessel. The rupturing of these plaques or thromboses can result in myocardial infarction or stroke, where ruptured material blocks myocardial or brain arteries.

A major contributing risk factor in the formation of stenotic lesions is hypertension. Affecting over a quarter of the UK's adult population, the presence of hypertension raises the risk of cardiovascular events, including stroke, myocardial infarction and heart failure [5]. Another linked comorbidity is diabetes, with prolonged exposure to increased blood sugar destructive to the vessel endothelium. Nine in ten diabetics have *diabetes mellitus,* more commonly known as type-two diabetes, which is a glucose dysregulation linked to obesity, smoking and lifestyle choice, rather than a failure in insulin production by the pancreas. Those with diabetes are known to be at a 2 to 3-fold higher risk of contracting a heart or circulatory disease. One third of those with diabetes die of CVD [5].

Additional risk factors for CVD include drug use, high cholesterol, lack of exercise and exposure to air pollution. Some therapeutic agents, including cancer treatments, often induce cardiotoxicity as a side effect, leading to CVD being the leading cause of death of cancer patients in remission, other than secondary malignancies [12].

### 1.1.3 The genetic component of CVD

Whilst lifestyle choices play a role in CVD development, there is a genetic component to many cardiovascular conditions. Approximately 620,000 adults in the UK possess a genetic mutation that increases their risk of CVD [5]. This involves conditions such as familial hypercholesterolaemia which affects 1 in 250 [5,13]. Although drastically underdiagnosed, the condition is defined by high levels of lipids within the blood, which increase the risk of lipid deposition in the vessel walls [14].

The major class of genetic inherited CHD are cardiomyopathies, which are comprised of three main categories; hypertrophic, dilated and restrictive [15,16]. Hypertrophic cardiomyopathies are characterised by the thickening of the cardiac muscle in the absence of increased loading [16]. In contrast, those affected with dilated cardiomyopathy present with a phenotype of a dilated left ventricle, in addition to myocyte death and cardiac fibrosis [15]. Finally, restrictive cardiomyopathies are characterised by impairment of ventricular filling, usually due to a stiffening of the ventricular walls, and are typically less prevalent than other types of cardiomyopathies [17].

In addition to these inherited conditions observed in adults, there are further mutations detected in the embryo and new-born children. Congenital heart defects affect cardiac

morphogenesis, and are prevalent in one in every 150 live births in the UK [5]. Whilst genetics plays a role in the dysfunctional development of the cardiac system, it is also important to note that maternal infections and substance abuse can also induce these developmental defects [18]. These conditions can be classified as cyanotic or acyanotic. Acyanotic defects include obstructive lesions, incorporating vessel stenosis and septal defects, whilst cyanotic defects result in reduced peripheral blood flow, and include conditions such as Tetralogy of Fallot and Transposition of the Great Arteries [18]. Whilst some of these are reversable by surgical intervention, others result in the lifelong need for therapeutics. The lifespan for those born with congenital heart disorders is steadily increasing, and was approximately 57 years in 2007 [19]. Despite great advances in the care of these patients, half of adults born with a congenital heart defect will die from heart failure [19].

### 1.1.4 A summary of heart failure

Heart failure can be defined as the inability of the heart to fill ventricularly or eject blood to the body [20]. The condition can be identified as chronic or acute, depending on the timeline of onset. Approximately 920,000 people in the UK are defined as living with heart failure [5]. Despite affecting a relatively low proportion of the UK population, the cost of treating heart failure to the NHS is over £2 billion a year, and accounts for over 5% of all emergency medical admissions to hospitals [21]. The main aetiologies for developing heart failure are ischaemic, hypertensive and rheumatic heart disease, as well as chronic obstructive pulmonary disease (COPD)[4]. Ischemic heart disease and COPD disproportionately affect those in high-income countries, whilst low-income areas are more likely to be affected by the other factors, and myocarditis [4].

Regardless of aetiology, the treatments for heart failure remain broadly similar, with the aim of reducing pressure on the heart whilst maintaining good circulation.

### 1.1.5 Current therapeutics for heart failure

Treatments for heart failure target several different mechanisms in order to reduce the pressure, or pre-load, on the heart through a reduction in blood pressure. Blood pressure can be modulated through traditional methods, including diuretics, salt restriction in the diet to prevent water retention [20], and Angiotensin-converting enzyme inhibitors (ACEi). These

methods are particularly useful in congestive heart failure. Nitrates are suggested in patients unresponsive to these methods [20].

Other treatments provide an additive effect to these therapeutics, including digoxin to improve cardiac output, anticoagulants to prevent thrombosis, and ionotropic agents to reduce congestion and improve system perfusion [20].

Although these therapeutic interventions provide a relief to some of the symptoms of heart failure, they are not effective without complementary lifestyle changes. A major strain on the heart is smoking, which causes a reduction in lung capacity and increases in heart rate and blood pressure [22]. Without cessation of smoking, in collaboration with therapeutic agents, there is often a progression in disease and a decrease in prognosis. Similarly, high-salt diets are also detrimental to heart failure treatments, with high sodium intake linked to fluid retention in heart failure patients, resulting in swelling and symptoms of breathlessness [23].

In spite of this battery of treatments, some patients may ultimately need surgical interventions to manage their conditions. This can include the implantation of a cardioverter-defibrillator or a ventricular assist device. These allow shocks to the heart to prevent sudden death in the case of arrythmias and those who have survived a cardiac arrest [24]. Although preventing sudden death, it can cause a progression to haemodynamic failure and a worsening of left ventricular function, and ultimately a quicker progression into the end stages of heart failure [24]. Whilst demonstrating an improvement to mortality rates, it can lead to a worsening of patient quality of life, with increased hospitalisations.

The use of left ventricular assist devices is seen as a bridge to transplantation. These devices provide mechanical assistance to the heart by pumping the blood directly from the lungs into the aorta [25]. In addition to improved survival rates both prior and post-transplantation, there is also an improvement to patient quality of life [24]. These devices are available for those who are not suitable for transplant, with the most adverse risk being device failure. However, these devices do not substitute for a heart transplant.

Heart transplantation is the final, and only, treatment for those in end-stage heart failure. This option is limited by the number of hearts available on transplant lists, as well as the suitability of the patient for transplantation. Guidelines prevent those with current substance abuse, lack of compliance with treatments including dietary and lifestyle, and uncontrollable mental disease from receiving a transplant [24]. Those with a comorbidity, including renal or hepatic failure, or less than 5 years after the remission of cancer, are also counted as unsuitable candidates for transplant [24]. Post-transplantation patients are at risk of allograft rejection as well as complications of immunosuppression, including renal failure,

hypertension and malignancies [24]. Prognosis post-transplantation reveals a 1 to 5-year survival rate of 80%, which is reduced to 50% at 10 years [24,25].

No single treatment provides a quality of life equivalent to those without heart failure. Although surgical options allow for preservation of life in the short-term, there is no long-term option for those unsuitable for cardiac transplantation which, in itself, is not a life-long solution, with the need for further interventions. Severe renal dysfunction is subsequently observed in 30% of patients [26].

The number of donor hearts remains stagnant, whilst the number of those with heart failure is increasing with an aging population and obesity crisis [20]. Therefore, new treatments for heart disease must be devised to treat and cure heart failure.

### 1.1.6 New technologies to treat heart failure

In addition to conventional therapies, new technologies are emerging with regulatory approval for the treatment of heart failure. One such device is the HeartMate3, which has a similar function to left ventricular assist devices, although is less invasive to fit, preventing the need for open heart surgery [27]. Furthermore, its increased lifespan prevents the need for replacement within 2 years of insertion, and reduces the risk of stroke, bleeding or gastrointestinal haemorrhage [28].

Other recently approved devices include an upgraded cardiac contractility modulation device, which enables the improvement of the left ventricular contractile strength, whilst also improving biological markers associated with heart failure [29].

New biologics are also being developed. These are based on the need to regenerate areas of the heart that have been damaged, preventing full mechanical function. These ideas revolve around the theory that a patient's cells could be utilised to propagate a new population of cardiac cells, either through stem cell differentiation or somatic cell reprogramming.

### 1.1.7 Stem cells

Embryonic stem cells are an undifferentiated, pluripotent population of self-renewing cells isolated from the inner cell mass of the blastocyst [30], that have the potential to differentiate into any cell type of the organism. These cells typically express markers including *Nanog* and *Oct4.* Studies have used both mouse and human embryonic stem cells (mESCs, hESCs), both derived from the inner cell mass of the blastocyst [31]. ESCs have the

capacity to form all three germ layers; endoderm, mesoderm and ectoderm, from which cardiomyocytes emerge from the mesodermal layer [31].

There are also several types of adult stem cells within the body after development, although these are specialised to allow replacement of damaged cells and growth, as well as to maintain tissue homeostasis [30]. These cells are capable of self-renewal and are multipotent, giving rise to tissues belonging from one embryonic germ layer [32]. More specified stem cells include neural, hematopoietic and skin, which can only form cells of those lineages [30].

Embryonic stem cells can be directed towards certain lineages through modulation of the cell culture conditions. Culture conditions of cardiac differentiation have been well-described in both mESC and hESCs [33–40]. These ESCs progress from a pluripotent stem cell state to a mesodermal lineage, often through embryoid body formation, before differentiating into cardiac progenitor cells, and finally into fully committed cardiomyocytes [37].

These differentiated ESCs have been shown to form contractile cardiac structures, including sarcomeres, and express genes known to be transcribed in contracting cardiac cells, such as myosin heavy chain, *Nkx2.5* and *MEF2C* [31,39,41].

Several studies have reported positive results when implanting these cells into failing hearts. One study, performed in a sheep myocardial infarction model, demonstrated that mESCs engrafted into the myocardium can become functional cardiomyocytes that integrate into the infarct zone, as well as being functionally beneficial, as determined by an increased left ventricular ejection fraction [42]. In addition to infarct models, injection of mESCs committed to a cardiac fate (ESC-CM), with a cardiac ventricular phenotype, have also been injected into mouse hearts injured through doxycycline toxicity. In agreement with previous studies, ESC-CM injection resulted in increased ejection fractions and stroke volume, as well as a reduction in the number of apoptotic cardiomyocytes in the heart [43].

Several risks underlie the use of hESCs including allogenic effects when grafting these cells, with a risk of an immune response and transplant rejection [44,45].

*In vivo* clinical trials have occurred in the case of adult bone marrow stem cells, where a patient's adult stem cells are taken from the bone marrow (BMC) and implanted into the heart. A review of 50 such trials between 2003 and 2011 demonstrates an average increase of 3.96% in LVEF, and a reduction in infarct area by 4.03% [46]. The outcome of long-term studies reveals variable results in the longevity of this improvement.

This was true of the BOOST and Repair AMI trial, that found no improvement in LVEF in comparison to control patients [47,48]. By contrast, the TOPCARE-Ami trial, which utilised the same technique in 59 patients, showed an 11% increase in LVEF after 5 years [49].

The C-Cure study reprogrammed BM-MSCs to cardiopoietic cells (CPCs), as noted by the translocation of MEF2C to the nucleus and delivered them to the endomyocardium of 48 patients. Six months after treatment there was a 7% increase in LVEF in comparison to the standard care group [50].

The conflicting results shown in the clinicals are not unsurprising. Many studies on endogenous cardiac stem cells, marked by C-KIT or SCA-1 expression, have been published in the last 20 years. These non-clinical studies spawned the use of bone marrow stem cells expressing C-KIT in clinical trials for the use of regenerating the heart[22]. However, the majority of the founding studies that these trials were based on have been recently retracted[22].

Studies have shown that SCA-1 positive cells contribute to the heart vasculature rather than to form cardiomyocytes[23]. Indeed, bone marrow cells expressing C-KIT also do not convert to cardiomyocytes when implanted into the heart[22].

A low retention rate is observed in bone marrow stem cells implanted into the heart, as well as a low level of new cardiomyocytes formed [24]. The positive results witnessed in clinical studies delivering bone marrow stem cells into patients could be due to the secretion of paracrine factors from implanted cells into the injured myocardium.

Recent studies have delineated that the adult heart renews approximately 1% a year at 20 years of age, declining to 0.5% by the age of 50, mostly through mitosis [23,24]. This indicates that the heart does not have a population of endogenous stem cells capable of repairing large scale ischemic injury.

In contrast pluripotent stem cells engrafted into the damaged myocardium have been shown to secrete a range of cytokines, growth factors and chemokines that could induce cardiac repair, with noticeable increase in vascular endothelial growth factors and basic fibroblast growth factor [54]. Both of these factors are widely used in protocols, inducing the differentiation of ESCs to a cardiac fate. These paracrine factors may provide neovascularisation, as well as the ability to modulate post-infarction inflammation and cardiac contractility, not only in the engrafted cells, but also in the endogenous cells of the heart [54].

Despite these improvements in cardiac function by the implantation of pluripotent stem cells into the heart, further progress must be made to enable greater regeneration and function. Therefore, it is important to identify a new population of cells that are able to generate a positive cardiac outcome.

### 1.1.8 Somatic reprogramming

A new avenue of exploration is the use of somatic cells. Somatic cells have a defined lineage and are unable to form other cell fates. Over 50 years ago, it was established that a nucleus from somatic cells can be transferred into enucleated oocytes to create cells with the same genotype. However, it was not until 2013 that this technique was used in the creation of human nuclear transfer-ESCs (NT-ESCs), that match the donor somatic cell DNA [56]. This technique, in theory, allows for the creation of pluripotent stem cells in patients that have declining number of endogenous multipotent stem cells, that may also have additional age-related impairments. Whilst ground-breaking in nature, this still relied on the donation of oocytes from donors and is restricted to research rather than therapeutic purposes. This technique is currently used in UK laboratories to create human ESCs[57].

The discovery of the 'Yamanaka factors' in 2006 allowed patient somatic cells to be reprogrammed back to a pluripotent state without the need for donor cell material [58]. Overexpression of *Oct4, Klf2, Sox2* and *c-Myc* (OKSM) in adult human dermal fibroblasts led to a conversion rate of approximately 0.02%, with subsequent induced pluripotent cells (iPSC) capable of forming all cell lineages [58,59]. Alternative protocols have since been suggested to avoid the use of *c-Myc,* which is oncogenic, and include *Oct3/4, Sox2*, *Nanog* and *Lin* [60,61]. To therapeutically utilise this method, further advances have enabled the non-viral delivery of these factors, in methods that do not cause genomic instability through integration into the genome [62].

Regardless of the mechanism employed to generate them, iPSCs can be successfully directed towards cardiomyocyte lineages through existing cardiac differentiation protocols, successfully forming all cardiac lineages, including atrial, ventricular and nodal-like cells [63]. However, the resultant populations are heterogeneous, and whilst mimicking the expected features and molecular profiles of cardiomyocyte differentiation, the cardiomyocytes are relatively immature in phenotype [63]. These cells are small, with ultra-structures resembling foetal cardiomyocytes, and changes in electrophysiological properties, including calcium handling [64]. This has driven research into methods that allow the

maturation of these iPSC derived cardiomyocytes, through the use of small molecules, 3-D growth methods and manipulation of the environment [65].

This ground-breaking research has spawned new efforts to eliminate the need to proceed through a pluripotent stage and allow direct reprogramming of somatic cells. Successful direct reprogramming would allow the conversion of cardiac fibroblasts in the heart into functioning myocardium [66].

The theory of direct reprogramming was first proven with muscle cells, where overexpression of *Myod*, known as the master regulator of muscle cell fate, was sufficient to convert fibroblasts to muscle myoblasts [67]. This has proven more difficult for cardiac cells, where to date, no master regulator has been discovered.

The pivotal study that successfully transdifferentiated fibroblasts to functional cardiomyocytes was described by Ieda et al in 2010 [68]. Murine fibroblasts formed beating cardiomyocytes 4 to 5 weeks after transduction with 3 genes, *Gata4, Mef2c* and *Tbx5* (*GMT*), with an efficiency of approximately 4-6% [68]. Another group showed the addition of *Hand2 (GHMT)* was able to improve the efficiency of this transdifferentiation to 6.8%[69]. Both groups showed delivery of transduced fibroblasts, or retroviruses containing *GHMT*, into a murine myocardial infarction model reduced scar size, with cardiac fibroblasts able to form cardiomyocyte-like cells [68,69].

Various other attempts to use cardiac transcription factors to reprogram fibroblasts have been made, including the addition of human *Myocd, Srf, Smarcd3,* and mouse *Mesp1*[70]. Optimisation of cell culture conditions also proved to be essential to enhancing the efficiency of transdifferentiation, with the use of valproic acid, a histone deacetylase inhibitor, and a JAK inhibitor shown to induce a pro-cardio effect [70]. The reduction of serum within the media has also been shown to aid organisation of the cardiac structure [70].

Another combination of transcription factors that are capable of reprogramming fibroblasts to a cardiac lineage are *Mesp1* and *Ets2*. The dual expression was capable of inducing *Nkx2.5* expression in human dermal fibroblasts, suggesting differentiation into a cardiac progenitor cell fate [71]. Expression of *Mesp1* in human dermal fibroblasts, in conjunction with other transcription factors, including the *GMT* factors and *Myocd*, enabled the expression of cardiac genes, with cells exhibiting calcium transients [72]. The addition of the transcription activator *Essrg*, and *Zfpm2*, a modulator of the GATA proteins, was sufficient to enable enhanced cardiac reprogramming and sarcomeric formation, alongside calcium transients and action potentials [73].

These results show it is possible to induce cardiomyocyte cells from a fibroblast population, in both human and murine models. However, as yet, they are not suitable for therapeutic use due to their heterogenic populations and the use of viral vectors for transduction. Nevertheless, they open new opportunities to optimise treatments specific to the patient, utilising their own somatic cells.

### 1.1.9 Disease in a dish

Disease simulation has, to date, been centred on broad data sets, finding the average benefit of treatments based on a specific patient cohort. This method of finding suitable treatments is particularly true of pharmaceutical trials, where trials are often limited in patient numbers due to safety concerns. Novel idiopathic reactions are often identified when new treatments receive market authorisation, whilst the beneficial effects of the treatment are not always achieved in all patients. Many therapeutics are not a 'one size fits all', with patients having to trial many treatments to find those that provide maximum efficacy with the least side effects. In cancer care, genetic sequencing has paved the way for new treatments, based on the genetic mutations that initiated the cancer formation itself. DNA sequencing of the tumour allows for directed treatment against that specific kind of cancer, for example, *HER2* positive cancers can now be treated with chemotherapy and antibody-based treatment, improving patient survival rates [74,75]. This success has not yet been realised in cardiac care where treatments are based on the categories of cardiovascular disease, with the aim to minimise symptoms.

The ability to reprogram a patient's somatic cells into cardiomyocytes (induced cardiomyocytes [iCM]), either directly or through a pluripotent intermediate step, provides new tools for modelling the disease in each patient. These iCMs contain all the genetic information of the patient, including any mutations [76]. To date, models of several cardiac syndromes have been possible in iPSC cells, including Long QT syndrome, familial dilated cardiomyopathy and arrhythmogenic cardiomyopathies [77]. iCMs can mirror the cardiac defects *in vitro* that are observed within a patient *in vivo.* To date, iCMs have been utilised to test new experimental drug treatments, with several showing a reversal of some disease characteristics [78]. The approach of using iCMs has been particularly useful in short QT syndrome (SQTS) patients, where pharmacological studies on cells derived from patients with different types of SQTS established that only one type of antiarrhythmic drug was effective. In addition, iCMs from patients with Timothy Syndrome have been used to trial the

effects of different therapeutic treatments to restore electrical and calcium signalling properties. These studies show the effectiveness of personalised medicine, whereby iCMs can be utilised for modelling drug therapies prior to a patient receiving a treatment, to increase the chance of it being efficacious [79].

iPSCs open new opportunities to model cardiomyopathies, where there are no animal models that recapitulate the phenotype observed in the patients, including many arrythmias. The modelling can also supplement traditional animal models and confirm the results are relatable to humans, with many physiological differences in the heart, including size, number of beats per minute and ion channel composition [80]. Due to the low efficiency of reprogramming to date, it is important to identify new methods and combinations of factors that improve reprogramming, prevent heterogeneous cell populations and allow more reproduceable results.

### 1.1.10 Cardiac differentiation in murine models

To enable more efficient reprogramming of somatic cells, it is essential to understand the mechanisms governing heart formation *in vivo.* The heart is the first organ to develop and function in the embryo. To enable development of the embryo, a process termed gastrulation occurs, using morphogenetic movements to allow the transition from the pre-gastrula embryo, to one with three germ layers: the endoderm, mesoderm and ectoderm. The process of gastrulation is thought to begin at E5.5 in the mouse, and arises through the enlistment of epiblast cells to the primitive streak, a transient structure in the early embryo, which is clearly discernible at E6.5 [81]. The embryo has established asymmetry at this stage, with the primitive streak found at the posterior side of the embryo, and can be denoted by *Wnt3* and *Brachyury (T*) expression at E6.0 [81,82]. Cells undergo epithelial to mesenchymal transition and ingress towards the basal side of the epiblast, invading the space between the epiblast and visceral endoderm. This forms a sheet that extends both sides of the primitive streak. The localisation of the cells within the primitive streak forecast the part of the foetus they contribute to. The first lineage that emerges from the primitive streak is the extraembryonic mesoderm, that gives rise to the mesoderm that forms the visceral yolk sac and amnion [82]. This arises from the posterior end of the primitive streak, due to the high *BMP4* signal found in the extraembryonic mesoderm that helps to pattern these cells [82]. Cells from the middle and anterior streak zones produce the lateral plate mesoderm, from which the heart subsequently forms.

Cells migrate from the anterior lateral region to form the cardiac crescent, or first heart field, at E7.5. These cells are cardiogenic at this stage, however, are not fully committed to this fate [83], with the expression of genes including *Nkx2.5* and *Gata4* [84]. These cells are multipotent and are capable of differentiating into endothelial, smooth muscle or myocardial cells, which can form atrial, ventricular and conduction cells [85]. The cells of the cardiac crescent then migrate to fuse at the midline, forming the linear heart tube. This primitive structure will go on to form the left ventricle and parts of the atria. This initial rudimentary structure forms first to allow nutrient supply to the growing embryo that would be inadequately served by diffusion.

In addition to cells of the first heart field, another group contributes to the formation of the heart. Cells found in the same region of the anterior lateral plate mesoderm go on to form the pharyngeal mesoderm and the second heart field, which can be identified through the expression of *Islet-1* [85,86].

Clonal analysis of doxycycline induced *Mesp1-rtTA/TetO-Cre/Rosa-Confetti* mouse hearts established that doxycycline administration at an early time point (E6.25 and E6.75) resulted in preferential labelling of the left ventricle, consistent with the *Mesp1*-derived first heart field, whilst at a later time point (E7.25) favoured labelling of the second heart field cells [87]. These cells are able to form several cell types, including cardiomyocytes, endothelial cells and smooth muscle cells.

In agreement, an additional study employed the mosaic analysis with double markers (MADM) system in mice, that relied on Cre-mediated recombination to restore fluorescent proteins, whilst daughter cells can be identified based on colour [88]. This study found that cells from a single clone populated distinct areas, for example, the left ventricle. However, none spanned both the left and right ventricle, suggesting *Mesp1* positive cells are already pre-programmed to a specific heart compartment at or before E6.25 [88].

The anterior part of the second heart field enables formation of the outflow tract and right ventricle, through the contribution of myocardial cells [85,86]. It further donates smooth muscle cells to both major vessels, the pulmonary artery and the aorta. The posterior second heart field, however, contributes cells to the walls of the atria and atrial septum [85]. It should also be noted that *Mesp1* positive cells from the pharyngeal mesoderm form the craniofacial muscles.

The heart undergoes further morphological changes post-cardiogenesis to enable the full function of the organ. These include cardiac looping and wedging to create an inflow and

outflow limb [85]. At the end of this series of events the heart is fully developed and resembles the expected cardiac morphology.

### 1.1.11 Role of *Mesp1* in cardiac differentiation

*Mesp1* expression, in the cells of the primitive streak, is one of the first genetic indicators that a cell may contribute to the formation of the heart. *Mesp1* is a basic helix loop helix (bHLH) transcription factor, transiently expressed between E6.5 and E7.5 in the early mouse embryo [89]. As a bHLH transcription factor, it can recognise and bind to the DNA [90] at canonical E-box motifs; CACGTG [91,92].

Labelling experiments tracking *Mesp1* positive cells have shown they contribute to all cardiac lineages and to almost all cells in the heart [91,93]. Lineage tracing has also shown *Mesp1* positive cells contribute to the craniofacial muscles, as well as cells that form the embryonic liver and haematopoietic stem cells [94–96]. *Mesp1* expression has also been transiently observed at E8.5 in the first somites [91].

The isolation and single cell RNA sequencing of cardiac progenitor cells isolated at E7.25 from *Mesp1-rtTA/tetO-H2B–GFP* mice, and subsequent SPRING analysis, identified five distinct destination cell types within the embryo for *Mesp1* positive cells [97]. These were comprised of endothelial or endocardial progenitors, marked by *Etv2*, cardiomyocyte progenitors marked by *Bmp4* and *Gata4*, and endodermal cells expressing *Sox17.* The final two destination cell types were the cranio-pharyngeal progenitors, denoted by expression of *Foxc-2* and *Tcf-21,* and that of the anterior secondary heart field, which expressed *Wnt2b* and *Tcf-21*[97]. This data confirmed the contribution of *Mesp1* positive cells to all cardiac lineages, as well as to the craniofacial muscles as described above.

*Mesp1* knockout is found to be embryonically lethal at E10.5, presenting with a phenotype of severe cardia bifida [98]. Cardiac cells were present but failed to migrate to the midline and form a linear heart tube [98]. This function of *Mesp1* is conserved throughout chordates, with knockdown of the *Mesp1 Ciona* ortholog, resulting in the failure of the formation of the heart tube in the juvenile [99].

The formation of cardiogenic cells without the presence of *Mesp1* indicates that another factor can compensate for its action during gastrulation, and the migration of mesoderm bilaterally, from the primitive streak. *Mesp2* is a homolog of *Mesp1* and is found 23kb away on the same chromosome in humans, and 16kb in mice [91]. The two homologs both contain a well conserved basic helix loop helix region, with *Mesp2* containing an extra 123 amino acid residues and a C-terminal degradation domain (Figure 1) [96]. *Mesp2* expression is limited before E8.5, with low expression in the primitive streak, after which it can be found highly expressed in the presomitic mesoderm [96]. *Mesp2*$^{-/-}$ mice do not show defects in cardiogenesis, instead presenting with a phenotype of defective somitogenesis and segmentation, affecting the formation of the embryo body [100].

The roles of these two homologs show overlap in the formation of the heart, with *Mesp1*$^{-/-}$ mice showing an upregulation of *Mesp2,* which is capable of initiating the migration of cells from the primitive streak and differentiating into cardiac and vascular cell types [101]. Ablation of both *Mesp1* and *Mesp2* negates this compensatory effect with embryonic lethality by E9.5 [100]. Defects in gastrulation are observed, with cells prevented from migrating from the primitive streak, which is thickened in these embryos [102]. Furthermore, there is no embryonic mesoderm except for minimal axial mesoderm that does not expand [102].

*Mesp2*$^{-/-}$ mice can be rescued through the insertion of four copies of the *Mesp1* gene into the *Mesp2* locus, suggesting *Mesp1* can also compensate for some mechanisms of *Mesp2* [98]. These experiments illustrate a redundant role for *Mesp1* and *Mesp2* in gastrulation and the migration of cells from the primitive streak. The later roles for both genes are essential. *Mesp1* expression allows for the migration of multipotent cardiac progenitors from the cardiac crescent to the midline, whilst *Mesp2* is necessary for functional somitogenesis and segmentation.

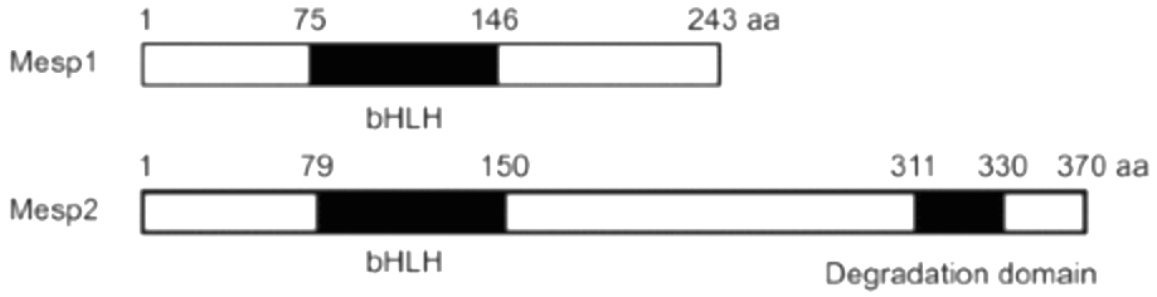



Figure 1: A schematic of the structure of MESP1 and MESP2 taken from Liang 2015

### 1.1.12 *Mesp1* overexpression can induce cardiovascular phenotypes

Embryonic stem cells have been utilised to mimic *in vivo* embryogenesis and can accurately depict the processes that occur in the early embryo, progressing from a pluripotent embryonic stem cell state to a primitive streak-like phase [103]. Through the addition of cytokines, these cells can be manipulated to a mesodermal lineage, which mimics the primitive streak stage *in vivo,* and are able to go on to form multipotent cardiac progenitors (MCP), and finally beating cardiomyocytes [104]. These stages can be monitored through the expression of stage-specific gene markers, with pluripotent cells identified through their *Oct4* and *Nanog* expression, whilst those that mimic cells of the primitive streak can be defined by expression of *Brachyury* (*T*) [81,105]. MCP stage cells express markers including *Nkx2.5*, denoting their commitment to a cardiac fate, whilst beating cardiomyocytes express cardiac structural genes, including *myosin heavy chain 6 (Myh6)*[106].

*Mesp1* expression during cardiomyocyte differentiation of mESCs is limited to a 24-hour time period, as found *in-vivo,* at mesoderm stage of differentiation. Dependent upon differentiation protocols, this typically occurs between day 3 and 4 of the protocol. *Mesp1* overexpression by doxycycline induction in mESC lines shows a large increase in the number of beating cardiac cells derived, with a corresponding 5-fold increase in cardiomyocyte gene expression [107–109]. The embryoid bodies of *Mesp1* induced mESCs were significantly larger during the first 48 hours post *Mesp1* induction, due to inhibition of apoptosis in comparison to wildtype embryos [110]. However, by 72 hours post-induction, wildtype and *Mesp1* induced cells had equivalent rates of growth. This slight advantage in growth rate did not correlate to the significant increase in beating cardiomyocytes, therefore, it is assumed that *Mesp1* expression induces a cardiac phenotype through inducing other gene expression rather than a selective mechanism [110].

The emergence of beating cells during cardiomyocyte differentiation from mESCs also occurs approximately 24 hours faster in cell populations where *Mesp1* is overexpressed [108]. This effect is limited to cell populations that have a transient expression of *Mesp1,* with constitutive expression in ESCs causing a decrease in the cardiogenic potential of the population, resulting in less cardiac cells derived [108].

In agreement with *in vivo* cardiogenesis, *Mesp1* positive cells are capable of forming all lineages of the heart, including endothelial and smooth muscle cells. This suggests that these *Mesp1* enriched populations can form cells of both the first and second heart fields [108].

Further studies have identified the need to inhibit *Wnt* signalling during the differentiation process to encourage mature cardiac phenotypes. This is evidenced by the addition of DKK1, an inhibitor of *Wnt* in the β-catenin dependent pathway, elevating levels of beating cells with more mature cardiac phenotypes [111].

This cardiogenic effect is mirrored in *Xenopus*, where overexpression of *Mesp1* at a 2-cell embryo stage was sufficient to cause the formation of regions that displayed ectopic beating, whilst also expressing *myosin light chain* [109]. This demonstrated that *Mesp1* expression was sufficient to cause the formation of ectopic cardiomyocytes. However, despite this cardiogenic promoting effect, *Mesp1* is not capable of transdifferentiating somatic cells to a cardiac phenotype [71].

More recently, studies have shown that dependent upon culture conditions, and the temporal overexpression of *Mesp1*, ESCs are capable of forming cardiac, skeletal muscle and hematopoietic lineages [112]. A reduction in media serum content, in coordination with overexpression of *Mesp1* for 5 days, starting at day 3 of culture, enables a skeletal muscle phenotype with the formation of myotubes [112]. Conversely, an earlier induction for 24 hours at day 2 of differentiation, allows for the establishment of hematopoietic progenitor cell populations, with an increase in markers such as *Tal1* [112]. *Mesp1* overexpression for 24 hours at day 3 gave the expected cardiac phenotype [112]. These experiments mimic *in vivo* labelling experiments identifying that *Mesp1* positive cells contribute to each of these 3 anlages [87,90,99], and the possibility of a *Mesp1* positive multipotent cell population *in vivo* capable of forming these 3 lineages [87,114,115].

### 1.1.13 *Mesp1*- Mechanism of action

To understand the mechanism by which *Mesp1* induces its cardiogenic actions, pre-conditioned media from *Mesp1* induced mESCs was supplemented into the media of wildtype mESCs. Wildtype mESCs grown with conditioned media showed no additional increase in the number of cardiac cells derived [108]. This indicates that *Mesp1* does not work through a paracrine mechanism, and in fact works cell autonomously.

It is also possible *Mesp1* is capable of inducing its own expression, with chromatin immunoprecipitation (ChIP) experiments demonstrating that *Mesp1* is able to bind to a region upstream of its own transcription initiation site [92,108].

Expression of *Mesp1* in the nucleus has an effect on a subset of genes, affecting the expression of 424 genes, 12 hours after overexpression, with 276 upregulated and 148

downregulated [108]. This equates to less than 1.5% of the genes present in the mouse genome. The rapidly upregulated genes included those that are involved in cardiogenesis, such as *Mef2c, Hand2* and *Myocardin* [108], whilst those that were downregulated included the endodermal marker *Foxd3*, and *Nkx6.3,* known to control gastric differentiation. Therefore, *Mesp1* appears to repress genes involved in endoderm cell fates[108]. In addition to these factors, *Mesp1* has also been shown to upregulate *Nkx2.5* and *Gata4,* which are thought to be the master regulators of cardiogenesis [108]. ChIP experiments have illustrated that MESP1 binds upstream of these cardiogenic factors (*Hand2, Gata4, Myocardin* and *Nkx2.5*) in some known enhancer regions, suggesting it is possible MESP1 induces expression of these genes, and its cardiogenic effect, through direct binding upstream [108].

MESP1 binding is also associated with changes in the epigenetic landscape of the cell, with enriched H3K27 acetylation in areas where MESP1 is bound, potentially detailing a role for MESP1 regulating the chromatin structure of targeted genes [92].

In addition to upregulating genes associated with cardiogenesis, MESP1 is also capable of modulating other genes involved in embryogenesis. ChIP experiments have identified that MESP1 binds upstream of genes involved in endoderm formation, including *Sox17* and *Goosecoid* [92,108]. RT-PCR analysis indicates these genes are downregulated as quickly as 12 hours post *Mesp1* expression, suggesting a role for *Mesp1* in the repression of other germ layer fates, including endodermal lineages [110].

*Mesp1* expression also results in the increased expression of *Ripply2*, a known repressor of *Mesp2*. MESP1 is seen to bind upstream of *Ripply2,* and subsequently represses *Mesp2* expression. Therefore, it is possible that *Mesp1* can not only autoregulate itself, but also indirectly regulate its homolog *Mesp2* [108].

### 1.1.14 Role of *Mesp1* in Cell Migration and Polarity

*In vivo, Mesp1* knockout murine models identify a role for *Mesp1* in facilitating the migration of cells from the cardiac crescent to the midline [89]. *Mesp1* overexpression in mESCs, show an increase in the cell migration speed in comparison to wildtype or *Mesp2* overexpression cells [116]. These cells were also found to be polarised towards the leading edge, and unidirectional, unlike wildtype or *Mesp2* overexpression cells [116]. Comparison of these cell types identified two differentially regulated genes: *Prickle1* and *RasGRP3*, with expression of both enriched in cells in which *Mesp1* was overexpressed[116].

MESP1 is capable of controlling cell migration speed through its interaction with genomic regulatory regions of *RasGRP3*, a Guanine Nucleotide Exchange Factor that activates the ERK signalling cascade through RAS activation [116]. MESP1 is capable of binding directly within the first intron of *RasGRP3* and is thought to cause an upregulation in expression. RASGRP3 can then in turn control cell migration speed, which is modulated through ERK signalling [116]. *Mesp1*$^{-/-}$ cells show a decrease in phospho-ERK in comparison to wildtype embryos, whilst inhibition of *ERK* signalling in *Mesp1* overexpressing mESCs eliminates the increased speed of migration previously observed [116].

In addition to controlling cell migratory speed, *Mesp1* is capable of regulating cell migration orientation. mESCs, in which *Mesp1* was overexpressed, showed unidirectional migration, with increased F-actin stress fibres on the leading edge [116]. These were not observed in wildtype or mESCs in which *Mesp2* was overexpressed, which changed their direction of migration multiple times. Furthermore, observation of the Golgi apparatus in these cells also showed that *Mesp1* overexpressing mESCs were polarised, with the Golgi found near to the leading edge of the cell, which was also not observed in the wildtype mESCs or those overexpressing *Mesp2* [116]. ChIP experiments have shown MESP1 to be capable of binding upstream of *Prickle1,* with expression of *Prickle1* increased in mESCs where *Mesp1* is overexpressed. Overexpression of *Prickle1* in mESCs was sufficient to ensure unidirectional cell migration, and was capable of rescuing the varied direction of migration in mESCs overexpressing *Mesp2* [116]. Knockout of *Prickle1* in *Mesp1* overexpressing cells ablated this unidirectional migratory effect [116], confirming *Mesp1* is capable of controlling cell orientation and migration direction through its modulation of *Prickle1*. Interestingly, *Prickle1* may also play a role in later-stage heart development, with mutations in the gene causing defects in later-stage second heart field development [117].

### 1.1.15 Role of *Mesp1* in epithelial to mesenchymal transition

The role of *Mesp1* in epithelial to mesenchymal transition (EMT) has been described in several studies [87,111,118]. *In vivo,* microarray analysis of doxycycline-induced *Mesp1-rTA/TetO-H2B-GFP* mouse embryos at E6.5 showed *Mesp1* positive progenitor cells are enriched in genes essential for EMT, including *Snail1*[87]. Overexpression of *Mesp1* in mESCs has been shown to promote induction of EMT, with the upregulation of mesenchymal genes, such as *N-cadherin* and *Fibronectin*, whilst inhibiting epithelial genes including *Claudin3, Occludin* and *E-cadherin* [111,118]. EMT is controlled by the transcription factors *Snail1* and *Twist*, both of which are upregulated in mESCs overexpressing *Mesp1*. *Snail1* was

evaluated to assess if it was capable of recapitulating the effect of *Mesp1* on EMT and consequential cardiac differentiation. Overexpression of *Snail1* in mESCs showed similar effects to *Mesp1* on the reduction of epithelial genes such as *E-cadherin,* however, did not produce $FLK1^{+}$ cells that are characteristic of those that go on to form cardiac lineages. Knockout of *Mesp1* in mESCs shows similar *Snail1* expression to wildtype cells, which is likely compensated for through expression of *Mesp2* [111]. This mechanism explains why *Mesp1* deficient embryos are capable of forming cardiac tissue, however, fail to migrate to form the correct heart structure.

### 1.1.16 Activators of *Mesp1* expression

As previously described, initiation of *Mesp1* expression occurs during gastrulation. Its transient expression lasts 24 hours before being rapidly down regulated by E7.5 [89]. It is known to be capable of autoregulating its own transcription [108]. However, its expression is also controlled by other transcription factors found within the primitive streak.

One of these factors is *T,* a T-box transcription factor expressed first in the epiblast extraembryonic ectoderm, before being expressed in the forming primitive streak [119]. Expression of *T* is induced by *Wnt3* signalling, and is found to be enriched in *Mesp1* positive progenitor cells at E6.5 [87]. *T* and *Mesp1* form a regulatory loop, with T capable of binding upstream of *Mesp1* and inducing its expression [92,120,121], the resulting MESP1 protein is then able to bind upstream of *T* and inhibit its transcription [108,119]. Knockdown of *T* is shown to cause a reduction in *Mesp1* expression, whilst *T* is rapidly downregulated post induction of *Mesp1* [108,121].

Another *Wnt* induced T-box transcription factor is *Eomesodermin* (*Eomes*). EOMES is capable of targeting T binding motifs in the *Mesp1* promoter, inducing its expression in the primitive streak [122,123]. This process is only possible in the presence of low levels of Activin, with higher concentrations leading to the induction of definitive endodermal genes [123]. Ablation of *Eomes* subsequently impairs the production of cardiac mesoderm [123]. Recent research has shown that EOMES is capable of binding upstream of the *Mesp1* transcription start site, in complex with other proteins. This includes the histone modifiers GCN5 and WDR5, as well as a large intergenic non-coding RNA 1405 (LINC1405), with the complex binding at a conserved T-binding motif upstream of the *Mesp1* transcription start site, enabling initiation of transcription [124]. MESP1 could also be capable of regulating the expression of *Eomes*, as it is capable of binding directly upstream of the *Eomes* transcription

start site [92]. MESP1 has previously been shown to be capable of binding upstream of *Wnt3*[92], and therefore could also affect expression of both *Wnt3* and *Eomes.*

These studies indicate that *Mesp1* is a central part of the induction of cardiac mesoderm through its role in an autoregulatory loop. Both *Eomes* and *T* are induced by Wnt and Nodal signalling and stimulate the expression of *Mesp1*[125,126]. Reduction in *Wnt, T and Eomes* is necessary to then stimulate progression from a mesodermal to cardiac lineage. Therefore, MESP1 could be capable of binding upstream of both transcription factors and *Wnt3a* to reduce their expression, allowing progression to a mesodermal and eventual cardiac lineage.

*Wnt* signalling is essential to controlling gastrulation in the early embryo. Ablation of *Wnt3* results in the absence of the primitive streak and failure of gastrulation to proceed[127]. *Wnt3* is initially expressed in the visceral endoderm at E5.5. Visceral endoderm, in the distal region of the embryo, secretes Nodal and Wnt inhibitors [128]. This then migrates, and by E6.5, is located at the anterior side, and is known as the anterior visceral endoderm (AVE; Figure 2). *Wnt3* is also expressed in the posterior epiblast, where the primitive streak forms, and induces expression of *Nodal* and *Brachyury* (depicted in Figure 2)[128]. *Nodal* is also found in the epiblast at E5.5, with low levels necessary to specify mesodermal cells, whilst increased concentration specifies an endodermal lineage[129]. Nodal is capable of regulating the *Wnt, FGF* and *BMP* pathways essential to mesoderm formation [129].

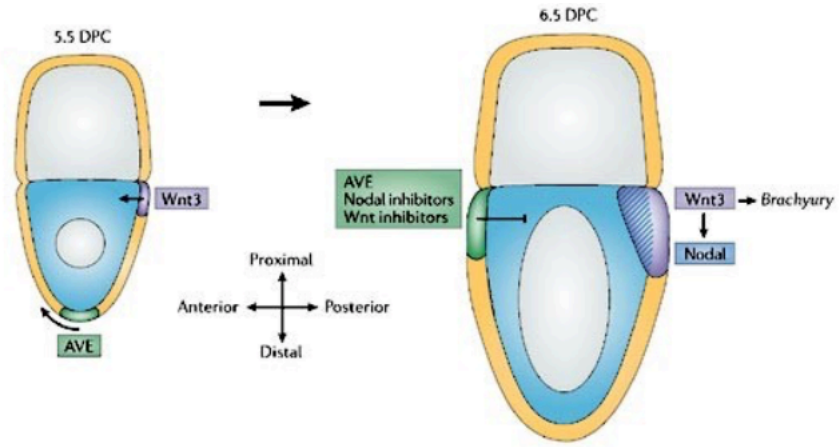



Figure 2: A schematic indicating Wnt3 signalling in the early mouse embryo at E5.5-E6.5 taken from Kimelman et al 2006.

### 1.1.17 Binding partners of MESP1

Whilst the role of *Mesp1* as a transcription factor has been well characterised, there is limited knowledge of its protein interactions with other transcription factor proteins. The bHLH proteins are known to dimerise to other bHLH transcription factors including E12 and E47 [130]. To date, only one other binding partner of *MESP1* has been identified.

The function of *Mesp1* is well conserved through the chordate family, and its function is most apparent in the *Ciona.* A rudimentary model, *Ciona* contain one ortholog of *Mesp1, Cs-Mesp,* which is essential to its cardiogenesis [99]. Experiments revealed that *Cs-Mesp* is expressed in a pair of blastomeres at approximately B7.5, which replicate to form 2 divergent fates [131]. The first set remain in the tail to form tail muscles, whilst the second form the basis of the heart [131,132]. Specification of the heart is mediated through FGF signalling, through the Ets family of transcription factors. In *Ciona* this is limited to a single ortholog *Ci-Ets1/2* and is essential for differentiation and migration of the heart cells [131]. Expression of *Ci-Ets1/2* mimics that of *Cs-Mesp* at B7.5, with *Cs-Mesp* found to directly regulate its transcription [131]. *Ci-Ets1/2* is activated through FGF signalling, which allows phosphorylation of *Ci-Ets1/2.* This subsequently, either directly or indirectly, regulates cardiac genes such as *Ci-FoxF* and *Ci-Nkx* [131].

This process is partially conserved within higher order chordates. *Mesp1* has been found to induce the expression of the mouse paralog of *Ci-Ets1/2, Etv2.* To do so, MESP1 must first directly bind to CAMP-Responsive Element Binding Protein 1 (CREB1). This interaction is mediated through the bHLH of MESP1 and a bZIP domain in CREB1[133]. This protein interaction is then capable of binding to the proximal promoter of *Etv2,* though the cAMP response elements (CRE) motif, enabling modulation of its expression [133]. Mutation of the CRE motif partially mitigates MESP1's activity [133], suggesting that its interaction with other motifs and binding partners may, in part, mediate its role in early development. These *Mesp1 Etv2* double positive cells are of a mesodermal fate, however, by contrast to most *Mesp1* positive cells, do not contribute to a cardiac fate, but to a haematopoietic or endothelial cell lineage[133].

*Etv2* is a member of the *ETS* family of transcription factors. Additional research into the role of another *ETS* family member, *Ets2* in early embryogenesis utilising mESC models, revealed that knockdown of its expression caused a reduction, not only in *Mesp1* expression, but also in other cardiac specification markers such as *Nkx2.5* and *Mef2c* [71]. These cells were also incapable of forming beating cardiomyocytes, with a corresponding reduction in

contractile gene expression [71]. Therefore, it appears that *Ets2* is also involved in early cell patterning decisions, 2 of which, forming haematopoietic and cardiac lineages, overlap with that of *Mesp1*.

### 1.1.18 Roles in reprogramming

The role of *Mesp1* in cardiogenesis is well-described, playing an essential part in the formation of the linear heart tube [89,93]. The expression of *Mesp1* alone is insufficient to induce a cardiac phenotype in somatic cells, however, research has focused on utilising *Mesp1* alongside other partners to enable this transdifferentiation.

The addition of *Mesp1* to other known cardiogenic factors, including *Gata4, Mef2c* and *Tbx5* (*GMT*), has limited impact on the ability of these factors to produce immature cardiac phenotypes from somatic cells in direct reprogramming [68,70].

Other work in hIPSCs has shown that *MESP1*, in conjunction with *BAF60C*, also known as *SMARCD3*, and *GATA4* were the most successful in forward programming cells to a cardiac fate [134]. The role of *SMARCD3* was shown to be partially redundant, with only the combination of *GATA4* and *MESP1* necessary to induce a cardiac lineage. Singularly, they were found to be insufficient in inducing a cardiac lineage [134].

One further study has identified that *Mesp1* could be used with *Ets2* to reprogram somatic cells to a cardiac phenotype [71]. Neither transcription factor is capable of reprogramming the somatic cells alone, however, forced expression of both in human dermal fibroblasts resulted in cardiac progenitor like cells, expressing known cardiac genes including *Nkx2.5* and *Mef2c* [71]. These cells had an immature cardiomyocyte phenotype with partially functional contractile units, and only 2.3% of the cells capable of spontaneous beating [71]. Therefore, it is apparent that *Mesp1,* in conjunction with other genes, may be capable of reprogramming somatic cells to a cardiac lineage.

## 1.2 Aims and hypothesis

The role of *Mesp1* during gastrulation allows for cell patterning to 3 distinct lineages: craniofacial muscle, haematopoietic and cardiac. The mechanism by which *Mesp1* directs these cell fates is currently unknown, however, it could be suggested that it is in part due to MESP1 exploiting different binding partners to progress discrete lineages in cells.

Hypothesis: MESP1 can induce varying cell types by interacting with different binding partners during differentiation, and that these binding partners, in conjunction with MESP1, could be used to reprogram somatic cells to a cardiac fate.

To examine this hypothesis, the following aims and objectives will be answered in this body of work.

Aims:

1. To identify MESP1-protein interacting partners in mESCs at mesoderm stage of differentiation.
2. To ascertain the effect of knockdown of *Mesp1* and binding partners on cardiac differentiation in mESCs.
3. To determine the effect of overexpression of *Mesp1* and its binding partners in fibroblasts and examine the effect on cell fate.

# 2 Chapter 2

## 2.1 Methods

### 2.1.1 Cell maintenance

#### Human embryonic kidney cells

Human embryonic kidney (HEK) cells were kindly donated by Dr Garcia-Maya, King's College London. Cells were maintained in 10% FBS (Gibco, 10500064), 1X penicillin streptomycin (Sigma-Aldrich, P4458) and low glucose DMEM (Sigma-Aldrich, D5546-500ML) at 37 degrees centigrade at 5% $CO_2$.

#### 3T3 cells

3T3 cells were kindly donated by Dr Mankoo, King's College London. Cells were maintained in Glutamax supplemented (Gibco, 10569010) with 10% FBS, 1X penicillin streptomycin at 37 degrees centigrade at 5% $CO_2$.

#### Mouse Embryonic Fibroblasts

Primary mouse embryonic fibroblasts were kindly donated by Professor Zammit, at King's College London. Cells were maintained in Glutamax supplemented with 10% FBS, 1X penicillin streptomycin at 37 degrees centigrade 5% $CO_2$ and 0.1mM betamercaptoethanol (Sigma-Aldrich, M6250).

#### Human Dermal Fibroblasts

Primary human dermal fibroblasts were kindly donated by Dr Shaw, King's College London. Cells were isolated by Mr Soldin, St George's Hospital, Tooting, under ethics approval:
Study Title: Molecular mechanisms of Tissue Repair
REC reference: 14/NS/1073
R&D reference: RJ115/N032
Cells were maintained at 5% $CO_2$, at 37 degrees centigrade in Glutamax supplemented with 10% FBS, 1X penicillin streptomycin at 37 degrees 5% $CO_2$.

A human BJ fibroblast cell line was kindly donated by Professor Ilic, King's College London. Cells were maintained at 5% $CO_2$, at 37 degrees centigrade in Glutamax supplemented with 10% FBS and 1X penicillin streptomycin.

### 2.1.2 mESC – differentiation

The mouse CCE embryonic stem cell line, derived from the 129/Sv mouse strain was kindly donated by Dr Grigoriadis, King's College London. Cells were maintained in 5% CO2 at 37 degrees centigrade for the duration of maintenance and differentiation.

Upon defrost, cells were plated in a 6 well plate that were previously coated with 0.1% gelatin (Sigma Aldrich, G1393) for 20 minutes before aspiration, in 2i media (see recipe below) supplemented with 10% FBS, 1:1000 leukaemia inhibitory factor and 1:1000 G418 (Generon, G418-H). Twenty-four hours post thaw, cell media was replaced with 2i media with LIF (Sigma-Aldrich, ESG1106) and G418, without FBS. mESCs were kept below 70% confluency and sequentially passaged using accutase, until reaching this confluency in two T75cm$^2$ flasks.

To allow the directed differentiation of mESCs to a cardiac fate, cells were dissociated and replated in 10cm petri dishes, in SFD media, at a density of 75,000 cells per cm$^2$. 48 hours post plating, embryoid bodies were dissociated and resuspended in SFD supplemented with 0.125ng/ml bone morphogenic protein 4 (BMP4) (R and D systems, 314-BP-010), vascular endothelial growth factor (VEGF) (R and D systems, 293-VE-010/CF), Activin A (R and D systems, 338-AC-010/CF) , and replated in 10cm petri dishes at a density of 75,000 cells per cm$^2$. It should be noted that BMP4 concentration is batch dependent and must be empirically derived with each new batch. 48 hours post cytokine addition embryoid bodies were dissociated, resuspended in Stem cell pro media, supplemented with VEGF, human fibroblast growth factor b (hFGFb) (R and D systems, 233-FB-025), human fibroblast growth factor 10 (hFGF10) (R and D systems, 345-FG-025), and plated on gelatinised tissue culture plates.

Cell media was subsequently replaced twice a day until day eight post differentiation.

### 2.1.3 mESC media

#### Stem cell maintenance media (2i)

1:1, DMEM/F-12 (Gibco, 11320-033 and Neurobasal medium (Gibco, 21103-049), 1x penicillin streptomycin, 0.05% bovine serum albumin (BSA) (Gibco, K15-013), 0.1mM betamercaptoethanol (Sigma, M6250), 0.5X N-2 supplement (Gibco, 17502-048), 0.5X B-27 supplement (Gibco, 12587010), 1X non-essential amino acid solution (Sigma, M-7145), 1mM sodium pyruvate (Sigma, 28636), 2mM stabilized glutamine (Gibco, 25030-081), 0.1%

sodium bicarbonate (Sigma, S-8875) with 1X leukaemia inhibitory factor (LIF) (Merck Millipore, ESG1106) and 1μM inhibitor of the MEK pathway (MEKi; Sigma, PD0325901), 3μM inhibitor of glycogen synthase kinase-3 (iGSK3β) (Sigma, CHIR99021).

#### SFD Media

IMDM (Thermofisher Scientific, 12440053) and Ham's F-12 (Thermofisher Scientific, 21765029) in a 3:1 ratio, 0.05% BSA, 2mM stabilized glutamine, 1X B-27 supplement, 1X N-2 supplement, 1X penicillin streptomycin, 0.45mM 1- Thioglycerol (Sigma, M6145), 0.05mg/ml ascorbic acid (Sigma, 33034).

#### Stem Cell Pro Media

Stem cell pro media (Gibco,10639011) plus 1X of supplied supplement, 2mM glutamine, 100μg/ml ascorbic acid, 5ng/ml hVEGF, 10 ng/ml human fibroblast growth factor b (hFGFb) and 25 ng/ml human fibroblast growth factor 10 (hFGF10).

#### Videos of cells

Moving images of cells taken during differentiation were taken on a Leica MZ16F microscope using a 20X lens.

### 2.1.4 qPCR

Cells samples were dissociated and centrifuged to result in a cell pellet. Cell pellets were then resuspended and stored in Trizol, at -80 degrees centigrade. RNA was extracted using RNA mini prep columns (Zymo, 62051) according to manufacturer's instructions. RNA was resuspended in 30μl of RNAase free water, and 500 ng converted into cDNA using Superscript II (Thermofisher, 18064022) according to manufacturer's instructions. cDNA was diluted to a standard concentration of 1ng/ μl and stored at -20 degrees centigrade.

All qPCRs were executed at the following settings: 1 minute at 95 degrees centigrade before 40 cycles of 30 seconds at 95 degrees centigrade, 30 seconds at 60 degrees centigrade and 30 seconds at 72 degrees centigrade.
All primers (Table 1) were tested for efficiencies of between 90-110% and analysed using the delta CT method.

The analysis of the verification of the embryonic stem cell differentiation timeline was performed by comparison to a standard curve. A cDNA mix of differentiation time

points was used to make a standard curve from 20ng-0.016ng to assess transcript levels of each gene normalized to the housekeeping gene *Gapdh*. Reactions were performed using Luna qPCR master mix (NEB) on a Stratagene MX3005P (Agilent Technologies) and results analysed using MxPro software. 2ng of cDNA of each time point of 1 differentiation of both induced and uninduced samples was analysed in a time course for each gene and two housekeeping genes.

All other qPCRs were normalised to a house keeping gene, *Gapdh*, before analysis through the delta CT method and normalised to either day 0 or day 4 (mesoderm stage, 88 hours).

Table 1: List of qPCR primers

| Gene | Forward | Reverse | Species |
|---|---|---|---|
| *Amot* | agagaaggagtacgaggggt | ggcgacacattgaaatccga | mouse |
| *Arid1a* | tgcaaccaacctcaatgtgg | agtcccaacccaagatccag | mouse |
| *Arid3b* | aaggttgtcagaggaggagc | cccaatgttgcttgtggtca | mouse |
| *Brachyury* | gcttcaaggagctaactaacgag | ccagcaagaaagagtacatggc | mouse |
| *Chd3* | tttcacagagctgcacaca | aacggaagccaataagggga | mouse |
| *Col1a1* | gcttcacctacagcaccctt | gtccgaattcctggtctggg | mouse |
| *cTNT* | gcggaagagtgggaagagacagac | gcacggggcaaggacacaag | mouse |
| *Desmin* | tattgacctggagcgcagaa | tcatactgagcccggatgtc | mouse |
| *Eomes* | tcgtggaagtgacagaggac | aaaggcttccgggacaactac | mouse |
| *Ep400* | accgagcccaataacttcc | cactttcctggagctctgga | mouse |
| *Fubp1* | aatccgagcagtaccacaca | ggcagaggaggtgaacagat | mouse |
| *FSP1* | aggaggccctggatgtaattg | attgtccctgttgctgtccaa | mouse |
| *GAPDH* | gtgaaggtcggtgtgaacg | aattgatgttagtggggtctcg | mouse |
| *Gata4* | ccacgggccctccatccat | ggcccccacgtcccaagtc | mouse |
| *Gatad2a* | agggtggagccgttatgtg | gttctgcacactttcagcca | mouse |
| *Islet1* | agcagcaacccaacgacaaaacta | gtatctgggagctgcgaggacat | mouse |

| | | | |
|---|---|---|---|
| *Lbd1* | acttatgtcccgccacaag | gcaacaccaacaacagcaac | mouse |
| *Mbd3* | gcggaagaggtgggagtgc | ttcgacttccgcaccggaa | mouse |
| *Mesp1* | gtccaggtttctagaagagcc | cagaatcgtgggacccatc | mouse |
| *Myh6* | cagtacatgctgacagatcg | tagcctggataatctggtcc | mouse |
| *MyoD* | tccgctacatcgaaggtctg | ccgctgtaatccatcatgcc | mouse |
| *Myogenin* | acaggccttgctcagctc | cgctgtgggagttgcatt | mouse |
| *Nanog* | atggaaaggcttccagatgc | ataagcaggttaagacctcg | mouse |
| *N Cadherin* | ttacagcgcagtcttaccga | gagggaagcttctcacagca | mouse |
| *Nestin* | agtgatgccccttcaccttg | gctcgctctctactttcccc | mouse |
| *Nkx2.5* | ctccgatccatcccacttta | agtgtggaatccgtcgaaag | mouse |
| *Oct4* | gaagcagaagaggatcaccttg | ttctaaggctgagctgcaag | mouse |
| *PDGB* | ctaagattggagagaagagc | aagacaacagcatcacaagg | mouse |
| *Sall4* | acacttgtgcttaccgagga | aggtagcagggcactgggaa | mouse |
| *Smarcc2* | tacgaagtcagtggtggtgc | taattgcctgtcacgaccgtag | mouse |
| *Supt5h* | cgcatcaaagcccgaatgag | catctttgaggggaaccgct | mouse |
| *Snai1* | cctgttgctgtccaagttgc | ccctggatgtgatggtgtcc | mouse |
| *Tbx5* | ctaccccgcgcccactctcat | tgcggtcggggtccaacact | mouse |
| *Trim33* | gcacagtgcccctcagtatt | aacaagtcccaccacagcaa | mouse |
| *V5 tag* | tcctcggtctcgattctagc | acgctgttcccatcggaagg | mouse |

| | | | |
|---|---|---|---|
| *Wdr5* | cctctggccagtgtctgaag | actacagcaaggggaagtgc | mouse |
| Nestin | cctcaaccctcaccactctatttt | gctttttactgtccccgagttctc | Human |
| Runx1 | cactctgaccatcaccgtct | cgctcggaaaaggacaaact | Human |
| Col1a1 | agtggtttggatggtgccaa | aggggctccagggcg | Human |
| Desmin | gacctggagcgcagaattga | ggcagtgaggtctggcttag | Human |
| FSP1 | atggtgtccaccttccacaag | tgttgctgtccaagttgctc | Human |
| N Cadherin | caacggggactgcacagatg | tgtttggcctggcgttcttt | Human |
| RPLO | tctacaaccctgaagtgcttgat | caatctgcagacagacactgg | Human |
| Runx1 | catcgctttcaaggtggtgg | ttccctcttccacttcgacc | Human |
| Snai1 | tcccagatgagcattggcag | gcgagctgcaggactctaat | Human |
| Tal1 | cacctggtcctgctgaacg | gtcgcggccctttaagtct | Human |

### 2.1.5 Rapid immunoprecipitation mass spectrometry

Cells from 5 separate differentiation protocols were collected at mesoderm stage (88 hours of differentiation). Cells were identified as 'induced' where V5- tagged *Mesp1* expression had been induced through the addition of doxycycline, 'uninduced' in those where no doxycycline had been added, and 'IgG control' samples. Cells were fixed in 11% paraformaldehyde mix (11% formaldehyde, 0.1M sodium chloride, 1mM EDTA, 0.05M Hepes pH 7.9) for 8 minutes. Fixative was added at a 1:10 dilution with the total media amount. Fixative was quenched using 2.5M glycine solution (final concentration of 0.125M) applied for 5 minutes before cells were centrifuged at 4 degrees centigrade for 10 minutes at 800xg. Cells were resuspended in PBS/IGEPAL solution (1 X PBS, 0.5% IGEPAL CA-630) before re-centrifugation. Finally, cells were resuspended in PBS/IGEPAL solution with 1% phenylmethylsulfonyl fluoride (Sigma, Cat#10837091001) before further centrifugation and pellets snap frozen and stored at minus 80 degrees centigrade. Samples were pooled into two replicates of doxycycline-induced cells and two replicates of uninduced cells. Cells were sent to Active Motif for Rapid immunoprecipitation mass spectrometry. V5- MESP1 and its binding partners were isolated by immunoprecipitation of the V5-tag or an IgG control to ensure specific binding before binding partners were identified using mass spectrometry

### 2.1.6 Plasmids

Potential Mesp1 binding partners, Eomes and Mbd3, were purchased in expression vectors from Vector Builder. Both contained N-terminal FLAG-tags. *Mesp1* was purchased with an N-terminal V5 tag from the same company. N-terminal FLAG tagged pcDNA3 *Wdr5* was a gift from Kai Ge (Addgene plasmid #15552). Plasmids containing mouse *Sall4* open reading frame was generously given by Professor Malcolm Logan, mouse *Amot* by Prof. Ling Yi Chen (Nankai University, China) and *Trim33* by Dr Florence Cammas (University of Montpellier). These were cloned into pcDNA3.1 vector containing N-terminal FLAG tag.

### 2.1.7 Cell transfection

2.5 x $10^5$ HEK cells were plated in each well of a six well gelatinised plate. 24 hours post plating, cell media was changed, and cells were co-transfected using 9µl of Lipofectamine 2000 per well in 250µl OptiMEM, in accordance with manufacturer's instructions. Plasmid co-transfection amount was previously optimised and variable amounts

were used according to the protein expression efficiency, as assessed by western blotting. Cells were lysed 24 hours post transfection.

### 2.1.8 Co-Immunoprecipitation

#### Cell lysate

24 hours post transfection, media was aspirated and 250µl of cold RIPA buffer (50mM Hepes (pH 7.6), 1mM EDTA, 0.7% DOC, 1% IGEPA, 0.5M LiCl) was added per well and incubated for 5 minutes on ice. Cells were subsequently scrapped into cold 1.5ml microfuge tubes, with 2 wells combined into 1 tube, before a further 10-minute incubation on ice. Cell lysate was centrifuged at 8000RPM for 10 minutes at 4 degrees centigrade. Supernatant was transferred to a new cold 1.5ml microfuge tube, with the cell debris pellet discarded. 50µl of cell lysate was aliquoted from each sample and stored at -20 degrees centigrade as the whole lysate (W) sample. 200µl of each sample was further aliquoted and formed those immunoprecipitated with the FLAG antibody (FLAG IP) (Sigma-Aldrich, F1804) and IgG control antibody (IgG IP) (Cambridge Biosciences, 1265-100) and stored at 4 degrees centigrade.

Magnetic IgG beads (Thermofisher Scientific, 10003D) were vortexed, and 25µl for each FLAG and IgG control immunoprecipitation sample, plus 10µl extra was aliquoted. Beads were immobilised on a cold magnetic strip and washed twice in 500µl of RIPA buffer with protease inhibitors (Sigma-Aldrich, 11836170001). Beads were resuspended in RIPA buffer to the original volume aliquoted. 1.5, 3 or 5µg of either IgG or FLAG antibody was added to the corresponding sample, and an additional 150µl of RIPA buffer to allow the beads to rotate. Samples were left to rotate at 4 degrees centigrade for a minimum of 4 hours.

Beads were subsequently immobilised on a magnetic strip, supernatant removed, and 200µl of RIPA buffer added. This was repeated twice before beads were resuspended in 25µl of RIPA buffer.

200µl of cell supernatant was added to the FLAG and IgG conjugated magnetic beads and left to rotate overnight at 4 degrees centigrade. The following day samples were immobilised on a magnetic strip and supernatant collected. This constituted the flow through (FT) for the FLAG and IgG immunoprecipitation samples. Beads were resuspended in 500µl of RIPA buffer and left to rotate at room temperature for 5 minutes, before being immobilised and supernatant discarded. This was repeated twice more. Beads were resuspended in 40µl l 1x Laemmli buffer (62.5 mM Tris-HCl, pH 6.8 10% glycerol, 1%

LDS) with 10% betamercaptoethanol before being heated at 95 degrees centigrade for 10 minutes. Finally, beads were immobilised on a magnetic strip and supernatant transferred to a new 1.5ml microfuge tube and stored at -80 degrees centigrade.

### 2.1.9 Western blotting

Co-immunoprecipitation samples for western blotting were prepared as follows: Flow through (FT) and whole samples (W) were defrosted on ice, and 10 µl of sample was added to 30µl 1x Laemmli buffer with 10% betamercaptoethanol before being boiled at 95 degrees centigrade for 10 minutes. Samples were centrifuged briefly.

Western blotting technique was optimised to each protein dependent on size and protein structure. All samples were run at 120V, on either 4-20% or 12% polyacrylamide gel (Mini-PROTEAN TGX Gels, BioRad), before samples were transferred using the Transblot Turbo system (BioRad) on the mixed molecular weight programme (7 minutes, 25V, 2.5A). Membranes were blocked in 5% milk in TBST (50 mM Tris-Cl, pH 7.6; 150 mM NaCl, 0.1% Tween 20) for 1 hour at room temperature, before primary antibodies were added overnight at 4 degrees centigrade at the stated dilutions (antibodies list). Membranes were washed for 15 minutes in TBST three times, before secondary antibodies were added in 5% milk TBST (antibodies list) and incubated at room temperature for 2 hours. Membranes were washed three times in PBS for 15 minutes, before imaging (BioRad, ChemiDoc XRS+).

### 2.1.10 Antibodies

Anti-FLAG – Sigma, F1804, 1:500
Anti-FLAG – NEB, 1479S, 1:500
Anti-V5 – Cell signalling, 1:500
Anti-V5 – Thermofisher Scientific, MA515253, 1:500
Alpha tubulin – Abcam, ab4074, 1:1000
GAPDH - Proteintech, 60004-1-Ig
Anti-Mouse HRP conjugated – Pierre (Now Thermofisher)
Anti-Rabbit HRP conjugated - Pierre (Now Thermofisher)
Anti- Mesp1 – Biorbyt, orb1329
Anti- Wdr5 – Sigma-Aldrich, ab22512
Anti- Eomes (Tbr2) – Sigma-Aldrich, ab15894
Anti-Mbd3 – Sigma-Aldrich, ab157464

### 2.1.11 siRNA

SiRNA (Flexitube, Qiagen) targeted against *Eomes, Mbd3, Mesp1 and Wdr5* (Table 2) were resuspended in 100 µl of RNAse free water to a final concentration of 25nmol. A mix of four siRNA against each target was mixed in equal volumes to a final solution of 5nmol (Table 2).

Mouse embryonic stem cells on day two of differentiation were plated at a concentration of 180,000 cells per well in a 12 well non tissue culture treated plate in SFD media supplemented with cytokines. Per well, 9 µl of RNAiMax was diluted in 150 µl of OptiMEM, whilst separately 2 µl of siRNA mix was diluted in 150 µl OptiMEM and incubated for 5 minutes at room temperature. RNAiMax and siRNA solutions were mixed and incubated for a further 20 minutes at room temperature, before 250 µl was added per well. At day 4 of differentiation, embryoid bodies were dissociated using accutase. Cells from each well were divided into three further wells in a 12 well plate and plated in Stemcell Pro media. One 6 well plate well was taken as a sample at day 4, alongside samples at day 5, 6 and 8.

Table 2 siRNA sequences

| Gene | Target sequence | Catalogue number |
|---|---|---|
| *Mesp1* | TGCCTGGTGTGTATTTATTTA | SI02709616 |
| | CAGAAACAGCATCCCAGGAAA | SI02687657 |
| | AGGGCTCAGGATAAAGCTACA | SI00177562 |
| | ACCGATTGTGCTAGTGTCAAA | SI00177555 |
| *Mbd3* | AAGTCACTTTCCTTCAATAAA | SI02740045 |
| | CAGGACCATGGACTTGCCCAA | SI00206850 |
| | ACCGGTGACCAAGATCACCAA | SI00206843 |
| | CGCAAAGATGTTGATGAACAA | SI00206836 |
| *Eomes* | CCGGTGCTATTAAATGAATTT | SI00994252 |
| | ACCACTGAAGAGTACAGTAAA | SI00994245 |
| | CTGCGGCAAAGCGGACAATAA | SI00994238 |
| | CACGGATATCACCCAGCTAAA | SI00994231 |
| *Wdr5* | CTGGAAAGAGGTAGTGGCCAA | SI02745638 |
| | TAGCGTGGTCATCAGATTCTA | SI02720396 |
| | ATCGAAGAAGCCATTGTTAAA | SI02676310 |
| | CTCACTGGTATCACTCAGTAA | SI00231882 |
| Negative | Proprietary sequence not given | SI1027280 |

### 2.1.12 Viral vectors

Mouse and human orthologues of potential binding partners in viral vectors were purchased from Vector builder and DNA prepared (Figure 3).

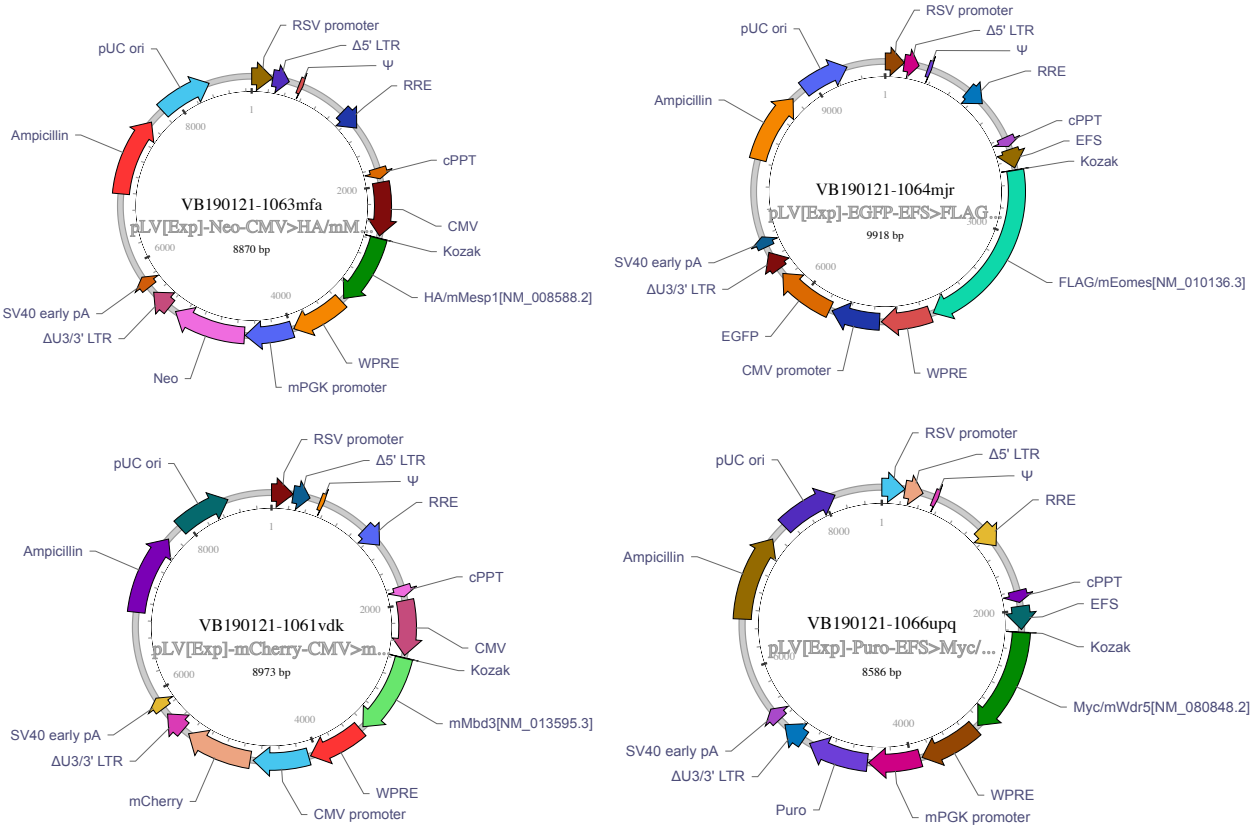
*Mesp1* orthologues contain: CMV promoter, N-terminal HA tagged gene with Neomycin resistance

*Mbd3* orthologues contain: CMV promoter mCherry selection

*Eomes* orthologues contain: N-terminal FLAG tagged gene with EFGP selection

*Wdr5* orthologues contain: N-terminal MYC tagged gene with Puromycin resistance

Figure 3: Vector maps of lentiviral vectors



#### Packaging

To package the vectors, HEK cells were plated in 10cm tissue culture dishes at 60% confluency. Per well a mix of plasmid DNA was diluted in 150µl of OptiMEM with a final concentration of packaging plasmids 2µg of RSV (Addgene, 12253) 6 µg PMDLG (Addgene, 12251), 4 µg of PMD2G (Addgene, 12259) and 8 µg of the plasmid of interest. 60µl of Lipofectamine was diluted in 150 µl of OptiMEM and incubated for 5 minutes at room temperature. Lentiviral packaging vectors, gene of interest vector and lipofectamine mix were combined and incubated for a further 20 minutes at room temperature before 250 µl was added to each plate.

Cell media was changed 24 hours post transfection, and media containing virus packaged vectors collected 48 hours post cell media exchange. Media was centrifuged to remove cell debris and syringe filtered through a 0.45uM filter. Viruses were then stored at -80 degrees centigrade.

#### Transduction

To transduce, cells were plated at 20,000 cells per well in complete media (Glutamax DMEM, 10% FBS, 1X pen/strep) in a 6 well tissue culture treated plate. 24 hours post seeding, Polyethylenimine (PEI) was diluted 1 in 10 in water, before 8 µl was added to 1.5ml of complete media, (Glutamax DMEM, 10% FBS, 1 X pen/strep) and 0.5ml of the selected virus media. Media was removed from each well and replaced with viral mix. Cell media was replaced 24 hours later. This process was repeated 72 hours post initial transduction to improve efficiency.

### 2.1.13 EDU Assay

To analyse cell proliferation, virally transduced 3T3 cells were plated at 5000 cells a well in a 96 well plate. The next day cell media (GlutaMAX supplemented with 10% FBS and 1X penicillin streptomycin) was refreshed with 100 µl complete media with 1:1000 dilution of EDU per well (Thermofisher, C10086) and incubated for two hours. Media was subsequently removed, and cells fixed for 10 minutes in 4% paraformaldehyde, before being washed in 1 X PBS. Samples were permeabilised in 0.1% Triton PBS. Primary antibodies were added in gold buffer (10 mM Tris, 155 mM NaCl, 2 mM EGTA, 2 mM MgCl2, pH 7.2) with 1% BSA overnight at 4 degrees, before three washes in PBS. Cells were incubated with Red azide 568 (Thermofisher, C10086) according to manufacturer's instructions for 30

minutes before being washed three times in PBS. Secondary antibodies (Thermofisher, A-10680) in 1% BSA gold buffer were added for a further 2 hours in dark conditions at room temperature, before samples were washed three times in PBS, and the addition of DAPI (1:1000) for 10 minutes. Finally, samples were washed three times in PBS and stored at 4 degrees until imaging. Samples were imaged using a Zeiss Axiovert 200 M microscope using a Zeiss AxioCam HRm and AxioVision software version 4.4 (Zeiss). Images were analysed using FIJI ImageJ software. Images split into separate colour channels and EDU positive (Azide 568 stained) quantified in comparison to all DAPI stained nuclei. Data was analysed using GraphPad Prism version 8. A minimum of two wells from three transductions of each cell line, total number of cells 6260, were analysed using one way ANOVA followed by Dunnett's multiple comparisons test.

### 2.1.14 Embryoid body size

Embryoid body size was assessed at day 4 of differentiation at mesoderm stage. A 12 well tissue culture plate was coated in 0.1% gelatin for 20 minutes, before being aspirated and drying for a further 10 minutes. Using a cut tip, 100µl was taken from each replicate of each siRNA condition and plated into a well of the 12 well plate. Samples were imaged on a light microscope at 20x magnification, with at least 3 images being taken in each well. Images were analysed using FIJI Image J software to determine embryoid body area. Background below 50 pixels was subtracted. To measure embryoid body area cells were outlined using the free hand tool and measured using ROI manager. Statistical analysis was performed using GraphPad Prism version 9 with conditions compared using nested one-way ANOVA with multiple comparisons. Three images were taken in 3 wells for 2 biological replicates.

### 2.1.15 Selection

#### MTT Assay

To evaluate cell viability, cells were plated at 5000 cells a well in a 96 well plate. Cells were treated with either neomycin or puromycin at a range of concentrations for 1 week, with media changed every 2 days. On day 7, media was exchanged with 100 µl of 50% serum free media, and 50% MTT reagent (Abcam, ab211091) per well before being incubated for 4 hours at 37 degrees centigrade at 5% $C0_2$. Post incubation, 150 µl of MTT solvent (Abcam, ab211091) was added per well, before the plate was covered and placed on

an orbital shaker for 15 minutes. Plates were read immediately at 590nm (PoleSTAR omega plate reader, BMG Labtech). Three wells were measured for three transductions of each cell line.

#### FACS sorting

To analyse transduction of lentiviral vectors with fluorescent reporters, fibroblasts that had been transduced were expanded to approximately 12 million cells. Before FACS sorting, cell media was exchanged 2 hours before FACS sorting, before being removed and centrifuged at 1200RPM for 3 minutes. Preconditioned media was removed and kept at 37 degrees centigrade in a water bath, with any cell pellet discarded. Cells were dissociated, cells centrifuged at 1200RPM for 3 minutes and resuspended in 300 µl of FACS buffer (GlutaMAX, 3% FBS, 5mM EDTA) per 2 million cells before being filtered through a 100 µm cell strainer. Cells were sorted using a BD FACS Aria 3, with a 130µm nozzle, into a 15ml tube with complete media. Sorted cells were centrifuged and plated in preconditioned media in 1 well of a 12 well tissue culture plate. Full GlutaMAX media was exchanged 24 hours later, and cells split 48 hours post sort.

### 2.1.16 Cell and Nuclear morphology

To examine cell and nuclear morphology, 3T3 cells transduced with lentiviral vectors were plated at a concentration of 5000 cells per 96 well tissue culture plate well in triplicate. Cells were fixed 24 hours post plating in 4% PFA for 10 minutes before washing in 1X PBS. Samples were permeabilised in 0.1% Triton PBS for 10 minutes before the addition of 1:500 alpha tubulin (E7, Developmental Studies Hybridoma Bank) in 1% BSA gold buffer overnight at 4 degrees centigrade. Cells were washed 3 times in 1X PBS before 1:1000 dilution of secondary antibody in 1% BSA gold buffer was added. The conjugated fluorophore varied depending on fluorescent reporters within cells. Images were acquired on a Zeiss Axiovert 200 M microscope using a Zeiss AxioCam HRm and AxioVision software version 4.4 (Zeiss). Images were analysed using FIJI ImageJ software. Images were split using adjust, colour, colour channel split. DAPI stained nuclei were converted to 8bit images before threshold was adjusted. Background below 50 pixels was subtracted, before processing through watershed. Nuclei area and circularity were analysed using the analyse particle command. To measure corresponding cell area and circularity, cells were outlined using the free hand tool and measured using ROI manager. Data was analysed using GraphPad Prism Version 8, with conditions compared using one-way ANOVA followed by

Dunnett's multiple comparisons in comparison to EV. Three images were taken in 3 wells of each cell line, with a minimum of 130 cells counted per cell line.

### 2.1.17 Apoptosis and Necrosis Assays

To assess the rate of apoptosis and necrosis in transduced 3T3 cells, cells were plated at a concentration of 1000 cells per well in a 96-well plate. Media was supplemented with Annexin V assay (Promega, JA1011) according to manufacturer's instructions. The plate was read immediately to note the fluorescent at 485nm at 1500 Gain (PoleSTAR omega plate reader, BMG Labtech) and luminescent (Mithras LKB 940, Berthold Technology) baseline levels. Plates were subsequently incubated at 37 degrees centigrade at 5% $CO_2$ for 72 hours, with luminescence and fluorescence measured every 24 hours for a total of 3 days.

Values were corrected for a media only control, before comparison to empty vector transduced cells to evaluate effect of the transduced gene in comparison to control. Data was analysed using GraphPad Prism Version 8, with conditions compared using one-way ANOVA. Three wells were measured for each cell line.

# 3 Chapter 3

## 3.1 Identification of three potential MESP1 binding partners that can alter embryonic stem cell cardiac differentiation

Cell tracing in the early embryo has shown that expression of *Mesp1* during gastrulation is a key marker that cells will progress into 1 of 3 fates: cardiac, haematopoietic and craniofacial muscle lineages. This can be recapitulated in mESCs, where alterations in cell culture conditions and overexpression time periods can result in these 3 lineages [135].

Whilst time controlled overexpression of *Mesp1* in mESCs is capable of restricting pluripotent cells to specific lineages, it is not sufficient to reprogram lineage restricted cells such as fibroblasts to other cell fates [71].

It is unknown *in vivo* how MESP1 can direct these 3 fates, with expression restricted to between E6.25 and E7.5 during mouse gastrulation. It is possible that MESP1 works in collaboration with other key factors during this expression period, that are changeable dependent upon cell location and cell patterning. However, these factors are, as yet, unknown with only one binding partner of MESP1 currently delineated.

MESP1 is capable of binding with CREB1, forming a heterodimer mediated by the bHLH domain of MESP1 and the bZIP domain of CREB1[133]. The formation of the dimer allows for the activation of *Etv2*, through the CRE motif found in the proximal promoter of the gene [133]. Mouse progenitor cells that expressed both genes went on to contribute to haematopoietic and endothelial lineages, suggesting a role for the complex in the restriction of cells to this mesodermal lineage [133].

In addition to this known binding partner, members of the bHLH transcription factor family are known to be capable of forming heterodimers with E12 or E47 proteins [89], enabling transcription of genes through the binding of Ebox motifs [111,112]. Heterodimerisation between MESP1 and E47 has been shown to enhance transcriptional activation, as opposed to a homodimerization of either protein [136].

To identify new binding partners of MESP1, it is essential to capture those that are expressed at a similar timepoint to peak *Mesp1* expression. Therefore, only those proteins found in the nucleus at mesoderm stage of ESC differentiation, that mimics the primitive streak, would be capable of binding to MESP1 *in vivo.* To examine which proteins are capable of binding, a mass spectrometry (mass spec) approach is necessary to identify putative candidates. Whilst traditional mass spec would identify these candidates, a novel method that enables isolation of proteins found only in the nucleus would prove more

stringent. Rapid immunoprecipitation mass spectrometry (RIME) utilises the power of mass spectrometry, whilst including preliminary steps to fix cells with paraformaldehyde and isolate the nucleus [137,138]. This method utilises immunoprecipitation to isolate endogenous proteins, identifying protein complexes that are transient in nature, and of low abundance within the cell [137,138]. This protocol has successfully been deployed in other mESC projects for the isolation of DNA sequences associated with specific histone modifications and transcription factors, and can also be used for protein protein interactions [137].

Identification of these binding partners is essential to further the field of somatic cell reprogramming. Pioneering work by Yamanaka [58] demonstrated that overexpression of a set of transcription factors was sufficient to reprogram somatic cells to iPSCs. Conversion of somatic cells to another lineage has also been proven possible in the field of skeletal muscle, where expression of *Myod* in fibroblasts is sufficient to transform fibroblasts into myoblasts and multinucleated myotubes [139]. No such master regulator has yet been identified in cardiac differentiation, therefore, like the reprogramming of somatic cells to iPSCs, a set of proteins may be necessary to induce this transdifferentiation.

Attempts at reprogramming fibroblasts to a cardiac lineage have proven moderately successful. Efforts to reprogram mouse cardiac fibroblasts using a combination of transcription factors including *Gata4*, *Mef2c* and *Tbx5* [68] found approximately 2.5% of cells exhibited cardiac structural proteins *Myosin heavy chain* and *Troponin T.* This result was variable between research groups, with some studies finding no evidence of cardiac gene induction [140]. Later studies showed that the addition of chromatin modifier, *Smarcd3*, and transcription factors *Mesp1, Myocardin and Serum Response Factor* (*SRF*) increased the efficiency of this reprogramming in mouse embryonic fibroblasts (MEFS), from 0.05% to 2.4%[70].

The use of histone modifiers and transcription factors in reprogramming is seen to have a dual effect; the ability to change the epigenetic landscape of the cell through opening up the chromatin and enabling transcription factor access. This can be through enabling the binding of other transcription factors or directly recruiting histone modifiers [141]. Reprogramming a somatic cell from a defined lineage requires the removal of highly repressive histone marks, to allow the opening and expansion of previously compacted closed chromatin. This complex system commands the need for histone modification proteins to enable the transdifferentiation from one cell lineage to another. The reprogramming of fibroblasts to a pluripotent state by the Yamanaka factors is facilitated by members of the

SWI/SNF family of chromatin remodelling factors [141], the activity of which opens chromatin structures facilitating transcription factor binding and ultimately its activity[141]. This suggests that the ability to open and maintain this chromatin structure is essential for reprogramming cells.

It is, therefore, important to identify transcription factors and histone modification factors that are bound to MESP1 at mesoderm stage. Mesoderm stage in ESCs mimics mesodermal cells found in the primitive streak in the early embryo. The cells are marked by the expression of genes including *Mesp1* and *T.* During cardiac differentiation of ESCs this is typically observed on day 3, or 4, of differentiation dependent on the protocol used. These putative candidates, either singularly or in concert with MESP1, may prove capable of reprogramming somatic cells to a cardiac progenitor fate.

### 3.1.1 Identifying MESP1 binding partners at mesoderm stage in mESCs

To identify potential protein interacting partners of MESP1, a doxycycline inducible *V5-tagged Mesp1* mESC line was generated and cultured. A protocol was optimised to direct cells towards a cardiac fate (S. Cutty, unpublished) and validated (Figure 4, Figure 5). During embryoid body stage, doxycycline was administered to half of population of cells to induce expression of *V5-tagged Mesp1* (Figure 6).

Cells were collected at mesoderm stage, approximately 40 hours post-addition of doxycycline. In total, cells from four separate differentiations were pooled, for two replicates of doxycycline induced cells, two replicates of uninduced cells and an IgG control. Cells were sent to Active Motif to complete RIME[137,138] based on an antibody targeting the V5 tag, co-immunoprecipitating any protein bound to V5-tagged MESP1 (Figure 7).

Figure 4

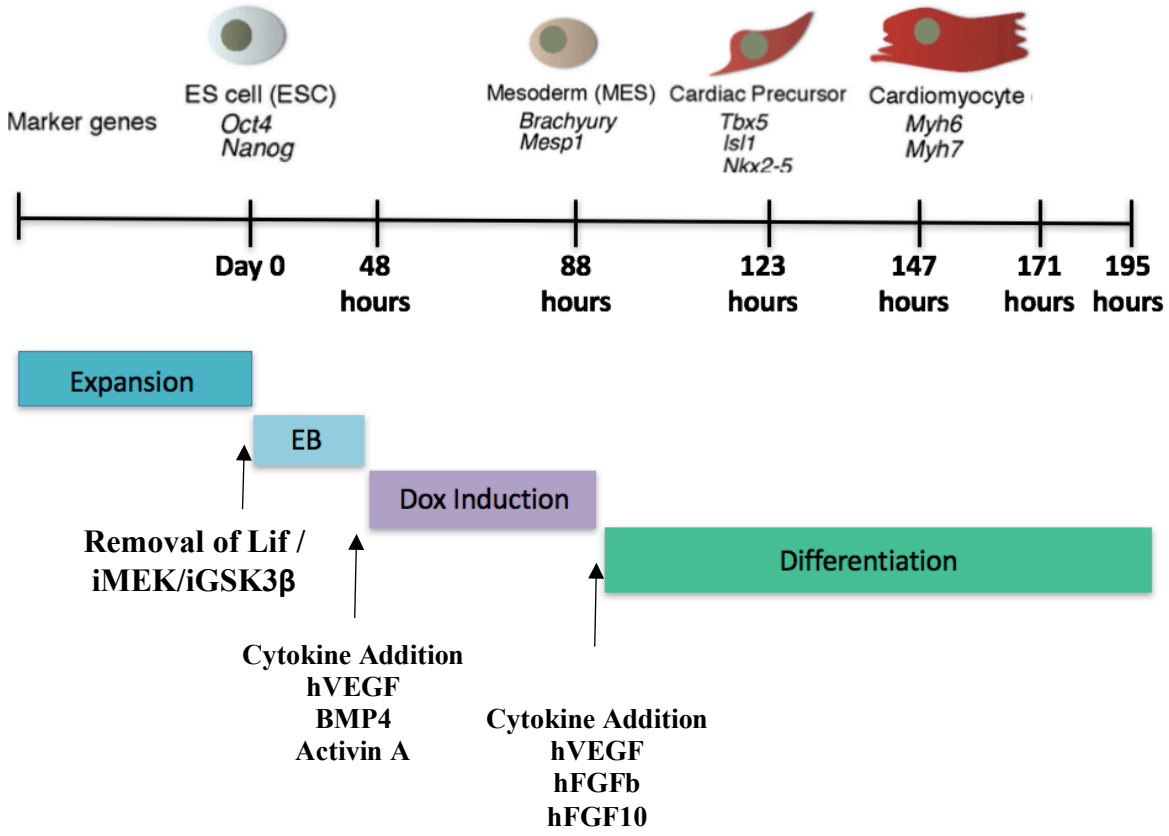



Figure 4: A schematic illustrating the differentiation process of mESCs to cardiomyocytes. A schematic of the differentiation of mESCs to cardiomyocytes. mESCs are expanded in adherent culture until Day 0, upon which cells are plated in petri dishes to allow embryoid (EB) formation with the removal of inhibitors (Lif, iMEK and iGSK). Addition of Doxycycline allows induction V5-tagged Mesp1 expression, with cytokines directing cells towards a cardiac lineage. Samples were collected at Day 0, 48 hours, 88 hours (mesoderm stage), 123 hours (cardiac progenitor stage) and 147 hours (cardiomyocyte stage) post differentiation.

Figure 5

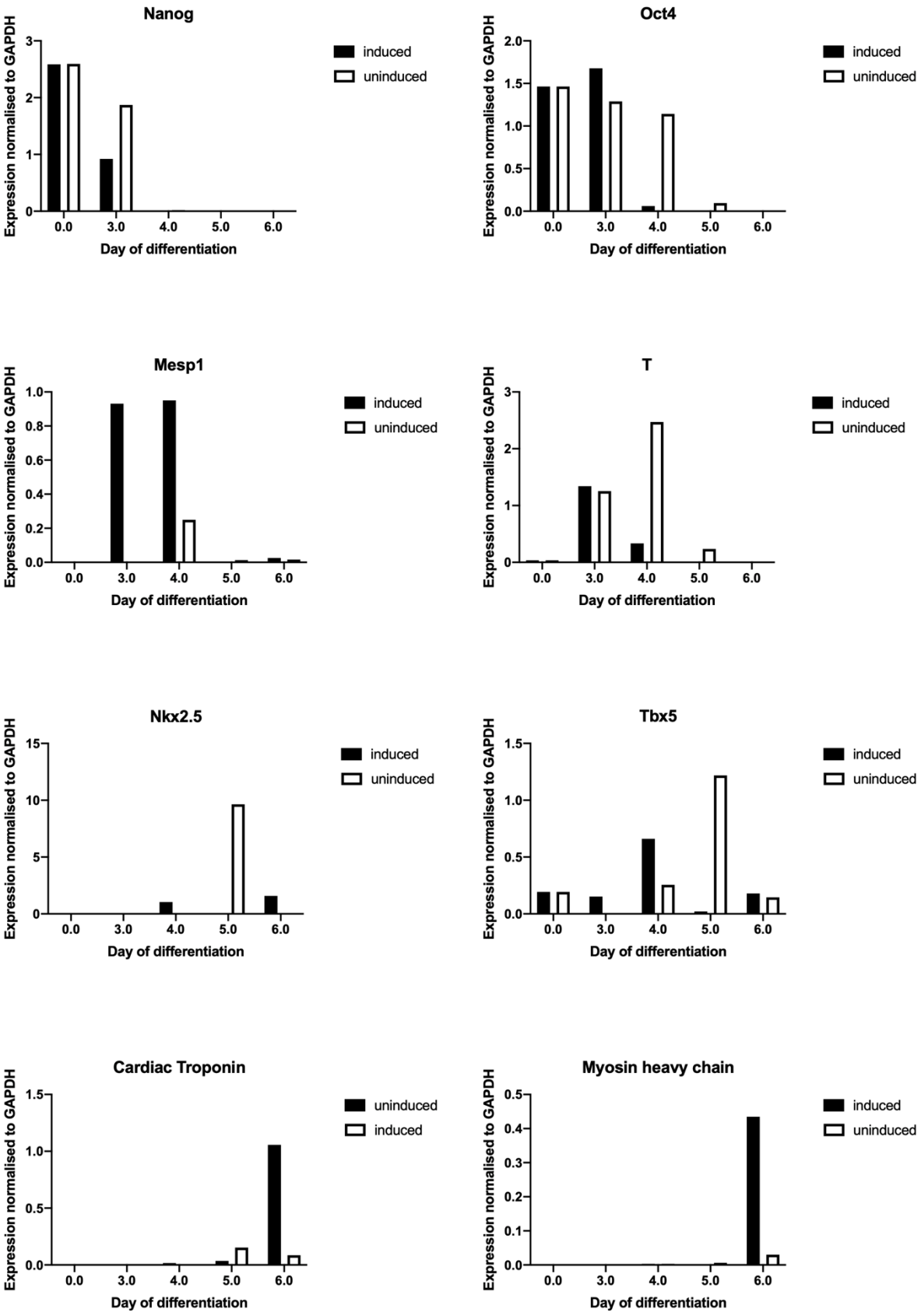



Figure 5: Verification of mESC Differentiation qPCR analysis confirms differentiation profile of mESC from stem cell state, denoted by high levels of *Oct4* and *Nanog*, to a primitive streak phenotype indicated by peaks in *Brachyury* and *Mesp1*, before differentiation into first cardiac progenitors as seen by increased *Nkx2.5* and *Tbx5*, and then cardiomyocytes (*Cardiac troponin*, *myosin heavy chain* 6). For all genes: Relative expression was calculated in comparison to a standard curve (20ng-0.016ng) and normalised to *GAPDH*. Samples were isolated from 1 differentiation.

Figure 6

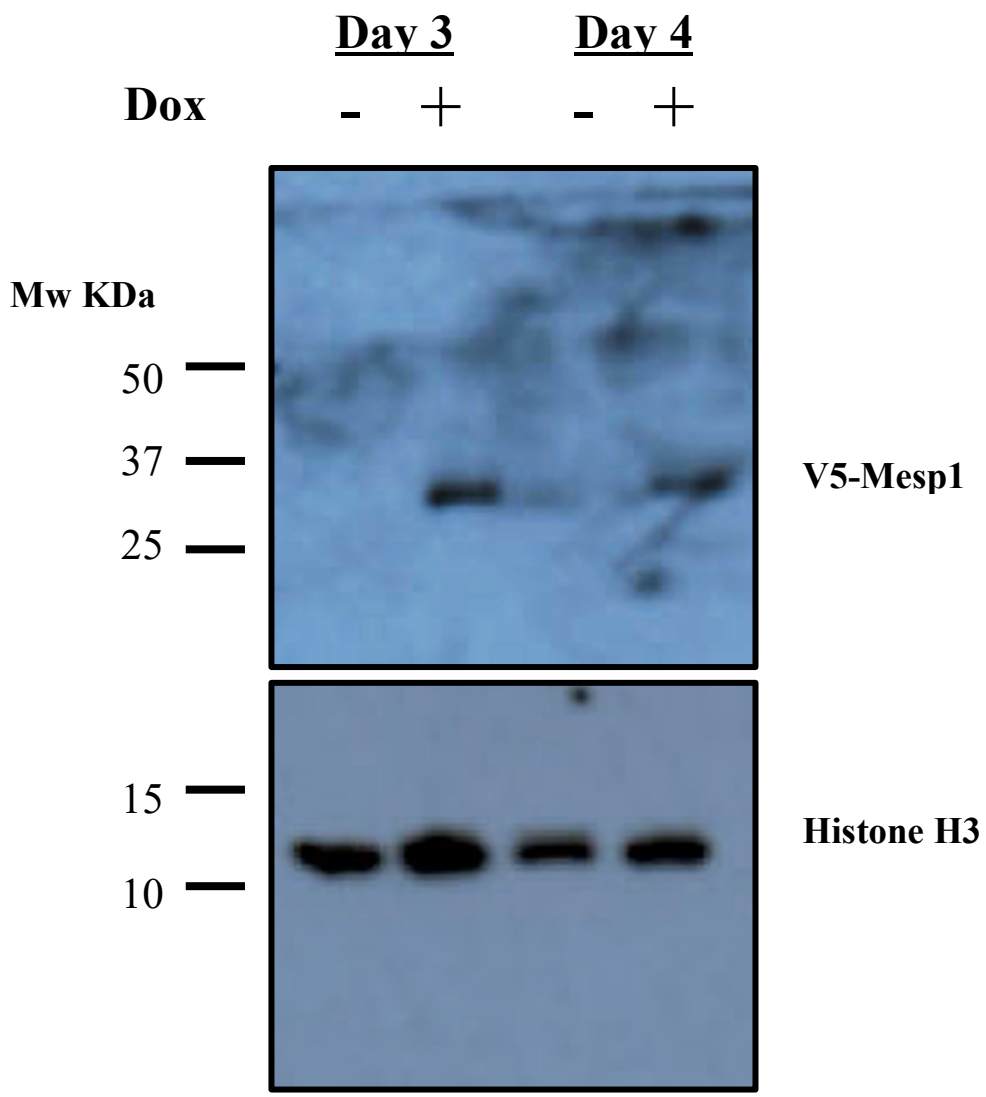



Figure 6: Western blot confirming V5-MESP1 expression. A western blot of samples taken at day 3 and day 4 of differentiation, of doxycycline (dox) induced (+) and uninduced (-) cells. Immunoblotting of the V5-tag is used to successfully identify V5-MESP1 induction.

Results from RIME (see Materials & Methods) were received in a format that identified proteins in the induced samples and uninduced samples normalised against the IgG control from spectral counts (Section 8) (Figure 7B).

The data produced some conflicting results between the two replicates; however, some data was as expected and could be rationalised. These conflicting results could be due to protein interactions that were transient in nature and of low abundance, or slight variations in the time of cell harvest for each differentiation.

In order to extract putative candidates, lists were compared to identify proteins that had enriched spectral counts in the induced samples, pinpointing MESP1, ANGIOMOTIN (AMOT) and WD REPEAT DOMAIN 5 (WDR5) as unique to the enriched samples (Table 3). It should be noted that whilst both MESP1 and AMOT were present in both induced samples, WDR5 was only present in replicate 1.

The remaining list was interrogated for proteins enriched in 1 of the induced cell replicates, as denoted by having at least 5 spectral counts more than the uninduced sample. These proteins were further condensed to those that were histone modifiers or transcription factors (Table 3), before evaluating those that had a similar spatiotemporal pattern *in vivo* to *Mesp1*, or similar phenotypes when mutated or ablated, in mice, zebrafish or humans (Table 4).

The analysis process resulted in a final list of 14 proteins. These included 2 transcription factors (SALL4 and EOMES), 11 histone modifiers (ARID3B, CHD3, EP400, FUBP1, GATAD2A, LBD1, MBD3, SMARCC2A, SUPT5H, TRIM33 and WDR5), as well as AMOT, which was found uniquely in the induced samples. AMOT is known to be involved in tight junction maintenance and migration in endothelial cells [142] (Table 3).

Figure 7

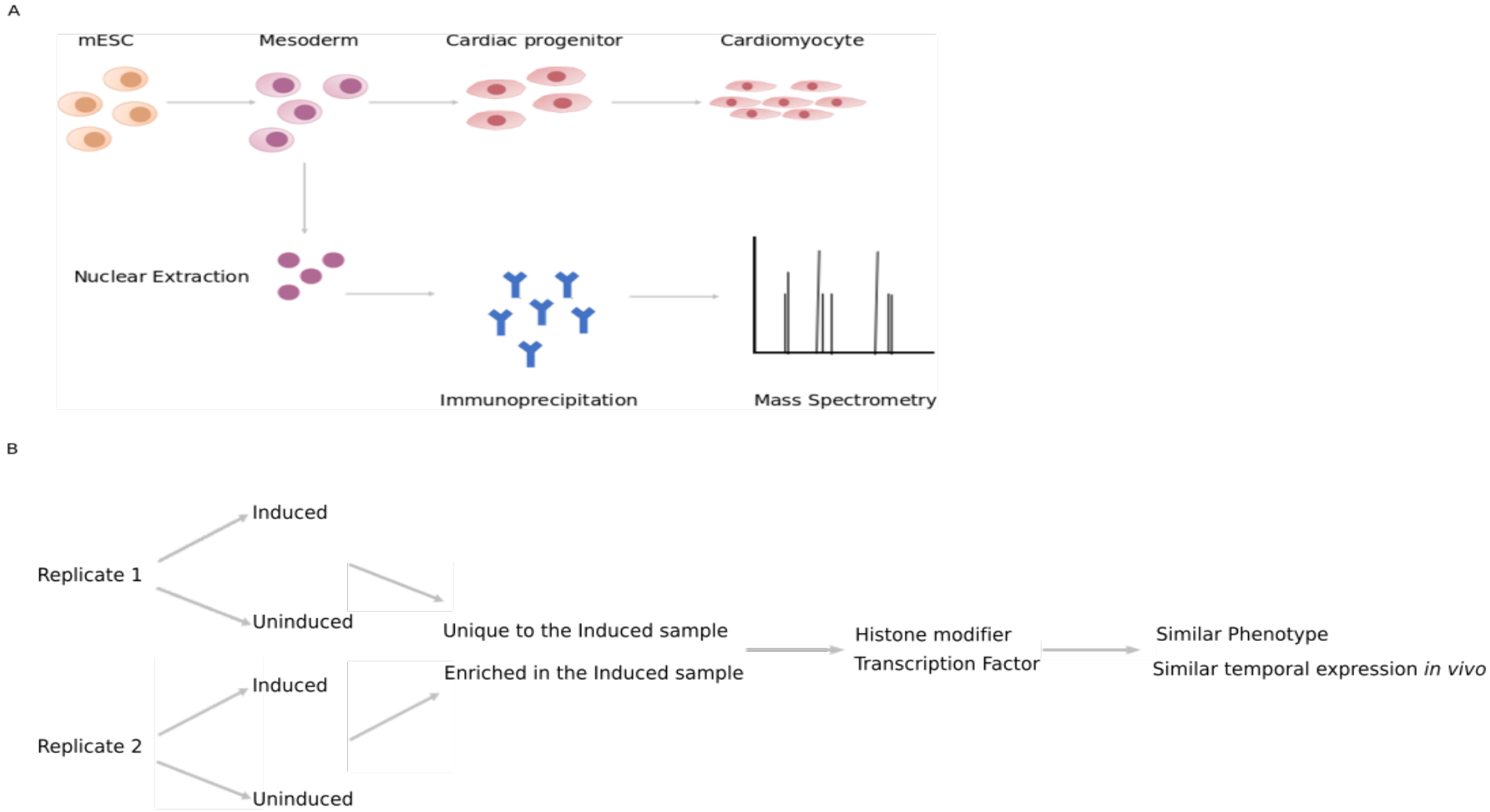



Figure 7: A schematic of RIME preparation and results analysis A) A schematic illustrating cells undergoing cardiac differentiation, with samples first taken at mesoderm stage and fixed, before nuclear extraction and Rapid immunoprecipitation mass spectrometry undertaken. Cells from the same differentiation, not fixed, were differentiated into cardiomyocytes to assess success of differentiation. B) A diagram clarifying RIME data analysis, with induced and uninduced from each replicate compared to identify proteins unique or enriched in the induced samples. These were further limited to those that were histone modifiers or transcription factors, before analysing *in vivo* spatiotemporal patterns and phenotypes.

Table 3

| Protein | Replicate 1 | | Replicate 2 | |
|---|---|---|---|---|
| | uninduced | induced | uninduced | induced |
| ***MESP1*** | 0 | **8** | 0 | **6** |
| ***AMOT*** | 0 | **6** | 0 | **6** |
| ***ARID1A*** | 6 | **12** | 7 | 7 |
| ***ARID3B*** | 5 | **10** | 6 | 7 |
| ***CHD3*** | 0 | **6** | 6 | 0 |
| ***EOMES*** | 9 | **15** | 5 | **18** |
| ***EP400 337KDA*** | 0 | **5** | 6 | 0 |
| ***EP400 37KDA*** | 0 | **6** | 6 | 0 |
| ***FUBP1*** | 0 | **5** | 5 | **18** |
| ***GATAD2A*** | 0 | **8** | 6 | 0 |
| ***LDB1 47KDA*** | 0 | **6** | 6 | 7 |
| ***LDB1 43KDA*** | 0 | **6** | 6 | 7 |
| ***MBD3*** | 5 | **11** | 8 | 7 |
| ***SALL4*** | 12 | **26** | 22 | 14 |
| ***SMARCC2A*** | 0 | **10** | 5 | 5 |
| ***SUPT5H*** | 0 | **6** | 9 | 0 |
| ***TRIM33*** | 0 | **5** | 6 | 0 |
| ***WDR5*** | 0 | **6** | 0 | 0 |

Table 3: Spectral counts of each protein identified by RIME. A table of spectral counts identified by rapid immunoprecipitation mass spectrometry (RIME) found in both replicate 1 and 2, in the induced and uninduced samples after being normalised to the IgG control.

Table 4

| | Mouse | | Zebrafish | | Human | |
|---|---|---|---|---|---|---|
| **Gene** | **Expression** | **Phenotype** | **Expression** | **Phenotype** | **Expression** | **Phenotype** |
| ***Mesp1*** | Primitive streak E6.5-E7.5[89] | Lethal E10.5 [98] | Margin Presumptive paraxial mesoderm [143] | Cardia bifida [144] | Heart and brain[145] | Cardiac defects with SNPs[146] |
| ***Amot*** | Mesoderm E6.5-E7.5 [147] | Embryonically lethal E7.5 (50% Embryos)[148] | Anterior axial hypoblast[149] | Dorsal aorta abnormalities[148] | Adult heart[145] | Cancer[150] |
| ***Arid1a*** | Embryo mesenchyme E9-15[151] | Lethal E9.5-13.5 arrested cardiac development[152] | Not Described | Not Described | Low tissue specificity [145] | Coffin-Siris syndrome 2[153] |
| ***Arid3b*** | Head and limb mesenchyme E11-12.25[154] | Lethal E12.5 arrested heart looping dilated heart[154] | Not Described | Not Described | Bone marrow, placenta, testis, blood, heart muscle[145] | Present in the nucleus of some testicular cancers[155] |
| ***Chd3*** | Embryo conceptus ubiquitous E3.5-5 Alimentary, urinary, reproductive systems E13-15[156] | Hyperactivity Decreased bone mineral content [157] | Whole organism[158] | Decreased brain size Cardiac oedema[159] | Foetal heart[145] | Dermatomyositis.[160] developmental delays, macrocephaly, impaired speech [161] |
| ***Eomes*** | Embryo mesoderm E6.25-E8[162] | Abnormal primitive streak[163] | Enveloping layer Blastoderm[164] | abnormal gastrulation[164] | Blood, brain, lymphoid tissue [145] | Microcephaly-polymicrogyria[165], favorable prognosis marker in renal cancer [166] |
| ***EP400*** | Floor plate, mesoderm E8.5[167] | Prenatally lethal Embryonic Arrest[167] | Not Described | Not Described | Low tissue specificity – found in adult tissues [145] | Negative prognosis marker in liver cancer[168] |

| | | | | | | |
|---|---|---|---|---|---|---|
| ***Fubp1*** | cardiovascular system/ heart. E12.5- 14, E16 Embryo mesenchyme E10-12.5[169] | Cardiac hypertrophy Hemorrhage, Decreased fetal size[169] | Not Described | Not Described | Low tissue specificity – found in adult tissues[145] | Negative prognosis marker in liver cancer[170] |
| ***Mbd3*** | Extraembryonic component E6.25 to E7.25[171] | E8.5 embryos are severely retarded[171] | Whole organism[149,172] | Heart left right asymmetry disrupted[173] | Fetal left ventricle, Low tissue specificity – found in adult tissues[145] | Negative prognosis marker in liver cancer[174] |
| ***Sall4*** | Embryo mesenchyme E8.5-E9.5[175] | Septal defects Thin and disorganized myocardium[175] | Whole organism[149,176,177] | Decreased Pou5f[177] | Skeletal muscle, parathyroid gland, testis, thyroid gland[145] | RNA enriched in testicular cancers Okihiro disease [178] |
| ***Supt5h*** | Embryo mesenchyme E10.5-E11.5[179,180] | Lethal Pre-weaning[181,182] | Whole organism[183] | Decreased blood and macrophages[184] | Heart - adult and foetus Low tissue specificity – found in adult tissues[145] | Negative prognosis marker in liver cancer[185] |
| ***Trim33*** | Ubiquitous E10.5 [186–188] | No primitive streak[189] | Whole organism[190,191] | Not Described | Low tissue specificity – found in adult tissues[145] | Favourable prognosis in colorectal, head and neck cancers[185] |
| ***Wdr5*** | Muscle skeletal system E15[192] | Not Described | Whole organism[193] | Not Described | Low tissue specificity – found in adult tissues[145] | Not Described |

Table 4 Identification of gene temporal and spatial expression and mutation effect *in vivo*. A table of temporal and spatial expression of each gene identified by RIME and the effect of its mutation on development. Each gene was investigated in mouse, zebrafish and human using MGI, Zfin and the Human protein atlas, expression in zebrafish limited from sphere to before segmentation in alignment with *Mespaa* expression

### 3.1.2 Identifying potential binding partners with similar temporal expression in mESCs

In our cardiomyocyte differentiation protocol, endogenous *Mesp1* expression peaks when mESCs transition into mesoderm stage, at approximately 88 hours of differentiation, post-plating in embryoid body form (Figure 4). In order to interact with MESP1, the protein of the binding partner must be expressed a similar time point. It is anticipated that genes with a peak in expression at 88 hours would be more likely to bind than those with lower expression at this time point; an increase in transcript could correlate to an increase in protein levels at a similar time to that of MESP1. Due to its transient expression profile, genes must be expressed during this limited time frame to be capable of binding to MESP1. Therefore, it is expected that genes of a similar temporal expression pattern to *Mesp1* would be more likely to bind to MESP1 and direct cell fate.

To investigate the temporal expression of the potential binding partners, mESCs were differentiated towards a cardiac fate. RNA was isolated at each stage of the differentiation, ranging from 0-hour stem cells through to cardiomyocytes at 195 hours post differentiation and Reverse Transcription-qPCR (RT-qPCR) carried out. mRNA expression of each potential binding partner was elucidated over the time course.

RT-qPCR analysis was not possible for *Arid3b* and *Smarcc2a*, with expression levels undetectable in tests containing up to 20µg of cDNA from all differentiation stages.

For all remaining potential binding partners, genes were plotted in relation to day zero expression to identify those expressed at 88 hours (Figure 8). This allowed for the isolation of seven potential binding partners that have peaks in their expression at mesoderm stage, similar to *Mesp1* (Figure 9A-H). This list included the transcription factors *Sall4* and *Eomes*, as well the unique protein identified in the induced samples, *Amot*. It further contained the histone modifiers *Mbd3, Supt5h, Trim33,* and *Wdr5*.

The temporal expression of both *Eomes* and *Wdr5* peaked, in keeping with *Mesp1,* at 88 hours before reducing to prior levels (Figure 9). Interestingly, whilst the other binding partners saw a peak in expression at 88 hours, a second peak was also observed at 147 hours which denotes the transition between cardiac progenitor and cardiomyocyte stages (Figure 9).

Although similar temporal expression suggests an increased chance of interaction between 2 proteins, it does not confirm binding. Therefore, the 7 putative binding partners were assessed to ascertain if they were capable of protein-protein interaction with MESP1, though co-immunoprecipitation experiments.

Figure 8

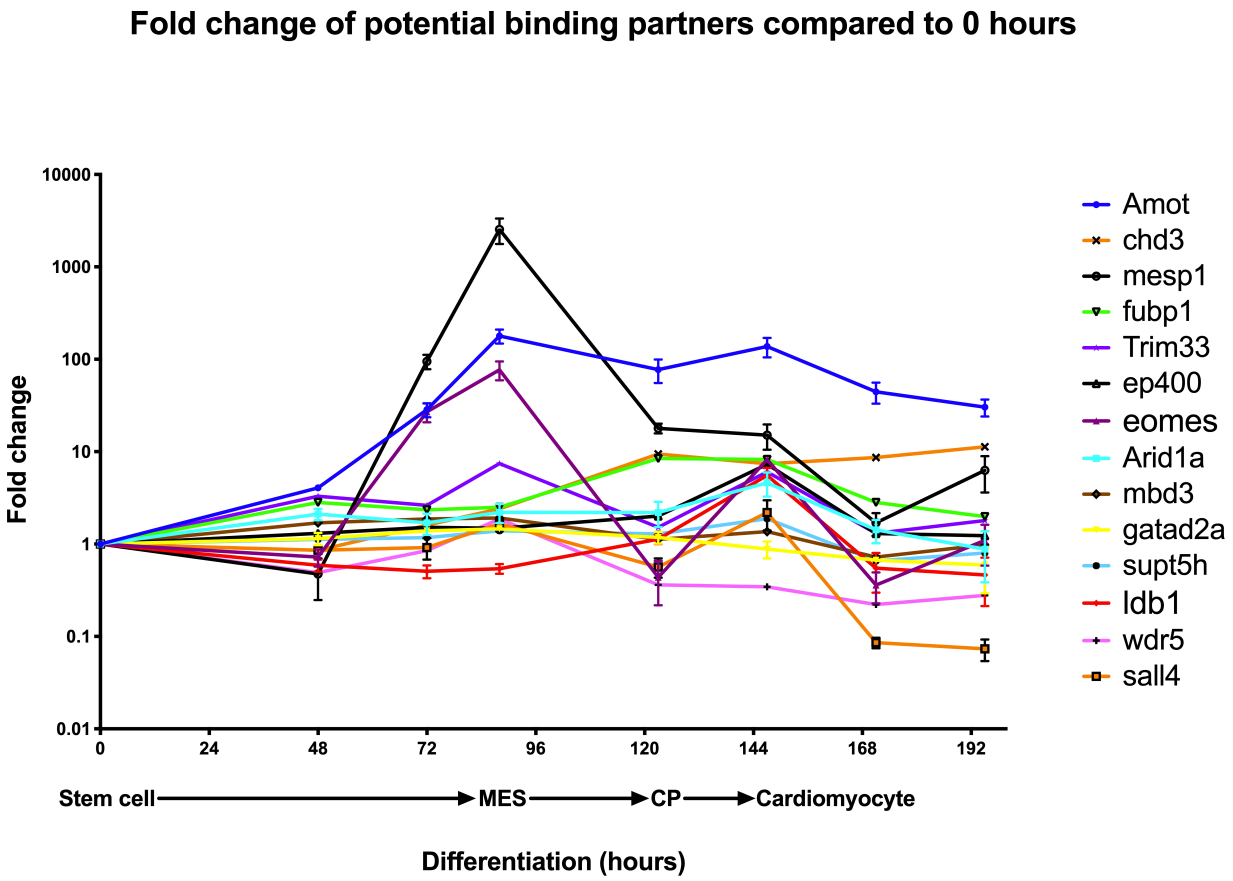



Figure 8 qPCR analysis shows seven genes with similar expression profiles to *Mesp1*. Temporal analysis of genes identified through RIME, over the cardiac differentiation time course, shows genes peaking at 88 hours, or MES stage of differentiation. Samples are from three differentiations, analysed using the ddCT method, with genes first normalised to a house keeping gene before normalisation to day zero sample.

Figure 9

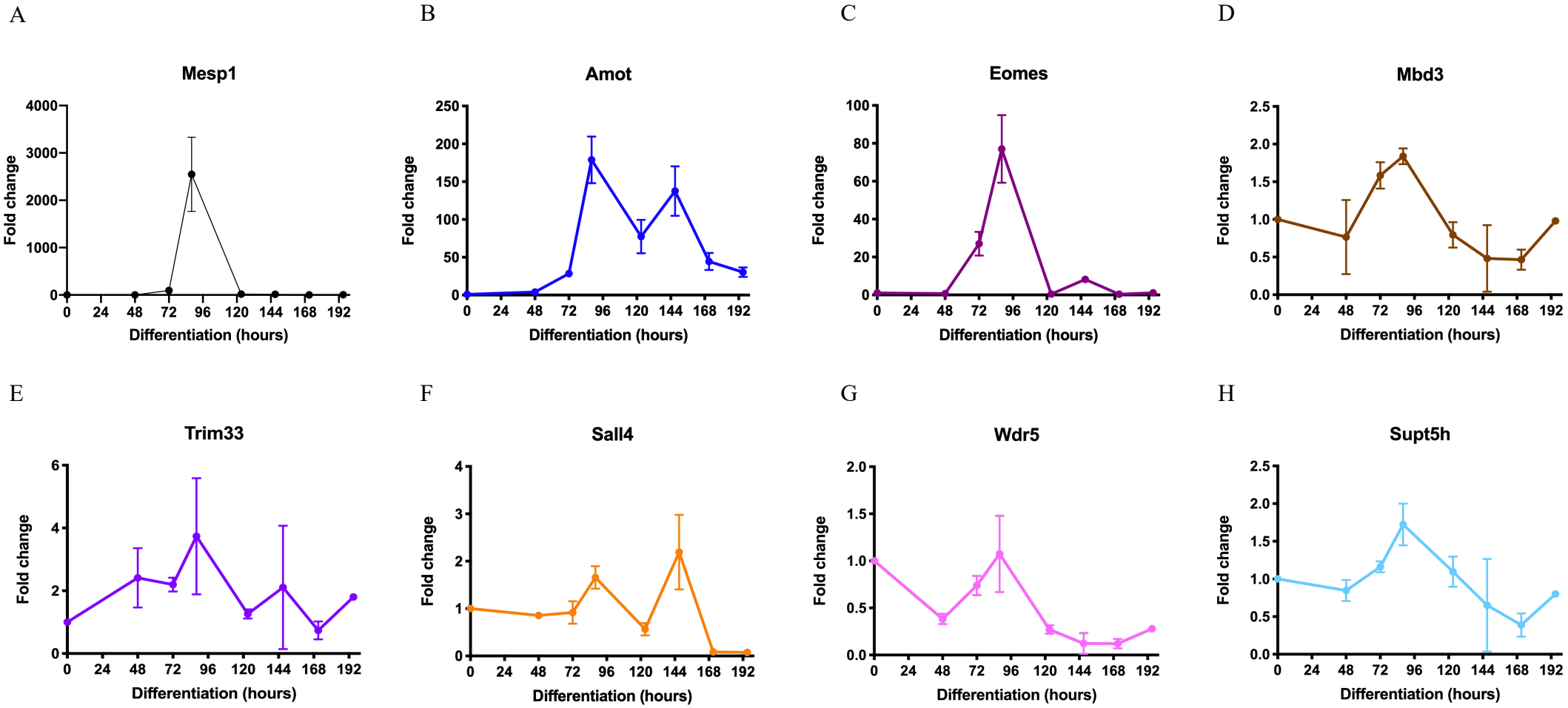



Figure 9A-H Graphs of potential binding partners with similar temporal patterns to *Mesp1*. Potential binding partners of MESP1 with similar temporal expression patterns, as denoted by a peak in expression at 88 hours of differentiation. Samples are from three differentiations, analysed using the ddCT method, with genes first normalised to a house keeping gene before normalisation to day zero sample.

### 3.1.3 Co-immunoprecipitation identifies two proteins capable of binding to MESP1 *in vitro*

To evaluate the ability of the potential binding partners to interact with MESP1, a co-immunoprecipitation and western blotting approach was undertaken.

Potential binding partners were individually cloned into FLAG tagged expression vectors, while *Mesp1* was cloned into a V5 tagged expression vector. Co-transfection of *V5-tagged Mesp1* and each FLAG tagged potential binding partner was optimised for maximum dual expression in HEK cells (data not shown).

Protein interaction was measured through immunoprecipitation of the FLAG-tagged binding partner and blotting for V5-MESP1. Both proteins should be visible in the whole cell lysate, whilst V5-MESP1 should only be found in the FLAG-tag immunoprecipitation (FLAG IP) lane if it interacts with the putative binding partner. An IgG control was included to control for non-specific binding, and the flow through samples should identify any unbound V5-MESP1 or FLAG-tagged binding partner.

### 3.1.4 Co-immunoprecipitation experiments reveal two putative binding partners

Blotting of the co-immunoprecipitation of FLAG-EOMES and V5-MESP1 identified both the 70kDa EOMES protein and the 28kDa MESP1 in the whole cell lysate, confirming transfection of the proteins in HEK cells (Figure 10A). FLAG IP successfully pulled down FLAG-EOMES with little residual protein identified left in the flow through (FT) lane. V5-MESP1 was also present in the immunoprecipitation lane, however some was also visible in the FT, indicating that not all V5-MESP1 bound to FLAG-EOMES in these conditions. Both proteins were also found in the IgG FT, and not in the IgG IP, indicating that binding between the proteins was specific. This result was replicated in all three repeats.

Co-immunoprecipitation experiments for FLAG-tagged MBD3 and V5-MESP1 proved technically challenging with a 4kDa size difference between the two proteins of interest. To be able to accurately blot for both proteins, a mix of HRP conjugated and fluorescently labelled secondary antibodies were necessary.

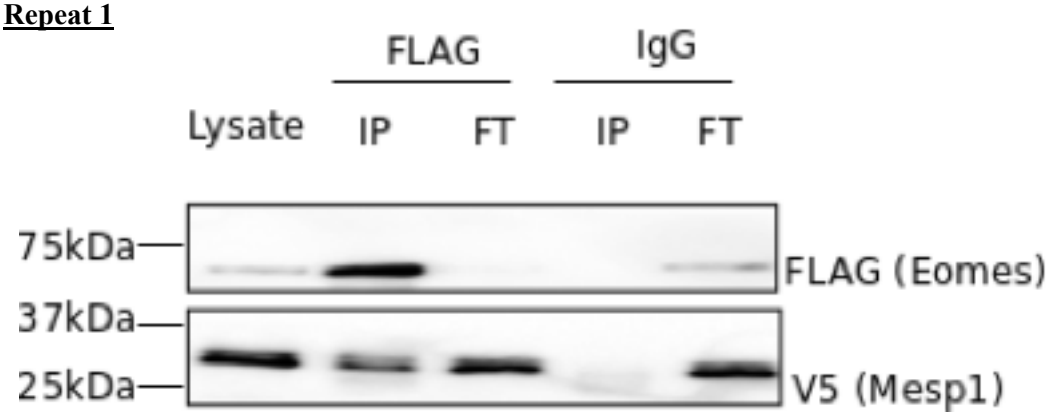
The whole cell lysate indicated both proteins were both transfected and pull down of FLAG tagged MBD3 was visible in the FLAG IP lane (Figure 10B). There appeared an equal amount of MBD3 in both the IP lane and the FT, suggesting not all MBD3 was successfully precipitated. V5-tagged MESP1 was present in the FLAG IP lane indicating that MBD3 and MESP1 do interact. Very little signal is seen in the IgG lane suggesting specific binding. This

pattern was observed when immunoblotting using HRP conjugated antibodies (repeat one and two) as well as fluorescently labelled secondary antibodies (repeat three).

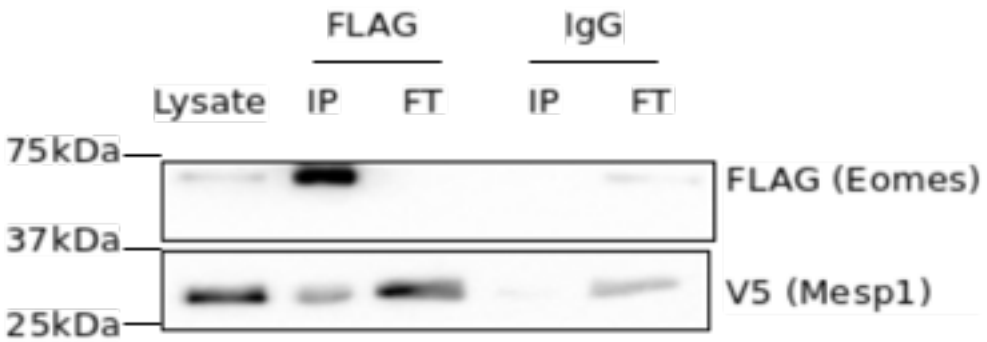
These results show that MESP1 is capable of binding to either EOMES or MBD3, either directly or in a complex with other factors in HEK cells not tested for in this experiment.

**Figure 10A**

**Repeat 1**




**Repeat 2**




**Repeat 3**

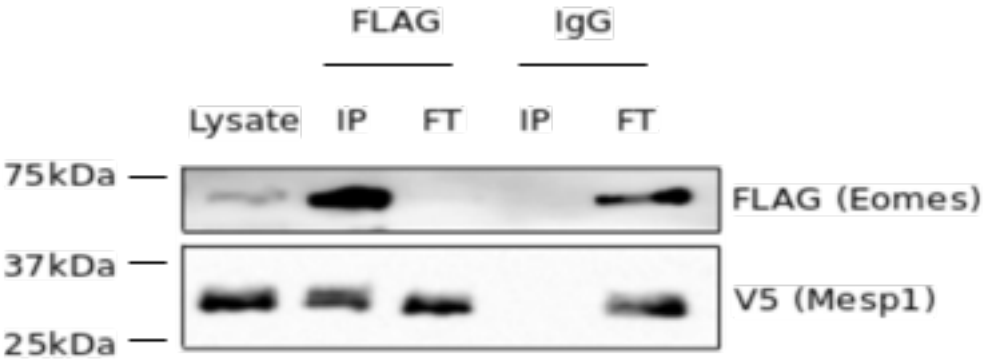
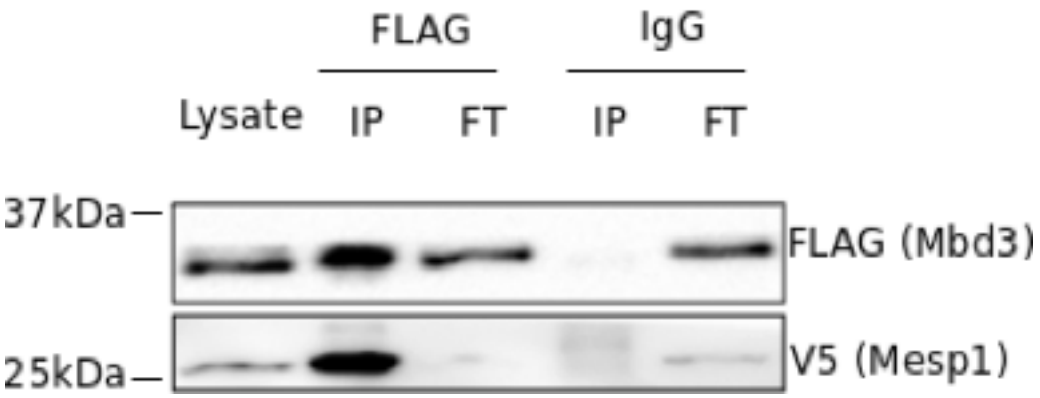
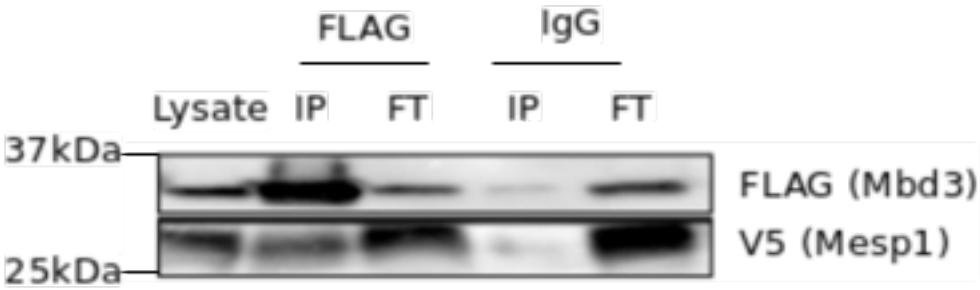


**Figure 10 B**

**Repeat 1**




**Repeat 2**




**Repeat 3**

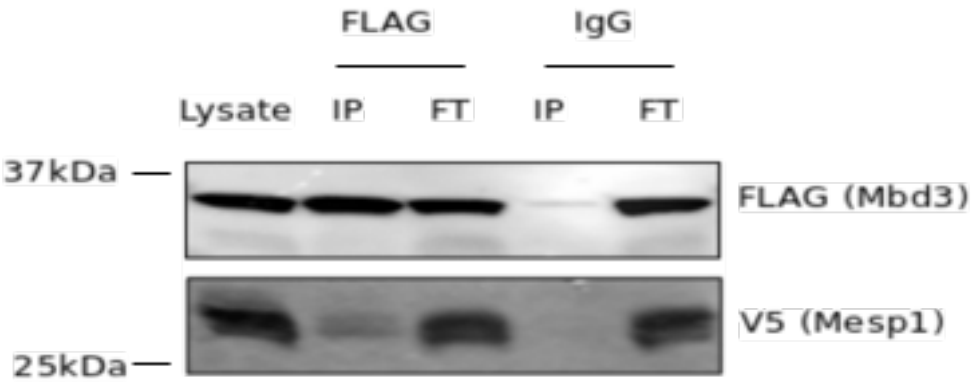



Figure 10 MESP1 interacts with EOMES and MBD3 A) Western blot of co-immunoprecipitation experiments between MESP1 and EOMES in HEK cells. FLAG-tagged EOMES was immunoprecipitated using the FLAG antibody (FLAG), with any protein not bound identified in flow through (FT) lanes. An IgG immunoprecipitation was used as a control (IgG) with flow through to identify non bound [proteins (FT). FLAG- tagged EOMES was blotted for using the FLAG tag antibody and V5-tagged MESP1 identified through blotting for the V5 tag. B) A western blot of coimmunoprecipitation experiments between MESP1 and MBD3. FLAG-tagged MBD3 was immunoprecipitated using the FLAG antibody, with an IgG control precipitation. Repeat one and 2 utilised HRP conjugated secondary antibodies, with repeat three V5-MESP1 identified through fluorescently conjugated secondary antibodies.

### 3.1.5 Co-immunoprecipitation experiments reveal four candidates that are indirect, or non-binding partners of MESP1

Analysis of the remaining putative binding partners, AMOT, SALL4, SUPT5H and TRIM33, revealed no binding between the FLAG-tagged binding partners and V5 tagged MESP1.

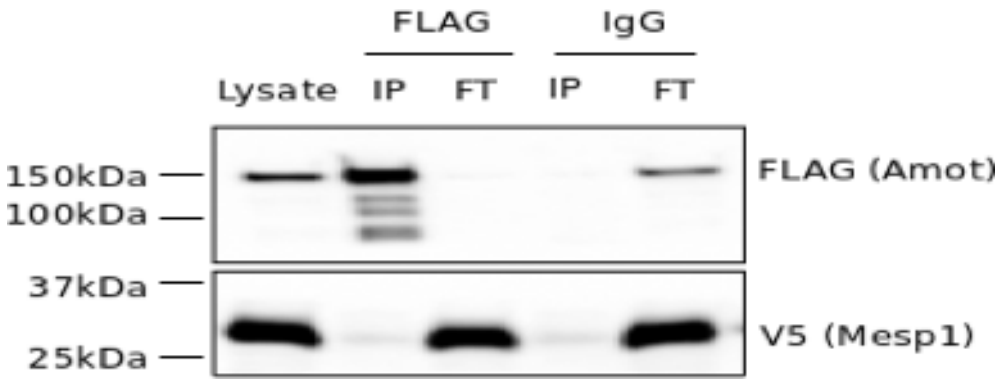
Results from AMOT immunoprecipitation, revealed no V5-tagged MESP1 in the IP lane, with it being clearly identified in the FT samples. There appears to be a small band where V5-tagged MESP1 would be found if successfully bound, however this is also apparently in the IgG control, indicating a non-specific binding event (Figure 11A). This pattern is also established for both SALL4 (Figure 11B) and TRIM33 (Figure 11C).

Immunoprecipitation of SUPT5H was found to be technically challenging with several repeats identifying V5-tagged MESP1 in the IgG control, with little to no FLAG-SUPT5H (Figure 11D). This was taken into account when analysing the results. Despite V5-MESP1 being present in the IP lane of two repeats (Figure 11D), there also appears to be an equal sized band in the IgG lane. The bands present also appear to be of a smaller weight than those indicative of V5-MESP1. Furthermore, no FLAG-tagged SUPT5H was present in the IgG samples in these two repeats suggesting a non-specific band. Therefore, it appears that SUPT5H does not bind to MESP1, when co-expressed in HEK cells.

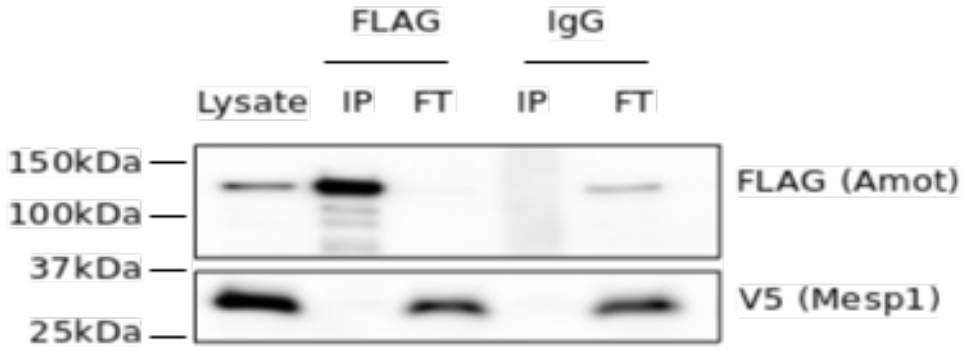
These results indicate that AMOT, SALL4, SUPT5H and TRIM33 are unable to directly bind to MESP1, or in a complex with MESP1 within the HEK cell environment. Further cell specific co-factors may be necessary to facilitate these interactions.

**Figure 11A**

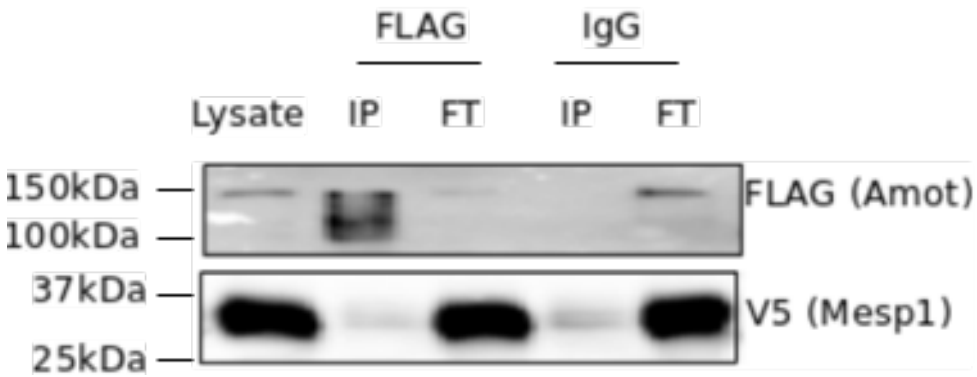
**Repeat 1**




**Repeat 2**

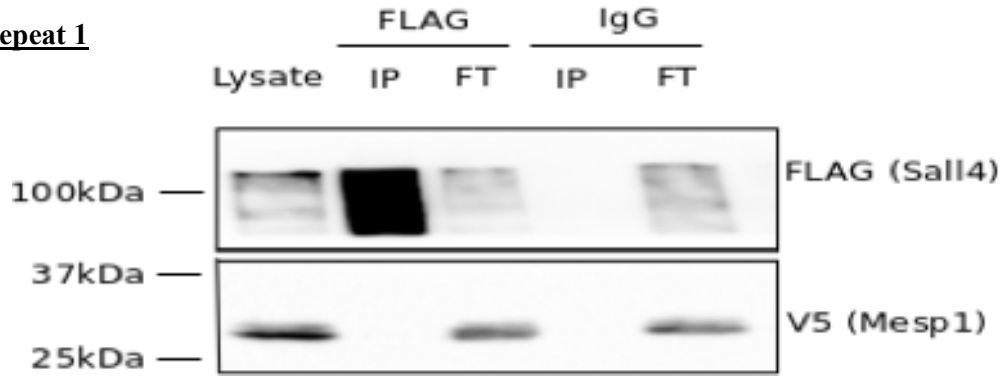



**Repeat 3**

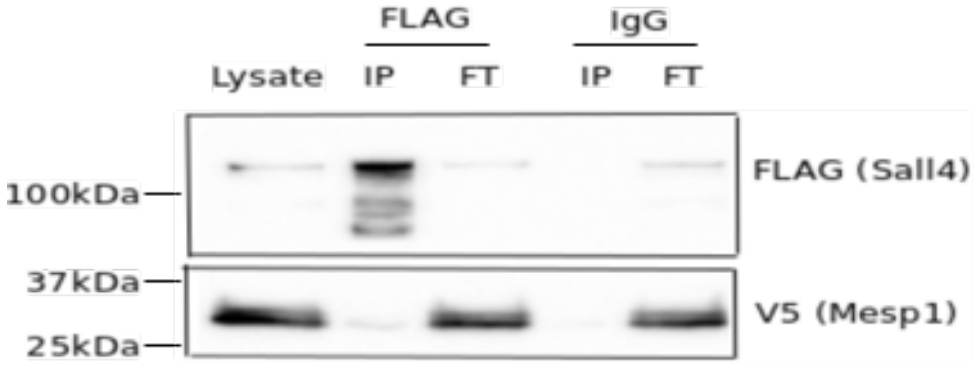


**B**

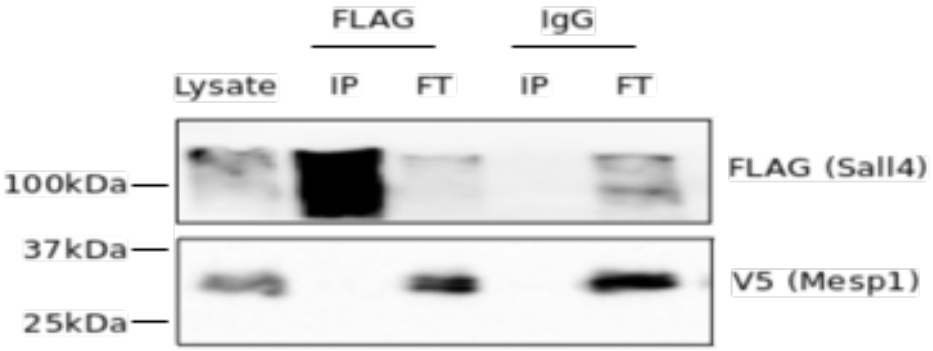
**Repeat 1**




**Repeat 2**

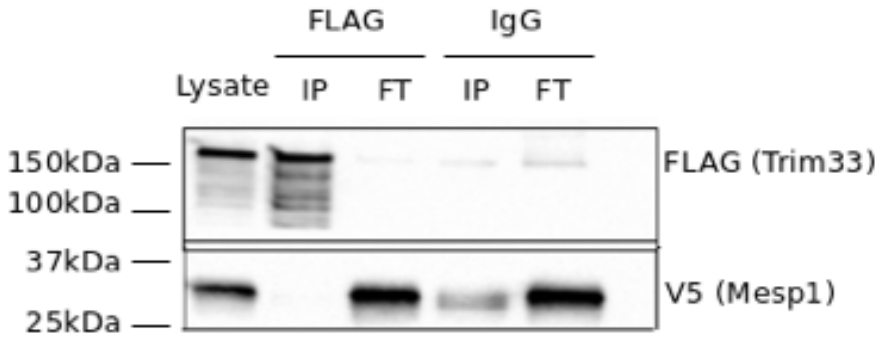



**Repeat 3**

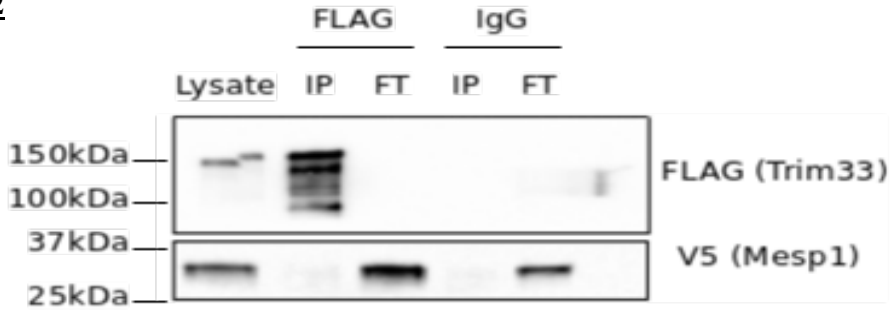


**C**

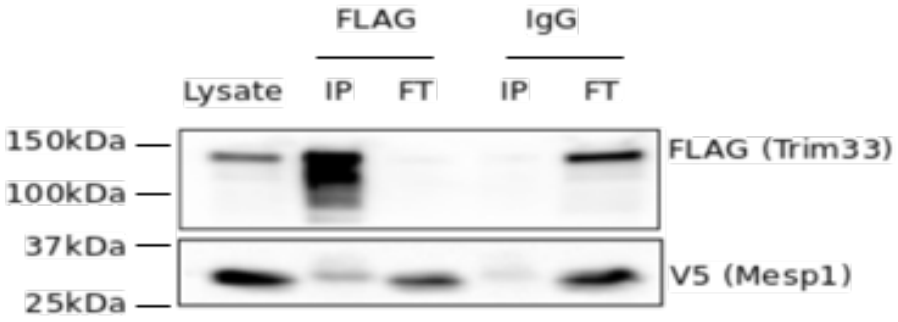
**Repeat 1**




**Repeat 2**

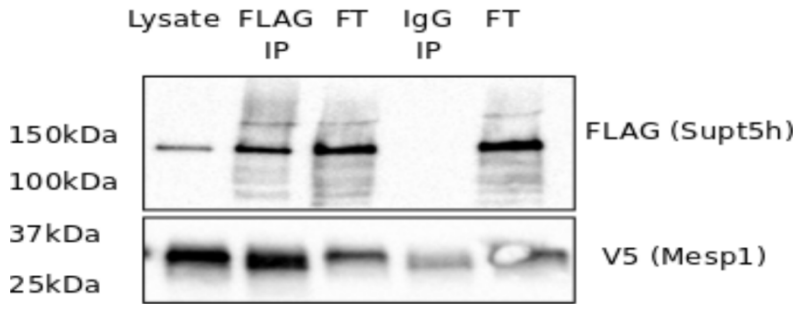



**Repeat 3**

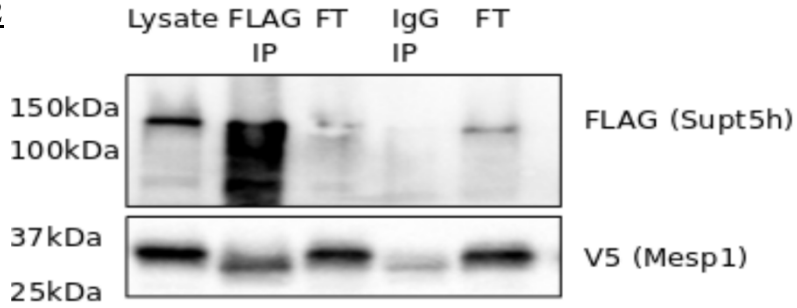


**D**

**Repeat 1**




**Repeat 2**




**Repeat 3**

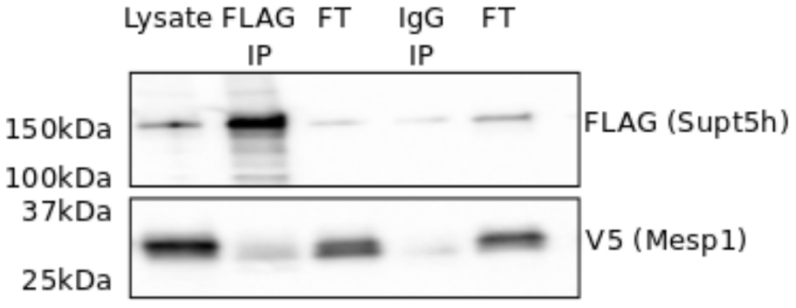



Figure 11A-D Western blot analysis shows no binding between MESP1 and AMOT, SALL4, TRIM33 or SUPT5H. Western blot analysis of co-immunoprecipitation experiments to identify binding between MESP1 and A)FLAG-tagged AMOT, B) FLAG-tagged SALL4 C) FLAG-tagged TRIM33 or D)FLAG-SUPT5H.A) The protein of interest was immunoprecipitated by FLAG tag antibody (FLAG) and blotted for bound proteins, any proteins not bound to FLAG-AMOT, FLAG-SALL4, FLAG-SUPT5H, FLAG-TRIM33 are identified in flow through (FT) lanes. An IgG immunoprecipitation was used as a specificity control. Three repeats were performed from three separate co-transfections.

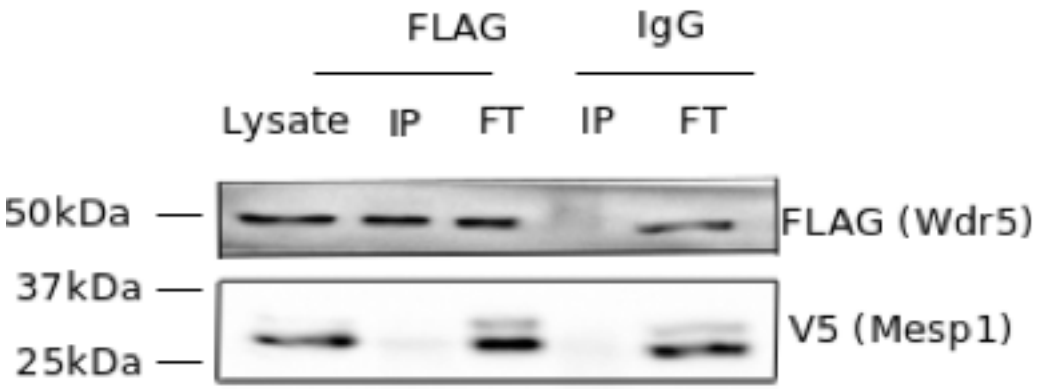
### 3.1.6 Literature reveals a potential indirect binding partner, that is capable of binding to an identified MESP1 binding partner

Blotting for MESP1 binding partners shows those capable of direct binding, however, would exclude those binding partners that are part of a complex. Direct immunoprecipitation of FLAG-tagged WDR5 indicates no direct binding to V5-MESP1, with MESP1 found solely in the FT lane (Figure 12). Despite this, literature provides direct evidence that WDR5 is able to bind with EOMES, in a complex with another histone modifier GCN5, which is mediated by a Large intergenic non-coding RNA (LINC)[124]. In addition to binding with EOMES, this complex is also pivotal to modulating *Mesp1* expression, binding at the *Mesp1* enhancer region, and facilitating transcription [124].

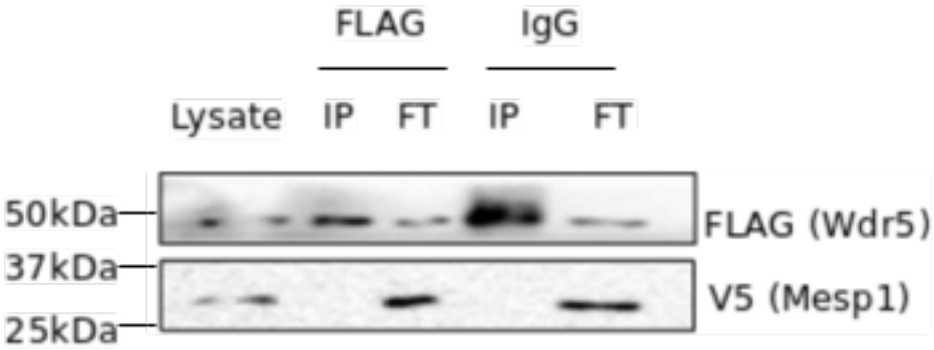
This provides evidence that WDR5 could be in a complex with EOMES, which can directly bind to MESP1.

**Figure 12**

**Repeat 1**




**Repeat 2**




**Repeat 3**

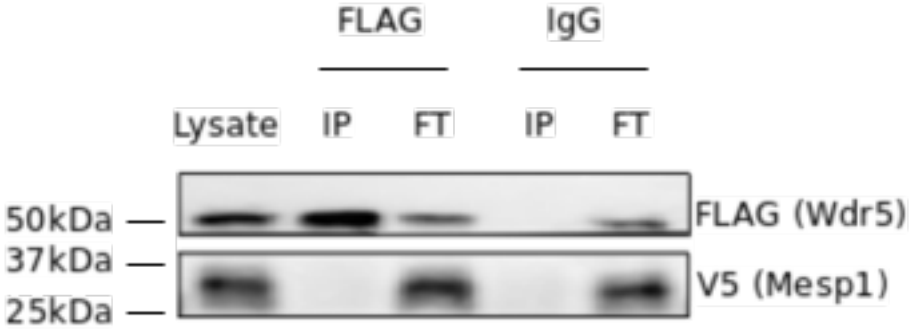



Figure 12 Western blot analysis indicates no direct binding between MESP1 and WDR5. Western blot analysis of co-immunoprecipitation experiments to identify binding between WDR5 and MESP1. FLAG -tagged WDR5 was immunoprecipitated by FLAG tag antibody (FLAG) and blotted for bound proteins, any proteins not bound to FLAG-WDR5 are identified in flow through (FT) lanes. An IgG immunoprecipitation was used as a specificity control. Three repeats were performed from three separate co-transfections.

### 3.1.7 Single cell RNA-seq identifies overlapping expression between binding partners and *Mesp1* at key developmental stages

In addition to *in vitro* binding, it is important to determine the likelihood that MESP1 and its potential binding partners are expressed in similar parts of the embryo at the same time point. Despite *in vitro* interaction, without an overlap in *in vivo* expression, it is likely these proteins would not be present in the same cells at a similar time and therefore could not interact to determine cell fate. It is of particular importance to observe this expression overlap in cells that will go on to form the heart, in which we anticipate MESP1 is a key driver of cardiac development.

To evaluate the *in vivo* likelihood of *Mesp1* and the putative binding partner interacting, published single cell RNA-seq data from early mouse embryos was analysed to identify if *Mesp1* and potential binding partners were found in the same spatiotemporal locations. The expression of *Eomes* and *Mbd3*, putative MESP1-binding partners that were found to bind in HEK cell lysates, and *Wdr5*, which may indirectly interact with MESP1 as part of a complex, were analysed over 3 time points across the peak expression of *Mesp1* in mouse embryos. It should be noted that whilst informative, the single cell RNA sequencing performed was limited in read depth to just 10,000 reads. This technique can fail to identify genes expressed at low levels, especially those that are transient in nature [194]. Furthermore, whilst it can identify if gene transcripts are present in the same cell, it cannot determine if the proteins are translated, or interact.

The expression of *Mesp1* is initiated at E6.5, at the onset of gastrulation. *Mesp1* is restricted to a subset of cells in the primitive streak (Figure 13B) with limited expression in parts of the epiblast [122,195]. *Eomes*, conversely, is broader in its expression, overlapping with *Mesp1* in the cells of the primitive streak (Figure 13C), as well as being expressed in cells of visceral endoderm and the epiblast. This is in keeping with its known role in endodermal formation and specification [125].

A similar overlap in expression of *Mesp1* and *Eomes* is also observed at E7.0 and E7.5 (Figure 14B, Figure 14C, Figure 15B and Figure 15C respectively). *Mesp1* is more restricted, being expressed highly in the primitive streak and nascent mesoderm at E7.0. Similarly, *Eomes* expression is found predominately in the primitive streak and nascent mesoderm, although it also has strong expression in the epiblast, extraembryonic ectoderm and visceral endoderm. By E7.5 *Mesp1* expression is spatially restricted at E7.5 to mesodermal lineages, in particular the nascent mesoderm. *Mesp1* expression is also evident in the pharyngeal and paraxial mesoderm as well as in the primordial germ cells. *Eomes* is

restricted to extraembryonic lineages with high expression observed in the extraembryonic ectoderm and endoderm. However, despite deviations in spatial expression, both transcripts are still found within the nascent mesoderm and primitive streak, corresponding with the majority of *Mesp1* expression.

In contrast to the restricted expression of *Mesp1* and *Eomes,* expression of *Wdr5* and *Mbd3* is diffuse throughout the embryos at each of the 3 time points. Expression of *Mbd3* is higher than that of *Wdr5,* however both transcripts overlap with *Mesp1* expression throughout the 3 time points (Figure 13D, Figure 13E, Figure 14D, Figure 14E, Figure 15D, Figure 15E).

This single cell RNA sequencing provides additional evidence that the interactions witnessed *in vitro* may be possible *in vivo.*

### 3.1.8 Summary

Rapid immunoprecipitation mass spectrometry, and literature analysis of expression and phenotypic data in embryos, resulted in the identification of 14 putative binding partners of MESP1 at mesoderm stage in mESCs. Further analysis using qPCR identified 7 candidates with similar temporal expression as *Mesp1* in mESCs at mesoderm stage, including *Amot*, two transcription factors, *Eomes and Sall4*, and four histone modifiers *Mdb3, Supt5h, Trim33* and *Wdr5*.

Whilst temporal expression was important to identify whether it was probable that putative binding partners could interact at mesoderm stage, co-immunoprecipitation experiments further clarified if they were capable of interacting with MESP1 when co-expressed *in* vitro. Immunoprecipitation of FLAG tagged EOMES and MBD3 identified them as capable of binding to Mesp1 *in vitro*, whilst the five other candidates did not appear to bind. Literature review suggested that WDR5 can form a complex with EOMES, and other histone modifiers to modulate gene expression [124], suggesting that WDR5 might indirectly interact with MESP1 via EOMES. Therefore, the *in vivo* expression of *Wdr5* alongside *Eomes* and *Mbd3* was analysed in recently published single cell RNA- seq data [196] to confirm expression overlap with *Mesp1* within the early embryo. Corresponding expression was identified in mesodermal lineages. *Mesp1* expressing cells are known to contribute to the developing heart and craniofacial muscles, as well as to haematopoietic stem lineages and cells that form the embryonic liver [94,135]. Cells that go on to form the rudimentary heart originate from *Mesp1* positive cells of the primitive streak, that egress and develop into the cardiac crescent.

The roles of *Mesp1* and *Eomes* in early embryo patterning and heart formation have been well examined, however there is little knowledge about the contribution of *Wdr5* and *Mbd3*. Understanding the roles of these two proteins in cardiogenesis in conjunction with MESP1 may be essential in future cell reprogramming effort.

Figure 13

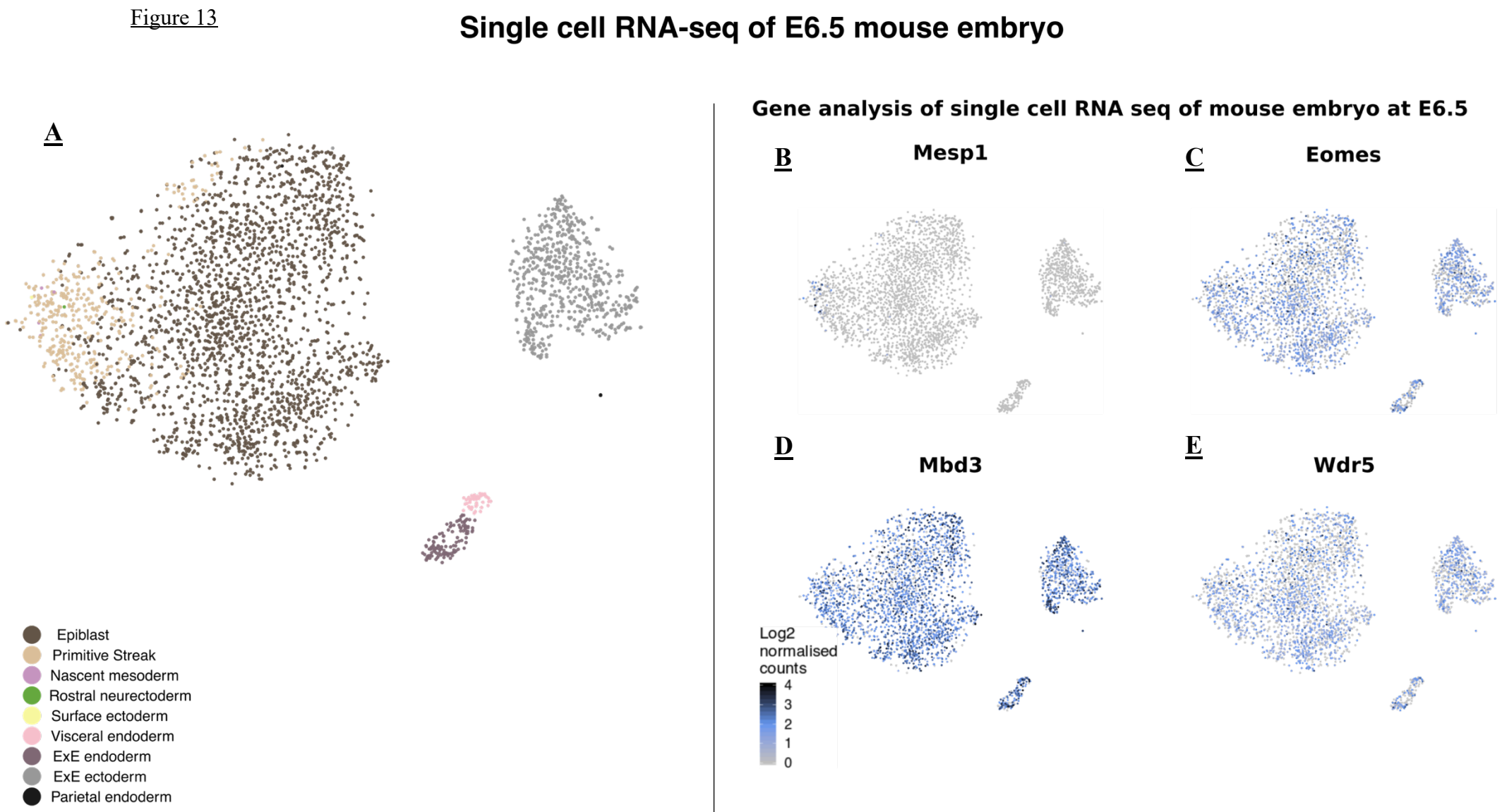



Figure 13 Single cell RNA seq analysis of E6.5 mouse embryos illustrates *in vivo* expression of *Mesp1* and binding partners. Single cell RNA seq of E6.5 mouse embryo indicates gene expression pattern of B) *Mesp1* C) *Eomes* D) *Mbd3* and E) *Wdr5* and each cell type throughout the early embryo (A). Analysis performed from published data and analysed on https://marionilab.cruk.cam.ac.uk/MouseGastrulation2018/

Figure 14

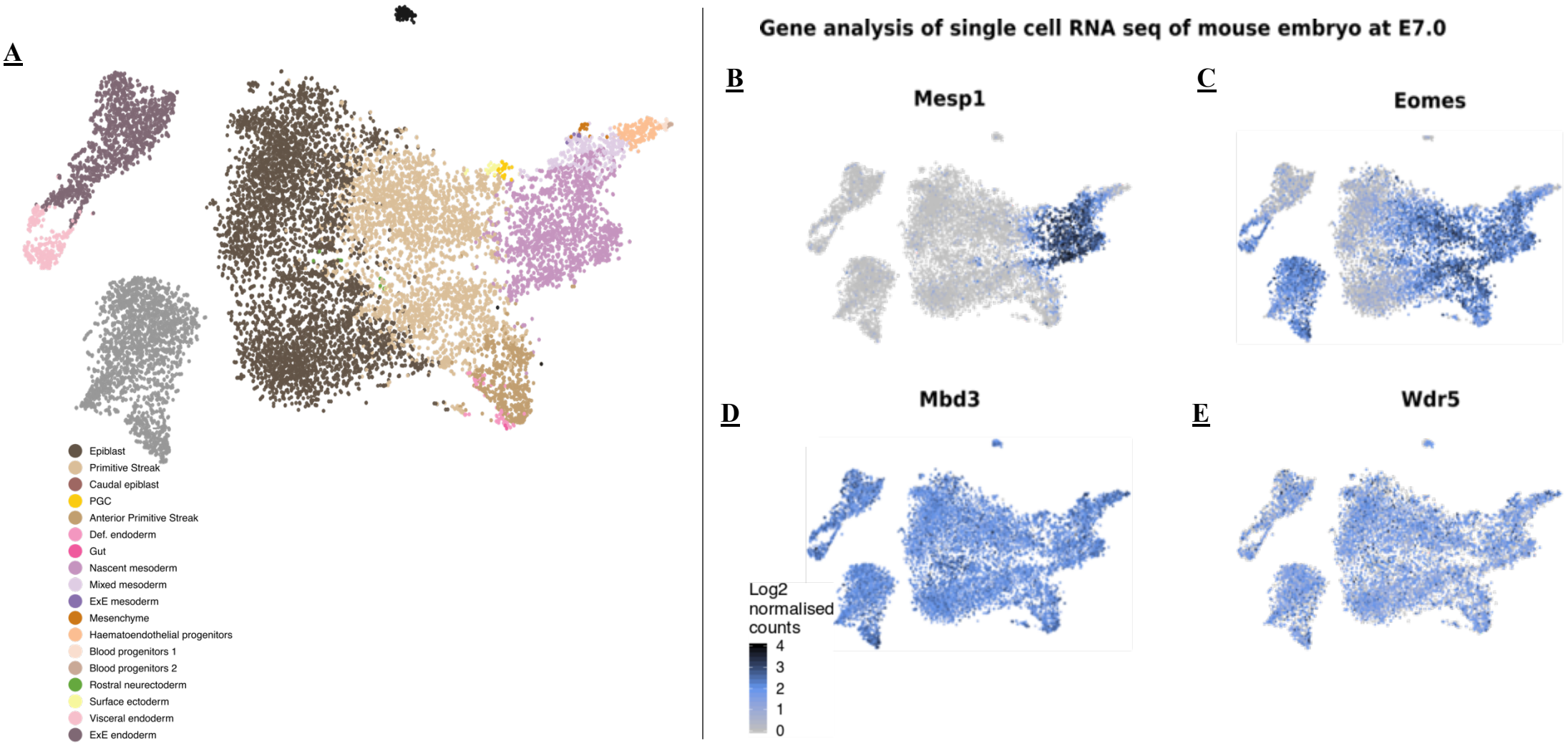



Figure 14 Single cell RNA seq analysis of E7.0 mouse embryos illustrates in vivo expression of *Mesp1* and binding partners. Single cell RNA seq of E7.0 mouse embryo indicates gene expression pattern of B) *Mesp1* C) *Eomes* D) *Mbd3* and E) *Wdr5* and each cell type throughout the early embryo (A). Analysis performed from published data and analysed on https://marionilab.cruk.cam.ac.uk/MouseGastrulation2018/

Figure 15

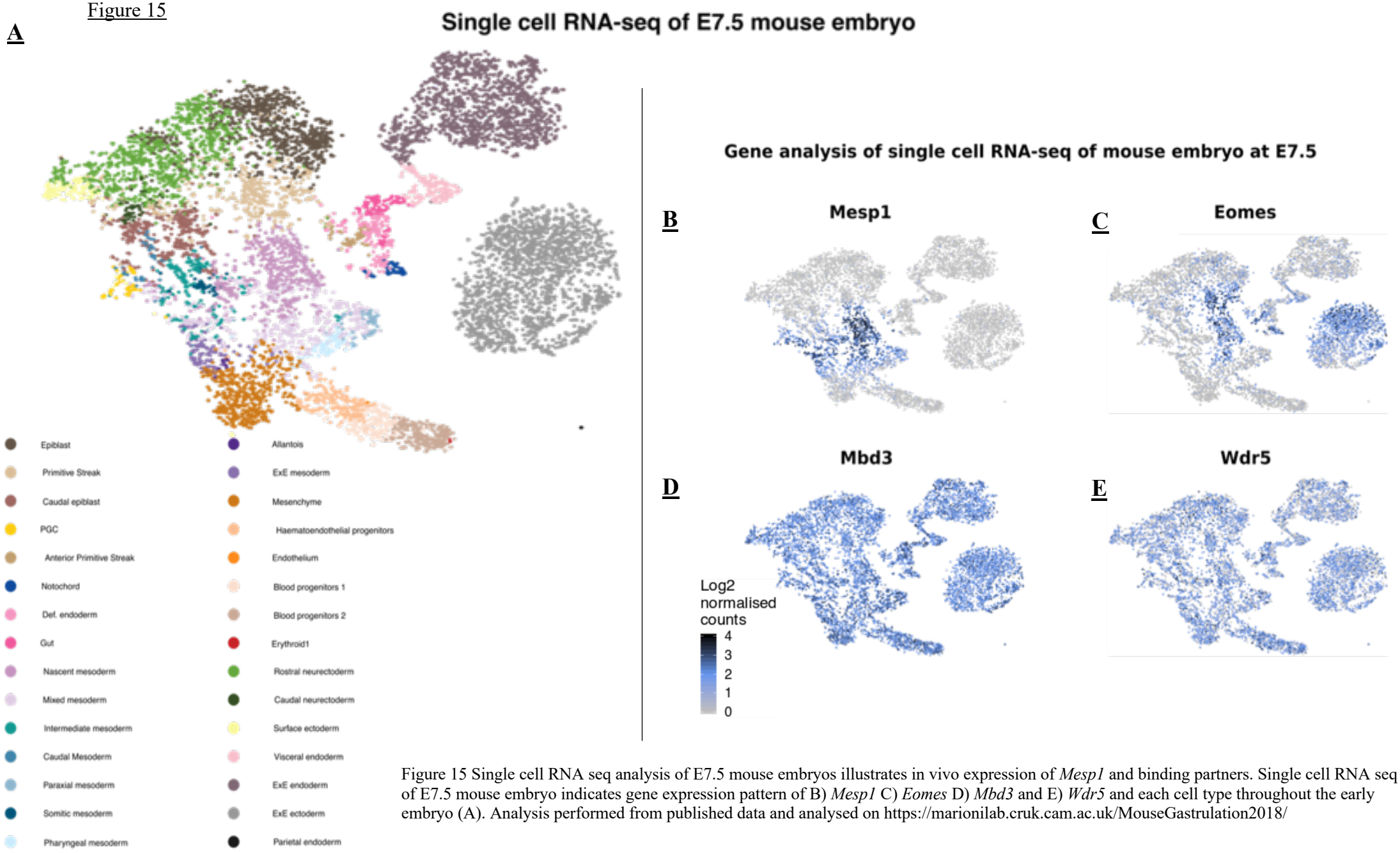



Figure 15 Single cell RNA seq analysis of E7.5 mouse embryos illustrates in vivo expression of *Mesp1* and binding partners. Single cell RNA seq of E7.5 mouse embryo indicates gene expression pattern of B) *Mesp1* C) *Eomes* D) *Mbd3* and E) *Wdr5* and each cell type throughout the early embryo (A). Analysis performed from published data and analysed on https://marionilab.cruk.cam.ac.uk/MouseGastrulation2018/

# 4 Chapter 4

## 4.1 Knockdown of *Mesp1* or its binding partners can affect expression of cardiac progenitor cell markers as well as expression of other binding partners

### 4.1.1 Introduction to stem cell cardiac differentiation

There are various practical limitations on tracking a developing embryo *in utero* in mammals, and the ethical implications of accessing human material often make it impossible to conduct experiments on embryos or foetuses. To compensate, experiments have been conducted on donated material from aborted embryos, as well as redundant preimplantation embryos from *in vitro* fertilisation procedures [197]. Whilst the mechanics of this developmental process have been described with detailed timelines of organ and system development, progress on delineating decision making in these cells was limited until recent single-cell RNA seq protocol developments [198]. For example, this has given us some insight into cell patterning decisions - why one mesoderm cell contributes to the craniofacial lineage, whilst another to the heart [97]. Genetic markers can give clues to which lineage the cell will undertake, with *Mesp1* positive cells indicating that a cell will go on to form a cardiac, craniofacial or haematopoietic anlage. However, further knowledge is needed to fully understand this process.

To further investigate this cell patterning, methods have been devised to investigate cellular changes at different time points, including single-cell RNA sequencing. Whilst this method is invaluable in tracking cells *in vivo*, it also involves increased animal usage. With an industry drive towards the principle of the 3Rs: reduce, replace and refine, other alternatives must be considered for these experiments.

One model which has proven essential in investigating these early cellular decisions is stem cells.

ESC modelling is essential to researching human developmental changes and estimating how they fit into the developmental paradigm. In addition to mapping developmental changes, the ability to adjust this model allows for its utilisation in mapping the effect of genetic changes on development. This is especially true of embryonically lethal genes, which account for approximately 25% of the mouse genome [199]. Studying the effects of these genes is limited to murine knockdown models. The use of ESCs allows either human or mouse genes to be deleted or mutated, and to track their role in differentiation to different lineages. Their effect on differentiation can be traced through genetic markers, or

the failure to form certain cellular structures, such as contracting myocytes. Downstream effects can be measured through known developmental gene hierarchies, assessing how one genetic mutation or deletion affects other genes. This process is exceptionally important in understanding patient mutations and resulting phenotypes.

Stem cells are also essential for understanding how each gene can affect a previously mapped process. Gene networks involved in transitions between the early embryo and a defined fate have been compiled using mouse and stem cell experiments, although are limited in size and depth of information.

It is important to know how the MESP1 binding partners identified in RIME and co-immunoprecipitation experiments feature in the dynamic transitions of stem cells towards a cardiac lineage, and what influence they have on other genetic markers. Often genes are characterised in knockout models, and therefore are absent from the stem cell stage. However, in this situation, their role at mesoderm stage, where they are present and capable of binding to MESP1, will be masked. Therefore, a knockdown model at mesoderm stage was devised to investigate the role of these genes on mesoderm and cardiac progenitor cell formation.

Aim: To investigate the effect each MESP1 binding partner has on cardiomyocyte directed differentiation of mESCs.

### 4.1.2 Embryoid body size affected by siRNA addition

To determine if each potential binding partner affected mesoderm formation during mESC differentiation, particularly the effect on morphology, a knockdown approach was undertaken. Knockdown of the gene at day 2 of differentiation allowed identification of the effect on mesoderm formation without affecting cell pluripotency. This is particularly important in the case of *Wdr5*, which is known to affect cell fate decisions in ESCs [200]. *Wdr5* is a downstream target of both OCT4 and NANOG and plays a key role in ESC self-renewal. Knockout of *Wdr5* in ESCs resulted in repression of ESC self-renewal and enhanced endoderm differentiation[200].

Overexpression of *Mesp1* in ESCs has previously been shown to transiently increase embryoid body (EB) size for up to 3 days post doxycycline induction of *Mesp1,* with EBs comprised of twice as many cells in comparison to wildtype cell EBs[110]. Furthermore, it has been reported that EB size could influence lineage-specific differentiation with

cardiogenesis enhanced in larger EBs (approximately 450 microns in diameter)[201,202]. Therefore, knockdown of *Mesp1*, or its potential binding partners, could affect a variation in EB size, which could reflect a change in the fate of cells during cardiac differentiation.

Each binding partner was knocked down singularly through a mix of 4 siRNAs for each gene. These were added to wildtype mESCs at day 2 of differentiation to ensure knockdown at mesoderm stage at day 4 of differentiation (Figure 4). The cells at day 2 of differentiation still express pluripotency markers *Nanog* and *Oct4.* However, these are decreased in comparison to day 0 due to the removal of Lif and iGSKβ and the addition of cytokines, including BMP4 and Activin A, to direct towards a cardiac lineage. Samples were isolated at day 4 and visually analysed.

Figure 16

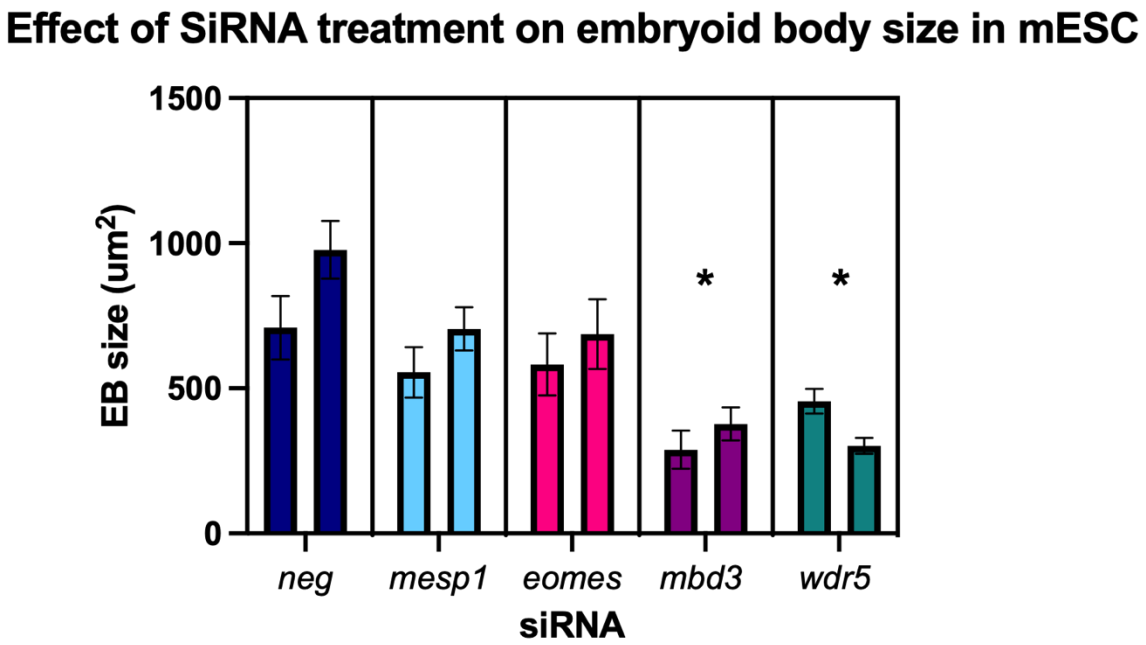



Figure 16: siRNA targeting *Mbd3* and *Wdr5* causes a reduction in embryoid body size. A graph of results from siRNA treatment of mESC treated at day two of differentiation, with embryoid body size measured at mesoderm stage on day four. Samples were taken from three wells from two differentiations with three images taken of different regions in each well. In total 405 embryoid bodies were analysed. Samples were analysed using nested one-way ANOVA with Dunnett's multiple comparison test compared to the negative control on GraphPad Prism. Comparison between repeats in each group $p<0.0567$ not significant. * $p<0.05$ compared to negative control samples. n=2 biological replicates

Whilst overexpression of *Mesp1* can induce a transient increase in EB size, the effect of knockdown was unknown. It was expected that a reduction in *Mesp1* may result in smaller

EB size. Knockdown of *Mesp1* and *Eomes,* two transcription factors known to be upregulated during mesoderm formation, did not show a statistically significant reduction in EB size (Figure 16). By contrast, knockdown of either histone modifier, *Mbd3* or *Wdr5*, resulted in a significant reduction in EB size, from an average of 800μm$^2$ to less than 400μm$^2$ (Figure 16) ($p<0.05$). This indicates these binding partners may affect the formation of mesoderm stage EBs, with EB size significantly reduced. EB formation mimics various stages of embryo development, including gastrulation, with the type and number of cell-cell contacts, size and number of cells involved in each sphere important in developmental processes [203]. With changes in EB size directly affecting cell lineage [202,204], these changes may directly correspond with changes in the cardiac differentiation process.

### 4.1.3 Knockdown of *Mesp1* results in a reduction in potential binding partner, mesodermal and cardiac progenitor cell marker expression

To establish if *Mesp1* knockdown results in changes to cardiac differentiation, siRNA directed at *Mesp1*, or a control siRNA, were added at day 2 of mESC cardiac differentiation to ensure maximal knockdown at mesoderm stage on day 4. Effects of knockdown were assessed through qPCR analysis of samples in 2 differentiations from day 3 to day 6, which represents the progression of mESCs undergoing cardiac differentiation from mesoderm stage to the start of cardiomyocyte formation. Inclusion of a further biological replicate was precluded due to the Covid-19 pandemic. Particular focus was given to mesodermal and cardiac progenitor markers, as well as the effect on the other potential binding partners. Day 3 samples from knockdown 1 were degraded due to mechanical failure of a -80˚C freezer and were not included in the analysis.

*Mesp1* knockdown was efficient, with transcripts reduced to between 10-20% of original levels during peak expression on day 4 of differentiation. Expression followed a similar pattern to control samples, peaking between day 3 and 4, before being switched off between day 4 and 5. No expression was observed post-day 5, at cardiac progenitor stage (Figure 17A).

To delineate the effect of *Mesp1* knockdown on the transcript levels of its potential binding partners, the temporal expression of each was analysed. EOMES is known to induce *Mesp1* expression, with peak expression typically found on day 4 of differentiation, coinciding with that of *Mesp1*. Expression of *Eomes* in knockdown samples was severely reduced, to between 15-40% of control cells on day 4. Temporal expression remained similar to that of control samples (Figure 17B).

*Mbd3* expression was similarly reduced in knockdown samples to between 20-50% of the levels found in control samples at day 4. Temporal transcript levels of *Mbd3* varied between replicates, however, this reduction was observed in both differentiations at day 4 (Figure 17C).

The expression profile of the final potential binding partner, *Wdr5*, varies between the 2 differentiations. Despite this deviation, expression of *Wdr5* is reduced at day 4 of differentiation, to approximately 50% of that found control samples (Figure 17D).

In order to assess the effect of *Mesp1* knockdown on mesoderm differentiation, *Brachyury (T)* expression was also analysed. As anticipated, expression of *T* peaked at day 4 of differentiation, associated with the cell's progression from a stem cell state to a

mesodermal lineage. This is observed in both replicates. At day 4, *Mesp1* knockdown resulted in a reduction of expression to 0.2 and 0.4 in differentiation 1 and 2 respectively. *T* expression is also greatly reduced in both knockdowns at day 5 and 6, with expression in differentiation 2 less than 25% of that seen in control samples, whilst expression was completely ablated in differentiation 1 (Figure 17E).

In addition to mesodermal markers, the temporal expression of cardiac progenitor cell markers *Gata4* and *Nkx2.5* were analysed to assess the effect of *Mesp1* knockdown on cardiac differentiation.

Typically, *Gata4* is expressed at the onset of cardiac progenitor cell stage, at day 5 of differentiation, as observed in control samples for both differentiations. Expression then decreases to baseline levels at day 6. In differentiation 2, transcripts were not quantifiable in *Mesp1* day 3 knockdown samples.

Transcript levels peaked at day 4 in knockdown samples from both differentiations. Differentiation 1 knockdown cells contained 3 times the level of *Gata4* compared to differentiation 1 control cells. In contrast, differentiation 2 knockdown cells expressed approximately half the amount of *Gata4* transcript in comparison to differentiation 2 control cells.

In contrast to control differentiations, both knockdown samples showed a decline in transcript expression at day 5 and day 6(Figure 17F).

Typically, in mESC differentiation, *Nkx2.5* regulation is tightly controlled, and is upregulated at day 5 of differentiation, before being switched off in early cardiomyocyte progression [205]. Control samples in this experiment also exhibit the same expression profile, with peak expression at day 5 of differentiation.

Knockdown samples show the highest expression of *Nkx2.5* at day 4. *Nkx2.5* expression is reduced by day 6 in knockdown samples.

Peak expression levels are equivalent between knockdown and control samples in differentiation 1, although peak expression is observed 24 hours earlier in knockdown samples. Whilst expression levels peak on day 5 for control samples, expression in differentiation 1 knockdown samples is approximately halved at this timepoint. A similar pattern is also observed in differentiation 2, with peak expression of *Nkx2.5* found on day 4 in comparison to control samples, where peak expression occurs on day 5(Figure 17G).

Knockdown of *Mesp1* results in the reduction of expression levels of each potential binding partner at day 4, as well as mesodermal marker *Brachyury.* In addition, a reduction in cardiac progenitor markers *Nkx2.5* and *Gata4* is observed at day 5 of differentiation, with a

temporal shift in peak expression to 24 hours earlier. It is known that EOMES and BRACHYURY are capable of inducing expression of *Mesp1*. However, these results indicate that MESP1 could also induce expression of these mesodermal markers. Peak expression of cardiac progenitor markers *Gata4* and *Nkx2.5* were also reduced on day 5. This result is not unexpected, with *Mesp1* known to work upstream of these factors to induce cardiac differentiation. Further replicates are needed to discern the effect of *Mesp1* knockdown on the temporal patterns and amplitude of expression of these cardiac progenitor markers, as well as that of *Wdr5,* which showed variations in expression between differentiation 1 and 2.

Figure 17

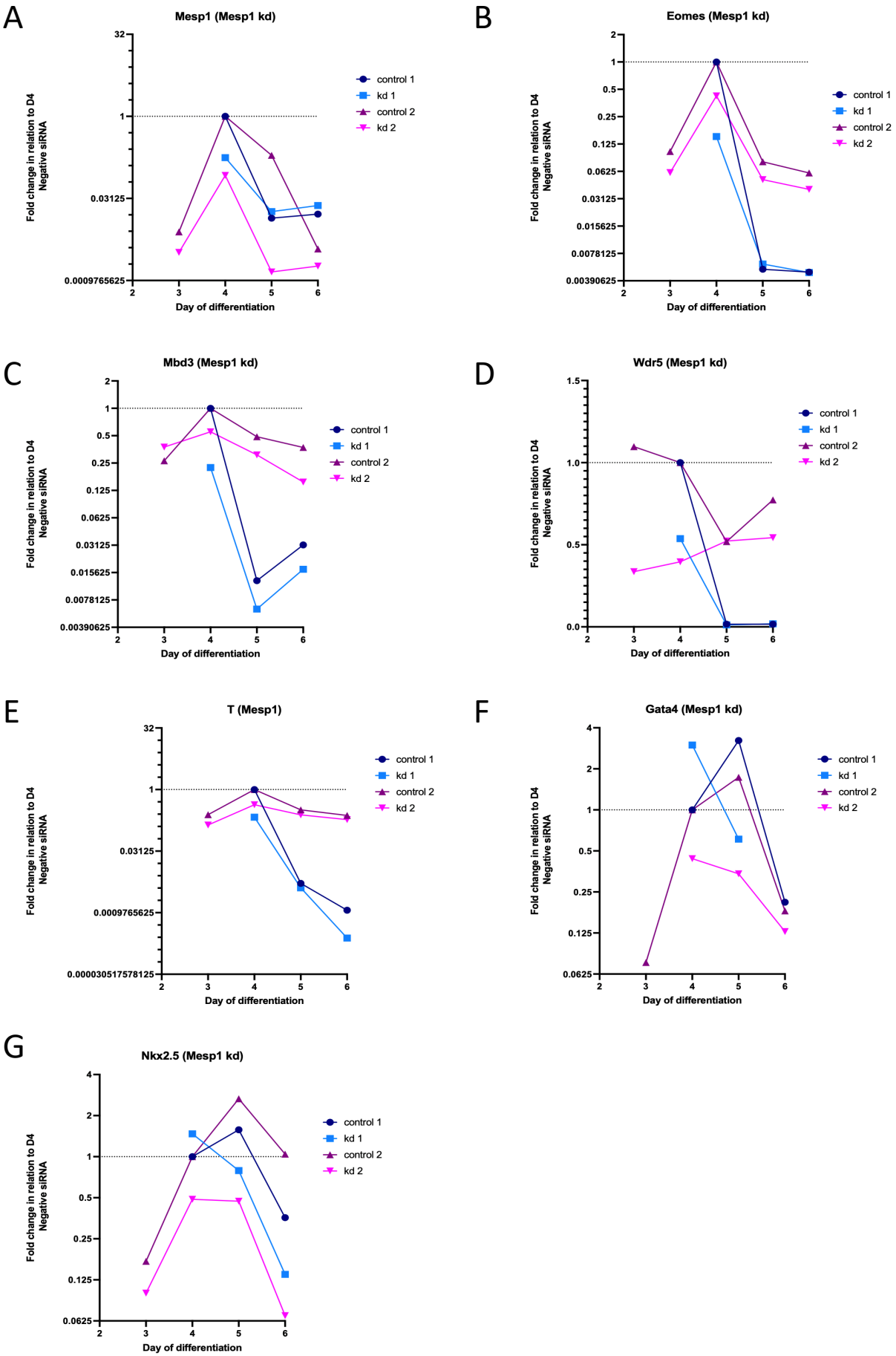



Figure 17A-G: Knockdown of *Mesp1* results in the reduction of potential binding partners (A-D), mesodermal marker *T* (E), and cardiac progenitor markers (F-G). Results from two different stem cell differentiations.

### 4.1.4 Knockdown of *Eomes* results in a decrease in expression of *Mbd3,* mesodermal markers and the cardiac progenitor marker *Nkx2.5*

In addition to considering the role of *Mesp1* in cardiac differentiation, it is also important to assess the role of its binding partners, in particular, to evaluate if knockdown of the binding partners could affect *Mesp1* expression and cardiac differentiation. This would aid in distinguishing the effect of the binding partners on mESC differentiation and if they are essential to producing cardiac progenitor cells.

*Eomes* is found upstream of *Mesp1* in the cardiac hierarchy of genes, with EOMES capable of inducing *Mesp1* expression [123]. *Eomes* itself is known to be essential in development, with ablation resulting in prenatal lethality at gastrulation [122]. Furthermore, knockout of *Eomes* in hESCs results in an inability to form cardiac cells [206]. Whilst this has been examined in mESCs, key markers for cardiac differentiation were not studied [125]. It is therefore critical to examine the role of *Eomes* knockdown on *Mesp1* expression, as well as how its expression can affect transcript levels of the other potential binding partners and cardiac differentiation markers.

Knockdown was undertaken in an identical manner to *Mesp1.* Typically, *Eomes* expression peaks at day 4 of differentiation, coinciding with mesoderm stage, as identified in control samples. *Eomes* knockdown shows a reduction in transcript levels to below 20% of that found in control samples at day 4 in both differentiations.

Temporal expression patterns of *Eomes* remain similar in both control and knockdown samples. These results indicate a successful knockdown of *Eomes* in mESCs at mesoderm stage of cardiac differentiation (Figure 18A).

The effect of knockdown on *Mesp1* and its other potential binding partners was also evaluated. Knockdown samples maintain the expected temporal *Mesp1* expression pattern, with peak expression observed at day 4 of differentiation, however, *Mesp1* levels were severely reduced to 20-30% of the transcript levels seen in control samples at the same time point, confirming the known induction of *Mesp1* by EOMES (Figure 18B).

There is no known link between *Eomes* expression and *Mbd3. Mbd3* expression in both control samples peaks at mesoderm stage on day 4. This trend is true of the knockdown samples, which also peak at day 4, although at 30% of the control level.

The pattern of *Mbd3* expression in differentiation 2 control and knockdown samples is similar, with a reduction from peak expression at day 4, whilst remaining steady at days 5 and 6. Control levels remain twice those seen in knockdown samples. Expression levels in

differentiation 2 samples are approximately twice those found in differentiation 1 samples at days 5 and 6 (Figure 18C).

The effect of *Eomes* on *Wdr5* is also an important variable to investigate, as a complex involving EOMES and WDR5 is known to act upstream of *Mesp1*[124]. Results of the 2 differentiations show conflicting trends. Knockdown in differentiation 1 results in peak transcript levels at day 4, at 1.7-fold of control samples, while expression is reduced to near 0 at day 5 and 6 in both control and knockdown samples. By contrast, differentiation 2 samples on day 4 exhibit approximately half the expression of that found in control samples. *Wdr5* transcript levels are reduced by half on day 5 in both control and knockdown samples, before an increase at day 6(Figure 18D). Due to the conflicting nature of these results, it would be imprudent to draw any conclusions without a third repeat.

Recent studies have shown that *Eomes* works in combination with mesodermal marker *T*, to control the exit from pluripotency towards a mesoderm or definitive endoderm state [125]. Therefore, it would be interesting to see if changes in *Eomes* expression affect expression of *T. T* typically peaks on day 4, which is observed in control samples. This expression is severely reduced in *Eomes* knockdown samples, to 30% of control samples, at day 4 in both differentiations. Expression in both control and knockdown samples is reduced at days 5 and 6 (Figure 18E).

*Eomes* is known to be capable of inducing *Mesp1* expression and cardiac differentiation in both an ESC and *in vivo* model[123,207]. In addition to the effect of *Eomes* knockdown on mesodermal markers, it is important to validate that this knockdown also recapitulates past published data in a reduction of the expression of cardiac progenitor cell markers.

*Gata4* expression peaks at day 5 of differentiation, as detected in both control samples, with expression peaking at 1.5 and 15-fold in differentiation 1 and 2 respectively. *Gata4* expression is highest a day earlier in both knockdown samples, with transcript levels decreasing after day 4 to near absent levels at day 6. It should be noted that peak transcript levels on day 4 in differentiation 1 knockdown samples is nearly 20-fold that of control, whilst in differentiation 2 it remains at 50% of control sample levels. Despite this, both differentiations show a similar pattern with knockdown samples peaking a day earlier than control samples (Figure 18F). However, due to the variation in transcript levels between differentiations, a further replicate is needed to confirm the effect of *Eomes* knockdown.

Experiments in *Eomes* knockout hESCs have illustrated a reduction in *Nkx2.5* positive cells after 8 days of culture [206]. Therefore, similar changes may also occur in mESCs when *Eomes* is knocked down at mesoderm stage.

*Nkx2.5* expression is characteristically upregulated at day 5 of differentiation, corresponding with a transition from mesoderm to cardiac progenitor state. This pattern is observed in control samples for both differentiations, peaking at 1.4-2.6-fold levels seen in control cells at day 4 of differentiation.

Conversely, peak expression in both knockdown replicates is seen at day 4, peaking at 0.9 and 0.3-fold that found in control samples in differentiation 1 and 2 respectively. Expression of the transcript is reduced in knockdown samples at day 5 and further by day 6 of differentiation. Expression in all samples is reduced at day 6, in line with expected temporal patterns (Figure 18G).

This knockdown mESC model does confirm previous data that indicates *Eomes* works upstream of *Mesp1* to induce its expression, which is severely reduced upon *Eomes* knockdown. It further agrees a role for *Eomes* in cardiac progenitor cell differentiation, with a reduction in *Eomes* resulting in a decrease in the cardiac progenitor cell marker *Nkx2.5.* In addition to endorsing these findings, data produced suggests a role for *Eomes* in the regulation of *Mbd3* expression, which is reduced upon *Eomes* knockdown. The mechanism behind how *Eomes* induces this expression is not known. It is possible that a reduction in *Eomes,* and associated reduction in *Mesp1,* may affect *Mbd3* expression. It may also be possible that *Eomes* is capable of directly regulating *Mbd3.* The effect of *Eomes* knockdown on the expression of *Wdr5* and *Gata4* cannot be discerned from these results, due to the variations in expression observed between the 2 differentiations.

Figure 18

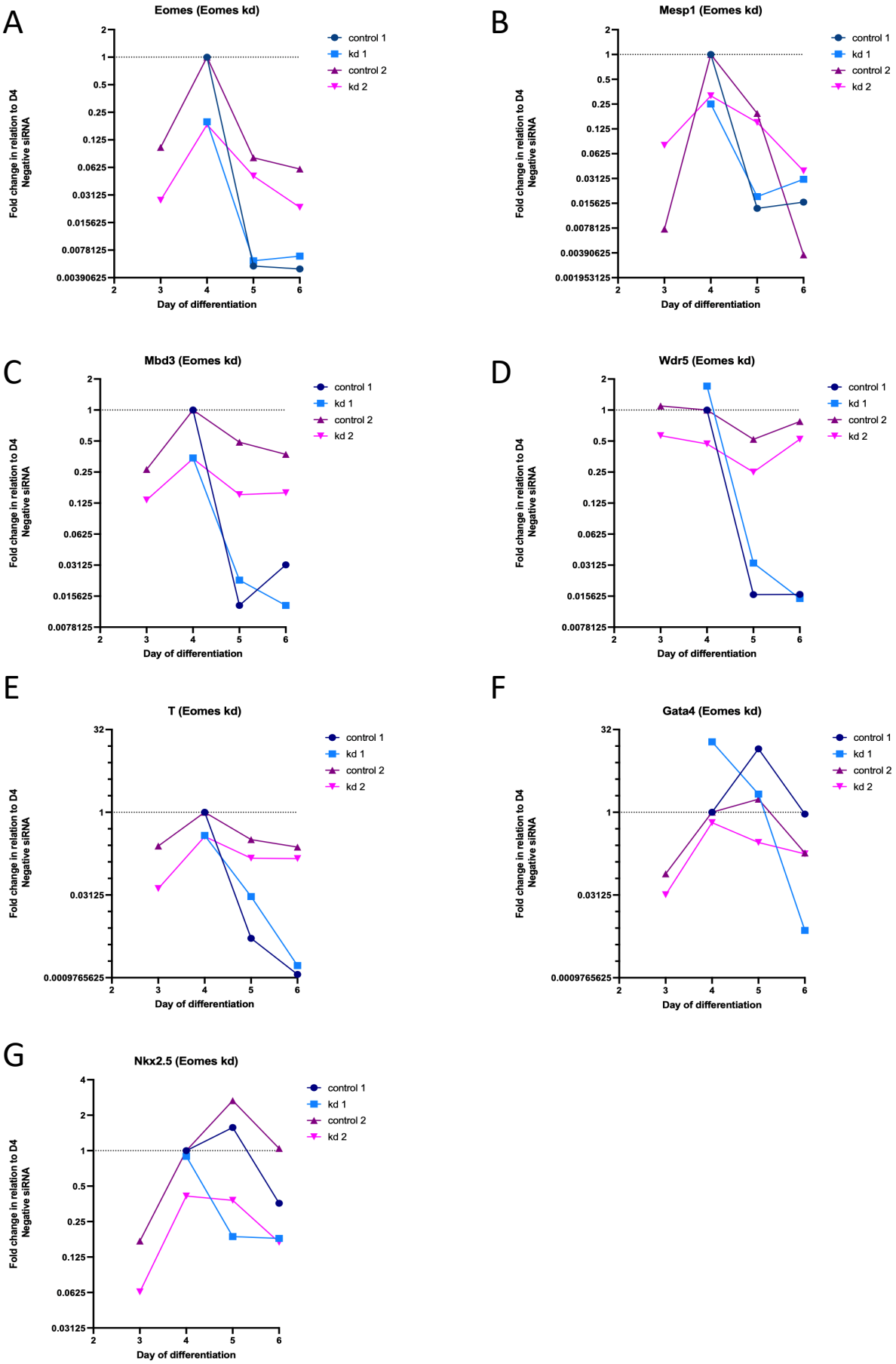



Figure 18A-G: Knockdown of *Eomes* results in a reduction of *Mesp1* (A), and other potential binding partner *Mbd3* (C). There was also a reduction in mesodermal marker *T* (E) and cardiac progenitor cell markers (F-G). Results from two different stem cell differentiations.

### 4.1.5 Knockdown in *Mbd3 results* in a decrease in *Mesp1,* whilst inducing a temporal shift in mesodermal and cardiac progenitor markers

*Mbd3* has been reported to have several roles in histone modification, embryonic development and cellular differentiation and reprogramming [171,208–212].

Previous results show MBD3 may potentially bind to MESP1 (Figure 10B). As a histone modifier, it may also change the expression patterns of other genes through changes to the epigenetic landscape of the cell.

*Mbd3* is ubiquitously found at high levels throughout ESCs. On day 4 of differentiation, *Mbd3* expression is decreased to 65% and 35% of control levels in differentiation 1 and 2 knockdown samples respectively. Control samples show a peak of expression at day 4 that approximately halves by day 5, with a further reduction at day 6.

Whilst differentiation 1 knockdown samples also decrease in expression at day 5, there is a return to peak levels at day 6. This contrasts with differentiation 2, where peak expression in knockdown samples is seen at day 5 before decreasing at day 6 (Figure 19A). The variation in knockdown efficiency, and return to control levels by day 6, could suggest the siRNAs are only active for 3 days, with knockdown most efficient at mesoderm stage, day three and four of differentiation (Figure 19A).

In addition to the effect of *Mesp1* knockdown on its binding partners, it is also important to evaluate the effect of its binding partners on *Mesp1*.

As previously shown, expression of *Mesp1* is tightly restricted, with a peak at day 4 of differentiation in control samples, which is reduced until day 6. By contrast, in *Mbd3* knockdown samples, there is no peak at day 4 in either differentiation replicate.

Differentiation 1 knockdown samples exhibit a peak in expression at day 5, although this is limited to 25% of the amount seen in day 4 control samples. This equates to approximately 50% of the level observed in day 5 control cells.

This reduction in transcript level is also observed in differentiation 2 knockdown samples, however, varies in temporal pattern. Highest expression is observed between day 3 and day 4 although is limited to below 15% of that witnessed in control samples at the same time point. Transcript level declines at day 5 but returns to a similar level at day 6. This is in contrast to the reduction observed at day 6 in other samples (Figure 19B).

These results indicate that although expression of *Mbd3* is not completely ablated by siRNA addition, the knockdown is sufficient to reduce *Mesp1* expression at mesoderm stage of differentiation.

As well as the effect of *Mbd3* reduction on *Mesp1* transcript levels, it is also important to consider the effect a reduction in *Mbd3* may have on the expression of other potential binding partners.

Typically, *Eomes* peak expression is observed between day 3 and 4 of differentiation, before reducing at day 5 and 6. This pattern is observed at day 4 in differentiation 2 samples; however, in knockdown samples, *Eomes* expression is limited to 50% of that seen in control samples. Expression of *Eomes* in differentiation 2 control and knockdown samples reduces from its peak at day 4 to 5% of peak expression at day 5 and 6.

This pattern is not observed in differentiation 1. Peak expression was observed at day 5, at 1.35-fold that found in day 4 control samples, before a reduction at day 6. Knockdown samples follow a similar pattern to that of differentiation 1 control, however at a lower amplitude. Knockdown expression of *Eomes* was 70% of control at day 4, and approximately 50% at day 5.

Despite the difference in temporal pattern of *Eomes* in differentiation 1 and 2, it is apparent in both differentiations that *Mbd3* knockdown results in a reduction of *Eomes* to approximately 50% of peak expression (Figure 19C).

Whilst knockdown of *Mbd3* appears to affect expression of *Mesp1* and *Eomes*, it is unknown what effect knockdown of one histone modifier may have on another in the same cell population.

*Wdr5* expression is similar to that of *Mbd3*, found ubiquitously throughout the cell population at high levels. As seen in *Mesp1* and *Eomes* knockdowns, *Wdr5* expression from the 2 differentiations show a variable pattern.

In differentiation 1, expression of *Wdr5* peaks at day 5, at approximately 1.45-fold that of day 4 expression. This same peak is observed in knockdown samples, however, is limited to 1.1-fold day 4 expression. Control and knockdown samples at day 6 show a reduction to approximately 15% of control levels at day 4. This pattern is also found in differentiation 2 knockdown samples, where the peak expression level at day 5 is similar to the level in differentiation 1. In contrast, differentiation 1 control samples show peak expression on day 3 of differentiation, before a reduction in expression until day 5. An increase in transcript level of approximately 50% was observed in these samples between days 5 and 6 (Figure 19D).

These results show an inconclusive picture for the effect of *Mbd3* knockdown on *Wdr5* expression. It is possible that there is a slight reduction in *Wdr5* expression at day 4 and

5 of differentiation, however, this would need to be substantiated through an additional differentiation replicate.

Previous investigations into the effect of *Mbd3* on cell lineage markers have been hindered by the embryonic lethality displayed in *Mbd3* knockout embryos [171], whilst *Mbd3* knockout ESCs fail to commit to developmental lineages [213]. Therefore, it is important to investigate the effects of *Mbd3* knockdown on cell lineage markers in mESCs that are directed towards a cardiovascular lineage. This would establish a post-pluripotency role for *Mbd3* in cardiac differentiation.

Expression of *T* characteristically peaks between day 3 and 4 of differentiation, reducing to near baseline from days 4 to 6. This pattern is clearly observed in control samples of differentiation 1 and 2. This temporal expression pattern is mimicked in knockdown samples of differentiation 2, however, peak expression is reduced by 25% at day 4. A more significant effect is seen in knockdown samples of differentiation 1, with a reduction in expression to 20% of control samples at day 4. Temporal expression imitated that of control samples (Figure 19E).

These results indicate that knockdown of *Mbd3* causes a reduction in the mesoderm marker, *T,* although this effect is variable between the 2 replicates.

*Gata4* expression ordinarily peaks at day 5 of differentiation before reducing to baseline levels at day 6. This is observed in control samples from both differentiations. In contrast, in knockdown cells, expression of *Gata4* is shifted by 24 hours to peak at day 4 of differentiation. Knockdown sample transcripts at this time point are equivalent to those seen in control samples, however, are severely reduced at day 5, failing to match the peak transcript levels witnessed in control samples at this time point (Figure 19F).

This suggests that *Mbd3* expression may play a role in the regulation of *Gata4*, with knockdown resulting in a change in the temporal expression, which is of reduced amplitude.

Typically, *Nkx2.5* expression peaks at day 5 of differentiation, coinciding with the transition from mesoderm to cardiac progenitor stage. This peak is observed in control cells in both differentiations, peaking at between 2.5 and 4.5-fold levels seen at day 4. This day 5 peak is not observed in knockdown cells, with peak expression witnessed at day 4 of differentiation. This expression is 1.2 and 2.8-fold greater than the expression seen in control cells at the same time point. However, expression was subsequently reduced by days 5 and 6, and failed to reach peak transcript levels witnessed in control samples at day 5 (Figure 19G).

Thus, in agreement with the effect of *Mbd3* knockdown on *Gata4*, the knockdown of *Mbd3* causes a temporal shift in *Nkx2.5* expression by 24 hours. This effect is similar to what was observed in the *Eomes* knockdown.

In summary, *Mbd3* knockdown causes a reduction in mesodermal markers *Mesp1, T* and *Eomes.* This knockdown also results in a progressive shift, and reduction in expression, of cardiac progenitor markers *Gata4* and *Nkx2.5* by 24 hours, indicating a role for *Mbd3* in cardiac progenitor cell differentiation. The role of *Mbd3* in modulating transcription of *Wdr5* is unclear, and further replicates would be needed to clarify this.

Figure 19

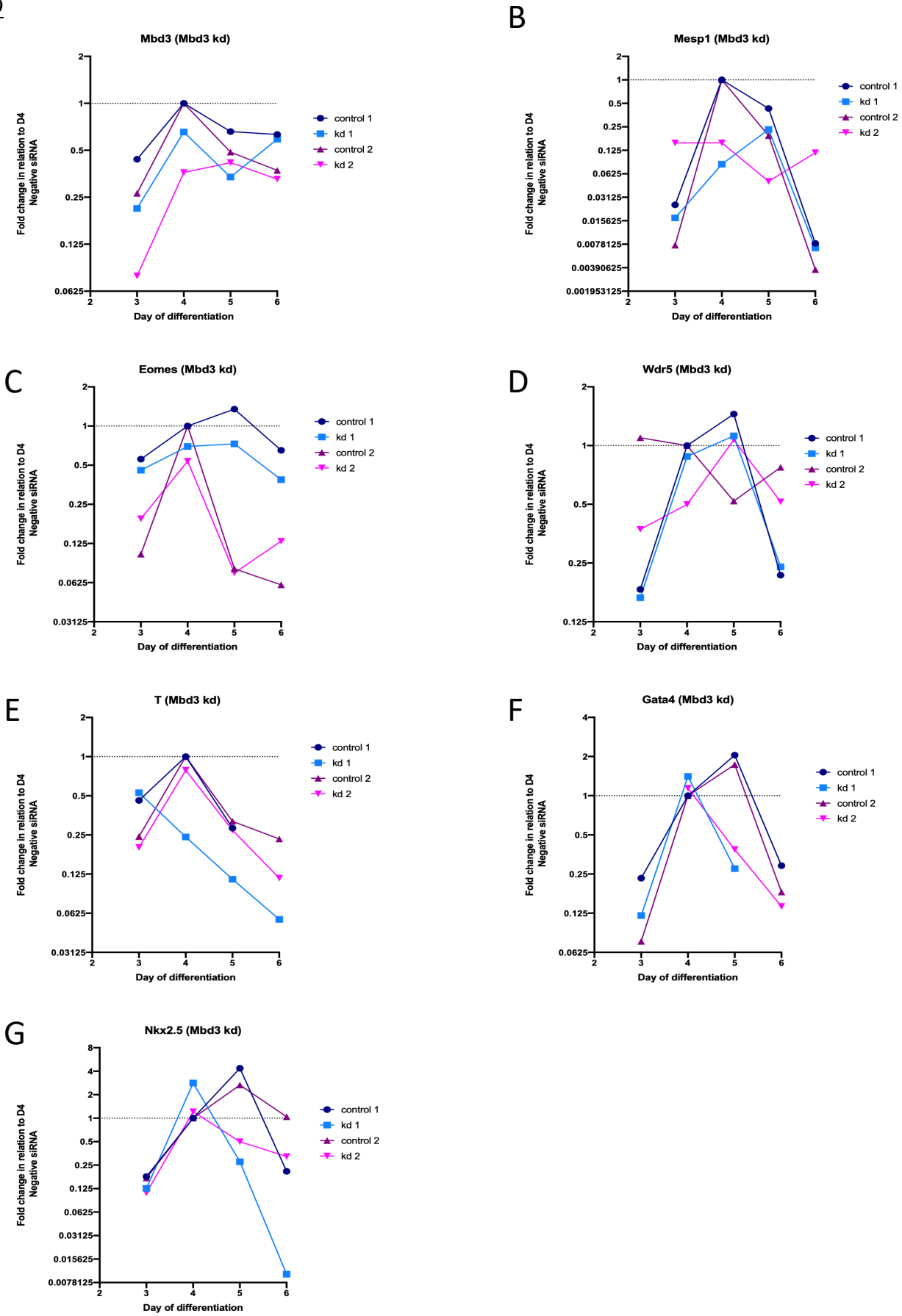



Figure 19A-G: Knockdown of *Mbd3* results in a reduction of mesodermal markers *Mesp1* (B), *Eomes* (C) and *T* (E) as well as a temporal shift in cardiac progenitor markers (F-G). A reduction in *Wdr5* was observed with variations in replicates (D). Results from two different stem cell differentiations.

### 4.1.6 Knockdown of *Wdr5* affects expression of mesodermal marker genes

The role of *Wdr5* is varied and includes functions in development and cell maintenance. This includes roles in mesoderm and endoderm patterning, as well as for maintaining pluripotency in conjunction with *Oct4* [200,214–216]. It is interesting, then, to look at the role of *Wdr5* post-stem cell phase, and its role in mesoderm and cardiac progenitor cell formation.

Knockdown of *Wdr5* at day 2 of differentiation resulted in a reduction in *Wdr5* expression at day 4, to approximately 40-55% of that seen at the same time point in control samples (Figure 20A).

As previously detected, there is some variation in expression profiles between the 2 differentiations, with differentiation 1 exhibiting peak expression at day 5, whilst the highest expression is found at day 3 in differentiation 2.

Despite this inconsistency, *Wdr5* expression is approximately halved in knockdown samples in comparison to controls at day 4, and whilst it would be preferable for this to be more efficient, it should be sufficient to see an effect on gene expression changes within the cell.

Whilst WDR5 has not been shown to bind directly to MESP1, it is capable of binding in a complex with EOMES. This complex can bind upstream of *Mesp1* and allow its upregulation [124]. Furthermore, in addition to binding to EOMES, WDR5 is also capable of binding to MBD3 isoform C, found in ESCs, which allows regulation of pluripotency [217]. Disturbance of these binding events could cause modifications in the expression profile of these genes.

Control samples exhibited the expected peak expression of *Mesp1* at day 4, that reduced to baseline levels by day 6. In contrast, at day 4, knockdown samples were limited to 10% and 15% of control expression, in differentiation 1 and 2 respectively.

The typical *Mesp1* expression profile was not seen in *Wdr5* knockdown samples. Peak expression was shifted 24 hours to day 5. Peak expression in knockdown samples was limited to between 0.2 and 0.25-fold of that observed in peak control samples (Figure 20B).

These results suggest that knockdown of *Wdr5* not only results in a reduction in *Mesp1* expression, but also causes a temporal shift by 24 hours, to peak at day 5 of differentiation.

In addition to *Mesp1* it is essential to discern the role of *Wdr5* on its known binding partner *Eomes.* As previously discussed, WDR5 is capable of binding to EOMES as part of a

complex with GCN5[124], however, it is not known if there is any interaction at a transcript level. Typically, *Eomes* expression peaks at day 4, with this profile observed in control samples from differentiation 2. In contrast, knockdown samples peak at nearly twice the levels of control samples at day 4. At day 5 this increases to 2.25-fold the peak control levels before reducing at day 6.

Diverging from the usual *Eomes* expression profile, control samples in differentiation 1 peak at day 5, at a slightly increased level to those seen at day 4, before reducing at day 6. In a similar pattern to that observed in control samples, differentiation 1 knockdown samples show amplified expression in comparison to peak control levels. *Eomes* expression increased to almost double that observed in day 5 control samples (Figure 20C).

Despite variations in temporal expression between replicates, it is apparent that *Eomes* expression increases beyond peak control levels when *Wdr5* is knocked down. This suggests a role for *Wdr5* in negatively regulating *Eomes* expression. Discrepancies in these results identify a need for a third replicate.

WDR5 has also been shown to interact with MBD3 isoform c, an isoform only found in ESCs. This complex is known to help regulate pluripotency in these cells[217] although its function post-stem cell phase is unknown.

Expression of *Mbd3* in both control samples follows a similar temporal pattern, with peak expression found on day 4. Expression is reduced at day 5, to between 50-70% expression, and is marginally reduced at day 6.

In contrast, knockdown samples from differentiation 2 show a double peak profile. Expression peaks at day 4, at almost 1.5-fold that seen in control samples, with a steep reduction in expression at day 5, before an increase at day 6. Transcript levels at day 6 are approximately 90% of peak control levels at day 4.

This pattern is not observed in knockdown samples from differentiation 1, where instead of a double peak profile, transcript levels rise gradually throughout the time span, to its peak on day 6, where expression is approximately 1.8-fold that seen in peak control samples at day 4 (Figure 20D).

Whilst these knockdown results diverge in terms of temporal pattern; both show an increase in expression at day 6 in comparison to control samples. Therefore, it is possible that *Wdr5* knockdown results in a change in *Mbd3* temporal expression. However, to delineate a pattern, a further replicate is necessary.

In addition to investigating its effects on binding partners it is also essential to identify the role of *Wdr5* in mesodermal and cardiac specification.

As previously discussed, expression of *T* is typically observed peaking at day 4 of differentiation, coinciding with mesoderm stage. This pattern is observed in both control differentiation samples.

A similar pattern is observed in both knockdown samples. Peak expression is reduced by 70% and 20%, in comparison to control levels on day 4, in differentiation 1 and 2 respectively (Figure 20E).

This data suggests that *Wdr5* may have a role in inducing expression of *T.* The reduction in *T* expression is variable between the knockdown samples, and therefore a further replicate is needed to confirm the mean reduction in *T* expression.

Mutations in *Wdr5* have been linked to conotruncal defects in human patients[218], and further studies in *Xenopus* show depletion of *Wdr5* can cause pericardial oedema and alterations in left-right patterning [219]. To date there is no link between *Wdr5* and cardiac progenitor genes, therefore, it is important to see if *Wdr5* knockdown affects the amplitude or temporal expression of these genes.

*Gata4* expression typically peaks at day 5, as observed in control samples from both differentiations, peaking at between 1.6 to 2-fold the expression level seen at day 4. Expression is then reduced by day 6 of differentiation.

There are conflicting results between the expression of *Gata4* in differentiation 1 and 2 in *Wdr5* knockdown samples. Knockdown samples in differentiation 1 show a peak in expression at day 5 of differentiation, correlating with that seen in control samples. Peak expression is approximately three-fold that seen at day 4 in control samples and is almost one third more than that seen in day 5 control samples. This would indicate that knockdown of *Wdr5* results in an increase in *Gata4* expression.

In differentiation 2, *Wdr5* knockdown samples similarly peak at day 5 of differentiation. This was seen to be 1.3-fold that observed in day 4 control samples. This is approximately 0.3-fold less than that seen in control samples at this stage. Both control and knockdown samples were reduced to similar levels at day 6 of differentiation (Figure 20F).

These results directly conflict with the high levels of *Gata4* transcript witnessed in day 5 differentiation 1 knockdown samples in comparison to control. Therefore, to ascertain what effect *Wdr5* has on *Gata4* expression, a further replicate will be necessary.

In addition to *Gata4*, another key cardiac progenitor marker is *Nkx2.5.* Similarly, expression is expected to peak at cardiac progenitor stage, or day 5 of differentiation, which is observed in control samples for both differentiations.

In agreement with results for *Gata4*, there is some variability in the results obtained from the 2 differentiations.

Control samples in differentiation 1 peak at day 5, at approximately 4.5-fold that seen at day 4. In *Wdr5* knockdown samples, this peak is amplified to over 6-fold that of control samples at day 4.

This effect is not observed in differentiation 2 samples. Peak expression of both control and knockdown samples is observed on day 5, with control samples 2.5-fold that seen on day 4. In contrast to differentiation 1, expression of *Nkx2.5* is reduced in knockdown samples (Figure 20G).

Variability in the transcript level of *Nkx2.5* identified in knockdown samples suggests a need for a third replicate to establish what effect *Wdr5* has on *Nkx2.5* transcript levels.

These results indicate that *Wdr5* expression is necessary for the correct temporal expression patterns of both *Eomes* and *Mesp1*. In contrast, *Mesp1* appears to be positively upregulated by *Wdr5*, whilst *Eomes* is negatively regulated. No clear effect can be discerned about the role of *Wdr5* on *Mbd3*, *Nkx2.5* or *Gata4*.

Figure 20

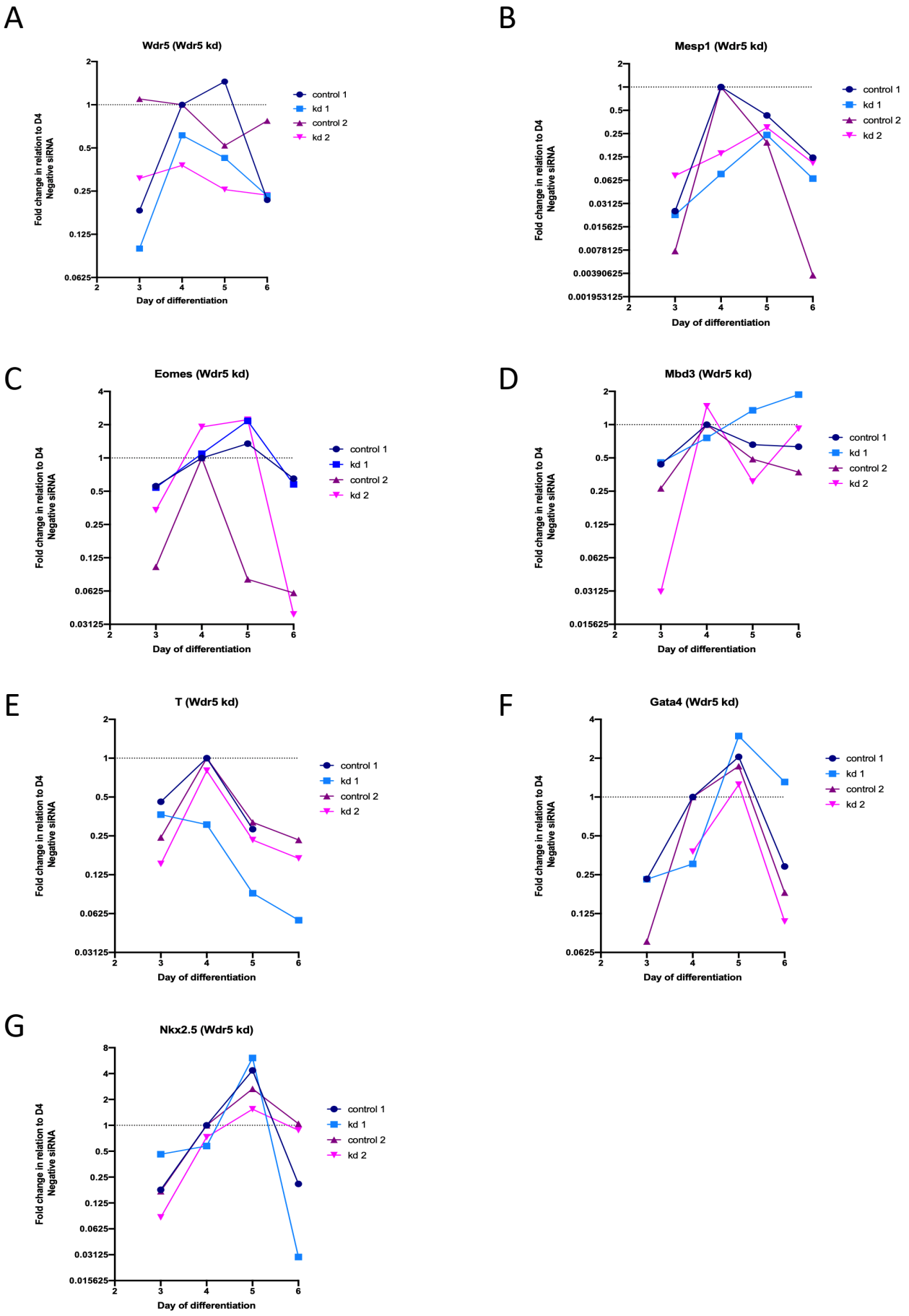



Figure 20A-G: Knockdown of *Wdr5* (A), resulted in a decrease in mesodermal markers *Mesp1* and *T* (B and E respectively). The results of the effect of *Wdr5* knockdown on *Mbd3* (D), and cardiac progenitor markers (F-G) is unclear and will need additional replicates to decipher the effect. Results from two different stem cell differentiations.

### 4.1.7 Summary

To date, experiments to evaluate the effects of knocking down *Mesp1, Eomes, Mbd3* and *Wdr5* at mesoderm stage had not been performed. These experiments show that knockdown of these factors at mesoderm stage causes multiple effects on cellular differentiation markers.

Knockdown of *Mesp1* was efficient, with transcript amount reduced to between 10 and 20% of control levels. It had clear effects on the expression of its binding partners *Eomes* and *Mbd3*, with a reduction in transcript levels. This was also observed in the case of fellow mesodermal marker *T*, which is reduced to below 40% of its control amount. These results bring further light to the regulation of mesodermal genes during gastrulation, with *T* and *Eomes* thought to induce *Mesp1* expression.

Knockdown of *Mesp1* also caused a shift in expression of cardiac progenitor markers *Gata4* and *Nkx2.5.* This is to be expected, as these genes are traditionally thought to be downstream of mesoderm specification. Therefore, alterations in transcription of *Mesp1* can change downstream gene expression that affects the ability of the cells to mature to a cardiomyocyte phenotype. The effect of *Mesp1* on *Wdr5* expression was undetermined due to differences in gene expression patterns between the two replicates, although peak *Wdr5* expression was reduced by half in knockdown samples at day 4. Therefore, further replicates will be needed to define the true effect of *Mesp1* on *Wdr5* expression.

Knockdown of fellow transcription factor, *Eomes,* was also successful with transcript levels reduced to approximately 20% of peak control level. In a similar pattern to *Mesp1* knockdown, reduction in *Eomes* expression led to a reduction in *Mesp1* to between 20-30% of control levels at peak expression, which was also observed in the expression of *Mbd3* and *T*. Correspondingly, there was also a shift in peak expression of *Gata4* and *Nkx2.5* to 24 hours earlier, at day 4. There was a reduction of expression at day 5, which is traditionally the peak expression time point in comparison to controls. Comparably, the effect of *Eomes* on *Wdr5* expression was also undeterminable, with further repeats needed. These results are not unexpected, due to known induction of *Mesp1* by *Eomes*, and its contribution to cardiac lineage specification. However, it is interesting to note the temporal shift and reduction in cardiac progenitor markers, which has not been previously examined.

Knockdown of *Mbd3* was not as efficient as that of the transcription factors, with transcript levels reduced to between 30 and 65% of that seen in control samples during peak

expression. Despite its limited knockdown, the effect of it was apparent in the expression of the mesodermal markers *Mesp1* and *T*, with a reduction in *Mesp1* transcript level to 30% in knockdown samples, and to between 20% and 70% for *T.* In agreement with knockdowns of *Mesp1* and *Eomes*, both cardiac progenitor cell markers *Gata4* and *Nkx2.5* show a temporal shift in knockdown samples. Both markers peak 24 hours prior to control samples. Similarly, to previous knockdowns, the effects on *Wdr5* expression vary between differentiations and will need at least a third replicate to identify a trend. These findings show that *Mbd3* is not just important in ESC pluripotency, but also in the progression of cells to a mesodermal lineage. These effects on mesodermal genetic markers could then instigate the variation in cardiac progenitor cell marker expression.

Knockdown of the histone modifier *Wdr5* was limited, with a reduction to approximately 40-55% of expression found in control samples at day 4. Knockdown resulted in a delay in peak *Mesp1* expression to day 5, with peak expression in differentiation knockdown samples limited to 30% of peak control expression on day 4. Equally, for the other mesodermal marker *T*, peak expression was reduced by 70% and 20% in differentiation 1 and 2 respectively. There were mixed results with regards to the effect on *Eomes* expression, with differentiation 1 control samples showing a shift in peak expression, although peak expression of knockdown samples remained higher than in control samples. Whilst a further replicate is needed to confirm results, it appears knockdown of *Wdr5* causes an increase in *Eomes* expression. Results from cardiac progenitor marker analysis were variable, and therefore, no conclusion can be taken from this data.

These experiments show that knockdown of transcription factors was more efficient than that of histone modifiers, perhaps in part due to the abundant nature of the histone modifiers, and lack of temporal restriction.

There were also variations between repeats in terms of temporal patterns, and knockdown efficiencies. This could be due to a variety of factors, including batches of cytokines used owing to their very limited lifespan, and a change in the cell culture room experiments were performed in between the replicates due to mechanical failure.

A third repeat of these knockdowns, with additional staining of cells, is necessary for firm conclusions to be made.

# 5 Chapter 5

## 5.1 Overexpression of *Mesp1* and its binding partners in fibroblasts can alter cell morphology and cardiac progenitor marker expression

### 5.1.1 Introduction to reprogramming

The field of cardiovascular disease has seen many therapeutic breakthroughs. However, the ability to cure end-stage heart failure remains limited to transplantation. Progress in this field is focused on the ability to repair damaged myocardium through direct reprogramming of cells *in situ* or implantation of cardiac cells originating from a patient's reprogrammed somatic cells. Studies of each approach have shown improvements in ventricular function, a delay in heart failure progression and reduction in ventricular remodelling [63]. Studies have shown that *in vivo* grafting of earlier stage cardiomyocytes have increased survival rates in comparison to more mature grafts in the heart [220], indicating that reprogramming to an earlier stage of cardiac differentiation may be beneficial.

Reprogramming of somatic cells, either directly or through a stem cell-like intermediary phase remains challenging, with low yields of cells successfully reprogrammed. Direct reprogramming using transcription factors has shown a success rate of between 1.6% to 27.6% with various assays, since studies first identified the transcription factors *Gata4, Mef2c,* and *Tbx5* (GMT) as capable of inducing fibroblasts to a cardiac lineage [73]. Further studies have since modified this protocol to include micro-RNAs, as well as histone modification proteins [51]. The addition of *Mesp1* and the histone modifier, *Smarcd3,* to the *GMT* reprogramming cocktail led to an increase in cardiac cells derived from fibroblasts [70], suggesting a role for *Mesp1* and histone modifiers in fibroblast reprogramming.

*Mesp1* and its potential binding partners affect cardiac differentiation in mESCs, altering the temporal expression of cardiac progenitor cell markers. Therefore, it is important to evaluate the effect of *Mesp1* and its binding partners on reprogramming fibroblasts to a cardiac progenitor cell fate.

Aim: To investigate the effect of overexpressing *Mesp1* and its binding partners in fibroblasts, on cell morphology, proliferation and cell lineage markers.

## 5.2 Increasing efficiency of viral transduction through repeated viral exposure

To investigate the effect of overexpression of *Mesp1* and its binding partners on fibroblast morphology and fate, lentiviral vectors for *Mesp1* and its binding partners were

purchased and packaged in HEK cells. To determine efficiency of transduction, fluorescent tags attached to the binding partners *Eomes* and *Mbd3* were utilised. Test transductions were assayed through FACS sorting of infected cells in comparison to untransduced cell lines. Primary human fibroblasts from two patients, 071113 and 240216, were transduced with a standard transduction protocol: 0.5mL of virus with 8µl of the cationic polymer polybrene for 24 hours, before cells were expanded and FACS sorted. Initial transduction in 071113 resulted in a 9.8% EGFP-*Eomes* positive population, whilst only 1.8% for mCherry-*Mbd3* (Figure 21). A similar efficiency was also observed in fibroblasts from patient 240216 with an 11.8% EGFP-*Eomes* positive population and 2.1% when transduced with mCherry-*Mbd3* (Figure 22).

The improvement of viral transduction efficiency would allow for an increased population of cells obtained from each transduction, reducing the need for large cell expansions, as well as a less heterogenic population. Therefore, to improve efficiency, a 3-pronged approach was undertaken; increasing the viral amount transduced, increasing the amount of cationic polymer to aid transduction and repeated transductions over a 72-hour time period.

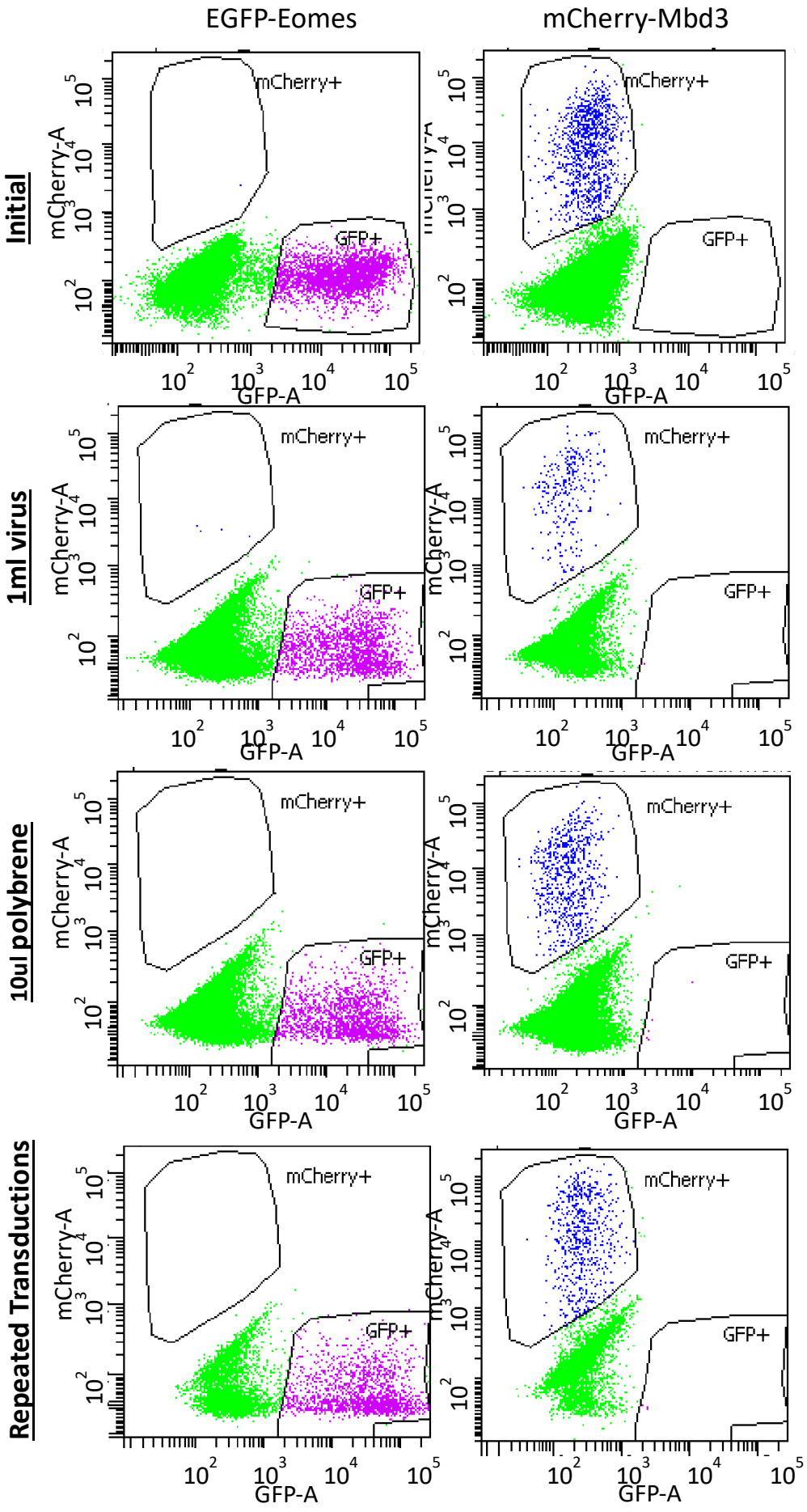
Increasing the viral amount from 0.5mL to 1mL in fibroblasts from patient 071113 resulted in a reduction in the EGFP positive population to 6.4%, whilst there was a marginal increase to 2% in the mCherry-*Mbd3* transduced cells (Figure 21). This reduction was also observed in patient 240216, where EGFP positive cells more than halved to 4%, whilst mCherry positive cells were reduced to 1.9% (Figure 22). This indicates that increasing the virus volume to cell ratio does not increase transduction efficiency.

Other efforts to increase viral transduction included increasing the amount of polybrene by 25% to 10uL. Polybrene is a cationic polymer that reduces the charge repulsion between the cell surface and lentivirus [221], however, exhibits cytotoxic properties at high concentrations. Increased polybrene in 071113 patient cells resulted in 8.1% EGFP positive cells, a reduction of 1.7% in comparison to the original protocol (Figure 21). mCherry positive cells increased from 1.8% to 2.7% (Figure 21). A similar trend was also observed in 240216 patient cells, where EGFP positive cells were reduced from 11.8% to 9.8% (Figure 22), whilst mCherry positive cells increased from 2.1% to 3% (Figure 22). This suggests a minimal improvement in transduction efficiency for mCherry-*Mbd3* positive cells when increasing polybrene, however a decrease for EGFP-*Eomes* transduced cells.

A final attempt to improve transduction efficiency embraced repeated transductions over a 72-hour period. Cells were plated and transduced according to the original protocol, with transduction at 24 hours post-plating. A further transduction was performed 72 hours post-plating, with media changed at 24 hours post-viral addition. Repeated transduction led to a greater than three-fold increase in viral efficiency in patient 071113, with EGFP positive cells increasing from 9.8% to 29.7% (Figure 22), whilst mCherry positive cells increased from less than 2% to 11.1% (Figure 22). This trend was replicated in 240216 patient cells. EGFP positive cells almost doubled from 11.8% to 20.8% (Figure 22), as did mCherry positive cells from 2.1 to 5.8% (Figure 22).

Optimisation of viral transduction is important for further experiments to prevent heterogenic populations that result in higher cell expansion rates to gain adequate cell numbers post-FACS sorting. These experiments suggest that repeated viral transductions over a 72-hour time period result in the highest FACS sorted positive population for both EGFP-*Eomes* and mCherry-*Mbd3* transduced cells, whilst increasing either viral or polybrene amount has minimal effect on transduction efficiency. Results summarised in Table 5.

Figure 21




**Patient 071113**

| **Initial Transduction** Transduction rates | |
|---|---|
| EGFP | mCherry |
| 9.8% | 1.8% |

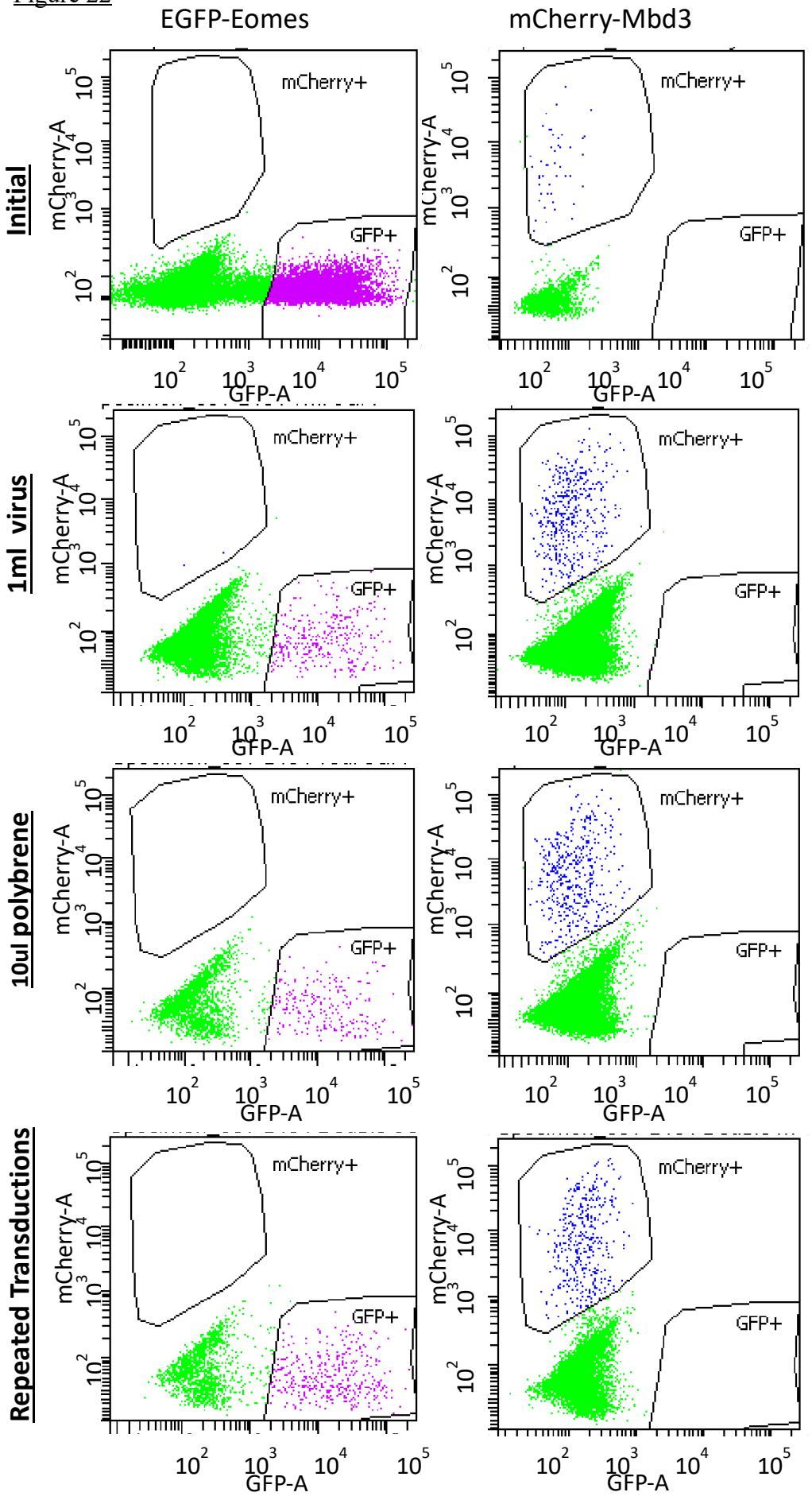
| **Transduction with 1ml virus** Transduction rates | |
|---|---|
| EGFP | mCherry |
| 6.4% | 2% |

| **Transduction with 10ul polybrene** Transduction rates | |
|---|---|
| EGFP | mCherry |
| 8.1% | 2.7% |

| **Transduction with repeated transductions** Transduction rates | |
|---|---|
| EGFP | mCherry |
| 29.7% | 11.1% |

Figure 21 Optimisation of viral transduction protocol on human dermal fibroblasts from patient 071113 illustrates necessity of repeated transductions. Optimisation of viral transduction of mCherry *Mbd3* and EGFP *Eomes* into patient 071113 primary Human dermal fibroblasts.

Figure 22




| Initial Transduction | |
|---|---|
| Transduction rates | |
| EGFP | mCherry |
| 11.8% | 2.1% |

| Transduction with 1ml virus | |
|---|---|
| Transduction rates | |
| EGFP | mCherry |
| 4% | 1.9% |

| Transduction with 10ul polybrene | |
|---|---|
| Transduction rates | |
| EGFP | mCherry |
| 9.8% | 3% |

| Transduction with repeated transductions | |
|---|---|
| Transduction rates | |
| EGFP | mCherry |
| 20.9% | 5.8% |

Figure 22 Optimisation of viral transduction protocol on human dermal fibroblasts from patient 260216 illustrates necessity of repeated transductions. Optimisation of viral transduction of mCherry *Mbd3* and EGFP *Eomes* into patient 240216 primary Human dermal fibroblasts.

Table 5

| **Primary cell line** | **Type of transduction** | **Efficiency of Transduction** | |
|---|---|---|---|
| | | **EGFP** | **mCherry** |
| **071113** | Initial | 9.8 | 1.8 |
| | 1ml | 6.4 | 2 |
| | 10ul | 8.1 | 2.7 |
| | Repeated transductions at 24/72 hours | 29.7 | 11.1 |
| **240216** | Initial | 11.8 | 2.1 |
| | 1ml | 4 | 1.9 |
| | 10ul | 9.8 | 3 |
| | Repeated transductions at 24/72 hours | 20.9 | 5.8 |

Table 5 A summary of the optimisation of transduction in primary dermal fibroblasts from two patients. A summary of FACS sorted cells when transduced with varying amounts of virus, polybrene or number of transductions in comparison to the original protocol used. Optimal results were obtained when using repeated transductions at 24 and 72 hours, in both sets of patient primary human dermal fibroblasts.

Figure 24

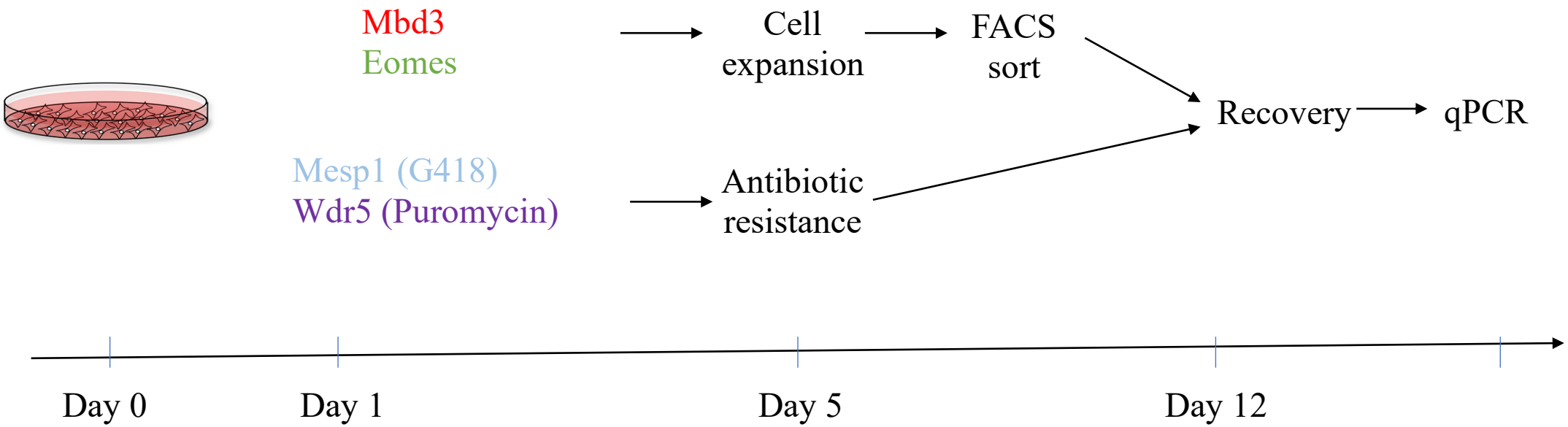



Figure 24 Cell transduction protocol. A schematic of optimised viral transduction where 3T3 cells are plated at day 0, with 0.5ml of virus added at day 1 and 3 of transduction with media exchanged at day 2 and 4. Those transduced with mCherry *Mbd3* or EGFP *Eomes* were expanded for 5 days before FACS sorting for fluorescent markers. Cells transduced with *Mesp1* or *Wdr5* underwent antibiotic resistance for 7 days. Post sorting all cells were allowed to recover for 7 days before qPCR analysis.

## 5.3 Optimising antibiotic resistance

In addition to fluorescently tagged lentiviral vectors, vectors for *Mesp1* and *Wdr5* included antibiotic resistance cassettes for neomycin (G418) and puromycin respectively. Whilst FACS sorting had allowed for the optimisation of the lentiviral transduction protocol, the resultant FACS sorted cells were of low viability, and due to their primary nature, were not capable of proliferating to the necessary population size for downstream experiments. Therefore, the decision was made to transfer to the 3T3 mouse fibroblast line, with greater proliferation potential, although more resistant to transduction. Due to their robust nature, these cells can survive the pressure of the flow cytometry sorting, and continue to proliferate in culture, unlike the primary human fibroblasts.

This exchange of fibroblast cell type occurred post-optimisation of transduction protocol, and prior to evaluating the optimal antibiotic concentration for selection.

To investigate the minimum concentration of puromycin necessary to cause cell death to untransduced cells, cells were plated in triplicate for an MTT assay. The assay gives a colorimetric reading of cell metabolism, with metabolically active cells forming an insoluble precipitate, giving a higher optical density (OD) reading. A reduction in cell viability results in a decreased amount of precipitate forming and a reduced OD reading. A reduction in OD was apparent between 10 and 20µg/mL of puromycin, reducing from an OD of 3 to less than 0.5 (Figure 25). This OD was constant between 20µg/mL and 50µg/mL suggesting that 20µg/mL was the lowest concentration necessary to kill untransduced cells.

The antibiotic neomycin (G418) is known to have a wide efficacy range of up to 900µg/mL [222]. To identify the correct concentration of antibiotic for transduced cell selection, transduced and untransduced cells were plated for an MTT assay in triplicate, before being treated over a range of 0-1000µg/mL. Both untransduced and transduced cells had a gradual decline in viability until 500µg/mL, with untransduced cells reducing from an OD of approximately 1.8 to approximately 0.6 (Figure 26). This viability level deteriorated further to an OD of less than 0.5 at 1000µg/mL. In contrast, whilst the OD of transduced cells declined from 1.5 to 1.0 when treated with 500µg/mL of neomycin, the OD of the transduced cells returned to approximately 1.2 when treated at 1000µg/mL. This indicates that the neomycin at 1000µg/mL was causing a greater reduction in cell viability of untransduced 3T3 cells than those transduced with *Neomycin-Mesp1*. However, as the cells appeared to tolerate the 1000µg/mL, cells were then treated with 2000µg/mL. At this concentration, untransduced cells became detached and died. Transduced cells also showed signs of cell

death and detachment, however, cells that remained attached grew in clusters and repopulated the plate (data not shown).

Based on cell metabolism readings, the optimal concentration for cell selection in 3T3 cells is 20μg/mL for puromycin, and 2000μg/mL for neomycin.

Figure 25

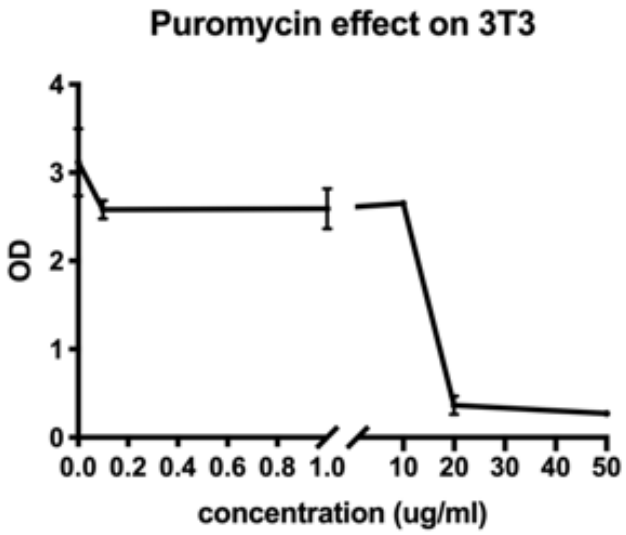



Figure 25 Kill curve of Puromycin on 3T3 cells. Effect of Puromycin on 3T3 cells over a concentration gradient. Optical density (OD) readings indicate that viability of wildtype 3T3 cells is severely decreased at 20ug/ml indicating this is the optimal concentration for transduced cells.

Figure 26

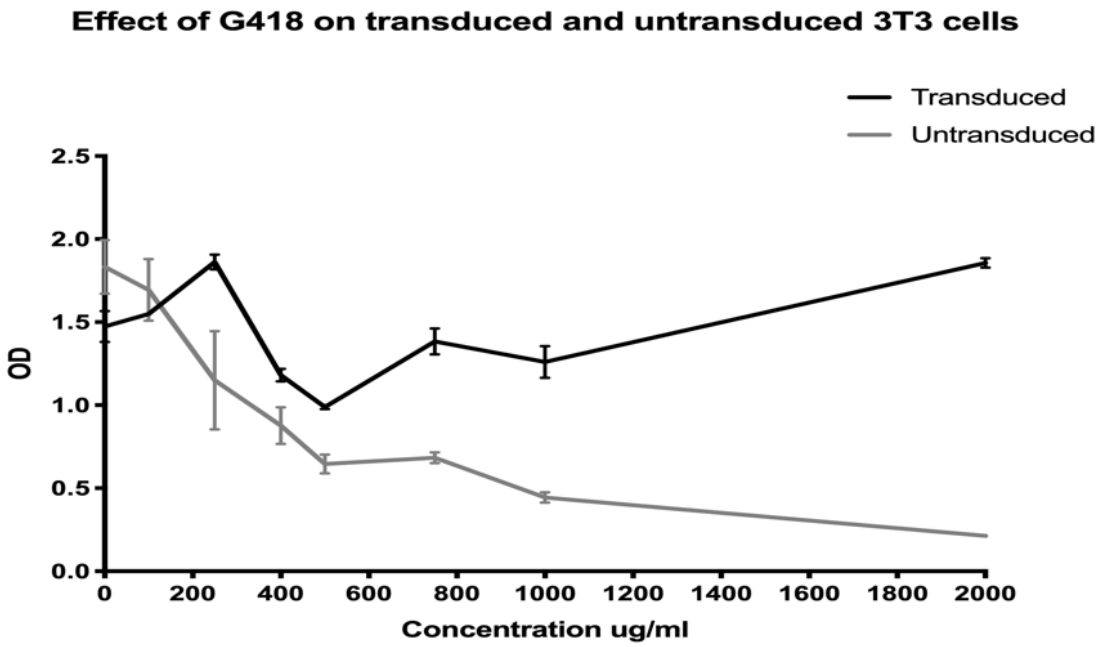



Figure 26 Kill curve of G418 on 3T3 cells. Effect of G418 on 3T3 cells over a concentration gradient. Optical density readings of cell viability show a decline in viability in untransduced cells in comparison to those transduced with Mesp1 virus containing a G418 resistance cassette.

## 5.4 Overexpression of viral vectors containing *Mesp1,* or its potential binding partners, confirmed through western blotting

Overexpression of genes within cells can be measured in several different assays. In addition to confirming transductions through FACS and antibiotic resistance sorting, protein expression in 3T3 cells must also be characterised.

To identify if overexpression of Myc-*tagged Wdr5* in 3T3 cells was successful, the MYC tag was immunoblotted for in control 3T3 cells, Myc-*Wdr5* transduced cells and those transduced with a control vector. The expression of the housekeeping gene, Alpha Tubulin, was constant throughout the three samples, whilst MYC-tagged WDR5 was only present in the transduced sample (Figure 27A). This relates to a near 25-fold average increase in MYC-WDR5 protein in comparison to untransduced 3T3 cells (Figure 27E).

*Mbd3* is a histone modifier, with ubiquitous expression in 3T3 cells. Immunoblotting of MBD3 indicates expression in untransduced cells, cells transduced with *Mbd3* and those transduced with a control vector (Figure 27B). Expression in 3T3 cells transduced with *Mbd3* showed an increased expression in comparison to control and untransduced samples.
A similar increase in alpha tubulin signal was also observed. Quantification of the *Mbd3* transduced cells, when normalised to the housekeeping gene, shows a minimal 1.7-fold average increase in comparison to untransduced cells (Figure 27E).

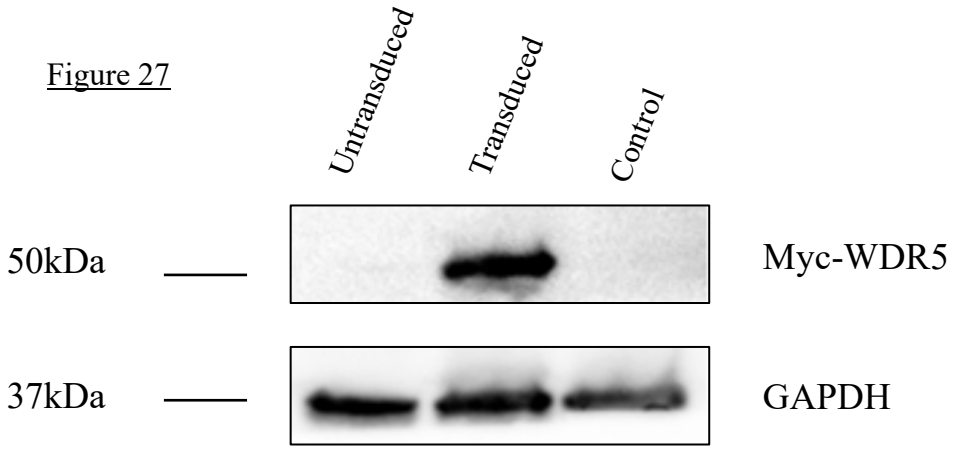
Transduction of *Eomes* in 3T3 cells results in a band at 70kDa, and a smaller band at approximately 78kDa (Figure 27C). There is a slight variation in expression of the housekeeping Alpha Tubulin between the 3 samples (Figure 27C). Quantification of the EOMES signal in all 3 samples relative to the housekeeping gene, and normalised to untransduced cells, identifies a 4.3-fold average increase in EOMES signal in transduced cells in comparison to untransduced.

To blot for MESP1, the attached HA tag was utilised to immunoblot. Immunoblotting revealed a HA signal at approximately 30kDa in the transduced cell sample, with minimal background at a similar size in the control and untransduced cells (Figure 27D). Alpha Tubulin showed variation in its signal between the samples, with the largest signal in the transduced cells (Figure 27D). Quantification, relative to the housekeeping gene and normalised to the untransduced cells, indicates a 12.3-fold average increase in MESP1 (Figure 27E)

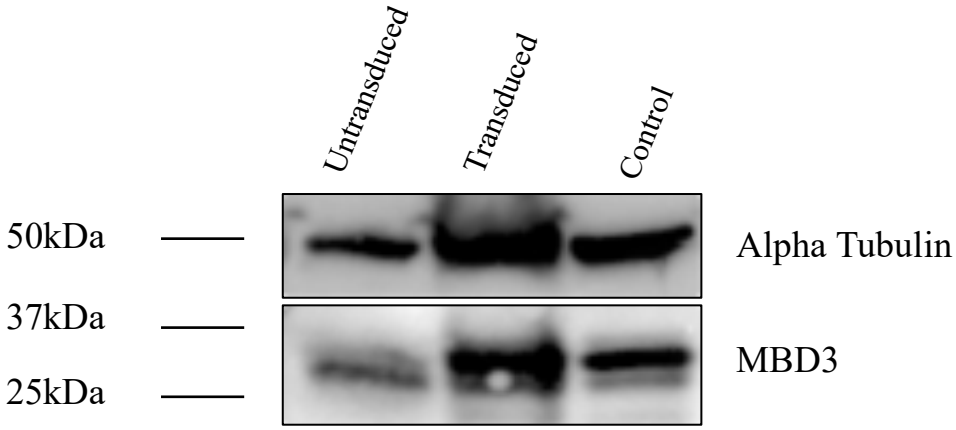
Protein quantification through western blotting indicates that MESP1, WDR5 and EOMES are all highly expressed when transduced into 3T3 cells in comparison to untransduced cells (Figure 27E). MBD3 shows minimal increases, however, this could be due

to protein degradation, or posttranslational modification of the protein preventing recognition by the antibody.

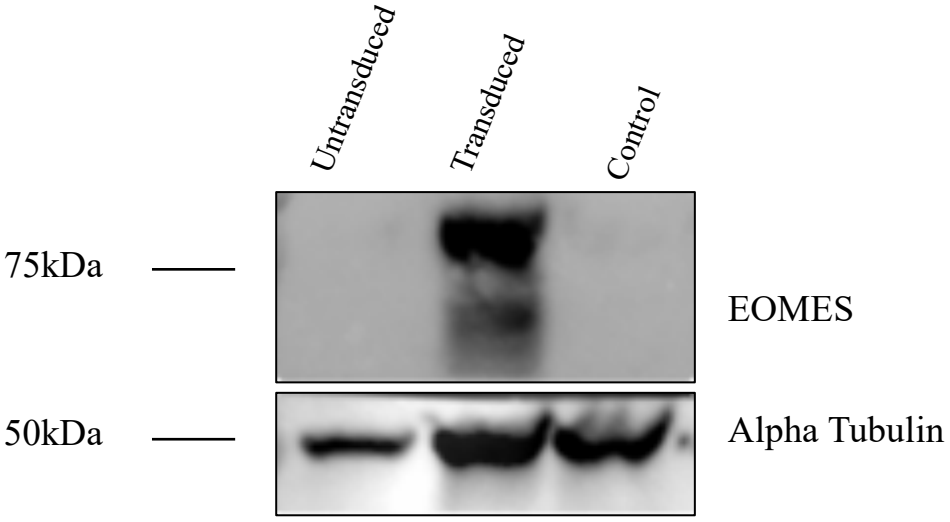
**A** Figure 27



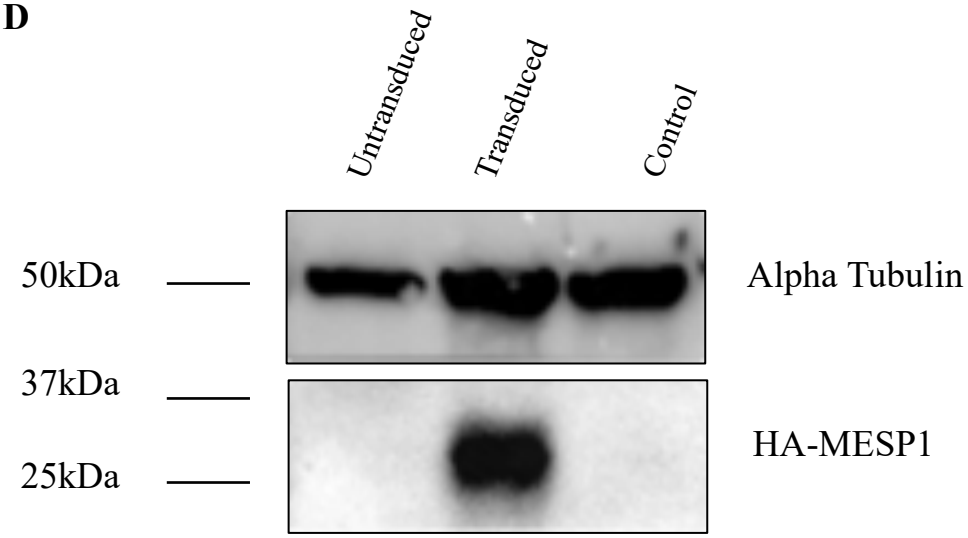

**B**




**C**



**D**




**E**

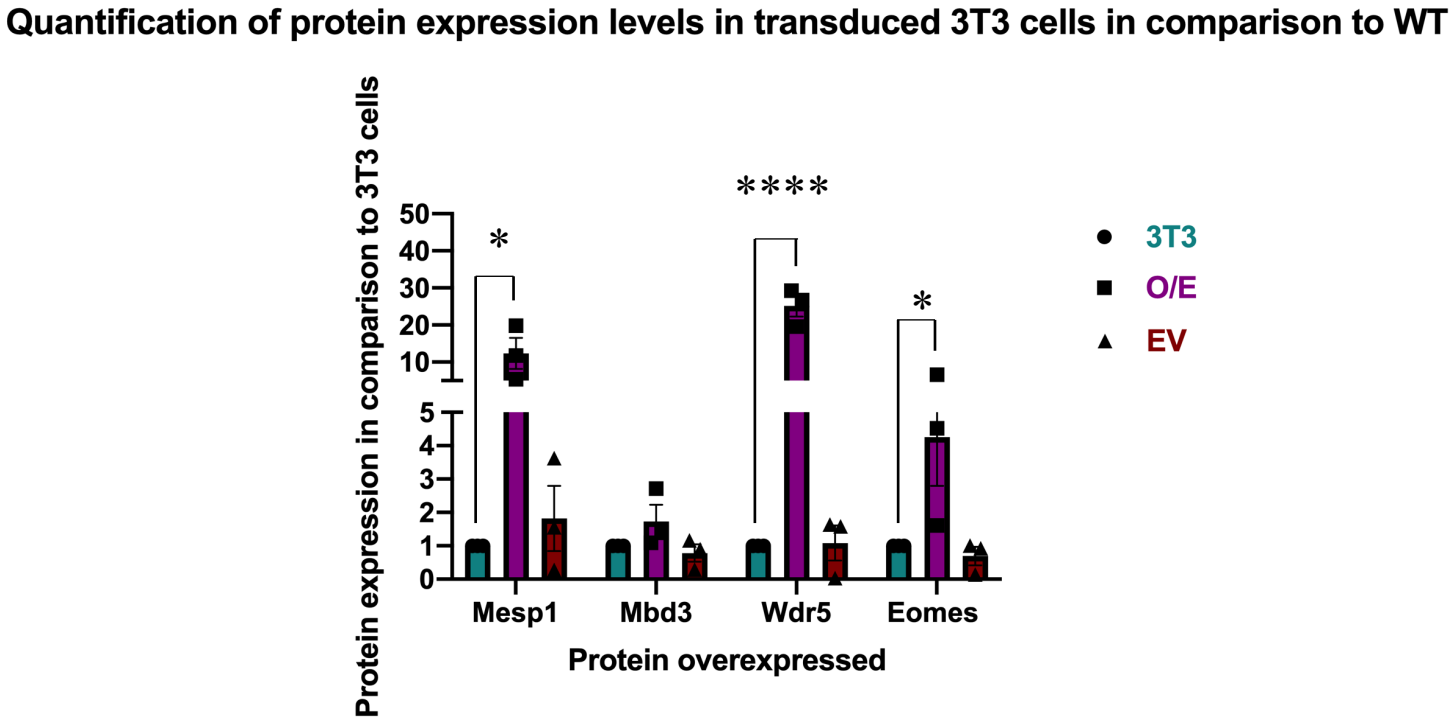



Figure 27A-E: Immunoblotting and quantification identifies successful viral transduction Western blot analysis of cell lysates taken from FACS and antibiotic resistance sorted cells identifies protein expression levels a) WDR5 b) MBD3 c) EOMES d) MESP1. E) quantification of western blots shows significant overexpression of MESP1, WDR5 and EOMES, although not MBD3. Analysed using one-way ANOVA, with three blots for each overexpression. *P<0.05, ****P<0.0001

## 5.5 Overexpression of potential binding partners effects cell proliferation

Overexpression of genes can cause changes in cell metabolism and gene regulation, as well as cell homeostasis. This can include changes to cell proliferation. To measure if overexpression of *Mesp1* or its binding partners affects cell proliferation, an EDU assay was utilised. EDU, a thymidine analogue, is integrated into the new strands of DNA created in proliferating cells. This can be identified through the addition of a fluorescent azide which cross links to the EDU in click-chemistry [223].

To negate for the effect of sheer stress on cell proliferation when cells undergo FACS sorting, untransduced 3T3 cells underwent FAC sorting before recovery and EDU assay. In comparison to unsorted wildtype 3T3 cells, there appears no change in cell proliferation when cells are FACS sorted, with approximately 50% of cells being proliferative (Figure 28).

In cells transduced with *Mbd3* or *Eomes*, there appears no change in cell proliferation levels in comparison to wildtype 3T3 cells. There is a slight decrease to approximately 40% in cells transduced with *Mesp1,* though not significant. In contrast, cells overexpressing *Wdr5* show a significant decrease in cellular proliferation levels, from over 50% in control cells to an average of 42%.

This indicates that *Wdr5* expression effects cellular processes in 3T3 cells as indicated by a change in cell proliferation rates. By what mechanism this occurs is unclear, as viral transduction alone has no effect.

Figure 28

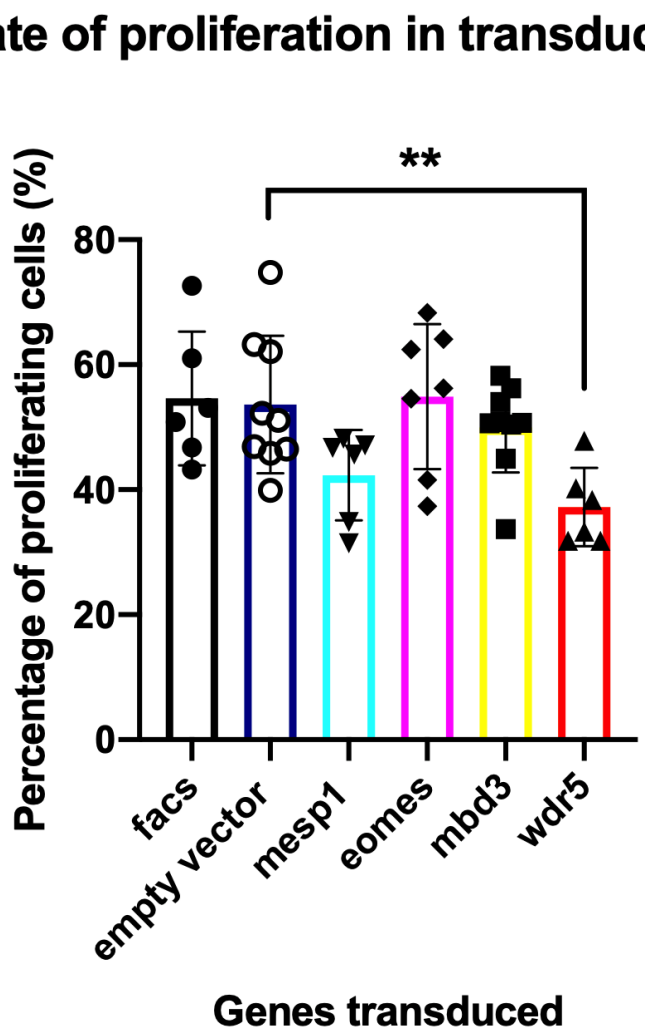
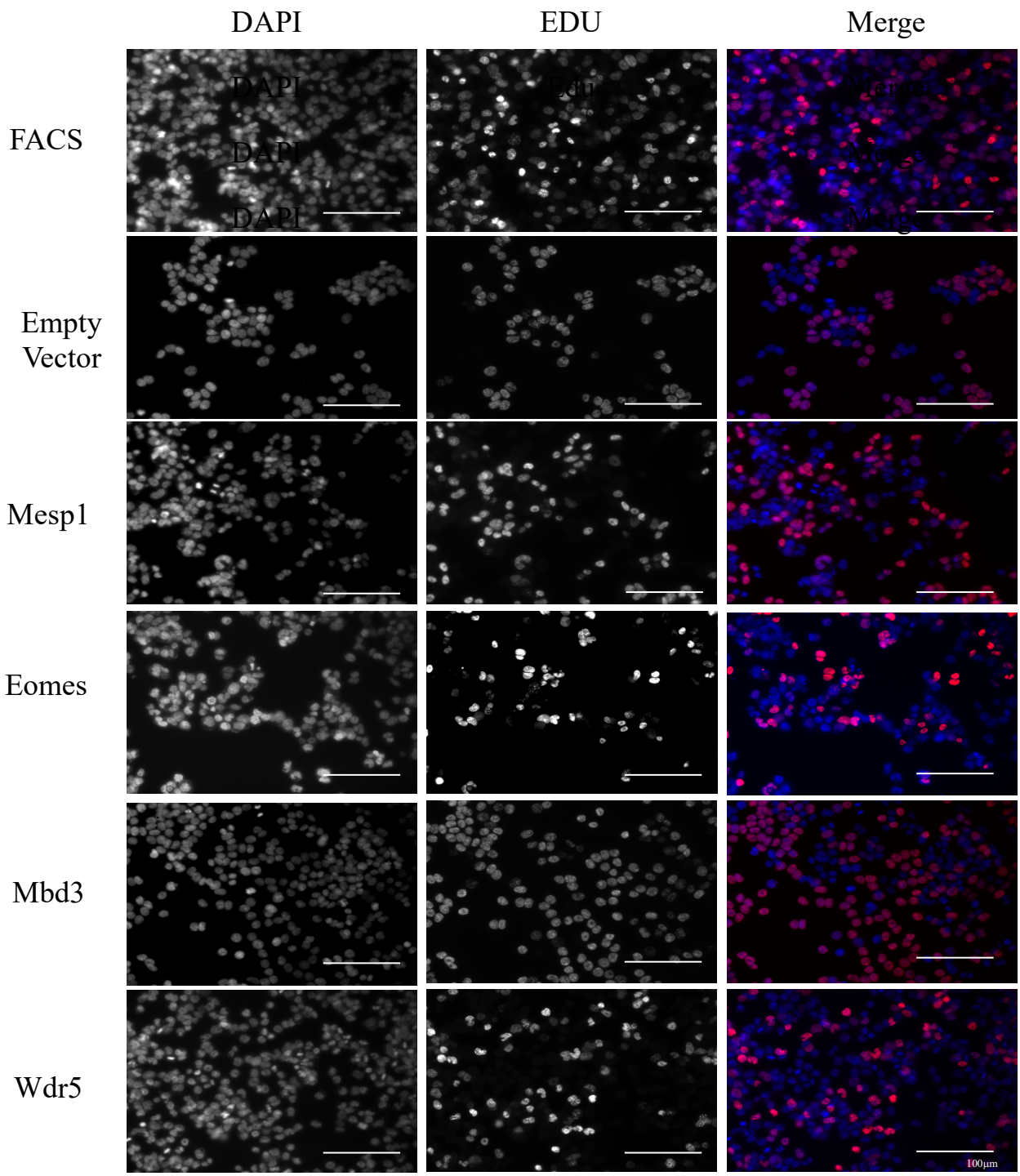



Figure 28 Transduction of *Wdr5* causes a reduction in cell proliferation. Transduction of *Mesp1* or its binding partners *Mbd3* and *Eomes* does not affect cellular proliferation as measured by an EDU assay, however transduction of *Wdr5* causes a reduction in proliferation. Data was obtained through analysing two wells from three transductions of each cell line, shown as an average ± SEM. **P<0.01. Results are from a minimum of two wells from three transductions, 6 wells in total for each. Total number of cells 6260. Results analysed using one way ANOVA ($p<0.05$) followed by Dunnett's multiple comparisons test (*Wdr5* $p<0.05$)

Representative images



## 5.6 Overexpression of binding partners in 3T3 cells does not affect cellular death.

To establish whether transduction of the genes affected cellular processes, such as necrosis and apoptosis, an Annexin V assay was performed to track these processes within the cells. Cells were measured at 3 hours post plating, when cells had adhered, and then at 24-hour time periods for 2 days. Plates were read to measure phosphatidylserine (PS) exposure to the outer leaflet of the cell membrane, the exposure of PS which only occurs during apoptosis. This allows the binding of two subunits of the Annexin V fusion proteins, which could be measured through a luminescent value using the Annexin V assay. In addition, necrosis could be measured through a fluorescent value, through a cell impermeable dye that binds to DNA, which can only occur when membrane integrity is lost.

Necrosis is an energy independent form of cell death, which in contrast to apoptosis results in karyolysis and loss of membrane integrity [224]. Examination of this process to establish if there was any change in average necrosis over 3 days, between cells transduced with *Mesp1* and its binding partners, in combination or singularly, showed no significant change in necrosis in comparison to empty vector controls (Figure 29A)

Apoptosis, by comparison to necrosis, is a cell programmed form of death that is utilised to maintain cellular homeostasis during development to ensure cell number population maintenance [224]. Analysis of the data showed that whilst there was some variation between the rate of apoptosis over the time course, there was no change in the average rate of apoptosis in comparison to empty vector controls (Figure 29B).

Figure 29

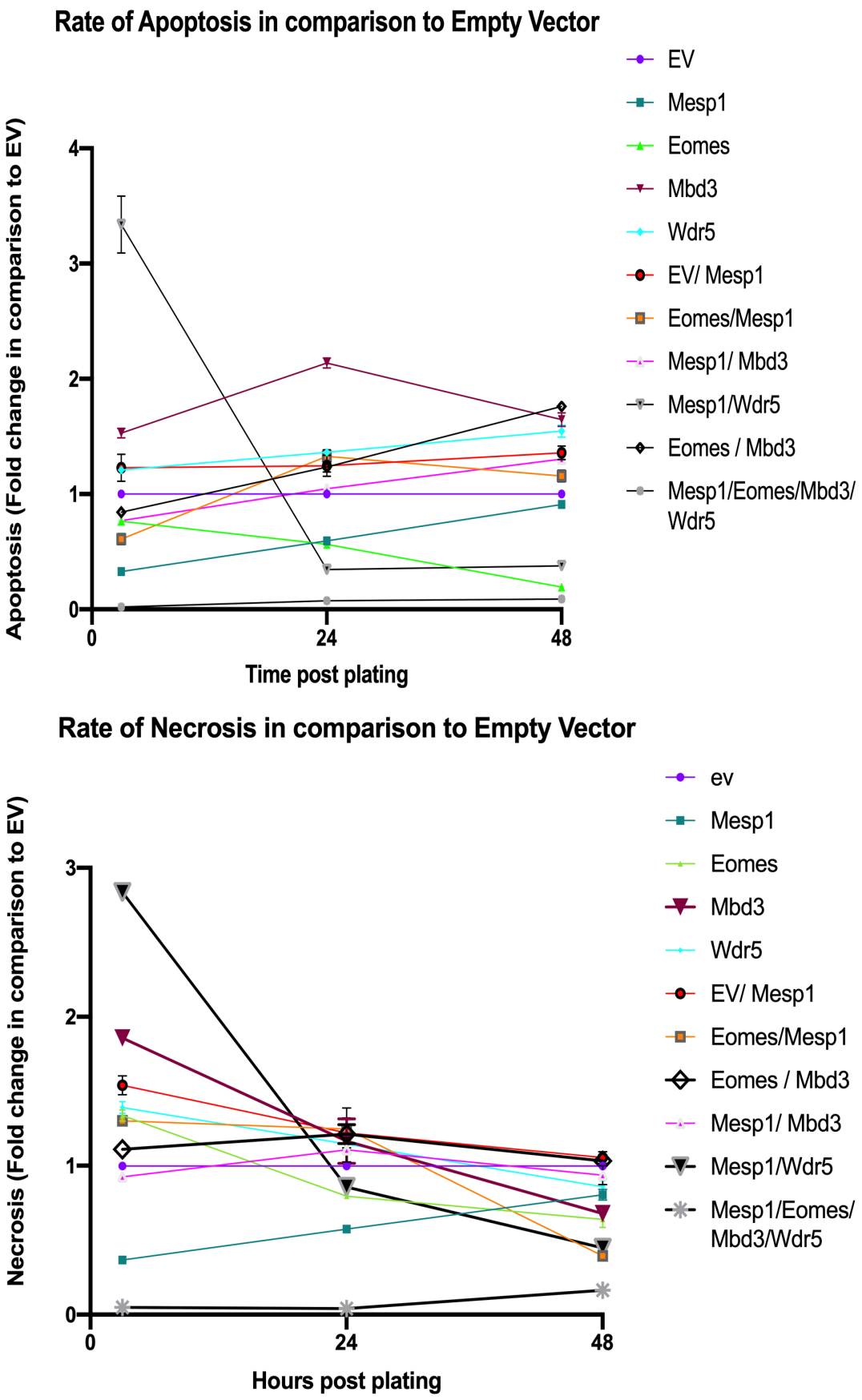



Figure 29 A-B Transduction of viral vectors does not affect cellular death processes A) Apoptosis B) Necrosis. Both measured over 48 hours, using an Annexin V assay, analysing three wells of each transduced cell line, shown as an average ± SEM., analysed by one-way ANOVA, results not significant.

## 5.7 Overexpression of potential binding partners alters cellular morphology.

It is important to evaluate if overexpression of a gene, or set of genes, elicits an effect on the cell transduced. One measurement that can provide information is cell morphology, whereby perturbations in gene expression can induce alterations in morphological features. Through the use of computer modelling and analysis of morphological features, prediction of cell type has been possible [79]. A change in cell shape may be the first indicator that overexpression of *Mesp1* and its potential binding partners may induce a change in cellular fate.

One of the potential binding partners of MESP1, WDR5, has been shown to be important in regulating cell morphology. Overexpression of *Wdr5* has previously been shown to alter both cellular and nuclear morphology, including nuclear circularity and stiffness[225]. The loss of *Wdr5* results in a reduction in the methylation of H3K4, resulting in a more compact chromatin confirmation[225] and changes to gene transcription. An additional cytoskeletal role for WDR5 has also been delineated with the protein essential for stabilising F-Actin and regulating cellular polarity[219].

To analyse the effects of overexpression of *Mesp1* and its binding partners on 3T3 cellular processes, cells were stained with Alpha Tubulin and DAPI to visualise cellular and nuclear shape.

### 5.7.1 Cell Area

To identify effects on cellular processes, Alpha Tubulin staining was implemented to allow visualisation of the cytoplasm of the cell. Cells were individually outlined using Fiji software to ensure accurate measurement of the cell area. Overexpression of individual binding partners and *Mesp1* did not result in any significant effect on cell area, with a slight decrease in median cell size and interquartile range in *Mesp1* transduced cells, although with a number of outliers preventing any significance (Figure 30).

A significant difference ($p<0.01$) was observed when *Mesp1* was transduced in addition to the control vector, *Mbd3* or *Wdr5*. These cells exhibited a significant decrease in cell size, with a reduction of median cell size from 4646 pixel squared in empty vector control to 3789, 3653 and 2711 in *Mesp1* plus control, *Mbd3* or *Wdr5* respectively. There was also a significant ($p<0.05$) reduction in cell area in cells transduced with *Eomes* and *Mbd3,* with a reduction from 4646 to 3916 in median cell area. This indicates that *Mesp1,* in

conjunction with its binding partners, has the ability to affect cell dynamics and area, however all genes transduced concurrently does not affect cell area.

Figure 30

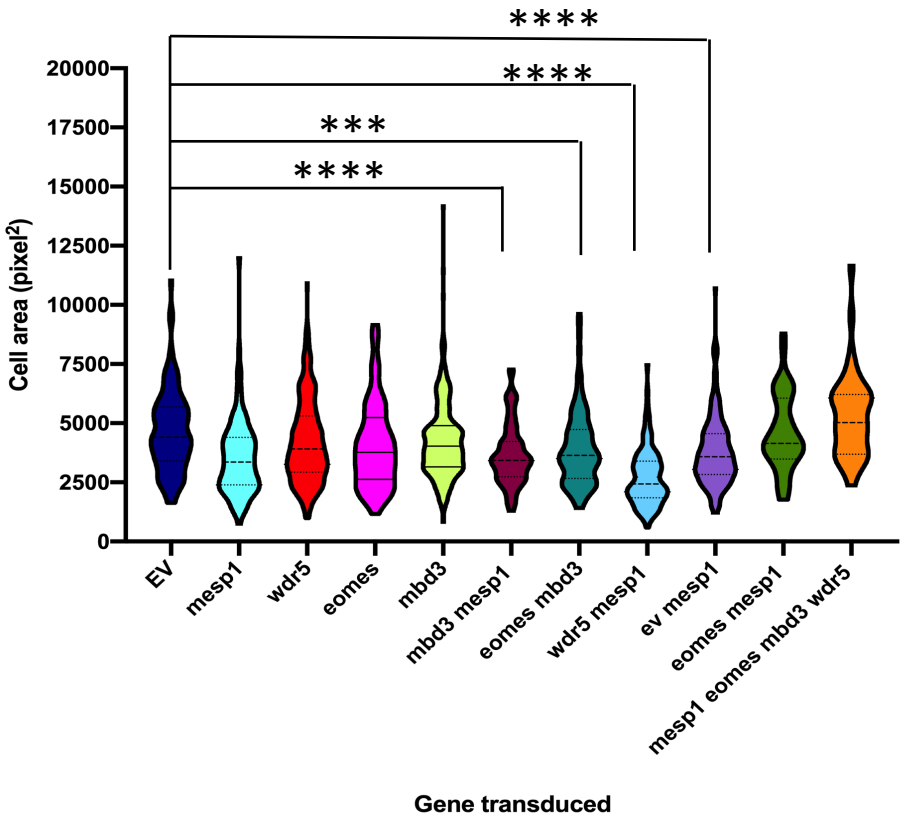



Figure 30 Overexpression of two binding partners simultaneously affects cell area. Overexpression of *Mesp1* and a potential binding partner or two binding partners results in a significant reduction of cell area. Measurements are from three wells with three images taken in each, with representative images. *P<0.05, **P<0.01, ***P<0.001 ****P<0.0001 Analysed using One way ANOVA (p<0.0001) with multiple Dunnett's tests (*P<0.05, **P<0.01, ***P<0.001 ****P<0.0001 in comparison to EV controls).

### 5.7.2 Cell circularity

Another measure of cellular dynamics is cell circularity, with changes in cell morphology indicative of intracellular changes. Cell circularity is given by a measurement between 0 and 1, where 1 is perfectly circular and 0 is a flat line. Mean cell circularity of cells transduced with empty vector controls was 0.6 (Figure 31). Cells transduced with *Mesp1* or *Wdr5* showed significantly increased circularity, with an average of 0.68 and 0.66 respectively ($p<0.001$). This was not observed in 3T3 cells transduced with either *Eomes* or *Mbd3,* that had an average circularity of 0.621 and 0.622 respectively. There was also no change when all genes were transduced simultaneously.

This trend was also observed when *Mesp1* was transduced in addition to empty vector controls, with an increased average cell circularity of 0.67 ($p<0.001$). The addition of *Wdr5* to *Mesp1* also led to a significant increase in average cell circularity, from 0.6 to 0.73($p<0.001$) (Figure 31).

There was no significant difference on the addition of *Mbd3* to *Mesp1* transduced cells, or the dual transduction of *Mbd3* and *Eomes* to 3T3 cells, all leading to an average cell circularity of 0.63 or below (Figure 31). However, the combination of *Mesp1* and *Eomes* did induce a significant change ($p<0.05$) in cell circularity, inducing a reduction in the circularity of the cells.

This indicates that *Mesp1* and *Wdr5* expression may have a role in cellular dynamics that can affect cell circularity, although the mechanism behind this is unidentified. The combination of the transcription factors *Mesp1* and *Eomes* also induces a slight change.

Figure 31

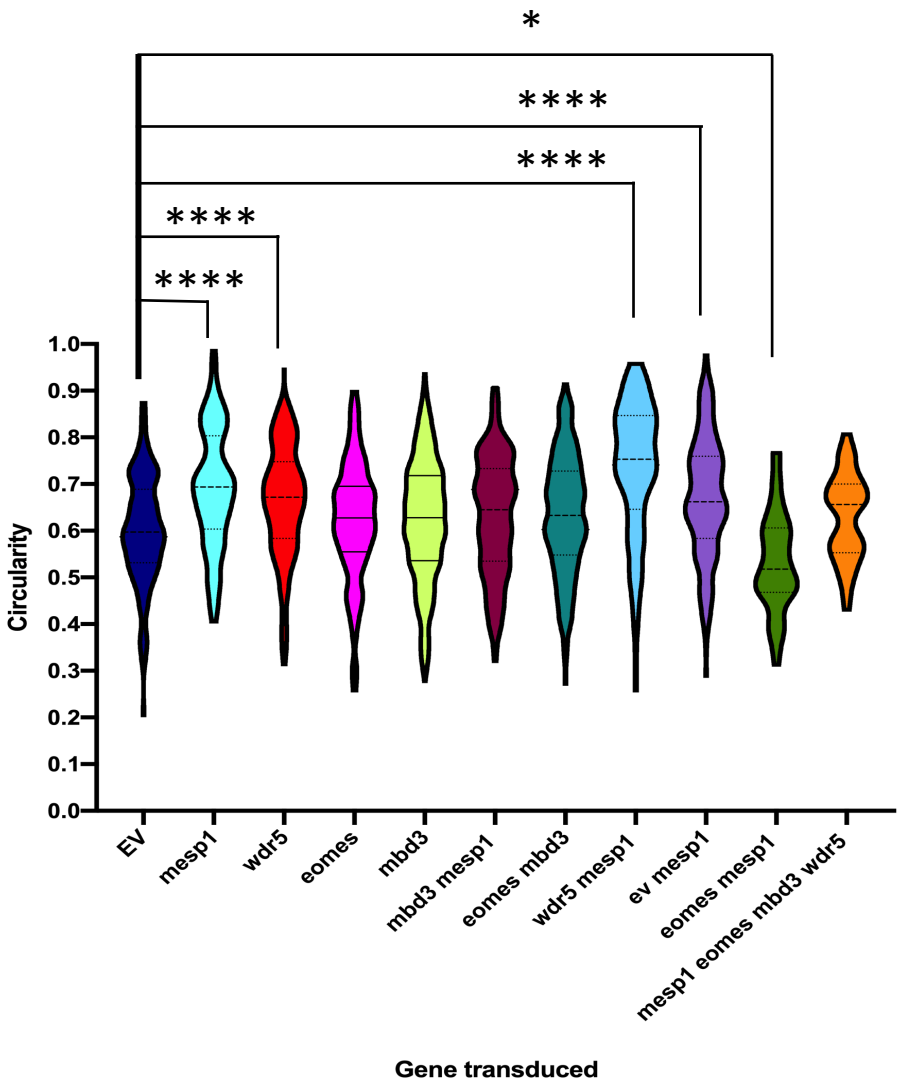



Figure 31 Overexpression of *Wdr5* and *Mesp1* results in increased circularity in 3T3 cells. Overexpression of *Mesp1* and *Wdr5* alone or in concert results in increased circularity in 3T3 cells in comparison to empty vector (EV) controls. Furthermore, the addition of *Mesp1* to EV controls also results in increased circularity. Results from three wells, with three images from each well with representative images. *P<0.05, **P<0.01, ***P<0.001 ****P<0.0001. Analysed using One way ANOVA (p<0.0001) with multiple Dunnett's tests (*P<0.05, **P<0.01, ***P<0.001 ****P<0.0001 in comparison to EV controls).

### 5.7.3 Nuclear Area

In addition to measuring cellular shape, it is also important to recognise any effects on the nucleus. Cell shape can have direct effects on intracellular signalling and nuclear morphology through cytoskeletal changes [226,227].

Empty vector transduced cells had an average nuclear size of 1847 pixels squared. This was significantly altered in cells with an overexpression of *Wdr5*, leading to a significant increase in nuclear area to 2280 pixels squared ($p<0.001$) (Figure 31).
This is in contrast to cells transduced with *Mbd3* where nuclear area is significantly reduced ($p<0.05$) (Figure 32) to an average of 1623 pixels squared.

A reduction in nuclear area was also observed in cells transduced with *Mesp1* and a binding partner, including *Mbd3* or *Wdr5,* with the average nuclear area reduced to 1249 and 1382 respectively ($p<0.001$) (Figure 32). This is a marked reduction in comparison to cells overexpressing *Mesp1* alone, where the average cell area is 1788 pixels squared. However, it should be noted that the addition of an empty vector to *Mesp1* expression also causes a reduction in nuclear area to 1422 ($p<0.001$) (Figure 32).

A similar reduction is also observed in cells overexpressing both *Eomes* and *Mbd3,* where nuclear area is reduced from 1847 to an average of 1510 ($p<0.01$). This is a reduction in nuclear area compared to expression of *Mbd3* alone. There was no significant change in nuclear area when *Mesp1* and *Eomes* were co-expressed, nor when all genes were transduced simultaneously.

These results indicate that overexpression of *Mesp1* and its histone modifying binding partners *Wdr5* and *Mbd3* affects nuclear area in transduced 3T3 cells. Whilst overexpression of either *Mesp1* or *Eomes* does not seem to significantly effect nuclear size alone, in combination with *Mbd3* nuclear size is reduced.

This suggests that overexpression of these proteins can cause changes in nuclear morphology.

Figure 32

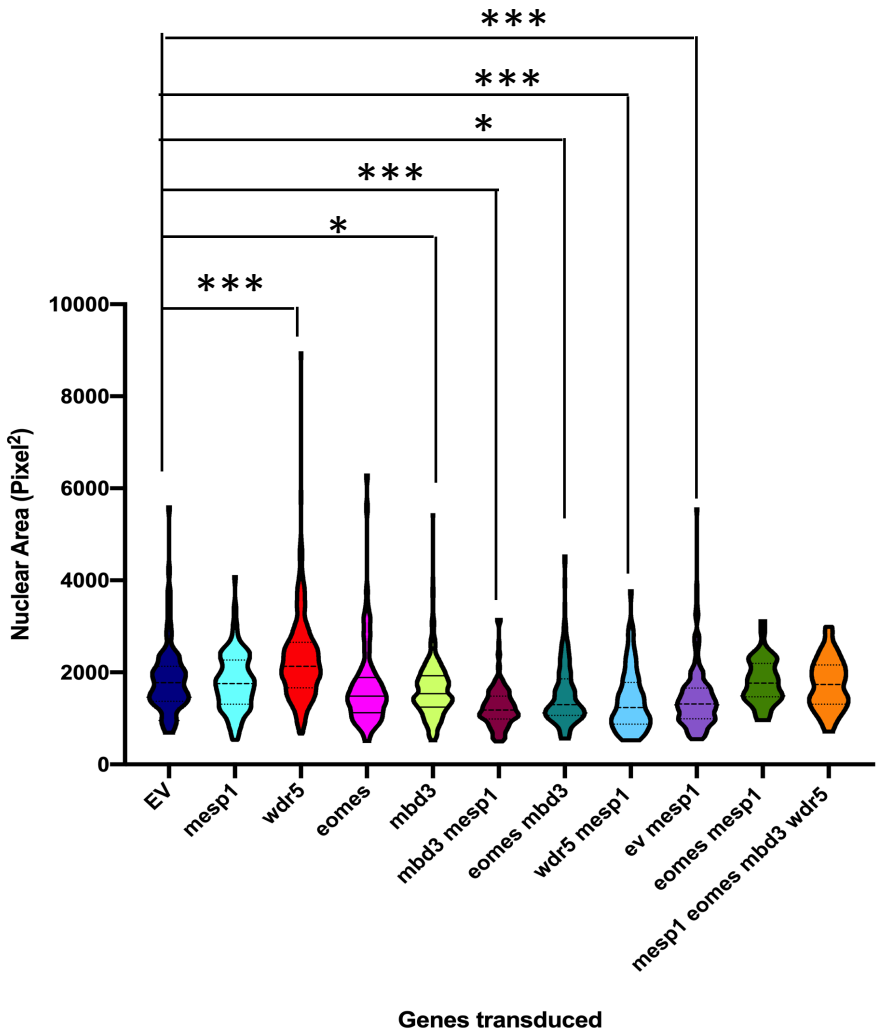



Figure 32 Overexpression of potential binding partners in 3T3 cells causes a change in nuclear area. Overexpression of *Wdr5* induces an increase in nuclear area in 3T3 cells, whilst expression of Mbd3 reduces nuclear area. Expression of *Mbd3* in combination with either *Eomes* and *Mesp1* decreases nuclear area, as does *Mesp1* in combination with *Wdr5* or EV controls. Results from three wells with three images in each well, representative images shown. *P<0.05, **P<0.01, ***P<0.001 ****P<0.0001. Analysed using One way ANOVA (p<0.001) with multiple Dunnett's tests (*P<0.05, **P<0.01, ***P<0.001 ****P<0.0001 in comparison to EV controls).

### 5.7.4 Nuclear circularity

Whilst nuclear size is an important indicator of the effect of the binding partners on cell morphology, it is also essential to ascertain any changes in nuclear circularity. Changes to nuclear circularity can be indicative of changes to chromatin texture, as well as nuclear deformity being linked to DNA damage [228,229].

Nuclear circularity of control empty vector transduced was on average 0.8. Overexpression of *Mesp1* or *Eomes* resulted in a significant decrease in circularity, to an average of 0.73 and 0.7 respectively (Figure 33). ($p<0.001$). This effect was exacerbated in 3T3 cells transduced with *Wdr5*, where average nuclear circularity was decreased from 0.8 to 0.63 (Figure 33). ($p<0.001$). This was similarly observed in cells where *Mesp1* and *Eomes* were co-expressed, with a significant reduction in nuclear circularity, though not to the extent seen in cells expressing *Mesp1* or *Eomes* singularly ($p<0.05$).

A similar effect was observed in cells overexpressing both *Mesp1* and *Wdr5,* where nuclear circularity decreased from an average of 0.8 to 0.74 (Figure 33). ($p<0.01$). A reduction in nuclear circularity was also observed when all genes were transduced simultaneously.

In contrast to its effects in nuclear area, expression of *Mbd3* alone, or in addition to *Mesp1,* showed no significant changes in nuclear circularity, with an average of 0.77 and 0.78 respectively (Figure 33).

This indicates that *Wdr5 and Mesp1* alone or in concert have an effect on nuclear circularity. Overexpression of *Eomes* alone seems to cause a reduction in circularity, whilst the addition of *Mbd3* negates this trend. Therefore, it is apparent that expression of *Mesp1,* and the histone modifier *Wdr5,* causes changes in nuclear morphology.

Figure 33

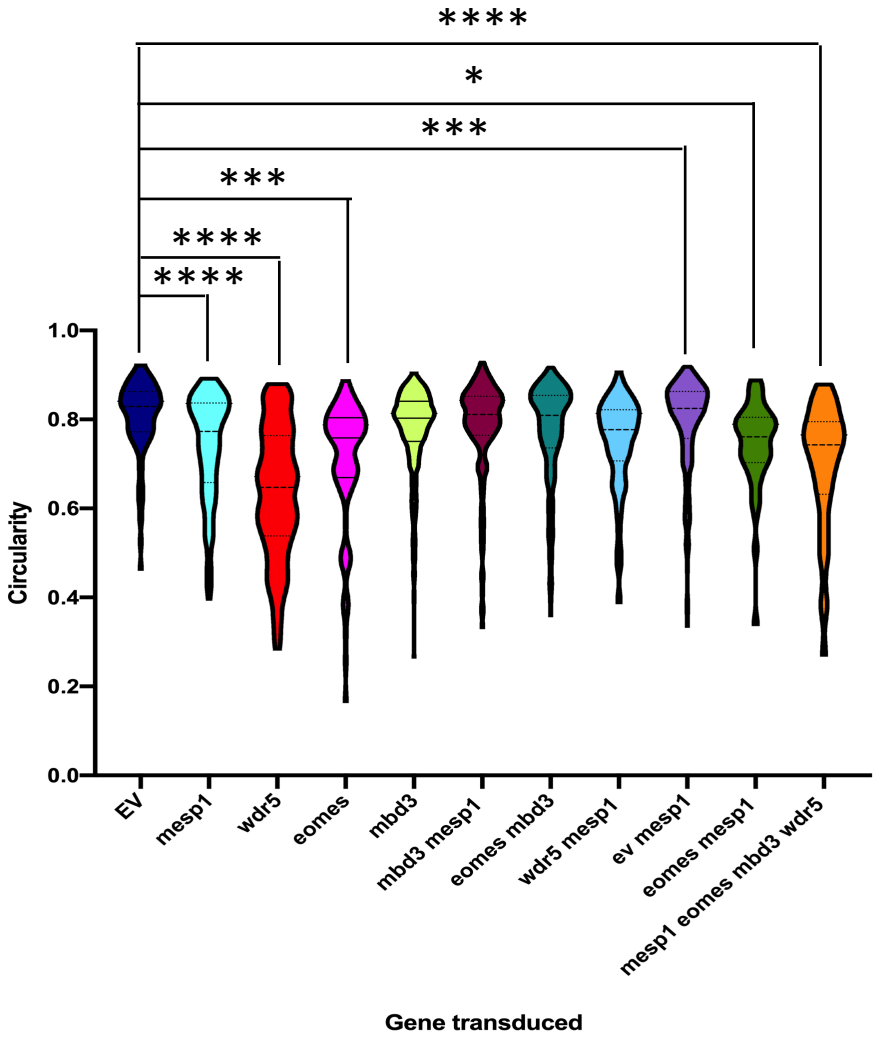



Figure 33 Overexpression of *Mesp1, Wdr5, Eomes* causes a decrease in nuclear circularity in 3T3 cells. Overexpression of *Mesp1, Wdr5* and *Eomes* in 3T3 cells causes a reduction in nuclear circularity. Overexpression of both *Mesp1* and *Wdr5* together also reduced nuclear circularity. Results are from three wells with three images in each well, representative images shown. *P<0.05, **P<0.01, ***P<0.001. Analysed using One way ANOVA (p<0.001) with multiple Dunnett's tests (*P<0.05, **P<0.01, ***P<0.001 ****P<0.0001 in comparison to EV controls).

### 5.7.5 Nuclear to cell ratio

Besides changes in cell and nuclear morphology, it is also important to understand the relationship between nuclear and cellular size. Changes to the ratio of cytoplasm to nucleus can indicate changes in cell function and also aberrant cell behaviours [230].

On average, cells transduced with a control vector maintained a nuclear to cell ratio of 0.43. This ratio was disturbed upon the transduction of either *Mesp1* or *Wdr5,* with an increased nuclear to cell size ratio of 0.52 and 0.6 respectively ($p<0.001$) (Figure 34). This trend was also seen to a smaller effect in cells where both genes were overexpressed, with the ratio of 0.52 ($p<0.01$) (Figure 34).

Cells transduced with *Eomes* or *Mbd3* show little variation in nuclear to cell ratio at 0.42 and 0.40 respectively, with similar values observed in cells transduced with *Mesp1* and *Mbd3, or Eomes* and *Mbd3*. However, cells transduced with both *Eomes* and *Mesp1* result in a significant increase in the nuclear to cell ratio ($p<0.001$). This could be in part due to the large variation in the nuclear to cell ratio in cells transduced with the transcription factors. However, there is a shift in the interquartile range of the majority of the cells, suggesting these factors induce an increase in the nuclear to cell ratio. There was no significant change when all genes were transduced simultaneously.

These findings indicate that changes in nuclear morphology, including size and circularity, induced by *Mesp1* and *Wdr5* overexpression, also lead to changes in the relationship between the nucleus and the cell. There is also a pattern indicating that *Eomes* and *Mesp1* co-expression leads to a decrease in cell and nuclear circularity, and an increase in the nuclear to cell ratio of the cell.

Figure 34

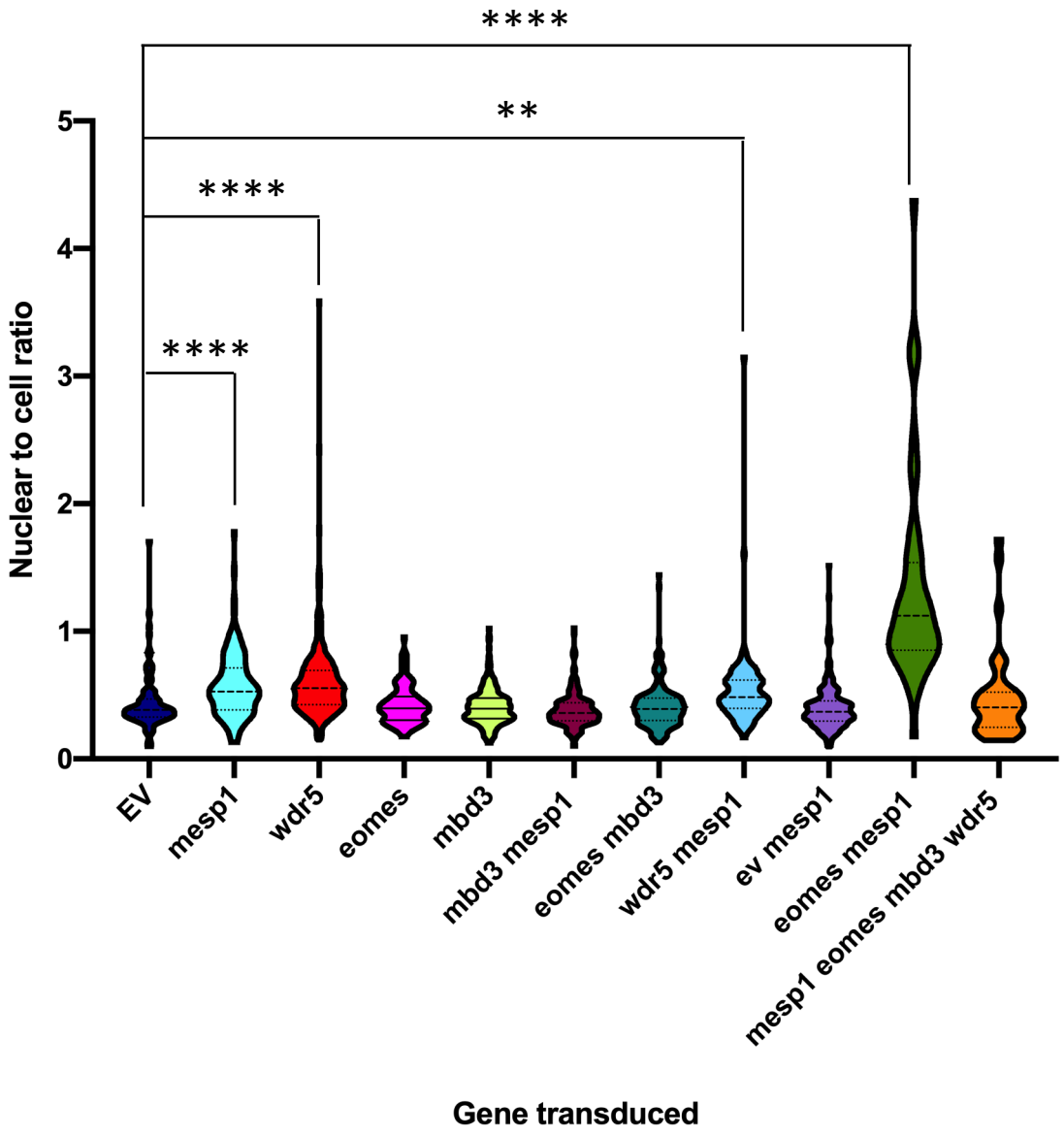
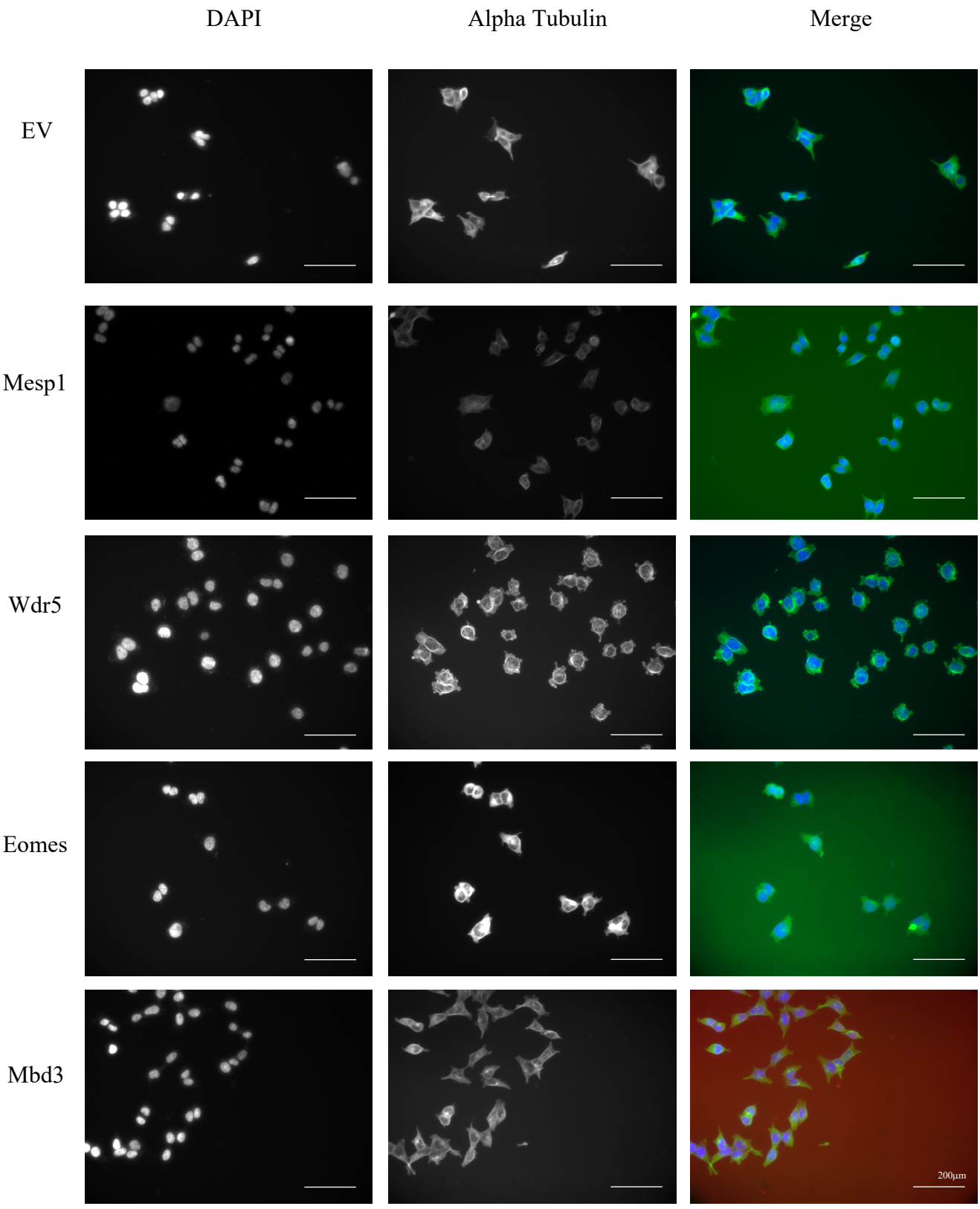
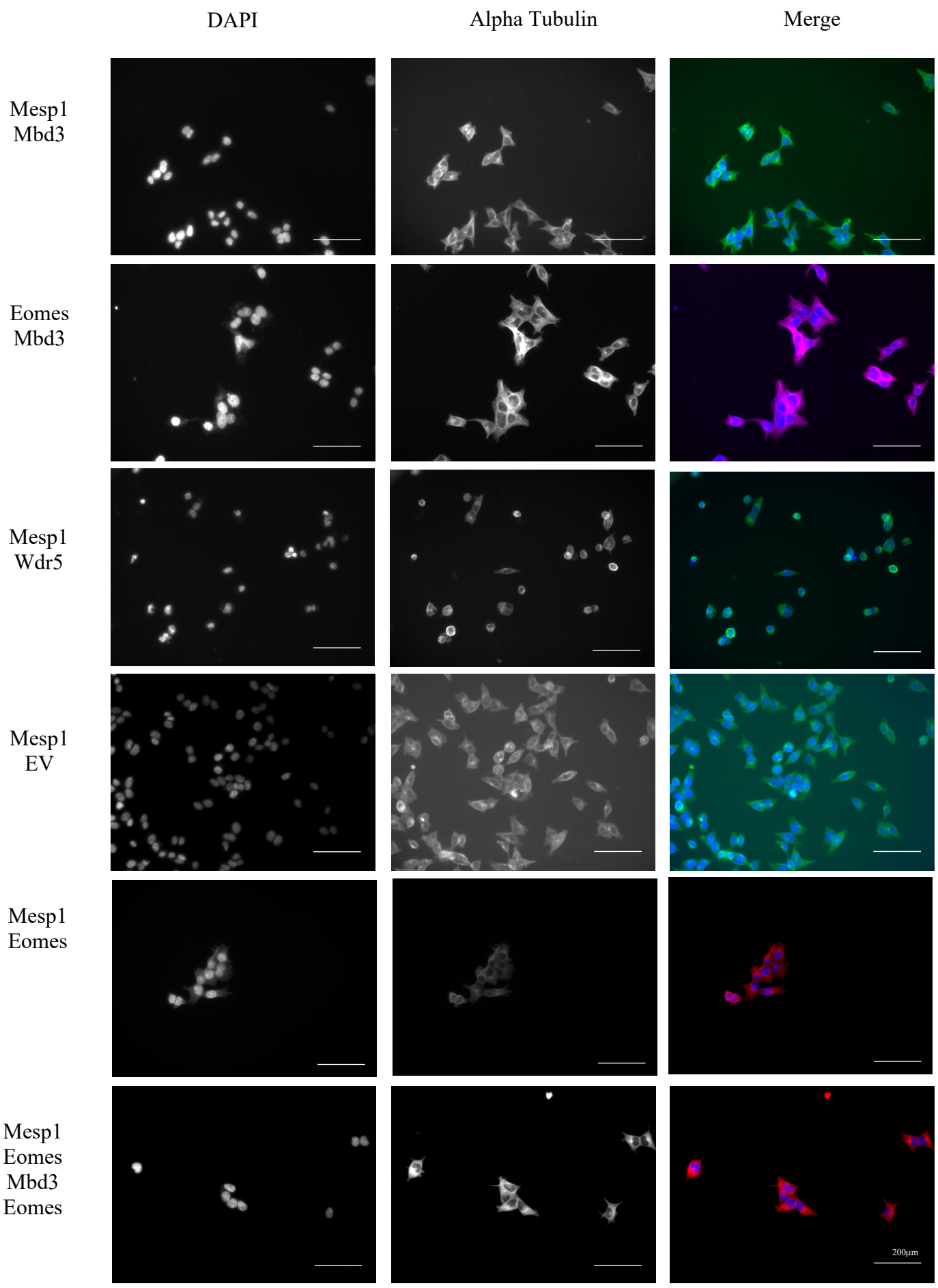



Figure 34 Overexpression of *Mesp1* and *Wdr5* affects nuclear to cell ratio. Overexpression of *Mesp1* or *Wdr5* induces an increase in nuclear to cell area ratio, which is also seen in double overexpression of *Mesp1* and *Wdr5* although to a lesser extent. Results are from three wells with three images in each well, representative images shown. *P<0.05, **P<0.01, ***P<0.001 ****P<0.0001. Analysed using One way ANOVA (p<0.001) with multiple Dunnett's tests (*P<0.05, **P<0.01, ***P<0.001 ****P<0.0001 in comparison to EV controls).

Representative images




DAPI
Alpha Tubulin
Merge
Mesp1
Mbd3
Eomes
Mbd3
Mesp1
Wdr5
Mesp1
EV
Mesp1
Eomes
Mesp1
Eomes
Mbd3
Eomes
200µm

### 5.7.6 Overexpression of a single potential binding partner alters gene expression

In order to understand the capability of *Mesp1* and its binding partners to aid in the reprogramming of somatic cells into a cardiac progenitor lineage, it is important to identify the changes gene overexpression has on lineage markers.

In order to achieve this, genes were overexpressed in singular and in combination with *Mesp1* and other binding partners. It is important to note that *Mesp1* is acknowledged as contributing to not only cardiac, but also haematopoietic and cranial facial muscle lineages[135]. To date, *Mesp1* alone is not sufficient to reprogram somatic cells to a cardiac lineage [71]. Therefore, genetic markers of fibroblasts, haematopoietic, and cardiac lineages were analysed using qPCR of RNA from lentivirally transduced 3T3 cells to assess the effect of overexpression on reprogramming genes.

## 5.8 Overexpression of potential MESP1 binding partners induces expression of *Mesp1*

Overexpression of *Mesp1* in 3T3 cells was efficient, with a statistically significant 150-fold increase in *Mesp1* expression in comparison to empty vector controls ($p<0.001$). There was little change in fibroblast marker expression, with levels of *Fsp-1* and *Col1a1* remaining similar to controls. Similarly, there was little effect on the expression of the early haematopoietic marker *Runx1*, which whilst elevated 2-fold, was not significantly increased. Overexpression of *Mesp1* shows little effect on the expression of its binding partners, *Mbd3, Eomes* or *Wdr5*, with levels of both histone modifiers and *Eomes* comparable to baseline levels in control transduced samples. There is a slight increase in the mesodermal marker *T*, which is known to act upstream of *Mesp1*, at approximately 2-fold baseline levels, which is also observed in cardiac progenitor marker *Nkx2.5*. Furthermore, there is a slight, but not statistically significant, increase in the cardiac progenitor marker *Gata4,* with expression 5-fold of that in control cells (Figure 35A).

Conversely, there is an increase in *Mesp1* observed when any of the potential binding partners are overexpressed. Overexpression of *Eomes* leads to a 50-fold increase in *Mesp1* ($p<0.05$), whilst overexpression of *Mbd3* and *Wdr5* resulted in a 400-fold ($p<0.001$) and 12-fold increase in *Mesp1* levels respectively (Figure 35 B, C,D).

Expression of either *Eomes* or *Mbd3* results in an increase in the cardiac progenitor marker *Gata4,* whilst transduction of *Wdr5* or *Mbd3* leads to an increase in *Nkx2.5* expression. However, expression of the cardiac progenitor markers is not significantly

increased, in part due to variable levels between replicates and low levels found in control samples.

Overexpression of a potential binding partner can also cause variations in the expression of another binding partner. Interestingly, overexpression of *Eomes* resulted in a reduction in the expression of *Wdr5*, though not to significant levels. By contrast, transduction of *Wdr5* resulted in a 1.6-fold increase in *Eomes* expression (Figure 35D). Expression of *Eomes* was also altered when *Mbd3* was overexpression, with a 5-fold increase in *Eomes* transcript levels in comparison to empty vector controls.

Figure 35

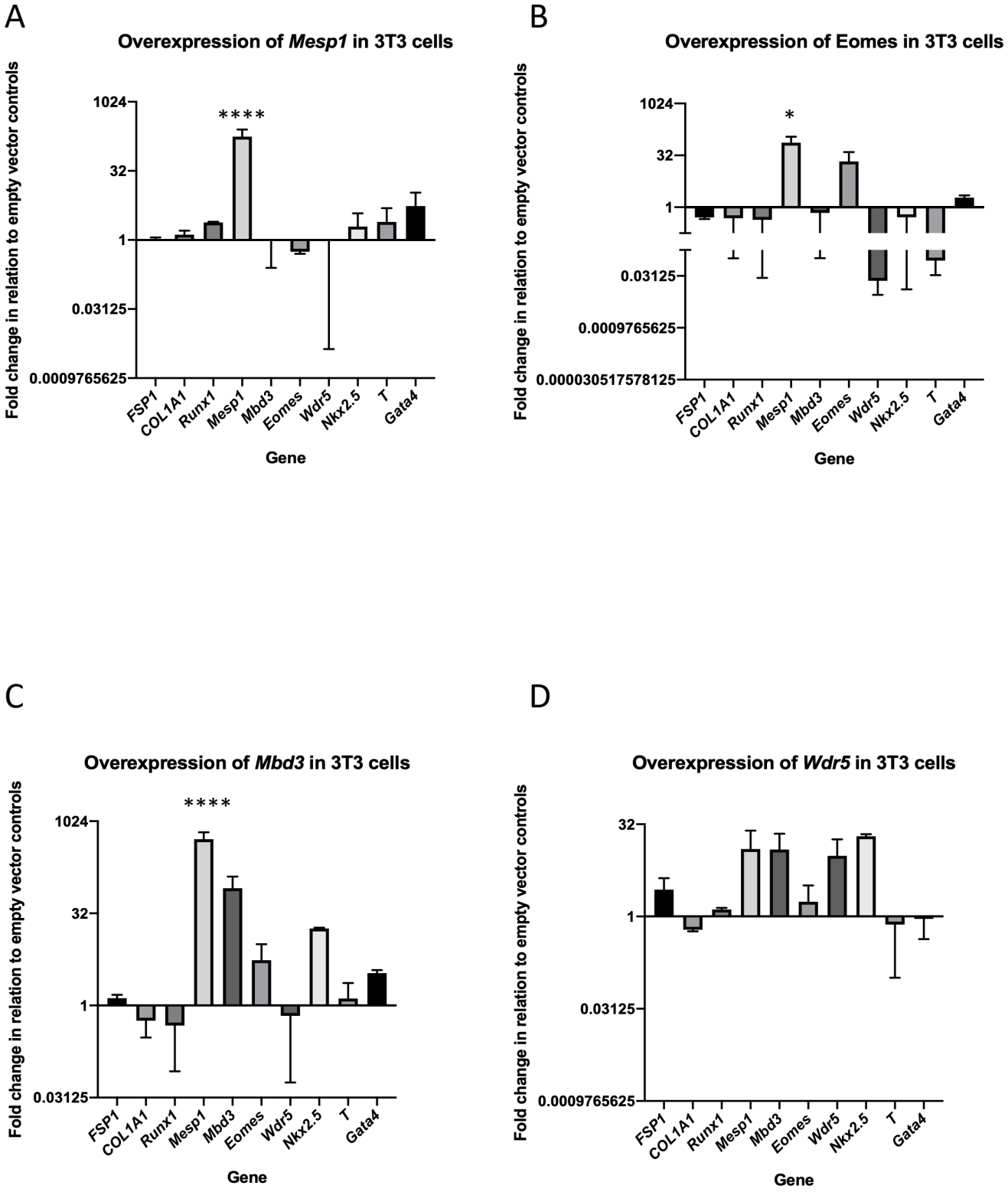



Figure 35 Overexpression of potential binding partners in 3T3 cells induces cardiac gene expression A) Overexpression of *Mesp1* induces expression of *Gata4, Nkx2.5* and *T*. B) Overexpression of *Eomes* in 3T3 cells induces expression of *Mesp1* and *Gata4*. C) Overexpression of *Mbd3* induces *Mesp1, Eomes, Nkx2.5* and *Gata4* in 3T3 cells. D) Overexpression of *Wdr5* in 3T3 cells induces expression of fibroblast marker *Fsp1, Mesp1, Mbd3,* and *Nkx2.5.* Expression is relative to empty vector control, normalised to a house keeping gene, average of three transductions ±SEM. Analysed by one-way ANOVA, a)$p<0.005$ b)ns $p<0.06$ c) $p<0.0001$ d)ns $p<0.19$ followed by Dunnett's multiple comparisons tests in comparison to EV . $p<0.001$ ****, $p<0.05$* n=3

## 5.9 Co-expression of *Mesp1* and a potential binding partner results in a variation of potential binding partner and marker gene expression

Analysis of qPCR data from cells overexpressing both *Mesp1* and an empty vector control indicate that *Mesp1* is highly expressed when compared to empty vector control alone, at approximately 800-fold control levels. In addition to changes in *Mesp1*, there is a 2-fold increase to the MESP1 binding partner *Mbd3*, whilst *Eomes* stays at control levels and *Wdr5* marginally decreases. There is an increase in fibroblast marker *FSP-1* to approximately 5-fold that of control cells, however, a reduction in *Col1A1* to baseline levels. *Nkx2.5* levels increase 3-fold, which is higher than that seen in *Mesp1* overexpression alone. There is also reduction in *T* and *Gata4* levels in comparison to samples when *Mesp1* is overexpressed singularly (Figure 36A).

Transduction of *Mesp1* and its binding partner *Mbd3* results in a 70,000-fold expression of *Mesp1* and 12-fold increase in *Mbd3* expression in comparison to control cells (Figure 36B). *Mesp1* is not endogenously present in 3T3 cells, therefore, sees dramatic increases when overexpressed. However, a lack of replicates means that the significance of this increase cannot be confirmed.

Whilst an overexpression of *Mesp1* and *Mbd3* has minimal effects on fibroblast markers, other potential binding partners *Eomes* and *Wdr5* are reduced. A reduction is also observed in mesodermal marker, *T,* and cardiac progenitor marker, *Gata4* (Figure 36B).

In contrast, a 6-fold increase in *Nkx2.5* is observed in comparison to control vector transduced 3T3 cells. Further replicates are needed to confirm if this is reproduceable and significant (Figure 36B).

In correlation with results from overexpression of *Mesp1* and *Mbd3,* a similarly large expression of *Mesp1* is observed during the dual transduction of *Mesp1* and *Wdr5* into 3T3 cells. Levels of *Mesp1* were observed to increase to 70,000-fold that seen in control vector cells (Figure 36C). *Wdr5* levels increased to 2.5-fold that seen endogenously, indicating there is overexpression in the cells. This dramatic fold increase in *Mesp1* can be explained by low endogenous levels in 3T3 cells, however, further replicates would be needed to identify if this was significant.

In agreement with previous results, expression of *Mesp1* and *Wdr5* causes a reduction in other potential binding partners, with a slight reduction observed in *Mbd3* and *Eomes* to below control levels. This reduction was also observed in mesodermal marker *T,* and the cardiac progenitor markers *Nkx2.5* and *Gata4.* Fibroblast marker *FSP-1* shows a marginal

increase to approximately 2.5-fold baseline level, whilst *Col1A1* remains steady at baseline levels.

The transcription factors, *Eomes* and *Mesp1*, are known to be essential to cardiac differentiation. Overexpression in 3T3 cells results in a slight increase in *Gata4* expression, a known cardiac progenitor cell gene. There was no corresponding increase in *Nkx2.5* expression, another cardiac progenitor cell marker (Figure 36D). There is an increase in *FSP-1* expression, a key fibroblast marker, which could suggest a transition towards a cardiac fibroblast lineage.

Figure 36

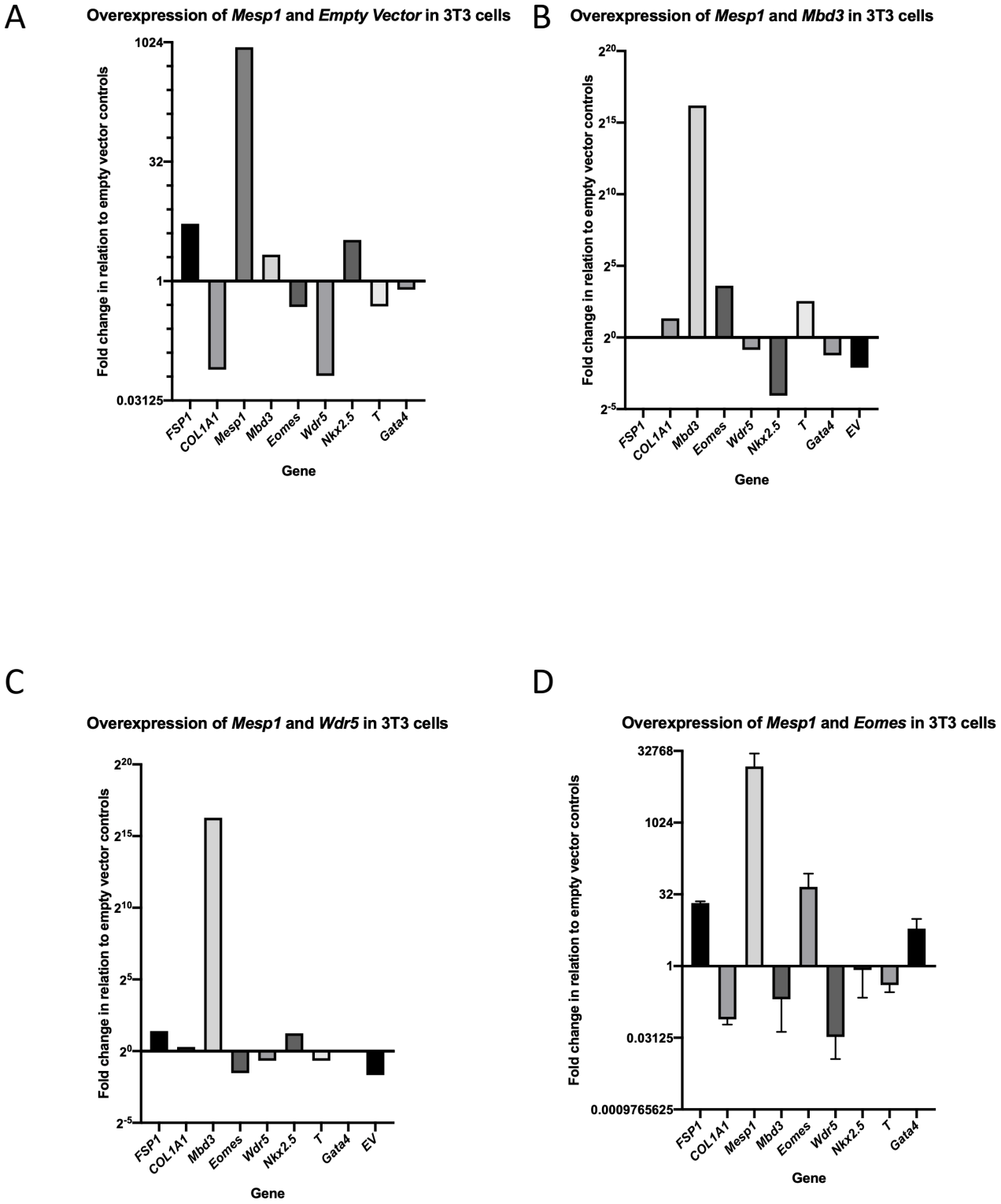



Figure 36 Overexpression of *Mesp1* and a potential binding partner induces fibroblast and cardiac marker genes. A) Overexpression of *Mesp1* and Empty Vector induces expression of *Nkx2.5* and *Fsp1.* B) Overexpression of *Mesp1* and *Mbd3* induces expression of *Col1a1* and *Nkx2.5* as well as their own expression. C) Overexpression of *Mesp1* and *Wdr5* induces their own expression and *Fsp1*. Results from 1 transduction, apart from *Mesp1* and *Eomes* overexpression n=3, analysed by one-way ANOVA, ns $p<0.29$, followed by Dunnett's multiple comparisons test in comparison to EV, ns.

## 5.10 Expression of *Mesp1* in combination with two potential binding partners induces *Gata4* expression

Transduction of *Mesp1* and *Mbd3,* in combination with either *Eomes or Wdr5,* results in an increase in *Gata4* expression to over 10-fold that seen in control samples (Figure 37A). *Mesp1* is highly expressed in both samples, at between 6000 and 30,000-fold that witnessed in control samples respectively. In samples where either *Eomes* or *Wdr5* was transduced in addition to *Mesp1,* expression of *Eomes* is between 3 to 5-fold higher that of control cells.

Levels of fibroblast markers vary between the 2 samples, with high levels of *Col1A1* observed in cells transduced with *Wdr5* in addition to *Mesp1* and *Mbd3* (Figure 37B). In contrast, both fibroblast markers are greatly reduced in cells transduced with *Eomes*, *Mesp1* and *Mbd3.*

Additional replicates are needed to confirm the significance of the increase observed in *Gata4* levels, as well as the reproducibility of these results.

## 5.11 Transduction of *Mesp1* and its potential binding partners results in a significant increase in *Mesp1* and fibroblast marker *FSP-1* levels

Transduction of *Mesp1* in combination with all of its potential binding partners results in a significant increase in *Mesp1* expression, to 2-fold that observed in control samples ($p<0.05$) (Figure 37C). An increase in *Mesp1* expression has been witnessed in each sample transduced with *Mesp1,* irrespective of the other genes co-expressed. *Mesp1* in combination with *Eomes, Mbd3* and *Wdr5* induces a statistically significant increase in *FSP-1* transcript levels, to 3-fold that observed in control samples ($p<0.001$). This increase in *FSP-1* was also observed when *Mesp1* was transduced alongside either *Eomes* or *Mbd3,* although not with *Wdr5.* Conversely, expression of *FSP-1* was increased by a similar amount when *Wdr5* was transduced alone. It should also be noted that *FSP-1* was also increased, to 5-fold control levels in *Mesp1* and *empty vector* transduced cells, although results were not statistically significant and were limited to 1 replicate. Further assays, including immunocytochemistry and qPCR with additional marker genes, are necessary to identify the effect of overexpression of *Mesp1* and its potential binding partners.

Figure 37

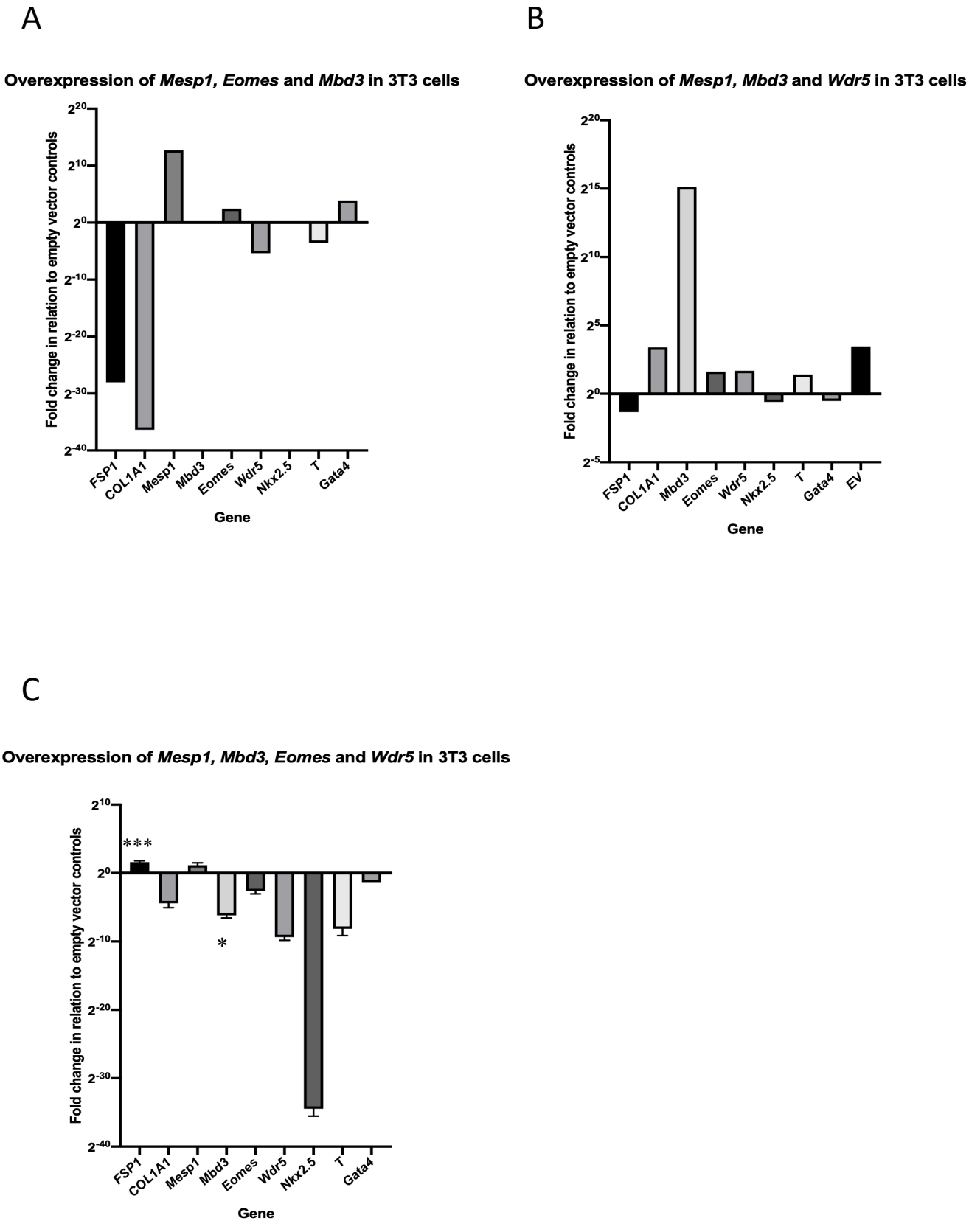



Figure 37 Overexpression of binding partners induces cardiac gene induction. A) Expression of *Eomes, Mbd3* and *Mesp1* in 3T3 cells induces *Gata4* expression in addition to *Mesp1* and *Eomes.* B) Overexpression of *Mbd3, Mesp1* and *Wdr5* in 3T3 induces expression of *Col1A1, Eomes* and *Nkx2.5* as well as *Gata4* C) Overexpression of *Mesp1* and its binding partners simultaneously induces expression of *Mesp1* and *Fsp-1*. Results from 1 transduction, apart from *Mesp1, Eomes, Mbd3* and *Wdr5* overexpression n=3, analysed by one-way ANOVA $p<0.001$, followed by Dunnett's multiple comparisons test in comparison to EV $p<0.001$ ****, $p<0.05$* n=3

## 5.12 Summary

It is important to be able to produce homogenous cell populations in the course of a reprogramming protocol. Experiments here indicate that repeated transductions result in the highest efficiency of lentiviral transduction. This protocol was then utilised to enable the investigation of the effects that *Mesp1,* and its binding partners, have on 3T3 cell morphology and genetic markers, that may indicate their role in reprogramming.

Data produced through lentiviral transduction indicates that *Mesp1* and its binding partners, either singularly or in conjunction with each other, have the ability to modulate cell morphology. *Mesp1* in combination with either *Mbd3* or *Wdr5* causes a reduction in cell area, whilst *Mesp1* and *Wdr5,* either in isolation or in combination, result in an increase in cell circuity. Similarly, *Mesp1,* in combination with *Wdr5* or *Mbd3,* results in a decrease in nuclear area. In contrast, *Wdr5* expression alone causes an increase in nuclear area, however, induces a significant decrease in the nuclear circularity of these cells. This is also true of *Mesp1* and *Eomes* transduced alone, and when *Mesp1* is transduced in conjunction with *Wdr5*. This leads to a final conclusion that *Mesp1* and *Wdr5,* alone or in concert, can cause alterations in nuclear morphology that result in the increase in nuclear size in comparison to cell size.

Overexpression of *Mesp1,* or each binding partner, results in modulation of the expression of other binding partner expression profiles, and the expression profiles of cardiac progenitor genes. These are altered when these binding partners are overexpressed in pairs, triplicate or all together, either with each other or in conjunction with *Mesp1*. There are variations between results, including those where *Mesp1* is transduced alone or in combination with an empty vector. This is likely to be due to variations in transduction efficiencies when producing the cell lines. Future experiments should focus on inducible cell lines. However, it is apparent that the genetic landscape of these cells is different to that of wildtype 3T3 cells, where transcripts of cardiac progenitor genes are not present.

# 6 Chapter 6

## 6. Discussion

This body of work focused on the characterisation of the binding partners of the bHLH transcription factor, *Mesp1*. To date, prior to the completion of this work, the role of 1 MESP1 binding partner, CREB1, was documented. This interaction mediated haematopoietic differentiation through its transcriptional activation of *Ets2* [133].

In agreement with other bHLH transcription factors, MESP1 was also capable of binding to the transcription factors E12 and E47, although the roles of these interactions have not been delineated [130].

It is known that changes in the temporal expression of *Mesp1* in an ESC model, in conjunction with varied differentiation protocols, affect the lineage derived. This ESC model has the ability to recapitulate the 3 lineages *Mesp1* positive cells contribute to *in vivo*; cardiac, craniofacial muscles and haematopoietic cells [87,91,135]. However, how *Mesp1* mediates these fates is currently unknown. We proposed that at peak *Mesp1* expression, mesoderm stage of ESC cardiac differentiation, MESP1 may bind to different proteins to promote these cell fates.

### 6.1 Discovery of two novel binding partners of MESP1

To enable discovery of these protein interactions, a doxycycline inducible *V5 tagged Mesp1*[111] mESC model was utilised, and cells were isolated at mesoderm stage. A modified mass spectrometry technique allowed for identification of the binding partners of MESP1. The list of these proteins was narrowed by those that were unique or enriched in the induced samples, and further by those that were transcription factors or histone modification proteins. This was subsequently reduced by comparing the expression and function of the gene to *Mesp1* through analysis of spatial expression *in vivo,* and the effect of gene knockdown. The resulting list revealed 14 putative binding partners found at mesoderm stage.

There was a clear discrepancy between the results from samples 1 and 2. This is not unexpected due to them being produced through the pooling of cells from multiple differentiations. Minute changes in timings can result in binary changes in protein-protein interactions, which can be transient in nature. This led to proteins such as MBD3 being

included, despite having an increased number of spectral counts in replicate 1 induced cells, and the inverse true in replicate 2. However, it is true that some proteins that could have been included in further steps were excluded. These proteins could be confirmed to be binding partners in future analysis. This, in part, was due to time and financial constraints, as attempting to prove physical interaction of all proteins identified by mass spectrometry would have exceeded the time limit for this project.

In order to verify if these proteins were capable of binding with MESP1 in an mESC model, the temporal expression of the genes was mapped over the course of cardiac differentiation. Those that had a similar peak in expression to *Mesp1,* around mesoderm stage, were considered more likely to bind. This is due to an increase in transcript, likely to subsequently cause an increase in protein levels at a similar stage to MESP1. Changes in transcript levels can denote changes in cell fate. For example, *Mesp1* increases in transcript at E6.5 to ensure the transition to a mesodermal lineage. Therefore, any gene that increases at a similar timepoint to *Mesp1* may be linked to the same function.

This allowed for a further narrowing of the number of genes to test in physical interaction studies with MESP1. However, this approach could have excluded some candidates. Expression of all genes except *Arid3b* and *Smarcc2a* were found in mESC samples across the differentiation time points. Therefore, excluding these 2 proteins, it is possible that any of the remaining 12 putative binding partners could be present and capable of interacting with MESP1.

To identify if the remaining 7 proteins were capable of binding to MESP1, each was cloned into an expression vector with a FLAG tag, before being co-expressed with *V5-tagged Mesp1* in HEK cells. Co-immunoprecipitation assays showed that 2 of the 7 proteins were capable of binding to MESP1 in HEK cells: EOMES and MBD3. However, it does not confirm if this was direct or through a complex.

This is an artificial environment to test for protein interaction and does not mimic the mESC environment at mesoderm stage, nor the *in vivo* environment in cells of the primitive streak. There are many key factors to proteins interacting, including protein conformation, and additional cellular proteins or other molecules that can mediate interaction [231]. The limitations of co-immunoprecipitation experiments in HEK cells were confirmed through literature searches that established the ability of WDR5 to bind to EOMES, in union with Linc1405 and another histone modification protein GCN5 [124]. Co-immunoprecipitation experiments have also shown that WDR5 is further capable of binding to MBD3 isoform-c in mESCs [217]. The direct binding of WDR5 to either MBD3 or EOMES could then facilitate

a complex with MESP1, as both of these 2 proteins have been shown to be capable of binding to MESP1 in HEK cells.

Therefore, it is possible some protein interactions between the 7 proteins and MESP1 were missed through using HEK cells for the assay. It is further possible that these proteins may work in conjunction with each other, in a complex, to stimulate downstream effects. Cell specific co-factors may be necessary to facilitate interaction between the proteins, that may only be present for example in pre-cardiac mesoderm. These interactions would not be identified in HEK cells that lack these cell specific co-factors. Conversely, if an interaction is observed between proteins such as MESP1 and EOMES, it may suggest no cell specific co-factors are necessary to enable this interaction but cannot confirm if other proteins are necessary to facilitate this complex. Additional work to identify these protein complexes could be achieved through fixation of mESCs at mesoderm stage of cardiac differentiation by paraformaldehyde. These samples could then be used for co-immunoprecipitation and western blotting with specific and efficient antibodies targeting the endogenous protein, or conversely the V5 antibody targeting V5-MESP1. This was not possible in the scope of this project due to the loss of the *V5-tagged Mesp1 mESC* line during freezing before this part of the project was started. Generation of a new cell line and subsequent experiments would allow for identification of previously unknown complexes, in addition to the new interactions we have described.

To verify if these interactions could occur *in vivo*, published single cell mouse RNA-seq data was interrogated to identify if these genes were co-expressed in the same cells at similar time points. This was limited to between E6.5 and E7.5, when peak *Mesp1* expression occurs. Results of this analysis showed overlap between *Mesp1* and both histone modification genes; *Mbd3* and *Wdr5.* This is not unexpected, due to the ubiquitous expression of histone modifiers throughout cells of the embryo. *Mesp1* was also found to overlap with *Eomes* in the primitive streak at E6.5. Expression also overlapped in the nascent mesoderm at E7 and E7.5, the region of the embryo from which the cells migrate to form the cardiac crescent, and eventually the heart [86,232]. Further work in mouse embryos at E6.5 and E7.5 to prove overlapping protein expression proved technically challenging, with antibodies against endogenous proteins giving non-specific signals (data not shown). As the RNA-seq analysis demonstrated both histone modifiers are ubiquitously expressed, further investigation would not provide further significant information.

Thus, this initial data identified 2 previously unknown binding partners of MESP1, EOMES and MBD3, that are capable of binding physically to MESP1 within a HEK cell

environment. Their transcripts have a similar temporal expression to *Mesp1* at mesoderm stage of mESC cardiac differentiation and are similarly located *in vivo* in the developing mouse embryo. In addition, a third protein, WDR5, was also identified. WDR5 has a similar temporal expression to *Mesp1* during mESC cardiac differentiation and is located in the same cells in the developing mouse embryo and can also interact with EOMES and MBD3. It is therefore possible that WDR5 may also indirectly interact with MESP1.

### 6.1.1 EOMES is capable of binding in a complex with MESP1

EOMES has been previously identified as being key to cardiac development, acting upstream of *Mesp1*. *Eomes*$^{-/-}$ ESCs fail to express the cardiac progenitor marker *Nkx2.5* or cardiac structural gene markers, such as *myosin light chain 7 (Myl7*)[122]. One mechanism by which EOMES modulates transcription of *Mesp1* is through binding in a complex with GCN5, WDR55 and LINC1405 upstream of the transcription start site, at an enhancer site for *Mesp1*[124]. Disruption of this complex through ablation of Linc1405 leads to modulation of the histone modification of this enhancer site, and reduction in both *Mesp1* expression, and disturbance of cardiac differentiation [124].

Whilst this previous role of modulating *Mesp1* expression was well characterised, this work shows that EOMES is also capable of binding to MESP1. It is possible that EOMES and MESP1 bind in a complex to the DNA, forming a feedback loop that can then induce the cardiac kernel of genes.

### 6.1.2 MBD3 is capable of binding in a complex with MESP1, whilst WDR5 may work in unison with other binding partners to affect cell fate

MBD3, is a histone modification protein that is essential to early development, with *Mbd3*$^{-/-}$ knockout embryos dying during early gastrulation [171]. A member of the NuRD complex, involved in chromatin remodelling, it was traditionally thought to be involved in gene silencing, although it has been shown to have roles in other biological processes, including DNA repair [233].

Through this work, it has been shown that MBD3 is capable of binding to MESP1 in HEK cells, though whether this is indirect or direct has not been fully delineated.

The NuRD complex varies in its composition and can include GATAD2A and CHD4, which were previously identified as potential binding partners of MESP1 using RIME (results chapter 1)[208]. It is then possible that MESP1 associates with the NuRD complex, potentially through its interaction with MBD3.

Whilst WDR5 was not found to be capable of binding to MESP1 in HEK cells, it is possible that it may bind in a complex through its interaction with MBD3 isoform C, through its N terminus [217]. This interaction is thought to be capable of co-regulating gene expression at enhancers and promoters [217]. In addition it may also bind via its interaction with EOMES, Linc1405 and GCN5, or through its association with MYC [215]. Therefore, whilst not being a direct binding partner, it is possible that WDR5 acts as a scaffold for further binding of other modification complexes, including the SET-MLL, NuRD or EOMES, GCN5, Linc 1405 complex.

## 6.2 Binding partners affect cardiac differentiation when disturbed at mesoderm stage

Studies to elucidate the effect of genes on differentiation traditionally knockdown or ablate genes in ESCs before initiating cardiac differentiation protocols. This is true in both *Eomes* and *Mesp1*, where ablation in mESCs results in a failure to differentiate cardiomyocytes [122]. Ablation of *Mbd3* in ESCs leads to cells that can maintain pluripotency without the addition of LIF to the media, and fail to differentiate into the 3 germ layers when plated in EB form[213]. *Wdr5* knockdown resulted in repression of self-renewal processes in ESCs, as well as a suppression of ectoderm differentiation, but an enrichment of genes associated with mesodermal development [200].

To fully evaluate the effect of these MESP1 binding partners on mesodermal differentiation, and consequent cardiac progression, genes for these candidates were knocked down at day 2 of differentiation to ensure maximal depletion by mesoderm stage (day 4). Unfortunately, due to the Covid-19 pandemic, a third repeat of these experiments was not possible, therefore, a full conclusion on the effects of these knockdowns could not be reached.

Knockdown of the transcription factors, *Eomes* or *Mesp1,* was more efficient than those of the histone modifiers, with less than 20% of transcript remaining. Knockdown of *Mbd3* fluctuated, with between 35% and 65% of original transcript levels remaining, whilst knockdown of *Wdr5* resulted in approximately half of the original levels of transcription. This was a result of a mix of 4 siRNAs added to the culture at day 2 of differentiation, whilst cells remained in EB form and in suspension. Variation in knockdown efficiency could be as a result of poor penetration into the EBs, with some central cells unable to take up the siRNAs, and therefore variable effects witnessed on qPCR analysis. Furthermore, transcripts of the histone modification genes, *Mbd3* and *Wdr5*, were already present in the cells before

the addition of the siRNAs, unlike that of *Eomes* or *Mesp1.* Therefore, it is conceivable an increased amount of siRNA was necessary to achieve a greater knockdown efficiency due to an increased amount of transcript to silence. Despite this, these knockdown efficiencies are in line with other published data that achieved approximately 40% knockdown [124].

Knockdown of both *Mbd3* and *Wdr5* resulted in significantly smaller EBs formed at day 4, or mesoderm stage, of differentiation. This is in agreement with *Mbd3*$^{-/-}$ ESCs, which show impaired differentiation of all 3 germ layers and proliferation [213], and *Wdr5* knockdown in ESCs that resulted in changes to cellular morphology [200].

Previous work had not classified the effect of *Mesp1* or *Eomes* knockdown on the effect of EB size. There is an observable reduction in EB size, though not significant when either transcription factor is knocked down. It was expected that this knockdown would have had an effect on EB size. However, the results of the statistical analysis on knockdown of the histone modification genes do reflect the cells observed under the microscope.

### 6.2.1 Knockdown of *Mesp1's* potential binding partners results in a reduction of *Mesp1* expression with a variable effect on the transcript levels of other mesodermal markers

With a high knockdown efficiency, *Mesp1* expression was reduced to below 25% of control expression at day 4, or mesoderm stage of differentiation, indicating an efficient knockdown of the gene. Gene temporal expression profiles followed a similar trend in both knockdown and control samples, with a lower expression observed in knockdown samples. This pattern was similarly observed when analysing *Eomes* and *T* expression in *Mesp1* knockdown samples, with transcript levels reduced by 50% in knockdown samples at day 4 of differentiation. A reduction in expression of *Eomes, T,* and *Mesp1* was also observed in *Eomes* knockdown samples. These findings correlate with published data that shows *Eomes* knockdown causes a reduction in *Mesp1* expression [125], although the study did not comment on expression of *T.* In *Eomes* null embryos, *T* expression is still present. In our mESC model expression of *T* is observed, however, reduced to 30% of control levels. It is further known that *Mesp1* expression can be induced by both T and EOMES [122,234]. MESP1 is also found in a regulatory feedback loop capable of repressing expression of *T* [108]. Expression of *T* and *Eomes* partially overlaps in the primitive streak and primitive epiblast [125]. It is therefore possible that *Mesp1* and *Eomes* are capable of regulating *T.* It has been shown that expression of *Mesp1* is induced by *T,* and subsequently causes a reduction of *T.* However, we observe that with reduced expression of *Mesp1* or *Eomes,* there

is a resultant reduction in *T* expression. This does not fit previously examined models, whereby after *Mesp1* is induced, *T* is repressed. It is possible that a small amount of *Mesp1* transcript also causes induction of *T,* before being repressive at peak transcription levels. The mechanism behind this action would need to be further investigated.

A variable effect on the expression of *T* is observed when either *Wdr5* or *Mbd3* are knocked down, with a reduction to between 25% and 80% of control levels at day 4. Therefore, further replicates are needed to confirm if these genes are capable of modulating the expression of *T.*

Discrepancies between replicates are also observed in the expression of *Eomes* when either *Wdr5* or *Mbd3* are knocked down. In *Wdr5* knockdown samples, there is a change in the temporal expression of *Eomes* between replicate 1 and 2. Expression of *Eomes* is conventionally highest at day 4, which is observed in replicate 2 control samples. However, replicate 1 shows a 24-hour shift in expression, with maximal expression witnessed at day 5. Despite this, the greatest expression of *Eomes* in both knockdown samples occurs at day 5 and is higher than that of control cells on any day of differentiation. This suggests *Wdr5* knockdown allows for an increase in *Eomes* expression, although further replicates would be needed to verify this finding. There is no information on the role of *Wdr5* on the regulation of *Eomes,* however, depletion in ESCs leads to increased differentiation and loss of histone activation markers at pluripotency genes, such as *Nanog* [200]. It is further known to interact with OCT4 to regulate pluripotency genes, and this loss of WDR5 could cause an acceleration in differentiation that results in a higher level of *Eomes* in comparison to control cells at a similar stage [200]. It is unknown why there are shifts in temporal expression between replicates, although there were changes in cell culture reagents, cytokines and culture rooms used in replicate 1 and 2 due to technical issues.

Replicate 1 of *Mbd3* knockdown also exhibits a 24-hour shift in *Eomes* expression, with control transcript levels highest at day 5 of differentiation. In replicate 2 this is not observed. In contrast to *Wdr5* knockdown, *Mbd3* knockdown samples show a reduction in *Eomes* expression to below 50% of control samples at their highest values. This indicates that *Mbd3* may be responsible for enhancing expression of *Eomes* through its histone modification role. To date, it is known that *Mbd3* knockout ESCs fail to differentiate or downregulate pluripotency genes such as *Oct4* [213]. It is therefore possible that these cells have failed to differentiate fully and thus lack full *Eomes* expression.

Knockdown of *Mbd3* resulted in a reduction in the expression of *Mesp1,* with expression limited to less than 25% of control expression at day 4 of differentiation.

Temporal expression remained similar to that of control cells. This reduction was also observed in *Wdr5* knockdown samples, with expression of *Mesp1* less than one tenth observed in control samples. In contrast, this also resulted in a temporal shift, with highest expression found at day 5 of differentiation. The sole effect of *Mbd3* or *Wdr5* on *Mesp1* expression has not been previously evaluated, and whilst a third repeat is necessary to confirm these results, they are suggestive that these histone modification genes may play a role in the expression of *Mesp1*. This effect could be in part due to the role of WDR5 in a complex that is known to bind upstream of *Mesp1* at an enhancer region, facilitating the transcription of *Mesp1*. A reduction in WDR5 could reduce formation of this complex and cause a corresponding reduction in *Mesp1* expression. The mechanism by which *Mbd3* could affect *Mesp1* is currently unknown. *Mbd3* is known to be part of the NuRD complex [235], which is essential in ESC fate decisions, as well as haematopoietic development [213]. MBD3 isoform c has also recently been shown to bind to WDR5, and it is possible that this complex can co-regulate gene expression at promoters and promoter-distal enhancers [217]. This could be 1 mechanism by which *Mbd3* or *Wdr5* could affect *Mesp1* expression.

### 6.2.2 Knockdown of *Mesp1* results in a reduction in *Mbd3* and *Wdr5* expression

Knockdown of the binding partners of MESP1 in ESCs can have an effect on ESC differentiation and transcription of the other binding partners. This is evident in that knockdown of *Mbd3, Wdr5* or *Eomes* can reduce expression of *Mesp1,* as discussed previously. Knockdown of *Mbd3, Mesp1* or *Wdr5* also leads to variation in *Eomes* transcript levels.

Knockdown of *Mesp1* in mESCs resulted in a reduction in *Mbd3* expression at day 4 of differentiation, resulting in a reduction in transcript level to below 50% of that observed in control samples. This was similarly observed when knocking down *Eomes.* This effect was not witnessed in samples where *Wdr5* was knocked down, with conflicting patterns between the 2 replicates. This suggests that whilst *Mesp1* and *Eomes* play a role in *Mbd3* expression, *Wdr5* does not. WDR5 is known to bind to MBD3 isoform c, which is only expressed in pluripotent stem cells and is switched off as differentiation progresses[217]. Therefore, it is likely that there is an effect on *Mbd3-isoform c* expression, although this would not be registered at these stages of differentiation. Further work to map the temporal expression of MBD3 isoform c, and the effect of *Wdr5* expression upon it, would need to focus on the initial pluripotent embryonic stages of differentiation. The effect of knocking these genes out

at ESC stage is already known [217], and therefore it is assumed this work would not provide additional insight into the role of these genes in cardiac differentiation.

The effect of knocking down *Mesp1* further showed an effect on the expression of *Wdr5.* Whilst the pattern of *Wdr5* expression varied between the replicates, both knockdown replicates showed a reduction in the amount of *Wdr5* expressed at day 4 of differentiation in comparison to control samples. This effect would need to be further investigated in a third replicate, to establish the pattern of *Wdr5* in these differentiations, and confirm the effect of *Mesp1* knockdown upon it. WDR5 works upstream of *Mesp1* in a complex that activates its transcription [124], however, it has not been shown if *Mesp1* has any effect on *Wdr5* expression. This data preliminarily suggests that *Mesp1* may also play a role in the expression of *Wdr5,* potentially through a feedback mechanism to enhance its own transcription. This is mirrored in *Eomes* expression, another key member of the complex that regulates *Mesp1*, that is reduced in *Mesp1* knockdown.

Finally, expression of *Wdr5* upon *Eomes* knockdown varied between the replicates, with expression in knockdown 1 showing an increase in *Wdr5,* whilst samples in knockdown 2 showed the opposite. To elucidate the effect of *Eomes* on *Wdr5* expression, further experiments would need to be performed. However, it has been shown that *Wdr5* knockdown causes an increase in *Eomes* transcript levels, this could be in part due to epigenetic regulation by WDR5.

### 6.2.3 Knockdown *of Mesp1* or *Eomes* reduces cardiac progenitor markers, whilst *Mbd3* knockdown causes a 24-hour temporal shift in expression.

Cardiac progenitor markers are the first sign mesodermal cells are transitioning to a cardiac cell lineage. This typically occurs around day 5 of cardiac differentiation and is marked by the expression of *Nkx2.5* and *Gata4.*

In knockdown samples of *Mesp1* or *Eomes,* both of these marker genes have a shift in expression by 24 hours, with peak expression found at day 4 of differentiation. Expression of these genes are reduced in comparison to control samples, where maximal expression is found at day 5. This indicates that reduction in mesodermal genes allows cardiac progenitor genes to be switched on at an earlier time point, however by a reduced amount. This could be through residual expression of *Mesp1* or *Eomes* activating these genes, however, due to a lack of transcript, this gene pathway was not fully amplified. This is to be expected, as both of these genes are known to be necessary for cardiac differentiation and activation of downstream cardiac transcription machinery [108,123].

Knockdown of *Mbd3 and Wdr5* also results in changes to cardiac progenitor gene expression. Knockdown of *Mbd3* resulted in a temporal shift in expression of both *Gata4* and *Nkx2.5* by 24 hours, to day 4 of differentiation. Similar to results seen in knockdown of *Eomes* and *Mesp1*, transcript levels of both *Gata4* and *Nkx2.5* were reduced in comparison to control levels. *Mbd3*$^{-/-}$ ESCs are known to have reduced differentiation capacity. Therefore, a reduction in transcript amount is expected, and could result from incomplete knockdown of *Mbd3* during siRNA treatment. This reduction may also be as a result of a reduction in *Mesp1* and *T* expression at day 4 of differentiation, which will affect the transcription of the downstream cardiac kernel of genes, including *Nkx2.5* and *Gata4.* It is interesting to note a shift in the temporal expression of both cardiac progenitor genes. The mechanism by which *Mbd3* plays a role in differentiation is unknown. *Gata4* positive cells are identified in the early embryo, at E4.5 in *Mbd3*$^{-/-}$ embryos, marking cells of the primitive endoderm. Therefore it is known that ablation of *Mbd3* does not prevent *Gata4* expression in other cells of the embryo, or at earlier stages [213].

Conversely, knockdown of *Wdr5* shows no effect on the temporal expression of either cardiac progenitor marker, with the highest expression of both knockdown samples seen at day 5 of differentiation. Expression of *Nkx2.5* in both knockdown replicates is less than that observed in control samples. Equally, knockdown of *Wdr5* results in conflicting patterns between replicates for *Gata4* expression, with knockdown replicate 1 expressing higher levels of *Gata4* than controls, whilst replicate 2 expresses lower levels. A third replicate will be required to analyse whether *Wdr5* does affect transcript levels of *Gata4.* As previously observed, knockdown of *Wdr5* resulted in a reduction in the mesodermal markers *Mesp1* and *T,* however an increase in *Eomes* expression. Therefore, it is not unexpected to see a reduction in transcript levels of the cardiac progenitor cell markers, and at a similar temporal expression due to the continued expression of *Eomes. Eomes* is sufficient to induce expression of *Mesp1 or Mesp2,* which is then capable of inducing cardiac differentiation[122]. *Mesp1* and *Mesp2* are redundant in the ability to induce differentiation of cells to a cardiac lineage [116,118]. Whilst there is a reduction in *Mesp1*, additional *Eomes* expression may be capable of rescuing this. This could potentially be through the induction of *Mesp2* which was not measured, to allow expression of cardiac progenitor markers, although at a reduced amount in comparison to controls.

## 6.3 Summary

In addition to proving physical interaction of each binding partner to MESP1, it was also critical to evaluate their role in cardiac differentiation, and the effect each had on other binding partners during differentiation. It is apparent that disturbance of *Mesp1* causes a decrease in the expression of each of its binding partners in comparison to control samples, suggesting *Mesp1* regulates expression of its partners, either directly or indirectly. Knockdown of the other binding partners also causes a fluctuation in the expression of other mesodermal genes such as *T* or *Eomes,* and these changes also affect cardiac progenitor gene expression. There are some variations between replicates, and a third replicate that could give insight as to the true pattern was unable to be performed due to COVID-19 disruptions. Further work should centre around this replicate to fully elucidate the effect of these knockdowns on each gene. Additional gene analysis could also be performed in relation to genes including *Mesp2,* to identify if there is any compensatory upregulation when *Mesp1* is knocked down, or any of the binding partners that when knocked down, reduced *Mesp1* expression.

### 6.3.1 Overexpression of binding partners in conjunction or independently of *Mesp1* affects cardiac gene expression in 3T3 cells

Previous work establishing the effect of *Mesp1* on cardiac transdifferentiation in fibroblast reprogramming has established that *Mesp1* alone is insufficient to reprogram fibroblasts to a cardiac state. However it is capable, in conjunction with *Ets2,* of converting 2.3% of a starting population of human dermal fibroblasts to a contractile cardiomyocyte like state [71].

In agreement with this paper, *Mesp1* and each binding partner were transduced singularly, and in conjunction, into fibroblasts. Initial work focused on using primary human fibroblasts to test for the effect of overexpression, and cell transduction protocols were optimised in these cells. Expression was optimised using FACS sorting to ensure reliable quantitative results, that gave an accurate overview of the number of cells transduced. Maximum efficiency was derived from cells that had been transduced twice in the course of 72 hours, increasing efficiency from 9.8-11.8% in EGFP-*Eomes* transduced cells, to 20-29%. This increase was also seen in mCherry-*Mbd3* transduced cells increasing from an initial rate of 1.8- 2.1% to 5.8-11.1%. This protocol was then used for all further transductions. Unfortunately, despite modifications in FACS sorting protocols, including increasing nozzle size and reducing speed of sort, the primary human fibroblasts were not viable post-sort and

did not recover in culture. Therefore, we exchanged cell types to work on 3T3 mouse fibroblasts, which are a more robust cell line. All future experiments utilised the same transduction protocol, with dual transductions over the course of 72 hours.

Transduction was confirmed using western blotting and quantification. Expression of *Eomes, Mesp1* and *Wdr5* were all significantly upregulated, with an increase of between 4 and 20-fold higher than that of 3T3 cells. Expression of *Mbd3* was on average 1.4-fold that seen in control cells, with quantification using an antibody targeting the endogenous protein, which is also present in wildtype cells. It is also now understood that the central dogma of biology, that DNA is made into mRNA, which is then processed to proteins, does not occur in a 1:1:1 ratio. In fact, despite small changes in protein levels, there can be large increases in mRNA [236,237]. Despite this increase not being statistically significant, subtle increases in protein amounts can have large scale downstream effects, especially in the case of histone modification genes and transcription factors that can affect gene regulatory networks [238,239].

Overexpression of *Wdr5* did cause a reduction in the rate of proliferation, however, there was no change in the proliferation rate when the other genes were overexpressed, nor was there a change in apoptosis or necrosis rates for any of the overexpressed genes. This indicates that *Wdr5* overexpression causes a reduction in cell division rate, although does not induce apoptosis.

To further establish the effect of overexpressing these genes in 3T3 cells, the cell and nuclear morphology was analysed. This is important as changes to cell and nuclear morphology can be characteristic of changes to cellular type.

Overexpression of each binding partner singularly did not affect cell area, however, a combination of *Mesp1,* in conjunction with either histone modification gene, *Mbd3* or *Wdr5,* did cause a reduction in cell area. This was also observed when *Eomes* was co-expressed with *Mbd3*. This indicates that a combination of a transcription factor and a histone modifier can affect cell area. However, this effect was also observed with *Mesp1* in union with an empty vector control. This could indicate an issue with an increase in vector amount, which results in changes to cell morphology. Alternatively, the integration of the empty vector in conjunction with *Mesp1* could also result in a change. This should be considered, as there is also a change observed in the empty vector control in addition to *Mesp1* in cell circularity and nuclear area. However, an increase in cell circularity is also observed in *Mesp1* overexpressing cells alone, and therefore this increase in cell circularity appears to be reliable.

*Mesp1* or *Wdr5* alone appear sufficient to cause a statistically significant increase in cell circularity. This ability to change cell morphology is not unexpected due to the known effects of WDR5, and its interaction with members of the cytoskeleton, including actin and γ-tubulin [240]. This effect is also observed when combining the 2 factors, or *Mesp1* in conjunction with an empty vector control.

The effect of *Wdr5* on cellular morphology is also apparent when analysing nuclear area, which is significantly increased in *Wdr5* transduced samples, though reduced in cells overexpressing *Wdr5* and *Mesp1*. This reduction is also observed in cells where *Mbd3* and *Mesp1* are co-expressed. In addition, *Mbd3* alone, or in conjunction with *Mesp1* or *Eomes,* also appears capable of causing a reduction in nuclear size. However, *Mesp1* in combination with empty vector controls, also causes a similar reduction. Therefore, it should be assumed that whilst *Wdr5* increases nuclear area, and *Mbd3* decreases nuclear area, the effects of a gene combinations may be caused by an increase in vector amount. Future experiments could use alternative control vectors to test this, or vectors encoding multiple genes. It should also be noted that the stoichiometry of genes transduced can have a large effect on the results. This has been seen in previous reprogramming papers where the interplay between the amount of *Essrb* and *Eomes* could result in 2 different cell types when reprogramming to a induced pluripotent stem cell or induced trophoblast cell [241]. Therefore, future experiments should attempt to vary the levels of these factors to discern their effects on cellular fate and morphology.

Finally, changes in nuclear circularity were also observed. This could denote changes in the epigenetic and chromatin landscape of the cell, as well as changes in the interaction between the cell and nuclear membranes. There was a reduction in nuclear circularity in *Mesp1, Wdr5* or *Eomes* overexpressing cells. This decrease was also observed in cells expressing both *Wdr5* and *Mesp1*, indicating a clear role for the transcription factor and histone modifier in eliciting changes to the nuclear shape.

This effect is similarly observed when looking at nuclear to cell ratio – where *Mesp1* and *Wdr5*, either singularly or in combination, causes an increase in the ratio. This confirms previous studies that suggest *Wdr5* is responsible for changes in the levels of the repressive mark H3Kme3, that directly impacts nuclear shape, as well as cell polarity and shape [225]. The localisation of MESP1 has not been fully delineated, with some bHLH proteins shuttling between the nucleus and the cytoplasm [242]. MESP1 binding of chromatin has also been shown to be correlated with activation marker H3K27ac [243]. This could be 1 mechanism

by which MESP1 changes the morphology of the nucleus. Further evaluation of nuclear chromatin structure and protein localisation would be necessary to identify a mechanism by which these changes occur. These changes, however, do suggest that overexpression of these proteins induces a morphological change suggestive of an alteration in genetic regulation. This can be indicative of changes in cellular function.

The effect of the overexpression of *Mesp1* in 3T3 cells was further explored through qPCR analysis. Increases in protein were confirmed by large scale increases in *Mesp1* mRNA, with a 150-fold increase in *Mesp1* expression in comparison to control cells. There was also an increase in cardiac progenitor markers *Gata4* and *Nkx2.5,* a 5 and 2-fold increase respectively. This is in agreement with previous work that showed *Mesp1* expression in fibroblast cells also induces an increase in *Nkx2.5* expression [71]. In agreement with results from the mESC knockdown experiments, where knockdown of a binding partner caused a reduction in *Mesp1* expression, overexpression of each binding partner resulted in the expression of *Mesp1*. This suggests a role for the binding partners in *Mesp1* regulation.

Additionally, overexpression of the binding partners resulted in upregulation of both *Nkx2.5* and *Gata4.* The levels and temporal expression of these genes were affected when the binding partners were knocked down in a mESC models.

Overexpression of *Wdr5,* either singularly or in conjunction with *Mesp1,* results in increased expression of fibroblast gene marker *Fsp-1.* There is an increase in *Mesp1* expression in both samples, however, only an increase in cardiac progenitor marker *Nkx2.5* in those that express *Wdr5* singularly. This could show a change from a dermal to a cardiac fibroblast, however, further staining and results would be necessary to justify that assumption. This increase in *Fsp-1* and *Nkx2.5* was similarly observed in samples overexpressing both *Eomes* and *Mbd3.*

Overall, these results appear to confirm the relationships observed between genes when knocked down in mESCs, with overexpression of each binding partner causing an increase in *Mesp1* expression. An increase in expression of the genes causes an increase in cardiac progenitor markers, whilst the reverse is true in mESC knockdowns.

These results should be carefully considered, however, as each of these genes, excluding *Mbd3*, were not present in wildtype 3T3 fibroblasts. Therefore, even a slight increase in gene transcript can cause a large-scale fold difference. To further validate if these results do indicate a change towards a cardiac phenotype, further investigation should be undertaken to measure cell calcium transients, the means by which the cell metabolises, for

example, using primarily glucose or fatty acids, and staining for cellular markers including *Nkx2.5*.

## 6.4 Covid-19 implications

The recent pandemic at the start of 2020 resulted in a quick shutdown period and an extended absence from the laboratory. As such, some cell lines did not recover from being frozen in non-optimal conditions, and cytokines for differentiation with a 3-month life span went out of date. This has resulted in expected work, such as cell staining and a further mESC differentiation, not being completed. This work would allow further insight into the effect of each gene knockdown on the genetic markers of cardiac differentiation, as well as to confirm and clarify existing findings.

## 6.5 Future work

To confirm protein-protein binding in a more *in vivo* relevant model, co-immunoprecipitation experiments of MESP1 from mESCs should be performed, with blotting for not only EOMES, MBD3 and WDR5, but also the other binding partners identified through mass spectrometry. This would enable validation of new complexes that may be present, however, were not identified through HEK cell co-expression and co-immunoprecipitation experiments. New technology, such as proximity ligation assays, could be used to further verify that these interactions occur within mESCs, and also potentially within embryos [244]. This technology utilises endogenous antibodies, targeting 2 proteins. Secondary antibodies that are conjugated to proximity ligation probes are then added. If these 2 probes are located in the immediate vicinity of one another they will link together[244]. Further reagents can then identify this link, and it can be visualised through microscopy.

It is also important to identify whether these proteins have additional effects on cardiac differentiation post-mesoderm stage. Several of the proteins identified through mass spectrometry, and verified through qPCR analysis of their temporal expression, showed an additional increase in expression at day 6, including *Mbd3.* This time point marks the transition between cardiac progenitor stage and progression into a cardiomyocyte fate. It would be interesting to know if knockdown of these genes at this stage, instead of at mesoderm stage, could affect cardiomyocyte formation. This would indicate a further role for these genes post-mesoderm specification.

Whilst overexpression of these genes in 3T3 cells show a change in the genetic profile of the cell, it is not sufficient to say they induce a cardiac fate. qPCR is a useful tool to

evaluate transcriptional changes within a cell, however, its results can be misleading in a situation where the gene was not previously expressed. To further analyse these results, cellular staining, to identify if NKX2.5 is expressed to a sufficient level to be identified by immunocytochemistry, as well as additional staining to confirm if any other cardiac genes have been induced. It is also worth noting that these cells did not appear to beat in culture, either at sparse or increased densities. It would be interesting to apply the same cytokines as found in mESC differentiation to transduced 3T3 cells and evaluate if these can enhance any cardiac gene expression or induce spontaneous cardiac beating.

## 6.6 Summary and overall model

This project has identified 2 novel binding partners for MESP1; EOMES and MBD3, with further evidence that WDR5 is also a potential partner of MESP1, however, not capable of direct interaction. Both MBD3 and EOMES can bind to MESP1, as evidenced through co-immunoprecipitation experiments. However, it is also known that WDR5 is capable of directly binding to both EOMES and MBD3[124,217]. Therefore, it is highly likely that MESP1, EOMES, MBD3 and WDR5 are capable of binding in a complex, which may direct differentiation of cells to a cardiac fate.

Knockdown of these binding partners in a mESC model reduces expression of *Mesp1* and can cause a temporal shift or reduction in cardiac progenitor markers. Therefore, whilst they are capable of inducing expression of cardiac genes in fibroblasts, in different combinations they can cause upregulation of fibroblast genes or other binding partners, in a similar pattern to that seen in mESC knockdown.

It is therefore logical to assume that *Mesp1,* in conjunction with these 3 partners, is capable of inducing a cardiac phenotype. Reduction of *Mesp1,* or the 3 partners, also causes a reduction in cardiac progenitor genes in a mESC model. Knockout of any of these binding partners also causes a reduction in the expression of mesodermal markers *T* and *Mesp1.* It is also known all 4 genes are necessary for correct morphogenetic development *in vivo* [89,122,216,245]. It is likely that these genes work in complex to enable cardiac differentiation at mesoderm stage of cardiac development, through promotion of *Mesp1* and *T,* which enables activation of the subsequent cardiac kernel of transcription factors, including *Nkx2.5* and *Gata4.* The loss of any of these factors impairs expression of cardiac progenitor markers and signals their importance in cardiac development.

Overall, it appears *Mesp1, Mbd3, Eomes* and *Wdr5* are capable of inducing expression of *T*, with knockdown of these partners causing a reduction in *T* in a mESC model. This is in direct contrast to previous work that found *Mesp1* inhibits *T* expression [108]. This discrepancy might be explained by levels of *Mesp1*. For instance, *T* can induce expression of *Mesp1* [120], and in a positive feedback mechanism, *Mesp1* can subsequently also cause upregulation of *T,* and consequently increase *Mesp1.* However, at higher concentrations of *Mesp1*, *T* is consequently downregulated, as cells progress from mesodermal to cardiac progenitor cell lineage.

A decrease in *Mesp1* or *Eomes* causes a reduction in cardiac progenitor markers in mESCs. *Mesp1* is known to act upstream of the cardiac kernel and induce cardiac differentiation, whilst *Eomes* is known to be essential in mESC cardiac differentiation [108,121,207]. Therefore, these studies confirm previous results that a reduction in *Mesp1* or *Eomes* causes a reduction in cardiac progenitor genes.

The effect of *Mbd3* on cardiac progenitor genes has not been described. However, in this work, we note there is a shift in cardiac progenitor gene temporal expression, with expression initiated 24 hours before those treated with control siRNA, at day 4 of differentiation. This is suggestive of the effect of *Mbd3* on chromatin architecture, with knockdown of *Mbd3* causing a change in chromatin structure that allows for the activation of cardiac progenitor genes at earlier time points. MBD3 has been found to be involved in nucleosome organisation near promoters, and shown to be enriched at promoters modified by H3K27ac [246]. However, a significant amount was also localised at regions marked by the repressive marker H3K27me3 [246]. In addition, *Mbd3*$^{-/-}$ fibroblasts transduced with OKSM factors were more efficiently reprogrammed than control cells [210]. This indicates *Mbd3* knockdown could result in changes to chromatin structure and potential changes in gene regulation, allowing upregulation of cardiac genes at earlier time points than in wildtype mESCs.

The importance of *Mesp1* in cardiac development is well described [110,118,247], however, the mechanism by which *Mesp1* can further induce haematopoietic and craniofacial muscle phenotypes is not fully elucidated [135]. By isolating the binding partners at mesoderm stage of development, we have identified those that may play a key role in cardiac development. Through these experiments, we have identified 2 novel binding partners of MESP1, and 1 further potential indirect partner. The roles of these partners have been identified during cardiac differentiation in a mESC model, and through their ability to induce the expression of cardiac genes in fibroblasts, though not a mature cardiac phenotype.

However, due to the nature of the experiments, the mESCs were directed towards a cardiac lineage through the addition of cardiac specific cytokines and growth factors. These binding partners may also be essential to the development of other mesodermal derived lineages, including haematopoietic and craniofacial muscle. It is also possible that due to the cardiac differentiation protocol, the binding partners identified are only involved in cardiac development. Further experiments are necessary to identify if these binding partners can be isolated in mESCs directed to other fates, and if they are capable of transdifferentiating fibroblasts to alternative lineages.

# 7 References


[1] Kaptoge S, Pennells L, De Bacquer D, Cooney MT, Kavousi M, Stevens G, et al. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. Lancet Glob Heal 2019;7:e1332–e1345. doi:10.1016/S2214-109X(19)30318-3.

[2] Heidenreich P a., Trogdon JG, Khavjou O a., Butler J, Dracup K, Ezekowitz MD, et al. Forecasting the future of cardiovascular disease in the United States: A policy statement from the American Heart Association. Circulation 2011;123:933–944. doi:10.1161/CIR.0b013e31820a55f5.

[3] Hinton W, McGovern A, Coyle R, Han TS, Sharma P, Correa A, et al. Incidence and prevalence of cardiovascular disease in English primary care: A cross-sectional and follow-up study of the Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC). BMJ Open 2018;8:1–9. doi:10.1136/bmjopen-2017-020282.

[4] Ziaeian B, Fonarow GC. Epidemiology and aetiology of heart failure. Nat Rev Cardiol 2016;13:368–378. doi:10.1038/nrcardio.2016.25.

[5] BHF. Cardiovascular Disease Statistics - BHF UK Factsheet. Br Hear Found 2018. doi:10.1017/CBO9781107415324.004.

[6] Dementieva Y, Green TL, Primerano DA, Wei L, Denvir J, Wehner P, et al. Identification of genes contributing to cardiovascular disease in overweight and obese individuals from West Virginia. W V Med J 2012;108:23–30.

[7] Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, et al. A multilocus genetic risk score for coronary heart disease: Case-control and prospective cohort analyses. Lancet 2010;376:1393–1400. doi:10.1016/S0140-6736(10)61267-6.

[8] Morris RW, Cooper JA, Shah T, Wong A, Drenos F, Engmann J, et al. Marginal role for 53 common genetic variants in cardiovascular disease prediction. Heart 2016;102:1640–1647. doi:10.1136/heartjnl-2016-309298.

[9] Said MA, van de Vegte YJ, Zafar MM, van der Ende MY, Raja GK, Verweij N, et al. Contributions of Interactions Between Lifestyle and Genetics on Coronary Artery Disease Risk. Curr Cardiol Rep 2019;21:1–8. doi:10.1007/s11886-019-1177-x.

[10] Ambrose JA, Singh M. Pathophysiology of coronary artery disease leading to acute coronary syndromes. F1000Prime Rep 2015;7:1–5. doi:10.12703/P7-08.

[11] Libby P, Theroux P. Pathophysiology of coronary artery disease. Circulation 2005;111:3481–3488. doi:10.1161/CIRCULATIONAHA.105.537878.

[12] Daher IN, Daigle TR, Bhatia N, Durand JB. The prevention of cardiovascular disease in cancer survivors. Texas Hear Inst J 2012;39:190–198.

[13] Andersson C, Vasan RS. Epidemiology of cardiovascular disease in young individuals. Nat Rev Cardiol 2018;15:230–240. doi:10.1038/nrcardio.2017.154.

[14] Qureshi N, Humphries SE, Seed M, Rowlands P, Minhas R. Identification and management of familial hypercholesterolaemia: What does it mean to primary care? Br J Gen Pract 2009;59:773–778. doi:10.3399/bjgp09X472674.

[15] Miles C, Fanton Z, Tome M, Behr ER. Inherited cardiomyopathies. BMJ 2019;365. doi:10.1136/bmj.l1570.

[16] Jacoby D, McKenna WJ. Genetics of inherited cardiomyopathy. Eur Heart J 2012;33:296–304. doi:10.1093/eurheartj/ehr260.

[17] Muchtar E, Blauwet LA, Gertz MA. Restrictive cardiomyopathy: Genetics,

pathogenesis, clinical manifestations, diagnosis, and therapy. Circ Res 2017;121:819–837. doi:10.1161/CIRCRESAHA.117.310982.
[18] Syamasundar P. Congenital Heart Defects – A Review. Congenit. Hear. Dis. - Sel. Asp., 2012, p. 2–44. doi:10.5772/27002.
[19] Stout KK, Broberg CS, Book WM, Cecchin F, Chen JM, Dimopoulos K, et al. Chronic Heart Failure in Congenital Heart Disease: A Scientific Statement from the American Heart Association. vol. 133. 2016. doi:10.1161/CIR.0000000000000352.
[20] Inamdar A, Inamdar A. Heart Failure: Diagnosis, Management and Utilization. J Clin Med 2016;5:62. doi:10.3390/jcm5070062.
[21] NICE. Putting NICE guidance into practice. Resource impact report : Chronic heart failure in adults. 2018.
[22] Deepa M. Gopal MM, Andreas P. Kalogeropoulos, MD P, Vasiliki V. Georgiopoulou M, Andrew L. Smith M, Douglas C. Bauer M, Anne B. Newman, MD M, et al. Cigarette smoking exposure and heart failure risk in older adults: The health , aging, and body composition study. Am Heart J 2012;164:236–242. doi:10.1038/jid.2014.371.
[23] Gupta D, Georgiopoulou V V., Kalogeropoulos AP, Dunbar SB, Reilly CM, Sands JM, et al. Dietary sodium intake in heart failure. Circulation 2012;126:479–485. doi:10.1161/CIRCULATIONAHA.111.062430.
[24] Friedrich EB, Böhm M. Management of end stage heart failure. Heart 2007;93:626–631. doi:10.1136/hrt.2006.098814.
[25] Rose EA, Gelijns, Annetine C. Moskowitz AJ, Heitjan, Daniel F. Stevenson, Lynne W. Dembitsky W, Long JW, Ascheim D, Tierney A, et al. Long-Term Use of a Left Ventricular Assist Device. N Engl J Med 2001;345:1435–1443. doi:10.1056/NEJMoa012175.
[26] Wilhelm MJ. Long-term outcome following heart transplantation: Current perspective. J Thorac Dis 2015;7:549–551. doi:10.3978/j.issn.2072-1439.2015.01.46.
[27] Chatterjee A, Feldmann C, Hanke JS, Ricklefs M, Shrestha M, Dogan G, et al. The momentum of HeartMate 3: A novel active magnetically levitated centrifugal left ventricular assist device (LVAD). J Thorac Dis 2018;10:S1790–S1793. doi:10.21037/jtd.2017.10.124.
[28] Mehra MR, Uriel N, Naka Y, Cleveland JC, Yuzefpolskaya M, Salerno CT, et al. A fully magnetically levitated left ventricular assist device - Final report. N Engl J Med 2019;380:1618–1627. doi:10.1056/NEJMoa1900486.
[29] Wiegn P, Chan R, Jost C, Saville BR, Parise H, Prutchi D, et al. Safety, Performance, and Efficacy of Cardiac Contractility Modulation Delivered by the 2-Lead Optimizer Smart System: The FIX-HF-5C2 Study. Circ Heart Fail 2020;13:e006512. doi:10.1161/CIRCHEARTFAILURE.119.006512.
[30] Zakrzewski W, Dobrzyński M, Szymonowicz M, Rybak Z. Stem cells: past, present, and future. Stem Cell Res Ther 2019;10:68. doi:10.1186/s13287-019-1165-5.
[31] Sanganalmath SK, Boli R. Cell therapy for heart failure: A comprehensive overview of experimental and clinical studies, current challenges, and future directions. Circ Res 2013;113:810–834. doi:10.1161/CIRCRESAHA.113.300219.Cell.
[32] Krampera M, Franchini M, Pizzolo G, Aprili G. Mesenchymal stem cells: From biology to clinical use. Blood Transfus 2007;5:120–129. doi:10.2450/2007.0029-07.
[33] Pucéat M. Protocols for cardiac differentiation of embryonic stem cells. Methods 2008;45:168–171. doi:10.1016/j.ymeth.2008.03.003.
[34] Kokkinopoulos I, Ishida H, Saba R, Coppen S, Suzuki K, Yashiro K. Cardiomyocyte

differentiation from mouse embryonic stem cells using a simple and defined protocol. Dev Dyn 2016;245:157–165. doi:10.1002/dvdy.24366.
[35] Hartman ME, Librande JR, Medvedev IO, Ahmad RN, Moussavi-Harami F, Gupta PP, et al. An optimized and simplified system of mouse embryonic stem cell cardiac differentiation for the assessment of differentiation modifiers. PLoS One 2014;9. doi:10.1371/journal.pone.0093033.
[36] Park SJ, Bae D, Moon SH, Chung HM. Modification of a purification and expansion method for human embryonic stem cell-derived cardiomyocytes. Cardiol 2013;124:139–150. doi:10.1159/000346390.
[37] Ardehali R, Ali SR, Inlay MA, Abilez OJ, Chen MQ, Blauwkamp TA, et al. Prospective isolation of human embryonic stem cell-derived cardiovascular progenitors that integrate into human fetal heart tissue. Proc Natl Acad Sci U S A 2013;110:3405–3410. doi:10.1073/pnas.1220832110.
[38] He JQ, Ma Y, Lee Y, Thomson JA, Kamp TJ. Human embryonic stem cells develop into multiple types of cardiac myocytes: Action potential characterization. Circ Res 2003;93:32–39. doi:10.1161/01.RES.0000080317.92718.99.
[39] Xu C, Police S, Rao N, Carpenter MK. Characterization and enrichment of cardiomyocytes derived from human embryonic stem cells. Circ Res 2002;91:501–508. doi:10.1161/01.RES.0000035254.80718.91.
[40] Mummery CL, Zhang J, Ng E, Elliott DA, Elefanty AG, Kamp TJ. Differentiation of Human ES and iPS Cells to Cardiomyocytes: A Methods Overview. Circ Res 2012;111:344–358. doi:10.1161/CIRCRESAHA.110.227512.Differentiation.
[41] Kehat I, Kenyagin-Karsenti D, Snir M, Segev H, Amit M, Gepstein A, et al. Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. J Clin Invest 2001;108:407–414. doi:10.1172/JCI200112131.
[42] Claudine Ménard P, Prof Albert A Hagège M, Onnik Agbulut M, Marietta Barro M, Miguel Cortes Morichetti M, Camille Brasselet M, et al. Transplantation of cardiac-committed mouse embryonic stem cells to infarcted sheep myocardium: a preclinical study. Lancet 2005;366. doi:DOI:https://doi.org/10.1016/S0140-6736(05)67380-1.
[43] Silva dos Santos D, Brasil GV, Ramos IPR, Mesquita FCP, Kasai-Brunswick TH, Christie MLA, et al. Embryonic stem cell-derived cardiomyocytes for the treatment of doxorubicin-induced cardiomyopathy. Stem Cell Res Ther 2018;9:30. doi:10.1186/s13287-018-0788-2.
[44] Cai J, Yi FF, Yang XC, Lin GS, Jiang H, Wang T, et al. Transplantation of embryonic stem cell-derived cardiomyocytes improves cardiac function in infarcted rat hearts. Cytotherapy 2007;9:283–291. doi:10.1080/14653240701247838.
[45] Zhao T, Zhang ZN, Rong Z, Xu Y. Immunogenicity of induced pluripotent stem cells. Nature 2011;474:212–216. doi:10.1038/nature10135.
[46] Jeevanantham V, Butler M, Saad A, Abdel-Latif A, Zuba-Surma EK, Dawn B. Adult Bone Marrow Cell Therapy Improves Survival and Induces Long-Term Improvement in Cardiac Parameters. Circulation 2012;126:551–568. doi:10.1161/CIRCULATIONAHA.111.086074.
[47] Meyer GP, Wollert KC, Lotz J, Pirr J, Rager U, Lippolt P, et al. Intracoronary bone marrow cell transfer after myocardial infarction: 5-year follow-up from the randomized-controlled BOOST trial. Eur Heart J 2009;30:2978–2984. doi:10.1093/eurheartj/ehp374.

[48] Assmus B, Rolf A, Erbs S, Elsässer A, Haberbosch W, Hambrecht R, et al. Clinical outcome 2 years after intracoronary administration of bone marrow-derived progenitor cells in acute myocardial infarction. Circ Hear Fail 2010;3:89–96. doi:10.1161/CIRCHEARTFAILURE.108.843243.
[49] Leistner DM, Fischer-Rasokat U, Honold J, Seeger FH, Schächinger V, Lehmann R, et al. Transplantation of progenitor cells and regeneration enhancement in acute myocardial infarction (TOPCARE-AMI): final 5-year results suggest long-term safety and efficacy. Clin Res Cardiol 2011;100:925–934. doi:10.1007/s00392-011-0327-y.
[50] Bartunek J, Behfar A, Dolatabadi D, Vanderheyden M, Ostojic M, Dens J, et al. Cardiopoietic stem cell therapy in heart failure: The C-CURE (cardiopoietic stem cell therapy in heart failURE) multicenter randomized trial with lineage-specified biologics. J Am Coll Cardiol 2013;61:2329–2338. doi:10.1016/j.jacc.2013.02.071.
[51] Doppler SA, Deutsch MA, Lange R, Krane M. Cardiac regeneration: Current therapies-future concepts. J Thorac Dis 2013;5:683–697. doi:10.3978/j.issn.2072-1439.2013.08.71.
[52] Bolli R, Chugh AR, Anversa P. Effect of Cardiac Stem Cells in Patients with Ischemic Cardiomyopathy : Initial Results of the SCIPIO Trial Study protocol 2019;393:7–11.
[53] Malliaras K, Makkar RR, Smith RR, Cheng K, Wu E, Bonow RO, et al. Intracoronary Cardiosphere-Derived Cells After Myocardial Infarction. J Am Coll Cardiol 2014;63:110–122. doi:10.1016/j.jacc.2013.08.724.
[54] Gnecchi M, Zhang Z, Ni A, Dzau VJ. Paracrine Mechanisms in Adult Stem Cell Signaling and Therapy. Circ Res 2008;103:1204–1219. doi:10.1161/CIRCRESAHA.108.176826.
[55] Broughton KM, Sussman MA. Empowering Adult Stem Cells for Myocardial Regeneration V2.0. Circ Res 2016;118:867–880. doi:10.1161/CIRCRESAHA.115.305227.
[56] Tachibana M, Amato P, Sparman M, Gutierrez NM, Tippner-Hedges R, Ma H, et al. Human embryonic stem cells derived by somatic cell nuclear transfer. Cell 2013;153:1228–1238. doi:10.1016/j.cell.2013.05.006.
[57] Gouveia C, Huyser C, Egli D, Pepper MS. Lessons learned from somatic cell nuclear transfer. Int J Mol Sci 2020;21. doi:10.3390/ijms21072314.
[58] Takahashi K, Yamanaka S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell 2006;126:663–676. doi:10.1016/j.cell.2006.07.024.
[59] Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell 2007;131:861–872. doi:10.1016/j.cell.2007.11.019.
[60] Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, et al. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat Biotechnol 2008;26:101–106. doi:10.1038/nbt1374.
[61] Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, et al. Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. Science (80- ) 2007;318:1917–1920. doi:10.1126/science.1151526.
[62] Raynaud CM, Ahmad FS, Allouba M, Abou-Saleh H, Lui KO, Yacoub M. Reprogramming for cardiac regeneration. Glob Cardiol Sci Pract 2014;2014:44. doi:10.5339/gcsp.2014.44.
[63] Budniatzky I, Gepstein L. Concise Review: Reprogramming Strategies for Cardiovascular Regenerative Medicine: From Induced Pluripotent Stem Cells to Direct

Reprogramming. Stem Cells Transl Med 2014;3:448–457. doi:10.5966/sctm.2013-0163.
[64] Vladar EK, Lee YL, Stearns T, Axelrod JD. Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes: Insights into Molecular, Cellular, and Functional Phenotypes. Circ Res 2015;117:37–54. doi:10.1016/bs.mcb.2015.01.016.Observing.
[65] Machiraju P, Greenway SC. Current methods for the maturation of induced pluripotent stem cellderived cardiomyocytes. World J Stem Cells 2019;11:33–43. doi:10.4252/wjsc.v11.i1.33.
[66] Bektik E, Fu J. Ameliorating the Fibrotic Remodeling of the Heart through Direct Cardiac Reprogramming. Cells 2019;8:679. doi:10.3390/cells8070679.
[67] Davis RL, Weintraub H, Lassar AB. Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell 1987;51:987–1000. doi:10.1016/0092-8674(87)90585-X.
[68] Ieda M, Fu J-D, Delagado-Olguin P, Vedantham V, Hayashi Y, G.Bruneau B, et al. Direct Reprogramming of Fibroblasts into Functional Cardiomyocytes by Defined Factors Masaki. Cell 2010;306:1895. doi:10.1016/j.cell.2010.07.002.Direct.
[69] Song K, Nam Y-J, Luo X, Qi X, Tan W, Huang GN, et al. Heart repair by reprogramming non-myocytes with cardiac transcription factors. Nature 2012;485:599–604. doi:10.1038/nature11139.
[70] Christoforou N, Chellappan M, Adler AF, Kirkton RD, Wu T, Addis RC, et al. Transcription Factors MYOCD, SRF, Mesp1 and SMARCD3 Enhance the Cardio-Inducing Effect of GATA4, TBX5, and MEF2C during Direct Cellular Reprogramming. PLoS One 2013;8. doi:10.1371/journal.pone.0063577.
[71] Islas JF, Liu Y, Weng K-C, Robertson MJ, Zhang S, Prejusa A, et al. Transcription factors ETS2 and MESP1 transdifferentiate human dermal fibroblasts into cardiac progenitors. Proc Natl Acad Sci 2012;109:13016–13021. doi:10.1073/pnas.1120299109.
[72] Wada R, Muraoka N, Inagawa K, Yamakawa H, Miyamoto K, Sadahiro T, et al. Induction of human cardiomyocyte-like cells from fibroblasts by defined factors. Proc Natl Acad Sci U S A 2013;110:12667–12672. doi:10.1073/pnas.1304053110.
[73] Fu JD, Stone NR, Liu L, Spencer CI, Qian L, Hayashi Y, et al. Direct reprogramming of human fibroblasts toward a cardiomyocyte-like state. Stem Cell Reports 2013;1:235–247. doi:10.1016/j.stemcr.2013.07.005.
[74] Ferrari A, Vincent-Salomon A, Pivot X, Sertier AS, Thomas E, Tonon L, et al. A whole-genome sequence and transcriptome perspective on HER2-positive breast cancers. Nat Commun 2016;7. doi:10.1038/ncomms12222.
[75] Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE, Davidson NE, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N Engl J Med 2005;353:1673–1684. doi:10.1056/NEJMoa052122.
[76] Tiscornia G, Vivas EL, Belmonte JCI. Diseases in a dish: Modeling human genetic disorders using induced pluripotent cells. Nat Med 2011;17:1570–1576. doi:10.1038/nm.2504.
[77] Zanella F, Lyon RC, Sheikh F. Modeling heart disease in a dish: From somatic cells to disease-relevant cardiomyocytes. Trends Cardiovasc Med 2014;24:32–44. doi:10.1016/j.tcm.2013.06.002.
[78] Jung CB, Moretti A, Mederos y Schnitzler M, Iop L, Storch U, Bellin M, et al. Dantrolene rescues arrhythmogenic RYR2 defect in a patient-specific stem cell model of catecholaminergic polymorphic ventricular tachycardia. EMBO Mol Med

2012;4:180–191. doi:10.1002/emmm.201100194.
[79] Cutiongco MFA, Jensen BS, Reynolds PM, Gadegaard N. Predicting gene expression using morphological cell responses to nanotopography. Nat Commun 2020;11:1–13. doi:10.1038/s41467-020-15114-1.
[80] Hamel V, Cheng K, Liao S, Lu A, Zheng Y, Chen Y, et al. De Novo Human Cardiac Myocytes for Medical Research: Promises and Challenges. Stem Cells Int 2017;2017. doi:10.1155/2017/4528941.
[81] Tam PPL, Behringer RR. Mouse gastrulation: The formation of a mammalian body plan. Mech Dev 1997;68:3–25. doi:10.1016/S0925-4773(97)00123-8.
[82] Rivera-Pérez JA, Hadjantonakis A-K. The Dynamics of Morphogenesis in the Early Mouse Embryo. Cold Spring Harb Perspect Biol 2015;7:a015867. doi:10.1101/cshperspect.a015867.
[83] Brade T, Pane LS, Moretti A, Chien KR, Laugwitz K. Embryonic heart progenitors and cardiogenesis. Cold Spring Harb Perspect Med 2013;3:1–18. doi:10.1101/cshperspect.a013847.
[84] Gittenberger-De Groot AC, Bartelings MM, Deruiter MC, Poelmann RE. Basics of cardiac development for the understanding of congenital heart malformations. Pediatr Res 2005;57:169–176. doi:10.1203/01.PDR.0000148710.69159.61.
[85] Schleich JM, Abdulla T, Summers R, Houyel L. An overview of cardiac morphogenesis. Arch Cardiovasc Dis 2013;106:612–623. doi:10.1016/j.acvd.2013.07.001.
[86] Meilhac SM, Lescroart F, Blanpain CD, Buckingham ME. Cardiac cell lineages that form the heart. Cold Spring Harb Perspect Med 2014;4:1–14. doi:10.1101/cshperspect.a013888.
[87] Lescroart F, Chabab S, Lin X, Rulands S, Paulissen C, Rodolosse A, et al. Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. Nat Cell Biol 2014;16:829–840. doi:10.1038/ncb3024.
[88] Devine WP, Wythe JD, George M, Koshiba-Takeuchi K, Bruneau BG. Early patterning and specification of cardiac progenitors in gastrulating mesoderm. Elife 2014;3:1–23. doi:10.7554/eLife.03848.
[89] Saga Y, Hata N, Kobayashi S, Magnuson T, Seldin MF, Taketo MM. MesP1: a novel basic helix-loop-helix protein expressed in the nascent mesodermal cells during mouse gastrulation. Development 1996;122:2769–2778. doi:10.1016/0169-328x(95)00282-w.
[90] Jones S. An overview of the basic helix-loop-helix proteins. Genome Biol 2004;5:2004. doi:10.1186/gb-2004-5-6-226.
[91] Bondue A, Blanpain C. Mesp1: A key regulator of cardiovascular lineage commitment. Circ Res 2010;107:1414–1427. doi:10.1161/CIRCRESAHA.110.227058.
[92] Soibam B, Benham A, Kim J, Weng KC, Yang L, Xu X, et al. Genome-wide identification of MESP1 targets demonstrates primary regulation over mesendoderm gene activity. Stem Cells 2015;33:3254–3265. doi:10.1002/stem.2111.
[93] Saga Y, Miyagawa-Tomita S, Takagi A, Kitajima S, Miyazaki J, Inoue T. MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. Development 1999;126:3437–3447.
[94] Asahina K, Tsai SY, Li P, Ishii M, Maxson RE, Socov HM, et al. Mesenchymal origin of hepatic stellate cells, submesothelial cells, and perivascular mesenchymal cells during mouse liver development. Hepatology 2009;49:998–1011. doi:10.1002/hep.22721.

[95] Harel I, Nathan E, Tirosh-Finkel L, Zigdon H, Guimarães-Camboa N, Evans SM, et al. Distinct Origins and Genetic Programs of Head Muscle Satellite Cells. Dev Cell 2009;16:822–832. doi:10.1016/j.devcel.2009.05.007.
[96] Liang Q, Xu C, Chen X, Li X, Lu C, Zhou P, et al. The roles of Mesp family proteins: functional diversity and redundancy in differentiation of pluripotent stem cells and mammalian mesodermal development. Protein Cell 2015;6:553–561. doi:10.1007/s13238-015-0176-y.
[97] Lescroart F, Wang X, Lin X, Swedlund B, Gargouri S, Sànchez-Dànes A, et al. Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. Science (80- ) 2018;359:1177–1181. doi:10.1126/science.aao4174.
[98] Saga Y. Genetic rescue of segmentation defect in MesP2-deficient mice by MesP1 gene replacement. Mech Dev 1998;75:53–66. doi:10.1016/S0925-4773(98)00077-X.
[99] Satou Y, Imai KS, Satoh N. The ascidian Mesp gene specifies heart precursor cells. Development 2004;131:2533–2541. doi:10.1242/dev.01145.
[100] Kitajima S, Takagi A, Inoue T, Saga Y. MesP1 and MesP2 are essential for the development of cardiac mesoderm. Development 2000;127:3215–3226.
[101] Haraguchi S, Kitajima S, Takagi A, Takeda H, Inoue T, Saga Y. Transcriptional regulation of Mesp1 and Mesp2 genes: Differential usage of enhancers during development. Mech Dev 2001;108:59–69. doi:10.1016/S0925-4773(01)00478-6.
[102] Saga Y, Kitajima S, Miyagawa-Tomita S. Mesp1 expression is the earliest sign of cardiovascular development. Trends Cardiovasc Med 2000;10:345–352. doi:10.1016/S1050-1738(01)00069-X.
[103] Siggia ED, Warmflash A. Modeling Mammalian Gastrulation With Embryonic Stem Cells. Physiol. Behav., vol. 176, 2018, p. 1–23. doi:10.1016/bs.ctdb.2018.03.001.
[104] Evans T. Embryonic stem cells as a model for cardiac development and disease. Drug Discov Today Dis Model 2008;5:147–155. doi:10.1016/j.ddmod.2009.03.004.
[105] Burridge PW, Sharma A, Wu JC. Genetic and Epigenetic Regulation of Human Cardiac Reprogramming and Differentiation in Regenerative Medicine. Annu Rev Genet 2016;49:461–484. doi:10.1146/annurev-genet-112414-054911.Genetic.
[106] Cripps RM, Olson EN. Control of cardiac development by an evolutionarily conserved transcriptional network. Dev Biol 2002;246:14–28. doi:10.1006/dbio.2002.0666.
[107] Lindsley RC. Canonical Wnt signaling is required for development of embryonic stem cell-derived mesoderm. Development 2006;133:3787–3796. doi:10.1242/dev.02551.
[108] Bondue A, Lapouge G, Paulissen C, Semeraro C, Iacovino M, Kyba M, et al. Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. Cell Stem Cell 2008;3:69–84. doi:10.1016/j.stem.2008.06.009.
[109] David R, Brenner C, Stieber J, Schwarz F, Brunner S, Vollmer M, et al. MesP1 drives vertebrate cardiovascular differentiation through Dkk-1-mediated blockade of Wnt-signalling. Nat Cell Biol 2008;10:338–345. doi:10.1038/ncb1696.
[110] Bondue A, Lapouge G, Paulissen C, Semeraro C, Iacovino M, Kyba M, et al. Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. Cell Stem Cell 2008;3:69–84. doi:10.1016/j.stem.2008.06.009.
[111] Lindsley RC, Gill JG, Murphy TL, Langer EM, Cai M, Mashayekhi M, et al. Mesp1 coordinately regulates cardiovascular fate restriction and epithelial-mesenchymal transition in differentiating ESCs. Cell Stem Cell 2008;3:55–68. doi:10.1016/j.stem.2008.04.004.
[112] Chan SS, Shi X, Toyama A, Arpke RW, Iacovino M, Kang J, et al. Mesp1 patterns

mesoderm into cardiac, hematopoietic, or skeletal myogenic progenitors in a context-dependent manner. Cell Stem Cell 2013;12:587–601. doi:10.1016/j.stem.2013.03.004.Mesp1.
[113] Harel I, Nathan E, Tirosh-Finkel L, Zigdon H, Guimarães-Camboa N, Evans SM, et al. Distinct Origins and Genetic Programs of Head Muscle Satellite Cells. Dev Cell 2009;16:822–832. doi:10.1016/j.devcel.2009.05.007.
[114] Chabab S, Lescroart F, Rulands S, Mathiah N, Simons BD, Blanpain C. Uncovering the Number and Clonal Dynamics of Mesp1 Progenitors during Heart Morphogenesis. Cell Rep 2016;14:1–10. doi:10.1016/j.celrep.2015.12.013.
[115] Chan SSK, Hagen HR, Swanson SA, Stewart R, Boll KA, Aho J, et al. Development of Bipotent Cardiac/Skeletal Myogenic Progenitors from MESP1+ Mesoderm. Stem Cell Reports 2016;6:26–34. doi:10.1016/j.stemcr.2015.12.003.
[116] Chiapparo G, Lin X, Lescroart F, Chabab S, Paulissen C, Pitisci L, et al. Mesp1 controls the speed, polarity, and directionality of cardiovascular progenitor migration. J Cell Biol 2016;213:463–477. doi:10.1083/jcb.201505082.
[117] Gibbs BC, Damerla RR, Vladar EK, Chatterjee B, Wan Y, Liu X, et al. Prickle1 mutation causes planar cell polarity and directional cell migration defects associated with cardiac outflow tract anomalies and other structural birth defects. Biol Open 2016;5:323–335. doi:10.1242/bio.015750.
[118] Bondue A, Tännler S, Chiapparo G, Chabab S, Ramialison M, Paulissen C, et al. Defining the earliest step of cardiovascular progenitor specification during embryonic stem cell differentiation. J Cell Biol 2011;192:751–765. doi:10.1083/jcb.201007063.
[119] Tam PPL, Loebel DAF. Gene function in mouse embryogenesis: get set for gastrulation. Nat Rev Genet 2007;8:368–381. doi:10.1038/nrg2084.
[120] David R, Jarsch VB, Schwarz F, Nathan P, Gegg M, Lickert H, et al. Induction of MesP1 by Brachyury(T) generates the common multipotent cardiovascular stem cell. Cardiovasc Res 2011;92:115–122. doi:10.1093/cvr/cvr158.
[121] Liu Y, Schwartz RJ. Transient mesp1 expression: A driver of cardiac cell fate determination. Transcription 2013;4:92–96. doi:10.4161/trns.24588.
[122] Costello I, Pimeisl I-M, Dräger S, Bikoff EK, Robertson EJ, Arnold SJ. The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. Nat Cell Biol 2011;13:1084–1091. doi:10.1038/ncb2304.
[123] van den Ameele J, Tiberi L, Bondue A, Paulissen C, Herpoel A, Iacovino M, et al. Eomesodermin induces Mesp1 expression and cardiac differentiation from embryonic stem cells in the absence of Activin. EMBO Rep 2012;13:355–362. doi:10.1038/embor.2012.23.
[124] Guo X, Xu Y, Wang Z, Wu Y, Chen J, Wang G, et al. A Linc1405/Eomes Complex Promotes Cardiac Mesoderm Specification and Cardiogenesis. Cell Stem Cell 2018;22:893-908.e6. doi:10.1016/j.stem.2018.04.013.
[125] Tosic J, Kim GJ, Pavlovic M, Schröder CM, Mersiowsky SL, Barg M, et al. Eomes and Brachyury control pluripotency exit and germ-layer segregation by changing the chromatin state. Nat Cell Biol 2019;21:1518–1531. doi:10.1038/s41556-019-0423-1.
[126] Arnold SJ, Robertson EJ. Making a commitment: Cell lineage allocation and axis patterning in the early mouse embryo. Nat Rev Mol Cell Biol 2009;10:91–103. doi:10.1038/nrm2618.
[127] Liu P, Wakamiya M, Shea MJ, Albrecht U, Behringer RR, Bradley A. Requirement for

Wnt3 in vertebrate axis formation. Nat Genet 1999;22:361–365. doi:10.1038/11932.
[128] Kimelman D. Mesoderm induction: From caps to chips. Nat Rev Genet 2006;7:360–372. doi:10.1038/nrg1837.
[129] Zorn AM, Wells JM. Vertebrate Endoderm Development and Organ Formation. Annu Rev Cell Dev Biol 2009;25:221–251. doi:10.1146/annurev.cellbio.042308.113344.
[130] Zhuang Y, Kim CG, Bartelmez S, Cheng P, Groudine M, Weintraub H. Helix-loop-helix transcription factors E12 and E47 are not essential for skeletal or cardiac myogenesis, erythropoiesis, chondrogenesis, or neurogenesis. Proc Natl Acad Sci U S A 1992;89:12132–12136. doi:10.1073/pnas.89.24.12132.
[131] Davidson B, Shi W, Beh J, Christiaen L, Levine M. FGF signaling delineates the cardiac progenitor field in the simple chordate, Ciona intestinalis. Genes Dev 2006;20:2728–2738. doi:10.1101/gad.1467706.
[132] Davidson B, Shi W, Levine M. Uncoupling heart cell specification and migration in the simple chordate Ciona intestinalis. Development 2005;132:4811–4818. doi:10.1242/dev.02051.
[133] Shi X, Zirbes KM, Rasmussen TL, Ferdous A, Garry MG, Koyano-Nakagawa N, et al. The transcription factor mesp1 Interacts with cAMPresponsive element binding protein 1 (creb1) and coactivates Ets variant 2 (Etv2) gene expression. J Biol Chem 2015;290:9614–9625. doi:10.1074/jbc.M114.614628.
[134] Hartung S, Schwanke K, Haase A, David R, Franz W-M, Martin U, et al. Directing Cardiomyogenic Differentiation of Human Pluripotent Stem Cells by Plasmid-Based Transient Overexpression of Cardiac Transcription Factors. Stem Cells Dev 2013;22:1112–1125. doi:10.1089/scd.2012.0351.
[135] Chan SSK, Shi X, Toyama A, Arpke RW, Dandapat A, Iacovino M, et al. Mesp1 patterns mesoderm into cardiac, hematopoietic, or skeletal myogenic progenitors in a context-dependent manner. Cell Stem Cell 2013;12:587–601. doi:10.1016/j.stem.2013.03.004.
[136] Werner P, Latney B, Deardorff MA, Goldmuntz E. MESP1 mutations in patients with congential heart defects. Hum Mutat 2016;37:302–314. doi:10.1002/humu.22947.
[137] Mohammed H, Taylor C, Brown GD, Papachristou EK, Carroll JS, D'Santos CS. Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. Nat Protoc 2016;11:316–326. doi:10.1038/nprot.2016.020.
[138] Mohammed H, D'Santos C, Serandour AA, Ali HR, Brown GD, Atkins A, et al. Endogenous Purification Reveals GREB1 as a Key Estrogen Receptor Regulatory Factor. Cell Rep 2013;3:342–349. doi:10.1016/j.celrep.2013.01.010.
[139] Choi J, Costa ML, Mermelstein CS, Chagas C, Holtzer S, Holtzer H. MyoD converts primary dermal fibroblasts, chondroblasts, smooth muscle, and retinal pigmented epithelial cells into striated mononucleated myoblasts and multinucleated myotubes. Proc Natl Acad Sci U S A 1990;87:7988–7992. doi:10.1073/pnas.87.20.7988.
[140] Chen JX, Krane M, Deutsch M-A, Wang L, Rav-Acha M, Gregoire S, et al. Inefficient Reprogramming of Fibroblasts into Cardiomyocytes Using Gata4, Mef2c, and Tbx5. Circ Res 2012;111:50–55. doi:10.1161/CIRCRESAHA.112.270264.
[141] Nashun B, Hill PW, Hajkova P. Reprogramming of cell fate: epigenetic memory and the erasure of memories past. EMBO J 2015;34:1296–1308. doi:10.15252/embj.201490649.
[142] Heller B, Adu-Gyamfi E, Smith-Kinnaman W, Babbey C, Vora M, Xue Y, et al. Amot recognizes a juxtanuclear endocytic recycling compartment via a novel lipid binding

domain. J Biol Chem 2010;285:12308–12320. doi:10.1074/jbc.M109.096230.
[143] Sawada a, Fritz a, Jiang YJ, Yamamoto a, Yamasu K, Kuroiwa a, et al. Zebrafish Mesp family genes, mesp-a and mesp-b are segmentally expressed in the presomitic mesoderm, and Mesp-b confers the anterior identity to the developing somites. Development 2000;127:1691–1702.
[144] Deshwar AR, Onderisin JC, Aleksandrova A, Yuan X, Burrows JTA, Scott IC. Mespaa can potently induce cardiac fates in zebrafish. Dev Biol 2016;418:17–27. doi:10.1016/j.ydbio.2016.08.022.
[145] Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science (80- ) 2015;347:1260419–1260419. doi:10.1126/science.1260419.
[146] Lahm H, Deutsch M-A, Dreßen M, Doppler S, Werner A, Hörer J, et al. Mutational analysis of the human MESP1 gene in patients with congenital heart disease reveals a highly variable sequence in exon 1. Eur J Med Genet 2013;56:591–598. doi:10.1016/j.ejmg.2013.09.001.
[147] Moore-Scott BA, Opoka R, Lin SCJ, Kordich JJ, Wells JM. Identification of molecular markers that are expressed in discrete anterior-posterior domains of the endoderm from the gastrula stage to mid-gestation. Dev Dyn 2007;236:1997–2003. doi:10.1002/dvdy.21204.
[148] Aase K, Ernkvist M, Ebarasi L, Jakobsson L, Majumdar A, Yi C, et al. Angiomotin regulates endothelial cell migration during embryonic angiogenesis. Genes Dev 2007;21:2055–2068. doi:10.1101/gad.432007.
[149] Thisse, B., Pflumio, S., Fürthauer, M., Loppin, B., Heyer, V., Degrave, A., Woehl, R., Lux, A., Steffan, T., Charbonnier, X.Q. and Thisse, C.Thisse, B., Pflumio, S., Fürthauer, M., Loppin, B., Heyer, V., Degrave, A., Woehl, R., Lux, A., Steffan, T., Charb C. Expression of the zebrafish genome during embryogenesis (NIH R01 RR15402)itle. ZFIN Direct Data Submiss 2001.
[150] Huang T, Zhou Y, Zhang J, Cheng ASL, Yu J, To KF, et al. The physiological role of Motin family and its dysregulation in tumorigenesis. J Transl Med 2018;16:1–13. doi:10.1186/s12967-018-1466-y.
[151] Flores-Alcantar A, Gonzalez-Sandoval A, Escalante-Alcalde D, Lomelí H. Dynamics of expression of ARID1A and ARID1B subunits in mouse embryos and in cells during the cell cycle. Cell Tissue Res 2011;345:137–148. doi:10.1007/s00441-011-1182-x.
[152] Chandler RL, Brennan J, Schisler JC, Serber D, Patterson C, Magnuson T. ARID1a-DNA Interactions Are Required for Promoter Occupancy by SWI/SNF. Mol Cell Biol 2013;33:265–280. doi:10.1128/mcb.01008-12.
[153] Kosho T, Okamoto N, Imai Y, Ohashi H, van Eerde AM, Chrzanowska K, et al. Genotype-phenotype correlation of coffin-siris syndrome caused by mutations in SMARCB1, SMARCA4, SMARCE1, and ARID1A. Am J Med Genet Part C Semin Med Genet 2014;166:262–275. doi:10.1002/ajmg.c.31407.
[154] Uribe V, Badóa-Careaga C, Domínguez JN, Domínguez JN, De La Pompa JL, Sanz-Ezquerro JJ. Arid3b is essential for second heart field cell deployment and heart patterning. Dev 2014;141:4168–4181. doi:10.1242/dev.109918.
[155] Lin C, Song W, Bi X, Zhao J, Huang Z, Li Z, et al. Recent advances in the ARID family: Focusing on roles in human cancer. Onco Targets Ther 2014;7:315–324. doi:10.2147/OTT.S57023.
[156] Xie J, Gao S, Schafer C, Colijn S, Muthukumar V, Griffin CT. The chromatin-remodeling

enzyme CHD3 plays a role in embryonic viability but is dispensable for early vascular development. PLoS One 2020;15:1–12. doi:10.1371/journal.pone.0235799.
[157] Diez-Roux G, Banfi S, Sultan M, Geffers L, Anand S, Rozado D, et al. A high-resolution anatomical atlas of the transcriptome in the mouse embryo. PLoS Biol 2011;9. doi:10.1371/journal.pbio.1000582.
[158] Pfefferli C, Müller F, Jaźwińska A, Wicky C. Specific NuRD components are required for fin regeneration in zebrafish. BMC Biol 2014;12:1–17. doi:10.1186/1741-7007-12-30.
[159] Fitzgerald TW, Gerety SS, Jones WD, Van Kogelenberg M, King DA, McRae J, et al. Large-scale discovery of novel genetic causes of developmental disorders. Nature 2015;519:223–228. doi:10.1038/nature14135.
[160] Snijders Blok L, Rousseau J, Twist J, Ehresmann S, Takaku M, Venselaar H, et al. CHD3 helicase domain mutations cause a neurodevelopmental syndrome with macrocephaly and impaired speech and language. Nat Commun 2018;9:1–12. doi:10.1038/s41467-018-06014-6.
[161] Hoffmeister H, Fuchs A, Erdel F, Pinz S, Gröbner-Ferreira R, Bruckmann A, et al. CHD3 and CHD4 form distinct NuRD complexes with different yet overlapping functionality. Nucleic Acids Res 2017;45:10534–10554. doi:10.1093/nar/gkx711.
[162] Ciruna BG, Rossant J. Expression of the T-box gene Eomesodermin during early mouse development. Mech Dev 1999;81:199–203. doi:10.1016/S0925-4773(98)00243-3.
[163] Russ AP, Wattler S, Colledge WH, Aparicio SAJR, Carlton MBL, Pearce JJ, et al. Eomesodermin is required for mouse trophoblast development and mesoderm formation. Nature 2000;404:95–99. doi:10.1038/35003601.
[164] Du S, Draper BW, Mione M, Moens CB, Bruce A. Differential regulation of epiboly initiation and progression by zebrafish Eomesodermin A. Dev Biol 2012;362:11–23. doi:10.1016/j.ydbio.2011.10.036.
[165] Baala L, Briault S, Etchevers HC, Laumonnier F, Natiq A, Amiel J, et al. Homozygous silencing of T-box transcription factor EOMES leads to microcephaly with polymicrogyria and corpus callosum agenesis. Nat Genet 2007;39:454–456. doi:10.1038/ng1993.
[166] Dielmann A, Letsch A, Nonnenmacher A, Miller K, Keilholz U, Busse A. Favorable prognostic influence of T-box transcription factor Eomesodermin in metastatic renal cell cancer patients. Cancer Immunol Immunother 2016;65:181–192. doi:10.1007/s00262-015-1786-1.
[167] Ueda T, Watanabe-fukunaga R, Ogawa H, Fukuyama H, Higashi Y, Nagata S, et al. Critical role of the p400/mDomino chromatin-remodeling ATPase in embryonic hematopoiesis. Genes to Cells 2007;12:581–592. doi:10.1111/j.1365-2443.2007.01080.x.
[168] Hong W, Hu Y, Fan Z, Gao R, Yang R, Bi J, et al. In silico identification of EP400 and TIA1 as critical transcription factors involved in human hepatocellular carcinoma relapse. Oncol Lett 2020;19:952–964. doi:10.3892/ol.2019.11171.
[169] Zhou W, Chung YJ, Parrilla Castellar ER, Zheng Y, Chung HJ, Bandle R, et al. Far upstream element binding protein plays a crucial role in embryonic development, hematopoiesis, and stabilizing myc expression levels. Am J Pathol 2016;186:701–715. doi:10.1016/j.ajpath.2015.10.028.
[170] Elman JS, Ni TK, Mengwasser KE, Jin D, Wronski A, Elledge SJ, et al. Identification of FUBP1 as a Long Tail Cancer Driver and Widespread Regulator of Tumor Suppressor and Oncogene Alternative Splicing. Cell Rep 2019;28:3435-3449.e5.

doi:10.1016/j.celrep.2019.08.060.
[171] Hendrich B, Guy J, Ramsahoye B, Wilson VA, Bird A. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. Genes Dev 2001;1:710–723. doi:10.1101/gad.194101.binds.
[172] Rauch, G.J., Lyons, D.A., Middendorf, I., Friedlander, B., Arana, N., Reyes, T., and Talbot WS. Submission and Curation of Gene Expression Data. ZFIN Direct Data Submiss 2003.
[173] Fillatre J, Fauny J-D, Fels JA, Li C, Goll M, Thisse C, et al. TEADs, Yap, Taz, Vgll4s transcription factors control the establishment of Left-Right asymmetry in zebrafish. Elife 2019;8:2–6. doi:10.7554/eLife.45241.
[174] Ranjan A, Sinha S, Ayabe R, Wach M, Ruff S, Gupta S, et al. Abstract C44: MBD3 stabilizes MYC, leading to metastatic outgrowth of pancreatic cancer in the liver. Model Syst., American Association for Cancer Research; 2019, p. C44–C44. doi:10.1158/1538-7445.PANCA19-C44.
[175] Sakaki-Yumoto M, Kobayashi C, Sato A, Fujimura S, Matsumoto Y, Takasato M, et al. The murine homolog of SALL4, a causative gene in Okihiro syndrome, is essential for embryonic stem cell proliferation, and cooperates with Sall1 in anorectal, heart, brain and kidney development. Development 2006;133:3005–3013. doi:10.1242/dev.02457.
[176] Paik EJ, Mahony S, White RM, Price EN, Dibiase A, Dorjsuren B, et al. A Cdx4-sall4 regulatory module controls the transition from mesoderm formation to embryonic hematopoiesis. Stem Cell Reports 2013;1:425–436. doi:10.1016/j.stemcr.2013.10.001.
[177] Dong X, Li J, He L, Gu C, Jia W, Yue Y, et al. Zebrafish Znfl1 proteins control the expression of hoxb1b gene in the posterior neuroectoderm by acting upstream of pou5f3 and sall4 genes. J Biol Chem 2017;292:13045–13055. doi:10.1074/jbc.M117.777094.
[178] Kohlhase J. Okihiro syndrome is caused by SALL4 mutations. Hum Mol Genet 2002;11:2979–2987. doi:10.1093/hmg/11.23.2979.
[179] Yokoyama S, Ito Y, Ueno-Kudoh H, Shimizu H, Uchibe K, Albini S, et al. A Systems Approach Reveals that the Myogenesis Genome Network Is Regulated by the Transcriptional Repressor RP58. Dev Cell 2009;17:836–848. doi:10.1016/j.devcel.2009.10.011.
[180] Brunskill EW, Potter AS, Distasio A, Dexheimer P, Plassard A, Aronow BJ, et al. A gene expression atlas of early craniofacial development. Dev Biol 2014;391:133–146. doi:10.1016/j.ydbio.2014.04.016.
[181] Skarnes WC, Rosen B, West AP, Koutsourakis M, Bushell W, Iyer V, et al. A conditional knockout resource for the genome-wide study of mouse gene function. Nature 2011;474:337–342. doi:10.1038/nature10163.
[182] International mouse phenotyping consortium. Supt5 Data Chart n.d. https://www.mousephenotype.org/data/charts?accession=MGI:1202400&allele_accession_id=MGI:4840996&zygosity=homozygote&parameter_stable_id=IMPC_VIA_001_001&pipeline_stable_id=MGP_001&procedure_stable_id=IMPC_VIA_001&parameter_stable_id=IMPC_VIA_001_001&phen.
[183] Krishnan K, Salomonis N, Guo S. Identification of Spt5 target genes in zebrafish development reveals its dual activity in vivo. PLoS One 2008;3. doi:10.1371/journal.pone.0003621.

[184] Yang Q, Liu X, Zhou T, Cook J, Nguyen K, Bai X. RNA polymerase II pausing modulates hematopoietic stem cell emergence in zebrafish. Blood 2016;128:1701–1710. doi:10.1182/blood-2016-02-697847.
[185] Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhori G, et al. A pathology atlas of the human cancer transcriptome. Science (80- ) 2017;357:eaan2507. doi:10.1126/science.aan2507.
[186] Wang E, Kawaoka S, Roe JS, Shi J, Hohmann AF, Xu Y, et al. The transcriptional cofactor TRIM33 prevents apoptosis in B lymphoblastic leukemia by deactivating a single enhancer. Elife 2015;4:e06377. doi:10.7554/eLife.06377.
[187] Mouse Genome Informatics. Trim33 expression E10.5 RNA insitu n.d. http://www.informatics.jax.org/assay/MGI:3509330#1_id.
[188] Gray PA, Fu H, Luo P, Zhao Q, Yu J, Ferrari A, et al. Mouse brain organization revealed through direct genome-scale TF expression analysis. Science (80- ) 2004;306:2255–2257. doi:10.1126/science.1104935.
[189] Carpenter AC, Rao S, Wells JM, Campbell K, Lang RA. Generation of mice with a conditional null allele for Trim33. Genesis 2008;46:329–333. doi:10.1002/dvg.20651.
[190] Ransom DG, Bahary N, Niss K, Traver D, Burns C, Trede NS, et al. The Zebrafish moonshine gene encodes transcriptional intermediary factor 1γ, an essential regulator of hematopoiesis. PLoS Biol 2004;2. doi:10.1371/journal.pbio.0020237.
[191] Boudinot P, van der Aa LM, Jouneau L, Pasquier L, Pontarotti P, Briolat V, et al. Origin and evolution of TRIM proteins: New insights from the complete TRIM repertoire of zebrafish and pufferfish. PLoS One 2011;6. doi:10.1371/journal.pone.0022022.
[192] Gori F, Zhu ED, Demay MB. Perichondrial expression of Wdr5 regulates chondrocyte proliferation and differentiation. Dev Biol 2009;329:36–43. doi:10.1016/j.ydbio.2009.02.006.
[193] Thisse, Bernard, Thisse C. Fast Release Clones: A High Throughput Expression Analysis. ZFIN Direct Data Submiss 2004.
[194] Haque A, Engel J, Teichmann SA, Lönnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Med 2017;9:1–12. doi:10.1186/s13073-017-0467-4.
[195] Tam PPL, Parameswaran M, Kinder SJ, Weinberger RP. The allocation of epiblast cells to the embryonic heart and other mesodermal lineages: The role of ingression and tissue movement during gastrulation. Development 1997;124:1631–1642.
[196] Pijuan-Sala B, Griffiths JA, Guibentif C, Hiscock TW, Jawaid W, Calero-Nieto FJ, et al. A single-cell molecular map of mouse gastrulation and early organogenesis. Nature 2019;566:490–495. doi:10.1038/s41586-019-0933-9.
[197] Dresser R. Stem cell research as innovation: Expanding the ethical and policy conversation. J Law, Med Ethics 2010;38:332–341. doi:10.1111/j.1748-720X.2010.00492.x.
[198] Wirka RC, Pjanic M, Quertermous T. Advances in transcriptomics investigating cardiovascular disease at unprecedented resolution. Circ Res 2018;122:1200–1220. doi:10.1161/CIRCRESAHA.117.310910.
[199] Wu Y, Zhang J, Peng B, Tian D, Zhang D, Li Y, et al. Generating viable mice with heritable embryonically lethal mutations using the CRISPR-Cas9 system in two-cell embryos. Nat Commun 2019;10. doi:10.1038/s41467-019-10748-2.
[200] Ang Y-S, Tsai S-Y, Lee D-F, Monk J, Su J, Ratnakumar K, et al. Wdr5 Mediates Self-Renewal and Reprogramming via the Embryonic Stem Cell Core Transcriptional

Network. Cell 2011;145:183–197. doi:10.1016/j.cell.2011.03.003.
[201] Hwang YS, Bong GC, Ortmann D, Hattori N, Moeller HC, Khademhosseinia A. Microwell-mediated control of embryoid body size regulates embryonic stem cell fate via differential expression of WNT5a and WNT11. Proc Natl Acad Sci U S A 2009;106:16978–16983. doi:10.1073/pnas.0905550106.
[202] Mohr JC, Zhang J, Azarin SM, Soerens AG, de Pablo JJ, Thomson JA, et al. The microwell control of embryoid body size in order to regulate cardiac differentiation of human embryonic stem cells. Biomaterials 2010;31:1885–1893. doi:10.1016/j.biomaterials.2009.11.033.
[203] Rungarunlert S, Techakumphu M, Pirity MK, Dinnyes A. Embryoid body formation from embryonic and induced pluripotent stem cells: Benefits of bioreactors. World J Stem Cells 2009;1:11. doi:10.4252/wjsc.v1.i1.11.
[204] Messena JM, Hwang NS, Coburn J, Elisseeff JH, Zhang Z. Size of the embryoid body influences chondrogenesis of mouse embryonic stem cells. J Tissue Eng Regen Med 2008;2:499–506. doi:10.1002/term.
[205] Dorn T, Goedel A, Lam JT, Haas J, Tian Q, Herrmann F, et al. Direct Nkx2 5 Transcriptional Repression of Isl1 Controls Cardiomyocyte Subtype Identity. Stem Cells 2014;33.
[206] Pfeiffer MJ, Quaranta R, Piccini I, Fell J, Rao J, Röpke A, et al. Cardiogenic programming of human pluripotent stem cells by dose-controlled activation of EOMES. Nat Commun 2018;9. doi:10.1038/s41467-017-02812-6.
[207] Costello I, Pimeisl I-M, Dräger S, Bikoff EK, Robertson EJ, Arnold SJ. The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. Nat Cell Biol 2011;13:1084–1091. doi:10.1038/ncb2304.
[208] Mor N, Rais Y, Sheban D, Peles S, Aguilera-Castrejon A, Zviran A, et al. Neutralizing Gatad2a-Chd4-Mbd3/NuRD Complex Facilitates Deterministic Induction of Naive Pluripotency. Cell Stem Cell 2018;23:412-425.e10. doi:10.1016/j.stem.2018.07.004.
[209] Luo M, Ling T, Xie W, Sun H, Zhou Y, Zhu Q, et al. NuRD blocks reprogramming of mouse somatic cells into Pluripotent stem cells. Stem Cells 2013;31:1278–1286. doi:10.1002/stem.1374.
[210] Rais Y, Zviran A, Geula S, Gafni O, Chomsky E, Viukov S, et al. Deterministic direct reprogramming of somatic cells to pluripotency. Nature 2013;502:65–70. doi:10.1038/nature12587.
[211] Jaffer S, Goh P, Abbasian M, Nathwani AC. Mbd3 promotes reprogramming of primary human fibroblasts. Int J Stem Cells 2018;11:235–241. doi:10.15283/ijsc18036.
[212] Dos Santos RL, Tosti L, Radzisheuskaya A, Caballero IM, Kaji K, Hendrich B, et al. MBD3/NuRD facilitates induction of pluripotency in a context-dependent manner. Cell Stem Cell 2014;15:102–110. doi:10.1016/j.stem.2014.04.019.
[213] Kaji K, Caballero IM, MacLeod R, Nichols J, Wilson VA, Hendrich B. The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. Nat Cell Biol 2006;8:285–292. doi:10.1038/ncb1372.
[214] Miller T, Krogan NJ, Dover J, Erdjument-Bromage H, Tempst P, Johnston M, et al. COMPASS: A complex of proteins associated with a trithorax-related SET domain protein. Proc Natl Acad Sci U S A 2001;98:12902–12907. doi:10.1073/pnas.231473398.

[215] Guarnaccia A, Tansey W. Moonlighting with WDR5: A Cellular Multitasker. J Clin Med 2018;7:21. doi:10.3390/jcm7020021.
[216] Wysocka J, Swigut T, Milne TA, Dou Y, Zhang X, Burlingame AL, et al. WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. Cell 2005;121:859–872. doi:10.1016/j.cell.2005.03.036.
[217] Ee LS, McCannell KN, Tang Y, Fernandes N, Hardy WR, Green MR, et al. An Embryonic Stem Cell-Specific NuRD Complex Functions through Interaction with WDR5. Stem Cell Reports 2017;8:1488–1496. doi:10.1016/j.stemcr.2017.04.020.
[218] Zaidi S, Choi M, Wakimoto H, Ma L, Jiang J, Overton JD, et al. De novo mutations in histone modifying genes in congenital heart disease. Nature 2013;498:220–223. doi:10.1038/nature12141.De.
[219] Kulkarni SS, Khokha MK. WDR5 regulates left-right patterning via chromatin-dependent and-independent functions. Dev 2018;145. doi:10.1242/dev.159889.
[220] Barreto S, Hamel L, Schiatti T, Yang Y, George V. Cardiac Progenitor Cells from Stem Cells: Learning from Genetics and Biomaterials. Cells 2019;8:1–55. doi:10.3390/cells8121536.
[221] Davis HE, Morgan JR, Yarmush ML. Polybrene increases retrovirus gene transfer efficiency by enhancing receptor-independent virus adsorption on target cell membranes. Biophys Chem 2002;97:159–172. doi:10.1016/S0301-4622(02)00057-1.
[222] Delrue I, Pan Q, Baczmanska AK, Callens BW, Verdoodt LLM. Determination of the Selection Capacity of Antibiotics for Gene Selection. Biotechnol J 2018;13:1–7. doi:10.1002/biot.201700747.
[223] Hasslund S, O'Keefe RJ, Awad HA. Proliferation Assays (BrdU and EdU) on Skeletal Tissue Sections. Methods Mol Biol 2014;1130:233–243. doi:10.1007/978-1-62703-989-5.
[224] Elmore S. Apoptosis: A review of programmed cell death. Toxicol Pathol 2007;35:495–516. doi:10.1016/j.pestbp.2011.02.012.Investigations.
[225] Wang P, Dreger M, Madrazo E, Williams CJ, Samaniego R, Hodson NW, et al. WDR5 modulates cell motility and morphology and controls nuclear changes induced by a 3D environment. Proc Natl Acad Sci U S A 2018;115:8581–8586. doi:10.1073/pnas.1719405115.
[226] Rangamani P, Lipshtat A, Azeloglu EU, Calizo RC, Hu M, Ghassemi S, et al. Decoding Information in Cell Shape. Cell 2013;154:1356. doi:10.1016/j.cell.2013.08.026.
[227] Chen B, Co C, Ho C-C. Cell Shape Dependent Regulation of Nuclear Morphology. Biomaterials 2016;67:129–136. doi:10.1016/j.physbeh.2017.03.040.
[228] Mukherjee A, Barai A, Singh RK, Yan W, Sen S. Nuclear Plasticity Increases Susceptibility to Damage During Con ned Migration 2020. doi:10.1101/2020/01.18.911529.
[229] Pantic I, Balcik Y. Correlation between nuclear envelope circularity and chromatin texture in spleen follicular cells Poster Communications : Correlation between nuclear envelope circularity and chromatin texture in spleen follicular cells 2016;2016.
[230] Webster M, Witkin KL, Cohen-Fix O. Sizing up the nucleus: nuclear shape, size and nuclear-envelope assembly. J Cell Sci 2009;122:1477–1486. doi:10.1242/jcs.037333.
[231] Koide A, Abbatiello S, Rothgery L, Koide S. Probing protein conformational changes in living cells by using designer binding proteins: Application to the estrogen receptor. Proc Natl Acad Sci U S A 2002;99:1253–1258. doi:10.1073/pnas.032665299.
[232] Buckingham M, Meilhac S, Zaffran S. Building the mammalian heart from two sources

of myocardial cells. Nat Rev Genet 2005;6:826–835. doi:10.1038/nrg1710.
[233] Basta J, Rauchman M. The Nucleosome Remodeling and Deacetylase Complex in Development and Disease. Transl Epigenetics to Clin 2017;165:37–72. doi:10.1016/B978-0-12-800802-7.00003-4.
[234] David R, Jarsch VB, Schwarz F, Nathan P, Gegg M, Lickert H, et al. Induction of MesP1 by Brachyury ( T ) generates the common multipotent cardiovascular stem cell. Cardiovasc Res 2011:115–122. doi:10.1093/cvr/cvr158.
[235] Le Guezennec X, Vermeulen M, Brinkman AB, Hoeijmakers WAM, Cohen A, Lasonder E, et al. MBD2/NuRD and MBD3/NuRD, Two Distinct Complexes with Different. Mol Cell Biol 2006;26:843–851. doi:10.1128/MCB.26.3.843.
[236] Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet 2012;13:227–232. doi:10.1038/nrg3185.
[237] Maier T, Güell M, Serrano L. Correlation of mRNA and protein in complex biological samples. FEBS Lett 2009;583:3966–3973. doi:10.1016/j.febslet.2009.10.036.
[238] Lee TI, Young RA. Transcriptional Regulation and Its Misregulation in Disease. Cell 2013;152:1237–1251. doi:10.1016/j.cell.2013.02.014.
[239] Swift J, Coruzzi GM. A matter of time — How transient transcription factor interactions create dynamic gene regulatory networks. Biochim Biophys Acta - Gene Regul Mech 2017;1860:75–83. doi:10.1016/j.bbagrm.2016.08.007.
[240] Kulkarni SS, Griffin JN, Date PP, Liem KF, Khokha MK. WDR5 Stabilizes Actin Architecture to Promote Multiciliated Cell Formation. Dev Cell 2018;46:595-610.e3. doi:10.1016/j.devcel.2018.08.009.
[241] Benchetrit H, Jaber M, Zayat V, Sebban S, Pushett A, Makedonski K, et al. Direct Induction of the Three Pre-implantation Blastocyst Cell Types from Fibroblasts. Cell Stem Cell 2019:1–12. doi:10.1016/j.stem.2019.03.018.
[242] Chen CMA, Kraut N, Groudine M, Weintraub H. I-mf, a novel myogenic repressor, interacts with members of the MyoD family. Cell 1996;86:731–741. doi:10.1016/S0092-8674(00)80148-8.
[243] Liu Y. Earlier and broader roles of Mesp1 in cardiovascular development. Cell Mol Life Sci 2017;74:1969–1983. doi:10.1007/s00018-016-2448-y.
[244] Sable R, Jambunathan N, Singh S, Pallerla S, Kousoulas KG, Jois S. Proximity ligation assay to study protein-protein interactions of proteins on two different cells. Biotechniques 2018;65:149–157. doi:10.2144/btn-2018-0049.
[245] Kaji K, Nichols J, Hendrich B. Mbd3, a component of the NuRD co-repressor complex, is reqiured for development of pluripotent cells. Development 2007;134:1123–1132. doi:10.1242/dev.02802.
[246] Shimbo T, Du Y, Grimm SA, Dhasarathy A, Mav D, Shah RR, et al. MBD3 Localizes at Promoters, Gene Bodies and Enhancers of Active Genes. PLoS Genet 2013;9:e1004028. doi:10.1371/journal.pgen.1004028.
[247] Wu SM. Mesp1 at the heart of mesoderm lineage specification. Cell Stem Cell 2008;3:1–2. doi:10.1016/j.stem.2008.06.017.

# 8 Appendix

## 8.1 Replicate one uninduced cells

| 31 elements included exclusively in Rep1 uninduced R1: | | | 176 common elements in Rep1 uninduced R1 and Rep1 uninduced R2: | | | 51 elements included exclusively in Rep1 uninduced R2: | | |
|---|---|---|---|---|---|---|---|---|
| Tubb5 | Tubulin beta-5 chain (Fragment): 7 kDa | 25 | Gm8994 | Uncharacterized protein: 47 kDa | 70 | Rasl2-9 | GTP-binding nuclear protein Ran, testis-specific isoform: 24 kDa | 9 |
| Tubb1 | Tubulin beta-1 chain: 50 kDa | 19 | Poldip3 | Polymerase delta-interacting protein 3: 46 kDa | 63 | Arpc4 | Actin-related protein 2/3 complex subunit 4: 20 kDa | 8 |
| Eno2 | Gamma-enolase: 47 kDa | 18 | Api5 | Apoptosis inhibitor 5: 57 kDa | 51 | Uhrf1 | E3 ubiquitin-protein ligase UHRF1: 88 kDa | 7 |
| Clint1 | Clathrin interactor 1: 69 kDa | 8 | Sec31a | Protein transport protein Sec31A: 134 kDa | 44 | Prpf4b | Serine/threonine-protein kinase PRP4 homolog: 117 kDa | 7 |
| Ccar1 | Cell division cycle and apoptosis regulator protein 1: 132 kDa | 6 | Sec31a | Protein transport protein Sec31A (Fragment): 101 kDa | 42 | Col20a1 | Collagen alpha-1(XX) chain: 141 kDa | 6 |
| Arid1a | AT-rich interactive domain-containing protein 1A: 242 kDa | 6 | Sf3b2 | Putative uncharacterized protein: 98 kDa | 36 | Lhx1 | LIM/homeobox protein Lhx1: 45 kDa | 6 |
| Abcf2 | ATP-binding cassette sub-family F member 2: 72 kDa | 6 | Spata5 | Spermatogenesis-associated protein 5: 97 kDa | 34 | Nudt21 | Cleavage and polyadenylation specificity factor subunit 5: 26 kDa | 6 |
| Psmd14 | 26S proteasome non-ATPase regulatory subunit 14: 35 kDa | 6 | Magoh | Protein mago nashi homolog: 17 kDa | 29 | Prpf31 | U4/U6 small nuclear ribonucleoprotein Prp31: 55 kDa | 6 |
| Etf1 | Eukaryotic peptide chain release factor subunit 1: 49 kDa | 6 | Sarnp | SAP domain-containing ribonucleoprotein: 24 kDa | 26 | Ncbp1 | Nuclear cap-binding protein subunit 1: 92 kDa | 6 |
| Gcn1 | eIF-2-alpha kinase activator GCN1: 293 kDa | 6 | Puf60 | Poly(U)-binding-splicing factor PUF60: 60 kDa | 26 | Cand2 | Cullin-associated NEDD8-dissociated protein 2: 136 kDa | 6 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Iqgap1 | Ras GTPase-activating-like protein IQGAP1: 189 kDa | 6 | Prtg | Protogenin: 131 kDa | 24 | Cbx5 | Chromobox protein homolog 5: 22 kDa | 6 |
| Mta3 | Metastasis-associated protein MTA3: 66 kDa | 6 | Chd4 | Chromodomain-helicase-DNA-binding protein 4: 218 kDa | 24 | Luc7l2 | Luc7l2 protein: 41 kDa | 6 |
| Mbd3 | Methyl-CpG-binding domain protein 3: 29 kDa | 5 | Acin1 | Apoptotic chromatin condensation inducer in the nucleus: 151 kDa | 24 | Cct7 | T-complex protein 1 subunit eta: 9 kDa | 6 |
| Ssrp1 | FACT complex subunit SSRP1: 81 kDa | 5 | Magohb | Mago nashi protein: 17 kDa | 24 | Eif4g2 | Eukaryotic translation initiation factor 4 gamma 2: 102 kDa | 6 |
| Ranbp2 | E3 SUMO-protein ligase RanBP2: 341 kDa | 5 | Acin1 | Apoptotic chromatin condensation inducer in the nucleus (Fragment): 144 kDa | 24 | Fxr2 | Fragile X mental retardation syndrome-related protein 2: 74 kDa | 6 |
| Cdc73 | Parafibromin: 61 kDa | 5 | Poldip3 | Polymerase delta-interacting protein 3 (Fragment): 22 kDa | 24 | Eif4g2 | Eukaryotic translation initiation factor 4 gamma 2: 102 kDa | 6 |
| Carm1 | Histone-arginine methyltransferase CARM1: 70 kDa | 5 | Srsf11 | Serine/arginine-rich-splicing factor 11 (Fragment): 56 kDa | 23 | Rbm26 | RNA-binding protein 26: 111 kDa | 6 |
| Rbm10 | RNA-binding protein 10: 103 kDa | 5 | Srsf11 | Putative uncharacterized protein: 53 kDa | 23 | Rbm26 | RNA-binding protein 26: 114 kDa | 6 |
| Aimp1 | Aminoacyl tRNA synthase complex-interacting multifunctional protein 1: 34 kDa | 5 | Chtop | Chromatin target of PRMT1 protein: 27 kDa | 22 | Asns | Asparagine synthetase [glutamine-hydrolyzing]: 64 kDa | 5 |
| Smchd1 | Structural maintenance of chromosomes flexible hinge domain-containing protein 1: 226 kDa | 5 | Pnn | Pinin: 82 kDa | 21 | Eomes | Eomesodermin homolog: 75 kDa | 5 |
| Snrpb2 | U2 small nuclear ribonucleoprotein B'': 25 kDa | 5 | Ddx46 | Probable ATP-dependent RNA helicase DDX46: 117 kDa | 20 | Srsf5 | Serine/arginine-rich splicing factor 5: 31 kDa | 5 |
| Pin1 | Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1: 18 kDa | 5 | Prmt5 | Protein arginine N-methyltransferase 5: 73 kDa | 20 | Sub1 | Activated RNA polymerase II transcriptional coactivator p15: 14 kDa | 5 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Smarcd1 | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 1: 58 kDa | 5 | Sf3a1 | Splicing factor 3A subunit 1: 89 kDa | 20 | Idh1 | Isocitrate dehydrogenase [NADP] cytoplasmic: 47 kDa | 5 |
| Srsf9 | Serine/arginine-rich splicing factor 9: 26 kDa | 5 | Rbm8a | RNA-binding protein 8A: 20 kDa | 20 | Hat1 | Histone acetyltransferase type B catalytic subunit: 49 kDa | 5 |
| Ptma | Prothymosin alpha: 12 kDa | 5 | Ctnna1 | Catenin alpha-1: 100 kDa | 20 | Cwc15 | Spliceosome-associated protein CWC15 homolog: 27 kDa | 5 |
| Fabp5 | Fatty acid-binding protein, epidermal: 15 kDa | 5 | Edc4 | Enhancer of mRNA-decapping protein 4: 142 kDa | 20 | Rfc4 | Replication factor C subunit 4: 40 kDa | 5 |
| Tfrc | Transferrin receptor protein 1: 86 kDa | 5 | Cdc5l | Cell division cycle 5-like protein: 92 kDa | 18 | Top2b | DNA topoisomerase 2-beta: 182 kDa | 5 |
| Atp6v1a | V-type proton ATPase catalytic subunit A: 68 kDa | 5 | Bclaf1 | Bcl-2-associated transcription factor 1: 106 kDa | 17 | Snrpd3 | Small nuclear ribonucleoprotein Sm D3: 14 kDa | 5 |
| Atp5j2 | ATP synthase subunit f, mitochondrial: 10 kDa | 5 | U2surp | U2 snRNP-associated SURP motif-containing protein: 118 kDa | 16 | Ncapd2 | Condensin complex subunit 1: 156 kDa | 5 |
| Gm5239 | MCG1031578: 14 kDa | 5 | Snrnp70 | U1 small nuclear ribonucleoprotein 70 kDa: 52 kDa | 16 | Psma5 | Proteasome subunit alpha type-5: 26 kDa | 5 |
| Dnaja1 | DnaJ homolog subfamily A member 1 (Fragment): 31 kDa | 5 | Msh6 | DNA mismatch repair protein Msh6: 151 kDa | 15 | Tcea1 | Transcription elongation factor A protein 1: 34 kDa | 5 |
| | | | Rnps1 | RNA-binding protein with serine-rich domain 1: 34 kDa | 15 | Mixl1 | Homeobox protein MIXL1: 25 kDa | 5 |
| | | | Prpf6 | Pre-mRNA-processing factor 6: 107 kDa | 15 | Ap2a1 | AP-2 complex subunit alpha-1: 108 kDa | 5 |
| | | | Sap18b | Histone deacetylase complex subunit SAP18: 20 kDa | 14 | Ubap2 | Ubiquitin-associated protein 2: 118 kDa | 5 |
| | | | Ddx42 | ATP-dependent RNA helicase DDX42: 102 kDa | 14 | U2af2 | Splicing factor U2AF 65 kDa subunit: 54 kDa | 5 |

| | | | | | |
|---|---|---|---|---|---|
| Zc3h11a | Zinc finger CCCH domain-containing protein 11A: 86 kDa | 14 | Txnl1 | Thioredoxin-like protein 1: 32 kDa | 5 |
| Nasp | Nuclear autoantigenic sperm protein: 84 kDa | 14 | Ppp4r3a | Serine/threonine-protein phosphatase 4 regulatory subunit 3A: 95 kDa | 5 |
| Srrt | Serrate RNA effector molecule homolog: 100 kDa | 14 | Arf3 | ADP-ribosylation factor 3: 21 kDa | 5 |
| Snw1 | SNW domain-containing protein 1: 61 kDa | 14 | Psmb2 | Proteasome subunit beta type-2: 23 kDa | 5 |
| Snw1 | SNW domain-containing protein 1: 61 kDa | 14 | Capza2 | F-actin-capping protein subunit alpha-2: 33 kDa | 5 |
| Sall4 | Sal-like protein 4: 113 kDa | 12 | Rab10 | Ras-related protein Rab-10: 23 kDa | 5 |
| Xpo1 | Exportin-1: 123 kDa | 12 | Capza1 | F-actin-capping protein subunit alpha-1: 33 kDa | 5 |
| Rbm39 | RNA-binding protein 39: 59 kDa | 12 | Gstp1 | Glutathione S-transferase P 1: 24 kDa | 5 |
| Pfas | Phosphoribosylformylglycinamidine synthase: 145 kDa | 12 | Lig1 | DNA ligase 1 (Fragment): 30 kDa | 5 |
| Sec24c | SEC24 related gene family, member C (S. cerevisiae), isoform CRA_a: 119 kDa | 12 | Arid3b | AT-rich interactive domain-containing protein 3B: 61 kDa | 5 |
| Rbm25 | RNA-binding protein 25: 100 kDa | 12 | Gtf2i | General transcription factor II-I: 108 kDa | 5 |
| Set | Protein SET: 33 kDa | 12 | Sars | Putative uncharacterized protein: 61 kDa | 5 |
| Sec23b | Protein transport protein Sec23B: 86 kDa | 12 | Pgam2 | Phosphoglycerate mutase 2: 29 kDa | 5 |
| L1td1 | LINE-1 type transposase domain-containing protein 1: 88 kDa | 12 | Sars | Serine--tRNA ligase, cytoplasmic: 58 kDa | 5 |
| Uncharacterized protein | Uncharacterized protein: 92 kDa | 12 | Gm20521 | Uncharacterized protein: 37 kDa | 5 |

| | | | | | |
|---|---|---|---|---|---|
| Zc3h14 | Zinc finger CCCH domain-containing protein 14: 82 kDa | 11 | Clta | Clathrin light chain: 25 kDa | 5 |
| Hdac1 | Histone deacetylase 1: 55 kDa | 11 | | | |
| Fkbp4 | Peptidylprolyl isomerase (Fragment): 30 kDa | 11 | | | |
| Ncbp3 | Nuclear cap-binding protein subunit 3: 70 kDa | 10 | | | |
| Ik | Protein Red: 66 kDa | 10 | | | |
| Smarca5 | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5: 122 kDa | 10 | | | |
| Smc2 | Structural maintenance of chromosomes protein 2: 134 kDa | 10 | | | |
| Snrpd1 | Small nuclear ribonucleoprotein Sm D1: 13 kDa | 10 | | | |
| Snrpa | U1 small nuclear ribonucleoprotein A: 32 kDa | 10 | | | |
| Wbp11 | WW domain-binding protein 11: 70 kDa | 10 | | | |
| Atxn2l | Ataxin-2-like protein: 111 kDa | 10 | | | |
| Srm | Spermidine synthase: 34 kDa | 10 | | | |
| U2af1 | Splicing factor U2AF 35 kDa subunit: 28 kDa | 10 | | | |
| Bub3 | Mitotic checkpoint protein BUB3: 37 kDa | 10 | | | |
| Anp32b | Acidic leucine-rich nuclear phosphoprotein 32 family member B: 31 kDa | 10 | | | |
| Snrpd2 | Small nuclear ribonucleoprotein Sm D2: 14 kDa | 10 | | | |
| Sf3a2 | Splicing factor 3A subunit 2: 51 kDa | 10 | | | |

| | | |
|---|---|---|
| Sssca1 | Sjoegren syndrome/scleroderma autoantigen 1 homolog: 21 kDa | 10 |
| Hdac2 | Histone deacetylase 2: 55 kDa | 10 |
| 2310022A10Rik | RIKEN cDNA 2310022A10 gene: 43 kDa | 10 |
| 2310022A10Rik | RIKEN cDNA 2310022A10 gene: 44 kDa | 10 |
| Cnot1 | CCR4-NOT transcription complex subunit 1: 267 kDa | 9 |
| Snrpa1 | U2 small nuclear ribonucleoprotein A': 28 kDa | 9 |
| Nop58 | Nucleolar protein 58: 60 kDa | 9 |
| Smu1 | WD40 repeat-containing protein SMU1: 58 kDa | 9 |
| Myef2 | Myelin expression factor 2: 63 kDa | 9 |
| Apex1 | DNA-(apurinic or apyrimidinic site) lyase: 35 kDa | 9 |
| Mta1 | Metastasis-associated protein MTA1: 79 kDa | 9 |
| Fyttd1 | UAP56-interacting factor: 36 kDa | 9 |
| Lig1 | DNA ligase 1: 102 kDa | 9 |
| Prpf40a | Pre-mRNA-processing factor 40 homolog A: 108 kDa | 9 |
| Prpf40a | Pre-mRNA-processing factor 40 homolog A: 106 kDa | 9 |
| Dnmt1 | DNA (cytosine-5)-methyltransferase 1: 183 kDa | 8 |
| Ctnnbl1 | Beta-catenin-like protein 1: 65 kDa | 8 |
| Prpf4 | U4/U6 small nuclear ribonucleoprotein Prp4: 58 kDa | 8 |
| Sec23ip | SEC23-interacting protein: 111 kDa | 8 |
| Luc7l3 | Luc7-like protein 3: 51 kDa | 8 |

| | | |
|---|---|---|
| Usp5 | Ubiquitin carboxyl-terminal hydrolase 5: 96 kDa | 8 |
| Snrnp40 | U5 small nuclear ribonucleoprotein 40 kDa protein: 39 kDa | 8 |
| Sf3b6 | Splicing factor 3B subunit 6: 15 kDa | 8 |
| Mfap1a | Microfibrillar-associated protein 1A: 52 kDa | 8 |
| Srsf10 | Serine/arginine-rich splicing factor 10: 31 kDa | 8 |
| Hmgcs1 | Hydroxymethylglutaryl-CoA synthase, cytoplasmic: 58 kDa | 8 |
| Tra2a | Transformer-2 protein homolog alpha: 32 kDa | 8 |
| Psmc1 | 26S proteasome regulatory subunit 4: 49 kDa | 8 |
| Actl6a | Actin-like protein 6A: 47 kDa | 8 |
| St13 | Hsc70-interacting protein: 42 kDa | 8 |
| Psmd6 | 26S proteasome non-ATPase regulatory subunit 6: 46 kDa | 8 |
| Ap1b1 | AP-1 complex subunit beta-1: 104 kDa | 8 |
| Pabpn1 | Polyadenylate-binding protein 2: 32 kDa | 8 |
| Luc7l2 | Putative RNA-binding protein Luc7-like 2: 47 kDa | 8 |
| Ubap2l | Ubiquitin-associated protein 2-like: 117 kDa | 8 |
| Clint1 | Clathrin interactor 1: 70 kDa | 8 |
| Snrpn | Small nuclear ribonucleoprotein-associated protein N: 25 kDa | 8 |
| Snrpb | Small nuclear ribonucleoprotein-associated protein B: 24 kDa | 8 |
| Sf3a2 | Splicing factor 3A subunit 2: 50 kDa | 8 |

| | | |
|---|---|---|
| Ubap2l | Ubiquitin associated protein 2-like, isoform CRA_g: 112 kDa | 8 |
| Wdhd1 | WD repeat and HMG-box DNA-binding protein 1: 120 kDa | 7 |
| Sec23a | Protein transport protein Sec23A: 86 kDa | 7 |
| Nsun2 | tRNA (cytosine(34)-C(5))-methyltransferase: 81 kDa | 7 |
| Ddx23 | DEAD (Asp-Glu-Ala-Asp) box polypeptide 23: 95 kDa | 7 |
| Smarce1 | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1: 47 kDa | 7 |
| Psmd1 | 26S proteasome non-ATPase regulatory subunit 1: 106 kDa | 7 |
| Dut | Deoxyuridine triphosphatase: 21 kDa | 7 |
| Glrx3 | Glutaredoxin-3: 38 kDa | 7 |
| Tomm34 | Mitochondrial import receptor subunit TOM34: 34 kDa | 7 |
| Ap2b1 | AP complex subunit beta: 101 kDa | 7 |
| Pum1 | Pumilio homolog 1: 100 kDa | 7 |
| Pum1 | Pumilio homolog 1: 127 kDa | 7 |
| Sssca1 | Sjoegren syndrome/scleroderma autoantigen 1 homolog: 10 kDa | 7 |
| Numa1 | Nuclear mitotic apparatus protein 1: 236 kDa | 6 |
| Crnkl1 | Crooked neck-like protein 1: 83 kDa | 6 |
| Ddb1 | DNA damage-binding protein 1: 127 kDa | 6 |

| | | |
|---|---|---|
| Cpsf6 | Cleavage and polyadenylation specificity factor subunit 6: 59 kDa | 6 |
| Hcfc1 | Host cell factor 1: 210 kDa | 6 |
| Prpf3 | U4/U6 small nuclear ribonucleoprotein Prp3: 77 kDa | 6 |
| Dis3 | Exosome complex exonuclease RRP44: 109 kDa | 6 |
| Plrg1 | Pleiotropic regulator 1: 57 kDa | 6 |
| C14orf166 | UPF0568 protein C14orf166 homolog: 28 kDa | 6 |
| Polr2a | DNA-directed RNA polymerase II subunit RPB1: 217 kDa | 6 |
| Bcas2 | Pre-mRNA-splicing factor SPF27: 26 kDa | 6 |
| Cherp | Calcium homeostasis endoplasmic reticulum protein: 108 kDa | 6 |
| Psip1 | PC4 and SFRS1-interacting protein: 60 kDa | 6 |
| Ctps1 | CTP synthase 1: 67 kDa | 6 |
| Rpa1 | Replication protein A 70 kDa DNA-binding subunit: 69 kDa | 6 |
| Snrpe | Small nuclear ribonucleoprotein E: 11 kDa | 6 |
| Sf3b4 | Splicing factor 3B subunit 4: 44 kDa | 6 |
| Sept7 | Septin-7: 51 kDa | 6 |
| Ctbp2 | C-terminal-binding protein 2: 49 kDa | 6 |
| Sart1 | U4/U6.U5 tri-snRNP-associated protein 1: 91 kDa | 6 |
| Psmb5 | Proteasome subunit beta type-5: 29 kDa | 6 |

| | | |
|---|---|---|
| Akr1b1 | Aldose reductase: 36 kDa | 6 |
| Ogt | UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit: 117 kDa | 6 |
| Cfdp1 | Craniofacial development protein 1: 33 kDa | 6 |
| Cbr1 | Carbonyl reductase [NADPH] 1: 31 kDa | 6 |
| Snrpf | Small nuclear ribonucleoprotein F: 10 kDa | 6 |
| Psmb1 | Proteasome subunit beta type-1: 26 kDa | 6 |
| Hmgn2 | Non-histone chromosomal protein HMG-17: 10 kDa | 6 |
| Srsf4 | Putative uncharacterized protein: 56 kDa | 6 |
| Luc7l | Putative RNA-binding protein Luc7-like 1: 44 kDa | 6 |
| Cdc37 | Hsp90 co-chaperone Cdc37: 45 kDa | 6 |
| Chd5 | Chromodomain-helicase-DNA-binding protein 5: 219 kDa | 6 |
| Psmd13 | 26S proteasome non-ATPase regulatory subunit 13: 43 kDa | 6 |
| Eif1a | Eukaryotic translation initiation factor 1A: 17 kDa | 6 |
| Sept11 | Septin-11: 50 kDa | 6 |
| Pklr | Pyruvate kinase: 59 kDa | 6 |
| Clta | Clathrin light chain: 23 kDa | 6 |
| Raly | RNA-binding protein Raly (Fragment): 23 kDa | 6 |
| Calml3 | Calmodulin-like protein 3: 17 kDa | 6 |
| Dst | Dystonin: 871 kDa | 6 |

| | | |
|---|---|---|
| Raly | RNA-binding protein Raly: 33 kDa | 6 |
| Dst | Dystonin: 836 kDa | 6 |
| Smc4 | Structural maintenance of chromosomes protein 4: 147 kDa | 5 |
| Pspc1 | Paraspeckle component 1: 59 kDa | 5 |
| Trrap | Transformation/transcription domain-associated protein: 437 kDa | 5 |
| Fkbp3 | Peptidyl-prolyl cis-trans isomerase FKBP3: 25 kDa | 5 |
| Psma2 | Proteasome subunit alpha type-2: 26 kDa | 5 |
| Psmb6 | Proteasome subunit beta type-6: 25 kDa | 5 |
| Dnmt3b | DNA (cytosine-5)-methyltransferase 3B: 97 kDa | 5 |
| Sec13 | Protein SEC13 homolog: 36 kDa | 5 |
| Psma7 | Proteasome subunit alpha type-7: 28 kDa | 5 |
| Sfn | 14-3-3 protein sigma: 28 kDa | 5 |
| Son | Protein SON: 269 kDa | 5 |
| Nap1l4 | Nucleosome assembly protein 1-like 4: 43 kDa | 5 |
| Son | Protein SON: 266 kDa | 5 |
| Son | Protein SON: 252 kDa | 5 |

## 8.2 Replicate one induced cells

| 52 elements included exclusively in Rep1 induced R1: | | | 203 common elements in Rep1 induced R1 and Rep1 induced R2: | | | 34 elements included exclusively in Rep1 induced R2: | | |
|---|---|---|---|---|---|---|---|---|
| Actb | Actin, cytoplasmic 1 (Fragment): 11 kDa | 20 | Poldip3 | Polymerase delta-interacting protein 3: 46 kDa | 74 | Krt85 | Keratin, type II cuticular Hb5: 56 kDa | 12 |
| Dhx9 | ATP-dependent RNA helicase A: 51 kDa | 17 | Gm8994 | Uncharacterized protein: 47 kDa | 70 | Cwc15 | Spliceosome-associated protein CWC15 homolog: 27 kDa | 7 |
| Eno2 | Gamma-enolase: 47 kDa | 12 | Sec31a | Protein transport protein Sec31A: 134 kDa | 61 | Srsf10 | Serine/arginine-rich splicing factor 10: 31 kDa | 7 |
| Fxr1 | Fragile X mental retardation syndrome-related protein 1: 51 kDa | 12 | Sec31a | Protein transport protein Sec31A (Fragment): 101 kDa | 56 | Gatad2b | Transcriptional repressor p66-beta: 65 kDa | 7 |
| Son | Protein SON: 252 kDa | 9 | Api5 | Apoptosis inhibitor 5: 57 kDa | 54 | Pdcd6 | Programmed cell death protein 6: 22 kDa | 7 |
| Glrx3 | Glutaredoxin-3: 38 kDa | 8 | Spata5 | Spermatogenesis-associated protein 5: 97 kDa | 54 | Ssbp3 | Single-stranded DNA-binding protein 3: 40 kDa | 7 |
| Tpr | Nuclear pore complex-associated intranuclear coiled-coil protein TPR: 267 kDa | 8 | Prtg | Protogenin: 131 kDa | 50 | Sssca1 | Sjoegren syndrome/scleroderma autoantigen 1 homolog: 10 kDa | 7 |
| Ap2b1 | AP complex subunit beta: 101 kDa | 7 | Sf3b2 | Putative uncharacterized protein: 98 kDa | 42 | Asns | Asparagine synthetase [glutamine-hydrolyzing]: 64 kDa | 6 |
| Tpr | Nucleoprotein TPR: 274 kDa | 7 | Chd4 | Chromodomain-helicase-DNA-binding protein 4: 218 kDa | 34 | Prpf4b | Serine/threonine-protein kinase PRP4 homolog: 117 kDa | 6 |
| Tlk2 | Serine/threonine-protein kinase tousled-like 2: 82 kDa | 7 | Acin1 | Apoptotic chromatin condensation inducer in the nucleus: 151 kDa | 32 | Gtf3c1 | General transcription factor 3C polypeptide 1: 237 kDa | 6 |
| Tlk2 | Serine/threonine-protein kinase tousled-like 2: 73 kDa | 7 | Acin1 | Apoptotic chromatin condensation inducer in the nucleus (Fragment): 144 kDa | 32 | Dhx38 | DEAH (Asp-Glu-Ala-His) box polypeptide 38: 141 kDa | 6 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Sub1 | Activated RNA polymerase II transcriptional coactivator p15: 14 kDa | 6 | Magoh | Protein mago nashi homolog: 17 kDa | 30 | Pqbp1 | Polyglutamine-binding protein 1: 31 kDa | 6 |
| Vcl | Vinculin: 117 kDa | 6 | Sarnp | SAP domain-containing ribonucleoprotein: 24 kDa | 28 | Smarca2 | Probable global transcription activator SNF2L2: 173 kDa | 6 |
| St13 | Hsc70-interacting protein: 42 kDa | 6 | Puf60 | Poly(U)-binding-splicing factor PUF60: 60 kDa | 28 | Clta | Clathrin light chain: 23 kDa | 6 |
| Dnajc7 | DnaJ homolog subfamily C member 7: 56 kDa | 6 | Zc3h14 | Zinc finger CCCH domain-containing protein 14: 82 kDa | 26 | Krt72 | Keratin, type II cytoskeletal 72: 57 kDa | 6 |
| Ap1b1 | AP-1 complex subunit beta-1: 104 kDa | 6 | Sall4 | Sal-like protein 4: 113 kDa | 26 | Pcnp | PEST proteolytic signal-containing nuclear protein: 19 kDa | 6 |
| Ptma | Prothymosin alpha: 12 kDa | 6 | Prph | Peripherin: 58 kDa | 26 | Myh11 | Myosin-11: 227 kDa | 6 |
| Otub1 | Ubiquitin thioesterase OTUB1: 28 kDa | 6 | Magohb | Mago nashi protein: 17 kDa | 26 | Tcf3 | Transcription factor E2-alpha: 68 kDa | 6 |
| Rabep1 | Rab GTPase-binding effector protein 1: 100 kDa | 6 | Poldip3 | Polymerase delta-interacting protein 3 (Fragment): 22 kDa | 26 | Myh11 | Myosin-11: 223 kDa | 6 |
| Dmap1 | DNA methyltransferase 1-associated protein 1: 53 kDa | 6 | Ddx46 | Probable ATP-dependent RNA helicase DDX46: 117 kDa | 25 | Thoc2 | THO complex subunit 2: 183 kDa | 5 |
| Sall1 | Sal-like 1 (Drosophila): 140 kDa | 6 | Srsf11 | Serine/arginine-rich-splicing factor 11 (Fragment): 56 kDa | 24 | Xrn2 | 5'-3' exoribonuclease 2: 109 kDa | 5 |
| Rbm6 | Putative uncharacterized protein: 114 kDa | 6 | Tubb5 | Tubulin beta-5 chain (Fragment): 7 kDa | 24 | Skiv2l2 | Superkiller viralicidic activity 2-like 2: 118 kDa | 5 |
| Rbm6 | RNA-binding motif protein 6: 128 kDa | 6 | Srsf11 | Putative uncharacterized protein: 53 kDa | 24 | Ctbp2 | C-terminal-binding protein 2: 49 kDa | 5 |
| Smc3 | Structural maintenance of chromosomes protein 3: 142 kDa | 5 | Chtop | Chromatin target of PRMT1 protein: 27 kDa | 22 | Ythdc1 | YTH domain-containing protein 1: 86 kDa | 5 |
| Psmd1 | 26S proteasome non-ATPase regulatory subunit 1: 106 kDa | 5 | Bclaf1 | Bcl-2-associated transcription factor 1: 106 kDa | 22 | Ncapd2 | Condensin complex subunit 1: 156 kDa | 5 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Ewsr1 | RNA-binding protein EWS: 68 kDa | 5 | Prmt5 | Protein arginine N-methyltransferase 5: 73 kDa | 22 | Xab2 | Pre-mRNA-splicing factor SYF1: 100 kDa | 5 |
| Psmd14 | 26S proteasome non-ATPase regulatory subunit 14: 35 kDa | 5 | Sf3a1 | Splicing factor 3A subunit 1: 89 kDa | 22 | Tomm34 | Mitochondrial import receptor subunit TOM34: 34 kDa | 5 |
| Ppig | Peptidyl-prolyl cis-trans isomerase G: 88 kDa | 5 | Pnn | Pinin: 82 kDa | 21 | Zc3h18 | RIKEN cDNA 5830416A07, isoform CRA_c: 108 kDa | 5 |
| Uba2 | SUMO-activating enzyme subunit 2: 71 kDa | 5 | U2surp | U2 snRNP-associated SURP motif-containing protein: 118 kDa | 21 | Taf15 | Putative uncharacterized protein: 59 kDa | 5 |
| Fbl | rRNA 2'-O-methyltransferase fibrillarin: 34 kDa | 5 | Tubb1 | Tubulin beta-1 chain: 50 kDa | 21 | Brd4 | Bromodomain-containing protein 4: 156 kDa | 5 |
| Ncbp1 | Nuclear cap-binding protein subunit 1: 92 kDa | 5 | Rbm8a | RNA-binding protein 8A: 20 kDa | 20 | Fubp1 | Far upstream element-binding protein 1 (Fragment): 15 kDa | 5 |
| SUMO1 | Small ubiquitin-related modifier 1: 12 kDa | 5 | Zc3h11a | Zinc finger CCCH domain-containing protein 11A: 86 kDa | 20 | Sumo3 | Small ubiquitin-related modifier 3: 12 kDa | 5 |
| Rhoa | Transforming protein RhoA: 22 kDa | 5 | Sap18b | Histone deacetylase complex subunit SAP18: 20 kDa | 18 | Tcf12 | Transcription factor 12: 76 kDa | 5 |
| Fabp5 | Fatty acid-binding protein, epidermal: 15 kDa | 5 | Snrnp70 | U1 small nuclear ribonucleoprotein 70 kDa: 52 kDa | 18 | Chd3 | Chromodomain helicase DNA-binding protein 3: 233 kDa | 5 |
| Fip1l1 | Pre-mRNA 3'-end-processing factor FIP1: 65 kDa | 5 | Sec23b | Protein transport protein Sec23B: 86 kDa | 18 | | | |
| Casc3 | Protein CASC3: 76 kDa | 5 | Eomes | Eomesodermin homolog: 75 kDa | 18 | | | |
| Wdr77 | Methylosome protein 50: 37 kDa | 5 | Cdc5l | Cell division cycle 5-like protein: 92 kDa | 18 | | | |
| Rps27l | 40S ribosomal protein S27-like: 9 kDa | 5 | Edc4 | Enhancer of mRNA-decapping protein 4: 142 kDa | 18 | | | |
| Eif1a | Eukaryotic translation initiation factor 1A: 17 kDa | 5 | Ddx42 | ATP-dependent RNA helicase DDX42: 102 kDa | 17 | | | |
| Capza2 | F-actin-capping protein subunit alpha-2: 33 kDa | 5 | Sec23a | Protein transport protein Sec23A: 86 kDa | 17 | | | |

| | | |
|---|---|---|
| Gstp1 | Glutathione S-transferase P 1: 24 kDa | 5 |
| Cstf2 | Cleavage stimulation factor subunit 2: 61 kDa | 5 |
| Lig1 | DNA ligase 1 (Fragment): 30 kDa | 5 |
| Pabpc5 | Poly A binding protein, cytoplasmic 5: 43 kDa | 5 |
| Pdlim5 | PDZ and LIM domain protein 5: 63 kDa | 5 |
| Ptbp3 | Polypyrimidine tract-binding protein 3: 57 kDa | 5 |
| Gm14214 | Uncharacterized protein (Fragment): 25 kDa | 5 |
| Gtf2i | General transcription factor II-I: 108 kDa | 5 |
| Fxr2 | Fragile X mental retardation syndrome-related protein 2: 74 kDa | 5 |
| Pgam2 | Phosphoglycerate mutase 2: 29 kDa | 5 |
| Krt16 | Keratin, type I cytoskeletal 16: 52 kDa | 5 |
| Gtf2i | General transcription factor II-I: 112 kDa | 5 |
| Rnps1 | RNA-binding protein with serine-rich domain 1: 34 kDa | 16 |
| Srrt | Serrate RNA effector molecule homolog: 100 kDa | 16 |
| Mta1 | Metastasis-associated protein MTA1: 79 kDa | 16 |
| Prpf6 | Pre-mRNA-processing factor 6: 107 kDa | 15 |
| Sec24c | SEC24 related gene family, member C (S. cerevisiae), isoform CRA_a: 119 kDa | 15 |
| Wbp11 | WW domain-binding protein 11: 70 kDa | 15 |
| Hdac2 | Histone deacetylase 2: 55 kDa | 15 |
| Fubp1 | Far upstream element-binding protein 1 (Fragment): 32 kDa | 15 |
| Nasp | Nuclear autoantigenic sperm protein: 84 kDa | 14 |
| Ncbp3 | Nuclear cap-binding protein subunit 3: 70 kDa | 14 |
| Rbm39 | RNA-binding protein 39: 59 kDa | 14 |
| Ctnna1 | Catenin alpha-1: 100 kDa | 14 |
| Snrpa | U1 small nuclear ribonucleoprotein A: 32 kDa | 14 |
| Bub3 | Mitotic checkpoint protein BUB3: 37 kDa | 14 |
| Hdac1 | Histone deacetylase 1: 55 kDa | 14 |
| 2310022A10Rik | RIKEN cDNA 2310022A10 gene: 44 kDa | 14 |

| | | |
|---|---|---|
| Snw1 | SNW domain-containing protein 1: 61 kDa | 14 |
| Snw1 | SNW domain-containing protein 1: 61 kDa | 14 |
| Fyttd1 | UAP56-interacting factor: 36 kDa | 13 |
| Xpo1 | Exportin-1: 123 kDa | 12 |
| Pfas | Phosphoribosylformylglycinamidine synthase: 145 kDa | 12 |
| Arid1a | AT-rich interactive domain-containing protein 1A: 242 kDa | 12 |
| Prpf4 | U4/U6 small nuclear ribonucleoprotein Prp4: 58 kDa | 12 |
| U2af1 | Splicing factor U2AF 35 kDa subunit: 28 kDa | 12 |
| Smarce1 | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1: 47 kDa | 12 |
| Snrpd2 | Small nuclear ribonucleoprotein Sm D2: 14 kDa | 12 |
| Actl6a | Actin-like protein 6A: 47 kDa | 12 |
| Uncharacterized protein | Uncharacterized protein: 92 kDa | 12 |
| 2310022A10Rik | RIKEN cDNA 2310022A10 gene: 43 kDa | 12 |
| Rbm25 | RNA-binding protein 25: 100 kDa | 11 |
| Set | Protein SET: 33 kDa | 11 |
| Mbd3 | Methyl-CpG-binding domain protein 3: 29 kDa | 11 |
| Pabpn1 | Polyadenylate-binding protein 2: 32 kDa | 11 |
| Rbm26 | RNA-binding protein 26: 111 kDa | 11 |

| | | |
|---|---|---|
| Rbm26 | RNA-binding protein 26: 114 kDa | 11 |
| Msh6 | DNA mismatch repair protein Msh6: 151 kDa | 10 |
| Ik | Protein Red: 66 kDa | 10 |
| Smarca5 | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5: 122 kDa | 10 |
| Snrpa1 | U2 small nuclear ribonucleoprotein A': 28 kDa | 10 |
| Snrpd1 | Small nuclear ribonucleoprotein Sm D1: 13 kDa | 10 |
| Ctnnbl1 | Beta-catenin-like protein 1: 65 kDa | 10 |
| Nop58 | Nucleolar protein 58: 60 kDa | 10 |
| Smu1 | WD40 repeat-containing protein SMU1: 58 kDa | 10 |
| Sec23ip | SEC23-interacting protein: 111 kDa | 10 |
| Luc7l3 | Luc7-like protein 3: 51 kDa | 10 |
| Srm | Spermidine synthase: 34 kDa | 10 |
| Sf3b6 | Splicing factor 3B subunit 6: 15 kDa | 10 |
| Cherp | Calcium homeostasis endoplasmic reticulum protein: 108 kDa | 10 |
| Trrap | Transformation/transcription domain-associated protein: 437 kDa | 10 |
| Tra2a | Transformer-2 protein homolog alpha: 32 kDa | 10 |
| Hmgn2 | Non-histone chromosomal protein HMG-17: 10 kDa | 10 |
| Sf3a2 | Splicing factor 3A subunit 2: 51 kDa | 10 |
| smarcc2 | SWI/SNF complex subunit SMARCC2: 133 kDa | 10 |

| | | |
|---|---|---|
| Arid3b | AT-rich interactive domain-containing protein 3B: 61 kDa | 10 |
| Snrpn | Small nuclear ribonucleoprotein-associated protein N: 25 kDa | 10 |
| Cnot1 | CCR4-NOT transcription complex subunit 1: 267 kDa | 9 |
| Numa1 | Nuclear mitotic apparatus protein 1: 236 kDa | 9 |
| Myef2 | Myelin expression factor 2: 63 kDa | 9 |
| Anp32b | Acidic leucine-rich nuclear phosphoprotein 32 family member B: 31 kDa | 9 |
| Bcas2 | Pre-mRNA-splicing factor SPF27: 26 kDa | 9 |
| Lhx1 | LIM/homeobox protein Lhx1: 45 kDa | 9 |
| Dbt | Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial: 53 kDa | 9 |
| Smarcd1 | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 1: 58 kDa | 9 |
| Pum1 | Pumilio homolog 1: 100 kDa | 9 |
| Sf3a2 | Splicing factor 3A subunit 2: 50 kDa | 9 |
| Dnmt1 | DNA (cytosine-5)-methyltransferase 1: 183 kDa | 8 |
| Smc2 | Structural maintenance of chromosomes protein 2: 134 kDa | 8 |
| Wdhd1 | WD repeat and HMG-box DNA-binding protein 1: 120 kDa | 8 |
| Atxn2l | Ataxin-2-like protein: 111 kDa | 8 |

| | | |
|---|---|---|
| Ddx23 | DEAD (Asp-Glu-Ala-Asp) box polypeptide 23: 95 kDa | 8 |
| Cpsf6 | Cleavage and polyadenylation specificity factor subunit 6: 59 kDa | 8 |
| Apex1 | DNA-(apurinic or apyrimidinic site) lyase: 35 kDa | 8 |
| Plrg1 | Pleiotropic regulator 1: 57 kDa | 8 |
| UPF0568 protein C14orf166 homolog | UPF0568 protein C14orf166 homolog: 28 kDa | 8 |
| Snrnp40 | U5 small nuclear ribonucleoprotein 40 kDa protein: 39 kDa | 8 |
| Mfap1a | Microfibrillar-associated protein 1A: 52 kDa | 8 |
| Psmc1 | 26S proteasome regulatory subunit 4: 49 kDa | 8 |
| Snrpb2 | U2 small nuclear ribonucleoprotein B'': 25 kDa | 8 |
| Gatad2a | Transcriptional repressor p66 alpha: 67 kDa | 8 |
| Srsf9 | Serine/arginine-rich splicing factor 9: 26 kDa | 8 |
| Mesp1 | posterior protein 1: 28 kDa | 8 |
| Chd5 | Chromodomain-helicase-DNA-binding protein 5: 219 kDa | 8 |
| Sssca1 | Sjoegren syndrome/scleroderma autoantigen 1 homolog: 21 kDa | 8 |
| Mta3 | Metastasis-associated protein MTA3: 66 kDa | 8 |
| Luc7l2 | Putative RNA-binding protein Luc7-like 2: 47 kDa | 8 |
| Ubap2l | Ubiquitin-associated protein 2-like: 117 kDa | 8 |

| | | |
|---|---|---|
| Prpf40a | Pre-mRNA-processing factor 40 homolog A: 108 kDa | 8 |
| Pum1 | Pumilio homolog 1: 127 kDa | 8 |
| Son | Protein SON: 269 kDa | 8 |
| Snrpb | Small nuclear ribonucleoprotein-associated protein B: 24 kDa | 8 |
| Son | Protein SON: 266 kDa | 8 |
| Prpf40a | Pre-mRNA-processing factor 40 homolog A: 106 kDa | 8 |
| Ccar1 | Cell division cycle and apoptosis regulator protein 1: 132 kDa | 7 |
| Crnkl1 | Crooked neck-like protein 1: 83 kDa | 7 |
| Smc4 | Structural maintenance of chromosomes protein 4: 147 kDa | 7 |
| Col20a1 | Collagen alpha-1(XX) chain: 141 kDa | 7 |
| Prpf3 | U4/U6 small nuclear ribonucleoprotein Prp3: 77 kDa | 7 |
| Dis3 | Exosome complex exonuclease RRP44: 109 kDa | 7 |
| Pspc1 | Paraspeckle component 1: 59 kDa | 7 |
| Sart1 | U4/U6.U5 tri-snRNP-associated protein 1: 91 kDa | 7 |
| Cfdp1 | Craniofacial development protein 1: 33 kDa | 7 |
| Rbm17 | Splicing factor 45: 45 kDa | 7 |
| Ogdhl | Oxoglutarate dehydrogenase-like: 117 kDa | 7 |
| Ubap2l | Ubiquitin associated protein 2-like, isoform CRA_g: 112 kDa | 7 |
| Nsun2 | tRNA (cytosine(34)-C(5))-methyltransferase: 81 kDa | 6 |

| | | |
|---|---|---|
| Ddb1 | DNA damage-binding protein 1: 127 kDa | 6 |
| Usp5 | Ubiquitin carboxyl-terminal hydrolase 5: 96 kDa | 6 |
| Srsf5 | Serine/arginine-rich splicing factor 5: 31 kDa | 6 |
| Hcfc1 | Host cell factor 1: 210 kDa | 6 |
| Ssrp1 | FACT complex subunit SSRP1: 81 kDa | 6 |
| Carm1 | Histone-arginine methyltransferase CARM1: 70 kDa | 6 |
| Dut | Deoxyuridine triphosphatase: 21 kDa | 6 |
| Rpa1 | Replication protein A 70 kDa DNA-binding subunit: 69 kDa | 6 |
| Snrpe | Small nuclear ribonucleoprotein E: 11 kDa | 6 |
| Supt5h | Transcription elongation factor SPT5: 121 kDa | 6 |
| Nudt21 | Cleavage and polyadenylation specificity factor subunit 5: 26 kDa | 6 |
| Rfc4 | Replication factor C subunit 4: 40 kDa | 6 |
| Sf3b4 | Splicing factor 3B subunit 4: 44 kDa | 6 |
| Bckdha | 2-oxoisovalerate dehydrogenase subunit alpha, mitochondrial: 50 kDa | 6 |
| Sec24b | Sec24 related gene family, member B (S. cerevisiae): 136 kDa | 6 |
| Sept7 | Septin-7: 51 kDa | 6 |
| Akr1b1 | Aldose reductase: 36 kDa | 6 |
| Vrtn | Vertnin: 83 kDa | 6 |

| | | |
|---|---|---|
| Psmb6 | Proteasome subunit beta type-6: 25 kDa | 6 |
| Hnrnpll | Heterogeneous nuclear ribonucleoprotein L-like: 64 kDa | 6 |
| Ogt | UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit: 117 kDa | 6 |
| Snrpf | Small nuclear ribonucleoprotein F: 10 kDa | 6 |
| Cand2 | Cullin-associated NEDD8-dissociated protein 2: 136 kDa | 6 |
| Bud31 | Protein BUD31 homolog: 17 kDa | 6 |
| Lig1 | DNA ligase 1: 102 kDa | 6 |
| Cbx5 | Chromobox protein homolog 5: 22 kDa | 6 |
| Srsf4 | Putative uncharacterized protein: 56 kDa | 6 |
| Khdrbs1 | KH domain-containing, RNA-binding, signal transduction-associated protein 1: 48 kDa | 6 |
| Luc7l | Putative RNA-binding protein Luc7-like 1: 44 kDa | 6 |
| Wdr5 | WD repeat-containing protein 5: 37 kDa | 6 |
| Ctps2 | CTP synthase 2: 66 kDa | 6 |
| Amot | Angiomotin: 121 kDa | 6 |
| KRT84 | Keratin, type II cuticular Hb4: 65 kDa | 6 |
| Clint1 | Clathrin interactor 1: 70 kDa | 6 |
| Ep400 | E1A-binding protein p400 (Fragment): 35 kDa | 6 |
| Raly | RNA-binding protein Raly (Fragment): 23 kDa | 6 |

| | | |
|---|---|---|
| Luc7l2 | Luc7l2 protein: 41 kDa | 6 |
| Dst | Dystonin: 871 kDa | 6 |
| Raly | RNA-binding protein Raly: 33 kDa | 6 |
| Gm20521 | Uncharacterized protein: 37 kDa | 6 |
| Ldb1 | LIM domain-binding protein 1: 47 kDa | 6 |
| Dst | Dystonin: 836 kDa | 6 |
| Chd3 | Chromodomain helicase DNA-binding protein 3 (Fragment): 214 kDa | 6 |
| Ldb1 | LIM domain-binding protein 1: 43 kDa | 6 |
| Polr2a | DNA-directed RNA polymerase II subunit RPB1: 217 kDa | 5 |
| Uhrf1 | E3 ubiquitin-protein ligase UHRF1: 88 kDa | 5 |
| Hmgcs1 | Hydroxymethylglutaryl-CoA synthase, cytoplasmic: 58 kDa | 5 |
| Top2b | DNA topoisomerase 2-beta: 182 kDa | 5 |
| Prpf31 | U4/U6 small nuclear ribonucleoprotein Prp31: 55 kDa | 5 |
| Pin1 | Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1: 18 kDa | 5 |
| Trim33 | E3 ubiquitin-protein ligase TRIM33: 124 kDa | 5 |
| Sf3b5 | Splicing factor 3B subunit 5: 10 kDa | 5 |
| Fam98b | Protein FAM98B: 45 kDa | 5 |
| U2af2 | Splicing factor U2AF 65 kDa subunit: 54 kDa | 5 |
| Ep400 | E1A-binding protein p400: 337 kDa | 5 |
| Crebbp | CREB-binding protein: 265 kDa | 5 |

| | | |
|---|---|---|
| Pklr | Pyruvate kinase: 59 kDa | 5 |
| Dnajc8 | DnaJ homolog subfamily C member 8: 30 kDa | 5 |

## 8.3 Replicate two uninduced cells

| 29 elements included exclusively in Rep2 uninduced R1: | | | 195 common elements in Rep2 uninduced R1 and Rep2 uninduced R2: | | | 58 elements included exclusively in Rep2 uninduced R2: | | |
|---|---|---|---|---|---|---|---|---|
| Hnrnpk | Heterogeneous nuclear ribonucleoprotein K (Fragment): 20 kDa | 34 | Poldip3 | Polymerase delta-interacting protein 3: 46 kDa | 87 | Eno2 | Gamma-enolase: 47 kDa | 13 |
| Tubb5 | Tubulin beta-5 chain (Fragment): 7 kDa | 23 | Gm8994 | Uncharacterized protein: 47 kDa | 73 | Prph | Peripherin: 58 kDa | 13 |
| Hnrnpd | Heterogeneous nuclear ribonucleoprotein D0 (Fragment): 25 kDa | 16 | Sec31a | Protein transport protein Sec31A: 134 kDa | 63 | Supt5h | Transcription elongation factor SPT5: 121 kDa | 9 |
| Gtf2i | General transcription factor II-I: 108 kDa | 9 | Api5 | Apoptosis inhibitor 5: 57 kDa | 61 | Gm17087 | Uncharacterized protein: 18 kDa | 8 |
| Gtf2i | General transcription factor II-I: 112 kDa | 8 | Sec31a | Protein transport protein Sec31A (Fragment): 101 kDa | 58 | Ctps1 | CTP synthase 1: 67 kDa | 7 |
| Safb | Scaffold attachment factor B1: 105 kDa | 7 | Spata5 | Spermatogenesis-associated protein 5: 97 kDa | 54 | Ncbp1 | Nuclear cap-binding protein subunit 1: 92 kDa | 7 |
| Uhrf1 | E3 ubiquitin-protein ligase UHRF1: 88 kDa | 7 | Sf3b2 | Putative uncharacterized protein: 98 kDa | 46 | Ap1g1 | AP-1 complex subunit gamma-1: 91 kDa | 7 |
| Raly | RNA-binding protein Raly (Fragment): 23 kDa | 7 | Magoh | Protein mago nashi homolog: 17 kDa | 39 | Lig1 | DNA ligase 1 (Fragment): 30 kDa | 7 |
| Luc7l2 | Luc7l2 protein: 41 kDa | 7 | Prtg | Protogenin: 131 kDa | 35 | Krt90 | Keratin 90: 58 kDa | 7 |
| Raly | RNA-binding protein Raly: 33 kDa | 7 | Acin1 | Apoptotic chromatin condensation inducer in the nucleus: 151 kDa | 35 | Son | Protein SON: 269 kDa | 7 |
| Vcl | Vinculin: 117 kDa | 6 | Acin1 | Apoptotic chromatin condensation inducer in the nucleus (Fragment): 144 kDa | 35 | Son | Protein SON: 266 kDa | 7 |
| Sf3b5 | Splicing factor 3B subunit 5: 10 kDa | 6 | Magohb | Mago nashi protein: 17 kDa | 34 | Mdn1 | Midasin: 630 kDa | 6 |
| Kdm1a | Lysine-specific histone demethylase 1A: 93 kDa | 6 | Poldip3 | Polymerase delta-interacting protein 3 (Fragment): 22 kDa | 33 | Psip1 | PC4 and SFRS1-interacting protein: 60 kDa | 6 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Rabep1 | Rab GTPase-binding effector protein 1: 100 kDa | 6 | Sarnp | SAP domain-containing ribonucleoprotein: 24 kDa | 30 | Rbm10 | RNA-binding protein 10: 103 kDa | 6 |
| Krt85 | Keratin, type II cuticular Hb5: 56 kDa | 6 | Puf60 | Poly(U)-binding-splicing factor PUF60: 60 kDa | 29 | Sart1 | U4/U6.U5 tri-snRNP-associated protein 1: 91 kDa | 6 |
| Gm20521 | Uncharacterized protein: 37 kDa | 6 | Chd4 | Chromodomain-helicase-DNA-binding protein 4: 218 kDa | 27 | Uba2 | SUMO-activating enzyme subunit 2: 71 kDa | 6 |
| Tpr | Nucleoprotein TPR: 274 kDa | 6 | Chtop | Chromatin target of PRMT1 protein: 27 kDa | 26 | Sec13 | Protein SEC13 homolog: 36 kDa | 6 |
| Asns | Asparagine synthetase [glutamine-hydrolyzing]: 64 kDa | 5 | Ddx46 | Probable ATP-dependent RNA helicase DDX46: 117 kDa | 26 | Smarcd1 | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 1: 58 kDa | 6 |
| Hmgcs1 | Hydroxymethylglutaryl-CoA synthase, cytoplasmic: 58 kDa | 5 | Cdc5l | Cell division cycle 5-like protein: 92 kDa | 26 | Trim33 | E3 ubiquitin-protein ligase TRIM33: 124 kDa | 6 |
| Gtf3c1 | General transcription factor 3C polypeptide 1: 237 kDa | 5 | Srsf11 | Serine/arginine-rich-splicing factor 11 (Fragment): 56 kDa | 26 | Ap2m1 | AP-2 complex subunit mu: 50 kDa | 6 |
| SUMO1 | Small ubiquitin-related modifier 1: 12 kDa | 5 | Srsf11 | Putative uncharacterized protein: 53 kDa | 26 | Ptma | Prothymosin alpha: 12 kDa | 6 |
| Rhoa | Transforming protein RhoA: 22 kDa | 5 | Pnn | Pinin: 82 kDa | 25 | Ogdhl | Oxoglutarate dehydrogenase-like: 117 kDa | 6 |
| Ipo4 | Importin-4: 119 kDa | 5 | Rbm8a | RNA-binding protein 8A: 20 kDa | 24 | Fxr2 | Fragile X mental retardation syndrome-related protein 2: 74 kDa | 6 |
| Ppp4r3a | Serine/threonine-protein phosphatase 4 regulatory subunit 3A: 95 kDa | 5 | U2surp | U2 snRNP-associated SURP motif-containing protein: 118 kDa | 23 | 2310022A10Rik | RIKEN cDNA 2310022A10 gene (Fragment): 21 kDa | 6 |
| Edc3 | Enhancer of mRNA-decapping protein 3: 56 kDa | 5 | Bclaf1 | Bcl-2-associated transcription factor 1: 106 kDa | 22 | Son | Protein SON: 252 kDa | 6 |
| Ctps2 | CTP synthase 2: 66 kDa | 5 | Prmt5 | Protein arginine N-methyltransferase 5: 73 kDa | 22 | Sub1 | Activated RNA polymerase II transcriptional coactivator p15: 14 kDa | 5 |
| Psma7 | Proteasome subunit alpha type-7: 28 kDa | 5 | Zc3h14 | Zinc finger CCCH domain-containing protein 14: 82 kDa | 22 | Larp1 | La-related protein 1: 121 kDa | 5 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Smarcc2 | SWI/SNF complex subunit SMARCC2: 133 kDa | 5 | Sall4 | Sal-like protein 4: 113 kDa | 22 | Fkbp3 | Peptidyl-prolyl cis-trans isomerase FKBP3: 25 kDa | 5 |
| Fubp1 | Far upstream element-binding protein 1 (Fragment): 15 kDa | 5 | Sec23b | Protein transport protein Sec23B: 86 kDa | 22 | Sept7 | Septin-7: 51 kDa | 5 |
| | | | Zc3h11a | Zinc finger CCCH domain-containing protein 11A: 86 kDa | 21 | Ythdc1 | YTH domain-containing protein 1: 86 kDa | 5 |
| | | | Sap18b | Histone deacetylase complex subunit SAP18: 20 kDa | 20 | Top2b | DNA topoisomerase 2-beta: 182 kDa | 5 |
| | | | Sf3a1 | Splicing factor 3A subunit 1: 89 kDa | 20 | Psmb5 | Proteasome subunit beta type-5: 29 kDa | 5 |
| | | | Rnps1 | RNA-binding protein with serine-rich domain 1: 34 kDa | 20 | St13 | Hsc70-interacting protein: 42 kDa | 5 |
| | | | Edc4 | Enhancer of mRNA-decapping protein 4: 142 kDa | 20 | Cnot2 | CCR4-NOT transcription complex subunit 2: 60 kDa | 5 |
| | | | Snrnp70 | U1 small nuclear ribonucleoprotein 70 kDa: 52 kDa | 18 | Ppig | Peptidyl-prolyl cis-trans isomerase G: 88 kDa | 5 |
| | | | Wbp11 | WW domain-binding protein 11: 70 kDa | 18 | Vrtn | Vertnin: 83 kDa | 5 |
| | | | Uncharacterized protein | Uncharacterized protein: 92 kDa | 18 | Fbl | rRNA 2'-O-methyltransferase fibrillarin: 34 kDa | 5 |
| | | | Snw1 | SNW domain-containing protein 1: 61 kDa | 18 | Psma2 | Proteasome subunit alpha type-2: 26 kDa | 5 |
| | | | Snw1 | SNW domain-containing protein 1: 61 kDa | 18 | Hnrnpll | Heterogeneous nuclear ribonucleoprotein L-like: 64 kDa | 5 |
| | | | Ddx42 | ATP-dependent RNA helicase DDX42: 102 kDa | 17 | Cbx5 | Chromobox protein homolog 5: 22 kDa | 5 |
| | | | Tubb1 | Tubulin beta-1 chain: 50 kDa | 17 | Ranbp3 | Ran-binding protein 3: 53 kDa | 5 |
| | | | Prpf6 | Pre-mRNA-processing factor 6: 107 kDa | 16 | Usp7 | Ubiquitin carboxyl-terminal hydrolase 7: 128 kDa | 5 |
| | | | Ncbp3 | Nuclear cap-binding protein subunit 3: 70 kDa | 16 | Dpy30 | Protein dpy-30 homolog: 11 kDa | 5 |

| | | | | | |
|---|---|---|---|---|---|
| Cnot1 | CCR4-NOT transcription complex subunit 1: 267 kDa | 16 | Znhit6 | Box C/D snoRNA protein 1: 52 kDa | 5 |
| Sec24c | SEC24 related gene family, member C (S. cerevisiae), isoform CRA_a: 119 kDa | 16 | Ppp1ca | Serine/threonine-protein phosphatase PP1-alpha catalytic subunit: 38 kDa | 5 |
| Srrt | Serrate RNA effector molecule homolog: 100 kDa | 16 | Tial1 | Nucleolysin TIAR: 43 kDa | 5 |
| 2310022A10Rik | RIKEN cDNA 2310022A10 gene: 44 kDa | 16 | Ppm1g | Protein phosphatase 1G: 59 kDa | 5 |
| Dhx9 | ATP-dependent RNA helicase A: 51 kDa | 16 | Capza2 | F-actin-capping protein subunit alpha-2: 33 kDa | 5 |
| Pfas | Phosphoribosylformylglycinamidine synthase: 145 kDa | 15 | Sall1 | Sal-like 1 (Drosophila): 140 kDa | 5 |
| Mta1 | Metastasis-associated protein MTA1: 79 kDa | 15 | Bckdk | [3-methyl-2-oxobutanoate dehydrogenase [lipoamide]] kinase, mitochondrial: 47 kDa | 5 |
| Xpo1 | Exportin-1: 123 kDa | 14 | Sfn | 14-3-3 protein sigma: 28 kDa | 5 |
| Rbm39 | RNA-binding protein 39: 59 kDa | 14 | Wbp11 | WW domain-binding protein 11: 23 kDa | 5 |
| Ik | Protein Red: 66 kDa | 14 | Calml3 | Calmodulin-like protein 3: 17 kDa | 5 |
| Rbm25 | RNA-binding protein 25: 100 kDa | 14 | Cct7 | T-complex protein 1 subunit eta: 9 kDa | 5 |
| Sec23a | Protein transport protein Sec23A: 86 kDa | 14 | Kxd1 | KxDL motif-containing protein 1: 27 kDa | 5 |
| Snrpa | U1 small nuclear ribonucleoprotein A: 32 kDa | 14 | Dst | Dystonin: 871 kDa | 5 |
| Nop58 | Nucleolar protein 58: 60 kDa | 14 | Rbm6 | RNA-binding motif protein 6: 128 kDa | 5 |
| Fyttd1 | UAP56-interacting factor: 36 kDa | 14 | Dst | Dystonin: 836 kDa | 5 |
| 2310022A10Rik | RIKEN cDNA 2310022A10 gene: 43 kDa | 14 | | | |

| | | |
|---|---|---|
| Msh6 | DNA mismatch repair protein Msh6: 151 kDa | 13 |
| Luc7l3 | Luc7-like protein 3: 51 kDa | 13 |
| Nasp | Nuclear autoantigenic sperm protein: 84 kDa | 12 |
| Ctnna1 | Catenin alpha-1: 100 kDa | 12 |
| Snrpa1 | U2 small nuclear ribonucleoprotein A': 28 kDa | 12 |
| Snrpd1 | Small nuclear ribonucleoprotein Sm D1: 13 kDa | 12 |
| Ctnnbl1 | Beta-catenin-like protein 1: 65 kDa | 12 |
| Sec23ip | SEC23-interacting protein: 111 kDa | 12 |
| Bcas2 | Pre-mRNA-splicing factor SPF27: 26 kDa | 12 |
| Snrpd2 | Small nuclear ribonucleoprotein Sm D2: 14 kDa | 12 |
| Lig1 | DNA ligase 1: 102 kDa | 12 |
| Hdac1 | Histone deacetylase 1: 55 kDa | 12 |
| Hdac2 | Histone deacetylase 2: 55 kDa | 12 |
| Clint1 | Clathrin interactor 1: 70 kDa | 12 |
| Rbm26 | RNA-binding protein 26: 111 kDa | 12 |
| Rbm26 | RNA-binding protein 26: 114 kDa | 12 |
| Set | Protein SET: 33 kDa | 11 |
| Smu1 | WD40 repeat-containing protein SMU1: 58 kDa | 11 |
| Atxn2l | Ataxin-2-like protein: 111 kDa | 11 |
| Plrg1 | Pleiotropic regulator 1: 57 kDa | 11 |
| Bub3 | Mitotic checkpoint protein BUB3: 37 kDa | 11 |
| Srsf10 | Serine/arginine-rich splicing factor 10: 31 kDa | 11 |

| | | |
|---|---|---|
| Trrap | Transformation/transcription domain-associated protein: 437 kDa | 11 |
| Tra2a | Transformer-2 protein homolog alpha: 32 kDa | 11 |
| Prpf40a | Pre-mRNA-processing factor 40 homolog A: 108 kDa | 11 |
| Prpf40a | Pre-mRNA-processing factor 40 homolog A: 106 kDa | 11 |
| Clint1 | Clathrin interactor 1: 69 kDa | 11 |
| Prpf4 | U4/U6 small nuclear ribonucleoprotein Prp4: 58 kDa | 10 |
| Myef2 | Myelin expression factor 2: 63 kDa | 10 |
| Srm | Spermidine synthase: 34 kDa | 10 |
| U2af1 | Splicing factor U2AF 35 kDa subunit: 28 kDa | 10 |
| Actl6a | Actin-like protein 6A: 47 kDa | 10 |
| Pabpn1 | Polyadenylate-binding protein 2: 32 kDa | 10 |
| Sf3a2 | Splicing factor 3A subunit 2: 51 kDa | 10 |
| Sssca1 | Sjoegren syndrome/scleroderma autoantigen 1 homolog: 21 kDa | 10 |
| Pum1 | Pumilio homolog 1: 100 kDa | 10 |
| Pum1 | Pumilio homolog 1: 127 kDa | 10 |
| Smarca5 | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5: 122 kDa | 9 |
| Wdhd1 | WD repeat and HMG-box DNA-binding protein 1: 120 kDa | 9 |
| Ddx23 | DEAD (Asp-Glu-Ala-Asp) box polypeptide 23: 95 kDa | 9 |

| | | |
|---|---|---|
| Eomes | Eomesodermin homolog: 75 kDa | 9 |
| Anp32b | Acidic leucine-rich nuclear phosphoprotein 32 family member B: 31 kDa | 9 |
| Sf3b6 | Splicing factor 3B subunit 6: 15 kDa | 9 |
| Mfap1a | Microfibrillar-associated protein 1A: 52 kDa | 9 |
| Prpf4b | Serine/threonine-protein kinase PRP4 homolog: 117 kDa | 9 |
| Sec24b | Sec24 related gene family, member B (S. cerevisiae): 136 kDa | 9 |
| Ap1b1 | AP-1 complex subunit beta-1: 104 kDa | 9 |
| Smc2 | Structural maintenance of chromosomes protein 2: 134 kDa | 8 |
| Nsun2 | tRNA (cytosine(34)-C(5))-methyltransferase: 81 kDa | 8 |
| Crnkl1 | Crooked neck-like protein 1: 83 kDa | 8 |
| Apex1 | DNA-(apurinic or apyrimidinic site) lyase: 35 kDa | 8 |
| Mbd3 | Methyl-CpG-binding domain protein 3: 29 kDa | 8 |
| Prpf3 | U4/U6 small nuclear ribonucleoprotein Prp3: 77 kDa | 8 |
| L1td1 | LINE-1 type transposase domain-containing protein 1: 88 kDa | 8 |
| UPF0568 protein C14orf166 homolog | UPF0568 protein C14orf166 homolog: 28 kDa | 8 |
| Snrnp40 | U5 small nuclear ribonucleoprotein 40 kDa protein: 39 kDa | 8 |

| | | |
|---|---|---|
| Cherp | Calcium homeostasis endoplasmic reticulum protein: 108 kDa | 8 |
| Snrpe | Small nuclear ribonucleoprotein E: 11 kDa | 8 |
| Dbt | Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial: 53 kDa | 8 |
| Snrpb2 | U2 small nuclear ribonucleoprotein B'': 25 kDa | 8 |
| Cand2 | Cullin-associated NEDD8-dissociated protein 2: 136 kDa | 8 |
| Bud31 | Protein BUD31 homolog: 17 kDa | 8 |
| Srsf9 | Serine/arginine-rich splicing factor 9: 26 kDa | 8 |
| Mixl1 | Homeobox protein MIXL1: 25 kDa | 8 |
| Luc7l | Putative RNA-binding protein Luc7-like 1: 44 kDa | 8 |
| Chd5 | Chromodomain-helicase-DNA-binding protein 5: 219 kDa | 8 |
| Luc7l2 | Putative RNA-binding protein Luc7-like 2: 47 kDa | 8 |
| Snrpn | Small nuclear ribonucleoprotein-associated protein N: 25 kDa | 8 |
| Snrpb | Small nuclear ribonucleoprotein-associated protein B: 24 kDa | 8 |
| Sf3a2 | Splicing factor 3A subunit 2: 50 kDa | 8 |
| Arid1a | AT-rich interactive domain-containing protein 1A: 242 kDa | 7 |
| Cpsf6 | Cleavage and polyadenylation specificity factor subunit 6: 59 kDa | 7 |
| Srsf5 | Serine/arginine-rich splicing factor 5: 31 kDa | 7 |

| | | |
|---|---|---|
| Dut | Deoxyuridine triphosphatase: 21 kDa | 7 |
| Nudt21 | Cleavage and polyadenylation specificity factor subunit 5: 26 kDa | 7 |
| Sf3b4 | Splicing factor 3B subunit 4: 44 kDa | 7 |
| Dhx38 | DEAH (Asp-Glu-Ala-His) box polypeptide 38: 141 kDa | 7 |
| Xab2 | Pre-mRNA-splicing factor SYF1: 100 kDa | 7 |
| Ubap2l | Ubiquitin-associated protein 2-like: 117 kDa | 7 |
| Sssca1 | Sjoegren syndrome/scleroderma autoantigen 1 homolog: 10 kDa | 7 |
| Numa1 | Nuclear mitotic apparatus protein 1: 236 kDa | 6 |
| Dnmt1 | DNA (cytosine-5)-methyltransferase 1: 183 kDa | 6 |
| Ccar1 | Cell division cycle and apoptosis regulator protein 1: 132 kDa | 6 |
| Col20a1 | Collagen alpha-1(XX) chain: 141 kDa | 6 |
| Ddb1 | DNA damage-binding protein 1: 127 kDa | 6 |
| Usp5 | Ubiquitin carboxyl-terminal hydrolase 5: 96 kDa | 6 |
| Dis3 | Exosome complex exonuclease RRP44: 109 kDa | 6 |
| Ssrp1 | FACT complex subunit SSRP1: 81 kDa | 6 |
| Ranbp2 | E3 SUMO-protein ligase RanBP2: 341 kDa | 6 |
| Smarce1 | SWI/SNF-related matrix-associated actin-dependent regulator of | 6 |

| | | |
|---|---|---|
| | chromatin subfamily E member 1: 47 kDa | |
| Pspc1 | Paraspeckle component 1: 59 kDa | 6 |
| Cwc15 | Spliceosome-associated protein CWC15 homolog: 27 kDa | 6 |
| Xrn2 | 5'-3' exoribonuclease 2: 109 kDa | 6 |
| Lhx1 | LIM/homeobox protein Lhx1: 45 kDa | 6 |
| Rpa1 | Replication protein A 70 kDa DNA-binding subunit: 69 kDa | 6 |
| Psmc1 | 26S proteasome regulatory subunit 4: 49 kDa | 6 |
| Glrx3 | Glutaredoxin-3: 38 kDa | 6 |
| Ewsr1 | RNA-binding protein EWS: 68 kDa | 6 |
| Rfc4 | Replication factor C subunit 4: 40 kDa | 6 |
| Bckdha | 2-oxoisovalerate dehydrogenase subunit alpha, mitochondrial: 50 kDa | 6 |
| Skiv2l2 | Superkiller viralicidic activity 2-like 2: 118 kDa | 6 |
| Nup155 | Nuclear pore complex protein Nup155: 155 kDa | 6 |
| Prpf31 | U4/U6 small nuclear ribonucleoprotein Prp31: 55 kDa | 6 |
| Pqbp1 | Polyglutamine-binding protein 1: 31 kDa | 6 |
| Pin1 | Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1: 18 kDa | 6 |
| Ogt | UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit: 117 kDa | 6 |

| | | |
|---|---|---|
| Cfdp1 | Craniofacial development protein 1: 33 kDa | 6 |
| Cbr1 | Carbonyl reductase [NADPH] 1: 31 kDa | 6 |
| Snrpf | Small nuclear ribonucleoprotein F: 10 kDa | 6 |
| Rbm17 | Splicing factor 45: 45 kDa | 6 |
| Gatad2a | Transcriptional repressor p66 alpha: 67 kDa | 6 |
| Pdcd6 | Programmed cell death protein 6: 22 kDa | 6 |
| Hmgn2 | Non-histone chromosomal protein HMG-17: 10 kDa | 6 |
| Srsf4 | Putative uncharacterized protein: 56 kDa | 6 |
| Khdrbs1 | KH domain-containing, RNA-binding, signal transduction-associated protein 1: 48 kDa | 6 |
| Ap2b1 | AP complex subunit beta: 101 kDa | 6 |
| U2af2 | Splicing factor U2AF 65 kDa subunit: 54 kDa | 6 |
| Ep400 | E1A-binding protein p400: 337 kDa | 6 |
| Mta3 | Metastasis-associated protein MTA3: 66 kDa | 6 |
| Dnajc8 | DnaJ homolog subfamily C member 8: 30 kDa | 6 |
| Pum2 | Pumilio homolog 2: 106 kDa | 6 |
| Msto1 | Protein misato homolog 1: 61 kDa | 6 |
| Clta | Clathrin light chain: 23 kDa | 6 |
| Tpr | Nuclear pore complex-associated intranuclear coiled-coil protein TPR: 267 kDa | 6 |

| | | |
|---|---|---|
| Ep400 | E1A-binding protein p400 (Fragment): 35 kDa | 6 |
| Arid3b | AT-rich interactive domain-containing protein 3B: 61 kDa | 6 |
| Chd3 | Chromodomain helicase DNA-binding protein 3: 233 kDa | 6 |
| Ldb1 | LIM domain-binding protein 1: 47 kDa | 6 |
| Chd3 | Chromodomain helicase DNA-binding protein 3 (Fragment): 214 kDa | 6 |
| Ldb1 | LIM domain-binding protein 1: 43 kDa | 6 |
| Nup93 | Nuclear pore complex protein Nup93: 93 kDa | 5 |
| Hells | Lymphocyte-specific helicase: 95 kDa | 5 |
| Ctbp2 | C-terminal-binding protein 2: 49 kDa | 5 |
| Akr1b1 | Aldose reductase: 36 kDa | 5 |
| Psmb6 | Proteasome subunit beta type-6: 25 kDa | 5 |
| Pklr | Pyruvate kinase: 59 kDa | 5 |

## 8.4 Replicate two induced cells

| 23 elements included exclusively in Rep2 induced R1: | | | 156 common elements in Rep2 induced R1 and Rep2 induced R2: | | | 63 elements included exclusively in Rep2 induced R2: | | |
|---|---|---|---|---|---|---|---|---|
| Tubb1 | Tubulin beta-1 chain: 50 kDa | 30 | Gm8994 | Uncharacterized protein: 47 kDa | 84 | Tubb5 | Tubulin beta-5 chain (Fragment): 7 kDa | 26 |
| Ruvbl1 | RuvB-like helicase (Fragment): 15 kDa | 19 | Poldip3 | Polymerase delta-interacting protein 3: 46 kDa | 82 | Fubp1 | Far upstream element-binding protein 1 (Fragment): 32 kDa | 18 |
| Ywhaq | 14-3-3 protein theta (Fragment): 34 kDa | 18 | Api5 | Apoptosis inhibitor 5: 57 kDa | 62 | Myh11 | Myosin-11: 227 kDa | 10 |
| Rasl2-9 | GTP-binding nuclear protein Ran, testis-specific isoform: 24 kDa | 11 | Sec31a | Protein transport protein Sec31A: 134 kDa | 41 | Nop58 | Nucleolar protein 58: 60 kDa | 9 |
| Sfn | 14-3-3 protein sigma: 28 kDa | 10 | Sec31a | Protein transport protein Sec31A (Fragment): 101 kDa | 38 | Arid1a | AT-rich interactive domain-containing protein 1A: 242 kDa | 9 |
| Eif3g | Eukaryotic translation initiation factor 3 subunit G: 36 kDa | 7 | Prtg | Protogenin: 131 kDa | 34 | Psmd13 | 26S proteasome non-ATPase regulatory subunit 13: 43 kDa | 9 |
| Copb2 | Coatomer subunit beta': 102 kDa | 6 | Acin1 | Apoptotic chromatin condensation inducer in the nucleus: 151 kDa | 32 | Fxr1 | Fragile X mental retardation syndrome-related protein 1: 51 kDa | 9 |
| Crnkl1 | Crooked neck-like protein 1: 83 kDa | 5 | Acin1 | Apoptotic chromatin condensation inducer in the nucleus (Fragment): 144 kDa | 32 | Nsun2 | tRNA (cytosine(34)-C(5))-methyltransferase: 81 kDa | 8 |
| Safb | Scaffold attachment factor B1: 105 kDa | 5 | Magoh | Protein mago nashi homolog: 17 kDa | 31 | Idh1 | Isocitrate dehydrogenase [NADP] cytoplasmic: 47 kDa | 8 |
| Asns | Asparagine synthetase [glutamine-hydrolyzing]: 64 kDa | 5 | Poldip3 | Polymerase delta-interacting protein 3 (Fragment): 22 kDa | 30 | Sf3a2 | Splicing factor 3A subunit 2: 50 kDa | 8 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Polr2a | DNA-directed RNA polymerase II subunit RPB1: 217 kDa | 5 | Spata5 | Spermatogenesis-associated protein 5: 97 kDa | 28 | Srsf5 | Serine/arginine-rich splicing factor 5: 31 kDa | 7 |
| Bcas2 | Pre-mRNA-splicing factor SPF27: 26 kDa | 5 | Sarnp | SAP domain-containing ribonucleoprotein: 24 kDa | 28 | Ncbp1 | Nuclear cap-binding protein subunit 1: 92 kDa | 7 |
| Ctps1 | CTP synthase 1: 67 kDa | 5 | Sf3b2 | Putative uncharacterized protein: 98 kDa | 27 | Ap1b1 | AP-1 complex subunit beta-1: 104 kDa | 7 |
| Aimp1 | Aminoacyl tRNA synthase complex-interacting multifunctional protein 1: 34 kDa | 5 | Chtop | Chromatin target of PRMT1 protein: 27 kDa | 26 | Tpr | Nuclear pore complex-associated intranuclear coiled-coil protein TPR: 267 kDa | 7 |
| Psmb5 | Proteasome subunit beta type-5: 29 kDa | 5 | Puf60 | Poly(U)-binding-splicing factor PUF60: 60 kDa | 26 | Arid3b | AT-rich interactive domain-containing protein 3B: 61 kDa | 7 |
| Pin1 | Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1: 18 kDa | 5 | Magohb | Mago nashi protein: 17 kDa | 26 | Ldb1 | LIM domain-binding protein 1: 47 kDa | 7 |
| Nsfl1c | NSFL1 cofactor p47: 41 kDa | 5 | Pnn | Pinin: 82 kDa | 24 | Tpr | Nucleoprotein TPR: 274 kDa | 7 |
| Ptma | Prothymosin alpha: 12 kDa | 5 | Rbm8a | RNA-binding protein 8A: 20 kDa | 23 | Ldb1 | LIM domain-binding protein 1: 43 kDa | 7 |
| Gga2 | ADP-ribosylation factor-binding protein GGA2: 66 kDa | 5 | Chd4 | Chromodomain-helicase-DNA-binding protein 4: 218 kDa | 21 | Dis3 | Exosome complex exonuclease RRP44: 109 kDa | 6 |
| Taf15 | Putative uncharacterized protein: 59 kDa | 5 | Zc3h14 | Zinc finger CCCH domain-containing protein 14: 82 kDa | 21 | Psip1 | PC4 and SFRS1-interacting protein: 60 kDa | 6 |
| KRT84 | Keratin, type II cuticular Hb4: 65 kDa | 5 | Rnps1 | RNA-binding protein with serine-rich domain 1: 34 kDa | 19 | Psmd1 | 26S proteasome non-ATPase regulatory subunit 1: 106 kDa | 6 |
| Calml3 | Calmodulin-like protein 3: 17 kDa | 5 | Zc3h11a | Zinc finger CCCH domain-containing protein 11A: 86 kDa | 19 | Fkbp3 | Peptidyl-prolyl cis-trans isomerase FKBP3: 25 kDa | 6 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Sssca1 | Sjoegren syndrome/scleroderma autoantigen 1 homolog: 10 kDa | 5 | Sap18b | Histone deacetylase complex subunit SAP18: 20 kDa | 18 | Prpf4b | Serine/threonine-protein kinase PRP4 homolog: 117 kDa | 6 |
| | | | Ncbp3 | Nuclear cap-binding protein subunit 3: 70 kDa | 17 | Glrx3 | Glutaredoxin-3: 38 kDa | 6 |
| | | | Bclaf1 | Bcl-2-associated transcription factor 1: 106 kDa | 16 | Etf1 | Eukaryotic peptide chain release factor subunit 1: 49 kDa | 6 |
| | | | Sf3a1 | Splicing factor 3A subunit 1: 89 kDa | 16 | Psma2 | Proteasome subunit alpha type-2: 26 kDa | 6 |
| | | | Cdc5l | Cell division cycle 5-like protein: 92 kDa | 16 | Psmb6 | Proteasome subunit beta type-6: 25 kDa | 6 |
| | | | 2310022A10Rik | RIKEN cDNA 2310022A10 gene: 43 kDa | 16 | Cbx5 | Chromobox protein homolog 5: 22 kDa | 6 |
| | | | 2310022A10Rik | RIKEN cDNA 2310022A10 gene: 44 kDa | 16 | Mixl1 | Homeobox protein MIXL1: 25 kDa | 6 |
| | | | Srsf11 | Serine/arginine-rich-splicing factor 11 (Fragment): 56 kDa | 16 | Rhoa | Transforming protein RhoA: 22 kDa | 6 |
| | | | Srsf11 | Putative uncharacterized protein: 53 kDa | 16 | Cdc37 | Hsp90 co-chaperone Cdc37: 45 kDa | 6 |
| | | | Snw1 | SNW domain-containing protein 1: 61 kDa | 16 | Pklr | Pyruvate kinase: 59 kDa | 6 |
| | | | Snw1 | SNW domain-containing protein 1: 61 kDa | 16 | Sept8 | Septin-8: 56 kDa | 6 |
| | | | Eomes | Eomesodermin homolog: 75 kDa | 15 | Amot | Angiomotin (Fragment): 76 kDa | 6 |
| | | | Fkbp4 | Peptidylprolyl isomerase (Fragment): 30 kDa | 15 | Smc3 | Structural maintenance of chromosomes protein 3: 142 kDa | 5 |
| | | | Ddx42 | ATP-dependent RNA helicase DDX42: 102 kDa | 14 | Hcfc1 | Host cell factor 1: 210 kDa | 5 |

| | | | | | |
|---|---|---|---|---|---|
| U2surp | U2 snRNP-associated SURP motif-containing protein: 118 kDa | 14 | Cherp | Calcium homeostasis endoplasmic reticulum protein: 108 kDa | 5 |
| Sall4 | Sal-like protein 4: 113 kDa | 14 | Lhx1 | LIM/homeobox protein Lhx1: 45 kDa | 5 |
| Snrnp70 | U1 small nuclear ribonucleoprotein 70 kDa: 52 kDa | 14 | Rfc4 | Replication factor C subunit 4: 40 kDa | 5 |
| Srrt | Serrate RNA effector molecule homolog: 100 kDa | 14 | Actl6a | Actin-like protein 6A: 47 kDa | 5 |
| Fyttd1 | UAP56-interacting factor: 36 kDa | 14 | Psmd14 | 26S proteasome non-ATPase regulatory subunit 14: 35 kDa | 5 |
| Ddx46 | Probable ATP-dependent RNA helicase DDX46: 117 kDa | 12 | Ncapd2 | Condensin complex subunit 1: 156 kDa | 5 |
| Nasp | Nuclear autoantigenic sperm protein: 84 kDa | 12 | Gatad2b | Transcriptional repressor p66-beta: 65 kDa | 5 |
| Xpo1 | Exportin-1: 123 kDa | 12 | Fbl | rRNA 2'-O-methyltransferase fibrillarin: 34 kDa | 5 |
| Rbm39 | RNA-binding protein 39: 59 kDa | 12 | Snrpb2 | U2 small nuclear ribonucleoprotein B'': 25 kDa | 5 |
| Ctnna1 | Catenin alpha-1: 100 kDa | 12 | Cfdp1 | Craniofacial development protein 1: 33 kDa | 5 |
| Wbp11 | WW domain-binding protein 11: 70 kDa | 12 | Snrpf | Small nuclear ribonucleoprotein F: 10 kDa | 5 |
| Prmt5 | Protein arginine N-methyltransferase 5: 73 kDa | 11 | Rbm17 | Splicing factor 45: 45 kDa | 5 |
| Ik | Protein Red: 66 kDa | 11 | Tcea1 | Transcription elongation factor A protein 1: 34 kDa | 5 |
| Snrpd1 | Small nuclear ribonucleoprotein Sm D1: 13 kDa | 10 | Ckap5 | Cytoskeleton-associated protein 5: 226 kDa | 5 |

| | | | | | |
|---|---|---|---|---|---|
| Snrpa | U1 small nuclear ribonucleoprotein A: 32 kDa | 10 | Kdm1a | Lysine-specific histone demethylase 1A: 93 kDa | 5 |
| Smu1 | WD40 repeat-containing protein SMU1: 58 kDa | 10 | Rabep1 | Rab GTPase-binding effector protein 1: 100 kDa | 5 |
| Usp5 | Ubiquitin carboxyl-terminal hydrolase 5: 96 kDa | 10 | Cdc42 | Cell division control protein 42 homolog: 21 kDa | 5 |
| Anp32b | Acidic leucine-rich nuclear phosphoprotein 32 family member B: 31 kDa | 10 | Vdac2 | Voltage-dependent anion-selective channel protein 2: 32 kDa | 5 |
| Uncharacterized protein | Uncharacterized protein: 92 kDa | 10 | Myl12a | MCG5400: 20 kDa | 5 |
| Cnot1 | CCR4-NOT transcription complex subunit 1: 267 kDa | 9 | Fn1 | Fibronectin: 273 kDa | 5 |
| Sec24c | SEC24 related gene family, member C (S. cerevisiae), isoform CRA_a: 119 kDa | 9 | Psmb7 | Proteasome subunit beta type-7: 30 kDa | 5 |
| Smc2 | Structural maintenance of chromosomes protein 2: 134 kDa | 9 | Rab14 | Ras-related protein Rab-14: 24 kDa | 5 |
| Myef2 | Myelin expression factor 2: 63 kDa | 9 | Gstp1 | Glutathione S-transferase P 1: 24 kDa | 5 |
| Ddb1 | DNA damage-binding protein 1: 127 kDa | 9 | Smarcc2 | SWI/SNF complex subunit SMARCC2: 133 kDa | 5 |
| Bub3 | Mitotic checkpoint protein BUB3: 37 kDa | 9 | Gm5239 | MCG1031578: 14 kDa | 5 |
| Sf3a2 | Splicing factor 3A subunit 2: 51 kDa | 9 | Pdlim5 | ENH isoform 1d: 62 kDa | 5 |
| Prpf40a | Pre-mRNA-processing factor 40 homolog A: 108 kDa | 9 | Ubap2l | Ubiquitin associated protein 2-like, isoform CRA_g: 112 kDa | 5 |
| Prpf40a | Pre-mRNA-processing factor 40 homolog A: 106 kDa | 9 | | | |

| | | |
|---|---|---|
| Edc4 | Enhancer of mRNA-decapping protein 4: 142 kDa | 9 |
| Prpf6 | Pre-mRNA-processing factor 6: 107 kDa | 8 |
| Smarca5 | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5: 122 kDa | 8 |
| Rbm25 | RNA-binding protein 25: 100 kDa | 8 |
| Numa1 | Nuclear mitotic apparatus protein 1: 236 kDa | 8 |
| Set | Protein SET: 33 kDa | 8 |
| Snrpa1 | U2 small nuclear ribonucleoprotein A': 28 kDa | 8 |
| Sec23a | Protein transport protein Sec23A: 86 kDa | 8 |
| Prpf4 | U4/U6 small nuclear ribonucleoprotein Prp4: 58 kDa | 8 |
| Sec23b | Protein transport protein Sec23B: 86 kDa | 8 |
| Srm | Spermidine synthase: 34 kDa | 8 |
| Apex1 | DNA-(apurinic or apyrimidinic site) lyase: 35 kDa | 8 |
| U2af1 | Splicing factor U2AF 35 kDa subunit: 28 kDa | 8 |
| Smarce1 | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1: 47 kDa | 8 |
| UPF0568 protein C14orf166 homolog | UPF0568 protein C14orf166 homolog: 28 kDa | 8 |

| | | |
|---|---|---|
| Mta1 | Metastasis-associated protein MTA1: 79 kDa | 8 |
| Snrpd2 | Small nuclear ribonucleoprotein Sm D2: 14 kDa | 8 |
| Srsf10 | Serine/arginine-rich splicing factor 10: 31 kDa | 8 |
| Tra2a | Transformer-2 protein homolog alpha: 32 kDa | 8 |
| Psmc1 | 26S proteasome regulatory subunit 4: 49 kDa | 8 |
| Cand2 | Cullin-associated NEDD8-dissociated protein 2: 136 kDa | 8 |
| U2af2 | Splicing factor U2AF 65 kDa subunit: 54 kDa | 8 |
| Hdac1 | Histone deacetylase 1: 55 kDa | 8 |
| Rbm26 | RNA-binding protein 26: 111 kDa | 8 |
| Rbm26 | RNA-binding protein 26: 114 kDa | 8 |
| Prdx6 | Peroxiredoxin-6: 22 kDa | 8 |
| Ctnnbl1 | Beta-catenin-like protein 1: 65 kDa | 7 |
| Atxn2l | Ataxin-2-like protein: 111 kDa | 7 |
| Ddx23 | DEAD (Asp-Glu-Ala-Asp) box polypeptide 23: 95 kDa | 7 |
| Luc7l3 | Luc7-like protein 3: 51 kDa | 7 |
| Cpsf6 | Cleavage and polyadenylation specificity factor subunit 6: 59 kDa | 7 |
| Mbd3 | Methyl-CpG-binding domain protein 3: 29 kDa | 7 |
| Snrnp40 | U5 small nuclear ribonucleoprotein 40 kDa protein: 39 kDa | 7 |
| Pspc1 | Paraspeckle component 1: 59 kDa | 7 |
| Sf3b6 | Splicing factor 3B subunit 6: 15 kDa | 7 |
| Sept11 | Septin-11: 50 kDa | 7 |

| | | |
|---|---|---|
| Sssca1 | Sjoegren syndrome/scleroderma autoantigen 1 homolog: 21 kDa | 7 |
| Hdac2 | Histone deacetylase 2: 55 kDa | 7 |
| Msh6 | DNA mismatch repair protein Msh6: 151 kDa | 6 |
| Pfas | Phosphoribosylformylglycinamidine synthase: 145 kDa | 6 |
| Smc4 | Structural maintenance of chromosomes protein 4: 147 kDa | 6 |
| Sec23ip | SEC23-interacting protein: 111 kDa | 6 |
| Ssrp1 | FACT complex subunit SSRP1: 81 kDa | 6 |
| Plrg1 | Pleiotropic regulator 1: 57 kDa | 6 |
| Ranbp2 | E3 SUMO-protein ligase RanBP2: 341 kDa | 6 |
| Mfap1a | Microfibrillar-associated protein 1A: 52 kDa | 6 |
| Cwc15 | Spliceosome-associated protein CWC15 homolog: 27 kDa | 6 |
| Dut | Deoxyuridine triphosphatase: 21 kDa | 6 |
| Vcl | Vinculin: 117 kDa | 6 |
| Rpa1 | Replication protein A 70 kDa DNA-binding subunit: 69 kDa | 6 |
| Snrpe | Small nuclear ribonucleoprotein E: 11 kDa | 6 |
| Sf3b4 | Splicing factor 3B subunit 4: 44 kDa | 6 |
| Sept7 | Septin-7: 51 kDa | 6 |
| St13 | Hsc70-interacting protein: 42 kDa | 6 |
| Psmd6 | 26S proteasome non-ATPase regulatory subunit 6: 46 kDa | 6 |

| | | |
|---|---|---|
| Psmb1 | Proteasome subunit beta type-1: 26 kDa | 6 |
| Psma5 | Proteasome subunit alpha type-5: 26 kDa | 6 |
| Smarcd1 | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 1: 58 kDa | 6 |
| Hmgn2 | Non-histone chromosomal protein HMG-17: 10 kDa | 6 |
| Srsf9 | Serine/arginine-rich splicing factor 9: 26 kDa | 6 |
| Srsf4 | Putative uncharacterized protein: 56 kDa | 6 |
| Luc7l | Putative RNA-binding protein Luc7-like 1: 44 kDa | 6 |
| Ap2b1 | AP complex subunit beta: 101 kDa | 6 |
| Arhgdia | Rho GDP-dissociation inhibitor 1: 23 kDa | 6 |
| Pabpn1 | Polyadenylate-binding protein 2: 32 kDa | 6 |
| Mesp1 | posterior protein 1: 28 kDa | 6 |
| Pdia6 | Protein disulfide-isomerase A6: 48 kDa | 6 |
| Ckb | Creatine kinase B-type: 43 kDa | 6 |
| Amot | Angiomotin: 121 kDa | 6 |
| Luc7l2 | Putative RNA-binding protein Luc7-like 2: 47 kDa | 6 |
| Ubap2l | Ubiquitin-associated protein 2-like: 117 kDa | 6 |
| Clint1 | Clathrin interactor 1: 70 kDa | 6 |
| Raly | RNA-binding protein Raly (Fragment): 23 kDa | 6 |

| | | |
|---|---|---|
| Pdlim5 | PDZ and LIM domain protein 5: 63 kDa | 6 |
| Vdac1 | Voltage-dependent anion-selective channel protein 1: 32 kDa | 6 |
| Sars | Putative uncharacterized protein: 61 kDa | 6 |
| Dst | Dystonin: 871 kDa | 6 |
| Sars | Serine--tRNA ligase, cytoplasmic: 58 kDa | 6 |
| Raly | RNA-binding protein Raly: 33 kDa | 6 |
| Dst | Dystonin: 836 kDa | 6 |
| Prpf3 | U4/U6 small nuclear ribonucleoprotein Prp3: 77 kDa | 5 |
| Akr1b1 | Aldose reductase: 36 kDa | 5 |
| Cbr1 | Carbonyl reductase [NADPH] 1: 31 kDa | 5 |
| Khdrbs1 | KH domain-containing, RNA-binding, signal transduction-associated protein 1: 48 kDa | 5 |
| Ubap2 | Ubiquitin-associated protein 2: 118 kDa | 5 |
| Rcc1 | Regulator of chromosome condensation: 45 kDa | 5 |
| Chd5 | Chromodomain-helicase-DNA-binding protein 5: 219 kDa | 5 |
| Rab10 | Ras-related protein Rab-10: 23 kDa | 5 |
| Serpinh1 | Serpin H1: 47 kDa | 5 |
| Psma7 | Proteasome subunit alpha type-7: 28 kDa | 5 |
| Snrpn | Small nuclear ribonucleoprotein-associated protein N: 25 kDa | 5 |

| | | |
|---|---|---|
| Snrpb | Small nuclear ribonucleoprotein-associated protein B: 24 kDa | 5 |