The role of the homeobox transcription factor Duxbl in rhabdomyosarcoma formation

INAUGURAL-DISSERTATION

Zur Erlangung des Doktorgrades der Naturwissenschaften

-Doctor rerum naturalium-

(Dr. rer. nat.)

vorgelegt dem

Fachbereich 08 – Biologie und Chemie der

Justus-Liebig-Universität Gießen

Durchgeführt am Max-Planck-Institut für Herz-und Lungenforschung,

W.G. Kerckhoff-Institut Bad Nauheim

vorgelegt von

Jiasheng Zhong

aus Shenzhen, China

Gießen (2021)

Die Untersuchungen zur vorliegenden Arbeit wurden am Max-Planck-Institut für Herz-und Lungenforschung (W. G. Kerckhoff-Institut) in Bad Nauheim unter Leitung von Prof. Dr. Dr. Thomas Braun durchgeführt.

Vom Fachbereich 08-Biologie und Chemie der Justus-Liebig-Universität Gießen angenommen

Erstgutachter:

Prof. Dr. Dr. Thomas Braun

Abteilung für Entwicklung und Umbau des Herzens

Max-Planck-Institut für Herz-und Lungenforschung

Ludwigstraße 43D-61231 Bad Nauheim

Zweitgutachter:

Prof. Dr. M. Lienhard Schmitz

Justus-Liebig-Universität Gießen

Biochemisches Institut

Friedrichstrasse 24-35392 Gießen

Declaration

"I declare that I have completed this dissertation single-handedly without the unauthorized help of a second party and only with the assistance acknowledged therein. I have appropriately acknowledged and referenced all text passages that are derived literally from or are based on the content of published or unpublished work of others, and all information that relates to verbal communications. I have abided by the principles of good scientific conduct laid down in the charter of the Justus Liebig University of Giessen in carrying out the investigations described in the dissertation."

"Part of results described in this dissertation has been published, in which Dr. Jens Preussner and I contributed equally and shared equal authorship to the published paper (Preussner et al., 2018). Dr. Jens Preussner conducted the bioinformatical analyses and provided the visualized data, and I conducted related cellular, animal and biochemical experiments. All results related to this published paper have been cited and described with modifications in this dissertation. Particularly, Dr. Jens Preussner's contribution to this dissertation has been notified in figure legends."

Bad Nauheim Jiasheng, Zhong

I Table of contents.

II Summary.

Skeletal muscles exhibit a remarkable regeneration capacity that is mediated by resident muscle stem cells (MuSCs). While normally quiescent they become activated upon injury, proliferate and differentiate into myofibers to repair damaged skeletal muscle. Rhabdomyosarcoma (RMS) tumors show features of skeletal muscle differentiation. Previous studies revealed that RMS tumors form at high incidence in mdx mice that undergo chronic muscle regeneration. However, unequivocal proofs on the cellular origin of these RMS tumors are lacking. Consequently, the underlying molecular mechanism of how RMS tumors are formed remain poorly understood.

In this study, I show that MuSC-specific deletion of the tumor suppressor *p53* is sufficient to induce formation of fluorescently lineage traced RMS tumors in conditions of chronic muscle regeneration when MuSCs are activated. The genetic lineage tracing approach demonstrated that MuSCs are the origin of RMS tumors in these mice. Importantly, the tracing approach also enabled separation of cells from RMS tumors that are either derived from transformed MuSC descendants or from stromal cells of non-MuSC origin. Whole-exome sequencing performed on these purified tumor cell populations from different tumors revealed discrete oncogenic copy number amplifications (CNA) associated with MuSC-derived RMS tumor formation.

I performed a comprehensive analysis on the function of a poorly investigated gene encoding for a putative transcriptional regulator called *Duxbl*, because this gene was recurrently amplified in many of the exome-sequenced RMS tumors. Inactivation of *Duxbl* in tumor cells harboring a *Duxbl* CNA resulted in cell death, indicating that *Duxbl* plays an important role in both promotion and maintenance of *Duxbl*-associated RMS tumors. In contrast, overexpression of *Duxbl* in MuSCs resulted in enhanced proliferation and inhibited myogenic differentiation.

I generated a mouse model enabling conditional overexpression of *Duxbl* and found that induction of *Duxbl* in *p53*-deficient MuSCs is sufficient to induce RMS tumor formation *in vivo*, proving that *Duxbl* can function as an oncogene. I discovered that overexpression of *Duxbl* in conjunction with *p53* inactivation induces expression of endogenous retrotransposons of the *ERVL* subclass and that *ERVL* is specially re-activated in *Duxbl*-induced RMS tumors. This study identifies *Duxbl* as a novel oncogene that is able to promote RMS tumor formation in *p53*-deficient MuSCs, probably by activation of *ERVL*.

Keywords: MuSCs, RMS, tumor propagating cells, *Duxbl*, *ERVL*

III Zusammenfassung.

Die Skelettmuskulatur verfügt über eine bemerkenswerte Regenerationsfähigkeit, die durch residente Muskelstammzellen (MuSCs) vermittelt wird. Im normalen Ruhezustand werden sie als Reaktion auf eine Verletzung aktiviert, proliferieren und differenzieren zu Muskelfasern, um geschädigte Skelettmuskeln zu reparieren. Rhabdomyosarkom (RMS)-Tumore enthalten Zellen, die Merkmale von Skelettmuskeldifferenzierung aufweisen. Vorangehende Studien haben eine hohe Inzidenz von RMS-Tumorbildung in mdx-Mäusen gezeigt, die durch chronischen Muskelregeneration charakterisiert sind. Es fehlten jedoch eindeutige Beweise für den zellulären Ursprung dieser RMS-Tumoren. Die molekularen Mechanismen, welche der RMS-Tumorbildung zugrundeliegende sind nach wie vor nur unzureichend verstanden.

In dieser Studie habe ich herausgefunden, dass bei der Aktivierung von MuSCs die muskelspezifische Deletion des Tumor Suppressors *p53* ausreicht, um eine RMS-Tumorbildung unter chronischen Muskelregenerationsbedingungen zu induzieren. Durch genetisches Lineage-Tracing konnte ich beweisen, dass MuSCs der Ursprung von RMS-Tumoren in diesen Mäusen waren. Zudem ermöglichte der Tracing Ansatz, Zellen aus RMS-Tumoren zu isolieren, die von transformierten MuSC-Nachkommen abstammen, und solche von Zellen mit nicht-MuSC-Ursprung zu unterscheiden. Die Sequenzierung der genomischen DNA gereinigter Tumorzellpopulationen aus verschiedenen Tumoren ergab, dass verschiedene diskrete Amplifikationen genomischer Bereiche mit der Bildung von RMS-Tumoren assoziiert ist.

Ich analysierte umfassend die Funktion eines bisher wenig untersuchten Gens, das für *Duxbl* kodiert und in RMS-Tumoren häufig amplifiziert ist. Man nimmt an, dass es sich bei *Duxbl* um einen Transkriptionsregulator handelt. In Tumorzellen, die *Duxbl* CNA enthalten, führte die Inaktivierung von *Duxbl* zum Zelltod, was darauf hindeutet, dass *Duxbl* eine wichtige Rolle bei der Förderung und Aufrechterhaltung von *Duxbl*-assoziierten RMS-Tumoren spielt. Umgekehrt erhöht die Überexpression von *Duxbl* in MuSCs die Proliferation und hemmt die myogene Differenzierung.

Ich habe ein Mausmodell generiert, in welchem *Duxbl* konditional überexpremiert werden kann. Es zeigte sich, dass die Induktion der *Duxbl* Expression in *p53*-defizienten MuSCs ausreichend war, um die Bildung von RMS-Tumoren *in vivo* zu induzieren. Dies beweist, dass *Duxbl* als Onkogen wirken kann. Weiterhin fand ich heraus, dass die

Überexpression von *Duxbl* und die Inaktivierung von *p53* die Expression der endogenen *ERVL*-Subklasse von Retrotransposons induziert und dass *ERVL* spezifisch in *Duxbl*-induzierten RMS-Tumoren aktiviert ist. Die vorliegende Studie beweist, dass *Duxbl* ein neuartiges Onkogen ist, das die RMS-Tumorbildung in *p53*-defizienten MuSCs fördert, möglicherweise durch Induktion der Aktivierung von *ERVL*.

Schlüsselwörter: MuSCs, RMS, Tumor-Vermehrungszellen, *Duxbl*, *ERVL*

1. Introduction.

1.1 Skeletal muscles.

1.1.1 Structures and functions of skeletal muscles.

Skeletal muscles are one of the most dynamic and plastic tissues of the human body that contains approximately 600 individual muscles (Chal and Pourquié, 2017b), which together make up roughly 45 % of the total body weight (Frontera and Ochala, 2015). An individual skeletal muscle comprises hundreds or thousands of multinucleated muscle fibers bundled together and wrapped in connective tissue coverings called epimysium. Bundles of muscle fibers are organized into so called fascicles that are surrounded by perimysium (Capers, 1960; Frontera and Ochala, 2015). This organization enables the nervous system to activate a subset of muscle fibers thereby triggering specific body movements (Pham and Puckett, 2020). Each muscle fiber is encased in a thin connective tissue layer of collagen and reticular fibers known as endomysium, which supports force transmission during muscle contractions (Purslow, 2008). The endomysium also contains extracellular fluid and nutrients to support the metabolic demands of muscle fibers (Purslow, 2008). Collectively, the epimysium, perimysium, and endomysium allows skeletal muscles to form a ropelike tendon or flat sheet-like aponeurosis, attaching skeletal muscles with movable structures such as bones, cartilages ligaments, and fibrous membranes.

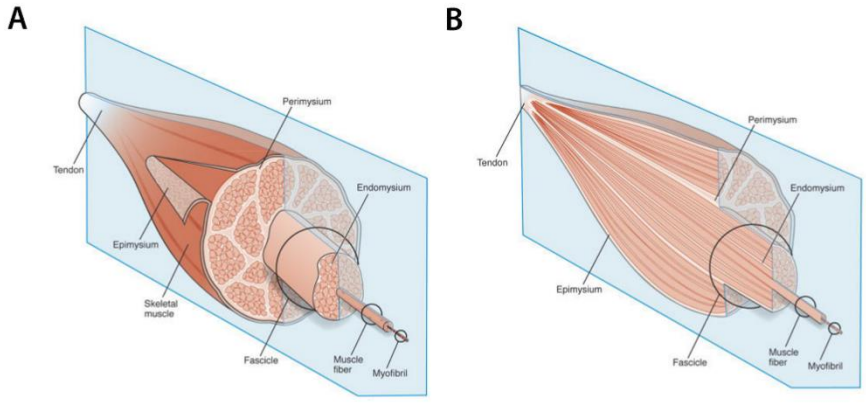

Figure 1. Schematic diagram of skeletal muscle tissue.

(A) Skeletal muscles are organized within extracellular matrix consisting of epimysium (surrounding the muscle), perimysium (surrounding muscle fascicles), and endomysium (surrounding muscle fibers). (B) Longitudinal cross-section of a skeletal muscle. The perimysium is connected to the tendon, whereas endomysium is contained within muscle fascicles (Gillies and Lieber, 2011).

1.1.2 Skeletal muscle development.

The formation of skeletal muscles during embryonic development can be divided into three main phases (Buckingham et al., 2003). In the first two phases, myoblasts are formed within anatomically discrete locations of the developing organism, while in the third phase the muscle grows through fusion of myoblasts. These processes are tightly controlled by timely expression of muscle-specific proteins (Buckingham et al., 2003; Fiorotto, 2012).

The first embryonic myogenic phase occurs when the dermomyotome is formed from the paraxial mesoderm (Aulehla and Pourquié, 2010). At this stage, future myoblasts delaminate from the paraxial mesoderm that organize into aggregates of epithelial cells called somites on either side of the neural tube and notochord. The progressive formation of somites proceeds in a head-to-tail arrangement (termed the segmentation clock) (Fiorotto, 2012). At this stage, a posterior to anterior gradient of the fibroblast growth factor (FGF) and activation of Wnt/β-catenin signaling maintains presomitic mesodermal cells in the undifferentiated state (Volckaert and De Langhe, 2015). Antagonism by retinoic acid (vitamin A) signaling leads to differentiation of the somites (Lamarche et al., 2015). The dorsal surface of the somites then forms the dermomyotome which contains muscle lineage progenitor cells (Buckingham et al., 2003).

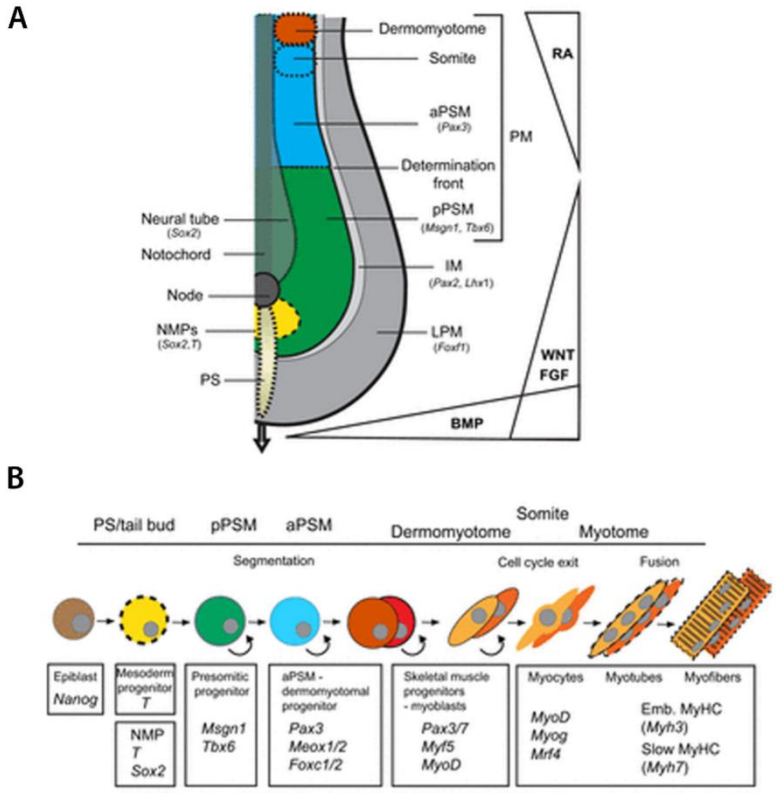

Figure 2. Embryonic development of skeletal muscles

(A) Spatial organization of mesoderm fate in the posterior region of an amniotes embryo. Mesoderm forms by ingression of epiblast cells at the level of the primitive streak (PS). Mesoderm subtypes (color-coded) are distinguished by their mediolateral position, whereby the axial mesoderm corresponds to the notochord. The lateral domains of the paraxial mesoderm (PM), intermediate mesoderm (IM) and lateral plate mesoderm (LPM), and the corresponding marker genes are shown. BMP, Wnt, FGF and retinoic acid (RA) signaling factors are distributed in gradients as shown to the right.

(B) Diagram recapitulating the differentiation of paraxial mesoderm toward skeletal muscles.

Adapted from (Chal and Pourquié, 2017a).

In the second phase, muscle lineage progenitor cells that express the myogenic homeobox transcription factors *Pax3* and *Pax7* differentiate into skeletal myoblasts (Buckingham and Relaix, 2007; Hutcheson et al., 2009; Tajbakhsh, 2009). The commitment of muscle lineage progenitors to become muscle precursor cells or myoblasts is initiated by highly regulated expression of myogenic regulatory factors (*MRFs*), transcription factors which include basic helix–loop–helix (bHLH) transcription factors *Myf5*, *MyoD*, *MyoG* and *Mrf4* (Hutcheson et al., 2009; Tajbakhsh, 2009). *Wnts* and Sonic hedgehog (Shh) promote expression of *MRFs* while negative regulators like the bone morphogenetic proteins (BMPs) and the Notch

signaling pathway inhibit *MRFs* expression and myogenic differentiation (Fiorotto, 2012). Such spatial and temporal regulation of *MRFs* expression ensures that progenitor cells undergo proliferation and/or differentiation at the right site and time.

The third phase of myogenesis comprises two waves. The first wave results in the formation of primary fibers that act as supporting scaffolds for forming muscle shape and orientation. During the second wave, proliferating fetal myoblasts fuse to form secondary myotubes along the primary myotubes (Buckingham, 1992; Buckingham et al., 2003; Murphy and Kardon, 2011). It is assumed that not all myogenic progenitors undergo differentiation. The progenitors that retain *Pax7* expression maintain the undifferentiated state and adopt a state of dormancy, also termed quiescence, between differentiated muscle fibers. Quiescent *Pax7*-expressing cells are considered as the developmental origin of muscle stem cells (Gros et al., 2005; Kassar-Duchossoy et al., 2005; Relaix et al., 2005; Schienda et al., 2006).

1.1.3 Skeletal muscle regeneration.

After embryonic muscle development, skeletal muscles can form *de novo* through muscle regeneration. Embryonic and regenerative myogenesis share many common features. In fact, the core embryonic myogenic program is believed to be re-activated in muscle stem cells during muscle regeneration (Tajbakhsh and Cossu, 1997). Like embryonic myogenesis, skeletal muscle regeneration is a highly regulated process. It involves various cellular responses, which can be divided into three distinguishable but overlapping phases: i) necrosis and inflammation of the injured muscle; ii) activation, proliferation, and differentiation of resident muscle stem cells (MuSCs) into muscle fibers and iii) remodeling and maturation of newly-formed muscle fibers.

In the first phase, skeletal muscle regeneration begins with necrosis of muscle fibers, which upon damage releases DAMPs (damage-associated molecular patterns) such as creatine kinase and *miR-133a* to the interstitial space. Release of DAMPs into the damaged tissue environment induces infiltration of different immune cells to the site of injury inducing inflammation (Angelini et al., 1968; Chazaud et al., 2003; Laterza et al., 2009). The complement system is the first defense line of innate immunity, which immediately activates seconds after injury (Zipfel and Reuter, 2009). It senses the injury and promotes activation and infiltration of mast cells, neutrophils, and macrophages to the lesion site (Frenette et al., 2000). Activated mast cells release pro-inflammatory cytokines such as *TNF-α* or *IL-1* to

enhance recruitment of immune cells (Radley and Grounds, 2006). The infiltrating neutrophils release enzymes and oxidative factors to remove fiber debris (Dumont et al., 2008). *CD68* and *CD163* positive macrophages remove fiber debris by phagocytosis and contribute to the termination of inflammation (Chazaud et al., 2009).

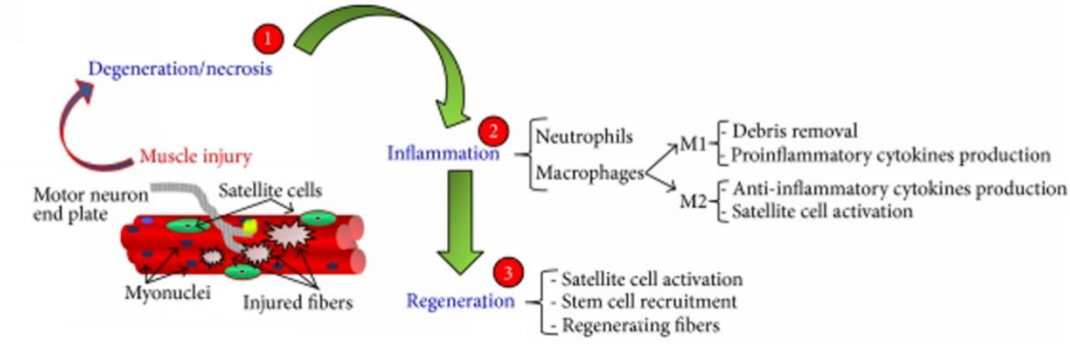

Figure 3. Schematic model-outlining phases of skeletal muscle regeneration.

Skeletal muscle regeneration occurs in three interlaced and time-dependent phases (text in blue). Necrosis of muscle fibers activates a transient muscle inflammation, leading to the activation of resident MuSCs, which proliferate and differentiate to replace damaged muscle myofibers.

Adapted from (Musarò, 2014), with modifications.

Cell proliferation mediates the second phase of skeletal muscle regeneration. Blocking cell proliferation through colchicine treatment or irradiation significantly inhibits skeletal muscle regeneration (Pietsch, 1961; Quinlan et al., 1995). Activation and proliferation of resident MuSCs is the key event of this phase. At this phase, MuSCs exit quiescence, re-enter the cell cycle, undergo differentiation, and finally fuse to existing damaged myofibers or form myofibers *de novo* (Goh and Millay, 2017). Newly-formed myofibers appear as early as four days after injury and display a distinguished morphology, including small caliber size and centrally located myonuclei. In addition, newly formed myofibers express unique myosin isoforms that are normally only expressed during embryonic muscle development such as embryonic myosin heavy chain (*eMHC*) (Etienne et al., 2020; Goh and Millay, 2017; Murphy et al., 2011b; Yin et al., 2013). Besides MuSCs, proliferation of other cell types concurs with skeletal muscle regeneration, such as the proliferation of fibro-adipogenic progenitors (FAPs), a muscle interstitial mesenchymal cell population that is required for the generation of

extracellular matrix, provides trophic support to the skeletal musculature (Biferali et al., 2019; Chapman et al., 2017; Liu et al., 2017).

During the last phase, newly-formed myofibers increase in size and centrally located myonuclei move to the periphery of the muscle fibers. This process initiates from the first week after injury, peaks during the second week and lasts for several weeks until the skeletal muscles have reached homeostasis and have fully recovered (Huard et al., 2002). Interestingly, the growth and maturation of skeletal muscles during regeneration requires muscle innervation in the last phase. Muscle innervation and nerve activity directly influence gene expression within multinucleated regenerating myofibers and therefore indirectly influence proliferation and differentiation of MuSCs (Mozdziak et al., 2001; Musarò, 2014).

1.2 Muscle stem cells.

1.2.1 Identification and characterization of MuSCs.

Skeletal muscle stem cells were first discovered half a century ago when Alexander Mauro and colleagues observed a group of mononucleated cells at the periphery of adult skeletal muscle myofibers through electron microscopy (Mauro, 1961). Based on their sublaminal location on the myofibers plasma membrane, these cells were named "satellite" cells (Mauro, 1961). It is now well established that satellite cells fulfill the hallmarks of stem cells such that they can either self-renew or undergo differentiation and are thus now more commonly referred to as muscle stem cells (MuSCs) (Collins et al., 2005; Relaix et al., 2021).

The first evidence that MuSCs exhibit stemness was provided in 1971 through tracing experiments wherein nuclei of MuSCs were labeled with [^{3}H] thymidine. These labelled nuclei were detected in myonuclei of mature myofibers, suggesting that MuSCs contribute to the formation of skeletal muscles (Moss and Leblond, 1971; Snow, 1977). In support of this notion, researchers later found that transplanting single myofibers containing as few as seven MuSCs could eventually generate hundreds of new MuSCs in the recipients (Collins et al., 2005).

Functional studies of MuSCs-specific genes have greatly extended the knowledge on the biology of MuSCs (Mauro, 1961; Relaix, 2006). The identification of MuSC-specific biomarkers has enabled to prospectively purify MuSCs. These biomarkers include myogenic homeobox transcription factors *Pax3*/*Pax7*, *MRFs*, and other membrane (surface) proteins. Several mouse models have been developed through genetic engineering wherein MuSCs

express fluorescent markers under the control of MuSC-specific promoters. Examples include a Pax3-GFP reporter mouse model as well as Pax7-ZsGreen and Pax7-nGFP reporter mouse models (Bosnakovski et al., 2008b; Montarras et al., 2005; Sambasivan et al., 2011). Additional and frequently used MuSC-specific markers include cell-surface attachment receptor *α7-intergin*, cluster of differentiation protein *CD34*, transmembrane heparan sulfate proteoglycans *syndecan-3* and *syndecan-4*, chemokine receptor *Cxcr4* and calcitonin receptor (Beauchamp et al., 2000; Burkin and Kaufman, 1999; Cornelison et al., 2001; Fukada et al., 2007; Ratajczak et al., 2003). A common method to purify MuSCs from skeletal muscle derived cell suspensions is fluorescent activated cell sorting (FACS) using antibodies to positively select for *α7-integrin* and *CD34* expressing MuSCs while excluding hematopoietic and fibrogenic cell types with antibodies against *CD45*, *CD11b*, *CD31*, and *Ly6a* (Pasut et al., 2012).

1.2.2 The regulators of skeletal muscle regeneration.

MuSCs play a crucial role in skeletal muscle regeneration in both physiological (e.g., extensive exercise) and pathological conditions (e.g., muscular dystrophy) (Parise et al., 2008; Ribeiro et al., 2019). While they normally reside in a quiescent state, MuSCs become activated at the lesion site in response to muscle injury, proliferate, differentiate and eventually form new myofibers. MuSCs are absolutely required for regeneration since their genetic ablation abolishes muscle regeneration (Murphy et al., 2011a; Sambasivan et al., 2011). Similar to embryonic myogenesis, *Pax* and *MRFs* transcription factors display important roles in the regulation of MuSCs behavior during regenerative myogenesis. The differential expression patterns of these factors reflect the activities of MuSCs and their state of differentiation.

Pax7 is predominantly expressed in quiescent and undifferentiated MuSCs. *Pax7* does not only serve as a marker of undifferentiated MuSCs. Genetic inactivation of *Pax7* in MuSCs significantly reduces MuSC proliferation and therefore inhibits muscle regeneration (Kuang et al., 2006; Seale et al., 2000). Conversely, forced overexpression of *Pax7* in MuSCs enhances proliferation rates (Günther et al., 2013). *Pax7* acts genetically upstream to the *MRFs* like *MyoD* and *Myf5* (Relaix et al., 2005; Seale et al., 2000). In chromatin immunoprecipitation (ChIP) assays it has been shown that *Pax7* directly binds a DNA region between -1.5 and -1.6 kb upstream of the *MyoD* transcriptional start site (TSS) thereby regulating its expression (Hu et al., 2008). Interestingly, genomic translocation resulting in fusions of *Pax7* with the

transcriptional activation domain of *Foxo1a* results in formation of alveolar rhabdomyosarcoma that concurs with an upregulation of *MyoD* in tumor cells (Barr, 2001).

Myf5 and *MyoD* are immediately expressed upon activation of MuSCs and contain a conserved basic DNA binding domain that binds to the consensus E-box sequences (CANNTG) on muscle-specific gene promoters (Davis et al., 1990; Le Grand and Rudnicki, 2007). *Myf5* and *MyoD* are mainly involved in muscle specification. Along this line, when overexpressed they can convert fibroblasts into muscle cells (Choi et al., 1990; Davis et al., 1987). *Myf5* mutant mice exhibit a significant decrease of MuSC numbers and a significant delay in skeletal muscle regeneration (Ustanina et al., 2007). Like *Myf5*, *MyoD* is also expressed as MuSCs become activated. Notably, once *Myf5* and *MyoD* are expressed, MuSCs are irreversibly committed to myogenic differentiation (Crescenzi et al., 1990; Halevy et al., 1995).

Notably, inactivation of either *Myf5* or *MyoD* alone does not result in absence of myogenesis (Rudnicki et al., 1992). Loss of *Myf5* abolishes the first wave of emerging muscle cells during development but this loss is efficiently compensated by *MyoD* expressing cells during embryonic myogenesis (Braun et al., 1992; Gensch et al., 2008). Primary myoblasts isolated from *MyoD* mutant mice display myogenic differentiation defects (Sabourin et al., 1999). Compound inactivation of *Myf5* and *MyoD* mice results in a complete lack of skeletal muscle formation suggesting functionally redundancy between these two *MRFs*, which is supported by the fact that *Myf5* and *MyoD* share overlapping DNA binding sites (Conerly et al., 2016b). Consistent with this, *Myf5* is highly expressed in *MyoD* mutant MuSCs (Rudnicki et al., 1992). However, functional differences between *MyoD* and *Myf5* exist since genetic knock-in of *Myf5* into the *MyoD* locus of *MyoD* mutant mice does not completely rescue delayed differentiation phenotypes (Haldar et al., 2014; Kablar et al., 1997). Along this line, *Myf5* does not robustly recruit PolII to induce gene transcription once bound to cognate loci, whereas *MyoD* does (Conerly et al., 2016a). Therefore, *Myf5* and *MyoD* exhibit functional divergence at identical binding sites including differential co-factor recruitment.

MyoG (*Myogenin*) governs the development of functional skeletal muscles, but it is dispensable for the expression of all other muscle determination genes (Dilworth and Blais, 2011; Liu et al., 2012; Venuti et al., 1995). Therefore, *MyoG* acts genetically downstream to *Pax7*, *MyoD* and *Myf5*. *MyoG* is expressed at the late stage of myogenic differentiation subsequent to *MyoD* and *Myf5* expression. Inactivation of *MyoG* abolishes terminal MuSC differentiation. *MyoG* deficient mice exhibit severe skeletal muscle deficiencies (Hasty et al., 1993; Nabeshima et al., 1993). Most *MyoG* mutant mice die at birth; surviving mice display

severe growth deficiencies and are 30% smaller than control mice (Knapp et al., 2006). However, when removing *MyoG* as early as 1 day after birth (P1) after most of the skeletal muscle has formed, mice survive and do not develop or display overt muscle abnormalities except for modest alterations in *MyoD* and *Mrf4* expression levels (Knapp et al., 2006; Meadows et al., 2008). No *MRF* members can compensate for the loss of *MyoG* (Rawls et al., 1995).

Mrf4 is the *MRF* member that is expressed latest during myogenesis and is essential for the maturation of myotubes. The expression of *Mrf4* in late myogenesis mediates the reorganization of myofilaments and migration of centrally located myonuclei to the periphery of myofibers (Megeney and Rudnicki, 1995).

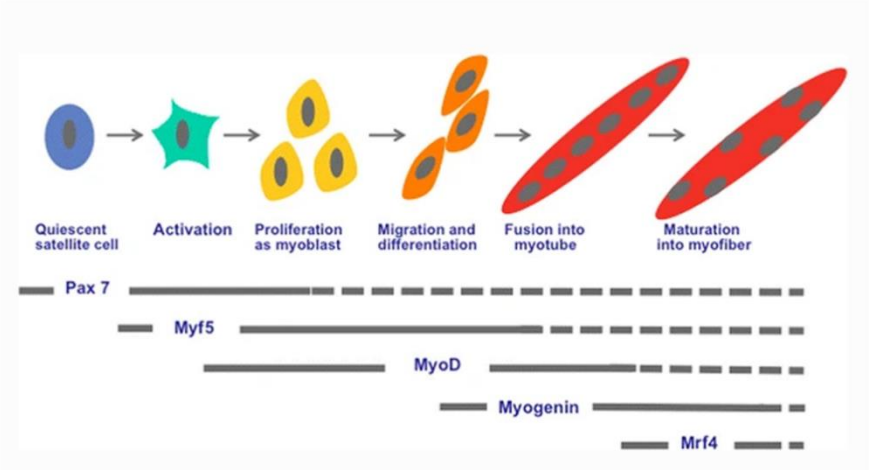

Figure 4. Interplay between the myogenic regulatory factors (MRFs) during skeletal muscle regeneration.

Pax7 is mainly expressed in quiescent MuSCs. Upon muscle injury, MuSCs become activated and express *Myf5* and *MyoD* and begin to proliferate. Later, *MyoD*-positive cells exit from the cell cycle and initiate *MyoG* expression, which is important for the differentiation of myoblasts and their fusion to form myofibers. *Mrf4* acts in the late regeneration phase and is required for the maturation of myofibers.

Adapted from(Meadows et al., 2008), with modifications.

1.3 Introduction to rhabdomyosarcoma.

1.3.1 The features and diagnosis of RMS tumors.

Tumor incidence is closely associated with the turnover of the tissue resident stem cells. Primary tumors develop more commonly in highly proliferative tissues such as in skin, intestine or blood (Ferlay et al., 2018). In contrast, the incidence of tumor formation is lower in soft tissues such as the skeletal muscle. Roughly 350 deaths related to soft tissue tumors are reported every year in the USA alone (Ferlay et al., 2018). Rhabdomyosarcoma (RMS) is the most common soft tissue tumor accounting for nearly half of all reported cases (Arndt and Crist, 1999). RMS tumors can present anywhere in the body but occurs most often in extremities, genitourinary system, head, and neck region (Arndt and Crist, 1999). Despite the significant advances that have been made to treat non-metastatic RMS tumor patients, the 5-year survival rate of patients with metastatic RMS tumors is lower than 30 % and has not improved in the past three decades (De Giovanni et al., 2009).

Besides anatomical location, RMS tumors diagnosis primarily relies on skeletal muscle differentiation features. RMS tumors harbor rhabdomyoblast-like cells, which are analyzed via immunohistochemistry staining and the expression of muscle lineage genes, including *Desmin*, *MyoD*, *MyoG*, and muscle-specific *actin* (Dias et al., 1990; Parham et al., 1991). RMS tumors are subdivided into two major subtypes, alveolar rhabdomyosarcoma (aRMS) and embryonal rhabdomyosarcoma (eRMS), counting roughly 60 % and 16 % of overall RMS tumors, respectively (Sultan et al., 2009). eRMS tumors most commonly consist of a heterogeneous mixture of immature mesenchymal cells and more differentiated elongated cells expressing myogenic differentiation markers. In contrast, aRMS tumors are composed of small, poorly differentiated and densely packed cells organized around spaces that are reminiscent of pulmonary alveoli. While also expressed in some eRMS tumors, *MyoG* expression is usually used as a specific marker for aRMS tumors to distinguish them from other RMS subtypes. In clinical analysis, *MyoG* has 75 % specificity for aRMS tumors and only 25 % specificity for eRMS tumors (Kumar et al., 2000). Other reports have also suggested a combination of gene expression patterns for RMS tumor diagnosis such as a combination of *EGFR*/*fibrillin-2* for eRMS tumors and *AP2beta*/*P-cadherin* for aRMS tumors, respectively (Wachtel et al., 2006). Overall, profound heterogeneity exists not only across RMS subtypes but also within one and the same RMS subtype.

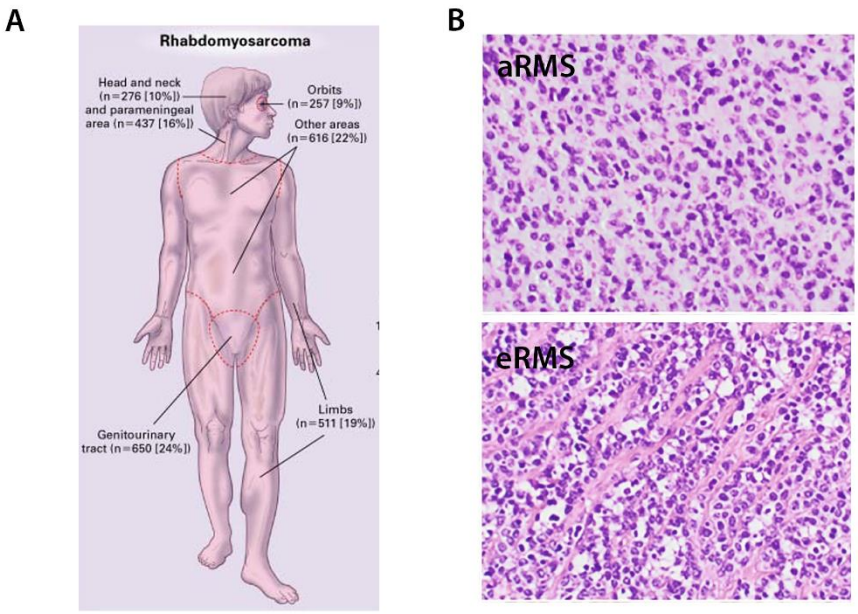

Figure 5. RMS tumor localization and histology.

(A) Primary sites of RMS tumors.

(B) Hematoxylin & Eosin staining of aRMS (left panel) and eRMS (right panel).

Adapt from (Arndt and Crist, 1999; Skapek et al., 2019) with modification.

1.3.2 The causes of RMS tumors.

Although classification of RMS tumor subtypes based on expression markers and histology serve prognostic and diagnosis value, it is too simplistic. Instead, recent studies have suggested to distinguish RMS tumors into two other clusters based on molecular markers. These distinctions are based on the finding that RMS tumors can be subdivide into two molecular subtypes: PAX-fusion positive (PFP) and PAX-fusion negative (PFN) RMS tumors, which show distinct genetic landscapes and methylation patterns (Davicioni et al., 2006; Seki et al., 2015; Shern et al., 2014; Wachtel et al., 2004).

Approximately 80 % of all aRMS tumors are PFP (Sorensen et al., 2002). These tumors comprise the genomic translocations t (2;13)(q35;q14) and t (1;13)(p36;q14), which lead to the generation of chimeric *Pax3-Foxo1* and *Pax7-Foxo1* genes, respectively (Sorensen et al., 2002). Occasional other rare translocations result in alternative *Pax* gene fusion products including *Pax3-Foxo4*, *Pax3-Ncoa1*, *Pax3-Ncoa2*, and *Pax3-Ino80D* (Wachtel and Schäfer, 2015). Fusion of the DNA binding domain of *Pax* genes to transcriptional transactivation domains of

other genes leads to generation of oncogenic and chimeric transcription factors that are capable of inducing aberrant expression of cognate genes. For example, two genes that are frequently upregulated in PFP tumors are *Cdk4* and *Mycn*, overexpression of which are alone sufficient to induce tumorigenic transformation (Shern et al., 2015). These translocation events are almost never observed in eRMS tumors (El Demellawy et al., 2017; Sorensen et al., 2002).

In contrast to PFP tumors, PFN tumors exhibit a much more complex genetic landscape and exhibit varying mutations. Some eRMS tumors, exhibiting mutations of *Ras* genes, are exclusive to PFN. Interestingly, overexpression of *Ras* in myoblasts appears to amongst the fastest way to induce RMS tumor transformation (Langenau et al., 2007). Indeed, *Ras* expression has been long recommended as a marker to stratify PFN patients based on the fact that 75 % of the high-risk and 45 % of the intermediate-risk RMS tumor patients contain *Ras*-related mutations (Chen et al., 2013). Notably however, inhibitors against the RAS/PI3K pathway do not significantly affect growth of RMS tumor cells derived from PFN xenografts indicating that the genetic etiology of such tumors might not only depend on *Ras* genes alone (Chen et al., 2013). Besides, many other mutations associated with PFN but not primarily with PFP have been identified including mutations in genes related to the *p53* axis, including mutations in *p53* itself, amplifications of its negative regulators *Mdm4*/*Mdm2*, loss of *Cdkn2a*, or hyperactivating mutations *Cdk4*, *Os9*, *Gli1* and/or *FGFs* and its receptor *FGFRs* (Chen et al., 2013; Shern et al., 2014).

1.3.3 The cellular origin of RMS tumors.

Despite the rapid improvement of sensitive sequencing technologies enabling identification of mutations and genomic alterations, the cellular origin in which these genetic alternations actually occur giving rise to the tumor often remains unclear. Much attention has been directed to adult stem cells owing to their inherent proliferation ability and multipotent potential (Charytonowicz et al., 2009). As for other cancer types, identifying the cellular origin of RMS tumors is challenging. Mesenchymal stem cells (MSCs) have long been considered as a potential cellular origin of RMS tumors. Fusion proteins found in RMS tumors can induce transcriptional and phenotypic changes that resemble sarcoma when overexpressed in MSCs (Tirode et al., 2007). Moreover, inhibiting expression of oncogenic fusions in RMS tumor cells can shift their gene expression pattern towards that of MSCs (Tirode et al., 2007). More recently, it was shown that overexpression of the sonic hedgehog effector *SmoM2* in endothelial

progenitors results in myogenic transdifferentiation and RMS tumor formation in mice, suggesting that RMS tumors can also originate from non-myogenic cells (Drummond et al., 2018).

Since the presence of myogenic features is a hallmark of RMS tumors, various reports have implicated MuSCs and/or their more differentiated descendants to be a predominant origin of RMS tumors. In support of this notion, it has been shown that overexpression of the major Hippo signaling effector *Yap1* in muscle cells can elicit RMS tumor formation in mice (Tremblay et al., 2014). Moreover, conditional mouse models wherein various tumor suppressors are deleted in the myogenic lineage under resting conditions eventually give rise to RMS tumors with varying penetrance (Rubin et al., 2011). Most interestingly, an increased incidence of spontaneous RMS tumor formation occurs in aged mdx mice, a mouse model of Duchenne muscular dystrophy (Chamberlain et al., 2007b; Hosur et al., 2012a). Later it was shown that *p53* deletion in mdx mice not only significantly increases RMS tumor incidence but also dramatically reduces the latency of RMS tumor formation (Camboni et al., 2012). Together, these results strongly suggest that muscle regeneration might provide an environment that promotes the initiation of RMS tumors from *p53*-deficient MuSCs.

1.4 Thesis aim and overview.

Previously, it was shown that RMS tumor formation occurs with high rate under conditions of continuous regeneration in *p53*-deficient (Camboni et al., 2012). However, definitive evidence on the cellular origin of RMS tumors was still lacking. Consequently, the causative mechanisms that transform the yet undefined cancer cell of origin of RMS tumors had remained elusive.

Based on the fact that MuSCs are the major source of muscle regeneration, the first aim was to test whether MuSCs are a cellular origin of RMS tumors by inducing *p53* definition specifically in MuSCs under regenerative and non-regenerative conditions. The second aim was to investigate the molecular mechanisms underlying RMS tumor formation. Based on the idea that oncogenic transformation occurs through acquisition of tumorigenic mutations (Ashkenazi et al., 2008), an additional aim was to identify these tumorigenic mutations by genomic sequencing of tumor propagating cells.

The last aim was to test whether such mutations in MuSCs play causative roles in RMS tumor formation by analyzing hallmarks of cancer including cell proliferation and differentiation through a series of *in vitro* and *in vivo* gain- and loss-of-function experiments.

2. Materials and methods.

2.1 Materials.

2.1.1 Antibodies.

Table 1. List of primary antibodies.

Antibody	Source and types	Company	Dilution
GAPDH	rabbit,polyclonal	Cell Signaling Techn. (Cat.2118)	WB (1:5000)
β-actin	mouse,monoclonal	Sigma (Cat.A-5441)	WB (1:5000)
Myogenin	rabbit,polyclonal	Sigma (Cat.HPA03809)	WB, IHC (1:1000)
Desmin	rabbit,polyclonal	Sigma (Cat.D8281)	WB, IHC (1:1000)
Pan-canherin	mouse,monoclonal	BD (Cat.610182)	WB, IF (1:1000)
V5-tag	rabbit,polyclonal	Abcam (Cat.ab9116)	WB, IF (1:1000)
MyoD1	rabbit,polyclonal	Abcam (Cat.ab64159)	WB, IHC (1:1000)

Table 2. List of secondary antibodies.

Antibody	Source and types	Company	Dilution
Anti-rabbit IgG	goat, HRP-conjugated	Pierce (Cat.1858415)	WB (1:5000)
Anti-mouse IgG	goat, HRP-conjugated	Pierce (Cat.1858413)	WB (1:5000)
Anti-rabbit IgG	goat, Alex594-conjugated	Invitrogen (Cat.A11012)	IF (1:1000)
Anti-mouse IgG	goat, Alex488-conjugated	Invitrogen (Cat.A11070)	IF (1:5000)

2.1.2 Primers.

All primers were synthesized and ordered from Sigma-Aldrich

Table 3. List of primers for quantitative real-time PCR (qRT-PCR).

Gene name	Primers sequences (5’-3’)	Annealing temp.in ℃
Duxbl	Forward:5’-GCATCTCTGAGTCTCAAATTATGACTTG-3’ Reverse:5’-GCGTTCTGCTCCTTCTAGCTTCT-3’	60
Yap1	Forward:5’-ACCCTCGTTTTGCCATGAAC-3’ Reverse:5’-TGTGCTGGGATTGATATTCCGTA-3’	60

β-actin	Forward:5'-TAGGCACCAGGGTGTGATGG-3' Reverse:5'-CATGGCTGGGGTGTTGAAGG-3'	60
GAPDH	Forward:5'-TCCATGACAACTTTGGCATTG-3' Reverse:5'-CAGTCTTCTGGGTGGCAGTGA-3'	60
Zscan4	Forward:5'-AAATGCCTTATGTCTGTTCCCTATG-3' Reverse:5'-TGTGGTAATTCCTCAGGTGACGAT-3'	58
Zfp3520	Forward:5'-AGAGGACAAGACCCAGTGCAG-3' Reverse:5'-GAGGTCCTCATCTGACCCAAG-3'	60
Zbed3	Forward:5'-AGAGGACAAGACCCAGTGCAG-3' Reverse:5'-GAGGTCCTCATCTGACCCAAG-3'	59
Eif1a	Forward:5'-AACAGGCGCAGAGGTAAA AA-3' Reverse:5'-GAGGTCCTCATCTGACCCAAG-3'	62
RikORF	Forward:5'-CCCAACTCAACCATCTCCAGC-3' Reverse:5'-TCCAAGGGACTCTTCAGGCATC-3'	60
LINE1-ORF1	Forward:5'-CTCGGCAGAAACCCTACAAG-3' Reverse:5'-CCATGTTTAGCGCTTCCTTC-3'	60
ERVL	Forward:5'-ATCTCCTGGCACCTGGTATG-3' Reverse:5'-AGAAGAAGGCATTTGCCAGA-3'	60
LINE1-ORF2	Forward:5'-GGAGGGACATTTCATTCTCATC-3' Reverse:5'-GCTGCTCTTGTATTTGGAGCATAGA-3'	60
MusD	Forward:5'-GGAGGGACATTTCATTCTCATC-3' Reverse:5'-GCTGCTCTTGTATTTGGAGCATAGA-3'	60
GLN	Forward:5'-CGTAAGGACCCTAGTGGCTG-3' Reverse:5'-GCACTCACTCTTCTTCACTCTG-3'	60
IAP-gag	Forward:5'-AATCTCAGAACCGCTCCATGA-3' Reverse:5'-TTTCTTAAAATGCCCAGGCTTT-3'	60
IAP-pol	Forward:5'-CTTGCCCTTAAAGGTCTAAAAGCA-3' Reverse:5'-GCGGTATAAGGTACAATTAAAAGATATGG-3'	60
IAP-5'UTR	Forward:5'-CGGGTCGCGGTAATAAAGGT-3' Reverse:5'-ACTCTCGTTCCCCAGCTGAA-3'	60
MERVK10C	Forward:5'-CAAATAGCCCTACCATATGTCAG-3' Reverse:5'-GTATACTTTCTTCTTCAGGTCCAC-3'	58
ERVK10CL	Forward:5'-GTGTGAGACACGCCTCTCCT-3'	60

TR	Reverse:5'-GGGAGAGCTTGA TTGCAGAG-3'	
ERVK10CGAG	Forward:5'-GTGTGAGACACGCCTCTCCT-3' Reverse:5'-GGGAGAGCTTGA TTGCAGAG-3'	60
ERVK10CPOL	Forward:5'-GCCACCAGAGACA TGGTTTT-3' Reverse:5'-CGGGCTTCTTTTCTTGTGAG-3'	60
ERVK10CENV	Forward:5'-TATCGCCTCAGGGTTAATGC-3' Reverse:5'-TGGATGCCACACAACTCATT-3'	60

Table 4. List of primers for genotyping.

Gene name	Primers sequences (5'-3')	Annealing temp.in ℃
$Pax7^{CreERT2}$	Forward:5'-ACTAGGCTCCACTCTGTCCTTC-3' Reverse:5'-GCAGATGTAGGGACATTCCAGTG-3'	58
ZsGreen	Forward:5'-CTGCATGTACCACGAGTCCA-3' Reverse:5'-GTCAGCTGCCACTTCTGGTT-3'	57
Rosa26 CAGG	RosaFA:5'-AAAGTCGCTCTGAGTTGTTAT-3' RosaRF:5'-GGAGCGGGAGAAATGGATATG-3' SpliAC:5'-CATCAAGGAAACCCTGGACTACTG-3'	57
Mdx	Forward:5'-GCGCGAAACTCATCAAATATGCGTG TTAGTGT-3' Reverse:5'-GATACGCTGCTTTAATGCCTTTAGTC ACTCAGATAGTTGAAGCCATTTTG-3'	60
tdTomato	Forward:5'-CTGTTCCTGTACGGCATGG-3' Reverse:5'-GGCATTAAAGCAGCGTATCC-3'	58
$p53^{flox}/p53^{\Delta}$	Forward:5'-CACAAAAACAGGTTAAACCCAG-3' Reverse-$p53^{flox}$:5'-AGCACATAGGAGGCAGAGAC-3' Reverse-$p53^{\Delta}$:5'-GAAGACAGAAAAGGGGAGGG-3'	55

Table 5. List of sequences for shRNA.

ShRNA name	Primers sequences (5'-3')	Resource
Yap1 shRNA#1	5'-GCAGACAGATTCCTTTGTTAA-3'	Sigma Aldrich
Yap1 shRNA #2	5'-CCACCAAGCTAGATAAAGAAA-3'	Sigma Aldrich

Yap1 shRNA #3	5'-CGGTTGAAACAACAGGAATTA-3'	Sigma Aldrich
Yap1 shRNA #4	5'-GCGGTTGAAACAACAGGAATT-3'	Sigma Aldrich
Yap1 shRNA #5	5'-CTGGTCAAAGATACTTCTTAA-3'	Sigma Aldrich
Duxbl shRNA #1	5'-GCAGGATAAACCTAGAGTTAA-3'	Sigma Aldrich
Duxbl shRNA #2	5'-GCTGAATGGATGCCTGACAAA-3'	Sigma Aldrich
Duxbl shRNA #3	5'-GCTTCAGTTATACTGCCTCTT-3'	Sigma Aldrich
Duxbl shRNA #4	5'-CCGCGCTTAGAAGATTGTACT-3'	Sigma Aldrich
Scrambled shRNA	5'-CCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGGGCGACTTAACCTTAGG-3'	Sigma Aldrich

2.1.3 Vectors.

Table 6. List of vectors.

Vector name	Description and resource
pLK0.1	Purchased from Sigma Aldrich
pGEM-T-easy	Purchased from Promega, Mannheim
pLVX-Dux-V5-T2A-mcherry	Duxbl ORF fused with C-terminal V5 tag was cloned into the commercial pLVX-T2A-mcherry vector for ectopic overexpression. Design by Dr. Johnny Kim and synthesized by vectorbuilder
pLVX-Duxbl-V5-T2A-mcherry	Dux ORF fused with C-terminal V5 tag was cloned into the commercial pLVX-T2A-mcherry vector for ectopic overexpression. Design by Dr. Johnny Kim and synthesized by vectorbuilder
pLVX-Dux4-V5-T2A-mcherry	Dux4 ORF fused with C-terminal V5 tag was cloned into the commercial pLVX-T2A-mcherry vector for ectopic overexpression. Design by Dr. Johnny Kim and synthesized by vectorbuilder
pLVX-DuxA-V5-T2A-mcherry	DuxA ORF fused with C-terminal V5 tag was cloned into the commercial pLVX-T2A-mcherry vector for ectopic overexpression. Design by Dr. Johnny Kim and synthesized by vectorbuilder

pLVX-DuxB-V5-T2A-mcherry	DuxB ORF fused with C-terminal V5 tag was cloned into the commercial pLVX-T2A-mcherry vector for ectopic overexpression. Design by Dr. Johnny Kim and synthesized by vectorbuilder
pBigT Duxbl	Duxbl ORF fused with C-terminal V5 tag was cloned into pBigT vector for Rosa26 targeting vector subclone. Kindly provide by Dr. André Schneider
pRosa26 Duxbl	Duxbl ORF fused with C-terminal V5 tag was cloned into pRosa26 vector for Rosa26 targeting. Kindly provide by Dr. André Schneider

2.1.4 Bacterial strains.

Table 7. List of bacterial strains.

Bacterial strains	Description
DH5α	F-φ80lacZΔM15Δ(lacZYA-argF) U169 deoR recA1 endA1 hsdR17(rK^-, mK^+) phoA supE44λ-thi-1 gyrA96 relA1
Stbl3	F−mcrB mrr hsdS20 (rB^-, mB^-) recA13 supE44 ara14 galK2 lacY1 proA2 rpsL20 (StrR) xyl5 λ−leu mtl1

2.1.5 Cell lines.

Table 8. List of cell lines.

Name	Description
HEK 293 ATCC ® CRL-1573™	Human Embryonic Kidney Cells 293
HEK 293T ATCC ® CRL-11268™	HEK293 cells constitutively express the simian virus 40 (SV40) large T antigen
C2C12 ATCC® CRL-1772™	Immortalized mouse skeletal myoblast cell line originally derived from MuSCs from the thigh muscle of a two month old female C3Hmouse donor 70h after a crush injury
Platinum-E(Plat-E) Cell Biolabs, Inc.RV-101	HEK293 cells constitutively express the Moloney Murine Leukemia Virus (M-MLV) viral gag, pol and env proteins

2.1.6 Medium for cell culture.

All media and reagent for cell culture were purchased from life technologies except as noted otherwise.

Table 9. List of medium for cell culture.

Name	Description
DMEM culture medium	Dulbecco's modified Eagle's medium (DMEM, high glucose, glutamineMax) supplemented with 10 % heat-inactivated fetal cow serum (FCS, Gibco), 100 U/mL of penicillin-streptomycin
Differentiation medium	Dulbecco's modified Eagle's medium (DMEM, high glucose, glutamineMax) supplemented with 2 % heat-inactivated horse serum (FCS, Gibco), 100 U/mL of penicillin-streptomycin
Muscle stem cell culture medium	Dulbecco's modified Eagle's medium (DMEM, high glucose, glutamineMax) supplemented with 10 % fetal cow serum (FCS, Gibco), 5 ng/ml bFGF, 100 U/mL of penicillin-streptomycin
Serum free ES (SFES) medium	Mixture of neurobasal and DMEM/F12 medium supplemented with 1 % N2 supplement and 0.5 % B27 supplement, 100 U/mL of penicillin-streptomycin
2i medium	SFES medium supplemented with 3µM CHIR99021 and 1µM PD03259010, 1000 U/mL LIF, 1 % heat-inactivated fetal cow serum (FCS, Gibco)
murine embryonic fibroblast (MEF) medium	Dulbecco's modified Eagle's medium (DMEM, high glucose, glutamineMax) supplemented with 20 % heat-inactivated horse serum (FCS, Gibco), 100 U/mL of penicillin-streptomycin, 1000 U/mL LIF, 0.1 mM β-mercaptoethanol, 1 mL MEM Non-Essential Amino Acids Solution
MEF-VC medium	MEF medium supplemented with vitamin C

2.1.7 Mice strains.

Table 10. List of mice strains.

Name	Description
C57BL/6J	Harlan-Winkelmann, Paderborn
$Pax7^{CRE/ERT2}$	The MuSCs specific Cre-recombinase mediates the excision of floxed inserted regions in MuSCs upon TAM treatment. Kind gift from Dr. Chenming Fan
CMVCre Tg	The Cre gene in this mice strain is derived by a human cytomegalovirus minimal promoter and express in all tissues. Kind gift from Spitznagel Birgit, MPI Bad Nauheim
Pax7ICN	The MuSCs specific Cre-recombinase mediates the excision of floxed-inserted regions in MuSCs during early postnatal development. Kind gift from Dr. Charles Keller
Pax7::ZsGreen	The *ZsGreen* reporter is inserted after endogenous *Pax7* promoter to lineage-traced MuSCs. Kind gift from Dr. Michael Kyba
$Rosa26^{tdTomato}$	The *tdTomato* reporter is inserted into *Rosa26* locus and under control of loxP-stop-loxP cassette and meditate by Cre-recombinase. Purchased from Jackson Laboratory Stock No: 007914
$p53^{flox/flox}$	The exon2 of *p53* gene is flanked by two loxP sites. Presenting of Cre-recombinase leads to deletion of exon2 and induced loss of function. Purchased from Jackson Laboratory Stock No: 008462
Mdx	This mice strain contains a point mutation in *dystrophin* gene that leads to amino acid coding changes from glutamine to STOP codon. The loss of dystrophin gene generates a constantly regeneration environment in skeletal muscle. Purchased from Jackson Laboratory Stock No: 001801
$Rosa26^{Duxbl}$(Duxbl)	The murine *Duxbl* ORF is inserted into *Rosa26* locus and under control of loxP-stop-loxP cassette and meditate by Cre-recombinase. Generated in this thesis.

2.1.8 Special materials and chemicals.

Table 11. List of materials and chemicals.

Name	Resource
BSA, Fraction V	Merck (Cat.112018)
Bromophenol blue	Merck (Cat.1081220005)

Cell culture dishes	Nunc and Greiner
Dispase	BD
Collagenase typeII	Worthington Biochemicals
DAPI	Invitrogen (Cat.D1306)
Ethidium bromide	AppliChem (Cat.A1152,0100)
Falcon®Cell strainer (100μM, 70μM,40μM)	BD Biosciences
MES running buffer	Invitrogen (Cat.NP0002)
Mowiol	Merck (Cat.475904)
NuPAGE®Novex®PAGE gel	Invitrogen
Paraformaldenhyde, PFA	Merck (Cat.1040051000)
Polybrene	Sigma (Cat.107689)
Protein standard marker	Invitrogen (LC5800)
Porteinase inhibitor cocktail tablets	Roche (Cat.04693116001)
PVDF membrane	Sigma (Cat. 3010040001)
Puromycine	Sigma (Cat.P8833)
Red Alert® buffer	Novagen (Cat.710783)
Super signal west® femto	Thermo fisher scientific (Cat.34095)
SYBR Green fluorescein qPCR mix(2x)	Thermo fisher scientific (Cat.k0241)
TRIzol® Reagent	Invitrogen (Cat.15596026)
M.O.M.® (Mouse on Mouse) Blocking Reagent	Vector Laboratories (Cat.MKB-2213)
Signal Enhancer HIKARI for Western Blotting and ELISA	Nacalai tesque (Cat. 02267-41)

2.1.9 Standard buffers and solutions.

Table 12. List of buffers and solutions.

Name	Description
Immunoprecipitation buffer	20 mM Tris-Hcl, pH7.4, 200 mM NaCl, 2 mM EDTA, 2 mM EGTA, 1 % Trition X-100 (v/v)
10x PBS (pH7.2)	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4
20x SSC (pH7.0)	3 M NaCl, 0.3 M Na-Citrate,
Tail lysis buffer	10 mM Tris, pH8.0, 100 mM NaCl, 10 mM EDTA, pH

	8.0, 0.5 % SDS, 200µg/mL proteinase K
Washing buffer for Southern blot	40 mM Na_2HPO_4, pH 6.76, 1 % SDS(w/v)
10xTBS	500 mM Tris-HCl, 1.5 M NaCl adjusted pH to 7.4
1xTBST	1xTBS, 0.1 % Tween-20
1xTAE buffer	40 mM Tris-HCl pH8.3, 1mM EDTA, 20 mM Acetic acid
TE buffer	10 mM Tris-HCl pH8.0, 1mM EDTA
Transfer buffer for Western blot	25 mM Tris base, 192 mM Glycine, 20 % Methanol
Sodium citrate buffer	10 mM Sodium citrate, 0.05 % Tween 20, pH 6.0

2.1.10 Enzymes.

Table 13. List of Enzymes.

Name	Resource
DNase I	Roche (Cat.04716728001)
Pfu DNA polymerase	Promega (Cat.M774A)
Proteinase K	Roth (Cat.7528.1)
T4 DNA ligase	Promega (Cat.M1801)
NheI restriction enzymes	New england biolabs (Cat.R0131S)
NotI restriction enzymes	New england biolabs (Cat.R0189S)
PacI restriction enzymes	New england biolabs (Cat.R0547S)
AscI restriction enzymes	New england biolabs (Cat.R0558S)
AflIII restriction enzymes	New england biolabs (Cat.R0541S)
EcoRV restriction enzymes	New england biolabs (Cat.R0195S)

2.1.11 Kits.

Table 14. List of Kits.

Name	Resource
Nucleobond AX500 Maxi Kit	Macherey-Nagel (Cat. 740414.10)
QIAEX II Gel Extraction Kit	Qiagen (Cat. 20021)
SuperScript™ II Reverse Transcriptase Kit	Invitrogen (Cat.18064-014)
Click-iT™ EdU Cell Proliferation Kit	Thermo fisher scientific (Cat.C10337)

Stem cells from skeletal muscle Kit	MACS miltenyi biotec (Cat.130-104-268)
BCA protein quantification Kit	Abcam (Cat.ab102536)

2.1.12 Equipment.

Table 15. List of equipment.

Name	Type	Resource
Fluorescence microscopes	Axiophot2 Z1 fluorescence microscope	Carl Zeiss Jena
PCR machine	Master cycler gradient PCR	Eppendorf
qRT-PCR machines	StepOnePlus™ Real-Time PCR system	Applied Biosystems
Spectrophotometer	NanoDrop 2000/2000c	Thermo fisher scientific
Ultrasonic Homogenizer	Bioruptor Sonopuls HD 2070/2200	Diagenode Bandelin
VersaDoc Image System	VersaDoc™ 3000	BioRad

2.1.13 Software.

Table 16. List of software.

Name	Resource
Axio Vision® 4.8	Carl Zeiss Imaging Solutioins, Jena
DNASTAR Lasergene®	DNASTAR Inc. USA
ImageJ	National Institutes of Health. USA
Photoshop 8.0	Adobe Systems Incorporated. USA
StepOnePlus Software v2.2.2	Applied Biosystems

2.2 Methods.

Standard molecular biology procedures were performed according to the following reference books except noted otherwise:

Current Protocols in Molecular Biology

F.M Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, K. Struhl; Wuley Interscience, 1989

Molecular Cloning: A Laboratory Manual, 2nd Edition

J.Sambrock, E.F.Frisch, T.Maniatis; Cold Spring Harbor Laboratory Press,1989

2.2.1 Sterilization of materials and solutions.

Normal glassware was sterilized with dry heat at 180 ℃. Sensitive glassware, plastic ware, solutions and medium were autoclaved at 121 ℃, 2.2 bars pressure for 30 mins. Non-autoclavable solutions or medium were sterilized by filter with 0.2 μm or 0.45 μm cellulose acetate filter.

2.2.2 Cloning and constructs.

The expression vectors are listed in Table 6. The shRNA vectors are selected from Sigma Mission® shRNA library and the targeting sequences are listed in Table 5. For other cloning, desired DNA fragments were amplified by PCR using specific primers followed by typical restriction enzyme-based cloning methods.

2.2.3 Cell culture.

All cells were cultured in cell incubator at 37 ℃ and 5 % CO_2 with required media listed in Table 9. For differentiation, cells were cultured in differentiation medium when 60 % confluence was reached. For cell freezing, trypsinized cell pellets were resuspended in freezing medium and aliquoted into cryovials. Cells were then stored in -80 ℃ or in liquid nitrogen for short-term and long-term storage, respectively. For cell thawing, aliquoted cells were quickly thawed in 37 ℃ water baths and washed thrice with 1x PBS before plating into culture dishes. Cells were then maintained in required medium.

2.2.4 Isolation of MuSCs.

MuSCs isolation was performed as described (Kim and Braun, 2014). Briefly, skeletal muscles collected from limb and trunk were minced into pieces and then digested with 100 CU

Dispase and 0.2 % type II collagenase for 30 mins upon continuous shaking. The mixture was consecutively filtered through 100 µm, 70 µm, and 40 µm cell strainers and applied to a discontinuous gradient Percoll consisting of 70 % v/v Percoll overlayed with 30 % v/v Percoll. Cells were collected at the 70/30 interphase and subjected to further steps. GFP fluorescence was used if MuSCs were labeled with *Pax7::ZsGreen*. Otherwise an antibody cocktail provided in a MuSCs selection kit (Miltenyi) was used. Purified MuSCs cells were cultured on Matrigel-coated microClear plates.

2.2.5 Cell transfection.

Cell transfection of desired vectors was performed according to the calcium phosphate transfection method (Kwon and Firestein, 2013). Desired number of cells was seeded in dishes or plates approximately 24 hours before transfection. Vector DNA dissolved in ddH_20 was precipitated by drop-wise addition of 2.5 M $CaCl_2$. The mixture was then applied into same volume of 2x HBS buffer and incubated at room temperature for 10 mins. The calcium phosphate-DNA complex was then drop-wise added to targeted cells. Medium was replaced the next day and the cells were harvested or cultured for the next experiment.

2.2.6 Preparation and infection of retrovirus or lentivirus.

Retroviruses or lentiviruses were differently used depending on the purposes. Retrovirus were used to for acute overexpression of ORFs in MuSCs, whereas lentiviruses were mainly used to induce shRNA-dependent genes silencing in MuSCs or for generation of stable cell lines. For virus packaging, retroviral expression vector or lentiviral expression vector were transfected into either Plat-E cells or HEK293T cells, respectively. Additional, the helper vector psPAX2 and envelope vector pMD2.G were added to the transfection mix when packaging lentiviral. Medium was replaced 12-16 hours after transfection. After an additional 24 hours of incubation, the culture medium was collected and filtered through a 0.45 µm protein low binding filter, the fileted supernatant was used as virus solution. For virus infection, virus solution was added to the targeted cells in the medium containing 4 µg/mL polybrene for 24 hours. Medium was replaced the next day and the cells were harvested or cultured for the following experiments. To generate stable cell lines, lentivirus infected cells were selected in

medium containing 2.5 mg/mL puromycin for at least one week. The medium was replaced every two days to remove debris of dead cells.

2.2.7 Genomic DNA isolation.

A small tail biopsy was cut and digested in 500 μL TENS buffer (100 mM EDTA, pH 8.0; 100 mM NaCl, 50 mM Tris-hydrochloric acid, pH 8.0 and 1 % SDS) with 200 μg/mL proteinase K at 56 °C overnight. The following day, samples were centrifuged at 12000 g for 5 mins and the supernatant was transferred into a new tube. Then 500 μL isopropanol was added to precipitate the DNA. Spin down the DNA at 12000 g for 5 mins and the pellet was washed with 1 mL 75 % ethanol. Spin the pellet down again at 12,000 g for 5 mins and dry for 5 mins before adding 300 μL autoclaved water. The tube was incubated at 56 °C for 1 hour or longer to elute and re-suspend the DNA.

2.2.8 Polymerase chain reaction based genotyping.

Polymerase chain reaction (PCR) was performed to determine mice genotypes. For each setup, 10 μL 2× Taq Master Mix (Vazyme biotech co,Ltd.), 1 μL 10 mM forward primer, 1 μL 10 mM reverse primer, 5 μL molecular level water and 3 μL genomic DNA were mixed. PCR reaction were performed in a Thermocycler PCR machine at a cycle program corresponding to the PCR conditions. Primer sets used for different genotypes and annealing temperature were list in table 4.

2.2.9 RNA samples preparation and cDNA synthesis.

Total RNA was isolated from cells or tissues by using Trizol reagent (life technologies) according to the manufacturer's protocol. For tissue samples, samples were placed in a 2 mL eppendorf tube with a sterile grinding ball (5 mm, Retsch, 22.455.0003). 1 mL Trizol reagent was added and the tube was placed in a tissue homogenizer (Retsch, MM301) for 10 mins at 30/s frequency for homogenization. For cultured cells, cells were washed twice with 1x PBS and then 1 mL Trizol reagent was added to the culture dish. Cells were harvested and collected to a 1.5 mL eppendorf tube by using cell scraper (Sarstedt, 83.1830). 200 μL chloroform was

added to the tube and samples were incubated for 5 mins at room temperature. Following, samples were centrifuged at 4 °C for 10 mins at 12,000 g. The white pellet appeared at bottom of the tube and was washed twice with 1 mL 75 % ethanol. After centrifugation, ethanol was removed and the tubes were put inside the hood until dry. The pellet was re-suspended with 30-50 µL DEPC water (0.1 % v/v DEPC in water). The dissolved samples were considered as isolated total RNA samples, the concentration was measured with a spectrophotometer (NanoDrop ND-2000c Spectrophotometer).

For cDNA synthesis, >500 ng of purified RNA was reversely transcribed with PrimeScript RT Reagent Kit (Takara bio, RR037B) according to the manufacturer's protocol. Random primers and Oligo DT primers were modified in a usage of in the 1:1 mixture, and the incubation time was adjusted to 30 mins to gain better reverse transcription products. After the reaction, samples were considered as cDNA for further use or stored in -20 °C.

2.2.10 Quantitative real time polymerase chain reaction (qRT-PCR).

qPCR was used to quantitatively determine the mRNA expression. During the amplification, fluorescent dyes can incorporate into double stranded DNA and were read as signal. For a typical qPCR setup, it contained 10 µL 2x SYBR Green® mixture, 1 µL of forward and reserved primers (2.5 pmol), and 5 µL of diluted cDNA (1:100 diluted) in a total volume of 25 µL. The PCR conditions were chosen according to the manufactory's protocol and the calculated annealing temperatures of primers. The PCR runs were performed on StepOnePlus real time PCR machine. The relative amount of the target genes and endogenous housekeeping gene was determined in same plates to avoid deviation. The relative expression level of each genes was calculated with the $2^{-\Delta\Delta CT}$ method as described.

2.2.11 Protein samples preparation.

For tissue samples, samples were briefly freezed with liquid nitrogen and crushed using a mortar and pestle into white powder, and then transferred into a 1.5 mL eppendorf tube containing 300 µL protein lysis buffer (0.1 M Tris/Hcl pH 8.0, 0.01 M EDTA, 10 % SDS) supplemented with 1x protease inhibitor (ThermoFisher, A32965). Samples were incubated on ice for 10 mins, following with a sonication at 20 kHz for 1 min. After, samples were

centrifuged at 12,000 rpm for 10 mins. For cultured cells, the dish was washed twice with 1x PBS and cell lysate was prepared in same way as described for tissues sample. For both, the supernatant containing the protein lysate was transferred to a fresh tube and concentration was measured using Bio-Rad assay (BioRad, Reagent A-500-0113, and Reagent B-500-0114) in a 96 wells plate according to the manufacturer's instruction. In brief, samples for a standard regression curve were prepared by using tittered standard concentration of bovine serum albumin (BSA) shown below. Protein concentration was determined by measured and calculated based on the standard curve. For measurement, 200 µL reagent B and 25 µL reagent A were mixed with prepared BSA solution or samples. Mixed samples were incubated for 5 mins at room temperature before subjecting for the measurement. The lysates were considered as isolated protein samples used for Western Blot or stored at -80 ℃.

Table 17. BSA standard solution preparation.

BSA (µL)	1	2	5	10	15	20	0	5
H_2O (µL)	19	18	15	10	5	0	20	15

2.2.12 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis for western blot were self-made according to recipes shown below. In brief, prepared resolving gel was filled in around 80 % of the NOVEX cassettes (1 mm, Invitrogen, NC2010) and about 1 mL pure ethanol was added on top. The filled cassettes were incubated for 30 mins until the mixtures became set. Following, the ethanol was pulled out, the cassettes were filled with the stacking gel and inserted with comb. For western blot assay, the electrophoresis was performed in a running chamber (NOVEXE electrophoresis Mini-Cell, Invitrogen, EI001) using the MES (2.5 mM MES, 2.5 mM Tris, 0.05 % SDS, 50 mM EDTA) buffer for 2 to 3 hours at 120 volts until the samples were separated accordingly. Protein location was detected with the Protein-Marker VI (Peqlab, 27-2311).

Table 18. Composition of gel electrophoresis 10 % Bis-Tris polyacrylamide gel.

Resolving gel	10 %
Rotiphorese® Gel 30 (37.5:1) (Roth, 3029.1)	2.3 mL
3.5 x Bis-Tris pH 6.5 -6.8	2 mL
Milli-Q H_2O	2.7 mL

10 % Ammoniumperoxodisulfat (APS -Merck, 1.01201)	25 µL
TEMED (Roth, 2367.1)	7 µL
Stacking gel	5 %
Rotiphorese® Gel 30 (37.5:1) (Roth, 3029.1)	0.29 mL
3.5 x Bis-Tris pH 6.5 -6.8	0.5 mL
Milli-Q H_2O	0.5 mL
10 % Ammoniumperoxodisulfat (APS -Merck, 1.01201)	0.96 mL
TEMED (Roth, 2367.1)	8 µL
TEMED (Roth, 2367.1)	3 µL

2.2.13 Protein immunoblotting.

Total protein lysate was mixed with 2x protein loading buffer and denatured by boiling at 95 ℃ for 5min before loading on prepared Bis-Tris gels. In brief, wet transfer was used to transfer the proteins from the gel onto a nitrocellulose (AmershamTM ProtranTM 0.45 µm NC -GE Healthcare, 10600002). The protein transfer was achieved with transfer buffer (12.5 mM Bicine, 12.5 mM Bis-Tris, 0.8 mM EDTA, 20 % Methanol). The transfer chamber (NOVEX X Cell II Blot Module -Invitrogen, EI9051) was filled with sponges that had been soaked in transfer buffer. Next, whatman filter paper was placed followed by the gel and then the membrane. Another whatman filter paper and soaked sponges were then added.

This chamber was filled with transfer buffer and run at 30 volts for 1.5 hour. To visualize the loading and a proper transfer, the blots were later stained with Red Alert (RedAlert™ Western Blot Stain –Millipore, 71078-50ML) staining solution on a shaker for 10 mins. The red Alert staining was then removed by washing in 1x TBST. To stain the membrane for the specific antibody, the unstained membrane was firstly blocked with 3 % skim milk dissolved in 1x TBST for 1 hour at room temperate. After blocking, the membrane was incubated with primary antibodies as listed in table 1 dissolved in 3 % skimmed milk in 1x TBST overnight at 4°C on a horizontal shaker. After the incubation, the membrane was washed thrice for 5 mins by 1x TBST on a horizontal shaker. Next, the membrane was incubated with a horse-radish-peroxidase (HRP-) coupled secondary antibody dissolved in 3 % skimmed milk at room temperate for 1 hour on a horizontal shaker. In the end, the membrane was washed in TBST 5 times, each 5 mins. For signal detection, the membrane was incubated with Super Signal West Femto Maximum Sensitivity Substrate (Thermo Scientific; 34096), according to the

manufacturer's instructions and exposed in the Chemiluminescence Analyzer (ChemiDoc™ MP Imaging System -BioRad, 731BR01764) to detect the protein signal. Quantification of band intensities was carried out using the Image J software.

2.2.14 Southern blot.

After digestion by restriction enzymes EcoRV, the genomic DNAs were separated in 1 % agarose gel containing ethidium bromide in 1x TAE buffer at 10 V/cm for 5 hours. Then gel was photographed under UV light and the DNA was transferred to a Hybond-N membrane (Amersham) in alkaline buffer (0.4 M NaOH) by capillary blot for overnight. The transferred DNA was then cross-linked to the membrane by using UV crosslinker with an energy of 1.5 J/cm^2. Then the membrane was first washed in 2x SSC buffer twice and dried on filter paper. For hybridization, the membrane was incubated in Church & Gilbert buffer containing 200 µg/mL denatured herring sperm DNA at 65 ℃ for 2 hours. The internal or 3' probes were ^{32}P-labelled and incubated with membrane at 65 ℃ overnight in a rotary shaker. On the next day, the membrane was washed thrice 15 min at 65 ℃ with fresh washing buffer and applied to X-ray film for exposure.

2.2.15 Frozen samples and cryosections preparation.

Tissues were harvested from mice and kept in a humidified chamber before subjecting for frozen process. 2-methylbutane was poured into an open glass beaker to a depth of approximately 10 cm. The beaker was then put into an insulated container filled with liquid nitrogen. The beaker was ready for sample frozen when the white pellet started forming at bottom. Tissues were then dipped in the beaker for approximately 10 s and put on dry ice for 10 mins for drying. Samples were embedded with OCT at -20 ℃ for 10 mins. Cryosections were obtained by using a microtome (Leica, RM2125RT) and collected on glass slides (Superfrost Ultra Plus-ThermoFisher, J3800AMNZ). Section slides were immediately used in following staining or stored at -80 ℃.

2.2.16 Hematoxylin and Eosin (H&E) staining.

Section slides were dried at room temperature for 10 mins before washing twice in ddH_20 and then were incubated in Hematoxylin buffer (Haemalaum, acidic Mayer -WALDECK, 2E-038) for 3 mins. Slides were put into box filled with the running tap water for at least 10 mins to allow stain developed. Then they were dipped up and down in pure ethanol for 5 times, ddH_20 for thrice and incubated for 2 mins in fresh Eosin buffer (Eosine Solution-WALDECK, 2C-140). After that, slides were dehydrated through 80 %, 95 % and 100 % ethanol for 30 s each and incubated with Xylene for 10 mins. Then slides were mounted with mowiol solution (Sigma, 81381), covered with cover slips, and photographed with a histology microscope (Keyence, BZ-9000).

2.2.17 Immunofluorescence staining.

Cells or tissues sections fixed in 4 % PFA for 10 mins at room temperature in plates or slides, washed thrice with 1x PBS, permeabilized with 0.1 % TritonX-100 in 1x TBST for 10 mins and blocked with 3 % BSA in 1x TBST or M.O.M.® for 1 hour at room temperature. The primary antibody was dissolved in 1 % BSA in 1x TBST or HIKARI A (for mouse spices antibodies) reagent and applied to cells overnight at 4 ℃. On next day, cells were first washed thrice for 10 mins with 1x TBST and then incubated with a fluorophore-conjugated secondary antibody or DAPI dissolved in 1 % BSA in 1x TBST or HIKARI B (for mouse spices antibodies) for 1 hour at room temperature in dark. After that, cells were washed thrice with 1x TBST. All the antibodies incubation performed in a humidified chamber. Slides were mounted with mowiol before covering with cover slips. Pictures were acquired with fluorescent light (ZEISS, Axio Vert.A1) microscopy or confocal microscope (Leica SP8 LIGHTNING).

2.2.18 Immunohistochemistry staining.

Cells or tissues sections were prepared and blocked as described before and applied to boiling sodium citrate buffer for 10 mins. Samples were then incubated with an unconjugated affinity purified F (ab) fragment for 1 hour at room temperature to blocked endogenous IgG. After washing thrice with 1x TBST for 5 mins, appropriately dissolved primary antibody was applied to samples in a humidified chamber for overnight incubation at 4 ℃. The sample was incubated with dissolved biotinylated secondary antibody for 30 mins at room temperature followed with dissolved Sav-HRP conjugated antibody for 1 hour at room temperature in dark.

Freshly prepared DAB substrate solution was applied to sample to reveal the color of antibody staining. The development process was limited in 2 mins for desired color intensity. In the final step, samples were subjected to Hematoxylin for cell nuclei staining.

2.2.19 Mice housing and care.

All animal experiments and transgenic manipulations were performed in accordance with the applicable technological guidelines and animal welfare regulations with the approval according to the regulations issued by the Committee for Animal Rights Protection of the State of Hessen. Mice were kept under a usual 12 hours light/dark cycle at 25 ℃ in cages of the laboratory facility of MPI Bad Nauheim.

2.2.20 Killing of the laboratory mice.

All mice were killed by CO_2 and followed by cervical dislocation according to laboratory mice euthanasia protocol with the approval according to the regulations issued by the Committee for Animal Rights Protection of the State of Hessen.

2.2.21 Tamoxifen (TAM) administration.

TAM was prepared as 20 mg/mL stock solution in corn oil and injected to mice at 100 mg/kg body weight. Mice received TAM administration at the age of 8 weeks for 10 consecutive days. Samples were harvested and analyzed according to laboratory mice protocol with the approval according to the regulations issued by the Committee for Animal Rights Protection of the State of Hessen.

2.2.22 Statistical analysis.

Statistical analysis was performed by unpaired, two-tailed Student's t-test, or by paired, one-tailed Student's t-test when a normal distribution was assumed, or by one-way ANOVA followed by Bonferroni's multiple comparison test. For all bar graphs, data are represented as

mean ± S.E.M, P values < 0.05 were considered significant. n represents the number of independent experiments. All calculations were performed using GraphPad Prism8.0software.

3. Results.

3.1 Inactivation of p53 in regenerating MuSCs is sufficient to induce RMS tumor formation in mice.

Previously it was shown that germline inactivation of the tumor suppressor *p53* in mdx mice, which undergo continuous muscle de- and regeneration, significantly increases the rate of RMS tumor formation (Camboni et al., 2012). Similar to human patients, these tumors were classified as RMS tumors based on the expression of myogenic markers including *Desmin*, *MyoD* and *MyoG*. (Camboni et al., 2012; Fernandez et al., 2010b; Hosur et al., 2012b). These observations raised the question whether RMS tumors might originate from cells of the muscle lineage, and specifically from muscle resident MuSCs under conditions of active muscle regeneration.

To test this idea, mice were generated that enable MuSC-specific deletion of *p53* in both non-regenerating (wild-type) or regenerating (mdx) mice by administration of tamoxifen (TAM), hereafter referred to as $Pax7^{Cre/ERT2}$; $p53^{flox/flox}$; $mdx^{+/+}$ and $Pax7^{Cre/ERT2}$; $p53^{flox/flox}$; $mdx^{-/-}$, respectively. Strikingly, all $Pax7^{Cre/ERT2}$; $p53^{flox/flox}$; $mdx^{-/-}$ mice developed tumors in or close to trunk and limb muscles around 20 weeks after TAM administration (Figure 6A and B). The mdx mice, in which only one *p53* allele was inactivated in MuSCs ($Pax7^{Cre/ERT2}$; $p53^{flox/+}$; $mdx^{-/-}$), developed tumors in average 10 weeks later than $Pax7^{Cre/ERT2}$; $p53^{flox/flox}$; $mdx^{-/-}$ mice (Figure 6A). Importantly, control mice ($Pax7^{Cre/ERT2}$; $p53^{flox/flox}$; $mdx^{+/+}$ and $Pax7^{Cre/ERT2}$; $p53^{+/+}$; $mdx^{-/-}$) never developed tumors up to an age of 52 weeks (Figure 6A). Notably, $Pax7^{Cre/ERT2}$; $p53^{flox/flox}$; $mdx^{-/-}$ mice sometimes developed several tumors in different muscles (Figure 6B). Immunohistochemical analysis of surgically excised tumors revealed hypercellularity and presence of rhabdomyoblasts of spindled morphology (Figure 6C). All tumors were positive for *Desmin*, *MyoG* and *MyoD*, myogenic markers that classify tumors as RMS in humans (Figure 6C) (Altmannsberger et al., 1985; Cessna et al., 2001; Sebire and Malone, 2003). Together, these data show that MuSCs-specific inactivation of *p53* in regenerating mdx mice is sufficient to induce formation of RMS tumors *in vivo*.

Because tumors never formed in non-regenerative, steady state conditions in $Pax7^{Cre/ERT2}$; $p53^{flox/flox}$; $mdx^{+/+}$ mice, it was speculated that continuous regeneration and hence enhanced proliferation cycles, would facilitate acquisition of spontaneous tumorigenic

mutations in the absence of *p53*. In support of this idea, cultured MuSCs purified from TAM-treated $Pax7^{Cre/ERT2}$; $p53^{flox/flox}$; $mdx^{-/-}$ mice displayed significantly enhanced EdU-incorporation rates and increased levels of *yH_2Ax* and *53bp1* compared to non-treated control mice (Figure 6D-F).

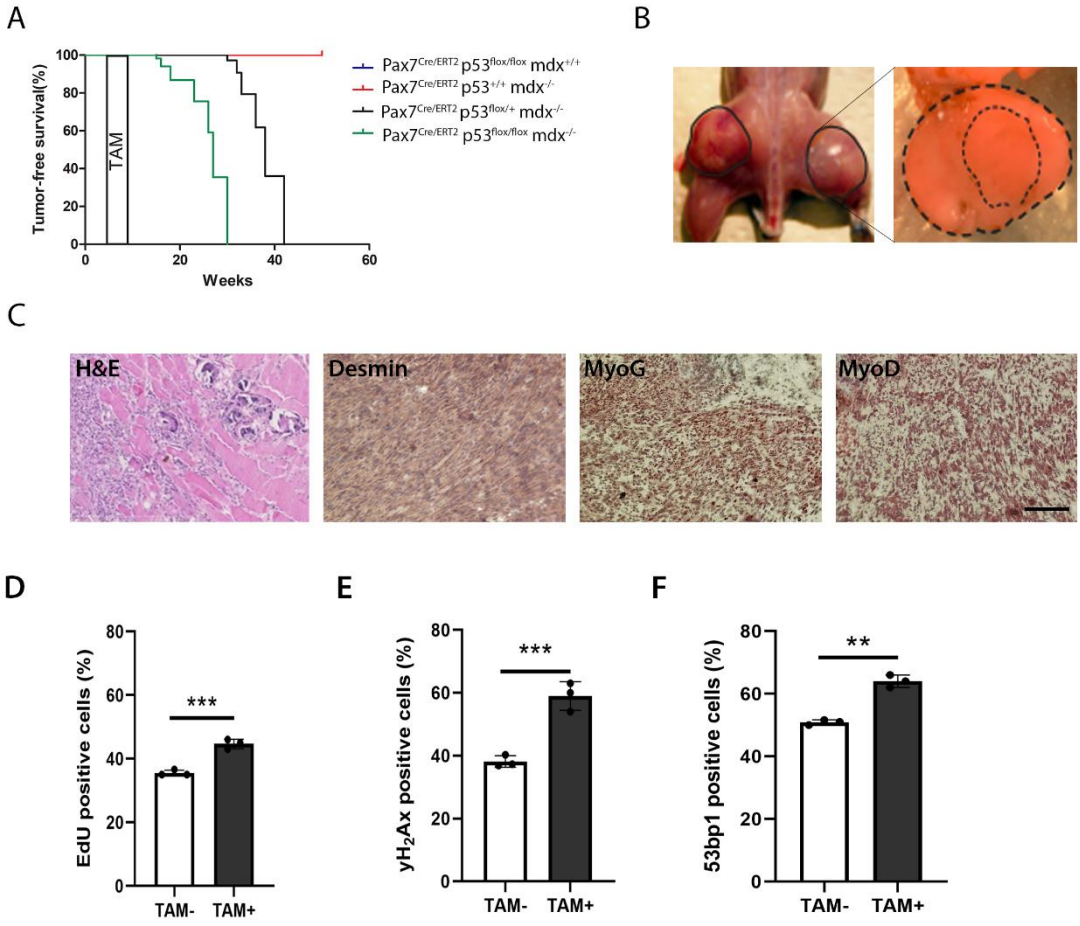

Figure 6. Loss of *p53* in regenerating MuSCs induces RMS tumor formation.

(A) Kaplan-Meier tumor-free survival curves are shown for indicated genotypes.

(B) Representative images of RMS tumors (left panel) and an isolated tumor (right panel).

(C) Representative immunohistochemical staining of cross-sectioned tumor with H&E or indicated antibodies

(D-F) Quantification of percentage of (D) EdU, (E) *yH_2Ax* and (F) *53bp1*-positive MuSCs isolated from TAM-treated $Pax7^{Cre/ERT2}$; $p53^{flox/flox}$; $mdx^{-/-}$ mice and cultured under growth conditions. Error bars indicate SD of the mean (t-test: *p<0.05; **p<0.01; ***p<0.001).

Figure adapted from (Preussner et al., 2018) with modification.

3.2 Lineage tracing identifies MuSCs as a cellular origin of RMS tumors.

To confirm that MuSCs were the cellular origin of RMS tumors in Pax7$^{Cre/ERT2}$; p53$^{flox/flox}$; mdx$^{-/-}$ mice, a *Rosa26::lsl Tomato* allele was introduced, enabling permanent fluorescent lineage tracing of MuSCs after TAM treatment. Furthermore, a *Pax7::ZsGreen* allele was introduced, a transgene from which the *ZsGreen* gene is driven by the *Pax7* promoter and thereby marks undifferentiated MuSCs with green fluorescence (Pax7$^{Cre/ERT2}$; p53$^{flox/flox}$; mdx$^{-/-}$; Pax7^{ZsGreen}; Rosa26^{Tomato}) (Figure 7A). After TAM treatment, cells originating from *Pax7::ZsGreen* expressing MuSCs are permanently labeled by red fluorescence through activation of the *Rosa26::lsl Tomato* allele (Figure 7A).

FACS-based purification of MuSCs isolated from Pax7$^{Cre/ERT2}$; p53$^{flox/flox}$; mdx$^{-/-}$; Pax7^{ZsGreen}; Rosa26^{Tomato} mice via the *Pax7::ZsGreen* reporter revealed that essentially all MuSCs displayed intense Tomato fluorescence due to activation of the *Rosa26::lsl Tomato* allele after TAM treatment, indicating the genetic strategy to be highly effective for fluorescent lineage tracing of MuSCs (Figure 7B). Consistently, strong Tomato fluorescence of the skeletal muscles, but not the liver, revealed prominent and specific contribution of MuSCs toward *de novo* myofiber formation in chronically regenerating mdx muscles (Figure 7C). Interestingly, Tomato fluorescence was also detected in the brain consistent with the expression of *Pax7* in the central nervous system (Bandín et al., 2013; Gruss and Walther, 1992; Wehr and Gruss, 1996) (Figure 7C).

Remarkably, all tumors in TAM-treated Pax7$^{Cre/ERT2}$; p53$^{flox/flox}$; mdx$^{-/-}$; Pax7^{ZsGreen}; Rosa26^{Tomato} mice were labelled by activation of *Rosa26::lsl Tomato allele* (Figure 7C) clearly demonstrating that MuSCs are the cellular origin of RMS tumors in these animals. It should be noted, however, that the experiment did not answer the question whether only undifferentiated *Pax7* expressing MuSCs or their more differentiated progeny (e.g. *Pax7*-negative, *Myf5*-positive MuSC descendants) contributed to RMS tumor formation, since all cells of the myogenic lineage are lineage-traced upon TAM administration.

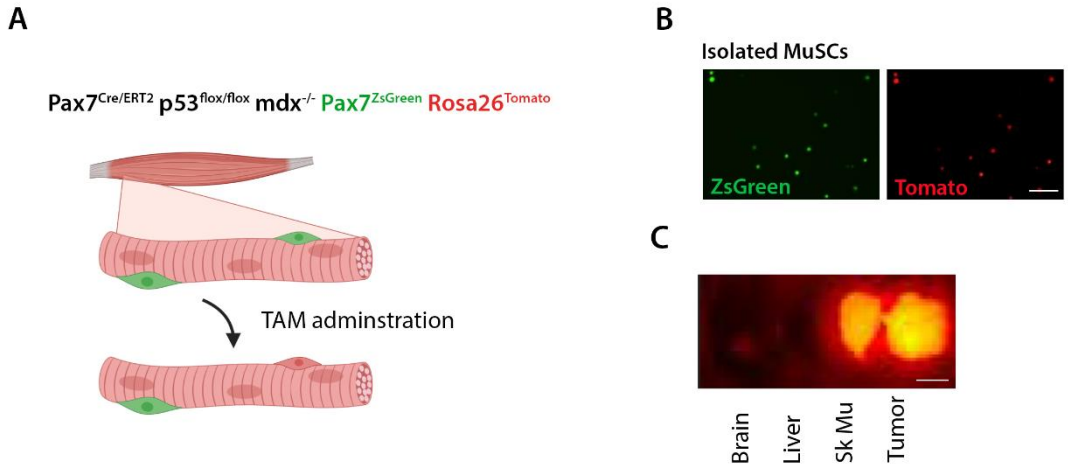

Figure 7. Lineage tracing reveals MuSCs as a cellular origin of RMS tumors.

(A) Genetic components of the mouse model. MuSCs expressing ZsGreen are lineage-traced *in vivo* via recombination of a *Rosa26::lsl Tomato* allele upon TAM administration.

(B) Representative immunofluorescence images of isolated MuSCs from $Pax7^{Cre/ERT2}$; $p53^{flox/flox}$; $mdx^{-/-}$; $Pax7^{ZsGreen}$; $Rosa26^{Tomato}$ mice.

(C) Representative immunofluorescence images of isolated tissues from $Pax7^{Cre/ERT2}$; $p53^{flox/flox}$; $mdx^{-/-}$; $Pax7^{ZsGreen}$; $Rosa26^{Tomato}$ mice.

Figure adapted from (Preussner et al., 2018) with modification.

3.3 RMS tumors comprise different cell populations.

Next, single cell isolates obtained from surgically excised RMS tumors were subjected to FACS analysis revealing that RMS tumors were composed of both lineage-traced and non-lineage-traced cells, accounting for 57.2 % and 42.8 % of cells within RMS tumors, respectively (Figure 8A). PCR-based genotyping confirmed that recombination of the *p53* locus was highly efficient and specific to lineage-traced tumor cells (TPCs Tom^{pos}) that were strictly *p53* deficient, whereas non-lineage-traced tumor cells (TPCs Tom^{neg}) only contained intact non-recombined, *p53* DNA (data not shown). Essentially, FACS-separated TPCs were either strictly lineage-traced containing recombined *p53* alleles, or non-lineage-traced containing intact *p53* alleles that did or did not express *Myf5*, *MyoD*, or *MyoG*, respectively (Figure 8B). These data further confirm efficient labeling of *p53*-deficient MuSCs. In contrast to TPCs Tom^{pos} , TPCs Tom^{neg} were highly enriched for *Cdkn1a* mRNA transcripts (also known as *p21*), a primary target of *p53* (Laptenko and Prives, 2006) (Figure 8B), emphasizing that loss of *p53* occurs specifically in MuSC-derived cells which express

myogenic markers. In summary, these results show that RMS tumors comprise different cell types that are and are not derived from *Pax7*-expressing MuSCs.

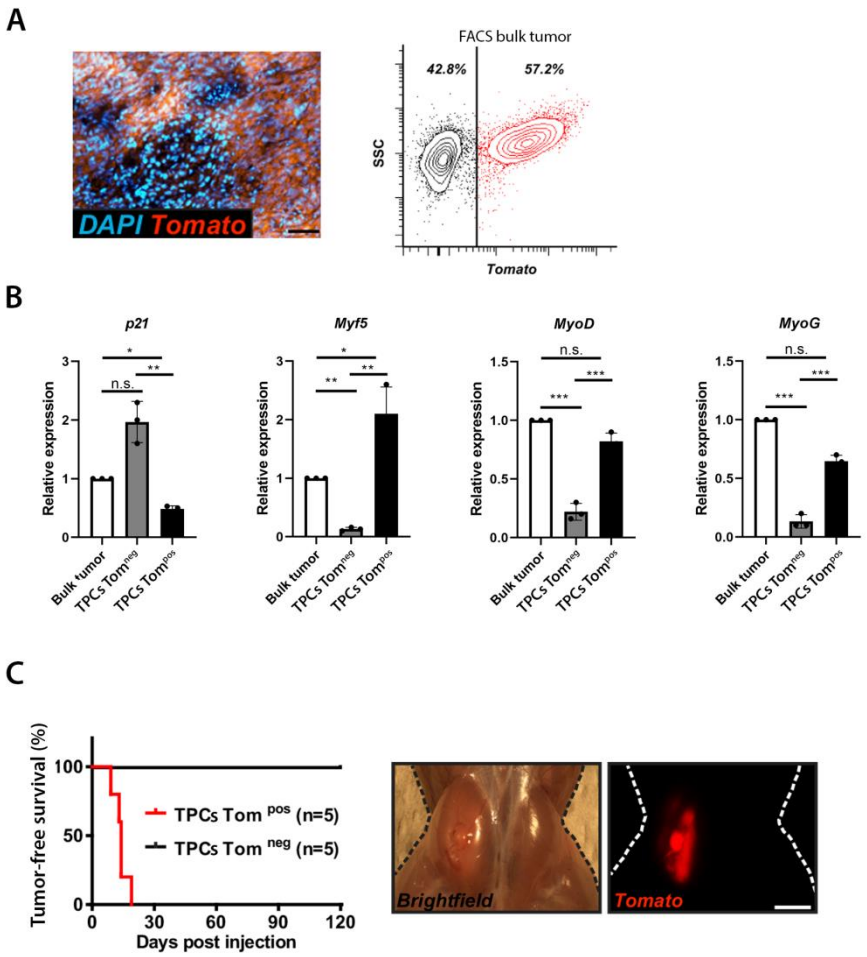

Figure 8. Composition of RMS tumors.

(A) Representative immunofluorescence images of cross-sectioned tumor (left panel) and FACS plot of bulk tumor.

(B) Transcriptional expression analysis of *p21* and myogenic markers in corresponding populations. Error bars indicate SD of the mean (t-test: *p<0.05; **p<0.01; ***p<0.001).

(C) Kaplan-Meier curves of tumor free survival of mice injected with TPC Tom^{pos} or TPC Tom^{neg} (left panel); Macroscopic image of Tomato fluorescence in secondary tumors (right panel).

Figure adapted from (Preussner et al., 2018) with modification.

Next, it was tested whether both of these two cell populations contributes to RMS tumor formation. To achieve this aim, FACS-separated TPC Tom^{pos} or TPC Tom^{neg} were injected into immunocompromised recipient mice (Figure 8C). Notably, only mice injected

with TPC Tom^{pos} developed tumors at site of injection and remarkably within only two weeks (Figure 8C). These data show that only lineage-traced tumors cells (TPC Tom^{pos}) derived from *Pax7*-expressing MuSCs behave as tumor-propagating cells. It is noteworthy that the immunocompromised mice were not subjected to regenerative pressures and contained intact *p53* alleles. Therefore, regeneration or *p53*-deficiency of the tumor microenvironment does not appear to be necessary for propagation of established RMS tumor cells *in vivo*.

3.4 Genomic analyses of purified TPCs reveal different types of mutations.

DNA copy-number variations (CNVs) are common in human RMS patients (Paulson et al., 2011).To identify CNVs that might be responsible for tumorigenic transformation of MuSCs, comparative whole-exome sequencing was performed on paired tumor-normal samples (TPCs versus liver) of TAM-treated Pax7$^{Cre/ERT2}$; p53$^{flox/flox}$; mdx$^{-/-}$ mice (Figure 9A). Genomic DNA isolated from individual whole tumors was compared to genomic DNA isolated from the liver of the same mice. Surprisingly, this approach failed to consistently detect CNVs in tumor DNA (Figure 9A left panel). Since tumors were composed of TPC Tom^{pos} and TPC Tom^{neg} (Figure 8A), it was hypothesized that the abundance of TPC Tom^{neg} within tumors might interfere with the analysis. Therefore, the analyses were repeated using genomic DNA purified from TPC Tom^{pos} cells only. Strikingly, this strategy led to the identification of distinct CNVs in nearly all analyzed samples (96 %, 21 out of 22) (Figure 9A and B). Positional mapping of the sequenced reads led to the identification of genomically amplified genes which have previously been shown to support RMS tumor formation, including *Jun* (Durbin et al., 2009) (4.5 %, 1 out of 22), *Met* (Fleischmann et al., 2003; Taulli et al., 2006) (23 %, 5 out of 22), *Yap1* (Tremblay et al., 2014) (36 %, 8 out of 22) and *Cdk4*/*Os9* and *Gli1* (Liu et al., 2014) (4.5 %, 1 out of 22). In addition, other CNV were identified that had not been associated with RMS tumors so far, including *Rras2* (Flex et al., 2014) (4.5 %, 1 out of 22) and *Kdm4c* (Soini et al., 2015) (17 %, 3 out of 22) (Figure 9C). A direct linkage between the identified genes was not found (except to the recombined *p53* gene), suggesting that for each individual tumor, a single discrete amplification was sufficient for transformation of MuSC. Interestingly, it appeared that CNV mutations did not occur at random sites since they often occurred at few genomic loci (Figure 9C).

qPCR analysis revealed that the copy number amplifications correlated with expression of the amplified genes therein (Figure 9D). For example, in TPC Tom^{pos} harboring CNV of the *Yap1* gene or *Duxbl* gene, designated $TPCs^{Yap1}$ and $TPCs^{Duxbl}$, respectively, *Yap1* expression was restricted to $TPCs^{Yap1}$ but was not expressed in $TPCs^{Duxbl}$ (Figure 9D). Likewise, *Duxbl* expression was significantly higher in $TPCs^{Duxbl}$ in comparison to $TPCs^{Yap1}$ (Figure 9D).

To test whether these expressed genes play functional roles in tumor cell propagation, short hairpin RNA (shRNA)-mediated knockdown was applied to individual cultured TPCs that were isolated from the primary tumors in which the CNV were found. Strikingly, shRNA-mediated knockdown of *Duxbl* in TPCs harboring a *Duxbl* CNV ($TPCs^{Duxbl}$) but not in TPCs harboring a *Yap1* CNV ($TPCs^{Yap1}$), and *vice versa*, resulting in cell death (Figure 9E). These data indicate that *Duxbl* expression is specifically required for the maintenance of $TPCs^{Duxbl}$ and that *Yap1* expression is specifically required for maintenance of $TPCs^{Yap1}$ *in vitro*. While the underlying mechanisms remain unclear, the data show that the genetic strategy combined with sequencing of purified TPCs allows identification of specific oncogenes that play important roles in RMS tumor propagation and/or maintenance. In addition, these data strongly suggest that disease-causing mutations need to be identified in each individual tumor to achieve specific therapeutic targeting.

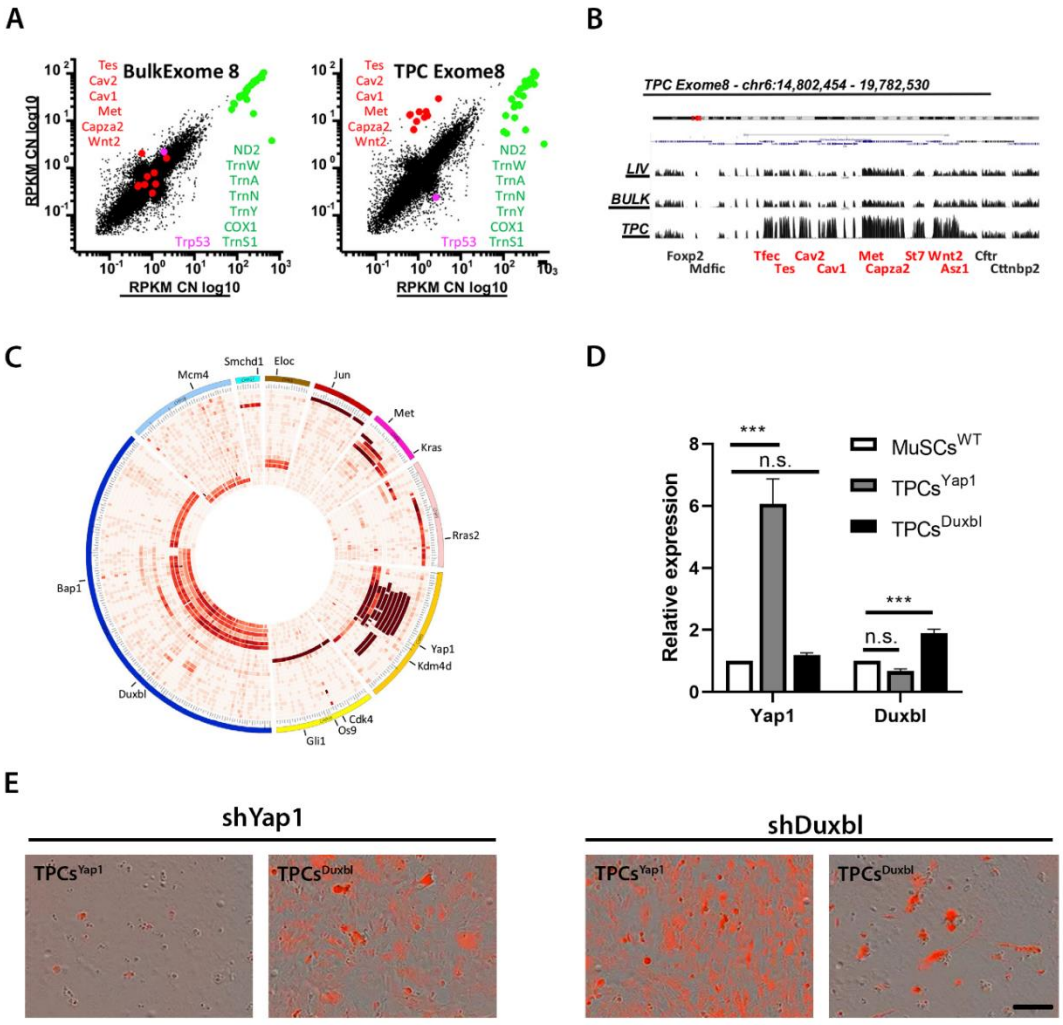

Figure 9. Identification of CNVs in purified TPCs.

(A) Scatterplots depicting log-scaled RPKM values of genomic DNA in bulk tumor (left panel) or TPCs (right panel) versus liver control. Note that same genes are highlighted in green for the mitochondria-encoded genes and red for distinct amplified genes. Magenta circle represents *p53*.

(B) Amplified genes highlighted in red in A displayed in physical genomic order.

(C) Cumulative CIRCOS plot of amplified genes from all 21 analyzed tumors. Note that genomic regions are displayed in physical order.

(D) qRT-PCR analysis of *Yap1* and *Duxbl* in TPCs^{Yap1} or TPCs^{Duxbl} versus wild-type MuSCs (MuSCs^{WT}) as control. Expression levels are normalized to m36b4 mRNA. Error bars indicate SD of the mean (t-test: *p<0.05; **p<0.01; ***p<0.001).

(E) Representative immunofluorescence images of TPCs^{Yap1} and TPCs^{Duxbl} four days after transduced with different shRNA against either *Yap1* or *Duxbl*.

Figure adapted from (Preussner et al., 2018) with modification. Bioinformatical analysis presented in (Figure 9 A-C) was performed by Dr. Jens Preussner.

3.5 Dux transcription factors define a molecular subtype of cancer.

Of the different amplified genes within the CNVs of all analyzed tumors, *Duxbl* was of particular interest, because CNVs of this gene were highly recurrent and *Duxbl* has never been associated with oncogenesis so far. In fact, *Duxbl* was only recently identified to belong to the *Dux* homeobox-containing transcription factors (*Dux* TFs) (Leidenroth and Hewitt, 2010b). In humans, three major *Dux* TFs were identified, known as *Dux4*, *DuxA* and *DuxB* (Leidenroth and Hewitt, 2010b). In mice, two major *Dux* TFs exist, designated *Dux* and *Duxbl* (Leidenroth and Hewitt, 2010a). All *Dux* TFs contain two identical homeobox domains responsible for DNA binding (Eidahl et al., 2016). The founding family member of the family, *Dux4*, is the best studied *Dux* TF because of its pathological role in human diseases, including facioscapular muscular dystrophy (FSHD) and cancer (Himeda and Jones, 2019; Jones et al., 2015). Under normal conditions, the *Dux4* gene locus (D4Z4) is heavily methylated and silenced (Das and Chadwick, 2016). Recently it was shown that genomic translocations of the *Dux4* gene result in expression of chimeric DUX4-IGH or DUX4-CIC fusion products, which are causative for inducing B-cell acute lymphoblastic leukemia (B-ALL) or sarcomas, respectively (Kawamura-Saito et al., 2006; Lilljebjorn and Fioretos, 2017; Yasuda et al., 2016a, b; Zhang et al., 2016) (Figure 10A). These observations suggest that *Dux* TFs might be involved in other types of cancer including RMS tumors.

To investigate this, RNA sequencing data from The Cancer Genome Atlas-Pan-Cancer database (TCGA-PANCAN) of over 10,000 tumors were rescreened for the expression of *Dux* TFs. Strikingly, 349 patients (3.49 %) were identified that displayed distinct *Dux4*, *DuxA* and *DuxB* expression either in alone or combination (Figure 10B). The cancer types of these patients were highly variable, comprising of 32 different types of somatic cancer, including RMS tumors, according to ICD-1O (International Classification of Diseases of Oncology) (Figure 10B). Interestingly, many tumors expressed genes that are highly associated with zygotic genome activation (ZGA), a developmental program driven by *Dux* TFs in preimplantation embryos (De Iaco et al., 2017) (Figure 10B). These results suggest that a ZGA-like program driven by *Dux* TFs might induce tumorigenesis in these patients. More importantly, these analyses support the idea that *Duxbl*, as a family member of *Dux* TFs, plays an important role in RMS tumor formation.

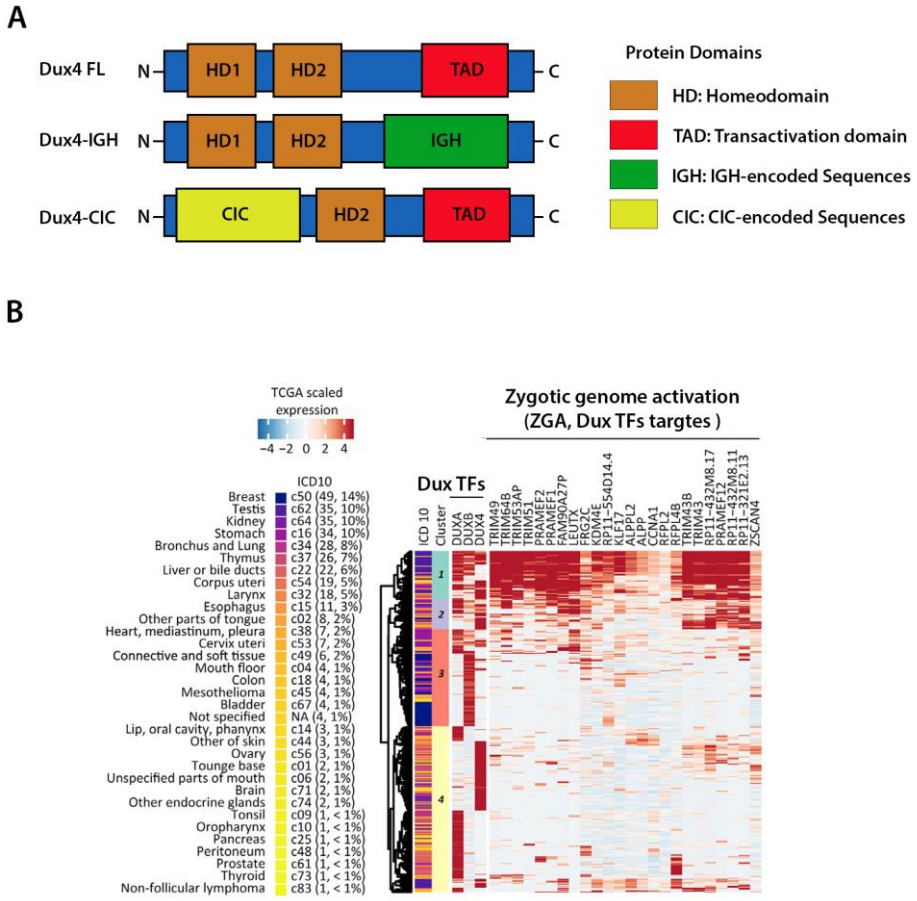

Figure 10. Dux TFs define a molecular subtype of cancer.

(A) Schematic representation of various forms of *Dux4*. In oncogenic DUX4-IGH or DUX-CIC chimeric proteins C-terminal or N-terminal portions of *Dux4* are replaced by either IGH- or CIC-encoded amino-acid sequences.

(B) Unsupervised cluster analysis of tumors from the full TCGA-PANCAN dataset revealing four subgroups driven by mRNA expression of *Dux* TFs and/or *Dux4*-dependent zygotic gene activation. Percentages indicate prevalence of color-coded tumor type across 349 *Dux* TFs and/or ZGA-positive tumors.

Figure adapted from (Preussner et al., 2018) with modification. Bioniformatical analysis presented in (Figure 10 B) was performed by Dr. Jens Preussner

3.6 Duxbl overexpression alters the myogenic program of MuSCs in vitro.

Next, a gain-of-function approach was used to investigate the impact of *Duxbl* overexpression on MuSCs. Wide-type MuSCs were isolated and infected with retroviruses harboring a *Duxbl* expression cassette or vehicle, hereafter referred to as MuSCs^{Duxbl} or MuSCs^{vehicle}, respectively. In comparison to MuSCs^{vehicle}, MuSCs^{Duxbl} displayed significantly enhanced proliferation rates as determined by live cell imaging (Figure11 A).

Consistent with this, MuSCs^{Duxbl} displayed significantly increased EdU positive cells (53 %) in comparison to MuSCs^{vehicle} (30 %) (Figure 11 B). Enhanced proliferation rates upon *Duxbl* overexpression were also detected in HEK293T and NIH3T3 cells (data not shown). These data show that *Duxbl* promotes cell proliferation of both myogenic and non-myogenic cells

To test whether *Duxbl* affects myogenic differentiation, MuSCs^{Duxbl} or MuSCs^{vehicle} were expanded to confluency and subsequently exposed to low mitogen conditions for induction of myogenic differentiation (Figure 11C). Cell fusion and differentiation indices were determined by counting numbers of myonuclei in *MF20* immunostained differentiated myotubes (Figure 11C). MuSCs^{vehicle} exhibited robust formation of multinucleated *MF20* positive myotubes, whereas differentiation of MuSCs^{Duxbl} was severely impaired (Figure 11C). Consistent with this observation, expression levels of *MyoD* and *MyoG* were significantly decreased in MuSCs^{Duxbl} (Figure 11D). In addition, *Duxbl* overexpression increased expression of the cell cycle promoting gene *Ccnd1* but reduced expression of the cell cycle inhibitor *p21* (Figure 11D) (Wu et al., 2014). Interestingly, MuSCs^{Duxbl} displayed increased *Ccnb1* expression levels while *Ccna1* expression levels were slightly increased, suggesting that the presence of *Duxbl* in MuSCs inhibits S to G2/M phase transition (Figure 11E). Taken together, these results indicate that *Duxbl* overexpression enhances cell proliferation at the expense of myogenic differentiation, providing a potential mechanism of *Duxbl*-dependent tumorigenesis.

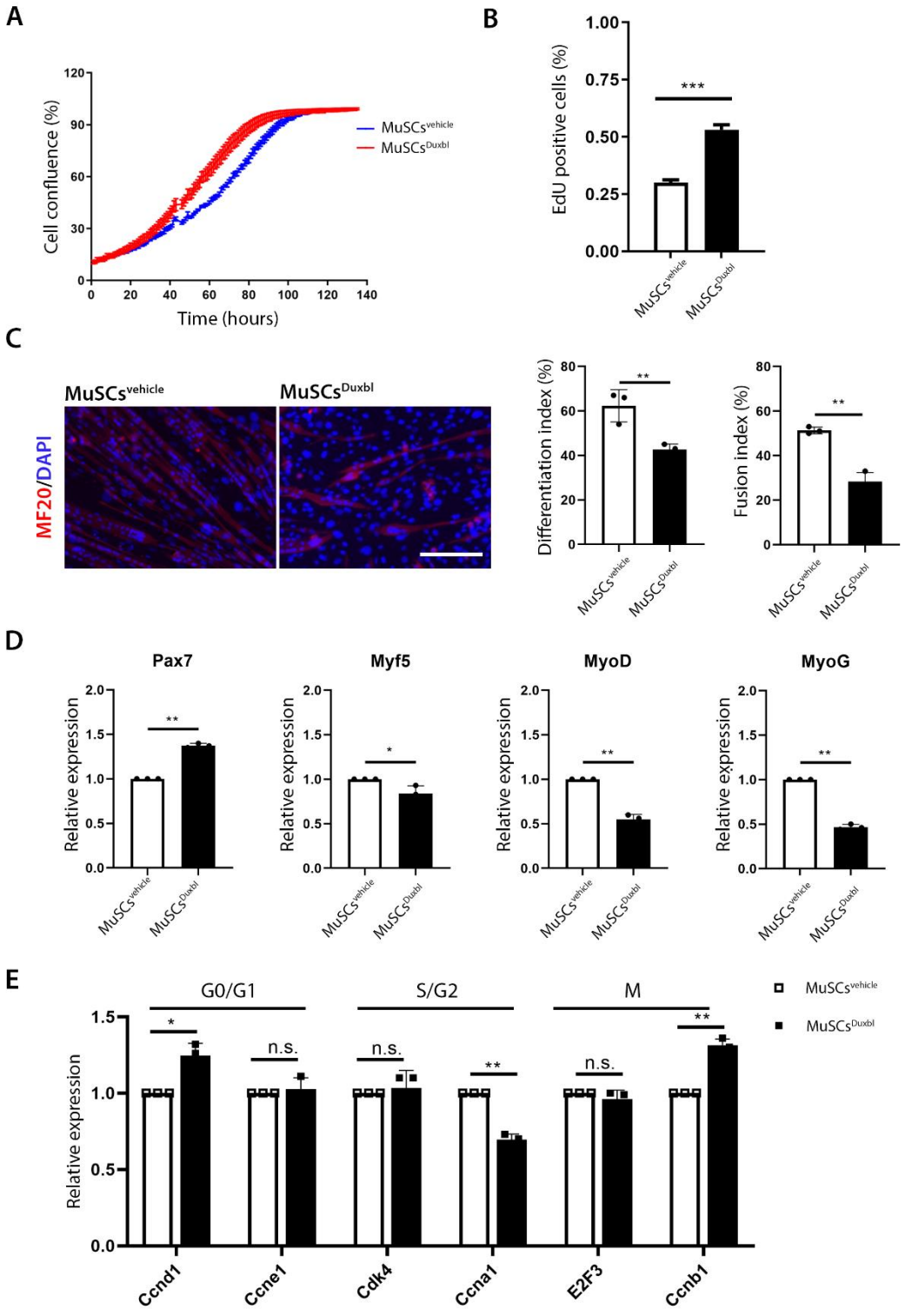

Figure 11. Duxbl overexpression enhances MuSCs proliferation and inhibits differentiation.

(A) Cell proliferation analyzed by live cell imaging over one-week culture. Error bars indicate SD of the mean.

(B) Quantitative analysis of EdU incorporation of MuSCs in growth conditions.

(C) Representative immunofluorescence images of differentiated MuSCs infected with vehicle or Duxbl- overexpressing retroviruses. MF20 and DAPI are represented in red and blue, respectively (left panel); MF20 positive fibers and DAPI positive

nuclei are counted and quantified into differentiation index and fusion index for each cell preparation (right panel).

(D-E) Transcriptional expression analysis of myogenic markers (D) and cell cycle genes (E) in corresponding cells. Error bars indicate SD of the mean (t-test: *p<0.05; **p<0.01; ***p<0.001).

3.7 Generation of Rosa26 transgenic mice for conditional Duxbl overexpression.

Next, a mouse model allowing conditional overexpression of *Duxbl* was generated to investigate the functional roles of *Duxbl in vivo*. To achieve this goal, a V5-epitope tagged *Duxbl* coding sequence was cloned into a generic shuttle vector (pBigT) containing a neomycin resistance cassette flanked by two loxP sites through NheI and NotI restriction sites (Figure 12A). From this intermediate shuttle vector, the *Duxbl* expression cassette was excised and cloned into a Rosa26 destination vector (pRosa26) through PacI and AscI restriction sites (Figure 12A). The resulting floxed-neoPA-*Duxbl* cassette is flanked with sequences permitting homologous recombination into the genomic *Rosa26* locus once transfected into targeted cells. The targeting vector also contains a diphtheria toxin A (DTA) cassette to prevent random transgene integration (Figure 12A).

Next, the constructed destination vector was linearized by AflIII enzyme restriction, and electroporated it into SV129/C57BL6 E14 mouse embryonic stem cells (ES). In total, 48 candidate ES colonies were obtained after G418 selection (Figure 12B). Southern blot assay revealed that 12 out of 48 (25 %) clones displayed the expected insertion of the floxed-neoPA-Duxbl cassette into the *Rosa26* locus (Figure 12B). PCR-based genotyping was used to validate the insertion and revealed that 4 out of 12 (33 %) candidates were homozygous for the $Duxbl^{V5}$ insertion (Figure 12C). Importantly, infection of candidate $Duxbl^{V5}$ ES clones with an adenovirus harbouring Cre recombinase resulted in high expression levels of Duxbl protein indicating functional activity of the engineered allele (Figure 12D). Injection of $Duxbl^{V5}$ ES cells into blastocysts resulted in chimeric founder mice, offspring of which were backcrossed for stable germline transmission of the conditional $Duxbl^{V5}$ overexpression allele.

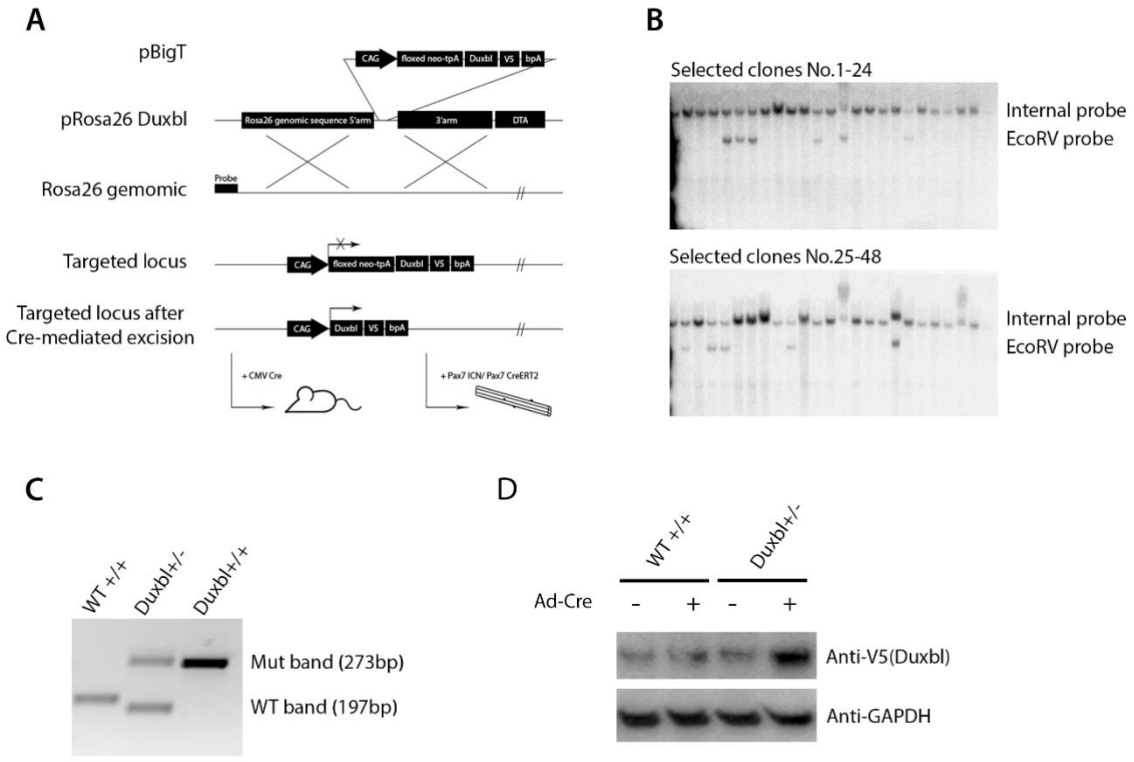

Figure 12. Transgenic Rosa26 knock in mouse model for conditional Duxbl expression.

(A) Schematic construction diagram of the conditional Duxbl mouse model.

(B) Southern blot of selected 48 ES colonies electroporated with constructed vector. Internal probe indicates the wild-type Rosa26 allele, the EcoRV probe indicates the insertion allele.

(C) PCR-based genotyping of representative ES colonies. The lower 197bp band indicates the wild-type Rosa26 allele, the upper 273bp indicates the $Duxbl^{V5}$ insertion allele.

(D) Western blot of representative ES colonies infected with Ad-Cre to evaluate exogenous *Duxbl* expression. GAPDH (lower panel) expression levels were used as loading controls.

3.8 Characterization of Duxbl overexpression mice.

To understand the impact of *Duxbl* overexpression *in vivo*, $Duxbl^{+/-}$ mice were bred with either $CMVCre^{+/-}$ mice and $Pax7^{Cre/ERT2}$ mice resulting ubiquitously expression and MuSCs-specific expression of *Duxbl* upon TAM administration, respectively. Interestingly, when breeding $Duxbl^{+/-}$ mice with $CMVCre^{+/-}$ mice, no offspring (65 offspring obtained from 5 independent crosses) harboring both the Cre-recombinase and the *Duxbl* allele was obtained, indicating that overexpression of *Duxbl* during development resulted in embryonic

lethality (Figure 13A). Since *Dux* TFs are normally only transiently expressed during very early stages of preimplantation development (Sugie et al., 2020), it is conceivable to speculate that preventing downregulation of *Duxbl* through sustained overexpression during preimplantation development might be responsible for embryonic lethality.

To investigate how *Duxbl* overexpression affects MuSC function, $Duxbl^{+/+}$ mice were bred with $Pax7^{Cre/ERT2}$ mice enabling overexpression of *Duxbl* specifically in MuSCs upon TAM administration. TAM-treated $Pax7^{Cre/ERT2}$ $Duxbl^{+/+}$ mice (TAM+) remained viable and displayed no obvious phenotype under physiological conditions compared with non-treated control mice (TAM-). No significant changes of body weight and TA muscle weight were apparent (Figure 13B). Interestingly, however, the percentage of muscle fibers with centrally located nuclei was significantly increased in TAM-treated $Pax7^{Cre/ERT2}$ $Duxbl^{+/+}$ mice compared with control mice indicating that overexpression of *Duxbl* induces precautious activation of MuSCs, which subsequently contribute to myofiber formation (Figure 13C). Consistently, *Pax7* and *MyoD* double positive MuSCs were found in single myofibers of TAM-treated $Pax7^{Cre/ERT2}$ $Duxbl^{+/+}$ mice but not in control mice (Figure 13D).

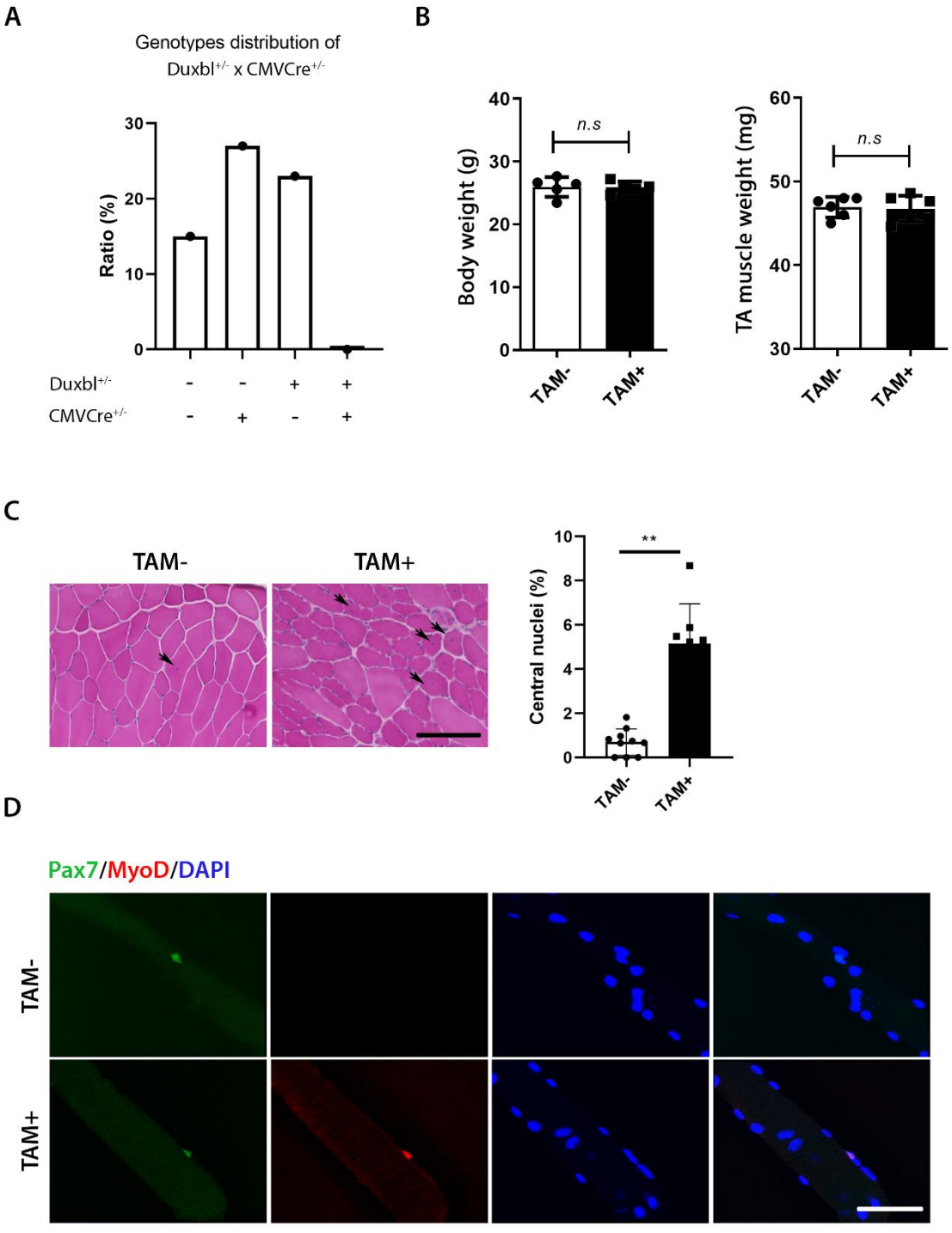

Figure 13. Characterization of Duxbl overexpression mice.

(A) Distribution of genotypes obtained from offspring through matings with Duxbl+/- and CMVCre+/- mice. Offsprings were obtained and quantified from five independent crosses.

(B) Body weight (left panel) and TA muscle weight (right panel) of control mice (TAM-) or TAM-treated Pax7$^{Cre/ERT2}$; Duxbl$^{+/+}$ mice at 12 weeks old.

(C) Representative histological H&E staining images (left panel) of cross-sectioned TA muscle from control mice (TAM-) or TAM-treated $Pax7^{Cre/ERT2}$; $Duxbl^{+/+}$ mice. Black arrows indicate myofibers with central nuclei. Percentage of TA muscle myofibers with central nuclei (right panel). Error bars indicate SD of the mean (t-test: n.s.: not significant; *p<0.05; **p<0.01; ***p<0.001).

(D) Representative immunofluorescence images of cross-sectioned TA muscle from control mice (TAM-) or TAM-treated $Pax7^{Cre/ERT2}$; $Duxbl^{+/+}$ mice. *Pax*7, *MyoD* and DAPI are represented in green, red and blue, respectively

3.9 Duxbl overexpression in adult MuSCs impairs muscles regeneration.

The observation that *Duxbl* overexpression inhibits myogenic differentiation of MuSCs (Figure 11), together with the findings that RMS tumors are often associated with expression of *Duxbl* from CNVs acquired during muscle regeneration (Figure 9), prompted to investigate how *Duxbl* overexpression affects MuSCs functions specifically during muscle regeneration. To induce muscle regeneration, cardiotoxin (CTX) was injected into tibialis anterior (TA) muscles of $Pax7^{Cre/ERT2}$; $Duxbl^{+/+}$ mice treated with and without TAM administration, respectively (Figure 14). In comparison to control mice, CTX mediated muscle injury of TAM+ mice led to a significant decrease in TA muscle sizes and pronounced fibrosis (Figure 14B). The increase of fibrosis in TAM+ mice was evaluated by histological H&E staining in cross-sectioned TA muscles (Figure 14C). In contrast to those of TAM+ mice, the injured TA muscles of TAM- mice showing minimal fibrosis were completely replaced by newly generated myofibers containing centrally located nuclei 14 days after injury (Figure 14C). In addition, significantly increased amounts of *eMHC* positive myofibers were visible in TAM+ mice compared to TAM- mice (Figure 14D). Consistent with the *in vitro* analyses, these data show that overexpression of *Duxbl* in MuSCs inhibits MuSC differentiation *in vivo* consequently impairing muscle regeneration.

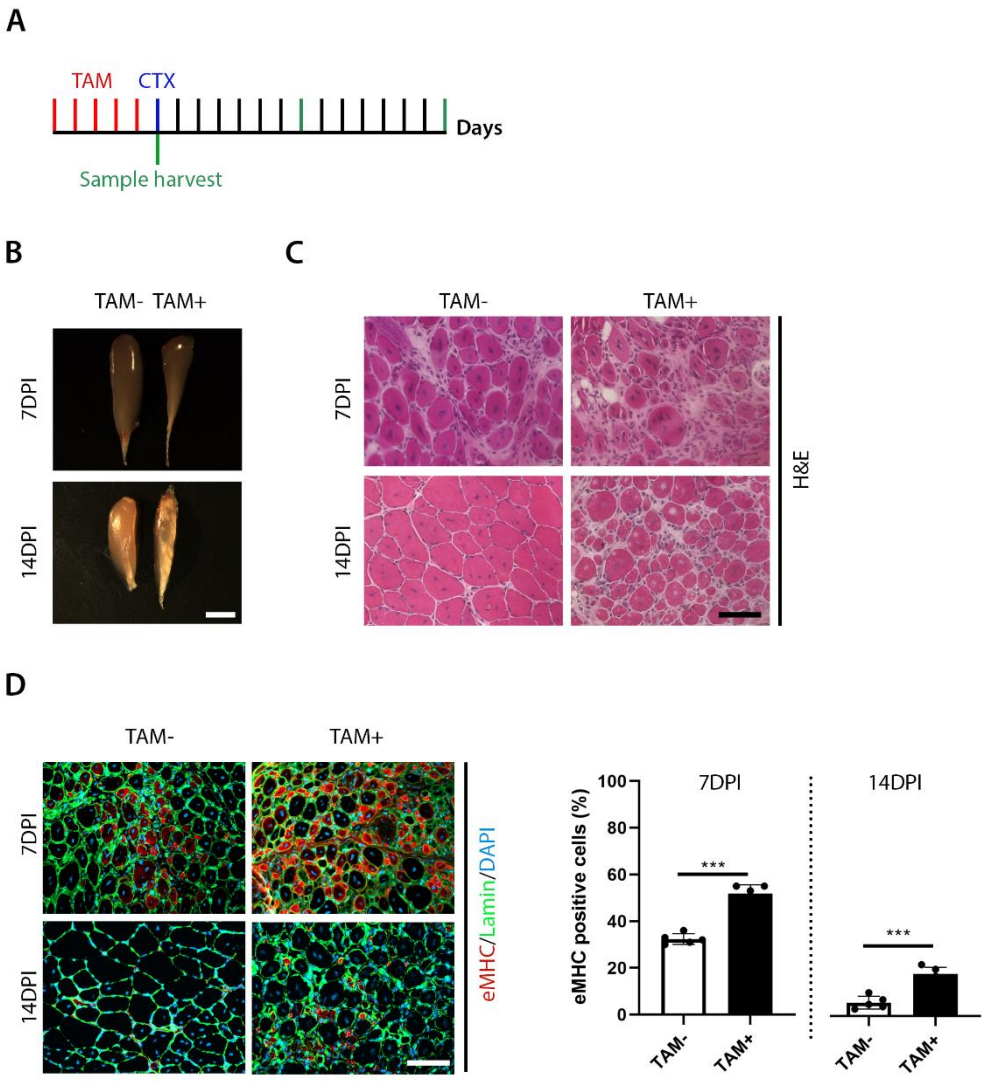

Figure 14. Regeneration of skeletal muscles after CTX injury in Duxbl-expressing mice.

(A) Schematic outline of CTX-induced injury in TAM-treated Pax7$^{Cre/ERT2}$; Duxbl$^{+/+}$ mice.

(B-D) Representative macroscopic images (B), histological H&E staining images (C), and immunofluorescence images (D left panel) of CTX-injured TA muscles from control or TAM–treated mice. Embryonic myosin heavy chain (*eMHC*), *Lamin* and DAPI are represented in red, green and blue, respectively (D left panel). The percentage of *eMHC* positive fibers in each group at 7 days (7DPI) or 14 days (14DPI) post injury. Error bars indicate SD of the mean (t-test: *p<0.05; **p<0.01; ***p<0.001).

3.10 Duxbl overexpression promotes RMS tumor formation in vivo.

Maintaining MuSCs in an activated but undifferentiated state by overexpression of *Duxbl* might promote tumorigenesis. To test this hypothesis, Pax7$^{Cre/ERT2}$; Duxbl$^{+/+}$ mice

were crossed to chronically regenerating mdx mice (Pax7$^{Cre/ERT2}$; Duxbl$^{+/+}$; mdx$^{-/-}$). In addition, Pax7$^{Cre/ERT2}$; Duxbl$^{+/+}$; p53$^{flox/flox}$ mice were generated, since loss of *p53* in regenerative MuSCs was sufficient for acquisition of *Duxbl* CNVs and subsequent RMS tumor formation (Figure 6A).

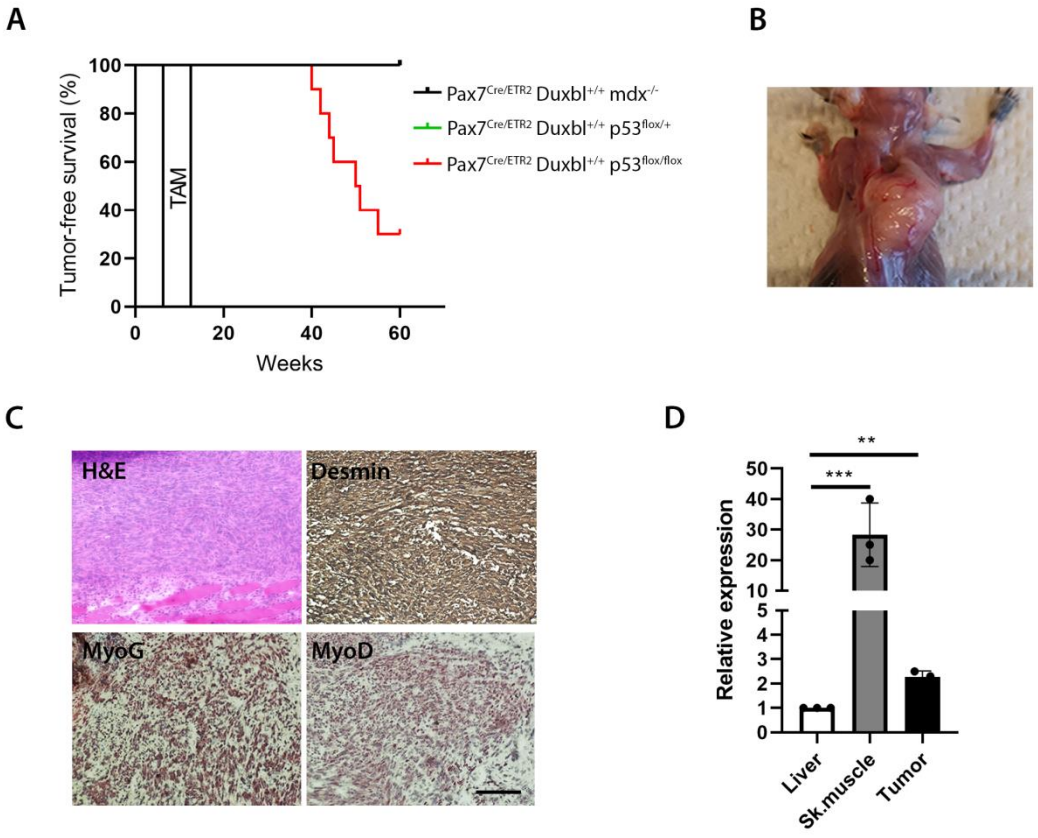

Figure 15. Introduction of Duxbl in p53-deficient MuSCs induces RMS tumor formation.

(A) Kaplan-Meier tumor-free survival curves are shown for indicated genotypes.

(B) Representative images of generated tumors on sites.

(C) Representative Immunohistochemistry staining of cross-sectioned tumor with H&E or indicated antibodies

(D) Quantitative analysis of *Duxbl* expression in liver, skeletal muscle and generated tumors of TAM-treated Pax7$^{Cre/ERT2}$; Duxbl$^{+/+}$; p53$^{flox/flox}$ mice. Error bars indicate SD of the mean (t-test: *p<0.05; **p<0.01; ***p<0.001).

Pax7$^{Cre/ERT2}$; Duxbl$^{+/+}$; p53$^{+/+}$ mice appeared completely normal and never developed tumors up to the age of 60 weeks (Figure 15A). In contrast, 70 % of all TAM-treated Pax7$^{Cre/ERT2}$; Duxbl$^{+/+}$; p53$^{flox/flox}$ mice developed tumors within 50 weeks after TAM

administration (Figure 15A and B). Importantly, TAM-treated Pax7$^{Cre/ERT2}$; p53$^{flox/flox}$ mice never developed tumors up to one year of age, indicating that increased expression of *Duxbl* is sufficient to induce RMS, when *p53* is absent (Figure 15A). qRT-PCR analysis confirmed that TAM-treated Pax7$^{Cre/ERT2}$; Duxbl$^{+/+}$; p53$^{flox/flox}$ mice exhibit high *Duxbl* transcript levels in both skeletal muscles and in tumors. Interestingly, tumor formation upon *Duxbl* overexpression only occurred in the absence of both *p53* alleles (Figure 15A). All tumors were classified as eRMS tumors harboring cells of spindle-like morphology and prominent expression of myogenic markers including *Desmin*, *MyoG* and *MyoD* (Figure 15C). These data show that overexpression of *Duxbl* in *p53*-deficient MuSCs is sufficient to elicit RMS tumor formation. Moreover, these data suggest that activation of MuSCs is required for RMS formation upon loss of *p53* since MuSCs are activated in TAM-treated Pax7$^{Cre/ERT2}$; Duxbl$^{+/+}$ mice (Figure 13C-D)

3.11 ERVLs are activated in Duxbl-expressing cells.

Dux TFs are only expressed during pre-implantation development (Sugie et al., 2020). At this early developmental stage, *Dux* TFs act as critical regulators of zygotic genome activation (ZGA) when endogenous transcription from the developing zygote first begins (Sugie et al., 2020). Specifically, it has been shown that overexpression of the founding member of the *Dux* TF family, *Dux4* (or its murine homolog *Dux*), is sufficient to activate a core set of ZGA-associated genes (De Iaco et al., 2017). Interestingly, it was also shown that ZGA is associated with expression of endogenous retroviruses (*ERVs*), a subclass of transposable retrotransposons of unknown function (Fu et al., 2019). In recent years, *ERVs* have attracted much attention, since they are reactivated in many human diseases including cancer and facioscapular humeral dystrophy (FSHD), a severe muscular dystrophy (Kassiotis, 2014; Young et al., 2013). Along this line, it was recently shown that *Dux* is transiently expressed during skeletal muscle regeneration (Knopp et al., 2016). Most interestingly, overexpression of *Dux4*/*Dux* is sufficient to activate *ERVL*, one subtype of the *ERV* superfamily, in myoblasts and ES (Hendrickson et al., 2017; Young et al., 2013). Based on these findings, it was next tested whether *Duxbl*-induced RMS tumor formation during muscle regeneration might be associated with *ERVL* activation.

To investigate this, it was first tested via qRT-PCR whether *Dux*, *Duxbl* and *ERVL* are expressed in regenerating muscle upon CTX-induced damage. Consistent with a previous

report, transient *Dux* expression was detectable in regenerating muscles beginning one day after injury (Knopp et al., 2016). *Dux* expression levels peaked at 30 hours after injury during the early regeneration phase, and subsequently rapidly decreased back to almost undetectable levels 14 days after injury when regeneration was completed (Figure 16A). Most interestingly, *Duxbl* expression levels increased upon muscle damage shortly after *Dux* expression (Figure 16A). Furthermore, increasing levels of *ERVL* expression were detected 36 hours after injury and 12 hours after initiation of *Dux* TFs expression, reaching a 2.4-fold peak over baseline levels 3 days after injury (Figure 16A). *ERVL* expression levels steadily dropped thereafter to pre-injury levels when regeneration was completed (Figure 16A). These data indicate that the *Dux* TFs-*ERVL* axis is transiently activated during muscle regeneration. Next, expression levels of *Dux*, *Duxbl* and *ERVL* were compared between TA muscles of physiologically resting wild-type mice and mdx mice undergoing continuous muscle regeneration. Interestingly, *Dux*, *Duxbl*, and *ERVL* expression levels were significantly increased in muscles of mdx mice compared to controls, confirming that the *Dux* TFs-*ERVL* axis was activated under conditions of chronic muscle muscles regeneration (Figure 16B). Importantly, these data show that chronic activation of the *Dux* TFs-*ERVL* axis alone is not sufficient to induce tumorigenesis since mdx mice did not form tumors at least up to the age of 60 weeks.

Next, *ERVL* expression levels were investigated in tumor cells obtained from TAM-treated $Pax7^{Cre/ERT2}$; $p53^{flox/flox}$; $mdx^{-/-}$ mice, which harbor increased genomic copy numbers of either *Duxbl* or *Yap1* ($TPCs^{Duxbl}$ or $TPCs^{Yap1}$). Remarkably, a significant increase of *ERVL* expression was found in $TPCs^{Duxbl}$ compared to wild-type MuSCs ($MuSCs^{WT}$), whereas *ERVL* expression levels were not changed in $TPCs^{Yap1}$ (Figure 16C). Considering that the tumors were additionally *p53*-deficient, these data suggest that reactivation of *ERVL* through *Duxbl* expression might depend on inactivation of *p53*. In support of this notion, we also found enhanced expression levels of *ERVL* in tumors isolated from TAM-treated $Pax7^{Cre/ERT2}$; $Duxbl^{+/+}$; $p53^{flox/flox}$ mice but not in the liver and lung wherein *p53* alleles were still intact and *Duxbl* was not expressed (Figure 16D).

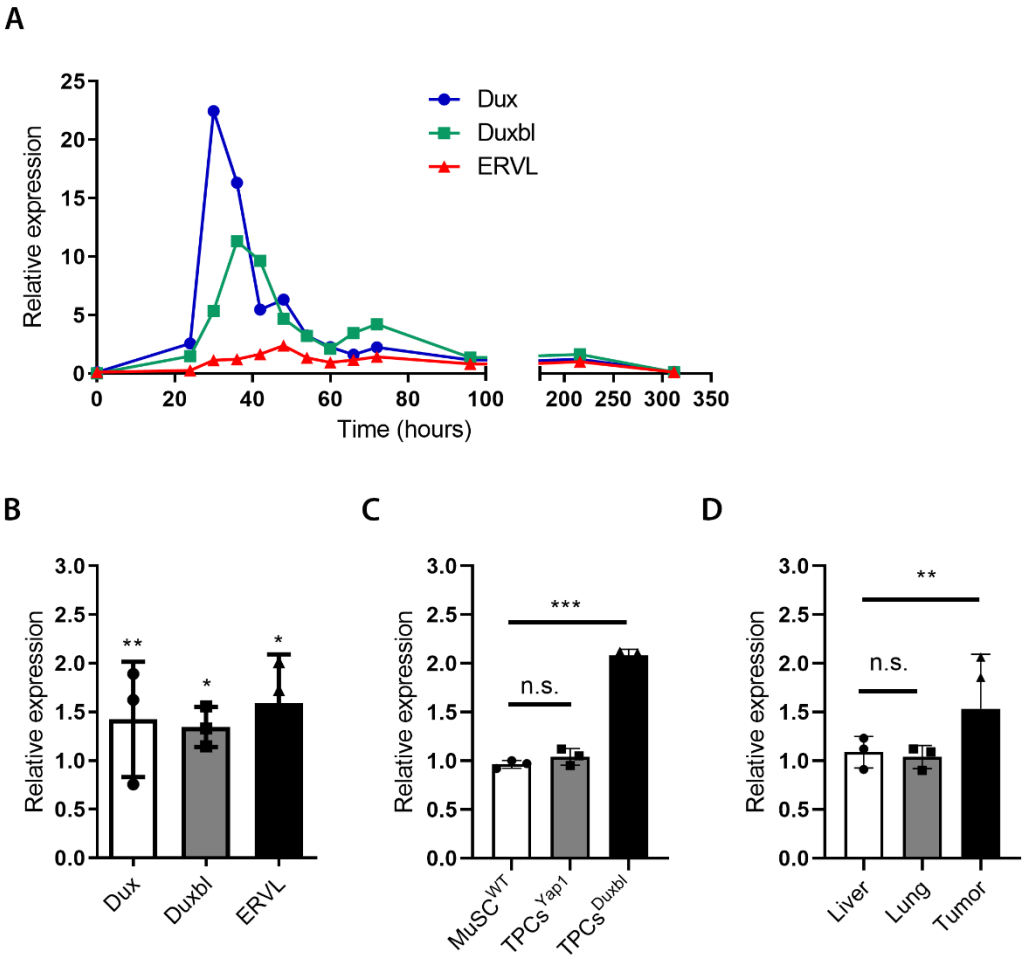

Figure 16. ERVL expression is enhanced in Duxbl-expressing tissues.

(A-B) Quantitative analysis of *Dux*, *Duxbl* and *ERVL* expression in CTX damaged T.A. muscles of WT mice (A) or in T.A. muscles of regenerating mdx mice.

(C-D) Quantitative analysis of *ERVL* expression in WT MuSCs, $TPCs^{Yap1}$ and $TPCs^{Duxbl}$ (C) or in different tissues of TAM-treated $Pax7^{Cre/ERT2}$; $Duxbl^{+/+}$; $p53^{flox/flox}$ mice (D). Error bars indicate SD of the mean (t-test: *p<0.05; **p<0.01; ***p<0.001).

3.12 Dux TFs overexpression activates ERVL-LTR reporter constructs.

To further investigate the regulation of *ERVL* activation by *Dux* TFs, and *Duxbl*, a fluorescence based *ERVL* reporter assay was performed by using a vector in which an *ERVL-LTR* was linked to the coding sequences of the red fluorescent *tdTomato* gene. It has previously been shown that *LTRs* can function as transcriptional promoters of corresponding *ERVs* and neighboring genes (Bénit et al., 1999; Dunn et al., 2006; Lee et al., 2019). Using

this setup it was then tested whether overexpression of *Duxbl* activates the *ERVL-LTR::tdTomato* reporter system in HEK293T cells (Figure 17A).

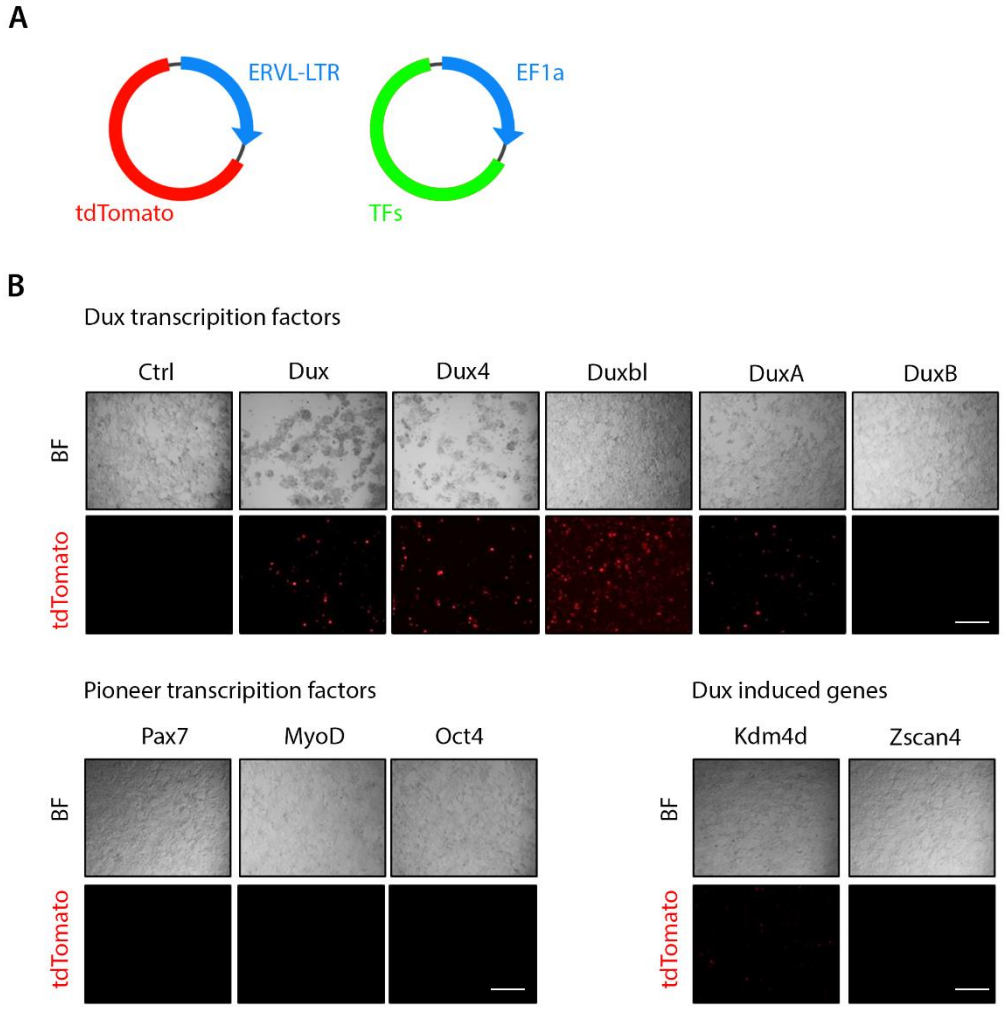

Figure 17. Dux TFs mediate activation of ERVL-LTR reporter.

(A) Diagram of *ERVL-LTR* reporter and transcription factors (TFs) constructs.

(B) Representative images showed the effect of different TFs on the activation of *ERVL-LTR* reporter. Note that the *tdtomato* positive cells reflected the activation of reporter and cells were imaged two days after transduction.

Consistent with previous reports, *Dux* or *Dux4* overexpression was sufficient to induce *ERVL* expression reflected by a prominent *tdTomato* signal 48 hours after transfection (Figure 17B) (Dunn et al., 2006). Notably, *tdTomato* signal was highest 24 hours after *Dux* and *Dux4* transfection but cell numbers and fluorescence signal decreased thereafter as a consequence of pronounced *Dux*/*Dux4*-induced cell death (Knopp et al., 2016). Strikingly, overexpression of *Duxbl* resulted in a massive *tdTomato* signal clearly

demonstrating that *Duxbl* overexpression is sufficient to activate *ERVL* expression in HEK293T cells (Figure 17B). Similar to *Duxbl* overexpression, human *DuxA* and *DuxB* did not elicit cell death (Figure 17B), suggesting that human *DuxA* might function similarly to murine *Duxbl* (Figure 17B). To test whether *ERVL* activation depends solely on *Dux* TFs, the *ERVL-LTR*::*tdTomato* reporter was transfected with three other different transcription factors, *Pax7*, *MyoD* and *Oct4*, all of which play critical roles in establishment of cell identity and development and harbor pioneering activity (Pan et al., 2002; Tapscott, 2005; Zammit et al., 2006). Furthermore, *Kdm4d* or *Zscan4*, two direct targets of *Dux/Dux4* during ZGA (Sugie et al., 2020), were also evaluated in the *ERVL* reporter assay. None of these genes activated the *ERVL-LTR*::*tdTomato* reporter, in contrast to *Dux4*/*Dux* and *Duxbl*/*DuxA* (Figure 17B).

3.13 Inactivation of p53 is required for Duxbl-mediated ERVL activation.

p53 function is inhibited in HEK239T cells due to the presence of the SV40 T antigen (Bajpai and Terskikh, 2007). Considering that *ERVL* expression was only detected in *p53*-deficient MuSCs and tumor cells *in vivo* (Figure 15), the *ERVL-LTR*::*tdTomato* reporter assay was repeated in HEK293 cells that do not harbor the SV40 T antigen, leaving *p53* functional (Lin et al., 2014). Interestingly, overexpression of *Duxbl* failed to activate the *ERVL*-LTR::*tdTomato* reporter in HEK293 cells with normal p53 activity (Figure 18A). To determine whether this effect was an entity of HEK293 cells, a C2C12 myoblast cell line was generated stably expressing a shRNA targeting *p53* mRNA to knockdown endogenous *p53* expression (Figure 18B). As control, a C2C12 myoblast cell line was also generated carrying a non-targeting shRNA (Figure 18B). These myoblast lines were designated $C2^{shp53}$ and $C2^{shCtrl}$. Western blot analysis using protein extracts harvested from these cells showed that UV-treatment led to *p53* activation in $C2^{shCtrl}$ myoblasts, whereas *p53* expression was not detected in $C2^{shp53}$ myoblasts, demonstrating efficient *p53* knockdown (Figure 18B). Importantly, activation of the *ERVL-LTR*::*tdTomato* reporter was only detected in $C2^{shp53}$ but not in $C2^{shCtrl}$ cells (Figure 18C). Increased expression of endogenous *ERVL* in $C2^{shp53}$ was confirmed by qRT-PCR (Figure 18D). Taken together, these data show that *Duxbl* activates *ERVL* in somatic cells in the absence of *p53*, or conversely, that inactivation of *p53* is mandatory for *Duxbl*-mediated *ERVL* activation. More importantly, these findings, together with the observation that *ERVL* expression was only detected in RMS tumors when the *Duxbl*

gene was amplified and *p53* was absent, suggest that *Duxbl*-mediated *ERVL* activation is probably the cause for promotion of RMS tumor formation *in vivo* by *Duxbl*.

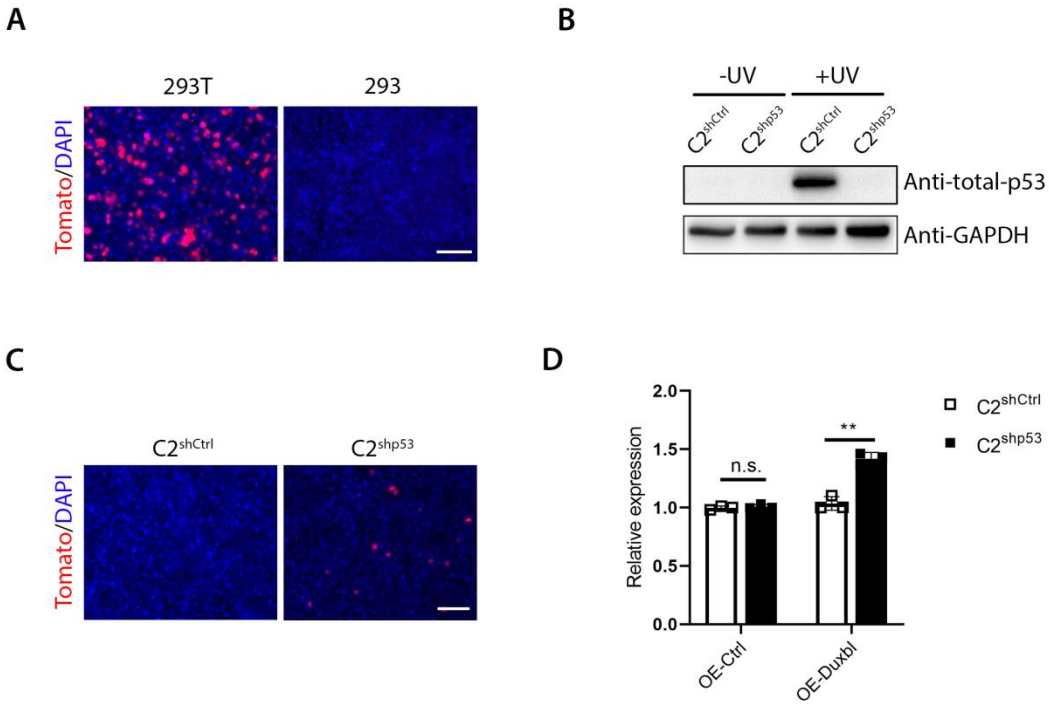

Figure 18. p53 shapes Duxbl-mediated ERVL activation.

(A) Representative images showed the *ERVL-LTR* reporter assay with *Duxbl* introduction in either HEK293T or HEK293 cells.

(B) Western blot analysis showed shRNA mediated knock down of *p53* in myoblast cells. Note that the total cell lysates were prepared 24 hours after UV treatment (20 mJ/cm2 UVB).

(C) Representative images showing the *ERVL-LTR* reporter assay with *Duxbl* introduction in myoblast cells infected with lentivirus encoding shCtrl or shp53. (A and C) Note that the *tdtomato* positive cells reflected activation of the reporter and cells were imaged two days after transduction.

(D) Quantitative analysis of *ERVL* expression in *Duxbl* or *Ctrl*-expressing myoblast cells infected with lentivirus coding for different shRNAs. Error bars indicate SD of the mean (t-test: *p<0.05; **p<0.01; ***p<0.001).

4. Discussion.

Despite significant progress in the understanding of cancer development and better option for treatment, cancer still remains one of the most common causes of deaths worldwide (Collaborators, 2016). Among the major open questions are the cellular origins of different cancer types and the causative mechanisms of cancer formation. Only after pinpointing the cellular origins of cancer it will possible to identify the genuine mechanisms that transform healthy cells to tumor forming cells.

MuSCs have been proposed a long time ago to be the cellular origin of RMS tumors (Camboni et al., 2012; Chamberlain et al., 2007b; Duchenne, 1867; Fernandez et al., 2010a). Surprisingly however, unequivocal proof has been lacking. In this study, combined *in vivo* mutagenesis and lineage tracing constituted the basis for genetic mouse model of RMS tumor formation that helped to answer some of the open questions. Whole exome sequencing of MuSCs-derived RMS tumors in these mice allowed identification of known and novel oncogenic copy number amplifications of genes that are responsible for tumorigenic transformation of MuSCs. Based on the comprehensive analysus of one of the identified oncogene, *Duxbl*, I propose that activation of the retrotransposon *ERVL* in *p53*-deficient MuSCs might be an underlying cause of RMS tumor development.

4.1 MuSCs as a cellular origin of RMS tumors.

It is generally accepted that the emergence of cancer cells is a consequence of accumulating mutations in a previously healthy cell. Such mutations are more likely to occur in dividing cells (White and Lowry, 2015). However, once a tumor is established, it is difficult to determine the cellular origin for various reasons. First, tumors are heterogeneous containing many different cell types, such as vascular cells or fibroblasts that mainly supply trophic and metabolic support to the growth of cancer cells. In fact, it has been proposed that for some cancer types, cancer cells only comprise 0.1 % cells of the whole tumor (Sharpe et al., 2013). This creates problems when analyzing tumors in bulk, particularly when genomic DNA is analyzed for oncogene identification. Indeed, in this study it was found that the abundance of tumor supporting cells within tumors (which are not tumorigenic themselves) interferes with such analyses. Second, tumor development is a transformative process. By the time a tumor has been established, the cancer cells therein may have lost most/all molecular and cellular features of the parental cell types from which they originally derived. In fact, many amplified or

expressed genes in terminal tumor cell appear to be secondary consequences of tumor establishment but not the causes (Visvader, 2011).

Adult stem cells (ASCs) have long been considered as a prominent cellular origin of cancer because of their stem cell properties. Many ASC types are long-lived and can reside in an organ throughout the lifespan (Koyuncu et al., 2015). Therefore, they can persist long enough to accumulate carcinogenic mutations (Chen et al., 2017; Mani et al., 2020). In addition, ASCs are often multipotent, providing another explanation why different cell types can be found within one tumor and/or the plastic nature of cancer cells (Bhartiya et al., 2013). Lastly, ASCs can self-renew and thus they harbor a proliferative potential, an obvious essential feature of tumor cell expansion.

Thus, it is conceivable to assume that MuSCs are especially vulnerable to tumorigenic transformation, because they own features that are common to cancer cells and do not need to acquire them *de novo*. Under normal physiological resting conditions, dormant MuSCs are long-lived but they can proliferate to self-renew or undergo differentiation, which reflects their inherent plastic nature. In this study, it was shown that loss of *p53* in regenerating MuSCs elicits RMS tumor formation with 100 % penetrance. Importantly, mdx mice containing intact *p53* never developed RMS tumors up to one year of age. In addition, MuSC-specific loss of *p53* alone did not result in RMS tumor formation. Therefore, the non-proliferative state of dormancy seems to protect MuSCs from tumorigenic transformation when *p53* activity is lost. This is consistent with the observations that RMS tumors usually occurs in early childhood during which the skeletal muscles are still undergoing hyperplastic growth and myoblasts are still expanding (Ognjanovic et al., 2009). Along this line, the low turnover rate of adult skeletal muscles provides a reasonable explanation for the low frequency of RMS tumor formation in adults (Gerber et al., 2013; Tremblay et al., 2014).

Notably, this study is not the first to address the cellular origin of RMS tumors. It has been shown that other cell types can transform to RMS tumor forming cells including mesenchymal stem cells and endothelial progenitor cells under certain conditions (Drummond et al., 2018; Drummond and Hatley, 2018; Sun et al., 2015). In this study, MuSCs were investigated as a potential source of RMS tumors under conditions of regeneration, based on two previous reports. Chamberlain and colleagues showed that patients suffering from Duchenne muscular dystrophy (DMD) (the dystrophic disease which is mimicked in mdx mice) own an increased predisposition towards spontaneous RMS tumor development, even though the incidence rate is relatively low (Chamberlain et al., 2007a). Furthermore, Camboni and

colleagues showed that *p53* knockout mice are permissive for RMS tumor development under conditions of chronic muscle regeneration (Camboni et al., 2012). However, the cellular origin in these mice had remained unidentified.

This question was answered in this study by combining conditional inactivation of *p53* in MuSCs alone and by genetic lineage tracing. Essentially, all TAM-treated $Pax7^{Cre/ERT2}$; $p53^{flox/flox}$; $mdx^{-/-}$ mice developed RMS tumors and all of them were fluorescently lineage-traced, unequivocally demonstrating that MuSCs are the cellular origin of RMS tumors in these mice. Importantly, all generated RMS tumors consisted of both non lineage-traced traced cells as well as of MuSC-derived lineage-traced cells. Only transplantation of the MuSC-derived tumors cells, but not the non-lineage-traced cells generated tumors upon transplantation, showing that only transformed MuSCs were *bona fide* tumorigenic cells in RMS tumors. This is important since it suggest that the true tumor cells and those of the tumor supporting stroma probably need to be targeted differently when considering therapeutic interventions.

4.2 The complex mechanisms of RMS tumor formation.

The causes of cancerogenic transformation are manifold, depending not only on the cancer types but also on the cell types from which they are derived, reflecting variations of responses towards treatment of cancer patients. Likewise, Walther and colleagues identified different genetic alterations from 19 RMS tumor patients, including genomic amplifications of *Mycn*, *Cdk4*, and *Mir17hg*, deletions of *Nf1*, *Cdkn2a*, and *Cdkn2b*, and loss of homozygosity of *Ascc3* or *Odz3* genes (Walther et al., 2016). Most interestingly, it has also been described that two separate RMS tumors that develop in one patient can harbor distinct genetics signatures (Nishimura et al., 2013), suggesting that different genetic alterations can lead to the tumor formation of the same cancer type.

In view of the different copy number amplifications that were identified in this study, there is no a master oncogene controlling the transformation of MuSCs, meaning that a (single) common gene was not identified across samples with the exception of *p53* itself. The known functions of the amplified genes also vary between samples ranging from functions in Hippo-signaling, MET signaling, JNK signaling, NF-kappaB signaling to Ras signaling. Interestingly, RMS tumor cells carrying different amplifications displayed significantly different characteristics and behaviors. For example, RMS tumor cells in which either the *Yap1* or *Met*

gene was amplified appeared more aggressive and fast growing, while RMS tumors carrying genomic *Ras* amplifications appeared more heterogeneous by morphological H&E analysis and cells purified from these tumors did not grow well in culture (data not shown).

It did not escape my attention that some amplified genes act differently to what has been previously described. For example, Nguyen and colleagues showed that *Yap1* is required for *Ras*-induced cellular transformation and tumor formation *in vivo* (Nguyen et al., 2014). *Yap1* expression is also frequently detected in *Ras*-induced tumors because RAS can prevent YAP1 protein ubiquitination (Nguyen et al., 2014). Furthermore, it has been shown that silencing of *Yap1* in *Ras*-induced RMS tumor cells can lead to cell senescence followed by cell death (Linardic et al., 2005; Slemmons et al., 2015). These studies suggest an important interaction between *Yap1* and *Ras*. In this study however, *Ras* amplifications were not detected in any of the RMS tumors containing *Yap1*-amplificaions and *vice versa*. A possible explanation is that RMS tumor cells might develop further pro-cancerogenic features and mutations if they are maintained long enough in culture. Such a mechanism may enable tumor cells to develop escape routes and bypass selective pressures during therapy. In support of this idea, seminal studies from Vogelstein and colleagues have revealed that cancer genomes evolve over time enabling tumor maintenance or survival (Vogelstein et al., 1988).

Essentially, this study shows that the mechanisms of RMS tumor formation and maintenance are diverse and thus remain incompletely understood. Further investigations are necessary to understand the heterogeneous underpinnings of RMS tumor pathology at both molecular and cellular levels, the outcomes of which might benefit the development of personalized cancer treatment strategies.

4.3 The roles of Duxbl in RMS tumor formation.

A significant result of this study was the identification of a recurrent genomic amplification on chromosome 14qA3 (5 out of 22, 23 %) in RMS tumors. This subset of RMS tumors was of particular interest because this genomic locus contained a very poorly described and triplicated gene array consisting of the genes *Plac9*, *Tmem254*, *Cphx* and *Duxbl*. None of these had been previously described as an oncogene. In fact, this genomic region was not annotated at the time of discovery. I focused the study on *Duxbl* because sequence homology suggested that it belongs to the family of *Dux* TFs of which *Dux4* is the founding member.

While under healthy conditions *Dux4* is present as highly methylated and multicopy silenced gene array (known as D4Z4 repeats), it is robustly expressed in disease conditions such as the muscular dystrophy FSHD and in different cancer contexts (Geng et al., 2012). For cancer, it was shown that expression of *Dux4* fusion products occurring through genomic translocations can induce tumor formation (Chew et al., 2019). For example, in B-cell acute lymphoblastic leukemia (B-ALL), the transactivation domain of *Dux4* is replaced with a fragment of the *Igh* gene. This results in expression of a novel out-of-frame DUX4-IGH fusion oncoprotein. This oncoprotein can induce expression of an alternative *ERG* isoform (*ERGalt*) in B-cells, inducing leukemic transformation (Lilljebjorn and Fioretos, 2017; Yasuda et al., 2016a, b; Zhang et al., 2016). Similarly, genomic translocation of *Dux4* can result in the generation of a chimeric DUX4-CIC fusion protein that can induce sarcoma formation (Kawamura-Saito et al., 2006).

It seems likely that all *Dux* TFs (*DuxA*, *DuxB*, *Dux4* in humans and *Dux* and *Duxbl* in mice) may exhibit oncogenic potential since all *Dux* TFs share highly similar DNA binding domains (Leidenroth and Hewitt, 2010). This is evident for *Dux4* and *Dux* that both can induce the expression of a core set of genes and *ERVs* in developing zygotes (Eidahl et al., 2016a; Hendrickson et al., 2017). Moreover, both *Dux4* and *Dux* can induce rapid cell death when overexpressed in somatic cells, indicating orthologous function for at least these two *Dux* TFs.

However, it is worth noting that *Duxbl* does not seem to harbor a transactivation domain like *Dux4* or *Dux*, suggesting that *Duxbl* possibly does not function as a typical transcriptional activator but rather as a repressor. This putative antagonistic function of *Duxbl* to *Dux*/*Dux4* is supported by the observation that *Duxbl* overexpression does not elicit cell death but instead promotes myoblast cell proliferation and inhibits myogenic differentiation, which will favor tumorigenic transformation (Wu et al., 2014). Since the DNA binding domains between *Dux* TFs are highly similar, it is tempting to speculate that *Duxbl* might function as a competitive inhibitor to *Dux*/*Dux4* of DNA binding.

Most interestingly, it was found during the course of this thesis that *Dux4*/*Dux*, *DuxA*, *DuxB*, and *Duxbl* are all expressed in early zygotes during zygotic genome activation (Sugie et al., 2020). Several independent tumors contained genomic amplification of *Kdm4d* or *Mbd3l2*, which are activated by *Dux4*/*Dux* during ZGA. Interestingly, *Kdm4d* and *Mbd3l2* encode for epigenetic modifiers regulating H3K9 demethylation and 5mC oxidation during ZGA, respectively (Liu et al., 2018; Marcho et al., 2015; Tan and Shi, 2012; Xu et al., 2021). Essentially, this study indicates that *Dux* TFs and ZGA-associated genes are involved in tumor formation and that reactivation of early zygotic genes induces a gain tin plasticity through

epigenetic mechanisms for initiation of tumor development. Importantly, these findings are more than of academic interest, since around 10% of RMS patients are positive for *Dux* TFs and/or ZGA related genes.

4.4 The activation of ERVL in Duxbl-driven RMS tumor formation.

An unexpected finding was the genetic interaction of *Duxbl* and *p53* in regulating the expression of transposable elements (*TEs*). Making up about half of mammalian genomes, *TEs* are remnants of ancestral virus genomes and were widely considered as non-functional "junk DNA". Only in recent years *TEs* have moved into focus of intense research efforts, since they are expressed in several diseases including cancer and notably also during preimplantation development (Lander et al., 2001; Waterston et al., 2002). Interestingly, many ZGA-associated genes are controlled by neighboring *TEs*, the *LTRs* of which can serve as transcriptional promoters (Fu et al., 2019; Macfarlan et al., 2012). Based on their sequences, *TEs* are classified into long terminal repeat (*LTR*) retrotransposons (also known as endogenous retroviruses (*ERVs*), non-LTR retrotransposons represented by long interspersed elements (*LINEs*) and short interspersed elements (*SINEs*) (Mager and Stoye, 2015). *TEs* that own transposase activity can be mobilized resulting in insertional mutagenesis or chromosomal alteration (Maksakova et al., 2006). In humans, it is estimated that roughly 0.1 % of all spontaneous mutations arise from *TEs*-dependent insertions, which can result in activation of oncogenes or inactivation of tumor suppressors (Maksakova et al., 2006). Consistent with the observations that *ERVs* can serve as alternative promoters, it has been shown that *ERVs* can also induce oncogenesis through aberrant activation of neighboring genes. One example is a variant of human T-cell leukemia (HTLV-1) in which *Hbz* and *Tax* are aberrantly expressed by neighboring *ERVs* (Laverdure et al., 2016). In this study, it was shown that *Duxbl* expression enhances transcription of *ERVL*, which was similar to what had been reported for *Dux4* and *Dux*. It remains to be shown in future studies whether different *Dux* TFs play different roles in activation of different *TEs*.

Since in contrast to *Dux4*/*Dux, Duxbl* does not seem to harbor a transcriptional activation domain, it is tempting to speculate that direct binding of *Duxbl* to cognate *TE*-coding DNA sequences requires cooperation with other transcriptional factors to activate expression. Direct examination of *Duxbl* binding on *ERVL* by ATAC-seq and ChIP-seq experiments will be helpful in this regard.

The requirement of *p53* inactivation for *Duxbl*-dependent induction of *ERVL* expression is fascinating. This observation, together with the finding that RMS tumors only form upon loss of *p53*, gives rise to the notion that *p53* expression defines a defense line against *Duxbl*-induced *ERVL* activation in RMS tumor formation. Notably, it has been reported that nearly half of all human cancers express *TEs* (Rodić et al., 2014). Importantly, the same report showed that *TEs* were more commonly reactivated in *p53*-deficient human cancers suggesting that *p53* counteracts the expression of *TEs* (Rodić et al., 2014). Interestingly, it appeared that one intact *p53* allele is sufficient to suppress *Duxbl*-mediated RMS tumor development, which might also be related to *p53*-mediated regulation of *ERVL*.

Given the pleiotropic molecular processes in which *p53* is involved, it is likely that inactivation of *p53* may provide a favorable environment allowing activation of *TEs*. For example, ,*p53* deficient cells contain less DNA methylation marks at a genome scale, increasing global accessibility of *TEs* distributed across the genome (Crichton et al., 2014; Younger and Rinn, 2017). Notably, it has been reported that *p53* is required for *Dux4*-mediated apoptosis both *in vitro* and *in vivo* although this hypothesis remains debated (Barro et al., 2010; Bosnakovski et al., 2017a; Bosnakovski et al., 2017b; Bosnakovski et al., 2008a; Wallace et al., 2011). It remains to be shown how exactly *p53* participates in the regulation of *TEs* and whether or not expression of specific *TEs* can induce tumorigenesis.

4.5 Perspective.

This study clearly demonstrates that regenerating MuSCs are cells of origin for RMS tumors. *Duxbl* exhibits tumorigenic potential and induces RMS tumor formation in *p53*-deficient MuSCs. From this study I propose that activation of a Duxbl-p53-ERVL axis is a causative mechanism that is sufficient to promote RMS tumor formation

It is still necessary to obtain a better understanding of role of *ERVL* in RMS tumor formation. In this regard, *Duxbl/p53* mice carrying *ERVL* reporters will reveal whether *ERVL* becomes activated before, during, or after tumor formation. More importantly, it is not clear whether *ERVL* activation alone is sufficient to induce RMS tumor formation. It will be interesting to test the behavior of MuSCs when endogenous *TEs* are overexpressed or alternatively, when they are inactivated. However, this will be challenging, since there is a high sequence conservation across hundreds to thousands of highly repetitive *TEs* that are distributed

across the genome (Crichton et al., 2014). Therefore, it will be difficult to identify specific *TEs* that may or may not have oncogenic potential. Clearly, molecular profiling of *p53*-deficient MuSCs overexpressing *Duxbl* will provide insight into the regulatory network driven by the *Duxbl-p53-ERVL* axis. Furthermore, it will be interesting to identify potential protein interactors of *Duxbl* in proteomic analyses. These future investigations will help to obtain a deeper mechanistic understanding as to how *Duxbl* transforms *p53*-deficient MuSCs to tumor forming cells. Such studies will aid in the development of diagnostic strategies and therapeutic treatments of *Duxbl* associated cancers.

IV. List of figures.

V. List of tables.

VI. List of abbreviation.

Abbreviation	Meaning
%	Percent
APS	Ammoniumperoxodisulfat
ASCs	Adult stem cells
ATAC	Assay for Transposase-Accessible Chromatin using sequencing
aRMS	Alveolar rhabdomyosarcoma
α7-integrin	Integrin alpha 7
BSA	Bovine serum albumin
BMP	Bone morphogenetic protein pathway
°C	Degrees celcius
cDNA	Complementary DNA
CO_2	Carbon dioxide
ChIP	Chromatin immunoprecipitation assay
CNVs	DNA copy-number variations
cm	Cenetimetre
CTX	Cardiotoxin
DAPI	4',6-diamidino-2-phenylindole
DEPC	Diethylpyrocarbonate
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
DMD	Duchenne muscular dystroph
DTA	Diphtheria toxin A
DPI	Days post injury
dNTP	DeoxyriboNucleotide TriPhosphate
Dux	Double homeobox
DuxA	Double homeobox protein A
DuxB	Double homeobox protein B
Dux4	Double homeobox protein 4
Duxbl	Double homeobox B-like protein
EDTA	Ethylenediaminetetraacetic acid
eRMS	embryonal rhabdomyosarcoma
eMHC	Embryonic form of myosin heavy chain
ERVs	Endogenous retroviruses
ERVL	Endogenous retroviruses type L
FBS	Fetal bovine serum
EdU	5-ethynyl-2'-deoxyuridine
FGF	Fibroblast growth factor
FAPs	Fibro-adipogenic progenitors
FACS	Fluorescence-activated cell sorting
g	Gram
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GFP	Green Fluorescent Protein
GOF	Gain-Of-Function
H&E	Hematoxylin and eosin staining
IF	Immunofluorescence
In Situ	In place(Latin)

In vitro	In the glass (Latin)
In vivo	Within the living (Latin)
LTR	Long terminal repeat
LINEs	Long interspersed elements
KO	Knockout
L	Litre
LOF	Loss of function
LoxP	Locus of cross-over (x)of the bacteriophage P1
IL-1	Interleukin-1
M	Molar
Mg	Milogram
mL	Millilitre
Mm	Millimetre
mM	Millimolar
miR	Micro-RNA
mESC	Embryonic stem cell
MuSCs	Muscle stem cells
mRNA	messenger RNA
MRFs	Myogenic regulatory factors
Myf5	Myogenic factor 5
MyoD	Myoblast determination protein 1
Myogenin	Myogenic factor 4
Mdx	Mice with point mutation in its DMD gene
MSC	Mesenchymal stem cells
Ng	Nagogram
Nm	Manometers
Oct4	Octamer-binding transcription factor 4
OCT	Optimal cutting temperature medium
PBS	Phosphate-buffered saline
bHLH	basic helix–loop–helix
Pax3	Paired box gene 3
Pax7	Paired box gene 7
p53	Tumor Protein P53
p21	Cyclin Dependent Kinase Inhibitor 1A
PFP	PAX-fusion positive
PFN	PAX-fusion negative
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
PH	Pulmonary hypertension
qRT-PCR	Quantitative real-time polymerase chain reaction
RNA	Ribonucleic acid
RNA-seq	RNA-sequencing
ROSA26	Reverse oriented splice acceptor, clone 26
RMS	Rhabdomyosarcoma
Ras	Ras GTPase
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
shRNA	Short hairpin RNA
Shh	Sonic hedgehog pathway

SNVs	Single nucleotide variants
SINEs	Short interspersed elements
TBST	Tris buffered saline with tween-20
TE	Tris-EDTA
TEs	Transposable element
TNF-α	Tumor Necrosis Factor-α
TF	Transcription factor
TEMED	Tetramethylethylenediamine
Tomato or tdTomato	Tomato fluorescent protein
TAM	Tamoxifen
TPC	Tumor-propagating cells
USA	United States of America
µg	Microgram
µL	Microliter
V	Volts
V5	V5 tag peptide
WT	Wild type
Wnt/β-catenin	Wnt/β-catenin signaling pathways
WGS	Whole genome sequencing
yH2Ax	H2A histone family member X
Yap1	Yes-associated protein 1
ZsGreen	Zoanthus green fluorescent protein
ZGA	Zygotic gene activation
53bp1	Tumor suppressor p53-binding protein 1

VII Acknowledgements.

This Ph.D. thesis is the result of a challenging journey, upon which many people have kindly contributed and given their supports. At the end of this thesis, it is my pleasure to acknowledge all those who made this thesis possible.

First, I would like to thank all the audiences for this thesis no matter on purposed or just by chance. It is my great honor to have your time. If you have any questions or advice, please contact and share your ideas with me by szzhong83@hotmail.com anytime.

A special thanks to the Director Prof. Dr. **Thomas Braun** and associate colleagues, Dept. Cardiac Development and Remodeling, Max Planck Institute for Heart and Lung Research, for providing me this chance to work here and countless supports from the first day me being in Germany and the end of the thesis. This thesis is prepared under the supervision of Dr. **Johnny Kim**. I am sincerely grateful for his guidance and encouragement. Mission impossible is complete mainly because of his constant helpings. From whom more than the subject, I learned what a research scholar should aim to become a proper researcher. His passion for science and attitude toward learning has always been a sample for me. I am also grateful for having this thesis been critically assessed and approved by my outstanding Doctoral Committee members, Prof. Dr. **Thomas Braun**, Prof. Dr. **Lienhard Schmitz**, Prof. Dr. **Reinhard Dammann**, and Prof. **Dr. Sandra.Hake**.

The presented results are also contributed by the hard work of my lab and staff members. Special thanks to Dr. **Krishna Sreenivasan**, Dr. **Giovanni Maroli**, Dr. **Jens Preussner**, Dr. **Mario Looso**, Dr. **Stefan Günther**, Mr. **Felipe Lüttmann** and Mr. **YanPu Chen**. I am deeply delighted to have worked with you and sincerely look forward for cooperation in the future.

I will be forever grateful to my post-graduation supervisor Prof. Dr. **Deming Gou**, and the colleagues Dr. **Kang Kang**, Dr. **Huiling Qiu**, Mrs. **Yuna Wang**, Mrs. **Lan Luo**, Mrs. **Zhixiong Tang** for the startups. It was their efforts that instilled in my interest for pursuing academics. This Ph.D. journey would not have been possible without supports and cares from the lovely friends and family:

Finally, I gratefully acknowledge the support, encouragement and love from **my parents**, Mr. **Jiayuan Zhong**, Mrs. **Qiong Wu**, and Mrs. **Yangzi Yang**. I can not thank you enough

VIII Curriculum vitae.

Der Lebenslauf wurde aus der elektronischen Version der Arbeit entfernt.

The curriculum vitae was removed from the electronic version of the paper.

Publications List

Connect-four: genomic analyses of regenerative stem cells identifies zygotic Dux factors as tumor initiators. Preussner J, **Zhong JS**, Looso M, Braun T, Kim J. Mol Cell Oncol. 2019 Feb 8;6(2).

Oncogenic Amplification of Zygotic Dux Factors in Regenerative p53-Deficient Muscle Stem Cells Defines a Molecular Cancer Subtype. Preussner J, **Zhong JS**, Sreenivasan K, Günther S, Engleitner T, Künne C, Looso M, Braun T, Kim J. Cell Stem Cell. 2018 Dec 6;23(6):794-805.e4. **First co-author**

Regulatory axis of miR-195/497 and HMGA1-Id3 governs muscle cell proliferation and differentiation. Qiu HL, **Zhong JS**, Luo L, Tang ZX, Liu N, Kang K, Li L, Gou DM. Int J Biol Sci. 2017 Jan 15;13(2):157-166. **First co-author**

miR-34b Modulates Skeletal Muscle Cell Proliferation and Differentiation. Tang ZX, Qiu HL, Luo L, Liu N, **Zhong JS**, Kang K, Gou DM. J Cell Biochem. 2017 Dec;118(12):4285-4295.

A PCR-Based Method to Construct Lentiviral Vector Expressing Double Tough Decoy for miRNA Inhibition. Qiu HL, **Zhong JS**, Luo L, Liu N, Kang K, Qu JL, Peng WD, Gou DM. PLoS One. 2015 Dec 1;10(12):e0143864. **First co-author**

Hypoxia inducible factor-1 mediates expression of miR-322: potential role in proliferation and migration of pulmonary arterial smooth muscle cells. Zeng Y, Liu HT, Kang K, Wang ZW, Hui G, Zhang XY, **Zhong JS**, Peng WD, Ramchandran R, Raj U, Gou DM. Sci Rep. 2015 Jul 13;5:12098.

A direct real-time polymerase chain reaction assay for rapid high-throughput detection of highly pathogenic North American porcine reproductive and respiratory syndrome virus in China without RNA purification. Kang K, Yang KL, **Zhong JS**, Tian YX, Zhang LM, Zhai JX, Zhang L, Song CX, Gou CY, Luo L, Gou DM. J Anim Sci Biotechnol. 2014 Oct 2;5(1):45.

Identification of microRNA-Like RNAs in the filamentous fungus Trichoderma reesei by solexa sequencing. Kang K, **Zhong JS**, Jiang L, Liu G, Gou CY, Wu Q, Wang Y, Luo L, Gou DM. PLoS One. 2013 Oct 2;8(10):e76288. **First co-author**

A novel real-time PCR assay of microRNAs using S-Poly(T), a specific oligo(dT) reverse transcription primer with excellent sensitivity and specificity. Kang K, Zhang XY, Liu HT, Wang ZW, **Zhong JS**, Huang ZT, Peng X, Luo J, Deming Gou. PLoS One. 2012;7(11):e48536.

IX References.

Altmannsberger, M., Weber, K., Droste, R., and Osborn, M. (1985). Desmin is a specific marker for rhabdomyosarcomas of human and rat origin. Am J Pathol *118*, 85-95.

Angelini, C., Di Mauro, S., and Margreth, A. (1968). Relationship of serum enzyme changes to muscle damage in vitamin E deficiency of the rabbit. Sperimentale *118*, 349-369.

Arndt, C.A., and Crist, W.M. (1999). Common musculoskeletal tumors of childhood and adolescence. N Engl J Med *341*, 342-352.

Ashkenazi, R., Gentry, S.N., and Jackson, T.L. (2008). Pathways to tumorigenesis--modeling mutation acquisition in stem cells and their progeny. Neoplasia *10*, 1170-1182.

Aulehla, A., and Pourquié, O. (2010). Signaling gradients during paraxial mesoderm development. Cold Spring Harb Perspect Biol *2*, a000869.

Bajpai, R., and Terskikh, A. (2007). CHAPTER 19 - Genetic Manipulation of Human Embryonic Stem Cells: Lentivirus Vectors. In Human Stem Cell Manual, J.F. Loring, R.L. Wesselschmidt, and P.H. Schwartz, eds. (Oxford: Academic Press), pp. 255-266.

Bandín, S., Morona, R., Moreno, N., and González, A. (2013). Regional expression of Pax7 in the brain of Xenopus laevis during embryonic and larval development. Front Neuroanat *7*, 48.

Barr, F.G. (2001). Gene fusions involving PAX and FOX family members in alveolar rhabdomyosarcoma. Oncogene *20*, 5736-5746.

Barro, M., Carnac, G., Flavier, S., Mercier, J., Vassetzky, Y., and Laoudj-Chenivesse, D. (2010). Myoblasts from affected and non-affected FSHD muscles exhibit morphological differentiation defects. Journal of Cellular and Molecular Medicine *14*, 275-289.

Beauchamp, J.R., Heslop, L., Yu, D.S., Tajbakhsh, S., Kelly, R.G., Wernig, A., Buckingham, M.E., Partridge, T.A., and Zammit, P.S. (2000). Expression of CD34 and Myf5 defines the majority of quiescent adult skeletal muscle satellite cells. J Cell Biol *151*, 1221-1234.

Bénit, L., Lallemand, J.-B., Casella, J.-F., Philippe, H., and Heidmann, T. (1999). ERV-L Elements: a Family of Endogenous Retrovirus-Like Elements Active throughout the Evolution of Mammals. Journal of Virology *73*, 3301-3308.

Bhartiya, D., Boheler, K.R., and Rameshwar, P. (2013). Multipotent to pluripotent properties of adult stem cells. Stem Cells Int *2013*, 813780.

Biferali, B., Proietti, D., Mozzetta, C., and Madaro, L. (2019). Fibro-Adipogenic Progenitors Cross-Talk in Skeletal Muscle: The Social Network. Frontiers in physiology *10*, 1074.

Bosnakovski, D., Gearhart, M.D., Toso, E.A., Recht, O.O., Cucak, A., Jain, A.K., Barton, M.C., and Kyba, M. (2017a). p53-independent DUX4 pathology in cell and animal models of facioscapulohumeral muscular dystrophy. Dis Model Mech *10*, 1211-1216.

Bosnakovski, D., Toso, E.A., Recht, O.O., Cucak, A., Jain, A.K., Barton, M.C., and Kyba, M. (2017b). p53 is not necessary for DUX4 pathology. bioRxiv, 118315.

Bosnakovski, D., Xu, Z., Ji Gang, E., Galindo, C.L., Liu, M., Simsek, T., Garner, H.R., Agha-Mohammadi, S., Tassin, A., Coppée, F., *et al.* (2008a). An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. The EMBO Journal *27*, 2766-2779.

Bosnakovski, D., Xu, Z., Li, W., Thet, S., Cleaver, O., Perlingeiro, R.C., and Kyba, M. (2008b). Prospective isolation of skeletal muscle stem cells with a Pax7 reporter. Stem Cells *26*, 3194-3204.

Braun, T., Rudnicki, M.A., Arnold, H.H., and Jaenisch, R. (1992). Targeted inactivation of the muscle regulatory gene Myf-5 results in abnormal rib development and perinatal death. Cell *71*, 369-382.

Buckingham, M. (1992). Making muscle in mammals. Trends Genet *8*, 144-148.

Buckingham, M., Bajard, L., Chang, T., Daubas, P., Hadchouel, J., Meilhac, S., Montarras, D., Rocancourt, D., and Relaix, F. (2003). The formation of skeletal muscle: from somite to limb. J Anat *202*, 59-68.

Buckingham, M., and Relaix, F. (2007). The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. Annual review of cell and developmental biology *23*, 645-673.

Burkin, D.J., and Kaufman, S.J. (1999). The alpha7beta1 integrin in muscle development and disease. Cell Tissue Res *296*, 183-190.

Camboni, M., Hammond, S., Martin, L.T., and Martin, P.T. (2012). Induction of a regenerative microenvironment in skeletal muscle is sufficient to induce embryonal rhabdomyosarcoma in p53-deficient mice. The Journal of pathology *226*, 40-49.

Capers, C.R. (1960). Multinucleation of skeletal muscle in vitro. J Biophys Biochem Cytol *7*, 559-566.

Cessna, M.H., Zhou, H., Perkins, S.L., Tripp, S.R., Layfield, L., Daines, C., and Coffin, C.M. (2001). Are Myogenin and MyoD1 Expression Specific for

Rhabdomyosarcoma?: A Study of 150 Cases, With Emphasis on Spindle Cell Mimics. The American Journal of Surgical Pathology *25*, 1150-1157.

Chal, J., and Pourquié, O. (2017a). Making muscle: skeletal myogenesis in vivo and in vitro. Development *144*, 2104-2122.

Chal, J., and Pourquié, O. (2017b). Making muscle: skeletal myogenesis in vivo and in vitro. Development *144*, 2104-2122.

Chamberlain, J.S., Metzger, J., Reyes, M., Townsend, D., and Faulkner, J.A. (2007a). Dystrophin-deficient mdx mice display a reduced life span and are susceptible to spontaneous rhabdomyosarcoma. The FASEB Journal *21*, 2195-2204.

Chamberlain, J.S., Metzger, J., Reyes, M., Townsend, D., and Faulkner, J.A. (2007b). Dystrophin-deficient mdx mice display a reduced life span and are susceptible to spontaneous rhabdomyosarcoma. Faseb j *21*, 2195-2204.

Chapman, M.A., Mukund, K., Subramaniam, S., Brenner, D., and Lieber, R.L. (2017). Three distinct cell populations express extracellular matrix proteins and increase in number during skeletal muscle fibrosis. Am J Physiol Cell Physiol *312*, C131-c143.

Charytonowicz, E., Cordon-Cardo, C., Matushansky, I., and Ziman, M. (2009). Alveolar rhabdomyosarcoma: Is the cell of origin a mesenchymal stem cell? Cancer Letters *279*, 126-136.

Chazaud, B., Brigitte, M., Yacoub-Youssef, H., Arnold, L., Gherardi, R., Sonnet, C., Lafuste, P., and Chretien, F. (2009). Dual and beneficial roles of macrophages during skeletal muscle regeneration. Exercise and sport sciences reviews *37*, 18-22.

Chazaud, B., Sonnet, C., Lafuste, P., Bassez, G., Rimaniol, A.C., Poron, F., Authier, F.J., Dreyfus, P.A., and Gherardi, R.K. (2003). Satellite cells attract monocytes and use macrophages as a support to escape apoptosis and enhance muscle growth. J Cell Biol *163*, 1133-1143.

Chen, F., Liu, Y., Wong, N.K., Xiao, J., and So, K.F. (2017). Oxidative Stress in Stem Cell Aging. Cell Transplant *26*, 1483-1495.

Chen, X., Stewart, E., Shelat, A.A., Qu, C., Bahrami, A., Hatley, M., Wu, G., Bradley, C., McEvoy, J., Pappo, A., *et al.* (2013). Targeting oxidative stress in embryonal rhabdomyosarcoma. Cancer Cell *24*, 710-724.

Choi, J., Costa, M.L., Mermelstein, C.S., Chagas, C., Holtzer, S., and Holtzer, H. (1990). MyoD converts primary dermal fibroblasts, chondroblasts, smooth muscle, and retinal pigmented epithelial cells into striated mononucleated myoblasts and multinucleated myotubes. Proc Natl Acad Sci U S A *87*, 7988-7992.

Collaborators, G.M.a.C.o.D. (2016). Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet *388*, 1459-1544.

Collins, C.A., Olsen, I., Zammit, P.S., Heslop, L., Petrie, A., Partridge, T.A., and Morgan, J.E. (2005). Stem Cell Function, Self-Renewal, and Behavioral Heterogeneity of Cells from the Adult Muscle Satellite Cell Niche. Cell *122*, 289-301.

Conerly, M.L., Yao, Z., Zhong, J.W., Groudine, M., and Tapscott, S.J. (2016a). Distinct Activities of Myf5 and MyoD Indicate Separate Roles in Skeletal Muscle Lineage Specification and Differentiation. Dev Cell *36*, 375-385.

Conerly, Melissa L., Yao, Z., Zhong, Jun W., Groudine, M., and Tapscott, Stephen J. (2016b). Distinct Activities of Myf5 and MyoD Indicate Separate Roles in Skeletal Muscle Lineage Specification and Differentiation. Developmental Cell *36*, 375-385.

Cornelison, D.D., Filla, M.S., Stanley, H.M., Rapraeger, A.C., and Olwin, B.B. (2001). 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.

Crescenzi, M., Fleming, T.P., Lassar, A.B., Weintraub, H., and Aaronson, S.A. (1990). MyoD induces growth arrest independent of differentiation in normal and transformed cells. Proc Natl Acad Sci U S A *87*, 8442-8446.

Crichton, J.H., Dunican, D.S., Maclennan, M., Meehan, R.R., and Adams, I.R. (2014). Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline. Cell Mol Life Sci *71*, 1581-1605.

Das, S., and Chadwick, B.P. (2016). Influence of Repressive Histone and DNA Methylation upon D4Z4 Transcription in Non-Myogenic Cells. PLoS One *11*, e0160022.

Davicioni, E., Finckenstein, F.G., Shahbazian, V., Buckley, J.D., Triche, T.J., and Anderson, M.J. (2006). Identification of a PAX-FKHR gene expression signature that defines molecular classes and determines the prognosis of alveolar rhabdomyosarcomas. Cancer Res *66*, 6936-6946.

Davis, R.L., Cheng, P.F., Lassar, A.B., and Weintraub, H. (1990). The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. Cell *60*, 733-746.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987-1000.

De Giovanni, C., Landuzzi, L., Nicoletti, G., Lollini, P.L., and Nanni, P. (2009). Molecular and cellular biology of rhabdomyosarcoma. Future Oncol *5*, 1449-1475.

De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nat Genet *49*, 941-945.

Dias, P., Parham, D.M., Shapiro, D.N., Webber, B.L., and Houghton, P.J. (1990). Myogenic regulatory protein (MyoD1) expression in childhood solid tumors: diagnostic utility in rhabdomyosarcoma. Am J Pathol *137*, 1283-1291.

Dilworth, F.J., and Blais, A. (2011). Epigenetic regulation of satellite cell activation during muscle regeneration. Stem Cell Res Ther *2*, 18.

Drummond, C.J., Hanna, J.A., Garcia, M.R., Devine, D.J., Heyrana, A.J., Finkelstein, D., Rehg, J.E., and Hatley, M.E. (2018). Hedgehog Pathway Drives Fusion-Negative Rhabdomyosarcoma Initiated From Non-myogenic Endothelial Progenitors. Cancer Cell *33*, 108-124.e105.

Drummond, C.J., and Hatley, M.E. (2018). A Case of mistaken identity: Rhabdomyosarcoma development from endothelial progenitor cells. Molecular & Cellular Oncology *5*, e1448246.

Duchenne (1867). The Pathology of Paralysis with Muscular Degeneration (Paralysie Myosclerotique), or Paralysis with Apparent Hypertrophy. British medical journal *2*, 541-542.

Dumont, N., Bouchard, P., and Frenette, J. (2008). Neutrophil-induced skeletal muscle damage: a calculated and controlled response following hindlimb unloading and reloading. American journal of physiology Regulatory, integrative and comparative physiology *295*, R1831-1838.

Dunn, C.A., Romanish, M.T., Gutierrez, L.E., van de Lagemaat, L.N., and Mager, D.L. (2006). Transcription of two human genes from a bidirectional endogenous retrovirus promoter. Gene *366*, 335-342.

Durbin, A.D., Somers, G.R., Forrester, M., Pienkowska, M., Hannigan, G.E., and Malkin, D. (2009). JNK1 determines the oncogenic or tumor-suppressive activity of the integrin-linked kinase in human rhabdomyosarcoma. J Clin Invest *119*, 1558-1570.

Eidahl, J.O., Giesige, C.R., Domire, J.S., Wallace, L.M., Fowler, A.M., Guckes, S.M., Garwick-Coppens, S.E., Labhart, P., and Harper, S.Q. (2016). Mouse Dux is myotoxic and shares partial functional homology with its human paralog DUX4. Hum Mol Genet *25*, 4577-4589.

El Demellawy, D., McGowan-Jordan, J., de Nanassy, J., Chernetsova, E., and Nasr, A. (2017). Update on molecular findings in rhabdomyosarcoma. Pathology *49*, 238-246.

Etienne, J., Liu, C., Skinner, C.M., Conboy, M.J., and Conboy, I.M. (2020). Skeletal muscle as an experimental model of choice to study tissue aging and rejuvenation. Skeletal Muscle *10*, 4.

Ferlay, J., Colombet, M., Soerjomataram, I., Dyba, T., Randi, G., Bettio, M., Gavin, A., Visser, O., and Bray, F. (2018). Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. Eur J Cancer *103*, 356-387.

Fernandez, K., Serinagaoglu, Y., Hammond, S., Martin, L.T., and Martin, P.T. (2010a). Mice lacking dystrophin or alpha sarcoglycan spontaneously develop embryonal rhabdomyosarcoma with cancer-associated p53 mutations and alternatively spliced or mutant Mdm2 transcripts. Am J Pathol *176*, 416-434.

Fernandez, K., Serinagaoglu, Y., Hammond, S., Martin, L.T., and Martin, P.T. (2010b). Mice Lacking Dystrophin or α Sarcoglycan Spontaneously Develop Embryonal Rhabdomyosarcoma with Cancer-Associated p53 Mutations and Alternatively Spliced or Mutant Mdm2 Transcripts. The American Journal of Pathology *176*, 416-434.

Fiorotto, M. (2012). The making of a muscle. Biochem (Lond) *34*, 4-11.

Fleischmann, A., Jochum, W., Eferl, R., Witowsky, J., and Wagner, E.F. (2003). Rhabdomyosarcoma development in mice lacking Trp53 and Fos: tumor suppression by the Fos protooncogene. Cancer Cell *4*, 477-482.

Flex, E., Jaiswal, M., Pantaleoni, F., Martinelli, S., Strullu, M., Fansa, E.K., Caye, A., De Luca, A., Lepri, F., Dvorsky, R., *et al.* (2014). Activating mutations in RRAS underlie a phenotype within the RASopathy spectrum and contribute to leukaemogenesis. Hum Mol Genet *23*, 4315-4327.

Frenette, J., Cai, B., and Tidball, J.G. (2000). Complement activation promotes muscle inflammation during modified muscle use. Am J Pathol *156*, 2103-2110.

Frontera, W.R., and Ochala, J. (2015). Skeletal muscle: a brief review of structure and function. Calcif Tissue Int *96*, 183-195.

Fu, B., Ma, H., and Liu, D. (2019). Endogenous Retroviruses Function as Gene Expression Regulatory Elements During Mammalian Pre-implantation Embryo Development. Int J Mol Sci *20*.

Fukada, S., Uezumi, A., Ikemoto, M., Masuda, S., Segawa, M., Tanimura, N., Yamamoto, H., Miyagoe-Suzuki, Y., and Takeda, S. (2007). Molecular signature of quiescent satellite cells in adult skeletal muscle. Stem Cells *25*, 2448-2459.

Geng, Linda N., Yao, Z., Snider, L., Fong, Abraham P., Cech, Jennifer N., Young, Janet M., van der Maarel, Silvere M., Ruzzo, Walter L., Gentleman, Robert C., Tawil, R., *et*

al. (2012). DUX4 Activates Germline Genes, Retroelements, and Immune Mediators: Implications for Facioscapulohumeral Dystrophy. Developmental Cell *22*, 38-51.

Gensch, N., Borchardt, T., Schneider, A., Riethmacher, D., and Braun, T. (2008). Different autonomous myogenic cell populations revealed by ablation of Myf5-expressing cells during mouse embryogenesis. Development *135*, 1597-1604.

Gerber, N.K., Wexler, L.H., Singer, S., Alektiar, K.M., Keohan, M.L., Shi, W., Zhang, Z., and Wolden, S. (2013). Adult rhabdomyosarcoma survival improved with treatment on multimodality protocols. Int J Radiat Oncol Biol Phys *86*, 58-63.

Gillies, A.R., and Lieber, R.L. (2011). Structure and function of the skeletal muscle extracellular matrix. Muscle Nerve *44*, 318-331.

Goh, Q., and Millay, D.P. (2017). Requirement of myomaker-mediated stem cell fusion for skeletal muscle hypertrophy. eLife *6*.

Gros, J., Manceau, M., Thomé, V., and Marcelle, C. (2005). A common somitic origin for embryonic muscle progenitors and satellite cells. Nature *435*, 954-958.

Gruss, P., and Walther, C. (1992). Pax in development. Cell *69*, 719-722.

Günther, S., Kim, J., Kostin, S., Lepper, C., Fan, C.-M., and Braun, T. (2013). Myf5-Positive Satellite Cells Contribute to Pax7-Dependent Long-Term Maintenance of Adult Muscle Stem Cells. Cell Stem Cell *13*, 590-601.

Haldar, M., Karan, G., Watanabe, S., Guenther, S., Braun, T., and Capecchi, M.R. (2014). Response: Contributions of the Myf5-independent lineage to myogenesis. Dev Cell *31*, 539-541.

Halevy, O., Novitch, B.G., Spicer, D.B., Skapek, S.X., Rhee, J., Hannon, G.J., Beach, D., and Lassar, A.B. (1995). Correlation of terminal cell cycle arrest of skeletal muscle with induction of p21 by MyoD. Science *267*, 1018-1021.

Hasty, P., Bradley, A., Morris, J.H., Edmondson, D.G., Venuti, J.M., Olson, E.N., and Klein, W.H. (1993). Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene. Nature *364*, 501-506.

Hendrickson, P.G., Doráis, J.A., Grow, E.J., Whiddon, J.L., Lim, J.W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., *et al.* (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat Genet *49*, 925-934.

Himeda, C.L., and Jones, P.L. (2019). The Good, The Bad, and The Unexpected: Roles of DUX4 in Health and Disease. Developmental Cell *50*, 525-526.

Hosur, V., Kavirayani, A., Riefler, J., Carney, L.M., Lyons, B., Gott, B., Cox, G.A., and Shultz, L.D. (2012a). Dystrophin and dysferlin double mutant mice: a novel model for rhabdomyosarcoma. Cancer genetics *205*, 232-241.

Hosur, V., Kavirayani, A., Riefler, J., Carney, L.M.B., Lyons, B., Gott, B., Cox, G.A., and Shultz, L.D. (2012b). Dystrophin and dysferlin double mutant mice: a novel model for rhabdomyosarcoma. Cancer Genetics *205*, 232-241.

Hu, P., Geles, K.G., Paik, J.-H., DePinho, R.A., and Tjian, R. (2008). Codependent Activators Direct Myoblast-Specific MyoD Transcription. Developmental Cell *15*, 534-546.

Huard, J., Li, Y., and Fu, F.H. (2002). Muscle Injuries and Repair: Current Trends in Research. *84*, 822-832.

Hutcheson, D.A., Zhao, J., Merrell, A., Haldar, M., and Kardon, G. (2009). Embryonic and fetal limb myogenic cells are derived from developmentally distinct progenitors and have different requirements for beta-catenin. Genes Dev *23*, 997-1013.

Jones, T.I., King, O.D., Himeda, C.L., Homma, S., Chen, J.C.J., Beermann, M.L., Yan, C., Emerson, C.P., Miller, J.B., Wagner, K.R., *et al.* (2015). Individual epigenetic status of the pathogenic D4Z4 macrosatellite correlates with disease in facioscapulohumeral muscular dystrophy. Clinical Epigenetics *7*, 37.

Kablar, B., Krastel, K., Ying, C., Asakura, A., Tapscott, S.J., and Rudnicki, M.A. (1997). MyoD and Myf-5 differentially regulate the development of limb versus trunk skeletal muscle. Development *124*, 4729-4738.

Kassar-Duchossoy, L., Giacone, E., Gayraud-Morel, B., Jory, A., Gomès, D., and Tajbakhsh, S. (2005). Pax3/Pax7 mark a novel population of primitive myogenic cells during development. Genes Dev *19*, 1426-1431.

Kassiotis, G. (2014). Endogenous retroviruses and the development of cancer. J Immunol *192*, 1343-1349.

Kawamura-Saito, M., Yamazaki, Y., Kaneko, K., Kawaguchi, N., Kanda, H., Mukai, H., Gotoh, T., Motoi, T., Fukayama, M., Aburatani, H., *et al.* (2006). Fusion between CIC and DUX4 up-regulates PEA3 family genes in Ewing-like sarcomas with t(4;19)(q35;q13) translocation. Hum Mol Genet *15*, 2125-2137.

Kim, J., and Braun, T. (2014). Skeletal muscle stem cells for muscle regeneration. Methods Mol Biol *1213*, 245-253.

Knapp, J.R., Davie, J.K., Myer, A., Meadows, E., Olson, E.N., and Klein, W.H. (2006). Loss of myogenin in postnatal life leads to normal skeletal muscle but reduced body size. Development *133*, 601-610.

Knopp, P., Krom, Y.D., Banerji, C.R.S., Panamarova, M., Moyle, L.A., den Hamer, B., van der Maarel, S.M., and Zammit, P.S. (2016). DUX4 induces a transcriptome more characteristic of a less-differentiated cell state and inhibits myogenesis. Journal of Cell Science *129*, 3816-3831.

Koyuncu, S., Irmak, D., Saez, I., and Vilchez, D. (2015). Defining the General Principles of Stem Cell Aging: Lessons from Organismal Models. Current Stem Cell Reports *1*, 162-169.

Kuang, S., Chargé, S.B., Seale, P., Huh, M., and Rudnicki, M.A. (2006). Distinct roles for Pax7 and Pax3 in adult regenerative myogenesis. J Cell Biol *172*, 103-113.

Kumar, S., Perlman, E., Harris, C.A., Raffeld, M., and Tsokos, M. (2000). Myogenin is a specific marker for rhabdomyosarcoma: an immunohistochemical study in paraffin-embedded tissues. Mod Pathol *13*, 988-993.

Kwon, M., and Firestein, B.L. (2013). DNA Transfection: Calcium Phosphate Method. In Neural Development: Methods and Protocols, R. Zhou, and L. Mei, eds. (Totowa, NJ: Humana Press), pp. 107-110.

Lamarche, É., Lala-Tabbert, N., Gunanayagam, A., St-Louis, C., and Wiper-Bergeron, N. (2015). Retinoic acid promotes myogenesis in myoblasts by antagonizing transforming growth factor-beta signaling via C/EBPβ. Skeletal Muscle *5*, 8.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Langenau, D.M., Keefe, M.D., Storer, N.Y., Guyon, J.R., Kutok, J.L., Le, X., Goessling, W., Neuberg, D.S., Kunkel, L.M., and Zon, L.I. (2007). Effects of RAS on the genesis of embryonal rhabdomyosarcoma. Genes Dev *21*, 1382-1395.

Laptenko, O., and Prives, C. (2006). Transcriptional regulation by p53: one protein, many possibilities. Cell Death Differ *13*, 951-961.

Laterza, O.F., Lim, L., Garrett-Engele, P.W., Vlasakova, K., Muniappa, N., Tanaka, W.K., Johnson, J.M., Sina, J.F., Fare, T.L., Sistare, F.D., *et al.* (2009). Plasma MicroRNAs as sensitive and specific biomarkers of tissue injury. Clin Chem *55*, 1977-1983.

Laverdure, S., Polakowski, N., Hoang, K., and Lemasson, I. (2016). Permissive Sense and Antisense Transcription from the 5' and 3' Long Terminal Repeats of Human T-Cell Leukemia Virus Type 1. J Virol *90*, 3600-3610.

Le Grand, F., and Rudnicki, M.A. (2007). Skeletal muscle satellite cells and adult myogenesis. Curr Opin Cell Biol *19*, 628-633.

Lee, H.-E., Jo, A., Im, J., Cha, H.-J., Kim, W.-J., Kim, H.H., Kim, D.-S., Kim, W., Yang, T.-J., and Kim, H.-S. (2019). Characterization of the Long Terminal Repeat of the Endogenous Retrovirus-derived microRNAs in the Olive Flounder. Scientific Reports *9*, 14007.

Leidenroth, A., and Hewitt, J.E. (2010a). A family history of DUX4: phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene. BMC Evol Biol *10*, 364.

Leidenroth, A., and Hewitt, J.E. (2010b). A family history of DUX4: phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUXgene. BMC Evolutionary Biology *10*, 364.

Lilljebjorn, H., and Fioretos, T. (2017). New oncogenic subtypes in pediatric B-cell precursor acute lymphoblastic leukemia. Blood *130*, 1395-1401.

Lin, Y.-C., Boone, M., Meuris, L., Lemmens, I., Van Roy, N., Soete, A., Reumers, J., Moisse, M., Plaisance, S., Drmanac, R., *et al.* (2014). Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. Nature Communications *5*, 4767.

Linardic, C.M., Downie, D.L., Qualman, S., Bentley, R.C., and Counter, C.M. (2005). Genetic modeling of human rhabdomyosarcoma. Cancer Res *65*, 4490-4495.

Liu, C., Li, D., Hu, J., Jiang, J., Zhang, W., Chen, Y., Cui, X., Qi, Y., Zou, H., Zhang, W., *et al.* (2014). Chromosomal and genetic imbalances in Chinese patients with rhabdomyosarcoma detected by high-resolution array comparative genomic hybridization. Int J Clin Exp Pathol *7*, 690-698.

Liu, N., Garry, Glynnis A., Li, S., Bezprozvannaya, S., Sanchez-Ortiz, E., Chen, B., Shelton, John M., Jaichander, P., Bassel-Duby, R., and Olson, Eric N. (2017). A Twist2-dependent progenitor cell contributes to adult skeletal muscle. Nature Cell Biology *19*, 202-213.

Liu, Q.C., Zha, X.H., Faralli, H., Yin, H., Louis-Jeune, C., Perdiguero, E., Pranckeviciene, E., Muñoz-Cànoves, P., Rudnicki, M.A., Brand, M., *et al.* (2012). Comparative expression profiling identifies differential roles for Myogenin and p38α MAPK signaling in myogenesis. J Mol Cell Biol *4*, 386-397.

Liu, X., Wang, Y., Gao, Y., Su, J., Zhang, J., Xing, X., Zhou, C., Yao, K., An, Q., and Zhang, Y. (2018). H3K9 demethylase KDM4E is an epigenetic regulator for bovine embryonic development and a defective factor for nuclear reprogramming. Development *145*.

Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature *487*, 57-63.

Mager, D.L., and Stoye, J.P. (2015). Mammalian Endogenous Retroviruses. Microbiol Spectr *3*, Mdna3-0009-2014.

Maksakova, I.A., Romanish, M.T., Gagnier, L., Dunn, C.A., van de Lagemaat, L.N., and Mager, D.L. (2006). Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLoS Genet *2*, e2.

Mani, C., Reddy, P.H., and Palle, K. (2020). DNA repair fidelity in stem cell maintenance, health, and disease. Biochim Biophys Acta Mol Basis Dis *1866,* 165444.

Marcho, C., Cui, W., and Mager, J. (2015). Epigenetic dynamics during preimplantation development. Reproduction *150*, R109-120.

Mauro, A. (1961). Satellite cell of skeletal muscle fibers. J Biophys Biochem Cytol *9*, 493-495.

Meadows, E., Cho, J.-H., Flynn, J.M., and Klein, W.H. (2008). Myogenin regulates a distinct genetic program in adult muscle stem cells. Developmental Biology *322*, 406-414.

Megeney, L.A., and Rudnicki, M.A. (1995). Determination versus differentiation and the MyoD family of transcription factors. Biochem Cell Biol *73*, 723-732.

Montarras, D., Morgan, J., Collins, C., Relaix, F., Zaffran, S., Cumano, A., Partridge, T., and Buckingham, M. (2005). Direct isolation of satellite cells for skeletal muscle regeneration. Science *309*, 2064-2067.

Moss, F.P., and Leblond, C.P. (1971). Satellite cells as the source of nuclei in muscles of growing rats. Anat Rec *170*, 421-435.

Mozdziak, P.E., Pulvermacher, P.M., and Schultz, E. (2001). Muscle regeneration during hindlimb unloading results in a reduction in muscle size after reloading. Journal of Applied Physiology *91*, 183-190.

Murphy, M., and Kardon, G. (2011). Origin of vertebrate limb muscle: the role of progenitor and myoblast populations. Curr Top Dev Biol *96*, 1-32.

Murphy, M.M., Lawson, J.A., Mathew, S.J., Hutcheson, D.A., and Kardon, G. (2011a). Satellite cells, connective tissue fibroblasts and their interactions are crucial for muscle regeneration. Development *138*, 3625-3637.

Murphy, M.M., Lawson, J.A., Mathew, S.J., Hutcheson, D.A., and Kardon, G. (2011b). Satellite cells, connective tissue fibroblasts and their interactions are crucial for muscle regeneration. *138*, 3625-3637.

Musarò, A. (2014). The Basis of Muscle Regeneration. Advances in Biology *2014*, 612471.

Nabeshima, Y., Hanaoka, K., Hayasaka, M., Esumi, E., Li, S., Nonaka, I., and Nabeshima, Y. (1993). Myogenin gene disruption results in perinatal lethality because of severe muscle defect. Nature *364*, 532-535.

Nguyen, H.T., Hong, X., Tan, S., Chen, Q., Chan, L., Fivaz, M., Cohen, S.M., and Voorhoeve, P.M. (2014). Viral small T oncoproteins transform cells by alleviating hippo-pathway-mediated inhibition of the YAP proto-oncogene. Cell reports *8*, 707-713.

Nishimura, R., Takita, J., Sato-Otsubo, A., Kato, M., Koh, K., Hanada, R., Tanaka, Y., Kato, K., Maeda, D., Fukayama, M., *et al.* (2013). Characterization of genetic lesions in rhabdomyosarcoma using a high-density single nucleotide polymorphism array. Cancer science *104*, 856-864.

Ognjanovic, S., Linabery, A.M., Charbonneau, B., and Ross, J.A. (2009). Trends in childhood rhabdomyosarcoma incidence and survival in the United States, 1975-2005. Cancer *115*, 4218-4226.

Pan, G.J., Chang, Z.Y., SchÖLer, H.R., and Pei, D. (2002). Stem cell pluripotency and transcription factor Oct4. Cell Research *12*, 321-329.

Parham, D.M., Webber, B., Holt, H., Williams, W.K., and Maurer, H. (1991). Immunohistochemical study of childhood rhabdomyosarcomas and related neoplasms. Results of an Intergroup Rhabdomyosarcoma study project. Cancer *67*, 3072-3080.

Pasut, A., Oleynik, P., and Rudnicki, M.A. (2012). Isolation of muscle stem cells by fluorescence activated cell sorting cytometry. Methods Mol Biol *798*, 53-64.

Paulson, V., Chandler, G., Rakheja, D., Galindo, R.L., Wilson, K., Amatruda, J.F., and Cameron, S. (2011). High-resolution array CGH identifies common mechanisms that drive embryonal rhabdomyosarcoma pathogenesis. Genes, Chromosomes and Cancer *50*, 397-408.

Pham, S., and Puckett, Y. (2020). Physiology, Skeletal Muscle Contraction. In StatPearls (Treasure Island (FL)).

Pietsch, P. (1961). The effects of colchicine on regeneration of mouse skeletal muscle. *139*, 167-172.

Preussner, J., Zhong, J., Sreenivasan, K., Günther, S., Engleitner, T., Künne, C., Glatzel, M., Rad, R., Looso, M., Braun, T., *et al.* (2018). Oncogenic Amplification of Zygotic Dux Factors in Regenerating p53-Deficient Muscle Stem Cells Defines a Molecular Cancer Subtype. Cell Stem Cell *23*, 794-805.e794.

Purslow, P.P. (2008). The Extracellular Matrix of Skeletal and Cardiac Muscle. In Collagen: Structure and Mechanics, P. Fratzl, ed. (Boston, MA: Springer US), pp. 325-357.

Quinlan, J.G., Lyden, S.P., Cambier, D.M., Johnson, S.R., Michaels, S.E., and Denman, D.L. (1995). Radiation inhibition of mdx mouse muscle regeneration: dose and age factors. Muscle & nerve *18*, 201-206.

Radley, H.G., and Grounds, M.D. (2006). Cromolyn administration (to block mast cell degranulation) reduces necrosis of dystrophic muscle in mdx mice. Neurobiology of disease *23*, 387-397.

Ratajczak, M.Z., Majka, M., Kucia, M., Drukala, J., Pietrzkowski, Z., Peiper, S., and Janowska-Wieczorek, A. (2003). Expression of functional CXCR4 by muscle satellite cells and secretion of SDF-1 by muscle-derived fibroblasts is associated with the presence of both muscle progenitors in bone marrow and hematopoietic stem/progenitor cells in muscles. Stem Cells *21*, 363-371.

Rawls, A., Morris, J.H., Rudnicki, M., Braun, T., Arnold, H.-H., Klein, W.H., and Olson, E.N. (1995). Myogenin's Functions Do Not Overlap with Those of MyoD or Myf-5 during Mouse Embryogenesis. Developmental Biology *172*, 37-50.

Relaix, F. (2006). Skeletal muscle progenitor cells: from embryo to adult. Cell Mol Life Sci *63*, 1221-1225.

Relaix, F., Bencze, M., Borok, M.J., Der Vartanian, A., Gattazzo, F., Mademtzoglou, D., Perez-Diaz, S., Prola, A., Reyes-Fernandez, P.C., Rotini, A., *et al.* (2021). Perspectives on skeletal muscle stem cells. Nature Communications *12*, 692.

Relaix, F., Rocancourt, D., Mansouri, A., and Buckingham, M. (2005). A Pax3/Pax7-dependent population of skeletal muscle progenitor cells. Nature *435*, 948-953.

Rodić, N., Sharma, R., Sharma, R., Zampella, J., Dai, L., Taylor, M.S., Hruban, R.H., Iacobuzio-Donahue, C.A., Maitra, A., Torbenson, M.S., *et al.* (2014). Long interspersed element-1 protein expression is a hallmark of many human cancers. Am J Pathol *184*, 1280-1286.

Rubin, B.P., Nishijo, K., Chen, H.I., Yi, X., Schuetze, D.P., Pal, R., Prajapati, S.I., Abraham, J., Arenkiel, B.R., Chen, Q.R., *et al.* (2011). Evidence for an unanticipated relationship between undifferentiated pleomorphic sarcoma and embryonal rhabdomyosarcoma. Cancer Cell *19*, 177-191.

Rudnicki, M.A., Braun, T., Hinuma, S., and Jaenisch, R. (1992). Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development. Cell *71*, 383-390.

Sabourin, L.A., Girgis-Gabardo, A., Seale, P., Asakura, A., and Rudnicki, M.A. (1999). Reduced differentiation potential of primary MyoD-/- myogenic cells derived from adult skeletal muscle. J Cell Biol *144*, 631-643.

Sambasivan, R., Yao, R., Kissenpfennig, A., Van Wittenberghe, L., Paldi, A., Gayraud-Morel, B., Guenou, H., Malissen, B., Tajbakhsh, S., and Galy, A. (2011). Pax7-expressing satellite cells are indispensable for adult skeletal muscle regeneration. Development *138*, 3647-3656.

Schienda, J., Engleka, K.A., Jun, S., Hansen, M.S., Epstein, J.A., Tabin, C.J., Kunkel, L.M., and Kardon, G. (2006). Somitic origin of limb muscle satellite and side population cells. Proc Natl Acad Sci U S A *103*, 945-950.

Seale, P., Sabourin, L.A., Girgis-Gabardo, A., Mansouri, A., Gruss, P., and Rudnicki, M.A. (2000). Pax7 is required for the specification of myogenic satellite cells. Cell *102*, 777-786.

Sebire, N.J., and Malone, M. (2003). Myogenin and MyoD1 expression in paediatric rhabdomyosarcomas. Journal of Clinical Pathology *56*, 412-416.

Seki, M., Nishimura, R., Yoshida, K., Shimamura, T., Shiraishi, Y., Sato, Y., Kato, M., Chiba, K., Tanaka, H., Hoshino, N., *et al.* (2015). Integrated genetic and epigenetic analysis defines novel molecular subgroups in rhabdomyosarcoma. Nat Commun *6*, 7557.

Sharpe, B., Beresford, M., Bowen, R., Mitchard, J., and Chalmers, A.D. (2013). Searching for prostate cancer stem cells: markers and methods. Stem Cell Rev Rep *9*, 721-730.

Shern, J.F., Chen, L., Chmielecki, J., Wei, J.S., Patidar, R., Rosenberg, M., Ambrogio, L., Auclair, D., Wang, J., Song, Y.K., *et al.* (2014). Comprehensive genomic analysis of rhabdomyosarcoma reveals a landscape of alterations affecting a common genetic axis in fusion-positive and fusion-negative tumors. Cancer Discov *4*, 216-231.

Shern, J.F., Yohe, M.E., and Khan, J. (2015). Pediatric Rhabdomyosarcoma. Crit Rev Oncog *20*, 227-243.

Skapek, S.X., Ferrari, A., Gupta, A.A., Lupo, P.J., Butler, E., Shipley, J., Barr, F.G., and Hawkins, D.S. (2019). Rhabdomyosarcoma. Nature Reviews Disease Primers *5*, 1.

Slemmons, K.K., Crose, L.E.S., Rudzinski, E., Bentley, R.C., and Linardic, C.M. (2015). Role of the YAP Oncoprotein in Priming Ras-Driven Rhabdomyosarcoma. PLOS ONE *10*, e0140781.

Snow, M.H. (1977). Myogenic cell formation in regenerating rat skeletal muscle injured by mincing. II. An autoradiographic study. Anat Rec *188*, 201-217.

Soini, Y., Kosma, V.M., and Pirinen, R. (2015). KDM4A, KDM4B and KDM4C in non-small cell lung cancer. Int J Clin Exp Pathol *8*, 12922-12928.

Sorensen, P.H., Lynch, J.C., Qualman, S.J., Tirabosco, R., Lim, J.F., Maurer, H.M., Bridge, J.A., Crist, W.M., Triche, T.J., and Barr, F.G. (2002). PAX3-FKHR and PAX7-FKHR gene fusions are prognostic indicators in alveolar rhabdomyosarcoma: a report from the children's oncology group. J Clin Oncol *20*, 2672-2679.

Sugie, K., Funaya, S., Kawamura, M., Nakamura, T., Suzuki, M.G., and Aoki, F. (2020). Expression of Dux family genes in early preimplantation embryos. Scientific Reports *10*, 19396.

Sultan, I., Qaddoumi, I., Yaser, S., Rodriguez-Galindo, C., and Ferrari, A. (2009). Comparing adult and pediatric rhabdomyosarcoma in the surveillance, epidemiology and end results program, 1973 to 2005: an analysis of 2,600 patients. J Clin Oncol *27*, 3391-3397.

Sun, X., Guo, W., Shen, J.K., Mankin, H.J., Hornicek, F.J., and Duan, Z. (2015). Rhabdomyosarcoma: Advances in Molecular and Cellular Biology. Sarcoma *2015*, 232010.

Tajbakhsh, S. (2009). Skeletal muscle stem cells in developmental versus regenerative myogenesis. Journal of internal medicine *266*, 372-389.

Tajbakhsh, S., and Cossu, G. (1997). Establishing myogenic identity during somitogenesis. Current Opinion in Genetics & Development *7*, 634-641.

Tan, L., and Shi, Y.G. (2012). Tet family proteins and 5-hydroxymethylcytosine in development and disease. Development *139*, 1895-1902.

Tapscott, S.J. (2005). The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. Development *132*, 2685-2695.

Taulli, R., Scuoppo, C., Bersani, F., Accornero, P., Forni, P.E., Miretti, S., Grinza, A., Allegra, P., Schmitt-Ney, M., Crepaldi, T., *et al.* (2006). Validation of met as a therapeutic target in alveolar and embryonal rhabdomyosarcoma. Cancer Res *66*, 4742-4749.

Tirode, F., Laud-Duval, K., Prieur, A., Delorme, B., Charbord, P., and Delattre, O. (2007). Mesenchymal Stem Cell Features of Ewing Tumors. Cancer Cell *11*, 421-429.

Tremblay, A.M., Missiaglia, E., Galli, G.G., Hettmer, S., Urcia, R., Carrara, M., Judson, R.N., Thway, K., Nadal, G., Selfe, J.L., *et al.* (2014). The Hippo transducer YAP1 transforms activated satellite cells and is a potent effector of embryonal rhabdomyosarcoma formation. Cancer Cell *26*, 273-287.

Ustanina, S., Carvajal, J., Rigby, P., and Braun, T. (2007). The Myogenic Factor Myf5 Supports Efficient Skeletal Muscle Regeneration by Enabling Transient Myoblast Amplification. STEM CELLS *25*, 2006-2016.

Venuti, J.M., Morris, J.H., Vivian, J.L., Olson, E.N., and Klein, W.H. (1995). Myogenin is required for late but not early aspects of myogenesis during mouse development. Journal of Cell Biology *128*, 563-576.

Visvader, J.E. (2011). Cells of origin in cancer. Nature *469*, 314-322.

Vogelstein, B., Fearon, E.R., Hamilton, S.R., Kern, S.E., Preisinger, A.C., Leppert, M., Nakamura, Y., White, R., Smits, A.M., and Bos, J.L. (1988). Genetic alterations during colorectal-tumor development. N Engl J Med *319*, 525-532.

Volckaert, T., and De Langhe, S.P. (2015). Wnt and FGF mediated epithelial-mesenchymal crosstalk during lung development. Developmental dynamics : an official publication of the American Association of Anatomists *244*, 342-366.

Wachtel, M., Dettling, M., Koscielniak, E., Stegmaier, S., Treuner, J., Simon-Klingenstein, K., Bühlmann, P., Niggli, F.K., and Schäfer, B.W. (2004). Gene expression signatures identify rhabdomyosarcoma subtypes and detect a novel t(2;2)(q35;p23) translocation fusing PAX3 to NCOA1. Cancer Res *64*, 5539-5545.

Wachtel, M., Runge, T., Leuschner, I., Stegmaier, S., Koscielniak, E., Treuner, J., Odermatt, B., Behnke, S., Niggli, F.K., and Schäfer, B.W. (2006). Subtype and prognostic classification of rhabdomyosarcoma by immunohistochemistry. J Clin Oncol *24*, 816-822.

Wachtel, M., and Schäfer, B.W. (2015). Unpeaceful roles of mutant PAX proteins in cancer. Semin Cell Dev Biol *44*, 126-134.

Wallace, L.M., Garwick, S.E., Mei, W., Belayew, A., Coppee, F., Ladner, K.J., Guttridge, D., Yang, J., and Harper, S.Q. (2011). DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. Ann Neurol *69*, 540-552.

Walther, C., Mayrhofer, M., Nilsson, J., Hofvander, J., Jonson, T., Mandahl, N., Øra, I., Gisselsson, D., and Mertens, F. (2016). Genetic heterogeneity in rhabdomyosarcoma revealed by SNP array analysis. Genes, Chromosomes and Cancer *55*, 3-15.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., *et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Wehr, R., and Gruss, P. (1996). Pax and vertebrate development. Int J Dev Biol *40*, 369-377.

White, A.C., and Lowry, W.E. (2015). Refining the role for adult stem cells as cancer cells of origin. Trends Cell Biol *25*, 11-20.

Wu, S.L., Li, G.Z., Chou, C.Y., Tsai, M.S., Chen, Y.P., Li, C.J., Liou, G.G., Chang, W.W., Chen, S.L., and Wang, S.H. (2014). Double homeobox gene, Duxbl, promotes myoblast proliferation and abolishes myoblast differentiation by blocking MyoD transactivation. Cell Tissue Res *358*, 551-566.

Xu, R., Li, C., Liu, X., and Gao, S. (2021). Insights into epigenetic patterns in mammalian early embryos. Protein & Cell *12*, 7-28.

Yasuda, T., Tsuzuki, S., Kawazu, M., Hayakawa, F., Kojima, S., Ueno, T., Imoto, N., Kohsaka, S., Kunita, A., Doi, K., *et al.* (2016a). Corrigendum: Recurrent DUX4 fusions in B cell acute lymphoblastic leukemia of adolescents and young adults. Nat Genet *48*, 1591.

Yasuda, T., Tsuzuki, S., Kawazu, M., Hayakawa, F., Kojima, S., Ueno, T., Imoto, N., Kohsaka, S., Kunita, A., Doi, K., *et al.* (2016b). Recurrent DUX4 fusions in B cell acute lymphoblastic leukemia of adolescents and young adults. Nat Genet *48*, 569-574.

Yin, H., Price, F., and Rudnicki, M.A. (2013). Satellite cells and the muscle stem cell niche. Physiol Rev *93*, 23-67.

Young, J.M., Whiddon, J.L., Yao, Z., Kasinathan, B., Snider, L., Geng, L.N., Balog, J., Tawil, R., van der Maarel, S.M., and Tapscott, S.J. (2013). DUX4 binding to retroelements creates promoters that are active in FSHD muscle and testis. PLoS Genet *9*, e1003947.

Younger, S.T., and Rinn, J.L. (2017). p53 regulates enhancer accessibility and activity in response to DNA damage. Nucleic Acids Res *45*, 9889-9900.

Zammit, P.S., Relaix, F., Nagata, Y., Ruiz, A.P., Collins, C.A., Partridge, T.A., and Beauchamp, J.R. (2006). Pax7 and myogenic progression in skeletal muscle satellite cells. Journal of Cell Science *119*, 1824-1832.

Zhang, J., McCastlain, K., Yoshihara, H., Xu, B., Chang, Y., Churchman, M.L., Wu, G., Li, Y., Wei, L., Iacobucci, I., *et al.* (2016). Deregulation of DUX4 and ERG in acute lymphoblastic leukemia. Nat Genet *48*, 1481-1489.

Zipfel, P.F., and Reuter, M. (2009). Complement Activation Products C3a and C4a as Endogenous Antimicrobial Peptides. International Journal of Peptide Research and Therapeutics *15*, 87.