# Extrinsic and Intrinsic Signalling Pathways that Regulate Stem Cell Developmental Potential

Feodor duPasquier Price

This thesis is submitted as a
partial fulfillment of the Ph.D. program in
Cellular and Molecular Medicine

Submitted on May 25th, 2012 to:
Faculty of Graduate and Postdoctoral Studies
& Department of Cellular and Molecular Medicine
Faculty of Medicine, University of Ottawa

© Feodor duPasquier Price, Ottawa, Canada, 2012

# Abstract

Instructive signals, whether external or internal, play critical roles in regulating the developmental potency or ability to self-renew of stem cells. External signals may range from secreted growth factors to extracellular matrix proteins found in the stem cell niche. Internal signals include activated signalling cascades and the eventual transcriptional mechanisms they initiate. In either fashion, stem cells are regulated in a complex temporal and context specific manner in order to maintain or maximise their unique characteristics.

Previous experiments suggest that Wnt3a plays a role in maintaining the pluripotent state of mouse embryonic stem (mES) cells. However, in the absence of leukemia inhibitory factor (LIF), Wnt signalling is unable to maintain ES cells in the undifferentiated state. This implies that maintaining the pluripotent state of mES cells is not the primary function of canonical Wnt signalling. To further characterize the role of Wnt3a in pluripotency and lineage specification undifferentiated and differentiated mES cells were induced with Wnt3a. Wnt3a induced the formation of a metastable primitive endoderm state and upon subsequent differentiation, the induction of large quantities of visceral endoderm. Furthermore, we determined that the ability of Wnt3a to induce a metastable primitive endoderm state was mediated by the T-box transcription factor *Tbx3*. Our data demonstrates a novel role for Wnt3a in promoting the interconversion of undifferentiated mES cells into a pluripotent primitive endoderm state.

Aging of skeletal muscle tissue is accompanied by fibrosis, atrophy and remodeling all of which negatively affect muscle performance. Whether this reduction in

skeletal muscle competency is directly attributed to a resident adult stem cell population called satellite cells remains largely unknown.

Here, we undertook an investigation into how age affects the transcriptional profile of satellite cells and their repopulating ability following transplantation. We determined that as satellite cells age, both their regenerative capacity and ability to colonize the satellite cell niche is reduced. Additionally, we identified satellite cell specific transcriptional profiles that differed with respect to age. Therefore, we conclude that intrinsic factors are an important determinant of satellite cell regenerative capacity during the aging process.

# Authorization

Regarding the use of material from "Paul M. Krzyzanowski, Feodor D. Price, Enrique M. Muro, Michael A. Rudnicki, and Miguel A. Andrade-Navarro. Integration of expressed sequence tag data flanking predicted RNA secondary structures facilitates novel non-coding RNA discovery. PLoS One (2011)" in Appendix B of this thesis:

The Public Library of Science (PLoS) applies the Creative Commons Attribution License (CC-BY) to all works we publish (read the human-readable summary or the full license legal code). Under this license, authors retain ownership of the copyright for their content, but allow anyone to download, reuse, reprint, modify, distribute, and/or copy the content as long as the original authors and source are cited. **No permission is required from the authors or the publishers.**

# Table of Contents

**Abstract** ..... i
**Authorization** ..... iii
**Table of Contents** ..... iv
**List of Figures** ..... vi
**List of Tables** ..... vii
**List of Abbreviations** ..... vii
**Contributions of Collaborators** ..... ix
**Acknowledgments** ..... x
**Chapter 1 – General Introduction** ..... 1
**1.1. Stem Cells** ..... 2
**1.2. Embryonic Stem Cells** ..... 5
*1.2.1 Transcriptional Networks Regulating the Epiblast and Primitive Endoderm* ..... 8
*1.2.2. The Metastable State of Embryonic Stem Cells* ..... 9
**1.3. Adult Stem Cells** ..... 11
*1.3.1. History and Origin of the Satellite cell: an Adult Muscle Stem cell* ..... 12
*1.3.2. The Role of Satellite Cells in Muscle Regeneration* ..... 20
*1.3.3. The Effect of Age on Skeletal Muscle Regeneration* ..... 25
**1.4. The Canonical Wnt signalling pathway** ..... 28
*1.4.1. Canonical Wnt Signalling and Pluripotency* ..... 29
*1.4.2. Wnt Signalling and Lineage Specification* ..... 32
**1.5 Hypothesis and Specific Aims** ..... 35
*1.5.1. A Role for Wnt3a in Pluripotency and Early Differentiation of Mouse Embryonic Stem Cells* ..... 35
*1.5.2. The Effect of Age on Satellite Cell Function* ..... 35
**Chapter 2 - Wnt3a induces a primitive endoderm metastable state in mouse embryonic stem cells** ..... 37
**2.1 Abstract** ..... 38
**2.2 Introduction** ..... 38
**2.3 Results** ..... 41
*2.3.1 Wnt3a Affects the Morphology and Growth Rate of Undifferentiated and Differentiating mES Cells* ..... 42
*2.3.2 Wnt3a Alters the Global Gene Expression Profile of Blastocyst Cell Types and the Primary Germ Layers.* ..... 45
*2.3.3 Differentiation mES Cells with Wnt3a Enhances the Formation of Visceral Endoderm* ..... 49
*2.3.4 Wnt3a Perturbs the Pluripotent Ground State of mES Cells by Inducing a Primitive Endoderm Metastable State* ..... 52
*2.3.5 A Mechanism that Links Wnt3a via Tbx3 to the Induction of a Metastable PE Population.* ..... 56
**2.4 Discussion** ..... 59
**2.5 Materials and Methods** ..... 69
**Chapter 3 - Transgenic Models to Study Myogenesis** ..... 86
**3.1 Abstract** ..... 87
**3.2 Introduction** ..... 87
*3.2.1 Adult Skeletal Muscle Development* ..... 87

*3.2.2 Satellite Cell Heterogeneity* ... *89*
*3.2.3 Generating Transgenic Reporter Mice for Pax7 and Myf5* ... *90*
**3.3 Results** ... **93**
*3.3.1 Pax7 Transgenic Models* ... *94*
*3.3.2 Myf5 Transgenic Models* ... *104*
**3.4 Discussion** ... **113**
**3.5 Materials and Methods** ... **118**
**Supplemental Figures – Tables** ... **125**
**Chapter 4 - Intrinsic Decline in Repopulating Capacity of Aging Satellite Cells** ... **136**
**3.1 Abstract** ... **137**
**3.2 Introduction** ... **137**
**3.3 Results** ... **140**
*3.3.1 Purification of Satellite Cells from Fetal, Adolescent, Adult and Aged Muscle* ... *140*
*3.3.2 M-cadherin Expression on Sorted Satellite Cell Populations Does Not Correlate with Differences in Functional Capacity In Vitro* ... *144*
*3.3.3 Increased Age Negatively Affects the Regenerative Potential of Engrafted Satellite Cells into mdx Muscle* ... *147*
*3.3.4 Genome Wide Expression Analysis of Satellite Cell Populations Identifies Distinct Age-Related Profiles* ... *151*
**3.4. Discussion** ... **155**
**3.5 Materials and Methods** ... **161**
**Chapter 5 - General Discussion** ... **176**
**5.1. Overview of findings** ... **177**
*5.1.1 Wnt3a Induces a Metastable Primitive Endoderm Population in mES Cells.* *178*
*5.1.2 Intrinsic Decline in Repopulating Capacity of Aging Satellite Cells* ... *179*
**5.2. Defining the mechanisms involved in modulating the metastable state** ... **180**
*4.2.1. Significance* ... *181*
*5.2.2. Future Directions* ... *182*
**5.3. Age-related effects on skeletal muscle regenerative capacity** ... **183**
*5.3.1. Significance* ... *183*
*5.3.2. Future Directions* ... *184*
**5.4. Biomedical Implications** ... **186**
*5.4.1 The Metastable State in Lineage Priming and Reprogramming* ... *186*
*5.4.2. Satellite Cells as a Cell Based Therapy to Treat Muscular Dystrophy* ... *187*
**5.5. Conclusions** ... **188**
**Appendix A - Integration of expressed sequence tag data flanking predicted RNA secondary structures facilitates novel non-coding RNA discovery.** ... **190**
**References** ... **204**

# List of Figures

**Chapter 1**

Figure 1. Stem Cell Potency as a function of Lineage Specification. 3
Figure 2. Primary Cell Types of the Blastocyst. 6
Figure 3. Characteristics of the Satellite Cell. 16
Figure 4. Satellite cell activation differentiation and fusion. 22
Figure 5. The Canonical Wnt Signalling Pathway 30

**Chapter 2**

Figure 1. Wnt3a overexpression elicits increased growth rate and the formation of cystic embryoid bodies following differentiation of mES cells. 43
Figure 2. Primitive endoderm genes induced by constitutive Wnt3a overexpression. 46
Figure 3. Continued Wnt3a expression during differentiation induces formation of visceral endoderm in mESCs. 50
Figure 4. Wnt3a alters the metastable state of mES cells by promoting the formation of primitive endoderm cells. 53
Figure 5. Sorted Wnt3a Nanog-GFP populations interconvert under self-renewing conditions. 57
Figure 6. Tbx3 mediates the Wnt3a induction of a primitive endoderm metastable population. 60
Figure 7. Wnt signalling alters the differentiation of mES cells and induces a primitive endoderm meta-stable state. 67
Figure S1. Validation of Wnt3a expressing mESC lines. 73
Figure S2. RNAseq quality controls. 75
Figure S3. Activation of Canonical Wnt Signalling 77
Figure S4. RNAseq of recombinant Wnt3a treated mES cells. 79

**Chapter 3**

Figure 1. Design and validation of *Pax7* transgenic vectors. 95
Figure 2. Analysis of YFP expression in differentiated ES cells. 99
Figure 3. Analysis of YFP expression in primary myoblasts treated with Cre recombinase. 102
Figure 4. Design and validation of a 2aYpetGluc expression vector. 105
Figure 5. Design and validation of *Myf5* transgenic vectors. 108
Figure 6. Creation and analysis of a Myf5 TdTom BAC transgenic. 111
Figure S1. Validation of YFP-ArtINT and Cre recombination. 126
Figure S2. Production and purification of Tat-Cre. 128
Figure S3. Functional validation of Tat-Cre recombinase. 130

**Chapter 4**

Figure 1. Age promotes the expression of Pax7 on satellite cells while modulating the composition of cell surface markers. 142
Figure 2. M-cadherin expression on freshly sorted satellite cells increases with age. 145
Figure 3. Old age negatively affects the regenerative capacity of satellite cells following transplantation into *mdx* TA muscle. 149
Figure 4. Genome wide transcriptional differences as a function of satellite cell age. 153
Figure 5. GO term enrichment in Fetal and Aged satellite cell populations. 156
Figure S1. Zsgreen + cells decrease with Age 172

Figure S2. Repopulating capacity of adolescent satellite cells remains unchanged irrespective of M-cadherin expression. ..... 174

# List of Tables

**Chapter 2**
Table SI. RNA-seq results for FGF signaling pathway members ..... 81
Table SII. Antibodies Used in this Study..... 82
Table SIII. Primers Used in this Study. ..... 83
**Chapter 3**
Table I: Summary of Myf5 BAC transgenic validation..... 125
Table SI – Cloning Primers used for Pax7 and Myf5 Targeting Vectors..... 132
Table SII – Antibodies used in this study ..... 133
**Chapter 4**
Table 1. Cell counts of Dystrophin+ or Pax7+ and Zsgreen+ Satellite cells..... 166
Table SI. Antibodies Used in this Study..... 171

# List of Abbreviations

| | |
|---|---|
| **APC** | Adenomatous poliposis coli |
| **AVE** | Anterior Visceral Endoderm |
| **BAC** | Bacterial Articifical Chromosome |
| **bFGF** | basic fibroblast growth factor |
| **BMP4** | bone morphogenic protien 4 |
| **BMPs** | Bone Morphogenic Proteins |
| **bp** | base pair |
| **cDNA** | carboxy-terminal |
| **CMV** | Cytomegalovirus |
| **DAPI** | 4',6-diamidino-2-phenylindole |
| **DMEM** | Dulbecco's modified Eagle's medium |
| **DNA** | deoxyribonucleic acid |
| **dVE** | Distal Visceral Endoderm |
| **E** | Embryonic day |
| **ECM** | extracellular matrix |
| **EDL** | Extensor Digitalis Longus |
| **EDTA** | ethyl diamine-tetra-acetic acid |
| **EMT** | epithelial to mesenchymal transition |
| **ES** | Embryonic Stem cell |
| **ESCV** | Embryonic Stem Cell targeting Vector |
| **FACS** | fluorescent-activated cell sorting |
| **FBS** | fetal bovine serum |
| **FGFs** | Fibroblast Growth Factors |
| **Fzd** | Frizzled |

| | |
|---|---|
| **G418** | Neomycin |
| **GAPDH** | glyceraldyde-3-phosphate dehydrogenase |
| **Gata4** | GATA-Binding Factor 4 |
| **Gata6** | GATA-Binding Factor 6 |
| **GFP** | Green Fluorescent Protein |
| **gLuc** | Gaussia Luciferase |
| **HSC** | Hematopoietic Stem Cell |
| **iPS cells** | induced pluripotent stem cells |
| **IRES** | Internal Ribosomal Entry Site |
| **Kb** | kilobase |
| **kDa** | kilodalton |
| **Ki-67** | proliferation-related Ki-67 antigen |
| **LB** | Luria broth |
| **Mcad** | M-Cadherin; Cadherin 15 |
| **MEM** | Minimal essential medium |
| **mRNA** | messenger RNA |
| **Myf5** | Myogenic factor 5 |
| **MyHC** | Myosin Heavy Chain |
| **N-terminal** | sodium chloride |
| **NaCl** | nuclear localization sequence |
| **NLS** | amino-terminal |
| **Oct4** | Octamer Binding Transcription Factor 4 or Pou5f1 |
| **P7zs** | Pax7-zsgreen |
| **Pax7** | Paired box transcription factor 7 |
| **PBS** | phosphate-buffered saline |
| **PCR** | polymerase chain reaction |
| **PFA** | paraformaldehyde |
| **PrE** | Primitive Endoderm |
| **Puro** | Puromycin |
| **RA** | Retinoic Acid |
| **RNA** | Ribonucleic acid |
| **RT** | reverse-transcription |
| **Sox17** | Sry containing Homeobox Transcription Factor 17 |
| **Sox7** | Sry containing Homeobox Transcription Factor 7 |
| **TA** | Tibialis anterior |
| **TdTom** | Tandem-dimer of Dsred (Tomato) |
| **Wnt** | Wingless-related MMTV site |
| **Wt** | Wild-type |
| **XEN** | Extraembryonic Endoderm Cell line |
| **YFP** | Yellow Fluorescent Protein |

## Contributions of Collaborators

The conception, performance, analysis and interpretation of experimental work outlined in this thesis was conducted by the thesis author unless otherwise specified. The writing and preparation of the thesis was conducted by the thesis author with editing help from Micheal Rudnicki, Julia von Maltzahn, Andrew Jones, Johnathan Smid, Sharlene Faulkes and Hang Yin.

I have been very fortunate to collaborate with the following people. Dr. Frank Grosveld and his technical staff at Erasmus University who conducted Solexa sequencing of total RNA for experiments conducted in Chapter 2. Dr. Hang Yin for running RNA-seq data through TopHat and Cufflinks and conducting bioinformatics analysis giving rise to heat map data found in figures 2, 3, 6, and S3 in Chapter 2. Alexander Grayston contributed to sectioning embryoid bodies for figure 3 in Chapter 2. Dr. Julia von Maltzahn transplanted satellite cell populations, aided in sectioning muscle, conducting IF and counting cells for figure 3 in Chapter 3. David Wilson aided in sectioning and staining muscle giving rise to data found in figure 3 in Chapter 3. Gareth Palidwor conducted RMA on microarray raw data and provided figure 4 in Chapter 3.

I sincerely appreciate the help my collaborators provided to these projects. It was a pleasure working with them.

# Acknowledgments

First and foremost I would like to thank my PhD supervisor Dr. Michael Rudnicki for instilling in me the necessary qualities to persist, perform and succeed in scientific research. Over the years in the Rudnicki lab I have had the pleasure to learn and collaborate with many Post docs (Iain Mckinnell, Kazuki Kuroda, Yoichi Kawabe, Julia von Maltzahn, Shihuan Kuang, Vahab Soleimani, Gianni Parise, Fabien Le Grand, Florian Bentzinger, and Hang Yin). I have learned valuable lessons from previous PhD students in the lab who have guided and inspired me to pursue my PhD projects (Jeff Ishibashi, Chet Holterman, Mark Gillespie, Mike Huh, Johnathan Smid, Vince Punch, Paul Krzyzanowski). It has been a pleasure developing my scientific skills and learning in an evolving environment with fellow graduate students (Andrew Jones, Alessandra Pasut, Will Wang, Greg Addicks, Francina Jackson, Negin Shahid, and Vanessa Seale). Finally it is very fulfilling to instil scientific knowledge and techniques into the minds of bright young students (Alexander Grayston).

During my PhD I have been guided by my committee members (Rashmi Kothary, Lisheng Wang and Luc Sabourin). They have not only taught me valuable lessons about science but also about life.

At the OHRI I have been surrounded by many exceptional researchers. It was a pleasure to observe and hopefully one day emulate them. To my family, although most of what I have worked on during this PhD is complete gibberish to you, your unwavering support was absolutely vital to my success. I am greatly appreciative.

To Christine Johannes… I love you dearly.

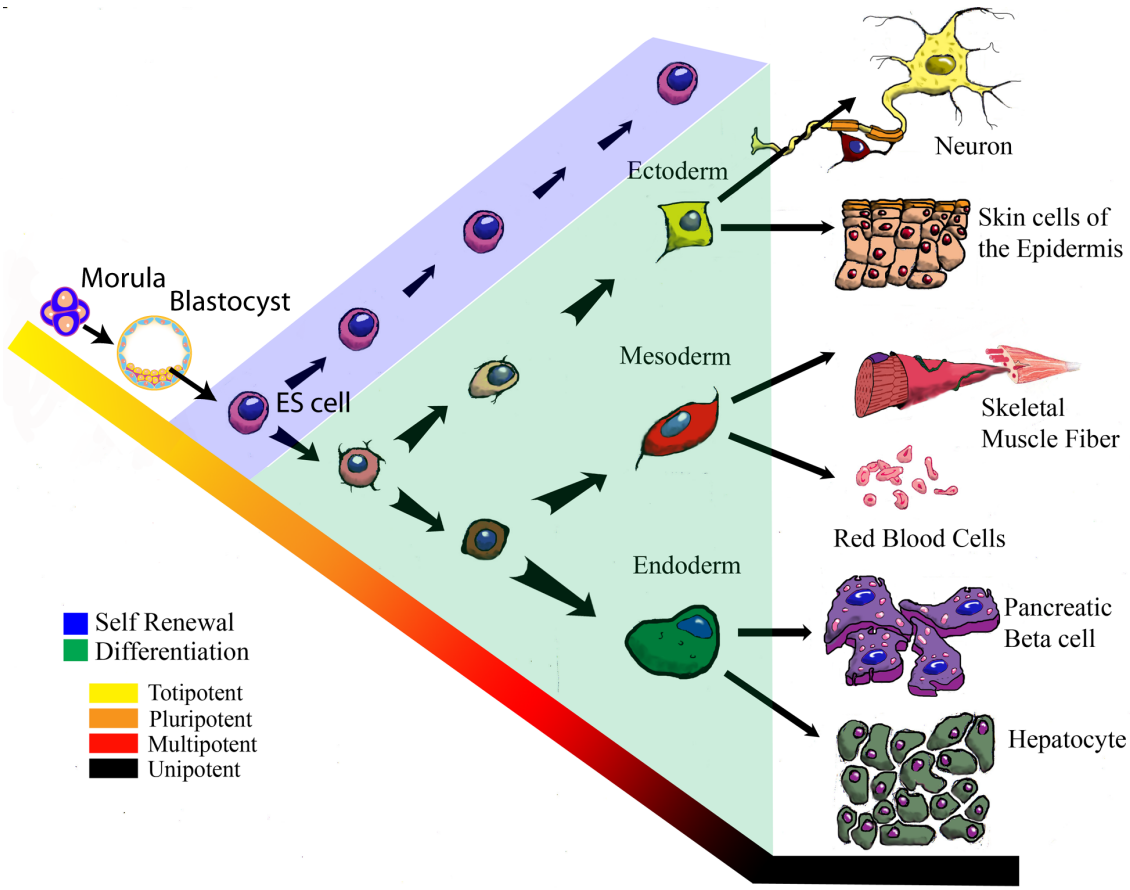
# Chapter 1 – General Introduction

## 1.1. Stem Cells

Stem cells are defined by certain characteristics, primarily an ability for long term self-renewal and the capacity to differentiate into multiple cell lineages. Stem cells are responsible for the development and maintenance of tissues and organs and self-renew or differentiate in response to a combination of biochemical signals and biomechanical stimuli provided by the stem cell niche. Stem cell division can be either symmetric or asymmetric. An asymmetric division results in the formation of two non-identical daughter cells; one commits to a specialized fate while the other repopulates the niche to maintain the stem cell pool. Conversely, differentiating daughter cells undergo symmetric divisions giving rise to a reservoir of precursor cells that contribute to embryonic development or adult tissue regeneration.

Stem cells can be isolated from either adult or embryonic tissue, and depending on a hierarchical state differ in their ability to give rise to multiple cell lineages. This diversity, or breadth of differentiation, is defined as potency. A hierarchical progression from a state of totipotency through to unipotency exists; whereby, at each level the ability to differentiate into multiple cell types is progressively diminished (**Fig 1**). Totipotency is reserved for the zygote as it has the ability to differentiate into all cell types of the embryo as well as the extra-embryonic lineages. Pluripotency is the ability to give rise to all the cells of the embryo proper, while multipotency represents a more restricted set of differentiated progeny (ex. cells of the hematopoietic lineage). Stem cells are broadly characterized as being either of embryonic or adult origin, the former being pluripotent and the latter being multipotent.

**Figure 1. Stem Cell Potency as a function of Lineage Specification.**
Stem cells progress from a state of Totipotency to Unipotency whereby at each stage they become progressively restricted in their ability to give rise to a variety of cell types. ES cells being pluripotent have the ability to self renew or differentiate into cells of the primary germ layers. Germ layer stem cells have a restricted differentiation potential and are deemed multipotent. The variation in potency is depicted as a gradient below a given cell type where yellow represents the greatest potency – Totipotent, while black represents the least potency – Unipotent.

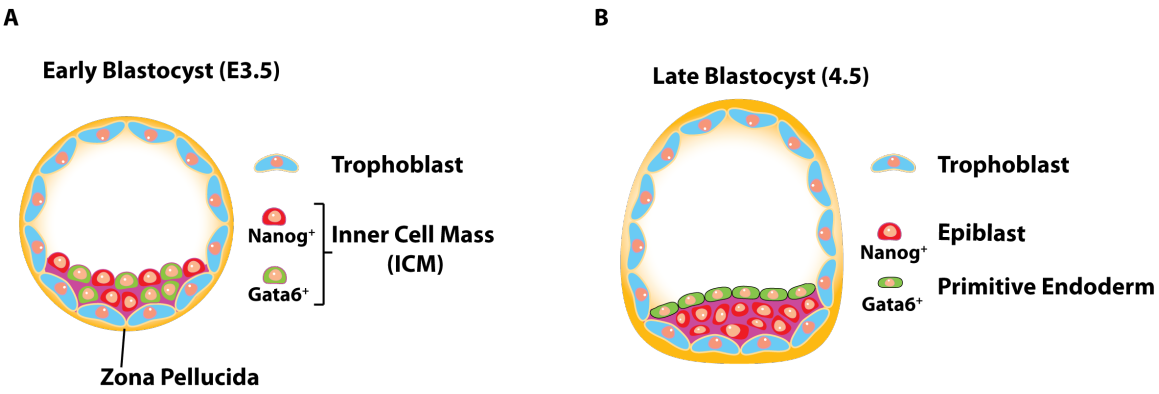

Morula
Blastocyst
ES cell
Ectoderm
Mesoderm
Endoderm
Neuron
Skin cells of
the Epidermis
Skeletal
Muscle Fiber
Red Blood Cells
Pancreatic
Beta cell
Hepatocyte
Self Renewal
Differentiation
Totipotent
Pluripotent
Multipotent
Unipotent

## 1.2. Embryonic Stem Cells

Embryonic stem (ES) cells are defined based on their origin – that being an early embryo. The process of cell specialization in the developing embryo is a highly coordinated process. This process is essential for converting cells with a great degree of similarity in the early morula to highly specialized cells in the embryo.

The first cell fate decision in the embryo occurs during the transition from a morula to an early blastocyst when the outer cells of the morula are epithelialized and become trophectoderm (TE) and the inner cells generate the pluripotent inner cell mass (ICM) (**Fig 2A**)[1]. The TE provides nutrients to the developing embryo and following gastrulation will give rise to the majority of the developing placenta[2]. As the blastocyst matures the ICM will segregate into the pluripotent epiblast (EPI) and the primitive endoderm (PrE). The PrE, a precursor of the extraembryonic endoderm (ExEn) lineage, forms a thin layer of cells that separates the epiblast from the blastocoel cavity (**Fig 2B**). Following implantation of the blastocyst into the uterine wall, the epiblast cells will divide to form a cup-like structure called the egg cylinder. During this period of time the PrE segregates into parietal endoderm (PE) and visceral endoderm (VE)[3]. The parietal endoderm associates with the trophectoderm and secretes Reichert's membrane forming the parietal yolk sac. In contrast the visceral endoderm maintains interactions with the developing embryo proper, forms the epithelial lining of the yolk sac and retains the ability to contribute cells to the definite gut tube[4-6]. Cells of the ExEn serve an important role during embryogenesis by providing nutritive support and orchestrate the determination of the anterior-posterior axis

**Figure 2. Primary Cell Types of the Blastocyst.**
**(A)** The developing embryo depicted at E3.5 is deemed a blastocyst. The blastocyst consists of two predominant cell types: the first being the Trophoblast (depicted here in blue), while the second cell type is the inner cell mass (ICM) composed of cells positive for *Gata6* (Green) and *Nanog* (Red).
**(B)** At E4.5 the blastocyst displays segregation of the primitive endoderm and the epiblast bringing the total number of distinct cell types in the developing embryo to three.


A
Early Blastocyst (E3.5)
Trophoblast
Nanog+
Inner Cell Mass
(ICM)
Gata6+
Zona Pellucida
B
Late Blastocyst (4.5)
Trophoblast
Epiblast
Nanog+
Primitive Endoderm
Gata6+

Mouse embryonic stem (mES) cells are derived from the inner cell mass of a pre-implantation blastocyst. mES cells can self-renew indefinitely *in vitro* while maintaining their ability to differentiate into all somatic cell types[7,8]. Leukemia inhibitory factor (LIF) was identified as being able to maintain mES in an undifferentiated state in the absence of MEFs[9]. mES cells when injected back into a blastocyst can contribute to tissues of the fetus and extraembryonic mesoderm and at a low frequency trophectoderm and primitive endoderm[10]. The ability to generate chimeric animals with the potential of contributing to the germline makes mES cells an important tool to study gene function[11].

*1.2.1 Transcriptional Networks Regulating the Epiblast and Primitive Endoderm*

The two primary transcription factors that define the epiblast and primitive endoderm are Nanog and Gata6 respectively. During blastocyst maturation (E3.5), Nanog and Gata6 are expressed in a mutually exclusive "salt and pepper" pattern (**Fig 2A**)[12]. Nanog is a homeodomain transcription factor present in pluripotent cells and developing germ cells[13,14]. Nanog$^{-/-}$ mice are embryonic lethal due to an overabundance of disorganized extraembryonic tissues and a lack of epiblast at E5.5[14]. Nanog$^{-/-}$ ES cells lose pluripotency and differentiate into the extraembryonic endoderm lineage. *In vitro*, constitutive expression of Nanog enhances self-renewal and maintains ES cells in the undifferentiated state under conditions in which pluripotency would otherwise be lost[13,15]. Nanog, along with other critical pluripotency transcription factors Oct4 and Sox2 are members in an autoregulatory loop that maintains the pluripotent state[16].

Gata6 is a zinc finger transcription factor first expressed at the 8 cell stage of the developing blastocyst[17]. Gata6 along with Gata4 are expressed in the primitive endoderm and are critical for extraembryonic endoderm development[18,19]. Gata6$^{-/-}$ mice are

embryonic lethal by E7.0 and fail to form a discernable primitive endoderm layer[20,21]. *In vitro*, overexpression of Gata6 or Gata4 is sufficient to differentiate mES cells into the primitive endoderm when in pluripotent culture conditions[22].

Interactions between Nanog and Gata6 are complex. Nanog is capable of binding the promoter and repressing Gata6 expression[23]. However, Nanog via a non cell autonomous mechanism is required for proper PrE formation as evidenced by Gata4 expression[24]. Recent experiments suggest initiation of the PrE and Gata6 expression is Nanog independent[25]. PrE formation is governed by the temporal expression of Gata6, followed by Sox17 and then Gata4[17,25,26]. Interestingly, the expression of Sox17 and Gata4 in PrE is dependent upon Nanog activating FGFs in a non cell autonomous manner. These experiments provide proof of the complex interplay between transcription factors and extrinsic growth factors involved in the proper maturation of PrE.

### *1.2.2. The Metastable State of Embryonic Stem Cells*

Metastability refers to the interconversion between transcriptional states in which mES cells reside while in self-renewing culture conditions[27]. The metastable state was first alluded to in experiments conducted by Furuwasa *et al.* looking at the distribution in expression of platelet endothelial cell adhesion molecule 1 (PECAM-1) and stage-specific embryonic antigen 1 (SSEA-1) in cultures of mouse embryonic stem cells[28]. In an attempt to determine the cell surface marker expression of pluripotent cells the authors discovered that sorted $PECAM^{+}/SSEA\text{-}1^{+}$ and $PECAM^{+}/SSEA\text{-}1^{-}$ were interchangeable following culture in self-renewing conditions.

Later experiments by Chambers *et al.* further characterized the metastable state of mES cells through the analysis of fluctuations in Nanog expression based on a Knock-in

GFP reporter line[27]. ES cells cycle between a Nanog-high or Nanog-low state. Transient down regulation of Nanog results in cells being "primed" for differentiation at the transcriptional level. However, it is important to note that these cells still remain pluripotent and were not committed to differentiation. The fact that purified Nanog GFP-low or Nanog GFP-high cells can interconvert renders these populations metastable.

Several papers have since presented data demonstrating that pluripotent mES cells display reversible oscillation in gene expression of Stella[29] and Rex1[30]. Expression of Stella (Dppa3) marks interchangeable subpopulations of pluripotent mES cells[29,31]. Stella$^+$ and Stella$^-$ cells, when differentiated, have similarities to the ICM and the epiblast respectively. Further characterization at the epigenetic level identified reversible differences between the Stella-GFP$^+$ and Stella-GFP$^-$ cells. These experiments demonstrate that mES cells contain a population of metastable cells that is most similar to the epiblast.

Analysis of undifferentiated mES cells has identified the expression of genes associated with various differentiation states including Gata6[23], Brachyury[32], Sox17[26], and Hex[33]. A recent report describes the formation of a metastable state with similarities to the primitive endoderm. Canham *et al.* demonstrate the expression of Hex, an early endodermal marker, is expressed in a population of Oct4$^+$ undifferentiated mES cells[33]. Interestingly, the Hex$^+$ and Hex$^-$ cells can interconvert in self-renewing conditions yet following differentiation *in vitro* or *in vivo* they differ in their lineage specification. Hex$^+$ cells predominantly contribute to the extraembryonic visceral and parietal endoderm while Hex$^-$ cells can generate high contribution chimeras. These experiments parallel early studies conducted on Nanog-low ES cells demonstrating an up regulation in

expression of primitive endoderm genes and further confirm the presence of a second metastable state that has characteristics of the primitive endoderm[34].

Based on the above experiments, two metastable populations currently exist that represent the cellular derivatives of the ICM, namely the epiblast (Epi) and the primitive endoderm (PE). Determining whether other metastable states exist (possibly a trophectoderm-like metastable state) or how to manipulate a given metastable state represent future areas of research.

## 1.3. Adult Stem Cells

Adult stem cells are distinct from embryonic stem cells with regards to their potency and origin. The existence of adult stem cells was first characterized by Till and McCulloch who identified the presence of self-renewing cells present in the bone marrow that can give rise to spleen nodules[35]. Small quantities of adult stem cells exist in most tissues throughout the body where they remain quiescent for long periods of time prior to being activated in response to disease or tissue injury. Adult stem cells represent undifferentiated cells found in a differentiated tissue that are capable of self-renewal and are multipotent in that they can differentiate to yield the specialized cell types of the tissue in which they reside. However, previous reports indicate that adult stem cells can differentiate into lineages other then their tissue of origin, for example transplanted bone marrow or enriched hematopoietic stem cells (HSCs) are reported to give rise to cells of the mesoderm[36-38], endoderm[39], and ectoderm[40,41].

Many different types of adult stem cells exist; the most notable being HSCs[42,43], however adult stem cells have been discovered in numerous organs including but not limited to neural[44-46], skeletal muscle[47-50], dermal[51], dental pulp[52], and the intestine[53]. As

opposed to embryonic stem cells, the origin of many adult stem cells remains unknown. Future experiments elaborating upon the origins and potency of adult stem cells are necessary in order to elucidate their full stem cell potential.

*1.3.1. History and Origin of the Satellite cell: an Adult Muscle Stem cell*

Half a century ago, a group of mononucleated cells that reside on the periphery of adult skeletal muscle myofibers were observed by electron microscopy[47]. These cells were named satellite cells due to their sub-laminar location and intimate association with the plasma membrane of myofibers.

The close proximity between satellite cells and their myofibers spawned the hypothesis that these cells may be an adult stem cell responsible for skeletal muscle growth and regeneration[47]. Indeed, thymidine-$^{3}$H tracing experiments indicated that satellite cells are mitotically quiescent in adult muscle but can quickly enter the cell cycle following muscle injury[54]. Electron microscopy demonstrated that following activation satellite cells undergo mitosis and contribute to myofiber nuclei[55,56]. Satellite cells give rise to proliferating myoblasts (myogenic progenitors cells), which were previously shown to be competent in generating multinucleated myotubes *in vitro*[54,57,58]. More definitive evidence demonstrated that myofiber necrosis is accompanied by satellite cell outgrowth, clonal expansion of the myogenic progenitor pool and later fusion to form functional regenerated myotubes[59,60]. These experiments support the notion that it is the satellite cells, rather than myonuclei, that are the principle component responsible for postnatal muscle growth and repair.

*In vivo*, satellite cells reside between the basal lamina and the sarcolemma in a specialized niche regulated by signals from the local and systemic milieu (**Fig. 3A**).

Numerous studies reveal that innervation, vascularity, hormones, nutrition and the extent of tissue injury profoundly influence the proliferation and differentiation of satellite cells[61-66]. These findings imply the satellite cell niche plays a critical role in regulating the functional capacity and cellular state of satellite cells.

Satellite cells fulfill the criteria of an adult stem cell in that they are able to both replicate themselves (self-renew) and differentiate into the functional progeny of skeletal muscle tissue. The first evidence of satellite cell self-renewal came from single fiber transplantation assays in which donor derived cells repopulated the satellite cell niche while giving rise to thousands of myonuclei[50]. The multipotency of satellite cells was demonstrated through their ability to differentiate into the myogenic, osteogenic and adipogenic lineages[48]. These observations demonstrate that satellite cells are *bona fide* adult muscle stem cells.

Asymmetric division is a common manner of stem cell self-renewal. Close examination of satellite cell divisions on single myofibers revealed that satellite cells can undergo either asymmetric division for self-renewal or symmetric division to expand the satellite cell population[49]. The satellite cell population is heterogeneous and can be classified in a hierarchical manner based on function and gene expression. Based on lineage tracing experiments a sub population of satellite cells having never expressed the myogenic transcription factor *Myf5* (satellite stem cells) are placed hierarchically above satellite cells that have expressed Myf5 at some point during development (satellite myogenic cells) (**Fig. 3B**). Satellite stem cells, upon asymmetric division (typically in a apical-basal orientation) will give rise to two daughter cells, only one of which has activated *Myf5*. Functional differences in regenerative potential exist between satellite stem cells and

satellite myogenic cells. Satellite stem cells are able to reconstitute the stem cell compartment therefore contributing to long-term muscle regeneration whereas satellite myogenic cells will preferentially differentiate upon transplantation *in vivo*. Of note, the fashion in which satellite cells divide, either symmetric or asymmetric, is dictated by the relative position of the daughter cells in relation to the myofiber. This finding strongly argues that the self-renewal of satellite cells is governed by the structure and signaling of their immediate niche[49]. Further investigation has revealed an important role for non-canonical Wnt signaling in the regulation of satellite cell self-renewal[67].

Satellite cells reside in a unique cellular niche allowing for their unequivocal identification by electron microscopy[47]. This technique also revealed other morphological characteristics of satellite cells: large nuclear-to-cytoplasmic ratio, few organelles, small nucleus, and condensed interphase chromatin. This morphology is consistent with the notion that satellite cells, in healthy unstressed muscles, are generally mitotically quiescent (G0 phase) and transcriptionally inactive[68]. In addition to electron microscopy, satellite cells can be identified by phase contrast microscopy on single myofiber explants[69]. The identification of satellite cells by fluorescence microscopy relies on the biomarkers of satellite cells and their niche (**Fig. 3A**). Common markers for satellite cells include: the paired box transcription factors Pax7[70] or Pax3[71], myogenic regulatory factor Myf5[72], cell adhesion protein M-Cadherin[73], tyrosine receptor kinase c-Met[74], cell surface attachment receptor α7-Intergin[75,76], cluster of differentiation protein CD34[77], transmembrane heparan sulfate proteoglycans Syndecan-3 and Syndecan-4[78], caveolae-forming protein Caveolin-1[76,79], calcitonin receptor[80] and nuclear envelope proteins lamin A/C and emerin[76]. Of these, Pax7 is the predominant biomarker for

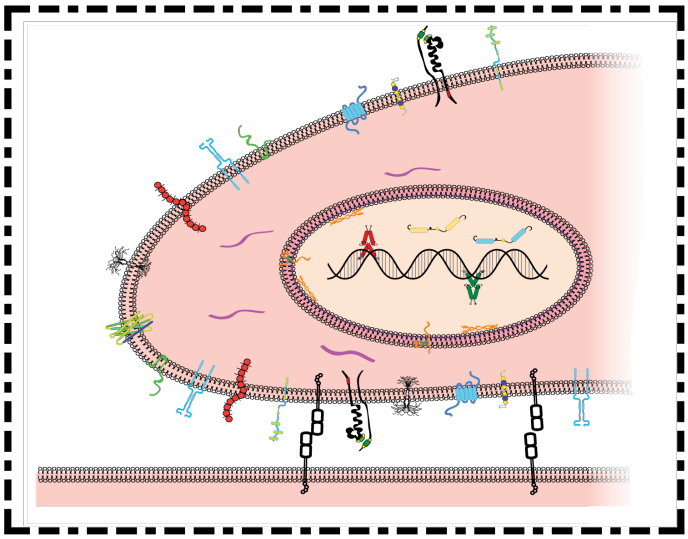
satellite cells as it specifically expressed in all quiescent satellite cells[70] across multiple species, including human[81], monkey[81], mouse[70], pig[82], chick[83], salamander[84], frog[85] and zebrafish[86]. Of note the Myf5 locus is active in about 90% of quiescent satellite cells, which suggests most satellite cells are committed to the myogenic lineage[77].

Satellite cells can be released from single myofiber explants by physical trituration[87] or enzymatic digestion[88]. Alternatively, satellite cells can be isolated from skeletal muscles by fluorescent-activated cell sorting (FACS). In the latter method, single cells are released from muscle tissue chunks by enzymatic digestion, followed by immunofluorescence labeling of specific satellite cell surface markers (positive selection) and definitive cell surface markers for non-satellite cell populations (negative selection). As Pax7 is specifically expressed in quiescent and proliferating satellite cells they can be faithfully identified by fluorescent protein expression in Tamoxifen injected animals carrying both a *Pax7*$^{CreER}$ allele and a fluorescent Cre reporter allele (e.g. *R26R-EYFP*)[89]. In addition, two transgenic mouse strains, which carry fluorescent proteins driven by Nestin- or Pax7-regulatory elements, are also useful for isolating satellite cells by FACS[90,91]. These FACS based methods allow for the isolation of the vast majority of satellite cells in high purity from skeletal muscle.

Protocols have been developed to isolate satellite cells in bulk utilizing their proliferation and adhesion characteristics under culturing conditions[92-94]. It is noteworthy that the satellite cell progeny cultured on regular collagen-coated plastic dishes and under activation conditions, also called satellite cell-derived primary myoblasts, are molecularly and functionally distinct from the satellite cells freshly isolated from muscles[80,95,96]. In addition, the gene expression profile of these cultured primary myoblasts also seems to

**Figure 3. Characteristics of the Satellite Cell.**
**(A)** Numerous proteins are expressed in satellite cells and have been used as markers to distinguish between surrounding cell types within skeletal muscle. Due to heterogeneity in satellite cell populations it is unlikely that all of these markers are expressed in a given satellite cell at a specific time. Never the less, this panel summarizes the cellular location of markers used to identify satellite cells.
**(B)** The satellite cell population is heterogeneous and can be classified in a hierarchical manner based on function and gene expression. Based on lineage tracing experiments a sub population of satellite cells having never expressed the myogenic transcription factor *Myf5* (satellite stem cells) are placed hierarchically above satellite cells that have expressed Myf5 at some point during development (satellite myogenic cells). Satellite stem cells, upon asymmetric division (typically in a apical-basal orientation) will give rise to two daughter cells, only one of which has activated *Myf5*. Functional differences in regenerative potential exist between satellite stem cells and satellite myogenic cells. Satellite stem cells are able to reconstitute the stem cell compartment therefore contributing to long-term muscle regeneration whereas satellite myogenic cells will preferentially differentiate upon transplantation *in vivo*.

A



**Satellite cell Markers**

Transcription Factors

Pax7 MyoD

Pax3 Myf5

Membrane Proteins

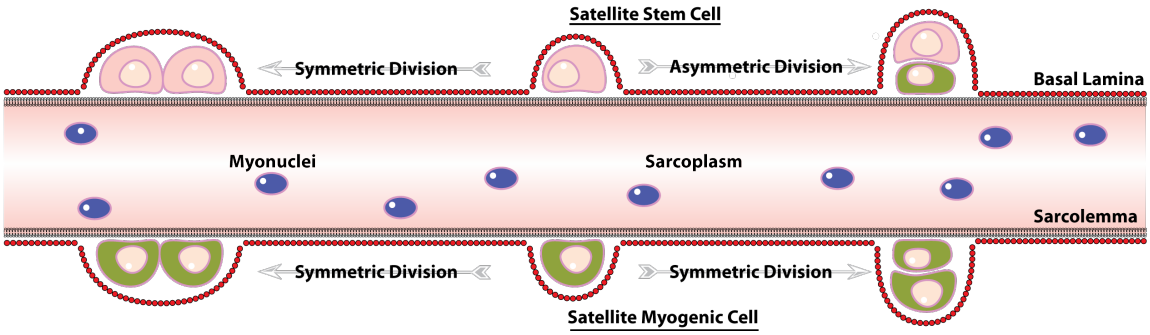
Itga7 Itgb1 CD34 Synd3/4

Lamin A/C Emerin Cav1

Mcad cMet Vcam1

CalcR Ncam1 CXCR4

B



differ from that of activated satellite cells *in vivo*[97]. These essential differences are possibly due to the lack of a satellite cell niche in *in vitro* culturing conditions.

Classical methods of developmental biology established that skeletal muscles within both the adult trunk and limbs develop from embryonic somites[98,99]. The generation of Pax3 and Pax7 knock-in reporter alleles allows precise tracing of Pax3/Pax7-expressing myogenic progenitor cells during muscle development in a temporal and spatial manner[96,100-103]. This technique together with electroporation labeling[104,105], retroviral labeling[106], Cre-LoxP based lineage tracing[106-108], and traditional quail-chick transplantation[105,106], shed light on the embryonic origins of adult satellite cells.

Accumulating evidence indicates that adult satellite cells originate from the dermomyotome[100,105,106,109], an epithelial structure formed on the dorsal part of the somite. The dermomyotome contains multipotent progenitor cells, which eventually give rise to multiple adult tissues/cell types, such as dermal fibroblasts, endothelial cells, vascular smooth muscles, brown fat tissue as well as to all skeletal muscles of the trunk and limbs[110].

Embryonic muscle development takes place in two successive stages. During the first stage, a group of post-mitotic mononucleated myocytes, expressing Myf5 and Mrf4, migrate out from the border regions of the dermomyotome and form primitive muscles beneath the dermomyotome (primary myotome)[105,109]. These primitive muscles are the source of fetal and adult trunk muscle. During the second stage of muscle growth, the central portion of the dermomyotome undergoes an epithelial-to-mesenchymal transition (EMT). During EMT, tightly packed epithelial cells tease apart and turn into a loose

mesenchymal state prior to assuming different developmental fates: cells in the medial dermomyotome will develop into brown fat, dermis and trunk muscle, while cells in the lateral dermomyotome will give rise to endothelia and limb muscles. The variation in cell fates arising from the same group of progenitor cells is presumably due to asymmetric cell division[104,111]. EMT is accompanied by the extensive cell migration [100,105,106,109]. With the breakdown of dermomyotome, a group of proliferating progenitor cells, expressing both Pax3 and Pax7, migrate from the central region of the dermomyotome into the previously formed primary myotome. Upon arrival, some progenitor cells continue to proliferate and replenish the progenitor pool. These cells, which were not observed in the primary myotome at earlier stages, account for the majority of proliferating cells in embryonic/fetal trunk muscle and will eventually take up the satellite cell niche and serve as postnatal satellite cells in trunk muscle [105,109].

During EMT, another group of proliferating progenitor cells, expressing Pax3 (but not Pax7 in mouse) delaminate from the ventral-lateral border of the dermomyotome and migrate to the limb bud mesenchyme[100,106,112]. These progenitor cells still maintain their multipotency as they give rise to the limb vascular, lymphatic endothelia and limb muscles[113,114]. At E11.5 of mouse embryo development, some progenitor cells start to express Pax7 in the anterior limb buds[103], which specifies these cells into the myogenic lineage[115]. Similar to the myotome-located progenitors, these $Pax3^{+}/Pax7^{+}$ progenitor cells in limb buds undergo proliferation/differentiation and subsequently a fraction of them withdraw from the cell cycle and become satellite cells. All together, these observations indicate that $Pax3^{+}/Pax7^{+}$ embryonic progenitor cells are the major source of adult satellite cells in trunk and limb muscles.

*1.3.2. The Role of Satellite Cells in Muscle Regeneration*

Muscle regeneration occurs in three sequential but overlapping stages: 1) the inflammatory response; 2) the activation, differentiation and fusion of satellite cells; and 3) the maturation and remodelling of newly formed myofibers. Muscle regeneration can be characterized by a series of morphological stages elucidated through histological and immunofluorescence staining. Muscle cross-sections of newly formed myofibers can be readily distinguished by their small calibre and centrally located myonuclei[116,117]. At the end of muscle regeneration, newly formed myofibers increase in size, and myonuclei move to the periphery of the muscle fibers. Under normal conditions, the regenerated muscles are morphologically and functionally indistinguishable from undamaged muscles. This process, in many but not all respects, recapitulates embryonic myogenesis.

Muscle degeneration begins with necrosis of damaged muscle fibers. This event is initiated by dissolution of the myofiber sarcolemma, which leads to increased myofiber permeability. Moreover, myofiber necrosis is accompanied by increased calcium influx or calcium release from the sarcoplasmic reticulum, which in turn activates calcium-dependent proteolysis and drives tissue degeneration [reviewed in[118-120]]. Myofiber necrosis also activates the complement cascade and induces the inflammatory response[121,122].

Following muscle degeneration a highly orchestrated regeneration process ensues. This process is contingent upon the proper expression of Pax7 (Pax3 in diaphragm muscle) and the myogenic regulatory factors (MRFs). The MRFs comprise a group of basic helix-loop-helix transcription factors including MyoD[123-125], Myf5[126,127], Myogenin and Mrf4. The majority of satellite cells express either MyoD or Myf5 by 24 hours

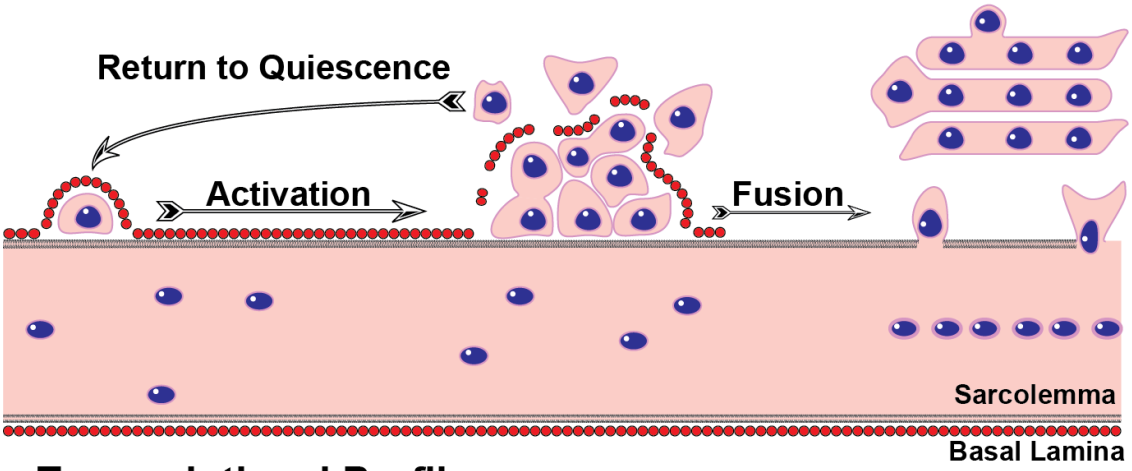
following injury and subsequently co-express both factors by 48 hours[72,128] (**Fig. 4**). The ability of satellite cells to up-regulate either MyoD or Myf5 suggests these two transcription factors may have different functions in adult myogenesis[129]. Evidence demonstrates that the proliferation of satellite cells and their progeny provide a sufficient source of new myonuclei for muscle repair[54,130,131].

*MyoD*$^{-/-}$ mutant mice display markedly reduced muscle mass[132]. This atrophy phenotype is potentially due to delayed differentiation[133,134]. Similarly, muscle regeneration is also impaired in *MyoD*$^{-/-}$ mice, which is characterized by an increased number of myoblasts within the damaged area[132]. These *MyoD*$^{-/-}$ myoblasts persist for extended periods of time, fail to differentiate and do not fuse into myotubes. Furthermore, when cultured in myogenic differentiation conditions, *MyoD*$^{-/-}$ myoblasts continue to proliferate and eventually give rise to a decreased number of differentiated mononucleated myocytes[123,124,134]. On the other hand, ectopic expression of MyoD in NIH-3T3 and 10T1/2 fibroblasts is sufficient to activate the complete myogenic program in these cells[125]. Taken together, these observations indicate that expression of MyoD is an important regulator of myogenic differentiation and in its absence, activated myoblasts have a propensity for proliferation and self-renewal[123].

In contrast to the *MyoD*$^{-/-}$ mice, *Myf5*$^{-/-}$ mutant mice show a myofiber hypertrophy phenotype[126] and the proliferation of *Myf5*$^{-/-}$ myoblasts is compromised[126,127]. These results implicate a distinct role for Myf5 in adult myoblast proliferation while MyoD is essential for differentiation. Together, the aforementioned observations suggest a hypothesis that satellite cells enter different myogenic programs depending on whether Myf5 or MyoD expression predominates[135]. Predominance of MyoD expression would

**Figure 4. Satellite cell activation differentiation and fusion.**
The myogenic program is orchestrated by key transcription factors that dictate the balance between proliferation and differentiation ultimately resulting in the transformation of individual satellite cells into a syncytial contractile myofiber. This process begins with mitotically quiescent satellite cells (G0 phase) residing in a sublaminar niche. Quiescent satellite cells are characterized by their expression of Pax7 and Myf5 but not MyoD or Myogenin. Damage to the environment surrounding satellite cells results in the deterioration of the basal lamina and their exit from the quiescent state (satellite cell activation). Proliferating satellite cells and their progeny are often referred to as myogenic precursor cells (MPC) or adult myoblasts. Adult myoblasts express the myogenic transcription factors MyoD and Myf5. Following proliferation, adult myoblasts begin differentiation by down regulating Pax7. The initiation of terminal differentiation and fusion begins with the expression of Myogenin, which in concert with MyoD will activate muscle specific structural and contractile genes such as myosin heavy chain (MyHC) and muscle creatine kinase (Mck). During regeneration activated satellite cells have the capability to return to quiescence to maintain the satellite cell pool. This ability is critical for long-term muscle integrity.

# Adult Skeletal Muscle Regeneration




## Transcriptional Profile

| Quiescence (Satellite Cell) | Proliferation (Primary Myoblast) | Differentiation (Myotube) |
|---|---|---|
| Pax7+ | Pax7+ | MyHC+ |
| Pax3+/- | MyoD+ | MCK+ |
| Myf5+/- | Myf5+ | MyoG+/- |

drive the program toward early differentiation, as exemplified by the behaviour of *Myf5*$^{-/-}$ myoblasts[129]. In contrast, predominance of Myf5 expression would direct the program into enhanced proliferation and delayed differentiation, as shown by the behaviour of *MyoD*$^{-/-}$ myoblasts[123]. Finally, myoblasts co-expressing both Myf5 and MyoD would exhibit the intermediate growth and differentiation propensities as shown by most of satellite cell-derived myoblasts.

After limited rounds of proliferation, the majority of myogenic precursors enter the myogenic differentiation program and begin to fuse to damaged myofibers or fuse to each other to form new myofibers. Terminal myogenic differentiation is an irreversible procedure and is driven by the sequential expression of key transcription factors (master regulators), which are destined to regulate their target genes[136,137]. The initiation of terminal differentiation starts with the expression of Myogenin, and Mrf4[72,124,138-140]. MyoD directly regulates the expression of Myogenin and the Mef2 family of co-factors (Mef2s)[141]. A large portion of target genes induced by MyoD, Myogenin and Mef2s are muscle-specific structural and contractile genes, such as genes encoding M-Cadherin and MyHC.

Complex regulation of MyoD activity affects the epigenetic and cell cycle states of differentiating myoblasts to promote growth arrest[142]. After exiting the cell cycle, myogenic cells undergo cell-to-cell fusion to repair damaged myofibers or form nascent multinucleated myofibers. The cellular events in this complex process have been extensively studied[143-147].

In summary, activation of satellite cells following muscle injury results in the expansion of the myogenic cell pool and the initiation of the myogenic program. This

program, orchestrated by key transcription factors, dictates the balance between proliferation and differentiation and drives the functional transformation from individual proliferating myogenic cells to a syncytial contractile myofiber.

*1.3.3. The Effect of Age on Skeletal Muscle Regeneration*

As organisms age they suffer from a decline in tissue regenerative capacity, which correlates with a reduction in physiological function[148]. This impaired regenerative potential is often attributed to a regression in the functional capacity of resident adult stem cells that are essential for tissue homeostasis and repair. In aged skeletal muscle, the progressive loss of muscle mass accompanied by a decrease in function is known as sarcopenia [149]. Atrophy of individual fibers occurs as a result of denervation combined with a reduction in the total number of myofibers[150-152]. Aged skeletal muscle undergoes changes in the contractile properties of the motor units resulting in a decrease in force and power[153,154]. The effect of age on skeletal muscle satellite cells can be attributed to both intrinsic and extrinsic factors. However, the precise mechanisms regulating this process remain largely unknown.

As mice mature, the total number of satellite cells in skeletal muscle rapidly decrease and subsequently plateau around 5% in adulthood[155]. The determination of total satellite cells in aged muscle has proven more difficult. Confounding factors including differences in muscle type assayed, species, methods of quantifying satellite cells and similarities in age groups being compared have rendered a decision on whether or not aged muscle contains fewer satellite cells difficult (reviewed[156]). However, irrespective of whether or not the total number of satellite cells decreases in old individuals, a more

important characteristic is whether or not the functional capacity of individual satellite cells changes with age.

In vitro studies on primary myoblasts isolated from aged muscle demonstrate an activation defect resulting in slower initial proliferation[157,158]. However, overall myogenic capacity of aged primary myoblasts appears comparable to those derived from adult, and following several passages *in vitro* the proliferation rate is similar as well[158,159]. Notably, a recent study thoroughly compared the gene expression profiles and proliferation/differentiation potential of satellite cells isolated from limb and facial muscles of adult and aged mice[160]. This study revealed heterogeneity at both the population and single cell level for satellite cells. Intriguingly, despite their heterogeneous gene expression profiles, satellite cells isolated from limb extensor digitorum longus (EDL) and facial (masseter) muscles are comparable in their ability to repair limb tibialis anterior (TA) muscle after transplantation. This finding suggests that although satellite cells from various muscles have distinct gene expression and behaviours *in vitro*, their regeneration potential *in vivo* might be largely determined by the host stem cell niche and microenvironment.

Early *in vivo* studies conducted with rat EDL muscles transplanted from old into young rats and vice versa concluded that the age of the host, not the age of the muscle being transplanted dictated the degree of functional recovery[161]. More recent heterochronic parabiotic studies demonstrated that systemic factors produced in young mice are able to rejuvenate age progenitors[162]. In aged muscles, the expression of Notch in satellite cells remains unchanged. However, the induction of Delta in myofibers is no longer responsive to muscle injury, which results in reduced satellite cell proliferation

and impaired muscle regeneration[88]. Interestingly, forced activation of Notch signaling was able to restore regenerative potential to aged muscle[88].

Of note, a recent study showed that Wnt signaling also plays a role in regulating muscle regeneration. Activated satellite cells from old mice have higher Wnt signaling activity and a greater tendency to assume a fibrogenic fate *in vitro*[163]. Heterochronic parabiotic experiments revealed that activating Wnt signaling in young mice lead to increased fibrogenesis in regenerating skeletal muscle. Whereas, abrogation of Wnt signaling, by a secreted form of Frizzled, diminished fibrogenesis in aged regenerating muscle. These experiments lead credence to a circulating systemic factor regulating fibrogenesis in skeletal muscle[163]. Due to the above experiments the age-related decline in the regenerative capacity of skeletal muscle is predominantly attributed to systemic changes that negatively impact the ability of satellite cells to maintain and repair muscle tissue.

Significant alterations occur with age to the satellite cell niche as well as the surrounding muscle environment (local niche). Characteristics of the satellite cell niche that change with age include ECM remodelling and stiffness[164-167], endothelial cells of the vascular[168,169], cells of the immune system[170], and association with fibroblasts[171]. Currently changes in the satellite cell niche and the effects this imparts on resident satellite cells at a transcriptional level are unknown.

It is apparent that cell intrinsic changes that occur as satellite cells age are reversible by local or circulating systemic factors[162]. Identifying these factors is difficult due to the complexities of serum and the inability to isolate and maintain satellite cells in their niche *in vitro*. As a result, to identify the effect of the circulation and the niche one

must isolate and characterize satellite cells from different stages of development. Based on their transcription profiles experiments can then proceed in order to identify the mechanisms that directly regulate the myogenic capacity of satellite cells as a function of age.

## 1.4. The Canonical Wnt signalling pathway

Wnt proteins constitute a large family of cysteine rich secreted glycoproteins that play critical roles during embryonic development. The name derives from the first 2 members of the family to be discovered: int-1 (mouse) and wingless (Drosophila). Wnt signaling controls diverse biological processes, such as cell proliferation, cell fate determination, cell adhesion, cell polarity, cell morphology and tumor genesis (reviewed [172,173]). The Wnt family consists of 19 members, which upon binding with Fzd (Frizzled) receptors and low-density lipoprotein receptor-related protein (LRP5, LRP6) pairs elicit various intracellular responses. There are three classical Wnt pathways: the Wnt/β-catenin or canonical pathway[174,175] and two non-canonical pathways, the Wnt/PCP (planar cell polarity) pathway[176,177] and the Wnt/Calcium pathway[178].

In the Wnt/β-catenin pathway, the level and subcellular localization of β-catenin determine the ultimate output (**Fig. 5**). Cytoplasmic β-catenin associates with both transmembrane cadherins and cytoskeleton-bound α-catenin, promoting cell adhesion and regulating cell shape[179]. Cytoplasmic levels of β-catenin are maintained by constitutive degradation of excess β-catenin by GSK3-β. In the cytoplasm, GSK3-β and β-catenin complex with Axin and adenomatous polyposis coli (APC), leading to β-catenin phosphorylation by GSK3-β. Phosphorylated β-catenin is bound by the β-TrCP component of the ubiquitin ligase complex and degraded by the proteosome. GSK3-β is

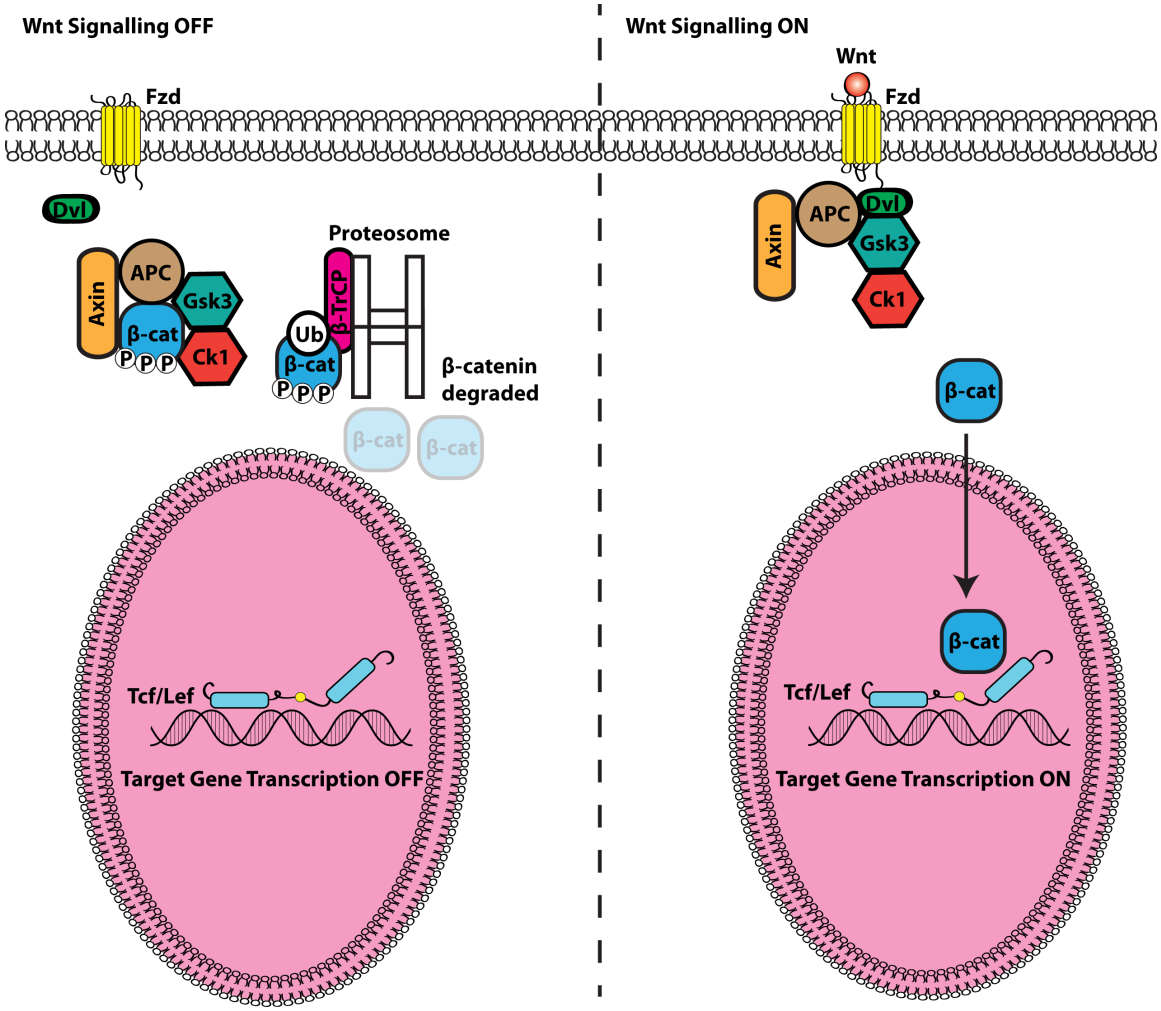
constitutively active but can be inhibited by Dishevelled (Dvl), which is activated by Wnt binding to a Frizzled and LRP receptor pair. Thus, when Wnt signaling is active, cytoplasmic β-catenin is stabilized and translocates into the nucleus to serve as a co-regulator for transcription factors (eg. Tcf/Lef), which in turn bind to target genes to modulate their expression[172]. Given this mode of action for the Wnt signalling pathway, mutations in different components can lead to both gain and loss of signalling activity, providing insight into the role Wnt/β-catenin signalling plays during development. Wnt3a is characterized as a canonical Wnt as it increases nuclear β-catenin levels and activates β-cateinin/Tcf reporters in multiple cell lines[180-182]. Wnt3a$^{-/-}$ embryos lack caudal somites, have a disrupted notochord, and fail to form a tailbud[44]. Activation of canonical Wnt signalling is known to play many roles in development and tissue homeostasis. However, our focus is on the role of canonical Wnt signaling with regards to pluripotency and early embryonic development.

*1.4.1. Canonical Wnt Signalling and Pluripotency*

Several studies indicated canonical Wnt signalling plays a role in maintaining pluripotency in both mouse and human ES cells. Activation of canonical Wnt signalling maintains pluripotency of mES cells when cultured on MEFs in the absence of LIF[183-185]. Wnt/β-catenin signalling can synergize with LIF/Stat3 to enhance pluripotency[186]. Furthermore, GSK-3β inhibitors (BIO and CHIR99021) greatly facilitate the establishment of de novo mES cell lines from C57BL/6 blastocysts[187-190].

However a role for Wnts to maintain pluripotency is not essential as β-catenin null ES cells are capable of generating chimeric embryos and Wnts in the absence of LIF and MEFs cannot maintain pluripotency[189,194,195]. These studies indicate that maintaining

**Figure 5. The Canonical Wnt Signalling Pathway.**
The canonical Wnt Signalling pathway is defined based on the activity of the intracellular protein β-catenin. In absence of a Wnt ligand, the destruction complex composed of glycogen synthase kinase 3-β (Gsk3-β), adenomatis polyposis coli (APC), Axin and casein kinase-1 (Ck-1) bind to cytoplasmic β-catenin. When bound by the destruction complex β-catenin is phosphorylated by Gsk3-β on key serine residues (Ser-33 and Ser-37). When phosphorylated β-catenin is recognized by β-transducing repeat containing protein (β-TrCP) and subsequently targeted for ubiquitin-mediated destruction. In the absence of a Wnt ligand, Tcf/Lef family transcription factors bind and modulate gene transcription. Wnt ligands bind Frizzled (Fzd) receptors resulting in the membrane localization of the destruction complex by Dishevelled (Dvl). Cytoplasmic β-catenin is therefore not phosphorylated and can migrate into the nucleus where it complexes with Tcf/Lef transcription factors to alter the modulation of gene transcription.


Wnt Signalling OFF
Wnt Signalling ON
Wnt
Fzd
Fzd
Dvl
Axin
APC
Gsk3
β-cat
Ck1
P P P
Proteosome
β-TrCP
Ub
β-cat
P P P
β-catenin degraded
β-cat
β-cat
Tcf/Lef
Target Gene Transcription OFF
Axin
APC
Dvl
Gsk3
Ck1
β-cat
β-cat
Tcf/Lef
Target Gene Transcription ON

the pluripotent state is not the primary purpose of Wnts[186,191,192]. The Tcf3 transcription factor, a downstream effector of canonical Wnt signalling, is able to bind the promoters of Oct4, Sox2, and Nanog[193]. Nanog has also been directly induced through activated Wnt signalling[194,195]. It was later discovered that Tcf3 represses the core circuitry of pluripotency. Yet upon activation of Wnt signalling and β-catenin nuclear translocation, repression is released and self-renewal is promoted[196-198]. Recently it was discovered Wnt3a stimulates self-renewal via a two-pronged mechanism involving the suppression of Tcf3 repression directly through Tcf3-β-catenin and indirectly through Tcf1 mechanisms[199].

Although Tcf transcription factors currently are the primary mediators of Wnt signaling, recent evidence indicates that β-catenin can directly interact with Oct4 and regulate pluripotency[194,200]. β-catenin is also functional in the cytoplasm, it is an integral component of adherens junctions involved in mediating cell-cell contacts via linking cadherins to α-catenin and the actin cytoskeleton[201]. Of note recent evidence describes a critical role for β-catenin dependent regulation of cadherin-mediated cell adhesion in the formation in vitro of definitive endoderm and the neuroepithelium[191].

As evidenced above, Wnts may act via multiple mechanisms in regulating the pluripotent state. To complicate matters Wnts likely cross-talk with other signalling molecules including BMPs and FGFs present in undifferentiated mES cells[15,202]. Thus, clarification on the roles played by canonical Wnt signalling in the undifferentiated state of mES cells remains an active area of research.

### *1.4.2. Wnt Signalling and Lineage Specification*

Numerous Wnts, Fzd receptors as well as soluble inhibitors (Dickopff (Dkk) and soluble frizzled related proteins (sFrp)) of the Wnt signalling pathway are expressed in the blastocyst and surrounding uterus during early embryonic development[203-207]. Additionally, active β-catenin has been detected in the nuclei of mouse blastocysts by immunofluorescence[203,208]. This data suggests that Wnts play a role in early embryonic development. What that specific function(s) is, either in maintaining pluripotency, guiding implantation, or in cell fate specification remains to be determined.

Irrespective of the role of Wnt signalling prior to implantation, studies have shown a role for Wnt signalling in early proximal distal patterning and anterior visceral endoderm (AVE) formation[209]. Analysis of adenomatous polyposis coli (APC) mutant mice ($Apc^{min/min}$), in which β-catenin is constitutively active, fail to form a pro-amniotic cavity or primitive streak and are embryonic lethal[210]. Incidentally, $Apc^{min/min}$ ES cells were unable to form teratomas[211]. $Apc^{min/min}$ mES cells had severe differentiation defects and maintained an undifferentiated appearance for considerable longer then wildtype and heterozygous controls. When they did differentiate $Apc^{min/min}$ ES cells gave rise to large amounts of visceral and parietal endoderm[211]. Heterozygous mice with constitutively active β-catenin, obtained through the removal of critical GSK3β phosphorylation sites in exon 3 (β-catΔex3/+), develop premature EMT, activate Brachyury in the epiblast and develop thickened visceral endoderm. Upon complementation assays with wild-type mice, mutant cells preferentially colonized the visceral endoderm, mesoderm and extraembryonic tissues[212].

In contrast, β-catenin knockout mice do not develop mesoderm and fail to gastrulate[192]. Mutant embryos appear phenotypically normal at E5.5, yet by E6.5 they

develop a visceral endoderm phenotype characterized by a lack of proximal-distal regionalization. In this case the distal visceral endoderm (dVE) forms yet fails to move proximally, a critical step in the creation of the anterior-posterior (AP) axis. Consequently, mutant mice fail to properly form the AP axis[192]. In these mice, β-catenin further regulates the molecular patterning of the dVE as specific markers such as Hex are not expressed. Failure to form the AP axis is explained by analyzing Dkk1 expression, which acts as an attractive guidance cue for the dVE. In contrast Wnt antagonism is required to shift the dVE proximally in order to properly form the AVE and the AP axis.

Following the migration of the dVE to the anterior side of the embryo, Wnt3 is initially expressed in posterior visceral endoderm and later the posterior epiblast to initiate the formation of the primitive streak[213]. To monitor the activation of canonical Wnt signaling many reporters have been created including Tcf/Lef-LacZ, Axin$^{LacZ}$, Lef-eGFP, and Tcf/Lef:H2B-GFP[214-217]. The earliest activation of Tcf based reporters occurs in the visceral endoderm and the posterior proximal epiblast, while the Axin$^{LacZ}$ reporter is activated earlier in the epiblast and primitive endoderm of the developing blastocyst[190,217].

Administering Wnts to differentiating mES cells must be done in both a dose and temporal dependent manner. Interestingly, mES cells in the presence of activated Wnt signalling can self-organize and undergo axis-formation in a fashion akin to embryonic development[218]. Induction of canonical Wnt signalling, typically via Wnt3a, has been used extensively to mimic primitive streak formation[219-221]. As is the case in vivo, treatment of differentiated mES cells with Wnt3a inhibits neural[222,223] and promotes mesodermal[224,225] lineages. Other lineages induced by canonical Wnt signalling include

the hematopoietic[226-228], cardiac[229,230], and osteogenic[231] lineages. These results reflect embryonic development and suggest that Wnts play a prominent role in the formation of mesodermal lineages.

## 1.5 Hypothesis and Specific Aims

### *1.5.1. A Role for Wnt3a in Pluripotency and Early Differentiation of Mouse Embryonic Stem Cells.*

Addition of Wnt3a following several days of mES cell differentiation activates T (Brachyury) and is implicated in early mesendoderm development. However in the presence of LIF, Wnt3a promotes the undifferentiated state of mES cells by regulating Tcf3 on the promoters of core pluripotency genes Oct4, Sox2 and Nanog. These dual roles for Wnt3a in both promoting and inhibiting differentiation of mES cells lead us to hypothesize that Wnt3a may play a role in cell specification while cells remained in the undifferentiated state. If indeed Wnt3a primes mES cells into distinct lineages, which lineages are they and how is Wnt3a accomplishing this role? To address these questions we created Wnt3a overexpressing mES cell lines and characterized them both in the undifferentiated and differentiated states.

### *1.5.2. The Effect of Age on Satellite Cell Function*

As an organism ages skeletal muscle undergoes a reduction in mass, contractile efficiency, and the ability to regenerate post trauma. Sarcopenia is the term used to describe these age related effects on skeletal muscle. Sarcopenia is typically attributed to variations in the local and systemic environment surrounding skeletal muscle. We hypothesized that in addition to local and systemic effects, intrinsic age-related

differences exist in satellite cells and that these intrinsic differences contribute to functional deficiencies observed in aged skeletal muscle. To discover age related intrinsic properties of satellite cells we propose to FACS sort pure populations of satellite cells from mice at different stages of development. Subsequently we will classify each satellite cell age group according to their genome wide mRNA expression profiles and assess their ability to differentiate in vitro and take part in muscle regeneration *in vivo*. In this fashion we aim to identify intrinsic mechanisms that contribute to age-related differences in satellite cell function.

# Chapter 2 - Wnt3a induces a primitive endoderm metastable state in mouse embryonic stem cells

## 2.1 Abstract

Activation of the canonical Wnt signalling pathway synergizes with leukemia inhibitory factor (LIF) to maintain pluripotency of mouse embryonic stem cells (mESCs). However, in the absence of LIF, Wnt signalling is unable to maintain ES cells in the undifferentiated state. To investigate the role of canonical Wnt signalling in pluripotency and lineage specification, we expressed Wnt3a in mESCs and characterized them in growth and differentiation. We found that Wnt3a induced the formation of a reversible metastable primitive endoderm state in mESC. Upon subsequent differentiation, Wnt3a stimulated mESCs gave rise to large quantities of visceral endoderm. Furthermore, we determined that the ability of Wnt3a to induce a metastable primitive endoderm state was mediated by *Tbx3*. Our data demonstrates a specific role for canonical Wnt signaling in promoting pluripotency while at the same time priming cells for subsequent differentiation into the primitive endoderm lineage.

## 2.2 Introduction

Two major events define early mammalian embryogenesis. First, the segregation of the morula into the trophoblast and inner cell mass (ICM) and second gastrulation, which is the formation of the three primary germ layers. ES cells are derived from the ICM of a pre-implantation blastocyst and thus exhibit a gene expression profile along with functional properties characteristic of pluripotent cells. ES cells retain the capacity to repopulate an embryo by contributing to all somatic cell lineages, including the germline[232]. For these reasons ES cells are characterized as naïve pluripotent cells. Recent evidence implies ES cells are heterogeneous and consist of subpopulations with

distinct functional and epigenetic states [29,30]. These metastable subpopulations share similarities to three different transient cell populations in the mouse embryo: the ICM [27,28,30], the epiblast cells of post-implantation embryos [29] and the primitive endoderm [33]. Moreover, the particular metastable state is maintained in a dynamic equilibrium allowing for the inter-conversion of the different populations. The molecular mechanisms that direct the generation of specific ESC metastable states have not been defined.

LIF/Stat-3 signalling is the predominant pathway involved in the maintenance of the pluripotent state in mESCs [233]. Currently the signaling pathways that regulate the epiblast and primitive endoderm metastable states remain largely unknown. The Wnt family of secreted glycoproteins plays diverse and critical roles in the proper specification and formation of organs and tissues [174,234]. Genetic studies implicate Wnt signalling in embryonic axis formation, mesodermal induction and carcinogenesis[224,234,235]. Previous experiments implicate Wnts in maintaining the potency of both embryonic and adult stem cells [172,174,236,237]. Numerous studies demonstrate that activating canonical Wnt signalling promotes mES cells to remain pluripotent[183,211]. It is therefore not surprising that the main components of the canonical Wnt pathway are actively expressed in ES cells in the undifferentiated state[238].

Under self-renewing conditions, Wnts synergize with LIF to maintain pluripotency [186,233]. In addition, core pluripotency factors (*Oct4, Sox2* and *Nanog*) are regulated by canonical Wnt signalling [194,239,240]. In fact Tcf3 (Tcf7L1), a key downstream effector of β-catenin mediated canonical Wnt signalling, binds to the promoters of *Oct4*, *Sox2*, and *Nanog* and limits their expression [193,197]. Consistent with the ability of *Tcf3* to repress pluripotency, *Tcf3*$^{-/-}$ mouse ESCs are delayed in their differentiation as embryoid bodies

and are able to self-renew even upon LIF withdrawal [198].

The canonical Wnt pathway is regulated by the expression level and subcellular localization of β-catenin. Intracellular β-catenin is phosphorylated by *GSK3-β* when bound in a complex with *Axin* and adenomatous polyposis coli (*APC*). Phosphorylated β-catenin is subsequently degraded by the proteasome [172]. Upon binding of Wnts to Frizzled and Lrp5/6 cell surface receptors, an intracellular cascade in which Dishevelled (*Dvl*) inhibits the phosphorylation of β-catenin by *GSK3-β* is activated [241,242]. Cytoplasmic β-catenin is stabilized and translocates into the nucleus to serve as a co-regulator for Tcf/Lef transcription factors, which in turn bind to target genes to modulate their expression [175].

Numerous experiments have identified Wnt3a as a canonical Wnt and demonstrate that it can functionally affect pluripotency [190,230,243]. Incidentally, a mutation in the β-catenin gene leads to a defect in anterior/posterior axis formation and the subsequent loss of mesoderm and head structures[192]. Wnt3a is expressed in the primitive streak of the gastrulating mouse embryo and is required for paraxial mesoderm development[243]. $Wnt3a^{-/-}$ mice lack a tail bud, display a disrupted notochord, lack caudal somites and have cells that adopt ectopic neural fates[244].

In embryonic development, the earliest activation of Wnt/β-catenin targets occurs in the visceral endoderm as evidenced by Tcf based reporters [215]. Many components of the Wnt signalling pathway are expressed in the blastocyst and in the uterine wall prior to implantation [203,204,207]. Although the $Axin^{LacZ}$ reporter is activated in the epiblast and primitive endoderm of the developing blastocyst [190,217], opposing views persist regarding

whether the Wnt/β-catenin pathway plays a functional role within the pre-implantation blastocyst [208,209,212].

The ability of Wnts to maintain pluripotency in mESCs is dependent on the LIF pathway as constitutively active β-catenin alone is unable to maintain self-renewal [186]. This is a critical point as it implies that the predominant role of the canonical Wnt signalling pathway is insufficient to maintain the pluripotent state of embryonic stem cells. This further raises the issue of whether canonical Wnt signalling has other functions in undifferentiated mESCs.

We investigated how canonical Wnt signalling regulates mESC pluripotency and cell fate commitment. We determined that canonical Wnt signalling reinforces the undifferentiated state by promoting pluripotency genes while at the same time activating a subset of primitive endoderm genes. Notably, activated canonical Wnt signalling induces a metastable ESC state in which primitive endoderm genes are up regulated. This subtype of primitive endoderm ESC gives rise to both a Nanog-GFP high population as well as a Nanog-GFP low population. The induction of primitive endoderm genes in the undifferentiated state comes at the expense of the neuroectodermal and mesendodermal potential. Wnt3a-expressing ESCs display a predisposition for the primitive endoderm lineage and as a result display pronounced visceral endoderm potential following differentiation. We further found that Wnt3a induction of the primitive endoderm metastable state is mediated by Tbx3. Together, our experiments reveal important insights into the early regulation of ESC pluripotency and commitment by canonical Wnt signalling.

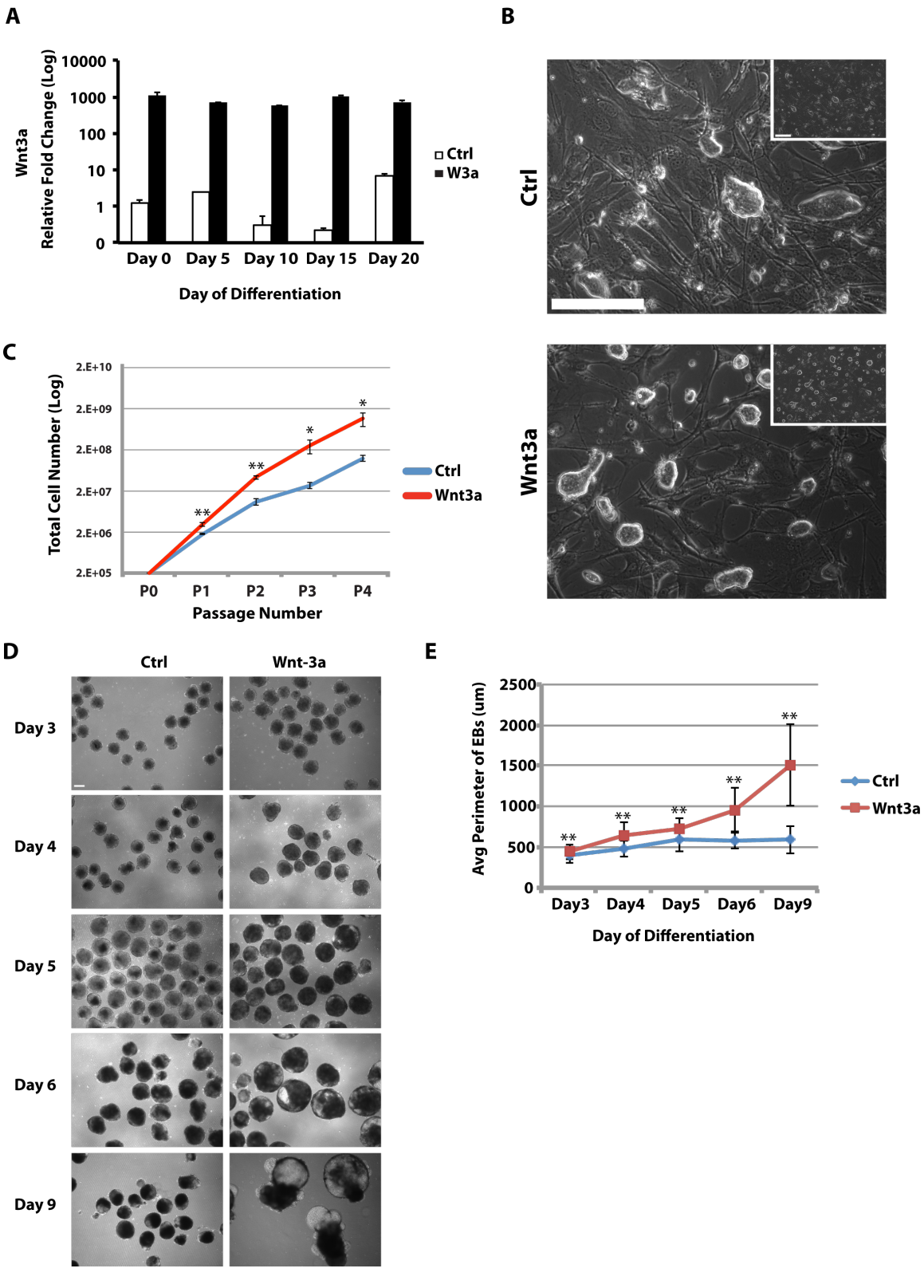
## 2.3 Results

*2.3.1 Wnt3a Affects the Morphology and Growth Rate of Undifferentiated and Differentiating mES Cells*

In order to better understand how Wnt signalling regulates pluripotency and early lineage specification, we created mES cell lines that constitutively express Wnt3a. In this fashion we would ensure constant and maintained activation of canonical Wnt signalling, an issue that is important when dealing with morphogens having different effects based on concentration. Mouse ES cell lines were infected with a self-inactivating retrovirus containing an EF1α promoter to drive the expression of Wnt3a followed by an IRES-Puro for selection (**Fig. S1**). We confirmed the constitutive expression of Wnt3a in undifferentiated D3 mES cells as well as during differentiation (**Fig. 1A**). To inhibit canonical Wnt signalling we treated wild type cells with Niclosamide[245] and Quercetin[246], inhibitors of Wnt/Frz and β-catenin/Tcf interactions respectively. Neither Niclosamide or Quercetin inhibited the gene expression of canonical Wnt signalling components in undifferentiated mES cells (Data not shown). Elevated levels of Wnt3a mRNA with respect to endogenous Wnt3a expression in mock control cell lines illustrate the effective and sustained over expression of Wnt3a in mES cells (**Fig. 1A**).

As previously described[247,248], we observed in undifferentiated mES cells expressing Wnt3a a significant increase in the rate of proliferation along with an increase in compacted colony morphology (**Fig. 1B and C**). However, a more striking phenotype was observed when Wnt3a-expressing mES cells were differentiated as embryoid bodies. These changes include an increase in embryoid body size relative to mock-infected control cells and the development of large cysts after 6 days of differentiation (**Fig. 1D and E**). Taken together, these results support the notion that canonical Wnt signalling

**Figure 1. Wnt3a overexpression elicits increased growth rate and the formation of cystic embryoid bodies following differentiation of mES cells.**
(**A**) qPCR expression values for Wnt3a from retrovirally infected D3 mES cells expressing Wnt3a (black) or infected with a mock control (white) over a 20 day time course. (**B**) Bright field images of undifferentiated control (upper panel) and Wnt3a infected (lower panel) mouse ES cells grown on MEFs in the presence of LIF. 100x magnification, inset 50x magnification, scale bar denotes 100 μm. **(C**) Growth kinetics of undifferentiated D3 mES cells over 4 passages either expressing Wnt3a or mock-infected control. (**D**) EB morphology and size is altered upon continued Wnt3a expression during 9 days of differentiation in the absence of LIF. Scale bar designates 100 μm. **E**) Quantification of EB perimeter size during differentiation. Data are presented as the mean ± SEM from biological triplicate experiments in D3 mESC lines, normalized to mock infected controls. P = paired, two-tailed Student's *t* –test where * < 0.05 and ** < 0.01. Abbreviations: mESC, murine embryonic stem cells; MEF, mouse embryonic fibroblasts; EBs, embryoid bodies.


A
Wnt3a
Relative Fold Change (Log)
10000
1000
100
10
1
0
Day 0
Day 5
Day 10
Day 15
Day 20
Day of Differentiation
Ctrl
W3a
B
Ctrl
Wnt3a
C
Total Cell Number (Log)
2.E+10
2.E+09
2.E+08
2.E+07
2.E+06
2.E+05
P0
P1
P2
P3
P4
Passage Number
Ctrl
Wnt3a
D
Ctrl
Wnt-3a
Day 3
Day 4
Day 5
Day 6
Day 9
E
Avg Perimeter of EBs (um)
2500
2000
1500
1000
500
0
Day3
Day4
Day5
Day6
Day9
Day of Differentiation
Ctrl
Wnt3a

does not cause overt precocious differentiation in undifferentiated mES cells. However, in the absence of LIF, Wnt3a perturbs the differentiation of mES cells and promotes the formation of cystic embryoid bodies.

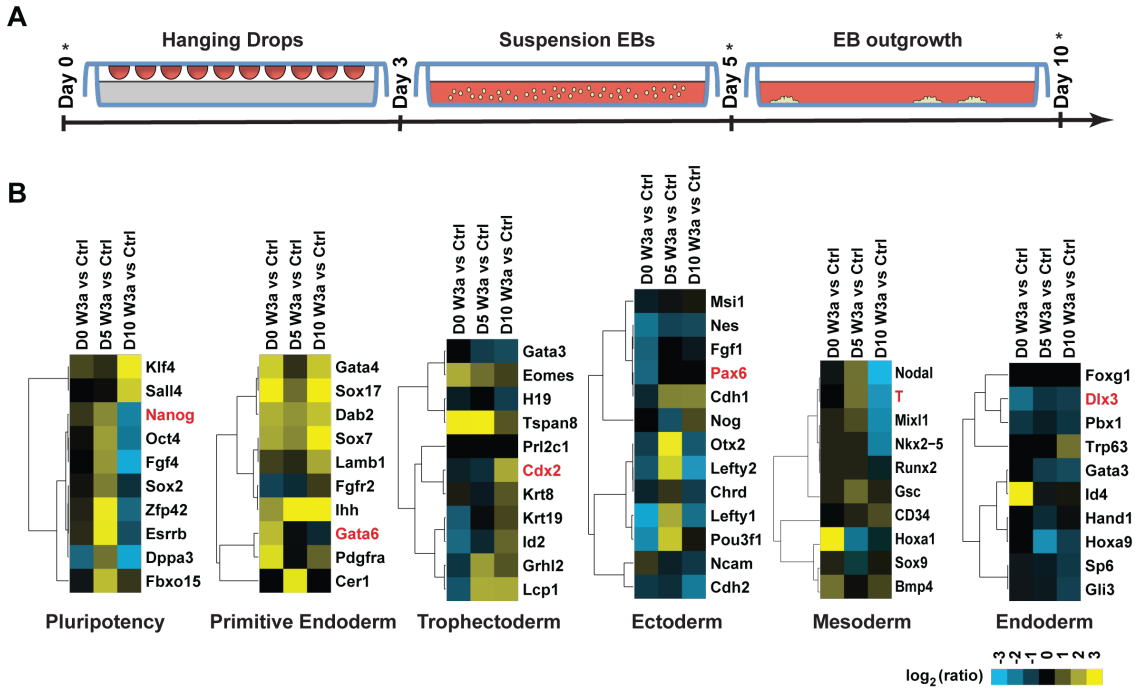
*2.3.2 Wnt3a Alters the Global Gene Expression Profile of Blastocyst Cell Types and the Primary Germ Layers.*

To gain insight into how Wnt3a affects pluripotency and lineage induction of ES cells, we conducted RNA-seq from total mRNA isolated at the undifferentiated state (Day 0), a mid differentiation time point (Day 5) and a late stage differentiation time point (Day 10) (**Fig. 2A**). In addition to our retroviral cell lines we treated parental D3 ES cells with 250 ng/ml of recombinant Wnt3a for three passages followed by 10 days of differentiation. Recombinant Wnt3a was present at the initiation of differentiation and further supplemented on Days 3, 5 and 7. As quality controls, we confirmed RNA-seq read quality and the distribution of reads prior to further gene analysis (**Fig. S2**). In addition, the transcriptional activation of canonical Wnt targets was confirmed in recombinant and retroviral cell lines (**Fig. S3**).

As transcription factors (TFs) are essential for defining cell-type specific phenotypes, we compiled panels of known TFs along with additional gene markers specific for each cellular lineage in question[17,249,250]. We chose to focus on core lineages involved in the formation of the blastocyst (Trophoblast, Primitive Endoderm and Epiblast) and the primary germ layers (Ectoderm, Mesoderm, Endoderm). In this fashion we could determine the broad effects of Wnt3a on the induction of a specific lineage.

We examined the RNA-seq data according to these core groups of genes to determine which cell types are being induced/reduced upon Wnt3a expression (**Fig. 2B**).

**Figure 2. Primitive endoderm genes induced by constitutive Wnt3a overexpression.** **(A)** RNAseq was conducted with total RNA collected on Days 0, 5, 10 from D3 mESCs **(B)** Global ratio heat maps representing the difference in expression between Wnt3a overexpressed and mock-infected controls for blastocyst and germ layer lineages are shown. Genes in red were chosen as positive controls for each lineage. Yellow represents genes that increase in expression relative to the control while blue represents a decrease. (**C**) qPCR validation was conducted for members of each lineage. Data are presented as the mean ± SEM of 3 biological samples originating from D3, J1 and R1 mESC lines, normalized to GAPDH and relative to mock infected controls. p = paired, two-tailed Student's *t* –test where * < 0.05 and ** < 0.01. Global *t*-tests were conducted per gene over the differentiation time course and significance accompanies the gene name listed under each lineage heading. * = $p < 0.05$ and ** = $p < 0.01$


A
Hanging Drops
Suspension EBs
EB outgrowth
Day 0 *
Day 3
Day 5 *
Day 10 *
B
D0 W3a vs Ctrl
D5 W3a vs Ctrl
D10 W3a vs Ctrl
Klf4
Sall4
Nanog
Oct4
Fgf4
Sox2
Zfp42
Esrrb
Dppa3
Fbxo15
Gata4
Sox17
Dab2
Sox7
Lamb1
Fgfr2
Ihh
Gata6
Pdgfra
Cer1
Gata3
Eomes
H19
Tspan8
Prl2c1
Cdx2
Krt8
Krt19
Id2
Grhl2
Lcp1
Msi1
Nes
Fgf1
Pax6
Cdh1
Nog
Otx2
Lefty2
Chrd
Lefty1
Pou3f1
Ncam
Cdh2
Nodal
T
Mixl1
Nkx2-5
Runx2
Gsc
CD34
Hoxa1
Sox9
Bmp4
Foxg1
Dlx3
Pbx1
Trp63
Gata3
Id4
Hand1
Hoxa9
Sp6
Gli3
Pluripotency
Primitive Endoderm
Trophectoderm
Ectoderm
Mesoderm
Endoderm
log2 (ratio)
-3 -2 -1 0 1 2 3
C
Pluripotency
Oct4 Sox2* Nanog
Primitive Endoderm
Sox17* Gata4* Gata6*
Trophectoderm
Cdx2** Eomes Id2
Ectoderm
Pax6** Nes** Ncam*
Mesoderm
T Mixl1 Gsc*
Endoderm
Gata3** Dlx3** Pbx1**
Relative Expression
Mock Wnt3a
Day 0
Day 5
Day 10

mRNA levels of key pluripotency factors *Sox2*, and *Nanog* were increased in Wnt3a-overexpressing undifferentiated ES cells. In contrast, the majority of the other pluripotency markers remained either unchanged, or in the case of *Dppa3* (Stella), were reduced. Upon removal of LIF and differentiation as embryoid bodies, pluripotency genes were up regulated on day 5 and down regulated by day 10.

The trophectoderm panel of genes displayed mixed induction and repression upon Wnt3a over-expression in the undifferentiated state, a trend that was maintained upon differentiation. Of the three blastocyst cell types, the primitive endoderm panel of lineage markers displayed a striking increase in expression levels for core transcription factors in the undifferentiated state (*Gata4* - 6 fold, *Gata6* - 5 fold, *Sox17* - 20 fold and *Sox7* - 5 fold by qPCR). Following differentiation, *Gata4*, *Sox7*, and *Sox17* were all strongly induced in Wnt3a expressing cell lines.

A consequence of activating the Wnt signalling pathway during ES cell differentiation is the induction of mesoderm and the reduction of neural ectoderm[218,219,222,223,251]. Consistent with these reports our qPCR data shows that Pax6, an early marker for neural development, had expression levels reduced by 70% following Wnt3a treatment in undifferentiated ES cells (**Fig. 2B and C**). Furthermore, 13/13 (Day 0), 8/13 (Day 5), and 11/13 (Day 10) ectodermal markers displayed reduced or unchanged expression levels in Wnt3a expressing cells. Endoderm markers as a whole were all either down regulated or unchanged in the undifferentiated (9/10) and differentiated (10/10 Day 5, 9/10 Day 10) states. Mesodermal genes displayed weak induction in undifferentiated cells (2/10) and during early differentiation (4/10), while on day 10 core mesodermal genes *T*, *Mixl1*, and *Nodal* were all strongly inhibited (>3 fold

down regulation). In concordance with our RNA-seq results, qPCR validation of RNA levels for core lineage markers followed similar patterns (**Fig. 2C**).

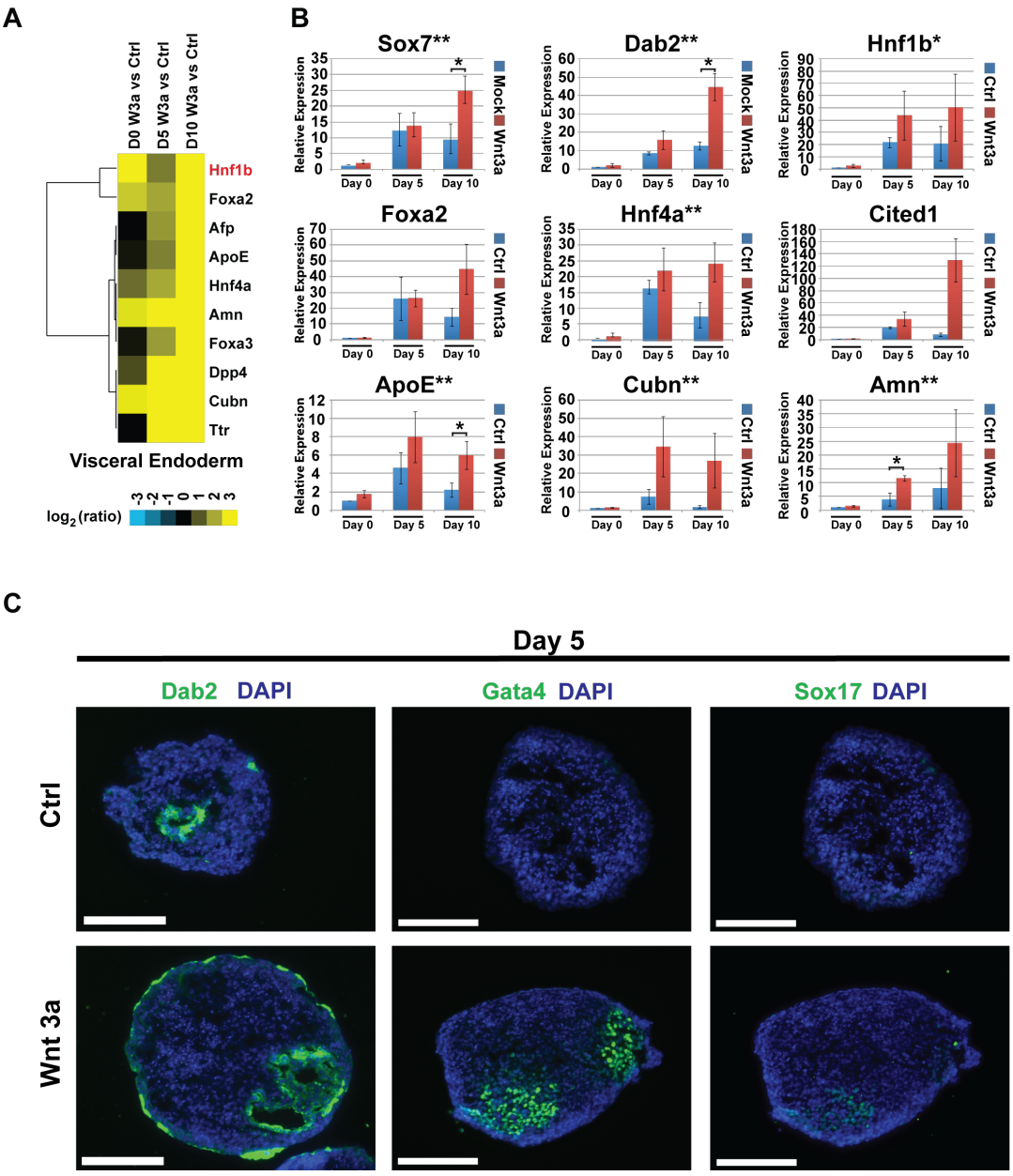
Recombinant Wnt3a affected differentiation in a similar yet subdued fashion when compared with retroviral over expression of Wnt3a. These changes include the induction of core pluripotency and primitive endoderm genes in the undifferentiated state while primitive and visceral endoderm markers were induced at later stages of differentiation (**Fig. S4**). Based on the above evidence, we conclude that Wnt3a alters the transcriptional profile of ES cells prior to differentiation and substantially during differentiation to promote primitive endoderm genes and prolong the expression of pluripotency markers.

### *2.3.3 Differentiation mES Cells with Wnt3a Enhances the Formation of Visceral Endoderm*

During blastocyst development the primitive endoderm is localized to a layer of cells that cover the epiblast and separates it from the blastocoel cavity. Following blastocyst implantation, the primitive endoderm will give rise to both the parietal endoderm to line the outer trophectoderm and the visceral endoderm to cover the epiblast[252]. Since Wnt3a overexpression induces up-regulation of primitive endoderm genes and the formation of cystic EBs, we next tested whether Wnt3a promotes visceral endoderm.

To assess visceral endoderm in our cultures, we examined the expression levels of genes enriched in visceral endoderm at the RNA and protein level. Primary visceral endoderm cell surface proteins *Amn*, *Cubn* and *ApoE* are all significantly up regulated based on our RNA-seq results (**Fig. 3A**) as well as by qPCR (**Fig. 3B**). *Dab2* a cell

**Figure 3. Continued Wnt3a expression during differentiation induces formation of visceral endoderm in mESCs.**
**(A)** Global ratio heat map representing the difference in mRNA expression between Wnt3a overexpressed and mock-infected controls for biomarkers of visceral endoderm. Genes in red were chosen as positive controls for each lineage. Yellow represents genes that increase in expression relative to the control while blue represents a decrease. **(B)** qPCR validation of visceral endoderm markers. Data are presented as the mean ± SEM of 3 biological samples originating from R1, J1 and D3 mESC lines, normalized to GAPDH and relative to mock infected controls. p = paired, two-tailed Student's *t* –test where * < 0.05 and ** < 0.01. Global *t*-tests were conducted per gene over the differentiation time course and significance accompanies the gene name listed under each lineage heading. * = $p < 0.05$ and ** = $p < 0.01$. **(C)** Immunofluorescence analysis for biomarkers of primitive and visceral endoderm in mES cells following 5 days of differentiation as embryoid bodies. Scale bars denote 100 μm.


A
D0 W3a vs Ctrl
D5 W3a vs Ctrl
D10 W3a vs Ctrl
Hnf1b
Foxa2
Afp
ApoE
Hnf4a
Amn
Foxa3
Dpp4
Cubn
Ttr
Visceral Endoderm
log2(ratio)
-3 -2 -1 0 1 2 3
B
Sox7**
Dab2**
Hnf1b*
Foxa2
Hnf4a**
Cited1
ApoE**
Cubn**
Amn**
Relative Expression
Day 0
Day 5
Day 10
Mock
Wnt3a
Ctrl
C
Day 5
Dab2 DAPI
Gata4 DAPI
Sox17 DAPI
Ctrl
Wnt 3a

membrane adapter protein along with the transcription factors *Sox7*, *Hnf4a*, *Hnf1b*, and *Foxa2* are all up regulated in Wnt3a expressing mESC lines.

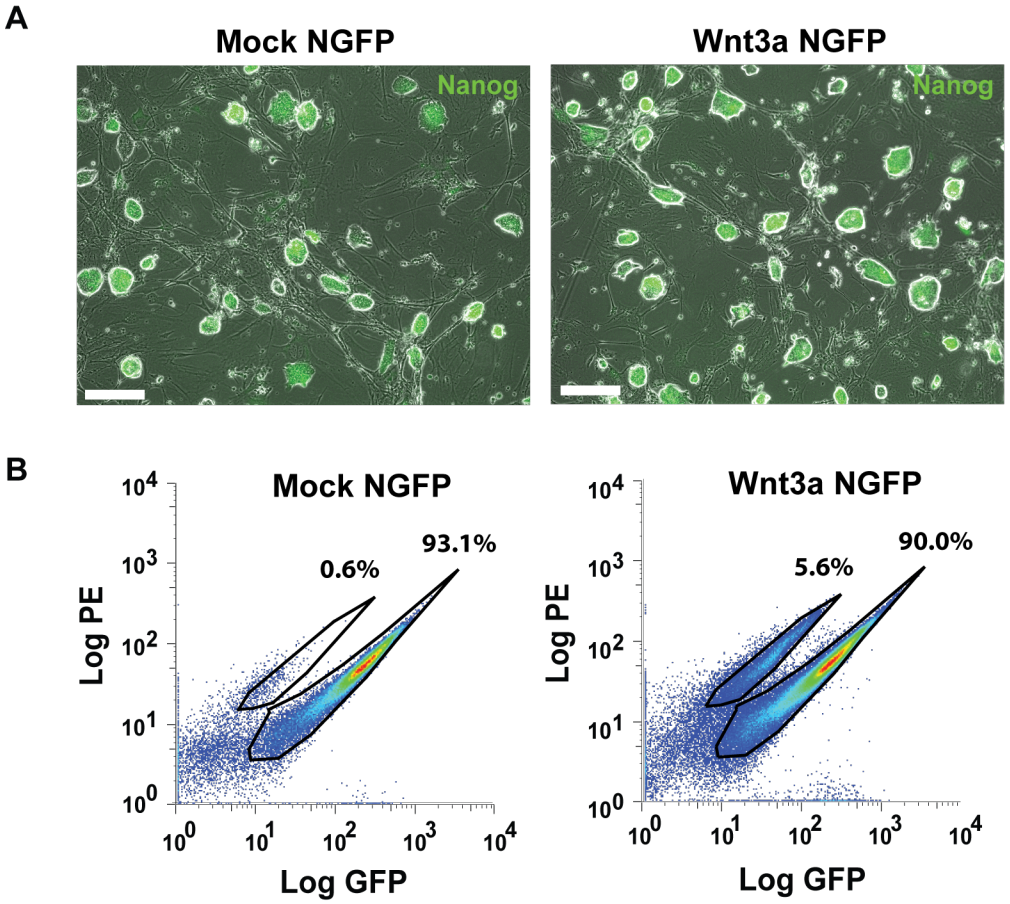
We confirmed the induction of visceral endoderm markers by immunofluorescence in differentiated embryoid bodies. Following 5 days of differentiation we observed an increase in Dab2, Sox17 and Gata4 positive cells in Wnt3a over-expressing EBs (**Fig. 3C**). In the presence of continued Wnt3a expression, the gene expression profile and protein complement of differentiating mES cells is enriched for primitive and visceral endoderm genes. Taken together these results indicate that upon differentiation the expression of Wnt3a directs cells into the visceral endoderm lineage.

### *2.3.4 Wnt3a Perturbs the Pluripotent Ground State of mES Cells by Inducing a Primitive Endoderm Metastable State*

Recent experiments demonstrate mES cells exist in a metastable state that fluctuates between ICM, epiblast and primitive endoderm-like states[29,33]. This equilibrium is maintained in proliferating ES cultures in the presence of LIF and can be characterized by the expression of either Nanog[27], Rex1[30], Stella[29], Hex[33] or Pecam1[28]. Our results demonstrate that Wnt3a can induce primitive endoderm genes in undifferentiated mES cells suggesting that the primitive endoderm-like metastable state is being activated.

To investigate whether Wnt3a affects the metastable state of mES cells, we analyzed the expression of GFP from a Nanog-GFP reporter cell line (**Fig. 4A**). Based on previous Nanog-GFP reporters there exist both high and low GFP populations while mES cells are maintained in the undifferentiated state[27]. Following retroviral infection of Wnt3a, we observed the induction of the GFP low population by FACS analysis. (**Fig.**

**Figure 4. Wnt3a alters the metastable state of mES cells by promoting the formation of primitive endoderm cells.**
**(A)** Nanog-GFP cells infected with Wnt3a or a mock control retrovirus continue to express high levels of GFP and maintain an undifferentiated phenotype. **(B)** FACS analysis of Nanog-GFP cells identifies a new population of GFP-low cells. Y-axis represents the Log of fluorescence in the phycoerythrin (PE) channel, while the X-axis represents the Log of green fluorescent protein (GFP). **(C)** qPCR analysis comparing the Nanog-GFP Wnt3a expressing cell fractions (GFP-high, GFP-low and unsorted) identifies a bias for genes specific to the primitive endoderm in the GFP-low fraction. Data are presented as the mean ± SEM of 3 biological triplicate samples in Nanog-GFP Wnt3a cell lines, normalized to GAPDH and relative to unsorted Nanog-GFP Wnt3a cells. P = paired two-tailed Student's *t* –test where * < 0.05 and ** < 0.01.


A
Mock NGFP
Nanog
Wnt3a NGFP
Nanog
B
Mock NGFP
0.6%
93.1%
Log PE
Log GFP
Wnt3a NGFP
5.6%
90.0%
Log PE
Log GFP
C
Unsorted
NanogGFP High
NanogGFP Low
Relative Expression
25
20
15
10
5
0
Oct4
Nanog
Pax6
T
Gata3
Tbx3
Gata4
Gata6
Sox17
Sox7
Hnf1b
ApoE
Cubn
Dab2

**4B**). During preimplantation development both *Nanog* and *Gata6* are co-expressed in cells of the morula. As the blastocyst matures *Nanog* and *Gata6* become mutually exclusive and specify cells of the epiblast and primitive endoderm respectively[12,253].

Since our results demonstrate an increase in primitive endoderm genes upon Wnt3a induction and previous reports dealing with the Nanog-GFP populations suggests the Nanog-GFP low cells express primitive endoderm genes[23], we speculated that Wnt3a may be inducing a population akin to the primitive endoderm.

To confirm the identity of both Nanog-GFP populations (high/low), they were FACs sorted and analyzed by qPCR. In contrast to Nanog-GFP high cells the Nanog-GFP low population displayed higher levels of primitive endoderm genes Gata4, Gata6, Sox17, Sox7, Cubn and Dab2 as well as increases in mesendoderm genes T, and Gata3 (**Fig. 4C**). A cell that is maintained in a transcriptionally poised state, often associated with epigenetic marks present at the promoters of lineage specific genes, is regarded as being in a lineage-primed state[254]. Up regulating primitive endoderm genes while maintaining, in a subset of cells, the expression of pluripotency genes suggests lineage priming. The question remains whether Wnt3a permanently enforces the primitive endoderm program on Nanog-GFP low cells and prevents their return to a Nanog-GFP high state.

Previous reports indicate that the activation of Sox17 in the primitive endoderm commits ES cells to differentiation[26]. As Sox17 is up regulated in the Nanog-GFP low population we questioned whether these cells were irreversibly committed to differentiation. Sorted Nanog-GFP Wnt3a high or low cells were cultured in the undifferentiated state for 2-6 days to determine their ability to interconvert. Nanog-GFP

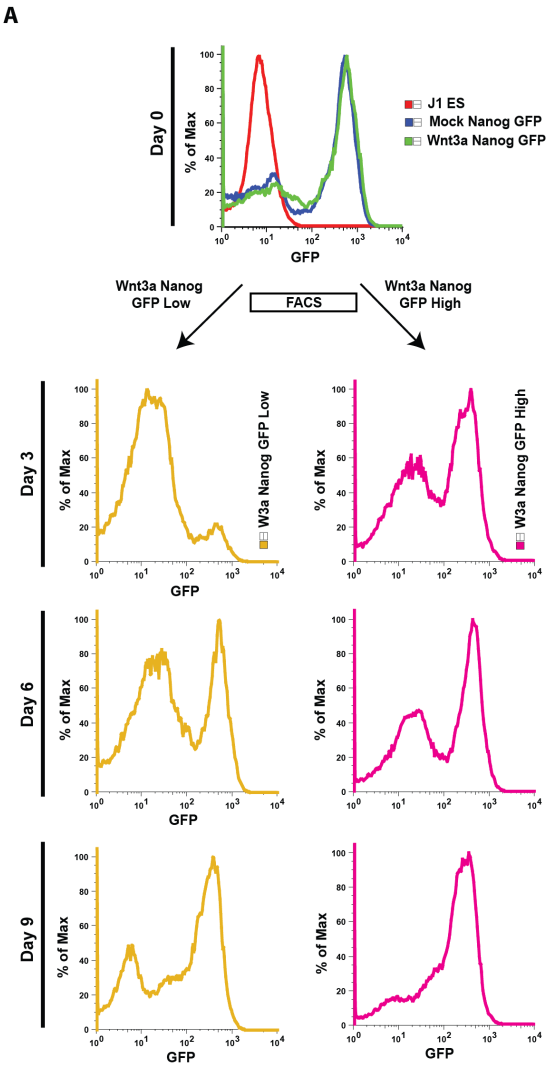
high cells maintain GFP expression while regenerating a GFP low population (**Fig. 5A**). Since previous reports indicate Sox17 activation in ES cells predisposes them for irreversible differentiation, we next tested whether the Nanog-GFP low cells could reactivate high levels of Nanog expression as evidenced by GFP. Two days post-sorting of Nanog-GFP low cells heterogeneous GFP colonies were observed. Following 6 days of growth post-sorting these colonies expand and recapitulate the pre-sorted distribution of cells (**Fig. 5B**). These results indicate that an equilibrium exists between the primitive endoderm and ICM metastable states and that this equilibrium can be modified but not exclusively restricted by Wnt3a.

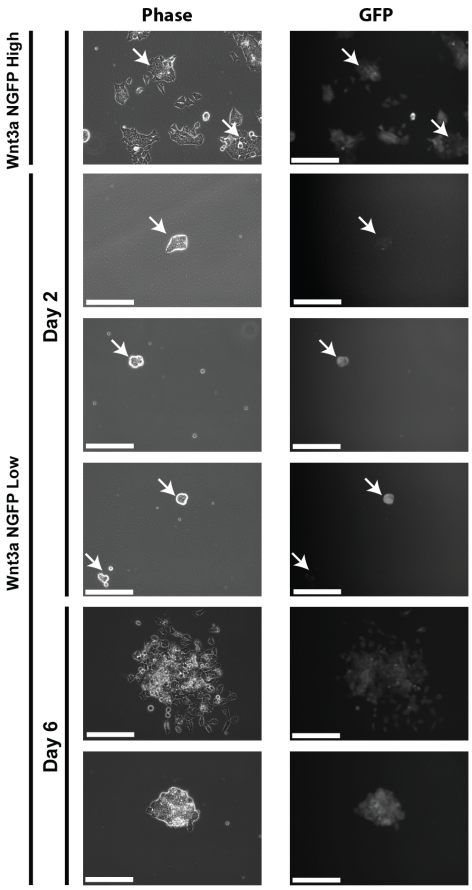
*2.3.5 A Mechanism that Links Wnt3a via Tbx3 to the Induction of a Metastable PE Population.*

Canonical Wnt signalling typically involves β-catenin binding with a Tcf/Lef family transcription factor on the promoter of Wnt target genes. To identify specific genes that are targeted by Wnt3a in undifferentiated mES cells we further analyzed our RNA-seq gene profiles obtained following retroviral and recombinant treatment. In this fashion we would gain mechanistic insight into how Wnt3a activates primitive endoderm genes.

We identified genes that were both up regulated or down regulated in unison by either Wnt3a retroviral overexpression and recombinant Wnt3a treatment. Based on this, we identified three transcription factors Tbx3, Pou2f3, and Cdx1 that were significantly up regulated upon Wnt3a overexpression or recombinant Wnt3a treatment (**Fig. 6A**). Up-regulation of these genes was subsequently confirmed by qPCR in ES cells overexpressing Wnt3a (**Fig. 6B**).

**Figure 5. Sorted Wnt3a Nanog-GFP populations interconvert under self-renewing conditions.**
**(A)** FACS analysis to assess the ability of sorted populations to return to a steady state equilibrium. Nanog-GFP-high and Nanog-GFP-low Wnt3a cells were cultured and analyzed at 3, 6, and 9 days post sort. Both populations are able to reconstitute each other following 9 days in culture implying that each population is metastable. The histogram represents one replicate. The time course of interconversion was conducted in duplicate. **(B)** Immunofluorescent validation of metastability for Nanog-GFP-high Wnt3a and Nanog-GFP-low Wnt3a populations. Scale bar denotes 100 μm.


A
Day 0
% of Max
GFP
J1 ES
Mock Nanog GFP
Wnt3a Nanog GFP
Wnt3a Nanog GFP Low
FACS
Wnt3a Nanog GFP High
Day 3
Day 6
Day 9
W3a Nanog GFP Low
W3a Nanog GFP High


B
Phase
GFP
Wnt3a NGFP High
Wnt3a NGFP Low
Day 2
Day 6

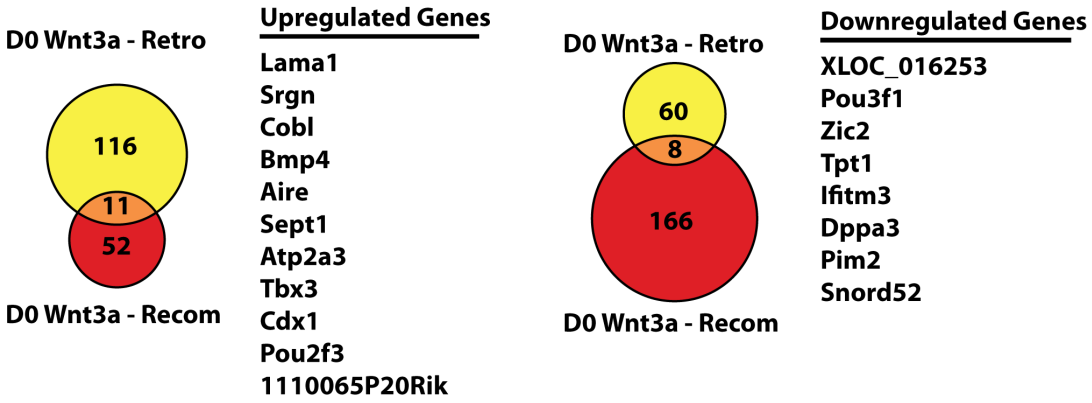
Of these three genes, *Tbx3* has recently been shown to play a role in pluripotency and is also involved in inhibiting ectoderm and trophectoderm while promoting primitive endoderm formation[255]. Taking into account these findings we used siRNA directed against *Tbx3* to determine whether it is required for induction of the Nanog-GFP low population by Wnt3a. We initially confirmed the reduction in *Tbx3* expression by siRNA in Nanog-GFP Wnt3a and mock cell lines (**Fig. 6C**). Nanog-GFP Wnt3a cells transfected with scrambled siRNA displayed a Nanog-GFP low population whereas the Nanog-GFP mock cells did not (**Fig. 6D**). Upon knockdown of Tbx3, the percentage of Nanog-GFP Wnt3a low cells underwent a ~25% decrease. Conversely when Tbx3 was knocked down we observed an increase in the percentage of Nanog-GFP high cells (**Fig. 6D**). These data suggest that following overexpression of Wnt3a, *Tbx3* mediates the induction of a primitive endoderm metastable state. In the absence of *Tbx3* the ability of Wnt3a to induce a Nanog-GFP low population is reduced.

## 2.4 Discussion

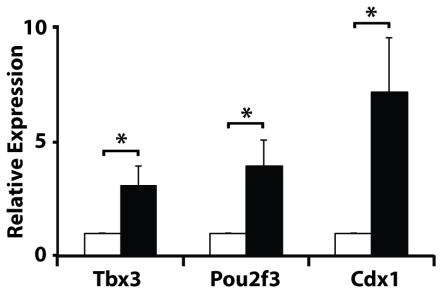
Embryonic stem cells were initially viewed as a homogeneous pluripotent population that has the capacity to give rise to all cell types of the embryo proper. This simplistic view has recently been challenged following the identification of distinct metastable states that exist in equilibrium within the undifferentiated state. Metastability confers to ESCs the unique ability to interconvert between two distinct cellular states. The first state is fully pluripotent whereas the metastable state is lineage primed but reversible. Cells that reside within a given metastable state display a unique epigenetic signature and are primed for differentiation into a specific lineage [29]. However, the

**Figure 6. Tbx3 mediates the Wnt3a induction of a primitive endoderm metastable population.**
**(A)** A Venn diagram compilation of genes either up or down regulated (>2 fold) between retroviral and recombinant treatment of Wnt3a in mES cells in the undifferentiated state. **(B)** Candidate transcription factors up regulated following Wnt3a induction were verified by qPCR. Data are presented as the mean ± SEM from one biological sample from J1, R1, and D3 mESC lines, normalized to GAPDH and relative to mock infected controls. P = two-tailed Student's *t* –test where * < 0.05 and ** < 0.01. **(C)** qPCR showing siRNA mediated knockdown of Tbx3 resulted in a 60% and 70% decrease in Tbx3 mRNA levels relative to siScrambled controls in the Nanog-GFP mock and Nanog-GFP Wnt3a cell lines respectively. **(D)** Following siRNA mediated knockdown of Tbx3 the % of Nanog-GFP low cells in the Wnt3a expressing cell line is reduced relative to the siScrambled control. **(E)** Similarly, the % of Nanog-GFP high cells is increased upon Tbx3 mediated knockdown relative to the siScrambled Nanog-GFP Wnt3a control. Data are presented as the mean ± SEM of 3 biological triplicate samples in Wnt3a Nanog-GFP cell lines, normalized to GAPDH and relative to mock infected controls. P = two-tailed Student's *t* –test where * < 0.05 and ** < 0.01.

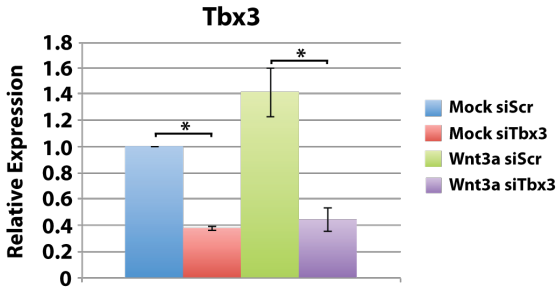
A


D0 Wnt3a - Retro
116
11
52
D0 Wnt3a - Recom
Upregulated Genes
Lama1
Srgn
Cobl
Bmp4
Aire
Sept1
Atp2a3
Tbx3
Cdx1
Pou2f3
1110065P20Rik
D0 Wnt3a - Retro
60
8
166
D0 Wnt3a - Recom
Downregulated Genes
XLOC_016253
Pou3f1
Zic2
Tpt1
Ifitm3
Dppa3
Pim2
Snord52

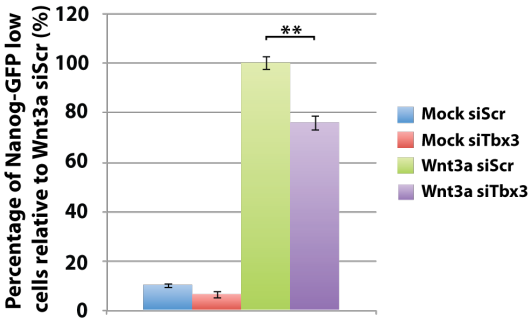
B


Relative Expression
10
5
0
*
*
*
Tbx3
Pou2f3
Cdx1

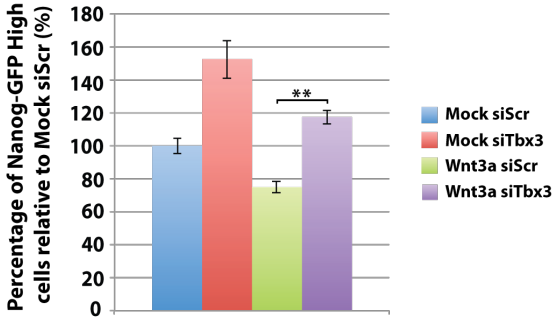
C


Tbx3
Relative Expression
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
*
*
Mock siScr
Mock siTbx3
Wnt3a siScr
Wnt3a siTbx3

D


Percentage of Nanog-GFP low cells relative to Wnt3a siScr (%)
120
100
80
60
40
20
0
**
Mock siScr
Mock siTbx3
Wnt3a siScr
Wnt3a siTbx3

E


Percentage of Nanog-GFP High cells relative to Mock siScr (%)
180
160
140
120
100
80
60
40
20
0
**
Mock siScr
Mock siTbx3
Wnt3a siScr
Wnt3a siTbx3

developmental signals and the molecular mechanisms that govern the formation and maintenance of metastable states have not been previously defined.

Here we demonstrate that induction of the canonical Wnt-signaling pathway in mESCs induces a pluripotent primitive endodermal metastable state. Moreover, we found that Wnt3a primes undifferentiated mESC for differentiation into the primitive endoderm lineage. Indeed, this early priming into the primitive endoderm culminates in a profound enrichment of visceral endoderm at the expense of other germ layer lineages in differentiating mESC cultures. Furthermore, we found that Wnt3a induction of a primitive endoderm metastable state is mediated by the T-box transcription factor *Tbx3*. Thus, these results provide a mechanistic framework that explains how activated canonical Wnt signalling can promote visceral endoderm development.

Three metastable states have been defined in mESC. The first is an ICM state characterized by cells that express *Stella*, *Rex1*, *Nanog*, and Pecam1. The second state is similar to the post-implantation epiblast (EpiSC) state and is defined by induction of *Fgf5* and *Brachyury*. The third and most recently identified is a primitive endoderm-like metastable state characterised by the expression of primitive endoderm markers. This state was first identified based on its low expression of Nanog and high expression of the primitive endoderm marker, *Gata6* [23]. Further studies using *Rex1, Hex* and *Sox17* support the idea for a primed primitive endoderm state in mESCs, yet differences exist regarding the ability of this population to reversibly convert back into the naive pluripotent state [26,30,33].

Our results demonstrate that Wnt3a is capable of inducing a primed primitive endodermal metastable state in mESCs. Wnt3a primes mESCs for primitive endoderm

specification (Figure 2B and C) and gives rise to a Nanog-GFP low population that up regulates expression of primitive endoderm genes (Figure 4B and C). Importantly, this population is capable of interconverting back to Nanog-GFP high cells following FACS isolation and *in vitro* culture (Figure 5A and B). Taken together these abilities support the characterization of this pluripotent primitive endoderm population as being metastable. Interestingly, activation of canonical Wnt signalling does not preclude the ability of Nanog GFP high and low cells to interconvert. Considering mESCs are metastable and fluctuate among different states surrounding the pluripotent state, a question remains as to whether activation of canonical Wnt signalling induces a subpopulation of primitive endoderm cells or prevents the return of these cells to the naïve pluripotent ground state. Irrespective of the mechanism, lineage priming into the primitive endoderm is apparent as robust differentiation into visceral endoderm occurs in the absence of LIF (Figure 3A-C).

Forced expression of Gata4 and Gata6 has been previously demonstrated to direct primitive endoderm differentiation in mESCs [22,256]. Interestingly, we found induction of primitive endodermal genes such as Gata6, Gata4, Sox7 and Sox17 in both the undifferentiated and differentiated states. In addition, we observed a broad increase in primitive endoderm markers following activation of the canonical signalling pathway. Therefore, although Wnt3a treated mESCs remain pluripotent, they become primed for entry into the primitive endoderm lineage. This is an important distinction, as our experiments support the notion that Wnt3a promotes and simultaneously maintains two functionally discrete cells types in the undifferentiated state.

In addition to the Gata4 and Gata6, inhibition of FGF signalling can reduce the presence of primitive endoderm populations within the heterogeneous undifferentiated state of mES cells *in vitro*[33]. Indeed mice carrying mutations in components of the FGF/Erk signalling pathway notably *Fgf4*, *Fgfr2* and *Grb2* exhibit a loss in primitive endoderm formation and suffer from peri-implantation lethality[12,257-259]. *In vitro* experiments involving FGF stimulation of the Erk1/2 signalling cascade in mES cells induces lineage commitment and abrogates self-renewal[260]. Paradoxically, we observed a reduction in *Fgfr2*, a receptor present on primitive endoderm cells, a slight increase in *Grb2* and no change in *Fgf4* upon stimulation of mES cells with Wnt3a (Table SI). During blastocyst development, the exogenous addition of *Fgf4* promotes the formation of primitive endoderm at the expense of the epiblast[261].

In mES cells, the addition of Fgf2 and activin results in the formation of cells with similarities to post implantation epiblast stem cells (EpiSCs) [67,68]. EpiSCs are able to proliferate indefinitely *in vitro* in a similar fashion to mES cells however as opposed to mES cells EpiSC are significantly restricted in their ability to contribute to the germline. Prior research demonstrates that Wnt3a inhibits the formation of EpiSCs[262]. Following Wnt3a overexpression we observed a down regulation of epiblast markers suggesting the induction of a primitive endoderm metastable state is distinct from the epiblast metastable state.

The likelihood that Wnts and FGFs interact in some fashion in undifferentiated mES cells is possible. Taken into context, the Wnt signalling pathway involves multiple modalities leading to diversity and the opportunity at multiple levels of intracellular signalling for crosstalk with other signalling pathways. In fact ERK, a downstream

effector of FGF signalling, is involved along with the Wnt/β-catenin pathway in Wnt3a-induced proliferation of NIH3T3 cells[248]. The complexity of the undifferentiated state complicates the ability to parallel a role for FGF signalling in a pluripotent cell line cultured *in vitro* with embryonic development, where the ICM exists only transiently *in vivo*.

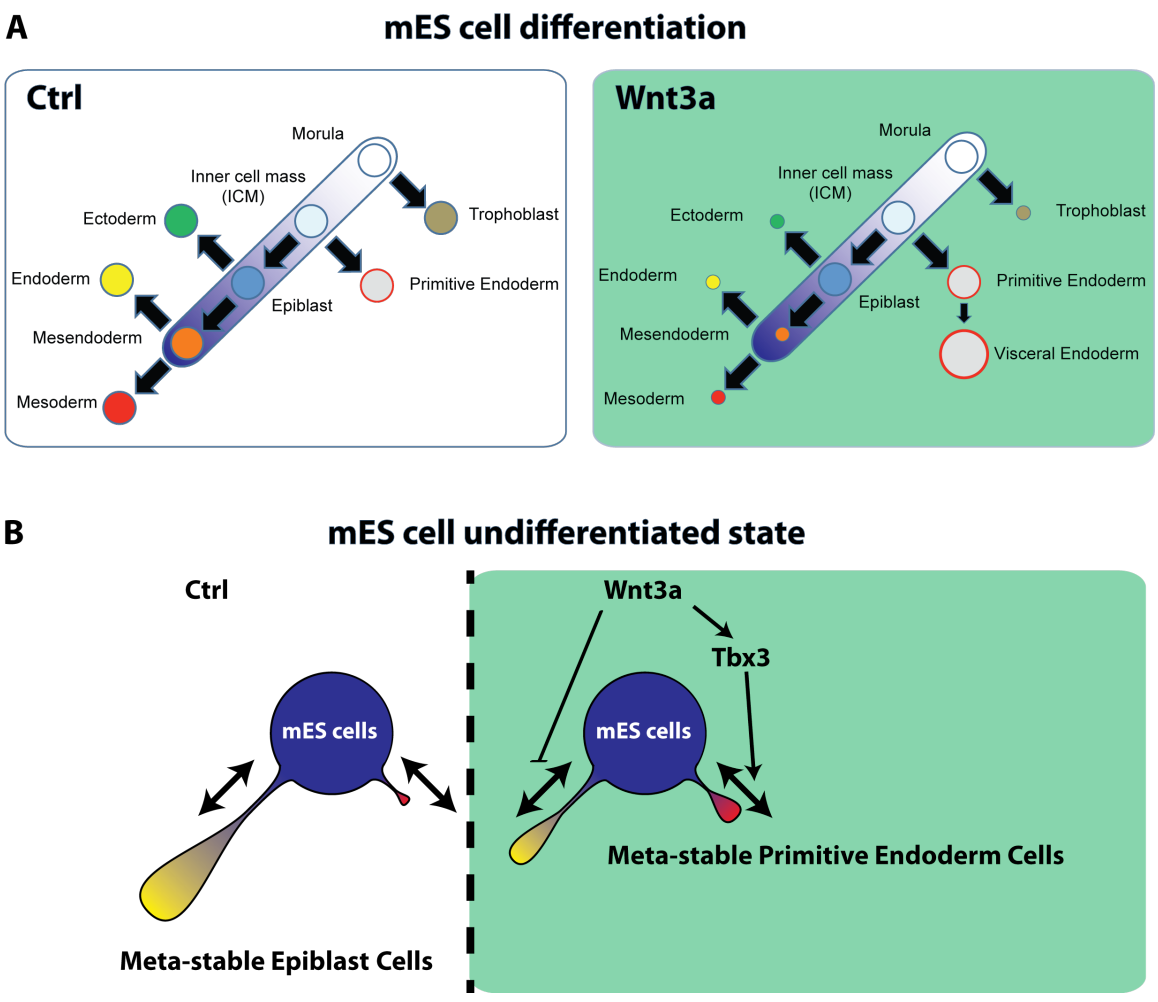
The Wnt signalling pathway is involved in numerous and at times contradictory roles in mES cells[183,263]. Wnts, as secreted morphogens elicit temporal and concentration effects[175]. In embryonic development and *in vitro* differentiation of mESCs, canonical Wnts are associated with the induction of brachyury (*T*) and the mesendodermal lineage [219,224,243]. Interestingly, prior to gastrulation, the earliest activation of downstream target genes of canonical Wnt signalling occur in the nascent visceral endoderm [215]. Notably, mice expressing a constitutively active form of β-catenin display an enlarged visceral endoderm and mESC lines used to derive these mice preferentially contribute to the visceral endoderm along with other extra-embryonic tissues in complementation assays [212]. Conversely, mice deficient for β-catenin lack appropriate organisation of the visceral endoderm resulting in defective anterior-posterior axis formation. *In vitro,* mESCs with an inactive form of adenomatous polyposis coli (APC) ($Apc^{Min/Min}$), a key regulator of β-catenin degradation, display increased levels of visceral endoderm upon differentiation [211,264]. Our experiments demonstrating the induction of a primitive endoderm metastable state elucidate how genetic studies perturbing canonical Wnt signalling affect visceral endoderm development.

The mechanisms involved in the induction and maintenance of a metastable state remain poorly understood. Of several conserved genes up regulated upon recombinant

Wnt3a treatment or retroviral overexpression, *Tbx3* was chosen due to its expression in the undifferentiated state and for its role in specification of primitive endoderm. Recent studies have shown that *Tbx3* along with *Klf4* are able to promote pluripotency in the absence of LIF [265]. *Tbx3* was previously identified as a downstream target of canonical Wnt signalling in murine and human liver tumours suggesting a direct mechanism exists for canonical Wnts to activate *Tbx3* [266]. *Tbx3* is required for extra-embryonic endoderm (ExEn) commitment and also functions to suppress trophectoderm and ectoderm differentiation in mESCs [255]. Overexpression of Tbx3 in mESCs maintained in the presence of LIF results in the induction of *Gata6*, *Sox17*, and *Dab2* suggesting that the primary role of Tbx3 is to induce the primitive endoderm. Thus, our finding that *Tbx3* mediates the induction of a primitive endoderm metastable state following activation of canonical Wnt signalling supports the hypothesis that Tbx3 is a canonical-Wnt target in mESCs.

In conclusion, our data reveals that activation of the canonical Wnt pathway by Wnt3a results in the induction of a metastable population of undifferentiated cells with characteristics of the primitive endoderm (**Fig. 7**). Upon LIF removal continued Wnt3a stimulation results in a bias of differentiation into the visceral endoderm lineage at the expense of ectoderm, mesoderm and endoderm. These results suggest canonical Wnt signalling, acting via *Tbx3*, is able to induce a primitive endoderm metastable state within undifferentiated mESCs. Further investigation into the specific mechanism by which the canonical Wnt pathway affects *Tbx3* is important to provide a better understanding of this effect. These findings provide important implications for studying pluripotency and

**Figure 7. Wnt signalling alters the differentiation of mES cells and induces a primitive endoderm meta-stable state.**
(**A**) mES cell differentiation in the absence (Left) and presence (Right) of constitutive activation of Wnt signalling. (**B**) Induction of a pluripotent primitive endoderm metastable state depends upon canonical Wnt activation and Tbx3 expression.


A
mES cell differentiation
Ctrl
Morula
Inner cell mass
(ICM)
Ectoderm
Trophoblast
Endoderm
Primitive Endoderm
Epiblast
Mesendoderm
Mesoderm
Wnt3a
Morula
Inner cell mass
(ICM)
Ectoderm
Trophoblast
Endoderm
Primitive Endoderm
Epiblast
Mesendoderm
Visceral Endoderm
Mesoderm
B
mES cell undifferentiated state
Ctrl
mES cells
Meta-stable Epiblast Cells
Wnt3a
Tbx3
mES cells
Meta-stable Primitive Endoderm Cells

regulation of the metastable state. Furthermore, applications consisting of ES cell derivation, cellular reprogramming and directed differentiation of pluripotent and induced pluripotent (iPS) cells currently employ Wnt3a or activation of canonical Wnt signalling[187]. Understanding the intricacies and effects of this signalling pathway on pluripotency is critical for therapeutic efficacy.

## 2.5 Materials and Methods

### Cell lines and Cell culture

The following mouse ES cell lines D3 Gl, J1, R1, TGFP and Millitrace Nanog GFP were maintained on a layer of irradiated MEFs (DR4) in GMEM high glucose medium (Wisent) supplemented with 15% fetal bovine serum (Hyclone), 1x non essential amino acids, 1x sodium pyruvate, 1x Glutamax, β-Mercaptoethanol, 1x Pen/Strep (Invitrogen), and 1000U ESGRO (chemicon) at 37$^{o}$C and 5% $CO_2$. Wnt3a retroviral cell lines used throughout the differentiation time course were made from each of the three ES cell lines (D3, J1 and R1). Prior to differentiation ES cell lines were cultured for two passages on 0.2% gelatin. For embryoid body formation, ES cells were seeded in 20 μl hanging drops at a density of $4x10^4$ cells/ml in ES cell medium without LIF. Embryoid bodies were transferred to petri dishes on Day 3 of differentiation and cultured in suspension until Day 5 at which time they were transferred to gelatine coated culture dishes for the remainder of differentiation. Plat-E cells, a derivative of HEK293T cells were used to generate retrovirus. Plat-E cells were maintained in DMEM high glucose medium supplemented with 10% fetal bovine serum (Hyclone), 1x Pen/Strep, 1 μg/ml Puromycin and 10 μg/ml Blastocydin at 37$^{o}$C and 5% $CO_2$. Plat-E cells at 80% confluence were transfected using Lipofectamine 2000 (Invitrogen) according to the manufacturer's

directions. Cells were fed with fresh DMEM 12 hrs post transfection. 48 hrs post tranfection retroviral supernatants were collected, centrifuged to remove cells and filtered (0.45 μm) for purification.

**Immunostaining**

Embryoid bodies at 5 days of differentiation were briefly rinsed in PBS and subsequently frozen in a 2:1 ratio of Tissue-Tek OCT to 30% sucrose in a dry ice/ethanol bath. Serial cryosections 12 μm in thickness were cut using a cryostat (Leicia CM 1850). Sections were fixed in 4% PFA for 10 min at room temperature followed by incubation in 100 mM Glycine for 10 min at room temperature. Fixed cells were permeabilized and blocked in 5% FBS, 0.2% Triton X-100 (Sigma) and 2% bovine serum albumin (Fisher Scientific) for 1 hr at room temperature. Cells were incubated with primary antibodies (View Table SII) for 1 hr at room temperature or alternatively 4$^{o}$C overnight. Secondary antibodies were diluted 1:1000 in PBS and incubated with fixed cells for 1 hr at room temperature. Nuclei were counter-stained with Hoechst 33342 (Sigma). Images were taken on a Carl Zeiss Axioplan 2 Microscope with an AxioCam HRm b/w (Zeiss) camera using Axiovision v3.2 (Zeiss) acquisition software. Photoshop was used to adjust the baseline fluorescence

**siRNA knockdown of Tbx3**

siRNA for Tbx3 was obtained from Ambion (s74777-4390771) and transfected into Nanog-GFP ES cells using lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Medium was changed 6 hours post transfection. FACS and RNA isolation was conducted 48 hrs after transfection.

**Fluorescent Activated Cell Sorting (FACS)**

Prior to FACS sorting Nanog-GFP reporter ES cells were trypsinized, pelleted, and filtered through 30 μm filters (Miltenyi Biotech). Cells were separated on a MoFlo cytometer (DakoCytomation) equipped with 3 lasers. Sorting gates were strictly defined based on both Side Scatter (SSC) vs GFP as well as PE vs GFP plots. Wild type J1 ES cells were used as a negative control for GFP.

**Gene expression analysis (qPCR)**

Total RNA was isolated from biological triplicate samples from each of the retroviral ES cell lines (D3, J1, R1). Total RNA was subjected to on column DNase digestion using an RNeasy mini Kit as per the manufacturer's instructions (Qiagen). 1 μg of RNA was reverse transcribed using iScript RT kit (BioRad) and subsequently diluted 1:5, with 1 μl of template used per qPCR reaction. SYBR Green real-time PCR reactions (BioRad) were performed in duplicate using an MX3000p PCR machine (Stratagene, La Jolla, CA) and 40 cycles of amplification (95$^{o}$C 10s; 60$^{o}$C 30s; 72$^{o}$C 30s). PCR primers for real-time PCR were designed using the online Primer3 software (http://frodo.wi.mit.edu/primer3/) with gene specific information obtained from Ensembl. All primer sequences are provided in Table SIII. Relative fold change in expression was normalized to GAPDH and calculated using the ΔΔCT method (CT values < 30). Primer specificity was validated by denaturation curve analysis (55–94$^{o}$C) resulting in a single peak. Amplification-curve plotting and calculation of cycle threshold (Ct) values were performed using the MX3000p software (v3; Stratagene), with further calculations performed using Microsoft Excel. Each experiment was performed independently on at least three occasions.

**Gene expression analysis (RNA-seq)**

Global gene expression was assessed in retroviral (D3 Wnt3a) cell lines and recombinant Wnt3a treated (D3) cells relative to mock infected (D3 mock) and bovine serum albumin (BSA) treated controls respectively. Total RNA was isolated and subjected to on column DNase digestion using an RNeasy mini Kit as per the manufacturer's instructions (Qiagen).

mRNA-Seq was performed at the Erasmus Center for Biomics (Rotterdam, The Netherlands) as per manufacturers instructions (Illumina). Raw Solexa reads were mapped to the mouse genome (mm9) by Tophat[267]. Gene expression values were calculated by Cufflinks with default parameters[268]. Differentially expressed genes between Wnt3a treated samples and corresponding mock or BSA controls were called by Cuffdiff[268-270]. Heatmaps were generated by GenePattern (http://www.broadinstitute.org/cancer/software/genepattern).

**Statistical Analysis**

Statistical analysis was conducted using the *Student's t-test* method of determining inference based on small samples. Two tailed paired *t-tests* were conducted between treated and control groups at a given day of differentiation or globally for all treated or control samples over the course of differentiation. * indicates a p value of $< 0.05$ while ** indicates a p value of $< 0.01$.

**Supplementary Figures and Tables**

**Figure S1. Validation of Wnt3a expressing mESC lines.**
qPCR expression of Wnt3a in R1, J1 and D3 mESC lines relative to mock infected control cell lines.

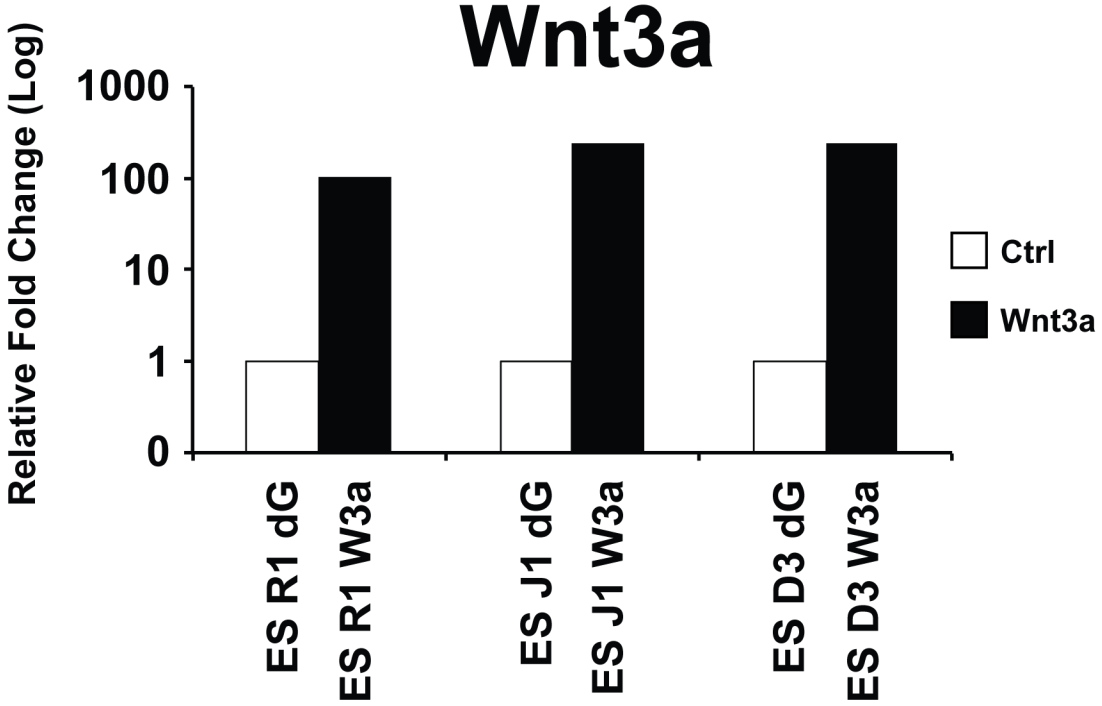
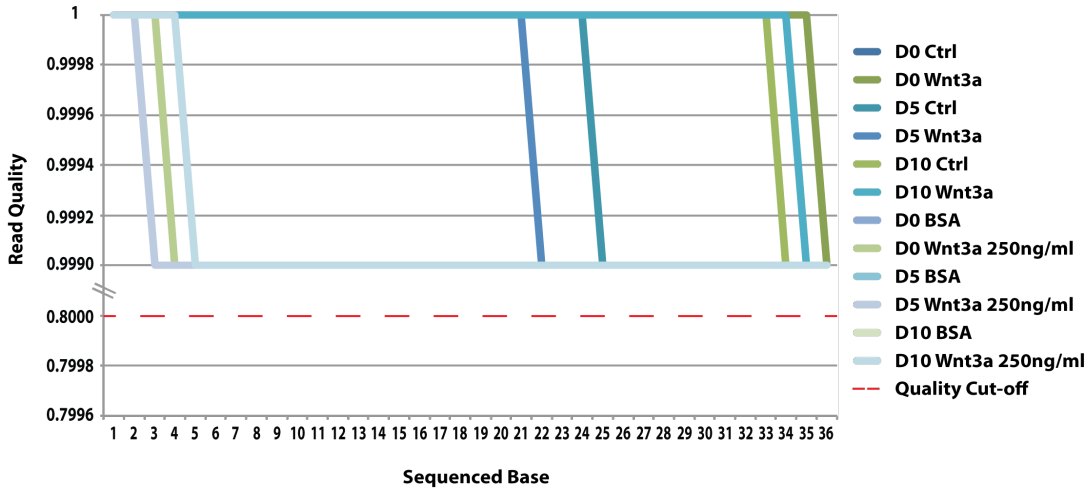

Wnt3a
Relative Fold Change (Log)
1000
100
10
1
0
ES R1 dG
ES R1 W3a
ES J1 dG
ES J1 W3a
ES D3 dG
ES D3 W3a
Ctrl
Wnt3a

**Figure S2. RNAseq quality controls.**
(**A**) Read quality as a function of the sequenced base demonstrates that high read quality (>0.99) is obtained for all samples in relation to an arbitrary cut off of 0.8 (**B**) Distribution of sequence reads in either exons, introns and intergenic regions. All samples sequenced samples displayed similar read distribution with the majority (>75%) of reads falling on exons.

A


Read Quality
1
0.9998
0.9996
0.9994
0.9992
0.9990
0.8000
0.7998
0.7996
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36
Sequenced Base
D0 Ctrl
D0 Wnt3a
D5 Ctrl
D5 Wnt3a
D10 Ctrl
D10 Wnt3a
D0 BSA
D0 Wnt3a 250ng/ml
D5 BSA
D5 Wnt3a 250ng/ml
D10 BSA
D10 Wnt3a 250ng/ml
Quality Cut-off

B

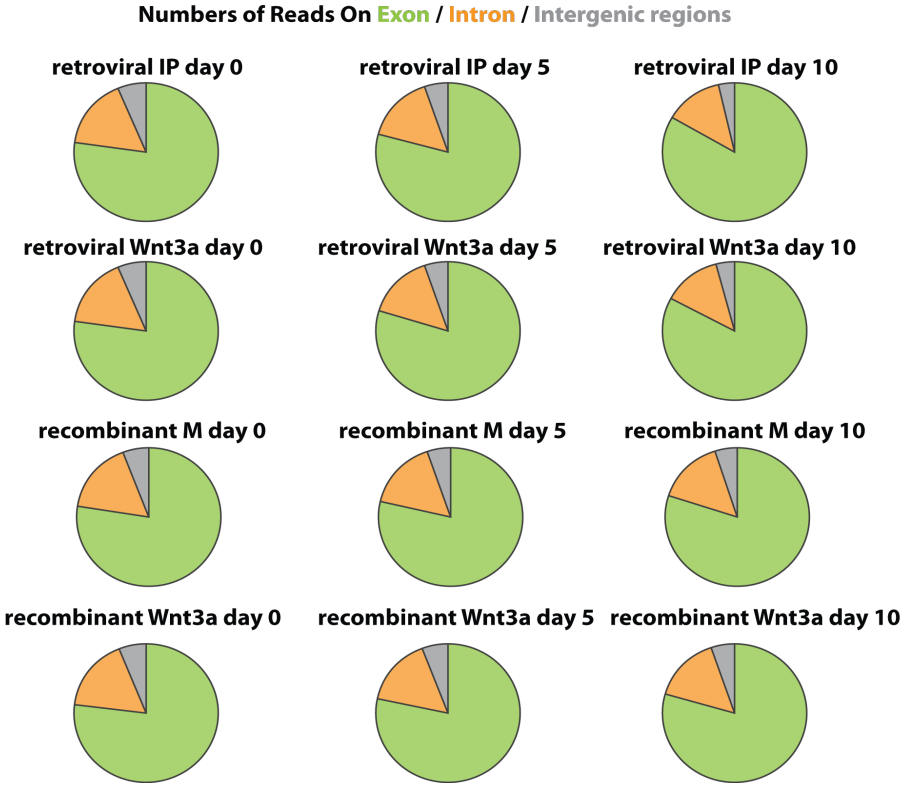

Numbers of Reads On Exon / Intron / Intergenic regions
retroviral IP day 0
retroviral IP day 5
retroviral IP day 10
retroviral Wnt3a day 0
retroviral Wnt3a day 5
retroviral Wnt3a day 10
recombinant M day 0
recombinant M day 5
recombinant M day 10
recombinant Wnt3a day 0
recombinant Wnt3a day 5
recombinant Wnt3a day 10

**Figure S3. Activation of Canonical Wnt Signalling**
(**A**) Activation of canonical Wnt signalling was determined by analyzing expression of Axin2 a known canonical Wnt signalling target. Affect of retroviral Wnt3a (**B**) and recombinant Wnt3a (**C**) on the predominant Tcf family members in the undifferentiated state. Abbreviations: FPKM, numbers of gene-matched Fragments Per 1 Kilo base length of transcript per 1 Million total reads

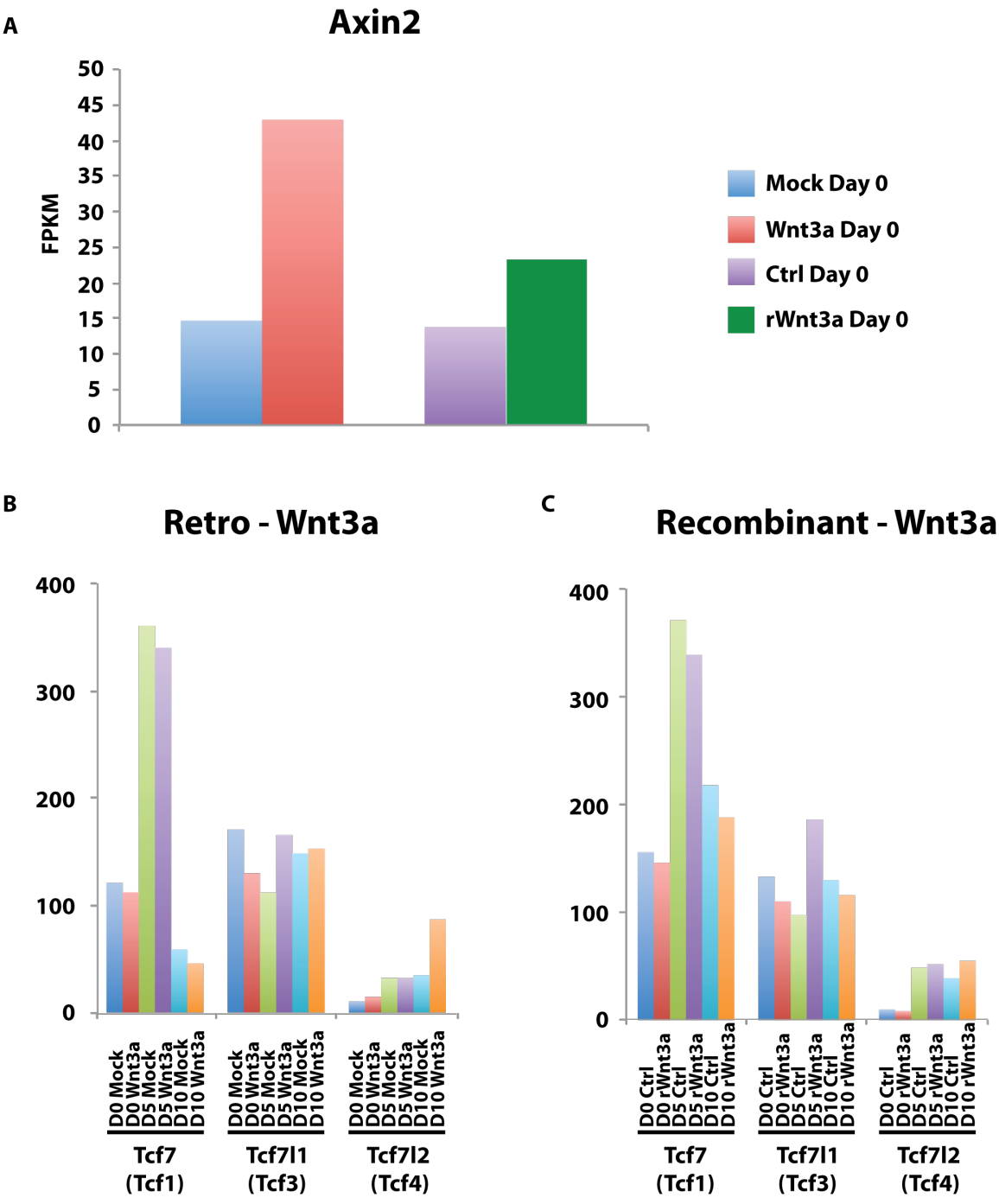
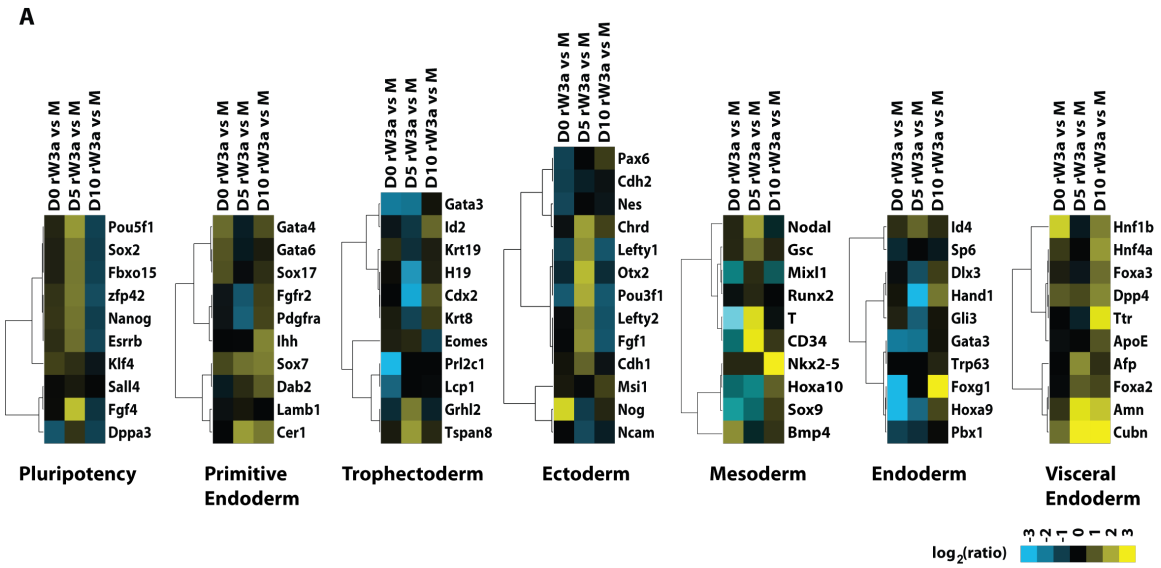

A
Axin2
FPKM
50
45
40
35
30
25
20
15
10
5
0
Mock Day 0
Wnt3a Day 0
Ctrl Day 0
rWnt3a Day 0
B
Retro - Wnt3a
400
300
200
100
0
D0 Mock
D0 Wnt3a
D5 Mock
D5 Wnt3a
D10 Mock
D10 Wnt3a
Tcf7
(Tcf1)
Tcf7l1
(Tcf3)
Tcf7l2
(Tcf4)
C
Recombinant - Wnt3a
400
300
200
100
0
D0 Ctrl
D0 rWnt3a
D5 Ctrl
D5 rWnt3a
D10 Ctrl
D10 rWnt3a
Tcf7
(Tcf1)
Tcf7l1
(Tcf3)
Tcf7l2
(Tcf4)

**Figure S4. RNAseq of recombinant Wnt3a treated mES cells.**
Global ratio heat map representing the difference in expression between recombinant Wnt3a (rWnt3a) and BSA (M) treated controls for biomarkers of blastocyst, primary germ layer and visceral endoderm genes. Yellow represents genes that increase in expression relative to the control while blue represents a decrease.


A
D0 rW3a vs M
D5 rW3a vs M
D10 rW3a vs M
Pou5f1
Sox2
Fbxo15
zfp42
Nanog
Esrrb
Klf4
Sall4
Fgf4
Dppa3
Pluripotency
Gata4
Gata6
Sox17
Fgfr2
Pdgfra
Ihh
Sox7
Dab2
Lamb1
Cer1
Primitive Endoderm
Gata3
Id2
Krt19
H19
Cdx2
Krt8
Eomes
Prl2c1
Lcp1
Grhl2
Tspan8
Trophectoderm
Pax6
Cdh2
Nes
Chrd
Lefty1
Otx2
Pou3f1
Lefty2
Fgf1
Cdh1
Msi1
Nog
Ncam
Ectoderm
Nodal
Gsc
Mixl1
Runx2
T
CD34
Nkx2-5
Hoxa10
Sox9
Bmp4
Mesoderm
Id4
Sp6
Dlx3
Hand1
Gli3
Gata3
Trp63
Foxg1
Hoxa9
Pbx1
Endoderm
Hnf1b
Hnf4a
Foxa3
Dpp4
Ttr
ApoE
Afp
Foxa2
Amn
Cubn
Visceral Endoderm
-3 -2 -1 0 1 2 3
log2(ratio)

## Supplementary Tables

### Table SI. RNA-seq results for FGF signaling pathway members

| | D0 Mock | D0 W3a | D5 Mock | D5 W3a | D10 Mock | D10 W3a | D0 BSA | D0 rW3a | D5 BSA | D5 rW3a | D10 BSA | D10 rW3a |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Retrovirus | | | | | | Recombinant | | | | | |
| Fgf1 | 11.74 | 3.36 | 0.00 | 0.00 | 3.65 | 2.58 | 5.52 | 5.18 | 0.00 | 2.16 | 1.62 | 0.00 |
| Fgf2 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 14.99 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Fgf3 | 0.00 | 11.89 | 138.55 | 107.94 | 0.00 | 7.41 | 5.78 | 11.25 | 72.00 | 22.03 | 21.85 | 40.87 |
| Fgf4 | 274.58 | 276.76 | 8.93 | 32.25 | 6.07 | 0.00 | 314.65 | 329.91 | 27.12 | 132.97 | 69.07 | 38.38 |
| Fgf5 | 8.97 | 0.00 | 15.62 | 135.37 | 0.00 | 0.00 | 14.02 | 0.00 | 81.58 | 327.41 | 62.33 | 34.18 |
| Fgf7 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 155.39 | 79.15 | 2.71 | 0.00 | 0.00 | 0.00 |
| Fgf8 | 0.00 | 0.00 | 52.24 | 99.29 | 0.00 | 0.00 | 17.45 | 0.00 | 78.62 | 107.61 | 71.03 | 46.96 |
| Fgf9 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Fgf10 | 5.69 | 16.05 | 79.79 | 81.13 | 30.56 | 26.70 | 48.13 | 43.01 | 136.14 | 74.08 | 68.21 | 56.83 |
| Fgf11 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 22.56 | 0.00 | 8.91 | 7.07 | 0.00 | 0.00 | 5.19 |
| Fgf13 | 12.71 | 8.65 | 17.66 | 16.76 | 7.44 | 3.53 | 12.26 | 11.54 | 19.02 | 13.77 | 25.87 | 18.64 |
| Fgf14 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.64 | 1.73 |
| Fgf15 | 63.68 | 25.29 | 194.38 | 146.81 | 6.50 | 6.19 | 77.76 | 49.61 | 150.15 | 72.80 | 70.72 | 100.82 |
| Fgf16 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Fgf17 | 101.32 | 67.29 | 0.00 | 7.05 | 0.00 | 0.00 | 32.14 | 20.37 | 4.87 | 9.67 | 13.47 | 6.66 |
| Fgf18 | 9.25 | 0.00 | 16.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 3.10 | 0.00 | 12.81 | 8.77 |
| Fgfr1 | 128.47 | 110.67 | 311.97 | 258.03 | 271.88 | 104.38 | 169.98 | 165.16 | 392.37 | 187.09 | 218.20 | 243.89 |
| Fgfr2 | 21.28 | 9.44 | 169.28 | 106.11 | 63.76 | 103.36 | 11.49 | 9.21 | 110.32 | 41.00 | 35.76 | 59.46 |
| Fgfr3 | 3.34 | 4.17 | 20.82 | 21.62 | 44.74 | 49.46 | 3.24 | 11.03 | 7.92 | 5.00 | 9.05 | 19.04 |
| Fgfr4 | 2.47 | 7.23 | 8.87 | 25.36 | 2.32 | 59.65 | 5.74 | 11.61 | 14.06 | 11.83 | 6.39 | 16.62 |
| Grb2 | 163.84 | 200.34 | 161.81 | 167.04 | 166.51 | 154.11 | 188.02 | 215.79 | 191.29 | 192.48 | 205.99 | 192.15 |

**Table SII. Antibodies Used in this Study.**

| Primary Antibodies | Dilution | Company | Cat# |
|---|---|---|---|
| Gata4 | 1:400 | Santacruz Biotech | sc25310 |
| Sox17 | 5μg/ml | RnD systems | AF1924 |
| Dab2 | 1:500 | BD bioscience | 610464 |
| **Secondary Antibodies** | | | |
| Donkey anti-Mouse IgG 647 | 1:1000 | Invitrogen - Molecular Probes | A31571 |
| Donkey anti-Goat 488 | 1:1000 | Invitrogen - Molecular Probes | A11055 |

**Table SIII. Primers Used in this Study.**

| Primer Name | Primer Sequence |
|---|---|
| Wnt3a ex3 F | ATTGAATTTGGAGGAATGGT |
| Wnt3a ex4 R | CTTGAAGTACGTGTAACGTG |
| GAPDH F | ATGCCAGTGAGCTTCCCGTC |
| GAPDH R | CATCACCATCTTCCAGGAGC |
| Oct4 ex3 F | AGCCGACAACAATGAGAACC |
| Oct4 ex4 R | TGATTGGCGATGTGAGTGAT |
| Sox2 F Endo (Takahashi) | TAGAGCTAGACTCCGGGCGATGA |
| Sox2 R Endo (Takahashi) | TTGCCTTAAACAAGACCACGAAA |
| Nanog ex1 F | ATGCCTGCAGTTTTTCATCC |
| Nanog ex2 R | ACAGTCCGCATCTTCTGCTT |
| Sox17 ex4 F | CTCGGGGATGTAAAGGTGAA |
| Sox17 ex5 R | GCTTCTCTGCCAAGGTCAAC |
| Gata4 ex6 F | CAGCAGTGAAGAGATGCGCC |
| Gata4 ex7 R | GCGATGTCTGAGTGACAGGA |
| Gata6 ex6 F | CTACACAAGCGACCACCTCA |
| Gata6 ex7 R | CCTCTTGGTAGCACCAGCTC |
| Cdx2 ex2 F | AAGACAAATACCGGGTGGTG |
| Cdx2 ex3 R | CTGCGGTTCTGAAACCAAAT |
| Eomes ex1 F | CCCTATGGCTCAAATTCCAC |
| Eomes ex2 R | TGCATGTTATTGTCCGCTTT |
| Id2 ex1 F | CTGGACTCGCATCCCACTAT |
| Id2 ex3 R | CTCCTGGTGAAATGGCTGAT |
| Pax6 ex12 F | AGTCCCCAGTCAGACCTCCT |
| Pax6 ex13 R | ACTTGGACGGGAACTGACAC |
| Nestin ex1 F | GTCGCTTAGAGGTGCAGCAG |
| Nestin ex2 R | TTCCAGGATCTGAGCGATCT |
| Ncam1 ex17 F | GAAGAACGGACTCCAAACCA |
| Ncam1 ex18 R | CTTTTGTTTGTGTGGCATCG |
| T ex7 F | CAGCCCACCTACTGGCTCTA |
| T ex8 R | CCCCTTCATACATCGGAGAA |
| Mixl1 ex1 F | AGTTGCTGGAGCTCGTCTTC |
| Mix1 ex2 R | AGGGCAATGGAGGAAAACTC |
| Gsc ex2 F | GAAGCCCTGGAGAACCTCTT |
| Gsc ex3 R | CAGTCCTGGGCCTGTACATT |
| Gata3 ex2 F | GGCTACGGTGCAGAGGTATC |
| Gata3 ex3 R | GATGGACGTCTTGGAGAAGG |
| Dlx3 ex1 F | TACTCGCCCAAGTCGGAATA |
| Dlx3 ex2 R | GCCAGATACTGGGCTTTCTG |
| Pbx1 ex7 F | GCAGTCACTCAATGGGGATT |
| Pbx1 ex9 R | ACTGCCAGGGCCTTCTGTAG |

| | |
|---|---|
| Sox7 ex1 F | CTTCAGGGGACAAGAGTTCG |
| Sox7 ex2 R | GGGTCTCTTCTGGGACAGTG |
| Dab2 ex10 F | CTTTGCCTCAGAACCTCCAG |
| Dab2 ex11 R | TGTAGACAACAGGCGTCCAG |
| Hnf1b ex3 F | AGAGGGAGGCCTTAGTGGAG |
| Hnf1b ex4 R | GTGCGTCTGGTTGGAGCTAT |
| Foxa2 ex2 F | CATCCGACTGGAGCAGCTA |
| Foxa2 ex3 R | GCGCCCACATAGGATGAC |
| Hnf4a ex8 F | TACATCAACGACCGGCAGTA |
| Hnf4a ex9 R | ATTGGTGCCCATGTGTTCTT |
| Cited1 F | CCGATTTATCGGACTTCTGC |
| Cited1 R | CGATCCAGAGGAGCTAGTGG |
| ApoE ex2 F | GTGCTGTTGGTCACATTGCT |
| ApoE ex4 R | AGCTGTTCCTCCAGCTCCTT |
| Cubn ex66 F | AATGGTCCAGATGCCAACTC |
| Cubn ex67 R | TCCCGATTCAGCTACTCCAC |
| Amn ex2 F | CTCCAACTGGAACCAGAACC |
| Amn ex5 R | GATTGCGGAAGAGCAACG |
| Tbx3 ex1 F | GCCTCTGAAGACCATGGAAC |
| Tbx3 ex2 R | CATCCACCGAGAGTTGTGAA |
| Pou2f3 ex12 F | GCAGCAATGAACTCCTCCTC |
| Pou2f4 ex13 R | CCTATGTGGAAGGCACGACT |
| Cdx1 ex1 F | GACGCCCTACGAATGGATG |
| Cdx1 ex2 R | CTCTGTGAGCCCCAGGTTAG |

**Acknowledgments**

We thank Dr. Hang Yin for helpful discussion and guidance as well as conducting the RNA-seq data normalization and displaying data in the form of heat maps. Dr Frank Grosveld, along with his technical staff, at Erasmus University and Medical Center in Rotterdam, The Netherlands for carrying out Solexa sequencing. Andrew Jones for helpful discussion and technical assistance. Dr. James Ellis for the HSC1-eGFP and HSC1-eGFP IRES Puro vectors. David Wilson for help and technical advice regarding tissue sectioning. Paul Oleynik for cell sorting and Dr. Mads Kaern for the millitrace Nanog-GFP cell line. M.A.R. holds the Canada Research Chair in Molecular Genetics and is an International Research Scholar of the Howard Hughes Medical Institute. F.D.P was supported by the Canadian Stem Cell Network, Muscular Dystrophy Canada and a Doctoral Research Award from the Canadian Institutes of Health Research.

# Chapter 3 - Transgenic Models to Study Myogenesis

## 3.1 Abstract

Adult skeletal muscle possesses a remarkable regenerative capacity due primarily to satellite cells, a resident population of adult stem cells that are important in muscle regeneration. Recent lineage tracing experiments suggest that the satellite cell population is heterogeneous and that the expression of *Myf5*, a core muscle regulatory factor distinguishes between a committed progenitor cell and a satellite stem cell. The mechanisms that govern the maintenance of this hierarchy as well as the developmental cues that lead to these satellite cell populations remain largely unknown. The creation of reporter cell lines and mouse models will help to understand and study these distinct satellite cell populations. This approach will benefit the isolation, live cell imaging and subsequent gene expression profiling of the two satellite cell populations. With these goals in mind attempts were made to create both a Pax7 YFP and Myf5 TdTomato reporter mouse line along with the intermediate targeted embryonic stem cell lines. Both embryonic stem cell targeting vectors along with targeted cell lines were successfully created. Chimeric mice from blastocyst injected targeted cell lines along with transgenic mice derived from BAC transgenesis were generated. However, these genetic tools did not demonstrate appropriate fluorescence concurrent with gene expression of *Pax7* and *Myf5* in mice. Reagents created over the course of this work will be valuable towards future efforts in this area.

## 3.2 Introduction

### *3.2.1 Adult Skeletal Muscle Development*

Adult skeletal muscle is capable of a remarkable degree of regeneration, suggesting the presence of a stem cell population either resident within muscle or capable of migrating to muscle. The major component of adult skeletal muscle is the myofiber; a giant syncytial cell containing hundreds of myonuclei within a continuous cytoplasm. Under physiological conditions the ability of adult muscle to undergo regeneration is largely attributed to a distinct subpopulation of myogenic cells, termed satellite cells, located between the basal lamina and sarcolemma of mature skeletal muscle fibers[47,271]. Satellite cells are a distinct lineage of myogenic stem cells that remain mitotically quiescent under normal physiological conditions. Upon muscle damage or in a state of disease, satellite cells activate and proliferate giving rise to two distinct populations. The first will return to the niche and replenish the stem cell pool. The second will continue to proliferate giving rise to myogenic progenitors that contribute to muscle regeneration via a process of differentiation and fusion[49,130,272]. Satellite cells are characterized by a panel of nuclear markers, *Pax7* being the most prominent[70]. *Pax7* is a paired-box transcription factor expressed during the development of cephalic neural crest derivatives, skeletal muscle and the central nervous system[273]. Embryonic skeletal muscle development in *Pax7*$^{-/-}$ mice appears normal yet upon birth mice exhibit a marginally reduced muscle mass and usually survive only a few weeks[101,274]. Analysis of affected skeletal muscle in postnatal *Pax7*$^{-/-}$ mice indicates a severe reduction in the number of satellite cells suggesting a deficiency in satellite self-renewal[70,274]. Experiments in double knockouts of *Pax7* and its closely related homologue *Pax3* displayed compromised development of somites and embryonic lethality at mid-gestation, suggesting a redundancy in functions of both genes during somitogenesis[275]. *Pax7* and *Pax3* initiate their expression patterns

during the formation of the somites and subsequent differentiation into the dermomyotome and sclerotome. Myotomal myogenesis initiates in the double knockout mice, however the resulting resident muscle progenitor cells either apoptose or deviate from the myogenic program and become cartilage[109]. In adult satellite cells *Pax3* expression is restricted to the diaphragm and is unable to compensate for a loss of *Pax7* in limb skeletal muscle[103,276].

Myogenic commitment and differentiation of myoblasts is dependent on a group of basic helix-loop-helix transcription factors comprising *MyoD*, *Myf5*, *Myogenin* and *Mrf4*. These myogenic regulator factors (MRFs) are subdivided into the primary MRFs (*MyoD*, *Myf5*) responsible for myogenic commitment and the secondary MRFs (*Myogenin*, *Mrf4*) required for myogenic differentiation. Upon expression of *Myogenin* and *Mrf4* myoblasts will differentiate by up regulating myosin heavy chain and fuse to form multinucleated myotubes[142,277,278].

*3.2.2 Satellite Cell Heterogeneity*

Our lab recently identified two satellite cell populations resident within adult skeletal muscle via lineage tracing experiments using the *Myf5*-Cre ROSA26-YFP mouse line. Satellite cells are either *Myf5* lineage negative ($Pax7^+/Myf5^-$ - satellite stem cell) or have expressed *Myf5* ($Pax7^+/Myf5^+$ - satellite progenitor cell) during development. The $Pax7^+/Myf5^-$ population is inherently different from $Pax7^+/Myf5^+$ with regards to gene expression and the ability to participate in muscle regeneration, therefore demonstrating the functional existence of satellite cell heterogeneity[49,67]. The $Pax7^+/Myf5^-$ population constitutes ~10% of total satellite cells in adult skeletal muscle and is referred to as a satellite stem cell due to its ability to give rise to both a $Myf5^+$ or $Myf5^-$ daughter cell

upon asymmetric division. The $Pax7^{+}/Myf5^{-}$ population can undergo both symmetric and asymmetric divisions giving rise to $Pax7^{+}/Myf5^{-}$ or $Pax7^{+}/Myf5^{+}$ daughter cells. This segregation is based on the orientation of the daughter cells with respect to the muscle fiber at the time of division: either apical basal, which results in asymmetric divisions or planar, which results in symmetric division. Upon in vitro culture satellite cells become cycling primary myoblasts and lose many characteristics attributed to the satellite cell *in vivo*[50]. Currently, both $Pax7^{+}/Myf5^{-}$ or $Pax7^{+}/Myf5^{+}$ satellite cell populations are difficult to isolate at high purity and in the case of $Pax7^{+}/Myf5^{-}$ are present only at low numbers limiting their ability to be analyzed. The creation of Knock-in ES cell lines for both Pax7 and Myf5 would offer the opportunity to generate and isolate large quantities of enriched progenitor cells, therefore resolving two of the major hurdles limiting the analysis of these adult satellite cell populations. In addition these ES cell lines can be used to create transgenic mice a useful tool to study the role of these cell types during muscle regeneration.

### *3.2.3 Generating Transgenic Reporter Mice for Pax7 and Myf5*

Embryonic stem (ES) cells have been shown to faithfully recapitulate embryonic development *in vivo*[7,8,233]. The use of this model system facilitates the mechanistic analysis of early developmental processes that regulate tissue differentiation and function. Upon differentiation, ES cells give rise to a heterogeneous population of cell lineages. Therefore, methods to isolate and purify specific cell types are critical to study lineage specification. A genetic reporter couples the production of a specific protein (fluorophore, LacZ, cell surface protein etc…) to gene expression allowing the identification/isolation of a cell type expressing a specific gene in a temporal fashion by

microscopy or flow cytometry[279,280]. Many fluorescent reporters exist that fluoresce in the blue, green, yellow, red and infrared spectrum facilitating the isolation of multiple cell types from one sample[280].

In order to study the formation of skeletal muscle Knock-In targeting vectors were created for two critical transcription factors expressed in myogenic progenitors, *Pax7* and *Myf5*. Naturally occurring genetic recombination involves a segment of DNA being moved from one place in the genome to another in either a homologous or non-homologous fashion. This process commonly arises during DNA repair, meiosis or in B cells of the immune system[281-283]. The process of homologous recombination is harnessed to create a genetic Knock-In involving the introduction of foreign DNA into a specific gene locus[284,285]. Controlling the specific integration site is advantageous as endogenous promoter sequences can remain unchanged and a reporter construct can be inserted in frame into the native loci. This technique is typically conducted in ES cells because it offers the option to create a transgenic mouse line in addition to an ES cell line for *in vitro* differentiation. We chose to introduce two different fluorophores, a yellow fluorophore (YFP or Ypet – emission peak 527 nm) into the Pax7 locus and a red fluorophore (TdTomato – emission peak 581 nm) into the Myf5 locus. Bacterial artificial chromosomes (BACs), containing genomic regions flanking the start site of either *Pax7* or *Myf5* were used to create reporter constructs and transgenic animals because they are capable of carrying large promoter and enhancer regions, thus increasing the chances of faithfully recapitulating gene expression[286]. The advent of recombineering methods to manipulate BAC DNA sequences further increased their usefulness in creating transgenic vectors[287]. The creation of a transgenic mouse differs from that of a Knock-in and

involves the random integration of a targeted BAC into the genome. Although this technique is less specific regarding integration it is advantageous as no exogenous DNA sequences are integrated near a gene of interest and therefore two functional alleles remain. Recombineering is a process incorporating the *Red* genes of bacteriophage λ to enable the site-specific homologous recombination of linear double-stranded DNA (dsDNA) fragments to be inserted into cloned DNA in a BAC or plasmid[288-290]. Red recombineering involves three proteins *exo*, *bet* and *gam*. *Exo* encodes a 5' to 3' exonuclease, which digests back the 5' strand of linear dsDNA to generate 3' overhangs. These 3' overhangs are subsequently coated with *bet*, a pairing protein that promotes annealing between complimentary DNA strands. *Gam* further improves the efficiency of recombination by inhibiting RecBCD activity in *E. coli*, an exonuclease that digests linear dsDNA. Together these three proteins (often abbreviated as α, β, and γ respectively) comprise the components of recombineering[291,292]. In order to create targeting vectors we used a recombineering system that expressed the recombineering components from a defective prophage integrated into the *E. coli* chromosome[293,294]. The expression of α β γ is regulated by a strong λ $P_L$ promoter, which itself is under the control of the temperature-sensitive λ *cI857* repressor. When cultured at 32$^{o}$C expression of α β γ is undetectable yet upon transfer to 42$^{o}$C expression is induced at very high levels.

Devising a strategy to create a reporter construct targeting a gene that is expressed at low levels or where the promoter or enhancer regions are unknown, as is potentially the case with Pax7, requires methods to enhance protein expression. One method to resolve these issues is to produce multiple copies of a reporter from a given promoter.

Multicistronic cassettes, in which more then one coding region is contained in a given mRNA transcript, often accomplish this by using either bi-directional promoters or an internal ribosomal entry site (IRES). However, these methods can be problematic due to "transcriptional interference" involving regulatory elements situated in close proximity resulting in either promoter suppression or transcriptional silencing[295-297]. To circumvent these issues we chose a method using 2a self-cleaving peptides to link multiple fluorophores to a secreted luciferase in order to report on the activity of the Pax7 promoter. 2a peptides mediate protein cleavage and are common to members of the Picornaviridae virus family including foot-and-mouth disease (FMDV), equine rhinitis A virus (ERAV), porcine teschovirus-1 and the insect virus Thosea asigna virus (TaV)[298-300]. The 2a consensus motif consists of the following amino acids (Asp-Val/Ile-Glu-X-Asn-Pro-Gly-Pro). Cleavage occurs due to a ribosomal skip mechanism that impairs normal peptide bond formation between the terminal Gly-Pro amino acids; the result of which has no affect on translation of amino acids downstream of the terminal Proline, and ultimately gives two distinct peptides[301].

The purpose of the following experiments was to develop various targeting vectors to create Knock-in ES cell lines for both *Pax7* and *Myf5*. These reporter lines would be used to study the specification of embryonic myogenic progenitors in vitro while transfer of the targeted ES cell lines into blastocysts would enable the creation of reporter mouse strains. Optimization of the initial targeting vectors included different fluorophores and the inclusion of 2a self-cleaving peptides while later attempts at creating a Myf5 reporter mouse involved creating a BAC transgenic animal.

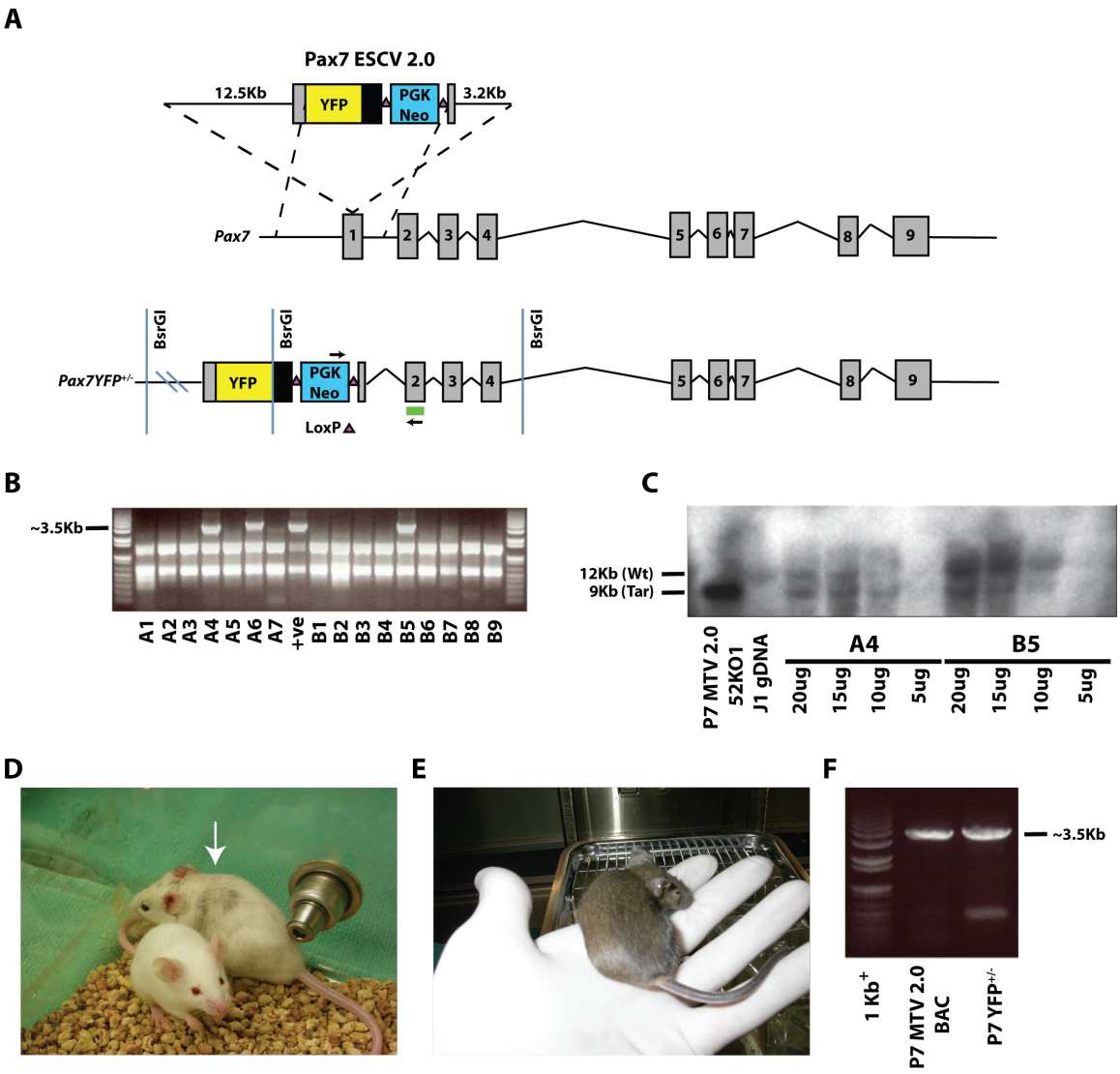
## 3.3 Results

*3.3.1 Pax7 Transgenic Models*

There are multiple steps involved in the creation of a Knock-in targeting vector using BACs. The first is the creation of a mini targeting vector (MTV) which contains a reporter, a selectable marker, and short (~100-200 bp) regions of homology (a 5' region and a 3' region) that are specific for the gene being targeted. The second step involves insertion of the MTV into a BAC by homologous recombination involving the recombineering machinery. In order to create the final ESC targeting vector a region containing homology arms (typically ~5-15 kb for a long arm and ~1-3 kb for a short arm) flanking the recombined fluorophore and selectable marker need to be retrieved from the BAC. This process is accomplished through recombineering into a targeting vector (PGKdtaBpa) that expresses a marker for negative selection, in this case diptheria toxin (DTA). Variations on this technique include using the targeted BAC as a vector to electroporate ES cells or pronuclear injection of the targeted BAC to create a transgenic mouse. These methods differ from a genetic Knock-in where the modified DNA is incorporated specifically into the genome; a characteristic that is beneficial when the regulatory sequences for a given gene are unknown, as is the case for *Pax7* and to a lesser extent *Myf5*.

The design and basis for our original Knock-in targeting vector for *Pax7* (P7 ESCV 2.0) (**Fig. 1A**) is based on the Brachyury (T) – GFP construct which was successful at recapitulating the expression of a mesoderm marker[226]. This construct consists of a mini gene locus, comprising the cDNA of YFP, followed by a splice donor site, an artificial intron, a splice acceptor site, a stop codon and an exon encoding the SV40 polyadenylation signal to prevent downstream transcription at the targeted loci of

**Figure 1. Design and validation of *Pax7* transgenic vectors.**
(**A**) Schematic structure of the mouse *Pax7* gene locus and the Pax7 ESCV 2.0 targeting vector. Pax7 ESCV 2.0 includes a YFP coding sequence in yellow followed by an artificial intron (ArtINT) in black comprising splice donor and acceptor sites and a SV40 polyadenylation signal. The *Pax7* gene locus is composed of grey boxes for exons and black lines for introns. Dashed lines represent regions that undergo homologous recombination upon entry of the targeting vector into ES cells. Purple triangles represent LoxP sites and flank a blue PGKNeo cassette. Vertical blue lines represent BsrGI restriction enzyme digestions sites, while black arrows designate primers and a green rectangle the hybridization probe for southern blotting. (**B**) Long range PCR screening of Pax7 ESCV 2.0 targeted J1 ES cell lines. Amplicon of a BAC targeted by Pax7 MTV 2.0 serves as a positive control with a band that migrates at ~3.5 Kb. A 1 Kb+ ladder is used as a marker. Clones A4, A6 and B5 display a band at ~3.5 Kb. (**C**) Southern blot validation of clones A4 and B5. A Pax7 MTV 2.0 targeted BAC was used as a positive control while J1 genomic DNA was a negative control. A wild type digested band has a size of ~12 Kb while the targeted (Tar) band has a size of ~9 Kb. (**D**) Chimeric mice derived from blastocyst injection of the J1 B5 Pax7 ESCV 2.0 cell line. White arrow designates the chimera. (**E**) Confirmation of germline transmission by the appearance of a full agouti mouse derived from a cross between a P7YFP$^{+/-}$ chimeric male and an FVB female. (**F**) Long range PCR confirmation of genomic DNA from a fully agouti Pax7YFP$^{+/-}$ mouse produces an amplicon of ~3.5 Kb. Pax7 MTV2.0 target BAC serves as a positive control and produces an amplicon of ~3.5 Kb.


A
Pax7 ESCV 2.0
12.5Kb
YFP
PGK Neo
3.2Kb
Pax7
1 2 3 4 5 6 7 8 9
BsrGI
BsrGI
BsrGI
Pax7YFP+/-
YFP
PGK Neo
LoxP
B
~3.5Kb
A1 A2 A3 A4 A5 A6 A7 +ve B1 B2 B3 B4 B5 B6 B7 B8 B9
C
12Kb (Wt)
9Kb (Tar)
P7 MTV 2.0
52KO1
J1 gDNA
A4
B5
20ug 15ug 10ug 5ug 20ug 15ug 10ug 5ug
D
E
F
~3.5Kb
1 Kb+
P7 MTV 2.0 BAC
P7 YFP+/-

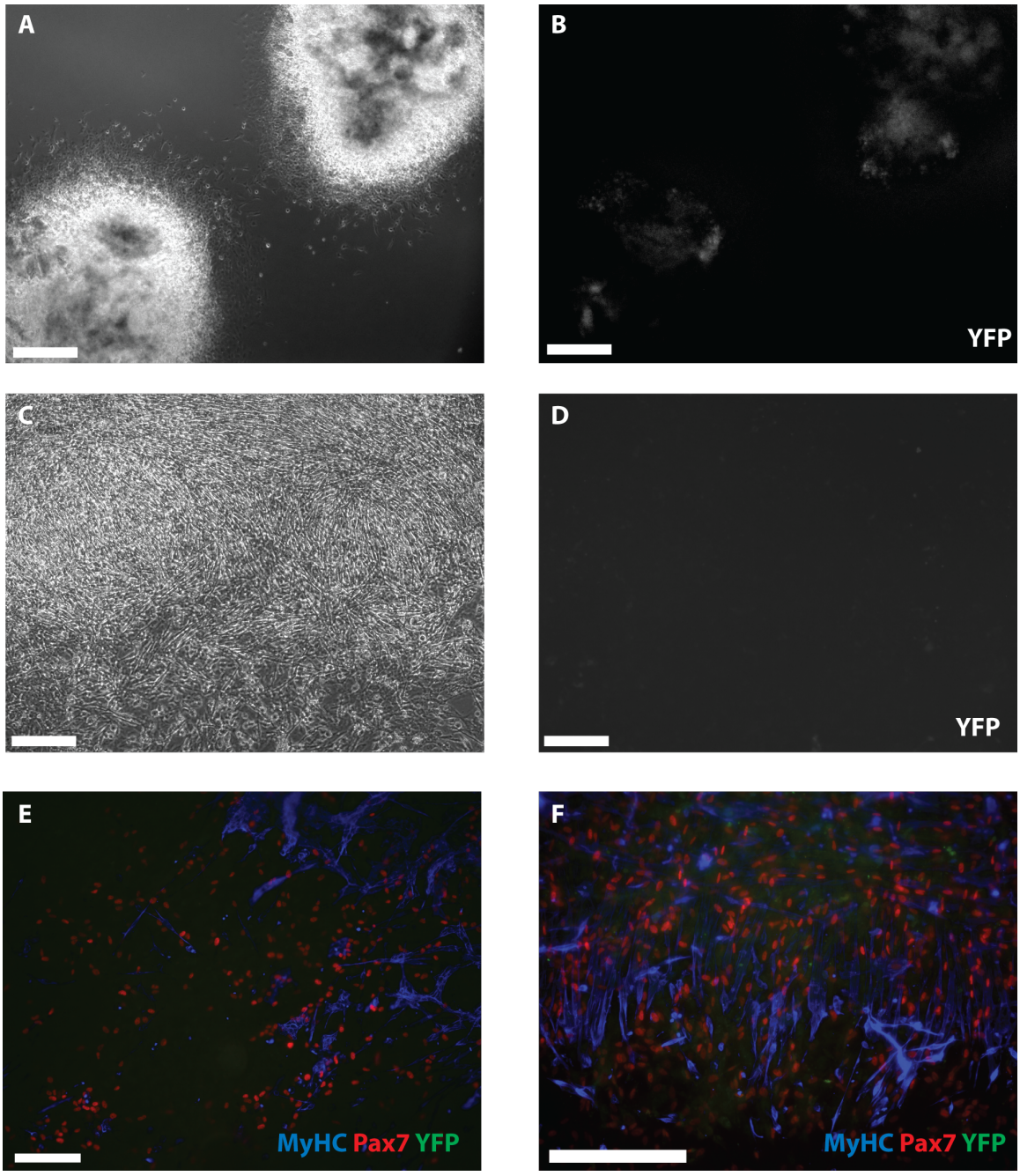
*Pax7*. The rationale for using a mini gene with a stop codon following the artificial intron is based on the notion that primary transcripts with stop codons preceding intronic sequences can be recognized as aberrant messages resulting in nonsense mediated decay[302]. The use of homologous recombination enabled both constructs to be clean insertions (no removal of endogenous or addition of exogenous DNA) therefore reducing any spurious alterations of the gene loci, which could result in decreased transcription of YFP. The Pax7 ESCV 2.0 is designed to target the first exon of *Pax7* specifically at the translational start site for the gene. A mini targeting vector (P7 MTV 2.0) was created capable of expressing YFP in mammalian cells along with LoxP sites (**Fig. S1**) flanking a PGK-Neo cassette that can be excised upon Cre recombinase treatment. The P7 MTV 2.0 was recombined into a Pax7 BAC containing 48 kb upstream and 36 kb downstream of the gene loci. A region of the targeting BAC encompassing ~12.5 kb upstream and ~3 kb downstream from the targeting site was subsequently retrieved into the PGKdtaBpA vector. Following verification by restriction enzyme digestion and DNA sequencing (data not shown) the targeting construct was electroporated into J1 ES cells and selected in 200 μg/ml of G418. Clones were picked 7-10 days following electroporation and passaged into 96 well dishes for expansion and validation by long range PCR and Southern blotting.

Long range PCR was conducted using primers that originated in the Neo cassette and terminated in the second intron of *Pax7*. These primers flank the short arm of homology and amplification is therefore specific for an insertion between the first and second exon of the *Pax7* locus. The long range PCR results in an amplicon that is ~3.5 kb (**Fig. 1B**). Of the 192 clones picked 3 (1.5%) displayed a 3.5 kb band. Clones that

displayed a ~3.5 kb band via PCR were further verified by Southern blot. Southern blot validation of positive Pax7 ESCV 2.0 clones was conducted by digestion of gDNA with BsrG1 followed by electrophoresis on a 0.8% agarose gel and overnight transfer to a Hybond NX membrane. The membrane was subsequently detected with a probe specific to a region encompassing exon 3 of *Pax7*, which lies between the insertion site of the small arm of homology and a BsrG1 digestion site. Successfully targeted ES cell clones, contain an additional BsrG1 restriction site, appear as two bands, a 12 kb band (Wt) and a mutant band of ~9 kb, whereas a non targeted ES cell line will have a single band at 12 kb (**Fig. 1C**). The A4 and B5 clones, which satisfied the above criteria, were sent to Yves deRepentigny (Kothary Lab) for blastocyst injection. The B5 clone provided 2 chimeric males (**Fig. 1D**) that were fertile while the A4 clone produced few chimeras none of which were fertile. Offspring from a cross between a B5 chimeric male with a Wt FVB female results in mice with an agouti coat colour and positive for the targeted construct by long range PCR indicative of germ line transmission (**Fig. 1E,F**). To validate YFP expression, founders (henceforth called P7YFP$^{+/-}$) were subsequently backcrossed with sv129 females or, in the event that the PGKneo cassette was affecting the expression of YFP driven by the *Pax7* promoter, crossed with a CMVcre$^{+/+}$ mouse that expresses Cre under the constitutive cytomegalovirus promoter [303,304].

While mouse work progressed, the J1 B5 P7YFP$^{+/-}$ ES cell line was differentiated *in vitro* to assess YFP expression. Prior to differentiation, J1 B5 P7YFP$^{+/-}$ ES cells were treated for 48 hrs to remove the PGKneo cassette with Tat-Cre, a soluble catalytically active form of Cre that has an N-terminal Tat protein transduction domain to facilitate cell entry (**Fig. S1** and **Fig. S2**)[304,305]. Following Tat-Cre treatment ES cells were

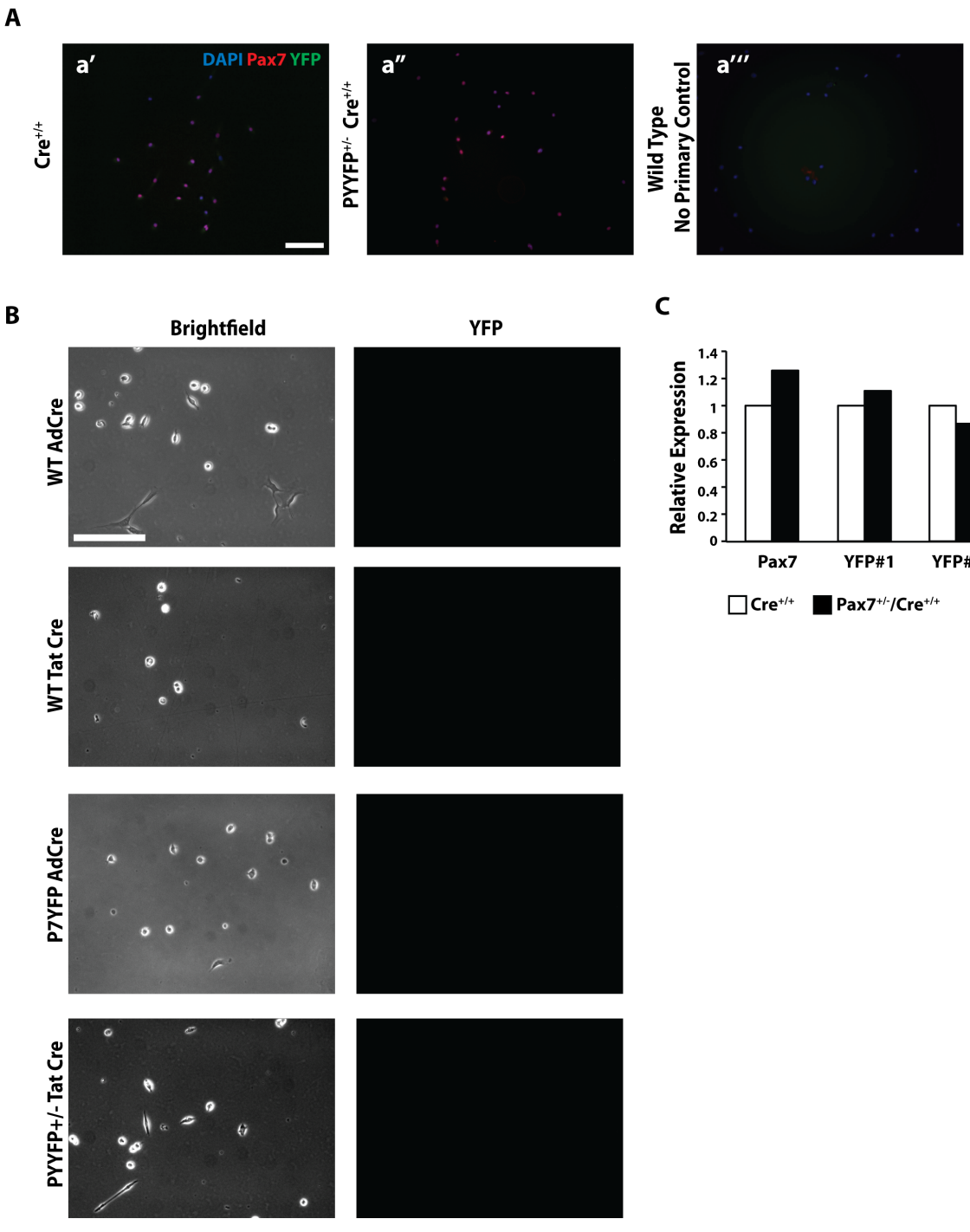
**Figure 2. Analysis of YFP expression in differentiated ES cells.**
Differentiated TGFP cells on Day 6 brightfield (**A**) and GFP (**B**). Assessment of YFP expression from J1 B5 P7YFP$^{+/-}$ ES cells following 25 days of differentiation. Live differentiated cells brightfield (**C**) and the YFP channel (**D**). Fixed differentiated cells were stained for YFP, Pax7 and Myosin heavy chain, magnification 100x (**E**) and at higher magnification 200x (**F**). Scale bar denotes 100 μm.


A
B
YFP
C
D
YFP
E
MyHC Pax7 YFP
F
MyHC Pax7 YFP

differentiated as embryoid bodies with subsequent adherence culture for a total of 25 days and assessed for the presence of YFP. As a fluorescent control the Brachyury GFP (TGFP) cell line was used to assess the production of GFP upon differentiation (**Fig. 2A and B**). Upon differentiation of Pax7YFP$^{+/-}$ ES cells no YFP was detected in live cells (**Fig. 2 C,D**) nor was YFP observed upon fluorescent staining of skeletal muscle positive for Pax7 and Myosin heavy chain (**Fig. 2 E,F**). These results indicated that upon endogenous activation of the *Pax7* promoter YFP is either not being expressed or expressed at very low levels such that it cannot be detected upon immunofluorescent staining. The best cell type to assess the expression of *Pax7* is the primary myoblast. Primary myoblasts were isolated from adult P7YFP$^{+/-}$/Cre$^{+/+}$ limb musculature and cultured *in vitro*. No YFP expression was observed in live P7YFP$^{+/-}$/Cre$^{+/+}$ or Cre$^{+/+}$ myoblasts (data not shown). Furthermore, no YFP was observed upon antibody staining for YFP in Cre$^{+/+}$ or P7YFP$^{+/-}$/Cre$^{+/+}$primary myoblasts (**Fig. 3a' and a''**), single muscle fibers or whole muscle sections (data not shown). In addition primary myoblasts from P7YFP$^{+/-}$ were either transfected with a Cre expressing plasmid treated with an adenovirus expressing Cre or treated with Tat-Cre, No YFP was observed in P7YFP$^{+/-}$ in any of the above-mentioned conditions (**Fig. 3B**). Finally, qPCR analysis of cDNA produced from proliferating P7YFP$^{+/-}$ primary myoblasts treated with Tat-Cre and compared to wild type myoblasts displayed no enrichment in YFP signal (**Fig. 3C**). Considering *Pax7* is robustly expressed in primary myoblasts these results suggest that YFP is not being produced upon *Pax7* activation.

In an attempt to eliminate any effects of the artificial intron present in the original *Pax7* ESCV 2.0 vector and increase fluorescent protein production from the *Pax7*

**Figure 3. Analysis of YFP expression in primary myoblasts treated with Cre recombinase.**
(**A**) Analysis of YFP expression in $Cre^{+/+}$ (**a'**) or $P7YFP^{+/-}/Cre^{+/+}$ primary myoblasts (**a''**). A control without primary antibodies is included to provide relative non-specific background staining of the secondary antibodies (**a'''**) scale bar denotes 100 μm. (**B**) Wild type or $P7YFP^{+/-}$ primary myoblasts were treated with AdCre (MOI 20) or Tat-Cre (10 μM) and assessed for YFP expression scale bar denotes 100 μm. (**C**) RTPCR was conducted for Pax7 and YFP on cDNA isolated from either $Cre^{+/+}$ or $P7YFP^{+/-}/Cre^{+/+}$ primary myoblasts.

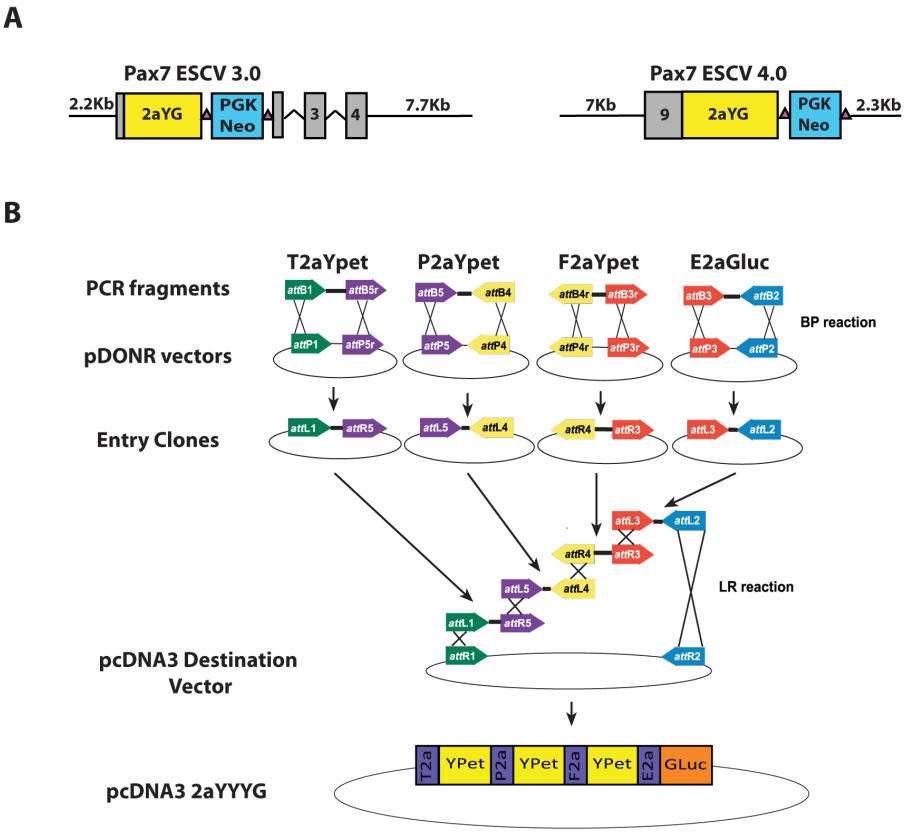

A
a′
DAPI Pax7 YFP
Cre+/+
a″
PYYFP+/- Cre+/+
a‴
Wild Type
No Primary Control
B
Brightfield
YFP
WT AdCre
WT Tat Cre
P7YFP AdCre
PYYFP+/- Tat Cre
C
Relative Expression
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Pax7
YFP#1
YFP#2
Cre+/+
Pax7+/-/Cre+/+

promoter two subsequent vectors were designed – Pax7 ESCV 3.0 and Pax7 ESCV 4.0 (**Fig. 4A**). Both of these vectors incorporate a 2a self-cleaving peptide interspersed between three Ypet (enhanced YFP) proteins and a secretable *Gaussia* Luciferase[301,306,307]. These modifications were made to increase the observable fluorescence in the event that Pax7 transcription was weak and to provide a method to track targeted cells *in vivo*[308]. In addition both of these new vectors did not incorporate splice donor or acceptor sites and were targeted to either the second exon or ninth exon of Pax7 as opposed to the first exon (**Fig. 1A** and **Fig. 4A**). The 2aYpetgLuc vector created was confirmed by DNA sequencing (data not shown) and temporally cloned into a pCDNA3 vector to validate protein production regulated by a constitutive CMV promoter (**Fig. 4B**). The pcDNA3 YYYG vector when transfected into 293T cells successfully expressed robust levels of Ypet and GLuc (**Fig. 4C,D**). Work on these vectors was ceased shortly after confirmation of the 2aYpetgLuc in 293T cells to focus on the creation of a BAC transgenic for Myf5 (outlined below) and other experiments detailed in Chapters 2 and 4.

*3.3.2 Myf5 Transgenic Models*

The *Myf5* targeting construct M5 ESCV 2.0 contained TdTomato, a dsRed variant, followed by an artificial intron cassette and was fashioned in an identical way to the original P7 ESCV (**Fig 5A**). The *Myf5* MTV (M5 MTV 2.0) was targeted to a *Myf5* BAC (clone 229I10) containing 99 kb upstream of the translational start site and 60 kb downstream of the stop codon for *Myf5*. A region of the targeting BAC encompassing ~14.5 kb upstream and ~3 kb downstream from the targeting site was subsequently retrieved into the PGKdtaBpA vector. Following verification by restriction enzyme

**Figure 4. Design and validation of a 2aYpetGluc expression vector.**
(**A**) Pax7 ESCV 3.0 and 4.0 targeting vectors include triple Ypet coding sequences along with a secreted *Gaussia* Luciferase and polyadenylation signal all separated by 2a self cleaving peptide sequences. (**B**) Schematic outlining the multisite gateway process involved in creating the 2aYpetGluc (YYYG) expression construct. (**C**) Validation of Ypet expression from the pcDNA3 2aYYYG vector in 293T cells. Magnification 50x, scale bar represents 100 μm. (**D**) Validation of secreted *Gaussia* luciferase from the pcDNA3 2aYYYG construct, untransfected cells were used as a negative control while cells transfected with the CMVGluc plasmid serve as a positive control.

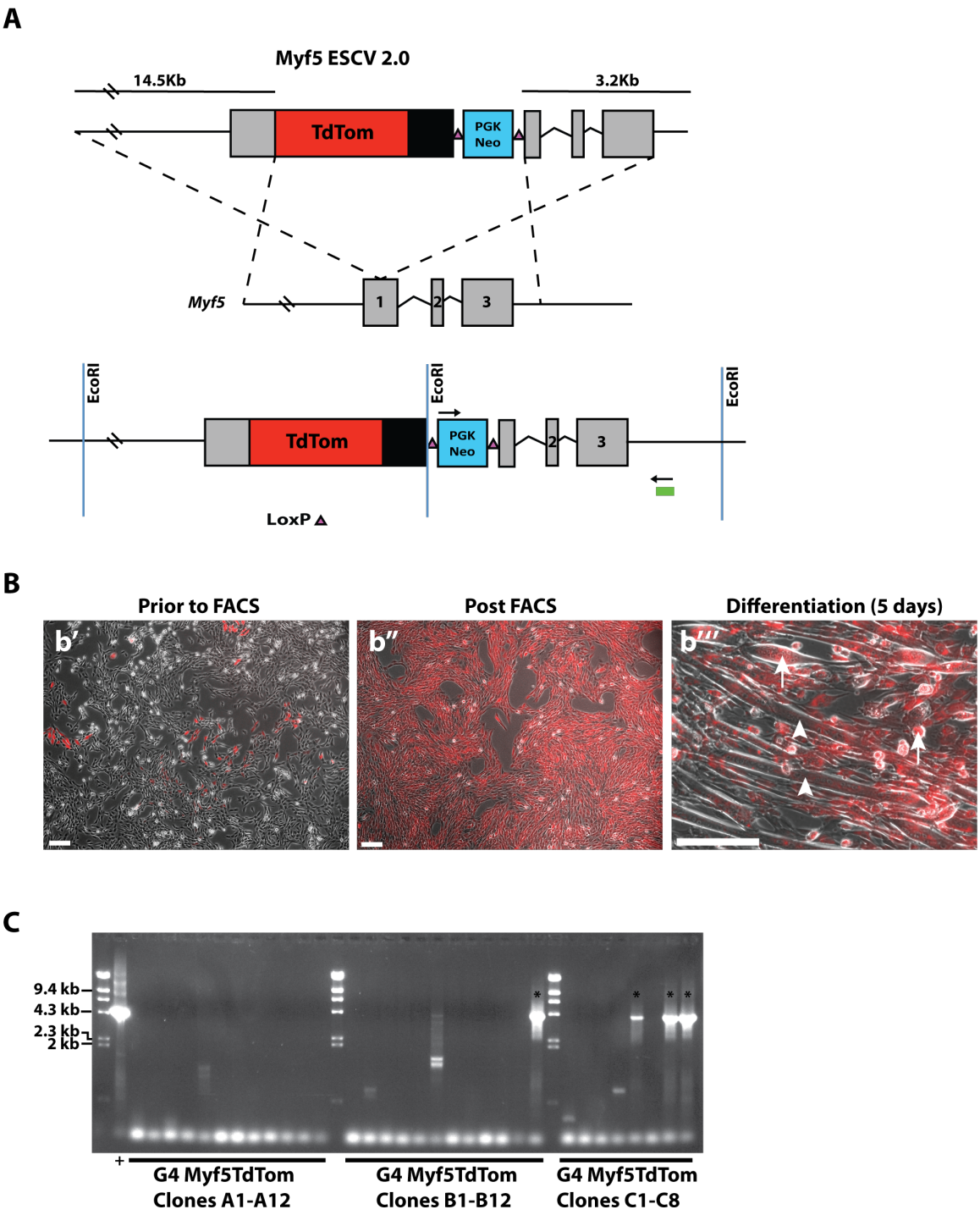

A
Pax7 ESCV 3.0
2.2Kb
2aYG
PGK Neo
3
4
7.7Kb
Pax7 ESCV 4.0
7Kb
9
2aYG
PGK Neo
2.3Kb
B
T2aYpet
P2aYpet
F2aYpet
E2aGluc
PCR fragments
pDONR vectors
BP reaction
Entry Clones
LR reaction
pcDNA3 Destination Vector
pcDNA3 2aYYYG
T2a
YPet
P2a
YPet
F2a
YPet
E2a
GLuc
C
YPet
D
Relative Luciferase Units Log (RLU)
1.E+08
1.E+06
1.E+04
1.E+02
1.E+00
Neg Ctrl
pCMV GLuc
YYYG Clone 1
YYYG Clone 2

digestion and DNA sequencing (data not shown) the M5 ESCV 2.0 targeting construct was electroporated into G4 ES cells and selected in 200 μg/ml of G418. Concurrently, as a means to evaluate the expression of TdTom in response to *Myf5* induction we transfected the M5 MTV 2.0 targeted BAC into C2C12 myoblasts[309]. After transfection of the M5 MTV 2.0 and selection in neomycin numerous cells displayed TdTom fluorescence. FACS sorting provided us with a purified population of TdTom C2C12 cells (**Fig. 5B'-B''**). Upon differentiation these cells down regulated the expression of TdTom thus recapitulating the expression pattern of *Myf5* from an undifferentiated to differentiated state (**Fig. 5B'''**).

Encouraging results in C2C12 cells led to the targeting of M5 ESCV 2.0 in the G4 mouse ES cell line and the creation of a transgenic BAC reporter mouse using the BAC targeted intermediate of M5 MTV 2.0. Mouse G4 ES cells were successfully targeted using the M5 ESCV 2.0 and clones (G4 M5TdTom) were verified using long range PCR (**Fig. 5C**). Of the 384 clones picked 23 (5.9%) displayed a positive 3.5 kb amplicon. Preliminary *in vitro* differentiation of G4 M5TdTom clones did not result in visible skeletal muscle or in TdTom positive cells after 25 days of differentiation (data not shown). In light of progress made in the creation of a *Myf5* BAC transgenic work on the M5TdTom clones was suspended.

In order to create a M5 BAC transgenic reporter mouse our targeted *Myf5* BAC (targeted with M5 MTV 2.0) was transformed into the EL350 *E. coli* strain in order to express Cre and remove the LoxP flanked PGK Neo cassette. Bacterial clones containing *Myf5* BACs that have their PGK Neo cassette removed will die upon exposure to Kanamycin while those that have not undergone proper Cre expression/excision will

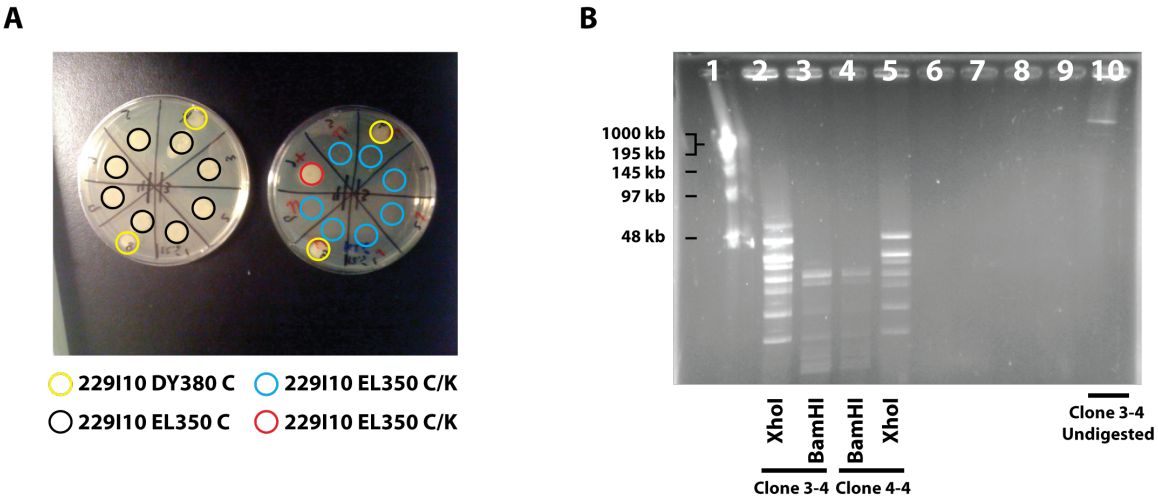
**Figure 5. Design and validation of *Myf5* transgenic vectors.**
(**A**) Schematic structure of the mouse *Myf5* gene locus along with the Myf5 ESCV 2.0 targeting vector. Myf5 ESCV 2.0 includes a Tdtomato (TdTom) coding sequence in red followed by an artificial intron (ArtINT) in black comprising splice donor and acceptor sites and a SV40 polyadenylation signal. Purple triangles represent LoxP sites and flank a blue PGKNeo cassette. Vertical blue lines represent EcoRI restriction enzyme digestions sites, while black arrows designate primers and a green rectangle the hybridization probe for Southern blotting. (**B**) Verification of TdTom fluorescence in C2C12 cells post electroporation of the M5 MTV 2.0 target BAC and selection (**b'**), post FACS enrichment for TdTom expressing cells (**b''**), following differentiation BAC integrated C2C12s fuse to form multinucleated myotubes (**b'''**). Prior to fusion, myoblasts express TdTom (white arrows) yet terminally differentiated myotubes down regulate their expression of TdTom (white arrow heads). Images b' and b'' magnification of 50x while b''' is 200x. Scale bars denote 100 μm. (**C**) Long range PCR screening of Myf5 ESCV 2.0 targeted G4 ES cell lines. Amplicon of a BAC targeted by Myf5 MTV 2.0 serves as a positive control with a band that migrates at ~4.3 kb. A λ HindIII ladder is used as a marker. Asterisks label positive clones (B12, C5, C7, C8).


A
Myf5 ESCV 2.0
14.5Kb
3.2Kb
TdTom
PGK Neo
Myf5
1
2
3
EcoRI
EcoRI
EcoRI
TdTom
PGK Neo
2
3
LoxP
B
Prior to FACS
Post FACS
Differentiation (5 days)
b'
b''
b'''
C
9.4 kb
4.3 kb
2.3 kb
2 kb
+
G4 Myf5TdTom Clones A1-A12
G4 Myf5TdTom Clones B1-B12
G4 Myf5TdTom Clones C1-C8

survive in Kanamycin containing media (**Fig. 6A**). Two bacterial clones correctly targeted by M5 MTV 2.0, showing positive excision of the PGK Neo cassette, were subsequently digested with restriction enzymes XhoI or BamHI to estimate their restriction digest patterns while an undigested version was electrophoresed to determine the state of degradation of the BAC prior to pronuclear injection (**Fig. 6B**).

Following two rounds of pronuclear injection 6 founders were identified from a pool of 73 offspring (5 male, 1 female) that contained the TdTom sequence (**Fig 6C**). Differing views exist regarding the efficiency of BAC pronuclear injection including the state of the BAC (circular vs linear) and the injection medium for resuspension. We injected a circular BAC dissolved in polyamine injection buffer according to a comprehensive study summarizing the efficiency of differing techniques for BAC pronuclear injection[310]. As such, a circular BAC will be linearized and integrate randomly in the genome. In order to ensure the maximum amount of *Myf5* regulatory sequences are contained upon integration primers were designed that originate in the backbone of the BAC and cross into the most 5' and 3' sequences of *Myf5*. In addition to TdTom, founders were genotyped at their 5' and 3' extremities to ensure the entire region of *Myf5* is intact upon BAC integration. Following genotype confirmation embryos were isolated at various embryonic time points (E9.5-16.5) and assessed for TdTom fluorescence under the dissecting microscope. *Myf5* is expressed in the somites of developing embryos resulting in an easily recognizable pattern of expression. No positive TdTom fluorescent embryos were detected at any of the embryonic time points. No TdTom expression was detected via confocal microscopy in intercostal muscles (E13.5) and diaphragm muscles (E16.5), containing *Myf5* positive cells. One embryo, from founder 958, was fixed, frozen and

**Figure 6. Creation and analysis of a Myf5 TdTom BAC transgenic.**
(**A**) Transfer of the Myf5 MTV 2.0 targeted BAC into the EL350 *E. coli.* strain for subsequent Cre removal of the LoxP flanked PGKNeo cassette. Positive Myf5 targeted BAC clones in EL350 that underwent PGKNeo removal are highlighted by blue circles, while red circles indicate a BAC clone that remains resistant to Kanamycin. C indicates agar containing Chloramphenicol (12.5 μg/ml), K indicates Kanamycin (12.5 μg/ml). (**B**) Pulse field gel electrophoresis to determine the quality of Myf5 MTV 2.0 BAC prior to pronuclear injection. Non digested BAC DNA is run in lane 10. Lambda ladder PFG marker is run in lane 1. Lanes 2-5 show BAC clones 3-4 and 4-4 following digestion with either XhoI or BamHI. (**C**) TdTom genotyping of Myf5 TdTom BAC transgenic founders along with their offspring following a cross to CD-1 females or in the case of #962 a male. (**D**) Yellow arrowheads mark mice from which primary myoblasts were isolated and assessed for Tdtom expression. Scale bar denotes 100 μm.


A
229I10 DY380 C
229I10 EL350 C/K
229I10 EL350 C
229I10 EL350 C/K
B
1 2 3 4 5 6 7 8 9 10
1000 kb
195 kb
145 kb
97 kb
48 kb
XhoI
BamHI
BamHI
XhoI
Clone 3-4 Clone 4-4
Clone 3-4
Undigested
C
Myf5
1
2
3
Wt amplicon = ~650bp
Myf5 MTV 2.0
Targeted BAC
1
TdTom
2
3
TdTom amplicon = ~1Kb
Tdtom
Wt
904 962 923 958 934 949
+ -
1 2
934
947
3
904
934
947
962
4
5
923
958
958
904
D
Brightfield
TdTomato
TdTom
947-1
947-2
962-3
923-4
958-5

sectioned to assess the presence of TdTom: no TdTom fluorescence was detected in sagittal sections. Offspring from crosses between *Myf5* TdTom founders and CD-1 females were genotyped for TdTom and selected pups were chosen for primary myoblast isolation. Live imaging of primary myoblasts isolated from wild type and TdTom mice positive at the *Myf5* locus displayed no TdTomato fluorescence (**Fig. 6D** and summarized in **Table I**). At this point further experiments on this project were suspended.

## 3.4 Discussion

The intention of this research was to create genetic reporters for two critical transcription factors (Pax7 and Myf5) involved in skeletal muscle development and regeneration. When this project was conceived no fluorescent reporters existed for either Pax7 or Myf5. By using two fluorophores with distinct emission spectra (YFP/YPet or TdTom) under the control of the Pax7 and Myf5 promoters respectively we could effectively sort both of these populations from one pool of cells. This method would provide a powerful approach to isolate myogenic populations from differentiating ES cells or mice. However, neither the targeted ES cells lines, the Pax7YFP transgenic mouse line, nor the Myf5TdTom BAC transgenic mouse line expressed a fluorophore upon the endogenous activation of either Pax7 or Myf5.

Prior to recombination occurring in ES cells the Pax7YFP construct was sequenced to confirm it's identity. Each component (YFP, artificial intron, polyA sequence) contained no mutations at the DNA level and they resided in frame for translation beginning at the Pax7 start site. Had translation begun, a functional YFP protein would have been produced. True enough upon transfer of the YFPArtINT into a pcDNA3 vector and subsequent transfection into 293T cells YFP fluorescence was

observed. If the YFPArtINT squences were functional when placed in a plasmid, it is possible that the targeting of this construct to the BAC or in ES cells went awry. C2C12 cells, an *in vitro* model system for adult skeletal muscle differentiation, do not express significant amounts of Pax7. Primary myoblasts isolated from adult skeletal muscle do express Pax7 yet senesce after multiple passages and are not amenable to BAC electroporation. Therefore, we were only able to test the functionality of the Myf5TdTom construct in C2C12 cells. To ensure the correct targeting of the Pax7YFP MTV into the Pax7 BAC multiple DNA sequencing reactions were performed. Having previously confirmed the sequence of the Pax7YFP MTV, upon BAC recombination we sequenced across the insertion sites of the regions of homology by using a reverse primer originating at the start of the YFP coding sequence and a forward primer in the NEO cassette. These reactions demonstrated that the recombineering step between the Pax7YFP MTV and BAC (52K01) had been successfully targeted in frame to the first exon of Pax7. Following the retrieval recombineering step to create the Pax7YFP ESCV a complete set of sequencing reactions were conducted to confirm the sequence of YFP, ArtINT, PgkNeo, and the beginning and end sequences of the short and long arm of homology. All of these sequencing reactions confirmed that our targeting construct was correct. Following homologous recombination of the Pax7YFP ESCV 2.0 into ES cells, screening via long range PCR and Southern blot confirmed that targeting had occurred at the Pax7 locus. Because Pax7 is not expressed in undifferentiated ES cells confirmation of YFP expression necessitated differentiation into the skeletal muscle lineage. While the Pax7YFP B5 ES cell derived Pax7YFP$^{+/-}$ mouse chimera was being created, *in vitro* differentiation of the ES cell line was conducted.

The purification and validation of Tat-cre along with it's published use in ES cells to remove loxP flanked sequences without affecting pluripotency enabled the treatment of Pax7YFP B5 ES cells prior to differentiation. By treating with Tat-cre we effectively removed the PGKNeo sequences decreasing any interference from the strong PGK promoter, which could negatively affect the expression of YFP. Upon differentiation skeletal muscle was observed based on morphology and confirmed via immunofluorescent detection of Pax7 and Myosin heavy chain protein. YFP fluorescence was not observed in live cells adjacent to spindle shaped skeletal muscle nor was it detected using a GFP antibody validated by staining Myf5-Cre ROSA26YFP myoblasts (Data not shown). Based on an inability to detect YFP in live cells before or after treatment with Tat-Cre or in fixed cells expressing *Pax7* it is likely that the YFP protein is not being produced during ES cell differentiation. The absence of YFP expression is presumably due to the same reasons that affect expression of YFP in the Pax7YFP$^{+/-}$ mouse.

We isolated primary myoblasts from Pax7YFP$^{+/-}$ mice to test for the expression of our targeted reporter because primary myoblasts endogenously express *Pax7* at high levels. Live primary myoblasts did not express YFP confirming that our reporter was unable to recapitulate endogenous *Pax7* expression. All results positively support the correct targeting of our construct into the *Pax7* locus in ES cells, expression of YFPArtINT in 293T cells and correct sequencing of the targeting vector. The general size and critical enhancer/repressor regions of the *Pax7* promoter currently remain unknown. As such, it is possible that a critical *Pax7* enhancer was disrupted in our targeting construct. Considering that no YFP mRNA is produced in primary myoblasts another

possible explanation could potentially involve alternate splicing that occurs upon introduction of an exogenous splice donor and acceptor site contained within the ArtINT sequence introduced with YFP. In summary, potential alterations affecting the enhancer/repressor efficiency, alternative splicing, and the disruption of the DNA tertiary structure could negatively impact the transcription or mRNA stability of YFP and explain a lack of YFP expression from the Pax7YFP$^{+/-}$ Knock-in mice.

The Myf5TdTom constructs followed an identical validation process throughout cloning to that of Pax7. The Myf5TdTom vector is built in an identical fashion to the Pax7 vectors except that a different fluorophore (TdTomato) was used along with sequences specific for the *Myf5* locus. One advantage inherent to *Myf5* is that validation of BAC driven expression of TdTom could be carried out in C2C12 cells. Encouraging results demonstrated that TdTom was indeed expressed in C2C12 cells upon introduction of the Myf5TdTom vector. TdTom expression was reduced upon differentiation and fusion of myotubes suggesting appropriate regulation by the *Myf5* promoter. Following integration and long range PCR screening of Myf5TdTom clones displaying a positive band indicating specific integration at the *Myf5* locus were observed. However, upon differentiation G4 ES cell lines did not produce skeletal muscle following our routine differentiation procedure. Without the generation of skeletal muscle it was not surprising that no TdTom was ever detected in differentiating targeted G4 cell lines. These results could potentially be due to the same factors that affected the Pax7YFP clone that shared a similar targeting vector layout. Positive data from transfection of the targeted M5 MTV 2.0 into C2C12s would suggest that abnormalities in the karyotype or phenotypic

alterations in the differentiation capacity of the G4 targeted cell lines could account for the absence of skeletal muscle following differentiation.

We chose to focus our efforts on the creation of a *Myf5* BAC transgenic mouse mainly due to positive data from the introduction of this targeted BAC into C2C12 cells. This mouse would have been crossed with a newly acquired Pax7 reporter (Pax7zsGreen, discussed in depth in Chapter 4) offering the possibility to address questions raised early during graduate work on the differences between embryonic and adult muscle progenitors.

The *Myf5* TdTom BAC was successfully prepared for pro-nuclear injection following removal of the PGKNeo cassette and confirmation of general quality by pulse-field gel electrophoresis. An appropriate ratio of founders was derived after pronuclear injection with 6 founders containing the furthest extremities of the BAC in addition to the TdTom reporter. It was important to verify that the BAC along with the targeted sequences had integrated in their entirety. Therefore, the greatest amount of *Myf5* regulatory sequence existed to drive expression of TdTom. Ultimately upon numerous attempts spanning varying time points and cell types no TdTom expression was observed at either the protein or RNA level. Regulation at the *Myf5* locus involves numerous enhancers that regulate expression in a temporal and spatial manner. It has been proposed that this complexity allows *Myf5* to respond to diverse inductive signals through binding of upstream regulators en route to skeletal muscle commitment[311].

Recent research in the lab has identified a critical enhancer in the *Myf5* promoter that is situated at -111 kb and is regulated by *Pax7*. *Pax7* is able to bind this element and drive expression of *Myf5* in the adult[311]. Currently it remains unclear if *Pax7* regulates this

element in earlier fetal or embryonic stages. In the adult upon the absence of the -111 kb element quiescent satellite cells do not show any *Myf5* expression. Incidentally the BAC used to create the Myf5TdTom transgenic mouse contains -99 kb of sequence upstream of the *Myf5* translational start site and so does not contain this -111 kb element. It is possible that the absence of the -111 kb enhancer accounts for the loss of TdTom expression observed in this transgenic line. It is equally possible that factors (splicing, effects on enhancers/repressors, effects on tertiary structure etc…) involved in the suppression of YFP expression in the Pax7 mouse and cell lines could be at play here as well.

It would be beneficial for future attempts to create genetically engineered cell lines or mice for *Pax7* and *Myf5* to use previously confirmed reporter constructs, target multiple sites of a given gene, and focus initially on BAC transgenesis in order to expedite this process and ensure the highest rate of success.

## 3.5 Materials and Methods

### Cloning of Transgenic vectors

The *Pax7* (52K01, RPCI-22) and *Myf5* (229I10, RPCI-22) bacterial artificial chromosomes (BACs) were obtained from the Children's Hospital Oakland Research Institute (CHORI) or The Center for Applied Genomics (TCAG). To create the *Pax7* mini targeting vector (MTV 2.0) the following cloning steps were conducted. Linking the 5' region of *Pax7* to a YFP coding region that lacked a stop codon was accomplished by overlapping PCR using the 52K01 BAC and a eYFP plasmid (peYFP-N1, Clontech) with the *Pax7* 5' Domain FWD primer (0.05 μM final concentration) a Pax7-eYFP Primer

(0.005 μM final concentration) and the eYFP REV primer (0.05 μM final concentration). The artificial intron region was PCR amplified from the pBK-CMV plasmid (Stratagene) with ArtINT pA FWD and ArtINT pA REV. The *Pax7* 3' region, which includes the start codon of *Pax7*, was PCR amplified with the *Pax7* 3' Domain FWD and *Pax7* 3' Domain REV primers. Each PCR fragment was TA cloned (Invitrogen K2050-01) and sequenced (Stemcore) to assure the correct region was amplified. Next a 5 way ligation was conducted with the following fragments SalI-Pax7eYFP-KpnI KpnI-ArtINT-EcoRI EcoRI-PGKNeo-BamHI BamHI-Pax7 3'-NotI using the PGKNeo cassette from the pl452 vector (NCI-Frederick) and the pl452 backbone. The P7 MTV vector was then sequenced to confirm its identity. The Myf5 MTV 2.0 was constructed in a similar fashion to that of Pax7. Linking the 5' region of *Myf5* to a TdTom coding region that lacked a stop codon was accomplished by overlapping PCR using the 229I10 BAC and a TdTomato plasmid (kind gift from Roger Tsien) with the *Myf5* 5' Domain FWD (0.05 μM final concentration) a Myf5-Tdtom primer (0.005 μM final concentration) and the TdTom REV primer (0.05 μM final concentration). The artificial intron region was PCR amplified from the pBK-CMV plasmid (Stratagene) with ArtINT pA FWD and ArtINT pA REV. The Myf5 3' region was PCR amplified with the Myf5 3' Domain FWD and Myf5 3' Domain REV primers. The previously mentioned sequencing and ligation procedures were conducted to create the Myf5 MTV 2.0.

The DY380 *E. coli* strain was transformed with either the Pax7 BAC (52K01) or the Myf5 BAC (229I10) and recombineering was conducted according to Liu et al[312]. Briefly DY380 52K01 or 229I10 *E. coli* were grown overnight at 32$^{o}$C with shaking at 225 rpm. The following morning 1 ml of the overnight culture was transferred to 20 ml of LB and

grown for 2 hrs until an OD of 0.5. Cells were subsequently transferred to a 42$^{o}$C water bath and incubated for 15 min to induce the α,β,γ recombineering machinery. The cells were quick cooled in ice water and centrifuged (4000 rpm, 0$^{o}$C. 5 min). The pellet was resuspended in 888 μl of ice cold water and this was repeated 3 times. Electroporation was conducted by resuspending DY380 cells containing either a Pax7 or Myf5 BAC in 50 μl of ice cold water with ~10 ng of Sal1 linearized MTV targeting vector and pulsing at 1.75 kV, 25 μF and 200 Ω resulting in a time constant of ~4 ms. Electroporated cells were diluted in warm LB medium and cultured for 1 hr at 32$^{o}$C followed by plating on LB agar plates with 50 μg/ml kanamycin and 20 μg/ml chloramphenicol. To retrieve regions from the targeted BACs into the ESCV (PGKdtaBpa), the following retrieval primers were used. Pax7 ESCV 2.0 (long arm, ~12.5 kb, Pax7 2.0 Ret 5' FWD/REV and short arm ~3 kb, Pax7 2.0 Ret 3' FWD/REV) while for Myf5 ESCV 2.0 (long arm ~8 kb, Myf5 2.0 Ret 5' FWD/REV, short arm ~2 kb, Myf5 2.0 Ret 3' FWD/REV). The PGKdtaBpa vector was digested with SalI and EcoRI and a 3-way ligation was performed with backbone SalI-PacI-5'Ret-HindIII and HindIII-3'Ret-EcoRI. Prior to recombineering to retrieve the targeted portion of the BAC the PGKdtaBpa containing the retrieval arms is linearized by digestion with HindIII. All primers are located in Supplementary Table 3.1.

**Creation of 2aYpet GLuciferase vector**

The creation of the 2a linked triple Ypet and Gaussia luciferase vector was conducted according to the methods outlined in the Multisite Pro 4.0 Gateway Kit (Invitrogen, 12537-104). Briefly, Ypet cDNA (Addgene 14032) was amplified with specific attB primers (Sup Table 3I) with T2a, P2a, and F2a. The *Gaussia* luciferase was amplified

with attB primers containing the E2a sequence. These four PCR products were run on 1% agarose gels and purified using a Gel Extraction Kit (Qiagen). Once purified a BP reaction was conducted between the PCR fragments and a specific pDONR vector using BP clonase. This vector was transformed into competent TOP10 *E.coli* (Invitrogen), amplified and sequenced to confirm it's identity. Subsequently a four way LR reaction is conducted to recombine the entry clones with the destination vector, which in this case is a pcDNA3 vector containing attR sites.

**Preparation of Myf5 MTV 2.0 targeted BAC for pronuclear injection.**

M5 MTV 2.0 targeted 229I10 BAC DNA was purified from DY380 *E. coli* and transformed into EL350 *E. coli* in order to remove the loxP flanked PGKNeo cassette. The EL350 *E. coli* strain contains a temperature inducible promoter driving Cre recombinase. When cultured at 37$^{o}$C, Cre is induced catalyzing the removal of the PGKNeo cassette present in the M5 TdTom reporter. Following the removal and confirmation of PGKNeo the targeted BAC was amplified in 1L of LB broth with 20 μg/ml of Chloramphenicol at 37$^{o}$C. Purification of BAC DNA was conducted on Nucleobond PC500 Columns (Macherey-Nagel, 740574). For verification targeted BACs were digested with either XhoI or BamHI and electrophoresed on 1% agarose gels at 4.5 v/cm, 16$^{o}$C for 24 hours using a CHEF Mapper XA pulse field gel electrophoresis box (BioRad, 170-3670) with a lambda ladder (NEB, N0340S). For pronuclear injection purified BAC DNA was adjusted to a concentration of 1 ng/μl in polyamine (PA) buffer (10 mM Tris pH 7.5, 0.1 mM EDTA, 30 μM Spermine, 70 μM Spermidine, 100 mM NaCl)[310].

**Cell culture**

The mouse ES cell lines J1 and G4 were maintained on a layer of irradiated MEFs (DR4) in DMEM high glucose medium (Invitrogen) supplemented with 15% fetal bovine serum (Hyclone), 1x non essential amino acids, 1x sodium pyruvate, 1x Glutamax, β-Mercaptoethanol, 1x Pen/Strep (Invitrogen), and 1000U ESGRO (chemicon) at 37$^{o}$C and 5% $CO_2$. Prior to differentiation ES cell lines were cultured for two passages on 0.2% gelatine (Sigma G1890) to remove the majority of MEFs. For embryoid body formation ES cells were seeded in 20 μl hanging drops at a density of $4x10^4$ cells/ml in ES cell medium without LIF. Embryoid bodies were transferred to suspension after 3 days and cultured in 60 mm petri dishes (VWR) on an analog shaker (Thomas Scientific). Following 2 days of culture in suspension embryoid bodies were plated on 0.2% gelatine coated 10 cm dishes for the remainder of the differentiation protocol. The mouse myoblast cell line C2C12 was cultured in DMEM (Wisent) with 10% FBS. To induce differentiation C2C12 cells were cultured in DMEM with 5% horse serum (Hyclone) for 5 days. ROSA26YFP fibroblasts were cultured in the same medium as C2C12s. All cell culture was conducted in Steri-cycle 370 incubators (Thermo) at 37$^{o}$C and 5% $CO_2$.

**Electroporation of ES cell for the creation of Pax7YFP/ Myf5Tdtom knock-in cell lines**

Mouse embryonic stem cells (J1, G4) were washed with phosphate buffered saline (PBS), typsinized with 0.25% trypsin (Invitrogen), and resuspended in ice cold Embryomax electroporation buffer (Chemicon). 25-40 μg of PacI-linearized embryonic stem cell targeting vector (ESCV) was added to ES cells at a concentration of ~$1.0x10^7$ cells/ 700 μl of electroporation buffer and pulsed twice in a 0.4 cm Gene Pulser cuvette (Biorad) at 220 V, 0 Ω, and 500 μF using a ECM 630 electroporator (BTX, Genetronics).

Electroporated cells were mixed with prewarmed medium and plated on irradiated DR4 MEFS for 24 hrs prior to selection with Geneticin (G418, Invitrogen) for ten days at a concentration of 250 μg/ml. Clones were picked for expansion between 7-10 days post selection.

**Long range PCR and Southern Blot**

Genomic DNA from potential clones was isolated using the DNeasy 96 well kit (Qiagen). Following isolation gDNA was typically diluted 1:100 prior to long range PCR. Long range PCR was conducted using the Qiagen long range PCR kit (Qiagen) with primers for *Pax7* or *Myf5* (**Table SI**). For Southern blot analysis gDNA was prepared from ES cell clones by incubation overnight in tail buffer (100 mM Tris pH8.5, 5 mM EDTA, 0.2%SDS, 200 mM NaCl, 100 μg/ml proteinase K) at 55$^{o}$C. Between 10 to 20 μg of gDNA was digested overnight at 37$^{o}$C with BsrG1 in NEB buffer 2 and BSA. Digested gDNA was loaded on a 0.8% agarose gel and run at 110 V for 3 hrs. Next the gel was stained with EtBr for 10 mins at room temperature. The gel was subsequently exposed to UV light to nick the DNA (GS gene linker BioRad) and incubated for 1 hour in denaturation buffer (0.5N NaOH and 1.5 M NaCl) and 1 hour in neutralization buffer (1 M Tris and 1.5 M NaCl pH 7.5). The gel was transferred overnight in 2xSSC (300 mM NaCl and 30 mM NaCitrate) onto a Hybond N+ membrane (Amersham, RPN303-T). The gel was crosslinked and incubated in hybridization buffer (4x SSC, 0.1% SDS, 5x Denhardts, ssDNA). A Pax7 probe was generated by PCR and radiolabelled with P$^{32}$ dCTP using the Rediprime Kit (Amersham) and purified on a G50 micro column (GE, Illustra Probe Quant). The radiolabelled probe was incubated with the membrane in hybridization solution at 65$^{o}$C overnight. The blot was washed with solutions containing

increased SDS and decreasing salt concentrations. Finally the membrane was exposed to Kodak MS film (Kodak) and developed 24-48 hours later.

**Cre removal of PGKNeo**

Excision of the PGKNeo cassette was conducted by either *in vivo* or *in vitro* based methods. The *in vivo* based method involved crossing the Pax7YFP$^{+/-}$ mouse with a CMVcre$^{+/+}$ mouse (BALB/c-Tg(CMV-cre)1Cgn/J: 003465: Jackson Laboratories[303]). The *in vitro* based methods involved transfection, infection or transduction with activated Cre recombinase. A PGK-nlsCre plasmid containing a Cre recombinase containing a nuclear localization sequence was transfected into 150,000 Pax7YFP$^{+/-}$ cells from the B5 clone in one well of a 6 well dish containing irradiated MEFS. The cells were transfected using Lipofectamine 2000 (Invitrogen) at a ratio of 4 μg DNA/10 μl of Lipofectamine 2000. For adenoviral infection of primary myoblasts, 300,000 cells per well of a 6 well dish were seeded and infected with Ad-Cre at a MOI of 20 overnight according to established techniques[313]. Transduction of Tat-Cre was carried out at a concentration of 10 μM/ml in standard ES medium overnight at 37$^{o}$C and 5% $CO_2$.

**Immunofluorescence**

Culture cells were fixed in 4% PFA, quenched with 100 mM Glycine and washed repeatedly with PBS. Cells were subsequently blocked and permeabilized in BP buffer (5% FBS, 0.2% Triton-X100, 2% BSA and 0.1% sodium azide). Incubation with primary antibodies was conducted in BP buffer for 1-24 hrs at 4$^{o}$C. Cells were washed 3x with PBS and incubated with secondary antibodies in the dark at room temperature for 1 hour. Cells were washed with PBS, incubated with Hoechst 33342 dye (Sigma, B2261) in PBS for 10 min at room temperature, washed again with PBS and coverslipped with

fluorescent mounting medium (Dako, S3023). Images were taken with a Carl Zeiss Axioplan 2 Microscope with an AxioCam HRm b/w (Zeiss) camera using Axiovision v3.2 (Zeiss) acquisition software.

**Isolation of primary myoblasts**

Primary myoblasts were isolated from adult C57BL/6 mice as described previously[314] and cultured on collagen-coated plates (rat-tail collagen, Roche-Boehringer) in Ham's F-10 media (Wisent) with 20% fetal bovine serum (Hyclone), 1% penicillin/streptomycin, and 2.5 ng/μl human recombinant bFGF (Invitrogen).

**Gaussia Luciferase Assay**

*Gaussia* Luciferase (GLuc) was examined using the Dual-Luciferase reporter assay system (Promega) with certain modifications. GLuc is naturally secreted into the cell culture medium and can be analyzed for luciferase activity. In addition, GLuc catalyzes the same substrate as Renilla Luciferase therefore only the use of the Stop & Glo Reagent was necessary to analyze luciferase levels. Luciferase analysis was conducted on a Lumistar Optima fluorometer (BMG Labtech).

**Table I: Summary of Myf5 BAC transgenic validation**

| Founder | Sex | Genotype | | | Age (n) | Fluorescence TdTom | | | RTPCR | $1^{o}$ myoblasts |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Tdtom | 5' | 3' | | TdTom | Confocal (n) | ICC (n) | TdTom(n) | TdTom |
| 923 | M | yes | yes | yes | E10.5 | Neg (0) | | | | |
| 934 | M | yes | no | yes | E10.5 | Neg (4) | | | | |
| 934 | M | yes | no | yes | E 9.5 | Neg (0) | | | | |
| 904 | M | yes | yes | yes | E 9.5 | Neg (12) | | | | |
| 947 | M | yes | yes | yes | E 12.5 | Neg (9) | | | | Neg |
| 958 | M | yes | yes | yes | E13.5 | Neg (12) | Intercostal Neg (4) | Neg (2) | | Neg |
| 934 | M | yes | no | yes | E16.5 | Neg (6) | Diaphram Neg (4) | | Neg (4) | |
| 904 | M | yes | yes | yes | E16.5 | Neg (12) | Diaphram Neg (4) | | Neg (4) | |
| 923 | M | yes | yes | yes | E16.5 | Neg (9) | Diaphram Neg (4) | | Neg (4) | Neg |
| 962 | F | yes | | | | | | | | Neg |
| **Total** | | | | | | **64** | **16** | **2** | **12** | **4** |

## Supplemental Figures – Tables

**Figure S1. Validation of YFP-ArtINT and Cre recombination.**
(**A**) Confirmation of the ability of Cre recombinase to remove DNA sequence between LoxP sites in derivatives of the PL452 vector. Vectors transformed into the EL350 *E. coli* strain containing a temperature sensitive promoter to drive Cre recombianse. The parental PL452 plasmid serves as a positive control while the Pax7 MTV 2.0 plasmid undergoes a similar decrease in plasmid size upon excision of the PGKNeo cassette by Cre recombinase. (**B**) Validation of YFP fluorescence upon cloning the YFP-ArtINT segment into pcDNA3 and subsequent transfection into 293T cells. Magnification 100x, scale bar denotes 100 μm.

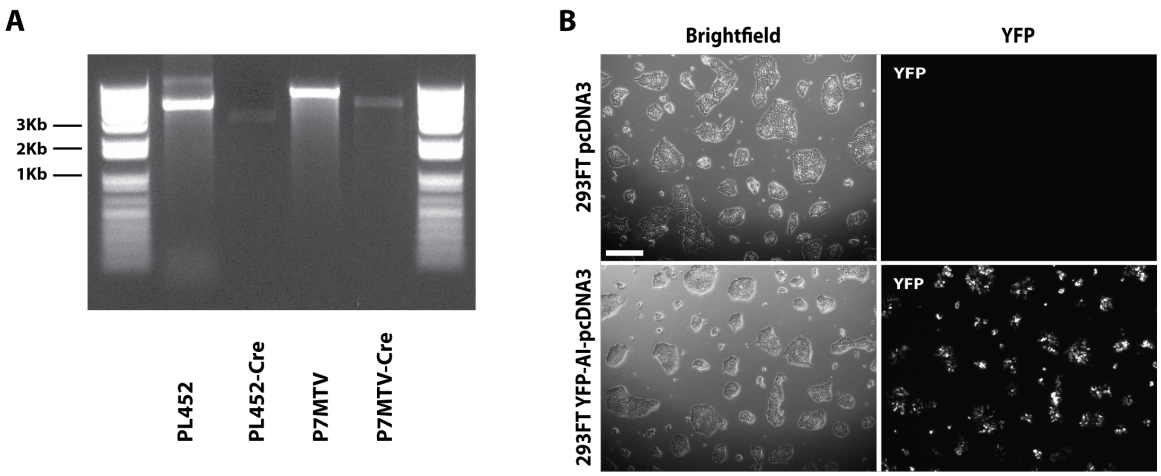

A
3Kb
2Kb
1Kb
PL452
PL452-Cre
P7MTV
P7MTV-Cre
B
Brightfield
YFP
293FT pcDNA3
YFP
293FT YFP-AI-pcDNA3
YFP

**Figure S2. Production and purification of Tat-Cre.**
(**A**) Coomassie stained gel of His-Tat-nls-Cre (HTNC or Tat-Cre) in the E. coli Tuner cell line. (**B**) HTNC appears as an enriched band at ~41 kDa upon IPTG induction. Binding optimization between HTNC and the probond nickel beads in the presence of increasing concentrations (10 mM-500 mM) of imidazole. B – Beads, FT – Flow through. (**C**) Determination of optimal elution fractions for collection and subsequent dialysis. (**D**) Coomassie stained gel to verify purity of purified and concentrated HTNC. (**E**) Western blot comparison of pTriEX HTNC to pHIT4nlsCre upon expression in 293T cells.

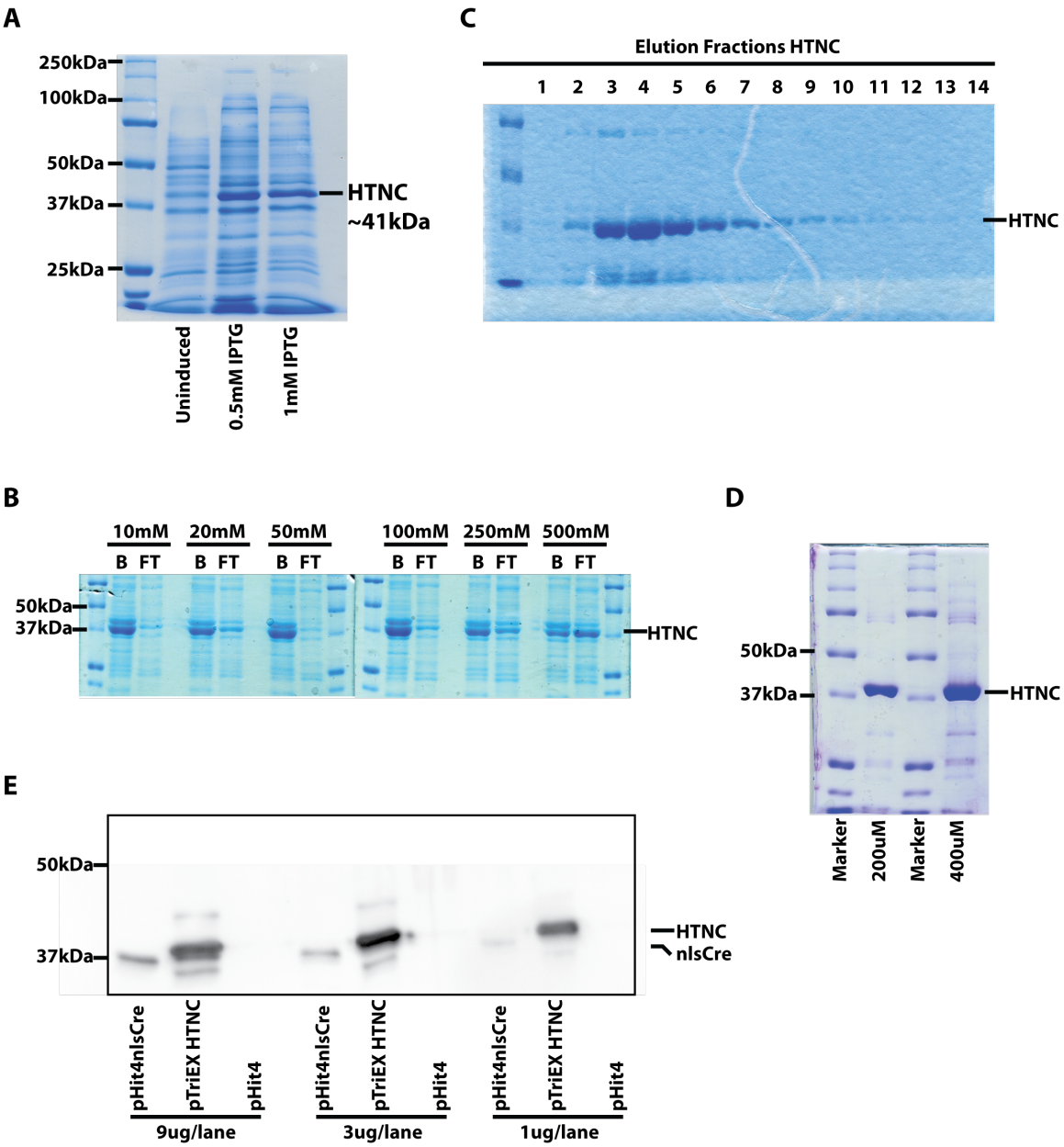
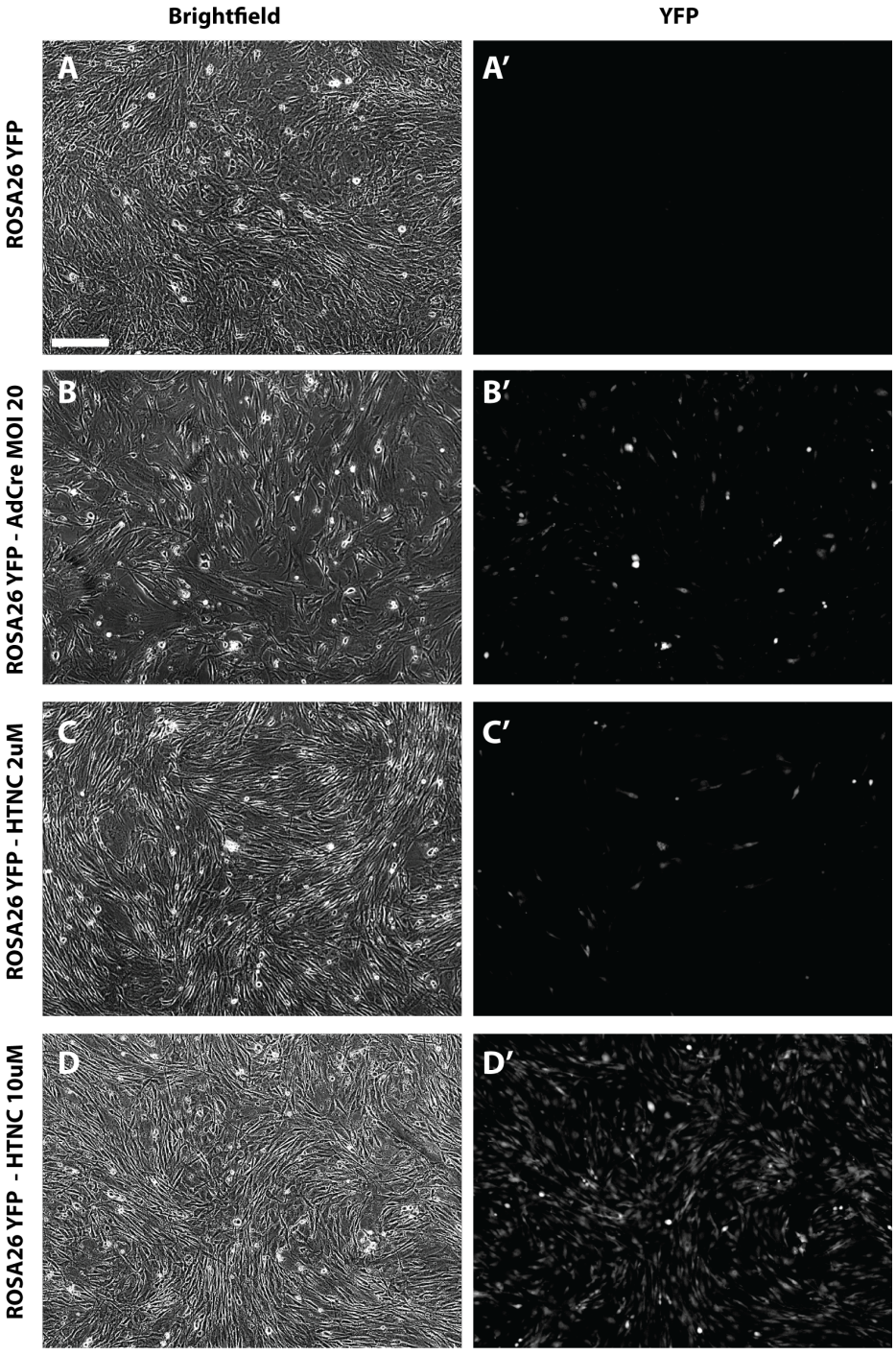

A
250kDa
100kDa
50kDa
37kDa
25kDa
HTNC
~41kDa
Uninduced
0.5mM IPTG
1mM IPTG
C
Elution Fractions HTNC
1 2 3 4 5 6 7 8 9 10 11 12 13 14
HTNC
B
10mM 20mM 50mM 100mM 250mM 500mM
B FT B FT B FT B FT B FT B FT
50kDa
37kDa
HTNC
D
50kDa
37kDa
HTNC
Marker
200uM
Marker
400uM
E
50kDa
37kDa
HTNC
nlsCre
pHit4nlsCre
pTriEX HTNC
pHit4
9ug/lane
pHit4nlsCre
pTriEX HTNC
pHit4
3ug/lane
pHit4nlsCre
pTriEX HTNC
pHit4
1ug/lane

**Figure S3. Functional validation of Tat-Cre recombinase.**
Determining the ability of purified Tat-Cre to functionally recombine LoxP sites in mammalian cells. (**A** and **A'**) ROSA26YFP fibroblasts in the absence of Cre do not express YFP, upon Cre expression a stop codon flanked by LoxP sites is removed allowing expression of YFP. (**B** and **B'**) AdCre at an MOI of 20 is used as a positive control. Two concentrations of HTNC 2 μM/ml (**C** and **C'**) and 10 μM/ml (**D** and **D'**) are assessed for recombination as evidenced by YFP. Scale bar denotes 100 μm.


Brightfield
YFP
ROSA26 YFP
A
A′
ROSA26 YFP - AdCre MOI 20
B
B′
ROSA26 YFP - HTNC 2uM
C
C′
ROSA26 YFP - HTNC 10uM
D
D′

## Table SI – Cloning Primers used for Pax7 and Myf5 Targeting Vectors

| Primer Name | Primer Sequence |
|---|---|
| **Pax7 MTV 2.0** | |
| Pax7 5' Domain FWD | GATGTCGACCTTAACTTCTGGCCGAGCCC |
| Pax7-eYFP | CCTCGCCCTTGCTCACCATTCCTGCGCACGCTGGAGACG |
| eYFP REV | GATGGTACCCTTGTACAGCTCGTCCATGC |
| ArtINT pA FWD | GATGGTACCAGGTAAGTGTACCCAATTCGCCC |
| ArtINT pA REV | GATGAATTCTAAGATACATTGATGAGTTTGG |
| Pax7 3' Domain FWD | GATGGATCCATGGCGGCGCTGCCCGG |
| Pax7 3' Domain REV | GATGCGGCCGCAGAGTGGGAAACCCTGG |
| Pax7 Ret 2.0 5' FWD | GATGTCGACTTAATTAAGAGGAGAGGCTCCAAGACCT |
| Pax7 Ret 2.0 5' REV | GATAAGCTTTTTTCCCTCACTTGGACAGC |
| Pax7 Ret 2.0 3' FWD | GATAAGCTTTGTTCCCAGACGCTACACAG |
| Pax7Ret 2.0 3' REV | GATGAATTCGCAGGAAGGACAGAGTGGAG |
| **Myf5 MTV 2.0** | |
| Myf5 5' Domain FWD | GATAAGCTTGCGTCAAAGGGACCAGTAAA |
| Myf5-TdTom | CCCACCTGCTTCTCTGAAGGATGGTGAGCAAGGGCGAGGAGGTC |
| TdTom REV | GATGGTACCCCACCTTGTACAGCTCGTCCATGCCG |
| ArtINT pA FWD | GATGGTACCAGGTAAGTGTACCCAATTCGCCC |
| ArtINT pA REV | GATGAATTCTAAGATACATTGATGAGTTTGG |
| Myf5 3' Domain FWD | GATGGATCCGACATGACGGACGGCTGCC |
| Myf5 3' Domain REV | GATGCGGCCGCACCTTGGGGAGTCTCTGG |
| Myf5 Ret 2.0 5' FWD | GATGTCGACTTAATTAATGCTTGAGACTGTATGGCTGTT |
| Myf5 Ret 2.0 5' REV | GATAAGCTTGGCTTAGCTCTGAGCAGGTG |
| Myf5 Ret 2.0 3' FWD | GATAAGCTTCCAATCATGAGCCAGGTTTT |
| Myf5 Ret 2.0 3' REV | GATGAATTCTGGCTTCTTTGTGTCCCTCT |
| **2aYpet Gluc primers** | |
| attB1GSGT2aYpet F | ggggACAAGTTTGTACAAAAAAGCAGGCTCCGGCTCTGGAGAGGG CAGAGGAAGTCTTCTAACATGCGGTGACGTGGAGGAGAATCCCG GCCCTATGGTGAGCAAAGGCGAAGA |
| attB5rYpet R | ggggACAACTTTTGTATACAAAGTTGTCTTATAGAGCTCGTTCATGC |
| attB5GSGP2aYpet F | ggggACAACTTTGTATACAAAAGTTGGAGGCTCTGGAGCGACAAAC TTTAGTCTCCTTAAACAGGCGGGAGATGTGGAAGAAAATCCAGGC CCCATGGTGAGCAAAGGCGAAGA |
| attB4Ypet R | ggggACAACTTTGTATAGAAAAGTTGGGTGCTTATAGAGCTCGTTCATGC |
| attB4rGSGF2aYpet F | ggggACAACTTTTCTATACAAAGTTGGCGGCTCTGGAGTGAAACAGA CTTTGAATTTTGACCTTCTCAAGTTGGCGGGAGACGTGGAGTCCA ACCCAGGGCCTATGGTGAGCAAAGGCGAAGA |
| attB3rYpet R | ggggACAACTTTATTATACAAAGTTGTCTTATAGAGCTCGTTCATGC |
| attB3GSGE2agLuc F | ggggACAACTTTGTATAATAAAGTTGGAGGCTCTGGACAATGTACTAA CTACGCTTTGTTGAAACTCGCTGGCGATGTTGAAAGTAACCCTGGG CCCATGGGAGTCAAAGTTCTGTTTG |
| attB2gLuc R | ggggACCACTTTGTACAAGAAAGCTGGGTACGCGTGGGGATACCCCCTAG |

| | |
|---|---|
| **LRPCR** | |
| Neo FWD | TGCTCCTGCCGAGAAAGTATCC |
| P7 ESCV2.0 LRPCR REV | CAAAGTCCCTCTACACTGCAGGAAA |
| M5 ESCV2.0 LRPCR REV | GCCAGCAAGGACTACACAGC |
| **Pax7 Southern Probe** | CCACTGGCTCAGCTACCTTC |
| | AGCAGCCCTAGTTGTCTGGA |
| **TdTom Genotyping** | |
| Myf5 5'2 F | GATAAGCTTGCGTCAAAGGGACCAGTAAA |
| Myf5 3'2 R | GATGCGGCCGCACCTTGGGGAGTCTCTGG |
| TdTom REV | GATGGTACCCCACCTTGTACAGCTCGTCCATGCCG |
| **qPCR Primers** | |
| Pax7 ex8 F | GACTCCATCAAGCCAGGAGA |
| Pax7 ex9 R | CACGTTTTTGGCCAGGTAAT |
| YPF #1 F | GCACGACTTCTTCAAGTCCGCCATGCC |
| YFP #1 R | GCGGATCTTGAAGTTCACCTTGATGCC |
| YFP #2 F | ACGTAAACGGCCACAAGTTC |
| YFP #2 R | AAGTCGTGCTGCTTCATGTG |

**Table SII – Antibodies used in this study**

| Primary Antibodies | Dilution | Company | Cat# |
|---|---|---|---|
| Pax7 | 1:10 | Developmental Studies Hybridoma Bank | PAX7 |
| Myosin Heavy Chain | 1:10 | Developmental Studies Hybridoma Bank | MF-20 |
| YFP | 1:400 | Invitrogen | A-6455 |
| Cre | 1:400 | EMD Biosciences | 69050-3 |
| **Secondary Antibodies** | | | |
| anti-Rabbit A488 | 1:1000 | Invitrogen - Molecular Probes | A-21206 |
| anti-Mouse IgG1 A568 | 1:1000 | Invitrogen - Molecular Probes | A21124 |
| Goat anti-Rabbit HRP | 1:5000 | Biorad | 170-6515 |

## Supplementary Methods

### Isolation and Verification of Tat-Cre

The pTriEx HTNC plasmid containing a His-Tat-nls-Cre construct was created by Michael Peitz and obtained from Addgene (13763)[305]. The pTriEx HTNC plasmid was transformed into the *Escherichia coli* strain TUNER(DE3)pLacI (Novagen). Cultures of LB (500 ml) containing 100 μg/ml of ampicillin and 34 μg/ml of chloramphenicol were

grown to an $OD_{600}$ of 0.7 and protein expression was induced with isopropyl β-D-thiogalactoside (IPTG) at a final concentration of 0.5 mM for 3-4 hours at 37°C (**Fig. 2**). Cells were harvested by centrifugation at 6000 rpm for 10 mins (Avanti J-25 centrifuge, Beckman Coulter) frozen in $LiqN_2$ and stored at -80°C. Pellets were thawed and resuspended in phosphate tris buffer (PTB) (50 mM $NaH_2PO_4$, 5 mM Tris (pH 7.8), all from Sigma) at 5 ml/g of wet weight. Cell membranes were disrupted with 2 mg/ml of lysozyme (Sigma, L6876) and benzonase (VWR, CA 80108-808) at 1 μl per ml of PTB shaking at room temperature for 15 min. Samples were sonicated (M500 Sonic Dismembrator, Fisher Scientific) in triplicate for 30 secs at 15% amplitude while kept on ice. 1 ml of cold TSB (2 M L-Tartaric Acid, 1x PTB and 20 mM Imidazole both from Sigma) followed by centrifugation for 25 min at 30000 g at 4°C. The supernatant was incubated with 2 ml of 50% Ni-NTA slurry (Probond Ni resin, Invitrogen, 46-0019) per liter of initial *E. coli* culture for 1 hr at 4°C. The bound slurry was loaded into a plugged 20 ml syringe and washed twice with 5 resin volumes of wash buffer (500 mM NaCl, 1xPTB, 15 mM imidazole). Tat-Cre protein is eluted with 3 resin volumes of elution buffer (500 mM NaCl, 1xPTB and 250 mM imidazole). The eluate is filtered through a 0.22 μm filter (VWR ca28145-477). For concentration of Tat-Cre dialyze the eluate using 10000 MWCO (Thermo, Slide-A-Lyzer 66380) against high salt buffer (600 mM NaCl, 20 mM HEPES pH 7.4) at 4°C overnight. The following day the dialysis buffer is changed to 25% Glycerol (500 mM NaCl, 20 mM HEPES, 25% Glycerol) for 4 hours at 4°C, followed by 50% Glycerol (500 mM NaCl, 20 mM HEPES, 50% Glycerol) overnight at 4°C. Concentrated Tat-Cre is quantified via Bradford Assay (Biorad Protein Assay, 500-0006) and stored at -20°C (**Fig. S2**). ROSA26-YFP fibroblasts were isolated

as a by product from hindlimb muscle of 6-8 week old mice and used to assess the efficiency of Tat-Cre and Ad-Cre[132]. Ad-Cre at an MOI of 20 was used as a positive control for assessing the functional ability of Tat-Cre. For adenoviral infection of ROSAYFP fibroblasts, 300,000 cells per well of a 6 well dish were seeded and infected with Ad-Cre according to established techniques. Tat-Cre was included in standard culture medium for 24hrs at varying concentrations to assess YFP expression (**Fig. S3**).

**Acknowledgements**

We thank Yves deRepentigny for blastocyst injections and BAC pronuclear injections. Nathaniel Santos at CHEO for the use of their Pulse Field Electrophoresis machine. Michael Peitz for useful protocols to produce Tat-Cre. Past Rudnicki lab members who worked on gene targeting related projects Josee Coulombe and Ayesha Misquith. M.A.R. holds the Canada Research Chair in Molecular Genetics and is an International Research Scholar of the Howard Hughes Medical Institute. This work was supported by grants to M.A.R. from the Canadian Stem Cell Network. F.D.P was supported by the Canadian Stem Cell Network and a Doctoral Research Award from the Canadian Institutes of Health Research.

# Chapter 4 - Intrinsic Decline in Repopulating Capacity of Aging Satellite Cells

## 3.1 Abstract

The long term regenerative capacity of adult skeletal muscle is attributed to a subset of satellite cells dubbed "satellite stem cells" that reside in the satellite cell niche and have never expressed the transcription factor Myf5. Aging of skeletal muscle tissue is accompanied by fibrosis, atrophy and remodeling all of which negatively affect muscle performance. Whether this reduction in skeletal muscle competency is directly attributed to the satellite cell remains largely unknown.

Here, we undertook an investigation into how age affects the transcriptional profile of satellite cells and their repopulating ability following transplantation. Aged satellite cells displayed a propensity for heightened expression of genes involved in vascular development and the expression of bZIP transcription factors. In contrast, fetal satellite cells upregulated a complex set of genes involved in cell cycle regulation, skeletal muscle development and differentiation. Purified satellite cells from fetal, adolescent, adult and aged mice were transplanted into *mdx* mice to assess their ability to take part in muscle regeneration. We determined that as satellite cells age, both their regenerative capacity and ability to colonize the satellite cell niche is reduced. Therefore, we conclude that intrinsic factors are an important determinant of satellite cell regenerative capacity during the aging process.

## 3.2 Introduction

The growth, maintenance, and regeneration of skeletal muscle is attributed to the satellite cell; a mitotically quiescent cell that resides between the basal lamina and sarcolemma of the muscle fiber[47,68,96]. During muscle regeneration satellite cells activate, proliferate and differentiate into new myofibers or fuse to existing myofibers. Satellite

cells can self-renew and it is this characteristic that allows adult skeletal muscle to undergo regeneration throughout the lifetime of an organism[49,50,315].

Intriguingly, as an organism ages a decrease in skeletal muscle function is concomitant with a decrease in satellite cell numbers from ~30% of all muscle nuclei at birth to ~15% in adolescence and ~5% in adulthood[155]. This suggests that muscle function is negatively impacted by a loss of satellite cells.

The functional and structural decline of skeletal muscle is one of the first hallmarks of aging in many organisms[316,317]. This progressive and generalized loss of muscle mass along with muscle function is characterized as sarcopenia[318]. Various mechanisms contribute to a sarcopenic phenotype including protein synthesis, proteolysis, neuromuscular integrity and muscle fat content[149]. The onset of sarcopenia has classically been attributed to alterations in both the local and systemic environments surrounding skeletal muscle[161,162,319]. To date little is known regarding a role for satellite cells in the development of sarcopenia.

A growing body of evidence suggests the satellite cell population is heterogeneous and that satellite cells differ with regards to their gene expression profiles[77,96,160,320-322], propensity for differentiation[50,323], capacity for self-renewal[49,324,325] and lineage potential[48,326]. Our lab recently identified a sub population (~10%) of satellite cells that have never expressed Myf5, based on lineage tracing experiments using a *Myf5$^{Cre}$; R26R-YFP* mouse line[49]. As Myf5 is a myogenic regulatory factor, this suggests that YFP$^-$ satellite cells are less committed with regards to cell fate then the YFP$^+$ satellite cells.

These two distinct satellite cell subpopulations also differ in terms of their

regenerative potential. $Pax7^{+}/YFP^{-}$ satellite cells, through asymmetric division, can self-renew and replenish the stem cell pool, as well as produce more committed myogenic progenitors following transplantation. Conversely, $Pax7^{+}/YFP^{+}$ satellite cells directly undergo myogenic differentiation when transplanted into regenerating muscles of $Pax7^{-/-}$ mice. These findings indicate that functional heterogeneity exists within the satellite cell population. This can be further characterized based on a hierarchy with a $Pax7^{+}/YFP^{-}$ (satellite stem cell) having more functional potency and therefore residing above the $Pax7^{+}/YFP^{+}$ (satellite committed progenitor).

In addition to heterogeneity within the satellite cell population there exists heterogeneity amongst myogenic progenitors of different developmental stages. As a consequence of age, embryonic and fetal myoblast populations display differences in their transcriptional profiles and respond differently to growth factor signaling[327]. To date it is unknown whether or not satellite cells are prone to age-related intrinsic differences.

Traditionally primary myoblasts, the progeny of muscle satellite cells expanded *in vitro,* were used to treat muscle degenerative diseases[328-331]. The transplantation of primary myoblasts results in minimal levels of physiologically relevant regeneration[332]. Until recently the satellite cell, due to its scarcity and the difficulties in isolating a pure population, was not a viable option for direct transplantation. Although many proteins have been used to identify satellite cells, Pax7 a paired-box transcription factor remains the best characterized and most consistent[70,273,274]. In the absence of Pax7 expressing satellite cells muscle regeneration is severely compromised[333,334].

The recent creation of a Pax7zsgreen BAC reporter mouse facilitates the isolation of pure populations of Pax7 expressing cells via fluorescence activated cell sorting

(FACS)[90]. Freshly sorted Pax7$^{+}$ satellite cells, when transplanted into *mdx* mice, give rise to dystrophin positive myotubes and can repopulate the sub laminar satellite cell niche. To date no direct comparison of transplantation capacity has been conducted to compare freshly sorted pure populations of satellite cells across various stages of development.

In this study we identify differences in the transcriptional profiles and regenerative capacity of satellite cells isolated from mice of different stages of development. FACS sorted satellite cells were purified from Pax7Zsgreen mice ranging from fetal (E17.5) to aged (>1yr) developmental stages. We demonstrated that as mice age the expression of cell surface markers M-cadherin, CD34 and PDGFRα is altered. When transplanted into 6-8 week old *mdx* mice, satellite cells from young mice preferentially persist within the endogenous satellite cell niche following transplantation. In addition young satellite cells demonstrate an increase over old satellite cells in their ability to contribute to muscle regeneration. Whole genome expression analysis identified transcriptional profiles for cell cycle regulation, extracellular components of the satellite cell niche, and myogenic development that are altered between satellite cells as they age. Together, these results suggest that intrinsic differences exist between young and old muscle satellite cells and allude to the potential mechanisms regulating development, regeneration and sarcopenia.

## 3.3 Results

### *3.3.1 Purification of Satellite Cells from Fetal, Adolescent, Adult and Aged Muscle*

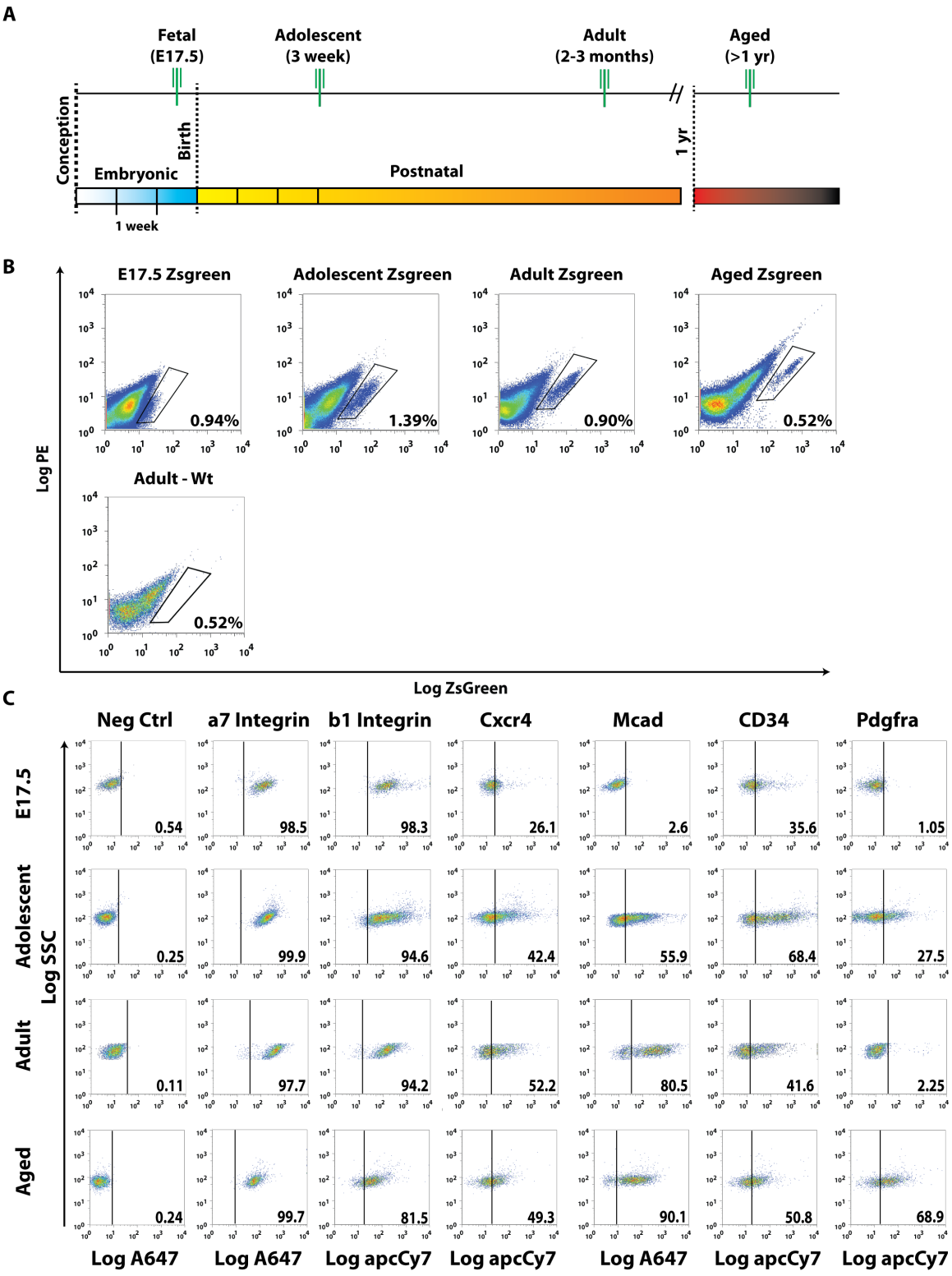
The decline in skeletal muscle function with age has been suggested to involve humoral[162] and extrinsic factors[88,163]. To determine whether age related intrinsic differences exist between young and old satellite cell populations we isolated fetal

(E17.5), adolescent (post natal day 21), adult (2-4 month) and aged (>1yr) Zsgreen-expressing cells by FACS from Pax7-Zsgreen reporter mice (**Fig 1A**).

Fetal muscle satellite cells were obtained from dissected and enzymatically digested fore and hind limbs of E17.5 Pax7-Zsgreen$^{+}$ embryos. Hind limb skeletal muscle from adolescent, adult, and aged mice was enzymatically digested from Pax7-Zsgreen$^{+}$ mice. We observed brighter Zsgreen fluorescence resulting in better separation of sorted Zsgreen cells as mice age suggestive of an increase in Pax7 expression (**Fig. 1B**). In addition we see a 2.5 fold decrease in the total number of Zsgreen$^{+}$ cells obtained from hind limb skeletal muscle when progressing from adolescent to aged mice (**Fig. S1**). A direct comparison between fetal and post natal Zsgreen cells is not possible due to the composition of fetal digests originating from the whole limb. In summary, as satellite cells age they increase their expression of Pax7 and are found in lower quantities within hind limb skeletal muscle.

To determine whether the cell surface complement of satellite cells alters with age we FACS sorted Zsgreen$^{+}$ cells and tested them against a panel of known satellite cell surface markers. Cell surface markers known to be present on Satellite cells include CD34[77,335], M-cadherin[73], β1 integrin[75], α7 integrin[320,336], CXCR4[337], Syndecan 3/4[78], and c-met[72]. We chose to analyze the following markers on satellite cells with respect to age: α7 integrin, β1 integrin, CXCR4, M-cadherin, CD34, and Pdgfrα (**Fig. 1C**). Canonical cell surface markers α7 and β1 integrin are consistently found on ~95% of sorted Zsgreen$^{+}$ cells irrespective of age. Increases in the presence of cell surface markers from fetal through to aged Zsgreen$^{+}$ cells were observed for CD34 (36% - 51%), CXCR4 (26%

**Figure 1. Age promotes the expression of Pax7 on satellite cells while modulating the composition of cell surface markers.**
(**A**) Fluorescence-activated cell sorting profile of digests from Pax7 Zsgreen mice at E17.5 days (Fetal), 3 weeks (Adolescent), 3-4 months (Adult) and >1 year old (Aged). Gating of PE on the y-axis and Zsgreen on the x-axis allows for the best separation of dim Zsgreen cells. (**B**) When analyzed according to FSC and SSC (indicators of size and granularity of cells respectively), $Zsgreen^{+}$ cells are largely homogeneous. (**C**) Cell surface profile of Zsgreen+ cells from different age groups of hind limb skeletal muscle digests. Abbreviations: PE, phycoerythrin; FSC, forward scatter; SSC, side scatter; CXCR4, C-X-C chemokine receptor type 4; Mcad, Muscle-cadherin; PDGFRα, platelet derived growth factor receptor α.


A
Fetal (E17.5)
Adolescent (3 week)
Adult (2-3 months)
Aged (>1 yr)
Conception
Birth
1 yr
Embryonic
Postnatal
1 week
B
E17.5 Zsgreen
Adolescent Zsgreen
Adult Zsgreen
Aged Zsgreen
0.94%
1.39%
0.90%
0.52%
Adult - Wt
0.52%
Log PE
Log ZsGreen
C
Neg Ctrl
a7 Integrin
b1 Integrin
Cxcr4
Mcad
CD34
Pdgfra
E17.5
Adolescent
Adult
Aged
Log SSC
0.54
98.5
98.3
26.1
2.6
35.6
1.05
0.25
99.9
94.6
42.4
55.9
68.4
27.5
0.11
97.7
94.2
52.2
80.5
41.6
2.25
0.24
99.7
81.5
49.3
90.1
50.8
68.9
Log A647
Log A647
Log apcCy7
Log apcCy7
Log A647
Log apcCy7
Log apcCy7

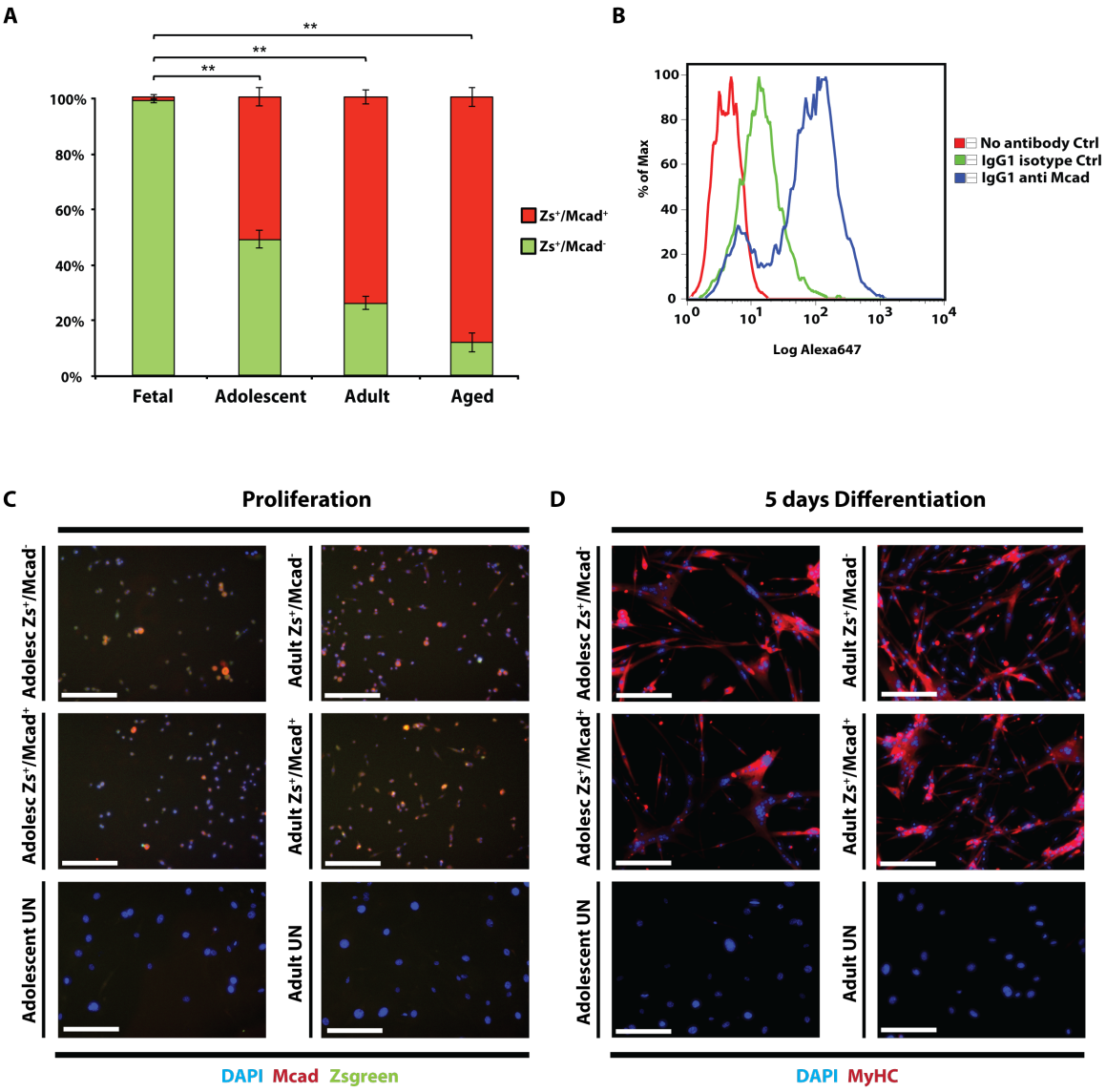
49%), M-cadherin (2% - 90%), and Pdgfrα (1% - 69%). The age-related diversity in expression of known satellite cell surface proteins suggests that there exist intrinsic differences between young and old satellite cells. We next set out to investigate whether these differences would act as predictors for differences in the functional capacity of the sorted satellite cell populations.

*3.3.2 M-cadherin Expression on Sorted Satellite Cell Populations Does Not Correlate with Differences in Functional Capacity In Vitro*

We choose to examine M-cadherin as its expression underwent the most drastic increase when comparing fetal to aged satellite cells. M-cadherin belongs to a family of $Ca^{2+}$ dependent cell adhesion molecules with morphoregulatory functions[338]. M-cadherin is expressed during embryonic myogenesis and is down regulated shortly before birth[339]. Quiescent satellite cells express M-cadherin as do primary myoblasts upon *in vitro* culture[340]. M-cadherin undergoes a transient increase in expression during the fusion of myoblasts into multinucleated myotubes[339,341]. To further study the significance responsible for the large variation in M-cadherin expression between fetal and aged satellite cells we sorted $Pax7^{+}/Mcad^{-}$ or $Pax7^{+}/Mcad^{+}$ populations from each of the stages of mouse development. The proportion M-cadherin on satellite cells varies significantly between fetal (2% ±0.66%), adolescent (50% ±3.21%), adult (74% ±2.47), and aged (88% ±3.51%) (**Fig. 2A**).

To ensure authentic M-cadherin detection we use an IgG1 isotype control to determine non-specific staining (Fig. 2B). Fetal Pax7-Zsgreen cells are principally M-cad negative (~98%) while aged Pax7-Zsgreen cells are the opposite (~90% M-cad positive). This striking difference in the expression of M-cad between young and old

**Figure 2. M-cadherin expression on freshly sorted satellite cells increases with age.** (**A**) Quantification of the proportion of M-cadherin protein on Zsgreen+ satellite cells sorted from Fetal, Adolescent, Adult and Aged mice. Zsgreen+/Mcad+ cells are depicted in red while Zsgreen+/Mcad- are in green. Data represent mean ± SE for Fetal (n=4), Adolescent (n=11), Adult (n=14), and Aged (n=9). ** $p<0.01$. (**B**) Fluorescence-activated cell sorting histogram comparing Zsgreen+ cells either unstained (red) an IgG1 isotype antibody (green) or an M-cadherin antibody (blue). (**C**) Immunofluorescence of Zsgren+/Mcad-, Zsgreen+/Mcad+ or unsorted (UN) cells isolated from Adolescent or Adult satellite cells and expanded in proliferation conditions *in vitro*. M-cadherin is pseudocolored in red, Zsgreen is in green while DAPI is in blue. (**D**) Immunofluorescence conducted on differentiated primary myoblasts isolated from Zsgreen+/Mcad-, Zsgreen+/Mcad+ or unsorted (UN) cells of adolescent or adult origin. MyHC (red) and DAPI (blue) mark intracellular myosin and nuclei respectively. Abbreviations: DAPI, 4',6-diamidino-2-phenylindone; MyHC, myosin heavy chain. Scale bars denote 100 μm.


A
**
**
**
100%
80%
60%
40%
20%
0%
Fetal
Adolescent
Adult
Aged
Zs+/Mcad+
Zs+/Mcad-
B
% of Max
100
80
60
40
20
0
10^0
10^1
10^2
10^3
10^4
Log Alexa647
No antibody Ctrl
IgG1 isotype Ctrl
IgG1 anti Mcad
C
Proliferation
Adolesc Zs+/Mcad-
Adult Zs+/Mcad-
Adolesc Zs+/Mcad+
Adult Zs+/Mcad+
Adolescent UN
Adult UN
DAPI Mcad Zsgreen
D
5 days Differentiation
Adolesc Zs+/Mcad-
Adult Zs+/Mcad-
Adolesc Zs+/Mcad+
Adult Zs+/Mcad+
Adolescent UN
Adult UN
DAPI MyHC

satellite cells implies the existence of phenotypic differences between satellite cells of different ages.

In order to determine functional differences between M-cadherin$^{+}$ or M-cadherin$^{-}$ cell fractions we purified adolescent and adult ZsGreen$^{+}$/Mcad$^{-}$ and Zsgreen$^{+}$/Mcad$^{+}$ cells and cultured them *in vitro*. Cultured primary myoblasts from ZsGreen$^{+}$/Mcad$^{-}$ or Zsgreen$^{+}$/Mcad$^{+}$ satellite cell populations uniformly express M-cadherin and Zsgreen upon activation and proliferation *in vitro* (**Fig. 2C**). This suggests that irrespective of their M-cadherin expression *in vivo*, once removed from the endogenous satellite cell niche, all primary myoblasts express M-cadherin.

Next we sought to investigate whether ZsGreen$^{+}$/Mcad$^{-}$ or Zsgreen$^{+}$/Mcad$^{+}$ cells differed in their ability to undergo myogenic differentiation. Immunofluorescent staining with an antibody specific for myosin heavy chain was performed on cultured primary myoblasts from adolescent and adult (ZsGreen$^{+}$/Mcad$^{-}$ or Zsgreen$^{+}$/Mcad$^{+}$) cell populations after 5 days in differentiation conditions (**Fig. 2D**). Following differentiation both the Pax7$^{+}$/Mcad$^{+}$ and Pax7$^{+}$/Mcad$^{-}$ populations derived from adolescent or adult satellite cells were capable of forming differentiated myotubes. This is in stark contrast to the unsorted population where only a small proportion of cells expressed myosin heavy chain. These data suggest that adult and adolescent satellite cells harbour similar myogenic capacity *in vitro* and this ability to differentiate is accomplished irrespective of whether or not the original satellite cell population expressed M-cadherin.

### *3.3.3 Increased Age Negatively Affects the Regenerative Potential of Engrafted Satellite Cells into mdx Muscle*

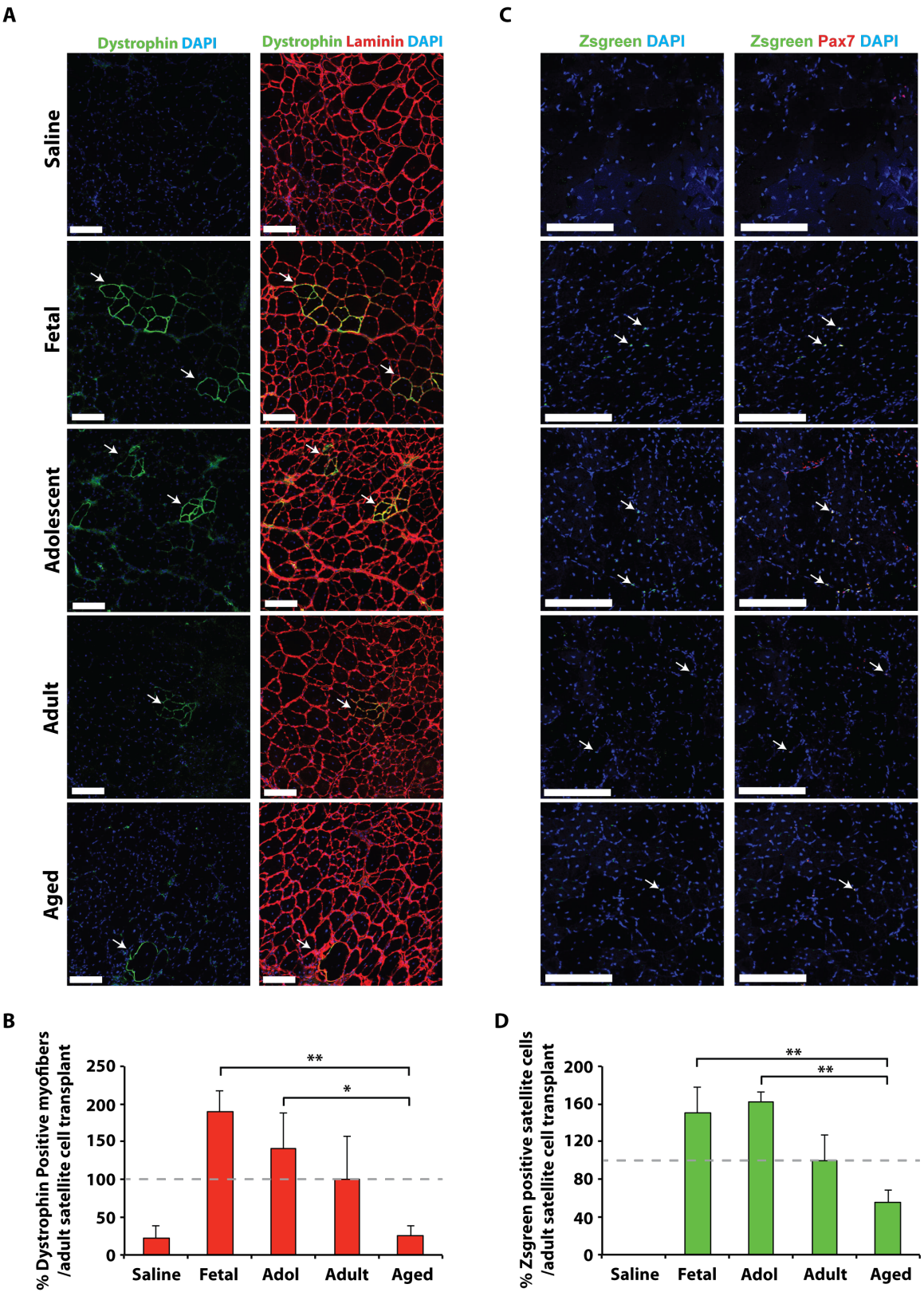
Previous studies suggest the cellular milieu dictates to what extent satellite cells

participate in regeneration[88,162]. Primary myoblasts have been used with limited success as a cell source for muscle regeneration[328,342,343]. In comparison, satellite cells present on intact fibers following transplantation re-occupy the host niche and demonstrate a physiological benefit[50,96]. To investigate the ability of young and old satellite cells to participate in muscle regeneration we transplanted freshly sorted pooled satellite cell populations from a mix of male and female Pax7Zsgreen/129sv mice into *mdx* mice.

In order to study the intrinsic aspects of satellite cell aging we transplanted into dystrophin deficient 6-8 week old *mdx* mice to eliminate any differences provided by the host environment. Prior to transplantation the tibialis anterior (TA) muscle was injured with cardiotoxin to prime the environment for regeneration and an osmotic pump dispensing the immunosuppressive FK506 was implanted. Two days later 10,000 freshly sorted satellite cells from fetal, adolescent, adult or aged muscle preparations were transplanted into the cardiotoxin injured TA muscle. Three weeks following transplantation treated TA muscles were dissected and serially sectioned through the mid belly of the muscle.

Transplanted cells were assessed for their regenerative ability based on the percentage of dystrophin$^{+}$ fibers out of the total number of fibers in a muscle section. Representative sections were immunostained for dystrophin and laminin to determine this ratio. Fetal and adolescent satellite cell populations display a greater ability to give rise to dystrophin+ fibers in comparison to adult, aged and saline control (**Fig. 3A**). Fetal (189% ± 28%) and Adolescent (141% ± 47%) satellite cells displayed an increase in the percentage of dystrophin positive fibers relative to Adult satellite cells (100% ± 58%) (**Fig. 3B**). Strikingly, the Aged population demonstrated a 75% decrease relative to Adult

**Figure 3. Old age negatively affects the regenerative capacity of satellite cells following transplantation into *mdx* TA muscle.**
10,000 freshly sorted Zsgreen$^{+}$ satellite cells from Fetal, Adolescent, Adult or Aged mice were transplanted into the TA muscle of *mdx* mice ($n \geq 3$ for each time point). (**A**) Dystrophin (green), laminin (red) and DAPI (blue) staining in *mdx* TA muscle 3 weeks following cardiotoxin treatment and transplantation of either Mcad$^{-}$ or Mcad$^{+}$ Zsgreen$^{+}$ satellite cells. (**B**) Quantification of the percentage of Dystrophin positive fibers post transplantation of age related satellite cell populations. Immunofluorescence of Pax7 (red) and Zsgreen (green) 3 weeks post transplantation into cardiotoxin treated TA muscle of *mdx*$^{-/-}$ mice (C). Quantification of double positive Pax7$^{+}$/Zsgreen$^{+}$ cells in relation to total Pax7$^{+}$ satellite cells (D). * p value < 0.05. ** p value < 0.01. Scale bar denotes 100 μm. Abbreviations: DAPI, 4',6-diamidino-2-phenylindone; MyHC, myosin heavy chain.


A
Dystrophin DAPI
Dystrophin Laminin DAPI
Saline
Fetal
Adolescent
Adult
Aged
C
Zsgreen DAPI
Zsgreen Pax7 DAPI
B
% Dystrophin Positive myofibers /adult satellite cell transplant
250
200
150
100
50
0
**
*
Saline
Fetal
Adol
Adult
Aged
D
% Zsgreen positive satellite cells /adult satellite cell transplant
200
160
120
80
40
0
**
**
Saline
Fetal
Adol
Adult
Aged

satellite cells. Taken together these data suggest a negative correlation between the age of satellite cells and their ability to contribute to muscle regeneration.

Functional satellite cells are of critical importance in order to obtain a significant long-term amelioration of the dystrophic phenotype[50,96]. Post transplantation it was important to determine the number of transplanted Zsgreen$^{+}$ cells that had taken up the satellite cell niche. Isolated TA muscle sections were stained for Zsgreen to enumerate transplanted cells and Pax7/DAPI to confirm their continued satellite cell identity. Interestingly fetal, adolescent, adult and aged samples irrespective of their M-cadherin expression contribute to the satellite cell pool (**Fig. 3C**). Relative to adult Zsgreen$^{+}$ cells (100%), aged cells displayed the least ability to persist in the satellite cell niche (56% ±12%) following transplantation. Fetal and adolescent Zsgreen$^{+}$ cells demonstrated an increased ability, 151% ± 27% and 161% ±11% respectively, to colonize and persist in the satellite cell niche 3 weeks post transplantation (**Fig. 3D**) and **Table 1**. Our experiments provide compelling evidence that satellite cells, isolated at various stages during development, have different cell intrinsic attributes that can functionally influence muscle regeneration.

### *3.3.4 Genome Wide Expression Analysis of Satellite Cell Populations Identifies Distinct Age-Related Profiles*

We performed genome wide expression analysis to gain insight into the mechanisms that govern the intrinsic functional differences observed between young and old satellite cells. Microarray analysis was conducted on total RNA from freshly sorted and pooled populations of fetal (Zsgreen$^{+}$/Mcad$^{-}$, n=4), adolescent (Zsgreen$^{+}$/Mcad$^{-}$ and

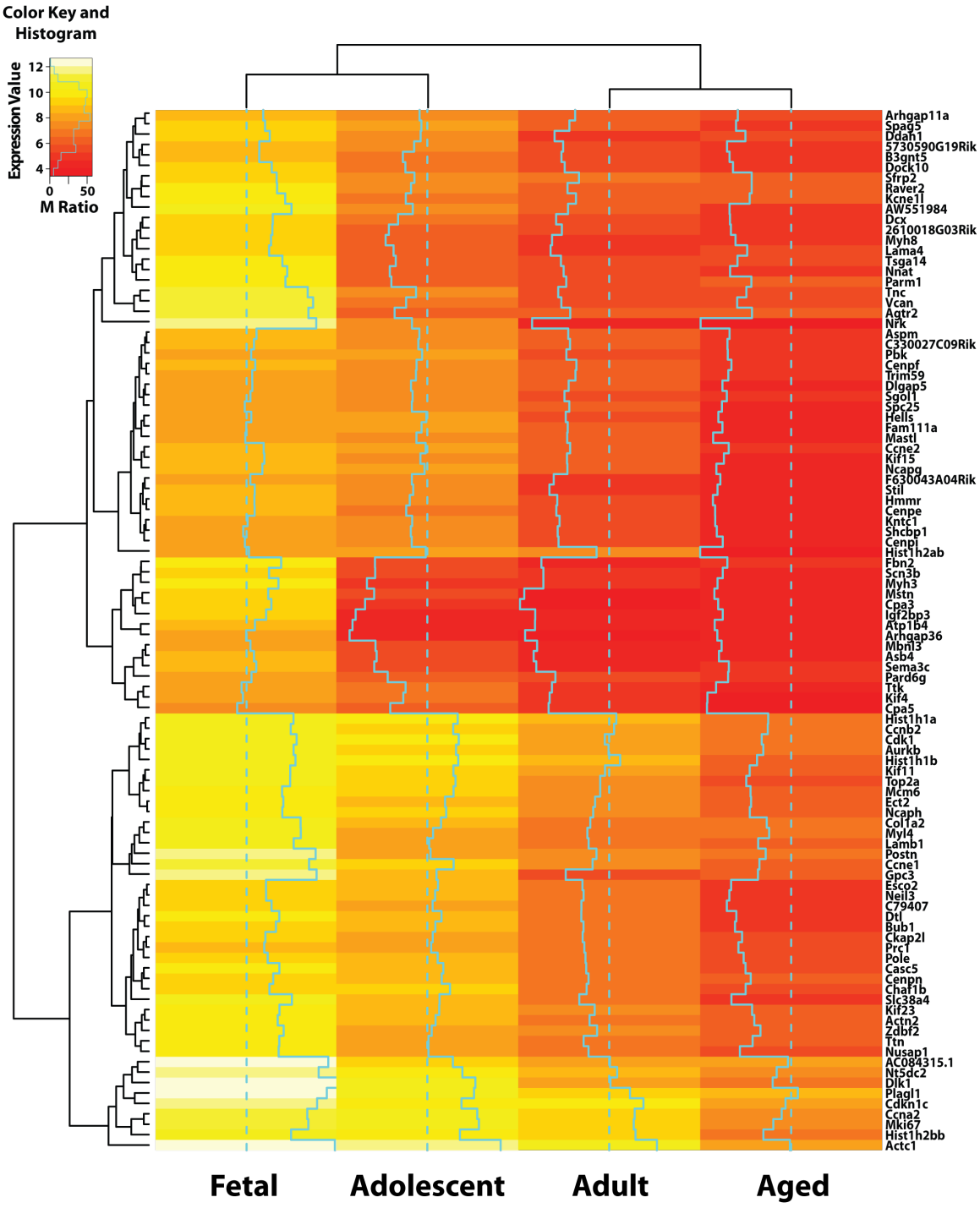
Zsgreen$^{+}$/Mcad$^{+}$, n=6), adult (Zsgreen$^{+}$/Mcad$^{+}$, n=8) and aged (Zsgreen$^{+}$/Mcad$^{+}$, n=7) satellite cells.

Hierarchical clustering of the data sets resulted in fetal Zsgreen$^{+}$ cells being most similar to adolescent while adult was associated with the aged data set (**Fig. 4**). As M-cadherin expression did not correlate with any significant difference in the regenerative capacity of sorted satellite cells (**Fig. S2**), both adolescent populations were averaged.

To lend credibility to our microarray data we examined known genes and their respective families involved in satellite cell development, activation and maintenance. Our FACS data indicates that Pax7 expression increases with age, an observation that is corroborated by our microarray results (+2.7 fold). In addition, myogenic regulatory factors including Myf5 (+2.9 fold), Myod (+1.47 fold) and Myogenin (-4.7 fold) were altered as a function of age. Myosin heavy chain 3 (42 fold) and 8 (17 fold), genes specific for embryonic and perinatal skeletal muscle respectively, were both upregulated in fetal satellite cells[344]. In contrast, betaglycan (5.5 fold) and B-cell lymphoma 6 (*Bcl6*) (11.5 fold), genes known to be upregulated in aged skeletal muscle, were enriched in aged satellite cells[345,346]. The aforementioned results demonstrate markers for young and old satellite cells are recognized by our microarray and support further analysis to determine putative mechanisms responsible for age-related functional differences in satellite cells.

GO term analysis of genes upregulated (>2 fold) in Fetal vs. Aged Zsgreen$^{+}$ cells identified broad categories of cell cycle regulation, cell division, cytoskeletal arrangement, extracellular matrix components and striated muscle development (**Fig.**

**Figure 4. Genome wide transcriptional differences as a function of satellite cell age.** A heat map of log2 transformed RMA values from fetal, adolescent, adult and aged microarray data sets. Genes expressed at least 10 fold higher in fetal compared to aged satellite cells. Rows represent gene identifiers for probe sets while columns are satellite cell samples. Red corresponds to low expression, yellow to high expression; note the expression is log2 transformed. Both the rows (gene identifiers) and columns (samples) are hierarchically clustered by similarity. Branch distance is indicative of similarity. The blue line through the color key is a histogram indicating the number of genes with a given expression value per condition and therefore serves as a linear indication for the distribution of expression values within a sample.


Color Key and Histogram
Expression Value
12
10
8
6
4
0
50
M Ratio
Arhgap11a
Spag5
Ddah1
5730590G19Rik
B3gnt5
Dock10
Sfrp2
Raver2
Kcne1l
AW551984
Dcx
2610018G03Rik
Myh8
Lama4
Tsga14
Nnat
Parm1
Tnc
Vcan
Agtr2
Nrk
Aspm
C330027C09Rik
Pbk
Cenpf
Trim59
Dlgap5
Sgol1
Spc25
Hells
Fam111a
Mastl
Ccne2
Kif15
Ncapg
F630043A04Rik
Stil
Hmmr
Cenpe
Kntc1
Shcbp1
Cenpi
Hist1h2ab
Fbn2
Scn3b
Myh3
Mstn
Cpa3
Igf2bp3
Atp1b4
Arhgap36
Mbnl3
Asb4
Sema3c
Pard6g
Ttk
Kif4
Cpa5
Hist1h1a
Ccnb2
Cdk1
Aurkb
Hist1h1b
Kif11
Top2a
Mcm6
Ect2
Ncaph
Col1a2
Myl4
Lamb1
Postn
Ccne1
Gpc3
Esco2
Neil3
C79407
Dtl
Bub1
Ckap2l
Prc1
Pole
Casc5
Cenpn
Chaf1b
Slc38a4
Kif23
Actn2
Zdbf2
Ttn
Nusap1
AC084315.1
Nt5dc2
Dlk1
Plagl1
Cdkn1c
Ccna2
Mki67
Hist1h2bb
Actc1
Fetal
Adolescent
Adult
Aged

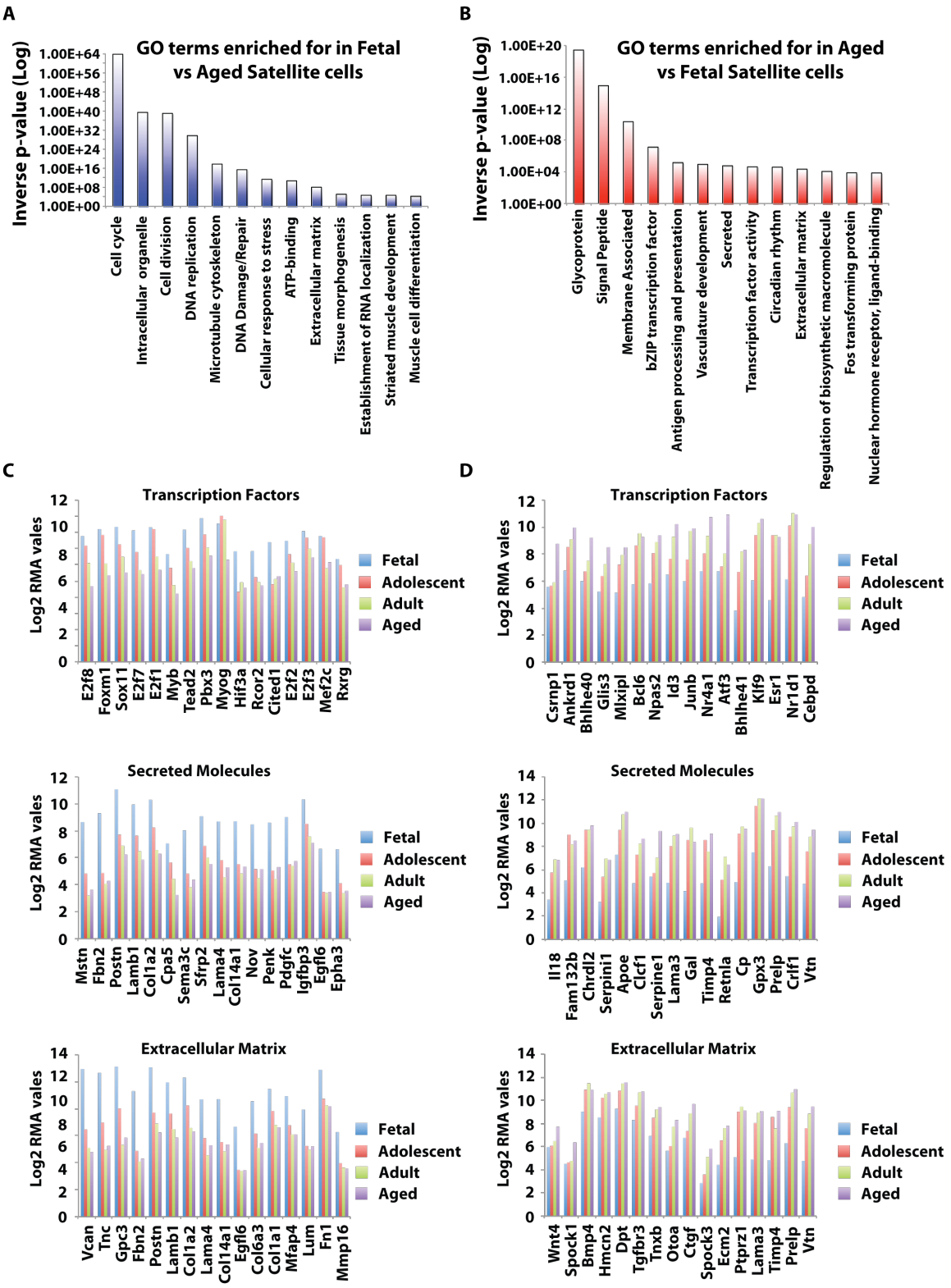
**5A**). In contrast, GO term analysis conducted on genes up regulated (>2 fold) in the aged data set demonstrated enrichment for membrane bound proteins, bZIP transcription factors, vascular development, circadian rhythm and ligand binding nuclear hormone receptors (**Fig. 5B**).

To determine any age-related changes in the components of the satellite cell niche we examined the expression of secreted and extracellular matrix (ECM) related proteins. Fetal satellite cells are enriched (>30 fold) for the ECM related proteins Versican (*Vcan*), Glypican 3 (*Gpc3*) and Tenascin C (*Tnc*). In addition to changes in the ECM fetal satellite cells express high levels of Carboxypeptidase a3 (*Cpa3* – 51 fold), Myostatin (*Mstn* – 45fold), Platelet derived growth factor c (*Pdgfc* – 12fold) and Periostin (*Postn* – 12 fold) (**Fig 5C and Table II**). In contrast, Vitronectin (26 fold), Laminin a3 (20 fold), Connective tissue growth factor (7.6 fold) are ECM related proteins that are upregulated in aged satellite cells. Interleukin 18 (11 fold), and ApoE (13 fold) are secreted proteins enriched for in aged satellite cells (**Fig. 5D and Table III**). These age-related variations in the ECM components surrounding satellite cells elude to a satellite cell intrinsic role in remodelling the niche and are indicative of greater complexity in regulation of satellite cell function over time.

## 3.4. Discussion

Here we isolate pure populations of Pax7 Zsgreen$^{+}$ cells from fetal, adolescent, adult and aged mice to characterize age-related intrinsic differences between satellite cells. Our *in vitro* experiments did not identify any age-associated differences between primary myoblast populations. However, the age of donor satellite cells significantly affected their ability to contribute to muscle regeneration and colonize the satellite cell

**Figure 5. GO term enrichment in Fetal and Aged satellite cell populations.**
(**A**) GO term enrichment for genes >2 fold up regulated in fetal satellite cells. (**B**) GO term enrichment for genes >2 fold up regulated in aged satellite cells. (**C**) Genes up regulated in fetal satellite cells comprising the categories of transcription factors, secreted molecules and extracellular matrix molecules. (**D**) Genes up regulated in aged satellite cells comprising the categories of transcription factors, secreted molecules and extracellular matrix molecules.


A
GO terms enriched for in Fetal vs Aged Satellite cells
Inverse p-value (Log)
1.00E+64
1.00E+56
1.00E+48
1.00E+40
1.00E+32
1.00E+24
1.00E+16
1.00E+08
1.00E+00
Cell cycle
Intracellular organelle
Cell division
DNA replication
Microtubule cytoskeleton
DNA Damage/Repair
Cellular response to stress
ATP-binding
Extracellular matrix
Tissue morphogenesis
Establishment of RNA localization
Striated muscle development
Muscle cell differentiation
B
GO terms enriched for in Aged vs Fetal Satellite cells
Inverse p-value (Log)
1.00E+20
1.00E+16
1.00E+12
1.00E+08
1.00E+04
1.00E+00
Glycoprotein
Signal Peptide
Membrane Associated
bZIP transcription factor
Antigen processing and presentation
Vasculature development
Secreted
Transcription factor activity
Circadian rhythm
Extracellular matrix
Regulation of biosynthetic macromolecule
Fos transforming protein
Nuclear hormone receptor, ligand-binding
C
Transcription Factors
Log2 RMA vales
E2f8
Foxm1
Sox11
E2f7
E2f1
Myb
Tead2
Pbx3
Myog
Hif3a
Rcor2
Cited1
E2f2
E2f3
Mef2c
Rxrg
Secreted Molecules
Mstn
Fbn2
Postn
Lamb1
Col1a2
Cpa5
Sema3c
Sfrp2
Lama4
Col14a1
Nov
Penk
Pdgfc
Igfbp3
Egfl6
Epha3
Extracellular Matrix
Vcan
Tnc
Gpc3
Fbn2
Postn
Lamb1
Col1a2
Lama4
Col14a1
Egfl6
Col6a3
Col1a1
Mfap4
Lum
Fn1
Mmp16
D
Transcription Factors
Csrnp1
Ankrd1
Bhlhe40
Glis3
Mlxipl
Bcl6
Npas2
Id3
Junb
Nr4a1
Atf3
Bhlhe41
Klf9
Esr1
Nr1d1
Cebpd
Secreted Molecules
Il18
Fam132b
Chrdl2
Serpini1
Apoe
Clcf1
Serpine1
Lama3
Gal
Timp4
Retnla
Cp
Gpx3
Prelp
Crlf1
Vtn
Extracellular Matrix
Wnt4
Spock1
Bmp4
Hmcn2
Dpt
Tgfbr3
Tnxb
Otoa
Ctgf
Spock3
Ecm2
Ptprz1
Lama3
Timp4
Prelp
Vtn
Fetal
Adolescent
Adult
Aged

niche following transplantation. To clarify the cell intrinsic differences that arise with age we conducted microarray analysis on each sorted population. The resulting transcriptional profiles spanning multiple developmental time points provide a rich data set to further pursue mechanistically how age affects the functional ability of satellite cells.

We observe a decline with age in total satellite cells present within limb musculature. These results serve to clarify previous experiments in mice as well as reports in humans examining multiple different muscle types[88,90,157,347-350]. The discrepancy in the functional ability of skeletal muscle between young and old animals in combination with a decrease in the prevalence of total satellite cells indicate the functional ability of satellite cells are paramount to muscle regeneration. However, the proportion of satellite stem cells with respect to age remains to be determined, a characteristic that is important considering there exists a functional heterogeneity within the satellite cell pool with regards to Myf5 activation.

Following transplantation of freshly sorted Zsgreen$^{+}$ satellite cells into *mdx* mice of between 6-8 weeks of age, the fetal population demonstrated the best ability to contribute to muscle regeneration based on dystrophin staining. These experiments eliminate any confounding factors due to the systemic environment and to our knowledge are the first to address whether age affects the myogenic capacity of pure satellite cell populations upon transplantation. The ability of fetal satellite cells to provide a greater degree of differentiation compared to aged satellite cells is novel and presents a new avenue to explore regarding intrinsic age-related properties of the satellite cell.

Transplantation of freshly sorted Zsgreen$^{+}$ cells from various ages resulted in significant numbers of Zsgreen$^{+}$/Pax7$^{+}$ cells present in host TA muscles. The presence of a cell that is both Zsgreen$^{+}$ and Pax7$^{+}$ demonstrates that donor derived Zsgreen$^{+}$ cells can colonize the endogenous satellite cell niche. In addition to being able to contribute to muscle differentiation post transplantation the ability to re-colonize the satellite cell niche is paramount to ensure long-term physiological muscle regeneration. These findings constitute evidence that intrinsic properties of aged satellite cells decrease their regenerative capacity. Whether this regenerative deficiency is due to an inability to proliferate, or possibly to colonize the endogenous niche remains to be determined. In summary, we conclude that satellite cells isolated from young mice have both a greater capacity to contribute to muscle differentiation and ability to colonize the endogenous satellite cell niche following transplantation.

The age related differences we observe with respect to cell surface markers, regenerative capacity, and colonization of the endogenous satellite cell niche following transplantation do not explain how fetal satellite cells are functionally different from their aged brethren. To address this issue we conducted a global gene expression comparison amongst the different isolated satellite cell populations.

Our GO term analysis identified several key differences between fetal and aged satellite cells. Fetal cells have enriched genes involved in the cell cycle, cell division and DNA replication while aged cells are deficient in these genes. These trends are indicative of fetal cells being activated and dividing while aged cells are quiescent. During embryonic and fetal stages satellite cells must rapidly divide to generate sufficient myogenic progenitors to account for the significant growth in muscle mass. In the adult

satellite cells are predominantly quiescent. It is not surprising that fetal cells are more activated then adolescent, adult or aged satellite cells. However, it is interesting to note that satellite cells become progressively more quiescent as they age.

Quiescence is often associated with adult stem cell populations[351,352]. Sprouty1 an inhibitor of receptor tyrosine kinases is known to be involved in the return of activated satellite cells to quiescence[353]. Of interest aged satellite cells express ~3 fold higher levels of Sprouty1 then fetal satellite cells. This coupled with a reduction in cell cycle genes (Ccne1– 28 fold, Ccne2 – 17 fold, Cdk1 – 17 fold) suggests abundant quiescent signals are present in aged satellite cells. We speculate that a defect in activation or an over abundance in quiescence is responsible for the drastic inability of aged satellite cells to contribute to muscle regeneration following transplantation. Prior to testing this hypothesis validation via qPCR of gene expression changes observed following Fetal and Aged microarray is necessary.

The local and systemic milieu that surrounds the satellite cell niche has profound effects on the functional characteristics of satellite cells[340]. Having discovered that satellite cells have intrinsic functional properties and transcriptional profiles that differ with age we believe the examination of categories including secreted proteins, proteins present in the extracellular matrix, and transcription factors specific for a given satellite cell age will aid in identifying the mechanisms responsible for these intrinsic differences. Secreted proteins/enzymes along with growth factors enriched in fetal satellite cells potentially enhance their ability to contribute to muscle regeneration post transplantation. A classical example of this would be myostatin, a growth and differentiation factor that when genetically absent gives rise to hypertrophic bulls[354]. Conversely the treatment of

adult satellite cells with secreted and/or ECM proteins enriched in aged satellite cells would lead to an increase in quiescence.

Our data implies that sarcopenia arises due to a combination of changes within satellite cells as they age in addition to alterations in the systemic environment. The ability of young satellite cells to participate in muscle regeneration, while also contributing to the satellite cell niche in a fashion that is significantly better then aged or adult satellite cells, implies an age dependant functional heterogeneity exists in satellite cells. The question remains as to whether or not these differences are permanent or reversible. Previous experiments dealing with parabiotic coupling of old and young mice would suggest alterations to aged satellite cells are reversible as exposure to young serum ameliorated muscle function[162].

The discovery of satellite cell intrinsic aspects that alter with age coupled with genome wide expression analysis of these populations has enabled the identification of specific pathways inherently different in young versus old satellite cells. Future work to validate changes in gene expression by qPCR followed by experimental validation either in vitro or in vivo of key regulatory molecules and pathways is essential. Determining the mechanisms that are responsible for age-related differences in the functional capacity of satellite cells will aid in reversing the regenerative deficiency of aged satellite cells.

## 3.5 Materials and Methods

### Animal Section

The following is a list of the mice used to conduct this research: 129S2/SvPasCrl (Charles River laboratories, St. Constant QC). Pax7Zsgreen mice were graciously provided by Dr. Michael Kyba. C57BL/10ScSn-Dmd$^{mdx}$/J (Jackson Laboratories, Bar

Harbour Maine - #001801), Blc10 (Charles River, St. Constant QC ). All mice were housed under standard conditions and allowed free access to food and water ad libitum. The Animal Care Committee at the University of Ottawa approved of all studies.

**Immunofluorescence**

E17.5 Embryos and TA muscles were frozen in Tissue-Tek OCT in a dry ice/ethanol bath and stored at -80$^{o}$C prior to sectioning. Serial cryosections 12 μm in thickness were cut using a cryostat (Leicia CM 1850). Cells cultured on collagen coated slides and muscle sections were fixed in 4% PFA for 10 min at room temperature followed by incubation in 100 mM Glycine for 10 min at room temperature. Fixed cells were permeabilized and blocked in 5% FBS, 0.2% Triton X-100 (Sigma) and 2% bovine serum albumin (Fisher Scientific) for 1 hr at room temperature. Cells were incubated with primary antibodies (View **supplementary Table I**) for 1 hour at room temperature or alternatively 4$^{o}$C overnight. Secondary antibodies were diluted 1:1000 in PBS and incubated with fixed cells for 1 hr at room temperature. Nuclei were counter-stained with Hoechst 33342 (Sigma). Standard images were taken on a Carl Zeiss Axioplan 2 Microscope with an AxioCam HRm b/w (Zeiss) camera using Axiovision v3.2 (Zeiss) acquisition software. Confocal images were taken on a LSM710 using the Zen 2009 LE software and Adobe Photshop CS5.

**FACS isolation**

Skeletal muscles from heterozygous Pax7-ZsGreen mice were isolated at defined stages: E17.5 (Fetal - whole forelimb and hindlimb), post natal day 21 (Adolescent - hindlimb), 2-3 month old (Adult - hindlimb) and >1 year old (Aged - hindlimb) mice. Muscles were aseptically dissected, minced and digested in a dispase/collagenase solution (1 mg/ml

collagenase Cat# 11 088 831 001, 4 mg/ml Dispase II Cat# 04 942 078 001 – Roche) for 25 minutes at 37°C. The digested muscle slurry was briefly triturated after 15 min of incubation and returned to the incubator. The slurry was subsequently diluted with PBS containing 10% fetal bovine serum and 2 mM EDTA (FACS sorting medium), filtered through 70 μm Netwell cell strainers (Costar), and pelleted at 1700 rpm for 5 min (Thermo IEC Centra CL2). Cell pellets were gently resuspended in FACS buffer and incubated/agitated with primary antibodies and when appropriate secondary antibodies for 9 min on ice. Cells were subsequently washed in FACS buffer, pelleted and filtered through 30 μM filters (Miltenyi Biotech) prior to FACS sorting. Cells were separated on a MoFlo cytometer (DakoCytomation) equipped with 3 lasers. Sorting gates were strictly defined based on age of sample or in the case of a negative control for Pax7zsgreen, wildtype 129sv mice. As a negative control for M-cadherin the same procedure was followed with a mouse IgG1 isotype control antibody (BD bioscience; 554121).

**Cell Culture and Differentiation Assays**

FACS purified Zsgreen$^{+}$ cells were propagated on collagen coated dishes in Hams F-10 (Wisent, 318-051-CL) supplemented with 20% fetal bovine serum (FBS, Hyclone), 1x penicillin/streptomycin (Invitrogen) and 2.5 ng/ml human bFGF (Millipore) at 37$^{o}$C and 5% $CO_2$. Differentiation was induced by culture in DMEM (Wisent, 319-016-CL) with 5% horse serum (Hyclone) for 5 days at 37$^{o}$C and 5% $CO_2$.

**Myogenic Progenitor Transplantation**

For transplantation, the dystrophin deficient mouse model *mdx* (#001801, Jackson laboratory, Bar Harbor, ME.)[355,356] immunosuppressed with 5 mg/kg Tacrolimus (LC laboratories, F4900-FK506) administered daily via Alzet osmotic pumps (Durect

Corporation, 0000298) was used. FACS-purified Zsgreen+/Mcad- or Zsgreen+/Mcad+ cells (10,000) were injected in a volume of 25 μl with a 29g syringe into the right TA muscle of *mdx* mice. The TA muscles of *mdx* mice were injured with injections of 50 μl of 10 μM cardiotoxin (Sigma) 48 hrs prior to transplantation to provide an environment for satellite cell based regeneration. Injected TA muscles were harvested 3 weeks later for analysis via serial sectioning and staining for either Zsgreen/Pax7 or dystrophin/laminin. Analyses and quantification of the dystrophin-positive fibers was done as previously described[90]. Briefly, dystrophin positive muscle fibers within transverse sections from the mid-belly of TA muscle (n=3) were enumerated and compared between saline injected and cell transplanted cohorts. Analysis and quantification of donor derived satellite cells were enumerated based on double positive staining for both Zsgreen and Pax7.

**Microarray and Bioinformatic Analysis**

FACS Purified Zsgreen+ cells were obtained from pooled populations of fetal (Zsgreen+/Mcad-, n=4), adolescent (Zsgreen+/Mcad- and Zsgreen+/Mcad+, n=6), adult (Zsgreen+/Mcad+, n=8) and aged (Zsgreen+/Mcad+, n=7) muscle. Prior to submission for microarray analysis, total RNA was extracted from each replicate using the Picopure RNA isolation kit (Applied Biosystems, KIT0204) and concentrated using the RNA Clean and Concentrator-5 kit (Zymo Research) according to the manucfacturer's instructions. Sample RNA quality was assessed on the Agilent Bioanalyzer 2100 using the RNA 6000 nano chip (Agilent Technologies, Santa Clara). Target preparation was achieved using the WT Expression kit (Ambion) and the Genechip terminal labelling kit (Affymetrix). The microarray data was collected at Stemcore (Ottawa, ON) using Genechip mouse gene 1.0 ST arrays (Affymetrix, Santa Clara, CA). R-2.10.0 with the

XPS library was used to import CEL files, RMA normalize and perform DABG (detection above background) calls. Anti-genomic background and the metacore exon_level definition were used. The RMA expression values are log2 transformed. Probe annotation was extracted from ensembl v62 using the biomart R package and used to map transcript cluster ids to genes. The resulting five Mogene v1.0ST microarray data sets are referred to as Fetal, Adolescent (Adlscnt), Adult and Aged for subsequent methods of analysis. Genes that met the criteria of a log-fold change of greater than one (that is, 2-fold cut-off) were used to derive Tables 4.II and 4.III. Figure 4.5 was constructed from functional annotation clustering data mined with DAVID (http://david.abcc.ncifcrf.gov/home.jsp)[357,358].

## Acknowledgments

We thank Dr. Julia vonMaltzahn for helpful discussion and guidance as well as transplantating FACS sorted satellite cells into injured *mdx* muscle. Julia aided in sectioning and staining work done on the age-related satellite cell populations. Paul Oleynik conducted cell sorting and provided guidance regarding FACS isolation and analysis. We thank David Wilson for TA sectioning and staining. Jennifer Ritchie for mouse husbandry. We thank Gareth Palidwor for bioinformatic work conducted on the sorted satellite cell microarray data. Dr. M. Kyba for the Pax7Zsgreen mice. M.A.R. holds the Canada Research Chair in Molecular Genetics and is an International Research Scholar of the Howard Hughes Medical Institute. F.D.P was supported by the Canadian Stem Cell Network and a Doctoral Research Award from the Canadian Institutes of Health Research.

**Table 1. Cell counts of Dystrophin+ or Pax7+ and Zsgreen+ Satellite cells**

| Sample | Dys + | Lam + fibers | % Dys + | Average | STD |
|---|---|---|---|---|---|
| Saline 1 | 4 | 828 | 0.48 | | |
| Saline 2 | 1 | 575 | 0.17 | | |
| Saline 3 | 12 | 1071 | 1.12 | 0.59 | 0.48 |
| Fetal 1 | 97 | 2002 | 4.85 | | |
| Fetal 2 | 105 | 1727 | 6.08 | | |
| Fetal 3 | 91 | 1950 | 4.67 | 5.20 | 0.77 |
| Adolescent 1 | 64 | 1207 | 5.30 | | |
| Adolescent 2 | 54 | 1956 | 2.76 | | |
| Adolescent 3 | 34 | 959 | 3.55 | 3.87 | 1.30 |
| Adult 1 | 12 | 572 | 2.10 | | |
| Adult 2 | 16 | 351 | 4.56 | | |
| Adult 3 | 8 | 507 | 1.58 | 2.74 | 1.59 |
| Aged 1 | 14 | 2144 | 0.65 | | |
| Aged 2 | 8 | 1634 | 0.49 | | |
| Aged 3 | 19 | 2037 | 0.93 | 0.69 | 0.22 |

| Sample | Zsgreen + | Pax7 + | % Zsgreen + | Average | STD |
|---|---|---|---|---|---|
| Saline 1 | 0 | 0 | 0.00 | | |
| Saline 2 | 0 | 0 | 0.00 | | |
| Saline 3 | 0 | 0 | 0.00 | 0.00 | 0.00 |
| Fetal 1 | 152 | 342 | 44.44 | | |
| Fetal 2 | 167 | 348 | 47.99 | | |
| Fetal 3 | 192 | 310 | 61.94 | 51.46 | 9.25 |
| Adolescent 1 | 455 | 893 | 50.95 | | |
| Adolescent 2 | 161 | 281 | 57.30 | | |
| Adolescent 3 | 214 | 373 | 57.37 | 55.21 | 3.68 |
| Adult 1 | 81 | 204 | 39.71 | | |
| Adult 2 | 72 | 184 | 39.13 | | |
| Adult 3 | 125 | 530 | 23.58 | 34.14 | 9.15 |
| Aged 1 | 47 | 276 | 17.03 | | |
| Aged 2 | 89 | 352 | 25.28 | | |
| Aged 3 | 40 | 239 | 16.74 | | |
| Aged 4 | 18 | 103 | 17.48 | 19.13 | 4.11 |

## Table II. Preferentially Expressed Genes in Fetal Zsgreen+ Satellite Cells

| RMA Values | | | | | | Fold Change | |
|---|---|---|---|---|---|---|---|
| Fetal | AdIscnt | Adult | Aged | Gene | Description | (F/Ag) | Ensembl |
| **Cell Cycle** | | | | | | | |
| 10.74 | 8.90 | 6.70 | 5.94 | Ccne1 | cyclin E1 Gene | 28.029 | ENSMUSG00000002068 |
| 8.36 | 7.42 | 5.42 | 3.74 | Ncapg | non-SMC condensin I complex, subunit G Gene | 24.644 | ENSMUSG00000015880 |
| 9.39 | 5.95 | 5.09 | 4.81 | Tsga14 | testis specific gene A14 Gene | 23.796 | ENSMUSG00000029790 |
| 8.83 | 7.83 | 6.14 | 4.27 | C79407 | expressed sequence C79407 Gene | 23.608 | ENSMUSG00000047534 |
| 8.55 | 3.90 | 3.78 | 4.06 | Igf2bp3 | insulin-like growth factor 2 mRNA binding protein 3 Gene | 22.534 | ENSMUSG00000029814 |
| 9.23 | 8.28 | 6.13 | 4.76 | Dtl | denticleless homolog (Drosophila) Gene | 22.178 | ENSMUSG00000037474 |
| 9.05 | 7.97 | 6.24 | 4.69 | Bub1 | budding uninhibited by benzimidazoles 1 homolog (S. cerevisiae) Gene | 20.574 | ENSMUSG00000027379 |
| 9.78 | 8.89 | 7.34 | 5.44 | Kif11 | kinesin family member 11 Gene | 20.277 | ENSMUSG00000012443 |
| 8.46 | 7.20 | 5.40 | 4.12 | Kif15 | kinesin family member 15 Gene | 20.250 | ENSMUSG00000036768 |
| 7.96 | 6.48 | 4.86 | 3.67 | Cenpe | centromere protein E Gene | 19.587 | ENSMUSG00000045328 |
| 9.19 | 7.61 | 6.50 | 4.94 | Nusap1 | nucleolar and spindle associated protein 1 Gene | 19.090 | ENSMUSG00000027306 |
| 7.37 | 6.33 | 4.51 | 3.28 | Kif4 | kinesin family member 4 Gene | 17.017 | ENSMUSG00000034311 |
| 10.71 | 10.19 | 8.81 | 6.64 | Mki67 | antigen identified by monoclonal antibody Ki 67 Gene | 16.749 | ENSMUSG00000031004 |
| 8.43 | 7.47 | 5.56 | 4.37 | Ccne2 | cyclin E2 Gene | 16.735 | ENSMUSG00000028212 |
| 8.55 | 8.10 | 6.19 | 4.49 | Esco2 | establishment of cohesion 1 homolog 2 (S. cerevisiae) Gene | 16.684 | ENSMUSG00000022034 |
| 10.11 | 9.16 | 7.33 | 6.05 | Cdk1 | cyclin-dependent kinase 1 Gene | 16.627 | ENSMUSG00000019942 |
| 9.23 | 8.38 | 6.38 | 5.23 | Casc5 | cancer susceptibility candidate 5 Gene | 16.023 | ENSMUSG00000027326 |
| 7.80 | 7.51 | 5.34 | 3.84 | Hells | helicase, lymphoid specific Gene | 15.499 | ENSMUSG00000025001 |
| 9.38 | 8.52 | 6.74 | 5.44 | Ncaph | non-SMC condensin I complex, subunit H Gene | 15.348 | ENSMUSG00000034906 |
| 8.21 | 6.88 | 5.20 | 4.28 | 5730590G19Rik | RIKEN cDNA 5730590G19 gene | 15.196 | ENSMUSG00000046591 |
| 7.61 | 6.78 | 5.02 | 3.68 | Kntc1 | kinetochore associated 1 | 15.195 | ENSMUSG00000029414 |
| 7.49 | 7.00 | 5.41 | 3.57 | Mastl | microtubule associated serine/threonine kinase-like Gene | 15.093 | ENSMUSG00000026779 |
| 11.16 | 9.30 | 9.30 | 7.27 | Cdkn1c | cyclin-dependent kinase inhibitor 1C (P57) Gene | 14.824 | ENSMUSG00000037664 |
| 9.37 | 8.89 | 7.10 | 5.48 | Mcm6 | minichromosome maintenance deficient 6 Gene | 14.803 | ENSMUSG00000026355 |
| 9.96 | 9.02 | 7.52 | 6.13 | Aurkb | aurora kinase B Gene | 14.169 | ENSMUSG00000020897 |
| 7.75 | 6.77 | 4.72 | 3.94 | F630043A04Rik | RIKEN cDNA F630043A04 gene Gene | 13.992 | ENSMUSG00000021965 |
| 7.85 | 6.72 | 5.48 | 4.05 | Dlgap5 | discs, large (Drosophila) homolog-associated protein 5 Gene | 13.910 | ENSMUSG00000037544 |
| 8.50 | 7.25 | 5.52 | 4.74 | Spag5 | sperm associated antigen 5 Gene | 13.553 | ENSMUSG00000002055 |
| 9.83 | 9.22 | 8.90 | 6.14 | Hist1h2bb | histone cluster 1, H2bb Gene | 12.892 | ENSMUSG00000075031 |
| 9.24 | 8.16 | 6.81 | 5.56 | Kif23 | kinesin family member 23 Gene | 12.754 | ENSMUSG00000032254 |
| 8.06 | 7.15 | 5.46 | 4.41 | Aspm | asp (abnormal spindle)-like, microcephaly associated (Drosophila) Gene | 12.564 | ENSMUSG00000033952 |
| 8.45 | 7.75 | 6.28 | 4.86 | Prc1 | protein regulator of cytokinesis 1 Gene | 12.046 | ENSMUSG00000038943 |
| 10.66 | 10.11 | 8.85 | 7.07 | Ccna2 | cyclin A2 Gene | 12.002 | ENSMUSG00000027715 |
| 8.63 | 8.08 | 6.34 | 5.13 | Pole | polymerase (DNA directed), epsilon Gene | 11.340 | ENSMUSG00000007080 |
| 8.00 | 7.00 | 5.88 | 4.50 | Cenpf | centromere protein F Gene | 11.259 | ENSMUSG00000026605 |
| 8.72 | 8.71 | 6.40 | 5.23 | Chaf1b | chromatin assembly factor 1, subunit B (p60) Gene | 11.254 | ENSMUSG00000022945 |
| 7.75 | 6.80 | 5.27 | 4.27 | Sgol1 | shugoshin-like 1 (S. pombe) Gene | 11.187 | ENSMUSG00000023940 |
| 7.81 | 5.53 | 4.81 | 4.34 | Pard6g | par-6 partitioning defective 6 homolog gamma (C. elegans) Gene | 11.093 | ENSMUSG00000056214 |
| 9.81 | 8.88 | 7.81 | 6.36 | Ccnb2 | cyclin B2 Gene | 10.906 | ENSMUSG00000032218 |
| 7.46 | 6.77 | 5.52 | 4.02 | Spc25 | SPC25, NDC80 kinetochore complex component Gene | 10.875 | ENSMUSG00000005233 |
| 8.92 | 8.29 | 6.50 | 5.49 | Cenpn | centromere protein N Gene | 10.723 | ENSMUSG00000031756 |
| 7.47 | 6.76 | 4.97 | 4.06 | Cenpi | centromere protein I Gene | 10.602 | ENSMUSG00000031262 |
| **Intracelllular Signal Transduction** | | | | | | | |
| 11.11 | 6.77 | 3.62 | 2.93 | Nrk | Nik related kinase Gene | 290.875 | ENSMUSG00000052854 |
| 10.71 | 5.88 | 5.57 | 5.55 | Agtr2 | angiotensin II receptor, type 2 Gene | 35.690 | ENSMUSG00000068122 |
| 8.89 | 6.06 | 4.84 | 4.43 | Dcx | doublecortin Gene | 22.033 | ENSMUSG00000031285 |
| 8.87 | 5.59 | 5.14 | 4.48 | 2610018G03Rik | RIKEN cDNA 2610018G03 gene Gene | 21.041 | ENSMUSG00000031112 |
| 7.93 | 4.97 | 3.70 | 3.87 | Asb4 | ankyrin repeat and SOCS box-containing 4 Gene | 16.701 | ENSMUSG00000042607 |
| 9.43 | 8.46 | 6.84 | 5.51 | Ect2 | ect2 oncogene Gene | 15.174 | ENSMUSG00000027699 |
| 7.53 | 3.59 | 3.26 | 3.78 | Arhgap36 | Rho GTPase activating protein 36 Gene | 13.430 | ENSMUSG00000036198 |
| 7.38 | 6.71 | 4.92 | 3.76 | Shcbp1 | Shc SH2-domain binding protein 1 Gene | 12.326 | ENSMUSG00000022322 |
| 7.30 | 6.45 | 4.63 | 3.72 | Ttk | Ttk protein kinase Gene | 11.930 | ENSMUSG00000038379 |
| 8.40 | 7.16 | 5.82 | 4.83 | Arhgap11a | Rho GTPase activating protein 11A Gene | 11.894 | ENSMUSG00000041219 |
| 7.79 | 7.29 | 5.36 | 4.32 | Pbk | PDZ binding kinase Gene | 11.088 | ENSMUSG00000022033 |
| 9.45 | 7.78 | 6.94 | 5.99 | Zdbf2 | zinc finger, DBF-type containing 2 Gene | 10.969 | ENSMUSG00000027520 |
| 8.84 | 6.43 | 5.28 | 4.58 | Dock10 | dedicator of cytokinesis 10 Gene | 19.140 | ENSMUSG00000038608 |

**Secreted/Extracellular Matrix associated**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 10.95 | 6.47 | 5.11 | 4.80 | Vcan | versican Gene | 71.170 | ENSMUSG00000021614 |
| 10.68 | 6.98 | 4.95 | 5.27 | Tnc | tenascin C Gene | 42.491 | ENSMUSG00000028364 |
| 9.87 | 8.18 | 6.03 | 4.51 | Slc38a4 | solute carrier family 38, member 4 Gene | 41.174 | ENSMUSG00000022464 |
| 11.13 | 8.03 | 5.33 | 5.86 | Gpc3 | glypican 3 Gene | 38.553 | ENSMUSG00000055653 |
| 9.32 | 4.89 | 4.09 | 4.30 | Fbn2 | fibrillin 2 Gene | 32.544 | ENSMUSG00000024598 |
| 11.08 | 7.72 | 6.90 | 6.25 | Postn | periostin, osteoblast specific factor Gene | 28.444 | ENSMUSG00000027750 |
| 8.70 | 4.89 | 4.19 | 4.06 | Scn3b | sodium channel, voltage-gated, type III, beta Gene | 24.857 | ENSMUSG00000049281 |
| 7.95 | 6.66 | 4.90 | 3.56 | Hmmr | hyaluronan mediated motility receptor (RHAMM) Gene | 20.998 | ENSMUSG00000020330 |
| 9.53 | 6.46 | 5.86 | 5.44 | Kcne1l | potassium voltage-gated channel, Isk-related family, member 1-like | 17.038 | ENSMUSG00000090122 |
| 9.96 | 7.65 | 6.50 | 5.87 | Lamb1 | laminin B1 Gene | 17.005 | ENSMUSG00000002900 |
| 9.54 | 5.72 | 5.36 | 5.52 | Parm1 | prostate androgen-regulated mucin-like protein 1 Gene | 16.141 | ENSMUSG00000034981 |
| 10.31 | 8.25 | 6.59 | 6.31 | Col1a2 | collagen, type I, alpha 2 Gene | 16.033 | ENSMUSG00000029661 |
| 7.08 | 5.66 | 4.46 | 3.25 | Cpa5 | carboxypeptidase A5 Gene | 14.147 | ENSMUSG00000029788 |
| 8.20 | 6.31 | 5.12 | 4.41 | B3gnt5 | UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 5 Gene | 13.845 | ENSMUSG00000022686 |
| 8.05 | 4.84 | 3.83 | 4.37 | Sema3c | Semaphorin 3C Gene | 12.781 | ENSMUSG00000028780 |
| 7.93 | 7.10 | 5.56 | 4.30 | C330027C09Rik | RIKEN cDNA C330027C09 gene Gene | 12.404 | ENSMUSG00000033031 |
| 9.09 | 6.90 | 6.01 | 5.54 | Sfrp2 | secreted frizzled-related protein 2 Gene | 11.735 | ENSMUSG00000027996 |

**Striated Muscle Development**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 12.20 | 10.01 | 7.62 | 6.61 | Dlk1 | delta-like 1 homolog (Drosophila) Gene | 48.113 | ENSMUSG00000040856 |
| 9.85 | 6.82 | 5.45 | 4.41 | AW551984 | expressed sequence AW551984 Gene | 43.459 | ENSMUSG00000038112 |
| 9.18 | 4.47 | 4.19 | 3.80 | Myh3 | myosin, heavy polypeptide 3, skeletal muscle, embryonic Gene | 41.892 | ENSMUSG00000020908 |
| 8.66 | 4.84 | 3.21 | 3.64 | Mstn | myostatin Gene | 32.548 | ENSMUSG00000026100 |
| 12.07 | 11.30 | 9.99 | 7.49 | Actc1 | actin, alpha, cardiac muscle 1 Gene | 23.790 | ENSMUSG00000068614 |
| 10.05 | 9.14 | 8.12 | 5.81 | Hist1h1b | histone cluster 1, H1b Gene | 18.885 | ENSMUSG00000058773 |
| 7.97 | 6.84 | 4.50 | 3.79 | Stil | Scl/Tal1 interrupting locus Gene | 18.103 | ENSMUSG00000028718 |
| 8.81 | 5.43 | 4.64 | 4.71 | Myh8 | myosin, heavy polypeptide 8, skeletal muscle, perinatal Gene | 17.055 | ENSMUSG00000055775 |
| 10.33 | 7.85 | 6.44 | 6.43 | Myl4 | myosin, light polypeptide 4 Gene | 14.894 | ENSMUSG00000061086 |
| 7.75 | 4.91 | 3.87 | 4.00 | Mbnl3 | muscleblind-like 3 (Drosophila) Gene | 13.419 | ENSMUSG00000036109 |
| 9.23 | 7.61 | 6.25 | 5.59 | Ttn | titin Gene | 12.468 | ENSMUSG00000051747 |
| 9.15 | 7.97 | 6.55 | 5.67 | Actn2 | actinin alpha 2 Gene | 11.098 | ENSMUSG00000052374 |

**DNA Binding**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 7.68 | 7.48 | 6.91 | 2.91 | Hist1h2ab | histone cluster 1, H2ab Gene | 27.191 | ENSMUSG00000061615 |
| 9.78 | 9.00 | 7.10 | 5.26 | Top2a | topoisomerase (DNA) II alpha Gene | 22.908 | ENSMUSG00000020914 |
| 11.73 | 8.86 | 7.59 | 7.39 | AC084315.1 | retrotransposon-derived protein PEG10 isoform 1 | 20.183 | ENSMUSG00000092035 |
| 7.98 | 3.73 | 3.77 | 3.65 | Atp1b4 | ATPase, (Na+)/K+ transporting, beta 4 polypeptide Gene | 20.057 | ENSMUSG00000016327 |
| 8.55 | 7.96 | 6.25 | 4.40 | Neil3 | nei like 3 (E. coli) Gene | 17.793 | ENSMUSG00000039396 |
| 11.65 | 9.94 | 8.62 | 7.90 | Plagl1 | pleiomorphic adenoma gene-like 1 Gene | 13.509 | ENSMUSG00000019817 |
| 9.11 | 6.72 | 5.42 | 5.54 | Raver2 | ribonucleoprotein, PTB-binding 2 Gene | 11.858 | ENSMUSG00000035275 |
| 9.95 | 9.11 | 7.92 | 6.39 | Hist1h1a | histone cluster 1, H1a Gene | 11.845 | ENSMUSG00000049539 |

**Misc**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 9.63 | 5.64 | 5.23 | 4.43 | Nnat | neuronatin Gene | 36.698 | ENSMUSG00000067786 |
| 8.82 | 4.38 | 3.00 | 3.83 | Cpa3 | carboxypeptidase A3, mast cell Gene | 31.785 | ENSMUSG00000001865 |
| 11.24 | 9.35 | 7.96 | 6.81 | Nt5dc2 | 5'-nucleotidase domain containing 2 Gene | 21.423 | ENSMUSG00000071547 |
| 8.74 | 7.21 | 4.77 | 5.22 | Ddah1 | dimethylarginine dimethylaminohydrolase 1 Gene | 11.470 | ENSMUSG00000028194 |
| 7.84 | 6.97 | 5.83 | 4.35 | Trim59 | tripartite motif-containing 59 Gene | 11.265 | ENSMUSG00000034317 |
| 7.54 | 7.26 | 5.44 | 4.05 | Fam111a | family with sequence similarity 111, member A Gene | 11.233 | ENSMUSG00000024691 |
| 8.52 | 7.83 | 6.20 | 5.05 | Ckap2l | cytoskeleton associated protein 2-like Gene | 11.036 | ENSMUSG00000048327 |

## Table II. Preferentially Expressed Genes in Aged Zsgreen+ Satellite Cells

| RMA Values | | | | | | Fold Change | |
|---|---|---|---|---|---|---|---|
| Fetal | Adolescent | Adult | Aged | Gene | Description | (Ag/F) | Ensembl |
| **Circadian Rhythm** | | | | | | | |
| 6.09 | 10.08 | 10.99 | 10.91 | Nr1d1 | nuclear receptor subfamily 1, group D, member 1 Gene | 28.12 | ENSMUSG00000020889 |
| 3.76 | 6.64 | 8.14 | 8.27 | Bhlhe41 | basic helix-loop-helix family, member e41 Gene | 22.68 | ENSMUSG00000030256 |
| 5.80 | 8.03 | 8.85 | 9.35 | Npas2 | neuronal PAS domain protein 2 Gene | 11.70 | ENSMUSG00000026077 |
| 5.96 | 6.69 | 7.46 | 9.17 | Bhlhe40 | basic helix-loop-helix family, member e40 Gene | 9.24 | ENSMUSG00000030103 |
| 8.07 | 8.34 | 8.76 | 9.74 | Per1 | period homolog 1 (Drosophila) Gene | 3.19 | ENSMUSG00000020893 |
| 5.49 | 6.68 | 6.64 | 7.15 | Arntl | aryl hydrocarbon receptor nuclear translocator-like Gene | 3.16 | ENSMUSG00000055116 |
| 5.41 | 6.73 | 7.09 | 6.60 | Per3 | period homolog 3 (Drosophila) Gene | 2.28 | ENSMUSG00000028957 |
| 6.42 | 6.44 | 7.10 | 7.48 | Cry2 | cryptochrome 2 (photolyase-like) Gene | 2.07 | ENSMUSG00000068742 |
| **Transcription factors** | | | | | | | |
| 4.78 | 6.37 | 8.71 | 9.97 | Cebpd | CCAAT/enhancer binding protein (C/EBP), delta | 37.04 | ENSMUSG00000071637 |
| 6.09 | 10.08 | 10.99 | 10.91 | Nr1d1 | nuclear receptor subfamily 1, group D, member 1 Gene | 27.78 | ENSMUSG00000020889 |
| 4.54 | 9.36 | 9.36 | 9.23 | Esr1 | estrogen receptor 1 (alpha) Gene | 25.64 | ENSMUSG00000019768 |
| 6.02 | 9.36 | 10.26 | 10.56 | Klf9 | Kruppel-like factor 9 Gene | 23.26 | ENSMUSG00000033863 |
| 3.76 | 6.64 | 8.14 | 8.27 | Bhlhe41 | basic helix-loop-helix family, member e41 Gene | 22.73 | ENSMUSG00000030256 |
| 6.70 | 7.05 | 8.00 | 10.88 | Atf3 | activating transcription factor 3 Gene | 18.18 | ENSMUSG00000026628 |
| 6.70 | 8.02 | 9.32 | 10.71 | Nr4a1 | nuclear receptor subfamily 4, group A, member 1 Gene | 16.13 | ENSMUSG00000023034 |
| 5.93 | 7.56 | 9.68 | 9.87 | Junb | Jun-B oncogene Gene | 15.38 | ENSMUSG00000052837 |
| 6.44 | 7.62 | 9.25 | 10.18 | Id3 | inhibitor of DNA binding 3 Gene | 13.33 | ENSMUSG00000007872 |
| 5.80 | 8.03 | 8.85 | 9.35 | Npas2 | neuronal PAS domain protein 2 Gene | 11.76 | ENSMUSG00000026077 |
| 5.72 | 8.60 | 9.50 | 9.25 | Bcl6 | B-cell leukemia/lymphoma 6 Gene | 11.49 | ENSMUSG00000022508 |
| 5.10 | 7.18 | 7.90 | 8.46 | Mlxipl | MLX interacting protein-like Gene | 10.31 | ENSMUSG00000005373 |
| 5.19 | 6.33 | 7.22 | 8.47 | Glis3 | GLIS family zinc finger 3 Gene | 9.71 | ENSMUSG00000052942 |
| 5.96 | 6.69 | 7.46 | 9.17 | Bhlhe40 | basic helix-loop-helix family, member e40 Gene | 9.26 | ENSMUSG00000030103 |
| 6.74 | 8.49 | 9.06 | 9.93 | Ankrd1 | ankyrin repeat domain 1 (cardiac muscle) Gene | 9.17 | ENSMUSG00000024803 |
| 5.54 | 5.62 | 5.86 | 8.73 | Csrnp1 | cysteine-serine-rich nuclear protein 1 Gene | 9.09 | ENSMUSG00000032515 |
| **Vascular Development** | | | | | | | |
| 5.93 | 7.56 | 9.68 | 9.87 | Junb | Jun-B oncogene Gene | 15.29 | ENSMUSG00000052837 |
| 7.29 | 9.43 | 10.76 | 10.96 | Apoe | apolipoprotein E Gene | 12.76 | ENSMUSG00000002985 |
| 3.42 | 5.76 | 6.88 | 6.83 | Il18 | interleukin 18 Gene | 10.67 | ENSMUSG00000039217 |
| 6.74 | 7.34 | 8.86 | 9.67 | Ctgf | connective tissue growth factor Gene | 7.61 | ENSMUSG00000019997 |
| 8.56 | 8.55 | 10.11 | 11.37 | Socs3 | suppressor of cytokine signaling 3 Gene | 7.03 | ENSMUSG00000053113 |
| 7.06 | 7.88 | 9.38 | 9.84 | Id1 | inhibitor of DNA binding 1 Gene | 6.85 | ENSMUSG00000042745 |
| 6.20 | 8.37 | 7.89 | 8.87 | Epas1 | endothelial PAS domain protein 1 Gene | 6.36 | ENSMUSG00000024140 |
| 8.29 | 9.52 | 10.68 | 10.74 | Tgfbr3 | transforming growth factor, beta receptor III Gene | 5.45 | ENSMUSG00000029287 |
| 7.54 | 9.45 | 9.99 | 9.94 | Plxnd1 | plexin D1 Gene | 5.29 | ENSMUSG00000030123 |
| 4.68 | 5.20 | 6.05 | 6.98 | Cx3cl1 | chemokine (C-X3-C motif) ligand 1 Gene | 4.93 | ENSMUSG00000031778 |
| 4.61 | 5.21 | 5.62 | 6.67 | Zc3h12a | zinc finger CCCH type containing 12A Gene | 4.17 | ENSMUSG00000042677 |
| 7.60 | 8.74 | 9.52 | 9.60 | Foxo1 | forkhead box O1 Gene | 4.01 | ENSMUSG00000044167 |
| 5.58 | 7.23 | 6.38 | 7.57 | Flt1 | FMS-like tyrosine kinase 1 Gene | 3.95 | ENSMUSG00000029648 |
| 9.01 | 10.92 | 11.46 | 10.90 | Bmp4 | bone morphogenetic protein 4 Gene | 3.70 | ENSMUSG00000021835 |

**Secreted/Extracellular Matrix**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 4.77 | 7.58 | 8.84 | 9.45 | Vtn | vitronectin Gene | 25.58 | ENSMUSG00000017344 |
| 5.44 | 8.85 | 9.72 | 10.11 | Crlf1 | cytokine receptor-like factor 1 Gene | 25.39 | ENSMUSG00000007888 |
| 6.30 | 9.39 | 10.65 | 10.95 | Prelp | proline arginine-rich end leucine-rich repeat Gene | 25.16 | ENSMUSG00000041577 |
| 7.50 | 11.50 | 12.12 | 12.12 | Gpx3 | glutathione peroxidase 3 Gene | 24.61 | ENSMUSG00000018339 |
| 4.94 | 9.10 | 9.71 | 9.52 | Cp | ceruloplasmin Gene | 23.79 | ENSMUSG00000003617 |
| 1.95 | 5.13 | 7.10 | 6.44 | Retnla | resistin like alpha Gene | 22.56 | ENSMUSG00000061100 |
| 4.83 | 8.55 | 7.57 | 9.08 | Timp4 | tissue inhibitor of metalloproteinase 4 Gene | 19.07 | ENSMUSG00000030317 |
| 4.16 | 8.56 | 9.61 | 8.38 | Gal | galanin Gene | 18.71 | ENSMUSG00000024907 |
| 4.88 | 8.03 | 8.93 | 9.06 | Lama3 | laminin, alpha 3 Gene | 18.12 | ENSMUSG00000024421 |
| 5.09 | 9.01 | 9.44 | 9.10 | Ptprz1 | protein tyrosine phosphatase, receptor type Z, polypeptide 1 Gene | 16.10 | ENSMUSG00000068748 |
| 5.41 | 5.72 | 7.03 | 9.33 | Serpine1 | serine (or cysteine) peptidase inhibitor, clade E, member 1 Gene | 15.12 | ENSMUSG00000037411 |
| 4.84 | 7.29 | 8.24 | 8.68 | Clcf1 | cardiotrophin-like cytokine factor 1 Gene | 14.24 | ENSMUSG00000040663 |
| 7.29 | 9.43 | 10.76 | 10.96 | Apoe | apolipoprotein E Gene | 12.76 | ENSMUSG00000002985 |
| 3.22 | 5.39 | 6.95 | 6.85 | Serpini1 | serine (or cysteine) peptidase inhibitor, clade I, member 1 Gene | 12.39 | ENSMUSG00000027834 |
| 6.20 | 9.43 | 9.44 | 9.80 | Chrdl2 | chordin-like 2 Gene | 12.17 | ENSMUSG00000030732 |
| 5.08 | 9.02 | 8.17 | 8.50 | Fam132b | family with sequence similarity 132, member B Gene | 10.70 | ENSMUSG00000047443 |
| 3.42 | 5.76 | 6.88 | 6.83 | Il18 | interleukin 18 Gene | 10.67 | ENSMUSG00000039217 |
| 4.42 | 6.53 | 7.59 | 7.77 | Ecm2 | extracellular matrix protein 2, female organ and adipocyte specific Gene | 10.21 | ENSMUSG00000043631 |
| 2.85 | 3.59 | 5.12 | 5.81 | Spock3 | sparc/osteonectin, cwcv and kazal-like domains proteoglycan 3 Gene | 7.81 | ENSMUSG00000054162 |
| 6.74 | 7.34 | 8.86 | 9.67 | Ctgf | connective tissue growth factor Gene | 7.61 | ENSMUSG00000019997 |
| 5.64 | 6.04 | 7.71 | 8.29 | Otoa | otoancorin Gene | 6.26 | ENSMUSG00000034990 |
| 4.59 | 6.41 | 6.86 | 7.20 | Mamdc2 | MAM domain containing 2 Gene | 6.11 | ENSMUSG00000033207 |
| 4.85 | 7.21 | 7.09 | 7.39 | Entpd2 | ectonucleoside triphosphate diphosphohydrolase 2 Gene | 5.83 | ENSMUSG00000015085 |
| 6.91 | 8.49 | 9.21 | 9.41 | Tnxb | tenascin XB Gene | 5.65 | ENSMUSG00000033327 |
| 8.29 | 9.52 | 10.68 | 10.74 | Tgfbr3 | transforming growth factor, beta receptor III Gene | 5.45 | ENSMUSG00000029287 |
| 9.26 | 10.83 | 11.44 | 11.54 | Dpt | dermatopontin Gene | 4.88 | ENSMUSG00000026574 |
| 5.10 | 5.90 | 6.77 | 7.31 | Lgals3bp | lectin, galactoside-binding, soluble, 3 binding protein Gene | 4.63 | ENSMUSG00000033880 |
| 8.52 | 10.22 | 10.54 | 10.67 | Hmcn2 | hemicentin 2 Gene | 4.43 | ENSMUSG00000055632 |
| 4.95 | 5.86 | 6.52 | 7.02 | Crispld2 | cysteine-rich secretory protein LCCL domain containing 2 Gene | 4.22 | ENSMUSG00000031825 |
| 9.01 | 10.92 | 11.46 | 10.90 | Bmp4 | bone morphogenetic protein 4 Gene | 3.70 | ENSMUSG00000021835 |
| 4.50 | 4.64 | 4.77 | 6.36 | Spock1 | sparc/osteonectin, cwcv and kazal-like domains proteoglycan 1 Gene | 3.62 | ENSMUSG00000056222 |
| 5.96 | 6.07 | 6.47 | 7.70 | Wnt4 | wingless-related MMTV integration site 4 Gene | 3.35 | ENSMUSG00000036856 |

## Supplementary Tables and Figures

### Table SI. Antibodies Used in this Study

| Primary Antibodies | Dilution | Company | Cat# |
|---|---|---|---|
| Pax7 (msIgG1) | 1:10 | Developmental Studies Hybridoma Bank | PAX7 |
| Myosin Heavy Chain (msIgG2b) | 1:10 | Developmental Studies Hybridoma Bank | MF-20 |
| Alexa647-a7 Integrin | 1μg/Million Cells | UBC ablab | a7 |
| Biotin-b1 Integrin | 1μg/Million Cells | eBioscience | 13-0291 |
| Biotin-Cxcr4 | 1μg/Million Cells | BD Bioscience | 551968 |
| Biotin-Pdgfra | 1μg/Million Cells | eBioscience | 13-1401-82 |
| Biotin-CD34 | 1μg/Million Cells | eBioscience | 13-0341-85 |
| Mcadherin (IgG1) | 1μg/Million Cells | Millipore | 05-852 |
| Laminin (Rb Poly) | 1:1000 | Sigma | L9393 |
| Dystrophin | 1:1 | Developmental Studies Hybridoma Bank | 7A10 |
| Mouse IgG1 Isotype | 1μg/Million Cells | BD Bioscience | 554121 |
| ZsGreen | 1:200 | Clontech | 632474 |
| **Secondary Antibodies** | | | |
| Goat anti-Mouse IgG1 (A488) | 1:1000 | Invitrogen - Molecular Probes | A21121 |
| Goat anti-Mouse IgG1 (A568) | 1:1000 | Invitrogen - Molecular Probes | A21124 |
| Goat anti mouse IgG1 (A647) | 1:1000 | Invitrogen - Molecular Probes | A21240 |
| Streptavidin ApcCy7 | 1:1000 | BD Bioscience | 554063 |

**Figure S1. Zsgreen + cells decrease with Age**
Quantification of Zsgreen + cells relative to total cells in age-related muscle digests. Data represent mean ± SE for fetal (n=7), adolescent (n=7), adult (n=7), and aged (n=7). ** $p<0.01$.

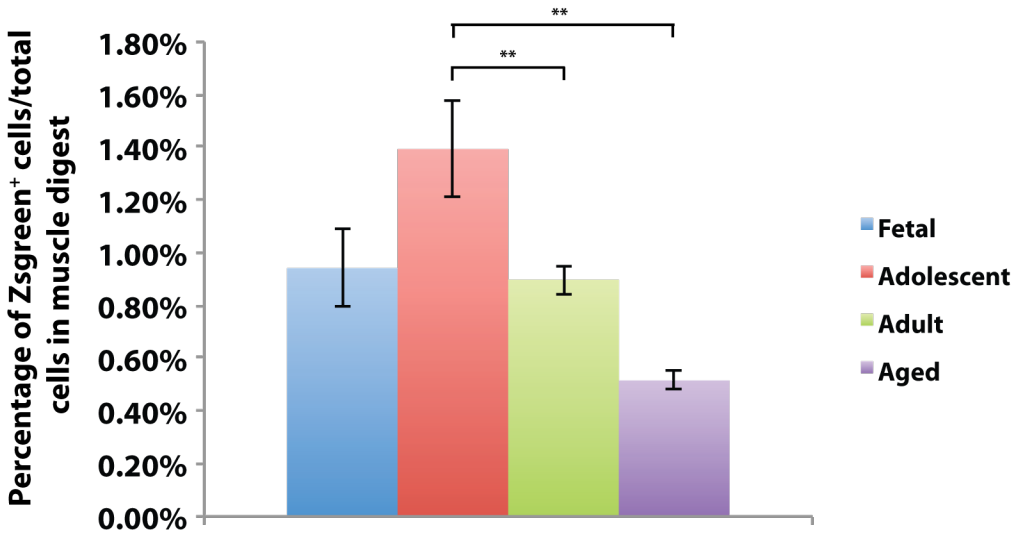
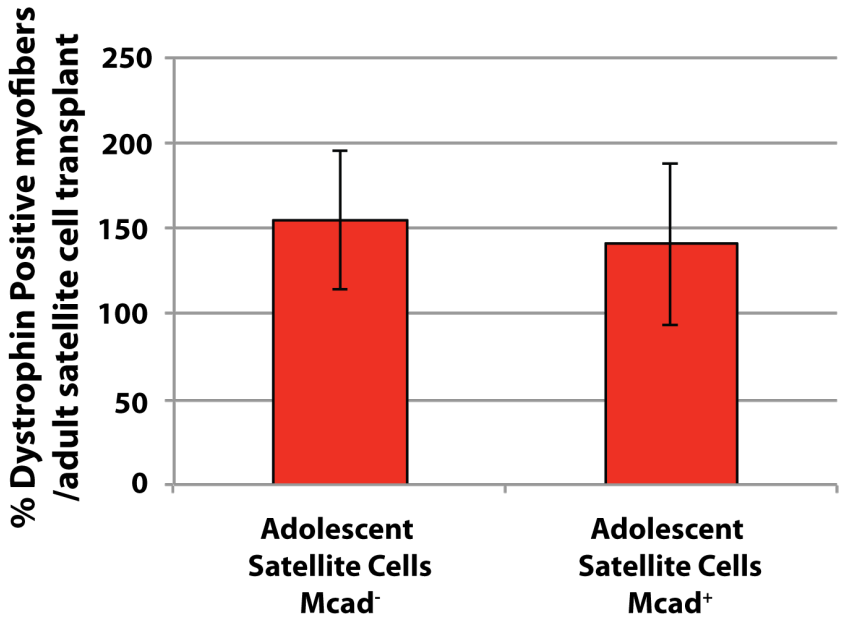
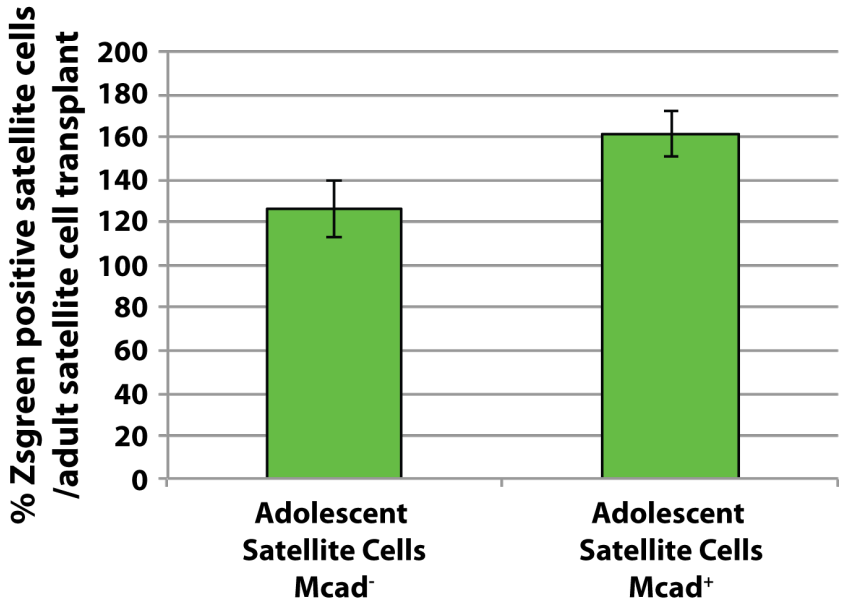

Percentage of Zsgreen+ cells/total cells in muscle digest
1.80%
1.60%
1.40%
1.20%
1.00%
0.80%
0.60%
0.40%
0.20%
0.00%
**
**
Fetal
Adolescent
Adult
Aged

**Figure S2. Repopulating capacity of adolescent satellite cells remains unchanged irrespective of M-cadherin expression.**
Freshly sorted Zsgreen+ and either Mcad- or Mcad+ satellite cells (10,000) from adolescent mice were transplanted into cardiotoxin treated TA muscle of *mdx* mice between the ages of 6-8 weeks (n=3 for each time point). *mdx* TA muscle was sectioned 3 weeks following transplantation and quantification of the percentage of dystrophin positive fibers (**A**) or Pax7+/Zsgreen+ cells in relation to total Pax7+ satellite cells (**B**) was conducted. Both percentages are relative to adult donor Zsgreen+ Mcad+ cells. * p value < 0.05. ** p value < 0.01.

A


% Dystrophin Positive myofibers /adult satellite cell transplant
250
200
150
100
50
0
Adolescent Satellite Cells Mcad-
Adolescent Satellite Cells Mcad+

B


% Zsgreen positive satellite cells /adult satellite cell transplant
200
180
160
140
120
100
80
60
40
20
0
Adolescent Satellite Cells Mcad-
Adolescent Satellite Cells Mcad+

# Chapter 5 - General Discussion

## 5.1. Overview of findings

The primary aim of my graduate work was to generate, isolate and characterise myogenic progenitors both from the embryonic and adult states of mouse development. The experiments outlined in Chapter 2 were designed to test the ability of Wnts to promote skeletal muscle formation. Our hypothesis was based on experiments that demonstrate Wnt3a can induce myogenesis in developing embryonic somites *in vivo* [243,359]. Surprisingly, we determined that the continued expression of Wnt3a in mES cells resulted in the induction of a metastable primitive endoderm state and that following differentiation mES cells were predisposed for entry into the visceral endoderm.

In Chapter 3, we attempted to create genetic reporters for two key myogenic genes *Pax7* and *Myf5*. We were unable to create either Knock-In mice or transgenic mice that expressed our fluorophores of interest under the control of either the Pax7 or Myf5 promoters. However, tools including a multifunctional 2a triple Ypet *Gaussia* luciferase reporter along with experimental data outlining the shortfalls of the P7YFP and M5TdTom vectors create a foundation for future research into the creation of myogenic reporters.

In Chapter 4, satellite cells at different stages of development were isolated and characterised based on their in vivo regenerative capacity and genome expression profiles. We discovered an intrinsic decline in the regenerative capacity and repopulating ability associated with aged satellite cells. Furthermore, we identified gene pathways that distinguish fetal from aged satellite cells. This microarray data will be useful in future

experiments to identify mechanisms that underlie the observed functional differences between satellite cells isolated from different stages of development.

### *5.1.1 Wnt3a Induces a Metastable Primitive Endoderm Population in mES Cells*

Although previous research has identified a role for canonical Wnt signalling in maintaining pluripotency in mES cells; surprisingly, in the absence of functional β-catenin (β-cat$^{\Delta/\Delta}$) mouse ES cells retain their ability to self-renew[183,191,193,360]. These experiments indicate that although canonical Wnt signalling promotes pluripotency, it is not exclusively required. Presently we have demonstrated that Wnt3a is able to induce a primitive endoderm metastable state and potentiate the differentiation of visceral endoderm in mES cells. Our results elucidate novel roles for Wnt3a signalling in the undifferentiated and differentiated states of mES cells.

Culture of undifferentiated mES cells overexpressing Wnt3a exhibited a slight phenotype in colony morphology and a significant increase in their growth rate. Following LIF removal and aggregation, embryoid bodies expressing Wnt3a increased in diameter and subsequently developed a striking cystic phenotype. Global gene expression analysis by RNA-seq identified a bias for both primitive endoderm and pluripotency genes in Wnt3a overexpressing cells when maintained in the undifferentiated state. In differentiation conditions, Wnt3a induced genes specific for the visceral endoderm lineage. To corroborate our RNA-seq results we demonstrated by qPCR and immunofluorescent staining that primitive and visceral endoderm markers were enriched in Wnt3a expressing cultures.

Induction of both primitive endoderm and pluripotency genes suggested that Wnt3a could be inducing a primitive endoderm metastable state. Indeed, Wnt3a treatment

of Nanog-GFP cells enriched a GFP low population containing cells that express elevated levels of primitive endoderm genes. Interestingly, this distinct population of Nanog low cells maintained the ability to interconvert back to Nanog GFP high cells following culture *in vitro*. This ability to interconvert is a hallmark of a metastable state. As this population was increased in Wnt3a expressing cells, this data implicates Wnt3a in the induction of a primitive endoderm metastable population.

To determine the mechanism responsible for Wnt3a induction of a primitive endoderm metastable state we compared genes that were induced following Wnt3a overexpression and Wnt3a recombinant protein treatment. Tbx3 was induced in cells expressing or treated with Wnt3a. Prior studies implicate Tbx3 in maintaining pluripotency and promoting primitive endoderm[255,265]. Based on siRNA knockdown of Tbx3, our results demonstrated that Wnt3a activation of Tbx3 mediates the induction of the primitive endoderm metastable state. Our data in combination with reports that Wnt3a can directly activate the promoter of Tbx3 in human liver tumours suggests that Tbx3 is a direct target of canonical Wnt signalling[266]. These findings present novel insight into the role played by Wnt3a in pluripotency and provide evidence of mechanisms that govern the induction of a primitive endoderm metastable state.

### *5.1.2 Intrinsic Decline in Repopulating Capacity of Aging Satellite Cells*

Although age negatively affects the ability of skeletal muscle to function and regenerate, the predominant cellular and molecular mechanisms that regulate this process remain unknown. Presently, we demonstrated that purified populations of satellite cells, isolated from a Pax7Zsgreen reporter mouse, display age-related differences in

regenerative capacity and global gene expression signatures. This functional variability led us to conclude that there exist age-related intrinsic differences between satellite cells.

This age associated variation in expression of M-cadherin renders it a promising marker to distinguish satellite cells based on their age. However, the presence of M-cadherin does not correlate with any changes in the functional capacity of adolescent satellite cells *in vivo* or the myogenic capacity of adolescent or adult satellite cell derived primary myoblasts *in vitro*.

The age of donor satellite cells significantly affected their ability to contribute to muscle regeneration and repopulate the endogenous satellite cell niche following transplantation. Fetal satellite cells showed improvement relative to aged satellite cells in their ability to differentiate into Dystrophin$^{+}$ myofibers and repopulate the satellite cell niche.

Microarray analysis, conducted on each age-associated population of satellite cells, elucidated cell intrinsic differences in global gene expression. Fetal satellite cells displayed an inherent increase in genes associated with cell cycle, cell division and DNA replication indicative of proliferating cells. Aged satellite cells had down regulated expression of cell cycle related genes and displayed enhanced expression of bZIP transcription factors along with increased expression of genes involved in vascular development and nuclear hormone receptor synthesis. These findings provide novel insight into the intrinsic molecular regulation of satellite cells throughout the aging process.

## 5.2. Defining the mechanisms involved in modulating the metastable state

*4.2.1. Significance*

Our research into the role Wnt3a plays in affecting the interconversion of two cellular states increases our understanding of the mechanisms involved in the regulation of metastability. Hayashi *et al.* determined that Stella GFP+ cells are enriched following culture on MEFs and diminished when cultured on gelatin[29]. This implies that the metastable state can be influenced by environmental cues. Our research provides direct evidence that a secreted growth factor can modulate the metastable state.

Wnt3a plays multiple roles in embryonic development. Our data links canonical Wnt signalling with the formation of the nascent visceral endoderm. This research provides a novel explanation for why Tcf-based reporters are first activated in visceral endoderm[216] and why $Apc^{-/-}$ ES cells give rise to teratomas biased for visceral endoderm differentiation[211]. Additionally, high-throughput sequencing of RNA/cDNA molecules (RNA-Seq) in mES cells (following overexpression or treatment with recombinant Wnt3) bestows an advantage on the quantification of absolute mRNA levels, which is critical for comparing expression levels between genes yet inaccessible by the microarray approach.

The analysis of the metastable state and the mechanisms at play provide key information on how two cellular states interconvert. With the advent of induced pluripotent cells (iPSCs) the ability to convert a somatic cell back to a pluripotent state is now possible. However, many unresolved questions remain regarding reprogramming. Considering Wnt3a has been used to enhance cellular reprogramming could it be doing so in a Tbx3 dependent manner? This is possible as Tbx3 is highly expressed in the pluripotent state and regulated by LIF expression[255,265]. If indeed this is the case, potential

implications for primitive endoderm priming must be considered if reprogramming with canonical Wnt activators.

*5.2.2. Future Directions*

Our data indicates that Tbx3 mediates Wnt3a induction of a primitive endoderm metastable state. However, it remains to be determined whether or not β-catenin or Tcf/Lef transcription factors directly bind the Tbx3 promoter. Future experiments to address this issue, whether by direct chromatin immunoprecipitation (ChIP) or ChIP-seq for global binding sites in mES cells would be beneficial. Additional information regarding the complex members for nuclear β-catenin following Wnt stimulation would provide a better understanding of the mechanism by which Wnt3a regulates the activation of primitive endoderm genes.

We observe activation of core primitive endoderm transcription factors (Gata6, Gata4, Sox17) following Wnt3a induction. Tbx3 is able to bind and activate the promoter of Gata6[255], which would provide an explanation for why we observe increased expression following Wnt3a treatment. However future experiments to identify why Sox17 and Gata4 are induced are necessary. Recent reports suggest a novel role for Nanog in activating Sox17 and Gata4 in late stage blastocysts[24,25]. This is in contrast to previous reports that implicate Nanog in the repression of Gata6 and the primitive endoderm[12,361]. Considering Nanog is induced following Wnt3a overexpression and is a known downstream target of Wnt signalling in mES cells[196] further experiments to characterize this role in mES cells would be beneficial.

In addition to Tbx3 our RNA-seq data identified the upregulation of Cdx1 following both overexpression and recombinant treatment of Wnt3a. Cdx1 is a known

downstream target of canonical Wnt signalling in mES cells[362] and is modulated by Oct4[363]. Future experiments involving siRNA knockdown and/or overexpression of Cdx1 in mES cells would clarify if this transcription factor is playing a role in the induction of primitive endoderm.

To date the metastable state has been characterized based on the oscillations of multiple transcription factors. Experiments to unify this data are required. Single cell analysis via qPCR or standardized cell surface markers to sort and characterize the various metastable populations would greatly aid this emerging area of research. In characterizing the sorted Nanog-GFP high and low populations we discovered enrichment for the cell surface marker Cubillin in the metastable primitive endoderm. In future experiments we plan to sort cells based on Cubillin from NanogGFP[27], StellaGFP[29], and Sox17GFP[364] to further characterize this metastable population.

## 5.3. Age-related effects on skeletal muscle regenerative capacity

*5.3.1. Significance*

The age-dependent decline in muscle maintenance and regenerative capacity is attributed to a combination of changes in the systemic and local milieu surrounding the satellite cell niche and to changes in the satellite cells themselves. To date the principle avenue of research to identify these age-related changes affecting muscle function has focused on humoral factors in the circulation[162]. These experiments have identified roles for Notch[325] and Wnt[163] signalling pathways in the regulation of regenerative capacity.

The age related differences we observe in satellite cells with respect to cell surface markers, regenerative capacity, and repopulating ability strongly support the notion that intrinsic properties play a functional role in regulating myogenic capacity.

Our results are unique as they assay freshly sorted $Pax7^{+}$ satellite cells according to their age-dependent regenerative capacity following transplantation. The satellite cell intrinsic regenerative phenotype we observe, along with differences in genome expression profiles are the first of their kind and provide evidence to mechanistically address this observed age-related phenotype.

*5.3.2. Future Directions*

Extracellular components of the satellite cell niche along with signalling molecules provided by the local and systemic environment play a critical role in affecting the cellular state of satellite cells. The regenerative capacity of satellite cells rapidly degrades following their removal from the satellite cell niche and subsequent culture *in vitro*[340]. Our research identifies expression, at the mRNA level, for components of the satellite cell niche that are directly secreted by satellite cells. Interestingly, these components differ with age.

Our future goals are to experimentally address the mechanistic implications of genes that are enriched in fetal satellite cells to determine whether they impart a similar function effect to adult satellite cells. Experiments will involve treatment of sorted adult satellite cells with secreted growth factors and ECM molecules specific for each age group prior to transplantation into *mdx* mice.

In fetal satellite cells we observed the robust upregulation in expression for three ECM molecules namely; Versican (Vcan), Glypican3 (Gpc3) and Tensascin C (Tnc). Vcan is a chondroitin sulphate proteoglycan and a main component of the extracellular matrix[365]. Vcan is expressed in adult skeletal muscle and is present in the satellite cell niche (Personal communication Dr. N. Stupka). Additionally versican is a target of Pax3

and implicated in neural crest cell migration[366,367]. Gpc3 is a heparin sulphate proteoglycan that functions as a low-affinity co-receptor for FGF2[368]. Additionally Gpc3 can bind to Wnt and stabilize Wnt:Frizzled interactions[369,370]. Tnc is known to antagonize the adhesive effects of fibronectin by suppressing fibronectin-mediated activation of RhoA and focal adhesion kinase[371,372] and to generate an environmental niche for neural stem cells[373]. Furthermore, Tnc is implicated along with Periostin in modulating the extracellular meshwork architechture[374]. The expression of these ECM proteins in fetal satellite cells and their broad roles in cellular growth, growth factor binding and ECM remodelling along with their availability in recombinant form make them suitable candidates to test their effect on satellite cell function.

In addition to ECM components our microarray identified several secreted growth factors enriched in fetal satellite cells. Periostin (Pstn), Platelet derived growth factor c (Pdgfc), Myostatin (Mstn), soluble frizzled related protein 2 (Sfrp2). Pstn is a well known factor involved in inducing epithelial to mesenchymal transitions (EMTs), while the mitogen Pdgfc serves as a chemoatractant for mesenchymal cells[375]. Mstn, is a growth and differentiation factor that serves to inhibit muscle differentiation[354]. Sfrp2 was previously identified as being upregulated in young skeletal muscle, a finding that correlates with additional research demonstrating the an increase in Wnt signalling in old age negatively affects muscle regeneration[163,376]. In this fashion we can test molecules enriched in either fetal or aged satellite cells that modulate the regenerative and repopulating capacity of adult satellite cells.

The discovery that satellite cells have age-related intrinsic differences in regenerative capacity raises the question of whether or not satellite stem cells decrease

with age. Previous research has shown that satellite stem cells display a functional advantage over satellite myogenic cells with regards to their regenerative capacity. Future experiments to quantify the proportion of satellite stem cells throughout different stages of development will validate the notion that satellite cell function is modulated with age.

Our research serves to elaborate on the components and role of the satellite cell niche. A better understanding of the interplay between the satellite cell and its niche provides insight into the modulation of satellite cell quiescence, proliferation and differentiation.

## 5.4. Biomedical Implications

### *5.4.1 The Metastable State in Lineage Priming and Reprogramming*

Stem cells whether of adult, embryonic or induced origin, have great potential for therapeutic use. The characteristic of pluri- or multipotency coupled with the ability for self-renewal offer unique opportunities for cell therapy, disease modeling, toxicity testing, and developmental biology.

From a clinical perspective, the ability to derive a cell type that gives rise to a defined set of lineages and yet maintains the ability to return to pluripotency would be highly desirable. Recent research in mES cells suggests the pluripotent state is heterogeneous and composed of metastable subpopulations of cells that are pluripotent yet primed for epiblast or primitive endoderm differentiation[27-30,33]. Starting with a heterogeneous population of mES cells to derive specific cell lineages is not ideal. Having the ability to manipulate the metastable state of mES cells and subsequently select those cells prior to differentiation is advantageous.

Our research expands on the role of Wnt3a in maintaining pluripotency and directing differentiation. Understanding how Wnt3a regulates the interconversion of two cellular states will benefit research in the areas of directed differentiation and cellular reprogramming and ultimately lead to improvements in the field of regenerative medicine.

*5.4.2. Satellite Cells as a Cell Based Therapy to Treat Muscular Dystrophy*

Duchenne muscular dystrophy (DMD) is an X-linked genetic disorder that results in the loss of a functional dystrophin protein[16,17]. DMD is the most common form of muscular dystrophy and is characterized by progressive muscle wasting and weakness resulting in mortality in the second decade of life[377].

If cell therapy is to be used to treat muscular dystrophy, it will likely require the propagation of cells that maintain the ability to colonize the satellite cell niche. Primary myoblasts are not a suitable cell source for transplantation because they are drastically different from satellite cells regarding their gene expression profiles and *in vivo* regenerative capacity[50,96,328]. Satellite cells are the predominant cellular source for muscle regeneration and therefore, are a highly desirable cell type for use in treating DMD[333,334,378]. However, methods to propagate satellite cells while maintaining their regenerative capacity intact or procedures to rapidly influence satellite cell function without extended culture *in vitro* are lacking.

We demonstrated that transplantation of freshly sorted satellite cells into *mdx* mice resulted in robust colonization of the satellite cell niche and contribution to regenerating myofibers. More importantly we determined that the functional regenerative capacity of satellite cells diminishes with increased age.

We predict that treatment of freshly isolated adult satellite cells with secreted growth factors and extracellular glycoproteins is a feasible way to affect cellular function. In addition, the use of ECM or secreted proteins is non-invasive to the cell and is amenable to delivery *in vivo* and use *in vitro*. Large scale production of recombinant proteins and growth factors is feasible and can complement cell based therapies. The identification of a secreted protein that alters adult satellite cells to better mimic the functional capacity of fetal satellite cells is highly desirable as a therapeutic tool to treat muscular dystrophy.

## 5.5. Conclusions

Identification of the cellular mechanisms which govern the functional ability of a given cell type are tremendously important. In embryonic stem cells, activation of the canonical Wnt signalling pathway by Wnt3a was sufficient to alter the pluripotent state and prime cells for differentiation into the primitive endoderm lineage. Elucidating the mechanisms that regulate this process provided insight into the signalling pathways that affect lineage priming and pluripotency.

Correct transcriptional activation depends upon multiple processes not limited to the modulation of chromatin structure, methylation, and the availability of necessary transcription factors to bind at a given DNA sequence. Successful creation of transgenic mice, altered at a given genetic locus, can be affected by alterations in DNA tertiary structure, the frame of translation, exon/intron boundaries and unforeseen changes to the targeted DNA sequence. For these reasons caution along with the use of previously verified transgenic tools and methodology are strongly suggested. Future constructs used

to create Pax7 and Myf5 reporter lines will benefit from the knowledge outlined within this thesis.

In adult satellite cells, determining what factors or pathways affect their functional capacity to regenerate is of critical importance. We determined that age plays a significant role in this process and that cell intrinsic qualities dictate to what degree satellite cells will contribute to myogenic regeneration. Both of these projects provide insight into the functional abilities of stem cells and show potential for future therapeutic applications.

# Appendix A - Integration of expressed sequence tag data flanking predicted RNA secondary structures facilitates novel non-coding RNA discovery.

**Purpose**

To develop a means to incorporate expressed sequence tag information into a pipeline for predicting novel non-coding RNAs

**Submission Information**

Submitted: February 14, 2011

Accepted May 4, 2011

Published June 15, 2011

Paul M. Krzyzanowski, Feodor D. Price, Enrique M. Muro, Michael A. Rudnicki, Miguel A. Andrade-Navarro. Integration of Expressed Sequence Tag Data Flanking Predicted RNA Secondary Structures Facilitates Novel Non-Coding RNA Discovery. PLoS ONE, **2011** June;6(6):e20561.

**Contributions of Co-Authors**

Feodor Price conducted all wet lab experiments contained within the manuscript.

**Preface and Discussion of Figure 3**

We hypothesized that the combination of known cDNA-based Expressed Sequence Tag (EST) data with other computation methods to predict non-coding RNAs would enhance the prediction efficiency of microRNAs (miRNAs). EST termini were determined to be significantly associated with the 3' termini of miRNAs and therefore enrich the quality of miRNA predictions. In order to assess this method *in vitro*, murine skeletal muscle myoblasts were analyzed for their expression of non-coding RNA hairpins by a custom miRNA microarray. Known as well as novel predicted miRNAs were identified as being enriched in our miRNA microarray data sets. Validation by Northern blotting confirmed the presence of known and novel miRNAs in myoblast samples. Our results provide a novel method to enhance the efficiency of miRNA predication through incorporating EST data.

OPEN ACCESS Freely available online

PLoS one

# Integration of Expressed Sequence Tag Data Flanking Predicted RNA Secondary Structures Facilitates Novel Non-Coding RNA Discovery

**Paul M. Krzyzanowski[1,3]*, Feodor D. Price[1], Enrique M. Muro[2], Michael A. Rudnicki[1], Miguel A. Andrade-Navarro[2]**

**1** Sprott Center for Stem Cell Research, Ottawa Hospital Research Institute, Ottawa, Canada, **2** Max Delbrück Center for Molecular Medicine, Berlin, Germany, **3** Ontario Cancer Institute, Toronto, Canada

## Abstract

Many computational methods have been used to predict novel non-coding RNAs (ncRNAs), but none, to our knowledge, have explicitly investigated the impact of integrating existing cDNA-based Expressed Sequence Tag (EST) data that flank structural RNA predictions. To determine whether flanking EST data can assist in microRNA (miRNA) prediction, we identified genomic sites encoding putative miRNAs by combining functional RNA predictions with flanking ESTs data in a model consistent with miRNAs undergoing cleavage during maturation. In both human and mouse genomes, we observed that the inclusion of flanking ESTs adjacent to and not overlapping predicted miRNAs significantly improved the performance of various methods of miRNA prediction, including direct high-throughput sequencing of small RNA libraries. We analyzed the expression of hundreds of miRNAs predicted to be expressed during myogenic differentiation using a customized microarray and identified several known and predicted myogenic miRNA hairpins. Our results indicate that integrating ESTs flanking structural RNA predictions improves the quality of cleaved miRNA predictions and suggest that this strategy can be used to predict other non-coding RNAs undergoing cleavage during maturation.

**Citation:** Krzyzanowski PM, Price FD, Muro EM, Rudnicki MA, Andrade-Navarro MA (2011) Integration of Expressed Sequence Tag Data Flanking Predicted RNA Secondary Structures Facilitates Novel Non-Coding RNA Discovery. PLoS ONE 6(6): e20561. doi:10.1371/journal.pone.0020561

**Editor:** Matthew E. Hudson, University of Illinois, United States of America

**Received** February 14, 2011; **Accepted** May 4, 2011; **Published** June 15, 2011

**Copyright:** © 2011 Krzyzanowski et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The authors acknowledge support from the Ontario Government Scholarships in Science and Technology (PMK), the High Performance Computing Virtual Laboratory (PMK), the Canadian Institute of Health Research (FDP, MAR); the Canadian Stem Cell Network (PMK, EMM, MAR, MAA); the Canada Research Chair programme (MAA); the Helmholtz Germany's National Genome Research Network (Bundesministerium für Bildung und Forschung) (EMM, MAA); and The Helmholtz Alliance on Systems Biology (Helmholtz-Gemeinschaft Deutscher Forschungszentren) (EMM, MAA). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: paul.krzyzanowski@utoronto.ca

## Introduction

Over the past decade, microRNAs (miRNAs) have emerged to be important and powerful factors in most biological processes. miRNAs are small (~21 nt) species of non-coding RNA (ncRNA) molecules that regulate cognate mRNAs exhibiting sequence complementarity. miRNAs are first transcribed as long primary transcripts (pri-miRNAs) containing hairpin-like structures which are excised by DGCR8/Drosha dependent endonuclease cleavage to form pre-miRNAs [1,2,3], which are then exported into the cytoplasm. Pre-miRNAs exported to the cytoplasm are processed by the Dicer ribonuclease to form short dsRNA duplexes, which ultimately supply the RISC complex with mature miRNA strands [4,5,6]. Mature miRNAs are thought to coordinate rapid transcriptional transitions by repressing large numbers of mRNA transcripts in parallel [7]. Based on phylogenetic evidence, it is thought that miRNAs arose before vertebrate speciation and play roles to establish tissue identity [8,9,10]. During cell growth and differentiation, miRNAs have been shown to function during key transitions between cell states where they facilitate transcriptional reprogramming. Due to their ability to control numerous target transcripts simultaneously, miRNAs have emerged as regulators of gene expression in an increasing number of biological processes involving development and tissue regeneration (e.g. [11,12,13,14]).

Increases in abundance of three major muscle-specific miRNAs (mir-1, mir-133, and mir-206) have been observed during muscle cell differentiation in the mouse, *Drosophila*, and the zebrafish [15,16]. In mammals, mir-206 can be activated by MyoD expression [17] and remains active in terminally differentiated muscle fibers [18], yet is repressed in developing somites of chicken embryos in response to FGF4 signalling [19]. The induction of myogenic differentiation in C2C12 cells leads to mir-1 upregulation [20], which likely requires Myf5 based on observations of the developing chick myotome [21]. Interestingly, although Myf5 is one of the key regulators of early muscle development, it itself has been shown to be post-transcriptionally regulated by miRNAs in neurons [22], raising the possibility that it is under similar miRNA control in a myogenic context.

These miRNAs are necessary during myogenesis. Completely malformed musculature results when mir-1 is absent during *Drosophila* larvae development [23], and interference with mir-206 in mouse maintains C2C12 cells in a cycling state [24], implying that it is a key requirement for differentiation to proceed. In agreement to this, mir-206 is nearly absent in proliferating porcine satellite cells but is greatly induced during murine C2C12

differentiation [25,26], suggesting that miRNAs participate in muscle development.

In summary, there is abundant evidence indicating that miRNAs play an important role in specifying the myogenic lineage and it is likely that several undiscovered miRNAs participate in this process. To identify additional miRNAs expressed specifically during myogenic differentiation, we developed and tested a novel strategy of high-throughput tissue-specific miRNA prediction.

Methods employed to predict miRNAs have followed a general theme to identify genomic sites with the ability to produce simple, unbranched RNA hairpin structures and to identify putative RNA hairpins with properties making them distinct from spontaneously arising hairpins [27,28,29], with some methods considering conservation data [30,31] or sequence reads derived from small RNA sequencing [32,33]. Hybrid approaches to ncRNA prediction attempt to identify novel miRNAs residing within known regions of transcription.

One of the most abundant but underappreciated reservoirs of data to facilitate this work is Expressed Sequence Tag (EST) data previously generated to characterize the transcriptome with over 60 million EST sequences made publicly available in the NCBI dbEST [34]. Several groups have directly examined EST data for ncRNA prediction, attempting to identify novel miRNAs by examining ESTs for hairpin-like structures [35,36], with others monitoring sites of EST transcription to infer expression of nearby miRNAs [37]. The expression between intronic miRNAs and the protein coding transcripts they lie within is correlated [38], and in fact ESTs have illustrated alternative processing of coding transcripts containing intronic-miRNAs, as in the case of mir-126 within EGFL7 [39].

miRNA precursor molecules are polyadenylated [40,41,42,43,44], and can be successfully amplified using oligo-dT primers [45]. In fact, the usage of poly-T primers in EST library construction allows for observation of the 3′ ends of pri-miRNA cleavage products of miRNA biogenesis, with reverse transcription ending at the 5′ terminus of a miRNA hairpin (Figure 1A). As many examples of this effect are seen such as loci of murine mir-145, mir-126 and mir-206 (Figure 1B), we investigated the utility of adding EST data in a miRNA prediction pipeline.

To this end, we developed a methodology that combines techniques from ncRNA prediction with EST sequences. Our novel reinterpretation of preexisting public EST data identifies traces of miRNAs and provides expression predictions from tissue-specific EST libraries. This method constitutes a de facto novel contribution to techniques of predicting tissue-specific ncRNAs.

## Methods

### Source data

We used mm8, mm9 and hg18 releases of the mouse and human genomes. Four-way conservation data used in RNAz included human (hg18), rat (rn4), dog (canFam2) and mouse (mm8 or mm9, where appropriate). Where required, miRNA, EST annotations, and multiZ annotations were obtained from

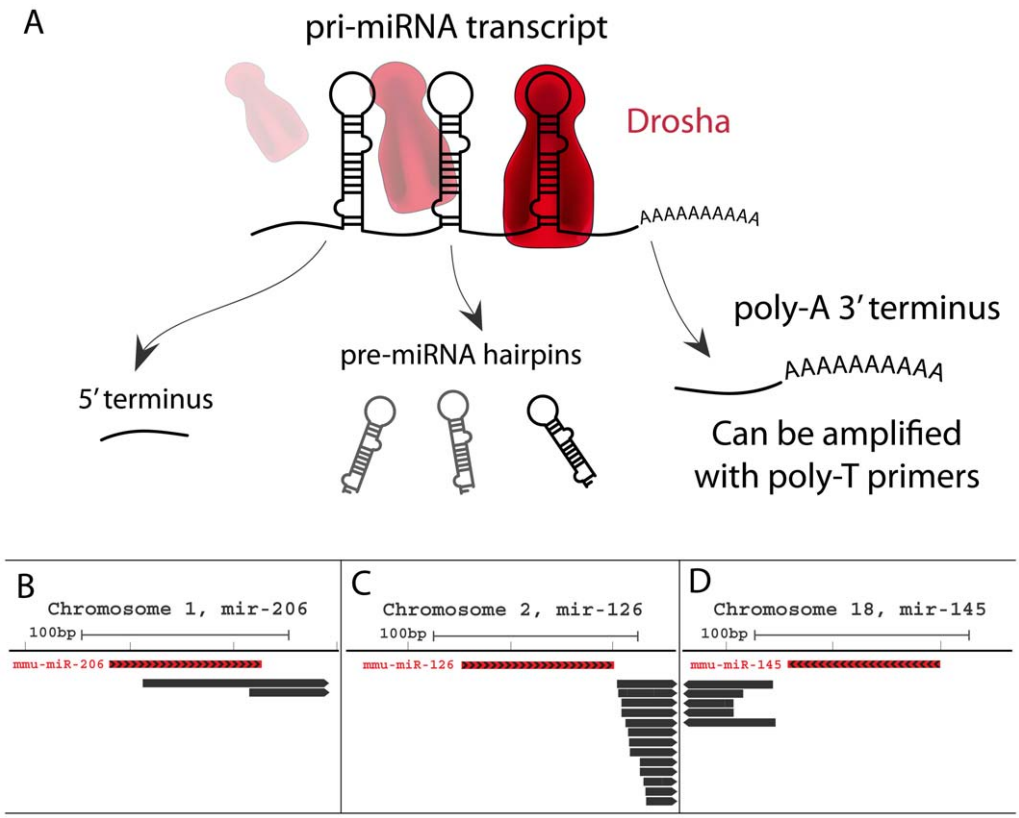



**Figure 1. Rationale for miRNA-EST model of prediction.** A: Model of miRNA biogenesis and generation of EST evidence of miRNA expression. RNA polymerase II transcribes the primary miRNA transcript (pri-miRNA), which is cleaved by the Drosha-DGCR8 complex. Cleaved pre-miRNAs are further processed, while intervening fragments are degraded. The remaining polyadenylated 3′ fragments of the pri-miRNA may be amplified by polyT primers commonly used in the generation of EST libraries. B, C, D: Examples of known miRNAs with ESTs in proximity to their 3′ terminus (Mouse genome, mm9).
doi:10.1371/journal.pone.0020561.g001

the appropriate releases of the UCSC genome annotation datasets [46].

## Generating ncRNA predictions

ncRNA predictions were generated using three methods: RNALfold [47], RNAz [48], and alignment of high-throughput sequencing fragments derived from a small RNA library provided in GEO record GSM314552 [33]. For details of the implementation of each of these methods, see the supporting information.

## Integrating EST data with known miRNAs

To identify and visualize relationships between miRNAs and ESTs, 3′ or 5′ coordinates of miRNA annotations on human and mouse genomes (hg18 and mm8, respectively) were identified and 2 kb sequence windows centered on this location were retrieved. To compute a normalized score of EST occupancy in this region, coordinates of EST termini mapping within each sequence window were identified and a score for each site was incremented by $1/n$, where $n$ equals the number of ESTs mapped to the miRNA associated sequence window. Scores for all miRNAs were aggregated, binned in 10 nt windows, and plotted. Significance scores for peaks observed in each genome assembly were determined by bootstrapping. Briefly, for each miRNA window, window coordinates were shuffled to random locations on the same chromosome and EST termini were identified in $\pm 1$ kb genomic windows. This procedure was replicated 10,000 times. EST accumulation was scored as described above, and the scores for each genome assembly and test (upstream/downstream of miRNA) were aggregated. The number of times a replicate generated a central peak in excess of the observations was counted and used to determine a P-value.

## Generating EST supported miRNA predictions

To screen candidate miRNA hairpins, we filtered hairpin structures based on their normalized $\Delta G$ value and the distance to a downstream EST terminus, as true miRNAs were observed to have lower than average normalized free energy (NFE) than the entire set of genomic hairpins computed with RNALfold. We thus examined the effect of filtering predicted structures based on precision/recall values computed using NFEs between 0 and $-1$ kcal/mol and EST proximities between 0 and 200 nt (See Figure S1). We retained predicted hairpin structures exhibiting a maximum NFE of $-0.44$ kcal/mol and with one or more EST termini within the structure, permitting the generation of mature miRNA and up to 14 nt downstream of the 3′ end. Predictions generated by RNAz were considered on either strand as RNAz did not preserve information regarding strandedness. For mapped Illumina sequencing fragments, which approximate mature miRNA sequences, not pre-miRNA hairpins, we determined the maximal distance between sequencing fragments and ESTs. We examined the distances between both 5′ and 3′ termini of known mature miRNAs and the ends of their corresponding pre-miRNAs (Figure S2). As we were not able to retrieve the strandedness of the mature miRNA sequences from the dataset due to its experimental design, we assumed that the mature miRNAs may reside in either arm. In the result, we were forced to use the greater of the two 'd' values which considers the possibility that the predicted mature miRNA resides in the 5′ arm of the pre-miRNA. The mean of the larger distances was 51 nt, which was used as a cutoff for considering sequence fragment/EST intersections.

## Statistics

Enrichment statistics for EST supported miRNA predictions were determined by shuffling candidate miRNA annotations to random locations within the same chromosome and strand. Candidate miRNAs were then predicted by intersecting EST data according to the methods described above.

Precision and recall of miRNA predictions were computed using as positives a set of known miRNAs or the subset of those supported by transcript data if such data were being used in prediction. MiRNA predictions were considered supported if they had EST annotations within 14 nt for those predicted computationally with RNALfold or RNAz, or within 51 nt for those predicted using high-throughput sequencing. For high-throughput sequencing data, a known miRNA was considered as transcript-supported if there was at least one Illumina sequencing fragment mapped within the known miRNA structure. For miRNA predictions generated by RNALfold and RNAz, a known miRNA was considered as supported if at least 75% of the nucleotides in the predicted hairpin and miRNA structure overlapped.

## miRNA predictions

13267 miRNA predictions were generated using RNA hairpin predictions in the mouse genome (mm8). Supporting dbEST annotations (retrieved September 9, 2007) were integrated with RNA structures as explained above to add source tissue predictions. A summary of miRNA predictions generated for mouse (mm9) is in Table S1. For immediate validation purposes, miRNAs with predicted expression in muscle (32), myotube (88), embryo (822), and embryonic stem cell (674) were selected.

## Design and analysis of a miRNA microarray

To detect expression of putative miRNAs, a custom oligonucleotide tiling microarray with 60mer probes (Agilent) was designed to tile over selected known and predicted miRNA hairpins with probes overlapping by 20 nt. Control probe sequences including positive and negative controls for known miRNAs, other ncRNAs, randomized negative control sequences, and sequences designed against known protein coding genes were designed (Table S2). In addition to the numbers of muscle, myotube, embryonic, and embryonic stem cell miRNA predictions (32, 88, 822, and 674 respectively) referenced in the previous paragraph, space permitted the addition of candidates with normalized hairpin free energies that met relaxed cutoff values. RNA preparation and microarray normalization were done according to the manufacturer's recommended protocols.

## miRNA target predictions

miRNA predictions were calculated using the TargetScan protocol [49]. The algorithm was replicated locally with custom scripts, which produce results highly correlated with published miRNA-target scores ($r>0.99$; Data not shown).

## Cell cultures

The mouse ES cell line J1 was maintained on a layer of irradiated MEFs (DR4) in DMEM high glucose medium (Invitrogen) supplemented with 15% fetal bovine serum (Hyclone), $1\times$ non essential amino acids, $1\times$ sodium pyruvate, $1\times$ Glutamax, β-Mercaptoethanol, $1\times$ Pen/Strep (Invitrogen), and 1000U ESGRO (chemicon) at 37°C and 5% $CO_2$. Prior to RNA isolation ES cells were trypsinized and preplated for 1 hr on 0.1% gelatin plates to remove the majority of MEFs. Mouse C2C12 cells and HEK293T cells were maintained in DMEM high glucose medium supplemented with 10% fetal bovine serum (Hyclone)

and 1× Pen/Strep at 37°C and 5% $CO_2$. For differentiation, C2C12 cells were grown to confluence and subsequently cultured in DMEM high glucose with 2% horse serum (Hyclone) for 3 days prior to RNA isolation. R1, C2C12, and HEK293T cells are available from ATCC as ATCC Numbers SCRC-1011, CRL-1772, and CRL-11268, respectively.

### Northern Blotting

To validate putative miRNAs by Northern blotting, total RNA was isolated using the mirVana kit according to the manufacturer's instructions (Ambion). 5 ug of total RNA was resolved with 12.5% urea-polyacrylamide gels and electroblotted onto Hybond NX membranes (GE). 60mer 5′ biotinilated oligonucleotides complementary to short hairpin RNA sequences were used as probes (Table S3). Membranes were developed using the light shift chemiluminescent kit with a streptavidin horse radish peroxidase secondary antibody according to the manufacturer's instructions (Pierce). Membranes were exposed to Biomax MR film (Kodak) for visualization.

## Results

### EST termini are significantly associated with 3′ termini of miRNAs

To determine whether ESTs could be used as a general data source in novel miRNA prediction, we examined whether a trend of association exists between the EST termini around known miRNAs in human and mouse. We observed a distinct clustering of EST termini in sequence regions aligned with the 3′ ends of known miRNA in the mouse and human genomes, respectively (Figure 2). By comparison, no clustering of EST termini was observed when examining the 5′ ends of known miRNAs indicating the validity of our rationale for interpreting EST data (Figure 1A).

One important issue in genome wide screens is ascertaining the degree of background signal present. To estimate the extent of this, the ncRNA-EST analysis was repeated numerous times using randomly shuffled miRNA positions. The results in Figure 2 show that 3′ termini of miRNAs are associated with ESTs above expected background levels. This result is independent of the genome assembly version (Figure S3). Together, these observations support our model of concurrent miRNA and EST generation presented in Figure 1. With this information, we inferred that miRNA predictions might be generated by identifying genomic regions potentially generating structured RNA with corresponding EST evidence in proximity of their 3′ terminus.

### EST data enrich the quality of microRNA predictions

To investigate whether improvments in miRNA predictions can be attained with this strategy, we considered whether potentially structured RNA regions demonstrating supporting EST evidence would benefit miRNA prediction pipelines, considering three different methods to identify miRNA candidates: simple hairpins generated by RNALfold [47], conserved RNA structures (RNAz) [48], and mapped high-throughput sequence fragments residing in hairpins.

RNALfold is a method of generating structured RNA predictions that identifies energetically stable RNA structures within subregions of very long sequences. After calibrating settings to discard predicted RNA structures unlikely to be miRNAs (See Methods and Figure S4 for details), we obtained approximately 3.3 million predicted hairpins in the mouse genome (mm9). In this set of RNA hairpins, known miRNAs were found to have lower than average Normalized Free Energy (NFE) (in agreement to [50]). This was reproduced through three different NFE computations, the most significant ($P<9\times10^{-80}$) when using the nucleotides contained within the duplexed region excluding those in the single stranded tails and terminal RNA loop (See Figure S5). We therefore refined this set of 3.3 million predicted hairpins by retaining those with both lower than average NFE (under −0.44 kcal/mol) and ESTs within 14 nt of the 3′ terminus (see Methods for determination of these values), yielding 41,347 candidates.

The association between ESTs and predicted structured RNA regions is significantly higher than expected if one were to assume a random distribution of genomic annotations ($P<0.0001$; Table 1). The rates of known miRNA prediction were examined with and without including known ESTs in the protocol for three different methods of predicting ncRNAs (Table 2). Using the set of RNALfold generated hairpin predictions, known miRNAs were predicted with a recall rate of almost 85% but a precision rate of only 0.01%. Upon incorporation of EST data, 81 out of 91 (89%) annotated miRNAs with ESTs within 14 nt were recalled by using this criteria while the precision rate increased to 0.20%, likely due to the loss of randomly arising hairpins not associated with EST evidence. Therefore, by considering known miRNA annotations recoverable with EST data, we effectively increased the recall rate by approximately 5% while simultaneously increasing the precision by 20 fold.

In a similar fashion, we generated predictions of structured RNAs exhibiting interspecies conservation between mouse, human, dog and rat using RNAz, a method that considers genomic alignments when generating candidate ncRNAs [48]. RNAz based ncRNA predictions were again highly significant (Table 1). However, the RNAz based results generated a higher Signal-to-Noise ratio than raw hairpins generated by RNALfold, presumably because interspecies conservation eliminated many potential false positives from the set of putative ncRNA hairpins. Applying EST data to RNAz predictions produced a modest increase in both precision by 22% (0.23% to 0.28%) and recall by 24% (31% to 38%) (Table 2). The lower recall rate observed with the RNAz implementation may be due to either some known miRNAs being located in genomic regions with lower conservation scores or to some miRNAs being species specific.

To investigate the utility of other types of transcript data for analysis of small RNA sequence libraries, we examined the influence on miRNA predictions of using published data derived from an embryonic stem cell library generated with Illumina sequencing technology [33]. Sequence fragments deposited in GEO (Accession: GSM314552) were aligned to the mouse genome, identifying 4239 small (~21 nt) sequence windows from which RNA expression was captured. After restricting these sequences to windows predicted to be within putative RNA hairpins, 3446 potential mature miRNA sequences with experimental support remained (see Methods for details). In a similar fashion to the other analyses discussed above, known miRNAs were intersected with EST windows, but a distance of 51 nt between sequencing fragments and ESTs was used to account for sequenced miRNAs corresponding to mature miRNAs, due to mature miRNA windows being nested within pre-miRNAs which are predicted computationally (Figure S2). Without EST data, the analysis yielded rates of recall and precision of 75.2% and 10.1%, respectively. When putative small RNA sequence windows were intersected with EST annotations to retain locations with both forms of experimental support, we observed the rates of recall and precision rise to 81.8% and 20.0% (Table 2).

Together, the results from the above three benchmarks strongly argue that improvements in the prediction rates of ncRNAs, and

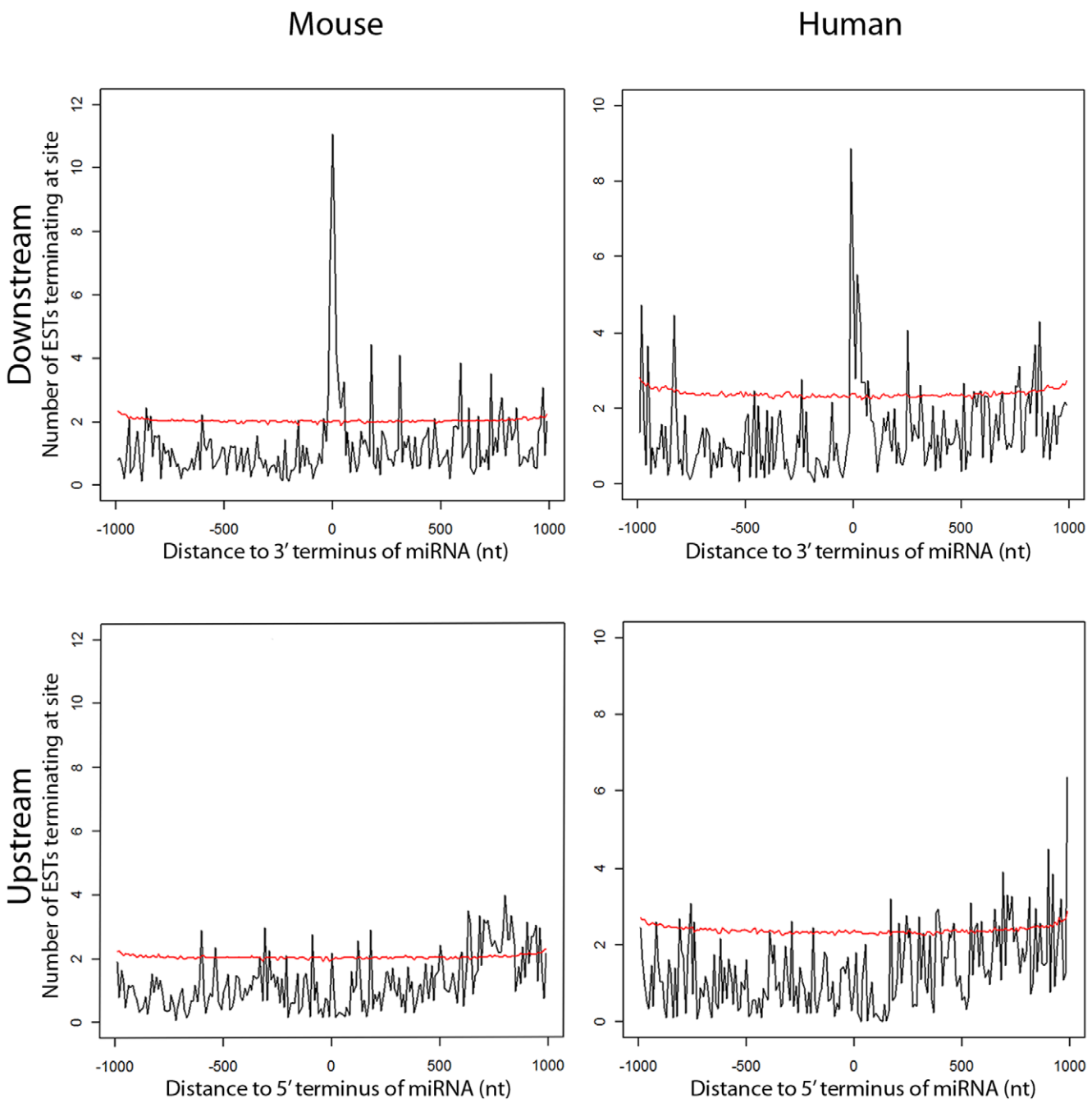



**Figure 2. EST termini associate with 3′ end of annotated miRNA in mouse and human genomes.** Y-axes measure the normalized density of observed EST termini within each genomic window surrounding known miRNAs. Top: Accumulation of EST termini adjacent to 3′ ends of known miRNAs in the mouse (mm8) and human (hg18) genomes. Bottom: This effect is not observed at the 5′ ends. The background signal obtained with randomly distributed ESTs and miRNAs is shown (red lines).
doi:10.1371/journal.pone.0020561.g002

specifically ncRNAs undergoing processing or cleavage, such as miRNAs, can be attained if future ncRNA prediction algorithms incorporate EST data into their protocols.

## Myoblast expressed miRNA-like hairpins are identified by a miRNA microarray

Beyond indicating the extent of transcription, records in the dbEST database provide valuable information describing the tissues, cell types, and conditions in which ESTs were observed. We reasoned that these data could be used to predict tissue-specific context in which our predicted miRNA could be expressed.

To demonstrate that this method could be used to produce tissue-specific results of biological significance we predicted novel myogenic miRNAs. Cross-referencing information from genomic sites containing predicted RNA hairpins and EST evidence derived from muscle or myoblast tissues, we identified 196 potential sites of miRNAs expressed in myogenic cell types. An overview of various numbers of predictions produced in muscle,

**Table 1.** ncRNA-EST associations are significantly higher than expected by chance.

| Genome | Object type | n objects | n ESTs | Intersecting @14 nt | Bootstrap mean | S/N Ratio | Bootstrap P-value |
|---|---|---|---|---|---|---|---|
| mm8 | RNALfold[1] | 3299070 | 4561569 | 41347 | 33182 | 1.24 | <10−4 |
| | RNAz[2] | 79094 | 4561569 | 17245 | 4176 | 4.13 | <10−4 |
| | Illumina[3] | 3442 | 4561569 | 112 | 25 | 4.48 | <10−4 |
| mm9 | RNALfold[1] | 3340947 | 4364783 | 40872 | 32496 | 1.26 | <10−4 |
| | RNAz[2] | 62057 | 4364783 | 12617 | 3039 | 4.15 | <10−4 |
| | Illumina[3] | 3446 | 4364783 | 88 | 18 | 4.89 | <10−4 |

[1]Hairpins predicted using RNALfold.
[2]Hairpins predicted using RNAz.
[3]Mature miRNA sequences predicted using Illumina high-throughput sequencing data.
doi:10.1371/journal.pone.0020561.t001

myoblast, or other tissues is shown in Table S1. We used a tiling microarray to examine the expression of each genomic site containing predicted miRNA hairpins with likely expression in differentiating myoblasts or mES cells.

The custom tiling microarray was designed to detect sequences for each putative miRNA hairpin site, along with multiple control sequences corresponding to known miRNAs. In total, 13840 probes tiling over putative miRNAs were designed. RNA from differentiating C2C12 myoblasts and J1 embryonic stem (ES) cells was isolated and hybridized to the microarrays in duplicate, for a total of eight hybridizations. Correlations between replicates are shown in Table S4.

After data normalization and analysis, 148 probe sets with >2 fold enrichment in muscle RNA versus ES derived RNA were identified, including muscle control probe sets designed against known muscle miRNAs mir-206, mir-133a/b (Figure S6), as well as the let-7 family miRNAs. In contrast, control miRNAs derived from the mir-291 cluster were strongly expressed in the RNA derived from ES cells, as expected. Taken together, the microarray data showed correct and robust reciprocal expression of major control miRNAs in each cell type, suggesting that numerous miRNA-like hairpins are differentially expressed in differentiating myoblasts.

## Known miRNAs expressed in differentiating myoblast cultures

Several known miRNAs that were captured by our prediction method were also detected as differentially expressed in myoblasts compared to ES cells. Several putative miRNA probe sets correspond to known miRNAs mir-24-2, mir-351, mir 26a-2, mir-92-2 and mir-7-2. According to these probes, miRNAs mir-24-2, mir-351, and mir-26a-2 were potentially upregulated in differentiating myoblasts, while mir-92-2 and mir-7-2 loci were downregulated (Table 3). Individual probes in each probe set exhibiting differential signal were manually verified to overlap with the sites of their corresponding mature miRNAs, with the exception of mir-7-2, suggesting that differential expression of mature miRNAs was observed. The changes in expression for probes at the mir-7-2 locus are likely due to non-specific hybridization.

To better understand the context by which these microarray signals could have arisen, the genomic contexts of these candidates were examined. Of highest relevance to the observed expression in myoblasts, the miRNA hairpin mapping to mir-26a-2 was found to be intronic to a long splice variant of *Ctdsp2* (carboxy-terminal domain, RNA polymerase II, polypeptide A, small phosphatase 2); it appears that a shorter *Ctdsp2* variant would be produced if this miRNA were excised from the longer transcript. *Ctdsp2* is able to induce mesodermal lineage commitment in a *Xenopus* system [51], while mir-26a-2 is known to be expressed during myogenesis, where it represses Enhancer of Zeste (Ezh2) transcript levels [52].

mir-351 is potentially part of a miRNA cluster with mir-503 and mir-322 lying within 2 kb upstream, with numerous ESTs mapped in this region supporting this possibility. However, no contiguous regions of expression appear to span the genomic region defined

**Table 2.** Considering EST annotations improves rates of miRNA prediction.

| | | mm9 | | | | | |
|---|---|---|---|---|---|---|---|
| | | Illumina[1] | | RNALfold[2] | | RNAz[3] | |
| | | w/o ESTs | w/ESTs | w/o ESTs | w/ESTs | w/o ESTs | w/ESTs |
| A | Raw sequencing objects | 3446 | 3446 | 3340947 | 3340947 | 62057 | 62057 |
| B | Sequencing objects considered | 3446 | 540 | 3340947 | 40782 | 62057 | 12617 |
| C | Unique miRNAs detected | 348 | 108 | 393 | 81 | 144 | 35 |
| D | Total miRNA predictable | 463 | 132 | 463 | 91 | 463 | 91 |
| | Precision (C/B) | 10.10% | 20.00% | 0.01% | 0.20% | 0.23% | 0.28% |
| | Recall (C/D) | 75.16% | 81.82% | 84.88% | 89.01% | 31.10% | 38.46% |

[1]Mature miRNA sequences predicted using Illumina high-throughput sequencing data.
[2]Hairpins predicted using RNALfold.
[3]Hairpins predicted using RNAz.
doi:10.1371/journal.pone.0020561.t002

**Table 3.** Hairpins with miRNA overlap exhibiting differential expression on microarray.

| Hairpin ID | Fold Increase | | miRNA | Location |
|---|---|---|---|---|
| | C2C12 | ES | | |
| hp1643988 | 141.56 | - | mir-24-2 | chr8:87098910-87099035 |
| hp3202291 | 108.27 | - | mir-351 | chrX:49297888-49297981 |
| hp907540 | 94.12 | - | mir-26a-2 | chr10:126398456-126398589 |
| hp1980840 | - | 1.52 | mir-92-2 | chrX:48986474-48986552 |
| hp85655 | - | 1.72 | mir-7-2 | chr7:78761795-78761882 |

doi:10.1371/journal.pone.0020561.t003

by these three miRNAs. mir-24-2 is the downstream most member of miRNA cluster containing mir-27a and mir-23a. This cluster is downstream from *Zswim4*, an uncharacterized zinc finger containing protein.

## Prediction of targets for known miRNAs

To identify potential targets for the miRNAs expressed at higher levels in differentiating myoblasts, the TargetScan method of identifying miRNA targets was used. Potential target transcripts, which mature sequences for mir-24-2, mir-26a, and mir-351 might be repressing, were predicted and rather than considering targets individually, scores were combined multiplicatively to identify potential transcripts that might be synergistically repressed by all three (Table 4), a method that penalizes targets where one or more scores are abnormally low. Immediately striking was the high rank for potential targeting of the Dystrophin (*Dmd*) transcript.

Dystrophin is transcribed from a large genomic region (~3 megabases), contains 79 exons, and undergoes numerous alternative splicing patterns [53]. As the second best target, Dystrophin was targeted most robustly by mir-351 (TargetScan score = −0.45). Upon closer inspection, the Dystrophin 3′UTR targeted by all three miRNAs was a longer Dmd transcript (ENSMUST00000114000) that terminates approximately 1 kb beyond the main transcription stop site, and which has a single EST as evidence of transcription (CK385805) in the current version of dbEST. Despite the low number of ESTs associated with this Ensembl transcript, it is not possible to conclude that the *in vivo* expression of this longer Dmd transcript is proportionally low. In contrast, Dystrophin's shorter 3′UTR is targeted only by mature mir-26a (Data not shown). As different Dystrophin splice variants are expressed in different tissues, with the 3′ region of the transcript undergoing alternative splicing to produce tissue-specific transcripts in brain, cardiac and muscle fibers [53], it is possible that miRNAs may be involved to enforce transcription of some Dmd isoforms over others in certain contexts.

## Novel miRNA-like hairpins are expressed in differentiating myoblasts

Several putative ncRNA hairpins with strong expression scores were selected to determine whether the expression and molecular size of the RNAs were consistent with miRNA presence. Northern blotting was used to visualize the size of RNA species responsible for microarray signals, using the respective microarray probe sequences. Several of the selected probe sets indeed detected expression of small RNAs in the ~20–30 nt size range, suggesting that mature miRNAs were being generated (Figure 3A). Microarray data detected three predicted ncRNA hairpins at higher levels in myoblast — hp566131, hp1621227, and hp30491555 — and Northern blotting supplied information regarding their size and also confirmed their differential expression (Figure 3B). Identifiers with hp- prefixes denote RNA hairpins predicted in the course of this analysis. These results indicate that the microarray signals were due to small RNA species, likely derived from predicted miRNA hairpins, which are expressed in myogenic cells undergoing differentiation at higher levels than in embryonic stem cells.

# Discussion

We showed that addition of EST data in ncRNA prediction protocols can significantly improve prediction rates, using miRNAs as a representative class of ncRNAs. In doing this, we show that taking into account predicted behaviours of one type of transcript, inferences can be made as to what might be observed in data produced for entirely disparate purposes. In this study, we combined the features of EST generation (priming poly-A transcript ends) with knowledge of the RNA cleavage activity during miRNA biogenesis, to develop an approach that uses EST records and information about predicted RNA structures to predict miRNAs as a class of hairpin derived ncRNAs. Our results

**Table 4.** Top 10 predicted targets for synergistic repression by mir-24, mir-26a, and mir-351.

| | | Targetscan Scores | | |
|---|---|---|---|---|
| Symbol | Description | miR-24 | miR-26a | miR-351 |
| Plod2 | Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2 Precursor | −0.3234 | −0.7219 | −0.2086 |
| Dmd | Dystrophin | −0.352 | −0.2867 | −0.4483 |
| Tbc1d30 | TBC1 domain family member 30 | −0.4755 | −0.5532 | −0.1713 |
| Ret | Proto-oncogene tyrosine-protein kinase receptor ret Precursor (C-ret) | −0.3827 | −0.5018 | −0.1895 |
| 4732454E20Rik | Protein FAM26D | −0.7775 | −0.3364 | −0.1225 |
| Irf4 | Interferon regulatory factor 4 | −0.1166 | −0.516 | −0.5252 |
| Papola | Poly(A) polymerase alpha | −0.5424 | −0.229 | −0.2491 |
| Zfp697 | Zinc finger protein 697 | −0.492 | −0.3602 | −0.1722 |
| Slc25a35 | Solute carrier family 25 member 35 | −0.3222 | −0.2178 | −0.4084 |
| St8sia4 | CMP-N-acetylneuraminate-poly-alpha-2,8-sialyltransferase | −0.3275 | −0.335 | −0.2319 |

doi:10.1371/journal.pone.0020561.t004

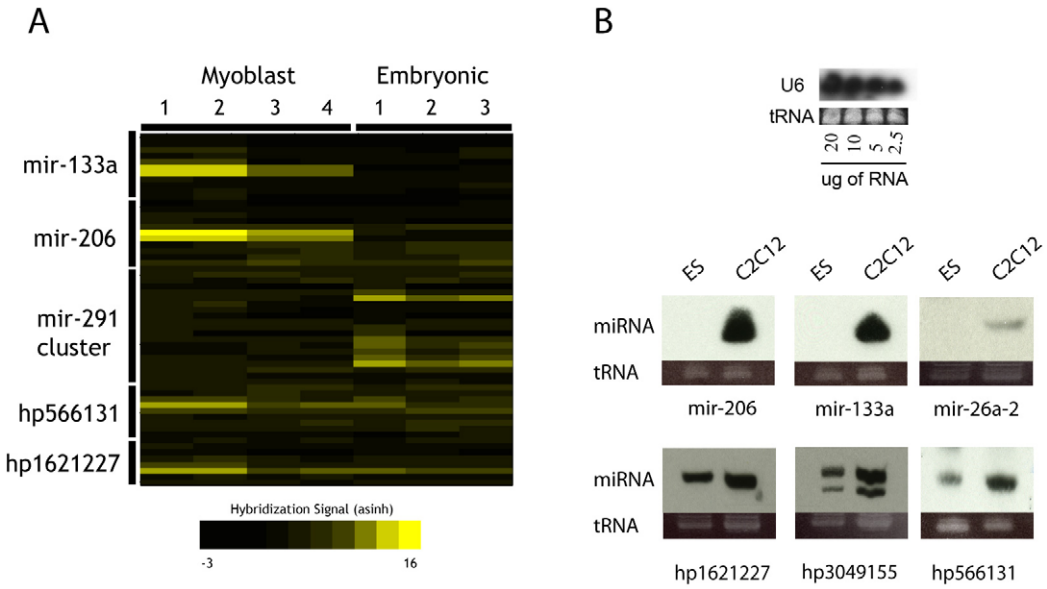



**Figure 3. Detection of known and predicted miRNAs.** A: Detection of control miRNAs via microarray in C2C12 and mES RNA samples. Numbers for each sample type indicate replicates. B: Northern blotting of several known and predicted miRNAs exhibiting enrichment in differentiating C2C12 myoblasts.
doi:10.1371/journal.pone.0020561.g003

show that future development of non-coding gene prediction strategies should combine secondary sources of biological data with primary sources of information such as sequence composition, motifs, and predicted RNA secondary structures when knowledge of RNA transcript behaviour facilitiates the development of an integrative model. We emphasize that our method is not an alternative miRNA prediction method but rather contributes an analytical step that can improve the quality of existing miRNA prediction strategies. For this reason, we explicitly showed that integrating non-overlapping EST and miRNA annotations enable improving miRNA prediction methods.

We furthermore expect that including EST data in ncRNA prediction strategies is a generalizable approach to other models of ncRNAs undergoing cleavage during their production or function. We can illustrate this latter point with the Clec2d 3′UTR, which has been reported to contain a self-cleaving hammerhead ribozyme [54]: upon manual inspection, we observed that the ribozyme cleavage site lies in close proximity to a cluster of ESTs (Data not shown), suggesting that novel ribozymes might be predicted by considering co-occurrence of ESTs and covariation models for specific RNA families.

It must be noted that adding EST evidence, as with the addition of any data, imposes a constraint on the analysis such that some miRNA predictions are called as false negatives. This trade-off must be considered. However, the significant increases of precision and recall for miRNA predictions that pass the integration step suggest that follow up experiments using these first tier candidates can experience lower false discovery rates. Nonetheless, those interested can still investigate second tier candidates but now with the additional knowledge that false discovery rates in these predictions will be lower due to the lack of EST evidence to support them.

On a practical level, we also illustrated how pre-existing data can produce compelling improvements in the performance of current high-throughput sequencing technologies. The increased precision and recall rates demonstrated in Table 2 illustrate how rates of ncRNA validation can be doubled when considering the results of the small RNA high-throughput sequence data. The combination of these data can justify small scale validation trials when genome wide analysis is not being pursued, as the doubling of precision with small RNA sequencing data to support miRNA predictions imply that manual validation attempts can be halved for similar outcomes. This combination is also one where two separate experimentally generated sets of data supported computational predictions and yielded the most dramatic improvements in precision rates, clearly highlighting the benefit of working within a single model (miRNA biogenesis) and saturating it with as much biological data as available. The improved results from this example also hint at potential synergistic effects when using independent sources of high-throughput data (i.e. compensating for technique-specific biases), which may become addressable on a broad scale as mutually compatible data sets accumulate.

A second practical contribution of ESTs towards ncRNA prediction is in supplying probable tissues of expression in which validation of ncRNAs can be pursued, for example when a specific tissue is of interest. Integrating data this way can also simplify experimental designs by minimizing sample size requirements, limiting resources consumed in pursuing specific questions.

Next-generation sequencing is emerging as a successor to EST library sequencing with RNA-seq becoming commonly available. However, as far as we know, no database of next-generation sequence transcript data exists that is comparable to dbEST. The wide distribution of conditions sampled in dbEST and the standardized format of annotations facilitate the discovery of transcripts that possibly occur in very particular conditions such as the ones we have presented here. The analysis of deep-sequencing data in an EST-like fashion still involves curation of primary sequence reads, assembly of short reads into contigs, and finally alignment to genomic positions, each step potentially requiring manual tuning. Should a database of deep-sequencing transcript data similar to dbEST be generated in the future, we have no doubt that our method will be successfully applied to it.

## On known miRNAs expressed in myoblast samples

Using a custom miRNA microarray allowed us to verify the performance of our method by capturing known miRNAs expressed in myoblast samples. We observed that three known miRNAs, mir24-2, mir-351, and mir-26a-2 were upregulated in myoblasts and showed robust expression. Expressed at 141 times the level observed in the embryonic stem cell sample, mir-24-2 was the most strongly overexpressed mature miRNA, and has previously been captured during myogenic differentiation [55]. mir-24-2 lies in a cluster alongside miRNAs mir-23a and mir-27a upstream of a gene encoding ZSWIM4, an uncharacterized zinc-finger containing protein. This miRNA cluster is activated in other cellular differentiation related contexts with similar effects, such as during the formation of multiple blood cell lineages during hematopoiesis [56] and during hepatic stem cell differentiation [57], being potentially associated with regulation of cell fate choice. While superficially it seems that the mir-24 containing cluster is a generalized differentiation controlling locus, mir-24 has been observed to be specifically expressed in porcine satellite cells in a reciprocal fashion to mir-206 [25]. The placement of mir-24 in a rare myogenic population may hint at a role in regulating stem cell entry into differentiation, possibly by the control of timing or proportions of cells entering the process. Similarly, mir-26a-2 has previously been demonstrated to function in repressing the polycomb group gene Enhancer of Zeste (*Ezh2*) during myogenesis [52]. Ezh2 is known to be involved in myotome development, where it functions to inhibit progress through differentiation [58]. Remarkably, *Ezh2* is also active in an analogous system – epidermal differentiation – where it has recently been shown to contribute towards control of differentiation rates throughout multistage epidermal shedding [59]. Thus, the mir-26a-2/Ezh2 system may function as another control mechanism to balance cell differentiation rates and speed.

Finally, mir-351 was upregulated in myoblasts at levels 108-fold over ES. In the mouse, mir-351 lies in a region of EST activity which includes mir-503, mir-322, and an uncharacterized transcript, which appears to undergo splicing (dbEST identifier AK021262). Taken together, these elements may represent a multipart transcript, and interestingly four additional miRNA coding sites (mir-542, mir-450a-2, mir-450a-1, and mir-450b) reside nearby, which do not have any evidence of being transcribed based on known dbEST sequence records. The function of mir-351 and this region is unknown, but mir-351 has been observed in SAGE tags derived from embryonic heart [60]. With this data and given the density of miRNA features and transcriptional activity, it may be worthwhile to investigate this region for potential involvement in muscle tissues.

We were interested in identifying predicted co-repression of target transcripts by all three mature miRNAs observed in C2C12s (mir24-2, mir-351, and mir-26a-2), reasoning that they might act in concert to exert effects on transcripts comparable to much more highly expressed miRNAs that are solitary actors such as mir-206. Considering combined miRNA repression scores, the best target identified was *Plod2*, encoding a collagen cross-linking enzyme with two isoforms exhibiting tissue specific expression in adult, with strongest protein expression observed in muscle tissues [61,62]. *Plod2* is an attractive target as the integrity of muscle tissue is strongly dependent on extracellular matrix components, and particular collagen subtypes are associated with myoblasts in different stages of proliferation [63]. What is especially enigmatic of *Plod2* is that both isoforms are transcribed in muscle yet only the longer isoform undergoes translation [62]. If this is a case of the shorter isoform being specifically repressed by miRNAs, it would be in contrast to the model of gene dysregulation by miRNA binding site loss via 3′UTR truncation, which has been used to explain dysregulation of oncogenic mRNAs [64]. Given the expression of *Plod2* isoforms in muscle and their differential translation, combined with potential targeting by three known miRNAs with expression in muscle, exploring miRNA mediated control of *Plod2* expression may be promising.

The second best target identified was Dystrophin (*Dmd*), encoding a myotubular structural protein critically linked to the manifestation of Duchenne Muscular Dystrophy [65]. Dystrophin is not expressed during early differentiation of myoblasts until fusion into myotubes has begun [66]. Closer inspection of the Dmd 3′UTR regions identified two transcriptional termination sites, with the variant targeted by all three miRNAs being longer than the major termination region covered robustly by known ESTs. If this interaction is true, it may hint at a case where active miRNAs are specifically responsible in eliminating Dmd transcripts created from abnormal termination events, a role that has been reported in cases where alternative isoforms are specifically targeted to control their abundance levels [67,68,69]. In all, our results concerning mir24-2, mir-351, and mir-26a-2 upregulation in myoblasts represent added evidence for miRNA based regulation of myogenic genes and suggest that deeper involvement of ncRNAs in myogenesis will be revealed.

## General conclusion

We demonstrate that considering EST data in ncRNA prediction algorithms improves results when predicting tissue-specific miRNAs, and propose that EST data be included in future ncRNA prediction pipelines, especially of those that are polyadenylated and cleaved. In addition, we identified several known miRNAs with expression in differentiating myoblasts, representing potential participants of differentiation during this process. Understanding the role of these miRNAs and ncRNAs may provide insight into how endogenous capacities for muscle regeneration are determined.

# Supporting Information

**Figure S1 Calibration of filter based on Normalized Free Energy and miRNA-EST distances.** Cutoffs for Normalized Free Energy (NFE) and miRNA-EST distance were calibrated to filter the set of hairpins generated from the genomewide scan of RNALfold. Precision/Recall curves shown for each miRNA-EST distance ($d$) cutoff, with individual points corresponding to NFE cutoffs at each value of $d$ (Points for $d = 14$ indicated in black with NFE values decreasing from 0 to $-1$ counterclockwise). Circled point corresponds to $d = 14$ and NFE = $-0.44$ kcal/mol. NFE values range from 0 to $-1$ kcal/mol and miRNA to EST distances range from 0 nt to 200 nt.
(TIF)

**Figure S2 Determination of mature miRNA placement within pre-miRNA windows.** (A) 553 mature miRNA sequences were mapped within their parent pre-miRNA hairpin annotations and distances between termini were measured. $d_1$ and $d_2$ correspond to the shorter and longer distances, respectively. (B) Distributions of the values. The mean of the $d_1$ and $d_2$ are shown (12.7 nt and 51.3 nt, respectively).
(TIF)

**Figure S3 Distributions of EST peaks near miRNAs are similar across genome versions.** Distributions of miRNA-ESTs are highly similar across different versions of the mouse

genome (mm8 and mm9). EST occupancy surrounding miRNAs was calculated as described in Methods.
(TIF)

**Figure S4 Calibration of real-time RNALfold filter parameters.** Precision and recall values were calculated by applying RNALfold on genomic DNA sequences centered on known miRNAs and filtering resultant RNA structures. Filtering parameters used were: hairpin size (Degree of blue shading); hairpin NFE (Degree of red shading); and RNALfold window size (Proportional to point size). Parameters used in this study and corresponding precision and recall are shown with a white X. For computing precision/recall, miRNA-hairpin annotations had a minimum overlap of 80%.
(TIF)

**Figure S5 Distributions of Normalized Free Energies in sets of miRNAs and set of genome-wide predictions.** Three distributions are shown: The red, black, and blue distributions correspond to values derived by normalizing the total minimum free energy (MFE) by the number of nucleotides in the hairpin sections as highlighted in the inset figure.
(TIF)

**Figure S6 Control miRNAs are expressed correctly on microarray.** Heatmap showing correct expression of control miRNA probe sets for muscle (mir-133a and mir-206) and ES cell control (mir-293 containing cluster). Probe sets were composed of multiple probes tiling over each pre-miRNA hairpin sequence, and only probes overlapping mature miRNA sequences yielded expression signals.
(TIF)

**Table S1 Overview of predicted miRNA hairpins in different tissues.** 13952 unique putative miRNA hairpins were annotated in the mouse genome (mm9), based on predictions generated from RNA structures with supporting EST annotations. Counts above do not correspond to this value due to some miRNA predictions having ESTs derived from multiple sources, and sources with under 10 predictions being excluded due to space limitations.
(XLS)

**Table S2 Table of RNA control sequences included on the miRNA microarray.** Individual miRNAs with specific roles included where possible. In cases of clustered miRNAs where one member of a cluster was identified with a specific function (e.g. ES miRNA), additional members of the cluster were tiled as well.
(XLS)

**Table S3 Probe sequences used in Northern blotting.**
(XLS)

**Table S4 Pearson correlation between replicates of miRNA tiling microarray.** All samples were performed in duplicate.
(XLS)

## Acknowledgments

Special thanks go to Gareth Palidwor and Chris Porter (Ottawa Hospital Research Institute) for feedback during implementation of algorithms and for review of the manuscript.

## Author Contributions

Conceived and designed the experiments: PMK MAA. Performed the experiments: PMK FDP. Analyzed the data: PMK MAA. Contributed reagents/materials/analysis tools: EMM MAR. Wrote the paper: PMK MAA.

## References

1. Yeom KH, Lee Y, Han J, Suh MR, Kim VN (2006) Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. Nucleic Acids Res 34: 4622–4629.
2. Han J, Lee Y, Yeom KH, Nam JW, Heo I, et al. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell 125: 887–901.
3. Wang Y, Medvid R, Melton C, Jaenisch R, Blelloch R (2007) DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. Nat Genet 39: 380–385.
4. Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U (2004) Nuclear export of microRNA precursors. Science 303: 95–98.
5. Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, et al. (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science 293: 834–838.
6. O'Toole AS, Miller S, Haines N, Zink MC, Serra MJ (2006) Comprehensive thermodynamic analysis of 3′ double-nucleotide overhangs neighboring Watson-Crick terminal base pairs. Nucleic Acids Res 34: 3338–3344.
7. Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, et al. (2005) Combinatorial microRNA target predictions. Nat Genet 37: 495–500.
8. Tanzer A, Stadler PF (2004) Molecular evolution of a microRNA cluster. J Mol Biol 339: 327–335.
9. Christodoulou F, Raible F, Tomer R, Simakov O, Trachana K, et al. (2010) Ancient animal microRNAs and the evolution of tissue identity. Nature 463: 1084–1088.
10. Molnar A, Schwach F, Studholme DJ, Thuenemann EC, Baulcombe DC (2007) miRNAs control gene expression in the single-cell alga Chlamydomonas reinhardtii. Nature 447: 1126–1129.
11. Conaco C, Otto S, Han JJ, Mandel G (2006) Reciprocal actions of REST and a microRNA promote neuronal identity. Proc Natl Acad Sci U S A 103: 2422–2427.
12. Chen JF, Mandel EM, Thomson JM, Wu Q, Callis TE, et al. (2006) The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. Nat Genet 38: 228–233.
13. Georgantas RW, 3rd, Hildreth R, Morisot S, Alder J, Liu CG, et al. (2007) CD34+ hematopoietic stem-progenitor cell microRNA expression and function: a circuit diagram of differentiation control. Proc Natl Acad Sci U S A 104: 2750–2755.
14. Xu C, Lu Y, Pan Z, Chu W, Luo X, et al. (2007) The muscle-specific microRNAs miR-1 and miR-133 produce opposing effects on apoptosis by targeting HSP60, HSP70 and caspase-9 in cardiomyocytes. J Cell Sci 120: 3045–3052.
15. Brennecke J, Stark A, Cohen SM (2005) Not miR-ly muscular: microRNAs and muscle development. Genes Dev 19: 2261–2264.
16. Nguyen HT, Frasch M (2006) MicroRNAs in muscle differentiation: lessons from Drosophila and beyond. Curr Opin Genet Dev 16: 533–539.
17. Rosenberg MI, Georges SA, Asawachaicharn A, Analau E, Tapscott SJ (2006) MyoD inhibits Fstl1 and Utrn expression by inducing transcription of miR-206. J Cell Biol 175: 77–85.
18. Anderson C, Catoe H, Werner R (2006) MIR-206 regulates connexin43 expression during skeletal muscle development. Nucleic Acids Res 34: 5863–5871.
19. Sweetman D, Rathjen T, Jefferson M, Wheeler G, Smith TG, et al. (2006) FGF-4 signaling is involved in mir-206 expression in developing somites of chicken embryos. Dev Dyn 235: 2185–2191.
20. Nakajima N, Takahashi T, Kitamura R, Isodono K, Asada S, et al. (2006) MicroRNA-1 facilitates skeletal myogenic differentiation without affecting osteoblastic and adipogenic differentiation. Biochem Biophys Res Commun 350: 1006–1012.
21. Sweetman D, Goljanek K, Rathjen T, Oustanina S, Braun T, et al. (2008) Specific requirements of MRFs for the expression of muscle specific microRNAs, miR-1, miR-206 and miR-133. Dev Biol 321: 491–499.
22. Daubas P, Crist CG, Bajard L, Relaix F, Pecnard E, et al. (2009) The regulatory mechanisms that underlie inappropriate transcription of the myogenic determination gene Myf5 in the central nervous system. Dev Biol 327: 71–82.
23. Sokol NS, Ambros V (2005) Mesodermally expressed Drosophila microRNA-1 is regulated by Twist and is required in muscles during larval growth. Genes Dev 19: 2343–2354.
24. Kim HK, Lee YS, Sivaprasad U, Malhotra A, Dutta A (2006) Muscle-specific microRNA miR-206 promotes muscle differentiation. J Cell Biol 174: 677–687.
25. McDaneld TG, Smith TP, Doumit ME, Miles JR, Coutinho LL, et al. (2009) MicroRNA transcriptome profiles during swine skeletal muscle development. BMC Genomics 10: 77.
26. Rao PK, Kumar RM, Farkhondeh M, Baskerville S, Lodish HF (2006) Myogenic factors that regulate expression of muscle-specific microRNAs. Proc Natl Acad Sci U S A 103: 8721–8726.

27. Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, et al. (2003) The microRNAs of Caenorhabditis elegans. Genes Dev 17: 991–1008.
28. Sewer A, Paul N, Landgraf P, Aravin A, Pfeffer S, et al. (2005) Identification of clustered microRNAs using an ab initio prediction method. BMC Bioinformatics 6: 267.
29. Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, et al. (2005) Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet 37: 766–770.
30. Washietl S, Hofacker IL, Lukasser M, Huttenhofer A, Stadler PF (2005) Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. Nat Biotechnol 23: 1383–1390.
31. Washietl S, Pedersen JS, Korbel JO, Stocsits C, Gruber AR, et al. (2007) Structured RNAs in the ENCODE selected regions of the human genome. Genome Res 17: 852–864.
32. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, et al. (2008) Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res 18: 610–621.
33. Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R (2008) Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev 22: 2773–2785.
34. Boguski MS, Lowe TM, Tolstoshev CM (1993) dbEST–database for "expressed sequence tags". Nat Genet 4: 332–333.
35. Zhang BH, Pan XP, Wang QL, Cobb GP, Anderson TA (2005) Identification and characterization of new plant microRNAs using EST analysis. Cell Res 15: 336–360.
36. Li SC, Pan CY, Lin WC (2006) Bioinformatic discovery of microRNA precursors from human ESTs and introns. BMC Genomics 7: 164.
37. Gu J, He T, Pei Y, Li F, Wang X, et al. (2006) Primary transcripts and expressions of mammal intergenic microRNAs detected by mapping ESTs to their flanking sequences. Mamm Genome 17: 1033–1041.
38. Baskerville S, Bartel DP (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. Rna 11: 241–247.
39. Kim YK, Kim VN (2007) Processing of intronic microRNAs. Embo J 26: 775–783.
40. Pawlicki JM, Steitz JA (2009) Subnuclear compartmentalization of transiently expressed polyadenylated pri-microRNAs: processing at transcription sites or accumulation in SC35 foci. Cell Cycle 8: 345–356.
41. Eis PS, Tam W, Sun L, Chadburn A, Li Z, et al. (2005) Accumulation of miR-155 and BIC RNA in human B cell lymphomas. Proc Natl Acad Sci U S A 102: 3627–3632.
42. Song L, Han MH, Lesicka J, Fedoroff N (2007) Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. Proc Natl Acad Sci U S A 104: 5437–5442.
43. Xiong H, Qian J, He T, Li F (2009) Independent transcription of miR-281 in the intron of ODA in Drosophila melanogaster. Biochem Biophys Res Commun 378: 883–889.
44. Houbaviy HB, Dennis L, Jaenisch R, Sharp PA (2005) Characterization of a highly variable eutherian microRNA gene. Rna 11: 1245–1257.
45. Cai X, Hagedorn CH, Cullen BR (2004) Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. Rna 10: 1957–1966.
46. Kuhn RM, Karolchik D, Zweig AS, Wang T, Smith KE, et al. (2009) The UCSC Genome Browser Database: update 2009. Nucleic Acids Res 37: D755–761.
47. Hofacker IL, Priwitzer B, Stadler PF (2004) Prediction of locally stable RNA secondary structures for genome-wide surveys. Bioinformatics 20: 186–190.
48. Washietl S, Hofacker IL, Stadler PF (2005) Fast and reliable prediction of noncoding RNAs. Proc Natl Acad Sci U S A 102: 2454–2459.
49. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, et al. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell 27: 91–105.
50. Bonnet E, Wuyts J, Rouze P, Van de Peer Y (2004) Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. Bioinformatics 20: 2911–2917.
51. Zohn IE, Brivanlou AH (2001) Expression cloning of Xenopus Os4, an evolutionarily conserved gene, which induces mesoderm and dorsal axis. Dev Biol 239: 118–131.
52. Wong CF, Tellam RL (2008) MicroRNA-26a targets the histone methyltransferase Enhancer of Zeste homolog 2 during myogenesis. J Biol Chem 283: 9836–9843.
53. Sironi M, Cagliani R, Pozzoli U, Bardoni A, Comi GP, et al. (2002) The dystrophin gene is alternatively spliced throughout its coding sequence. FEBS Lett 517: 163–166.
54. Martick M, Horan LH, Noller HF, Scott WG (2008) A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA. Nature 454: 899–902.
55. Sun Q, Zhang Y, Yang G, Chen X, Zhang Y, et al. (2008) Transforming growth factor-beta-regulated miR-24 promotes skeletal muscle differentiation. Nucleic Acids Res 36: 2690–2699.
56. Lal A, Pan Y, Navarro F, Dykxhoorn DM, Moreau L, et al. (2009) miR-24-mediated downregulation of H2AX suppresses DNA repair in terminally differentiated blood cells. Nat Struct Mol Biol 16: 492–498.
57. Rogler CE, Levoci L, Ader T, Massimi A, Tchaikovskaya T, et al. (2009) MicroRNA-23b cluster microRNAs regulate transforming growth factor-beta/bone morphogenetic protein signaling and liver stem cell differentiation by targeting Smads. Hepatology 50: 575–584.
58. Caretti G, Di Padova M, Micales B, Lyons GE, Sartorelli V (2004) The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. Genes Dev 18: 2627–2638.
59. Ezhkova E, Pasolli HA, Parker JS, Stokes N, Su IH, et al. (2009) Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. Cell 136: 1122–1135.
60. Ge X, Wu Q, Wang SM (2006) SAGE detects microRNA precursors. BMC Genomics 7: 285.
61. Pornprasertsuk S, Duarte WR, Mochida Y, Yamauchi M (2004) Lysyl hydroxylase-2b directs collagen cross-linking pathways in MC3T3-E1 cells. J Bone Miner Res 19: 1349–1355.
62. Salo AM, Sipila L, Sormunen R, Ruotsalainen H, Vainio S, et al. (2006) The lysyl hydroxylase isoforms are widely expressed during mouse embryogenesis, but obtain tissue- and cell-specific patterns in the adult. Matrix Biol 25: 475–483.
63. Alexakis C, Partridge T, Bou-Gharios G (2007) Implication of the satellite cell in dystrophic muscle fibrosis: a self-perpetuating mechanism of collagen overproduction. Am J Physiol Cell Physiol 293: C661–669.
64. Mayr C, Bartel DP (2009) Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell 138: 673–684.
65. Klamut HJ, Zubrzycka-Gaarn EE, Bulman DE, Malhotra SB, Bodrug SE, et al. (1989) Myogenic regulation of dystrophin gene expression. Br Med Bull 45: 681–702.
66. Nudel U, Robzyk K, Yaffe D (1988) Expression of the putative Duchenne muscular dystrophy gene in differentiated myogenic cell cultures and in the brain. Nature 331: 635–638.
67. Ghosh T, Soni K, Scaria V, Halimani M, Bhattacharjee C, et al. (2008) MicroRNA-mediated up-regulation of an alternatively polyadenylated variant of the mouse cytoplasmic {beta}-actin gene. Nucleic Acids Res 36: 6318–6332.
68. Legendre M, Ritchie W, Lopez F, Gautheret D (2006) Differential repression of alternative transcripts: a screen for miRNA targets. PLoS Comput Biol 2: e43.
69. Laneve P, Di Marcotullio L, Gioia U, Fiori ME, Ferretti E, et al. (2007) The interplay between microRNAs and the neurotrophin receptor tropomyosin-related kinase C controls proliferation of human neuroblastoma cells. Proc Natl Acad Sci U S A 104: 7957–7962.

# References


1 Gardner, R. L. Cell allocation and lineage in the early mouse embryo. *Ciba Found Symp* **144**, 172-181; discussion 181-176, 208-111 (1989).

2 Kunath, T., Strumpf, D. & Rossant, J. Early trophoblast determination and stem cell maintenance in the mouse--a review. *Placenta* **25 Suppl A**, S32-38, doi:10.1016/j.placenta.2004.01.015
S0143400404000293 [pii] (2004).

3 Gardner, R. L. Origin and differentiation of extraembryonic tissues in the mouse. *Int Rev Exp Pathol* **24**, 63-133 (1983).

4 Kadokawa, Y., Kato, Y. & Eguchi, G. Cell lineage analysis of the primitive and visceral endoderm of mouse embryos cultured in vitro. *Cell Differ* **21**, 69-76 (1987).

5 Kwon, G. S., Viotti, M. & Hadjantonakis, A. K. The endoderm of the mouse embryo arises by dynamic widespread intercalation of embryonic and extraembryonic lineages. *Dev Cell* **15**, 509-520, doi:S1534-5807(08)00296-7 [pii]
10.1016/j.devcel.2008.07.017 (2008).

6 Gardner, R. L. Investigation of cell lineage and differentiation in the extraembryonic endoderm of the mouse embryo. *J Embryol Exp Morphol* **68**, 175-198 (1982).

7 Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-156 (1981).

8 Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* **78**, 7634-7638 (1981).

9 Smith, A. G. *et al.* Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688-690, doi:10.1038/336688a0 (1988).

10 Beddington, R. S. & Robertson, E. J. An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development* **105**, 733-737 (1989).

11 Gossler, A., Doetschman, T., Korn, R., Serfling, E. & Kemler, R. Transgenesis by means of blastocyst-derived embryonic stem cell lines. *Proc Natl Acad Sci U S A* **83**, 9065-9069 (1986).

12 Chazaud, C., Yamanaka, Y., Pawson, T. & Rossant, J. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Dev Cell* **10**, 615-624, doi:S1534-5807(06)00125-0 [pii]
10.1016/j.devcel.2006.02.020 (2006).

13 Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643-655, doi:S0092867403003921 [pii] (2003).

14 Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631-642, doi:S0092867403003933 [pii] (2003).

15 Ying, Q. L., Nichols, J., Chambers, I. & Smith, A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**, 281-292, doi:S009286740300847X [pii] (2003).

16 Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947-956, doi:S0092-8674(05)00825-1 [pii]
10.1016/j.cell.2005.08.020 (2005).

17 Guo, G. *et al.* Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev Cell* **18**, 675-685, doi:S1534-5807(10)00110-3 [pii]
10.1016/j.devcel.2010.02.012 (2010).
18 Arceci, R. J., King, A. A., Simon, M. C., Orkin, S. H. & Wilson, D. B. Mouse GATA-4: a retinoic acid-inducible GATA-binding transcription factor expressed in endodermally derived tissues and heart. *Mol Cell Biol* **13**, 2235-2246 (1993).
19 Morrisey, E. E., Ip, H. S., Lu, M. M. & Parmacek, M. S. GATA-6: a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. *Dev Biol* **177**, 309-322, doi:10.1006/dbio.1996.0165
S0012-1606(96)90165-7 [pii] (1996).
20 Cai, K. Q., Capo-Chichi, C. D., Rula, M. E., Yang, D. H. & Xu, X. X. Dynamic GATA6 expression in primitive endoderm formation and maturation in early mouse embryogenesis. *Dev Dyn* **237**, 2820-2829, doi:10.1002/dvdy.21703 (2008).
21 Morrisey, E. E. *et al.* GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev* **12**, 3579-3590 (1998).
22 Fujikura, J. *et al.* Differentiation of embryonic stem cells is induced by GATA factors. *Genes Dev* **16**, 784-789, doi:10.1101/gad.968802 (2002).
23 Singh, A. M., Hamazaki, T., Hankowski, K. E. & Terada, N. A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem Cells* **25**, 2534-2542, doi:2007-0126 [pii]
10.1634/stemcells.2007-0126 (2007).
24 Messerschmidt, D. M. & Kemler, R. Nanog is required for primitive endoderm formation through a non-cell autonomous mechanism. *Dev Biol* **344**, 129-137, doi:S0012-1606(10)00265-4 [pii]
10.1016/j.ydbio.2010.04.020 (2010).
25 Frankenberg, S. *et al.* Primitive Endoderm Differentiates via a Three-Step Mechanism Involving Nanog and RTK Signaling. *Dev Cell* **21**, 1005-1013, doi:S1534-5807(11)00471-0 [pii]
10.1016/j.devcel.2011.10.019 (2011).
26 Niakan, K. K. *et al.* Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes Dev* **24**, 312-326, doi:24/3/312 [pii]
10.1101/gad.1833510 (2010).
27 Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230-1234, doi:nature06403 [pii]
10.1038/nature06403 (2007).
28 Furusawa, T., Ohkoshi, K., Honda, C., Takahashi, S. & Tokunaga, T. Embryonic stem cells expressing both platelet endothelial cell adhesion molecule-1 and stage-specific embryonic antigen-1 differentiate predominantly into epiblast cells in a chimeric embryo. *Biol Reprod* **70**, 1452-1457, doi:10.1095/biolreprod.103.024190
biolreprod.103.024190 [pii] (2004).
29 Hayashi, K., Lopes, S. M., Tang, F. & Surani, M. A. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* **3**, 391-401, doi:S1934-5909(08)00398-6 [pii]
10.1016/j.stem.2008.07.027 (2008).

30 Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. & Niwa, H. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* **135**, 909-918, doi:135/5/909 [pii]
10.1242/dev.017400 (2008).
31 Payer, B. *et al.* Generation of stella-GFP transgenic mice: a novel tool to study germ cell development. *Genesis* **44**, 75-83, doi:10.1002/gene.20187 (2006).
32 Suzuki, A. *et al.* Maintenance of embryonic stem cell pluripotency by Nanog-mediated reversal of mesoderm specification. *Nat Clin Pract Cardiovasc Med* **3 Suppl 1**, S114-122, doi:ncpcardio0442 [pii]
10.1038/ncpcardio0442 (2006).
33 Canham, M. A., Sharov, A. A., Ko, M. S. & Brickman, J. M. Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. *PLoS Biol* **8**, e1000379, doi:10.1371/journal.pbio.1000379 (2010).
34 Hamazaki, T., Oka, M., Yamanaka, S. & Terada, N. Aggregation of embryonic stem cells induces Nanog repression and primitive endoderm differentiation. *J Cell Sci* **117**, 5681-5686, doi:jcs.01489 [pii]
10.1242/jcs.01489 (2004).
35 Becker, A. J., Mc, C. E. & Till, J. E. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature* **197**, 452-454 (1963).
36 Gussoni, E. *et al.* Dystrophin expression in the mdx mouse restored by stem cell transplantation. *Nature* **401**, 390-394, doi:10.1038/43919 (1999).
37 Ferrari, G. *et al.* Muscle regeneration by bone marrow-derived myogenic progenitors. *Science* **279**, 1528-1530 (1998).
38 Orlic, D. *et al.* Transplanted adult bone marrow cells repair myocardial infarcts in mice. *Ann N Y Acad Sci* **938**, 221-229; discussion 229-230 (2001).
39 Petersen, B. E. *et al.* Bone marrow as a potential source of hepatic oval cells. *Science* **284**, 1168-1170 (1999).
40 Brazelton, T. R., Rossi, F. M., Keshet, G. I. & Blau, H. M. From marrow to brain: expression of neuronal phenotypes in adult mice. *Science* **290**, 1775-1779, doi:9028 [pii] (2000).
41 Mezey, E., Chandross, K. J., Harta, G., Maki, R. A. & McKercher, S. R. Turning blood into brain: cells bearing neuronal antigens generated in vivo from bone marrow. *Science* **290**, 1779-1782, doi:9025 [pii] (2000).
42 Spangrude, G. J., Heimfeld, S. & Weissman, I. L. Purification and characterization of mouse hematopoietic stem cells. *Science* **241**, 58-62 (1988).
43 Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* **273**, 242-245 (1996).
44 Temple, S. Division and differentiation of isolated CNS blast cells in microculture. *Nature* **340**, 471-473, doi:10.1038/340471a0 (1989).
45 Reynolds, B. A. & Weiss, S. Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. *Science* **255**, 1707-1710 (1992).
46 Lewis, P. D. Mitotic activity in the primate subependymal layer and the genesis of gliomas. *Nature* **217**, 974-975 (1968).
47 Mauro, A. Satellite cell of skeletal muscle fibers. *J Biophys Biochem Cytol* **9**, 493-495 (1961).

48 Asakura, A., Komaki, M. & Rudnicki, M. Muscle satellite cells are multipotential stem cells that exhibit myogenic, osteogenic, and adipogenic differentiation. *Differentiation* **68**, 245-253 (2001).
49 Kuang, S., Kuroda, K., Le Grand, F. & Rudnicki, M. A. Asymmetric self-renewal and commitment of satellite stem cells in muscle. *Cell* **129**, 999-1010, doi:S0092-8674(07)00513-2 [pii]
10.1016/j.cell.2007.03.044 (2007).
50 Collins, C. A. *et al.* Stem cell function, self-renewal, and behavioral heterogeneity of cells from the adult muscle satellite cell niche. *Cell* **122**, 289-301, doi:S0092-8674(05)00455-1 [pii]
10.1016/j.cell.2005.05.010 (2005).
51 Toma, J. G. *et al.* Isolation of multipotent adult stem cells from the dermis of mammalian skin. *Nat Cell Biol* **3**, 778-784, doi:10.1038/ncb0901-778
ncb0901-778 [pii] (2001).
52 Gronthos, S., Mankani, M., Brahim, J., Robey, P. G. & Shi, S. Postnatal human dental pulp stem cells (DPSCs) in vitro and in vivo. *Proc Natl Acad Sci U S A* **97**, 13625-13630, doi:10.1073/pnas.240309797
240309797 [pii] (2000).
53 Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003-1007, doi:nature06196 [pii]
10.1038/nature06196 (2007).
54 Snow, M. H. Myogenic cell formation in regenerating rat skeletal muscle injured by mincing. II. An autoradiographic study. *Anat Rec* **188**, 201-217, doi:10.1002/ar.1091880206 (1977).
55 Moss, F. P. & Leblond, C. P. Nature of dividing nuclei in skeletal muscle of growing rats. *J Cell Biol* **44**, 459-462 (1970).
56 Reznik, M. Thymidine-3H uptake by satellite cells of regenerating skeletal muscle. *J Cell Biol* **40**, 568-571 (1969).
57 Konigsberg, I. R. Clonal analysis of myogenesis. *Science* **140**, 1273-1284 (1963).
58 Yaffe, D. Cellular aspects of muscle differentiation in vitro. *Curr Top Dev Biol* **4**, 37-77 (1969).
59 Bischoff, R. Regeneration of single skeletal muscle fibers in vitro. *Anat Rec* **182**, 215-235, doi:10.1002/ar.1091820207 (1975).
60 Konigsberg, U. R., Lipton, B. H. & Konigsberg, I. R. The regenerative response of single mature muscle fibers isolated in vitro. *Dev Biol* **45**, 260-275, doi:0012-1606(75)90065-2 [pii] (1975).
61 d'Albis, A. *et al.* Regulation by thyroid hormones of terminal differentiation in the skeletal dorsal muscle. I. Neonate mouse. *Dev Biol* **123**, 25-32, doi:0012-1606(87)90423-4 [pii] (1987).
62 Hansen-Smith, F. M., Picou, D. & Golden, M. H. Muscle satellite cells in malnourished and nutritionally rehabilitated children. *J Neurol Sci* **41**, 207-221 (1979).
63 Jirmanova, I. & Thesleff, S. Ultrastructural study of experimental muscle degeneration and regeneration in the adult rat. *Z Zellforsch Mikrosk Anat* **131**, 77-97 (1972).
64 McGeachie, J. K. & Grounds, M. D. Initiation and duration of muscle precursor replication after mild and severe injury to skeletal muscle of mice. An autoradiographic study. *Cell Tissue Res* **248**, 125-130 (1987).

65 Mulvaney, D. R., Marple, D. N. & Merkel, R. A. Proliferation of skeletal muscle satellite cells after castration and administration of testosterone propionate. *Proc Soc Exp Biol Med* **188**, 40-45 (1988).
66 Phillips, W. D. & Bennett, M. R. Elimination of distributed synaptic acetylcholine receptor clusters on developing avian fast-twitch muscle fibres accompanies loss of polyneuronal innervation. *J Neurocytol* **16**, 785-797 (1987).
67 Le Grand, F., Jones, A. E., Seale, V., Scime, A. & Rudnicki, M. A. Wnt7a activates the planar cell polarity pathway to drive the symmetric expansion of satellite stem cells. *Cell Stem Cell* **4**, 535-547, doi:S1934-5909(09)00148-9 [pii]
10.1016/j.stem.2009.03.013 (2009).
68 Schultz, E., Gibson, M. C. & Champion, T. Satellite cells are mitotically quiescent in mature mouse muscle: an EM and radioautographic study. *J Exp Zool* **206**, 451-456, doi:10.1002/jez.1402060314 (1978).
69 Bischoff, R. A satellite cell mitogen from crushed adult muscle. *Dev Biol* **115**, 140-147, doi:0012-1606(86)90235-6 [pii] (1986).
70 Seale, P. *et al.* Pax7 is required for the specification of myogenic satellite cells. *Cell* **102**, 777-786, doi:S0092-8674(00)00066-0 [pii] (2000).
71 Buckingham, M. *et al.* The formation of skeletal muscle: from somite to limb. *J Anat* **202**, 59-68 (2003).
72 Cornelison, D. D. & Wold, B. J. Single-cell analysis of regulatory gene expression in quiescent and activated mouse skeletal muscle satellite cells. *Dev Biol* **191**, 270-283, doi:S0012-1606(97)98721-2 [pii]
10.1006/dbio.1997.8721 (1997).
73 Irintchev, A., Zeschnigk, M., Starzinski-Powitz, A. & Wernig, A. Expression pattern of M-cadherin in normal, denervated, and regenerating mouse muscles. *Dev Dyn* **199**, 326-337, doi:10.1002/aja.1001990407 (1994).
74 Allen, R. E., Sheehan, S. M., Taylor, R. G., Kendall, T. L. & Rice, G. M. Hepatocyte growth factor activates quiescent skeletal muscle satellite cells in vitro. *J Cell Physiol* **165**, 307-312, doi:10.1002/jcp.1041650211 (1995).
75 Burkin, D. J. & Kaufman, S. J. The alpha7beta1 integrin in muscle development and disease. *Cell Tissue Res* **296**, 183-190 (1999).
76 Gnocchi, V. F., White, R. B., Ono, Y., Ellis, J. A. & Zammit, P. S. Further characterisation of the molecular signature of quiescent and activated mouse muscle satellite cells. *PLoS One* **4**, e5205, doi:10.1371/journal.pone.0005205 (2009).
77 Beauchamp, J. R. *et al.* Expression of CD34 and Myf5 defines the majority of quiescent adult skeletal muscle satellite cells. *J Cell Biol* **151**, 1221-1234 (2000).
78 Cornelison, D. D., Filla, M. S., Stanley, H. M., Rapraeger, A. C. & Olwin, B. B. Syndecan-3 and syndecan-4 specifically mark skeletal muscle satellite cells and are implicated in satellite cell maintenance and muscle regeneration. *Dev Biol* **239**, 79-94, doi:10.1006/dbio.2001.0416
S0012-1606(01)90416-6 [pii] (2001).
79 Volonte, D., Liu, Y. & Galbiati, F. The modulation of caveolin-1 expression controls satellite cell activation during muscle repair. *FASEB J* **19**, 237-239, doi:04-2215fje [pii]
10.1096/fj.04-2215fje (2005).
80 Fukada, S. *et al.* Molecular signature of quiescent satellite cells in adult skeletal muscle. *Stem Cells* **25**, 2448-2459, doi:2007-0019 [pii]
10.1634/stemcells.2007-0019 (2007).

81 McLoon, L. K. & Wirtschafter, J. Activated satellite cells in extraocular muscles of normal adult monkeys and humans. *Invest Ophthalmol Vis Sci* **44**, 1927-1932 (2003).
82 Patruno, M., Caliaro, F., Martinello, T. & Mascarello, F. Expression of the paired box domain Pax7 protein in myogenic cells isolated from the porcine semitendinosus muscle after birth. *Tissue Cell* **40**, 1-6, doi:S0040-8166(07)00077-8 [pii]
10.1016/j.tice.2007.08.006 (2008).
83 Halevy, O. *et al.* Pattern of Pax7 expression during myogenesis in the posthatch chicken establishes a model for satellite cell differentiation and renewal. *Dev Dyn* **231**, 489-502, doi:10.1002/dvdy.20151 (2004).
84 Morrison, J. I., Loof, S., He, P. & Simon, A. Salamander limb regeneration involves the activation of a multipotent skeletal muscle satellite cell population. *J Cell Biol* **172**, 433-440, doi:jcb.200509011 [pii]
10.1083/jcb.200509011 (2006).
85 Chen, Y., Lin, G. & Slack, J. M. Control of muscle regeneration in the Xenopus tadpole tail by Pax7. *Development* **133**, 2303-2313, doi:dev.02397 [pii]
10.1242/dev.02397 (2006).
86 Hammond, C. L. *et al.* Signals and myogenic regulatory factors restrict pax3 and pax7 expression to dermomyotome-like tissue in zebrafish. *Dev Biol* **302**, 504-521, doi:S0012-1606(06)01299-1 [pii]
10.1016/j.ydbio.2006.10.009 (2007).
87 Rosenblatt, J. D., Lunt, A. I., Parry, D. J. & Partridge, T. A. Culturing satellite cells from living single muscle fiber explants. *In Vitro Cell Dev Biol Anim* **31**, 773-779, doi:10.1007/BF02634119 (1995).
88 Conboy, I. M., Conboy, M. J., Smythe, G. M. & Rando, T. A. Notch-mediated restoration of regenerative potential to aged muscle. *Science* **302**, 1575-1577, doi:10.1126/science.1087573
302/5650/1575 [pii] (2003).
89 Lepper, C., Conway, S. J. & Fan, C. M. Adult satellite cells and embryonic muscle progenitors have distinct genetic requirements. *Nature* **460**, 627-631, doi:nature08209 [pii]
10.1038/nature08209 (2009).
90 Bosnakovski, D. *et al.* Prospective isolation of skeletal muscle stem cells with a Pax7 reporter. *Stem Cells* **26**, 3194-3204, doi:2007-1017 [pii]
10.1634/stemcells.2007-1017 (2008).
91 Day, K., Shefer, G., Richardson, J. B., Enikolopov, G. & Yablonka-Reuveni, Z. Nestin-GFP reporter expression defines the quiescent state of skeletal muscle satellite cells. *Dev Biol* **304**, 246-259, doi:S0012-1606(06)01455-2 [pii]
10.1016/j.ydbio.2006.12.026 (2007).
92 Doumit, M. E. & Merkel, R. A. Conditions for isolation and culture of porcine myogenic satellite cells. *Tissue Cell* **24**, 253-262, doi:0040-8166(92)90098-R [pii] (1992).
93 Musaro, A. & Barberi, L. Isolation and culture of mouse satellite cells. *Methods Mol Biol* **633**, 101-111, doi:10.1007/978-1-59745-019-5_8 (2010).
94 Rando, T. A. & Blau, H. M. Primary mouse myoblast purification, characterization, and transplantation for cell-mediated gene therapy. *J Cell Biol* **125**, 1275-1287 (1994).
95 Fan, Y., Maley, M., Beilharz, M. & Grounds, M. Rapid death of injected myoblasts in myoblast transfer therapy. *Muscle Nerve* **19**, 853-860, doi:10.1002/(SICI)1097-4598(199607)19:7<853::AID-MUS7>3.0.CO;2-8 [pii]
10.1002/(SICI)1097-4598(199607)19:7<853::AID-MUS7>3.0.CO;2-8 (1996).

96 Montarras, D. *et al.* Direct isolation of satellite cells for skeletal muscle regeneration. *Science* **309**, 2064-2067, doi:1114758 [pii]
10.1126/science.1114758 (2005).
97 Pallafacchina, G. *et al.* An adult tissue-specific stem cell in its niche: a gene profiling analysis of in vivo quiescent and activated muscle satellite cells. *Stem Cell Res* **4**, 77-91, doi:S1873-5061(09)00119-6 [pii]
10.1016/j.scr.2009.10.003 (2010).
98 Scaal, M. & Christ, B. Formation and differentiation of the avian dermomyotome. *Anat Embryol (Berl)* **208**, 411-424, doi:10.1007/s00429-004-0417-y (2004).
99 Armand, O., Boutineau, A. M., Mauger, A., Pautou, M. P. & Kieny, M. Origin of satellite cells in avian skeletal muscles. *Arch Anat Microsc Morphol Exp* **72**, 163-181 (1983).
100 Kassar-Duchossoy, L. *et al.* Pax3/Pax7 mark a novel population of primitive myogenic cells during development. *Genes Dev* **19**, 1426-1431, doi:19/12/1426 [pii]
10.1101/gad.345505 (2005).
101 Mansouri, A., Stoykova, A., Torres, M. & Gruss, P. Dysgenesis of cephalic neural crest derivatives in Pax7-/- mutant mice. *Development* **122**, 831-838 (1996).
102 Relaix, F. *et al.* The transcriptional activator PAX3-FKHR rescues the defects of Pax3 mutant mice but induces a myogenic gain-of-function phenotype with ligand-independent activation of Met signaling in vivo. *Genes Dev* **17**, 2950-2965, doi:10.1101/gad.281203
17/23/2950 [pii] (2003).
103 Relaix, F., Rocancourt, D., Mansouri, A. & Buckingham, M. Divergent functions of murine Pax3 and Pax7 in limb muscle development. *Genes Dev* **18**, 1088-1105, doi:10.1101/gad.301004
18/9/1088 [pii] (2004).
104 Ben-Yair, R. & Kalcheim, C. Lineage analysis of the avian dermomyotome sheet reveals the existence of single cells with both dermal and muscle progenitor fates. *Development* **132**, 689-701, doi:dev.01617 [pii]
10.1242/dev.01617 (2005).
105 Gros, J., Manceau, M., Thome, V. & Marcelle, C. A common somitic origin for embryonic muscle progenitors and satellite cells. *Nature* **435**, 954-958, doi:nature03572 [pii]
10.1038/nature03572 (2005).
106 Schienda, J. *et al.* Somitic origin of limb muscle satellite and side population cells. *Proc Natl Acad Sci U S A* **103**, 945-950, doi:0510164103 [pii]
10.1073/pnas.0510164103 (2006).
107 Kanisicak, O., Mendez, J. J., Yamamoto, S., Yamamoto, M. & Goldhamer, D. J. Progenitors of skeletal muscle satellite cells express the muscle determination gene, MyoD. *Dev Biol* **332**, 131-141, doi:S0012-1606(09)00873-2 [pii]
10.1016/j.ydbio.2009.05.554 (2009).
108 Lepper, C. & Fan, C. M. Inducible lineage tracing of Pax7-descendant cells reveals embryonic origin of adult satellite cells. *Genesis* **48**, 424-436, doi:10.1002/dvg.20630 (2010).
109 Relaix, F., Rocancourt, D., Mansouri, A. & Buckingham, M. A Pax3/Pax7-dependent population of skeletal muscle progenitor cells. *Nature* **435**, 948-953, doi:nature03594 [pii]
10.1038/nature03594 (2005).
110 Buckingham, M. Myogenic progenitor cells and skeletal myogenesis in vertebrates. *Curr Opin Genet Dev* **16**, 525-532, doi:S0959-437X(06)00159-6 [pii]

10.1016/j.gde.2006.08.008 (2006).
111 Cinnamon, Y., Ben-Yair, R. & Kalcheim, C. Differential effects of N-cadherin-mediated adhesion on the development of myotomal waves. *Development* **133**, 1101-1112, doi:dev.02291 [pii]
10.1242/dev.02291 (2006).
112 Otto, A., Schmidt, C. & Patel, K. Pax3 and Pax7 expression and regulation in the avian embryo. *Anat Embryol (Berl)* **211**, 293-310, doi:10.1007/s00429-006-0083-3 (2006).
113 He, L. *et al.* Three different fates of cells migrating from somites into the limb bud. *Anat Embryol (Berl)* **207**, 29-34, doi:10.1007/s00429-003-0327-4 (2003).
114 Kardon, G., Campbell, J. K. & Tabin, C. J. Local extrinsic signals determine muscle and endothelial cell fate and patterning in the vertebrate limb. *Dev Cell* **3**, 533-545, doi:S1534580702002915 [pii] (2002).
115 Hutcheson, D. A., Zhao, J., Merrell, A., Haldar, M. & Kardon, G. Embryonic and fetal limb myogenic cells are derived from developmentally distinct progenitors and have different requirements for beta-catenin. *Genes Dev* **23**, 997-1013, doi:gad.1769009 [pii]
10.1101/gad.1769009 (2009).
116 Hall-Craggs, E. C. & Seyan, H. S. Histochemical changes in innervated and denervated skeletal muscle fibers following treatment with bupivacaine (marcain). *Exp Neurol* **46**, 345-354, doi:0014-4886(75)90140-5 [pii] (1975).
117 Whalen, R. G., Harris, J. B., Butler-Browne, G. S. & Sesodia, S. Expression of myosin isoforms during notexin-induced regeneration of rat soleus muscles. *Dev Biol* **141**, 24-40, doi:0012-1606(90)90099-5 [pii] (1990).
118 Alderton, J. M. & Steinhardt, R. A. How calcium influx through calcium leak channels is responsible for the elevated levels of calcium-dependent proteolysis in dystrophic myotubes. *Trends Cardiovasc Med* **10**, 268-272, doi:S1050-1738(00)00075-X [pii] (2000).
119 Armstrong, R. B. Initial events in exercise-induced muscular injury. *Med Sci Sports Exerc* **22**, 429-435 (1990).
120 Belcastro, A. N., Shewchuk, L. D. & Raj, D. A. Exercise-induced muscle injury: a calpain hypothesis. *Mol Cell Biochem* **179**, 135-145 (1998).
121 Orimo, S., Hiyamuta, E., Arahata, K. & Sugita, H. Analysis of inflammatory cells and complement C3 in bupivacaine-induced myonecrosis. *Muscle Nerve* **14**, 515-520, doi:10.1002/mus.880140605 (1991).
122 Tidball, J. G. Inflammatory cell response to acute muscle injury. *Med Sci Sports Exerc* **27**, 1022-1032 (1995).
123 Sabourin, L. A., Girgis-Gabardo, A., Seale, P., Asakura, A. & Rudnicki, M. A. Reduced differentiation potential of primary MyoD-/- myogenic cells derived from adult skeletal muscle. *J Cell Biol* **144**, 631-643 (1999).
124 Cornelison, D. D., Olwin, B. B., Rudnicki, M. A. & Wold, B. J. MyoD(-/-) satellite cells in single-fiber culture are differentiation defective and MRF4 deficient. *Dev Biol* **224**, 122-137, doi:10.1006/dbio.2000.9682
S0012-1606(00)99682-9 [pii] (2000).
125 Hollenberg, S. M., Cheng, P. F. & Weintraub, H. Use of a conditional MyoD transcription factor in studies of MyoD trans-activation and muscle determination. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 8028-8032 (1993).
126 Gayraud-Morel, B. *et al.* A role for the myogenic determination gene Myf5 in adult regenerative myogenesis. *Dev Biol* **312**, 13-28, doi:S0012-1606(07)01274-2 [pii]
10.1016/j.ydbio.2007.08.059 (2007).

127 Ustanina, S., Carvajal, J., Rigby, P. & Braun, T. The myogenic factor Myf5 supports efficient skeletal muscle regeneration by enabling transient myoblast amplification. *Stem Cells* **25**, 2006-2016, doi:2006-0736 [pii]
10.1634/stemcells.2006-0736 (2007).
128 Cooper, R. N. *et al.* In vivo satellite cell activation via Myf5 and MyoD in regenerating mouse skeletal muscle. *J Cell Sci* **112 ( Pt 17)**, 2895-2901 (1999).
129 Montarras, D., Lindon, C., Pinset, C. & Domeyne, P. Cultured myf5 null and myoD null muscle precursor cells display distinct growth defects. *Biol Cell* **92**, 565-572 (2000).
130 Snow, M. H. An autoradiographic study of satellite cell differentiation into regenerating myotubes following transplantation of muscles in young rats. *Cell Tissue Res* **186**, 535-540 (1978).
131 Walker, B. E. An investigation of skeletal muscle regeneration with radioautography. *Anat Rec*, 350 (1960).
132 Megeney, L. A., Kablar, B., Garrett, K., Anderson, J. E. & Rudnicki, M. A. MyoD is required for myogenic stem cell function in adult skeletal muscle. *Genes Dev* **10**, 1173-1183 (1996).
133 White, J. D. *et al.* Myotube formation is delayed but not prevented in MyoD-deficient skeletal muscle: studies in regenerating whole muscle grafts of adult mice. *J Histochem Cytochem* **48**, 1531-1544 (2000).
134 Yablonka-Reuveni, Z. *et al.* The transition from proliferation to differentiation is delayed in satellite cells from mice lacking MyoD. *Dev Biol* **210**, 440-455, doi:10.1006/dbio.1999.9284
S0012-1606(99)99284-9 [pii] (1999).
135 Rudnicki, M. A., Le Grand, F., McKinnell, I. & Kuang, S. The molecular regulation of muscle stem cell function. *Cold Spring Harb Symp Quant Biol* **73**, 323-331, doi:sqb.2008.73.064 [pii]
10.1101/sqb.2008.73.064 (2008).
136 Penn, B. H., Bergstrom, D. A., Dilworth, F. J., Bengal, E. & Tapscott, S. J. A MyoD-generated feed-forward circuit temporally patterns gene expression during skeletal muscle differentiation. *Genes Dev* **18**, 2348-2353, doi:18/19/2348 [pii]
10.1101/gad.1234304 (2004).
137 Berkes, C. A. & Tapscott, S. J. MyoD and the transcriptional control of myogenesis. *Semin Cell Dev Biol* **16**, 585-595, doi:S1084-9521(05)00087-X [pii]
10.1016/j.semcdb.2005.07.006 (2005).
138 Grounds, M. D., Garrett, K. L., Lai, M. C., Wright, W. E. & Beilharz, M. W. Identification of skeletal muscle precursor cells in vivo by use of MyoD1 and myogenin probes. *Cell Tissue Res* **267**, 99-104 (1992).
139 Smith, C. K., 2nd, Janney, M. J. & Allen, R. E. Temporal expression of myogenic regulatory genes during activation, proliferation, and differentiation of rat skeletal muscle satellite cells. *J Cell Physiol* **159**, 379-385, doi:10.1002/jcp.1041590222 (1994).
140 Yablonka-Reuveni, Z. & Rivera, A. J. Temporal expression of regulatory and structural muscle proteins during myogenesis of satellite cells on isolated adult rat fibers. *Dev Biol* **164**, 588-603, doi:S0012-1606(84)71226-7 [pii]
10.1006/dbio.1994.1226 (1994).
141 Cao, Y. *et al.* Global and gene-specific analyses show distinct roles for Myod and Myog at a common set of promoters. *EMBO J* **25**, 502-511, doi:7600958 [pii]
10.1038/sj.emboj.7600958 (2006).

142 Charge, S. B. & Rudnicki, M. A. Cellular and molecular regulation of muscle regeneration. *Physiol Rev* **84**, 209-238, doi:10.1152/physrev.00019.2003
84/1/209 [pii] (2004).
143 Knudsen, K. A. & Horwitz, A. F. Tandem events in myoblast fusion. *Dev Biol* **58**, 328-338, doi:0012-1606(77)90095-1 [pii] (1977).
144 Wakelam, M. J. The fusion of myoblasts. *Biochem J* **228**, 1-12 (1985).
145 Przybylski, R. J. & Blumberg, J. M. Ultrastructural aspects of myogenesis in the chick. *Lab Invest* **15**, 836-863 (1966).
146 Lipton, B. H. & Konigsberg, I. R. A fine-structural analysis of the fusion of myogenic cells. *J Cell Biol* **53**, 348-364 (1972).
147 Rash, J. E. & Fambrough, D. Ultrastructural and electrophysiological correlates of cell coupling and cytoplasmic fusion during myogenesis in vitro. *Dev Biol* **30**, 166-186, doi:0012-1606(73)90055-9 [pii] (1973).
148 Pollina, E. A. & Brunet, A. Epigenetic regulation of aging stem cells. *Oncogene* **30**, 3105-3126, doi:10.1038/onc.2011.45
onc201145 [pii] (2011).
149 Fielding, R. A. *et al.* Sarcopenia: an undiagnosed condition in older adults. Current consensus definition: prevalence, etiology, and consequences. International working group on sarcopenia. *J Am Med Dir Assoc* **12**, 249-256, doi:S1525-8610(11)00019-3 [pii]
10.1016/j.jamda.2011.01.003 (2011).
150 Cartee, G. D. What insights into age-related changes in skeletal muscle are provided by animal models? *J Gerontol A Biol Sci Med Sci* **50 Spec No**, 137-141 (1995).
151 Carlson, B. M. Factors influencing the repair and adaptation of muscles in aged individuals: satellite cells and innervation. *J Gerontol A Biol Sci Med Sci* **50 Spec No**, 96-100 (1995).
152 Fahim, M. A. & Robbins, N. Ultrastructural studies of young and old mouse neuromuscular junctions. *J Neurocytol* **11**, 641-656 (1982).
153 Faulkner, J. A., Brooks, S. V. & Zerba, E. Muscle atrophy and weakness with aging: contraction-induced injury as an underlying mechanism. *J Gerontol A Biol Sci Med Sci* **50 Spec No**, 124-129 (1995).
154 Kadhiresan, V. A., Hassett, C. A. & Faulkner, J. A. Properties of single motor units in medial gastrocnemius muscles of adult and old rats. *J Physiol* **493 ( Pt 2)**, 543-552 (1996).
155 Allbrook, D. B., Han, M. F. & Hellmuth, A. E. Population of muscle satellite cells in relation to age and mitotic activity. *Pathology* **3**, 223-243 (1971).
156 Brack, A. S. & Rando, T. A. Intrinsic changes and extrinsic influences of myogenic stem cell function during aging. *Stem Cell Rev* **3**, 226-237, doi:SCR:3:3:226 [pii] (2007).
157 Schultz, E. & Lipton, B. H. Skeletal muscle satellite cells: changes in proliferation potential as a function of age. *Mech Ageing Dev* **20**, 377-383 (1982).
158 Barani, A. E., Durieux, A. C., Sabido, O. & Freyssenet, D. Age-related changes in the mitotic and metabolic characteristics of muscle-derived cells. *J Appl Physiol* **95**, 2089-2098, doi:10.1152/japplphysiol.00437.2003
95/5/2089 [pii] (2003).
159 Shefer, G., Van de Mark, D. P., Richardson, J. B. & Yablonka-Reuveni, Z. Satellite-cell pool size does matter: defining the myogenic potency of aging skeletal muscle. *Dev Biol* **294**, 50-66, doi:S0012-1606(06)00119-9 [pii]
10.1016/j.ydbio.2006.02.022 (2006).

160 Ono, Y., Boldrin, L., Knopp, P., Morgan, J. E. & Zammit, P. S. Muscle satellite cells are a functionally heterogeneous population in both somite-derived and branchiomeric muscles. *Developmental biology* **337**, 29-41, doi:10.1016/j.ydbio.2009.10.005 (2010).
161 Carlson, B. M. & Faulkner, J. A. Muscle transplantation between young and old rats: age of host determines recovery. *Am J Physiol* **256**, C1262-1266 (1989).
162 Conboy, I. M. *et al.* Rejuvenation of aged progenitor cells by exposure to a young systemic environment. *Nature* **433**, 760-764, doi:nature03260 [pii]
10.1038/nature03260 (2005).
163 Brack, A. S. *et al.* Increased Wnt signaling during aging alters muscle stem cell fate and increases fibrosis. *Science* **317**, 807-810, doi:317/5839/807 [pii]
10.1126/science.1144090 (2007).
164 Collinsworth, A. M., Zhang, S., Kraus, W. E. & Truskey, G. A. Apparent elastic modulus and hysteresis of skeletal muscle cells throughout differentiation. *Am J Physiol Cell Physiol* **283**, C1219-1227, doi:10.1152/ajpcell.00502.2001 (2002).
165 Engler, A. J. *et al.* Myotubes differentiate optimally on substrates with tissue-like stiffness: pathological implications for soft or stiff microenvironments. *J Cell Biol* **166**, 877-887, doi:10.1083/jcb.200405004
jcb.200405004 [pii] (2004).
166 Gao, Y., Kostrominova, T. Y., Faulkner, J. A. & Wineman, A. S. Age-related changes in the mechanical properties of the epimysium in skeletal muscles of rats. *J Biomech* **41**, 465-469, doi:S0021-9290(07)00398-3 [pii]
10.1016/j.jbiomech.2007.09.021 (2008).
167 Rosant, C., Nagel, M. D. & Perot, C. Aging affects passive stiffness and spindle function of the rat soleus muscle. *Exp Gerontol* **42**, 301-308, doi:S0531-5565(06)00327-5 [pii]
10.1016/j.exger.2006.10.007 (2007).
168 Ryan, N. A. *et al.* Lower skeletal muscle capillarization and VEGF expression in aged vs. young men. *J Appl Physiol* **100**, 178-185, doi:00827.2005 [pii]
10.1152/japplphysiol.00827.2005 (2006).
169 Brandes, R. P., Fleming, I. & Busse, R. Endothelial aging. *Cardiovasc Res* **66**, 286-294, doi:S0008-6363(05)00017-9 [pii]
10.1016/j.cardiores.2004.12.027 (2005).
170 Ashcroft, G. S., Mills, S. J. & Ashworth, J. J. Ageing and wound healing. *Biogerontology* **3**, 337-345, doi:5100237 [pii] (2002).
171 Snow, M. H. The effects of aging on satellite cells in skeletal muscles of mice and rats. *Cell Tissue Res* **185**, 399-408 (1977).
172 Logan, C. Y. & Nusse, R. The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol* **20**, 781-810, doi:10.1146/annurev.cellbio.20.010403.113126 (2004).
173 Archbold, H. C., Yang, Y. X., Chen, L. & Cadigan, K. M. How do they do Wnt they do?: regulation of transcription by the Wnt/beta-catenin pathway. *Acta Physiol (Oxf)*, doi:10.1111/j.1748-1716.2011.02293.x (2011).
174 Clevers, H. Wnt/beta-catenin signaling in development and disease. *Cell* **127**, 469-480, doi:S0092-8674(06)01344-4 [pii]
10.1016/j.cell.2006.10.018 (2006).
175 MacDonald, B. T., Tamai, K. & He, X. Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev Cell* **17**, 9-26, doi:S1534-5807(09)00257-3 [pii]
10.1016/j.devcel.2009.06.016 (2009).

176 Simons, M. & Mlodzik, M. Planar cell polarity signaling: from fly development to human disease. *Annu Rev Genet* **42**, 517-540, doi:10.1146/annurev.genet.42.110807.091432 (2008).
177 Gao, C. & Chen, Y. G. Dishevelled: The hub of Wnt signaling. *Cell Signal* **22**, 717-727, doi:S0898-6568(09)00368-4 [pii]
10.1016/j.cellsig.2009.11.021 (2010).
178 Saneyoshi, T., Kume, S., Amasaki, Y. & Mikoshiba, K. The Wnt/calcium pathway activates NF-AT and promotes ventral cell fate in Xenopus embryos. *Nature* **417**, 295-299, doi:10.1038/417295a
417295a [pii] (2002).
179 Nelson, W. J. & Nusse, R. Convergence of Wnt, beta-catenin, and cadherin pathways. *Science* **303**, 1483-1487, doi:10.1126/science.1094291
303/5663/1483 [pii] (2004).
180 Shibamoto, S. *et al.* Cytoskeletal reorganization by soluble Wnt-3a protein signalling. *Genes Cells* **3**, 659-670 (1998).
181 Haertel-Wiesmann, M., Liang, Y., Fantl, W. J. & Williams, L. T. Regulation of cyclooxygenase-2 and periostin by Wnt-3 in mouse mammary epithelial cells. *J Biol Chem* **275**, 32046-32051, doi:10.1074/jbc.M000074200
M000074200 [pii] (2000).
182 Lee, J. S., Ishimoto, A. & Yanagawa, S. Characterization of mouse dishevelled (Dvl) proteins in Wnt/Wingless signaling pathway. *J Biol Chem* **274**, 21464-21470 (1999).
183 Sato, N., Meijer, L., Skaltsounis, L., Greengard, P. & Brivanlou, A. H. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nat Med* **10**, 55-63, doi:10.1038/nm979
nm979 [pii] (2004).
184 Dravid, G. *et al.* Defining the role of Wnt/beta-catenin signaling in the survival, proliferation, and self-renewal of human embryonic stem cells. *Stem Cells* **23**, 1489-1501, doi:2005-0034 [pii]
10.1634/stemcells.2005-0034 (2005).
185 Singla, D. K., Schneider, D. J., LeWinter, M. M. & Sobel, B. E. wnt3a but not wnt11 supports self-renewal of embryonic stem cells. *Biochem Biophys Res Commun* **345**, 789-795, doi:S0006-291X(06)00935-1 [pii]
10.1016/j.bbrc.2006.04.125 (2006).
186 Ogawa, K., Nishinakamura, R., Iwamatsu, Y., Shimosato, D. & Niwa, H. Synergistic action of Wnt and LIF in maintaining pluripotency of mouse ES cells. *Biochem Biophys Res Commun* **343**, 159-166, doi:S0006-291X(06)00400-1 [pii]
10.1016/j.bbrc.2006.02.127 (2006).
187 Gertsenstein, M. *et al.* Efficient generation of germ line transmitting chimeras from C57BL/6N ES cells by aggregation with outbred host embryos. *PLoS One* **5**, e11260, doi:10.1371/journal.pone.0011260 (2010).
188 Sato, H., Amagai, K., Shimizukawa, R. & Tamai, Y. Stable generation of serum- and feeder-free embryonic stem cell-derived mice with full germline-competency by using a GSK3 specific inhibitor. *Genesis* **47**, 414-422, doi:10.1002/dvg.20514 (2009).
189 Kiyonari, H., Kaneko, M., Abe, S. & Aizawa, S. Three inhibitors of FGF receptor, ERK, and GSK3 establishes germline-competent embryonic stem cells of C57BL/6N mouse strain with high efficiency and stability. *Genesis* **48**, 317-327, doi:10.1002/dvg.20614 (2010).

190 Berge, D. T. *et al.* Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. *Nat Cell Biol* **13**, 1070-1075, doi:10.1038/ncb2314 ncb2314 [pii] (2011).
191 Lyashenko, N. *et al.* Differential requirement for the dual functions of beta-catenin in embryonic stem cell self-renewal and germ layer formation. *Nat Cell Biol* **13**, 753-761, doi:10.1038/ncb2260 ncb2260 [pii] (2011).
192 Huelsken, J. *et al.* Requirement for beta-catenin in anterior-posterior axis formation in mice. *J Cell Biol* **148**, 567-578 (2000).
193 Cole, M. F., Johnstone, S. E., Newman, J. J., Kagey, M. H. & Young, R. A. Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes Dev* **22**, 746-755, doi:22/6/746 [pii] 10.1101/gad.1642408 (2008).
194 Takao, Y., Yokota, T. & Koide, H. Beta-catenin up-regulates Nanog expression through interaction with Oct-3/4 in embryonic stem cells. *Biochem Biophys Res Commun* **353**, 699-705, doi:S0006-291X(06)02748-3 [pii] 10.1016/j.bbrc.2006.12.072 (2007).
195 Morkel, M. *et al.* Beta-catenin regulates Cripto- and Wnt3-dependent gene expression programs in mouse axis and mesoderm formation. *Development* **130**, 6283-6294, doi:10.1242/dev.00859 130/25/6283 [pii] (2003).
196 Pereira, L., Yi, F. & Merrill, B. J. Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal. *Mol Cell Biol* **26**, 7479-7491, doi:MCB.00368-06 [pii] 10.1128/MCB.00368-06 (2006).
197 Tam, W. L. *et al.* T-cell factor 3 regulates embryonic stem cell pluripotency and self-renewal by the transcriptional control of multiple lineage pathways. *Stem Cells* **26**, 2019-2031, doi:2007-1115 [pii] 10.1634/stemcells.2007-1115 (2008).
198 Yi, F., Pereira, L. & Merrill, B. J. Tcf3 functions as a steady-state limiter of transcriptional programs of mouse embryonic stem cell self-renewal. *Stem Cells* **26**, 1951-1960, doi:2008-0229 [pii] 10.1634/stemcells.2008-0229 (2008).
199 Yi, F. *et al.* Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. *Nat Cell Biol* **13**, 762-770, doi:10.1038/ncb2283 ncb2283 [pii] (2011).
200 Kelly, K. F. *et al.* beta-catenin enhances Oct-4 activity and reinforces pluripotency through a TCF-independent mechanism. *Cell Stem Cell* **8**, 214-227, doi:S1934-5909(10)00705-8 [pii] 10.1016/j.stem.2010.12.010 (2011).
201 Kemler, R. From cadherins to catenins: cytoplasmic protein interactions and regulation of cell adhesion. *Trends Genet* **9**, 317-321, doi:0168-9525(93)90250-L [pii] (1993).
202 Lanner, F. & Rossant, J. The role of FGF/Erk signaling in pluripotent cells. *Development* **137**, 3351-3360, doi:137/20/3351 [pii] 10.1242/dev.050146 (2010).
203 Kemp, C., Willems, E., Abdo, S., Lambiv, L. & Leyns, L. Expression of all Wnt genes and their secreted antagonists during mouse blastocyst and postimplantation development. *Dev Dyn* **233**, 1064-1075, doi:10.1002/dvdy.20408 (2005).

204 Mohamed, O. A., Dufort, D. & Clarke, H. J. Expression and estradiol regulation of Wnt genes in the mouse blastocyst identify a candidate pathway for embryo-maternal signaling at implantation. *Biol Reprod* **71**, 417-424, doi:10.1095/biolreprod.103.025692 biolreprod.103.025692 [pii] (2004).
205 Mohamed, O. A. *et al.* Uterine Wnt/beta-catenin signaling is required for implantation. *Proc Natl Acad Sci U S A* **102**, 8579-8584, doi:0500612102 [pii] 10.1073/pnas.0500612102 (2005).
206 Na, J., Lykke-Andersen, K., Torres Padilla, M. E. & Zernicka-Goetz, M. Dishevelled proteins regulate cell adhesion in mouse blastocyst and serve to monitor changes in Wnt signaling. *Dev Biol* **302**, 40-49, doi:S0012-1606(06)01106-7 [pii] 10.1016/j.ydbio.2006.08.036 (2007).
207 Lloyd, S., Fleming, T. P. & Collins, J. E. Expression of Wnt genes during mouse preimplantation development. *Gene Expr Patterns* **3**, 309-312, doi:S1567133X03000462 [pii] (2003).
208 Xie, H. *et al.* Inactivation of nuclear Wnt-beta-catenin signaling limits blastocyst competency for implantation. *Development* **135**, 717-727, doi:dev.015339 [pii] 10.1242/dev.015339 (2008).
209 Chazaud, C. & Rossant, J. Disruption of early proximodistal patterning and AVE formation in Apc mutants. *Development* **133**, 3379-3387, doi:dev.02523 [pii] 10.1242/dev.02523 (2006).
210 Moser, A. R. *et al.* Homozygosity for the Min allele of Apc results in disruption of mouse development prior to gastrulation. *Dev Dyn* **203**, 422-433, doi:10.1002/aja.1002030405 (1995).
211 Kielman, M. F. *et al.* Apc modulates embryonic stem-cell differentiation by controlling the dosage of beta-catenin signaling. *Nat Genet* **32**, 594-605, doi:10.1038/ng1045 ng1045 [pii] (2002).
212 Kemler, R. *et al.* Stabilization of beta-catenin in the mouse zygote leads to premature epithelial-mesenchymal transition in the epiblast. *Development* **131**, 5817-5824, doi:dev.01458 [pii] 10.1242/dev.01458 (2004).
213 Rivera-Perez, J. A. & Magnuson, T. Primitive streak formation in mice is preceded by localized activation of Brachyury and Wnt3. *Dev Biol* **288**, 363-371, doi:S0012-1606(05)00619-6 [pii] 10.1016/j.ydbio.2005.09.012 (2005).
214 Lustig, B. *et al.* Negative feedback loop of Wnt signaling through upregulation of conductin/axin2 in colorectal and liver tumors. *Mol Cell Biol* **22**, 1184-1193 (2002).
215 Ferrer-Vaquer, A. *et al.* A sensitive and bright single-cell resolution live imaging reporter of Wnt/ss-catenin signaling in the mouse. *BMC Dev Biol* **10**, 121, doi:1471-213X-10-121 [pii] 10.1186/1471-213X-10-121 (2010).
216 Currier, N. *et al.* Dynamic expression of a LEF-EGFP Wnt reporter in mouse development and cancer. *Genesis* **48**, 183-194, doi:10.1002/dvg.20604 (2010).
217 Mohamed, O. A., Clarke, H. J. & Dufort, D. Beta-catenin signaling marks the prospective site of primitive streak formation in the mouse embryo. *Dev Dyn* **231**, 416-424, doi:10.1002/dvdy.20135 (2004).
218 ten Berge, D. *et al.* Wnt signaling mediates self-organization and axis formation in embryoid bodies. *Cell Stem Cell* **3**, 508-518, doi:S1934-5909(08)00483-9 [pii] 10.1016/j.stem.2008.09.013 (2008).

219 Gadue, P., Huber, T. L., Paddison, P. J. & Keller, G. M. Wnt and TGF-beta signaling are required for the induction of an in vitro model of primitive streak formation using embryonic stem cells. *Proc Natl Acad Sci U S A* **103**, 16806-16811, doi:0603916103 [pii]
10.1073/pnas.0603916103 (2006).
220 Izumi, N., Era, T., Akimaru, H., Yasunaga, M. & Nishikawa, S. Dissecting the molecular hierarchy for mesendoderm differentiation through a combination of embryonic stem cell culture and RNA interference. *Stem Cells* **25**, 1664-1674, doi:2006-0681 [pii]
10.1634/stemcells.2006-0681 (2007).
221 Nakanishi, M. *et al.* Directed induction of anterior and posterior primitive streak by Wnt from embryonic stem cells cultured in a chemically defined serum-free medium. *FASEB J* **23**, 114-122, doi:fj.08-111203 [pii]
10.1096/fj.08-111203 (2009).
222 Aubert, J., Dunstan, H., Chambers, I. & Smith, A. Functional gene screening in embryonic stem cells implicates Wnt antagonism in neural differentiation. *Nat Biotechnol* **20**, 1240-1245, doi:10.1038/nbt763
nbt763 [pii] (2002).
223 Haegele, L. *et al.* Wnt signalling inhibits neural differentiation of embryonic stem cells by controlling bone morphogenetic protein expression. *Mol Cell Neurosci* **24**, 696-708, doi:S104474310300232X [pii] (2003).
224 Arnold, S. J. *et al.* Brachyury is a target gene of the Wnt/beta-catenin signaling pathway. *Mech Dev* **91**, 249-258, doi:S0925-4773(99)00309-3 [pii] (2000).
225 Lindsley, R. C., Gill, J. G., Kyba, M., Murphy, T. L. & Murphy, K. M. Canonical Wnt signaling is required for development of embryonic stem cell-derived mesoderm. *Development* **133**, 3787-3796, doi:dev.02551 [pii]
10.1242/dev.02551 (2006).
226 Fehling, H. J. *et al.* Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. *Development* **130**, 4217-4227 (2003).
227 Nostro, M. C., Cheng, X., Keller, G. M. & Gadue, P. Wnt, activin, and BMP signaling regulate distinct stages in the developmental pathway from embryonic stem cells to blood. *Cell Stem Cell* **2**, 60-71, doi:S1934-5909(07)00228-7 [pii]
10.1016/j.stem.2007.10.011 (2008).
228 Lengerke, C. *et al.* BMP and Wnt specify hematopoietic fate by activation of the Cdx-Hox pathway. *Cell Stem Cell* **2**, 72-82, doi:S1934-5909(07)00239-1 [pii]
10.1016/j.stem.2007.10.022 (2008).
229 Tran, T. H. *et al.* Wnt3a-induced mesoderm formation and cardiomyogenesis in human embryonic stem cells. *Stem Cells* **27**, 1869-1878, doi:10.1002/stem.95 (2009).
230 Kwon, C. *et al.* Canonical Wnt signaling is a positive regulator of mammalian cardiac progenitors. *Proc Natl Acad Sci U S A* **104**, 10894-10899, doi:0704044104 [pii]
10.1073/pnas.0704044104 (2007).
231 zur Nieden, N. I., Price, F. D., Davis, L. A., Everitt, R. E. & Rancourt, D. E. Gene profiling on mixed embryonic stem cell populations reveals a biphasic role for beta-catenin in osteogenic differentiation. *Mol Endocrinol* **21**, 674-685, doi:me.2005-0438 [pii]
10.1210/me.2005-0438 (2007).
232 Bradley, A., Evans, M., Kaufman, M. H. & Robertson, E. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**, 255-256 (1984).
233 Smith, A. G. Embryo-derived stem cells: of mice and men. *Annu Rev Cell Dev Biol* **17**, 435-462, doi:10.1146/annurev.cellbio.17.1.435

17/1/435 [pii] (2001).
234 Petersen, C. P. & Reddien, P. W. Wnt signaling and the polarity of the primary body axis. *Cell* **139**, 1056-1068, doi:S0092-8674(09)01493-7 [pii]
10.1016/j.cell.2009.11.035 (2009).
235 Giles, R. H., van Es, J. H. & Clevers, H. Caught up in a Wnt storm: Wnt signaling in cancer. *Biochim Biophys Acta* **1653**, 1-24, doi:S0304419X03000052 [pii] (2003).
236 Reya, T. & Clevers, H. Wnt signalling in stem cells and cancer. *Nature* **434**, 843-850, doi:nature03319 [pii]
10.1038/nature03319 (2005).
237 Fleming, H. E. *et al.* Wnt signaling in the niche enforces hematopoietic stem cell quiescence and is necessary to preserve self-renewal in vivo. *Cell Stem Cell* **2**, 274-283, doi:S1934-5909(08)00004-0 [pii]
10.1016/j.stem.2008.01.003 (2008).
238 Sato, N. *et al.* Molecular signature of human embryonic stem cells and its comparison with the mouse. *Dev Biol* **260**, 404-413, doi:S0012160603002562 [pii] (2003).
239 Hao, J., Li, T. G., Qi, X., Zhao, D. F. & Zhao, G. Q. WNT/beta-catenin pathway up-regulates Stat3 and converges on LIF to prevent differentiation of mouse embryonic stem cells. *Dev Biol* **290**, 81-91, doi:S0012-1606(05)00789-X [pii]
10.1016/j.ydbio.2005.11.011 (2006).
240 Storm, M. P. *et al.* Regulation of Nanog expression by phosphoinositide 3-kinase-dependent signaling in murine embryonic stem cells. *J Biol Chem* **282**, 6265-6273, doi:M610906200 [pii]
10.1074/jbc.M610906200 (2007).
241 Angers, S. & Moon, R. T. Proximal events in Wnt signal transduction. *Nat Rev Mol Cell Biol* **10**, 468-477, doi:nrm2717 [pii]
10.1038/nrm2717 (2009).
242 Metcalfe, C. & Bienz, M. Inhibition of GSK3 by Wnt signalling--two contrasting models. *J Cell Sci* **124**, 3537-3544, doi:124/21/3537 [pii]
10.1242/jcs.091991 (2011).
243 Yamaguchi, T. P., Takada, S., Yoshikawa, Y., Wu, N. & McMahon, A. P. T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. *Genes Dev* **13**, 3185-3190 (1999).
244 Takada, S. *et al.* Wnt-3a regulates somite and tailbud formation in the mouse embryo. *Genes Dev* **8**, 174-189 (1994).
245 Chen, M. *et al.* The anti-helminthic niclosamide inhibits Wnt/Frizzled1 signaling. *Biochemistry* **48**, 10267-10274, doi:10.1021/bi9009677 (2009).
246 Shan, B. E., Wang, M. X. & Li, R. Q. Quercetin inhibit human SW480 colon cancer growth in association with inhibition of cyclin D1 and survivin expression through Wnt/beta-catenin signaling pathway. *Cancer Invest* **27**, 604-612, doi:911184314 [pii]
10.1080/07357900802337191 (2009).
247 Vijayakumar, S. *et al.* High-frequency canonical Wnt activation in multiple sarcoma subtypes drives proliferation through a TCF/beta-catenin target gene, CDC25A. *Cancer Cell* **19**, 601-612, doi:S1535-6108(11)00120-6 [pii]
10.1016/j.ccr.2011.03.010 (2011).
248 Yun, M. S., Kim, S. E., Jeon, S. H., Lee, J. S. & Choi, K. Y. Both ERK and Wnt/beta-catenin pathways are involved in Wnt3a-induced proliferation. *J Cell Sci* **118**, 313-322, doi:jcs.01601 [pii]
10.1242/jcs.01601 (2005).

249 Kurimoto, K. *et al.* An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res* **34**, e42, doi:34/5/e42 [pii]
10.1093/nar/gkl050 (2006).
250 Seguin, C. A., Draper, J. S., Nagy, A. & Rossant, J. Establishment of endoderm progenitors by SOX transcription factor expression in human embryonic stem cells. *Cell Stem Cell* **3**, 182-195, doi:S1934-5909(08)00327-5 [pii]
10.1016/j.stem.2008.06.018 (2008).
251 Yoshikawa, Y., Fujimori, T., McMahon, A. P. & Takada, S. Evidence that absence of Wnt-3a signaling promotes neuralization instead of paraxial mesoderm development in the mouse. *Dev Biol* **183**, 234-242, doi:10.1006/dbio.1997.8502
S0012-1606(97)98502-X [pii] (1997).
252 Bielinska, M., Narita, N. & Wilson, D. B. Distinct roles for visceral endoderm during embryonic mouse development. *Int J Dev Biol* **43**, 183-205 (1999).
253 Plusa, B., Piliszek, A., Frankenberg, S., Artus, J. & Hadjantonakis, A. K. Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. *Development* **135**, 3081-3091, doi:135/18/3081 [pii]
10.1242/dev.021519 (2008).
254 Mansson, R. *et al.* Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. *Immunity* **26**, 407-419, doi:S1074-7613(07)00213-0 [pii]
10.1016/j.immuni.2007.02.013 (2007).
255 Lu, R., Yang, A. & Jin, Y. Dual functions of T-box 3 (Tbx3) in the control of self-renewal and extraembryonic endoderm differentiation in mouse embryonic stem cells. *J Biol Chem* **286**, 8425-8436, doi:M110.202150 [pii]
10.1074/jbc.M110.202150 (2011).
256 Shimosato, D., Shiki, M. & Niwa, H. Extra-embryonic endoderm cells derived from ES cells induced by GATA factors acquire the character of XEN cells. *BMC Dev Biol* **7**, 80, doi:1471-213X-7-80 [pii]
10.1186/1471-213X-7-80 (2007).
257 Cheng, A. M. *et al.* Mammalian Grb2 regulates multiple steps in embryonic development and malignant transformation. *Cell* **95**, 793-803, doi:S0092-8674(00)81702-X [pii] (1998).
258 Feldman, B., Poueymirou, W., Papaioannou, V. E., DeChiara, T. M. & Goldfarb, M. Requirement of FGF-4 for postimplantation mouse development. *Science* **267**, 246-249 (1995).
259 Arman, E., Haffner-Krausz, R., Chen, Y., Heath, J. K. & Lonai, P. Targeted disruption of fibroblast growth factor (FGF) receptor 2 suggests a role for FGF signaling in pregastrulation mammalian development. *Proc Natl Acad Sci U S A* **95**, 5082-5087 (1998).
260 Kunath, T. *et al.* FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* **134**, 2895-2902, doi:134/16/2895 [pii]
10.1242/dev.02880 (2007).
261 Yamanaka, Y., Lanner, F. & Rossant, J. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* **137**, 715-724, doi:137/5/715 [pii]
10.1242/dev.043471 (2010).
262 Hanna, J. *et al.* Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell Stem Cell* **4**, 513-524, doi:S1934-5909(09)00169-6 [pii]

10.1016/j.stem.2009.04.015 (2009).
263 Bakre, M. M. *et al.* Generation of multipotential mesendodermal progenitors from mouse embryonic stem cells via sustained Wnt pathway activation. *J Biol Chem* **282**, 31703-31712, doi:M704287200 [pii]
10.1074/jbc.M704287200 (2007).
264 Rubinfeld, B. *et al.* Binding of GSK3beta to the APC-beta-catenin complex and regulation of complex assembly. *Science* **272**, 1023-1026 (1996).
265 Niwa, H., Ogawa, K., Shimosato, D. & Adachi, K. A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature* **460**, 118-122, doi:nature08113 [pii]
10.1038/nature08113 (2009).
266 Renard, C. A. *et al.* Tbx3 is a downstream target of the Wnt/beta-catenin pathway and a critical mediator of beta-catenin survival functions in liver cancer. *Cancer Res* **67**, 901-910, doi:67/3/901 [pii]
10.1158/0008-5472.CAN-06-2344 (2007).
267 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111, doi:btp120 [pii]
10.1093/bioinformatics/btp120 (2009).
268 Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-515, doi:nbt.1621 [pii]
10.1038/nbt.1621 (2010).
269 Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **27**, 2325-2329, doi:btr355 [pii]
10.1093/bioinformatics/btr355 (2011).
270 Roberts, A., Trapnell, C., Donaghey, J., Rinn, J. L. & Pachter, L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol* **12**, R22, doi:gb-2011-12-3-r22 [pii]
10.1186/gb-2011-12-3-r22 (2011).
271 Muir, A. R., Kanji, A. H. & Allbrook, D. The structure of the satellite cells in skeletal muscle. *J Anat* **99**, 435-444 (1965).
272 Moss, F. P. & Leblond, C. P. Satellite cells as the source of nuclei in muscles of growing rats. *Anat Rec* **170**, 421-435, doi:10.1002/ar.1091700405 (1971).
273 Jostes, B., Walther, C. & Gruss, P. The murine paired box gene, Pax7, is expressed specifically during the development of the nervous and muscular system. *Mech Dev* **33**, 27-37 (1990).
274 Oustanina, S., Hause, G. & Braun, T. Pax7 directs postnatal renewal and propagation of myogenic satellite cells but not their specification. *EMBO J* **23**, 3430-3439, doi:10.1038/sj.emboj.7600346
7600346 [pii] (2004).
275 Relaix, F. *et al.* Pax3 and Pax7 have distinct and overlapping functions in adult muscle progenitor cells. *J Cell Biol* **172**, 91-102, doi:jcb.200508044 [pii]
10.1083/jcb.200508044 (2006).
276 Kuang, S., Charge, S. B., Seale, P., Huh, M. & Rudnicki, M. A. Distinct roles for Pax7 and Pax3 in adult regenerative myogenesis. *J Cell Biol* **172**, 103-113, doi:jcb.200508001 [pii]
10.1083/jcb.200508001 (2006).
277 Hilfer, S. R., Searls, R. L. & Fonte, V. G. An ultrastructural study of early myogenesis in the chick wing bud. *Dev Biol* **30**, 374-391, doi:0012-1606(73)90095-X [pii] (1973).

278 Sweeney, L. J., Kennedy, J. M., Zak, R., Kokjohn, K. & Kelley, S. W. Evidence for expression of a common myosin heavy chain phenotype in future fast and slow skeletal muscle during initial stages of avian embryogenesis. *Dev Biol* **133**, 361-374, doi:0012-1606(89)90040-7 [pii] (1989).
279 Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol* **22**, 1567-1572, doi:nbt1037 [pii]
10.1038/nbt1037 (2004).
280 Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. A guide to choosing fluorescent proteins. *Nat Methods* **2**, 905-909, doi:nmeth819 [pii]
10.1038/nmeth819 (2005).
281 Paques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiol Mol Biol Rev* **63**, 349-404 (1999).
282 Stern, C. Somatic Crossing over and Segregation in Drosophila Melanogaster. *Genetics* **21**, 625-730 (1936).
283 Schatz, D. G. & Ji, Y. Recombination centres and the orchestration of V(D)J recombination. *Nat Rev Immunol* **11**, 251-263, doi:nri2941 [pii]
10.1038/nri2941 (2011).
284 Thomas, K. R. & Capecchi, M. R. Introduction of homologous DNA sequences into mammalian cells induces mutations in the cognate gene. *Nature* **324**, 34-38, doi:10.1038/324034a0 (1986).
285 Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A. & Kucherlapati, R. S. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature* **317**, 230-234 (1985).
286 O'Connor, M., Peifer, M. & Bender, W. Construction of large DNA segments in Escherichia coli. *Science* **244**, 1307-1312 (1989).
287 Court, D. L., Sawitzke, J. A. & Thomason, L. C. Genetic engineering using homologous recombination. *Annu Rev Genet* **36**, 361-388, doi:10.1146/annurev.genet.36.061102.093104
061102.093104 [pii] (2002).
288 Zhang, Y., Buchholz, F., Muyrers, J. P. & Stewart, A. F. A new logic for DNA engineering using recombination in Escherichia coli. *Nat Genet* **20**, 123-128, doi:10.1038/2417 (1998).
289 Muyrers, J. P., Zhang, Y., Testa, G. & Stewart, A. F. Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res* **27**, 1555-1557, doi:gkc282 [pii] (1999).
290 Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-6645, doi:10.1073/pnas.120163297
120163297 [pii] (2000).
291 Stahl, F. W. Recombination in phage lambda: one geneticist's historical perspective. *Gene* **223**, 95-102 (1998).
292 Poteete, A. R. What makes the bacteriophage lambda Red system useful for genetic engineering: molecular mechanism and biological function. *FEMS Microbiol Lett* **201**, 9-14, doi:S0378-1097(01)00242-7 [pii] (2001).
293 Yu, D. *et al.* An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A* **97**, 5978-5983, doi:10.1073/pnas.100127597
100127597 [pii] (2000).

294 Lee, E. C. *et al.* A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**, 56-65, doi:10.1006/geno.2000.6451
S0888-7543(00)96451-6 [pii] (2001).
295 Kadesch, T. & Berg, P. Effects of the position of the simian virus 40 enhancer on expression of multiple transcription units in a single plasmid. *Mol Cell Biol* **6**, 2593-2601 (1986).
296 Proudfoot, N. J. Transcriptional interference and termination between duplicated alpha-globin gene constructs suggests a novel mechanism for gene regulation. *Nature* **322**, 562-565, doi:10.1038/322562a0 (1986).
297 Yahata, K. *et al.* cHS4 insulator-mediated alleviation of promoter interference during cell-based expression of tandemly associated transgenes. *J Mol Biol* **374**, 580-590, doi:S0022-2836(07)01243-0 [pii]
10.1016/j.jmb.2007.09.054 (2007).
298 Ryan, M. D., King, A. M. & Thomas, G. P. Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *J Gen Virol* **72 ( Pt 11)**, 2727-2732 (1991).
299 Palmenberg, A. C. *et al.* Proteolytic processing of the cardioviral P2 region: primary 2A/2B cleavage in clone-derived precursors. *Virology* **190**, 754-762 (1992).
300 Donnelly, M. L. *et al.* The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. *J Gen Virol* **82**, 1027-1041 (2001).
301 Szymczak, A. L. *et al.* Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nat Biotechnol* **22**, 589-594, doi:10.1038/nbt957
nbt957 [pii] (2004).
302 Maquat, L. E. Nonsense-mediated mRNA decay. *Curr Biol* **12**, R196-197, doi:S0960982202007479 [pii] (2002).
303 Schwenk, F., Baron, U. & Rajewsky, K. A cre-transgenic mouse strain for the ubiquitous deletion of loxP-flanked gene segments including deletion in germ cells. *Nucleic Acids Res* **23**, 5080-5081, doi:l50367 [pii] (1995).
304 Gu, H., Zou, Y. R. & Rajewsky, K. Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. *Cell* **73**, 1155-1164, doi:0092-8674(93)90644-6 [pii] (1993).
305 Peitz, M., Pfannkuche, K., Rajewsky, K. & Edenhofer, F. Ability of the hydrophobic FGF and basic TAT peptides to promote cellular uptake of recombinant Cre recombinase: a tool for efficient genetic engineering of mammalian genomes. *Proc Natl Acad Sci U S A* **99**, 4489-4494, doi:10.1073/pnas.032068699
032068699 [pii] (2002).
306 Tannous, B. A., Kim, D. E., Fernandez, J. L., Weissleder, R. & Breakefield, X. O. Codon-optimized Gaussia luciferase cDNA for mammalian gene expression in culture and in vivo. *Mol Ther* **11**, 435-443, doi:S1525-0016(04)01508-4 [pii]
10.1016/j.ymthe.2004.10.016 (2005).
307 Nguyen, A. W. & Daugherty, P. S. Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat Biotechnol* **23**, 355-360, doi:nbt1066 [pii]
10.1038/nbt1066 (2005).
308 Tannous, B. A. Gaussia luciferase reporter assay for monitoring biological processes in culture and in vivo. *Nat Protoc* **4**, 582-591, doi:nprot.2009.28 [pii]

10.1038/nprot.2009.28 (2009).
309 Blau, H. M. *et al.* Plasticity of the differentiated state. *Science* **230**, 758-766 (1985).
310 Van Keuren, M. L., Gavrilina, G. B., Filipiak, W. E., Zeidler, M. G. & Saunders, T. L. Generating transgenic mice from bacterial artificial chromosomes: transgenesis efficiency, integration and expression outcomes. *Transgenic Res* **18**, 769-785, doi:10.1007/s11248-009-9271-2 (2009).
311 Carvajal, J. J., Keith, A. & Rigby, P. W. Global transcriptional regulation of the locus encoding the skeletal muscle determination genes Mrf4 and Myf5. *Genes Dev* **22**, 265-276, doi:22/2/265 [pii]
10.1101/gad.442408 (2008).
312 Liu, P., Jenkins, N. A. & Copeland, N. G. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res* **13**, 476-484, doi:10.1101/gr.749203 (2003).
313 Parks, R. J., Bramson, J. L., Wan, Y., Addison, C. L. & Graham, F. L. Effects of stuffer DNA on transgene expression from helper-dependent adenovirus vectors. *J Virol* **73**, 8027-8034 (1999).
314 Huh, M. S., Parker, M. H., Scime, A., Parks, R. & Rudnicki, M. A. Rb is required for progression through myogenic differentiation but not maintenance of terminal differentiation. *J Cell Biol* **166**, 865-876, doi:10.1083/jcb.200403004
jcb.200403004 [pii] (2004).
315 Zammit, P. S. *et al.* Muscle satellite cells adopt divergent fates: a mechanism for self-renewal? *J Cell Biol* **166**, 347-357, doi:10.1083/jcb.200312007
jcb.200312007 [pii] (2004).
316 Carlson, M. E. & Conboy, I. M. Loss of stem cell regenerative capacity within aged niches. *Aging Cell* **6**, 371-382, doi:ACE286 [pii]
10.1111/j.1474-9726.2007.00286.x (2007).
317 Augustin, H. & Partridge, L. Invertebrate models of age-related muscle degeneration. *Biochim Biophys Acta* **1790**, 1084-1094, doi:S0304-4165(09)00177-9 [pii]
10.1016/j.bbagen.2009.06.011 (2009).
318 Cruz-Jentoft, A. J. *et al.* Sarcopenia: European consensus on definition and diagnosis: Report of the European Working Group on Sarcopenia in Older People. *Age Ageing* **39**, 412-423, doi:afq034 [pii]
10.1093/ageing/afq034 (2010).
319 Zacks, S. I. & Sheff, M. F. Age-related impeded regeneration of mouse minced anterior tibial muscle. *Muscle Nerve* **5**, 152-161, doi:10.1002/mus.880050213 (1982).
320 Sacco, A., Doyonnas, R., Kraft, P., Vitorovic, S. & Blau, H. M. Self-renewal and expansion of single transplanted muscle stem cells. *Nature* **456**, 502-506, doi:nature07384 [pii]
10.1038/nature07384 (2008).
321 Harel, I. *et al.* Distinct origins and genetic programs of head muscle satellite cells. *Dev Cell* **16**, 822-832, doi:S1534-5807(09)00209-3 [pii]
10.1016/j.devcel.2009.05.007 (2009).
322 Sambasivan, R. *et al.* Distinct regulatory cascades govern extraocular and pharyngeal arch muscle progenitor cell fates. *Dev Cell* **16**, 810-821, doi:S1534-5807(09)00210-X [pii]
10.1016/j.devcel.2009.05.008 (2009).
323 Schultz, E. Satellite cell proliferative compartments in growing skeletal muscles. *Dev Biol* **175**, 84-94, doi:S0012-1606(96)90097-4 [pii]
10.1006/dbio.1996.0097 (1996).

324 Shinin, V., Gayraud-Morel, B., Gomes, D. & Tajbakhsh, S. Asymmetric division and cosegregation of template DNA strands in adult muscle satellite cells. *Nat Cell Biol* **8**, 677-687, doi:ncb1425 [pii]
10.1038/ncb1425 (2006).
325 Conboy, I. M. & Rando, T. A. The regulation of Notch signaling controls satellite cell activation and cell fate determination in postnatal myogenesis. *Dev Cell* **3**, 397-409, doi:S153458070200254X [pii] (2002).
326 Sherwood, R. I. *et al.* Isolation of adult mouse myogenic progenitors: functional heterogeneity of cells within and engrafting skeletal muscle. *Cell* **119**, 543-554, doi:S0092867404010360 [pii]
10.1016/j.cell.2004.10.021 (2004).
327 Biressi, S. *et al.* Intrinsic phenotypic diversity of embryonic and fetal myoblasts is revealed by genome-wide gene expression analysis on purified cells. *Dev Biol* **304**, 633-651, doi:S0012-1606(07)00032-2 [pii]
10.1016/j.ydbio.2007.01.016 (2007).
328 Gussoni, E. *et al.* Normal dystrophin transcripts detected in Duchenne muscular dystrophy patients after myoblast transplantation. *Nature* **356**, 435-438, doi:10.1038/356435a0 (1992).
329 Huard, J. *et al.* Human myoblast transplantation: preliminary results of 4 cases. *Muscle Nerve* **15**, 550-560, doi:10.1002/mus.880150504 (1992).
330 Karpati, G. *et al.* Myoblast transfer in Duchenne muscular dystrophy. *Ann Neurol* **34**, 8-17, doi:10.1002/ana.410340105 (1993).
331 Skuk, D. *et al.* Dystrophin expression in muscles of duchenne muscular dystrophy patients after high-density injections of normal myogenic cells. *J Neuropathol Exp Neurol* **65**, 371-386, doi:10.1097/01.jnen.0000218443.45782.81
00005072-200604000-00007 [pii] (2006).
332 Price, F. D., Kuroda, K. & Rudnicki, M. A. Stem cell based therapies to treat muscular dystrophy. *Biochim Biophys Acta* **1772**, 272-283, doi:S0925-4439(06)00173-6 [pii]
10.1016/j.bbadis.2006.08.011 (2007).
333 Lepper, C., Partridge, T. A. & Fan, C. M. An absolute requirement for Pax7-positive satellite cells in acute injury-induced skeletal muscle regeneration. *Development* **138**, 3639-3646, doi:138/17/3639 [pii]
10.1242/dev.067595 (2011).
334 Sambasivan, R. *et al.* Pax7-expressing satellite cells are indispensable for adult skeletal muscle regeneration. *Development* **138**, 3647-3656, doi:138/17/3647 [pii]
10.1242/dev.067587 (2011).
335 Ieronimakis, N. *et al.* Absence of CD34 on murine skeletal muscle satellite cells marks a reversible state of activation during acute injury. *PLoS One* **5**, e10920, doi:10.1371/journal.pone.0010920 (2010).
336 LaBarge, M. A. & Blau, H. M. Biological progression from adult bone marrow to mononucleate muscle stem cell to multinucleate muscle fiber in response to injury. *Cell* **111**, 589-601, doi:S0092867402010784 [pii] (2002).
337 Cerletti, M. *et al.* Highly efficient, functional engraftment of skeletal muscle stem cells in dystrophic muscles. *Cell* **134**, 37-47, doi:S0092-8674(08)00755-1 [pii]
10.1016/j.cell.2008.05.049 (2008).
338 Donalies, M., Cramer, M., Ringwald, M. & Starzinski-Powitz, A. Expression of M-cadherin, a member of the cadherin multigene family, correlates with differentiation of skeletal muscle cells. *Proc Natl Acad Sci U S A* **88**, 8024-8028 (1991).

339 Moore, R. & Walsh, F. S. The cell adhesion molecule M-cadherin is specifically expressed in developing and regenerating, but not denervated skeletal muscle. *Development* **117**, 1409-1420 (1993).
340 Kuang, S., Gillespie, M. A. & Rudnicki, M. A. Niche regulation of muscle satellite cell self-renewal and differentiation. *Cell Stem Cell* **2**, 22-31, doi:S1934-5909(07)00328-1 [pii]
10.1016/j.stem.2007.12.012 (2008).
341 Zeschnigk, M., Kozian, D., Kuch, C., Schmoll, M. & Starzinski-Powitz, A. Involvement of M-cadherin in terminal differentiation of skeletal muscle cells. *J Cell Sci* **108 ( Pt 9)**, 2973-2981 (1995).
342 Miller, R. G. *et al.* Myoblast implantation in Duchenne muscular dystrophy: the San Francisco study. *Muscle Nerve* **20**, 469-478, doi:10.1002/(SICI)1097-4598(199704)20:4<469::AID-MUS10>3.0.CO;2-U [pii] (1997).
343 Palmieri, B., Tremblay, J. P. & Daniele, L. Past, present and future of myoblast transplantation in the treatment of Duchenne muscular dystrophy. *Pediatr Transplant* **14**, 813-819, doi:10.1111/j.1399-3046.2010.01377.x (2010).
344 Weydert, A., Barton, P., Harris, A. J., Pinset, C. & Buckingham, M. Developmental pattern of mouse skeletal myosin heavy chain gene transcripts in vivo and in vitro. *Cell* **49**, 121-129, doi:0092-8674(87)90762-8 [pii] (1987).
345 Beggs, M. L. *et al.* Alterations in the TGFbeta signaling pathway in myogenic progenitors with age. *Aging Cell* **3**, 353-361, doi:ACE135 [pii]
10.1111/j.1474-9728.2004.00135.x (2004).
346 Edwards, M. G. *et al.* Gene expression profiling of aging reveals activation of a p53-mediated transcriptional program. *BMC Genomics* **8**, 80, doi:1471-2164-8-80 [pii]
10.1186/1471-2164-8-80 (2007).
347 Gibson, M. C. & Schultz, E. Age-related differences in absolute numbers of skeletal muscle satellite cells. *Muscle Nerve* **6**, 574-580, doi:10.1002/mus.880060807 (1983).
348 Collins, C. A., Zammit, P. S., Ruiz, A. P., Morgan, J. E. & Partridge, T. A. A population of myogenic stem cells that survives skeletal muscle aging. *Stem Cells* **25**, 885-894, doi:2006-0372 [pii]
10.1634/stemcells.2006-0372 (2007).
349 Brack, A. S., Bildsoe, H. & Hughes, S. M. Evidence that satellite cell decrement contributes to preferential decline in nuclear number from large fibres during murine age-related muscle atrophy. *J Cell Sci* **118**, 4813-4821, doi:118/20/4813 [pii]
10.1242/jcs.02602 (2005).
350 Shefer, G., Rauner, G., Yablonka-Reuveni, Z. & Benayahu, D. Reduced satellite cell numbers and myogenic capacity in aging can be alleviated by endurance exercise. *PLoS One* **5**, e13307, doi:e13307 [pii]
10.1371/journal.pone.0013307 (2010).
351 Nishikawa, S. I., Osawa, M., Yonetani, S., Torikai-Nishikawa, S. & Freter, R. Niche required for inducing quiescent stem cells. *Cold Spring Harb Symp Quant Biol* **73**, 67-71, doi:sqb.2008.73.024 [pii]
10.1101/sqb.2008.73.024 (2008).
352 Abou-Khalil, R. & Brack, A. S. Muscle stem cells and reversible quiescence: the role of sprouty. *Cell Cycle* **9**, 2575-2580, doi:12149 [pii]
10.4161/cc.9.13.12149 (2010).
353 Shea, K. L. *et al.* Sprouty1 regulates reversible quiescence of a self-renewing adult muscle stem cell pool during regeneration. *Cell Stem Cell* **6**, 117-129, doi:S1934-5909(09)00638-9 [pii]

10.1016/j.stem.2009.12.015 (2010).
354 Grobet, L. *et al.* A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. *Nat Genet* **17**, 71-74, doi:10.1038/ng0997-71 (1997).
355 Bulfield, G., Siller, W. G., Wight, P. A. & Moore, K. J. X chromosome-linked muscular dystrophy (mdx) in the mouse. *Proc Natl Acad Sci U S A* **81**, 1189-1192 (1984).
356 Sicinski, P. *et al.* The molecular basis of muscular dystrophy in the mdx mouse: a point mutation. *Science* **244**, 1578-1580 (1989).
357 Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13, doi:gkn923 [pii]
10.1093/nar/gkn923 (2009).
358 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57, doi:nprot.2008.211 [pii]
10.1038/nprot.2008.211 (2009).
359 Tajbakhsh, S. *et al.* Differential activation of Myf5 and MyoD by different Wnts in explants of mouse paraxial mesoderm and the later activation of myogenesis in the absence of Myf5. *Development* **125**, 4155-4162 (1998).
360 Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519-523, doi:nature06968 [pii]
10.1038/nature06968 (2008).
361 Chen, Y., Du, Z. & Yao, Z. Roles of the Nanog protein in murine F9 embryonal carcinoma cells and their endoderm-differentiated counterparts. *Cell Res* **16**, 641-650, doi:7310067 [pii]
10.1038/sj.cr.7310067 (2006).
362 Lickert, H. *et al.* Wnt/(beta)-catenin signaling regulates the expression of the homeobox gene Cdx1 in embryonic intestine. *Development* **127**, 3805-3813 (2000).
363 Rousso, S. Z. *et al.* Negative autoregulation of Oct3/4 through Cdx1 promotes the onset of gastrulation. *Dev Dyn* **240**, 796-807, doi:10.1002/dvdy.22588 (2011).
364 Kim, I., Saunders, T. L. & Morrison, S. J. Sox17 dependence distinguishes the transcriptional regulation of fetal from adult hematopoietic stem cells. *Cell* **130**, 470-483, doi:S0092-8674(07)00777-5 [pii]
10.1016/j.cell.2007.06.011 (2007).
365 Rahmani, M. *et al.* Versican: signaling to transcriptional control pathways. *Can J Physiol Pharmacol* **84**, 77-92, doi:10.1139/y05-154 (2006).
366 Henderson, D. J., Ybot-Gonzalez, P. & Copp, A. J. Over-expression of the chondroitin sulphate proteoglycan versican is associated with defective neural crest migration in the Pax3 mutant mouse (splotch). *Mech Dev* **69**, 39-51 (1997).
367 Mayanil, C. S. *et al.* Microarray analysis detects novel Pax3 downstream target genes. *J Biol Chem* **276**, 49299-49309, doi:10.1074/jbc.M107933200
M107933200 [pii] (2001).
368 Timpl, R. Proteoglycans of basement membranes. *Experientia* **49**, 417-428 (1993).
369 Filmus, J., Capurro, M. & Rast, J. Glypicans. *Genome Biol* **9**, 224, doi:gb-2008-9-5-224 [pii]
10.1186/gb-2008-9-5-224 (2008).
370 Capurro, M. I., Xiang, Y. Y., Lobe, C. & Filmus, J. Glypican-3 promotes the growth of hepatocellular carcinoma by stimulating canonical Wnt signaling. *Cancer Res* **65**, 6245-6254, doi:65/14/6245 [pii]

10.1158/0008-5472.CAN-04-4244 (2005).
371 Wenk, M. B., Midwood, K. S. & Schwarzbauer, J. E. Tenascin-C suppresses Rho activation. *J Cell Biol* **150**, 913-920 (2000).
372 Huang, W., Chiquet-Ehrismann, R., Moyano, J. V., Garcia-Pardo, A. & Orend, G. Interference of tenascin-C with syndecan-4 binding to fibronectin blocks cell adhesion and stimulates tumor cell proliferation. *Cancer Res* **61**, 8586-8594 (2001).
373 Garcion, E., Halilagic, A., Faissner, A. & ffrench-Constant, C. Generation of an environmental niche for neural stem cell development by the extracellular matrix molecule tenascin C. *Development* **131**, 3423-3432, doi:10.1242/dev.01202
131/14/3423 [pii] (2004).
374 Kii, I. *et al.* Incorporation of tenascin-C into the extracellular matrix by periostin underlies an extracellular meshwork architecture. *J Biol Chem* **285**, 2028-2039, doi:M109.051961 [pii]
10.1074/jbc.M109.051961 (2010).
375 Chiariello, M., Marinissen, M. J. & Gutkind, J. S. Regulation of c-myc expression by PDGF through Rho GTPases. *Nat Cell Biol* **3**, 580-586, doi:10.1038/35078555
35078555 [pii] (2001).
376 Scime, A. *et al.* Transcriptional profiling of skeletal muscle reveals factors that are necessary to maintain satellite cell integrity during ageing. *Mech Ageing Dev* **131**, 9-20, doi:S0047-6374(09)00151-1 [pii]
10.1016/j.mad.2009.11.001 (2010).
377 Emery, A. E. Population frequencies of inherited neuromuscular diseases--a world survey. *Neuromuscul Disord* **1**, 19-29 (1991).
378 Murphy, M. M., Lawson, J. A., Mathew, S. J., Hutcheson, D. A. & Kardon, G. Satellite cells, connective tissue fibroblasts and their interactions are crucial for muscle regeneration. *Development* **138**, 3625-3637, doi:138/17/3625 [pii]
10.1242/dev.064162 (2011).