DNA Methylation Dynamics during Myogenesis

Elvira Carrió Gaspar

DNA Methylation Dynamics during Myogenesis

Les dinàmiques de metilació de l'ADN durant la miogènesi

Memòria presentada per

Elvira Carrió Gaspar

per optar al títol de

Doctora en Biologia per la Universitat de Barcelona

Programa de Genètica, Departament de Genètica

Tesi realitzada sota la direcció de

la **Dra. Mònica Suelves Esteban** i el **Dr. Miguel Angel Peinado Morales**

A l'Institut de Medicina Predictiva i Personalitzada del Càncer (IMPPC).

Dra. Mònica Suelves Esteban Dr. Miguel Angel Peinado Morales

Dra. Lluïsa Vilageliu Arqués Elvira Carrió Gaspar

(Tutora) Barcelona, 2015

Als meus pares i al Pau

A la Sofia

Al Lau

Epigenetics folds the genome, shaping the cellular identity

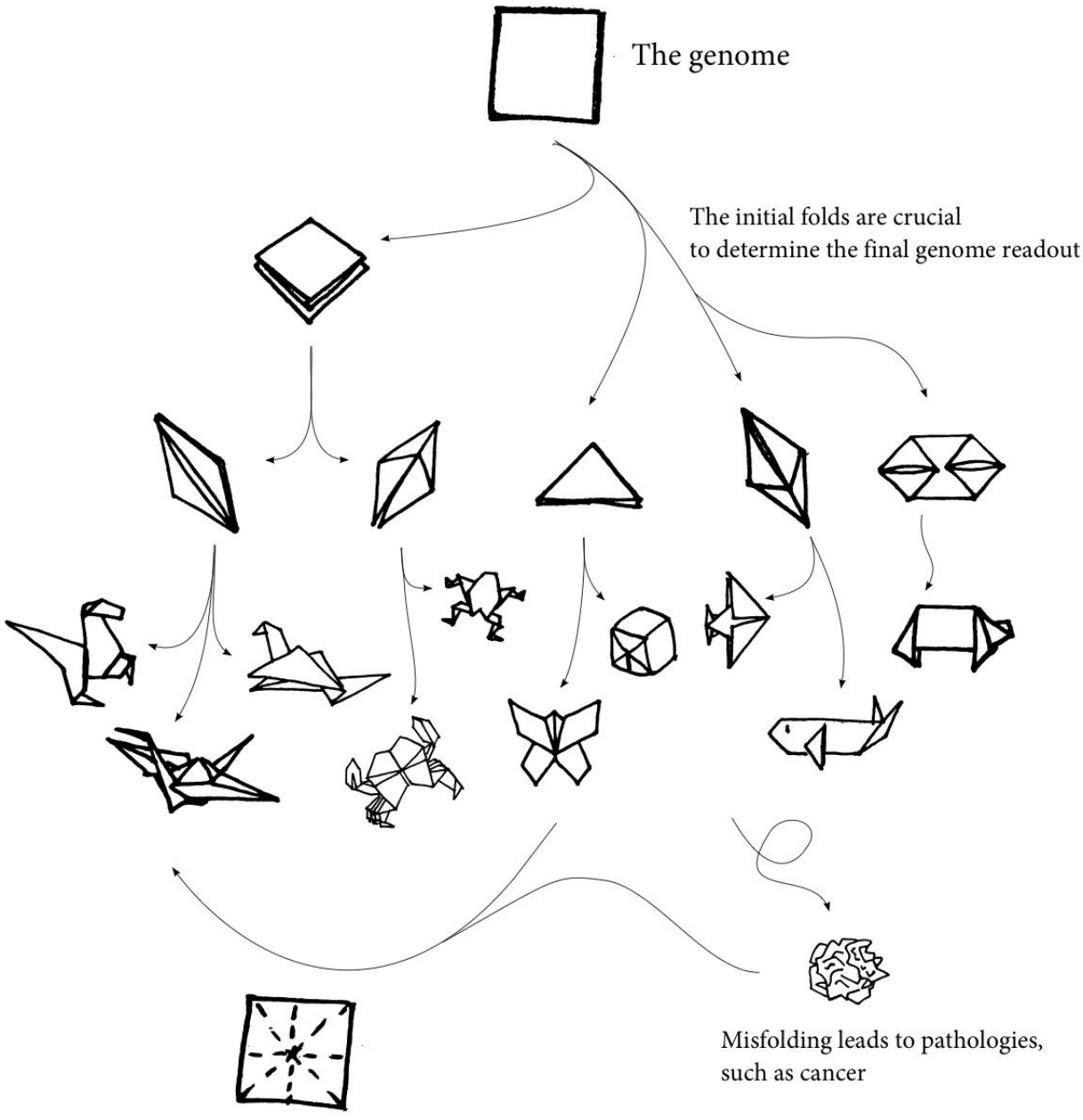


Figure index

Table index

Abbreviations

5caC:	5-Carboxylcytosine
5fc:	5-Formylcytosine
5hmc:	5-Hydroximetilcytosina
5mc:	5-Metylcitosine
Ac:	Acetylation
AIMS:	Amplification of Intermethylated Sites
aRMS:	alveolar Rhabdomyosarcoma
ATP:	Adenosine triphosphate
Aza:	5'Aza2 deoxycytidine
BER:	Base excision repair
bFgf:	basic Fibroblast growth factor
BMP:	Bone morphogenic proteins
Bp:	Base pair
C:	Cytosine
ChIP:	Chromatin immunoprecipitation
CpGi:	CpG island
DM:	Differentiation medium
DMR:	Differentially methylated region
DNA:	Desoxyribonucleic acid
Dnmt:	DNA methyltransferase
EB:	Embryoid body
EBM:	EB differentiation medium
EEE:	Early epaxial enhancer
eRMS:	embryonal Rhabdomyosarcoma
ESC:	Embryonic stem cells
FACS:	Fluorescence-Activated Cell Sorting
FBS:	Fetal bovine serum
G:	Guanine
GM:	Growth medium
H:	Histone
HAT:	Histone acetyltransferase
HDAC:	Histone deacetyltransferase
HS:	Horse serum
ICM:	Inner cell mass
i*Pax7:*	inducible *Pax7*
iPs:	induced pluripotent cells
K:	Lysine
Kb:	Kilo base
Lif:	Leukemia inhibitory factor
LINE:	Long interspersed nucleotide element
LTR:	Long terminal repeat
Mb:	Mega base
MB:	Myoblasts
Mbd:	Methyl-binding domain protein

Me:	Methylated
MeDip:	Methylated DNA immunoprecipitation
MEF2:	Myocyte enhancer 2
MEFs:	Mouse embryonic fibroblast
MF:	Myofiber
MRF:	Myogenic regulatory factor
MT:	Myotube
NC:	Notochord
Nfix:	Nuclear Factor I/X
NT:	Neural tube
PcG:	Polycomb group
PCR:	Polymerase chain reaction
PdgfaR:	Platelet-derived growth factor receptor alpha
PGC:	Primordial germ cells
PRC1/2:	Polycomb repressor complex 1 and 2
RMS:	Rhabdomyosarcoma
RNA:	Ribonucleic acid
SAM:	S-adenosylmethionine
SB region	Sodium bisulphite region
Seq:	Sequencing
Shh:	Sonic hedgehog
shRNA:	Small hairpin RNA
SINE:	Short interspersed nucleotide element
siRNA:	Small interfering RNA
Smug1	Single-strand-selective monofunctional uracil-DNA glycosylase 1
Tdg:	Thymine-DNA glycosylase
Tet:	Ten eleven translocation
TF:	Transcription factor
TRABs:	Transcription balancing sequences
Trx:	Trithorax group
TSS:	Transcription start site
Usf1:	Upstream stimulatory factor 1

INTRODUCTION

1 CELL DIFFERENTIATION: FROM PLURIPOTENCY TO LINEAGE SPECIFICATION

Cell differentiation entails lineage choices leading to a more mature cell type by the activation and preservation of the gene expression program characteristic of each cell type. *In vivo,* stem cells trigger this dynamic process during two main stages: embryonic development and adult tissue maintenance (**Figure I1**).

During development, after the fusion of sperm and egg gametes, the diploid zygote initiates a serie of cell divisions that result in a multicellular embryo. *In utero*, the blastocyst implants and all three embryonic germ layers are formed, which ultimately will form the organs of the adult organism. Importantly, cell differentiation allows the pluripotent embryo to give rise to all somatic cell types of a complex organism. In the adult stage, cell differentiation occurs when the multipotent adult stem cell (somatic stem cells) replenishes the functional cells of the tissue during homeostasis maintenance and tissue regeneration (Loebel *et al.,* 2003).

Embryonic and adult stem cells are able to self-renew and have a high proliferative rate, which allows the expansion of the population prior to cell differentiation. Diversification, proliferation and self-renewal potentials get lost as the differentiation progresses, narrowing the cell fate possibilities while acquiring specific functional characteristics. *In vitro* embryonic stem cells (ESCs) are derived from the inner cell mass of the blastocyst and are able to self-renew in culture in the presence of leukemia inhibitory factor (Lif) (Evans and Kaufman, 1981). Upon induction, ESCs can differentiate into several lineages of different germ layers as hematopoietic cells (Wiles and Keller, 1991; Kennedy *et al.,* 1997; Müller *et al.,* 2000), cardiac muscle (Klug *et al.,* 1996), neuronal cells (Mohn and Schübeler, 2009) and endothelial cells (Vittet *et al.,* 1996).

Although the differentiation process in mammals has been traditionally considered unidirectional (stem cell -> progenitor cell -> differentiated cell), recent studies have demonstrated that differentiated cells can be converted into progenitor cells through

a dedifferentiation process giving rise to induced pluripotent cells (iPS) and switch to a different lineage program (reprogramming) to regenerate tissue lineages (reviewed in Eguizabal *et al.,* 2013).

At the regulatory level, cell differentiation comprehends changes in the cellular identity directed by transcription factor (TF) networks, which establish gene expression patterns in response to developmental cues. Once established, core TF networks maintain robust lineage-restriction and ensure unidirectional development towards defined differentiated cell-types. In parallel, the chromatin state (how genomic DNA is packaged along with histones) reinforces cell-fate decisions and establishes barriers against reversion to previous cellular states (Reik *et al.,*2007).

Deregulation of cell differentiation leads to severe impaired regeneration potential, aging and developmental pathologies. Moreover, tumoral processes are characterized by the loss of a differentiated signature in favor of a more undifferentiated and proliferative phenotype. A better understanding of the regulatory pathways orchestrating cell differentiation processes might allow to modulate stem cell fate and ultimately provide new therapeutic approaches to enhance tissue regeneration.

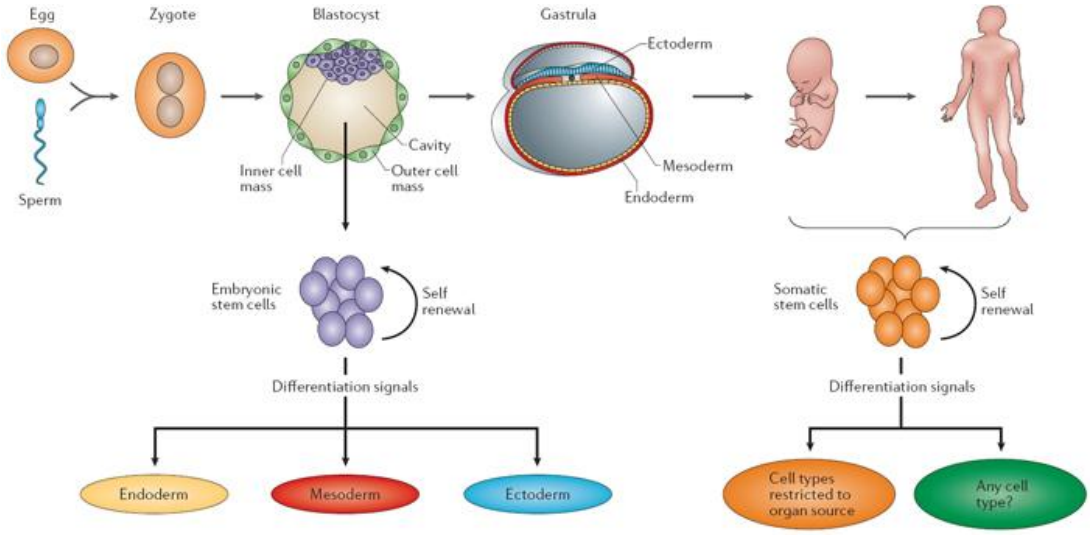


Figure I1. Embryonic and somatic stem cells differentiation process. Embryonic differentiation can be mimicked *in vitro* giving rise to the three germ layers, which *in vivo* will ultimately form the complete organism. Somatic stem cells are present in fetal and adult tissues and maintain the tissue homeostasis while they can be differentiated *in vitro* too. Reproduced from O'Connor and Crystal, 2006.

2 SKELETAL MYOGENESIS

The human body has 650 skeletal muscles representing the 40 % of the body weight. Skeletal muscle is one of the most dynamic and plastic tissues of the human body contributing significantly to multiple body functions. From a mechanical point of view, the main function of skeletal muscle is to convert chemical energy into mechanical energy to generate force, maintain the posture and produce movement. From a metabolic perspective, the role of skeletal muscle includes the contribution to basal energy metabolism, serving as storage for important substrates such as amino acids and carbohydrates, the production of heat for the maintenance of body temperature, and the consumption of the majority of oxygen and fuel used during physical activity and exercise. Morphologically, the adult skeletal muscle is mainly composed of terminally differentiated, postmitotic, contractile and plurinucleated myofibers, wrapped by the sarcolema membrane, and a small pool of muscle stem cells, known as satellite cells (SCs), located beneath the basal lamina of each myofiber (Mauro *et al.,* 1961) (**Figure I2.A-B**).

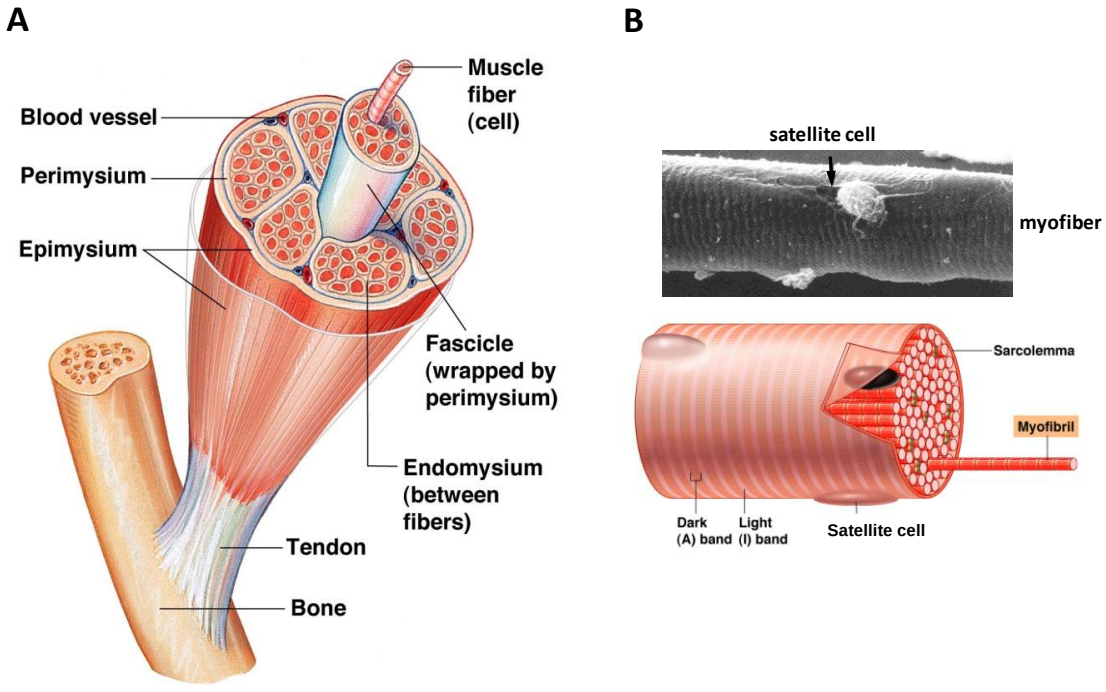


Figure I2. Skeletal muscle morphology. A. Scheme of the skeletal muscle tissue structure. **B**. Detailed morphology of a skeletal muscle fiber and the satellite cells. Illustrations were obtained from Pearson Education.

Skeletal myogenesis comprehends the process of building muscle, which occurs early during development (embryonic and fetal stages), and postnatally to allow muscle growth, maintenance of tissue homeostasis and muscle regeneration after injury. Skeletal myogenesis is a dynamic process in which mononucleated and undifferentiated myoblasts first proliferate and, upon cell cycle withdrawal, differentiate and fuse each other to form multinucleated myotubes, which ultimately give rise to mature myofibers. The myogenic process involves the transformation of progenitor cells to terminally differentiated cells, which have acquired new and more specialized functions. Muscle formation or myogenesis is orchestrated by the same transcription factors in similar regulatory network throughout life (see section 2.3).

2.1 Developmental myogenesis

During embryogenesis skeletal muscles arise from the dermomyotome in the somites, except for some head muscles. Migrating myogenic cells leaving the dermomyotome will form the limb muscles (**Figure I3**). Starting from the very beginning, the pre-patterned embryo subsequently develops the ectoderm, mesoderm, and endoderm. The mesoderm is anatomically separated into paraxial, intermediate, and lateral mesoderm, with respect to the position from the midline. In the course of development, local oscillations in gene expression and morphogen gradients induce pairwise condensations of paraxial mesoderm into somites, which develop progressively on either side of the axis, following an anterior-to-posterior developmental gradient from head to tail (Tajbakhsh and Buckingham, 2000). Spatiotemporal somitogenesis involves expression of genes related directly or indirectly to the Notch and Wnt pathways, as well as morphogen gradients of Fgf, Wnt and retinoic acid (Brent and Tabin, 2002).

Considering the dorso-ventral axis (in relation to the notochord and defined by their source of innervations) the parts of the somite are named epaxial or hypaxial if they are dorsal and ventral, respectively (Ordahl and Le Douarin, 1992, Huang and Christ, 2000). Dorso-ventral patterning of the somite results from a balance between

gradients of dorsalizing and ventralizing signals corresponding to Sonic hedgehog (Shh), Wnt and Bone morphogenic proteins (BMPs) (Cossu and Borello, 1999).

At the end, in the embryo, the sclerotome is the source of the axial skeleton, the myotome contains skeletal muscle precursors, and the dermomyotome gives rise to dorsal dermis and all the skeletal muscles of the trunk and limbs, as well as endothelial and smooth muscle cells of blood vessels and brown fat (Brent and Tabin, 2002).

In mice, the embryonic myogenesis forms the primary muscle fibers and takes place between embryonic day (E) 8.5 and E13.5. It is followed by a secondary myogenesis during fetal development giving rise to the bulk of skeletal muscle fibers present at birth (Kelly and Zacks, 1969). These successive waves of myogenesis are undertaken by embryonic and fetal myoblasts respectively, sharing a common myogenic regulatory network but also having distinct growth factor responses and different proliferative and differentiation capacities (Cossu and Molinaro, 1987;Harris *et al.,* 1989; Condon *et al.,* 1990; Barbieri *et al.,* 1990;Pin and Konieczny, 2002).

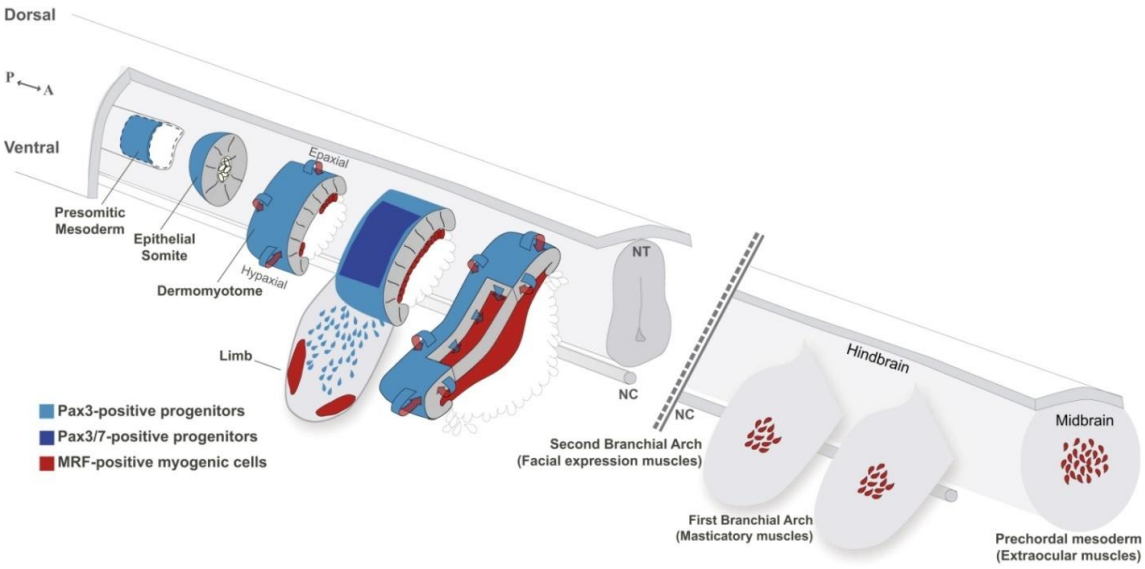


Figure I3. Skeletal muscle arises from dermomyotome in the somites. Schematic representation of skeletal muscles source: somites for the body muscles and first and second branchial arches and prechordal mesoderm for head muscles. Somites mature following an anterior (A) to posterior (P) developmental gradient. NT, neural tube; NC, notochord. Reproduced from (Buckingham and Rigby, 2014).

2.2 Adult myogenesis

Toward the end of fetal muscle development, progenitor cells begin to be enveloped under a basal lamina that forms around the muscle fibers (Relaix *et al.,* 2005). This is the characteristic location of the myogenic stem cells and therefore were named satellite cells (Mauro *et al.,* 1961). This population is in charge of postnatal myogenesis, involved in the postnatal growth, tissue homeostasis and tissue regeneration. Remarkably, adult skeletal muscle has a great regenerative capacity even following multiple rounds of injury due to the satellite cells properties, which maintain a constant pool size (Montarras *et al.,* 2013). Without external stimuli, satellite cells are in a mitotically inactivate state (quiescent). In the adult muscle, their abundance is very low representing the 2-7% of all the nuclei of the muscle tissue, but upon activation they get amplified and give rise to proliferating myoblasts, which then fuse together to form new myofibers and/or to fuse to the existing muscle fibers during muscle regeneration (Bischoff *et al.*,1975; Lipton and Schultz, 1979). Satellite cells can be defined histologically and phenotypically, by the expression of surface markers including M-cadherin, c-Met, CD34, syndecan-4 and, more importantly, the paired box homeodomain-containing transcription factor Pax7 (Oustanina *et al.,* 2004).

Many environmental cues can activate satellite cells, such as adhesion molecules, growth factors and cytokines released by the neighboring cells—a local milieu composed of fibroblasts, interstitial cells, resident macrophages and microvasculature-related cells, namely endothelial cells, pericytes and mesoangioblasts (Gopinath and Rando, 2008). Extracellular cues are transmitted to the muscle cell nucleus through signaling cascades, among which the p38 MAPK and the IGF1-AKT pathways are through to play a major role during terminal differentiation (Lluís *et al.,* 2005; Schiaffino and Mammucari 2011).

Muscle regeneration relay on the satellite cell population, however, some situations as muscular dystrophies, cachexia and sarcopenia exhaust this population leading to

muscular wasting. To improve the regeneration of the skeletal muscle pioneer studies tried to transplant adult myoblasts in injured tissue without success (Mendell *et al.,* 1995; Partridge *et al.,* 1998; Tremblay *et al.,* 1993). As an alternative, great interest has been growing around stem cells based therapies in the muscle regeneration research field, being the inducible ESC and iPS cell technologies a new attractive sources of myoblasts (Darabi, Santos, *et al.,* 2011).

2.3 Myogenic regulatory network

The establishment of embryonic progenitors in the myotome, a well as the activation of the satellite cells giving rise to myoblasts, involves the upregulation of the basic helix-loop-helix myogenic regulatory factors or MRFs: Myf5, MyoD, Myf6 and Myogenin (reviewed in Moncaut *et al*., 2013) (**Figure I4.A-B**). In addition, *Pax3* and *Pax7,* lie upstream of MRFs and their expression characterizes the skeletal muscle stem cell (reviewed in Lagha *et al.,* 2008). In the following sections, I will review the upstream and downstream regulation of the main myogenic regulatory genes, focusing on governing regulatory elements.

2.3.1 *Pax* genes

Pax3 and Pax7 are important upstream regulators during developmental and adult myogenesis (Buckingham and Relaix, 2007). They belong to the highly conserved *Pax* gene family (Horst *et al.,* 2006), which play key roles during tissue specification and organ development and are characterized by the presence of a paired domain that confers sequence-specific binding to DNA. Unlike the MRFs, *Pax3* and *Pax7* are not tissue specific, being also expressed in neurectoderm, in subdomains of the brain and in the neural tube. Spontaneous mutations in *Pax* genes in mice and humans give rise to developmental defects, while haploinsufficiency can also lead to dysmorphic phenotypes. Mutations in *Pax3* can lead to Waardenburg syndrome and

rhabdomyosarcomas, whereas mutations in *Pax7* can develop melanomas, neuroblastomas, as well as rhabdomyosarcomas. *Pax3* mutant mice lack limb muscles and show cell death at later stages in the hypaxial domain of the somite (Goulding *et al.,* 1994), whereas *Pax7* mutants mice are compensated by *Pax3* and develop normally (Relaix *et al.,* 2006).

In the embryo, *Pax3* is transcribed in the presomitic paraxial mesoderm just prior to segmentation and then is transcribed throughout the epithelial somite before becoming restricted to the dermomyotome (**Figure I3**). Pax3 is required for the formation of hypaxial muscles of the trunk and for the delamination and migration of myogenic progenitor cells to other myogenic sites, such as the limbs (Tajbakhsh and Buckingham, 2000). The migrating *Pax3* expressing cells became *Myf5* and *MyoD* expressing cells and lead to further myogenesis, while they give rise to the satellite cell pool too (Schienda *et al.,* 2006). Specifically, in the hypaxial somite, Pax3 directly controls *Myf5* expression through a regulatory sequence named -57.5 kb enhancer (Bajard *et al.,* 2006). In the epaxial somite, however, Pax3 does not directly target *Myf5,* but can indirectly affect it through the Dmrt2 transcription factor (Sato, Rocancourt, *et al.,* 2010). In *Myf5/Myf6/Pax3* triple mutants, *MyoD* expression is compromised and skeletal muscle does not form in the trunk and limbs, indicating that *Pax3*, as well as *Myf5/Myf6*, lies genetically upstream of *MyoD* (Kassar-Duchossoy *et al*., 2004, Tajbakhsh *et al.,* 1997).

In the fetal muscle, *Pax3* transcription is downregulated while *Pax7* becomes the dominant factor in all myogenic progenitor cells. In the limb, *Pax7* is initially coexpressed with *Pax3,* and genetic tracing experiments show that all later *Pax7* expressing cells in the fetal limbs are derived from cells that had expressed *Pax3* (Hutcheson and Kardon, 2009). Activation of fetal-specific muscle genes depends on the Nfix transcription factor, where *Nfix* is a potential Pax7 target (Messina *et al.,* 2010). Postnatal and adult satellite cells are marked by *Pax7* expression, with continuing transcription of *Pax3* in trunk muscles such as the diaphragm and some limb muscles (Montarras *et al.,* 2013). Before birth, Pax7 is not essential for myogenesis, presumably because Pax3 can compensate their functions. After birth, on

the other hand, *Pax7* mutants lose their satellite cells and Pax3 cannot compensate even in trunk muscles such as the diaphragm, perhaps because the protein is present at too low a level or because of divergent Pax3 and Pax7 functions by this stage (Soleimani *et al.,* 2012). *Pax7* is upstream regulated by Notch singling cascade through the direct binding of Notch1 to the Rbpj binding sites located in a distal regulatory region of *Pax7* (Wen *et al.,* 2012).

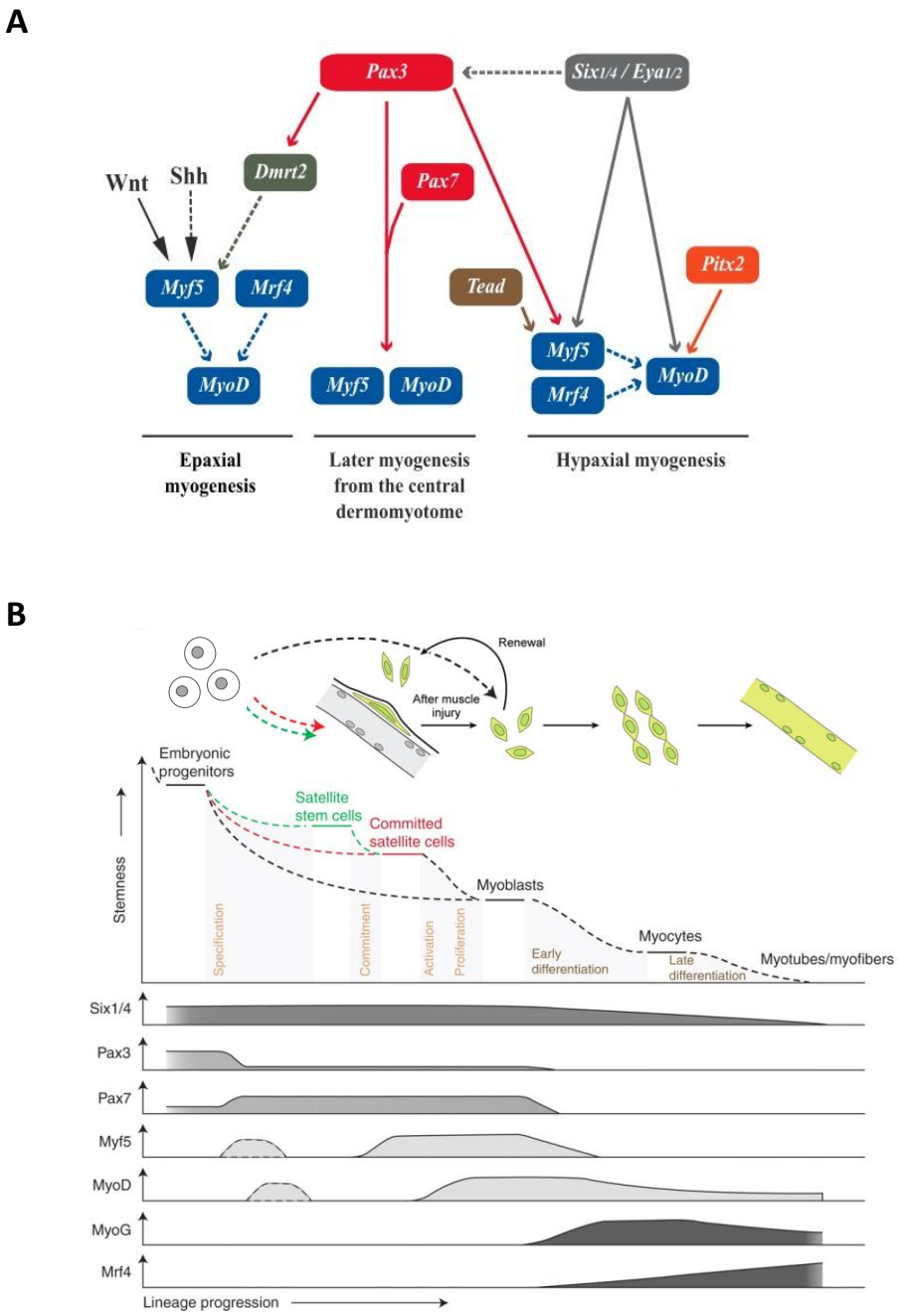


Figure I4. Gene regulatory networks that govern myogenesis. A. Scheme of gene regulatory network involved in embryonic myogenesis. Reproduced from (Buckingham and Rigby, 2014).**B**. Scheme of the expression pattern of the principal myogenic transcription factor during postnatal myogenesis. Adapted from (Bentzinger *et al.,* 2012) and (Buckingham and Relaix, 2007).

2.3.2 Myogenic regulatory factors

In the different scenarios, the myogenic process is defined by the expression of the same highly conserved basic helix–loop–helix transcription factors Myf5, MyoD, Myf6/Mrf4 and Myogenin (**Figure I4.A-B**). MyoD, Myf5 and Myf6 function as myogenic determination factors; in the absence of all three, no skeletal muscle forms. The fourth member, Myogenin, acts as a differentiation factor and, together with Myf6 and MyoD, controls the differentiation of myoblasts into mature muscle fibers (Moncaut *et al.,* 2012). All MRFs have a great myogenic potential, being able to induce myoblast traits in non-muscle cell lines (Davis *et al.,* 1987; Tapscott *et al.,* 1988; Braun *et al.,* 1989; Rhodes and Konieczny, 1989; Braun *et al.,* 1990; Miner and Wold 1990).

The basic domain of the MRFs mediates DNA binding, whereas the helix–loop–helix motif is required for heterodimerization with ubiquitously-expressed E proteins. The MRFs in cooperation to E proteins (the E2A gene products, E12 and E47, and HEB) and myocyte enhancer factor 2 (MEF2) transcription regulators, bind to E-boxes (CANNTG) and MEF2-boxes *(C/T* TA(A/T)$_4$TA*G/A***)** respectively, located in muscle promoters. (Murre *et al.,* 1989; Pollock and Treisman, 1991; Andrés *et al.,* 1995; Lluís *et al.,* 2006; Molkentin and Olson, 1996). Thereby, once the cascade of MRFs has been activated, the MRFs regulate their own transcription and induce the transcription of many myogenic genes, including those encoding the contractile proteins such as the Myosines.

Myf5* and *Myf6

Myf5 is the first MRF expressed during mammalian skeletal development. Myf5 is essential for the determination and differentiation of skeletal muscle during embryogenesis (Francetic and Li 2011; Pownall *et al.,*2002), and for the tissue homeostasis and regeneration potential of the adult muscle (Günther *et al.,* 2013). *Myf6* is located in the same locus than *Myf5* and plays an early role in myogenic determination (Kassar-Duchossoy *et al.,* 2004), but is also subsequently activated in differentiating myotube unlike *Myf5*. *Myf5* and *Myf6* genes are separated by 7kb and

their spatiotemporal control during development is very complex, with well defined regulatory sequences in intra- and intergenic regions distributed over more than 110 kb upstream of the gene (**Figure I5**) (Carvajal *et al.,* 2001; Hadchouel *et al.,* 2003; Chang *et al.,* 2004; Zammit *et al.,* 2004; Carvajal *et al.,* 2008; Ribas *et al.,* 2011).

During development, the first known expression time point of *Myf5* is at embryonic day 8.5 in the epaxial dermomyotome. This early activation depends on the canonical Wnt signaling and Shh, targeting the early epaxial enhancer (EEE), located at -5,5 kb from the transcription start site (TSS), through Lef/Tcf and Gli binding sites, respectively (Borello *et al.,* 2006; Gustafsson *et al.,* 2002; Teboul *et al.,*2003). Pax3, through the Pax3/Dmrt2 cascade, also regulates the epaxial enhancer by direct binding of Dmrt2 transcription factor (Sato, Rocancourt, *et al.,* 2010).

Another important regulatory region is located at -57.5 kb and contains a 145 bp long enhancer required for the *Myf5* expression in the hypaxial somite from E10.5 and in the developing limbs and hypoglossal chord. This small enhancer is regulated by Six1/4 proteins (Giordani *et al.,* 2007) together with Pax3 (Bajard *et al.,* 2006). A similar mechanism dependent on Six1/4 and Pax3 activation regulates the -111 kb enhancer in the hypaxial somite/myotome (Bajard *et al.,* 2006). In adult satellite cells, is Pax7 who regulates -57.5 and 111 kb enhancers by direct binding (Ribas *et al.,* 2011, Soleimani *et al.,* 2012). Another hypaxial enhancer is located within *Myf5* gene and is sufficient to lead the expression of *Myf5* in the ventral part of the somites (Summerbell *et al.,* 2000). Part of the epaxial myotome is targeted by another enhancer element located at -17 kb from *Myf5* TSS where Upstream stimulatory factor 1 (Usf-1) binds through the E-Box motif (Chang *et al.,* 2004).

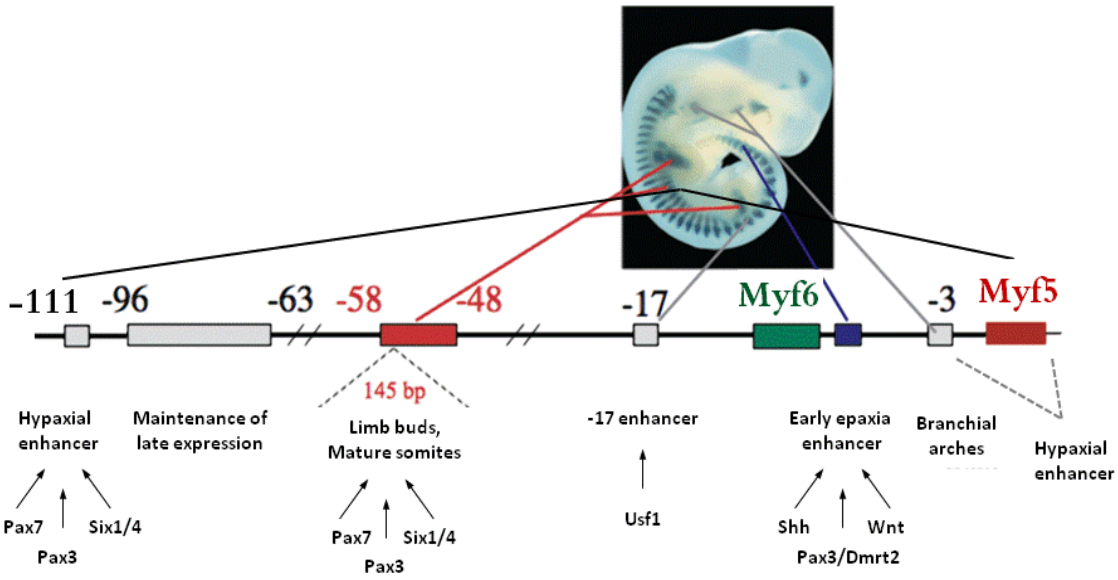


Figure I5. *Myf5/Myf6* regulatory regions. Distinct regulatory elements distributed along the locus control the spatiotemporal expression of *Myf5* and *Myf6*. Adapted from (Lagha *et al*. 2008).

Regarding *Myf6* regulation, it shares some of the mentioned *Myf5* enhancers, however, the number of regulatory regions is more limited**.** The equilibrium between the enhancers and the promoters of *Myf5* and *Myf6* is maintained due to the transcription balancing sequences, one of those is located in the proximal branchial arch element (PAE) (Carvajal *et al.,* 2008). Very interestingly the *Myf5/Myf6* locus was recently identified as a super-enhancer region in myoblast, myotubes and skeletal muscle samples (Hnisz *et al.,* 2013).

In summary, the complex regulation of *Myf5/Myf6* locus allows the expression of *Myf5* and *Myf6* genes at different times and distinct anatomical muscle locations, integrating into a single muscle cell fate the diverse myogenic inductive signals.

MyoD

MyoD is transcribed after the onset of *Myf5* expression in the hypaxial and epaxial dermomyotome. Initial activation of *MyoD* depends genetically on Myf5, Six proteins and also on Pax3 because in *Myf5/Myf6/Pax3* mutants, no skeletal muscle forms in the

trunk and limbs (Tajbakhsh *et al.,* 1997). Later on, Myo*D* can activate itself through a feedback mechanism.

Two enhancer elements have been identified for *MyoD*: the proximal enhancer, necessary for early myotomal expression, and the distal enhancer, which acts slightly later in the myotome, and also during the activation of adult muscle satellite cells in muscle regeneration (Asakura *et al.,* 1995). The distal enhancer is located at -20 kb of the TSS (Tapscott *et al.,*2005; Goldhamer *et al.,* 1995) and contain two MEF3 binding sites (5'-GACAGTAATTTTATCCTGCT-3', 5'-GGTCTTCTCCGGTTTC-3') bound by Six proteins (Relaix *et al.,* 2013; Goldhamer *et al.,* 1995) and four E-boxes without embedded CpGs. Mutation of the MEF3 sites results in almost a complete abolition of *MyoD* expression (Relaix *et al.,* 2013), while mutations on the E-boxes does not affect *MyoD* embryonic expression (Tapscott *et al.,* 2005; Goldhamer *et al.,* 1995). Moreover, Pax3 and Pax7 have been shown to bind the distal enhancer in myogenic cell cultures but there are no data on such a role of Pax3 and Pax7 in the embryo (Cao, Yu, *et al.,* 2010). Pax7 also activates *MyoD* by direct targeting its promoter (Hu *et al.,* 1992).
MyoD is expressed in proliferating myoblast, including activated satellite cells, and despite it is important to initiate myogenic differentiation, it gets downregulated throughout terminal differentiation. Importantly, *MyoD* is the first discovered reprogramming gene since it was demonstrated that *MyoD* expression in primary fibroblasts, neurons, fat tissue or liver, as well as in melanoma and neuroblastoma cell lines, converted these cells into skeletal muscle cells (Weintraub *et al.,* 1989).

Recent genome-wide analyses of MyoD binding in C2C12 myoblast cell lines and primary myoblasts have revealed that MyoD not only binds to promoters of genes regulated during muscle differentiation, but also to a large number of sites that are not associated with regional gene transcription, in both myoblasts and myotubes (Soleimani *et al.,* 2012; Cao *et al.,* 2006). Although the function of these additional binding events is not clear yet, the genome-wide MyoD binding is associated with histone acetylation, suggesting that MyoD could play a role in modulating the epigenetic landscape by recruiting histone acetyltransferases (HATs) to many regions throughout the genome.

Myogenin

Myogenin is regulated by MyoD and Myf5 and it is crucial for the differentiation of muscle cells since disruption of *Myogenin* in mice prevents muscle differentiation *in vivo* (Hasty *et al.,* 1993; Nabeshima *et al.,* 1993). During differentiation, Myogenin is recruited together with MyoD to activate the transcription of terminal differentiation genes (Cao *et al.,* 2006).

Myogenin presents a 1.5 kb regulatory region 5' upstream of the TSS, including a core region located at -184 to +18 (in relation to the TSS) able to recapitulate *Myogenin* expression (Yee and Rigby, 1993). This region contains two canonical E-box sequence recognized by the muscle regulatory factors (5'-CACATG-3', 5'-CAGTTG-3'), two overlapping non-canonical E-boxes (5'-CAACAG-3' and 5'-CAGCTT-3') (Heidt *et al.,* 2007), a Pbx/Meis site (5'-TTGATGTGCAG-3') (Berkes *et al.,* 2004), a MEF2 site (5'-CTATATTTAT-3') bound by MEF2 and a MEF3 binding site (5'-TCAGGTT-3') bound by the SIX proteins (Spitz *et al.,* 1998).

2.4 Terminal differentiation myogenic genes

Progression through the myogenic differentiation lead to the activation of genes coding for structural and enzymatic muscle proteins such as α-actins, troponins, tropomyosins, myosin heavy chains (MHC) or muscle creatine kinase (Ckm). These late differentiation genes are initially regulated by MyoD, which leads to an open chromatin state by recruiting histone acetyltransferases, and ultimately are activated by Myogenin (Cao *et al.,* 2006; Blais *et al.,* 2005; Jaynes *et al.,* 1988).

Ckm plays a central role in energy transduction in tissues with large fluctuating energy demands, such as skeletal muscle, heart, brain and spermatozoa. Its proximal promoter (-358 to +1) contains at least four active transcription factor binding sites: p53 (located between -177 and -81 to the TSS 5'-AGACATGCCT-AGACATGCCT-3', Jackson *et al.,* 1995), E-box (5'-CAGCTG-3'), CArG (bound by SRF (CC[A/T]TATA[A/T]GG)), and MPEX that recruits both Myc-associated zinc finger

protein (MAZ) and Krupple-like factor 3 (KLF3) (5'-CCCCTCCCTGGGG-'3), (Himeda *et al.,* 2010; Vincent *et al.,* 1993; Himeda *et al.,* 2011).

Ckm enhancer itself has the ability to drive high-level transcription of reporter genes in skeletal and cardiac muscle in both transgenic mice and cell culture, and to function with heterologous promoters (Jaynes *et al.,* 1986). It spans 206 bp, is located between 1,031 to 1,190 base pairs upstream of the TSS, and contians seven control elements: CArG (5'-CCCATGTAAGG-3', SRF binding motif), activator protein 2 (AP-2, 5'-GGCCTGGGG-3'), Six4/5 (MEF3/Trex motif 5'-CACCCGA-3'), AT-rich (TTATAATTA), left and right E-boxes (CATGTG and CACCTG) and myocyte enhancer factor 2 (MEF2, 5'-CTAAAAATAA-3') (Jaynes *et al.,* 1988; Donoviel *et al.,* 1996; Amacher *et al.,* 1993; Nguyen *et al.,* 2003, Himeda *et al.,* 2004; Shield *et al.,* 1996; Mueller and Wold, 1989). The combination of both, the 5'-enhancer and the proximal promoter, exhibits significant synergy in cell culture and transgenic mice, despite the promoter can function independently of the distal enhancer (Amacher *et al.,* 1993).

Myosins constitute a large superfamily of actin-dependent molecular motors. Phylogenetic analysis currently places myosins into 15 classes. Myosins class II are motor proteins able to convert energy from ATP hydrolysis into mechanical force upon interaction with actin filaments, leading to muscle contraction (Schiaffino and Reggiani, 2011). As downstream regulated genes of the myogenic regulatory pathway, Myosins contain the characteristic MEF2-box and E-box elements in their promoter region (Wheeler *et al.,* 1999).

In summary, the precise and temporal order of gene expression during development is critical to ensure proper lineage commitment, cell fate determination, and ultimately, organogenesis. Embryonic stem cells and tissue stem cells (e.g. satellite cells) require the maintenance of defined gene expression programs which must be stably adjusted upon differentiation. The most important mediators that modulate gene programs are specific transcription factors. However, over the last years, a large number of evidences have indicated that chromatin-based regulatory mechanisms play also an important role in the establishment and maintenance of transcriptional programs.

3 EPIGENETIC REGULATION

The chromatin organization is shaped by epigenetic marks, which, as the name suggest, refers to chemical modifications occurring above the DNA sequence (*epi-* (ἐπί) means *above* in greek). These marks modify the DNA sequence accessibility and, since they can be transferred to the daughter cells, they constitute a heritable code other than the genomic sequence that affects the genome readout.

The chromatinized genome can be seen as a template that can be flexibly modified in response to fate-determining and lineage-specific TFs to acquire properties characteristic of a given cell type. At the same time, the epigenetic code (epigenome) is constantly influenced by signals received from the microenvironment. Such signals, depending on their nature and origin, will cause either transient changes in gene activity or the deposition of chromatin marks that will not be transmitted to the cell progeny, or persistent chromatin alterations that will provoke long-lasting effects. Overall, the epigenome of a given cell at a specific time must reflect the developmental history of the cell and more or less recently received environmental signals.

The mechanisms involved in establishing the epigenome are DNA methylation of cytosines, (Bird and Wolffe, 1999), post-translational modifications of the histone tails (Jenuwein and Allis, 2001), ATP-dependent chromatin remodeling (Côté *et al*. 1994), histone variants exchange (Ahmad and Henikoff 2002), and non coding RNA-mediated pathways (Sheardown *et al.,* 1997) (**Figure I6**).

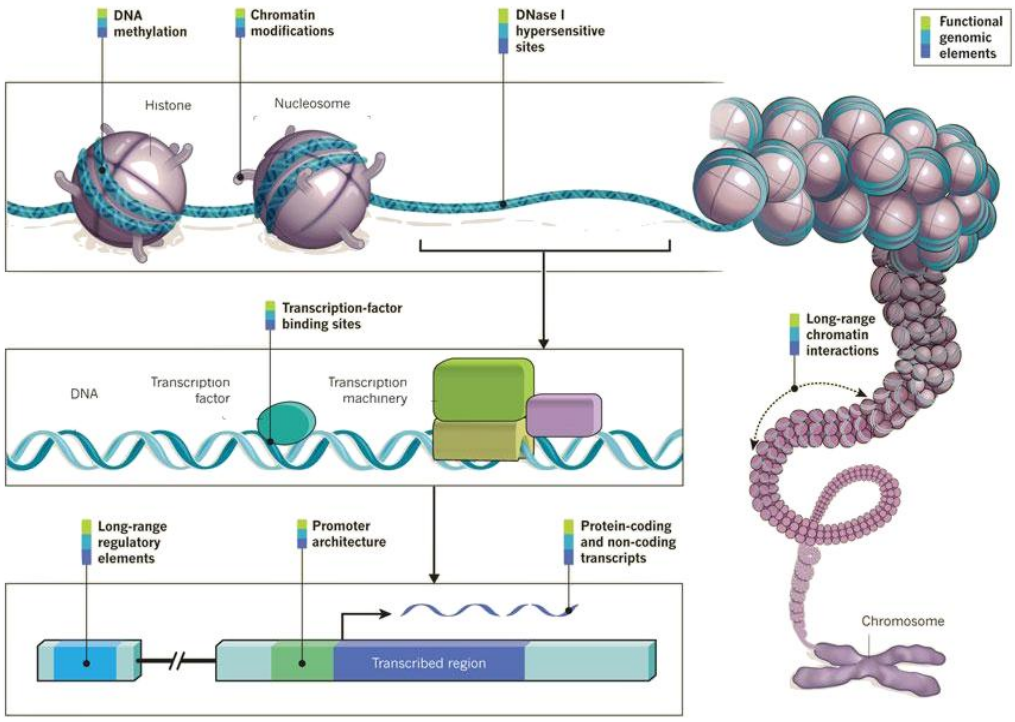


Figure I6. Epigenetic regulation. Epigenetic mechanisms are heritable chemical modifications which regulate genomic function beyond the DNA sequence itself. Reproduced from (Ecker *et*

From an evolutionary point of view, epigenetic regulation has co-evolved with increasing genome size along with organism complexity, as a necessary structuring mechanism (Lechner *et al.,* 2013). Concretely in vertebrates, sexual reproduction has result in an accumulation of repetitive and transposable elements that have enlarge the genome size during evolution (Bestor *et al.,* 2003), diminishing the proportion of protein coding sequences and regulatory RNA (Waterston *et al.,* 2002). Due to the enlarge genome size, transcriptional machinery faces the challenge of locating *cis*-regulatory regions in a cumulus of, apparently, non-functional DNA sequence. Importantly, the appearance of nucleosomes compared with naked DNA reduced the chance of aberrant transcription initiation (Struhl *et al.,* 1999; Workman and Kingston, 1998), as well as the advent of genome-wide DNA methylation enabled efficient transcriptional repression (Bird and Wolffe, 1999; Walsh *et al.,* 1998), being both mechanisms crucial to regulated gene expression. Additionally, the repressive epigenetic pathway mediated by Polycomb group proteins underwent also marked expansion over evolutionary time (Whitcomb *et al.,* 2007), enhancing specificity of transcription initiation and reducing random binding leading to inappropriate gene

regulation. At the end, for a unique underlying sequence, epigenetic mechanisms draw several epigenomes simplifying the binding recognitions of the transcriptional factors and giving rise to the diverse cell types of the organism.

Additionally, thanks to innovative technologies and to the unprecedented public collaborative consortiums efforts, important advances have been done in the last two decades regarding the epigenetic code of different cell types. In 2003 the human genome was completely sequenced and mapped by the consortium Human Genome Project after 13 years of international collaboration. It was quickly recognized that nearly 99% of the 3.3 billion nucleotides that constitute the human genome do not code for proteins (Lander *et al.,* 2001). Then, attention turned into identifying and annotating the functional elements encoded in the genome. With this purpose, The Encyclopedia of DNA Elements (ENCODE) Consortium was formed and, after a first pilot phase, in 2012 successfully ended the second phase of the ENCODE project. Indeed, the project provides important information about the organization and function of human genome from the perspective of annotation of coding and non-coding regions, chromatin accessibility, transcription factor binding, DNA methylation and interactions between parts of genome in three-dimensional space.

Excitingly, it claimed that 80.4% of the human genome displays some functionality in at least one cell type (Consortium, 2012), leaving obsolete the “junk DNA” definition given to the genome portion that does not code for proteins. Importantly, the ENCODE Project has laid the foundation stone in the exploration of our genome, but we are still far from the ultimate goal of understanding the function of the genome in every cell of each person, and across the life. However, nowadays there is no doubt that the genome functions are intimately related to epigenetic regulation, specific of every cell type and person and variable over the time.

3.1 DNA methylation

DNA methylation was the first discovered epigenetic modification (Holliday and Pugh, 1975; Riggs, 1975) and one of the best studied and most mechanistically understood. Chemically, it consists of the addition of a methyl group (-CH3) into the fifth position of cytosine nucleotide of the DNA molecule resulting in a modified base (5mC), frequently considered the fifth nucleotide. It is conserved among most plant, animal and fungal models (Feng *et al.,* 2010), showing different proportions among species (*A. thaliana* 14% of cytosines methylated, *M. musculus* 7.6%, *D. melanogaster* 0.034%, *E. coli* DH5α 2.3 %) (Capuano *et al.,* 2014).

3.1.1 Distribution of DNA methylation throughout the genome

In mammals, cytosine methylation is primarily restricted to the symmetrical CpG dinucleotide (Cytosine followed by a Guanine bound by a phosphate union) (Ramsahoye *et al.,* 2000; Ziller *et al.,* 2011), a dinucleotide globally underrepresented in the genome as a consequence of the spontaneous or enzymatic deamination of 5mC to thymine in the germ line (Bestor and Coxon, 1993).

Of note 5mC mapping experiments at base pair resolution revealed that methylation can also occur at non-CpG sites (predominantly CpA methylation) in oocytes (Tomizawa *et al.,* 2011, Haines, Rodenhiser, and Ainsworth 2001), and adult brain (Xie *et al.,* 2012), as well as in ES cells (Lister *et al.,* 2009, Stadler *et al.,* 2011). However, non-CpG is found at much lower frequencies than CpG methylation constituting only 0.02% of total 5mC in somatic cells, but is surprisingly high in hESCs and brain (25%) (Laurent *et al.,* 2010, Lister *et al.,* 2009, Guo *et al.,* 2014). Currently, it is unknown why this non-CpG methylation is more frequent in oocytes and brain, and if it plays any functional role.

Focusing on 5mC at CpG dinucleotides, it is present throughout the genome and shows a general bimodal distribution (**Figure I7**). This distribution is defined by the inverse

correlation between 5mC and CpG density: CpG-poor DNA, which comprises most of the genome (gene bodies, repetitive elements and the majority of inactive or repressed regulatory elements) shows high levels of 5mC, whereas high CpG-dense regions, termed CpG islands (Gardiner-Garden and Frommer, 1987) are largely resistant to DNA methylation. However, there are methylated regions with elevated CpG densities too, such as exons, some CpG island promoters of gamete specific genes, imprinted control regions, and a subset of endogenous retroviral elements; and demethylated regions with low CpG content too, such as active promoters, enhancers and boundary elements. Importantly, approximately 6% CpG islands become methylated in a tissue-specific manner during early development or in differentiated tissues (Straussman *et al.,* 2009).

CpG islands are defined as short regions of 200 bp, with more than 50% CG content and ratio greater than 0.6 of observed number of CG dinucleotides relative to the expected one. They are generally found at promoters of housekeeping and developmental regulatory genes (Deaton and Bird, 2011), representing 60% of the promoters, whereas tissue-specific genes contain archetypically CpG-poor promoters (Mohn and Schübeler, 2009). Interestingly, since CpG islands are depleted of DNA methylation they have been specially conserved in the course of evolution (Lister *et al.,* 2009; Meissner *et al.,* 2008).

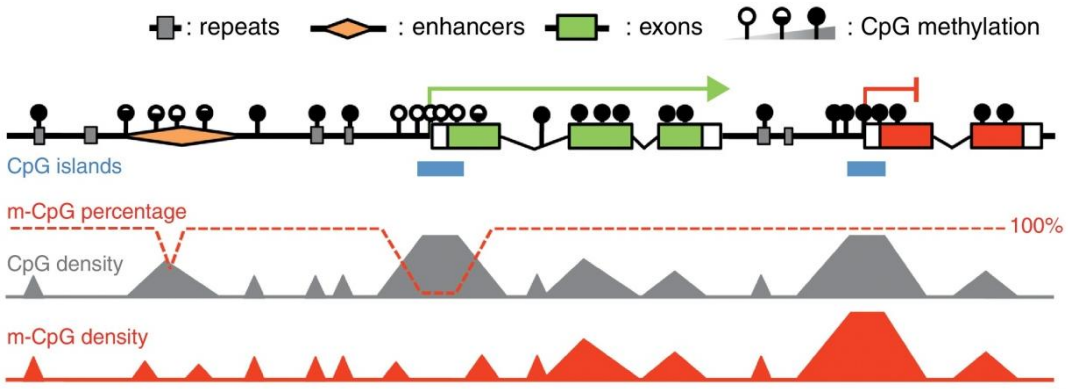


Figure I7. The mammalian CpG methylation landscape. Global bimodal distribution of 5mC, where CpG-rich regions are demethylated and CG-poor regions, representing the majority of the genome, are methylated. Punctual exceptions occurred, such as demethylated poor CG-content active enhancers, throughout composing the tissue-specific methylomes. Reproduced from (Baubec and

The surrounding regions to the CpG islands are susceptible to be methylated and are named CpG island shores. These regions are CpG poor and span around 2 kb at both sides of the islands (Irizarry *et al.,* 2009) and its DNA methylation seems also to correlate with gene expression, as it does with CpG island (Ji *et al.,* 2010). Some studies show that the methylation patterns of these regions are cell type-specific being sufficient to distinguish between tissues (Irizarry *et al.,* 2009). Interestingly, it has been reported that 70% of the differentially methylated regions in reprogrammed cells are associated with CpG island shores (Ji *et al.,* 2010).

Importantly, aberrant methylation patterns are clearly found in several cancer types. Typically, during tumoral processes CpG island located at the promoter regions of tumor suppressor genes become hypermethylated, whereas a general hypomethylation is observed in the bulk of the -genome (reviewed in Esteller, 2007). Hypermethylated promoters have been proposed as a new generation of biomarkers and hold great diagnostic and prognostic promise for clinicians (reviewed in Kelly *et al.,* 2010).

3.1.2 Regulatory implications of DNA methylation

DNA methylation is a critical epigenetic regulator in mammalian development, being crucial for X-inactivation (Heard *et al.,*1997), genetic imprinting, silencing of genomic elements, such as transposons ensuring genomic stability (Gaudet *et al.,* 2003; Walsh *et al.,*1998) and maintenance of constitutively repressed centromeric and pericentromeric DNA satellite repeats, which affects heterochromatin organization (Gopalakrishnan *et al*., 2009).

Traditionally, based in its involvement in all these functions, DNA methylation has been considered a heritable and stable silencing mark. However, thanks to improved genome-scale mapping of methylation, now it is known that the regulatory read out of

DNA methylation is context-specific and the relationship between DNA methylation and transcription is more nuanced than previously thought.

Centered on genes, 5mC in the immediate vicinity of the TSS is generally viewed as an inhibitor of transcriptional initiation, either by preventing the binding of certain transcription factors, or by recruiting methyl-binding proteins (Mbp) and histone deacetylases, and generating a repressed chromatin environment (Shen *et al.,* 2007; Hsieh, 1994; Jones and Takai, 2001, Cameron *et al.,* 1999). However, while some TF binding affinity is reduced due to DNA methylation (Iguchi-Ariga and Schaffner, 1989; Campanero *et al.,* 2000; Kim *et al.,* 2003) other TFs are not influenced by CpG methylation (Harrington *et al.,* 1988), and even a negative selection has been observed against TFBS containing "traffic lights", defined as the CpGs which is methylation state negatively correlates with gene expression (Medvedeva *et al.,* 2014).

Interestingly, recent studies have shown that not only promoters but also intragenic and intergenic regions are widely modulated during physiological processes and disease by DNA methylation. Regarding the intergenic regions, several genome-wide studies claimed that DNA methylation modulates intergenic enhancer regions during differentiation, pointing out the regulatory relevance of those regions and the need of further characterization of them (Stadler *et al.,* 2011; Schmidl *et al.,* 2009; Sérandour *et al.,* 2011). Concerning DNA methylation within gene bodies, it has been directly correlated with gene expression, playing the opposite role than in promoters. Higher levels of intragenic methylation have been associated with higher expression levels (Hellman and Chess, 2007; Laurent *et al.,* 2010), even though it is not yet clear if there is a causal relationship. Moreover, DNA methylation in the gene body has been related to alternative promoter usage (Archey *et al.,* 1999; Shmelkov *et al.,* 2004; Maunakea *et al.,* 2010; Cheong *et al.,* 2006), regulation of short non-coding RNAs such miRNAs and long non- coding RNA transcripts (Lyle *et al.,* 2000; Stöger *et al.,* 1993; Lujambio *et al.,* 2010), alternative RNA splicing (Shukla *et al.,* 2011; Anastasiadou *et al.,* 2011; Malousi *et al.,* 2008), reviewed in Brown *et al.*2012), as well as a regulatory mechanism of gene-embedded distal enhancer activity (reviewed in Kulis *et al.,* 2013).

As mentioned above, intragenic DNA methylation could modulate alternative and constitutive splicing RNA processing. Despite splicing occurs on the mRNA, chromatin structure may directly affect the co-transcriptional splicesome assembly. The splicing machinery recognizes exons due to their enrichment in nucleosomes and its high CG content. In this context, it has been observed that DNA methylation is increased in nucleosomes-bound DNA (Chodavarapu *et al.,* 2010), and since exons are enriched in nucleosomes, exons are more methylated that adjacent introns (Hellman and Chess 2007; Hodges *et al.,* 2009). Differences in the exon-intron methylation levels can impact in inclusion or exclusion of exons affecting the generation of splice variants (Shukla *et al.,* 2011; Gelfman *et al.,* 2013; Sati *et al.,* 2012).

3.1.3 Shaping the DNA methylation patterns

The DNA methylation patterns are established and maintained by the DNA methyltransferases (Dnmts) (**Figure I8**) and removed through different possible mechanism, which are still matter of debate in the DNA demethylation field.

Dnmt family comprises three active members in mammals (Dnmt1, Dnmt3a and Dnmt3b), which share a conserved catalytic domain that catalyzes the transfer of a methyl group from cofactor *S*-adenosylmethionine (SAM) to carbon 5 of the cytosine ring to generate 5mC (Christman, 2002). Dnmt1 maintains the DNA methylation patterns on the newly synthesized DNA strand during replication. It is recruited by UHRF1 to hemimethylated sites at replication forks via interaction with PCNA (Sharif *et al.,* 2007), and thanks to the Dnmt1 high affinity to hemimethylated CpGs and the 5mC symmetry it copies the DNA methylation profile to the nascent sequence (Song *et al.,* 2011; Song *et al.,* 2012).

New methylation patterns are established by *de novo* Dnmts (Dnmt3a, Dnmt3b, and cofactor Dnmt3L). However, Dnmt3a/3b are also thought to be involved in maintaining DNA methylation patterns globally by filling the gaps left out by Dnmt1 (Jackson *et al.,*

2004). Dnmt3a and Dnmt3b are highly homologous but possess distinct target specificities and expression patterns (Kato *et al.,* 2007; Watanabe *et al.,* 2002). Dnmt3b is more prevalent in early embryonic stages and is the main enzyme responsible for the acquisition of DNA methylation during implantation step (Borgel *et al.,* 2010), whereas Dnmt3a is expressed in later embryonic stages and differentiated cells. Dnmt3L physically interacts with the Dnmt3 enzymes, stimulates their *de novo* methylation activity, and promotes their recruitment to chromatin (Ooi *et al.,* 2007; Jia *et al.,* 2007).

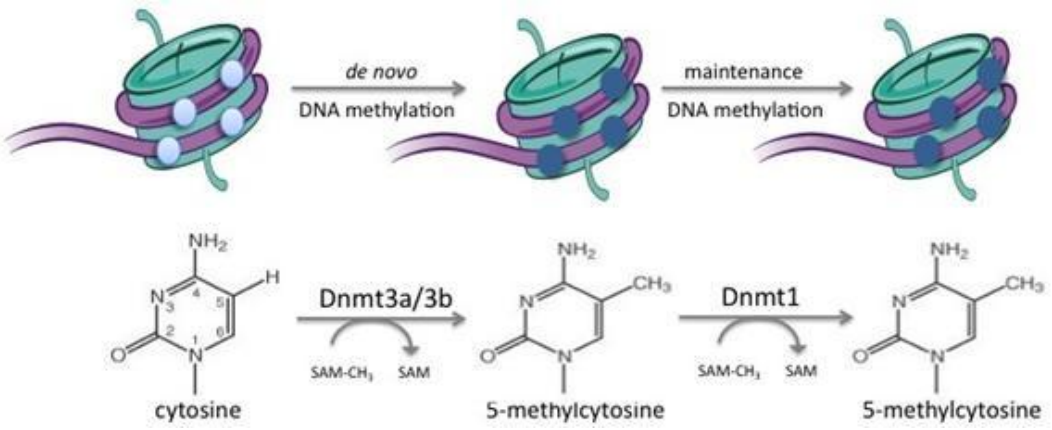


Figure I8. DNA methylation patterns are established and maintained by DNA methyltransferases. DNA methyltransferases transfer a methyl group from the cofactor *S*-adenosylmethionine to carbon 5 of the cytosine ring to generate 5mC.

DNA methylation modification can also be removed as a consequence of passive or active demethylation processes (**Figure I9**). DNA demethylation can simply be achieved through passive mechanisms, based on the lack of methylation maintenance during DNA replication, which results in a progressive dilution of DNA methylation over several rounds of DNA replication. However, loss of methylation has been observed independently of DNA replication in post-mitotic states in differentiation processes such as hematopoiesis (Wu and Zheng, 2014), hepatic lineage (Nagae *et al.,* 2011; Waterland *et al.,* 2009), osteogenic differentiation (Zhang *et al.,* 2011), neuronal differentiation (Guo, Ma, *et al.,* 2011; Ma *et al.,* 2009) and myogenesis (see section 5).
In addition, male and female pronuclear DNA in the preimplantation zygote is highly demethylated (Oswald *et al.,* 2000; Wang *et al.,* 2014; Guo, Li, *et al.,* 2014) exposing

the fact that an active mechanism might act also in this context. Notably, a constant in all these studies is that demethylation seems to mostly affect sequences with moderate CpG richness, suggesting that only methylation of CpG-poor regions is susceptible to rapid erasure in somatic cells whereas methylated CpG islands are probably demethylated only in zygotes or primordial germ cells.

The discovery of Ten eleven translocation family (Tet1, Tet2 and Tet3) and their capacity of hydroxylate 5mC into hydroxymethylcytosines (5hmC) brought back the already known 5hmC epigenetic mark (Penn *et al.,* 1972) to spotlight as a an intermediate in the cytosine demethylation reaction. Tet's can further catalyze the oxidation of 5hmC into 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (He *et al.,* 2011; Ito *et al.,* 2011), which were detectable in genomic DNA of various cell types and could be additional intermediates of demethylation process. Increased 5hmC levels at enhancer regions activated during cell differentiation suggested that Tet enzymes could be recruited to these regions to erase DNA methylation (Sérandour *et al.,* 2012). Interestingly, 5hmC signatures are enriched at sites of DNaseI hypersensitivity, which are indicative of genomic regions bound by regulatory proteins, whereas 5mC is generally depleted at sites of DNA–protein interaction (Wu *et al.,* 2011). Moreover, 5hmC is particularly enriched at bivalent domains (introduced in section 3.2.3,) near to transcription start sites, at CpG-rich proximal promoters, and at active and poised enhancers (Pastor *et al.,* 2011; Wu *et al.,* 2011; Szulwach *et al.,* 2011). Notably, both 5mC and 5hmC are also enriched within the gene bodies of actively transcribed genes (Ball *et al.,* 2009; Pastor *et al.,* 2011; Szulwach *et al.,* 2011; Wu *et al.,* 2010). This distribution suggests that 5hmC might be itself an epigenetic mark or a transitory state.

The importance of Tet enzymes was highlighted by the impaired differentiation and poor developmental level observed in the triple knock down ESCs (Dawlaty *et al.,* 2014). One possible scenario for the Tet-mediated demethylation is that the oxidized derivatives are not longer recognized by the methylation maintenance machinery and are replaced by unmodified cytosines during replication, which is supported by the lack of activity of Dnmt1 on hemi-hydroxymethylated DNA (Valinluck and Sowers,

2007). This scenario is transcription-dependent and might be involved in the massive demethylation of the preimplantation zygote (see section 4), where 5hmC, 5caC and 5fC signals on the paternal genome persist in the cleavage stage embryo and are gradually diluted during replication (Iqbal *et al.,* 2011; Inoue *et al.,* 2011; Inoue and Zhang, 2011).

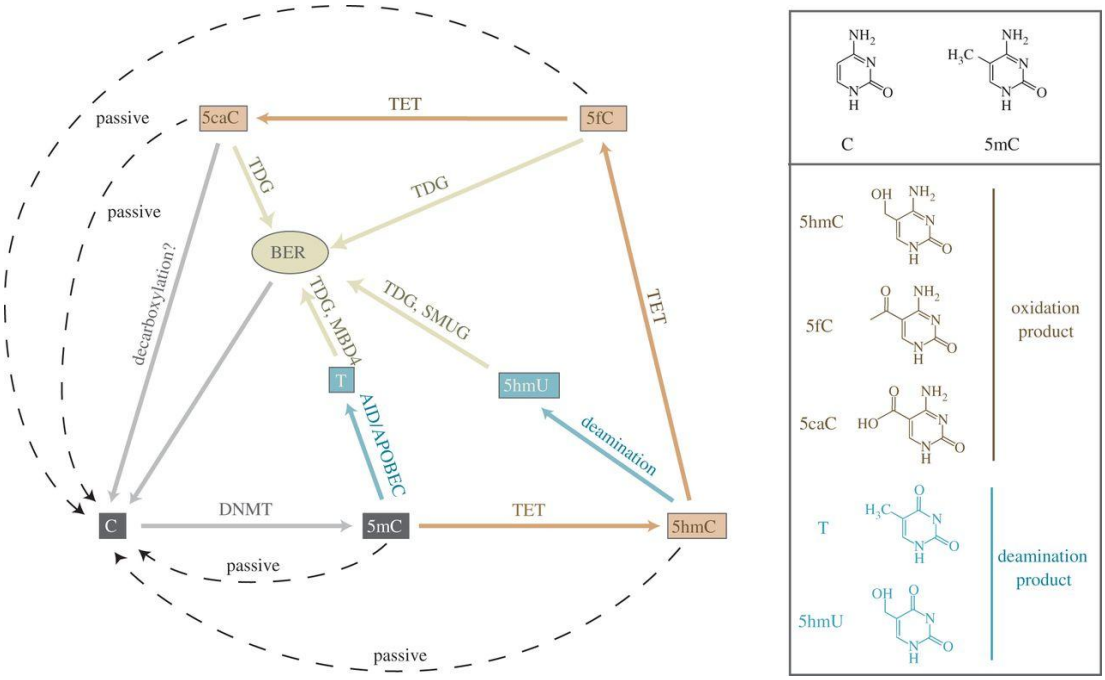


Figure I9. DNA demethylation pathways. DNA methylation can be lost passively due to a lack of maintenance after DNA replication (dashed line), or actively processed by enzymatic activity. Deamination processes are indicated in blue, oxidation in brown and glycosylation in beige. C = Cytosine, T= Thymine, 5mC = 5-methylcytosine, 5fC = 5-formylcytosine, 5 hmC = hydroxymethylcytosine and 5caC = 5-carboxylcytosine). Reproduced from (Seisenberger *et al.* 2013).

Another scenario could be that 5hmC, 5caC or 5fC are removed by DNA glycosylases such as thymine DNA glycosylase (Tdg), followed by base excision repair (BER) mechanism. Although no glycosylase has yet been identified to act on 5hmC, it has been shown that Tdg can efficiently remove 5caC and 5fC from DNA (He *et al.,* 2011; Maiti and Drohat, 2011). The important role of DNA glycosylases is supported by the observation that Tdg deficiency leads to embryonic lethality and hypermethylation of developmental genes in mice (Cortázar *et al.,* 2011; Cortellino *et al.,* 2011). The role of BER is further supported by the observations that BER components PARP1, APE1 and

XRCC1 are detected in the paternal pronucleus (Wossidlo *et al.,* 2010), and in PGCs at E11.5 (Hajkova *et al.,* 2010) at the time of demethylation. In addition, inhibition of BER activity with PARP or APE1 inhibitors results in an increase of DNA methylation on the zygotic paternal genome (Hajkova *et al.,* 2010).

A third possibility is that the interaction of AID/Apobec family of deaminases could directly deaminate 5mC, which creates T:G mismatches that can be repaired by DNA glycosylases, such as Tdg and Mbd4, followed by BER mechanism (Cortellino *et al.,* 2011). Alternatively, they could require prior oxidation of 5mC by the Tet proteins and deaminate 5hmC, thereby creating 5hmU that could be removed by the DNA glycosylases Tdg or Smug1 (single-strand-selective monofunctional uracil-DNA glycosylase 1) and BER mechanism. This latter model is supported by the fact that AID/ Apobec deaminases promote 5hmC-dependent demethylation in combination with BER enzymes in cultured cells and brain, and that Tdg and Smug1 have a glycosylase activity on 5hmU (Guo, Su, *et al.,* 2011; Cortellino *et al.,* 2011). These putative demethylating pathways, involving deaminases and glycosilases, are replication dependent, and might explain the demethylation processes where cells do not cycle.

In summary, a large variety of demethylation pathways have been proposed although a comprehensive experimental *in vivo* verification is still needed. This diversity of mechanisms might reflect the use of different pathways of demethylation depending on the cell type and the genomic target in order to ensure robust epigenetic regulation (Hackett *et al.,* 2012).

3.2 Histone code

3.2.1 Histone modifications

Histone proteins are also key players in epigenetics. The histones H2A, H2B, H3 and H4, all duplicated, form the nucleosomes around which 147 bp of genomic DNA is wrapped. The histones are globular proteins except for their N-terminal tails, which are unstructured. All histones are subjected to post-transcriptional modifications

(PTMs) in the histone tails, being phosphorylation, acetylation, sumoylation, methylation and ubiquitylation the main histone modifications occurring at different sites simultaneously and composing the histone code (**Figure I10**) (Strahl and Allis, 2000). Acetylation and methylation are the modifications best studied and are the ones addressed in this Thesis. A standard nomenclature is used to name the specific histone modifications, indicating the histone name, the modified residue followed by its location, and the type and number of modification; e.g. H3K4me3 stands for histone three lysine 4 trimethylation.

-

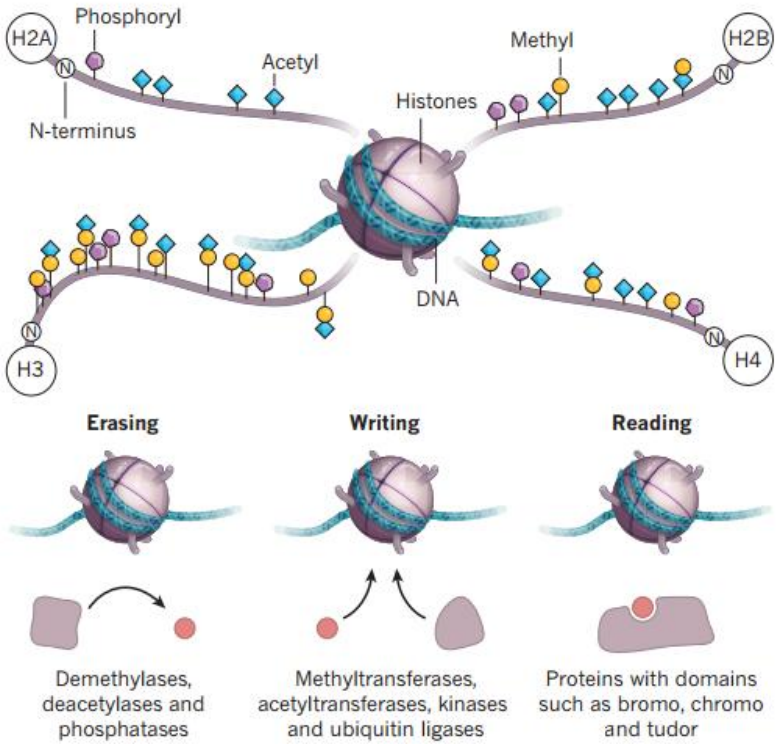


Figure I10. Histone modifications. The histone code is based on modifications of the nucleosomal histone tails which affect chromatin organization and therefore, gene transcription. The code is established and interpreted by specific enzymes (eraser and writers) and interpreted by other enzymes, which lead to changes in chromatin structure, called readers. Only the principal histone modifications are represented. Reproduced from

In relation to the transcriptional state, the genome can be divided into actively transcribed euchromatin and transcriptionally inactive heterochromatin. Acetylation and phosphorylation generally accompany active transcription; sumoylation, is usually found in heterochromatic regions, whereas methylation and ubiquitination are implicated in both activation and repression of transcription (Kouzarides, 2007).

Histone acetylation has the major potential to unfold chromatin since it neutralizes the basic charge of the lysine, making it more accessible for the transcription factors. Modified histones can recruit large protein complexes, called histone code readers, which interpret the modification and impose changes in chromatin structure (Jenuwein and Allis, 2001). Methylation is recognized by chromo-like domains of the Royal family (chromo, tudor, MBT) and non-related PHD domains, meanwhile acetylation is recognized by bromodomains and phosphorylation is recognized by a domain within 14-3-3 proteins. For instance, H3K27me3 recruits the chromodomain-containing Polycomb protein PRC1, which is associated with ubiquitin ligase activity specific for H2A, and the chromodomain-containing HP1 protein binds H3K9me3 and is associated with deacetylase activity and methyltransferase activity (**Table I1**).

Histone modification	Function	Signal	Writers	Erasers	Readers (domain)
H3K4me1	Marker of strong enhancer chromatin at distal loci. Found at transcribed regions including the promoter and end of genes	Peak/ Region	MLL, Set1, ASH1, SET7, SMYD2	LSD1/BCHC, JHDM1A/KDM2A	CHD1 (Chromo)
H3K4me2	Marker of active promoter and strong enhancer	Peak			
H3K4me3	Marker of transcription start sites and promoters, and stimulates the recruitment of the transcriptional and spliceosomal machinery. Inversely associated with and antagonistic to DNA methylation	Peak			
H3K9Ac	Marker of transcription and active chromatin at promoters and enhancers. Related to histone deposition	Peak	GCB5, PCAF, SRC1	SIRT6	BRD4 (Bromo)
H3K9me3	Marker of heterochromatin and inactive promoters	Peak/ Region	SUV39H1/2, G9a, CLL8, SETDB1, EuHMT1, Riz1	JMJD1A/KDM3A, JMJD1B/KDM3A	HP1 (Chromo)
					L3MBTL1/L2 (MBT)
H3K27me3	Marker of inactive promoter and repessed regions	Region	EZH2, EZH1	KDM6A/UTX, JMJD3/KDM6B	PRC1 (Chromo)
H3K36me3	Marker of transcriptional elongation enriched at the end of genes	Region	NSD1, SET2, SMYD2, NSD2	JHDM1B/KDM2b, JHDM3A/KDM4A	MRG15 (Chromo)
H3K27Ac	Marker of transcription and active chromatin at promoters and enhancers	Peak	p300	SMRT	TAF1 (Bromo)

Table I1. Histone marks define chromatin states. Example of some modifications, their associated function, the type of distribution designed as ChIP-seq signal, and enzymes which write, erase and read the specific modification (Consortium, 2012; Siggens and Ekwall, 2014).

3.2.2 Histone modifying enzymes

Most modifications have been found to be dynamic and can appear and disappear on chromatin within minutes after stimulus arrival at the cell surface. The enzymes involved in the removal or the addition of modifications are generally called histone code erasers and writers, respectively. These enzymes can also modify non-histone proteins.

The **acetylation** status of histones is maintained by the antagonistic action of histone acetyltransferases (HATs) and histone deacetylases (HDACs) enzymes. The HATs utilize acetyl CoA as cofactor and catalyze the transfer of an acetyl group to the ε-amino group of lysine side chains. In doing so, they neutralize the lysine's positive charge and this action has the potential to weaken the interactions between histones and DNA. There are two major classes of HATs: type-A and type-B.

Histone **deacetylation**, associated with transcriptional repression, is catalyzed by histone deacetylases which can be grouped into four subclases: Class I, Class II, Class III HDACs, also called Sirtuins (Narlikar *et al.,* 2002) and class IV, which has only a single member HDAC11. HDAC enzymes remove acetyl groups (O=C-CH3) from a ε-N-acetyl lysine amino acid on a histone, increasing the positive charge of histone tails and allowing the histones to wrap the DNA more tightly.

The histone **methylation** profile is mostly shaped by Trithorax group (TrxG) proteins, which confer transcriptionally permissive chromatin state (euchromatin) and by Polycomb group (PcG) proteins, resulting in a repressive chromatin states (heterochromatin). TrxG catalyzes the deposition of a trimethyl group on K4 of histone H3 (forming the H3K4me3 mark), and PcG is responsible for both di- and trimethylation of H3K27. Importantly, PcG and TrxG appear to be required for propagation of the repressed or activated transcriptional state during the cell-division cycle (Petruk *et al.,* 2012). Moreover, TrxG and PcG can recruit, directly or indirectly, the histone modifying proteins, such as HATs, HDACs, histone demethylases or Dnmts (Mills, 2010).

Most PcG proteins are part of the transcriptional-repressive complexes Polycomb Repressive Complexes (PRCs) (Levine *et al.,* 2002). The diversity of the PRC complexes is greater than previously anticipated, however two major complexes stand out: PRC1 and PRC2. PRC2 consists of three core PcG components: enhancer of zeste 2 (EZH2) or its close homolog EZH1, embryonic ectoderm development (EED), and suppressor of zeste 12 (SUZ12). As components of PRC2, EZH2 and EZH1 can catalyze mono-, di- and trimethylation of H3K27 (Margueron *et al.,* 2008; Shen *et al.,* 2008), and therefore are considered histone lysine methyltransferases (HKMTs). Usually HKMTs contain a so-called SET domain that harbors the enzymatic activity and catalyze the transfer of a methyl group from S-adenosylmethionine (SAM) to a lysine's ε-amino group. Once H3K27me3 is added it is recognize by the PRC1 complex, which contains member of the chromobox family (CBX) that bind to H3K27me3. Then, the RING1B (ring finger protein) catalyzes monoubiquitination of histone H2A at K112, a modification that impedes RNA polymerase II (RNAPII) elongation, resulting in transcriptional repression.

H3K9me3 is catalyzed by SUV39H1/2, which do not associate with PRC complexes. This modification is recognized by the chromodomain of HP1 and is heavily enriched at centromeric and pericentromeric DNA, where together with the DNA methylation machinery, it mediates the constitutive higher order heterochromatin structure necessary for mitotic spindle assembly. Importantly, both H3K9 and H3K27 methylation has been suggested to be prerequisite for subsequent DNA methylation, since SUV39H1/2 and EZH2 interact directly with Dnmt1, Dnmt3a, and Dnmt3b (Lehnertz *et al.,* 2003; Viré *et al.,* 2006).

For many years, histone methylation was considered a stable and static modification, but in 2004, the first lysine demethylase was identified. It was found to use FAD as co-factor, and it was termed as lysine-specific demethylase 1 (LSD1) (Shi *et al.,* 2004). The first enzyme identified as a tri-methyl lysine demethylase was JMJD2 that demethylates H3K9me3 and H3K36me3 (Whetstine *et al.,* 2006). The enzymatic activity of JMJD2 resides within a JmjC jumonji domain. Nowadays, many histone lysine demethylases are known and, except for LSD1, they all possess a catalytic jumonji domain (Whetstine *et al.,* 2006). The same as lysine methyltransferases, the

demethylases possess a high level of substrate specificity with respect to their target lysine.

3.2.3 Bivalent domains

In ESCs it was found the coexistence of two histone methylation modifications with opposite output, H3K27me3 and H3K4me3, (Azuara *et al.,* 2006;Bernstein *et al.,* 2006), in the named bivalent domains. Thanks to the ChIP on ChIP technology, it has been possible to identify broad H3K27me3 regions that also contain smaller, narrowly defined H3K4me3 regions. The majority of bivalent domains in ESCs directly overlap transcription start sites of genes that encode lineage-specific transcription factors, and approximately half of the bivalent domains contain binding sites for ≥1 of the pluripotency transcription factors (Bernstein *et al.,* 2006). Bivalent domains are enriched in CpG island and contain PRC2 complex, whereas PRC1 occupies less than half (39%) of bivalent promoters (Ku *et al.,* 2008). H3K27me3 is related to silent chromatin and H3K4me3 is involved in active chromatin. However, the combination of both marks in ESC domains results in repressed or very low expressed chromatin ready to be quickly activated, and therefore, these domains have been identified as poised chromatin regions. Upon differentiation, the bivalent domains tend to resolve preserving either the repressive mark H3K27me3, in the case of totally silenced genes, or the activating H3K4me3 modification, in the case of activated genes. Traditionally, it was thought that genes required for downstream developmental processes were actively repressed until needed. However, the discovery of bivalent domains pointed out that transcription factors that control certain differentiation processes are kept in a poised very low expression level until the precise expression time.

3.3 Other epigenetic mechanisms

Beside DNA methylation and histone modifications, two other important epigenetic mechanisms are worthy to briefly mention: the chromatin-remodeling factors and the non-coding RNA mediated pathways.

Chromatin-remodeling factors use energy derived from ATP hydrolysis to lose DNA-histone contacts and to allow the nucleosomes to move along a particular DNA sequence. One important chromatin-remodeling complex is SWI/SNF, a multisubunit complex that is highly conserved among eukaryotes. The mammalian SWI/SNF complex contain either the Brahma (BRM) or Brahma-related gene 1 (BRG1) ATPase as the catalytic subunit, which is associated with a variety of subunits called BRG1-associated factors (BAFs) (Kadam and Emerson, 2003; Simone, 2006). The SWI/SNF complex facilitates the binding and formation of the RNA polymerase II pre-initiation complex and promotes transcriptional elongation.

Non-coding RNAs (ncRNAs) can exert their regulatory functions by acting as epigenetic regulators of gene expression and chromatin remodeling. Although the detailed mechanisms are still unknown, ncRNAs might be the key event in the locus-specific recruitment of different readers, erasers and writers complexes, functioning as scaffolds to regulate epigenetic mechanisms within the cell (reviewed in Magistri *et al.,* 2012).

3.4 Regulatory elements are defined by epigenetic signatures

In the early twentieth century, microscope observations allowed to discern between different types of chromatin in the nucleus according to their staining properties: euchromatin and heterochromatin (Passarge, 1979), which later on were linked to the different transcriptional states. Classically, lighter-staining euchromatin or 'open' chromatin stands for accessible DNA, which lacks DNA methylation and present positive histone marks, as histone acetylation, allowing transcription factor binding

and subsequent transcriptional activation. Contrarily, in the densely staining heterochromatin, also called 'closed' chromatin, the genes are packed away from the transcription machinery, the DNA is typically methylated, lacks histone acetylation and is accompanied by negative histone marks such as H3K9me3 (**Figure I11**).

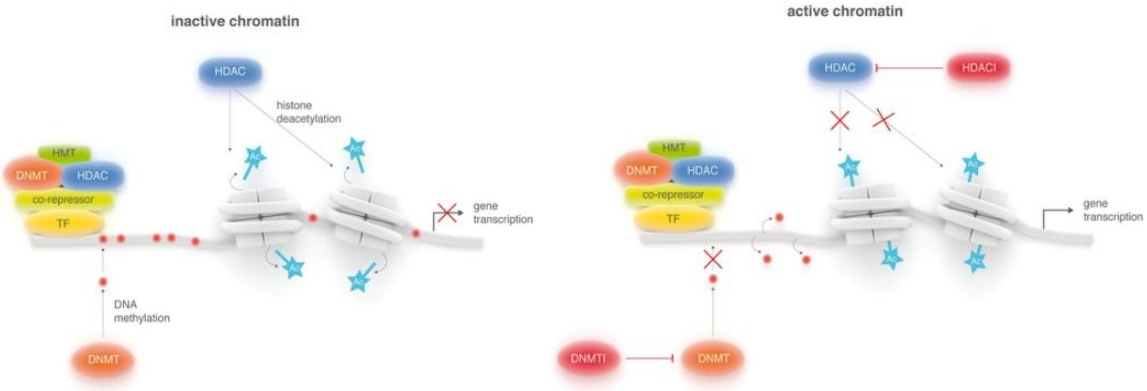


Figure I11. Representation of active and inactive chromatin states

There are two types of heterochromatin, the constitutive one, located mainly in centromeric and telomeric regions, and the facultative heterochromatin, which shares a compact structure but under specific developmental or environmental signaling cues loses its condensed structure and become transcriptional active (Oberdoerffer and Sinclair, 2007) . In the last years, several groups have reported genome-wide maps of histone modifications and DNA methylation patterns in pluripotent and differentiated cell types. From these data, a global picture of the architecture and regulatory dynamics of the chromatin is beginning to emerge (Rada-Iglesias *et al.,* 2011; Ernst *et al.,* 2011; Heintzman *et al.,* 2007; Creyghton *et al.,* 2010). Importantly, regulatory elements, namely promoters and enhancer regions, and their states (active, passive or poised) can be identified according to the histone profile, CpG content and DNA methylation levels (**Figure I12**).

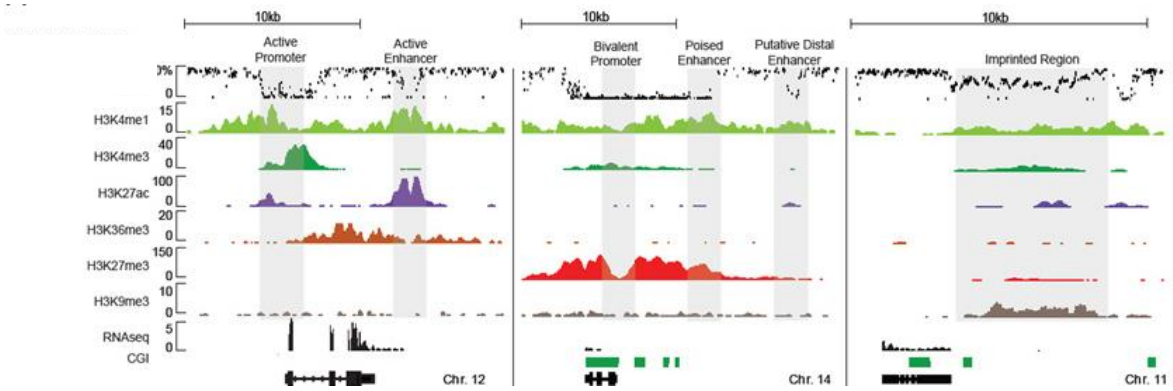


Figure I12. Epigenetic signature uncovers regulatory elements. The epigenetic marks together with the CpG content inform of the underlying element, indicated above and in grey. Representative image combining DNA methylation data (WGBS), histone marks (ChIP-seq) and RNA-seq data for undifferentiated hESC line at three independent loci. CpG islands (CGI) are indicated in green. Reproduced from (Gifford *et a*l. 2013).

3.4.1 Epigenetic signature of promoter regions

Promoters surround transcriptional start sites and serve as platform where the basal transcription machinery and RNA PolII are assembled during transcription initiation. Active genes promoters are enriched for H3K4me3, H3K9ac and H3K27ac and are found in highly expressed genes, as metabolic and housekeeping genes, which usually contain hypomethylated CpG islands within their promoters. Conversely, the promoters of silenced genes are repressed by the Polycomb group, and marked by H3K27me3 and H3K9me3.

Poised promoters, which are ready to be activated upon appropriated signaling, present bivalent domains in ESC, marked by the presence of both H3K4me3 and H3K27me3. Poised genes are generally associated with complex expression patterns and include key developmental transcription factors, morphogens, and cell surface molecules. In addition, several bivalent promoters seem to regulate transcription of lineage-specific microRNAs (Mikkelsen *et al.,* 2007).

Regarding DNA methylation, gene promoters are differentially regulated depending on their CpG content (Barrera *et al.,* 2008). Promoters with overlapping CpG islands are normally unmethylated and correspond to highly expressed genes in ESC and early differentiation cell stages. Since it is estimated that 95% of transcriptional start sites marked by H3K4me3 contain CpG islands, the presence of both characteristics usually indicates a promoter region (Bernstein *et al.,* 2006). Importantly, CpG island promoters corresponding to developmental genes are not usually repressed by DNA methylation in the non-expressing cells. They are instead silenced by H3K27me3, which may also protect them from spurious DNA methylation (Brinkman *et al.,* 2012; Bartke *et al.,* 2010). In contrast, low CpG density promoters as tissue-specific promoters, are preferentially silenced by DNA methylation (Xie *et al.,* 2013). Notably, in ESCs only a small percentage of low-CpG promoters contain significant levels of H3K4me3, and essentially none of them is marked by H3K27me3 (Mikkelsen *et al.,* 2007).

3.4.2 Epigenetic signature of enhancers and super-enhancers

Enhancers are regulatory regions, situated a variable distance from promoters, controlling gene expression in development and cell functions. They are typically 200 to 300 bp long and when they are active exhibit nucleosomal depletion and enrichment in DNaseI hypersensitivity sites as well as histone acetyltransferase p300 occupancy, suggesting an overall open chromatin structure (Heintzman *et al.,* 2007). Indeed, most enhancers are bound by multiple transcription factors and the Mediator complex, which is a large transcriptional coactivator necessary for RNA PolII recruitment to the promoters of enhancer-regulated target genes (Chen *et al.,* 2008). In addition, active enhancers are enriched in hm5C and depleted of 5mC (Hon *et al.,* 2014). Although active promoters and enhancers typically possess H3K27ac, enhancers are characteristically enriched in H3K4me1, in contrast to the H3K4me3 found at promoters (Ernst *et al.,* 2011; Heintzman *et al.,* 2007; Rada-Iglesias *et al.,* 2011). Therefore, enhancers can be distinguished from promoters by their unique histone modification signatures.

Inactive and poised enhancers exhibit high levels of H3K27me3 in addition to H3K4me1 and inactive and developmental poised enhancers are marked by 5mC and 5hmC, respectively (Kriaucionis and Heintz, 2009; Pastor *et al.,* 2011; Szulwach *et al.,* 2011; Ernst *et al.,* 2011). Enhancers are highly variable between lineages and cell types and confer cell type–specific or lineage-restricted gene expression (Xie *et al.,* 2012; Heintzman *et al.,* 2007; Gifford *et al.,* 2013; Hnisz *et al.,* 2013). Interestingly, only a small percentage of either active or poised enhancers overlap with CpG islands, and it is estimated that 94% of lineage-restricted enhancers are CpG-poor, depleted of H3K27me3 and its activity negatively correlate with DNA methylation (Xie *et al.,* 2012; Rada-Iglesias *et al.,* 2011; Xie *et al.,* 2013).

Super enhancers are a recently identified class of enhancers that play a major role in determining cellular identity (Hnisz *et al.,* 2013; Whyte *et al.,* 2013). They consist in large genomic regions (≥50 kb) possessing enhancer-like features: enrichment for H3K4me1, H3K27ac, Mediator complex, and transcription factors. However, unlike typical enhancers, they differ in size, transcription factor density and content, ability to activate transcription and sensitivity to perturbation. They contain unusually high levels of Mediator complex proteins and display enrichment for cohesins and Nipbl (Nipped-B homolog), a protein involved in enhancer–promoter communication (Hnisz *et al.,* 2013). The ≈200 identified super enhancers are often observed as clusters of smaller enhancers enriched for transcription factor motifs involved in cell identity (Whyte *et al.,* 2013). Furthermore, cancer cells show alterations in oncogene-related super-enhancers, as well as in other important genes in tumor pathogenesis (Hnisz *et al.,* 2013). Thus, super-enhancers play key roles in human cell identity in health and disease.

4. DNA METHYLATION DYNAMICS DURING CELLULAR DIFFERENTIATION

Stem cells are pluripotent cells with an open chromatin state, from which they can differentiate into many cell types and progressively develop a narrower potential. During development their gene-expression programs become more defined, restricted and, potentially 'locked in'. Due to the focus of this Thesis, I will expose the epigenetic changes occurring during development centered on DNA methylation.

The methylome is mainly preserved during development and heritable during DNA replication. However, DNA methylation undergoes dynamic remodeling during early embryogenesis to initially establish a globally demethylated state and subsequently, acquire progressively a lineage-specific methylome that maintains cellular identity and genomic stability (**Figure I13**). Concretely, two main DNA demethylation waves occur in pre-implantational embryo and in the primordial germ cells, which drop the methylation levels from 70%–85% to 30% and almost 5% respectively (Hon *et al.,* 2013; Ziller *et al.,* 2013). Later on, the genome became highly methylated again and DNA demethylation occur at lower extent across cell differentiation and specific hormone response (Kim *et al.,* 2009; Métivier *et al.,* 2008; Mohn *et al.,* 2008; Nagae *et al.,* 2011; de la Rica *et al.,* 2013).

After fertilization 5mC levels globally decrease from the first cleavage stage up to the early blastocyst, before being reacquired during and after implantation (Saitou *et al*., 2012). Immunofluorescence studies have shown that this global demethylation occurs asymmetrically on both parental genomes: the paternal DNA rapidly looses 5mC signal in the zygote before DNA replication whereas the maternal DNA looses its methylation more gradually. A number of factors have been implicated in this rapid demethylation of the male pronucleus, most prominently by Tet3 hydroxylation. The discovery of considerable amounts of 5hmC and 5fC in both pronuclei suggests that active DNA demethylation may also affect the maternal genome to some extent (Wang *et al.,* 2014). This reprogramming culminates in a globally demethylated genome in the inner

cell mass (ICM) of the pre-implantation embryo (by approx. E3.5 in mice), and correlates with the establishment of pluripotent cells, which can form embryonic stem cells *in vitro* (Smith and Meissner, 2013; Evans and Kaufman, 1981). The DNA methylation reprogramming is also coupled both to histone exchange and to novel chromatin regulation (Burton and Torres-Padilla, 2010).

The functionality of global demethylation after fertilization is to facilitate the activation of the pluripotency program in the embryo, which is supported by the observation that improper paternal demethylation is associated with delay in the activation of certain pluripotency genes and increase incidence in developmental failures (Gu *et al.,* 2011). Indeed, ESCs completely lacking DNA methylation are viable and competent for self-renewal, indicating that DNA methylation is totally dispensable for keeping proliferation capacity (Tsumura *et al.,* 2006). Interestingly, the global DNA demethylation after fertilization allows the expression of a large number of retrotransposon which may be required in the zygote (Evsikov *et al.,* 2004; Kigami *et al.,* 2003; Peaston *et al.,* 2004). Notably, not all the methylome is erased in this phase being the imprinted control regions faithfully maintained keeping the allele specific patterns.

ESCs, unlike differentiated cells, show extensive non-CpG methylation, most prominently at CpA dinucleotides. This methylation probably reflects a state of hyperactive *de novo* Dnmts activity. *De novo* non-CpG methylation is enhanced by Dnmt3L, which may direct *de novo* activity during pluripotency, but it is silenced upon differentiation (Arand *et al.,* 2012; Hu *et al.,* 2008).

A second wave of DNA methylation reprogramming occurs in the primordial germ cells (PGCs) (Saitou *et al.,* 2012). After implantation, the epiblast becomes the source of all embryonic lineages, including the precursors of germ cells that are called primordial germ cells. The PGCs precursors are highly methylated as somatic cells and over the following ~7 days lose 90% of their global methylation, representing the most comprehensive DNA methylation reprogramming event in the mammalian life cycle. They form a cluster of around 40 cells at embryonic day 7.25 (E7.25) and then

proliferate and migrate to colonize the genital ridges a by E10.5, where they continue to proliferate until E13.5 when enter into meiotic prophase in females and mitotic arrest in males. Genome-wide demethylation in PGCs is evidenced by a global loss of 5mC signal in immunofluorescence staining and is completed in both sexes by E13.5, soon after sex determination has taken place (Yamaguchi *et al.,* 2013).

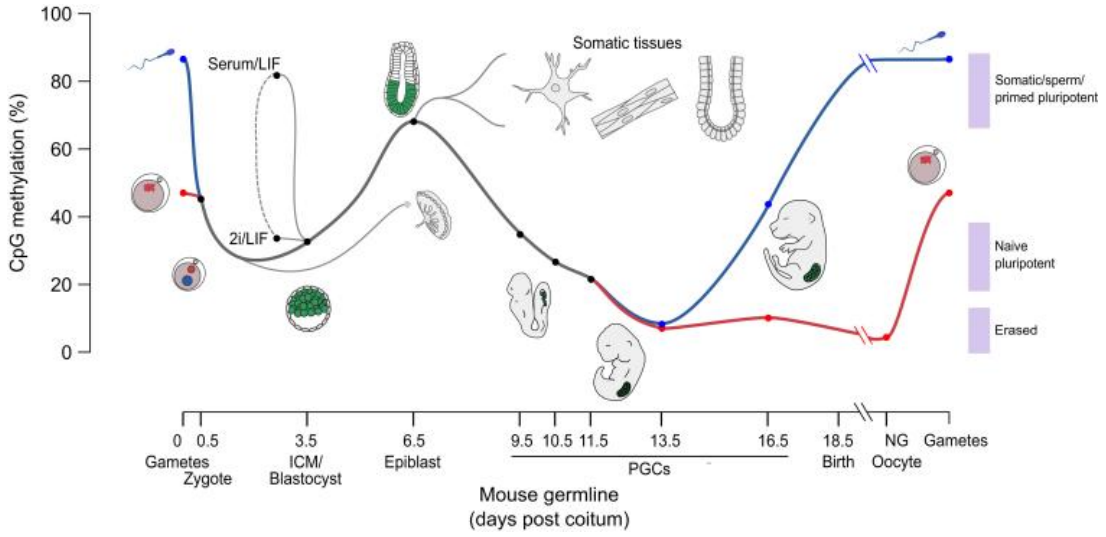


Figure I13. DNA methylation dynamics during the mammalian life cycle. Key developmental events are shown together with DNA methylation changes. The maternal and paternal genomes are colored red and blue, respectively. ESCs can be cultured from the ICM (dashed line). The placenta is derived from the blastocyst trophectoderm. Reproduced from (Lee *et al.*

Sequence-specific analysis revealed that 5mC erasure in PGCs includes genomic imprints, the Xist promoter and many transposable elements (Lane *et al.,* 2003; Hajkova *et al.,* 2002; Lees-Murdock *et al.,* 2003). Consequence of this global methylation erasure in the PGCs is ultimately the rapid acquisition of totipotency upon fertilization (Yamaji *et al.,* 2008), while it limits the potential for transgenerational transmission of epimutations and the re-establishment of the imprinting according to the sex of the cell. The mechanistic basis of DNA demethylation in PGCs has yet to be fully determined, and could potentially occur through the conversion to 5hmC coupled with passive depletion, but as well by direct replication-dependent loss of 5mC or active erasure associated with DNA repair pathways (Ooi *et al.,* 2007).

In the embryo, except for the special methylation dynamics of the PGCs, vital DNA methylation patterns are established to stably lock the pluripotency genes and lineage inappropriate genes, and to confer the tissue-specific properties to the somatic cells (Jackson *et al.,* 2004). Targeted gene disruption for each of the catalytically active Dnmts (1, 3a and 3b) results in a lethal phenotype demonstrating the essential role of DNA methylation in development (Li *et al.,* 1992; Okano *et al.,* 1999). Interestingly, Dnmt1 absence in ESCs inhibits differentiation, while lack of Dnmt3a/b does not, suggesting that DNA methylation maintenance, and not necessarily *de novo* silencing, is crucial for differentiation (Jackson *et al.,* 2004). Initiated in the early ICM and largely complete by approximately E6.5 hypermethylation affects most of CpGs in mammalian genome, and accumulates at pericentromeric repeats, transposable elements, extra-embryonic committed genes, pluripotency-associated and germline-specific genes, and low CpG density regions in a lineage-specific manner (Borgel *et al.,* 2010, Smith and Meissner, 2013).

During differentiation, CpG islands remain generally unmethylated due to an active and continuous exclusion of the DNA methyltransferases. Classical studies established that the unmethylated state of promoter CpG islands was strongly influenced by transcription factor binding. In the context of truncated or depleted transcription factor binding sites, CpG islands can progressively acquire heritable methylation (Macleod *et al.,* 1994; Lienert *et al.,* 2011). In addition, binding of the MLL family H3K4 methyltransferases protects promoters of developmental genes from DNA methylation (Otani *et al.,* 2009, Ooi *et al.,* 2007). Also the demethylation machinery, such as the Tet dioxygenases, cytosine deaminases and base excision repair enzymes, is tightly coupled to regulatory complexes associated with CpG islands, contributing to avoid CpG island hypermethylation. Specifically, Tet1 mostly binds CpG island promoters (Stadler *et al.,* 2011; Williams *et al.,* 2011) and its recruitment is not altered when DNA methylation is completely erased, suggesting that Tet1 function as a general epigenetic proofreader, which would keep the unmethylated state of CpG islands (Wu *et al.,* 2011; Williams *et al.,* 2011).

Considering lineage-specific DNA methylation signature, the recent whole-genome bisulphite sequencing of 17 adult mouse tissues has shown that the methylome is more than 6.7% variably between tissues (Hon *et al.,* 2013). Studies performed on

lineage-specific cell lines, such as neuronal progenitors (Mohn *et al.,* 2008), myeloid or lymphoid progenitors during hematopoiesis (Ji *et al.,* 2010), mesenchymal progenitors (Sorensen *et al.,* 2002), and on the three germ layers (Isagawa *et al.,* 2011;Gifford *et al.,* 2013; Hon *et al.,* 2013), have revealed that lineage-specific DNA methylation signatures are established in the early lineage commitment. This is evidenced by the fact that methylomes of different tissues cluster according to the germ layer origin (**Figure I14**). In contrast, little changes are observed at later stages of differentiation within the same lineage, suggesting that DNA methylation is crucial to define and maintain cell identity.

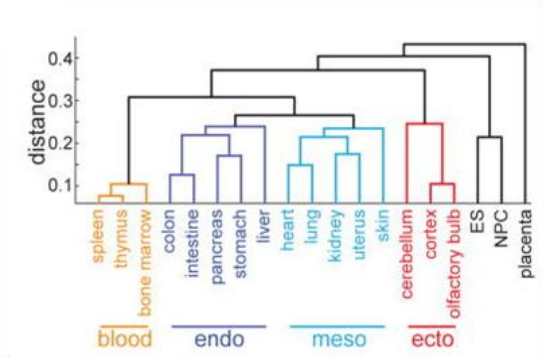


Figure I14. DNA methylomes correlated with germ layer origin. Dendrogram constructed from 1kb regions exhibiting significant tissue-specific methylation. Reproduced from (Hon *et al.,* 2013).

During differentiation DNA methylation level globally increases, while small tissues-specific demethylation occurs across cell differentiation too. This is evidenced by a recent study where the three germ layers were derived from ESC and analyzed by Reduced Representation Bisulphite Sequencing (RRBS) (Meissner *et al.,* 2008). They observed gain of methylation in CpG-poor intergenic regions showing overlap between lineages, whereas the observed loss of methylation occurred in a more lineage-specific fashion at intergenic regions too (Gifford *et al.,* 2013, Smith and Meissner, 2013), as anticipated by another study using MeDIP approach (Isagawa *et al.,* 2011). Tissue-specific DNA demethylation has been reported during hematopoiesis (Borgel *et al.,* 2010; Ji *et al.,* 2010; Bocker *et al.,* 2011; Shearstone *et al.,* 2011; Shearstone *et al.,* 2011), in liver cells (Nagae *et al.,* 2011;Waterland *et al.,* 2009;Brunner *et al.,* 2009), in neurons (Guo *et al.,* 2011; Meissner *et al.,* 2008; Stadler *et al.,* 2011; Mohn *et al.,* 2008), in mesenchymal progenitors (Sørensen *et al.,* 2010) and in muscle cells, detailed in section 5.

Importantly, the demethylated regions are enriched in tissue-specific binding motifs as observed in 21 human tissues (Nagae *et al.,* 2011). This suggests that tissue-specific transcription factors could bind upon demethylation to the corresponding TFBS or TFs might shape the methylation state of the underlying region (Stadler *et al.,* 2011; Feldmann *et al.,* 2013). Interestingly, these demethylated sequences are not restricted to promoter regions, disclosing distal regulatory elements as a crucial dynamically methylated regions (Stadler *et al.,* 2011; Meissner *et al.,* 2008).

Summing up, a global progressive gain of developmental DNA methylation outside CpG islands, and a tissue-specific demethylation are indispensable to canalize cell fate decisions, followed by the acquisition of cellular differentiation programs.

5. DNA METHYLATION DYNAMICS DURING MYOGENESIS

5.1 MyoD DNA demethylation: a crucial discovery by chance

In 1973, Peter Jones was screening for the effect of a chemotherapeutic drug on mouse embryonic fibroblasts (10T 1/2 cells) when he realized that treated cells turned into huge syncytium of multinucleated cells visible to the naked eye. This observation triggered a series of important discoveries that linked for the first time DNA methylation with cellular differentiation.

After confirming the myogenic phenotype of 10T 1/2 treated cells (Constantinides *et al.,* 1977) and demonstrating that the tested drug was a potent inhibitor of methylation (5-Azacytidine) (Jones *et al.,* 1983), MyoD was identified as the key gene involved in the reprogramming into muscle cells (Lassar *et al.,* 1986). Later on in 1990, P. Jones demonstrated for the first time and in a seminal work that a transcription factor (MyoD) was regulated by DNA methylation (Jones *et al.,* 1990). The same myoblastic conversion also occurred after the transfection of 10T1/2 fibroblasts with an antisense RNA against the maintenance DNA methylatransferase Dnmt1 (Szyf *et al.,* 1992). The drugs 5-Azacytidine and 5'Aza2 deoxycytidine are both cytosine analogs

which are incorporated into the replicating DNA in place of cytosine and trap DNA methyltransferases (Dnmts) leading to DNA demethylation (Jones *et al.,* 1983).

However, later analysis of 5-Azacytidine treated 10T1/2 fibroblasts revealed the unmethylated state of the CpG island surrounding MyoD promoter. Surprisingly at that moment, this CpG island was constitutively free of methylation in all the tissues of the organism (Bird, 1984), but methylated in 10T1/2 fibroblasts as an *in vitro* associated artifact and in tumoral cells (Jones *et al.,* 1990), raising the question whether MyoD promoter demethylation was indeed the signal required to initiate the myogenic program during embryonic development. This problem was solved when Brunk and colleagues., showed that another regulatory region in MyoD, the distal control element at -20 kb of the TSS, was specifically demethylated during somitogenesis in mice (Brunk *et al.,* 1996). They observed by sodium bisulphite conversion that three CpGs located in the enhancer regions were methylated in liver, heart and brain tissues (50-60%) and in fibroblast and neuroblastoma cells (50-80%), while they were almost totally demethylated in C2C12 myoblast cell line (<10%). Moreover, they analyzed the methylation state during somitogenesis observing an important loss dropping from 80% methylation in presomitic mesoderm until 45% methylation in the more mature somites (somites 13-14). Importantly, demethylation of MyoD distal enhancer was previous to gene activation, suggesting that demethylation of the distal enhancer is an active, regulated process that is essential for MyoD activation (Brunk *et al.,* 1996). Intriguingly, mutations of these differentially methylated CpGs did not prevent precocious activation of MyoD in trangenic mice, which implies that DNA methylation is not sufficient for MyoD inactivation (Brunk *et al.,* 1996).
Interestingly, MyoD CpG island promoter hypermethylation has been observed as a consequence of aging, at least in the hematopoietic lineage (Fernandez *et al.,* 2012) and in colon (Kawakami *et al.,* 2006), supporting the hypothesis that aging-related DNA methylation is an early step in carciongenesis (Toyota and Issa, 1999).

Later studies observed that not just a reprogramming through myogenic lineage was possible after DNA methylation impairment, but also an enhancement of the muscle differentiation was observed. Indeed, treatment of C2C12 myoblasts with 5-

Azacytidine promotes myogenesis, resulting in myotubes with enhanced maturity as compared to untreated myotubes (Hupkes *et al.,* 2011). A recent study analyzing the effect of 5-Azacytidine treatment on cell cycle regulation during proliferation and at the early stages of differentiation, showed that inhibition of DNA methylation increased the expression of checkpoint genes involved in cell cycle progression and up-regulated myogenic transcription factors (Montesano *et al.,* 2013). These results suggested that 5-Azacytidine-induced DNA demethylation could modulate cell cycle progression and enhance myogenesis, open novel clinical uses for 5-Azacytidine in the field of muscle pathologies. However, Montesano and collaborators did not analyze the methylation state of the myogenic genes and assumed that the effect of 5-Azacytidine was on DNA demethylation.

Intriguingly, Asano's lab observed an enhancement in myotube formation upon overexpression of Dnmt1 in myoblast cell line C2C12. By the use of methylation sensitive enzymes, they detected a positive correlation between MyoD expression and exon 1 and exon2 higher methylation levels (Takagi, *et al.,* 1995).

Finally, in 2007 H. Blau's laboratory showed that when human fibroblast and mouse muscle cells were fused to form non-dividing heterokaryons, active DNA demethylation was observed at the human MyoD promoter, which accompanied its activation and expression in the fibroblast cells (Zhang *et al.,* 2007).

5.2 Myogenin DNA demethylation

The correlation between muscle differentiation and DNA methylation was further underscored by the finding that *Myogenin* promoter became demethylated at the onset of C2C12 muscle differentiation. Lucarelli and collaborators reported that the methylation status of a single CpG site at 340 bp upstream from the TSS affected *Myogenin* transcription in mouse tissues and C2C12 cells. Addressing the methylation state using methylation sensitive restriction enzymes, they showed that muscle tissue and differentiated C2C12 were unmethylated correlating with *Myogenin* expression,

whereas in non-muscle tissue (spleen and brain) and proliferating MB, this CpG was methylated and the transcription repressed (Lucarelli *et al.,* 2001). Previously, the same lab had shown that C2C12 myoblast treated with 5'Aza2 deoxycytydine lead to an increase in Myogenin expression and a reduction of DNA methylation, addressed by northern blot hybridization and digestion with methyl sensitive restriction enzymes followed by southern blot hybridization, respectively (Scarpa *et al.,* 1996).

A decade later, Palacios' work showed a reduction in DNA methylation in *Myogenin* promoter upon C2C12 differentiation, and addressed the DNA methylation status during embryogenesis, showing by bisulphite sequencing that the *Myogenin* promoter became demethylated as the development proceeds (Palacios *et al.,* 2010). At day E9.5, posterior somites where methylated while in anterior somites, where the *Myogenin* gene was expressed, dropped 30% the DNA methylation level. Moreover, they proposed that the efficient activation of the Myogenin promoter requires demethylation following binding of Six1 and MEF2 proteins.

In parallel, Oikawa's work demonstrated that the methyl-CpG-binding protein CIBZ suppressed myogenic differentiation by directly binding to the methylated *Myogenin* promoter and inhibiting its expression in proliferating C2C12 cells. In C2C12 myoblast they observed a 42% of methylation, which slightly dropped to 32% in differentiating C2C12 myotubes and to 8% in C2C12 treated with 5'Aza2 deoxycytydine accompanied with an increase of gene expression. *In vivo* they observed higher levels of methylation at non-muscle tissue compared to skeletal tissue (brain, 62,5%; kidney, 68,5%; skeletal muscle, 39,2%). In addition, they showed that the binding of MyoD at the CpG free E-box located at the *Myogenin* promoter was independent of DNA methylation but, importantly DNA demethylation might avoid the binding of repressors (CIBZ) allowing the binding of the activator comple MyoD/Pbx/Meis complex (Oikawa *et al.,* 2011). Altogether suggesting that DNA methylation might modulate the competition between transcription factors and repressive DNA binding proteins.

Finally, Strom's lab also observed DNA demethylation in the *Myogenin* promoter region but at non-CpG regions (Fuso *et al.,* 2010), claiming that demethylation occurred more rapidly at non-CpG regions than at CpGs. However, these results await further validations and are not supported by the general knowledge restricting non-CpG methylation to ESC.

5.3 Genome wide studies addressing DNA methylation in muscle tissue

Recently, a small number of genome-wide DNA methylation reports compared skeletal muscle cells or muscle tissue with non-muscle samples focusing on promoter regions and/or CpG islands

Illingworth *et al.* compared DNA methylation in CpG islands of human skeletal muscle, brain, leukocytes and spleen using affinity purification of 5mC-enriched DNA fragments, followed by hybridization to microarrays (Illingworth *et al.,* 2008). Muscle and spleen had the highest levels of methylation and differential methylation between tissues. An average of 8% of the CpG islands were methylated in at least one sample and a 5% displayed differential methylation between tissues. The results indicated that CpG islands at promoters regions were the least susceptible to became methylated, while intragenic or intergenic regions were the most susceptible showing tissue-specific methylation patterns. Likewise, in a promoter screening of 1,505 CpGs surveying 808 gene promoters among 1,628 human samples, Fernandez and colleagues identified 183 muscle-specific differentially methylated CpGs (Fernandez *et al.,* 2012).

Using samples from normal second-trimester fetuses, Yuen and collaborators examined muscle, brain, kidney, lung and skin cells looking for differential methylation at 1,315 CpG sites located in promoters (GoldenGate Methylation Cancer Panel I, Illumina, CA, USA) (Yuen *et al.,* 2011). They found that 195 sites were significantly

different between the different fetal tissues. Muscle cells had the second highest number of tissue-specific DMRs and brain cells had the highest number.

Microarray-based DNA methylation profiling (HumanMethylation27 BeadChip, Illumina) was used by Berdasco and colleagues, to address DNA methylation plasticity of human adipose-derived stem cells (hASCs). They compared *in vitro* differentiated hASCs through myogenic- or osteogenic-lineages and the corresponding primary myocytes and osteocytes. The results showed that myogenic- and osteogenic- derived hASC cells that shared much of the DNA methylation landscape of primary myocytes and osteocytes respectively, representing that the plasticity of DNA methylation patterns has an important role in lineage commitment of adult stem cells (Berdasco *et al.,* 2012).

During this Thesis, two papers have been published addressing at genomic-scale the DNA methylation patterns in muscle cells. Last year, Ehrlich's lab addressed the myogenic methylome outside of the promoter regions and CpG islands performing a RRBS. They compared myogenic cells (myoblasts and myotubes) against 16 types of cell cultures derived from non-cancerous tissues other than muscle. They did not found DNA methylation changes between myoblasts and myotubes, while they found that 10,048 and 9,592 CpG sites displayed myogenic hypomethylation or hypermethylation, respectively, when myogenic cells were compared to non-myogenic cells. They also compared skeletal muscle tissue with 14 non-muscle tissue samples and determined that 11,255 CpGs displayed muscle-associated hypomethylation, while only 761 sites exhibited muscle-associated hypermethylation. Globally, their findings suggested *de novo* methylation before the myoblast stage and demethylation in mature muscle (Tsumagari *et al.,* 2013).

Very recently, it came up a paper where DNA methylation of human myoblast and myotubes was compred using the Infinium HumanMethylation 450 BeadChip array. Contrary to the previous genome-wide study, they found a tendency of increased of DNA methylation during myoblast differentiation. Gene ontology analysis revealed that hypermethylated genes were related to muscle contraction and muscle system

process, being some of them down-regulated during terminal differentiation process (Miyata *et al.,* 2014).

In addition, there is study that concludes that DNA methylation globally decreases very rapidly upon acute exercise (Barrès *et al.,*, 2012). Specifically, muscle biopsies of healthy men/women before and after 20 minutes of acute exercise were analyzed by luminometric methylation assay (LUMA) reveling a global decrease in DNA methylation (Barrès *et al.,* 2012). However, bisulphite sequencing analysis of target genes showed that methylation changes occurred mainly in a non-CpG context, which lead them to the interpretation that non-CpG methylation could play a specialized role mediating rapid or transient methylation remodeling. In parallel, other study evaluated the effect of 6 months of moderate aerobic exercise, in persons with and without diabetes type 2 family history, showing that 134 genes changed their DNA methylation status independent of the familiar history after six month long training period by MeDIP-ChIP analysis (Nitert *et al.,* 2012). Most of the genes showed decreased methylation (115 out of 134) and were involved in retinol metabolism, calcium-signaling pathway, and starch and sucrose metabolism.

Summing up, several studies have emerged in the last 20 years concerning the regulatory role of DNA methylation in myogenic differentiation (reviewed in Carrió et Suelves, 2015, submitted). Most of the studies have been focus on MyoD, Myogenin and genome-wide promoters regions pointing out the relevance of DNA demethylation in myogenic specification. However, very little is known about the methylation dynamics in the rest of the genome, and in particular in regulatory regions of other crucial muscle-specific genes to understand the contribution of epigenetic regulation, specifically DNA methylation, during myogenic cell fate determination and terminal differentiation.

OBJECTIVES

Myogenesis encompasses the formation of skeletal muscle by entailing cell fate decisions not only during development but also during muscle regeneration. The myogenic process is altered in many muscle pathologies where attractive stem cell based therapies are emerging, despite the mechanisms underlying muscle-lineage determination are not yet well understood. A large number of studies have highlighted the relevance of epigenetic mechanisms in the regulation of cell identity. Interestingly, a clear example of that was reported in the myogenic lineage, when a seminal work pointed out DNA methylation as a key regulatory mechanism involved in the silencing of the myogenic master regulatory transcription factor MyoD in non-muscle cells. These premises led us to hypothesize that DNA methylation, together with other epigenetic mechanisms and transcription factors, orchestrate the myogenic program and confer a unique signature to muscle identity.

To address this issue, we proposed:

- **To survey the global DNA methylation profiles during myogenesis**. We will characterize the DNA methylation changes associated with myogenic cell fate commitment and with myogenic differentiation.

- **To determine the epigenetic signature of specific genes implicated in myogenic cellular identity.** This study will be performed in isolated muscle stem cells, as well as in an inducible embryonic stem cell-based myogenic model.

- **To study the potential regulatory role of the DNA methylation in myogenesis.** We will analyze the associations between differential methylation and the underlying sequence, the histone code, the transcription factors occupancy and the gene expression modulation.

- **To assess the relevance of DNA demethylation in the myogenic progression.** We will analyze the effect of DNA demethylases disruption in an inducible stem cell-based myogenic model.

-

MATERIALS AND METHODS

1 SAMPLES AND CELL CULTURE

1.1 Murine primary myoblasts, myotubes and myofibers

Primary myoblasts (MBs) were isolated from muscles from three 6-to 8-week-old mice (kindly provided by P. Muñoz-Cánoves (UPF, Barcelona)) as previously described (Suelves *et al.,* 2007). The hind limbs were subjected to both mechanical and enzymatic dissociation with Pronase protease 1% (Calbiochem) at 37ºC. The filtered digest was centrifuged through an isotonic Percoll (Amersham) gradient (60% overlaid with 20%) and cells were collected from the gradient interface, resuspended in growth medium (GM: Ham's F-10 plus 20% FBS, 5 ng/ml bFgf, 100 U/ml penicillin and 100 µg/ml streptomycin) and grown on 100 mg Collagen I rat tail (Becton and Dickinson) coated tissue culture dishes at 37°C with 5% CO_2 in the incubator. To induce myotube formation, confluent proliferating primary myoblasts were grown on matrigel coated culture dishes (Basement Membrane Matrix, Becton and Dickinson) and switched to differentiation medium the next day (DM: DMEM plus 2% horse serum, 100 U/ml penicillin and 100 µg/ml streptomycin (Life technologies)). After 4 days in DM the myotubes (MTs) were harvested. While one hindleg was used to extract the primary myoblasts as mentioned, the quadriceps of the opposite hindleg was directly harvested representing the final skeletal muscle tissue, composed mainly by myofibers (MFs). This myofibers sample were used to perform the AIMS-seq analysis, however, single myofibers were extracted from quadriceps muscle biopsies from C56Bl/6j mice strain (Jackson Lab) in collaboration with S. Gutarra and P. Muñoz-Cánoves for the bisulphite sequencing analysis (**Figure M1**). We chosed to purify the single myofibers to avoid non-myogenic cells contaminations, present in the muscle tissue. Mouse protocols were approved by the Animal Care and Use Committee of the Barcelona Biomedical Research Park (PRBB), and the Ethical Committee for Animal Experimentation of the Government of Catalonia.

Figure M1. Single myofibers isolated from EDL muscle.
5x amplification

1.1.1 Murine quiescent and after-injury activated satellite cells isolation

Quiescent satellite cells were isolated from muscles from three 6-to 8-week- old Pax7$^{Cre\ -\ /YFP}$ mice generated by Bosnakovski and collaborators (Bosnakovski *et al.,* 2008) and kindly provided by P. Muñoz Cánoves (UPF, Barcelona). The extraction of muscle mononuclear cells was performed as explained in the previous section and the satellite cells population (cells expressing yellow fluorescent protein) was sorted by fluorescence-activated cell sorting (FACS) and directly frozen. Alike, *in vivo* activated satellite cells were isolated from Pax7 $^{Cre+/YFP+}$ mice 6 and 72h after CTX injury by FACS sorting (**Figure R27**). Intramuscular injection of 300 µl of 10−5 M CTX (Latoxan) was performed in the quadriceps muscle of the mice (Suelves et al., 2007) by V. Ruiz and M. Jardi (P. Muñoz-Canoves Lab, UPF, Barcelona). This concentration and volume were chosen to ensure maximum degeneration of the myofibers.

1.2 Murine cell lines

Murine embryonic stem cell lines were kindly provided by M. Carrió (cGR8) and F. Lluís (E14Tg2 and RD1); neuronal precursor cells (NPC) were also provided by F. Lluís (CRG, Barcelona) and cardiomyocite cells (HL1) by S. Pagans (UdG, Girona). MEFs and C2C12 cells were cultured in DMEM complemented with 10% Fetal Bovine Serum (FBS), 100 U/ml penicillin and100 µg/ml streptomycin (Life technologies). Three biological replicates of each experiment were performed and analyzed.

1.3 Human samples

The three human embryonic stem cell lines were kindly provided by M. Barrero (CNIO,, Madrid) and the three human myoblasts by N. Luna and E. Gallardo (Hospital de la Santa Creu i Sant Pau, UAB, Barcelona).

1.4 Rhabdomyosarcoma cell lines and Aza treatment

U29415 (*Pax7*CreER), U23674 (*Myf6*Cre) and U37125 (*Myf6*Cre) mouse primary tumor cell lines were generated in the laboratory of C. Keller (OHSU, Oregon, USA) and were described in (Abraham *et al.,* 2014) and (Rubin *et al.,* 2011). We amplified the cell lines in DMEM supplemented with 10% fetal bovine serum, penicillin (100 U/mL), streptomycin (100µg/mL (Invitrogen) at 37°C with 5% CO_2.

To evaluate if the aberrant expression pattern of U29415 and U23674 cell lines could be corrected, cells were treated with the demethylating drug 5-Aza-2'-Deoycytidine (Aza) Sigma-Aldrich). Aza is a cytosine analog that is incorporated into replicating DNA in place of cytosine and trap DNA methyltransferases (Dnmts), resulting in proteosomal degradation and heritable global demethylation as cells divide (Kelly *et al.*, 2010). Due to their different proliferative capacities U29415 were treated with 3 µM and U29415 with 20µM Aza, concentrations previously tested in C. Keller laboratory (Abraham *et al.,* 2014). The medium was changed daily and cells were collected after 24, 48 and 72 hours treatment.

1.5 Inducible *Pax7* ESC-derived myogenic model

Inducible *Pax7* ESC-derived myogenic model allows generating myogenic precursors by inducing *Pax7* expression in a ES cell line engineered at R. Perlingeiro's Lab (Darabi *et al.*, 2011). *iPax7* ESCs were co-cultured with irradiated mouse embryonic fibroblasts MEFs, obtained as described in (Jozefczuk *et al.,* 2012). The co-culture was maintained in knockout DMEM (Invitrogen) supplemented with 15% ES qualified FBS (Invitrogen), 1% penicillin/streptomycin (Invitrogen), 2 mM Glutamax, 0.1 mM b-mercaptoethanol, and 1,000 U/ml Lif (Millipore) by using daily fresh medium and at 37 C^{o} in 5% CO_2 culture conditions. To induce ESC differentiation ESCs were trypsinized, resuspended in EB differentiation media (EBM), and preplated for 30 minutes on gelatin-coated dishes to remove MEFs, wich get attached more quickly. After counting, $1x10^6$ cells were plated on 150 ml petri dishes and maintained in a moving incubator to induce EB

formation. EBM consisted of Iscove-modified DMEM (Life Technologies) supplemented with 15% ES qualified FBS (Life Technologies), penicillin/streptomycin, 2 mM Glutamax, 200 mg/ml iron-saturated transferring (Sigma-Aldrich), 4.5 mM monothioglycerol (Sigma), and 50 mg/ml ascorbic acid (Sigma-Aldrich).

To induce Pax7 transgene expression and therefore the myogenic cell fate, Doxycyclin was added to the culture at a final concentration of 1ug/µl beginning at day 3 of EB differentiation. Doxycycline stock solution was prepared by dissolving Doxycyclin (D9891, Sigma-Aldrich) in sterile PBS at a concentration of 1 mg/ml and was stored at 4Cº.

1.5.1 Fluorescence-Activated Cell Sorting (FACS) Analysis

To isolate the cells in the EBs corresponding to the mesoderm precursors, where the myogenic precursors arise from, FACS analysis was used after labeling the cells with CD140a (PdgfαR and CD309 (Flk) antibodies. Day 5 induced and non induced EBs were harvested, washed twice with PBS, and trypsinized for 1.5 minutes in a 37Cº water-bath with continuous shaking. Trypsin was inactivated by adding four volumes of PBS supplemented with 10% FBS (PBSF) and the cells were resuspended and filtered through a 70 mm strain to remove cell clumps. Cells were washed once with PBS and then incubated for 5 minutes in PBSF in the presence of Fc Block (1 ml/4million cells (E-Bioscience). Staining was performed by adding 0.5 ml of each antibody per 1 million cells and incubating on ice for 25 minutes. We used PE-conjugated anti-mouse CD140a (PdgfαR) and APC-conjugated anti-mouse CD309 (FLK-1) (both from E-Bioscience). Cells were washed twice with PBS and then resuspended in PBSF containing propidium iodide to exclude dead cells. Samples were analyzed and sorted using a FACS Aria II (BD Biosciences). After sorting the $PdgfαR^{+}/Flk^{-}$ population with and without Doxocycline day 5 EBs, the sorted cells were replated in monolayer culture on gelatinized flasks in EBM for 5 more days and then harvested, representing the iPax7 myoblasts precursor state. To induce the myotube formation, the iPax7 MBs precursors were cultured on DMEM (Gibco) supplemented with 2% Horse Serum

(Hyclone), 1% penicillin/streptomycin and 2 mM glutamax and, importantly, Doxocyline was removed to allow terminal myogenic differentiation, which physiologically only occurs upon Pax7 expression withdrawal.

1.6 Lentiviral infection

To evaluate the role of selected enzymes (Tet1, Tdg and Apobec2) on DNA demethylation, we infected the cells (i*Pax7*-ESC or i*Pax7* myoblasts precursors) with lentiviruses carrying shRNA targeting these enzymes. Lentiviral infection allows infecting cycling and no cycling cells and the shRNA-based strategy provides a long lasting repressive effect. *Escherichia coli* containing the vectors were obtained from the Biomedical Genomic Center at the University of Minnesota. The transfer vectors containing the shRNA were pLKO.1-puro vectors for Tdg and Apobec2 and pGIPZ for Tet1 (**Figure M.2**). Two shRNA were highly efficient targeting Apobec2 transcript: shRNA # 1 CCTGGCTTCCTGATTCTACTT and shRNA #4 GCTACCAGTCAACTTCTTCAA.

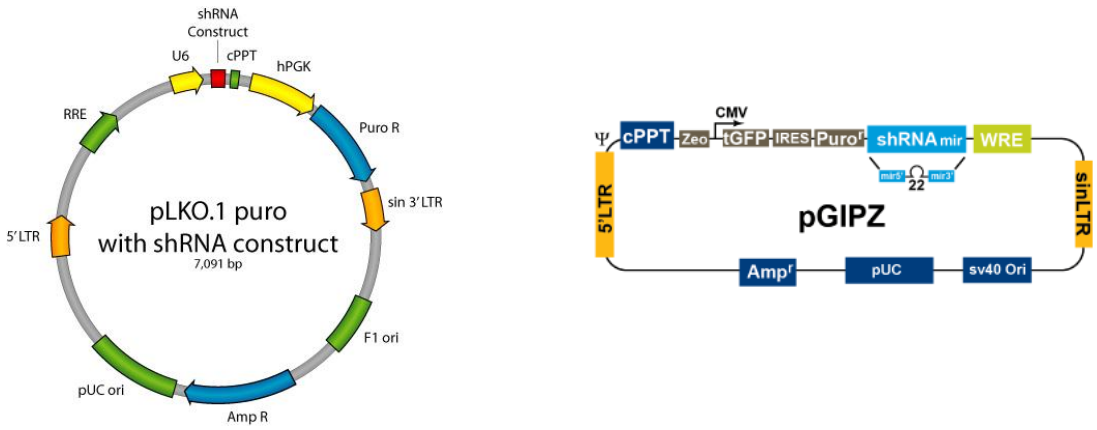


Figure M2. pLKO and pGIPZ vector maps.

Lentiviruses were produced by cotransfection of the transfer vector and the packaging constructs (pVSV-G, pREV, and pD8.74 (Follenzi *et al.,* 2000)) in the HEK 293T cells. Transfections were performed using Lipofectamine LTX with Plus Reagent (Life Technologies) following manufacturer instructions. Supernatants containing the lentiviral particles were collected 36 hours after transfection, passed through a 0.45

mm filter, and applied to the cells of interest. Tet1 shRNA experiments were done on iPax7ESC and Tdg and Apobec2 on iPax7 MB precursors. Cells were spin-infected at 1,100g for 1.5 hours at 30Cº and then incubated in the presence of the lentivirus for an additional 6 hours at 37Cº. After, in order to achieve a good transduction, a second spin infection with fresh viruses was performed. After 12 hours the medium was replaced by virus-free culture media. Four days after the infection the cells were harvest and analyzed. Cells transfected with pGIPZ vector were sorted by GFP while cells carrying pLKO vector, which contains a Puromycine resistance, could not be selected because iPax7 ESC-derived cells were already resistant to Puromycine due to the previous engineering. The efficiency of the pLKO insertion was assessed by evaluating the repressive effect of the shRNA compared to the control vector.

To amplify and purify all vectors bacteria were grown on 300µl LB medium with 100 µg/ml of ampicillin in a shaker at 37º C overnight and the GenElute™ High Pure Plasmid Midiprep Kit (Sigma-Aldrich) was used to extract genomic DNA following manufacturer conditions.

2 DNA METHYLATION ANALYSIS

2.1 Genomic DNA extraction

Genomic DNA was extracted with a standard phenol-chloroform protocol. Briefly, 1ml of lysis buffer (100mM Tris HCL pH 7.8, 1% SDS, 50mM EDTA pH 8.0) supplemented with 20 µl of 10mg/ml Proteinase K (Invitrogen) was used to lysate $5x10^6$ cells, overnight at 37ºC. Next day, 5 µl of 10mg/ml RNase A (Invitrogen) were added and incubated 1 hour before the phenol-chloroform extraction. Next, genomic DNA was ethanol precipitated, resuspended in PCR-grade pure water, quantified with a NanoDrop equipment (Thermo Scientific) and run in a 1% agarose gel stained with ethidium bromide to check its integrity.

2.2 AIMS-Seq method

AIMS-seq method was used to assess non-targeted genome-scale DNA methylation differences between samples.

2.2.1 AIMS-seq library construction

Differential methylation libraries were generated using a next generation sequencing adaptation of the Amplification of InterMethylated Sites (AIMS) technique (Frigola *et al.,* 2002). This method is a methyl-sensitive restriction enzyme based approach that combines the consecutive use of two neoisoschizomer enzymes, SmaI and XmaI, which recognize the same cleavage site (5'-CCCGGG-3') but present different methylation sensitivities and generate distinct end-fragments: SmaI digests unmethylated DNA leaving blunted ends, whereas XmaI digests DNA regardless methylation state and generates cohesive ends. Following 1µg genomic DNA digestion with both enzymes (SmaI from Roche, 6 hours digestion at 25ºC; XmaI from NEB, 6 hours digestion at 37ºC), cohesive adapters were ligated specifically to XmaI products (5'CCGGTCAGAGCTTTGCGAAT and 5'CCGAATTCGCAAAGCTCTGA were the adapter sequences), allowing the amplification of intermethylated fragments by conventional PCR. The amplified products were next-gen sequenced using Illumina GA2 platform at the Genomics Unit of CRG (Barcelona), and the high quality reads were mapped to the mm9 mouse genome (workflow scheme in **Figure R2.A)**.

2.2.2 Sequencing data processing, mapping and analysis

The global low quality reads were filtered out and the 5' or 3' end of the remaining reads were trimmed according to the quality. Adapter sequences were eliminated and reads were aligned to the UCSC/mm9 mouse reference sequence using the Bowtie alignment program (v0.11.3)(Langmead *et al.*, 2009). Only reads mapping uniquely inside amplicons between 20 and 2,000 base pairs were considered for further analysis

(**Figure R2.B**). Mapped reads of each sample were normalized using total-sum scaling method and respect the sample with the lowest number of mapped reads (normalization factor: 0.7439 for ESC, 1 for Myoblasts, 0.8683 for Myotubes, and 0.7713 for Myofibers). Differential DNA methylation at any given amplicon was assessed by comparing the number of reads per amplicon after normalization of the samples considered. Comparisons between samples were done using the DesSeq R package. R package allowed detecting differentially methylated regions (DMRs) from discrete count data based on a negative binomial distribution and assigned a single pooled empirical dispersion estimated from all the samples (Anders and Huber, 2010). Only amplicons with at least one read in at least one sample were considered for the detection of DMRs. A highly restrictive threshold of $\log_2$ Fold Change of Nº of reads ≥ 3 with a p-adj value < 0.01 was used to discern differentially methylated regions between samples. This threshold was chosen according to the empirical validation of several DMRs with the gold standard method to analyze cytosine methylation, the sodium bisulphite conversion method **Figure M3**).

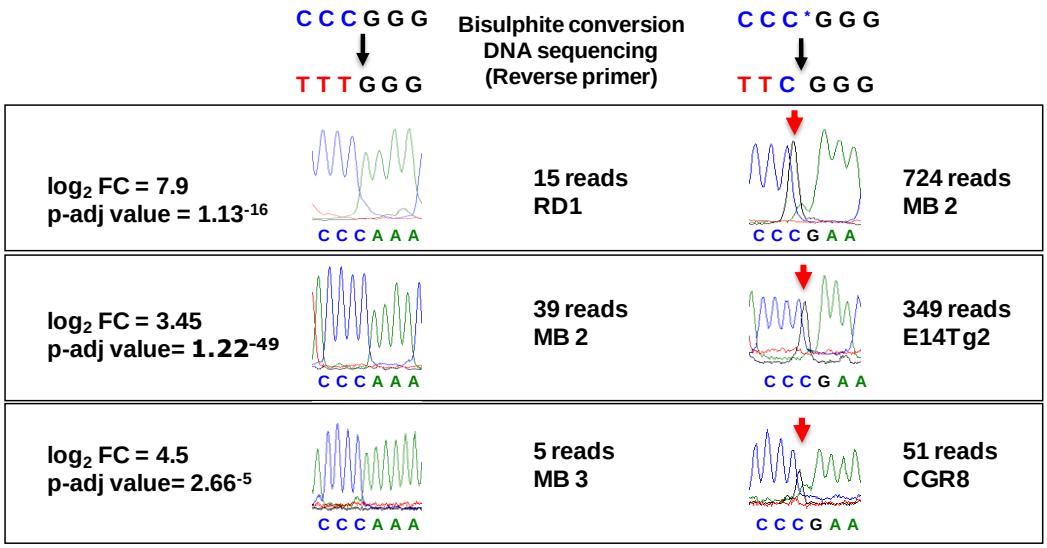


Figure M3. Sodium bisulphite sequencing validation of the DMRs. A threshold of $\log_2$ Fold Change ≥ 3 of difference in the number of reads between samples was used to discern differentially methylated regions. C* stands for methylated cytosine.

2.3 Targeted sodium bisulphite sequencing

Bisulphite sequencing method was used to asses DNA methylation state of specific sequences at single base resolution. The process of bisulphite treatment exploits the different sensitivities of cytosine and 5mC to deamination by bisulphite under acidic conditions—in which cytosine undergoes conversion to uracil, whereas 5mC remains unreactive. Bisulphite conversion was performed using 300 ng of genomic DNA with EZ DNA Methylation-Gold™ Kit (ZymoResearch, Ornage) according to manufacturer's instruction. Bisulphite sequencing was performed as previously described (Clark *et al.,* 2006). Briefly, for each region of interest, a first PCR amplification was performed using 1ųl of bisulphite-treated DNA in a final volume of 12,5 ųl, followed by a nested PCR. Each sample was analyzed in duplicated, purified with JETQUICK PCR Spin KIT (Genomed) and sequenced with BigDye Terminator v3.1 Cycle Sequencing Kit (Applied BioSystems). Specific primers for the methylation profiling are listed in **Appendix II Table 1**

2 RNA EXPRESSION ANALYSIS

Total RNA was isolated with miRNeasy Mini Kit (Qiagen) according to manufacturer's instruction, and retro-transcribed with SuperScript® III Reverse Transcriptase (Life Technologies). cDNA was amplified by qRT-PCR using LightCycler480 (Roche) with Fast Start DNA Master Sybr Green I mix (Roche) and expression results were normalized with *Gapdh* or *18S* housekeeping genes. Primer sequences are listed in **Appendix II Table 2.** Global gene expression was analyzed using SurePrint G3 Mouse 8x60K one color microarray (Agilent Technologies) at the Microarray Unit of CRG (Barcelona). Gene expression differences with log2 fold change ≥ 1.2 and with a FDR ≤ 0.01 were considered significant. Microarray data are in GEO accession number GSE63136.

4 PROTEIN EXPRESSION ANALYSIS

4.1 Immunofluorescence staining of satellite cells

Protein expression was assessed with immunofluorescence staining. After FACS sorting of quiescent and activated satellite cell populations, 25,000-50,000 cells were resuspend in 100ul PBS, loaded in a cytospin cone and centrifuge at 600 rpm for 3min in order to obtain a thin-layer preparation on a slide. The cells were fixed with 3.7% Formaldheyde (Sigma-Aldrich) for 10 minutes at room temperature. After removing the Formaldehid and let the preparation dry for 2 minutes and wash it three times with PBS, the samples were incubated for 30 minutes at room temperature with the blocking solution (Goat serum 15% in 0,01% Triton (Life technologies)). Then, cells were incubate with the primary antibodies (Ki67 (1:500, Abcam) or MyoD (1:250, clone MoAb 5.8A, BD Biosciences)) diluted on blocking solution for 1 hour at room temperature. After washing with PBS Tween (0.02%), Alexa fluor 647 goat-anti- rabbit (1:250, Life Technologies) was used as secondary antibody and incubated for 45 minutes at room temperature. After washing, preparations were mounted with Vectashield DAPI Fluorescence Mounting medium (Vector Labs) to counterstain nuclei.

4.2 Immunofluorescence staining of infected iPax7 myoblast precursors

Cells cultured on slides were fixed using 4% Paraformaldehyde, permeabilized with 0.1% Triton X-100 (Life technologies) for 10 minutes and blocked with 5% BSA reagent (Vector Labs) for 30 minutes, and then incubated with primary antibodies including Myogenin (1:250, clone F5D, BD Biosciences) and MHC (1:20, Developmental Studies HybridomaBank,. Alexa fluor 555 goat-anti-rabbitor (1:500, Invitrogen) was used as secondary antibody incubated for 45 minutes at room temperature. Preparations were mounted with Prolong mounting media with DAPI (Life Technologies) to counterstain nuclei.

5 CHROMATIN IMMUNOPRECIPITATION ASSAY

Chromatin immunoprecipitation (ChIP) assay coupled to quantitative PCR analysis is a robust technique for studying protein-DNA interactions (**Figure M4**). Immunoprecipitation assays were done using Magna ChIP TM A/G Assay Kit (Millipore) following manufacturer's protocol. Briefly, cultured cells were fixed in 1% Formaldehyde for 10 minutes at room temperature. Crosslinking reaction was terminated adding 0.125 M glycine for 5 minutes, and washing with ice-cold PBS containing protease inhibitors (1mM PMFS, 1µg/ml Aprotinin and1 µg/ml Pepstatin). Next, cells were lysated on ice with Lyse Buffer for 15 minutes, centrifuged and resuspended on nuclear lysis buffer. The chromatin was sheared into 200-500 bp fragments using standard protocol for a Covaris S2 sonicator. After checking the sonication level, the insoluble material was discarded by centrifugation and soluble chromatin was quantified using Nanodrop equipment. 300 µg of chromatin were used for each IP with Usf-1 antibody (C-20X, sc229X, Santa Cruz) and rabbit IgG as negative control (Millipore). The amount of immunoprecipitated chromatin and input material were measured using LightCycler480 (Roche) with Fast Start DNA Master Sybr Green I mix (Roche) with the primers listed on the **Appendix II Table 3**. The values were normalized to the chromatin input.

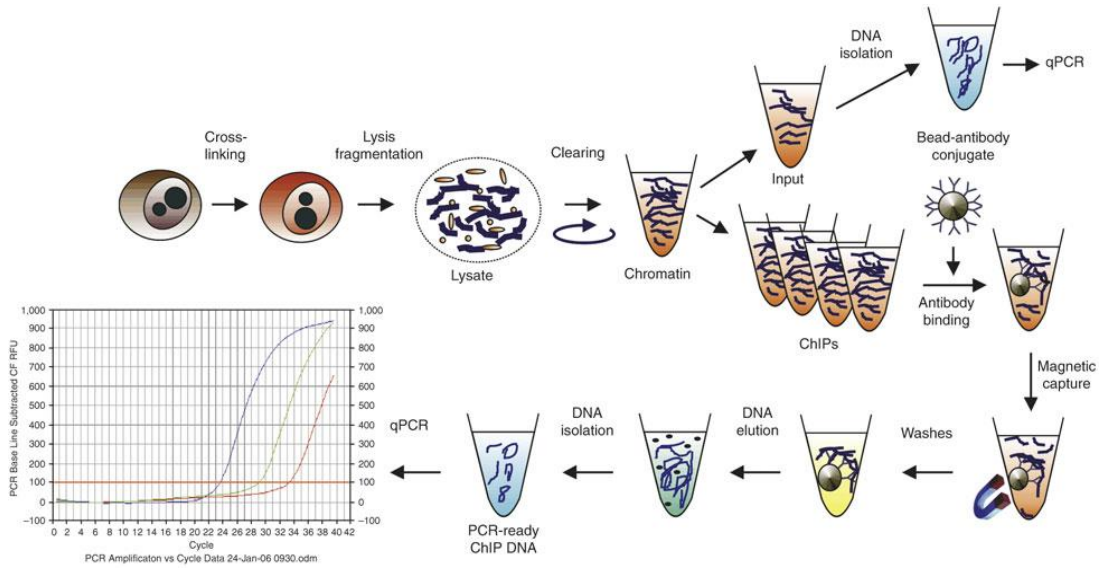


Figure M4 ChIP Assay(Dahl and Collas 2008)

6 BIOINFORMATICS ANALYSIS

6.1 Genomic analysis of publicly available ChIP-seq data sets

AIMS-Seq amplicons were annotated according to the overlap of the two flanking XmaI sites with selected genomic elements in the NCBI37/mm9 assembly. Genomic elements were downloaded from (http://hgdownload.soe.ucsc.edu/goldenPath/mm9/database/, the mm9 annotation database at the UCSC browser. We downloaded the public domain ENCODE data sets and correlated them with BED tools to the AIMS seq amplicons (Quinlan and Hall 2010). ENCODE ChIP-seq data for MB/MT and ESC were obtained from Dr. Hardison's lab and Dr. Ren's lab, respectively. Usf1 ChIP-seq data on C2C12 cells were from Myers' lab (ENCODE project) and H3K27Ac and P300 ChIP-seq data on C2C12 cells were from Dynlatch's lab (Blum et al, 2012). GEO accession numbers of the data are listed at **Appenidx II Table 4.**

6.2 Transcription factor binding motif analysis

Transcription factor motif enrichment of differentially methylated regions was performed by Ildefonso Cases (IMPPC, Barcelona) using transcription factor binding matrixes obtained from three databases UniPROBE mouse (Newburger and Bulyk 2009), JASPAR vertebrates (Mathelier *et al.,* 2014) and Jolma 2013 databases (Jolma *et al.,* 2013) and applying Fisher test corrected by Bonferroni.

6.3 Visualization tools

All genomic representations containing Reference genes, CpG islands and ChIP-seq data were integrated, explored and visualized using Integrative Genomics Viewer (Robinson *et al.,* 2011). Sodium bisulphite sequencing data where represented with Methylation plotter web tool http://gattaca.imppc.org:3838/methylation_plotter/,

(Mallona *et al.,* 2014). Aims-seq data were explored and visualized using our own R scripts and specific R packages: Gplot, MASS and Vioplot.

6.4 Statistical analysis

All stadistical test analysis were done with R language. Kruskal-Wallis Test of sodium bisulphite data were done with Methylation plotter web tool http://gattaca.imppc.org:3838/methylation_plotter/ (Mallona *et al.,* 2014).

RESULTS

1 GENOME-WIDE ANALYSIS OF DNA METHYLATION DYNAMICS DURING MYOGENESIS

In this chapter, we expose the results obtained by applying the Amplification of intermethylated sites followed by ultrasequencing (AIMS-seq) method to characterize the DNA methylation dynamics occurring during the myogenic lineage progression. Briefly, we established the *ex-vivo* myogenic model, analyzed it with AIMS-seq method and performed a first analysis classifying Positive and Negative AIMS-seq regions and a second analysis identifying differentially methylated regions (DMRs) between samples. Next, to fulfill the epigenetic landscape of this model and to evaluate the functional implications of the DMRs we correlated them with microarray expression data of the same samples generated in our lab and with histone marks and enzyme occupancy obtained by ChIP-seq from the ENCODE project and Dynlacht's lab. Finally, we analyzed in detail the most remarkable AIMS-seq DMRs in order to identify the regulatory role of DNA methylation at those regions.

As a whole, reduced but significant DNA methylation variations at CpG-poor regions occurred during myogenic progression, mainly during myogenic lineage commitment. Gain of DNA methylation was the predominant change, although hypomethylation events enriched at enhancer-type chromatin also took place during the myogenic process, suggesting the involvement of DNA methylation in the regulation of cell-type specific enhancers.

Importantly, one of the DMRs indentified by AIMS-seq was located within the super-enhancer region of the master transcription factor Myf5, which became demethylated only in muscle cells. Moreover, our results demonstrated that the binding of the transcription factor Upstream stimulatory factor 1 (Usf1) to *Myf5/Myf6* locus was methylation-dependent occurring upon DNA demethylation in myogenic cells.

Taken all together, this study provided a comprehensive picture of genome-wide DNA methylation dynamics during myogenic progression and identified regulatory regions

orchestrating myogenic cell fate and terminal differentiation. By the end of the writing of this thesis, the results exposed in the following chapter were under revision for publication at Stem Cells journal.

1.1 Establishment of the *ex-vivo* myogenic model

To study the myogenic progression we established a differentiation model based on four representative stages of the myogenic process: embryonic stem cells (ESCs), myoblasts (MBs), myotubes (MTs) and myofibers (MFs) (**Figure R1.A**). This model allowed us to address the changes occurring during the cell fate determination (ESC-MB), the myogenic early differentiation (MB-MT) and the terminal differentiation (MT-MF) by comparing the successive stages of the process.

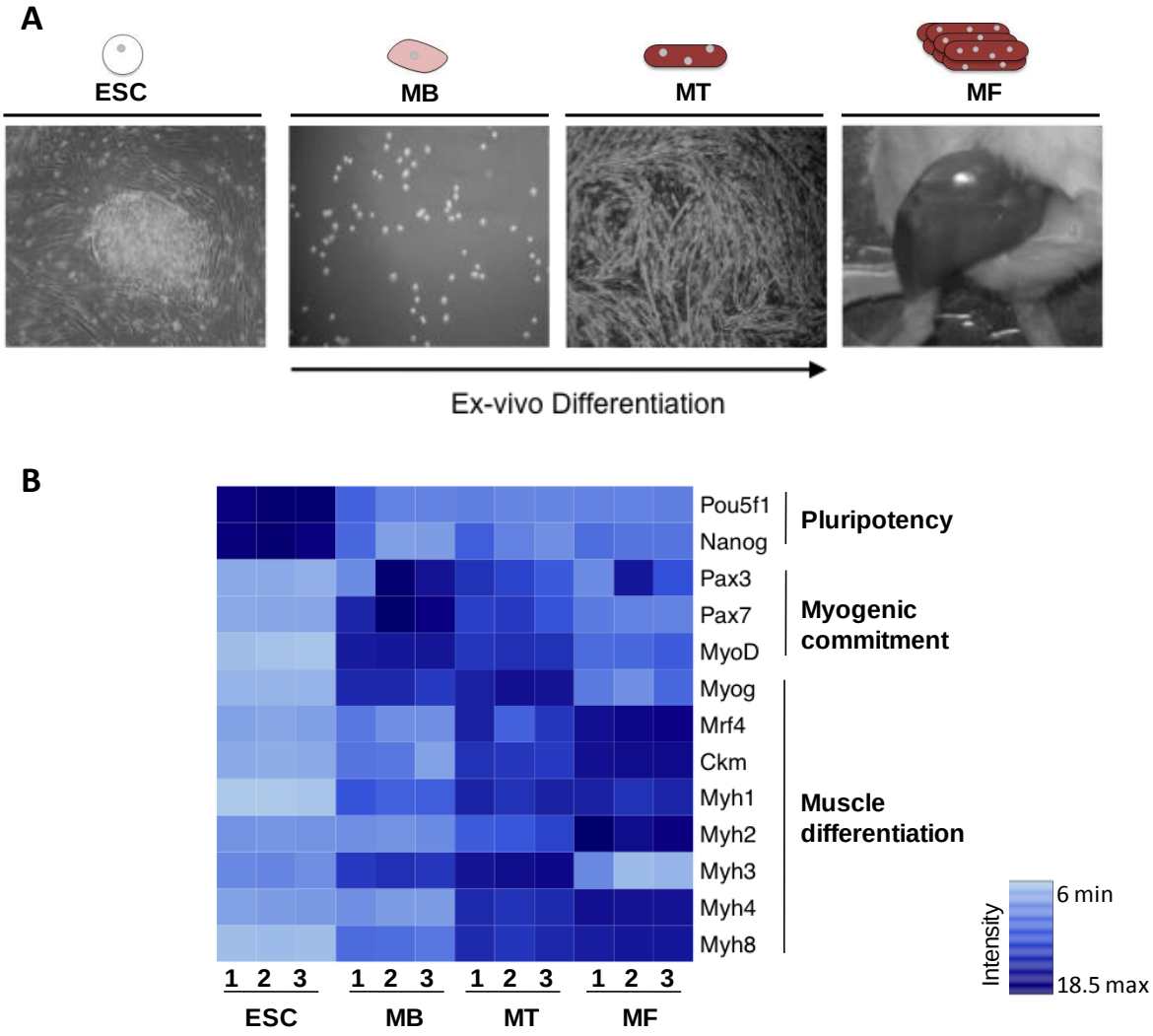


Figure R1. The myogenic model. A. Representative fields of the four stages of the model. Culture pictures were 20x amplified **B.** Heatmap of the intensities of representative differentiation markers at the successive points of the myogenic model obtained by expression microarray.

Summing up, we isolated primary satellite cells from hind legs of adult male mice and amplified them *in vitro* to generate the MBs population (protocol is described in detail in Material and methods section). After several amplification passages, we plated the MBs in high confluence and let them to differentiate by reducing drastically the serum concentration and by eliminating the bFgf of the medium. Consequently, the mononucleated MBs fused each other forming plurinucleated contractile MTs. As start and end differentiation stages of the myogenesis we brought into comparison pluripotent embryonic stem cell lines (RD1, E14Tg2 and cGR8) and adult skeletal muscle tissue, formed mainly by mature myofibers.

We generated biological triplicates of the samples and ensured by microarray expression analysis the expected expression profiles of several differentiation markers at every differentiation stage, prior to proceeding to the DNA methylation analysis. Specifically, we analyzed the expression of embryonic specific genes (*Pou5f1* and *Nanog*), myogenic lineage commitment genes (*Pax3* and *Pax7*), myogenic master regulatory genes (*MyoD, Myf5, Myf6* and *Myogenin*) and late myogenic differentiation genes (*Ckm, Mh1, Myh2, Myh3, Myh4* and Myh8) (**Figure R1.B**).

1.2 AIMS-seq as a method for identifying methylation signatures at genome-wide scale

We adapted the AIMS method, developed in our laboratory (Frigola *et al.,* 2002), to the next-generation DNA sequencing technology in order to assess non-targeted DNA methylation patterns during myogenesis. This methyl-sensitive restriction enzyme based (MSRE) method combines the use consecutive of two neoisoschizomers enzymes, SmaI and XmaI, which recognize the same cleavage site (5'-CCCGGG-3') but present different methylation sensitivities, and generate distinct ends: SmaI digests unmethylated DNA leaving blunted end-fragments, whereas XmaI digests regardless of the DNA methylation state and generates fragments with cohesive ends (**Figure 2A**). Following genomic DNA digestion with both enzymes, cohesive adaptors were ligated

only to XmaI products, allowing the selective amplification of intermethylated fragments by conventional PCR. Next, the amplified products were next generation sequenced, and the high quality reads were mapped to the mm9 mouse genome assembly. The ambiguously mapped reads and the ones mapping outside the established PCR criteria limits (20 pb to 2 kb) were discarded, reducing the 25 millions of reads per sample initially obtained to 10 millions of reads used to perform this study (**Figure R2.B**).

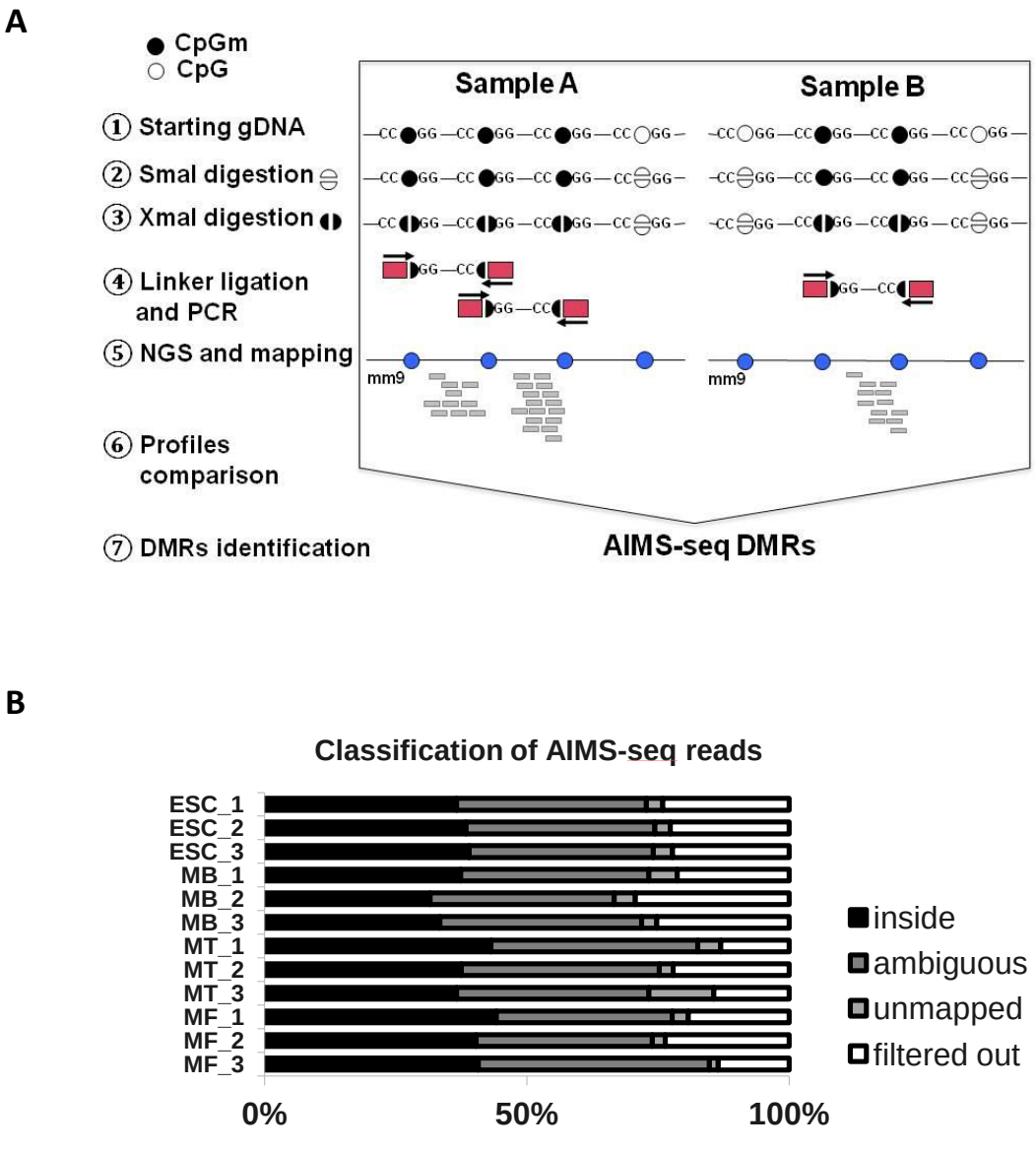


Figure R2. AIMS-seq method. A. AIMS-seq workflow scheme. The method is based on the use of methylation sensitive enzymes followed by adaptors ligation, PCR amplification and ultrasequencing. Mapped reads reflect the DNA methylation profile of the sample and the comparison of the profiles allows identifying DMRs between samples. **B**. Classification of the obtained AIMS-seq reads after filtering by quality ("filtered out"), mappability ("unmapped", "ambiguous") and amplicons size criteria ("inside", amplicon length range 20 bp to 2,000 bp).

AIMS-seq method interrogated 77,307 methyl-sensitive restriction enzyme cleavage sites that represented 50,745 amplicons, the majority of them smaller than 250 bp (average of 236 bp) (**Figure R3.F**). These regions were widely distributed across the mouse genome and covered all the chromosomes (**Figure R3.A, Appendix I Table 1**).

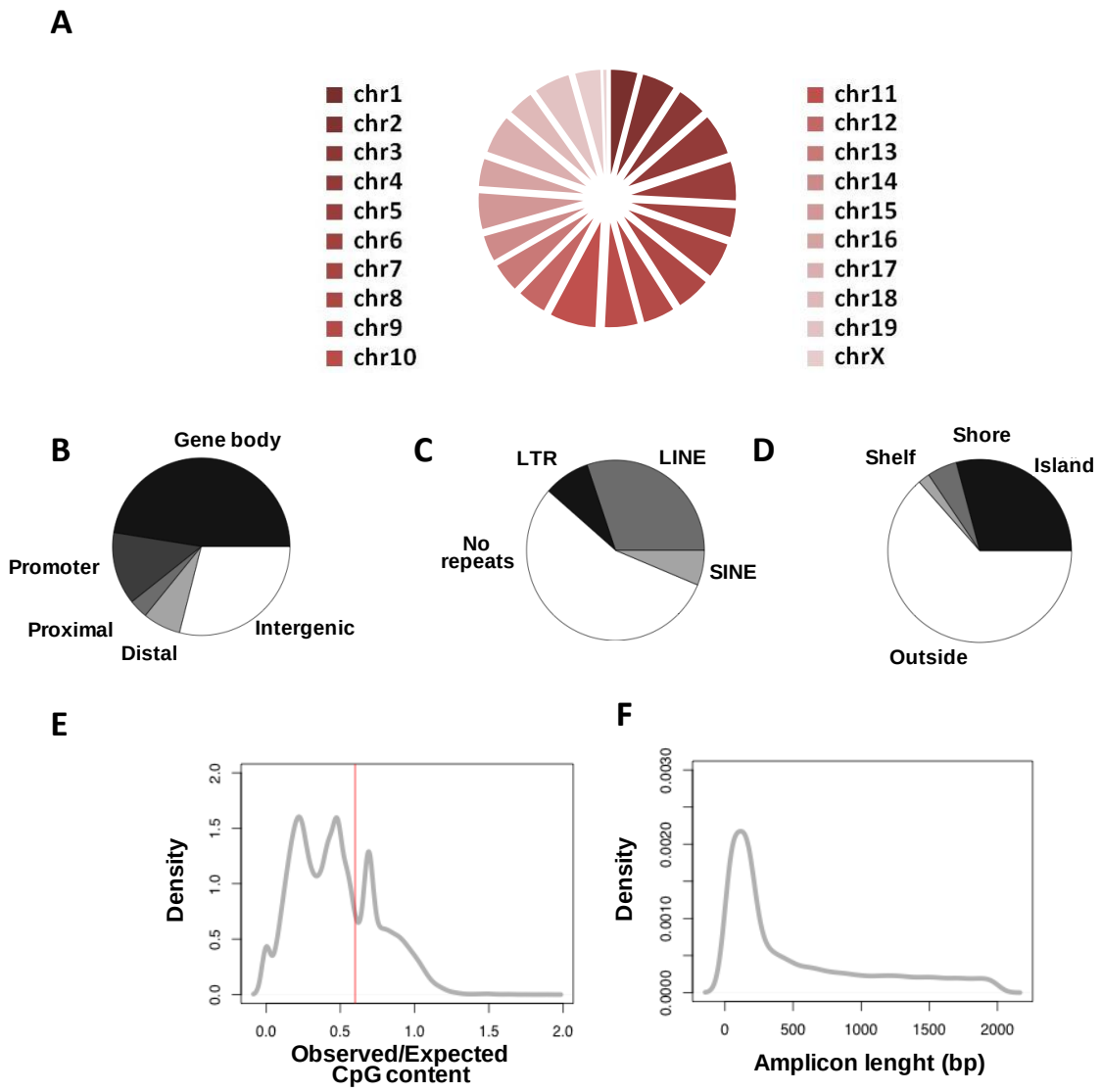


Figure R3. AIMS-seq virtual amplicons features. Classification according to their genomic distribution relative to chromosomes (**A**), to genes (**B**), to repetitive elements (**C**) and to CpG islands (**D**) of mm9 mouse genome assembly. **E**. Density plot of amplicons according to the Observed/Expected CpG content. As a reference, CpG island threshold of observed/expected CpG content is represented with a red line. **F.** Density plot of amplicons classification according to their base pair length.

A mutually exclusive classification of the AIMS-seq amplicons relative to the closest annotated gene showed that 47.5% amplicons were located in gene bodies and 13% at gene promoters (5'UTR – 2 kb upstream the TSS) surveying 11% of the mm9 Ensemble mouse genes. Interestingly, over 10% of the amplicons were located at regulatory regions (3.5% in proximal regulatory regions from 2kb to 10 kb upstream of the TSS and 7% at distal ones between 10 to 50 kb upstream of the TSS). The remaining 29% of the AIMS-seq amplicons mapped at intergenic regions, defined here as the regions downstream of a gene and the regions farther than 50 kb upstream of the TSS (**Figure R3.B**). Furthermore, the permissive distribution of the MSRE-based methods allowed the study of DNA methylation patterns at repetitive regions where DNA methylation may play a regulatory role by impairing the mobility of the transposons and therefore ensuring the genome stability (Yoder *et al.,* 1997; Bestor and Bourc'his, 2004). Here, we simplified the classification of repetitive elements in three main categories LINEs, SINEs and LTRs, where almost 50% of the AIMS-seq amplicons were located (**Figure R3.C**).

Considering the CpG island (CpGi) coverage of our method, 44% of annotated CpGi in the mouse genome (UCSC genome browser mm9) were interrogated by AIMS-seq, representing the 29% of the total amplicons. Whereas CpGi are regions typically devoid of DNA methylation, CpGi vicinity methylation levels have been shown to be especially dynamic during tissue specific differentiation (Doi *et al.,* 2009; Irizarry *et al.,* 2009). Amplicons located at CpGi shores (defined as the CpG island adjacent regions up to 2kb distance (Irizarry *et al.,* 2009) and CpG island shelves (defined as the regions from 2 kb to 4 kb distance at both sides of the CpG islands (Kulis *et al.,* 2012) represented the 5% and the 2% of all the amplicons, respectively. The regions more distant than 4 kb from a CpGi were named as "Outside CpGi regions" standing for the majority of the AIMS-seq amplicons (63%) (**Figure R3.D**). In order to obtain a more complete annotation of the AIMS-seq amplicons, we calculated the CpG observed/expected content of each amplicon and we distributed them according to this value. This classification revealed a multimodal distribution of the AIMS-seq amplicons in low, intermediate and high CpG content regions, coinciding with the

three types of promoters defined as HCP, ICP and LCP (High, Intermediate and Low content promoter) (Weber *et al.,* 2007), and ensuring their coverage (**Figure R3.E**).

To address the robustness of the method, three biological replicates were analyzed. We performed pair comparison between the three ESC cell lines (RD1, E14tg2 and cGR8) and the distribution was very similar in the three cell lines, with an adj-R square higher than 0.9 in all comparisons (p-value = 0.001, Chi squared test) **(Figure R4).**

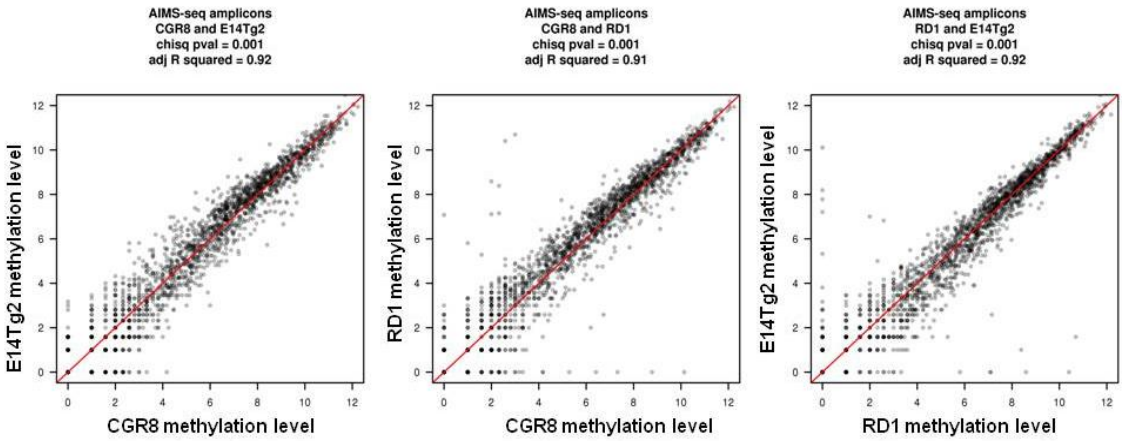


Figure R4.Biological replicates comparison. MA plot representing the comparison of the number of reads of 2,000 amplicons in three different ESC lines: cGR8, E14Tg2 and RD1. Lineal regression results (R^2) and Chi square test p-values are indicated at the top.

To test the reliability of the method, we compared the results obtained in ESC with AIMS-seq technique and the results obtained with RRBS methodology (Reduced Representation Bisulphite Sequencing) by Eric S. Lander's Lab (Meissner *et al.,* 2008). RRBS represents a robust method to interrogate methylation profiles at genome wide scale since it combines sodium bisulphite conversion after MSRE digestion with next-generation sequencing. Concretely, we selected the amplicons that had at least one AIMS-seq read in one sample and that embedded at least 3 CpG interrogated by RRBS, to ensure the analysis of informative regions. Next, we assessed the methylation state of those amplicons in an ESC line with both methodologies (AIMS-seq: log2 (number of reads per amplicon), RRBS: beta values) and we distributed them according to the methylation values observing a good correlation (**Figure R5.A**).

Finally, we analyzed the linear correlation between the methylation values obtained with both methodologies. Although AIMS-seq method was designed and used exclusively as a comparative method and not as an absolute DNA methylation measurement technique, we observed a positive correlation between the log2 of the number of reads per amplicon obtained with AIMS-seq and the RRBS methylation values of the amplicon-embedded CpGs **(Figure R5.B).** The ESC line used for this comparison was cGR8 (mouse strain 129S) for AIMS-seq values and ESC line from 129Sv Jae x C57BL/6 mouse strain for RRBS values.

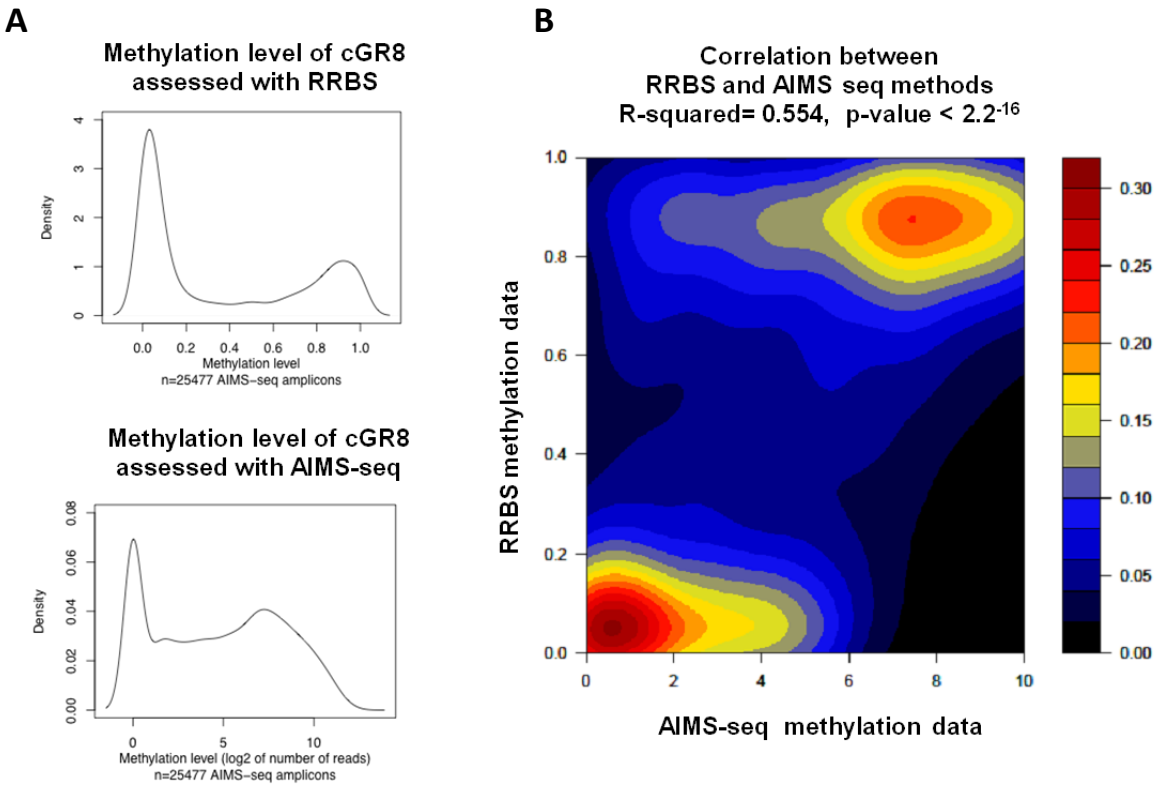


Figure R5. AIMS-seq compared to RRBS. A. Density plot of the methylation levels of ESCs measured with RRBS (top) and AIMS-seq (bottom). **B** Two dimensions Kernel density plot correlating AIMS-seq (log2 Nº of reads) and RRBS methylation values (β-values) per virtual amplicon. R^2 correlation test and Kendall p-value are shown at the top.

Taken all together, the AIMS-seq method assayed over 77,000 regions including a wide range of genomic elements, allowing a broader analysis of regions than classic methylation studies limited to high CpG density and promoter regions.

1.3 Genome-wide profiling of DNA methylation during myogenesis

We analyzed the samples described in section 1.1 with AIMS-seq method with the purpose of elucidate the DNA methylation dynamics associated to the myogenic lineage progression.

In a first analysis, we classified the genomic regions (amplicons) according to the number of mapped AIMS-seq reads. Amplicons with no reads representation in any sample were considered non-informative (Negative) and were either sequences that remained fully unmethylated in all the samples or that failed to be amplified. On the other hand, amplicons with at least five normalized reads in at least one sample were considered informative (Positive) and susceptible to be differentially methylated (**Figure 6A)**. 60% of AIMS-seq amplicons corresponded to non-informative regions showing a high and intermediate CpG density and overlapping in a 43% of cases with annotated CpG islands, whereas the AIMS-seq informative regions (40% of the amplicons) were almost exclusively found in low CpG content regions, and enriched in introns, intergenic regions, and SINES/LTR repetitive sequences (**Figure 6B-D**).

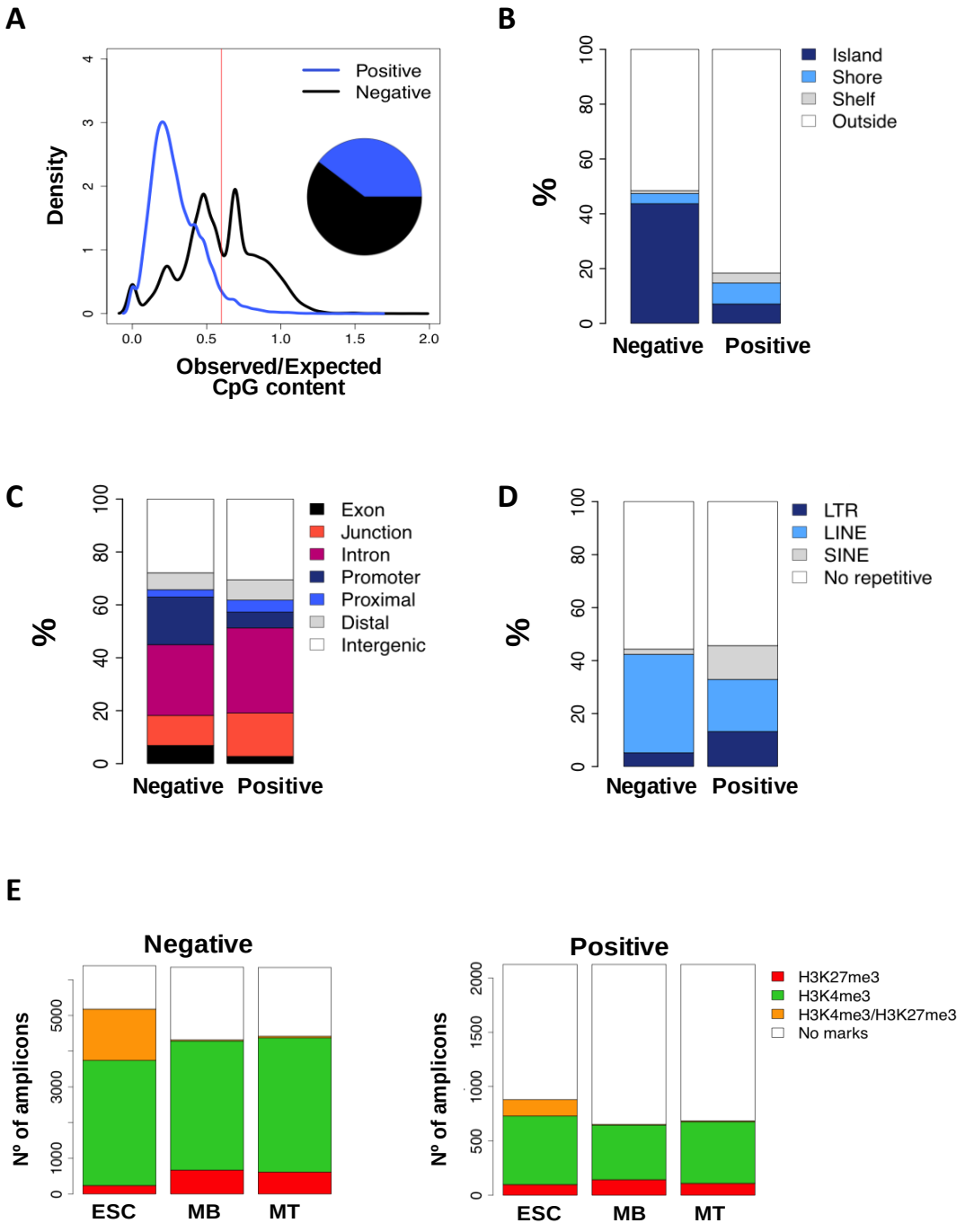


Figure R6. Aims-seq informative (Positive) and non informative (Negative) regions. **A.** Density plot of the distribution of the positive and negative regions according to the CpG Observed/Expected CpG content, the red line represent the CpG island threshold of observed/expected CpG content. The pie plot shows the proportion of AIMS-seq positive and negative of all AIMS-seq amplicons during myogenesis **B.** Positive and Negative amplicons distribution across the CpG island related classification. **C-D.** Positive and Negative amplicons distribution according to the gene-related and repetitive elements classifications. **E**. Chromatin state in ESCs, MBs and MTs Negative and Positive AIMS-seq regions.

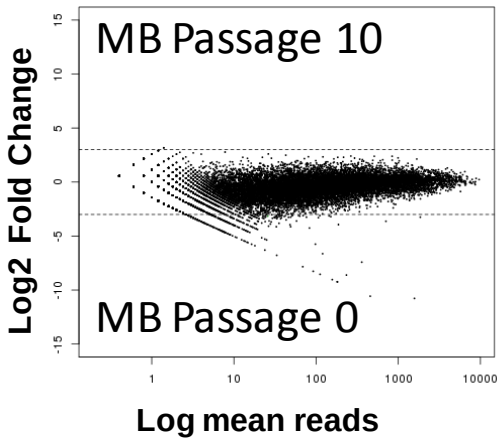
We compared the histone mark profiles of both populations (Positive and Negative) using the ChIP-seq data from ENCODE project and focusing on the bivalent marks (H3K4me3 together with H3K27me3). As shown in **Figure R6.E**, Negative regions were occupied more frequently with histone modifications than Positive regions and, remarkably, Negative regions were highly enriched in H3K4me3, despite no gene expression difference between Positive and Negative regions were observed (data not shown). Both populations contained bivalent domains in ESC that get resolved upon differentiation.

All together, these observations indicated that AIMS-negative and AIMS-positive regions represent two different genomic compartments showing different epigenetic profiles at DNA methylation and histone marks level.

1.4 Identification of differentially methylated amplicons during myogenesis

Before analyzing all the informative-AIMSseq data, we evaluated the effect of the passages in our cells. We compared the methylation profiles of primary myoblast freshly isolated with MBs cultured for up to 10 passages, without detecting significant differences between them (**Figure R7**).

Figure R7. Effect of cell passage on DNA methylation values. Scatterplot showing the correlation between the number of reads per amplicon obtained in myoblasts at passage zero and at passage 10**.** The absence of red/green dots shows that no significant differences between samples were observed.



Next, focusing on the AIMS-seq detected regions, we performed pair wise comparisons between samples (**Figure R8)**. An amplicon was considered differentially

methylated when $\log_2$ (Fold Change of number of reads) ≥ 3 (for details see Materials and Methods section 2.2). Doing Venn Diagrams of the differentially methylated amplicons of the pair-wise comparisons (**Appendix I Figure 1**), the regions showing a consistent DNA methylation change between differentiation stages were selected and represented the myogenic-associated differentially methylated regions (DMRs).

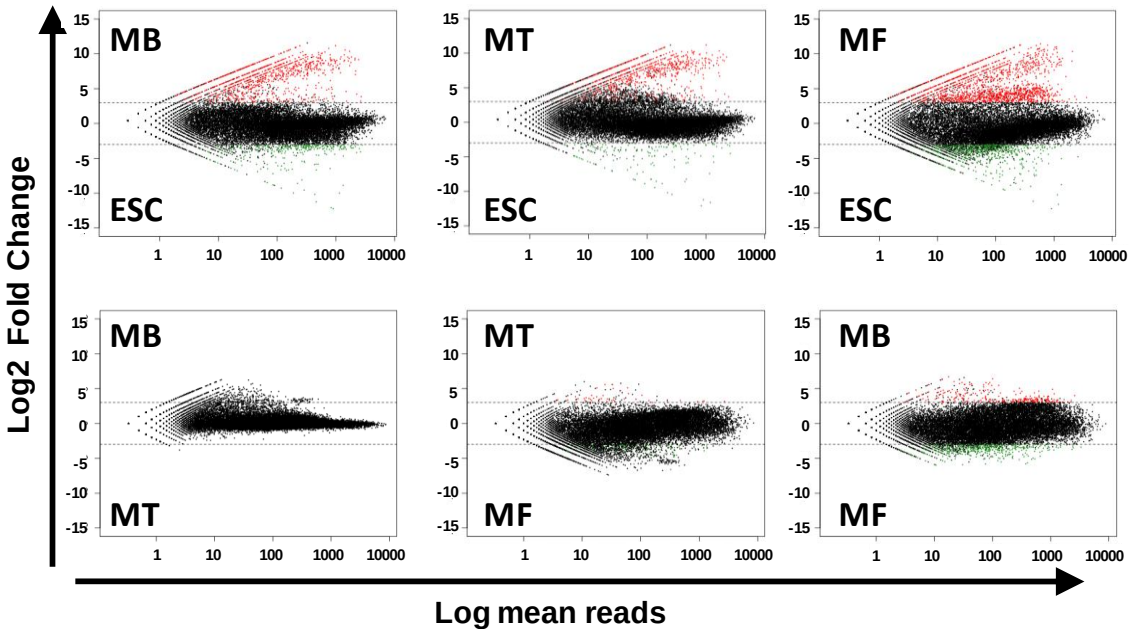


Figure R8. Detection of DMRs with AIMS-seq. MA plots of all pair-wise comparisons. Red and green dots show significant hypermethylated DMRs of the compared samples

We identified 982 DMRs corresponding to the 2% of the analyzed amplicons (**Figure R9**). Remarkably, the broad majority of the methylation changes took place during the lineage commitment step and was mainly hypermethylation events (715 out of 831 DMRs, 85.8%). Although no methylation changes were detected between myoblasts (MBs) and myotubes (MTs), notably loss of methylation was observed during the terminal differentiation process, representing at that point the 87.5% (132 out of 151 DMRs) of the methylation changes (**Figure R9.A**). Most of DMRs showed low - intermediate CpG density evidencing again the lack of methylation of the CpG islands (**Figure R9.B**). The hypermethylation changes between ESC and MB showed higher extent compared to the hypomethylation changes and the changes occurring during

terminal differentiation (**Figure R9.C-D**). Next, we compared the genomic distribution of all hypo- and hypermethylated DMRs, showing that promoters and exon-intron junctions were significantly enriched at hypomethylated DMRs (p-adj. value < 0.001, and p-adj. value < 0.012, respectively), whereas the intergenic regions were more frequently hypermethylated during the differentiation process (p-adj. value < 0.006) (**Figure R9.E**).

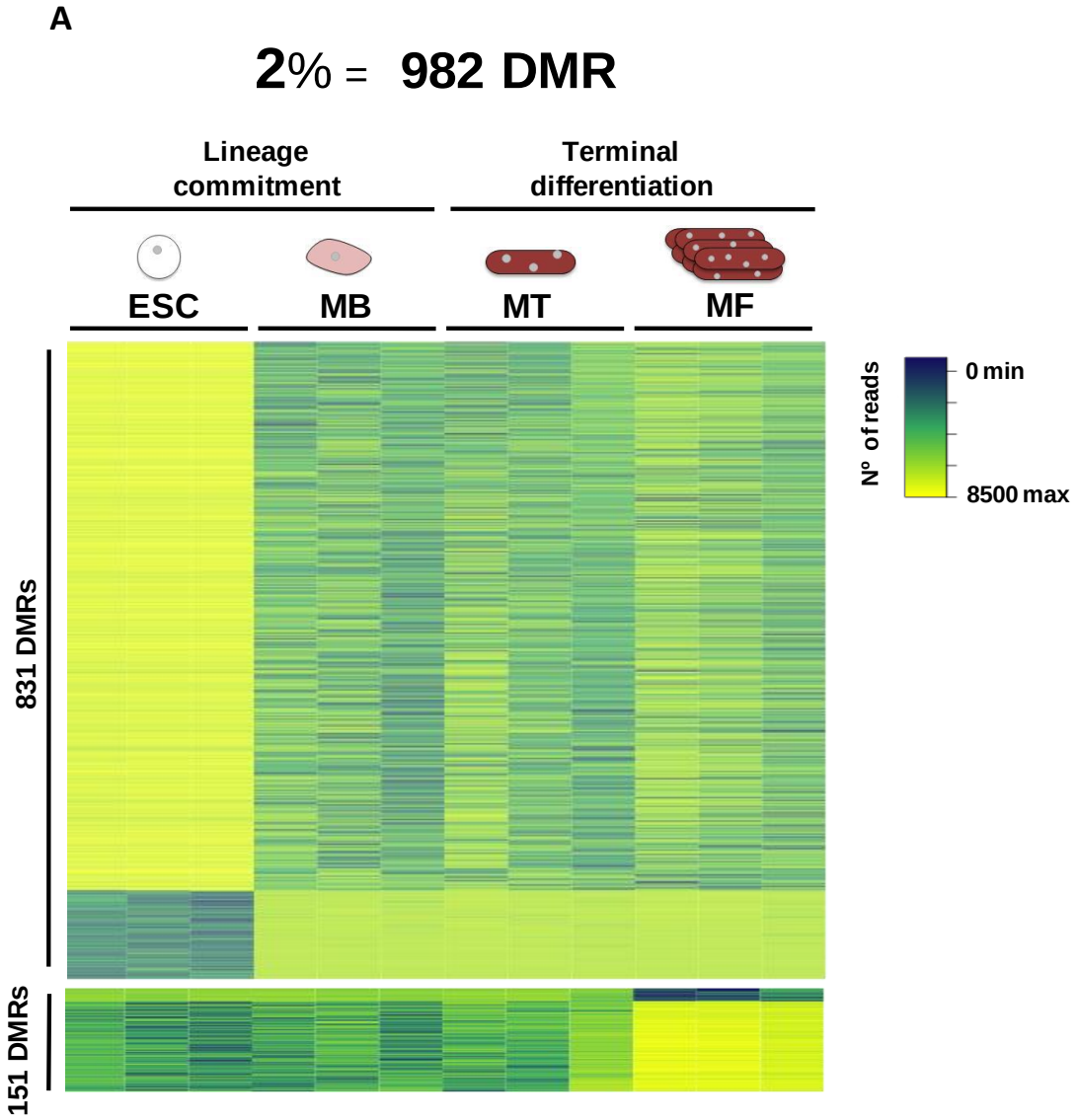


Figure R9. DMRs during myogenesis. A. Heatmap of the DMRs regions represented as number of reads per sample.

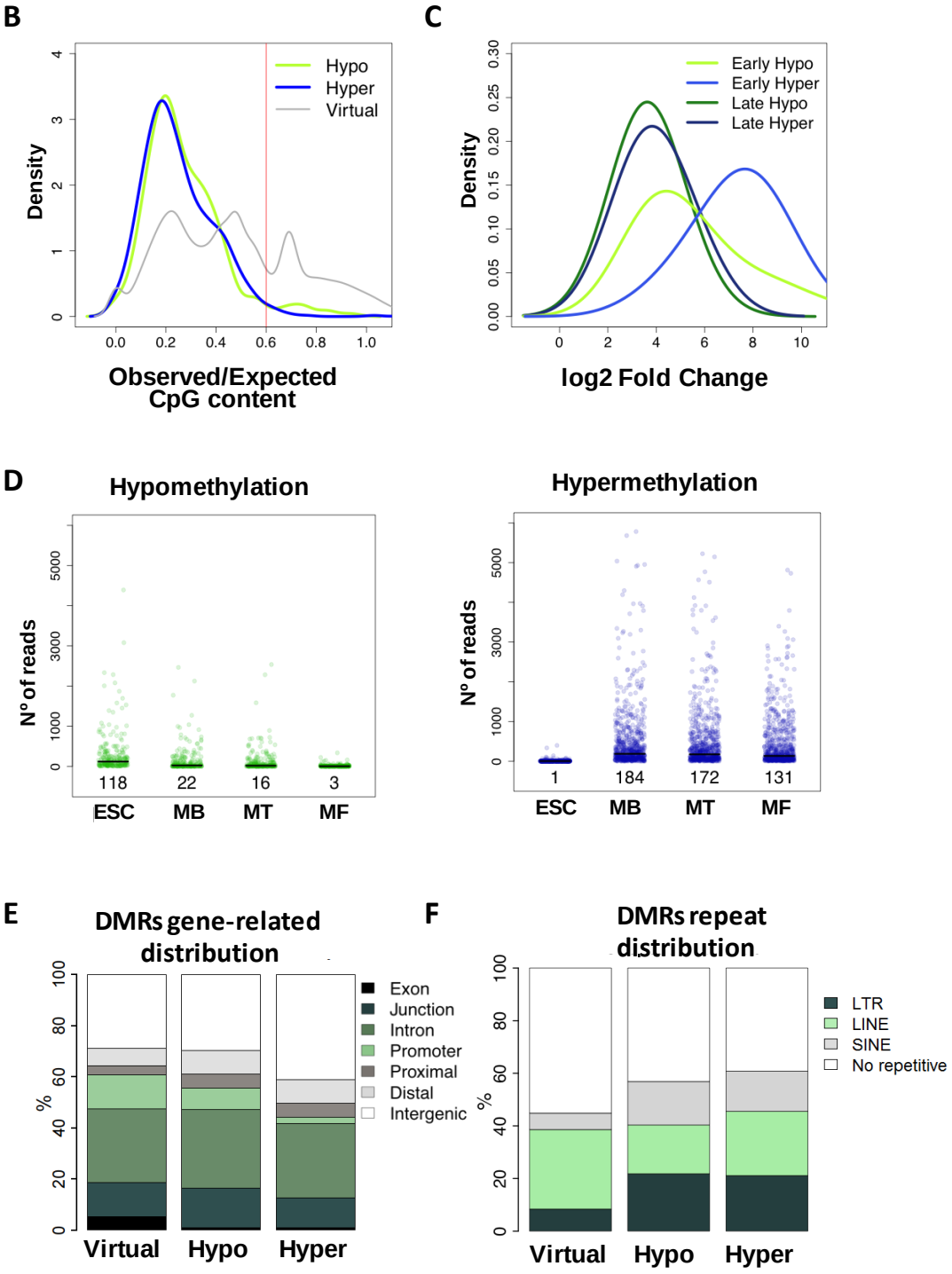


Figure R9. DMRs during myogenesis. B. Density plot of hypermethylated and hypomethylated DMRs according to the observed/expected CpG content. CpG island threshold of observed/expected CpG content is represented with a red line. **C.** Density plots showing the differences between hypermethylated DMRs and hypomethylated DMRs during lineage commitment (early) and terminal differentiation (late) according to the log2FoldChange of the reads. **D**. Stripchart of the number of reads at ESC, MB, MT and MF stage in hypomethylated (left) and hypermethylated DMRs (right). Average number of reads per samples is shown and a line is located at the indicated value. **E**-**F**. Hypo- and hypermethylated amplicons distribution depending on the gene-related classification and depending on the repetitive elements classification. As a reference, the total number of virtual amplicons in each classification is also shown.

Regarding the repetitive elements, no significant enrichment between hypo- and hypermethylation events was observed (**Figure R9.F).** For a detailed number of hypo- and hypermethylated amplicons classified according their genomic locations see **Appendix Table 3**.

1.5 Histone mark profiling of differentially methylated regions

Next, we evaluated the type of chromatin in which the DMRs were embedded taking advantage of public available ChIP-seq data. We focused on Pol II, H3K4me3, H3K27me3, H3K27Ac and p300 data to analyze active promoters (Pol II and H3K4me3), active enhancers (p300 and KH3K27Ac) and Polycomb-group repressed genes (H3K27me3). Processed ChIP-seq data from ESC were obtained from ENCODE Project (Consortium, 2012), whereas myoblasts and myotubes data were obtained from ENCODE Project (p300, H3K4me3 and H3K27me3) and, in their default, from GEO Omnibus database (H3K27Ac and p300) (Asp *et al.,* 2011; Blum *et al.,* 2012). The samples used for the ChIP-seq experiments were the ESC Bruce cell line and the murine myogenic cell line C2C12, which can be maintained in myoblast stage or differentiate into myotubes, as primary myoblasts do. We focused the analysis on the changes occurring between ESC and MB/MT, and consequently we just analyzed the DMRs occurring between them (116 DMRs hypomethyalted and 715 DMRs hypermethylated in MB/MT). When the ChIP-seq peak was found at less than 5 kb distance from the DMRs, they were considered to be targeted by a chromatin mark.

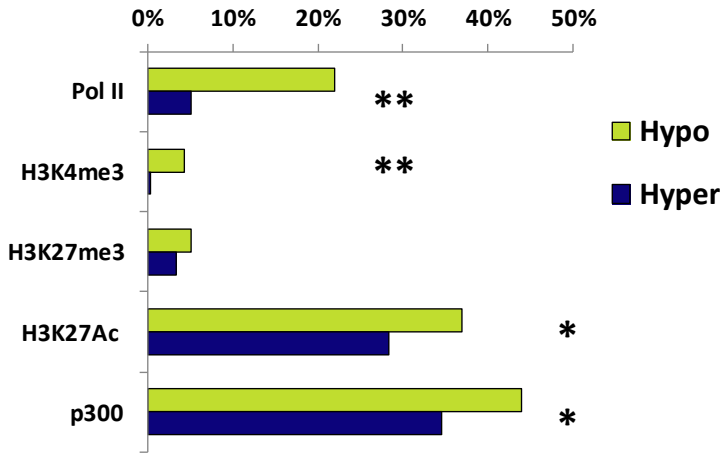


Figure R10. Percentage of DMRs that gain histone modifications or enzymes occupancy during myogenic lineage commitment. Statistical significance of the proportion test is indicated with * for p-value < 0.05 and ** for p-value <0.01 (Proportion test).

A general overview showed that most of DMRs were marked by at least one of the analyzed marks in ESCs, MBs or MTs (85% hypo DMRs and 70% hyper DMRs), highlighting the importance of epigenetic regulation on those regions. Notably, hypermethylated DMRs proportionally presented twice the number of regions without any mark than hypomethylated DMRs. This suggested that hyper methylated DMRs were less regulated by other chromatin marks than demethylated regions due to the robust silencing mediated by DNA methylation. In the other hand, hypomethylated regions might encoded for regulatory elements that need to be further fine-tuned by histone modifications.

In line with this hypothesis and focusing on the specific marks acquired upon myogenic commitment we observed that hypomethylated regions gain the active enhancer marks (H3K27Ac and p300) and the active promoter marks (Pol II and H3K4me3) in a significant larger extend compared to the hypermethylated regions (**Figure R10**). As it might be expected, enhancer marks were found more often than active promoter marks in both, hyper- and hypomethylated DMRs, correlating with the genomic distribution of the DMRs, mainly located at intergenic and intronic regions rather than in promoter regions. H3K27me3 modification was found at both hypo- and

hypermethylated DMRs regions too. The gain of this silencing mark at hypomethylated regions could be explained by the fact that some demethylated regions became silenced by the histone modifications until the right transcription time.

Summing up, the histone code observed in DMRs might point out that demethylated DMRs might be enriched for enhancer regions.

1.6 Transcription factor binding motifs identification in differentially methylated regions

To gain insight into the role of the detected DNA methylation changes, transcription factor binding sites (TFBS) were compared using UniPROBE mouse (Newburger and Bulyk, 2009), JASPAR vertebrates (Mathelier *et al.,* 2014), and Jolma 2013 databases (Jolma *et al.,* 2013).

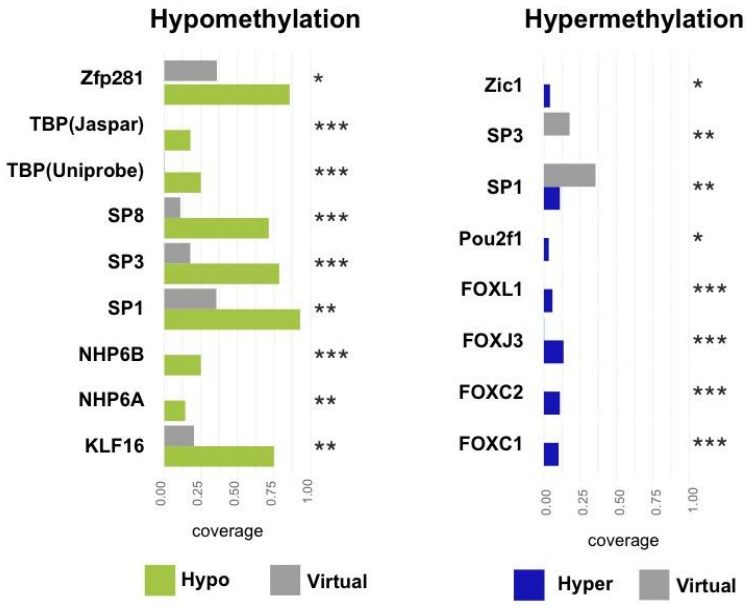


Figure R11. TFBS enrichment of DMRs. Comparison of hypo- and hypermethylted DMRs relative to the virtual amplicons. Statistical significance of Fisher test corrected by Bonferroni is indicated with * for p-adj. value < 0.1, ** for p-adj. value < 0.05 and *** for p-adj. value < 0.01.

As shown in **Figure R11**, hypermethylated DMRs presented significant enrichment of Forkhead transcription factors (p-adj value > 0.01), described as critical regulators of cellular fate determination, proliferation and differentiation, whereas hypomethylated DMRs were significantly enriched in Tata box binding protein (TBP) and Krüppel-like factor (KLF) proteins, as Sp1 and Sp3 among them. This binding regions are involved in gene expression regulation, since they represent core promoter binding motifs and TFBS relevant for muscle development, respectively (reviewed in Haldar *et al.,* 2007).

1.7 Correlation between DNA methylation and gene expression during myogenesis

In order to analyze the interplay between DNA methylation and gene expression changes we performed microarray gene expression analysis of the same samples. The unsupervised principal component analysis (PCA) of the expression data clustered clearly the four compared differentiation states, and interestingly, the PCA of the DMRs showed a tendency across the myogenic progression, pointing out that DNA methylation profiles might distinguish differentiation states (**Figure R12.A**).

To evaluate the correlation between gene expression and DNA methylation, we linked each DMR to the closest annotated gene. Genes linked to hypomethylated DMRs during lineage commitment were considered up-regulated if they increased their expression at any time of the myogenesic process, in order to be able to detect the cases where demethylation precedes gene expression activation. We observed that the majority of the DMRs were linked to genes without significant expression changes, in line with the results reported in other models (**Figure R12.B**) (Mohn *et al.,* 2008; Bock *et al.,* 2012). However, 32% of the DMRs correlated with gene expression changes (**Figure R12.C**) and notably 19% of DMRs showed a negative correlation with gene expression changes along myogenic progression (Fold Change ≥ ±1.2 and p-adj. value < 0.01) (**Figure R13 and R16**)**.** Interestingly, the negative correlation (19%) was higher than the positive correlation (13%) suggesting that DNA methylation could

negatively influence gene expression (**Figure R12.D**). To gain a more complete picture of the relationship between DMRs and gene expression changes we classified the correlated DMRs according to their genomic location. We observed the same correlation in all the genomic compartments.

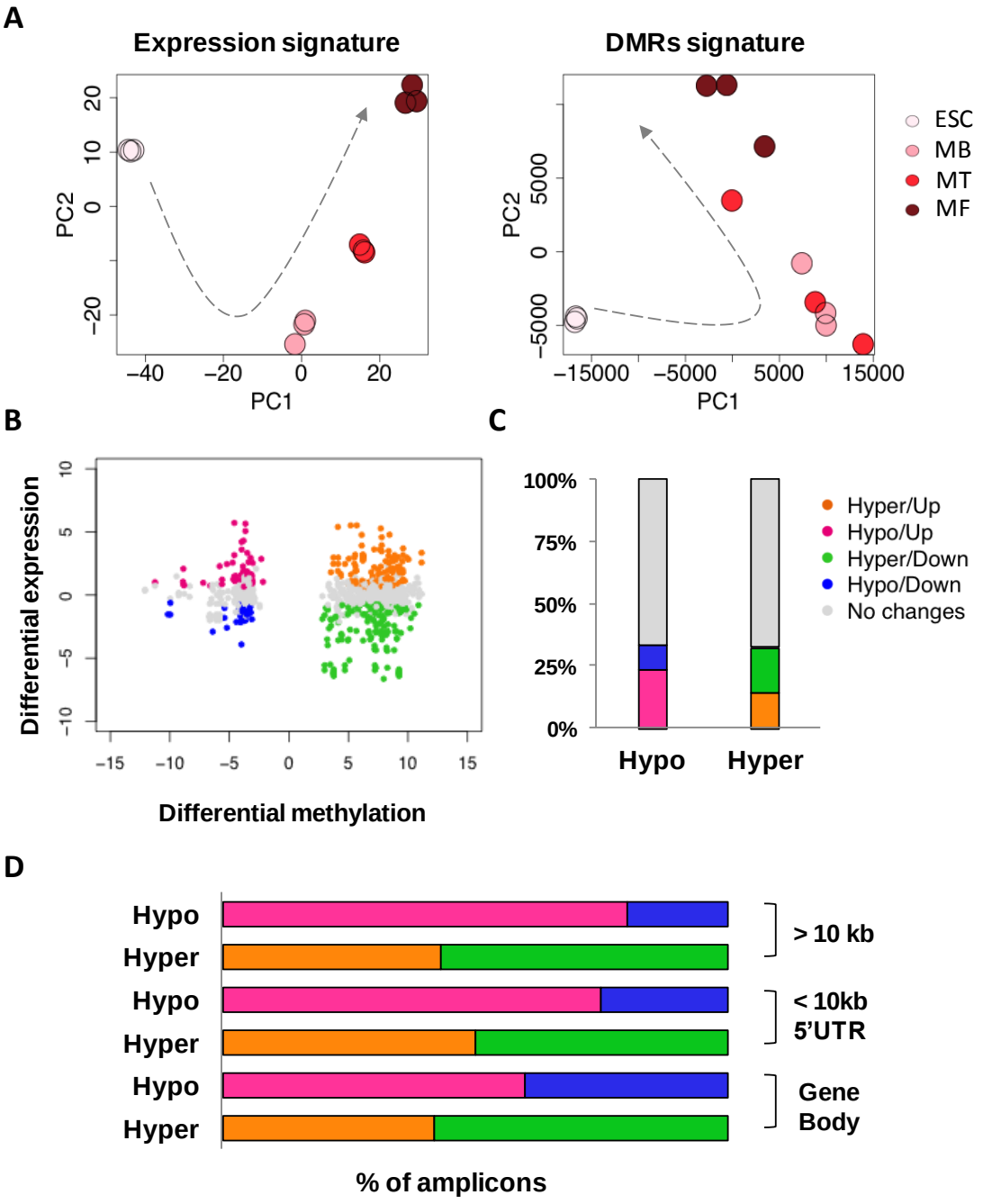


Figure 12. DMRs related to gene expression. A. PCA of the expression data (left) and PCA of the methylation data of the DMRs (right). **B.** Scatterplot of DMRs linked to expression changes occurring during the same differentiation stages. The four highlighted subsets represent the DMRs linked to significant expression changes, whereas the grey dots represent the DMRs linked to no variable probes. **C.** Percentage of up- and down- regulated and unchanged genes linked to DMRs. **D.** Percentage of up- and down- regulated and unchanged genes linked to DMRs and classified according their genomic location. Figure B,C, and D share the same color code, specified in the legend of figure C.

1.7.1 Hypermethylated genes during myogenesis

Many *de novo* methylated and silenced targets were lineage-specific factors and pluripotency-related genes (**Figure R13**). Among them, Rnf165 and Tox3 genes are essential for neuronal development (Kelly *et al.,* 2013; Tessema *et al.,* 2012), Rasip1 plays an essential role in vascular development (Xu and Chong, *et al.,* 2009), and the Zinc finger protein 423 (Zfp423) is a key initiator of adipogenic differentiation (Huang *et al.,* 2012). Regarding pluripotency-related genes, *MYBL2* contributes to the maintenance of pluripotent stem cell identity (Zhan *et al.,* 2012), *TRIML1* is expressed in embryos before implantation (Tian *et al.,* 2009), and *Dppa4* (Developmental pluripotency-associated 4) is a well studied pluripotency associated oncogene highly expressed in ESC and silenced upon lineage commitment (Tung *et al.,* 2013).

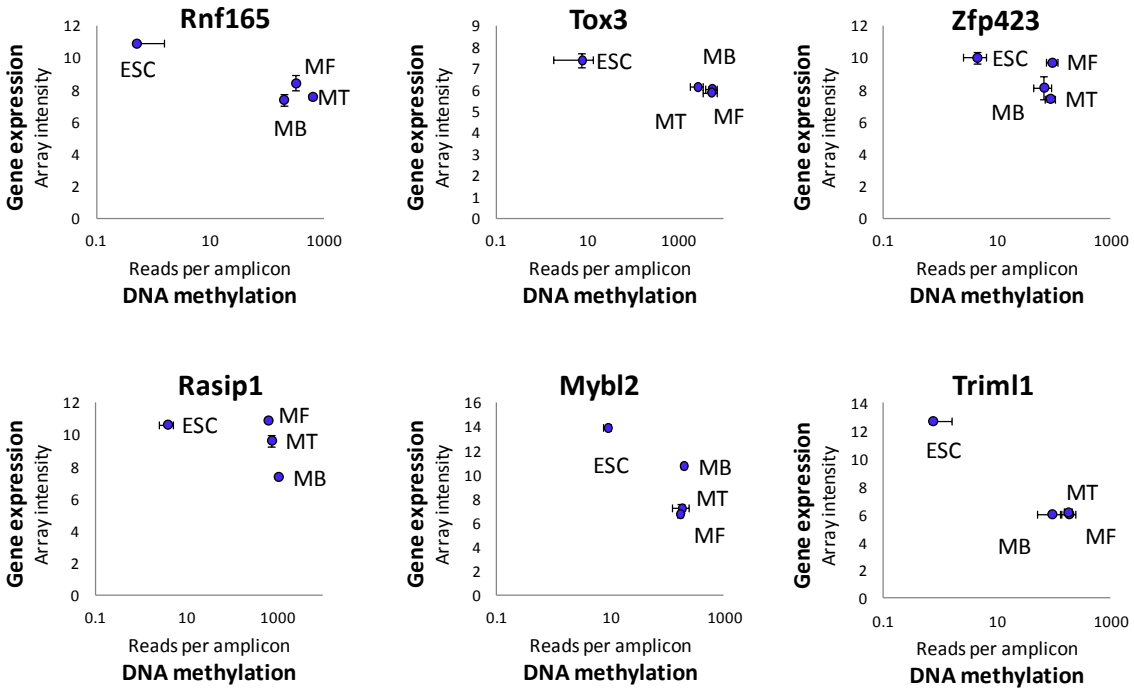


Figure R13. Gene expression downregulation upon lineage commitment hypermethylation. Scatterplot representing the DNA methylation values (reads per amplicon, log10 scale) related to the gene expression values (intensity) in ESCs, MBs, MTs and MFs of selected genes. The average of biological triplicates and the corresponding error bars are shown.

We analyzed the DMR located in the first exon of *Dppa4* gene (ID: 145136, chr16:48284157-48284241) in several non-pluripotent cell lines by sodium bisulphite conversion, and in all of them *Dppa4* was totally methylated, which correlated with the gene silencing observed in the differentiated muscle, cardiac, fibroblast and neuronal cells **(Figure R14.A).** Interestingly, among the hypermethylated DMRs, we found the amplicon ID: 30973 (chr3: 127233361-127234562) located ~5 kb downstream of the *Larp7* gene, which is expressed during all the myogenic progression. This caught our attention into the *miR302/367* cluster, located in an intron of *Larp7* gene. This cluster has been reported not only to be highly expressed in human and mouse ESC and during early embryogenesis (Suh *et al.,* 2004; Chen *et al.,* 2007; Card *et al.,* 2008), but also to promote somatic iPS reprogramming in the presence of the Yamanaka factors (Subramanyam *et al.,* 2011), and to directly reprogram both primary human and mouse cells into iPS cells (Anokye-Danso *et al.,* 2011). In fact, Oct4, Sox2 and Nanog have been shown to bind to the *miR-302* promoter region in human and mouse ESCs (Card *et al.,* 2008; Loh *et al.,* 2006). In order to address whether DNA methylation could regulate *miR-302/367* expression, we analyzed the methylation levels of the promoter region where Oct4, Sox2 and Nanog bind by sodium bisulphite conversion. As shown in **Figure R14.B,** all the CpGs inside the 1 kb promoter region were totally unmethylated in ESC, whereas they were fully or partially methylated in all differentiated tested cells correlating with reported miRNAs expression (Chen *et al.,* 2007).

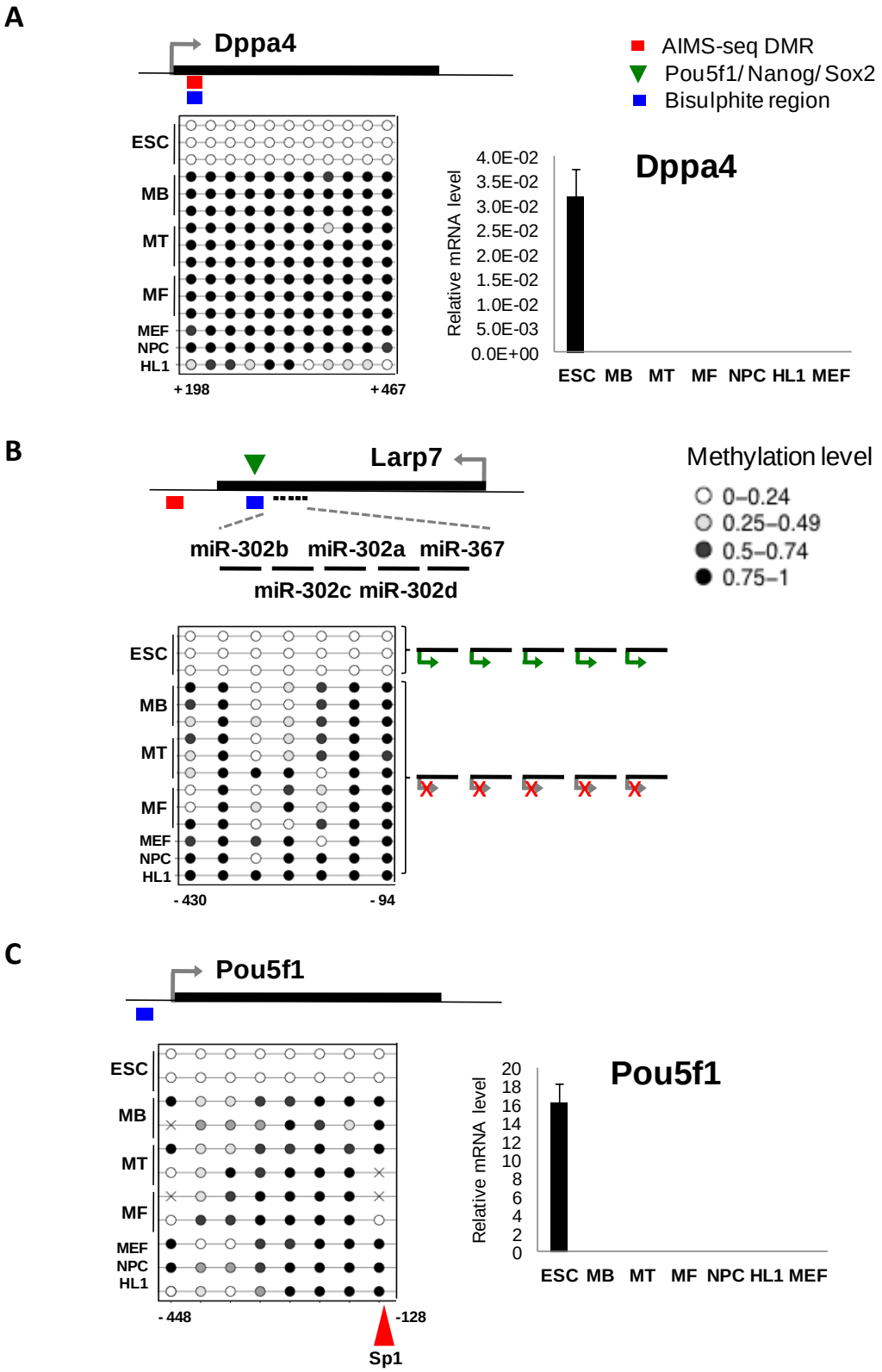


Figure R14. DNA methylation and gene expression profiles of *Dppa4, miR302/367* and *Pou5f1*. Scheme of the *Dppa4* (**A**), *miR302/367* (**B**) and *Pou5f1* (**C**) analyzed regions and the methylation levels of each sample assessed by sodium bisulphite sequencing. The results from biological triplicates are shown. Each circle represents a CpG dinucleotide, and its position in relation to the TSS is indicated below. The colour gradient represents the percentage of methylation indicated in the legend. *Dppa4* and *Pou5f*1 gene expression values were measured by qRT-PCR analysis and normalized with *18S* and *Gapdh*, respectively (n=3, mean ± SD) (**A** and **C** right panel). Scheme of the reported *miR302/367* expression (Chen *et al.*, 2007) in ESC and differentiated cells is shown.

To deepen in the gain of DNA methylation during development, we interrogated the methylation state of the pluripotency marker *Pou5f1* during myogenic differentiation. By sodium bisulphite sequencing we analyzed the *Pou5f1* CpG poor and TATA-less minimal promoter region (OkAzawa *et al.,* 1991). This promoter contains a putative Sp1 site susceptible to be regulated by DNA methylation (5'-(G/T)GGGCGG(G/A)(G/A)(C/T)-3')(Hattori *et al.,* 2004). In agreement with the DMRs linked to pluripotent genes, we observe a totally methylated profile at differentiated and silenced cells and a demethylated state at ESC coinciding with a high gene expression (**Figure R14.C**).

All together these results emphasize the importance of DNA methylation repressing other cell lineages and pluripotency genes during muscle-lineage commitment and terminal differentiation.

1.7.2 Hypomethylated genes during myogenesis

Regarding the demethylated DMRs many of them were located in introns. The most remarkable identified gene was the myogenic master regulator Myf5, analyzed in detail in the section 1.8. Moreover, we detected a DMR in the second intron (Amplicon ID: 53118, chr8: 87268373-87269496) of the Nuclear factor one X (*Nfix*), recently described as a key transcription factor essential for the transcriptional switch from embryonic to fetal myogenesis (Messina *et al.,* 2010) **(Figure R15.A**). Nfix belongs to the CTF/NF-I transcription factor family and is highly expressed in neurons, bone marrow cells and muscle progenitor cells (Gronostajski 2000; Holmfeldt *et al.,* 2013; Messina *et al.,* 2010). Interestingly, it has been reported that *Nfix* has 3 isoforms with different activities. Isoform Nfix1 and Nfix2 would act as transcriptional activators, while isoform Nfix3, the shortest one beginning in an alternative start codon, would act as a transcriptional repressor (Nebl and Cato, 1995).

We validated by sodium bisulphite sequencing the AIMS-seq result confirming that the region represented by the amplicon was fully methylated in ESC and became

demethylated upon myogenic determination (**Figure R15.A**). The gene expression analysis of *Nfix* showed that the ESC did not express any of the *Nfix* isoforms unlike the other samples (**Figure R15.B**). Next, we analyzed the expression of the different isoforms in muscle progenitor cells (MBs), neuronal precursor cells (NPCs) and cardiac cells (HL1s) using isoform specific primers, and we observed that the three isoforms were expressed in all cells, except in NPC, were the isoform Nfix2 was totally absent (**Figure R15.C**). In order to address if intronic DNA methylation could modulate alternative splicing we analyzed the methylation status of different regions, and as shown in **Figure R15.A** we found an intronic region totally methylated in NPCs and demethylated in muscle cells. Interestingly, this region contained an E-box sequence with a CpG dinucleotide (-CACGTG-), where it is bound the ubiquitous transcription factor Usf1 (Upstream stimulatory factor 1) in muscle cells (demethylated CpG) and not in ESC (methylated CpG) as the ChIP analysis showed (**Figure 15.D**). Interestingly, it has been reported that the binding of Usf1 can be affected by DNA methylation in renal cells (Aoki *et al.,* 2008).

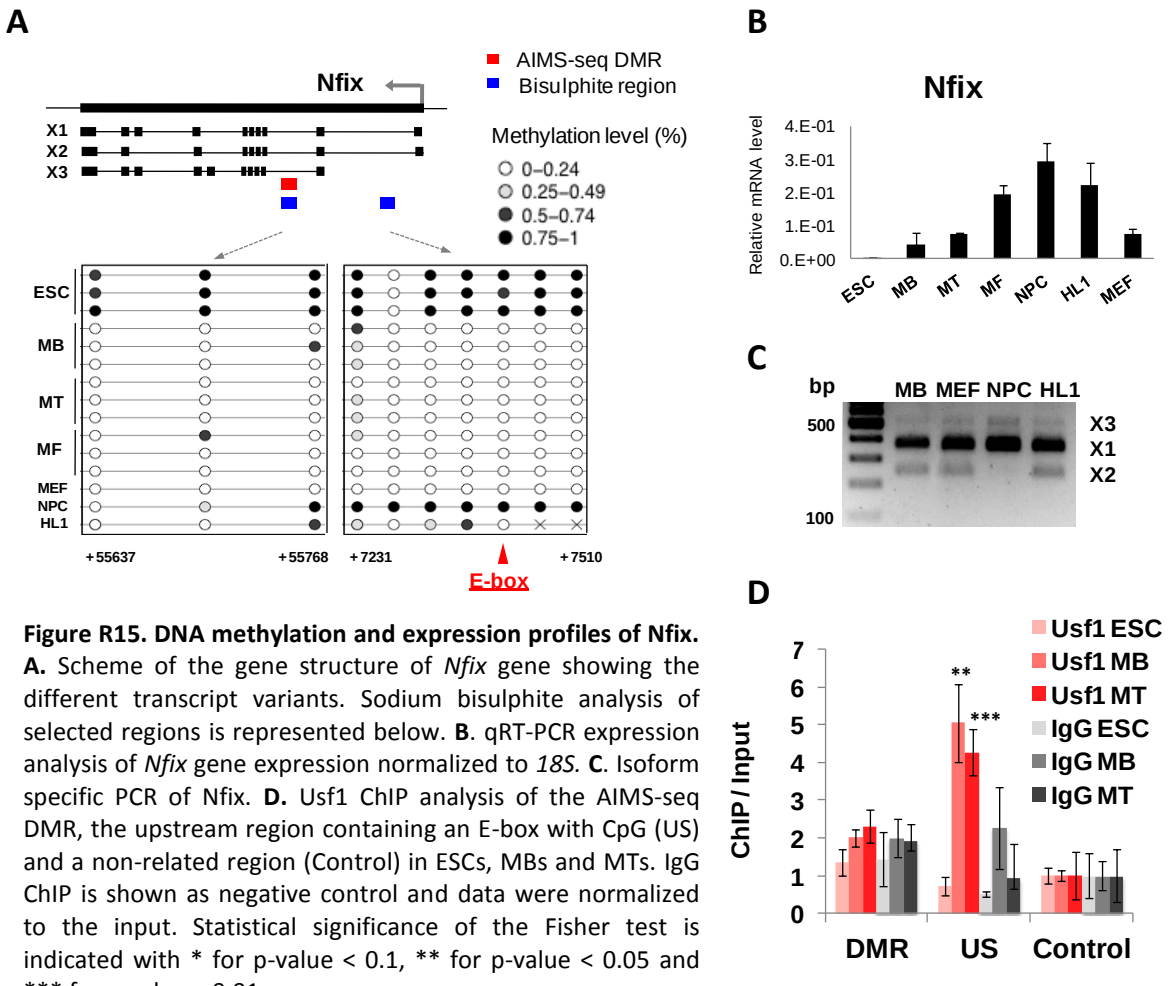


Figure R15. DNA methylation and expression profiles of Nfix. **A.** Scheme of the gene structure of *Nfix* gene showing the different transcript variants. Sodium bisulphite analysis of selected regions is represented below. **B**. qRT-PCR expression analysis of *Nfix* gene expression normalized to *18S.* **C**. Isoform specific PCR of Nfix. **D.** Usf1 ChIP analysis of the AIMS-seq DMR, the upstream region containing an E-box with CpG (US) and a non-related region (Control) in ESCs, MBs and MTs. IgG ChIP is shown as negative control and data were normalized to the input. Statistical significance of the Fisher test is indicated with * for p-value < 0.1, ** for p-value < 0.05 and *** for p-value < 0.01.

Taken together, we observed different expression profiles according to *Nfix* DNA methylation status. While *Nfix* was methylated in ESC and did not express any Nfix isoform, demethylated samples (myogenic cells, MEFs and HL1) expressed the three isoforms, and NPCs (differential intronic methylation) expressed only two isoforms, suggesting that *Nfix* alternative splicing could be modulated by DNA methylation.

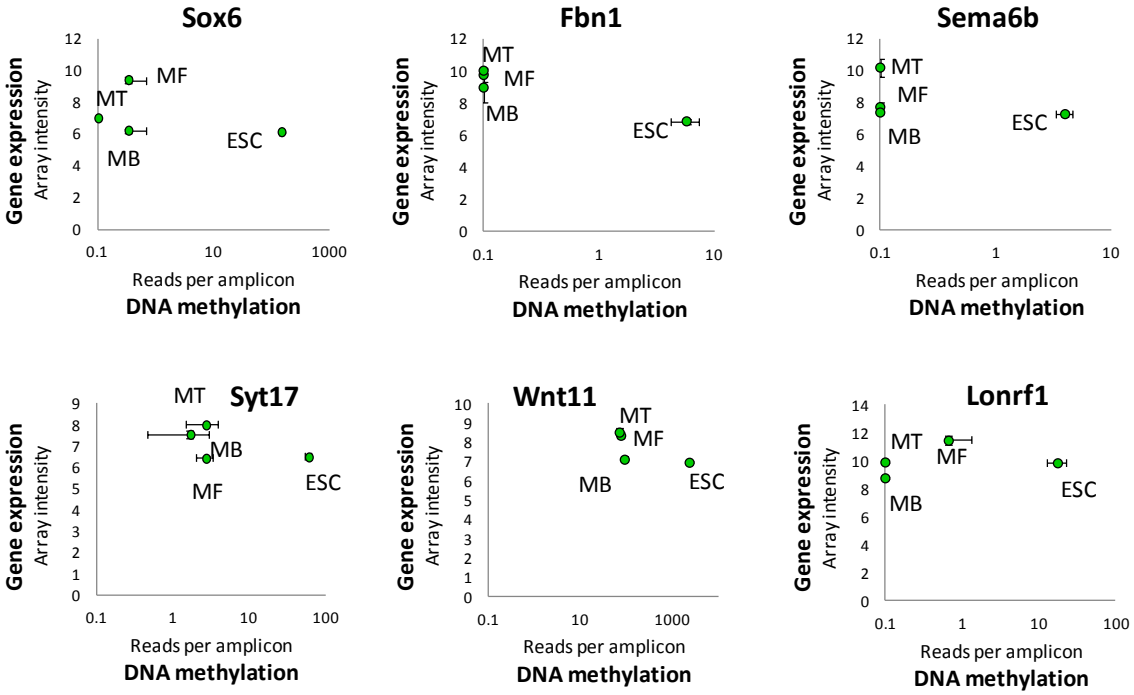


Figure R16. Gene expression upregulation upon lineage commitment hypomethylation. Scatterplot representing the DNA methylation values (reads per amplicon, log10 scale) related to the gene expression values (intensity) in ESCs, MBs, MTs and MFs of selected genes. The average of biological triplicates and the corresponding error bars are shown.

Other interesting demethylated genes are worthy of mention because they became upregulated at certain time during myogenesis and they are functionally related to the muscle formation (**Figure R16**). This is the case of the transcription factor Sox6, a Myf5 target (Wang *et al.,* 2011) that regulates expression of transcriptional regulators critical for muscle development (An *et al.,* 2011; Hagiwara *et al.,* 2007) and showed a loss of DNA methylation during lineage-specification in an intragenic region. This was also observed at *Sema6b* gene, a member of the semaphorine family implicated in axon guidance in neuro-muscular junction (Tamagnone *et al.,* 1999) and muscle

regeneration (Tatsumi *et al.,* 2009; McLoon 2009) that contains a DMR at the promoter region overlapping with a CpG island shore and with the binding site of its upstream regulator Parp alpha (Murad *et al.,* 2011). *Fbn1*, encoding for a glycoprotein that serve as a structural component of calcium-binding microfibrils and *Wnt11*, a member on Wnt family typically involved in organogenesis patterning, were upregulated during lineage commitment as well as demethylated in a distal and proximal amplicon, respectively. Altogether, these DMRs pointed out the repressive function of DNA methylation at the ESC stage and the need of DNA demethylation to allow gene transcription. However, further analyses are needed to ascertain the regulatory role of DNA methylation at these genes during myogenesis.

1.8 Differentiall methylation of *Myf5* super-enhancer

1.8.1 *Myf5* undergoes DNA demethylation during muscle-lineage specification

As previously mentioned, we found a very remarkable DMR at the muscle-lineage determination factor Myf5 among the demethylated genes showing up regulated gene expression. The identified AIMS-seq DMR (amplicon ID: 100707, chr10: 106921115-106921344) covered the intron-3'UTR junction of the *Myf5* gene and was overlapping with the hypaxial enhancer, which is sufficient to lead the expression of *Myf5* in the ventral part of the somites (Summerbell *et al.,* 2000). First of all, we validated this DMR by sodium bisulphitesequencing, and as we can see in **Figure R17,** the CpGs within the amplicon were highly methylated in ESCs and almost totally demethylated in MB, MT and MF. In addition, this demethylation was muscle-specific since did not take place in other cell lineages. This result suggests that the activation of this particular enhancer could be controlled by DNA methylation.

As explained above, *Myf5/Myf6* locus is tightly regulated by several enhancers distributed over more than 100kb upstream of the genes. Importantly, the generation of a super enhancer catalogue in 86 cell and tissue-types has identified *My5/Myf6* locus as a super-enhancer in myoblast, myotubes and skeletal muscle samples (Hnisz

et al., 2013). Very recently, has come up the concept of super-enhancers having crucial functions in defining cell identity (Whyte *et al.,* 2013). In order to address if the *Myf5/Myf6* super-enhancer could be epigenetically regulated by DNA methylation, we analyzed by sodium bisulphite sequencing the different enhancer elements described up to now: the -111 kb enhancer, the 17 kb enhancer, the early epaxial enhancer (EEE), and the proximal arch enhancer (PAE), as well as the minimal promoter region of *Myf5* (Summerbell *et al.,* 2000) and Myf6 (Black *et al.,*1995). The enhancer at -57.5 kb was not analyzed due the absolute lack of CpG dinucleotides at the core enhancer region of 145 bp, and the very poor CpG content of the surrounding sequence.

All the super-enhancer region was totally methylated at ESC stage, whereas in MB we observed an almost complete demethylation of all clustered enhancers, except for the -17 kb and the -111 kb, where the demethylation was partial (**Figure 17**). This profile was maintained in MT and MF in the EEE, and at a lesser extent in proximal arch enhancer, the -17 kb and the -111 kb enhancers, which showed a total or partial remethylation, respectively. Importantly, the demethylation of the *Myf5/Myf6* super-enhancer was muscle-lineage specific not taking place in MEF, NPC and HL1 cells. Subsequently, we analyzed the methylation status of the *My5* and *Myf6* promoter, being both CpG-poor regions. As expected, we also found very high levels of DNA methylation in both promoters in ESC. However, we observed a very different methylation pattern during muscle lineage progression: meanwhile *Myf5* promoter was highly demethylated in all muscle cell types (MB, MT and MF), *Myf6* was only demethylated in mature myofibers, where the gene is highly expressed (**Figure R18.A**).

In order to rule out the possibility that the complete genomic region was demethylated during the muscle-lineage commitment in a non-specific manner, we analyzed three regions without any specific enhancer activity described, two located upstream of the TSS of *Myf5* (at -4.4kb and -118kb) and one downstream of the gene. Importantly, the three control regions remained heavily methylated during the myogenic lineage progression, demonstrating the specificity of the reported DNA demethylation.

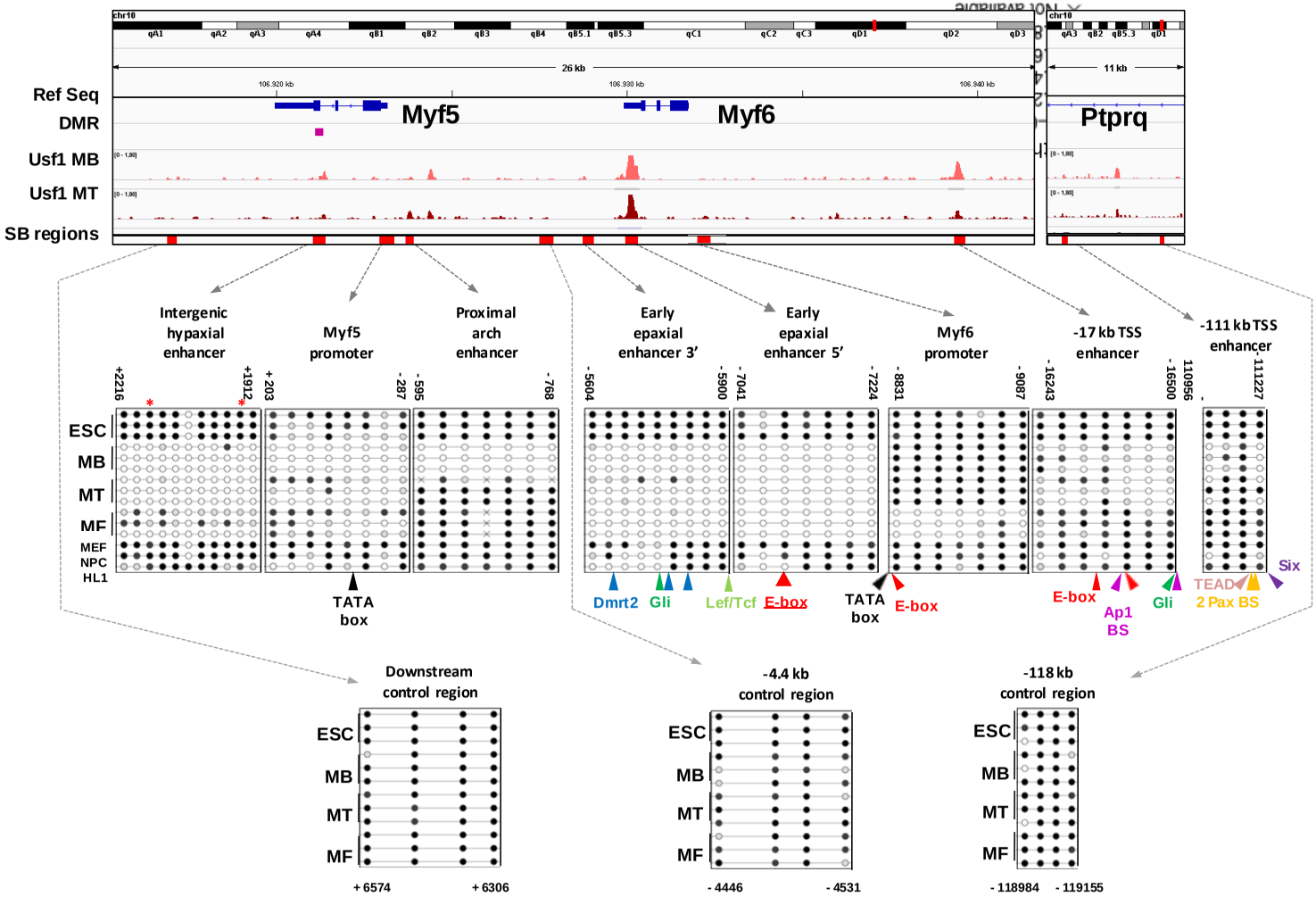


Figure R17. Regulatory role of DNA methylation in *Myf5/Myf6* locus. A. Scheme of the *Myf5/Myf6* and Ptprq loci showing the AIMS-Seq DMR, sodium bisulphite analyzed regions (SB) and ChIP peaks for Usf1 (ENCODE Project). SB analysis of the super enhancer regions and the transcription binding sites and E-boxes are indicated below. Each circle represents a CpG dinucleotide, and its position distance to the TSS of *Myf5* is indicated. The colour gradient represents the percentage of methylation indicated in the legend.

1.8.2 Usf-1 binds *Myf5* enhancers in a DNA demethylation dependent manner

Next, centered on *Myf5* regulation, we investigated whether DNA methylation could prevent its activation in ESCs. Usf1 is an ubiquitously expressed transcription factor important for mesoderm lineage determination (Deng *et al.,* 2013) and insulating functions (Huang *et al.,* 2007). It recognizes the canonical E-box sequence CANNTG and, interestingly, this binding has been postulated to be regulated by DNA methylation (Fujii *et al.,* 2006). In addition, it was reported that Usf1 bound to the E-box located at the -17 kb *Myf5* enhancer region in the myogenic cell line C2C12, suggesting that might be responsible for the transactivation of *Myf5*, although the DNA methylation state of the enhancer was never addressed, neither the binding profile in ESCs (Chang *et al.,* 2004). Importantly, Usf1 Chip-seq data from ENCODE project on C2C12 cells (Hardison's Lab) confirmed these results and also pointed out the EEE as a new region bound by Usf1 in the Myf5 locus (**Figure R17**).

After confirming by qPCR the expression of Usf1 in ESC, C2C12 MB and C2C12 MT (**Figure R18.B**) we performed ChIP assays on the enhancers that contained E-boxes and showed DNA demethylation during myogenic specification. An E-box located at -118 kb of the TSS of *Myf5*, and surrounded by deeply methylated CpGs in all the analyzed samples was used as a negative control for Usf1 binding. As shown in **Figure R18.C**, we detected a significant Usf1 binding at the EE enhancer at the MB stage (p-value= 0.0144,) and a slight binding at the -17 enhancer (p-value = 0.08693), and not binding in ESC. The EE enhancer showed a decrease in the Usf1 binding at MT stage (p-value = 0.06825) whereas the -17 lacked of Usf1 presence.

Of note, the E-box located at the EE enhancer contains a CpG dinucleotide inside, whereas the other analyzed E-boxes are not, suggesting that the effect of DNA methylation on the Usf1 binding may occurr through the direct mask of the binding motif and might prevent the Myf5 expression in non-muscle cells. This CpG-containing E-box is conserved in human genome and, importantly, when three primary human

MBs (hMBs) were compared to three human ES cell lines (hESCs), it also showed a loss of methylation in the myogenic lineage (**Figure R18.D**).

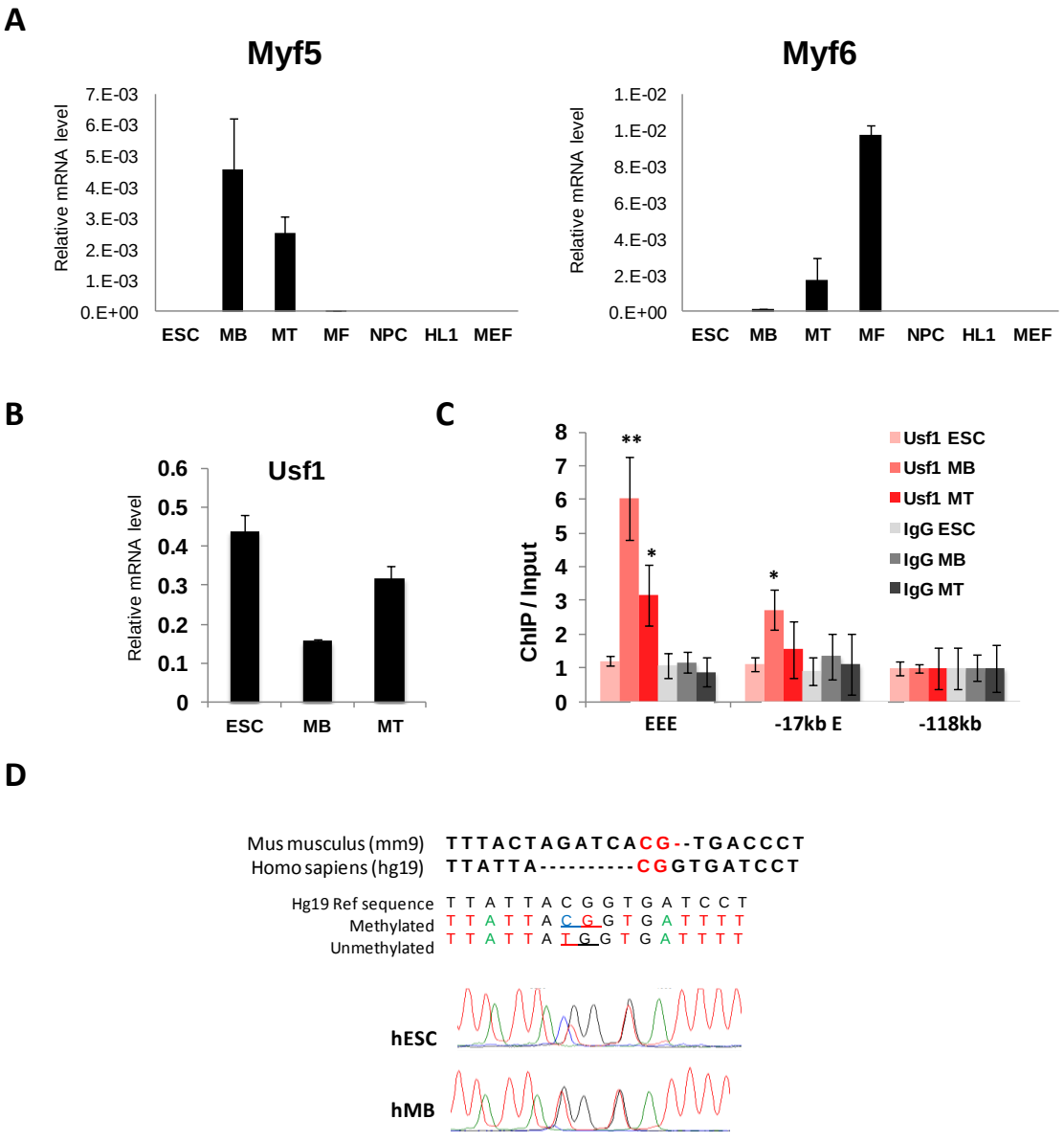


Figure R18. Regulatory role of DNA methylation in *Myf5/Myf6* locus. A qRT-PCR expression analysis of Myf5 and Myf6 normalized to 18S **B.** qRT-PCR expression analysis of Usf1 gene normalized to Sdha. **C.** Usf1 ChIP analysis of the EEE (containing an E-box with CpG), -17 kb enhancer and a non related region (Control) in ESC, MB and MT. IgG ChIP results are shown as negative control and data were normalized to the input. Statistical significance of the Fisher test is indicated with * for p-value < 0.05 and ** for p-value < 0.01. **D.** Methylation status of the conserved CpG embedded in the E-box located in EE enhancer of *Myf5* in hESC and hMB. Sodium bisulphite results of one out of three biological replicates

1.8.3 Chromatin marks at *Myf5/Myf6* locus

Finally, to complete the epigenetic regulatory characterization of the *Myf5/Myf6* locus, we took advantage of the ChIP-seq data available from the ENCODE project or, in default, from the GEO Omnibus database previously mentioned. Using Integrative genome viewer we visualized the positive marks (H3K4me3, H4K27Ac, H3K4me1, and p300) and the repressive mark H3K27me3 in the Myf5/Myf6 super-enhancer cluster (**Figure R19**). Notably, H3K27me3 was totally absent in ESCs, MBs and MTs, indicating that this locus is not repressed by Polycomb. As expected in a super-enhancer cluster, the entire region was highly enriched in H3K4me1, H3K27Ac and p300 in MB/MT whereas ESCs showed only slight levels of the poised enhancer mark H3K4me1.

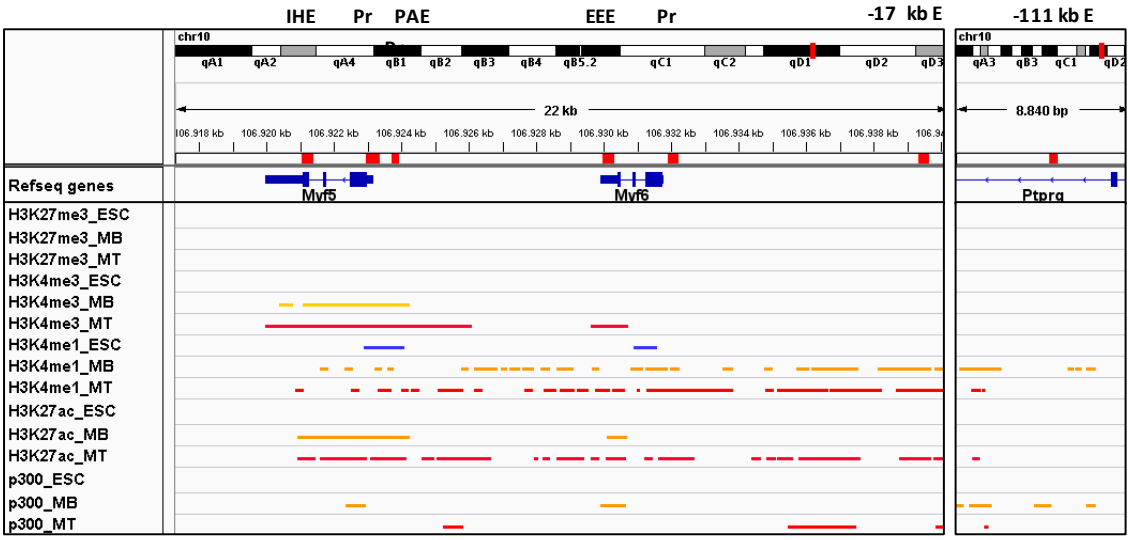


Figure R19. Chromatin marks at *Myf5/Myf6* locus. Scheme of the *Myf5/Myf6* locus were the regulatory regions are indicated as red boxes and the names of the specific regions are indicated on the top. ChIP-seq peaks for H3K4me1, H3k4me3, H3K27me3, H3K27Ac and p300 at ESCs, (coloured blue) MBs (colored orange) and MTs (colored red) are indicated. Names of the enhancers are indicated above; Pr means promoter.

Taken all together, we have characterized the DNA methylation dynamics during the muscle-lineage establishment and highlighted the importance of DNA methylation regulating the cell-identity *Myf5* super-enhancer in muscle cells.

2 EPIGENETIC PROFILE OF THE PRINCIPAL MYOGENIC REGULATORY GENES

In order to deepen into the regulatory implications of DNA methylation during myogenic lineage progression, we extended our analysis to the principal regulatory myogenic genes.

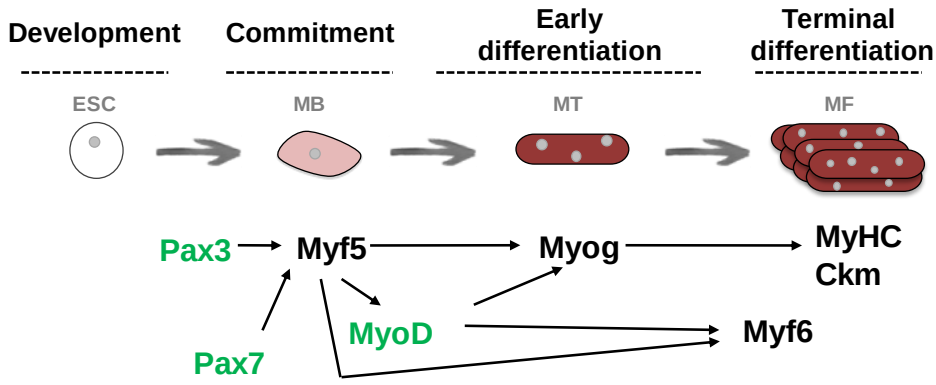


Figure R20. Scheme representing the temporal expression of key myogenic genes. Green colored genes present CpG island overlapping the promoter regions, whereas black colored genes have low CpG content promoters.

We analyzed by sodium bisulphite sequencing the DNA methylation state of reported regulatory regions of key genes implicated in the myogenic process, classified in three groups: developmental genes (*Pax3* and *Pax7*), myogenic regulatory factors (*MyoD* and *Myogenin*) and terminal differentiation genes (*Myh1, Myh4, Myh8*, and *Ckm*) (**Figure R20**). Likewise, these genes could be grouped according to the presence or absence of CpG islands at their promoter region, allowing us to infer in the impact of DNA methylation at diverse CpG content regions. As indicated in the chapter 1, we compared ESCs with MBs, MTs and MFs samples isolated from at least two different adult mice. To address the lineage specificity of the observed DNA methylation events, the methylation profile of non-myogenic cell lines (neuronal precursor cells (NPCs), embryonic fibroblasts (MEFs) and cardiomyocytes (HL1)) were analyzed too.

In addition to the myogenic model and thanks to the collaboration with Charles Keller's lab, we also analyzed the methylation profile of the developmental genes *Pax3* and *Pax7* at three Rhabdomyosarcoma cell lines. The results highlighted the aberrant methylation patterns associated with tumorigenesis.

Moreover, we analyzed the DNA methylation state of *MyoD* enhancer and *Myogenin* promoter in freshly isolated quiescent satellite cells and *in vivo* activated satellite cells, when they were not expressed yet, revealing an optimal methylation state for the ready to go latent myogenic adult stem cells.

Finally, taking advantage of the recent publication of the histone modifications profiles in myoblast and myotubes (Asp *et al.,* 2011; Blum *et al.,* 2012) together with the publicly available ENCODE data of multiple embryonic stem cells, we evaluated the histone modification patterns that go along with the myogenic progression. The combination of DNA methylation and histone modification analysis with the CpG island distribution exposed a different epigenetic profile guided by the underlying genetic sequence. By integrating all these data, we draw the epigenetic contour that together with the transcription factors orchestrates the muscle formation.

2.1. DNA methylation profiles of developmental genes: *Pax7* and *Pax3*

The Pax family of paired domain transcription factors play key roles during tissue specification and organ development and are remarkably conserved genes. Pax7 and Pax3 are involved in developmental and adult regenerative myogenesis and, as a distinctive trait of developmental genes, both contain well conserved CpG islands surrounding and overlapping the promoter region. We performed an accurate analysis of the DNA methylation state during myogenesis and at non-myogenic samples of this CpG islands and their shores, as well as of described enhancers located outside the CpG-rich-content regions. In parallel, we analyzed the mRNA expression of these developmental genes in the same samples, and as expected, *Pax7* was only expressed

in muscle cells whereas *Pax3* was also expressed in cardiac and neuronal cells **(Figure R21**).

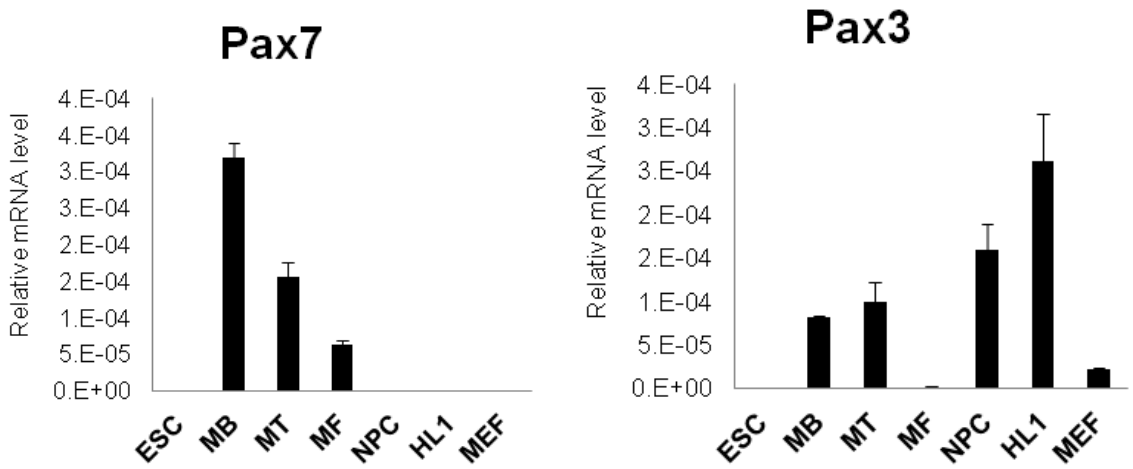


Figure R21. Expression profile of *Pax7* and *Pax3* during myogenesis and at non-myogenic cell lines. qRT-PCR expression analysis of *Pax7* and *Pax3* were normalized to *18S* expression. Data represent the average of three independent experiments ± SD.

Pax7 has four CpG islands located at the gene body and one of those covers the promoter region. We have analyzed the two CpG islands closest to the TSS and the surrounding region of the regulatory Rbpj binding site, located 7.4 kb upstream of the TSS. This binding site is of interest since satellite cells self renewal is positively regulated by the Notch signaling cascade through the direct binding to this region (Wen *et al.,* 2012). The binding site lacks CpG dinucleotides, therefore the methylation state of the closest CpGs was interrogated. We found the analyzed CpG islands totally demethylated in all samples whereas the CpGs surrounding the Rbpj binding site were fully methylated in all samples, except for the cardiac cell line (**Figure R22.A**).

Pax3 has five CpG islands, one in an intron, three upstream the gene and one CpG island coinciding with the TSS, the TATA box and the core binding sequence of the cAMP response element (CRE; 5'-TGACG-3') (Chen *et al.,*2005). Notably, when the embedded CpG in CRE is methylated in HeLa and PC12 cell lines, transcription binding is diminished (Iguchi-Ariga and Schaffner 1989). The sodium bisulphite analysis showed that *Pax3* presents no methylation at any of the five CpG islands in all cell

types, neither at the CpG island shores, consequently with the unmethylated status characteristic of the CpG islands (**Figure R22.B**).

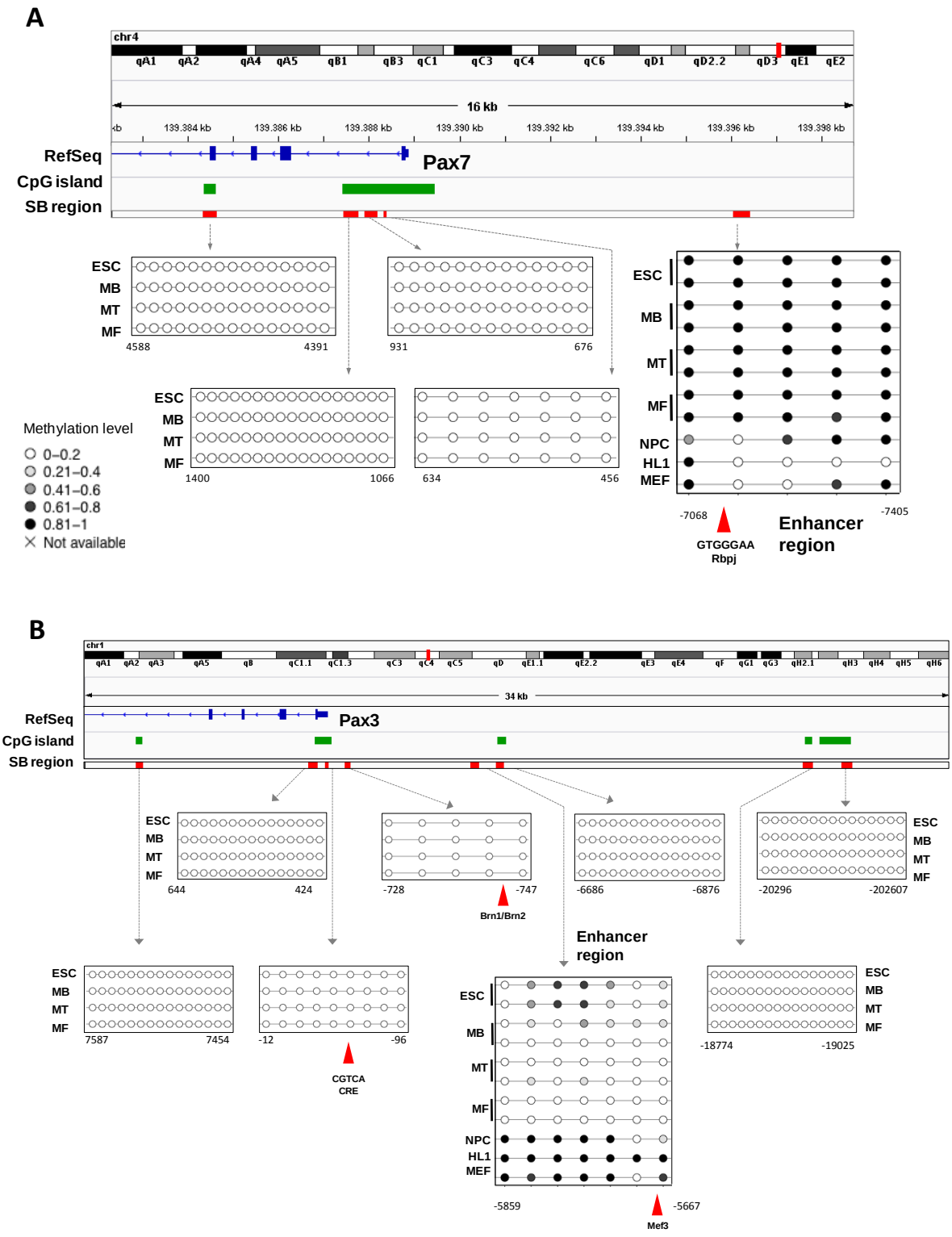


Figure R22. DNA methylation profiling of *Pax3* (A) and *Pax7* (B) locus during myogenesis. The DNA methylation levels of analyzed regions (determined by sodium bisulphite sequencing, SB region) are shown using circles charts. Each circle represents a CpG dinucleotide and its distance to the gene TSS is indicated below. The color gradient represents the percentage of methylation indicated in the legend. Transcription factor binding sites are indicated with red arrows.

Next, two CpG low content regions corresponding to the proximal promoter and the hypaxial enhancer at -5.7 kb upstream the TSS were analyzed too (Brown *et al.,* 2005). Whereas the promoter region remained unmethylated in all samples, the *Pax3* hypaxial enhancer region showed a significant loss of methylation in myogenic cells (MBs, MTs and MFs) compared to the cells that are not myogenic (ESCs, NPCs, MEFs and HL1).

2.1.1. DNA methylation patterns in Pax7 and Pax3 loci in Rhabdomyosarcoma cell lines

Tumoral processes are characterized by hypermethylation of CpG islands located in the 5'- upstream and coding regions of specific genes. We wondered if any alterations occurred at DNA methylation level in muscle developmental tumors. Thanks to our collaboration with Charles Keller's Lab (Pediatric Cancer Biology Program, Papé Family Pediatric Research Institute, Oregon Health & Science University, Portland, Oregon) we were able to analyze the methylation state of *Pax3* and *Pax7* CpG islands in murine alveolar rhabdomyosarcoma (aRMS) cell lines.

Most of the aRMS tumors are caused by the *Pax3/Fkhr* translocation, however the cellular origin of the tumors remains unknown. Our collaborators generated several mouse models where the *Pax3/Fkhr* fusion gene was expressed in different myogenic cell types, using the Cre/loxP system that survey different tumoral cellular origins, with the purpose of shedding light into the aRMS origin and create an accurate cellular model for this aggressive childhood cancer (Keller and Capecchi, 2005; Abraham *et al.,* 2014) (**Figure R23.A-B**). In addition, they generated mice with an estrogen receptor element (ER) induced by Tamoxifen in order to be able to distinguish postnatal from embryonic myogenesis. After detailed analyses of different murine models, the most susceptible lineages generating tumors were embryonic myogenic progenitors and postnatal satellite cells (Abraham *et al.,* 2014).

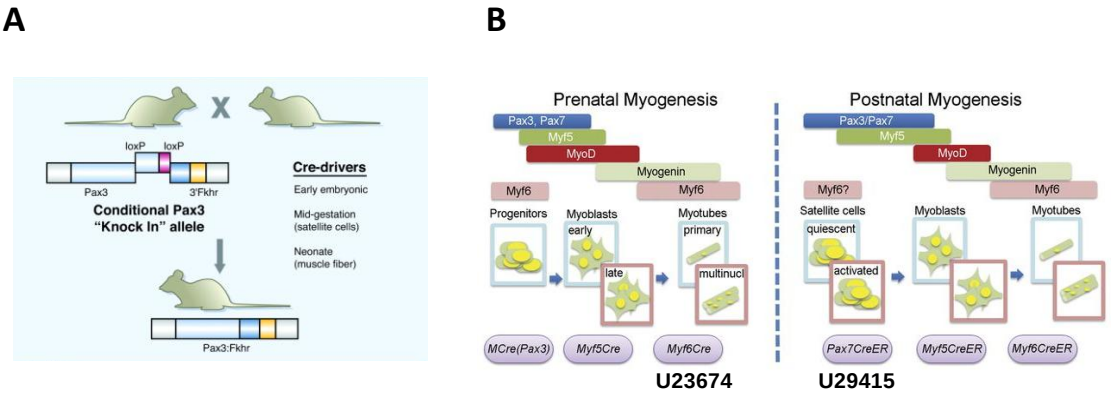


Figure R23. Alveolar Rhabdomyosarcoma cell lines generation. A. Scheme of the Cre/loxP-mediated conditional "knock-in" approach that drives the *Pax3: Fkhr* translocation under different Cre drivers. Obtained from (Keller and Capecchi, 2005). **B.** Representation of prenatal and postnatal myogenesis in the context of Cre drivers used to trigger tumors to generate aRMS cell lines. Obtained from (Abraham *et al*. 2014). In this study *Myf6*Cre (U23674) and *Pax7*CreER (U29415) cell lines have been analyzed.

The aRMS cell line that represents the myogenic progenitors cell origin (U23674 cell line, *Myf6*Cre) was isolated from a tumor of a mice that carried the Cre recombinase protein downstream the *Myf6* gene, leading to the *Pax3:Fkhr* expression in differentiating myogenic progenitors in the prenatal stage, when the first wave of myogenesis take place. Note that *Fkhr* gene can also been called *Fox1a.* The cell line that mimics the aRMS satellite cells origin (U29415 cell line, *Pax7*CreER) was obtained from tumors generated in mice that carry the Cre recombinase protein under the *Pax7* promoter induced by Tamoxifen. In these mice the *Pax3:Fkhr* expression occurs after the administration of Tamoxifen at day 30 post birth (adolescence) only in satellite cells ensuring the expression of the chimeric protein at the adult myogenic stem cells, and not during the developmental stage.

Importantly, during the amplification process of both tumor cell lines we detected a notably higher proliferation rate and a clear myogenic differentiation impairment of the U23674 cell line compared to U29415 cell line, in agreement with a very high expression of *Pax3:Fkhr* in U23674 cells compared to U29415 cells, where the expression was almost undetectable (**Figure R24**). Embryonic stem cells, proliferating primary myoblasts and differentiated primary myotubes expression profiles were analyzed in parallel as reference. Interestingly, the treatment with DNA methylation

inhibitor 5'Aza2 deoxycytidine (Aza) reduced dramatically *Pax3:Fkhr* expression in U23674 at 72h of treatment (**Figure R24**. Regarding *Pax7* expression after Aza treatment, no expression changes were detected neither at U29415 or at U23674 tumoral cell lines.

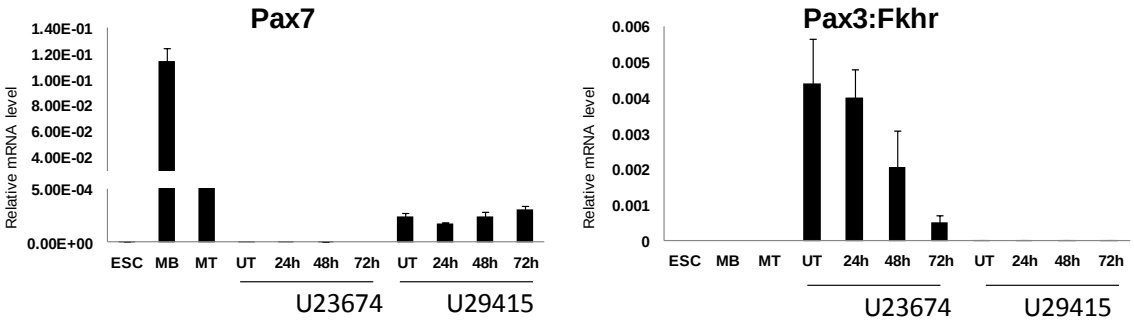


Figure R24. Expression analysis of *Pax3:Fkhr* and *Pax7* in aRMS cell lines treated with 5-Aza-2`Deoxcytidine. qRT-PCR expression analysis of U29415 and U23674 cell lines untreated (UT) and treated with Aza for 24, 48 and 72 hours, normalized to *18S* expression. As a reference, expression levels in ESC, MB and MT were analized.

5-Aza-2`Deoycytidine dose concentration for each cell line was established by our collaborator based on drug sensitive assays and, consequently, U23674 cells were treated with 3 µM and U29415 with 20 µM for 72 hours (for details see **Appendix I Figure 2A**).

In order to address the methylation status of the cell lines, we performed bisulphite sequencing analysis of *Pax3:Fkhr* locus in untreated cells. The results revealed a general unmethylated state at all five CpG islands and at the enhancer region in both aRMS cells lines, except from punctual partially methylated CpGs in both cell lines **(Figure R25.B)**. Nevertheless, in the *Pax7* promoter region, the U23674 showed an aberrant methylation profile at the two analyzed CpG islands. The intronic CpG island was almost totally methylated and some CpGs of the CpG island shore of the upstream CpG island were also hypermethylated **(Figure R25.A)**. This gain of methylation correlated with the extremely low *Pax7* expression observed in this cell line. Finally, we performed bisulphite sequencing analysis of the same regions after 5-Aza-

2`Deoycytidine treatment without observing major changes in DNA methylation levels (**Appendix I Figure 2B).**

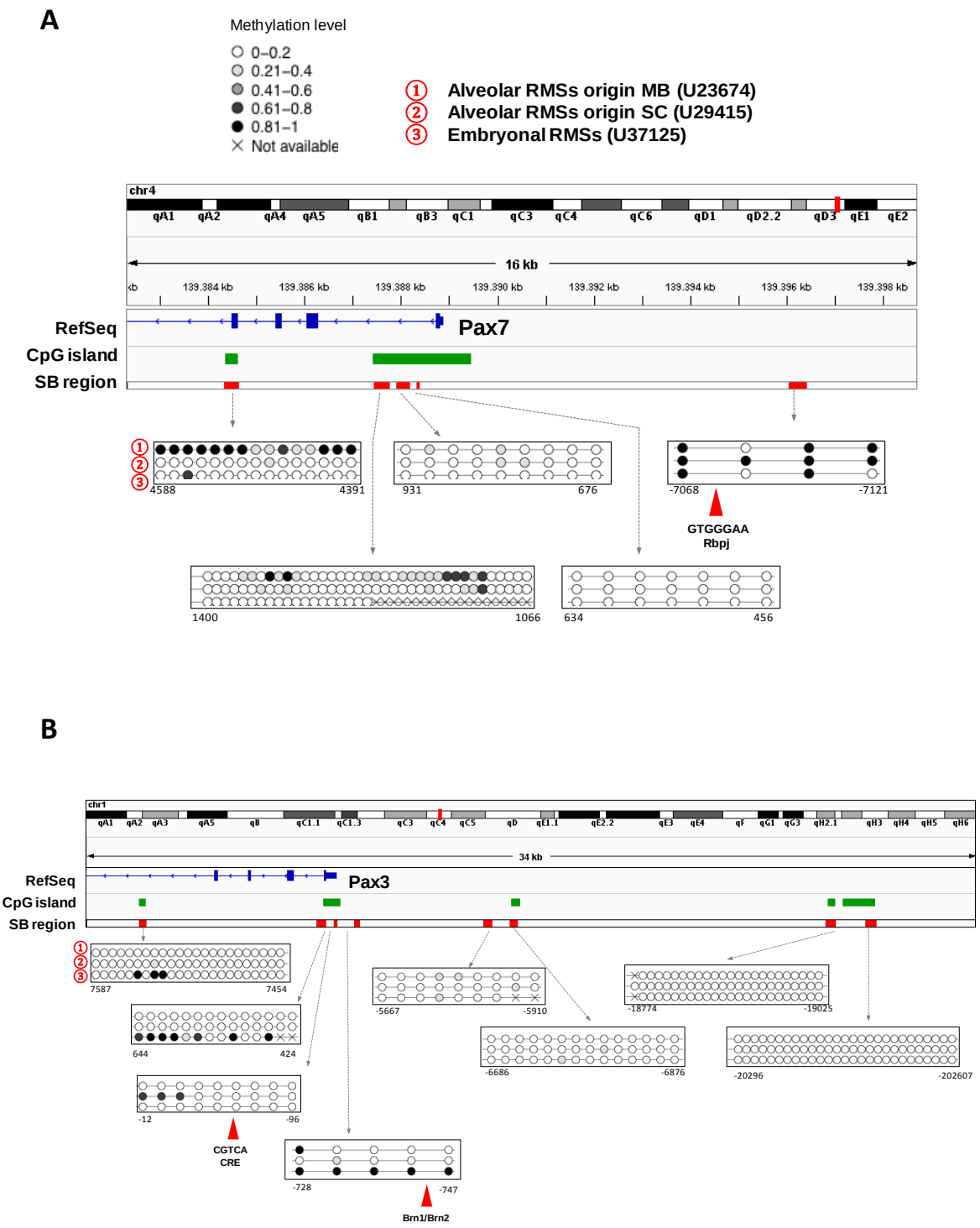


Figure R25. DNA methylation profiling of *Pax7* (A) and *Pax3* (B) loci in Rhabdomyosarcoma cell lines. The DNA methylation levels of analyzed regions (determined by sodium bisulphite sequencing, SB region) are shown using circles charts. Each circle represents a CpG dinucleotide and its distance to the gene TSS is indicated below. The color gradient represents the percentage of methylation indicated in the legend. Transcription factor binding sites are indicated with red arrows.

After an exhaustive analysis of different tumor types we concluded that the myogenic progenitors cells might be the most plausible cellular origin for aRMS, meanwhile the satellite cells derived tumors would resemble more eRMS (**Appendix I Figure 3**) (Abraham *et al.,* 2014). To provide an epigenetic clue to this hypothesis, we analyzed the methylation state of *Pax3:Fkhr* and *Pax7* genes in an embryonic RMS cell line (U37125, *Myf6*Cre line with *p53* deletion, (Rubin *et al.,* 2011)). The eRMS cell line U37125 showed hypermethylation of the *Pax3* CpG island that overlaps the promoter region coinciding with the results observed in human eRMS tumors (Kurmasheva *et al.,* 2005). However, these results do not correlated with the methylation pattern observed in U29415 cells, depriving of a simple link between the tumor originated from satellite cell and the eRMS tumor type (**Figure R25**).

2.2 DNA methylation profiles of myogenic regulatory factors

Muscle progenitors switch on a common muscle program, requiring the myogenic regulatory factors (MRFs) Myf5, MyoD, Mrf4/Myf6 and Myogenin, all of which are members of the basic helix-loop-helix family of transcription factors. Myf5 and MyoD are the two MRFs that determine the myogenic lineage commitment, being *Myf5* gene the first to be expressed. After the myogenic specification, Myogenin activates differentiation genes such as *Desmin* and *Myosines* while maintains the transcriptional activation of muscle specific genes initiated by MyoD (Blais *et al.,* 2005). Finally, Mrf4 acts as a differentiation factor in mature fibers.

Myf5, Mrf4 and *Myogenin* have low CpG content promoters whereas *MyoD* presents a CpG island overlapping the promoter region. The DNA methylation profile at *Myf5/Myf6* locus has been deeply described in section 1.6. Briefly, we concluded that DNA demethylation of *Myf5* super-enhancer regions occured only at myogenic samples allowing Usf1 binding and restricting the activation of *Myf5* to the myogenic lineage.

Here, we analyzed *MyoD* expression profile and methylation status of myogenic and non-myogenic samples by sodium bisulphite sequencing addressing the ESC stage for the first time. The results, in line with the previous studies, showed a clear methylated state in non-myogenic cells, whereas myogenic cells and cardiomyocytes were depleted of methylation (**Figure R26.A**). Next, we addressed *Myogenin* regulation analyzing its promoter region that, as mentioned in the introduction, contains several binding sites. As shown in **Figure R26.B**, we observed a lack of methylation exclusively in myogenic cells, whereas ESC and other cell lines showed high methylation rates.

Surprisingly, our results were in disagreement with the previous data published by Lucarelli and Oikawa showing DNA demethylation when the MBs turned into MTs (Lucarelli *et al.,* 2001);(Oikawa *et al.,* 2011). Unlike the cited works, we analyzed primary satellite cell derived MBs, MTs and ESCs ,as initial stage. To clarify this contradictory result and to discard the possibility of having myoblasts together with differentiated cells in proliferation conditions, we analyzed the methylation state of *Myogenin* promoter in freshly isolated quiescent satellite cells and in *in vivo* activated satellite cells (described in section 2.2.1).

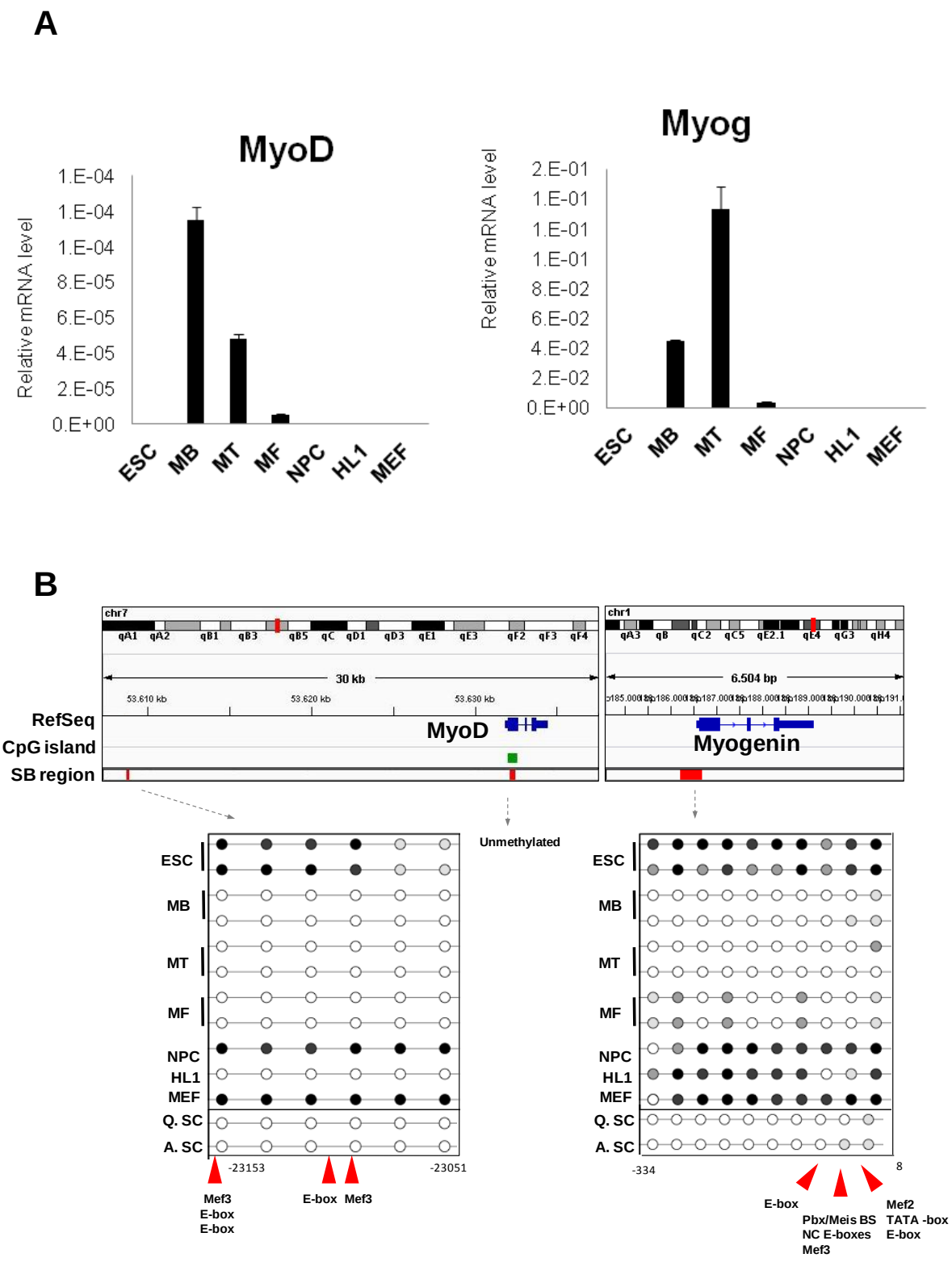


Figure R26 Gene expression and DNA methylation profiling of *MyoD* and *Myogenin*. A. qRT-PCR expression analysis of *MyoD* and *Myogenin* were normalized to *18S* expression (top). Data represent the average of three independent experiments ± SD **B.** The DNA methylation levels of analyzed regions (determined by sodium bisulphite sequencing, SB region) are shown using circles charts. Each circle represents a CpG dinucleotide and its distance to the gene TSS is indicated below. The color gradient represents the percentage of methylation indicated in the legend. Transcription factor binding sites are indicated with red arrows. *MyoD* CpG island is demethylated according to (Jones *et al.,* 1990).

2.2.1 Quiescent and activated satellite cells analysis

Quiescent satellite cells represent the dormant (no cycling) adult stem cell population of the skeletal muscle in charge to maintain the tissue homeostasis and to regenerate the tissue upon injury.

Our results clearly showed that important regulatory regions for *MyoD* and *Myogenin* were totally unmethylated in *in vitro* cultured MBs. In order to address for the first time how was *the in vivo* methylation state of these genes, we isolated quiescent satellite cells from transgenic *Pax7* $^{Cre+/YFP+}$ kindly provided by Dr. Muñoz Cánoves (UPF, Barcelona). Quiescent cells were isolated by flow cytometry after mechanical and chemical digestion of *Pax7* $^{Cre+/YFP+}$ muscles, (**Figure R27**).

We analyzed the same isolated fraction from mice injured with cardiotoxin (CTX), which leads to damage response and satellite cell mitotic expansion (*in vivo* activation). These activated cells where isolated at two different time point after injury (6h and 72 hours after CTX injection) to analyze the early activated cells and the transient amplification derived population. It is evident by the flow cytometry analysis that the complexity and size of the YFP^+ population increased when the satellite cells were proliferating.

We proceeded to perform gene expression analysis in order to analyze the expression profile of the isolated fractions and to evaluate whether it corresponded indeed to the quiescent population. Very low amounts of RNA were obtained from the sorted satellite cells due to their quiescen state,. therefore, prior to the expression analysis, we amplified the RNA of four biological replicates per condition following the protocol described by (Gonzalez-Roca *et al.,* 2010).

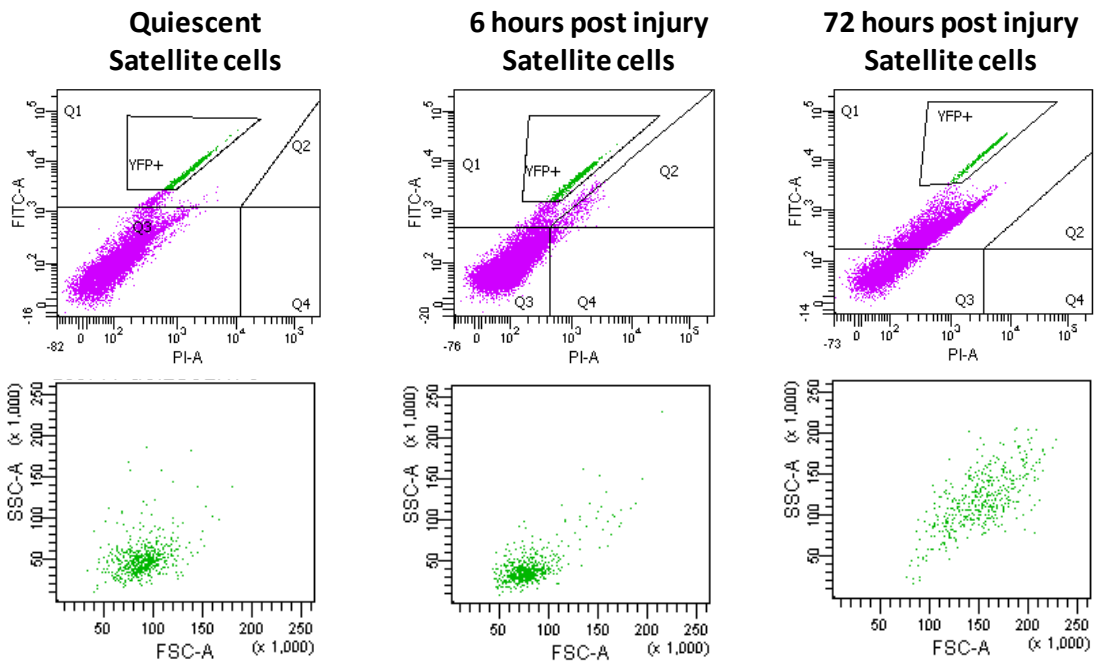


Figure R27. YFP^{+} sorting of quiescent and *in vivo* activated satellite cells. Satellite cells of Pax7 $^{Cre+/YFP}$ mice were isolated from not injured mice, representing the quiescent population (left plot) and from injured mice after 6h and 72h after CTX injury, representing the *in vivo* activated population (central and right plots). Upper plots show the FITC versus PI scatterplots evidencing the low amount of satellite cells (GFP^{+}) in the muscles. The bottom plots show the morphological differences between the different conditions. Forward scatter characteristics (FSC-A) measure the size of the cells and side scatter characteristics (SSC-A) measure the internal complexity.

The principal component analysis (PCA) of the expression values clustered the samples in three well-differentiated populations, indicative of a proper induction of the activation (**Figure R28**). To confirm the quiescent state of the satellite cells we evaluated the protein expression of MyoD and the proliferation marker Ki67 through immunostaining at quiescent satellite cells and at satellite cells 72h after CTX injury. As expected, quiescent satellite cells showed neither MyoD nor Ki67expression, whereas activated cells were positive for both markers (**Figure R29.A-B**). Once proved the quiescence state of the isolated satellite cells, we analyzed by sodium bisulphite sequencing a pool of genomic DNA from 7 mice. Importantly, we found that *MyoD* and *Myogenin* regulatory regions in the activated proliferating cells were totally demethylated. More interestingly, the quiescent satellite cells were fully demethylated too in both cases prior to the activation of gene expression (**Figure R25.B**).

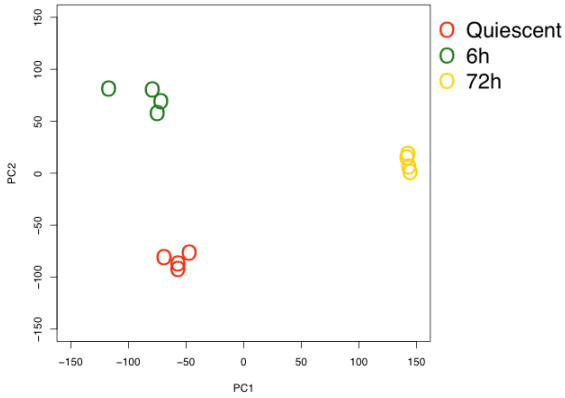


Figure R28. Expression signature of quiescent and *in vivo* activated satellite cells (6h and 72h after CTX injury). PCA of the expression microarray values obtained of the three different *Pax7* $^{cre+}/^{YFP+}$ mice: quiescent, 6 hour and 72 hours post CTX injection.

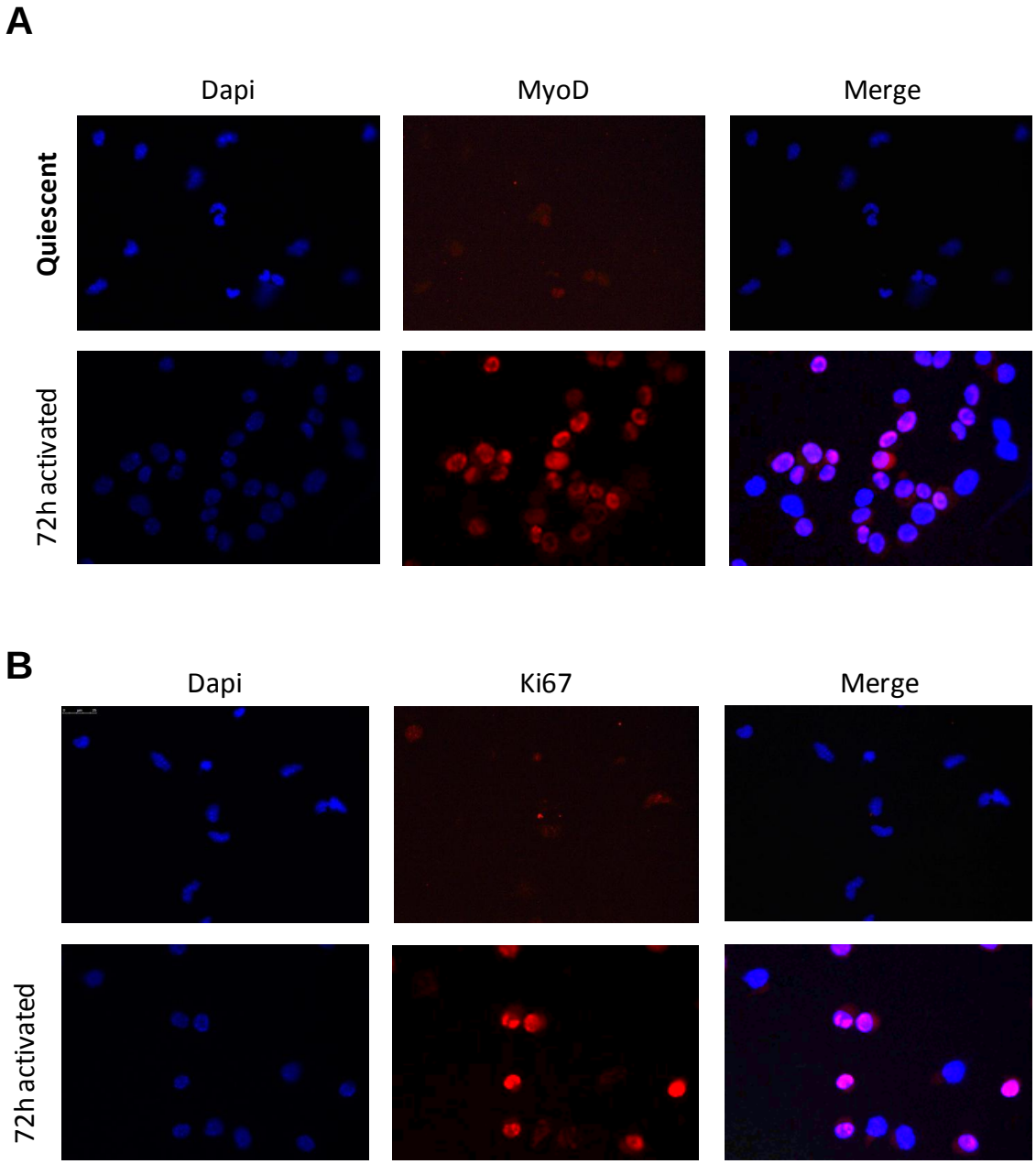


Figure R29. MyoD (A) and Ki67 (B) immunostaining of the quiescent satellite cells and *in vivo* activated satellite cells 72h after CTX injury. Activation markers (MyoD and Ki67) were stained with Alexa568 (red fluorescence). Dapi fluorescence stained the nucleus and colocalized with MyoD and Ki67 staining.

For *MyoD* these results could be interpreted in two different ways, in one hand the methylation state might reflect the poised-state of the satellite cells for a quick activation while, in the other hand, they might be a consequence of the prior expression of *MyoD* during developmental myogenesis in myogenic precursors which gave rise to SC. *Myogenin*, however, has never been reported as expressed in proliferating satellite cells precursors, suggesting that loss of methylation is a preparative step for a later expression. Finally, these results suggest that the reported methylated state of *Myogenin* at C2C12 MB by Lucarelli *et al.,* and Oikawa *et al.,* could be artifactual, probably due to the large scale *in vitro* cultures.

2.3 DNA methylation profile of terminal differentiation myogenic genes: *Myosins* and *Ckm*

The upstream MRFs recognize bHLH protein-binding sites located at the promoter regions of late myogenic genes (*Ckm, Myosines, Tropomyosines*, etc), which expression confers the muscle phenotype to the muscle fibers. In our study, we selected as terminal differentiation myogenic markers creatine kinase muscle (*Ckm)* gene and three Myosins class II genes located consecutively in the *Myh* cluster (chr11): the adult *Myh1* and *Myh4* and the perinatal *Myh8* (Cosgrove *et al.,* 2009). All the *MYH* class II and the *Ckm* genes present a CpG-poor content at their promoter regions, raising the possibility of being regulated by DNA methylation. At first, we proceeded to analyze the gene expression of *Myh1, Myh4, Myh8* and *Ckm* during the myogenic progression and in non-myogenic samples (**Figure R30**). For *Mhy1* and *Ckm* we observed an expression exclusively in late differentiated muscle cells, whereas for *Myh4* and *Myh8* we detected expression in the cardiomyocytes (HL1) too.

To perform the DNA methylation analysis, we selected the regulatory regions of the *Myosins* containing E-box- and MEF2 binding sites and the promoter and distal muscle specific enhancer of *Ckm* (**Figure R31**). Sodium bisulphite sequencing of the promoter regions revealed very high methylation levels in ESCs, which become almost totally demethylated in muscle cells (MBs/MTs/MFs). Interestingly, *Ckm* enhancer was

demethylated in all analyzed cells. Non-myogenic cells showed different methylation patterns, being MEFs always methylated whereas *Ckm* regions, *Myh1* and *Myh8* were demethylated in NPC and HL1. *Myh4*, in contrast, was methylated at NPC and HL1 cells.

Globally, DNA methylation analysis showed that muscle specific regulatory regions were heavily methylated in ESC and non-myogenic cells, whereas they were demethylated in muscle cells. Interestingly, our results pointed out that DNA demethylation occurs before gene expression, suggesting the acquisition of a transcriptional poised state.

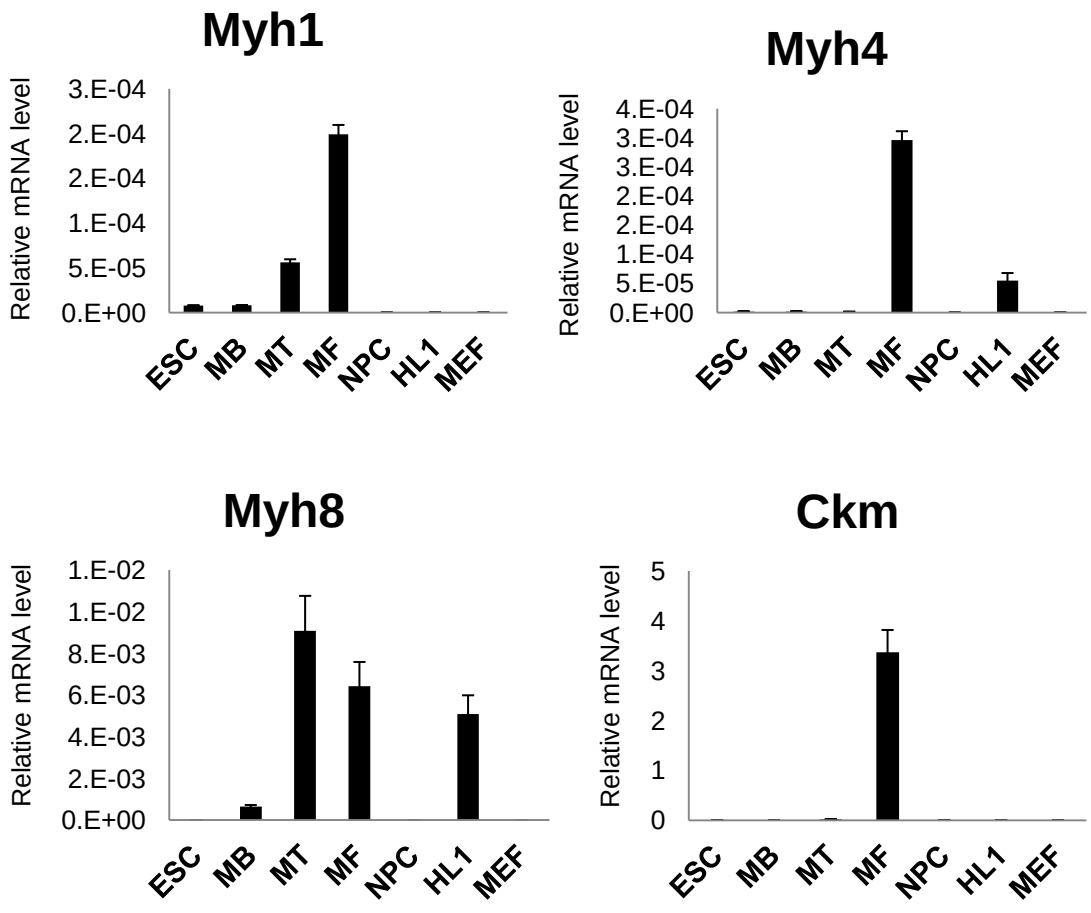


Figure R30. Expression profiling of terminal differentiation myogenic markers. qRT-PCR values of *Myh1, Myh4, Myh8* and *Ckm* expression normalized to *18S* expression during the myogenic progression. Data represent the average of three independent experiments ± SD. Non-myogenic cell lines were analyzed too (NPC, HL1 and Mef).

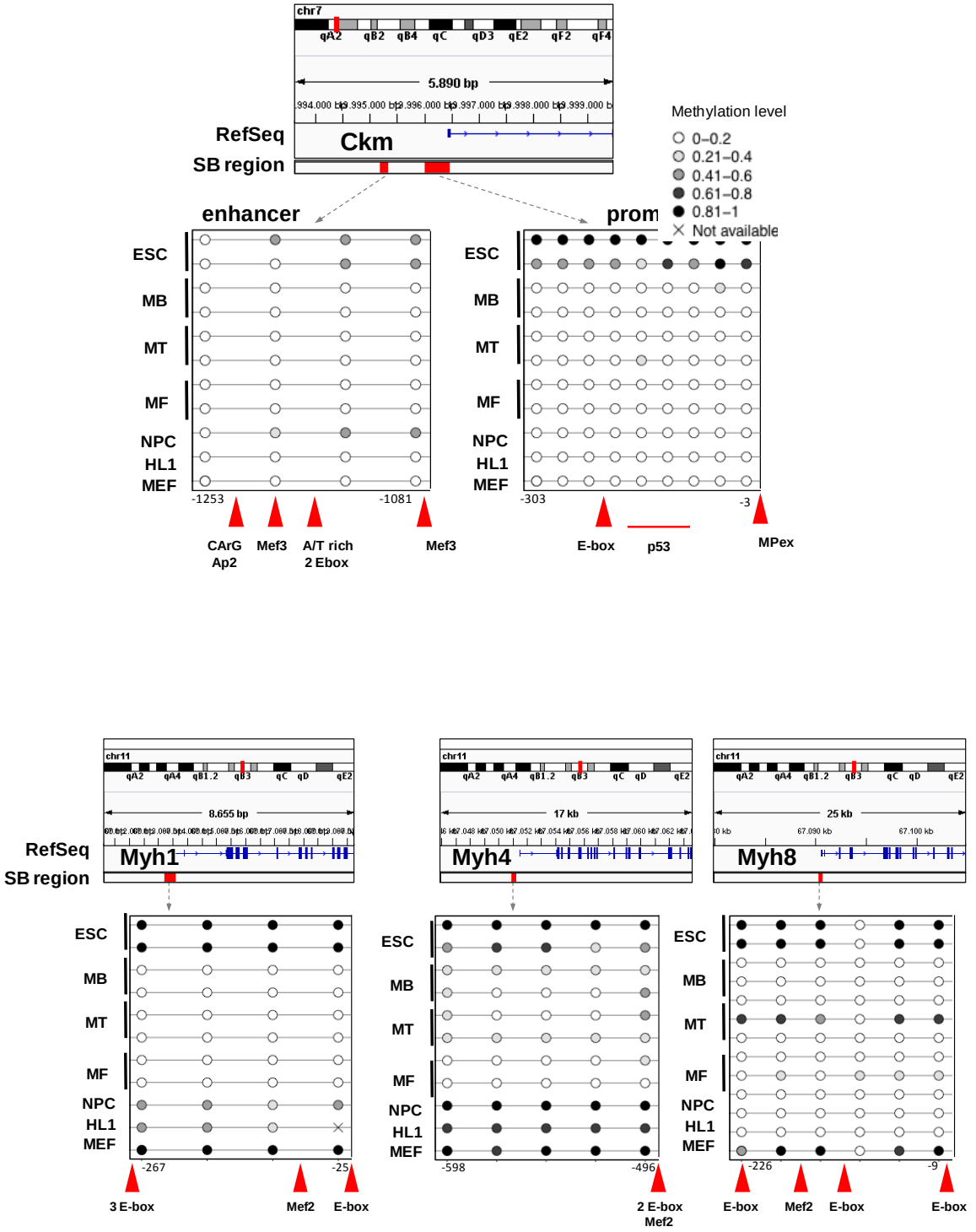


Figure R31. DNA methylation profiling of *Ckm, Mhy1, Myh4* and *Myh8* promoters during myogenesis. The DNA methylation levels of analyzed regions (determined by sodium bisulphite sequencing, SB region) are shown using circles charts. Each circle represents a CpG dinucleotide and its distance to the gene TSS is indicated below. The color gradient represents the percentage of methylation indicated in the legend. Transcription factor binding sites are indicated with red arrows.

2.4 Histone mark profiles of the principal myogenic regulatory genes

To complete characterize the epigenetic landscape involved in myogenic regulation, we analyzed the histone modifications profiles of the principal myogenic genes in ESCs, MBs and MTs using the same data sets as for *Myf5/Myf6* locus characterization (**Figure R32**). Firstly, we compared the histone profile of highly expressed genes in ESCs and silenced in MBs and MTs. We selected *Sox2* and *Fgf4* as genes with unmethylated CpG islands overlapping the promoter region and *Dppa4* and *Pou5f1* as genes with a poor-CpG content promoter that became hypermethylated upon differentiation. Sox2, Dppa4 and Pou5f1 are involved in embryonic stem cell pluripotency and Fgf4 is related to embryonic development. *Sox2* and *Fgf4* presented high levels of H3K4me3 in the promoter regions at ESCs and gained H3K27me3 mark in MB and MT cells correlating with gene silencing. *Dppa4* and *Pouf51* presented also the active mark H3K4me3 in the promoter region in ESCs, however in differentiating cells there was little (*Pou5f1*) or no gain (*Dppa4*) of H3K27me3.

Next, we analyzed the developmental genes *Pax3* and *Pax7* and the MRF *MyoD*, which all were expressed in MB stage and presented CpG island promoters. *Pax3, Pax7* and *MyoD* promoters showed a bivalent chromatin state (H3K4me3/H3K27me3) in ESCs, which resolves in the positive mark at MB and MT stage (H3K4me3) for *MyoD* and *Pax7,* whereas *Pax3* retained the bivalent state.

Analyzing the enhancers of *Pax3* and *MyoD*, an increase of H3K27Ac occurred in *Pax3* hypaxial enhancer at MB stage, whereas *MyoD* enhancer gained H3K4me1 occupancy in MBs and H3K27Ac and p300 in MTs. Since *MyoD* was already highly expressed in MB, this gain of active enhancer marks in MT might be involved in keeping an open chromatin state of the region to maintain the gene expression, rather than activate it.

After lineage determination leaded by *MyoD* and *Myf5* expression, the differentiation genes are expressed during the myogenic progression. These genes (*Myogenin, Mhy1, Myh4, Myh8* and *Ckm*) presented CpG-poor promoters that became demethylated in

the MB stage although the genes were still silenced. These myogenic differentiation genes were not bivalent at ESC stage and presented open chromatin marks (H3K4me3, H3K27Ac, H3K4me1 and p300) in the promoter and enhancer regions according to their expression timing. We hypothesized that the loss of DNA methylation prepared the gene to be later expressed, what we called transcriptional-poised demethylation, but ultimately was the acquisition of positive histone modifications what induced gene expression. Very interestingly, *Myh1*, *Myh4* and *Mhy8* gain H3K27me3 in MB cells after DNA demethylation, pointing out that the expression needs to be repressed until the terminal differentiation stage and this repression would be mediated by Polycomb repressive complex.

It has been proposed that housekeeping genes and tissue specific genes show different chromatin profiles (Ganapathi *et al.,* 2005) and specifically, it has been speculated that the histone modifications might regulate highly constitutively expressed genes, whereas tissue-specific genes are less occupied by histone marks (Poster presentation by Sílvia Pérez-Lluch, from Roderic Guigó's lab). After analyzing histone modifications of 12 housekeeping genes presenting CpG island promoter, (**Figure R33.A-B**), we observe high levels of H3K4me3 and lack of H3K27me3 in all analyzed genes in all samples. In addition, most of the genes showed binding of the coactivator p300 suggesting a constitutive open chromatin at the housekeeping promoters in the compared cell types. Therefore, we conclude that active genes, both modulated or consititutively expressed present high levels of active histone marks.

Collecting all these results together, the histone code correlates with gene expression, and importantly it revealed different signatures depending on the CpG content of the underlying sequence and the DNA methylation state, in line with the results observed in the genome wide approach.

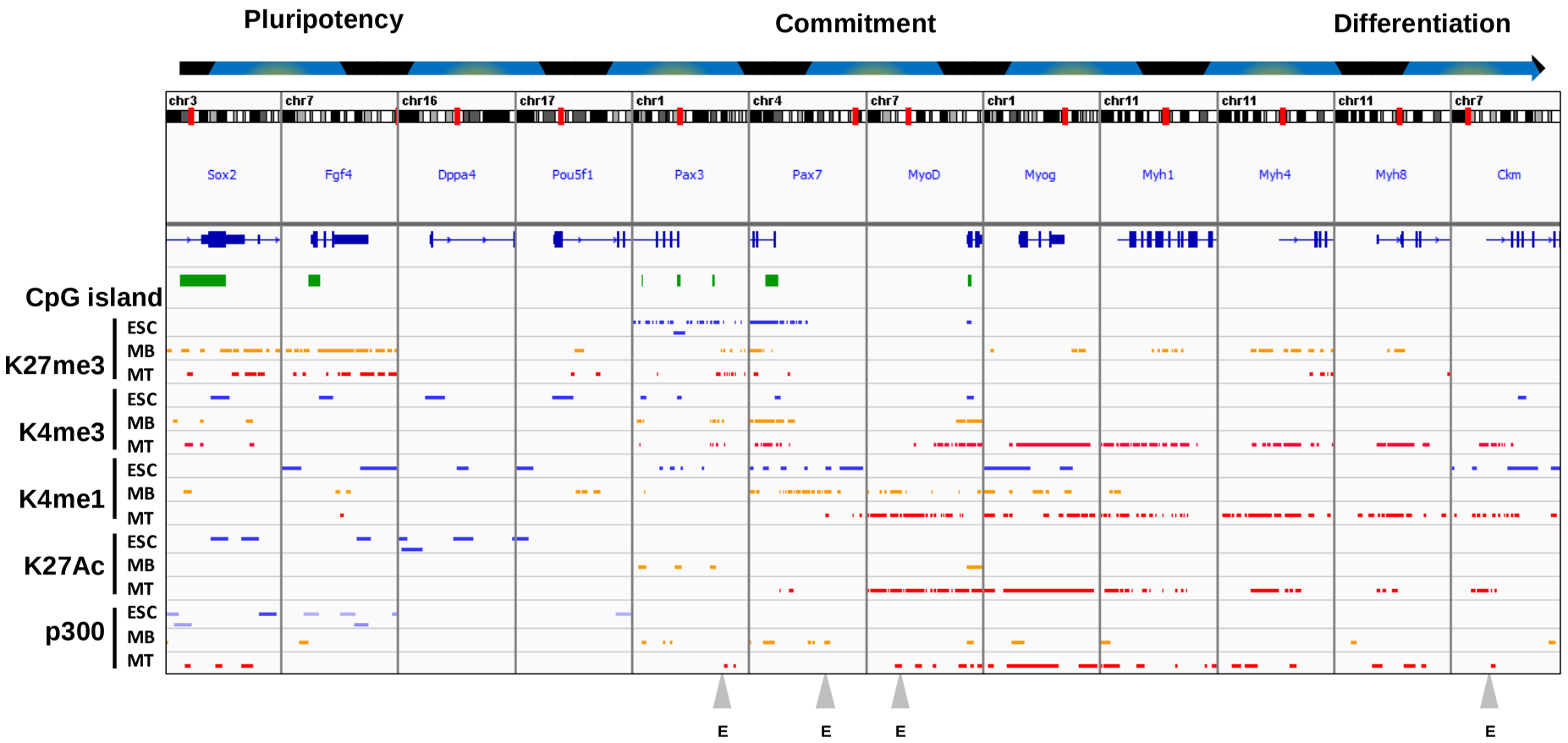


Figure R32. Chromatin marks of modulated genes during myogenesis. IGV visualization of the reference sequence of modulated genes during myogenesis, CpG islands and ChIP-seq peaks for H3K4me1, H3k4me3, H3K27me3, H3K27Ac and p300 at ESCs, (colored blue) MBs (colored orange) and MTs (colored red). Grey arrows indicate reported enhancer regions. The colored arrow indicates the differentiation progress, starting at embryonic expressed genes (blue) and ending at terminal myogenic differentiation expressed genes (red).

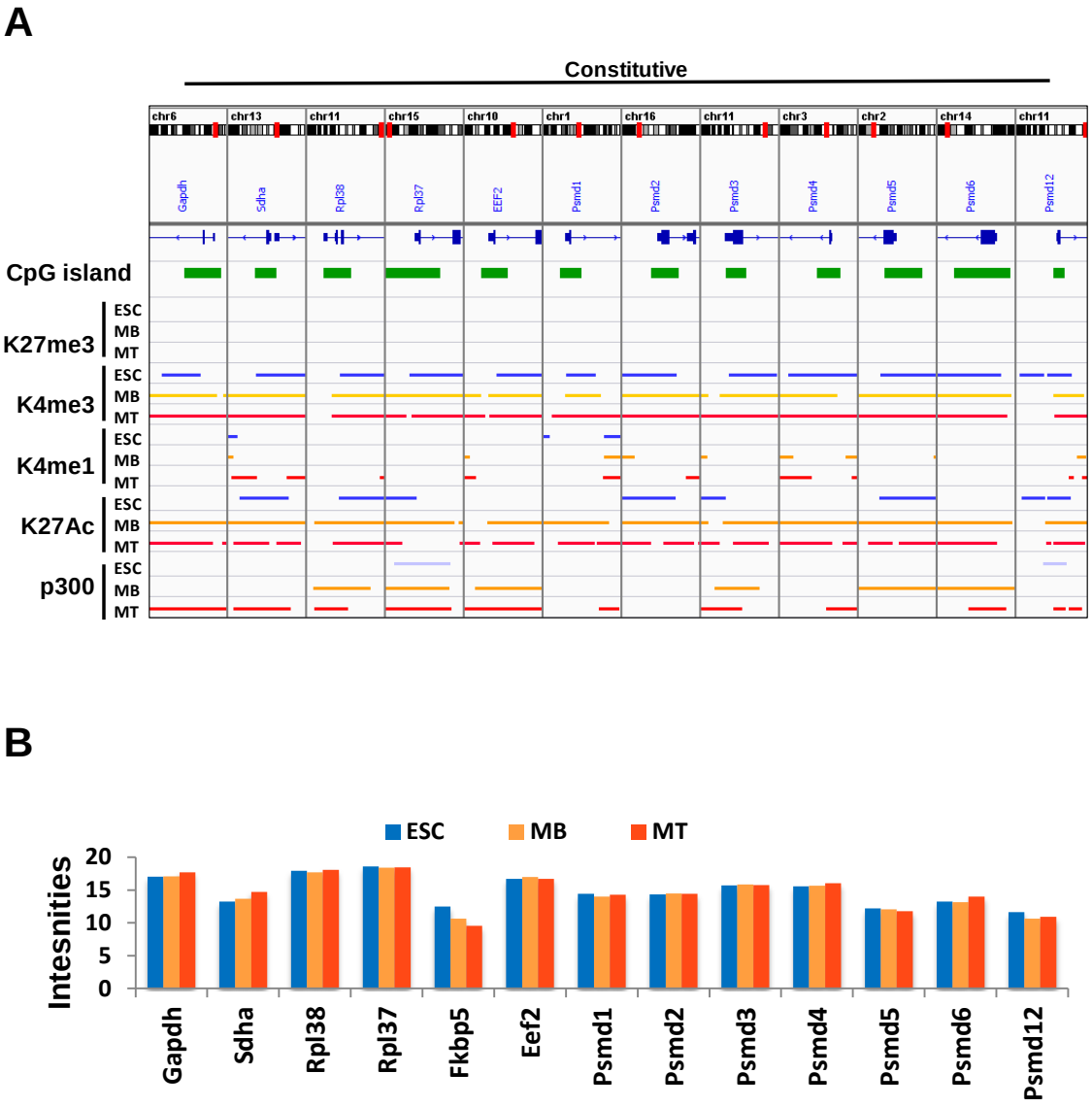


Figure R33. Chromatin marks of constitutively expressed genes. A IGV visualization of the reference sequence of modulated genes during myogenesis, CpG islands and ChIP-seq peaks for H3K4me1, H3K4me3, H3K27me3, H3K27Ac and p300 at ESCs (colored blue), MBs (colored orange) and MTs (colored red). **B.** Microarray expression intensity data of constitutively expressed genes in all analyzed samples.

3 STUDY OF THE DNA DEMETHYLATION DYNAMICS DURING THE MYOGENIC LINEAGE COMMITMENT AND THE UNDERLYING MECHANISMS

Our previous results identified the lineage commitment as the step when most of the DNA methylation changes occurr. Interested in deepen in the demethylation dynamics, we started a collaboration with Rita Perlingeiro's Lab (Lillehei Heart Institute, University of Minnesota, USA), because they had generated an inducible *Pax7* embryonic stem cell line (*iPax7*-ESC). This cell line gives rise to myogenic precursors with the ability to self-renew, expand, engrave, replenish the SC niche and is also able to respond to injury as resident satellite cells, generate myofibers and improve muscle function (Darabi, Santos, *et al.,* 2011). Therefore, *iPax7*-ESCs allows the generation of myogenic precursors that might be effective for the treatment of muscular dystrophies and provides a model that recapitulates *in vitro* the myogenic determination in a step wise manner.

Once validated the model through gene expression profiling of myogenic and developmental markers, we proceeded to evaluate the DNA methylation status of previously identified DMRs at successive time points. The methylation changes observed in ESC-derived MBs mimic the changes observed in primary MBs, reinforcing the accuracy of the model and exposing the importance of lineage-specific DNA methylation signature. Moreover, these results pointed out the model as an *in vitro* tool useful to study DNA demethylation mechanisms.

Thinking on an active mechanism driving DNA demethylation during myogenic differentiation, we proceeded to block putative DNA demethylases using shRNA strategies and to evaluate its effect on myogenic progression and DNA methylation.

3.1 ES-derived myogenesis after Pax7 induction

The ESC-derived myogenic model is based on the induction of *Pax7* gene in an engineered ESC line. This induction determines myogenic-committed pluripotency ESC into myoblast precursors, which can terminally differentiate in mature MTs (**Figure R34.A)**.

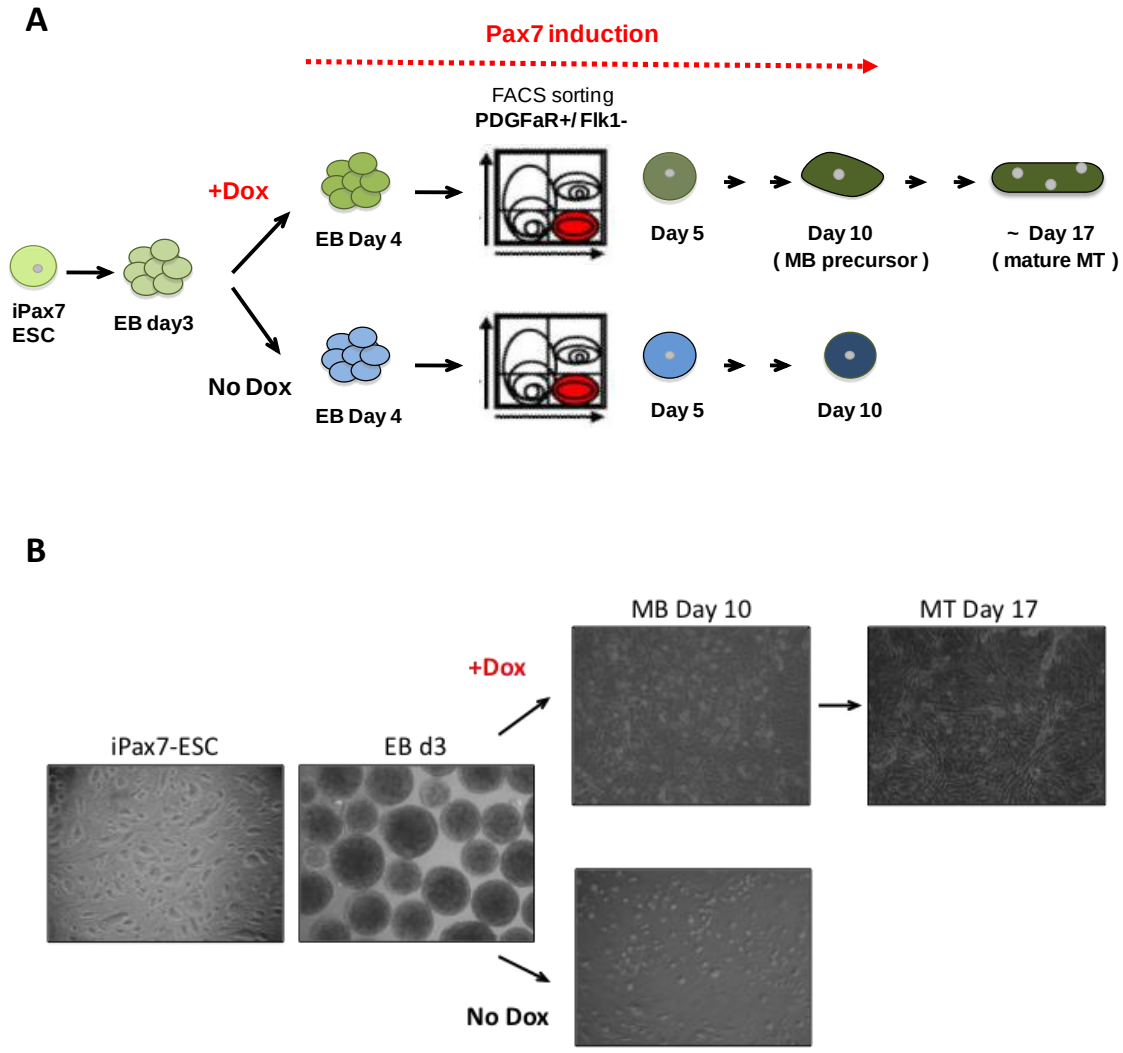


Figure R34. *Pax7*-induced ESC-derived myogenic model. A. Schematic overview of the *Pax7*-induced ESC-derived myogenic model. **B.** Representative fields of inducible *Pax7* ES cells at representative stages of the model progression. ESC, MB precursors at day 10 after EB differentiation and MT precursor at day 17 after EB differentiation were cultured in monolayer, whereas EB day 3 were grown in suspension. Representative fields of each state with 10x magnifications.

The inducible cell line was generated in Rita Perlingeiro's lab by introducing a P2lox vector carrying *Pax7* gene (1599bp) into a A2Lox.cre ES cell line, in which Cre enzyme is expressed upon Doxycycline (Dox) treatment, leading to *Pax7* expression (Darabi *et al.,* 2011).

iPax7-ESCs were amplified and co-cultured with irradiated MEFs in Lif (leukemia inhibitory factor) complemented medium to maintain their pluripotency capabilities. To induce embryoid bodies (EBs) differentiation *iPax7*-ESCs were grown in suspension without MEFs neither Lif (**Figure R34.B**). Next, the myogenic induction was performed by adding Dox at day 3 after embryoid body differentiation to activate *Pax7* expression.

Two days after *Pax7* induction (EB day 5), EBs were disaggregated to single cells and the PdgfαR^{+}/Flk^{-} fraction was sorted by flow cytometry (**Figure R35**). Platelet-derived growth factor alpha (PdgfαR^{+}) and lack of Fetal liver kinase 1 (Flk1^{-}) mark the paraxial mesoderm cells, where skeletal myogenic precursors arise from during development (Sakurai *et al.,* 2006). The sorted population was cultured in monolayer with Dox to maintain *Pax7* expression. After 2-3 passages this population reached the myogenic precursor stage (*iPax7*-MB precursor) representing day 10 after the EB differentiation. At that point, cells can be maintained as proliferating MBs or further differentiate to multinucleated myotubes (*iPax7*-MT) upon withdrawal of Dox. In our experiments myotubes were formed approximately at day 17 th after EB differentiation.

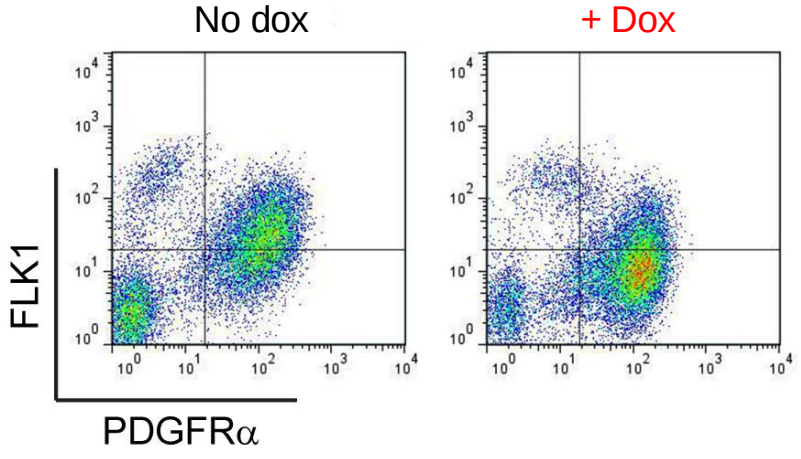


Figure R35. Dox induced *Pax7* expression during EB differentiation increased PdgfαR^{+}/Flk^{-} population. Representative FACS profile of *iPax7* ESCs at day 5 of EB differentiation. Dox was added to the EB medium from day 3 to day 5 of EB differentiation (right panel) or not added (left panel).

To control the specificity of the process, the same protocol was performed without Dox treatment. EBs cultured in suspension spontaneously differentiated into a heterogenic population where the PdgfαR^{+}/Flk^{-} population at day 5 after EB formation arise too (**Figure R35**). However, unlike Dox induced cells, the PdgfαR^{+}/Flk^{-} population was not the predominant fraction of the EBs at day 5 without Dox. Once the PdgfαR^{+}/Flk^{-} cells from not Dox induced EBs were sorted and cultured, the cells could not further proliferate neither differentiate, as previously documented (Darabi *et al.,* 2011). According to our collaborators, these cells have a cardiac potential and, eventually, cardiac beats were observed when EBs without Dox were further differentiate in a monolayer culture (Magli *et al.,* 2014).

3.1.1 Gene expression profiles during Pax7-induced ESC-derived myogenesis

We analyzed the expression levels of myogenic and pluripotency markers in four independent time courses treated or not with Dox. As positive control, not engineered ESCs and primary MBs were evaluated in parallel (**Figure R36**). Myogenic marker values were plotted in reference to *iPax7*-MBs (set to 1) whereas pluripotency markers were plotted in reference to *iPax7*-ESCs (set to 1). After Dox administration the *Pax7* induction occurred as expected, triggering an immediate *Myf5* expression followed by *MyoD* expression in *iPax7*-MBs (day 10). *Pax3* expression was not induced in this model because it is not a Pax7 downstream target and was not spontaneously expressed in EBs, as anticipated by our collaborators (Darabi *et al.,* 2008). Terminal differentiation markers (Myogenin, Mrf4, Myh8 and Ckm) were highly expressed in *iPax7*-MT stage, whereas pluripotency genes (*Pou5f1* and *Dppa4*) expression decreased during first stage of differentiation until negligible levels, as expected. The cells cultured without Dox did never express myogenic markers, but showed a reduction on the expression of pluripotency genes comparable to the one observed in the *Pax7*-induced cells. This result showed that,

even without myogenic induction, these cells tended to lose embryonic characteristics when cultured under differentiating conditions.

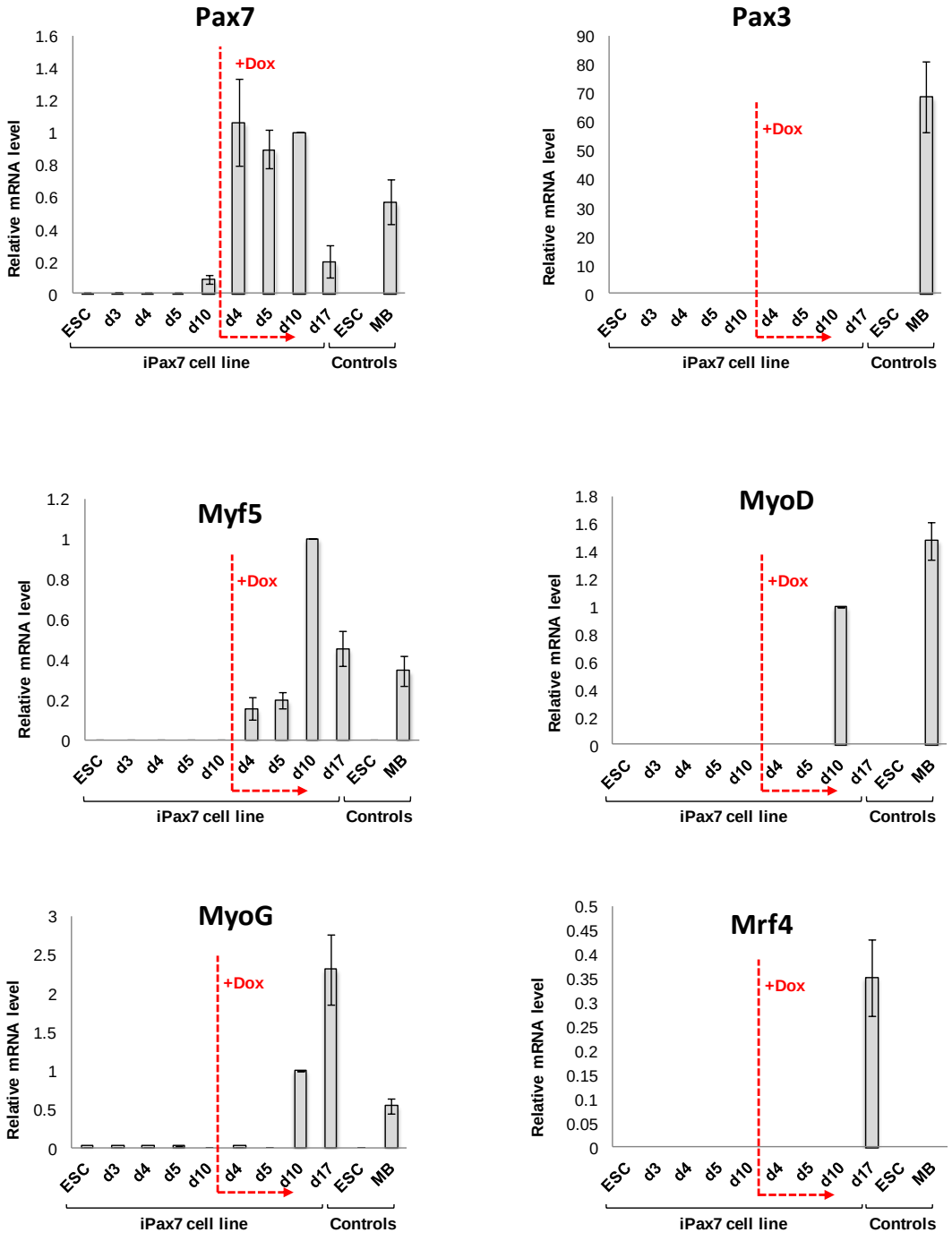
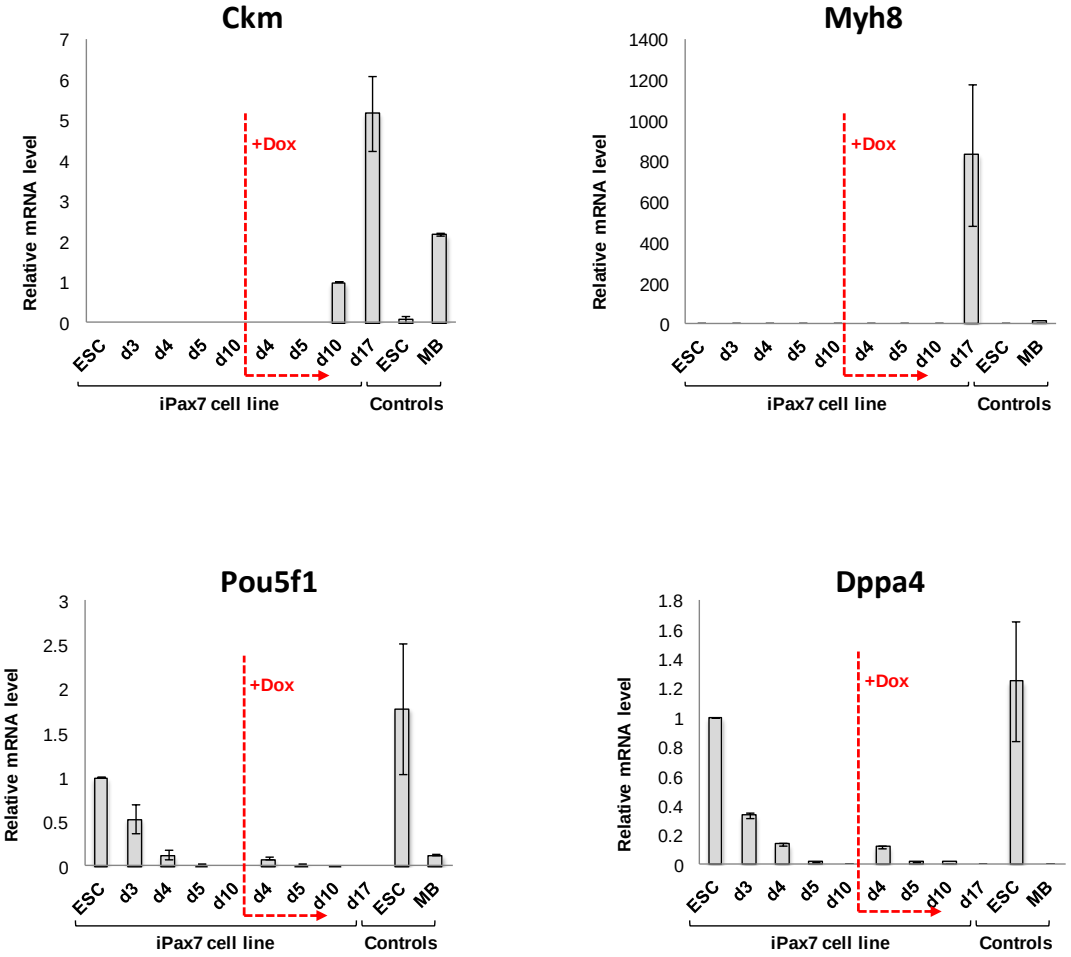


Figure R36 (continued on next page). Gene expression profiles during *iPax7* ESC-derived myogenesis. mRNA levels of myogenic (*Pax7, Pax3, Myf5*, *MyoD*, *Myogenin, Mrf4, Ckm* and *Myh8*) and embryonic markers *(Pou5f1* and *Dppa4*) at successive time points of *iPax7* ESC-derived myogenic model with and without Dox induction. As positive control not engineered ESCs and primary MBs were included to the analysis. Values were normalized with *Gapdh* expression and plotted in relation to *iPax7*-MB (for the myogenic markers) and *iPax7*-ESC (for the embryonic markers). Experiments were performed in quadruplicates, the average results are represented and standard error bars are shown.



3.1.2 DNA methylation profiles during Pax7-induced myogenesis

Once assessed the model at gene expression level, we proceeded to analyze the methylation status of the myogenic and developmental genes by sodium bisulphite sequencing. A preliminary analysis showed the expected complete unmethylated state of the CpG islands located at the *Pax3* and *Pax7* promoter regions (data not shown) evidencing the characteristic lack of methylation of CpG-rich regions. Subsequently, we centered our study on comparing the methylation status of the previously identified DMRs at *iPax7*-ESCs and at the derived myogenic progenitors.

In *iPax7*-ESCs, the *Pax3* hypaxial enhancer was moderately methylated and became completely demethylated at day three after EB formation before *Pax7* induction, despite the lack of *Pax3* expression at any stage of the model (**Figure R37**). This early event, independent of *Pax7* induction, could suggest that this enhancer region might play a role regulating *Pax3* in other lineages too. Contrarily, the *MyoD* enhancer turned

out to be highly methylated until *iPax7* MB stage, when it became totally demethylated coinciding with the increased gene expression (**Figure R38.A**). At day 17 (*iPax7*-MTs) the demethylated state persisted, despite the dramatic reduction of *MyoD* expression. *Myogenin* promoter showed a similar dynamic that *MyoD*, revealing a total loss of DNA methylation at *iPax7*-MB stage when *Myogenin* became slightly expressed (**Figure R38.B**).

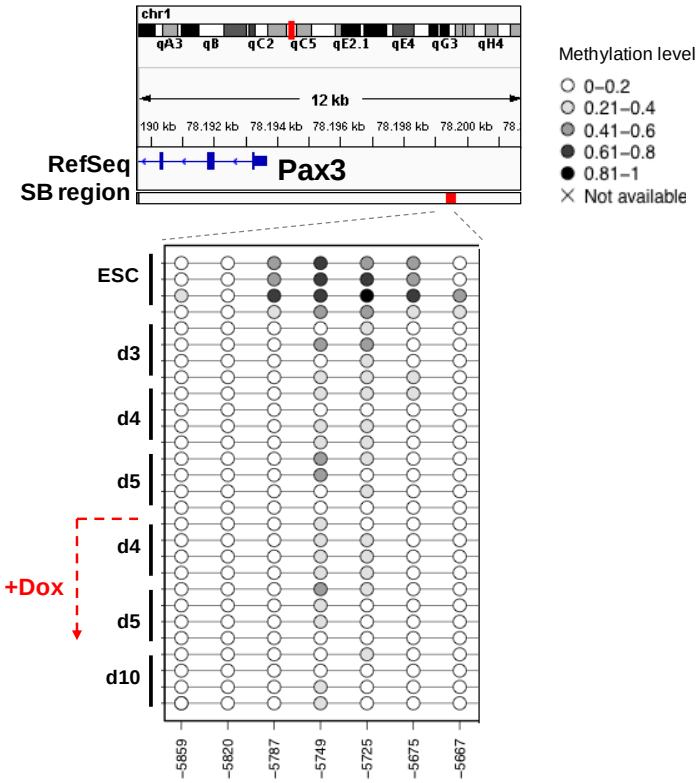


Figure R37. DNA methylation analysis of *Pax3* hypaxial enhancer region during *the iPax7* ESC-derived myogenic model. The DNA methylation levels of analyzed regions (n=4, determined by sodium bisulphite sequencing, SB region) are shown using circles charts. Each circle represents a CpG dinucleotide and its distance to the gene TSS is indicated below. The color gradient represents the percentage of methylation indicated in the legend.

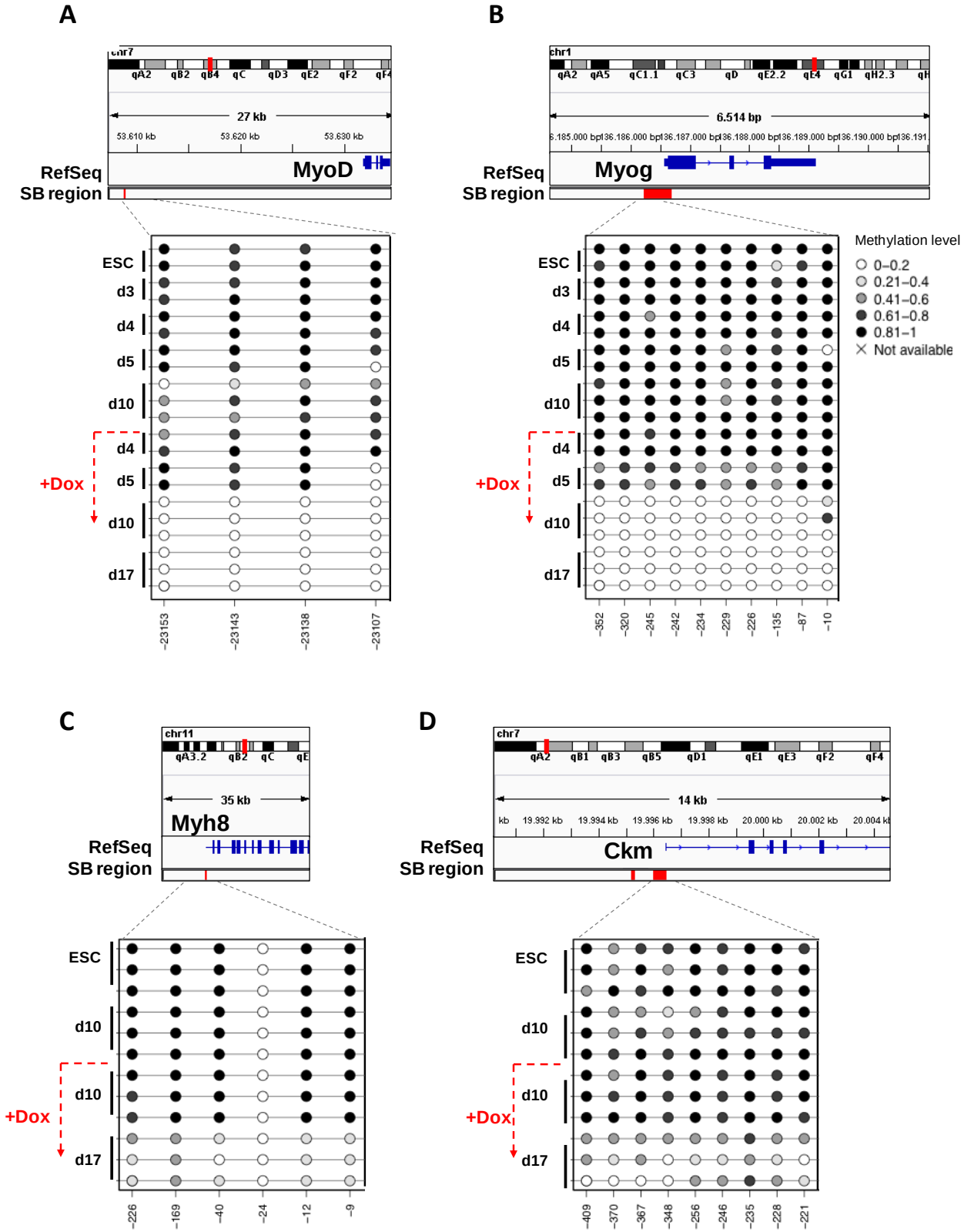


Figure R38. DNA methylation analysis of regulatory regions of *MyoD* (A), *Myogenin* (B), *Mhy8* (C) and *Ckm* (D) during the *iPax7* ESC-derived myogenic model. The DNA methylation levels of analyzed regions (n=3, determined by sodium bisulphite sequencing, SB region) are shown using circles charts. Each circle represents a CpG dinucleotide and its distance to the gene TSS is indicated below. The color gradient represents the percentage of methylation indicated in the

The demethylation was kept in *iPax7*-MTs, when the *Myogenin* gene reached the maximum expression level. However, *Myh8* and *Ckm* promoters were later on demethylated at *iPax7*-MT stage simultaneously with gene expression initiation (**Figure R38.C-D**). In summary, in the *iPax7* ESC-derived myogenic model we see a robust correlation of DNA demethylation coinciding with gene expression, except for *Pax3* gene. Importantly, we did not observe any loss of DNA methylation in the control cells, neither when the not induced PdgfαR^{+}/Flk^{-} cells were cultured in monolayer, indicating that the observed demethylation in myogenic genes is muscle-specific. Collectively, we observed three waves of DNA demethylation in our model: during early stages of development (Pax3), at *iPax7*-MB precursor stage (*MyoD* and *Myogenin*) and at *iPax7*-MT stage (Ckm and *Myh8*).

On the other hand, we analyzed the promoter regions of the pluripotency genes *Pou5f1* and *Dppa4,* wondering if they would show the same gain of methylation upon cell differentiation, as observed in primary myoblasts (**Figure R39.A-B**).

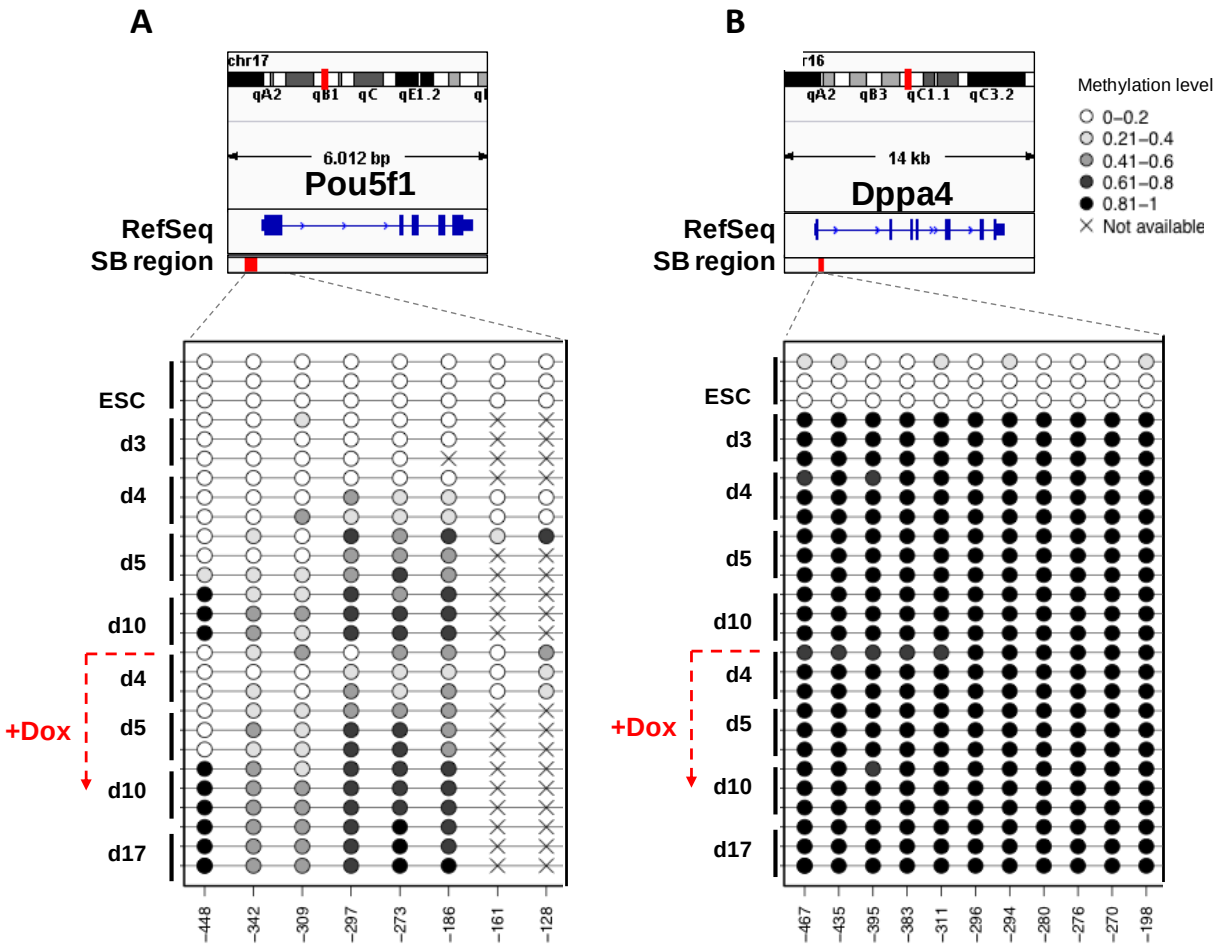


Figure R39. DNA methylation analysis of Pou5f1 (A) and Dppa4 (B) promoters during the *iPax7* ESC-derived myogenic model. The DNA methylation levels of analyzed regions (n=3, determined by sodium bisulphite sequencing, SB region) are shown using circles charts. Each circle represents a CpG dinucleotide and its distance to the gene TSS is indicated below. The color gradient represents the percentage of methylation indicated in the legend.

Pou5f1 showed a progressive gain of methylation starting at EB day 5 after EB differentiation independent of Dox induction, correlating with the transcriptional silencing observed at that stage. Interestingly, *Dppa4* became rapidly highly methylated at EB day 3 correlating with a dramatic reduced gene expression although the total silencing was at day 5 upon EB formation.

3.2 DNA demethylation mechanisms of muscle-specific genes during *Pax7*-induced myogenesis

DNA demethylation during lineage specification at myogenic regulatory regions has been clearly observed in primary myoblasts and in *iPax7* ESC-derived myogenic precursors. Next, we took advantage of the facilities that the *in vitro* inducible myogenic model offers and we addressed the driving mechanism of DNA demethylation.

3.2.1 Active demethylation mechanism during Pax7-induced myogenesis

At first we wondered if the observed demethylation could results from a passive process, in which the lack or the significant reduction of DNA methyltransferases (Dnmts) and the successive cell divisions would lead to the loss of this epigenetic mark.

We analyzed the mRNA levels of the Dnmts responsible of adding *de novo* methyl groups after replication (*Dnmt3a* and *Dnmt3b*) and *Dnmt1*, responsible of maintaining the existent methylation pattern (**Figure R40**). Again, we analyzed in parallel the expression in ES cells and primary MBs as positive controls and we established as reference *iPax7*-MBs expression values, set to 1. As shown in **Figure R40**, all stages expressed *Dnmt1* and *Dnmt3a* genes and we did not observed major expression changes that could explain the loss of methylation in myogenic regulatory regions. *De novo Dnmt3b* expression was dramatically dropped during

muscle-commitment, as anticipated due to its role in earlier developmental events according to its vital role during early development (Okano and Li 2002). These results together with the evidence of demethylation dynamics observed in differentiated myogenic cells (non-cycling cells) at *Ckm* and *Myh8* **(Figure 38.C-D**) suggested an active mechanism as the driving force of DNA demethylation.

As mentioned in the introduction, three families of enzymes could be involved in the active demethylation: Tet enzymes, deaminases AID/Apobec and o T/G glycosilases (see **Figure I9**). They represent three pathways that ultimately can modify 5mC, which is recognized by the base excision repair (BER) mechanism and replaced by an unmethylated cytosine. In order to evaluate the implication of all families in our model of DNA demethylation, we planned to knock-down a member of each family.

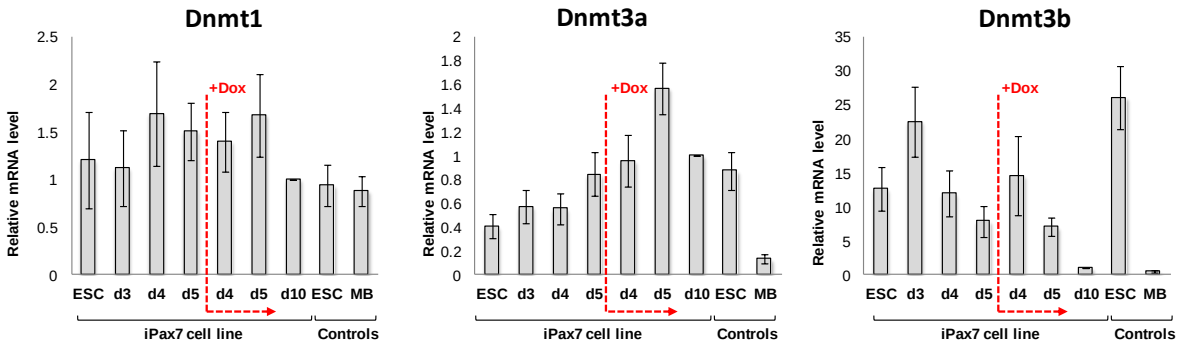


Figure R40. mRNA levels of DNMTs at successive time points of *iPax7* ESC-derived myogenic model with and without Dox induction. As positive control not engineered ESC and primary MB were included to the analysis. Values were normalized with *Gapdh* expression and plotted in relation to *iPax7*-MB. Experiments were performed in quadruplicates and standard error bars are shown.

3.2.2 Tet 1 knock-down in iPax7-ESC

The Ten-eleven translocation family has been related to DNA demethylation because of their capacity to oxidate 5mC into 5hmC, 5fC and 5caC, which are not longer recognized by Dnmt1 neither further modified by glycosylases coupled to the BER pathway.

The Tet proteins might play a dual role, on one hand preserving the unmethylated state of CpG islands (Jin *et al.,* 2014), whereas on the other hand they catalyze the demethylation of previous methylated regions (Vincent *et al.,*, 2013; Hackett *et al.,* 2013; Yamaguchi *et al.,*

2013). To evaluate their implications on myogenesis, we analyzed the gene expression pattern of the three members of the Tet family in the *iPax7* ESC-derived myogenic model (**Figure R41**). We observed increasing *Tet3* expression along the differentiation process, whereas *Tet1* and *Tet2* were mainly highly expressed in ESCs. Since the first detected DNA demethylation event occurred at day 3 after EB differentiation (*Pax3* enhancer, **Figure R37**) and *Tet1* was the highest expressed Tet gene in ESC, we hypothesized that Tet1 might be the enzyme starting the process. In addition, it has been reported in the literature that several muscle genes were dysregulated in *Tet1*$^{-/-}$ ESCs (Dawlaty *et al.,* 2011). To test it, we designed a stable knock-down experiment using shRNA against Tet1 mRNA in *iPax7*-ESC with the purpose of later on induce differentiation and evaluate the myogenic process.

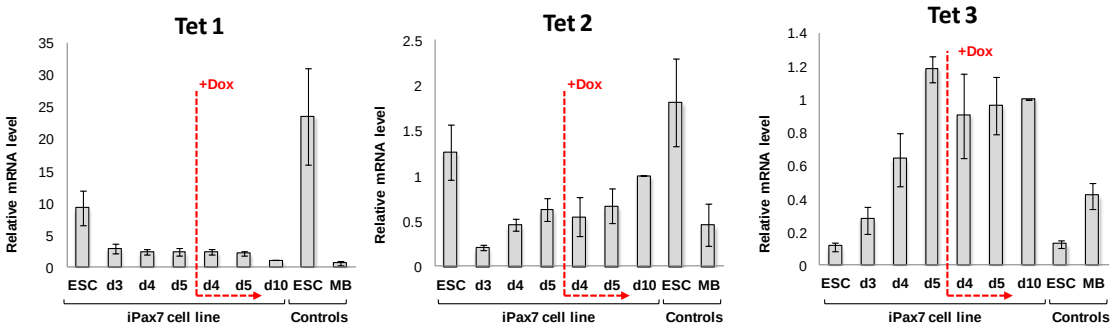


Figure R41. Gene expression levels of Tet family members at successive time points of iPax7 ESC-derived myogenic model with and without Dox induction. As positive control not engineered ESC and primary MB were included to the analysis. Values were normalized with *Gapdh* and plotted in relation to iPax7-MB. Experiments were performed in quadruplicates and

We infected *iPax7*-ESC with pGIPZ-GFP^{+} vector, a lentiviral construct that contained a bicistronic cassette expressing the sequence of interest as well as a green fluorescent protein (GFP) that allowed the selective analysis of the transduced cells. We evaluated eight different shRNA against Tet1 in order to choose the most effective sequence. Since the *iPax7*-ESCs were co-cultured with irradiated MEFs to avoid differentiation, we preplated and labeled the ESC whit the Stage-Specific Embryonic Antigen 1 (SSEA) Antibody before flow cytometry analysis, therefore ensuring the purification of *iPax7*-ESCs and not of transfected MEFs. Two days after the infection, the GFP^{+} SSEA^{+} population was very low, representing 0.5% of the total cells (**Figure R42.A**). We sorted this population and we proceed to expand part of them in co-culture with MEFs and, in parallel, we proceed to induce EB differentiation directly with the rest of cells. Despite the GFP^{+} fraction increased, the second flow cytometry analysis of

the ESCs revealed a low number of GFP+ (~ 5%) **(Figure R42.B)** and, unfortunately, the EBs generated from the GFP^+ $SSEA^+$ were totally depleted of the transduced population **(Figure R42.C**). After repeated attempts modifying the protocol without improving results, we did not proceed with this line of experiments.

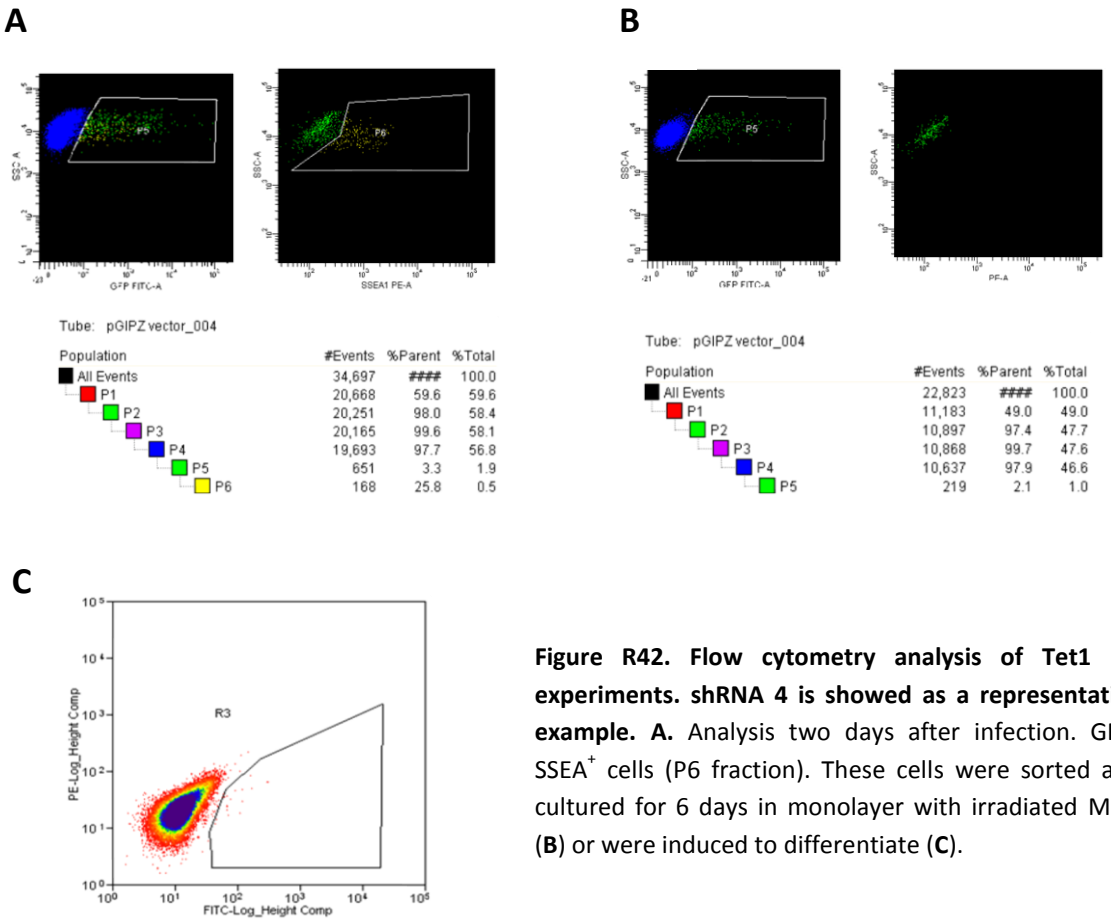


Figure R42. Flow cytometry analysis of Tet1 KD experiments. shRNA 4 is showed as a representative example. A. Analysis two days after infection. GFP^+ $SSEA^+$ cells (P6 fraction). These cells were sorted and cultured for 6 days in monolayer with irradiated Mefs (**B**) or were induced to differentiate (**C**).

3.2.3 Tdg knock-down in iPax7-ESC

The importance of T/G glycosilases during development is demonstrated by embryonal lethality of the KO mice (Cortázar *et al.,* 2011). To evaluate whether Tdg was involved in DNA demethylation during myogenesis we proceed to knock-down its expression. At first, we evaluated the expression of *Tdg* throughout the model confirming the transcription without major expression modulations (**Figure 43.A**). Next we proceed to down-regulate Tdg trying a set of five pLKO lentiviral vectors

containing different shRNAs against Tdg in *iPax7*-MBs precursors. Unfortunately, none of the tested shRNA showed a clear effect on *Tdg* expression after 3 trials (**Figure 43.B**). For a matter of time we did not persuade on trying new shRNAs, although it would be very interesting to study the role of Tdg on myogenic differentiation in the future.

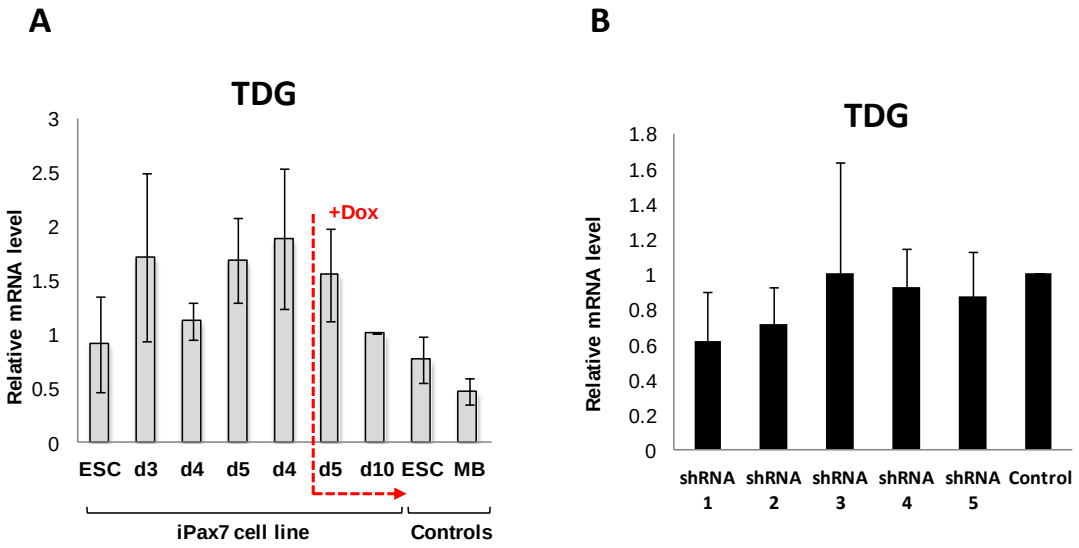


Figure R43. ***Tdg* expression during the *iPax7* ESC-derived myogenic model**. **A.** Gene expression levels of *Tdg* at successive time points of *iPax7* ESC-derived myogenic model with and without Dox induction are shown. Values were normalized with *Gapdh* expression and plotted in relation to *iPax7*-MB. Experiments were performed in quadruplicates and standard error bars are shown **B**. Average values of 5 different shRNAs against Tdg expression of three independent experiments plotted in relation to the control vector and normalized with *Gapdh.*

3.2.4 Apobec2 knock down in iPax7-ESC

Apobec family belongs to the activation induced cytidine deaminases (AID) enzymes, which besides of being involved in RNA editing, affects DNA methylation patterns by deaminating 5hmC (Guo *et al.,* 2011). Importantly, Apobec2 is the cardiac and skeletal muscle specific isoform. Prior to evaluate Apobec2 role in DNA demethylation during myogenesis, we analyzed its expression levels throughout the *iPax7*-ESC myogenic model and, as shown in Figure R45A, *Apobec2* expression levels were very high in MB stage (EB day 10) and derived MT (**Figure R44.A**).

Next, we tested four different pLKO lentiviral vectors containing different shRNA against Apobec2 in myoblasts. The shRNA 1 and 4 were the most effective and were the ones selected to perform the experiments in parallel with the control construct (**Figure R44.B**).

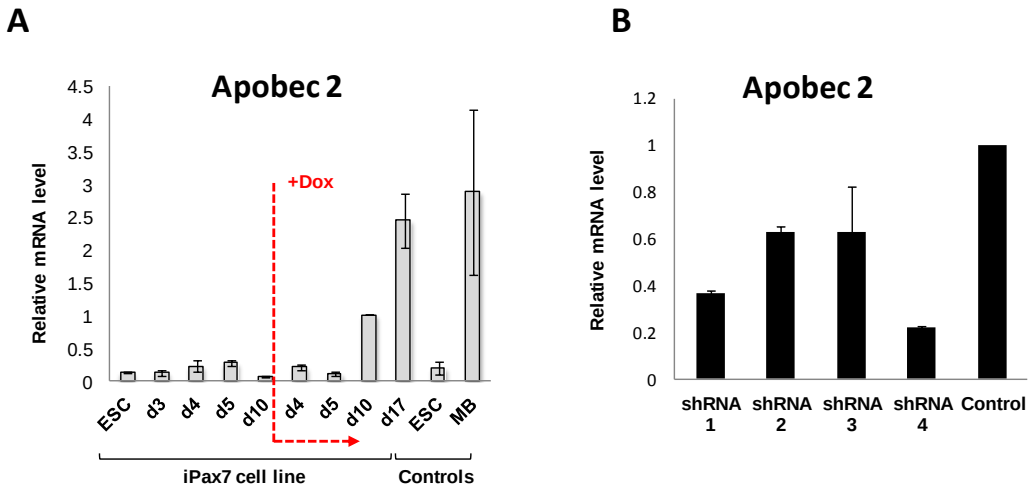


Figure R44. ***Apobec2* expression during the *iPax7* ESC-derived myogenic model**. **A.** Gene expression levels of *Apobec2* at successive time points of *iPax7* ESC-derived myogenic model with and without Dox induction are shown. Values were normalized with *Gapdh* expression and plotted in relation to iPax7-MB. Experiments were performed in quadruplicates and standard error bars are shown. **B**. Average values of 4 different shRNAs against *Apobec2* expression of two independent experiments plotted in relation to the control vector and normalized with *Gapdh.*

Since *Apobec2* is not highly expressed until MB stage, we decided to down-regulate its expression the day after platting PdgfαR^{+}/Flk^{-} sorted cells (EB day 6), and evaluate its effect on the second demethylation wave. However, although we knew from previous results that the second wave of demethylation occurred in *Pax7*-induced cells between EB day5 and MB precursor (EB day 10), we did not know the precise demethylation timing, neither if it was gradual or total at a certain point. In order to address that, we analyzed methylation and expression levels of daily time course starting from the EB day 5 PdgfαR^{+}/Flk^{-} sorted population until *iPax7* MB stage. We observed a perfect correlation between gradual DNA demethylation and increased gene expression for *MyoD* and *Myogenin* genes after Dox treatment at EB day 8 (**Figure R45.A-B)**. These results indicate that the original strategy to block *Apobec2* expression infecting EB day 6 cells with shRNAs was valid to analyze later on DNA methylation levels on target myogenic genes.

We proceeded to infect EB day 6 monolayers cells through lentiviral spin-infection, to harvest them and to evaluate the mRNA levels four days later (**Figure R46**). *Pax7* expression was

quantified to ensure a comparable myogenic induction between experiments. The shRNA 1 and 4 were highly effective, dropping the mRNA level to less than 10% and to negligible levels, respectively, in the triplicated experiments. The downregulation was even higher than when the shRNAs were tested due to an increment on the amount of virus on the triplicate independent experiments. Interestingly Apobec2 KDs showed a dramatic effect on *MyoD* and *Myogenin* expression, specially the construct carrying the shRNA 4. To test whether *Apobec2* could influence terminal myogenic differentiation, we induced myotube formation to the infected cells by culturing them in confluence without Dox. By immunostaining we could observe an absolute absence of the differentiation markers MHC and Myogenin in the KD compared to the controls (**Figure R47**).

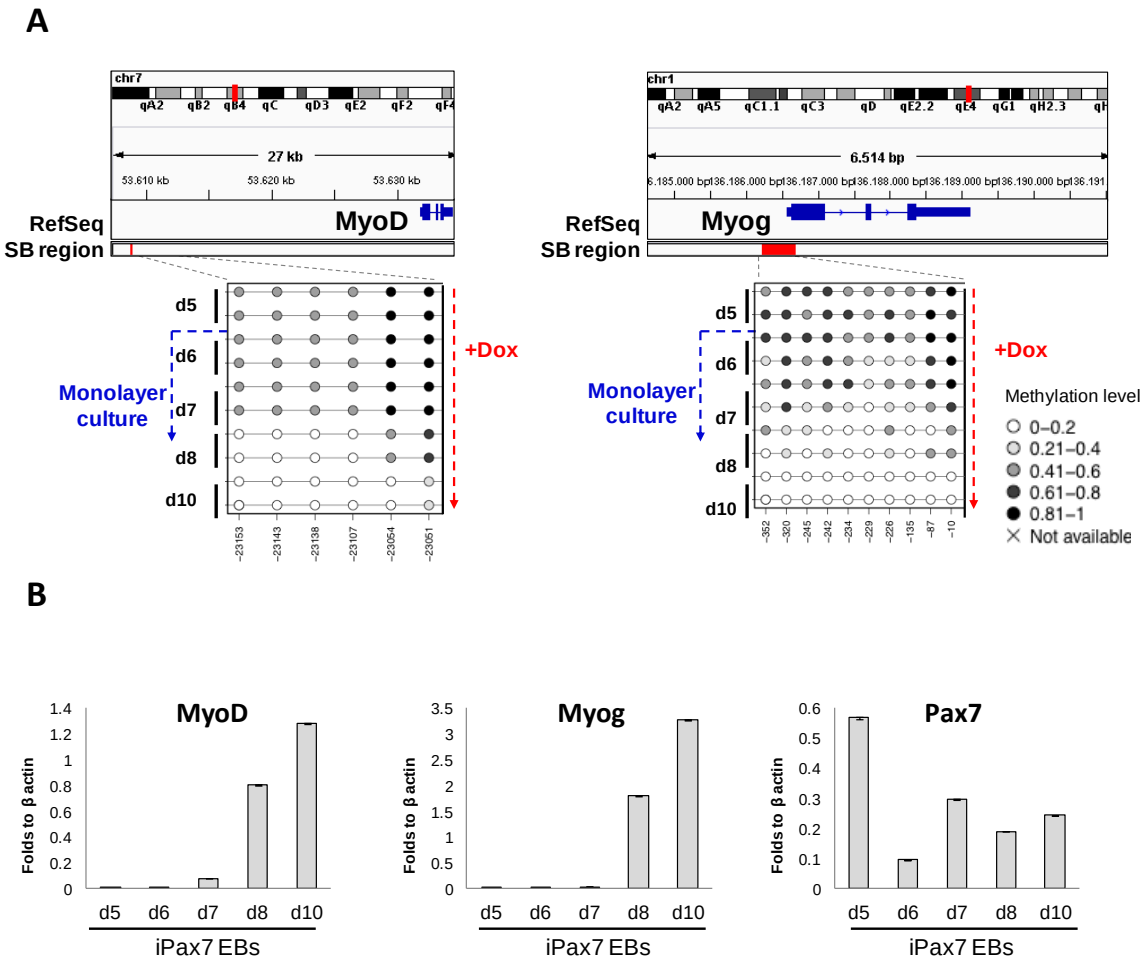


Figure R45. DNA methylation and gene expression analysis of *MyoD* and *Myogenin* in a daily time course after Dox induction in the *iPax7* ESC-derived myogenic model. A. The DNA methylation levels of analyzed regions (n=2, determined by sodium bisulphite sequencing, SB region) are shown using circles charts. Each circle represents a CpG dinucleotide and its distance to the gene TSS is indicated below. The color gradient represents the percentage of methylation indicated in the legend. **B.** qPCR expression analysis of *MyoD, Myogenin* and *Pax7* were normalized to *Gapdh* expression. The experiments were performed in duplicates.

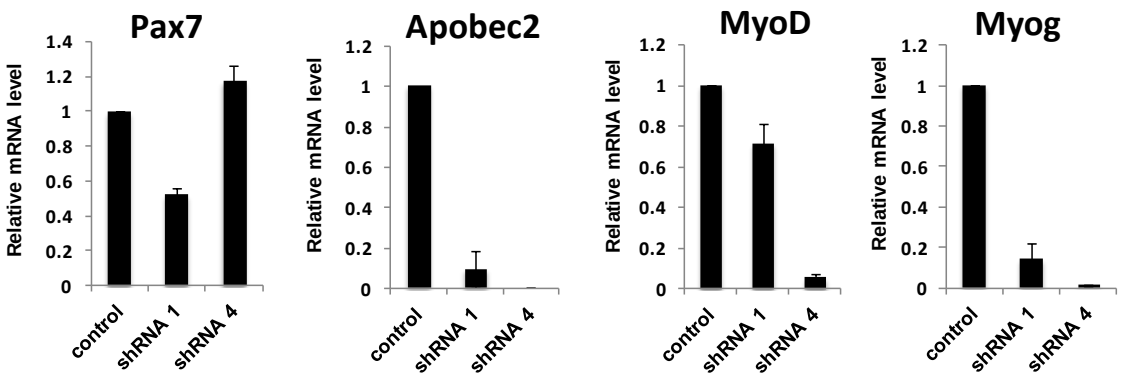


Figure R46. ***Apobec2* knock down affected myogenic progression**. qPCR expression analysis of *Pax7, Apobec2, MyoD and Myogenin* after Apobec2 shRNA experiment. Experiments were performed in triplicates and the average results normalized to 18S expression with standard error bars are shown.

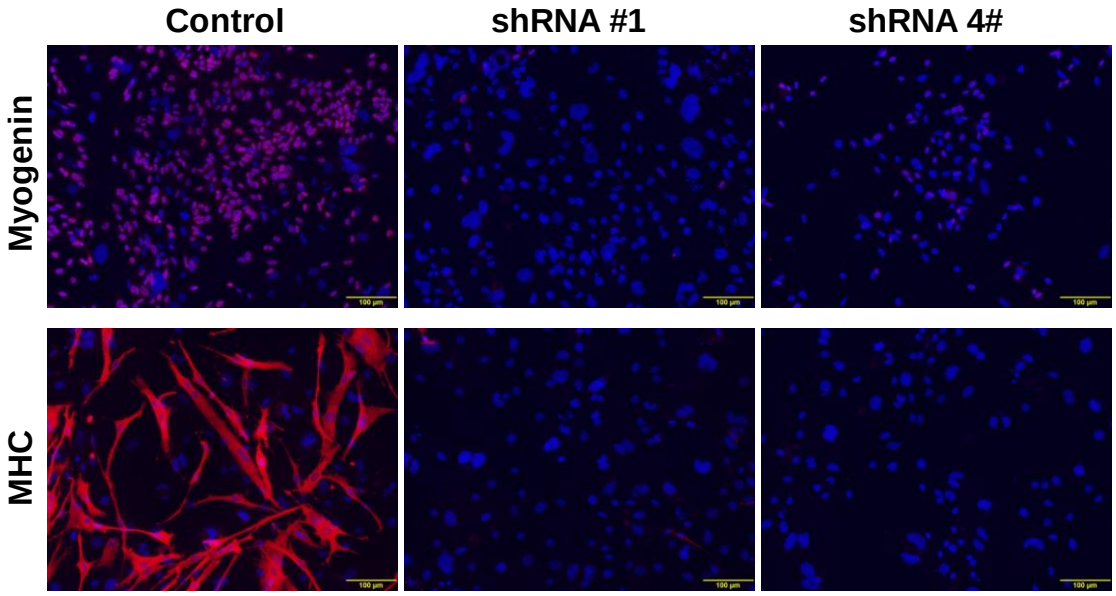


Figure R47. **Apobec2 knock down abolished myogenic differentiation.** Representative images of induced *iPax7*-MB derived cells under differentiation conditions stained for Myogenin (upper panel) and MHC (lower panel). Cells were co-stained with DAPI (blue). Scale bar was 100 µm. This experiment was repeated three times.

Finally, we analyzed the methylation status of *MyoD* enhancer and *Myogenin* promoter in *Apobec 2* knock down *iPax7*-MB in three independent experiments (**Figure R48.A**). *MyoD* promoter showed slight increased DNA methylation levels in the knock-down cells, despite no significant differences were detected. However and very importantly, both shRNAs against Apobec2 impaired *Myogenin* demethylation, being shRNA 4 the most effective one. All CpGs of the *Myogenin* promoter were significantly more methylated in shRNA 4

transfected cells compared to the control cells, and 4 out of 10 CpGs were also significantly more methylated in shRNA 1 transfected cells (Kruskal-Wallis test, p- val<0.05) (**Figure R48.B**). Although further studies are needed to clarify whether this lack of DNA demethylation depends on the deaminase activity, these results suggest that Apobec2 could participate actively in DNA demethylation process during myogenesis.

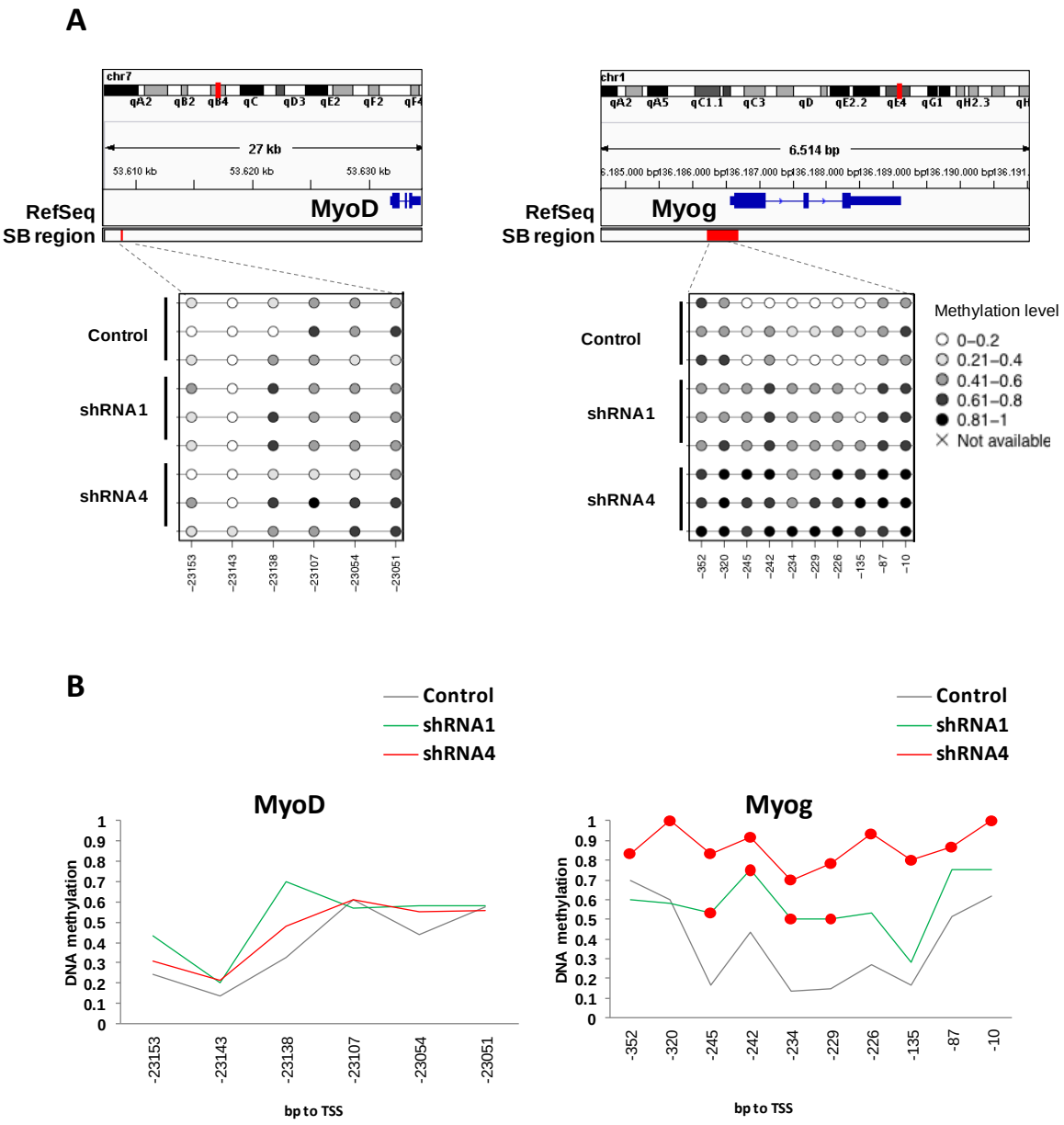


Figure R48. ***Apobec2* knock-down impairs myogenic associated DNA demethylation. A.** *MyoD* enhancer and *Myogenin* promoter DNA methylation analysis of cells infected with control, Apobec2 shRNA 1 and Apobec2 shRNA 4 constucts. The DNA methylation levels of analyzed regions (n=3, determined by sodium bisulphite sequencing, SB region) are shown using circles charts. Each circle represents a CpG dinucleotide and its distance to the gene TSS is indicated below. The color gradient represents the percentage of methylation indicated in the legend. **B.** Line plot representing the average DNA methylation values of three independent experiments. Red colored circles represent significant differences (p-value < 0.05) respect the control samples

DISCUSSION

Myogenesis is the differentiation process which encompasses the formation of skeletal muscle during development and tissue homeostasis throughout life. In addition, it also takes place upon severe acute injury in order to regenerate the damaged muscle tissue. Arising from a pluripotent stem cell during embryogenesis or from multipotent muscle stem cells after birth, the myogenic process comprehends the acquisition of a specialized and defined cell identity and the loss of pluri/multipotent and proliferative capacities. Importantly, myogenesis is altered in chronic muscle pathologies (dystrophies), in cancer (rhabdomyosarcoma), and in many clinical situations including chronic disease, malnutrition, disuse, cachexia and denervation and during aging (sarcopenia), leading in all cases to muscle wasting, mobility problems, low life quality and ultimately, in the most severe cases such as Duchenne muscular dystrophy (DMD), to premature death (reviewed in Emery, 2002; Bodine, 2013; Wall *et al.,* 2013; Arndt and Crist, 1999). While in some cases well known molecular mechanisms lead to the pathological outcome, such as chromosomal translocations in alveolar rhabdomyosarcoma (aRMS) or the non functional dystrophin protein in DMD, other muscle pathologies different potential mechanisms trigger muscle wasting.

Cellular identity is defined by specific transcriptional programs directed and maintained by transcription factors and epigenetic mechanisms, which generate the diverse cell types of the organism beginning from a common and unique underlying genomic sequence. Focusing on DNA methylation, it has been correlated with transcriptional gene silencing despite it might have other roles depending on the genomic context (see Introduction section 3.1.2). Recent studies correlated the epigenetic signatures with cell identity and differentiation stages in several lineages (hematopoietic, neuronal, skin and hepatic, see Introduction section 4). Strikingly, despite for seminal works studying the DNA methylation regulatory role on MyoD and Myogenin muscle transcription factors, no studies have investigated the DNA methylation dynamics at the genome-wide scale during myogenesis, with the exception of two papers published very recently (Miyata *et al.,* 2014; Tsumagari *et al.,* 2013) and discussed below.

1 GENOME-WIDE STUDY OF DNA METHYLATION DYNAMICS DURING MYOGENESIS

1.2 AIMS-seq method: technical considerations

Determination of DNA methylation can be performed with a broad range of methodologies. Most of them are based on one of these three main strategies: methylation sensitive endonucleases digestion, affinity-based methylated DNA enrichment and bisulphite conversion, which can be followed by gel-based analysis, array-probe hybridization and sequencing and ultra-sequencing methods to reveal the location of the 5mC residues (reviewed in Laird, 2010 and Jordà and Peinado, 2010).

Few years ago, our lab developed a method to identify differentially methylated sequences based on the use of methylation sensitive enzymes coupled to two-dimensional gel electrophoresis analysis, the AIMS method (reviewed in Jordà *et al.,* 2009). After demonstrating the high sensibility of AIMS method by performing serial dilutions of colorectal tumor DNA (Frigola *et al.,* 2002) successive works in our lab showed the capacity of AIMS method to detect DMRs in tumoral processes (Rodriguez *et al.,* 2006; Frigola *et al.,* 2005). Later on, we took advantage of the next generation sequencing possibilities, and developed the AIMS-seq method increasing therefore the DMRs detection potential (Forn et al, submitted). We mapped AIMS readout with next generation sequencing instead to use array-probe hybridization platforms because it is a more flexible method, need less input DNA and avoids hybridization artifacts.

In addition, AIMS-seq comparatively interrogated the DNA methylation state of CpG dinucleotides distributed across the genome and located in all the genomic compartments independently of the CpG density. This represents an advantage in front of affinity enrichment methods, as MeDIP-seq which is biased toward CpG rich regions (Weber *et al.,* 2007; Mohn *et al.,* 2009).

Considering the genome-scale bisulphite conversion methods, in humans it has been developed the 27K and 450K Illumina methylation platforms that allow genome-wide analysis of DNA methylation of bisulphite converted DNA. However, these platforms have not been yet developed for murine samples. Recently, it has been developed the whole genome bisulphite sequencing (WGBS) technique (Laurent *et al.,* 2010; Lister *et al.,* 2009), which potentially surveys the methylation sate of all genomic CpGs. However, it is an expensive technique that requires complex bioinformatic analyses.

A simplification of the WGBS technique is the reduced representation bisulphite sequencing (RRBS) method (Meissner *et al.,* 2005), based on the use of restriction enzymes followed by bisulphite treatment and next-generation sequencing, which has emerged as the gold standard method to address DMRs between samples (Consortium 2012). Therefore, we evaluated the robustness of our method by comparing the results obtained with AIMS-seq with the results obtained with RRBS on ESCs (Meissner *et al.,* 2008), and notably we observed a clear positive correlation (**Figure R5**).

Importantly, although methylation sensitive restriction enzyme based methods, as AIMS-seq, are considered indirect DNA methylation reporters, our lab recently demonstrated the accurate predictive value of the CpG methylation state inside the cleavage site for HpaII (5'-C<u>CG</u>G-3') in relation to the global methylation status of the embedding CpGs inside the CpG islands (Barrera and Peinado, 2012). This work allowed us to assume a homogeneous distribution of DNA methylation within the analyzed genomic regions. Finally, randomly selected AIMS-seq DMRs were correctly validated by bisulphite sequencing and replicates comparison proved the reproducibility of the method.

It's worthy to mention that none of the cited methods can discriminate between 5mC and its oxidative products 5hmC, 5fC and 5caC. However, since only 1% of methylation in somatic tissue and ~5% in ESC is 5hmC, and 0.03% and 0.01% is 5fC and 5caC respectively (Ito *et al.,* 2011; Zhang *et al.,* 2012; Pfaffeneder *et al.,* 2011), we considered very small the possible bias in the results obtained in this Thesis by AIMS-seq and bisulphite conversion methods.

Taken all together, the AIMS-Seq method is a simple, robust, reliable and cheap comparative method useful to identify non-targeted DMRs between samples.

3.1 Genome-wide identification of differentially methylated regions during myogenesis

Before identifying the DMRs along the myogenic process, we characterized the genomic regions susceptible to be analyzed by the AIMS-seq method. We classified the AIMS-seq amplicons as informative (Positive) when they were amplified by AIMS-seq and non-informative (Negative) when they were not amplified. Consequently, we considered the Positive compartment as the part susceptible to be regulated by DNA methylation, while the Negative compartment was carefully interpreted as regions depleted of methylation or not amplified due to technical issues. After characterizing Positive and Negative regions, the data supported the previous notion that CpG-rich and -poor regulatory elements undergo distinct modes of epigenetic regulation (**Figure R6**) (Meissner *et al.,* 2008; Weber *et al.,* 2007; Mikkelsen *et al.,* 2007). DNA methylation changes occur mainly at poor CpG-poor regions (Positive regions), at gene bodies and intergenic regions, while regions with rich and intermediate CpG content (Negative regions), comparatively enriched in LINEs and promoters, were more prone to be regulated by histone modifications, specially by the CpG island-associated histone modification H3K4me3 (Deaton and Bird, 2011).

Due to the increasing concern regarding whether DNA methylation could be altered during large-scale cell cultures (Meissner *et al.,* 2008; Smiraglia *et al.,* 2001; Shen *et al.,* 2006; Nazor *et al.,* 2012), prior to proceeding to DMRs identification during myogenesis, we compared the methylation profiles of primary freshly isolated MBs with cultured MBs, and importantly, we did not detect significant differences (**Figure R7**).

Comparing the DNA methylation profiles between ESCs, MBs, MTs and MFs, the methylome was mainly preserved during myogenic progression and only 2% of

amplicons were differentially methylated across the process. Interestingly, while prior reports proposed that CpG island shores were hotspots of informative DNA methylation differences between samples (Irizarry *et al.,* 2009), we observed that all DMRs, independently of CpG island proximity, correlated with cellular identity (**Figure R12.A**). Most of the changes occurring during muscle-lineage commitment consisted of hypermethylation events that were stably maintained during the muscle differentiation process, A smaller number of regions became hypomethylated throughout the myogenic progression too. Importantly, all the DNA methylation changes observed where unidirectional. These observations are in agreement with DNA methylation dynamics observed in neuronal and hepatocyte lineage commitment and terminal differentiation models (Mohn *et al.,* 2008; Brunner *et al.,* 2009; Kim *et al.,* 2011). The identified DMRs were found mainly at gene bodies and intergenic regions, being the demethylated DMRs enriched in promoters and exon-intron junctions, while hypermethylated DMRs were specially enriched at intergenic regions.

Globally, a low correlation between DMRs and gene expression changes of the closest annotated gene was found: 19% of the genes showed a negative correlation between expression and DNA methylation coinciding with a gain of active transcription marks (Pol II and H3K4me3) at hypomethylated regions. This modest inverse correlation, together with similar results observed in other models (Mohn *et al.,* 2008; Bock *et al.,* 2012), are explained by the fact that DNA methylation contributes to cell-type specific repression for a sizable number of genes throughout the genome (Hemberger *et al.,* 2009). In addition, DNA methylation not only might modulate the closest annotated gene, but might influence cell identity by regulating distal regulatory regions too.

Given the relatively small number of genes with overlapping DNA methylation and gene expression changes, we reasoned that genes exhibiting consistently negative association between these two properties may constitute strong candidates for a cell type specific functional role. Indeed, among the demethylated DMRs at promoter regions accompanied with increased gene expression, we identified genes related with muscle function, as Fbn1 and Wnt11, among others (**Figure 16**). Fnb1 is implicated in extracellular matrix structure and calcium ion binding (Davis and Summers, 2012) and Wnt11 is implicated in tissue development and muscle fibers elongation (Uysal-

Onganer and Kypta, 2012; Gros *et al.,* 2009). These observations were opposite to the very recently published Miyata and collaborators' work, where they reported gain of methylation at myogenic related promoters during differentiation, comparing DNA methylation levels between human MB, MT day 3, MT day 8 and MT day 15 samples (Miyata *et al.,* 2014). Using a 450 K platform and considering differences of 1.2 fold change as methylation changes (gain/loss of methylation) they found changes in a 0.18% CpGs (between MB and MT day 3), 0,89% CpGs (between MB and MT day 8), and a 2,9% CpGs (between MB and MT day 15), analyzing 1 sample per point. The reported changes can be considered extremely small, except for the last point corresponding to more mature myotubes. Notably, when they showed bisuphite sequencing results, few CpGs were analyzed with variable differences, pointing out that more complete analyses should be done to clarify these results. In fact, considering DNA methylation changes between myoblasts and myotubes, we were very surprised of not observing significant differences in our analyses, despite both stages represented two very different transcriptional programs (**Figure R12.A**). These results are supported by another recent study comparing DNA methylation between nine human myoblasts and myotubes samples using RRBS (Tsumagari *et al.,* 2013). This study shows no significant DNA methylation changes either, and it would indicate that identity-dependent methylation changes occurr mostly during early cell fate decisions, while fewer modifications take place later during terminal differentiation. In addition, Tsumagari's and our study coincide in identifying myogenic specific loss of DNA methylation at myogenic related genes, while Miyata's work showed gain of methylation at myogenic differentiation genes, such as contractile fibers related genes, and at Id binding sites. Since the genes identified in these studies are different, alternative roles of DNA methylation cannot be discarded. However, the reduced number of analyzed samples and the use of proliferating medium to keep alive terminally differentiated MTs by Miyata and colleagues lead us to warily consider their results as questionable.

Regarding the hypermethylated DMRs surrounding promoter regions that correlated with gene downregulation in muscle-committed cells, we detected pluripotency-related genes (Dppa4, Mybl2) and other lineage-specific factors related to neuronal

(Rnf1, (Tessema *et al.,* 2012; Kelly *et al.,* 2013)) , adipose (Zfp423 (Huang *et al.,* 2012)), gastrointestinal (Gpx2 (Brigelius-Flohé *et al.,* 2001)) and eritropoyetic (Gcnt2 (Inaba *et al.,* 2003)) lineages, between others. Interestingly, we found the pluripotency associated miRNA-302 (Anokye-Danso *et al.,* 2011), embedded in an intron and with the promoter region, which is bound by Pou5f1, Nanog, and Sox2, exclusively demethylated in ESC while heavily methylated in differentiated cells. Our data are consistent with previous reports analyzing neuronal, hematopoietic, skin and hepatocyte differentiation where DNA methylation repress inappropriate programs upon cell fate commitment (Mohn *et al.,* 2008; Bock *et al.,* 2012; Meissner *et al.,* 2008; Kim *et al.,* 2011).

1.3.1 DNA methylation changes outside promoter regions

Nowadays, the emerging picture of DNA methylation functions is becoming more and more complex and suggests that DNA methylation may also affect transcriptional splicing, the use of alternative promoters and would modulate enhancers activity (reviewed in Jones, 2012). It has been shown that exons are more highly methylated than introns and transitions in the degree of methylation occur at exon-intron boundaries. Indeed, genome-wide nucleosome-positioning data suggest that exons also show increased nucleosome-occupancy levels compared to introns and nucleosomes are preferential sites for DNA methylation (Schwartz *et al.,*2009; Chodavarapu et al. 2010). These observations suggest a role for DNA methylation regulating alternative splicing (Shukla *et al.,* 2011; Anastasiadou *et al.,* 2011; Malous *et al.,* 2008). Interestingly, at the gene body level, the identified DMRs are preferentially located in introns and exon-intron boundaries. Notably, we observed that the Nfix transcription factor, designated as master transcription factor for the embryonic to fetal myogenesis switch (Messina *et al.,* 2010; Hnisz *et al.,* 2013), contained a DMR in an intronic region too. This amplicon was found highly methylated in ESC, where Nfix is silenced, and demethylated in differentiated cells, except for NPC. The expression analysis of the different splicing isoforms revealed differences depending on the

methylation status, suggesting that DNA methylation could not only avoid spurious expression of Nfix in pluripotent cells, but also regulate the different splice variants as reported for other genes.

Interestingly, recent analyses of somatic methylomes (which did not included muscle tissue) have reported that tissue-specific DMRs are mainly found in *cis*-regulatory elements and, given that most of the DMRs showed hypomethylation in either one or two tissues, it was suggested that DNA methylation could control enhancer activity in a tissue-type dependent manner (Hon *et al.,* 2013). The idea that the methylation status of an enhancer could affect its activity was first reported more than 25 years ago by Saluz *et al.,* 1986. Importantly, enhancers are to a large extend cell-type specific being therefore differentially regulated across the different tissues(Heintzman *et al.,* 2007; Heintzman *et al.,* 2009). Indeed, Stadler and colleagues identified enhancers in the mouse genome as low methylated regions that were in a dynamic state depending on cell environment (Stadler *et al.,* 2011). In addition, DMRs have been reported within enhancer regions regulating differentiation specific genes at different subsets of T cells (Schmidl *et al.,* 2009). In line with these works, our results show that the demethylated regions in muscle cells were surrounded by enhancer-type chromatin features (p300, H3K27Ac), suggesting that the accessibility regulation to cell-type specific enhancers could be mediated by DNA methylation.

One of the reported mechanisms through which DNA methylation affects gene expression is by preventing the binding of transcriptional regulators to their sequence (Watt and Molloy 1988). In line this hypothesis, we have shown that demethylated DMRs are enriched at specific TFBSs such as Tata box and KLF**.** Importantly Sp proteins interact with Myogenin (Biesiada *et al.,* 1999) and MEF transcription factors (Krainc *et al.,* 1998; Park *et al.,* 2002), both essentials for myogenic differentiation, suggesting that DNA demethylation might generate a suitable landscape for TFs binding

1.3.2 DNA methylation changes in Myf5 super-enhancer

Recently, the concept of super-enhancer has emerged defined as large cluster of transcriptional enhancers close to cell-type specific genes, with very high transcription factor and co-activators density, and showing very high levels of both histone modifications H3K27Ac and H3K4me1 (Whyte *et al.,* 2013). The generation of a catalog of super-enhancers identified the *Myf5/Myf6* locus as a super-enhancer in myoblast, myotubes and skeletal muscle samples (Hnisz *et al.,* 2013), in line with pioneer functional studies that proved the regulatory implications of the multiple enhancers of the locus in the complex spatiotemporal regulation of *Myf5* and *Myf6* genes (reviewed in Carvajal *et al.,* 2008). Excitingly, an AIMS-seq DMR was identified in an exon-intron junction overlapping the *Myf5* intragenic enhancer (Summerbell *et al.,* 2000) and leading to the hypothesis that the entire locus might be regulated by DNA methylation. Of note, since *Myf5* presents only a splice variant, no implication of DNA methylation on the spliceosome was considered. Interestingly, an exhaustive analysis of the *Myf5/Myf6* locus revealed a methylation free state exclusively at enhancer regions in myogenic committed cells, while ESCs and non-myogenic differentiated cells were heavily methylated, correlating with the silencing of both myogenic-identity factors, Myf5 and Myf6 (**Figure R17**).

Most works describing the *Myf5/Myf6* enhancers have been done in embryonic stage (Carvajal *et al.,* 2008); however, despite Myf5 is expressed in MBs, we wondered if some demethylated regions might represent vestigial enhancers, defined as developmental enhancers demethylated during embryogenesis and kept in a demethylated state but inactive in adult stages (Hon *et al.,* 2013). The presence of active enhancer marks (H3K27Ac, p300, H3K4me1 and absence of H3K4me3) all over the locus (**Figure R20**) and the designation of *Myf5/Myf6* locus as a super-enhancer in skeletal myoblast, myotube and muscle tissue, granted an active state to the analyzed enhancers in postnatal myogenesis, highlighting the regulatory importance of these DMRs.

Our results showed that the ubiquitously expressed Usf1 TF bound to *Myf5* upon CpG-containing-E-box demethylation in muscle-committed cells, coinciding with transcriptional induction (**Figure R18**). Notably, the embedded CpG was conserved in the human genome and is demethylated in human myoblasts, while is methylated in

hESCs (**Figure R18.D**). Since Usf1 binding to E-boxes is impaired by DNA methylation (Fujii *et al.,* 2006, Aoki *et al.,* 2008; Li *et al.,* 2014) and the Usf1 binding leads to transcriptional activation through the recruitment of histone methyltransferases, resulting in H3K4me3 modification (Deng *et al.,* 2013; Huang *et al.,* 2007), we conclude that DNA demethylation of EE enhancer contributed to Myf5 activation. The lack of Usf1 bind at the highly methylated CpG-containt E-box at -118 control region, support this fact. In addition, Usf1 bound less efficiently to the CpG-free -17 kb enhancer, which might contribute to Myf5 activation too (Chang et al. 2004). Interestingly, this same association was observed at the Nfix intronic DMR, where the E-box contained a CpG demethylated bound by Usf1 in myogenic cells, and the binding correlated with gene transcription too. These results support the hypothesis that Usf1 is important for mesoderm lineage determination (Deng *et al.,* 2013) and highlighted the importance of DNA methylation at CpGs within TFBSs (Kim *et al.,* 2003; Deaton and Bird 2011).

S.T. Smale and colleagues considered that the process leading to the formation of tissue-specific enhancers probably starts at the earliest phase of the embryonic development. They showed in ESCs the existence of unmethylated windows within the tissue-specific enhancers, were the named pioneer factors were bound (Xu *et al.,* 2007; Xu, Watts, *et al.,* 2009). These enhancers will be in most cases remethylated upon differentiation in the lineages in which they are not functional, and maintenance of the demethylated window might require the replacement of the pioneer TF expressed in ES cells by the lineage-restricted TF (reviewed in Natoli, 2010). Our data in *Myf5/Myf6* super-enhancer locus, as well as in *Pax3* and *Myod* enhancers, show a completely different scenario where tissue-specific enhancers are totally methylated in ESC and non-myogenic cells and became demethylated upon myogenic lineage determination.

Specific transcription factors are essential in early cell type specification and differentiation. It has been reported that many of the developmental key transcription factors bind directly to a large number of genomic regions, as demonstrated in transcription factor binding studies for Pax5, HNF and MyoD (Delogu *et al.,* 2006; Odom *et al.,* 2004; Cao *et al.,* 2010). MyoD binds over 40% of the annotated genes at MBs and MTs where it leads to histone acetylation and therefore to an open

chromatin state. Noteworthy, it only acts as a direct transcription factor on 4% of the genes (Cao *et al.,* 2010). In contrast, Myogenin, selectively binds to the regulatory sequences of late muscle genes (Du *et al.,* 2012), and no global ChIP-seq for Myf5 has been performed yet. Considering the AIMS-seq DMRs, we concluded that DNA methylation did not directly interfere either MyoD or Myogenin binding, since they do not showed overlap with DMRs (data not shown). Interestingly, neither MyoD (Fong *et al.,* 2012) nor Myogenin (Cao *et al.,* 2006) present a CpG dinucleotide in its canonical E-box binding site, suggesting that other mechanisms than DNA methylation might modulate the activation of MyoD/Myogenin regulated genes, in line with Oikawa *et al.,* 2011 work, where they show the DNA-methylation-independent binding of MyoD on the *Myogenin* promoter.

However, many transcription factors (tissue-specific or ubiquitous) bind to the highly abundant E-boxes in the genome, leading us to suggest that DNA methylation could represent a guiding mechanism to reduce chromatin accessibility and the binding efficency of TFs to the E-box sequences. We suggest that the binding of ubiquitous TFs in a lineage–specific manner could be mediated by DNA methylation, taking place only in demethylated E-boxes in the right locus and tissue, as we reported by Usf1 binding at *Myf5* and *Nfix* locus, reinforcing the hypothesis that transcription factor binding can be strongly influenced by methylation of CpG sites within their recognition sequences (Feng *et al.,* 2010).

Considering the DMRs at repetitive sequences, they showed a significant enrichment in SINE and LTR elements, although no major differences were observed between hypo and hypermethylated regions. SINE sequences are mainly located in gene-rich regions (Jabbari and Bernardi, 1998), supporting the idea that DNA methylation in SINE elements may affect gene regulation. In addition, lncRNAs overlapping with LTR and SINE elements have been reported to influence developmental and homeostatic pathways, (Xie *et al.,* 2013; Wang *et al.,* 2013). This suggests that a detailed analysis of DNA methylation in lncRNAs regulatory regions may show how epigenetic mechanisms would affect LTR- and SINE-mediated gene regulation in a cell-specific manner.

Summing up, we have demonstrated that AIMS-seq technique allows untargeted survey of DNA methylation patterns. By applying AIMS-Seq to the study of myogenesis we have identified DMRs outside CpG rich/promoters regions allowing the discovery of previously unknown regulatory regions. Globally, our results point out a sequential and alternative process of hypermethylation and hypomethylation along myogenic cell fate and differentiation. In addition, we show that the terminal myogenic program is orchestrated almost independently of DNA methylation changes, suggesting that once DNA methylation signatures are set during cell fate determination, cell-type specific transcription factors ensure the expression switch of differentiation programs. All these observations reinforce the hypothesis that DNA methylation patterns shape cell fate commitment and are inherited and maintained throughout cell division establishing a molecular memory to keep cellular identity.

2 EPIGENETIC SIGNATURE OF THE PRINCIPAL MYOGENIC REGULATORY GENES

2.1 Epigenetic signature of CpG-rich and CpG-poor regulatory regions

Our results showing differentially methylation in *My5/Myf6* locus, together with the reported demethylated *MyoD* enhancer (Brunk *et al.,* 1996) and *Myogenin* promoter (Palacios *et al.,* 2010) in muscle cells prompted us to address if the expression of other relevant myogenic genes could be also modulated by DNA methylation. We analyzed in detail the developmental genes *Pax3* and *Pax7*, both highly expressed during muscle development, *MyoD* and *Myogenin* as specific-myogenic transcription factors, and *Ckm*, *Mhy1*, *Myh4* and *Myh8* as terminal differentiation expressed genes (see Results section 2). Regarding the CpG content, only *Pax3*, *Pax7* and *MyoD* present CpG island at their promoters and, as expected, they were totally unmethylated in all analyzed cell types. Interestingly, high levels of H3K27me3 were found in these regions in ESCs according to ENCODE ChIP-seq data, indicating that these genes are repressed by Polycomb proteins in ESCs. In fact, we observed that these genes displayed bivalent domains (H3K4me3/H3K27me3), which are typically enriched at CpG island regions and associated with a transcriptional poised state, which resolve upon differentiation

by the maintenance either of the positive mark (H3K4me3) or the negative one (H3K27me3) (Bernstein *et al.,* 2006). During myogenic lineage commitment, the Polycomb mark H3K27me3 got lost and the poised state declined toward the positive mark H3K4me3 and the genes were transcribed. Later on, during the differentiation process, the genes became silenced through the absence of positive marks.

Many genes are regulated by enhancer/promoter pairs, although they can be separated by several kb. Interestingly, early myogenic genes with CpG island promoters were modulated by DNA methylation changes at distal CpG-poor content regulatory enhancers. Notably, *MyoD* and *Pax3* enhancers, demethylated in muscle-committed cells, contained TFBSs with embedded CpGs, whereas this was not the case for *Pax7* distal enhancer, which was kept highly methylated in all cell types. These remarkable differences lead us to hypothesize that the lack of CpGs within TFBSs of *Pax7* enhancer might exclude DNA methylation as a direct regulator mechanism in this particular enhancer. In this context, promoters and enhancer regions of *Pax3* and *MyoD* genes showed distinct epigenetic regulation, with demethylation of enhancers and loss of H3K27me3 in promoters, suggesting a crosstalk between DNA methylation and histone modifications (**Figure R32**). Considering a tridimensional chromatin structure, where the promoter and the distal enhancers are close, we might hypothesize that the reduction of Ezh2 activity during cell differentiation, which catalyzes tri-methylation on H3K27, could lead to the Dnmts recruitment cessation, since both enzymes can cooperatively interact (Viré *et al.,* 2006).

The CpG-poor promoters showed a different epigenetic regulation. In a first stage, the genes *Myogenin*, *Ckm* and *Myosins*, silenced in ESCs by DNA methylation, became hypomethylated during the establishment of lineage commitment although gene silencing was maintained by H3K27me3 mark. In a second step, H3K27me3 was replaced by H3K4me3 and H3K27Ac marks, and together with p300 binding, resulting in the transcription of the genes. This two-step repressive mechanism, where DNA methylation is substituted by H3K27me3, has been observed when ESCs are derived to the three germ layers too (Gifford *et al.,* 2013) and represents a switch from a stable

repressed state (DNA methylation) to a more facultative state of repression (Smith and Meissner 2013).

In addition, the process of cellular differentiation requires the loss of pluripotency capacities. Promoters of pluripotency genes were occupied by H3K4me3, free of DNA methylation in ESCs and consequently highly expressed at that stage. Upon differentiation, pluripotency genes with CpG island promoters were silenced by H3K27me3 mark (such as Sox2 and Fgf4), whereas CpG-poor content promoters were repressed by DNA methylation (such as Dppa4 and Pou5f1) (**Figure R3**). The use of different silencing mechanisms, depending on CpG content, is consistent with previous reports showing that Polycomb binds preferentially unmethylated CpG islands (Ku *et al.,* 2008; Lynch *et al.,* 2012; Bartke *et al.,* 2010) which are mostly nucleosomes-deficient (Ramirez-Carrozzi *et al.,* 2009), whereas non CpG island promoters are nucleosome-rich and therefore preferentially targeted by Dnmts and susceptible to be regulated by DNA methylation (Chodavarapu *et al.,* 2010) while they do not often associate with H3K27me3 (Mikkelsen *et al.,* 2007; Mohn *et al.,* 2008).

Finally, genes constitutively and highly transcribed all over the differentiation process contained CpG island promoters (Saxonov *et al.,* 2006) and showed ubiquitous H3K4me3 and H3K27Ac occupancy comparable to the signals observed in myogenic modulated genes (**Figure R33**). This observstion discard a different regulation between modulated and ubiquitous expressed genes (Ganapathi *et al.,* 2005). The lack of methylation of these promoters is likely due to the prevalent presence of H3K4me3 on the CpG island which prevents the binding of the cofactor Dnmt3L to chromatin, protecting the promoter from *de novo* DNA methylation (Ooi *et al.,* 2007; Guenther *et al.,* 2007; Cedar and Bergman, 2009).

By linking our DNA methylation results with the ChIP-seq data we have categorize the gene regulation into six different epigenetic dynamics during myogenic progression, which are summarized in **Figure D1.**

Myogenin is a crucial myogenic differentiation factor, and as mentioned in the introduction its DNA methylation pattern has been addressed in several independent studies. Intriguingly, while Oikawa's, Lucarelli's and Palacios's work showed a loss of DNA methylation upon MB differentiation in the C2C12 cell line (Oikawa *et al.,* 2011; Lucarelli *et al.,* 2001; Palacios *et al.,* 2010) we observed a complete methylation at MBs state already.The analysis of the *Myogenin* promoter region in the quiescent satellite cells dissipated the possibility of an artifact due to the culture conditions and provided robust evidence that DNA demethylation of the *Myogenin* promoter occurs before MTs formation. We speculate that Oikawa's, Lucarelli's and Palacios's results were consequence of large scale culture conditions (Meissner *et al.,* 2008) or due to an incomplete bisulphite conversion, and therefore might be technical artifacts. In addition, we observed a partial re-methylation of Myogenin promoter in *ex-vivo* isolated myofibers. *Myogenin* is not expressed in skeletal muscle, like *MyoD* and *Myf5*, and also *Myf5* promoter showed a partial re-methylation in mature myofibers correlating with the lack of expression. Notably, the promoter region of *Myf6* (the only MRF expressed in mature tissue) was totally demethylated only in mature myofibers, linking once more time DNA demethylation with gene expression.

It is well known that MyoD and Myogenin proteins are absent in quiescent satellite cells (Gnocchi *et al.,* 2009). Interestingly, we found the *Myogenin* promoter and the *MyoD* enhancer fully demethylated in these dormant cells. The lack of methylation at the *MyoD* enhancer could be interpreted as a vestigial demethylated state consequence of *MyoD* expression in satellite cells precursors during embryogenesis (Hon *et al.,* 2013). However, Myogenin is a differentiation factor expressed in differentiated non-dividing muscle cells, pointing out that the loss of DNA methylation occurred before gene expression.

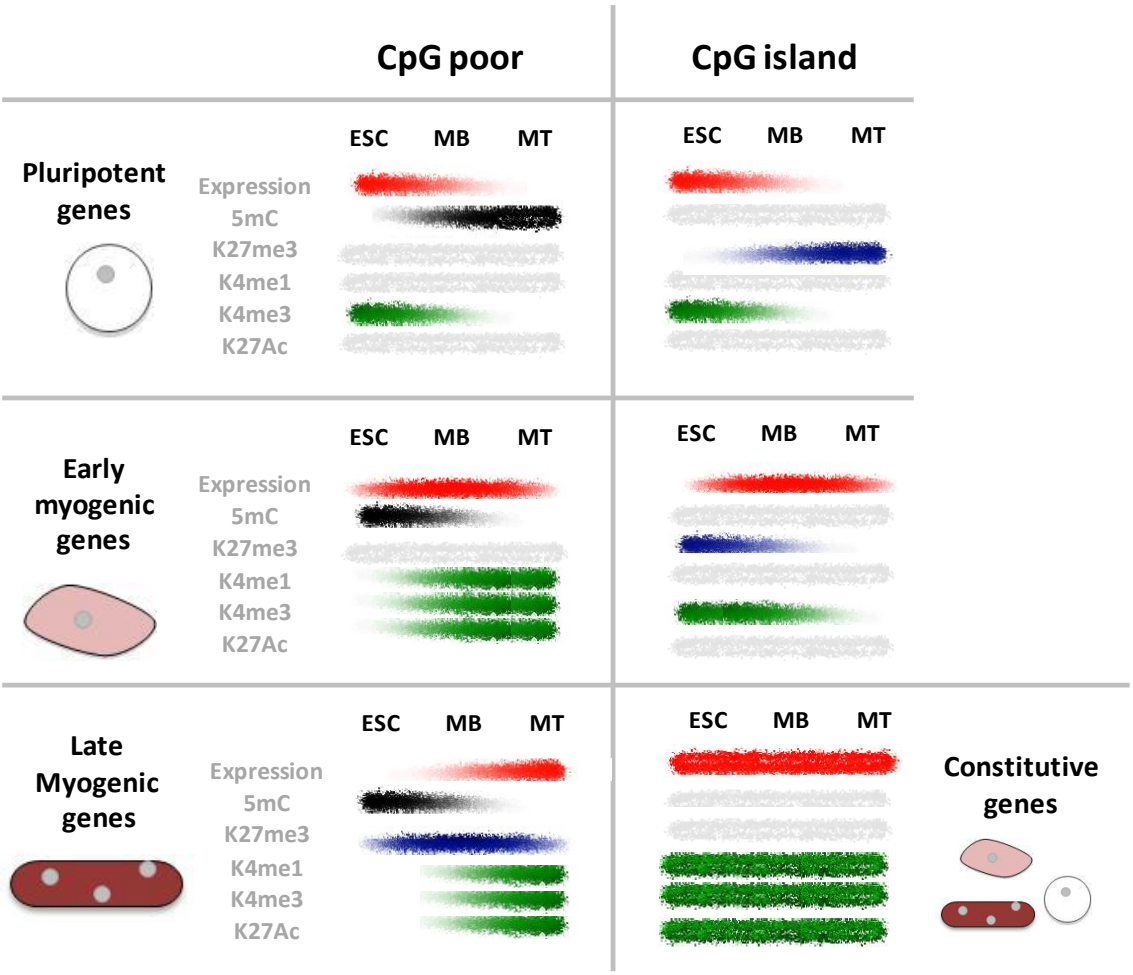


Figure D1. Integrative representation of the epigenetic signatures associated to CpG rich and poor sequences of genes modulated or constitutively expressed during myogenesis.

2.2 Inducible *Pax7* ESC-derived myogenic model mimics myogenic DNA methylation patterns

Importantly, the analysis of the DMRs in the *iPax7* ESC-derived myogenic model revealed very similar DNA methylation signatures compared to primary MT and MT. Collectively, these results ensured the identification of lineage-specific DMRs during myogenesis and reinforced the hypothesis that unique DNA methylation patterns correlate with defined cellular identities. The use of a stepwise model revealed that the myogenic differentiation process occurred coupled with a gradual loss of

methylation in myogenic regulatory regions. Initially, *Pax3* enhancer became demethylated before *Pax7* induction in line with its expression pattern, which is not restricted to the myogenic lineage. Despite the enhancer was demethylated, *Pax3* was not expressed in the model, because it is not a Pax7 target (Buckingham and Relaix, 2007), suggesting that other mechanisms are necessary to activate this gene in muscle precursors cells. Then, *MyoD* and *Myogenin* CpG-poor regulatory regions became demethylated in *iPax7*-myoblast precursors correlating with transcriptional initiation.

Finally, the late differentiation myogenic genes *Ckm* and *Myh8* became simultaneously demethylated and upregulated when the *iPax7* myogenic precursors differentiated and fused forming myotubes conversely to the results observed in primary MB, where demethylation of *Ckm* and *Myosine* promoters preceded gene expression. A plausible explanation for that difference is that *iPax7* myogenic precursors are less "myogenic committed" that primary myoblast (isolated from adult muscle tissue). These engineered cells are myogenic precursors with the capability to differentiate into myotubes, but might not be totally myogenic cells yet, which could be reflected in the methylation state of late differentiation markers only expressed in mature myotubes. Finally, gain of DNA methylation at pluripotency genes also took place following different time patterns in the inducible myogenic model, showing altogether that the acquisition of specialized transcription program together with the silencing of the pluripotency one is a gradual process modulated by epigenetic mechanisms.

Our epigenetic analysis supports the *iPax7* ESC-derived myogenic model as a *bona fide* model to generate myogenic precursors *in vitro* with therapeutic purposes. Moreover, we also show that it may be an appropriate system to study active DNA demethylation mechanisms.

Summing up, these results reinforced the hypothesis that CpG-rich and -poor regulatory elements present different epigenetic signatures, and that lineage-specific DNA demethylation together with pluripotency genes hypermethylation occur in parallel during early cell fate decisions (Mohn *et al.,* 2008; Kim *et al.,* 2011; Brunner *et al.,* 2009; Bock *et al.,* 2012; Gilsbach *et al.,* 2014). Importantly, most of the DNA

methylation changes observed during myogenesis, were stably maintained during the myogenic progression, which is consistent with the idea of epigenetic memory maintaining cell identity.

3 DNA METHYLATION, CAUSE OR CONSEQUENCE OF TRANSCRIPTIONAL ACTIVATION?

Recently, it has been proven that Polycomb-mediated repression is not required for gene silencing initiation during mouse ESC differentiation but is required for repression maintenance (Riising *et al.,* 2014). However, if DNA methylation patterns are a cause or a consequence of gene silencing remains controversial and several interpretations are possible. In one hand, DNA methylation could repress transcription by preventing the binding of TFs to their sequences (Iguchi-Ariga and Schaffner, 1989; Campanero *et al.,* 2000; Blattler and Farnham, 2013; Prendergast and Ziff, 1991) or by attracting methyl-DNA binding proteins, that can recruit chromatin remodelers to compact chromatin (reviewed in Buck-Koehntop and Defossez, 2013). However, in the other hand, loss of methylation could be considered as a consequence of transcriptional activity where TFs occupancy would actively drive DNA demethylation or passively impair Dnmts recognition leading to a progressive loss of methylation (Feldmann *et al.,* 2013; Stadler *et al.,* 2011).

Our results provided multiple evidences that DNA methylation might play a regulatory role rather than represent a passive transcriptional witness. First, we showed several DNA demethylation events that happened independently of gene transcription: the late differentiation myogenic genes *Ckm* and *Myh4* became demethylated already in primary MBs (**Figure R31**), *Myogenin* promoter region was fully demethylated in non-expressing quiescent satellite cells (**Figure R26**) and *Pax3* enhancer was demethylated in ESC-derived myogenic model (**Figure R37**), although the gene was not expressed. Our results, supported by other works (Palacios *et al.,* 2010; Brunk *et al.,*1996; Szyf *et*

al., 1992, Lindahl Allen *et al.,* 2009), would reinforce the idea that demethylation events occur before transcriptional activity. In the case of *Ckm* gene, the slight expression detected in i*Pax7* ESC-derived myoblast precursors before promoter demethylation, was likely due to the presence of few differentiated cells in proliferative conditions. Therefore, this result does not support demethylation as a consequence of transcriptional activation. Second, the dramatic and fast gain of DNA methylation at *Dppa4* promoter during the early differentiation in the inducible ESC-derived myogenic model coupled to a successive reduction of gene activity (**Figure R39**), pointed out that the gain of DNA methylation actively interfere in transcriptional activity, and not the other way around, where the progressive lack of expression would lead to a progressive gain of methylation. Third, Usf1 binding at *Myf5/Myf6* locus was impaired in methylated ESCs whereas we detected Usf1 occupancy upon loss of methylation (**Figure R18**). It could be argued that Usf1 might bind to the methylated TFBS and consequently lead to DNA demethylation; however, it has been shown that Usf1 does not bind methylated E-boxes (Aoki *et al.,* 2008; Li, Huang, *et al.,* 2014; Fujii *et al.,* 2006).

Contrarily, Schübeler lab's work showed that binding of CTCF and Rest transcription factors within methylated CpG-poor regions directly drives DNA demethylation (Feldmann *et al.,* 2013; Stadler *et al.,* 2011). Since the analyzed TFs were different and had different TFBS characteristics, these apparently contradictory results could be combined into an inclusive interpretation: DNA methylation interferes in the recognition of TFBS when the motif contains a CpG, which is highly frequent in TFBSs (Deaton and Bird, 2011), such as Usf1, MYC and YY1 (Fujii *et al.,* 2006; Aoki *et al.,* 2008; Perini *et al.,* 2005; Kim *et al.,* 2003; Fujii *et al.,* 2006), whereas it does not affect the binding of proteins that can recruit other proteins implicated in DNA demethylation process when the binding motif does not contain a CpG (CTCF and Rest TFBS). In other words, we hypothesize that DNA methylation directly regulates some TFs binding and therefore modulates gene expression; while at the same time, transcriptional activity regulation through TF binding and/or histone marks could also drive the loss of DNA methylation.

4 CPG ISLAND METHYLATION IN RHABDOMYOSARCOMA CELL LINES

Rhabdomyosarcoma (RMS) is the most common soft tissue sarcoma of childhood and it is presumed to originate from skeletal muscle because of its myogenic phenotype (reviewed in Keller and Guttridge 2013). RMS is composed of two main subtypes, embryonal RMS (eRMS) and alveolar RMS (aRMS). Whereas eRMS histologically resembles embryonic skeletal muscle, the aRMS subtype is more aggressive and has a poorer prognosis. Moreover, the genetic profile of eRMS is not well established, while aRMS is commonly associated with distinct chromosome translocations that fuse domains of the transcription factors Pax3 and Pax7 to the forkhead family member Fkhr (FOXO1A). Both eRMS and aRMS tumor cells express myogenic markers such as MyoD, but their ability to complete differentiation is impaired (reviewed in Keller and Guttridge, 2013).

The study of DNA methylation in RMS might shed light into its cell origin, which remains unknown, and could influence its pharmacological response. In collaboration with Charles Keller's laboratory we have investigated the expression and methylation state of *Pax3:Fkhr* and *Pax7* CpG islands in two different aRMS cell lines representing two different cell origins (prenatal precursor myoblast and postnatal satellite cell origin) with the aim of refining the range of possibilities for the cell of origin of the Pax3:Fkhr rhabdomysoarcomas.

After exhaustive analysis of the different cell lines of aRMS (tumor generation, survival rate, histological analysis and expression profiles) it was concluded that the prenatal precursor myoblasts might be the more plausible cellular origin for aRMS (Abraham *et al.,* 2014). Moreover, we observed in these cells CpGi methylation at the *Pax7* promoter region coinciding with a total abolishing of the gene expression, while the satellite cell line remained unmethylated and showed low, but consistent *Pax7* gene expression. It is well known that tumoral processes are accompanied by hypermethylation of several CpG islands (Jones and Baylin, 2002; Jaenisch and Bird, 2003), therefore our data support the prenatal precursor myoblasts cell origin hypothesis. In addition, *Pax7* silencing in prenatal precursor myoblasts set up the

question about why *Pax7* is hypermethylated, if it is advantageous for the cells and finally, if it has any functional implication in rhabdomyosarcoma tumor development. We evaluated if Pax7 expression could be re-expressed in the myoblast cell-of-origin cells by treating the cells with the DNA demethylating agent 5'Aza2 deoxycytidine. However and unfortunately, the treatment did not modify the methylation profile of Pax7 neither its expression. Interestingly, 5'Aza2 deoxycytidine treatment decreased Pax3:Fkhr expression in the cells, which is fully demethylated. A possible explanation to this observation would be that other repressive regulatory regions upstream Pax3 locus might be affected by the demethylating treatment, leading to the decrease in Pax3:Fkhr expression.

In addition, we tested whether the eRMS (U37125) could have some epigenetic similarities with the aRMS originated from satellite cells (U23674), given their expression analysis similarities (**Appendix I Figure 3**). The DNA methylation analysis of *Pax3* and *Pax7* promoters in an eRMS cell line showed hypermethylation at several CpGs in the *Pax3* CpG islands, but not methylation at all in *Pax7* promoter, revealing a lack of similarity between both DNA methylation profiles.

Although our work has contributed to clarify the alveolar RMS cell origin, further experiments would be needed to address if epigenetic drugs can be used as therapeutic agents against tumoral rhabdomyosarcoma progression.

5 ASSESSMENT OF THE DNA DEMETHYLATION MECHANISMS DURING MYOGENESIS

We have shown consistent DNA demethylation events occurring during myogenic lineage commitment and late differentiation stages of the myogenic process. They occurred independently of replication and in presence of maintenance Dnmt1, suggesting that they are not a consequence of lack of methylation and subsequent dilution of the mark, but instead, they respond to active enzymatic mechanisms.

A direct DNA demethylase capable of cleaving the carbon–carbon bond between the methyl-group and the deoxyribose of the cytosine has not been yet identified in mammals. However, recent work has explored indirect demethylation pathways that involve oxidation or deamination of 5mC coupled to glycosylases and base excision repair (BER) mechanisms, what ultimately leads to an unmethylated cytosine (reviewed in Seisenberger *et al.,* 2013). Although there are some evidences supporting a role for many of these DNA modifying pathways, they are often restricted to the specific biological system examined. The lack of a unifying mechanistic process has led to ongoing disputes over the relative importance of these DNA demethylation pathways (Chen and Riggs, 2011). In addition, most of the studies have focused on PGCs and zygote global DNA demethylation events, whereas lineage-specific loss of methylation has not been well addressed.

To assess the pathways involved in DNA demethylation during myogenesis and to evaluate the effect of impairing DNA demethylation on the lineage commitment progression we designed shRNA experiments to block the three putative enzymatic mechanisms (oxidation, deamination and glycosilation). Unfortunately and due to technical impairments briefly discussed below, we could not conclude the experiments for two out of three enzymes families tested.

5.1 Implications of Tet1 down-regulation during lineage-associated DNA demethylation

5-hydroxymethylcytosine mainly occurs in ES cells (5% of CpG), but also in certain types of neurons and other adult mouse tissues (reviewed in Wu and Zhang, 2011). However, the functional significance of 5hmC in ESCs is not totally clear. On one hand, Tet1 and 5hmC are enriched at transcriptional start sites of CpG-rich promoters and gene bodies in ES cells, where it promotes DNA demethylation and contributes to the maintenance of the unmethylated state of the CpG islands, mainly in bivalent promoters (Pastor *et al.,* 2011). On the other hand, the 5mC oxidative products

constitute epigenetic marks *per se* and are recognized by Polycomb repressive complex 2 (PRC2), leading to gene repression (Wu *et al.,* 2011; Williams *et al.,* 2011) and suggesting that Tet1 could also have an important functional role in the maintenance of pluripotency and development capacities. Moreover, in addition to ESCs, 5hmC is very abundant in brain tissue, but it is believed that 5hmC might play different roles in these two cell types (reviewed Santiago *et al.,* 2014).

Several knock-down (KD) and knock-out (KO) experiments of the different Tet enzymes (individually or combined) have been done to address their role in ESCs and neuronal lineage. Tet1 KO and KD ESCs skewed differentiation towards trophoctoderm (Dawlaty *et al.,* 2011; Ficz *et al.,* 2011), leading to global loss of 5hmC and affecting the gene expression profile. Very importantly, gene ontology of deregulated genes on ESC $Tet1^{-/-}$ showed that skeletal muscle development related genes were specially represented, as well as muscle contraction genes, suggesting that Tet1 might be involved in myogenic-related DNA demethylation (Dawlaty *et al.,* 2011). In addition, the level of 5hmC was quantified in muscles samples using an enzymatic assay. The results showed twice the amount of 5hmC in skeletal muscle compared to MB, MT, skin and sperm (Tsumagari *et al.,* 2013).

In brain, Tet1 KD and KO affected neuronal properties by exhibiting impaired hippocampal neurogenesis accompanied by poor learning and memory, suggesting that Tet1 plays an important role in regulating neural progenitor cell proliferation in adult mouse brain. (Dawlaty *et al.,* 2011; Rudenko *et al.,* 2013; Zhang *et al.,* 2013). Tet2 deficiency in ESCs showed reduced 5hmC (Huang *et al.,* 2014), while in NPCs showed defects in neuronal differentiation (Hahn *et al.,* 2013). $Tet3^{-/-}$ ESCs showed impair neuronal differentiation (Li, Yang, *et al.,* 2014), and cells with reduced Tet3 levels failed to convert 5mC to 5hmC in the paternal pronucleus (Wossidlo *et al.,* 2011). The double Tet1 and Tet2 KO showed developmental defects (Dawlaty *et al.,* 2013), while the triple KO could not support embryonic development and showed promoter hypermethylation and deregulation of genes implicated in differentiation, including skeletal muscle development genes as one of the top ten GO categories (Dawlaty *et al.,* 2014).

We chose *Tet1* to perform the KD experiments because it was the highest expressed in ESCs, embryoid bodies and sorted-mesodermal cells, had high affinity for CXXC regions while Tet2 does not (Xu *et al.,* 2012; Zhang *et al.,* 2010; Ko *et al.,* 2013), had been related with enhancer activation in differentiation processes (Sérandour *et al.,* 2012) and $TET1^{-/-}$ mice were viable and fertile but about 75% of the homozygous mutant pups had smaller body size and deregulated, including developmental skeletal muscle and muscle contraction genes (Dawlaty *et al.,* 2011).

Unfortunately, our Tet1 KD experiments were inconclusive since we were not able to amplify *iPax7*-ESCs stably infected with Tet1 shRNA (**Figure R42**). After modifying the infection conditions by increasing the amount of virus and trying several culture conditions of the infected population by reducing or completely abolishing the co-culture with irradiated MEFs, the infected population was still very low, and therefore we could not evaluate the efficiency of the shRNAs neither their effect. We speculate that ESCs could silence the insert as a defense mechanism, as previously reported (Xia *et al.,* 2007), or that Tet1 KD confer a dramatic decrease in the viability of the KD population, impairing its expansion.

5.2 Implications of Tdg down-regulation during lineage-associated DNA demethylation

As previously mentioned, thymine DNA glycosylase (Tdg) recognizes 5fC and 5caC Tet oxidated products, while leaving 5hmC intact (He *et al.,* 2011; Maiti and Drohat, 2011;Zhang *et al.,* 2012), as well as acts on AID/Apobec deaminated products, and excises the modified bases leading to C regeneration (Cortellino *et al.,* 2011; Cortázar *et al.,* 2011; Cortázar *et al.,* 2007). $Tdg^{-/-}$ mice are embryonic lethal (Cortellino *et al.,* 2011; He *et al.,* 2011; Cortázar *et al.,* 2011), and concerning ESC differentiation it has been reported that $Tdg^{-/-}$ ESCs fail to complete neuronal differentiation (Cortázar *et al.,* 2011), since the demethylation of the specific gene for the neuronal lineage (Lrrmt2) is impaired. In addition, Tdg selectively binds active chromatin, especially in pluripotency

genes, but decreases when the differentiation process take place and the methylation increase. Very importantly, few years ago, it was reported the implication of Tdg in myogenic associated DNA demethylation, showing a genome-wide loss of DNA methylation in differentiating myoblasts upon antisense morpholino-mediated Tdg downregulation (Jost *et al.,* 2001). Since the Tdg enzyme acts on Tet oxidative products it's blocking would affect the Tet-Tdg coupled pathway.

With these antecedents and after ensuring the Tdg expression in our inducible myogenic model, we proceeded to design the experiment to down-regulate Tdg using shRNAs. To avoid the problems we had in Tet1 shRNA ESC infection, and after a precise evaluation of the DNA demethylation timing occurring during the *iPax7* ESC-derived myogenic model, we changed our strategy and decided to infect the cells after *Pax7* induction. This new infection time point allowed us to address the effect of Tdg KD on the second wave of DNA demethylation, when MyoD enhancer and Myogenin promoter became demethylated. After infecting the myogenic lineage committed ESC-derived cell with several vectors carrying different shRNAs, we found that none of the constructs triggered a sufficient Tdg down-regulation. Unfortunately, due a lack of time of the stay performed in Rita Perlingeiro's lab, we could no persist on this line of research, which remains yet inconclusive (**Figure R43**).

5.3 Implications of Apobec2 down-regulation during lineage-associated DNA demethylation

AID/Apobec (activation-induced deaminase/apolipoprotein B-editing) family of proteins recognize m5C and hm5C and catalyze the deamination reactions generating T and 5-hmU respectively, and thereby initiating the mismatch repair mechanisms leading to active DNA demethylation (Cortellino *et al.,* 2011) (Guo, Su, *et al.,* 2011).This family of enzymes was originally identified as RNA editors (Conticello 2008). However, their involvement on epigenetic regulation is based in the founding of their 5mC deaminase activity (Liao *et al.,* 1999), their colocalization within a cluster of

pluripotency genes and its high expression in oocytes and primordial germ cells, which undergo extensively epigenetic reprogramming (Morgan *et al.,* 2004).The first AID role in mediating vertebrate active DNA demethylation came from zebra fish embryos (Rai *et al.,* 2008). Over-expression of AID and Mbd4 in zebra fish embryos caused demethylation of the bulk genome and of the injected methylated DNA fragments. Later on, this roles were supported by AID/Apobec1 experiments in brain, where they act downstream of Tet1 and resolves in a neuronal activity-induced active DNA demethylation and subsequent gene expression in the adult mouse brain *in vivo* (Guo, Ma, *et al.,* 2011).

Specifically, Apobec2 expression is restricted exclusively to cardiac and skeletal muscle tissues (Liao *et al.,* 1999), being the only member of the family expressed in muscle (Mikl *et al.,* 2005). Very recently, it has also been detected in retina (Powell *et al.,* 2013). The function of Apobec2 is not well known. Initially, it was proposed, by analogy with Apobec1, to be a RNA-editing enzyme (Anant *et al.,* 2001). Later on, Apobec2 mRNA editing activity was discarded while a low, but definite, intrinsic cytidine deaminase activity was detected (Liao *et al.,* 1999). However, it is still matter of controversy if Apobec2 has or not real cytidine deaminase activity (Mikl *et al.,* 2005; Sato *et al.,* 2010). The generation of Apobec2$^{-/-}$ mice showed no major implication in mouse health, fertility or survival up to 1 year of age, and the histological examination of heart and calf muscle also failed to reveal any abnormalities (Mikl *et al.,* 2005). However, detailed analysis of muscle in Apobec2-deficient mice showed a markedly increased ratio of slow to fast fibers in soleus muscle. In addition, animals exhibited a 15–20% reduction in body mass from birth onwards, with elderly mutant animals revealing clear histological evidence of a mild myopathy (Sato, Probst, *et al.,* 2010). These results suggested that Apobec2 was essential for normal muscle development and maintenance of fiber-type ratios; although its specific molecular function remains to be identified.

The expression analysis of Apobec2 in the inducible myogenic model showed a striking increase of Apobec2 mRNA levels in primary MB and *iPax7* ESC-derived myogenic precursors and myotubes, in line with its myogenic specificity and reinforcing the proper expression profile of this *in vitro* derived myoblasts. Interested in the role that

Apobec2 might play on myogenic-specific DNA demethylation, we performed shRNA KD experiments on early myogenic committed cells before the second wave of demethylation. Two out of four tested shRNA dramatically dropped Apobec2 mRNA levels and, most importantly, this resulted in a clear impairment of the differentiation process. Apobec2 down-regulation abolished the expression of terminal myogenic markers at the mRNA and protein level and blocked the myofibers formation, suggesting a relevant role of Apobec2 during myogenesis. These results were supported by experiments performed on C2C12 myoblast cells, were Apobec2 was required for myotube formation (Vonica *et al.,* 2011). Vonica and collaborators suggested that Apobec2 expression was induced by TGFβ signaling and had a direct involvement in left-right specification during embyronal development, while it is also required for TGFβ inhibition in myoblasts, generating a negative feedback loop to allow myogenic differentiation (Vonica *et al.,* 2011). However, this work did not addressed if Apobec2 played a role in DNA demethylation.

The bisulphite sequencing of *iPax7* ESC-derived myogenic precursors after Apobec2 shRNA infection, showed that *MyoD* enhancer presented similar methylation levels compared to the control vector, whereas *Myogenin* promoter showed a significant increase in DNA methylation in all the analyzed CpGs. Interestingly, our data suggested that Apobec2 would affect *Myogenin* DNA demethylation and *Myogenin* expression and therefore, myogenic differentiation. Notably, Apobec2 seems to play a role shaping the identity of the retina cells progenitor too. Powell and colleagues reported that Apobec2 was required for retina adult stem cells formation and hypothesized that it could participate in genome-wide active DNA demethylation of the retina progenitors following induced-injury (Powell *et al.,* 2012). However, after treating these cells with Apobec2 morpholinos, they found DNA methylation differences only in 66 DMRs of a total of 283,572 CpGs analyzed. In addition, 63.6% of the identified DMRs were more demethylated following knockdown of Apobec2a/2b, which represents the opposite to the expected result if Apobec2 would participate in DNA demethylation (Powell *et al.,* 2013). Therefore, they concluded that DNA methylation changes were independent of Apobec2 expression in zebra fish. Very recently, the same group reported that Apobec2 controls Pou6f2 binding to DNA during

development, providing a link between Apobec2 and retina cell regulation (Powell *et al.,* 2014).

Nevertheless, our results showed that lack of Apobec2 impaired myogenic differentiation and affected DNA demethylation of the crucial myogenic differentiation factor *Myogenin*. Since we and others (Palacios *et al.,* 2010) have shown that *Myogenin* promoter becomes demethylated before gene expression, we discarded the possibility of an impairment of DNA demethylation as a consequence of *Myogenin* down regulation. Palacios and colleagues showed that the lack of Six1 and MEF2A TFs leaded to a reduction of DNA demethylation on the *Myogenin* promoter and they proposed that these TFs were necessary for the recruitment of the DNA demethylation machinery. In this scenario, we speculated that Apobec2 could be recruited by Six1 and MEF2A and mediate DNA demethylation at the *Myogenin* promoter. However, further studies are needed to clarify whether the lack of DNA demethylation depends on the controversial deaminase activity of Apobec2 or if it is a consequence of disrupting the TGFβ inhibition or TFs guiding, which could ultimately lead to altered DNA methylation profiles.

6 RELEVANCE OF STUDYING THE DNA METHYLOME DURING MYOGENIC PROGRESSION

How DNA methylation, together with histone modifications, dynamically influences gene expression before, during and after cell-fate commitment is a fundamental question to understand cellular differentiation mechanisms. By analyzing the mouse genome, as a well-established model organism, this study provides a genome-scale and a target specific methylation analysis during myogenic progression, drawing the epigenetic contour that orchestrates myogenesis.

We have identified different epigenetic regulatory dynamics according to the underlying sequence, as well as identified non-promoter DMRs pointing out distinct functional roles of DNA methylation during myogenesis.

Importantly, the identification of common specific DNA methylation signatures along the *ex vivo* and *in vivo* myogenic process provides epigenetic markers of the different myogenic maturation stages. An example of the possible applications of these epimarkers is to provide a reference to ensure the efficient and safe myogenic induction in the context of novel therapeutic strategies involving stem/pluripotent cell reprogramming for muscle pathologies (Darabi, Santos, *et al.,* 2011;Darabi, Pan, *et al.,* 2011). In the same line, the identification of pluripotent specific methylation markers allows to evaluate the epigenetic reprogramming during iPS formation. In addition, we have demonstrated the validity of the inducible *Pax7* ESC-derived myogenic model to study DNA demethylation mechanisms.

According to the current knowledge, the etiology of most muscular diseases does not have an epigenetic basis *per se*; however it is crucial to understand the epigenetic mechanisms and the factors that regulate the muscle physiology to provide the knowledge to understand pathological processes and to address future therapies. The DNA methylation changes identified at tumoral cells (RMS) compared to healthy cells support the hypothesis that aberrant epigenetic events accompanied genetic alterations in cancer, and promotes the study of epigenetic drugs to face muscle tumoral pathologies. In addition, the methylation study of healthy young muscle stem cells represents a starting point to address if epigenetic differences are involved in the decline of the functional potential of the satellite cells in different clinical situations and during aging.

Finally, this study together with others differentiation models contributes to generate a global picture of the DNA methylation dynamics during lineage commitment and cellular differentiation, identifying and distinguishing at the epigenetic level the pluripotent and differentiated cell identity.

CONCLUSIONS

1. The myogenic process is characterized by a dynamic and specific DNA methylation signature established mainly during the lineage commitment process.

2. Muscle cell identity requires DNA demethylation of lineage-specific CpG-poor regulatory regions leading to a transcriptionally active or poised state.

3. Myogenic genes with CpG island promoters are stably unmethylated and are regulated by histone modifications during myogenesis.

4. During myogenesis, pluripotency genes with CpG poor promoters are repressed by DNA methylation and by Polycomb repressive complex 2 when they contain CpG island promoters.

5. *Myf5* super-enhancer activation is controlled by muscle-specific DNA demethylation of its Usf1 binding site, allowing Usf1 binding and enhancement of gene expression.

6. The inducible *Pax7* ESC-derived myoblast precursors recreate the DNA methylation signature of *in vivo* isolated muscle stem cells, supporting the myogenic inducible model as a *bona fide* strategy to generate myogenic precursors *in vitro* with therapeutic purposes.

7. Apobec2 down-regulation blocks the myogenic differentiation by impairing DNA demethylation of the *Myogenin* promoter and abolishes the expression of Myogenin and MHC proteins.

BIBLIOGRAPHY

Abraham, J., Y. Nuñez-Álvarez, S. Hettmer, E. Carrió, H. I. Chen, K. Nishijo, E. T. Huang, S. I. Prajapati, R. L. Walker, S. Davis, J. Rebeles, H. Wiebush, A. T. McCleish, S. T. Hampton, C. R. Bjornson, A. S. Brack, A. J. Wagers, T. A. Rando, M. R. Capecchi, F. C. Marini, B. R. Ehler, L. A. Zarzabal, M. W. Goros, J. E. Michalek, P. S. Meltzer, D. M. Langenau, R. D. LeGallo, A. Mansoor, Y. Chen, M. Suelves, B. P. Rubin, and C. Keller. 2014. "Lineage of origin in rhabdomyosarcoma informs pharmacological response." *Genes Dev* 28 (14):1578-91. doi: 10.1101/gad.238733.114.

Ahmad, K., and S. Henikoff. 2002. "Histone H3 variants specify modes of chromatin assembly." *Proc Natl Acad Sci U S A* 99 Suppl 4:16477-84. doi: 10.1073/pnas.172403699.

Amacher, S. L., J. N. Buskin, and S. D. Hauschka. 1993. "Multiple regulatory elements contribute differentially to muscle creatine kinase enhancer activity in skeletal and cardiac muscle." *Mol Cell Biol* 13 (5):2753-64.

An, C. I., Y. Dong, and N. Hagiwara. 2011. "Genome-wide mapping of Sox6 binding sites in skeletal muscle reveals both direct and indirect regulation of muscle terminal differentiation by Sox6." *BMC Dev Biol* 11:59. doi: 10.1186/1471-213X-11-59.

Anant, S., D. Mukhopadhyay, V. Sankaranand, S. Kennedy, J. O. Henderson, and N. O. Davidson. 2001. "ARCD-1, an apobec-1-related cytidine deaminase, exerts a dominant negative effect on C to U RNA editing." *Am J Physiol Cell Physiol* 281 (6):C1904-16.

Anastasiadou, C., A. Malousi, N. Maglaveras, and S. Kouidou. 2011. "Human epigenome data reveal increased CpG methylation in alternatively spliced sites and putative exonic splicing enhancers." *DNA Cell Biol* 30 (5):267-75. doi: 10.1089/dna.2010.1094.

Anders, S., and W. Huber. 2010. "Differential expression analysis for sequence count data." *Genome Biol* 11 (10):R106. doi: 10.1186/gb-2010-11-10-r106.

Andrés, V., M. Cervera, and V. Mahdavi. 1995. "Determination of the consensus binding site for MEF2 expressed in muscle and brain reveals tissue-specific sequence constraints." *J Biol Chem* 270 (40):23246-9.

Anokye-Danso, F., C. M. Trivedi, D. Juhr, M. Gupta, Z. Cui, Y. Tian, Y. Zhang, W. Yang, P. J. Gruber, J. A. Epstein, and E. E. Morrisey. 2011. "Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency." *Cell Stem Cell* 8 (4):376-88. doi: 10.1016/j.stem.2011.03.001.

Aoki, M., T. Terada, M. Kajiwara, K. Ogasawara, I. Ikai, O. Ogawa, T. Katsura, and K. Inui. 2008. "Kidney-specific expression of human organic cation transporter 2 (OCT2/SLC22A2) is regulated by DNA methylation." *Am J Physiol Renal Physiol* 295 (1):F165-70. doi: 10.1152/ajprenal.90257.2008.

Arand, J., D. Spieler, T. Karius, M. R. Branco, D. Meilinger, A. Meissner, T. Jenuwein, G. Xu, H. Leonhardt, V. Wolf, and J. Walter. 2012. "In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases." *PLoS Genet* 8 (6):e1002750. doi: 10.1371/journal.pgen.1002750.

Archey, W. B., M. P. Sweet, G. C. Alig, and B. A. Arrick. 1999. "Methylation of CpGs as a determinant of transcriptional activation at alternative promoters for transforming growth factor-beta3." *Cancer Res* 59 (10):2292-6.

Arndt, C. A., and W. M. Crist. 1999. "Common musculoskeletal tumors of childhood and adolescence." *N Engl J Med* 341 (5):342-52. doi: 10.1056/NEJM199907293410507.

Asakura, A., G. E. Lyons, and S. J. Tapscott. 1995. "The regulation of MyoD gene expression: conserved elements mediate expression in embryonic axial muscle." *Dev Biol* 171 (2):386-98. doi: 10.1006/dbio.1995.1290.

Asp, P., R. Blum, V. Vethantham, F. Parisi, M. Micsinai, J. Cheng, C. Bowman, Y. Kluger, and B. D. Dynlacht. 2011. "Genome-wide remodeling of the epigenetic landscape during myogenic differentiation." *Proc Natl Acad Sci U S A* 108 (22):E149-58. doi: 10.1073/pnas.1102223108.

Azuara, V., P. Perry, S. Sauer, M. Spivakov, H. F. Jørgensen, R. M. John, M. Gouti, M. Casanova, G. Warnes, M. Merkenschlager, and A. G. Fisher. 2006. "Chromatin signatures of pluripotent cell lines." *Nat Cell Biol* 8 (5):532-8. doi: 10.1038/ncb1403.

Bajard, L., F. Relaix, M. Lagha, D. Rocancourt, P. Daubas, and M. E. Buckingham. 2006. "A novel genetic hierarchy functions during hypaxial myogenesis: Pax3 directly activates Myf5 in muscle progenitor cells in the limb." *Genes Dev* 20 (17):2450-64. doi: 10.1101/gad.382806.

Ball, M. P., J. B. Li, Y. Gao, J. H. Lee, E. M. LeProust, I. H. Park, B. Xie, G. Q. Daley, and G. M. Church. 2009. "Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells." *Nat Biotechnol* 27 (4):361-8. doi: 10.1038/nbt.1533.

Barbieri, G., L. De Angelis, S. Feo, G. Cossu, and A. Giallongo. 1990. "Differential expression of muscle-specific enolase in embryonic and fetal myogenic cells during mouse development." *Differentiation* 45 (3):179-84.

Barrera, L. O., Z. Li, A. D. Smith, K. C. Arden, W. K. Cavenee, M. Q. Zhang, R. D. Green, and B. Ren. 2008. "Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs." *Genome Res* 18 (1):46-59. doi: 10.1101/gr.6654808.

Barrera, V., and M. A. Peinado. 2012. "Evaluation of single CpG sites as proxies of CpG island methylation states at the genome scale." *Nucleic Acids Res* 40 (22):11490-8. doi: 10.1093/nar/gks928.

Barrès, R., J. Yan, B. Egan, J. T. Treebak, M. Rasmussen, T. Fritz, K. Caidahl, A. Krook, D. J. O'Gorman, and J. R. Zierath. 2012. "Acute exercise remodels promoter methylation in human skeletal muscle." *Cell Metab* 15 (3):405-11. doi: 10.1016/j.cmet.2012.01.001.

Bartke, T., M. Vermeulen, B. Xhemalce, S. C. Robson, M. Mann, and T. Kouzarides. 2010. "Nucleosome-interacting proteins regulated by DNA and histone methylation." *Cell* 143 (3):470-84. doi: 10.1016/j.cell.2010.10.012.

Baubec, T., and D. Schübeler. 2014. "Genomic patterns and context specific interpretation of DNA methylation." *Curr Opin Genet Dev* 25:85-92. doi: 10.1016/j.gde.2013.11.015.

Bentzinger, C. F., Y. X. Wang, and M. A. Rudnicki. 2012. "Building muscle: molecular regulation of myogenesis." *Cold Spring Harb Perspect Biol* 4 (2). doi: 10.1101/cshperspect.a008342.

Berdasco, M., C. Melguizo, J. Prados, A. Gómez, M. Alaminos, M. A. Pujana, M. Lopez, F. Setien, R. Ortiz, I. Zafra, A. Aranega, and M. Esteller. 2012. "DNA methylation plasticity of human adipose-derived stem cells in lineage commitment." *Am J Pathol* 181 (6):2079-93. doi: 10.1016/j.ajpath.2012.08.016.

Berkes, C. A., D. A. Bergstrom, B. H. Penn, K. J. Seaver, P. S. Knoepfler, and S. J. Tapscott. 2004. "Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential." *Mol Cell* 14 (4):465-77.

Bernstein, B. E., T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber, and E. S. Lander. 2006. "A bivalent chromatin structure marks key developmental genes in embryonic stem cells." *Cell* 125 (2):315-26. doi: 10.1016/j.cell.2006.02.041.

Bestor, T. H. 2003. "Cytosine methylation mediates sexual conflict." *Trends Genet* 19 (4):185-90. doi: 10.1016/S0168-9525(03)00049-0.

Bestor, T. H., and D. Bourc'his. 2004. "Transposon silencing and imprint establishment in mammalian germ cells." *Cold Spring Harb Symp Quant Biol* 69:381-7. doi: 10.1101/sqb.2004.69.381.

Bestor, T. H., and A. Coxon. 1993. "Cytosine methylation: the pros and cons of DNA methylation." *Curr Biol* 3 (6):384-6.

Biesiada, E., Y. Hamamori, L. Kedes, and V. Sartorelli. 1999. "Myogenic basic helix-loop-helix proteins and Sp1 interact as components of a multiprotein transcriptional complex

required for activity of the human cardiac alpha-actin promoter." *Mol Cell Biol* 19 (4):2577-84.
Bird, A. P. 1984. "DNA methylation versus gene expression." *J Embryol Exp Morphol* 83 Suppl:31-40.
Bird, A. P., and A. P. Wolffe. 1999. "Methylation-induced repression--belts, braces, and chromatin." *Cell* 99 (5):451-4.
Bischoff, R. 1975. "Regeneration of single skeletal muscle fibers *in vitro*." *Anat Rec* 182 (2):215-35. doi: 10.1002/ar.1091820207.
Black, B. L., J. F. Martin, and E. N. Olson. 1995. "The mouse MRF4 promoter is trans-activated directly and indirectly by muscle-specific transcription factors." *J Biol Chem* 270 (7):2889-92.
Blais, A., M. Tsikitis, D. Acosta-Alvear, R. Sharan, Y. Kluger, and B. D. Dynlacht. 2005. "An initial blueprint for myogenic differentiation." *Genes Dev* 19 (5):553-69. doi: 10.1101/gad.1281105.
Blattler, A., and P. J. Farnham. 2013. "Cross-talk between site-specific transcription factors and DNA methylation states." *J Biol Chem* 288 (48):34287-94. doi: 10.1074/jbc.R113.512517.
Blum, R., V. Vethantham, C. Bowman, M. Rudnicki, and B. D. Dynlacht. 2012. "Genome-wide identification of enhancers in skeletal muscle: the role of MyoD1." *Genes Dev* 26 (24):2763-79. doi: 10.1101/gad.200113.112.
Bock, C., I. Beerman, W. H. Lien, Z. D. Smith, H. Gu, P. Boyle, A. Gnirke, E. Fuchs, D. J. Rossi, and A. Meissner. 2012. "DNA methylation dynamics during *in vivo* differentiation of blood and skin stem cells." *Mol Cell* 47 (4):633-47. doi: 10.1016/j.molcel.2012.06.019.
Bocker, M. T., I. Hellwig, A. Breiling, V. Eckstein, A. D. Ho, and F. Lyko. 2011. "Genome-wide promoter DNA methylation dynamics of human hematopoietic progenitor cells during differentiation and aging." *Blood* 117 (19):e182-9. doi: 10.1182/blood-2011-01-331926.
Bodine, S. C. 2013. "Disuse-induced muscle wasting." *Int J Biochem Cell Biol* 45 (10):2200-8. doi: 10.1016/j.biocel.2013.06.011.
Borello, U., B. Berarducci, P. Murphy, L. Bajard, V. Buffa, S. Piccolo, M. Buckingham, and G. Cossu. 2006. "The Wnt/beta-catenin pathway regulates Gli-mediated Myf5 expression during somitogenesis." *Development* 133 (18):3723-32. doi: 10.1242/dev.02517.
Borgel, J., S. Guibert, Y. Li, H. Chiba, D. Schübeler, H. Sasaki, T. Forné, and M. Weber. 2010. "Targets and dynamics of promoter DNA methylation during early mouse development." *Nat Genet* 42 (12):1093-100. doi: 10.1038/ng.708.
Bosnakovski, D., Z. Xu, W. Li, S. Thet, O. Cleaver, R. C. Perlingeiro, and M. Kyba. 2008. "Prospective isolation of skeletal muscle stem cells with a Pax7 reporter." *Stem Cells* 26 (12):3194-204. doi: 10.1634/stemcells.2007-1017.
Braun, T., E. Bober, G. Buschhausen-Denker, S. Kohtz, K. H. Grzeschik, H. H. Arnold, and S. Kotz. 1989. "Differential expression of myogenic determination genes in muscle cells: possible autoactivation by the Myf gene products." *EMBO J* 8 (12):3617-25.
Braun, T., E. Bober, B. Winter, N. Rosenthal, and H. H. Arnold. 1990. "Myf-6, a new member of the human gene family of myogenic determination factors: evidence for a gene cluster on chromosome 12." *EMBO J* 9 (3):821-31.
Brent, A. E., and C. J. Tabin. 2002. "Developmental regulation of somite derivatives: muscle, cartilage and tendon." *Curr Opin Genet Dev* 12 (5):548-57.
Brigelius-Flohé, R., C. Müller, J. Menard, S. Florian, K. Schmehl, and K. Wingler. 2001. "Functions of GI-GPx: lessons from selenium-dependent expression and intracellular localization." *Biofactors* 14 (1-4):101-6.
Brinkman, A. B., H. Gu, S. J. Bartels, Y. Zhang, F. Matarese, F. Simmer, H. Marks, C. Bock, A. Gnirke, A. Meissner, and H. G. Stunnenberg. 2012. "Sequential ChIP-bisulfite

sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk." *Genome Res* 22 (6):1128-38. doi: 10.1101/gr.133728.111.

Brown, C. B., K. A. Engleka, J. Wenning, M. Min Lu, and J. A. Epstein. 2005. "Identification of a hypaxial somite enhancer element regulating Pax3 expression in migrating myoblasts and characterization of hypaxial muscle Cre transgenic mice." *Genesis* 41 (4):202-9. doi: 10.1002/gene.20116.

Brown, S. J., P. Stoilov, and Y. Xing. 2012. "Chromatin and epigenetic regulation of pre-mRNA processing." *Hum Mol Genet* 21 (R1):R90-6. doi: 10.1093/hmg/dds353.

Brunk, B. P., D. J. Goldhamer, and C. P. Emerson. 1996. "Regulated demethylation of the myoD distal enhancer during skeletal myogenesis." *Dev Biol* 177 (2):490-503.

Brunner, A. L., D. S. Johnson, S. W. Kim, A. Valouev, T. E. Reddy, N. F. Neff, E. Anton, C. Medina, L. Nguyen, E. Chiao, C. B. Oyolu, G. P. Schroth, D. M. Absher, J. C. Baker, and R. M. Myers. 2009. "Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver." *Genome Res* 19 (6):1044-56. doi: 10.1101/gr.088773.108.

Buck-Koehntop, B. A., and P. A. Defossez. 2013. "On how mammalian transcription factors recognize methylated DNA." *Epigenetics* 8 (2):131-7. doi: 10.4161/epi.23632.

Buckingham, M., and F. Relaix. 2007. "The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions." *Annu Rev Cell Dev Biol* 23:645-73. doi: 10.1146/annurev.cellbio.23.090506.123438.

Buckingham, M., and P. W. Rigby. 2014. "Gene regulatory networks and transcriptional mechanisms that control myogenesis." *Dev Cell* 28 (3):225-38. doi: 10.1016/j.devcel.2013.12.020.

Burton, A., and M. E. Torres-Padilla. 2010. "Epigenetic reprogramming and development: a unique heterochromatin organization in the preimplantation mouse embryo." *Brief Funct Genomics* 9 (5-6):444-54. doi: 10.1093/bfgp/elq027.

Cameron, E. E., K. E. Bachman, S. Myöhänen, J. G. Herman, and S. B. Baylin. 1999. "Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer." *Nat Genet* 21 (1):103-7. doi: 10.1038/5047.

Campanero, M. R., M. I. Armstrong, and E. K. Flemington. 2000. "CpG methylation as a mechanism for the regulation of E2F activity." *Proc Natl Acad Sci U S A* 97 (12):6481-6. doi: 10.1073/pnas.100340697.

Cao, L., Y. Yu, S. Bilke, R. L. Walker, L. H. Mayeenuddin, D. O. Azorsa, F. Yang, M. Pineda, L. J. Helman, and P. S. Meltzer. 2010. "Genome-wide identification of Pax3-FKHR binding sites in rhabdomyosarcoma reveals candidate target genes important for development and cancer." *Cancer Res* 70 (16):6497-508. doi: 10.1158/0008-5472.CAN-10-0582.

Cao, Y., R. M. Kumar, B. H. Penn, C. A. Berkes, C. Kooperberg, L. A. Boyer, R. A. Young, and S. J. Tapscott. 2006. "Global and gene-specific analyses show distinct roles for Myod and Myog at a common set of promoters." *EMBO J* 25 (3):502-11. doi: 10.1038/sj.emboj.7600958.

Cao, Y., Z. Yao, D. Sarkar, M. Lawrence, G. J. Sanchez, M. H. Parker, K. L. MacQuarrie, J. Davison, M. T. Morgan, W. L. Ruzzo, R. C. Gentleman, and S. J. Tapscott. 2010. "Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming." *Dev Cell* 18 (4):662-74. doi: 10.1016/j.devcel.2010.02.014.

Capuano, F., M. Mülleder, R. Kok, H. J. Blom, and M. Ralser. 2014. "Cytosine DNA methylation is found in Drosophila melanogaster but absent in Saccharomyces cerevisiae, Schizosaccharomyces pombe, and other yeast species." *Anal Chem* 86 (8):3697-702. doi: 10.1021/ac500447w.

Card, D. A., P. B. Hebbar, L. Li, K. W. Trotter, Y. Komatsu, Y. Mishina, and T. K. Archer. 2008. "Oct4/Sox2-regulated miR-302 targets cyclin D1 in human embryonic stem cells." *Mol Cell Biol* 28 (20):6426-38. doi: 10.1128/MCB.00359-08.

Carvajal, J. J., D. Cox, D. Summerbell, and P. W. Rigby. 2001. "A BAC transgenic analysis of the Mrf4/Myf5 locus reveals interdigitated elements that control activation and maintenance of gene expression during muscle development." *Development* 128 (10):1857-68.

Carvajal, J. J., A. Keith, and P. W. Rigby. 2008. "Global transcriptional regulation of the locus encoding the skeletal muscle determination genes Mrf4 and Myf5." *Genes Dev* 22 (2):265-76. doi: 10.1101/gad.442408.

Cedar, H., and Y. Bergman. 2009. "Linking DNA methylation and histone modification: patterns and paradigms." *Nat Rev Genet* 10 (5):295-304. doi: 10.1038/nrg2540.

Chang, T. H., M. Primig, J. Hadchouel, S. Tajbakhsh, D. Rocancourt, A. Fernandez, R. Kappler, H. Scherthan, and M. Buckingham. 2004. "An enhancer directs differential expression of the linked Mrf4 and Myf5 myogenic regulatory genes in the mouse." *Dev Biol* 269 (2):595-608. doi: 10.1016/j.ydbio.2004.02.013.

Chen, A. E., D. D. Ginty, and C. M. Fan. 2005. "Protein kinase A signalling via CREB controls myogenesis induced by Wnt proteins." *Nature* 433 (7023):317-22. doi: 10.1038/nature03126.

Chen, C., D. Ridzon, C. T. Lee, J. Blake, Y. Sun, and W. M. Strauss. 2007. "Defining embryonic stem cell identity using differentiation-related microRNAs and their potential targets." *Mamm Genome* 18 (5):316-27. doi: 10.1007/s00335-007-9032-6.

Chen, X., H. Xu, P. Yuan, F. Fang, M. Huss, V. B. Vega, E. Wong, Y. L. Orlov, W. Zhang, J. Jiang, Y. H. Loh, H. C. Yeo, Z. X. Yeo, V. Narang, K. R. Govindarajan, B. Leong, A. Shahab, Y. Ruan, G. Bourque, W. K. Sung, N. D. Clarke, C. L. Wei, and H. H. Ng. 2008. "Integration of external signaling pathways with the core transcriptional network in embryonic stem cells." *Cell* 133 (6):1106-17. doi: 10.1016/j.cell.2008.04.043.

Chen, Z. X., and A. D. Riggs. 2011. "DNA methylation and demethylation in mammals." *J Biol Chem* 286 (21):18347-53. doi: 10.1074/jbc.R110.205286.

Cheong, J., Y. Yamada, R. Yamashita, T. Irie, A. Kanai, H. Wakaguri, K. Nakai, T. Ito, I. Saito, S. Sugano, and Y. Suzuki. 2006. "Diverse DNA methylation statuses at alternative promoters of human genes in various tissues." *DNA Res* 13 (4):155-67. doi: 10.1093/dnares/dsl008.

Chodavarapu, R. K., S. Feng, Y. V. Bernatavichute, P. Y. Chen, H. Stroud, Y. Yu, J. A. Hetzel, F. Kuo, J. Kim, S. J. Cokus, D. Casero, M. Bernal, P. Huijser, A. T. Clark, U. Krämer, S. S. Merchant, X. Zhang, S. E. Jacobsen, and M. Pellegrini. 2010. "Relationship between nucleosome positioning and DNA methylation." *Nature* 466 (7304):388-92. doi: 10.1038/nature09147.

Christman, J. K. 2002. "5-Azacytidine and 5-Aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy." *Oncogene* 21 (35):5483-95. doi: 10.1038/sj.onc.1205699.

Clark, S. J., A. Statham, C. Stirzaker, P. L. Molloy, and M. Frommer. 2006. "DNA methylation: bisulphite modification and analysis." *Nat Protoc* 1 (5):2353-64. doi: 10.1038/nprot.2006.324.

Condon, K., L. Silberstein, H. M. Blau, and W. J. Thompson. 1990. "Development of muscle fiber types in the prenatal rat hindlimb." *Dev Biol* 138 (2):256-74.

Consortium, ENCODE Project. 2012. "An integrated encyclopedia of DNA elements in the human genome." *Nature* 489 (7414):57-74. doi: 10.1038/nature11247.

Constantinides, P. G., P. A. Jones, and W. Gevers. 1977. "Functional striated muscle cells from non-myoblast precursors following 5-Azacytidine treatment." *Nature* 267 (5609):364-6.

Conticello, S. G. 2008. "The AID/Apobec family of nucleic acid mutators." *Genome Biol* 9 (6):229. doi: 10.1186/gb-2008-9-6-229.

Cortellino, S., J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, and A. Bellacosa. 2011. "Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair." *Cell* 146 (1):67-79. doi: 10.1016/j.cell.2011.06.020.

Cortázar, D., C. Kunz, Y. Saito, R. Steinacher, and P. Schär. 2007. "The enigmatic thymine DNA glycosylase." *DNA Repair (Amst)* 6 (4):489-504. doi: 10.1016/j.dnarep.2006.10.013.

Cortázar, D., C. Kunz, J. Selfridge, T. Lettieri, Y. Saito, E. MacDougall, A. Wirz, D. Schuermann, A. L. Jacobs, F. Siegrist, R. Steinacher, J. Jiricny, A. Bird, and P. Schär. 2011. "Embryonic lethal phenotype reveals a function of Tdg in maintaining epigenetic stability." *Nature* 470 (7334):419-23. doi: 10.1038/nature09672.

Cosgrove, B. D., A. Sacco, P. M. Gilbert, and H. M. Blau. 2009. "A home away from home: challenges and opportunities in engineering *in vitro* muscle satellite cell niches." *Differentiation* 78 (2-3):185-94. doi: 10.1016/j.diff.2009.08.004.

Cossu, G., and U. Borello. 1999. "Wnt signaling and the activation of myogenesis in mammals." *EMBO J* 18 (24):6867-72. doi: 10.1093/emboj/18.24.6867.

Cossu, G., and M. Molinaro. 1987. "Cell heterogeneity in the myogenic lineage." *Curr Top Dev Biol* 23:185-208.

Côté, J., J. Quinn, J. L. Workman, and C. L. Peterson. 1994. "Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex." *Science* 265 (5168):53-60.

Creyghton, M. P., A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young, and R. Jaenisch. 2010. "Histone H3K27ac separates active from poised enhancers and predicts developmental state." *Proc Natl Acad Sci U S A* 107 (50):21931-6. doi: 10.1073/pnas.1016071107.

Dahl, J. A., and P. Collas. 2008. "A rapid micro chromatin immunoprecipitation assay (microChIP)." *Nat Protoc* 3 (6):1032-45. doi: 10.1038/nprot.2008.68.

Darabi, R., K. Gehlbach, R. M. Bachoo, S. Kamath, M. Osawa, K. E. Kamm, M. Kyba, and R. C. Perlingeiro. 2008. "Functional skeletal muscle regeneration from differentiating embryonic stem cells." *Nat Med* 14 (2):134-43. doi: 10.1038/nm1705.

Darabi, R., W. Pan, D. Bosnakovski, J. Baik, M. Kyba, and R. C. Perlingeiro. 2011. "Functional myogenic engraftment from mouse iPS cells." *Stem Cell Rev* 7 (4):948-57. doi: 10.1007/s12015-011-9258-2.

Darabi, R., F. N. Santos, A. Filareto, W. Pan, R. Koene, M. A. Rudnicki, M. Kyba, and R. C. Perlingeiro. 2011. "Assessment of the myogenic stem cell compartment following transplantation of Pax3/Pax7-induced embryonic stem cell-derived progenitors." *Stem Cells* 29 (5):777-90. doi: 10.1002/stem.625.

Davis, M. R., and K. M. Summers. 2012. "Structure and function of the mammalian fibrillin gene family: implications for human connective tissue diseases." *Mol Genet Metab* 107 (4):635-47. doi: 10.1016/j.ymgme.2012.07.023.

Davis, R. L., H. Weintraub, and A. B. Lassar. 1987. "Expression of a single transfected cDNA converts fibroblasts to myoblasts." *Cell* 51 (6):987-1000.

Dawlaty, M. M., A. Breiling, T. Le, M. I. Barrasa, G. Raddatz, Q. Gao, B. E. Powell, A. W. Cheng, K. F. Faull, F. Lyko, and R. Jaenisch. 2014. "Loss of Tet enzymes compromises proper differentiation of embryonic stem cells." *Dev Cell* 29 (1):102-11. doi: 10.1016/j.devcel.2014.03.003.

Dawlaty, M. M., A. Breiling, T. Le, G. Raddatz, M. I. Barrasa, A. W. Cheng, Q. Gao, B. E. Powell, Z. Li, M. Xu, K. F. Faull, F. Lyko, and R. Jaenisch. 2013. "Combined deficiency of Tet1

and Tet2 causes epigenetic abnormalities but is compatible with postnatal development." *Dev Cell* 24 (3):310-23. doi: 10.1016/j.devcel.2012.12.015.

Dawlaty, M. M., K. Ganz, B. E. Powell, Y. C. Hu, S. Markoulaki, A. W. Cheng, Q. Gao, J. Kim, S. W. Choi, D. C. Page, and R. Jaenisch. 2011. "Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development." *Cell Stem Cell* 9 (2):166-75. doi: 10.1016/j.stem.2011.07.010.

de la Rica, L., J. Rodríguez-Ubreva, M. García, A. B. Islam, J. M. Urquiza, H. Hernando, J. Christensen, K. Helin, C. Gómez-Vaquero, and E. Ballestar. 2013. "PU.1 target genes undergo Tet2-coupled demethylation and Dnmt3b-mediated methylation in monocyte-to-osteoclast differentiation." *Genome Biol* 14 (9):R99. doi: 10.1186/gb-2013-14-9-r99.

Deaton, A. M., and A. Bird. 2011. "CpG islands and the regulation of transcription." *Genes Dev* 25 (10):1010-22. doi: 10.1101/gad.2037511.

Delogu, A., A. Schebesta, Q. Sun, K. Aschenbrenner, T. Perlot, and M. Busslinger. 2006. "Gene repression by Pax5 in B cells is essential for blood cell homeostasis and is reversed in plasma cells." *Immunity* 24 (3):269-81. doi: 10.1016/j.immuni.2006.01.012.

Deng, C., Y. Li, S. Liang, K. Cui, T. Salz, H. Yang, Z. Tang, P. G. Gallagher, Y. Qiu, R. Roeder, K. Zhao, J. Bungert, and S. Huang. 2013. "Usf1 and hSET1A mediated epigenetic modifications regulate lineage differentiation and HoxB4 transcription." *PLoS Genet* 9 (6):e1003524. doi: 10.1371/journal.pgen.1003524.

Doi, A., I. H. Park, B. Wen, P. Murakami, M. J. Aryee, R. Irizarry, B. Herb, C. Ladd-Acosta, J. Rho, S. Loewer, J. Miller, T. Schlaeger, G. Q. Daley, and A. P. Feinberg. 2009. "Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts." *Nat Genet* 41 (12):1350-3. doi: 10.1038/ng.471.

Donoviel, D. B., M. A. Shield, J. N. Buskin, H. S. Haugen, C. H. Clegg, and S. D. Hauschka. 1996. "Analysis of muscle creatine kinase gene regulatory elements in skeletal and cardiac muscles of transgenic mice." *Mol Cell Biol* 16 (4):1649-58.

Du, C., Y. Q. Jin, J. J. Qi, Z. X. Ji, S. Y. Li, G. S. An, H. T. Jia, and J. H. Ni. 2012. "Effects of myogenin on expression of late muscle genes through MyoD-dependent chromatin remodeling ability of myogenin." *Mol Cells* 34 (2):133-42. doi: 10.1007/s10059-012-2286-1.

Ecker, J. R., W. A. Bickmore, I. Barroso, J. K. Pritchard, Y. Gilad, and E. Segal. 2012. "Genomics: ENCODE explained." *Nature* 489 (7414):52-5. doi: 10.1038/489052a.

Eguizabal, C., N. Montserrat, A. Veiga, and J. C. Izpisua Belmonte. 2013. "Dedifferentiation, transdifferentiation, and reprogramming: future directions in regenerative medicine." *Semin Reprod Med* 31 (1):82-94. doi: 10.1055/s-0032-1331802.

Emery, A. E. 2002. "The muscular dystrophies." *Lancet* 359 (9307):687-95. doi: 10.1016/S0140-6736(02)07815-7.

Ernst, J., P. Kheradpour, T. S. Mikkelsen, N. Shoresh, L. D. Ward, C. B. Epstein, X. Zhang, L. Wang, R. Issner, M. Coyne, M. Ku, T. Durham, M. Kellis, and B. E. Bernstein. 2011. "Mapping and analysis of chromatin state dynamics in nine human cell types." *Nature* 473 (7345):43-9. doi: 10.1038/nature09906.

Esteller, M. 2007. "Epigenetic gene silencing in cancer: the DNA hypermethylome." *Hum Mol Genet* 16 Spec No 1:R50-9. doi: 10.1093/hmg/ddm018.

Evans, M. J., and M. H. Kaufman. 1981. "Establishment in culture of pluripotential cells from mouse embryos." *Nature* 292 (5819):154-6.

Evsikov, A. V., W. N. de Vries, A. E. Peaston, E. E. Radford, K. S. Fancher, F. H. Chen, J. A. Blake, C. J. Bult, K. E. Latham, D. Solter, and B. B. Knowles. 2004. "Systems biology of the 2-cell mouse embryo." *Cytogenet Genome Res* 105 (2-4):240-50. doi: 10.1159/000078195.

Feldmann, A., R. Ivanek, R. Murr, D. Gaidatzis, L. Burger, and D. Schübeler. 2013. "Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions." *PLoS Genet* 9 (12):e1003994. doi: 10.1371/journal.pgen.1003994.

Feng, S., S. E. Jacobsen, and W. Reik. 2010. "Epigenetic reprogramming in plant and animal development." *Science* 330 (6004):622-7. doi: 10.1126/science.1190614.

Fernandez, A. F., Y. Assenov, J. I. Martin-Subero, B. Balint, R. Siebert, H. Taniguchi, H. Yamamoto, M. Hidalgo, A. C. Tan, O. Galm, I. Ferrer, M. Sanchez-Cespedes, A. Villanueva, J. Carmona, J. V. Sanchez-Mut, M. Berdasco, V. Moreno, G. Capella, D. Monk, E. Ballestar, S. Ropero, R. Martinez, M. Sanchez-Carbayo, F. Prosper, X. Agirre, M. F. Fraga, O. Graña, L. Perez-Jurado, J. Mora, S. Puig, J. Prat, L. Badimon, A. A. Puca, S. J. Meltzer, T. Lengauer, J. Bridgewater, C. Bock, and M. Esteller. 2012. "A DNA methylation fingerprint of 1628 human samples." *Genome Res* 22 (2):407-19. doi: 10.1101/gr.119867.110.

Ficz, G., M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, and W. Reik. 2011. "Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation." *Nature* 473 (7347):398-402. doi: 10.1038/nature10008.

Follenzi, A., L. E. Ailles, S. Bakovic, M. Geuna, and L. Naldini. 2000. "Gene transfer by lentiviral vectors is limited by nuclear translocation and rescued by HIV-1 pol sequences." *Nat Genet* 25 (2):217-22. doi: 10.1038/76095.

Fong, A. P., Z. Yao, J. W. Zhong, Y. Cao, W. L. Ruzzo, R. C. Gentleman, and S. J. Tapscott. 2012. "Genetic and epigenetic determinants of neurogenesis and myogenesis." *Dev Cell* 22 (4):721-35. doi: 10.1016/j.devcel.2012.01.015.

Francetic, T., and Q. Li. 2011. "Skeletal myogenesis and Myf5 activation." *Transcription* 2 (3):109-114. doi: 10.4161/trns.2.3.15829.

Frigola, J., M. Ribas, R. A. Risques, and M. A. Peinado. 2002. "Methylome profiling of cancer cells by amplification of inter-methylated sites (AIMS)." *Nucleic Acids Res* 30 (7):e28.

Frigola, J., X. Solé, M. F. Paz, V. Moreno, M. Esteller, G. Capellà, and M. A. Peinado. 2005. "Differential DNA hypermethylation and hypomethylation signatures in colorectal cancer." *Hum Mol Genet* 14 (2):319-26. doi: 10.1093/hmg/ddi028.

Fujii, G., Y. Nakamura, D. Tsukamoto, M. Ito, T. Shiba, and N. Takamatsu. 2006. "CpG methylation at the Usf-binding site is important for the liver-specific transcription of the chipmunk HP-27 gene." *Biochem J* 395 (1):203-9. doi: 10.1042/BJ20051802.

Fuso, A., G. Ferraguti, F. Grandoni, R. Ruggeri, S. Scarpa, R. Strom, and M. Lucarelli. 2010. "Early demethylation of non-CpG, CpC-rich, elements in the myogenin 5'-flanking region: a priming effect on the spreading of active demethylation." *Cell Cycle* 9 (19):3965-76.

Ganapathi, M., P. Srivastava, S. K. Das Sutar, K. Kumar, D. Dasgupta, G. Pal Singh, V. Brahmachari, and S. K. Brahmachari. 2005. "Comparative analysis of chromatin landscape in regulatory regions of human housekeeping and tissue specific genes." *BMC Bioinformatics* 6:126. doi: 10.1186/1471-2105-6-126.

Gardiner-Garden, M., and M. Frommer. 1987. "CpG islands in vertebrate genomes." *J Mol Biol* 196 (2):261-82.

Gaudet, F., J. G. Hodgson, A. Eden, L. Jackson-Grusby, J. Dausman, J. W. Gray, H. Leonhardt, and R. Jaenisch. 2003. "Induction of tumors in mice by genomic hypomethylation." *Science* 300 (5618):489-92. doi: 10.1126/science.1083558.

Gelfman, S., N. Cohen, A. Yearim, and G. Ast. 2013. "DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure." *Genome Res* 23 (5):789-99. doi: 10.1101/gr.143503.112.

Gifford, C. A., M. J. Ziller, H. Gu, C. Trapnell, J. Donaghey, A. Tsankov, A. K. Shalek, D. R. Kelley, A. A. Shishkin, R. Issner, X. Zhang, M. Coyne, J. L. Fostel, L. Holmes, J. Meldrim, M.

Guttman, C. Epstein, H. Park, O. Kohlbacher, J. Rinn, A. Gnirke, E. S. Lander, B. E. Bernstein, and A. Meissner. 2013. "Transcriptional and epigenetic dynamics during specification of human embryonic stem cells." *Cell* 153 (5):1149-63. doi: 10.1016/j.cell.2013.04.037.

Gilsbach, R., S. Preissl, B. A. Grüning, T. Schnick, L. Burger, V. Benes, A. Würch, U. Bönisch, S. Günther, R. Backofen, B. K. Fleischmann, D. Schübeler, and L. Hein. 2014. "Dynamic DNA methylation orchestrates cardiomyocyte development, maturation and disease." *Nat Commun* 5:5288. doi: 10.1038/ncomms6288.

Giordani, J., L. Bajard, J. Demignon, P. Daubas, M. Buckingham, and P. Maire. 2007. "Six proteins regulate the activation of Myf5 expression in embryonic mouse limbs." *Proc Natl Acad Sci U S A* 104 (27):11310-5. doi: 10.1073/pnas.0611299104.

Gnocchi, V. F., R. B. White, Y. Ono, J. A. Ellis, and P. S. Zammit. 2009. "Further characterisation of the molecular signature of quiescent and activated mouse muscle satellite cells." *PLoS One* 4 (4):e5205. doi: 10.1371/journal.pone.0005205.

Goldhamer, D. J., B. P. Brunk, A. Faerman, A. King, M. Shani, and C. P. Emerson. 1995. "Embryonic activation of the myoD gene is regulated by a highly conserved distal control element." *Development* 121 (3):637-49.

Gonzalez-Roca, E., X. Garcia-Albéniz, S. Rodriguez-Mulero, R. R. Gomis, K. Kornacker, and H. Auer. 2010. "Accurate expression profiling of very small cell populations." *PLoS One* 5 (12):e14418. doi: 10.1371/journal.pone.0014418.

Gopalakrishnan, S., B. A. Sullivan, S. Trazzi, G. Della Valle, and K. D. Robertson. 2009. "Dnmt3b interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone code at centromeric regions." *Hum Mol Genet* 18 (17):3178-93. doi: 10.1093/hmg/ddp256.

Gopinath, S. D., and T. A. Rando. 2008. "Stem cell review series: aging of the skeletal muscle stem cell niche." *Aging Cell* 7 (4):590-8. doi: 10.1111/j.1474-9726.2008.00399.x.

Goulding, M., A. Lumsden, and A. J. Paquette. 1994. "Regulation of Pax-3 expression in the dermomyotome and its role in muscle development." *Development* 120 (4):957-71.

Gronostajski, R. M. 2000. "Roles of the NFI/CTF gene family in transcription and development." *Gene* 249 (1-2):31-45.

Gros, J., O. Serralbo, and C. Marcelle. 2009. "WNT11 acts as a directional cue to organize the elongation of early muscle fibres." *Nature* 457 (7229):589-93. doi: 10.1038/nature07564.

Gu, T. P., F. Guo, H. Yang, H. P. Wu, G. F. Xu, W. Liu, Z. G. Xie, L. Shi, X. He, S. G. Jin, K. Iqbal, Y. G. Shi, Z. Deng, P. E. Szabó, G. P. Pfeifer, J. Li, and G. L. Xu. 2011. "The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes." *Nature* 477 (7366):606-10. doi: 10.1038/nature10443.

Guenther, M. G., S. S. Levine, L. A. Boyer, R. Jaenisch, and R. A. Young. 2007. "A chromatin landmark and transcription initiation at most promoters in human cells." *Cell* 130 (1):77-88. doi: 10.1016/j.cell.2007.05.042.

Guo, F., X. Li, D. Liang, T. Li, P. Zhu, H. Guo, X. Wu, L. Wen, T. P. Gu, B. Hu, C. P. Walsh, J. Li, F. Tang, and G. L. Xu. 2014. "Active and passive demethylation of male and female pronuclear DNA in the Mammalian zygote." *Cell Stem Cell* 15 (4):447-58. doi: 10.1016/j.stem.2014.08.003.

Guo, J. U., D. K. Ma, H. Mo, M. P. Ball, M. H. Jang, M. A. Bonaguidi, J. A. Balazer, H. L. Eaves, B. Xie, E. Ford, K. Zhang, G. L. Ming, Y. Gao, and H. Song. 2011. "Neuronal activity modifies the DNA methylation landscape in the adult brain." *Nat Neurosci* 14 (10):1345-51. doi: 10.1038/nn.2900.

Guo, J. U., Y. Su, J. H. Shin, J. Shin, H. Li, B. Xie, C. Zhong, S. Hu, T. Le, G. Fan, H. Zhu, Q. Chang, Y. Gao, G. L. Ming, and H. Song. 2014. "Distribution, recognition and regulation of non-

CpG methylation in the adult mammalian brain." *Nat Neurosci* 17 (2):215-22. doi: 10.1038/nn.3607.

Guo, J. U., Y. Su, C. Zhong, G. L. Ming, and H. Song. 2011. "Hydroxylation of 5-methylcytosine by Tet1 promotes active DNA demethylation in the adult brain." *Cell* 145 (3):423-34. doi: 10.1016/j.cell.2011.03.022.

Gustafsson, M. K., H. Pan, D. F. Pinney, Y. Liu, A. Lewandowski, D. J. Epstein, and C. P. Emerson. 2002. "Myf5 is a direct target of long-range Shh signaling and Gli regulation for muscle specification." *Genes Dev* 16 (1):114-26. doi: 10.1101/gad.940702.

Günther, S., J. Kim, S. Kostin, C. Lepper, C. M. Fan, and T. Braun. 2013. "Myf5-positive satellite cells contribute to Pax7-dependent long-term maintenance of adult muscle stem cells." *Cell Stem Cell* 13 (5):590-601. doi: 10.1016/j.stem.2013.07.016.

Hackett, J. A., J. J. Zylicz, and M. A. Surani. 2012. "Parallel mechanisms of epigenetic reprogramming in the germline." *Trends Genet* 28 (4):164-74. doi: 10.1016/j.tig.2012.01.005.

Hadchouel, J., J. J. Carvajal, P. Daubas, L. Bajard, T. Chang, D. Rocancourt, D. Cox, D. Summerbell, S. Tajbakhsh, P. W. Rigby, and M. Buckingham. 2003. "Analysis of a key regulatory region upstream of the Myf5 gene reveals multiple phases of myogenesis, orchestrated at each site by a combination of elements dispersed throughout the locus." *Development* 130 (15):3415-26.

Hagiwara, N., K. Toyoda, R. Torisu, T. Inoue, K. Yasumori, S. Ibayashi, and Y. Okada. 2007. "Progressive stroke involving bilateral medial medulla expanding to spinal cord due to vertebral artery dissection." *Cerebrovasc Dis* 24 (6):540-2. doi: 10.1159/000111220.

Hahn, M. A., R. Qiu, X. Wu, A. X. Li, H. Zhang, J. Wang, J. Jui, S. G. Jin, Y. Jiang, G. P. Pfeifer, and Q. Lu. 2013. "Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis." *Cell Rep* 3 (2):291-300. doi: 10.1016/j.celrep.2013.01.011.

Haines, T. R., D. I. Rodenhiser, and P. J. Ainsworth. 2001. "Allele-specific non-CpG methylation of the Nf1 gene during early mouse development." *Dev Biol* 240 (2):585-98. doi: 10.1006/dbio.2001.0504.

Hajkova, P., S. Erhardt, N. Lane, T. Haaf, O. El-Maarri, W. Reik, J. Walter, and M. A. Surani. 2002. "Epigenetic reprogramming in mouse primordial germ cells." *Mech Dev* 117 (1-2):15-23.

Hajkova, P., S. J. Jeffries, C. Lee, N. Miller, S. P. Jackson, and M. A. Surani. 2010. "Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway." *Science* 329 (5987):78-82. doi: 10.1126/science.1187945.

Haldar, S. M., O. A. Ibrahim, and M. K. Jain. 2007. "Kruppel-like Factors (KLFs) in muscle biology." *J Mol Cell Cardiol* 43 (1):1-10. doi: 10.1016/j.yjmcc.2007.04.005.

Harrington, M. A., P. A. Jones, M. Imagawa, and M. Karin. 1988. "Cytosine methylation does not affect binding of transcription factor Sp1." *Proc Natl Acad Sci U S A* 85 (7):2066-70.

Harris, A. J., M. J. Duxson, R. B. Fitzsimons, and F. Rieger. 1989. "Myonuclear birthdates distinguish the origins of primary and secondary myotubes in embryonic mammalian skeletal muscles." *Development* 107 (4):771-84.

Hasty, P., A. Bradley, J. H. Morris, D. G. Edmondson, J. M. Venuti, E. N. Olson, and W. H. Klein. 1993. "Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene." *Nature* 364 (6437):501-6. doi: 10.1038/364501a0.

Hattori, N., K. Nishino, Y. G. Ko, J. Ohgane, S. Tanaka, and K. Shiota. 2004. "Epigenetic control of mouse Oct-4 gene expression in embryonic stem cells and trophoblast stem cells." *J Biol Chem* 279 (17):17063-9. doi: 10.1074/jbc.M309002200.

He, Y. F., B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, and G. L. Xu. 2011. "Tet-mediated formation of 5-

carboxylcytosine and its excision by Tdg in mammalian DNA." *Science* 333 (6047):1303-7. doi: 10.1126/science.1210944.

Heard, E., P. Clerc, and P. Avner. 1997. "X-chromosome inactivation in mammals." *Annu Rev Genet* 31:571-610. doi: 10.1146/annurev.genet.31.1.571.

Heintzman, N. D., G. C. Hon, R. D. Hawkins, P. Kheradpour, A. Stark, L. F. Harp, Z. Ye, L. K. Lee, R. K. Stuart, C. W. Ching, K. A. Ching, J. E. Antosiewicz-Bourget, H. Liu, X. Zhang, R. D. Green, V. V. Lobanenkov, R. Stewart, J. A. Thomson, G. E. Crawford, M. Kellis, and B. Ren. 2009. "Histone modifications at human enhancers reflect global cell-type-specific gene expression." *Nature* 459 (7243):108-12. doi: 10.1038/nature07829.

Heintzman, N. D., R. K. Stuart, G. Hon, Y. Fu, C. W. Ching, R. D. Hawkins, L. O. Barrera, S. Van Calcar, C. Qu, K. A. Ching, W. Wang, Z. Weng, R. D. Green, G. E. Crawford, and B. Ren. 2007. "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome." *Nat Genet* 39 (3):311-8. doi: 10.1038/ng1966.

Heidt, A. B., A. Rojas, I. S. Harris, and B. L. Black. 2007. "Determinants of myogenic specificity within MyoD are required for noncanonical E box binding." *Mol Cell Biol* 27 (16):5910-20. doi: 10.1128/MCB.01700-06.

Helin, K., and D. Dhanak. 2013. "Chromatin proteins and modifications as drug targets." *Nature* 502 (7472):480-8. doi: 10.1038/nature12751.

Hellman, A., and A. Chess. 2007. "Gene body-specific methylation on the active X chromosome." *Science* 315 (5815):1141-3. doi: 10.1126/science.1136352.

Hemberger, M., W. Dean, and W. Reik. 2009. "Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal." *Nat Rev Mol Cell Biol* 10 (8):526-37. doi: 10.1038/nrm2727.

Himeda, C. L., X. Chen, and S. D. Hauschka. 2011. "Design and testing of regulatory cassettes for optimal activity in skeletal and cardiac muscles." *Methods Mol Biol* 709:3-19. doi: 10.1007/978-1-61737-982-6_1.

Himeda, C. L., J. A. Ranish, J. C. Angello, P. Maire, R. Aebersold, and S. D. Hauschka. 2004. "Quantitative proteomic identification of six4 as the trex-binding factor in the muscle creatine kinase enhancer." *Mol Cell Biol* 24 (5):2132-43.

Himeda, C. L., J. A. Ranish, R. C. Pearson, M. Crossley, and S. D. Hauschka. 2010. "KLF3 regulates muscle-specific gene expression and synergizes with serum response factor on KLF binding sites." *Mol Cell Biol* 30 (14):3430-43. doi: 10.1128/MCB.00302-10.

Hnisz, D., B. J. Abraham, T. I. Lee, A. Lau, V. Saint-André, A. A. Sigova, H. A. Hoke, and R. A. Young. 2013. "Super-enhancers in the control of cell identity and disease." *Cell* 155 (4):934-47. doi: 10.1016/j.cell.2013.09.053.

Hodges, E., A. D. Smith, J. Kendall, Z. Xuan, K. Ravi, M. Rooks, M. Q. Zhang, K. Ye, A. Bhattacharjee, L. Brizuela, W. R. McCombie, M. Wigler, G. J. Hannon, and J. B. Hicks. 2009. "High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing." *Genome Res* 19 (9):1593-605. doi: 10.1101/gr.095190.109.

Holliday, R., and J. E. Pugh. 1975. "DNA modification mechanisms and gene activity during development." *Science* 187 (4173):226-32.

Holmfeldt, P., J. Pardieck, A. C. Saulsberry, S. K. Nandakumar, D. Finkelstein, J. T. Gray, D. A. Persons, and S. McKinney-Freeman. 2013. "Nfix is a novel regulator of murine hematopoietic stem and progenitor cell survival." *Blood* 122 (17):2987-96. doi: 10.1182/blood-2013-04-493973.

Hon, G. C., N. Rajagopal, Y. Shen, D. F. McCleary, F. Yue, M. D. Dang, and B. Ren. 2013. "Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues." *Nat Genet* 45 (10):1198-206. doi: 10.1038/ng.2746.

Hon, G. C., C. X. Song, T. Du, F. Jin, S. Selvaraj, A. Y. Lee, C. A. Yen, Z. Ye, S. Q. Mao, B. A. Wang, S. Kuan, L. E. Edsall, B. S. Zhao, G. L. Xu, C. He, and B. Ren. 2014. "5mC Oxidation by

Tet2 Modulates Enhancer Activity and Timing of Transcriptome Reprogramming during Differentiation." *Mol Cell* 56 (2):286-97. doi: 10.1016/j.molcel.2014.08.026.

Horst, D., S. Ustanina, C. Sergi, G. Mikuz, H. Juergens, T. Braun, and E. Vorobyov. 2006. "Comparative expression analysis of Pax3 and Pax7 during mouse myogenesis." *Int J Dev Biol* 50 (1):47-54. doi: 10.1387/ijdb.052111dh.

Hsieh, C. L. 1994. "Dependence of transcriptional repression on CpG methylation density." *Mol Cell Biol* 14 (8):5487-94.

Hu, J. S., E. N. Olson, and R. E. Kingston. 1992. "HEB, a helix-loop-helix protein related to E2A and ITF2 that can modulate the DNA-binding ability of myogenic regulatory factors." *Mol Cell Biol* 12 (3):1031-42.

Hu, Y. G., R. Hirasawa, J. L. Hu, K. Hata, C. L. Li, Y. Jin, T. Chen, E. Li, M. Rigolet, E. Viegas-Péquignot, H. Sasaki, and G. L. Xu. 2008. "Regulation of DNA methylation activity through Dnmt3L promoter methylation by Dnmt3 enzymes in embryonic development." *Hum Mol Genet* 17 (17):2654-64. doi: 10.1093/hmg/ddn165.

Huang, R., and B. Christ. 2000. "Origin of the epaxial and hypaxial myotome in avian embryos." *Anat Embryol (Berl)* 202 (5):369-74.

Huang, S., X. Li, T. M. Yusufzai, Y. Qiu, and G. Felsenfeld. 2007. "Usf1 recruits histone modification complexes and is critical for maintenance of a chromatin barrier." *Mol Cell Biol* 27 (22):7991-8002. doi: 10.1128/MCB.01326-07.

Huang, Y., L. Chavez, X. Chang, X. Wang, W. A. Pastor, J. Kang, J. A. Zepeda-Martínez, U. J. Pape, S. E. Jacobsen, B. Peters, and A. Rao. 2014. "Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells." *Proc Natl Acad Sci U S A* 111 (4):1361-6. doi: 10.1073/pnas.1322921111.

Huang, Y., A. K. Das, Q. Y. Yang, M. J. Zhu, and M. Du. 2012. "Zfp423 promotes adipogenic differentiation of bovine stromal vascular cells." *PLoS One* 7 (10):e47496. doi: 10.1371/journal.pone.0047496.

Hupkes, M., M. K. Jonsson, W. J. Scheenen, W. van Rotterdam, A. M. Sotoca, E. P. van Someren, M. A. van der Heyden, T. A. van Veen, R. I. van Ravestein-van Os, S. Bauerschmidt, E. Piek, D. L. Ypey, E. J. van Zoelen, and K. J. Dechering. 2011. "Epigenetics: DNA demethylation promotes skeletal myotube maturation." *FASEB J* 25 (11):3861-72. doi: 10.1096/fj.11-186122.

Hutcheson, D. A., and G. Kardon. 2009. "Genetic manipulations reveal dynamic cell and gene functions: Cre-ating a new view of myogenesis." *Cell Cycle* 8 (22):3675-8.

Iguchi-Ariga, S. M., and W. Schaffner. 1989. "CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation." *Genes Dev* 3 (5):612-9.

Illingworth, R., A. Kerr, D. Desousa, H. Jørgensen, P. Ellis, J. Stalker, D. Jackson, C. Clee, R. Plumb, J. Rogers, S. Humphray, T. Cox, C. Langford, and A. Bird. 2008. "A novel CpG island set identifies tissue-specific methylation at developmental gene loci." *PLoS Biol* 6 (1):e22. doi: 10.1371/journal.pbio.0060022.

Inaba, N., T. Hiruma, A. Togayachi, H. Iwasaki, X. H. Wang, Y. Furukawa, R. Sumi, T. Kudo, K. Fujimura, T. Iwai, M. Gotoh, M. Nakamura, and H. Narimatsu. 2003. "A novel I-branching beta-1,6-N-acetylglucosaminyltransferase involved in human blood group I antigen expression." *Blood* 101 (7):2870-6. doi: 10.1182/blood-2002-09-2838.

Inoue, A., L. Shen, Q. Dai, C. He, and Y. Zhang. 2011. "Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development." *Cell Res* 21 (12):1670-6. doi: 10.1038/cr.2011.189.

Inoue, A., and Y. Zhang. 2011. "Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos." *Science* 334 (6053):194. doi: 10.1126/science.1212483.

Iqbal, K., S. G. Jin, G. P. Pfeifer, and P. E. Szabó. 2011. "Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine." *Proc Natl Acad Sci U S A* 108 (9):3642-7. doi: 10.1073/pnas.1014033108.

Irizarry, R. A., C. Ladd-Acosta, B. Wen, Z. Wu, C. Montano, P. Onyango, H. Cui, K. Gabo, M. Rongione, M. Webster, H. Ji, J. B. Potash, S. Sabunciyan, and A. P. Feinberg. 2009. "The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores." *Nat Genet* 41 (2):178-86. doi: 10.1038/ng.298.

Irizarry, R. A., H. Wu, and A. P. Feinberg. 2009. "A species-generalized probabilistic model-based definition of CpG islands." *Mamm Genome* 20 (9-10):674-80. doi: 10.1007/s00335-009-9222-5.

Isagawa, T., G. Nagae, N. Shiraki, T. Fujita, N. Sato, S. Ishikawa, S. Kume, and H. Aburatani. 2011. "DNA methylation profiling of embryonic stem cell differentiation into the three germ layers." *PLoS One* 6 (10):e26052. doi: 10.1371/journal.pone.0026052.

Ito, S., L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, and Y. Zhang. 2011. "Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine." *Science* 333 (6047):1300-3. doi: 10.1126/science.1210597.

Jabbari, K., and G. Bernardi. 1998. "CpG doublets, CpG islands and Alu repeats in long human DNA sequences from different isochore families." *Gene* 224 (1-2):123-7.

Jackson, M., A. Krassowska, N. Gilbert, T. Chevassut, L. Forrester, J. Ansell, and B. Ramsahoye. 2004. "Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells." *Mol Cell Biol* 24 (20):8862-71. doi: 10.1128/MCB.24.20.8862-8871.2004.

Jaenisch, R., and A. Bird. 2003. "Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals." *Nat Genet* 33 Suppl:245-54. doi: 10.1038/ng1089.

Jaynes, J. B., J. S. Chamberlain, J. N. Buskin, J. E. Johnson, and S. D. Hauschka. 1986. "Transcriptional regulation of the muscle creatine kinase gene and regulated expression in transfected mouse myoblasts." *Mol Cell Biol* 6 (8):2855-64.

Jaynes, J. B., J. E. Johnson, J. N. Buskin, C. L. Gartside, and S. D. Hauschka. 1988. "The muscle creatine kinase gene is regulated by multiple upstream elements, including a muscle-specific enhancer." *Mol Cell Biol* 8 (1):62-70.

Jenuwein, T., and C. D. Allis. 2001. "Translating the histone code." *Science* 293 (5532):1074-80. doi: 10.1126/science.1063127.

Ji, H., L. I. Ehrlich, J. Seita, P. Murakami, A. Doi, P. Lindau, H. Lee, M. J. Aryee, R. A. Irizarry, K. Kim, D. J. Rossi, M. A. Inlay, T. Serwold, H. Karsunky, L. Ho, G. Q. Daley, I. L. Weissman, and A. P. Feinberg. 2010. "Comprehensive methylome map of lineage commitment from haematopoietic progenitors." *Nature* 467 (7313):338-42. doi: 10.1038/nature09367.

Jia, D., R. Z. Jurkowska, X. Zhang, A. Jeltsch, and X. Cheng. 2007. "Structure of Dnmt3a bound to Dnmt3L suggests a model for *de novo* DNA methylation." *Nature* 449 (7159):248-51. doi: 10.1038/nature06146.

Jin, C., Y. Lu, J. Jelinek, S. Liang, M. R. Estecio, M. C. Barton, and J. P. Issa. 2014. "Tet1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells." *Nucleic Acids Res* 42 (11):6956-71. doi: 10.1093/nar/gku372.

Jolma, A., J. Yan, T. Whitington, J. Toivonen, K. R. Nitta, P. Rastas, E. Morgunova, M. Enge, M. Taipale, G. Wei, K. Palin, J. M. Vaquerizas, R. Vincentelli, N. M. Luscombe, T. R. Hughes, P. Lemaire, E. Ukkonen, T. Kivioja, and J. Taipale. 2013. "DNA-binding specificities of human transcription factors." *Cell* 152 (1-2):327-39. doi: 10.1016/j.cell.2012.12.009.

Jones, P. 2011. "Out of Africa and into epigenetics: discovering reprogramming drugs." *Nat Cell Biol* 13 (1):2. doi: 10.1038/ncb0111-2.

Jones, P. A., and S. B. Baylin. 2002. "The fundamental role of epigenetic events in cancer." *Nat Rev Genet* 3 (6):415-28. doi: 10.1038/nrg816.

Jones, P. A., and D. Takai. 2001. "The role of DNA methylation in mammalian epigenetics." *Science* 293 (5532):1068-70. doi: 10.1126/science.1063852.

Jones, P. A., S. M. Taylor, and V. L. Wilson. 1983. "Inhibition of DNA methylation by 5-Azacytidine." *Recent Results Cancer Res* 84:202-11.

Jones, P. A., M. J. Wolkowicz, M. A. Harrington, and F. Gonzales. 1990. "Methylation and expression of the Myo D1 determination gene." *Philos Trans R Soc Lond B Biol Sci* 326 (1235):277-84.

Jones, P. A., M. J. Wolkowicz, W. M. Rideout, F. A. Gonzales, C. M. Marziasz, G. A. Coetzee, and S. J. Tapscott. 1990. "De novo methylation of the MyoD1 CpG island during the establishment of immortal cell lines." *Proc Natl Acad Sci U S A* 87 (16):6117-21.

Jordà, M., and M. A. Peinado. 2010. "Methods for DNA methylation analysis and applications in colon cancer." *Mutat Res* 693 (1-2):84-93. doi: 10.1016/j.mrfmmm.2010.06.010.

Jordà, M., J. Rodríguez, J. Frigola, and M. A. Peinado. 2009. "Analysis of DNA methylation by amplification of intermethylated sites (AIMS)." *Methods Mol Biol* 507:107-16. doi: 10.1007/978-1-59745-522-0_9.

Jost, J. P., E. J. Oakeley, B. Zhu, D. Benjamin, S. Thiry, M. Siegmann, and Y. C. Jost. 2001. "5-Methylcytosine DNA glycosylase participates in the genome-wide loss of DNA methylation occurring during mouse myoblast differentiation." *Nucleic Acids Res* 29 (21):4452-61.

Jozefczuk, J., K. Drews, and J. Adjaye. 2012. "Preparation of mouse embryonic fibroblast cells suitable for culturing human embryonic and induced pluripotent stem cells." *J Vis Exp* (64). doi: 10.3791/3854.

Kadam, S., and B. M. Emerson. 2003. "Transcriptional specificity of human SWI/SNF BRG1 and BRM chromatin remodeling complexes." *Mol Cell* 11 (2):377-89.

Kassar-Duchossoy, L., B. Gayraud-Morel, D. Gomès, D. Rocancourt, M. Buckingham, V. Shinin, and S. Tajbakhsh. 2004. "Mrf4 determines skeletal muscle identity in Myf5:Myod double-mutant mice." *Nature* 431 (7007):466-71. doi: 10.1038/nature02876.

Kato, Y., M. Kaneda, K. Hata, K. Kumaki, M. Hisano, Y. Kohara, M. Okano, E. Li, M. Nozaki, and H. Sasaki. 2007. "Role of the Dnmt3 family in *de novo* methylation of imprinted and repetitive sequences during male germ cell development in the mouse." *Hum Mol Genet* 16 (19):2272-80. doi: 10.1093/hmg/ddm179.

Kawakami, K., A. Ruszkiewicz, G. Bennett, J. Moore, F. Grieu, G. Watanabe, and B. Iacopetta. 2006. "DNA hypermethylation in the normal colonic mucosa of patients with colorectal cancer." *Br J Cancer* 94 (4):593-8. doi: 10.1038/sj.bjc.6602940.

Keller, C., and M. R. Capecchi. 2005. "New genetic tactics to model alveolar rhabdomyosarcoma in the mouse." *Cancer Res* 65 (17):7530-2. doi: 10.1158/0008-5472.CAN-05-0477.

Keller, C., and D. C. Guttridge. 2013. "Mechanisms of impaired differentiation in rhabdomyosarcoma." *FEBS J* 280 (17):4323-34. doi: 10.1111/febs.12421.

Kelly, A. M., and S. I. Zacks. 1969. "The histogenesis of rat intercostal muscle." *J Cell Biol* 42 (1):135-53.

Kelly, C. E., E. Thymiakou, J. E. Dixon, S. Tanaka, J. Godwin, and V. Episkopou. 2013. "Rnf165/Ark2C enhances BMP-Smad signaling to mediate motor axon extension." *PLoS Biol* 11 (4):e1001538. doi: 10.1371/journal.pbio.1001538.

Kelly, T. K., D. D. De Carvalho, and P. A. Jones. 2010. "Epigenetic modifications as therapeutic targets." *Nat Biotechnol* 28 (10):1069-1078. doi: 10.1038/nbt.1678.

Kennedy, M., M. Firpo, K. Choi, C. Wall, S. Robertson, N. Kabrun, and G. Keller. 1997. "A common precursor for primitive erythropoiesis and definitive haematopoiesis." *Nature* 386 (6624):488-93. doi: 10.1038/386488a0.

Kigami, D., N. Minami, H. Takayama, and H. Imai. 2003. "MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos." *Biol Reprod* 68 (2):651-4.

Kim, J., A. Kollhoff, A. Bergmann, and L. Stubbs. 2003. "Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, Peg3." *Hum Mol Genet* 12 (3):233-45.

Kim, M., T. W. Kang, H. C. Lee, Y. M. Han, H. Kim, H. D. Shin, H. S. Cheong, D. Lee, S. Y. Kim, and Y. S. Kim. 2011. "Identification of DNA methylation markers for lineage commitment of *in vitro* hepatogenesis." *Hum Mol Genet* 20 (14):2722-33. doi: 10.1093/hmg/ddr171.

Kim, M. S., T. Kondo, I. Takada, M. Y. Youn, Y. Yamamoto, S. Takahashi, T. Matsumoto, S. Fujiyama, Y. Shirode, I. Yamaoka, H. Kitagawa, K. Takeyama, H. Shibuya, F. Ohtake, and S. Kato. 2009. "DNA demethylation in hormone-induced transcriptional derepression." *Nature* 461 (7266):1007-12. doi: 10.1038/nature08456.

Klug, M. G., M. H. Soonpaa, G. Y. Koh, and L. J. Field. 1996. "Genetically selected cardiomyocytes from differentiating embronic stem cells form stable intracardiac grafts." *J Clin Invest* 98 (1):216-24. doi: 10.1172/JCI118769.

Ko, M., J. An, H. S. Bandukwala, L. Chavez, T. Aijö, W. A. Pastor, M. F. Segal, H. Li, K. P. Koh, H. Lähdesmäki, P. G. Hogan, L. Aravind, and A. Rao. 2013. "Modulation of Tet2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX." *Nature* 497 (7447):122-6. doi: 10.1038/nature12052.

Kouzarides, T. 2007. "Chromatin modifications and their function." *Cell* 128 (4):693-705. doi: 10.1016/j.cell.2007.02.005.

Krainc, D., G. Bai, S. Okamoto, M. Carles, J. W. Kusiak, R. N. Brent, and S. A. Lipton. 1998. "Synergistic activation of the N-methyl-D-aspartate receptor subunit 1 promoter by myocyte enhancer factor 2C and Sp1." *J Biol Chem* 273 (40):26218-24.

Kriaucionis, S., and N. Heintz. 2009. "The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain." *Science* 324 (5929):929-30. doi: 10.1126/science.1169786.

Ku, M., R. P. Koche, E. Rheinbay, E. M. Mendenhall, M. Endoh, T. S. Mikkelsen, A. Presser, C. Nusbaum, X. Xie, A. S. Chi, M. Adli, S. Kasif, L. M. Ptaszek, C. A. Cowan, E. S. Lander, H. Koseki, and B. E. Bernstein. 2008. "Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains." *PLoS Genet* 4 (10):e1000242. doi: 10.1371/journal.pgen.1000242.

Kulis, M., S. Heath, M. Bibikova, A. C. Queirós, A. Navarro, G. Clot, A. Martínez-Trillos, G. Castellano, I. Brun-Heath, M. Pinyol, S. Barberán-Soler, P. Papasaikas, P. Jares, S. Beà, D. Rico, S. Ecker, M. Rubio, R. Royo, V. Ho, B. Klotzle, L. Hernández, L. Conde, M. López-Guerra, D. Colomer, N. Villamor, M. Aymerich, M. Rozman, M. Bayes, M. Gut, J. L. Gelpí, M. Orozco, J. B. Fan, V. Quesada, X. S. Puente, D. G. Pisano, A. Valencia, A. López-Guillermo, I. Gut, C. López-Otín, E. Campo, and J. I. Martín-Subero. 2012. "Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia." *Nat Genet* 44 (11):1236-42. doi: 10.1038/ng.2443.

Kulis, M., A. C. Queirós, R. Beekman, and J. I. Martín-Subero. 2013. "Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer." *Biochim Biophys Acta* 1829 (11):1161-74. doi: 10.1016/j.bbagrm.2013.08.001.

Kurmasheva, R. T., C. A. Peterson, D. M. Parham, B. Chen, R. E. McDonald, and C. A. Cooney. 2005. "Upstream CpG island methylation of the Pax3 gene in human rhabdomyosarcomas." *Pediatr Blood Cancer* 44 (4):328-37. doi: 10.1002/pbc.20285.

Lagha, M., T. Sato, L. Bajard, P. Daubas, M. Esner, D. Montarras, F. Relaix, and M. Buckingham. 2008. "Regulation of skeletal muscle stem cell behavior by Pax3 and Pax7." *Cold Spring Harb Symp Quant Biol* 73:307-15. doi: 10.1101/sqb.2008.73.006.

Laird, P. W. 2010. "Principles and challenges of genomewide DNA methylation analysis." *Nat Rev Genet* 11 (3):191-203. doi: 10.1038/nrg2732.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, J. Szustakowki, and International Human Genome Sequencing Consortium. 2001. "Initial sequencing and analysis of the human genome." *Nature* 409 (6822):860-921. doi: 10.1038/35057062.

Lane, N., W. Dean, S. Erhardt, P. Hajkova, A. Surani, J. Walter, and W. Reik. 2003. "Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse." *Genesis* 35 (2):88-93. doi: 10.1002/gene.10168.

Lassar, A. B., B. M. Paterson, and H. Weintraub. 1986. "Transfection of a DNA locus that mediates the conversion of 10T1/2 fibroblasts to myoblasts." *Cell* 47 (5):649-56.

Laurent, L., E. Wong, G. Li, T. Huynh, A. Tsirigos, C. T. Ong, H. M. Low, K. W. Kin Sung, I. Rigoutsos, J. Loring, and C. L. Wei. 2010. "Dynamic changes in the human methylome during differentiation." *Genome Res* 20 (3):320-31. doi: 10.1101/gr.101907.109.

Lechner, M., M. Marz, C. Ihling, A. Sinz, P. F. Stadler, and V. Krauss. 2013. "The correlation of genome size and DNA methylation rate in metazoans." *Theory Biosci* 132 (1):47-60. doi: 10.1007/s12064-012-0167-y.

Lee, H. J., T. A. Hore, and W. Reik. 2014. "Reprogramming the methylome: erasing memory and creating diversity." *Cell Stem Cell* 14 (6):710-9. doi: 10.1016/j.stem.2014.05.008.

Lees-Murdock, D. J., M. De Felici, and C. P. Walsh. 2003. "Methylation dynamics of repetitive DNA elements in the mouse germ cell lineage." *Genomics* 82 (2):230-7.

Lehnertz, B., Y. Ueda, A. A. Derijck, U. Braunschweig, L. Perez-Burgos, S. Kubicek, T. Chen, E. Li, T. Jenuwein, and A. H. Peters. 2003. "Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin." *Curr Biol* 13 (14):1192-200.

Levine, S. S., A. Weiss, H. Erdjument-Bromage, Z. Shao, P. Tempst, and R. E. Kingston. 2002. "The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans." *Mol Cell Biol* 22 (17):6070-8.

Li, E., T. H. Bestor, and R. Jaenisch. 1992. "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality." *Cell* 69 (6):915-26.

Li, H. P., H. Y. Huang, Y. R. Lai, J. X. Huang, K. P. Chang, C. Hsueh, and Y. S. Chang. 2014. "Silencing of miRNA-148a by hypermethylation activates the integrin-mediated signaling pathway in nasopharyngeal carcinoma." *Oncotarget* 5 (17):7610-24.

Li, T., D. Yang, J. Li, Y. Tang, J. Yang, and W. Le. 2014. "Critical Role of Tet3 in Neural Progenitor Cell Maintenance and Terminal Differentiation." *Mol Neurobiol*. doi: 10.1007/s12035-014-8734-5.

Liao, W., S. H. Hong, B. H. Chan, F. B. Rudolph, S. C. Clark, and L. Chan. 1999. "Apobec-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family." *Biochem Biophys Res Commun* 260 (2):398-404. doi: 10.1006/bbrc.1999.0925.

Lienert, F., C. Wirbelauer, I. Som, A. Dean, F. Mohn, and D. Schübeler. 2011. "Identification of genetic elements that autonomously determine DNA methylation states." *Nat Genet* 43 (11):1091-7. doi: 10.1038/ng.946.

Lindahl Allen, M., C. M. Koch, G. K. Clelland, I. Dunham, and M. Antoniou. 2009. "DNA methylation-histone modification relationships across the desmin locus in human primary cells." *BMC Mol Biol* 10:51. doi: 10.1186/1471-2199-10-51.

Lipton, B. H., and E. Schultz. 1979. "Developmental fate of skeletal muscle satellite cells." *Science* 205 (4412):1292-4.

Lister, R., M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, and J. R. Ecker. 2009. "Human DNA methylomes at base resolution show widespread epigenomic differences." *Nature* 462 (7271):315-22. doi: 10.1038/nature08514.

Lluís, F., E. Ballestar, M. Suelves, M. Esteller, and P. Muñoz-Cánoves. 2005. "E47 phosphorylation by p38 MAPK promotes MyoD/E47 association and muscle-specific gene transcription." *EMBO J* 24 (5):974-84. doi: 10.1038/sj.emboj.7600528.

Lluís, F., E. Perdiguero, A. R. Nebreda, and P. Muñoz-Cánoves. 2006. "Regulation of skeletal muscle gene expression by p38 MAP kinases." *Trends Cell Biol* 16 (1):36-44. doi: 10.1016/j.tcb.2005.11.002.

Loebel, D. A., C. M. Watson, R. A. De Young, and P. P. Tam. 2003. "Lineage choice and differentiation in mouse embryos and embryonic stem cells." *Dev Biol* 264 (1):1-14.

Loh, Y. H., Q. Wu, J. L. Chew, V. B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K. Y. Wong, K. W. Sung, C. W. Lee, X. D. Zhao, K. P. Chiu, L. Lipovich, V. A. Kuznetsov, P. Robson, L. W. Stanton, C. L. Wei, Y. Ruan, B. Lim, and H. H. Ng. 2006. "The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells." *Nat Genet* 38 (4):431-40. doi: 10.1038/ng1760.

Lucarelli, M., A. Fuso, R. Strom, and S. Scarpa. 2001. "The dynamics of myogenin site-specific demethylation is strongly correlated with its expression and with muscle differentiation." *J Biol Chem* 276 (10):7500-6. doi: 10.1074/jbc.M008234200.

Lujambio, A., A. Portela, J. Liz, S. A. Melo, S. Rossi, R. Spizzo, C. M. Croce, G. A. Calin, and M. Esteller. 2010. "CpG island hypermethylation-associated silencing of non-coding RNAs transcribed from ultraconserved regions in human cancer." *Oncogene* 29 (48):6390-401. doi: 10.1038/onc.2010.361.

Lyle, R., D. Watanabe, D. te Vruchte, W. Lerchner, O. W. Smrzka, A. Wutz, J. Schageman, L. Hahner, C. Davies, and D. P. Barlow. 2000. "The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1." *Nat Genet* 25 (1):19-21. doi: 10.1038/75546.

Lynch, M. D., A. J. Smith, M. De Gobbi, M. Flenley, J. R. Hughes, D. Vernimmen, H. Ayyub, J. A. Sharpe, J. A. Sloane-Stanley, L. Sutherland, S. Meek, T. Burdon, R. J. Gibbons, D. Garrick, and D. R. Higgs. 2012. "An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment." *EMBO J* 31 (2):317-29. doi: 10.1038/emboj.2011.399.

Ma, D. K., M. H. Jang, J. U. Guo, Y. Kitabatake, M. L. Chang, N. Pow-Anpongkul, R. A. Flavell, B. Lu, G. L. Ming, and H. Song. 2009. "Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis." *Science* 323 (5917):1074-7. doi: 10.1126/science.1166859.

Macleod, D., J. Charlton, J. Mullins, and A. P. Bird. 1994. "Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island." *Genes Dev* 8 (19):2282-92.

MacLeod, M. C. 1993. "Identification of a DNA structural motif that includes the binding sites for Sp1, p53 and GA-binding protein." *Nucleic Acids Res* 21 (6):1439-47.

Magistri, M., M. A. Faghihi, G. St Laurent, and C. Wahlestedt. 2012. "Regulation of chromatin structure by long noncoding RNAs: focus on natural antisense transcripts." *Trends Genet* 28 (8):389-96. doi: 10.1016/j.tig.2012.03.013.

Magli, A., E. Schnettler, S. A. Swanson, L. Borges, K. Hoffman, R. Stewart, J. A. Thomson, S. A. Keirstead, and R. C. Perlingeiro. 2014. "Pax3 and Tbx5 specify whether PdgfRα+ cells assume skeletal or cardiac muscle fate in differentiating embryonic stem cells." *Stem Cells* 32 (8):2072-83. doi: 10.1002/stem.1713.

Maiti, A., and A. C. Drohat. 2011. "Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites." *J Biol Chem* 286 (41):35334-8. doi: 10.1074/jbc.C111.284620.

Mallona, I., A. Díez-Villanueva, and M. A. Peinado. 2014. "Methylation plotter: a web tool for dynamic visualization of DNA methylation data." *Source Code Biol Med* 9:11. doi: 10.1186/1751-0473-9-11.

Malousi, A., N. Maglaveras, and S. Kouidou. 2008. "Intronic CpG content and alternative splicing in human genes containing a single cassette exon." *Epigenetics* 3 (2):69-73.

Margueron, R., G. Li, K. Sarma, A. Blais, J. Zavadil, C. L. Woodcock, B. D. Dynlacht, and D. Reinberg. 2008. "Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms." *Mol Cell* 32 (4):503-18. doi: 10.1016/j.molcel.2008.11.004.

Marin, M., A. Karis, P. Visser, F. Grosveld, and S. Philipsen. 1997. "Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation." *Cell* 89 (4):619-28.

Mathelier, A., X. Zhao, A. W. Zhang, F. Parcy, R. Worsley-Hunt, D. J. Arenillas, S. Buchman, C. Y. Chen, A. Chou, H. Ienasescu, J. Lim, C. Shyr, G. Tan, M. Zhou, B. Lenhard, A. Sandelin, and W. W. Wasserman. 2014. "JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles." *Nucleic Acids Res* 42 (Database issue):D142-7. doi: 10.1093/nar/gkt997.

Maunakea, A. K., R. P. Nagarajan, M. Bilenky, T. J. Ballinger, C. D'Souza, S. D. Fouse, B. E. Johnson, C. Hong, C. Nielsen, Y. Zhao, G. Turecki, A. Delaney, R. Varhol, N. Thiessen, K. Shchors, V. M. Heine, D. H. Rowitch, X. Xing, C. Fiore, M. Schillebeeckx, S. J. Jones, D.

Haussler, M. A. Marra, M. Hirst, T. Wang, and J. F. Costello. 2010. "Conserved role of intragenic DNA methylation in regulating alternative promoters." *Nature* 466 (7303):253-7. doi: 10.1038/nature09165.

Mauro, A. 1961. "Satellite cell of skeletal muscle fibers." *J Biophys Biochem Cytol* 9:493-5.

McLoon, L. K. 2009. "A new role for satellite cells: control of reinnervation after muscle injury by semaphorin 3A. Focus on "Possible implication of satellite cells in regenerative motoneuritogenesis: HGF upregulates neural chemorepellent Sema3A during myogenic differentiation"." *Am J Physiol Cell Physiol* 297 (2):C227-30. doi: 10.1152/ajpcell.00256.2009.

Medvedeva, Y. A., A. M. Khamis, I. V. Kulakovskiy, W. Ba-Alawi, M. S. Bhuyan, H. Kawaji, T. Lassmann, M. Harbers, A. R. Forrest, V. B. Bajic, and FANTOM consortium. 2014. "Effects of cytosine methylation on transcription factor binding sites." *BMC Genomics* 15:119. doi: 10.1186/1471-2164-15-119.

Meissner, A., A. Gnirke, G. W. Bell, B. Ramsahoye, E. S. Lander, and R. Jaenisch. 2005. "Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis." *Nucleic Acids Res* 33 (18):5868-77. doi: 10.1093/nar/gki901.

Meissner, A., T. S. Mikkelsen, H. Gu, M. Wernig, J. Hanna, A. Sivachenko, X. Zhang, B. E. Bernstein, C. Nusbaum, D. B. Jaffe, A. Gnirke, R. Jaenisch, and E. S. Lander. 2008. "Genome-scale DNA methylation maps of pluripotent and differentiated cells." *Nature* 454 (7205):766-70. doi: 10.1038/nature07107.

Mendell, J. R., J. T. Kissel, A. A. Amato, W. King, L. Signore, T. W. Prior, Z. Sahenk, S. Benson, P. E. McAndrew, and R. Rice. 1995. "Myoblast transfer in the treatment of Duchenne's muscular dystrophy." *N Engl J Med* 333 (13):832-8. doi: 10.1056/NEJM199509283331303.

Messina, G., S. Biressi, S. Monteverde, A. Magli, M. Cassano, L. Perani, E. Roncaglia, E. Tagliafico, L. Starnes, C. E. Campbell, M. Grossi, D. J. Goldhamer, R. M. Gronostajski, and G. Cossu. 2010. "Nfix regulates fetal-specific transcription in developing skeletal muscle." *Cell* 140 (4):554-66. doi: 10.1016/j.cell.2010.01.027.

Mikkelsen, T. S., M. Ku, D. B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T. K. Kim, R. P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E. S. Lander, and B. E. Bernstein. 2007. "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells." *Nature* 448 (7153):553-60. doi: 10.1038/nature06008.

Mikl, M. C., I. N. Watt, M. Lu, W. Reik, S. L. Davies, M. S. Neuberger, and C. Rada. 2005. "Mice deficient in Apobec2 and Apobec3." *Mol Cell Biol* 25 (16):7270-7. doi: 10.1128/MCB.25.16.7270-7277.2005.

Mills, A. A. 2010. "Throwing the cancer switch: reciprocal roles of polycomb and trithorax proteins." *Nat Rev Cancer* 10 (10):669-82. doi: 10.1038/nrc2931.

Miner, J. H., and B. Wold. 1990. "Herculin, a fourth member of the MyoD family of myogenic regulatory genes." *Proc Natl Acad Sci U S A* 87 (3):1089-93.

Miyata, K., T. Miyata, K. Nakabayashi, K. Okamura, M. Naito, T. Kawai, S. Takada, K. Kato, S. Miyamoto, K. Hata, and H. Asahara. 2014. "DNA methylation analysis of human myoblasts during *in vitro* myogenic differentiation: *de novo* methylation of promoters of muscle-related genes and its involvement in transcriptional down-regulation." *Hum Mol Genet*. doi: 10.1093/hmg/ddu457.

Mohn, F., and D. Schübeler. 2009. "Genetics and epigenetics: stability and plasticity during cellular differentiation." *Trends Genet* 25 (3):129-36. doi: 10.1016/j.tig.2008.12.005.

Mohn, F., M. Weber, M. Rebhan, T. C. Roloff, J. Richter, M. B. Stadler, M. Bibel, and D. Schübeler. 2008. "Lineage-specific polycomb targets and *de novo* DNA methylation define restriction and potential of neuronal progenitors." *Mol Cell* 30 (6):755-66. doi: 10.1016/j.molcel.2008.05.007.

Mohn, F., M. Weber, D. Schübeler, and T. C. Roloff. 2009. "Methylated DNA immunoprecipitation (MeDIP)." *Methods Mol Biol* 507:55-64. doi: 10.1007/978-1-59745-522-0_5.

Molkentin, J. D., and E. N. Olson. 1996. "Combinatorial control of muscle development by basic helix-loop-helix and MADS-box transcription factors." *Proc Natl Acad Sci U S A* 93 (18):9366-73.

Moncaut, N., J. W. Cross, C. Siligan, A. Keith, K. Taylor, P. W. Rigby, and J. J. Carvajal. 2012. "Musculin and TCF21 coordinate the maintenance of myogenic regulatory factor expression levels during mouse craniofacial development." *Development* 139 (5):958-67. doi: 10.1242/dev.068015.

Moncaut, N., P. W. Rigby, and J. J. Carvajal. 2013. "Dial M(RF) for myogenesis." *FEBS J* 280 (17):3980-90. doi: 10.1111/febs.12379.

Montarras, D., A. L'honoré, and M. Buckingham. 2013. "Lying low but ready for action: the quiescent muscle satellite cell." *FEBS J* 280 (17):4036-50. doi: 10.1111/febs.12372.

Montesano, A., L. Luzi, P. Senesi, and I. Terruzzi. 2013. "Modulation of cell cycle progression by 5-Azacytidine is associated with early myogenesis induction in murine myoblasts." *Int J Biol Sci* 9 (4):391-402. doi: 10.7150/ijbs.4729.

Morgan, H. D., W. Dean, H. A. Coker, W. Reik, and S. K. Petersen-Mahrt. 2004. "Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming." *J Biol Chem* 279 (50):52353-60. doi: 10.1074/jbc.M407695200.

Mueller, P. R., and B. Wold. 1989. "In vivo footprinting of a muscle specific enhancer by ligation mediated PCR." *Science* 246 (4931):780-6.

Mummaneni, P., P. Yates, J. Simpson, J. Rose, and M. S. Turker. 1998. "The primary function of a redundant Sp1 binding site in the mouse aprt gene promoter is to block epigenetic gene inactivation." *Nucleic Acids Res* 26 (22):5163-9.

Murad, H., F. Kasies, R. Azroony, G. Alya, and A. Madania. 2011. "Effects of fenofibrate on Semaphorin 6B gene expression in rat skeletal muscle." *Mol Med Rep* 4 (3):575-80. doi: 10.3892/mmr.2011.463.

Murre, C., P. S. McCaw, H. Vaessin, M. Caudy, L. Y. Jan, Y. N. Jan, C. V. Cabrera, J. N. Buskin, S. D. Hauschka, and A. B. Lassar. 1989. "Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence." *Cell* 58 (3):537-44.

Métivier, R., R. Gallais, C. Tiffoche, C. Le Péron, R. Z. Jurkowska, R. P. Carmouche, D. Ibberson, P. Barath, F. Demay, G. Reid, V. Benes, A. Jeltsch, F. Gannon, and G. Salbert. 2008. "Cyclical DNA methylation of a transcriptionally active promoter." *Nature* 452 (7183):45-50. doi: 10.1038/nature06544.

Müller, M., B. K. Fleischmann, S. Selbert, G. J. Ji, E. Endl, G. Middeler, O. J. Müller, P. Schlenke, S. Frese, A. M. Wobus, J. Hescheler, H. A. Katus, and W. M. Franz. 2000. "Selection of ventricular-like cardiomyocytes from ES cells *in vitro*." *FASEB J* 14 (15):2540-8. doi: 10.1096/fj.00-0002com.

Nabeshima, Y., K. Hanaoka, M. Hayasaka, E. Esumi, S. Li, and I. Nonaka. 1993. "Myogenin gene disruption results in perinatal lethality because of severe muscle defect." *Nature* 364 (6437):532-5. doi: 10.1038/364532a0.

Nagae, G., T. Isagawa, N. Shiraki, T. Fujita, S. Yamamoto, S. Tsutsumi, A. Nonaka, S. Yoshiba, K. Matsusaka, Y. Midorikawa, S. Ishikawa, H. Soejima, M. Fukayama, H. Suemori, N. Nakatsuji, S. Kume, and H. Aburatani. 2011. "Tissue-specific demethylation in CpG-poor promoters during cellular differentiation." *Hum Mol Genet* 20 (14):2710-21. doi: 10.1093/hmg/ddr170.

Narlikar, G. J., H. Y. Fan, and R. E. Kingston. 2002. "Cooperation between complexes that regulate chromatin structure and transcription." *Cell* 108 (4):475-87.

Natoli, G. 2010. "Maintaining cell identity through global control of genomic organization." *Immunity* 33 (1):12-24. doi: 10.1016/j.immuni.2010.07.006.

Nazor, K. L., G. Altun, C. Lynch, H. Tran, J. V. Harness, I. Slavin, I. Garitaonandia, F. J. Müller, Y. C. Wang, F. S. Boscolo, E. Fakunle, B. Dumevska, S. Lee, H. S. Park, T. Olee, D. D. D'Lima, R. Semechkin, M. M. Parast, V. Galat, A. L. Laslett, U. Schmidt, H. S. Keirstead, J. F. Loring, and L. C. Laurent. 2012. "Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives." *Cell Stem Cell* 10 (5):620-34. doi: 10.1016/j.stem.2012.02.013.

Nebl, G., and A. C. Cato. 1995. "NFI/X proteins: a class of NFI family of transcription factors with positive and negative regulatory domains." *Cell Mol Biol Res* 41 (2):85-95.

Newburger, D. E., and M. L. Bulyk. 2009. "UniPROBE: an online database of protein binding microarray data on protein-DNA interactions." *Nucleic Acids Res* 37 (Database issue):D77-82. doi: 10.1093/nar/gkn660.

Nguyen, Q. G., J. N. Buskin, C. L. Himeda, C. Fabre-Suver, and S. D. Hauschka. 2003. "Transgenic and tissue culture analyses of the muscle creatine kinase enhancer Trex control element in skeletal and cardiac muscle indicate differences in gene expression between muscle types." *Transgenic Res* 12 (3):337-49.

Nguyen, Q. G., J. N. Buskin, C. L. Himeda, M. A. Shield, and S. D. Hauschka. 2003. "Differences in the function of three conserved E-boxes of the muscle creatine kinase gene in cultured myocytes and in transgenic mouse skeletal and cardiac muscle." *J Biol Chem* 278 (47):46494-505. doi: 10.1074/jbc.M308194200.

Nitert, M. D., T. Dayeh, P. Volkov, T. Elgzyri, E. Hall, E. Nilsson, B. T. Yang, S. Lang, H. Parikh, Y. Wessman, H. Weishaupt, J. Attema, M. Abels, N. Wierup, P. Almgren, P. A. Jansson, T. Rönn, O. Hansson, K. F. Eriksson, L. Groop, and C. Ling. 2012. "Impact of an exercise intervention on DNA methylation in skeletal muscle from first-degree relatives of patients with type 2 diabetes." *Diabetes* 61 (12):3322-32. doi: 10.2337/db11-1653.

O'Connor, T. P., and R. G. Crystal. 2006. "Genetic medicines: treatment strategies for hereditary disorders." *Nat Rev Genet* 7 (4):261-76. doi: 10.1038/nrg1829.

Oberdoerffer, P., and D. A. Sinclair. 2007. "The role of nuclear architecture in genomic instability and ageing." *Nat Rev Mol Cell Biol* 8 (9):692-702. doi: 10.1038/nrm2238.

Odom, D. T., N. Zizlsperger, D. B. Gordon, G. W. Bell, N. J. Rinaldi, H. L. Murray, T. L. Volkert, J. Schreiber, P. A. Rolfe, D. K. Gifford, E. Fraenkel, G. I. Bell, and R. A. Young. 2004. "Control of pancreas and liver gene expression by HNF transcription factors." *Science* 303 (5662):1378-81. doi: 10.1126/science.1089769.

Oikawa, Y., R. Omori, T. Nishii, Y. Ishida, M. Kawaichi, and E. Matsuda. 2011. "The methyl-CpG-binding protein CIBZ suppresses myogenic differentiation by directly inhibiting myogenin expression." *Cell Res* 21 (11):1578-90. doi: 10.1038/cr.2011.90.

Okano, M., D. W. Bell, D. A. Haber, and E. Li. 1999. "DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development." *Cell* 99 (3):247-57.

Okano, M., and E. Li. 2002. "Genetic analyses of DNA methyltransferase genes in mouse model system." *J Nutr* 132 (8 Suppl):2462S-2465S.

OkAzawa, H., K. Okamoto, F. Ishino, T. Ishino-Kaneko, S. Takeda, Y. Toyoda, M. Muramatsu, and H. Hamada. 1991. "The oct3 gene, a gene for an embryonic transcription factor, is controlled by a retinoic acid repressible enhancer." *EMBO J* 10 (10):2997-3005.

Ooi, S. K., C. Qiu, E. Bernstein, K. Li, D. Jia, Z. Yang, H. Erdjument-Bromage, P. Tempst, S. P. Lin, C. D. Allis, X. Cheng, and T. H. Bestor. 2007. "Dnmt3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA." *Nature* 448 (7154):714-7. doi: 10.1038/nature05987.

Ordahl, C. P., and N. M. Le Douarin. 1992. "Two myogenic lineages within the developing somite." *Development* 114 (2):339-53.

Oswald, J., S. Engemann, N. Lane, W. Mayer, A. Olek, R. Fundele, W. Dean, W. Reik, and J. Walter. 2000. "Active demethylation of the paternal genome in the mouse zygote." *Curr Biol* 10 (8):475-8.

Otani, J., T. Nankumo, K. Arita, S. Inamoto, M. Ariyoshi, and M. Shirakawa. 2009. "Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-Dnmt3-Dnmt3L domain." *EMBO Rep* 10 (11):1235-41. doi: 10.1038/embor.2009.218.

Oustanina, S., G. Hause, and T. Braun. 2004. "Pax7 directs postnatal renewal and propagation of myogenic satellite cells but not their specification." *EMBO J* 23 (16):3430-9. doi: 10.1038/sj.emboj.7600346.

Palacios, D., D. Summerbell, P. W. Rigby, and J. Boyes. 2010. "Interplay between DNA methylation and transcription factor availability: implications for developmental activation of the mouse Myogenin gene." *Mol Cell Biol* 30 (15):3805-15. doi: 10.1128/MCB.00050-10.

Park, S. Y., H. M. Shin, and T. H. Han. 2002. "Synergistic interaction of MEF2D and Sp1 in activation of the CD14 promoter." *Mol Immunol* 39 (1-2):25-30.

Partridge, T., Q. L. Lu, G. Morris, and E. Hoffman. 1998. "Is myoblast transplantation effective?" *Nat Med* 4 (11):1208-9. doi: 10.1038/3167.

Passarge, E. 1979. "Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago." *Am J Hum Genet* 31 (2):106-15.

Pastor, W. A., U. J. Pape, Y. Huang, H. R. Henderson, R. Lister, M. Ko, E. M. McLoughlin, Y. Brudno, S. Mahapatra, P. Kapranov, M. Tahiliani, G. Q. Daley, X. S. Liu, J. R. Ecker, P. M. Milos, S. Agarwal, and A. Rao. 2011. "Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells." *Nature* 473 (7347):394-7. doi: 10.1038/nature10102.

Peaston, A. E., A. V. Evsikov, J. H. Graber, W. N. de Vries, A. E. Holbrook, D. Solter, and B. B. Knowles. 2004. "Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos." *Dev Cell* 7 (4):597-606. doi: 10.1016/j.devcel.2004.09.004.

Penn, N. W., R. Suwalski, C. O'Riley, K. Bojanowski, and R. Yura. 1972. "The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid." *Biochem J* 126 (4):781-90.

Perini, G., D. Diolaiti, A. Porro, and G. Della Valle. 2005. "In vivo transcriptional regulation of N-Myc target genes is controlled by E-box methylation." *Proc Natl Acad Sci U S A* 102 (34):12117-22. doi: 10.1073/pnas.0409097102.

Petruk, S., Y. Sedkov, D. M. Johnston, J. W. Hodgson, K. L. Black, S. K. Kovermann, S. Beck, E. Canaani, H. W. Brock, and A. Mazo. 2012. "TrxG and PcG proteins but not methylated histones remain associated with DNA through replication." *Cell* 150 (5):922-33. doi: 10.1016/j.cell.2012.06.046.

Pfaffeneder, T., B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, and T. Carell. 2011. "The discovery of 5-formylcytosine in embryonic stem cell DNA." *Angew Chem Int Ed Engl* 50 (31):7008-12. doi: 10.1002/anie.201103899.

Pin, C. L., and S. F. Konieczny. 2002. "A fast fiber enhancer exists in the muscle regulatory factor 4 gene promoter." *Biochem Biophys Res Commun* 299 (1):7-13.

Pollock, R., and R. Treisman. 1991. "Human SRF-related proteins: DNA-binding properties and potential regulatory targets." *Genes Dev* 5 (12A):2327-41.

Powell, C., E. Cornblath, and D. Goldman. 2014. "Zinc-binding Domain-dependent, Deaminase-independent Actions of Apolipoprotein B mRNA-editing Enzyme, Catalytic Polypeptide 2 (Apobec2), Mediate Its Effect on Zebrafish Retina Regeneration." *J Biol Chem* 289 (42):28924-41. doi: 10.1074/jbc.M114.603043.

Powell, C., F. Elsaeidi, and D. Goldman. 2012. "Injury-dependent Müller glia and ganglion cell reprogramming during tissue regeneration requires Apobec2a and Apobec2b." *J Neurosci* 32 (3):1096-109. doi: 10.1523/JNEUROSCI.5603-11.2012.

Powell, C., A. R. Grant, E. Cornblath, and D. Goldman. 2013. "Analysis of DNA methylation reveals a partial reprogramming of the Müller glia genome during retina regeneration." *Proc Natl Acad Sci U S A* 110 (49):19814-9. doi: 10.1073/pnas.1312009110.

Pownall, M. E., M. K. Gustafsson, and C. P. Emerson. 2002. "Myogenic regulatory factors and the specification of muscle progenitors in vertebrate embryos." *Annu Rev Cell Dev Biol* 18:747-83. doi: 10.1146/annurev.cellbio.18.012502.105758.

Prendergast, G. C., and E. B. Ziff. 1991. "Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region." *Science* 251 (4990):186-9.

Quinlan, A. R., and I. M. Hall. 2010. "BEDTools: a flexible suite of utilities for comparing genomic features." *Bioinformatics* 26 (6):841-2. doi: 10.1093/bioinformatics/btq033.

Rada-Iglesias, A., R. Bajpai, T. Swigut, S. A. Brugmann, R. A. Flynn, and J. Wysocka. 2011. "A unique chromatin signature uncovers early developmental enhancers in humans." *Nature* 470 (7333):279-83. doi: 10.1038/nature09692.

Rai, K., I. J. Huggins, S. R. James, A. R. Karpf, D. A. Jones, and B. R. Cairns. 2008. "DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45." *Cell* 135 (7):1201-12. doi: 10.1016/j.cell.2008.11.042.

Ramirez-Carrozzi, V. R., D. Braas, D. M. Bhatt, C. S. Cheng, C. Hong, K. R. Doty, J. C. Black, A. Hoffmann, M. Carey, and S. T. Smale. 2009. "A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling." *Cell* 138 (1):114-28. doi: 10.1016/j.cell.2009.04.020.

Ramsahoye, B. H., D. Biniszkiewicz, F. Lyko, V. Clark, A. P. Bird, and R. Jaenisch. 2000. "Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a." *Proc Natl Acad Sci U S A* 97 (10):5237-42.

Reik, W. 2007. "Stability and flexibility of epigenetic gene regulation in mammalian development." *Nature* 447 (7143):425-32. doi: 10.1038/nature05918.

Relaix, F., J. Demignon, C. Laclef, J. Pujol, M. Santolini, C. Niro, M. Lagha, D. Rocancourt, M. Buckingham, and P. Maire. 2013. "Six homeoproteins directly activate Myod expression in the gene regulatory networks that control early myogenesis." *PLoS Genet* 9 (4):e1003425. doi: 10.1371/journal.pgen.1003425.

Relaix, F., D. Montarras, S. Zaffran, B. Gayraud-Morel, D. Rocancourt, S. Tajbakhsh, A. Mansouri, A. Cumano, and M. Buckingham. 2006. "Pax3 and Pax7 have distinct and overlapping functions in adult muscle progenitor cells." *J Cell Biol* 172 (1):91-102. doi: 10.1083/jcb.200508044.

Relaix, F., D. Rocancourt, A. Mansouri, and M. Buckingham. 2005. "A Pax3/Pax7-dependent population of skeletal muscle progenitor cells." *Nature* 435 (7044):948-53. doi: 10.1038/nature03594.

Rhodes, S. J., and S. F. Konieczny. 1989. "Identification of MRF4: a new member of the muscle regulatory factor gene family." *Genes Dev* 3 (12B):2050-61.

Ribas, R., N. Moncaut, C. Siligan, K. Taylor, J. W. Cross, P. W. Rigby, and J. J. Carvajal. 2011. "Members of the TEAD family of transcription factors regulate the expression of Myf5 in ventral somitic compartments." *Dev Biol* 355 (2):372-80. doi: 10.1016/j.ydbio.2011.04.005.

Riggs, A. D. 1975. "X inactivation, differentiation, and DNA methylation." *Cytogenet Cell Genet* 14 (1):9-25.

Riising, E. M., I. Comet, B. Leblanc, X. Wu, J. V. Johansen, and K. Helin. 2014. "Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide." *Mol Cell* 55 (3):347-60. doi: 10.1016/j.molcel.2014.06.005.

Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov. 2011. "Integrative genomics viewer." *Nat Biotechnol* 29 (1):24-6. doi: 10.1038/nbt.1754.

Rodriguez, J., J. Frigola, E. Vendrell, R. A. Risques, M. F. Fraga, C. Morales, V. Moreno, M. Esteller, G. Capellà, M. Ribas, and M. A. Peinado. 2006. "Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers." *Cancer Res* 66 (17):8462-9468. doi: 10.1158/0008-5472.CAN-06-0293.

Rubin, B. P., K. Nishijo, H. I. Chen, X. Yi, D. P. Schuetze, R. Pal, S. I. Prajapati, J. Abraham, B. R. Arenkiel, Q. R. Chen, S. Davis, A. T. McCleish, M. R. Capecchi, J. E. Michalek, L. A. Zarzabal, J. Khan, Z. Yu, D. M. Parham, F. G. Barr, P. S. Meltzer, Y. Chen, and C. Keller. 2011. "Evidence for an unanticipated relationship between undifferentiated pleomorphic sarcoma and embryonal rhabdomyosarcoma." *Cancer Cell* 19 (2):177-91. doi: 10.1016/j.ccr.2010.12.023.

Rudenko, A., M. M. Dawlaty, J. Seo, A. W. Cheng, J. Meng, T. Le, K. F. Faull, R. Jaenisch, and L. H. Tsai. 2013. "Tet1 is critical for neuronal activity-regulated gene expression and memory extinction." *Neuron* 79 (6):1109-22. doi: 10.1016/j.neuron.2013.08.003.

Saitou, M., S. Kagiwada, and K. Kurimoto. 2012. "Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells." *Development* 139 (1):15-31. doi: 10.1242/dev.050849.

Sakurai, H., T. Era, L. M. Jakt, M. Okada, S. Nakai, and S. Nishikawa. 2006. "In vitro modeling of paraxial and lateral mesoderm differentiation reveals early reversibility." *Stem Cells* 24 (3):575-86. doi: 10.1634/stemcells.2005-0256.

Saluz, H. P., J. Jiricny, and J. P. Jost. 1986. "Genomic sequencing reveals a positive correlation between the kinetics of strand-specific DNA demethylation of the overlapping estradiol/glucocorticoid-receptor binding sites and the rate of avian vitellogenin mRNA synthesis." *Proc Natl Acad Sci U S A* 83 (19):7167-71.

Santiago, M., C. Antunes, M. Guedes, N. Sousa, and C. J. Marques. 2014. "Tet enzymes and DNA hydroxymethylation in neural development and function - How critical are they?" *Genomics*. doi: 10.1016/j.ygeno.2014.08.018.

Sati, S., V. S. Tanwar, K. A. Kumar, A. Patowary, V. Jain, S. Ghosh, S. Ahmad, M. Singh, S. U. Reddy, G. R. Chandak, M. Raghunath, S. Sivasubbu, K. Chakraborty, V. Scaria, and S. Sengupta. 2012. "High resolution methylome map of rat indicates role of intragenic DNA methylation in identification of coding region." *PLoS One* 7 (2):e31621. doi: 10.1371/journal.pone.0031621.

Sato, T., D. Rocancourt, L. Marques, S. Thorsteinsdóttir, and M. Buckingham. 2010. "A Pax3/Dmrt2/Myf5 regulatory cascade functions at the onset of myogenesis." *PLoS Genet* 6 (4):e1000897. doi: 10.1371/journal.pgen.1000897.

Sato, Y., H. C. Probst, R. Tatsumi, Y. Ikeuchi, M. S. Neuberger, and C. Rada. 2010. "Deficiency in Apobec2 leads to a shift in muscle fiber type, diminished body mass, and myopathy." *J Biol Chem* 285 (10):7111-8. doi: 10.1074/jbc.M109.052977.

Saxonov, S., P. Berg, and D. L. Brutlag. 2006. "A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters." *Proc Natl Acad Sci U S A* 103 (5):1412-7. doi: 10.1073/pnas.0510310103.

Scarpa, S., M. Lucarelli, F. Palitti, D. Carotti, and R. Strom. 1996. "Simultaneous myogenin expression and overall DNA hypomethylation promote *in vitro* myoblast differentiation." *Cell Growth Differ* 7 (8):1051-8.

Schiaffino, S., and C. Mammucari. 2011. "Regulation of skeletal muscle growth by the IGF1-Akt/PKB pathway: insights from genetic models." *Skelet Muscle* 1 (1):4. doi: 10.1186/2044-5040-1-4.

Schiaffino, S., and C. Reggiani. 2011. "Fiber types in mammalian skeletal muscles." *Physiol Rev* 91 (4):1447-531. doi: 10.1152/physrev.00031.2010.

Schienda, J., K. A. Engleka, S. Jun, M. S. Hansen, J. A. Epstein, C. J. Tabin, L. M. Kunkel, and G. Kardon. 2006. "Somitic origin of limb muscle satellite and side population cells." *Proc Natl Acad Sci U S A* 103 (4):945-50. doi: 10.1073/pnas.0510164103.

Schmidl, C., M. Klug, T. J. Boeld, R. Andreesen, P. Hoffmann, M. Edinger, and M. Rehli. 2009. "Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity." *Genome Res* 19 (7):1165-74. doi: 10.1101/gr.091470.109.

Schwartz, S., E. Meshorer, and G. Ast. 2009. "Chromatin organization marks exon-intron structure." *Nat Struct Mol Biol* 16 (9):990-5. doi: 10.1038/nsmb.1659

Seisenberger, S., J. R. Peat, T. A. Hore, F. Santos, W. Dean, and W. Reik. 2013. "Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers." *Philos Trans R Soc Lond B Biol Sci* 368 (1609):20110330. doi: 10.1098/rstb.2011.0330.

Sharif, J., M. Muto, S. Takebayashi, I. Suetake, A. Iwamatsu, T. A. Endo, J. Shinga, Y. Mizutani-Koseki, T. Toyoda, K. Okamura, S. Tajima, K. Mitsuya, M. Okano, and H. Koseki. 2007. "The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA." *Nature* 450 (7171):908-12. doi: 10.1038/nature06397.

Sheardown, S. A., S. M. Duthie, C. M. Johnston, A. E. Newall, E. J. Formstone, R. M. Arkell, T. B. Nesterova, G. C. Alghisi, S. Rastan, and N. Brockdorff. 1997. "Stabilization of Xist RNA mediates initiation of X chromosome inactivation." *Cell* 91 (1):99-107.

Shearstone, J. R., R. Pop, C. Bock, P. Boyle, A. Meissner, and M. Socolovsky. 2011. "Global DNA demethylation during mouse erythropoiesis *in vivo*." *Science* 334 (6057):799-802. doi: 10.1126/science.1207306.

Shen, L., Y. Kondo, Y. Guo, J. Zhang, L. Zhang, S. Ahmed, J. Shu, X. Chen, R. A. Waterland, and J. P. Issa. 2007. "Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters." *PLoS Genet* 3 (10):2023-36. doi: 10.1371/journal.pgen.0030181.

Shen, X., Y. Liu, Y. J. Hsu, Y. Fujiwara, J. Kim, X. Mao, G. C. Yuan, and S. H. Orkin. 2008. "EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency." *Mol Cell* 32 (4):491-502. doi: 10.1016/j.molcel.2008.10.016.

Shen, Y., J. Chow, Z. Wang, and G. Fan. 2006. "Abnormal CpG island methylation occurs during *in vitro* differentiation of human embryonic stem cells." *Hum Mol Genet* 15 (17):2623-35. doi: 10.1093/hmg/ddl188.

Shi, Y., F. Lan, C. Matson, P. Mulligan, J. R. Whetstine, P. A. Cole, and R. A. Casero. 2004. "Histone demethylation mediated by the nuclear amine oxidase homolog LSD1." *Cell* 119 (7):941-53. doi: 10.1016/j.cell.2004.12.012.

Shield, M. A., H. S. Haugen, C. H. Clegg, and S. D. Hauschka. 1996. "E-box sites and a proximal regulatory region of the muscle creatine kinase gene differentially regulate expression in diverse skeletal muscles and cardiac muscle of transgenic mice." *Mol Cell Biol* 16 (9):5058-68.

Shmelkov, S. V., L. Jun, R. St Clair, D. McGarrigle, C. A. Derderian, J. K. Usenko, C. Costa, F. Zhang, X. Guo, and S. Rafii. 2004. "Alternative promoters regulate transcription of the gene that encodes stem cell surface protein AC133." *Blood* 103 (6):2055-61. doi: 10.1182/blood-2003-06-1881.

Shukla, S., E. Kavak, M. Gregory, M. Imashimizu, B. Shutinoski, M. Kashlev, P. Oberdoerffer, R. Sandberg, and S. Oberdoerffer. 2011. "CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing." *Nature* 479 (7371):74-9. doi: 10.1038/nature10442.

Siggens, L., and K. Ekwall. 2014. "Epigenetics, chromatin and genome organization: recent advances from the ENCODE project." *J Intern Med* 276 (3):201-14. doi: 10.1111/joim.12231.

Simone, C. 2006. "SWI/SNF: the crossroads where extracellular signaling pathways meet chromatin." *J Cell Physiol* 207 (2):309-14. doi: 10.1002/jcp.20514.

Smiraglia, D. J., L. J. Rush, M. C. Frühwald, Z. Dai, W. A. Held, J. F. Costello, J. C. Lang, C. Eng, B. Li, F. A. Wright, M. A. Caligiuri, and C. Plass. 2001. "Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies." *Hum Mol Genet* 10 (13):1413-9.

Smith, Z. D., and A. Meissner. 2013. "DNA methylation: roles in mammalian development." *Nat Rev Genet* 14 (3):204-20. doi: 10.1038/nrg3354.

Soleimani, V. D., V. G. Punch, Y. Kawabe, A. E. Jones, G. A. Palidwor, C. J. Porter, J. W. Cross, J. J. Carvajal, C. E. Kockx, W. F. van IJcken, T. J. Perkins, P. W. Rigby, F. Grosveld, and M. A. Rudnicki. 2012. "Transcriptional dominance of Pax7 in adult myogenesis is due to high-affinity recognition of homeodomain motifs." *Dev Cell* 22 (6):1208-20. doi: 10.1016/j.devcel.2012.03.014.

Song, J., O. Rechkoblit, T. H. Bestor, and D. J. Patel. 2011. "Structure of Dnmt1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation." *Science* 331 (6020):1036-40. doi: 10.1126/science.1195380.

Song, J., M. Teplova, S. Ishibe-Murakami, and D. J. Patel. 2012. "Structure-based mechanistic insights into Dnmt1-mediated maintenance DNA methylation." *Science* 335 (6069):709-12. doi: 10.1126/science.1214453.

Sorensen, P. H., J. C. Lynch, S. J. Qualman, R. Tirabosco, J. F. Lim, H. M. Maurer, J. A. Bridge, W. M. Crist, T. J. Triche, and F. G. Barr. 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 (11):2672-9.

Spitz, F., J. Demignon, A. Porteu, A. Kahn, J. P. Concordet, D. Daegelen, and P. Maire. 1998. "Expression of myogenin during embryogenesis is controlled by Six/sine oculis homeoproteins through a conserved MEF3 binding site." *Proc Natl Acad Sci U S A* 95 (24):14220-5.

Stadler, M. B., R. Murr, L. Burger, R. Ivanek, F. Lienert, A. Schöler, E. van Nimwegen, C. Wirbelauer, E. J. Oakeley, D. Gaidatzis, V. K. Tiwari, and D. Schübeler. 2011. "DNA-binding factors shape the mouse methylome at distal regulatory regions." *Nature* 480 (7378):490-5. doi: 10.1038/nature10716.

Strahl, B. D., and C. D. Allis. 2000. "The language of covalent histone modifications." *Nature* 403 (6765):41-5. doi: 10.1038/47412.

Straussman, R., D. Nejman, D. Roberts, I. Steinfeld, B. Blum, N. Benvenisty, I. Simon, Z. Yakhini, and H. Cedar. 2009. "Developmental programming of CpG island methylation profiles in the human genome." *Nat Struct Mol Biol* 16 (5):564-71. doi: 10.1038/nsmb.1594.

Struhl, K. 1999. "Fundamentally different logic of gene regulation in eukaryotes and prokaryotes." *Cell* 98 (1):1-4. doi: 10.1016/S0092-8674(00)80599-1.

Stöger, R., P. Kubicka, C. G. Liu, T. Kafri, A. Razin, H. Cedar, and D. P. Barlow. 1993. "Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal." *Cell* 73 (1):61-71.

Subramanyam, D., S. Lamouille, R. L. Judson, J. Y. Liu, N. Bucay, R. Derynck, and R. Blelloch. 2011. "Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells." *Nat Biotechnol* 29 (5):443-8. doi: 10.1038/nbt.1862.

Suelves, M., F. Lluís, V. Ruiz, A. R. Nebreda, and P. Muñoz-Cánoves. 2004. "Phosphorylation of MRF4 transactivation domain by p38 mediates repression of specific myogenic genes." *EMBO J* 23 (2):365-75. doi: 10.1038/sj.emboj.7600056.

Suelves, M., B. Vidal, A. L. Serrano, M. Tjwa, J. Roma, R. López-Alemany, A. Luttun, M. M. de Lagrán, A. Díaz-Ramos, M. A. Díaz, M. Jardí, M. Roig, M. Dierssen, M. Dewerchin, P. Carmeliet, and P. Muñoz-Cánoves. 2007. "uPA deficiency exacerbates muscular dystrophy in MDX mice." *J Cell Biol* 178 (6):1039-51. doi: 10.1083/jcb.200705127.

Suh, M. R., Y. Lee, J. Y. Kim, S. K. Kim, S. H. Moon, J. Y. Lee, K. Y. Cha, H. M. Chung, H. S. Yoon, S. Y. Moon, V. N. Kim, and K. S. Kim. 2004. "Human embryonic stem cells express a unique set of microRNAs." *Dev Biol* 270 (2):488-98. doi: 10.1016/j.ydbio.2004.02.019.

Summerbell, D., P. R. Ashby, O. Coutelle, D. Cox, S. Yee, and P. W. Rigby. 2000. "The expression of Myf5 in the developing mouse embryo is controlled by discrete and dispersed enhancers specific for particular populations of skeletal muscle precursors." *Development* 127 (17):3745-57.

Szulwach, K. E., X. Li, Y. Li, C. X. Song, H. Wu, Q. Dai, H. Irier, A. K. Upadhyay, M. Gearing, A. I. Levey, A. Vasanthakumar, L. A. Godley, Q. Chang, X. Cheng, C. He, and P. Jin. 2011. "5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging." *Nat Neurosci* 14 (12):1607-16. doi: 10.1038/nn.2959.

Szyf, M., J. Rouleau, J. Theberge, and V. Bozovic. 1992. "Induction of myogenic differentiation by an expression vector encoding the DNA methyltransferase cDNA sequence in the antisense orientation." *J Biol Chem* 267 (18):12831-6.

Sérandour, A. A., S. Avner, F. Oger, M. Bizot, F. Percevault, C. Lucchetti-Miganeh, G. Palierne, C. Gheeraert, F. Barloy-Hubler, C. L. Péron, T. Madigou, E. Durand, P. Froguel, B. Staels, P. Lefebvre, R. Métivier, J. Eeckhoute, and G. Salbert. 2012. "Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers." *Nucleic Acids Res* 40 (17):8255-65. doi: 10.1093/nar/gks595.

Sérandour, A. A., S. Avner, F. Percevault, F. Demay, M. Bizot, C. Lucchetti-Miganeh, F. Barloy-Hubler, M. Brown, M. Lupien, R. Métivier, G. Salbert, and J. Eeckhoute. 2011. "Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers." *Genome Res* 21 (4):555-65. doi: 10.1101/gr.111534.110.

Sørensen, A. L., B. M. Jacobsen, A. H. Reiner, I. S. Andersen, and P. Collas. 2010. "Promoter DNA methylation patterns of differentiated cells are largely programmed at the progenitor stage." *Mol Biol Cell* 21 (12):2066-77. doi: 10.1091/mbc.E10-01-0018.

Tajbakhsh, S., and M. Buckingham. 2000. "The birth of muscle progenitor cells in the mouse: spatiotemporal considerations." *Curr Top Dev Biol* 48:225-68.

Tajbakhsh, S., D. Rocancourt, G. Cossu, and M. Buckingham. 1997. "Redefining the genetic hierarchies controlling skeletal myogenesis: Pax-3 and Myf-5 act upstream of MyoD." *Cell* 89 (1):127-38.

Takagi, H., S. Tajima, and A. Asano. 1995. "Overexpression of DNA methyltransferase in myoblast cells accelerates myotube formation." *Eur J Biochem* 231 (2):282-91.

Tamagnone, L., S. Artigiani, H. Chen, Z. He, G. I. Ming, H. Song, A. Chedotal, M. L. Winberg, C. S. Goodman, M. Poo, M. Tessier-Lavigne, and P. M. Comoglio. 1999. "Plexins are a large family of receptors for transmembrane, secreted, and GPI-anchored semaphorins in vertebrates." *Cell* 99 (1):71-80.

Tapscott, S. J. 2005. "The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription." *Development* 132 (12):2685-95. doi: 10.1242/dev.01874.

Tapscott, S. J., R. L. Davis, M. J. Thayer, P. F. Cheng, H. Weintraub, and A. B. Lassar. 1988. "MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts." *Science* 242 (4877):405-11.

Tatsumi, R., Y. Sankoda, J. E. Anderson, Y. Sato, W. Mizunoya, N. Shimizu, T. Suzuki, M. Yamada, R. P. Rhoads, Y. Ikeuchi, and R. E. Allen. 2009. "Possible implication of satellite cells in regenerative motoneuritogenesis: HGF upregulates neural chemorepellent Sema3A during myogenic differentiation." *Am J Physiol Cell Physiol* 297 (2):C238-52. doi: 10.1152/ajpcell.00161.2009.

Teboul, L., D. Summerbell, and P. W. Rigby. 2003. "The initial somitic phase of Myf5 expression requires neither Shh signaling nor Gli regulation." *Genes Dev* 17 (23):2870-4. doi: 10.1101/gad.1117603.

Tessema, M., C. M. Yingling, M. J. Grimes, C. L. Thomas, Y. Liu, S. Leng, N. Joste, and S. A. Belinsky. 2012. "Differential epigenetic regulation of TOX subfamily high mobility group box genes in lung and breast cancers." *PLoS One* 7 (4):e34850. doi: 10.1371/journal.pone.0034850.

Tian, L., X. Wu, Y. Lin, Z. Liu, F. Xiong, Z. Han, Y. Zhou, Q. Zeng, Y. Wang, J. Deng, and H. Chen. 2009. "Characterization and potential function of a novel pre-implantation embryo-specific RING finger protein: TRIML1." *Mol Reprod Dev* 76 (7):656-64. doi: 10.1002/mrd.20997.

Tomizawa, S., H. Kobayashi, T. Watanabe, S. Andrews, K. Hata, G. Kelsey, and H. Sasaki. 2011. "Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes." *Development* 138 (5):811-20. doi: 10.1242/dev.061416.

Toyota, M., and J. P. Issa. 1999. "CpG island methylator phenotypes in aging and cancer." *Semin Cancer Biol* 9 (5):349-57. doi: 10.1006/scbi.1999.0135.

Tremblay, J. P., F. Malouin, R. Roy, J. Huard, J. P. Bouchard, A. Satoh, and C. L. Richards. 1993. "Results of a triple blind clinical study of myoblast transplantations without immunosuppressive treatment in young boys with Duchenne muscular dystrophy." *Cell Transplant* 2 (2):99-112.

Tsumagari, K., C. Baribault, J. Terragni, S. Chandra, C. Renshaw, Z. Sun, L. Song, G. E. Crawford, S. Pradhan, M. Lacey, and M. Ehrlich. 2013. "DNA methylation and differentiation: HOX genes in muscle cells." *Epigenetics Chromatin* 6 (1):25. doi: 10.1186/1756-8935-6-25.

Tsumagari, K., C. Baribault, J. Terragni, K. E. Varley, J. Gertz, S. Pradhan, M. Badoo, C. M. Crain, L. Song, G. E. Crawford, R. M. Myers, M. Lacey, and M. Ehrlich. 2013. "Early *de novo* DNA methylation and prolonged demethylation in the muscle lineage." *Epigenetics* 8 (3):317-32. doi: 10.4161/epi.23989.

Tsumura, A., T. Hayakawa, Y. Kumaki, S. Takebayashi, M. Sakaue, C. Matsuoka, K. Shimotohno, F. Ishikawa, E. Li, H. R. Ueda, J. Nakayama, and M. Okano. 2006. "Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b." *Genes Cells* 11 (7):805-14. doi: 10.1111/j.1365-2443.2006.00984.x.

Tung, P. Y., N. V. Varlakhanova, and P. S. Knoepfler. 2013. "Identification of DPPA4 and DPPA2 as a novel family of pluripotency-related oncogenes." *Stem Cells* 31 (11):2330-42. doi: 10.1002/stem.1526.

Uysal-Onganer, P., and R. M. Kypta. 2012. "Wnt11 in 2011 - the regulation and function of a non-canonical Wnt." *Acta Physiol (Oxf)* 204 (1):52-64. doi: 10.1111/j.1748-1716.2011.02297.x.

Valinluck, V., and L. C. Sowers. 2007. "Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase Dnmt1." *Cancer Res* 67 (3):946-50. doi: 10.1158/0008-5472.CAN-06-3123.

Vincent, C. K., A. Gualberto, C. V. Patel, and K. Walsh. 1993. "Different regulatory sequences control creatine kinase-M gene expression in directly injected skeletal and cardiac muscle." *Mol Cell Biol* 13 (2):1264-72.

Viré, E., C. Brenner, R. Deplus, L. Blanchon, M. Fraga, C. Didelot, L. Morey, A. Van Eynde, D. Bernard, J. M. Vanderwinden, M. Bollen, M. Esteller, L. Di Croce, Y. de Launoit, and F. Fuks. 2006. "The Polycomb group protein EZH2 directly controls DNA methylation." *Nature* 439 (7078):871-4. doi: 10.1038/nature04431.

Vittet, D., M. H. Prandini, R. Berthier, A. Schweitzer, H. Martin-Sisteron, G. Uzan, and E. Dejana. 1996. "Embryonic stem cells differentiate *in vitro* to endothelial cells through successive maturation steps." *Blood* 88 (9):3424-31.

Vonica, A., A. Rosa, B. L. Arduini, and A. H. Brivanlou. 2011. "Apobec2, a selective inhibitor of TGFβ signaling, regulates left-right axis specification during early embryogenesis." *Dev Biol* 350 (1):13-23. doi: 10.1016/j.ydbio.2010.09.016.

Wall, B. T., M. L. Dirks, and L. J. van Loon. 2013. "Skeletal muscle atrophy during short-term disuse: implications for age-related sarcopenia." *Ageing Res Rev* 12 (4):898-906. doi: 10.1016/j.arr.2013.07.003.

Walsh, C. P., J. R. Chaillet, and T. H. Bestor. 1998. "Transcription of IAP endogenous retroviruses is constrained by cytosine methylation." *Nat Genet* 20 (2):116-7. doi: 10.1038/2413.

Wang, L., J. Zhang, J. Duan, X. Gao, W. Zhu, X. Lu, L. Yang, G. Li, W. Ci, W. Li, Q. Zhou, N. Aluru, F. Tang, C. He, X. Huang, and J. Liu. 2014. "Programming and inheritance of parental DNA methylomes in mammals." *Cell* 157 (4):979-91. doi: 10.1016/j.cell.2014.04.017.

Wang, X., C. G. Duan, K. Tang, B. Wang, H. Zhang, M. Lei, K. Lu, S. K. Mangrauthia, P. Wang, G. Zhu, Y. Zhao, and J. K. Zhu. 2013. "RNA-binding protein regulates plant DNA methylation by controlling mRNA processing at the intronic heterochromatin-containing gene IBM1." *Proc Natl Acad Sci U S A* 110 (38):15467-72. doi: 10.1073/pnas.1315399110.

Wang, X., Y. Ono, S. C. Tan, R. J. Chai, C. Parkin, and P. W. Ingham. 2011. "Prdm1a and miR-499 act sequentially to restrict Sox6 activity to the fast-twitch muscle lineage in the zebrafish embryo." *Development* 138 (20):4399-404. doi: 10.1242/dev.070516.

Watanabe, D., I. Suetake, T. Tada, and S. Tajima. 2002. "Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis." *Mech Dev* 118 (1-2):187-90.

Waterland, R. A., R. Kellermayer, M. T. Rached, N. Tatevian, M. V. Gomes, J. Zhang, L. Zhang, A. Chakravarty, W. Zhu, E. Laritsky, W. Zhang, X. Wang, and L. Shen. 2009. "Epigenomic profiling indicates a role for DNA methylation in early postnatal liver development." *Hum Mol Genet* 18 (16):3026-38. doi: 10.1093/hmg/ddp241.

Waterston, R. H., K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S. E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M. R. Brent, D. G. Brown, S. D. Brown, C. Bult, J. Burton, J. Butler, R. D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A. T. Chinwalla, D. M. Church, M. Clamp, C. Clee, F. S. Collins, L. L. Cook, R. R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K. D. Delehaunty, J. Deri, E. T. Dermitzakis, C. Dewey, N. J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D. M. Dunn, S. R. Eddy, L. Elnitski, R. D. Emes, P. Eswara, E. Eyras, A. Felsenfeld, G. A. Fewell, P. Flicek, K. Foley, W. N. Frankel, L. A. Fulton, R. S. Fulton, T. S. Furey, D. Gage, R. A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T. A. Graves, E. D. Green, S. Gregory, R. Guigó, M. Guyer, R. C. Hardison, D. Haussler, Y. Hayashizaki, L. W. Hillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D. B. Jaffe, L. S. Johnson, M. Jones, T. A. Jones, A. Joy, M. Kamal, E. K. Karlsson, D. Karolchik, A. Kasprzyk, J. Kawai, E. Keibler, C. Kells, W. J. Kent, A. Kirby, D. L. Kolbe, I. Korf, R. S. Kucherlapati, E. J. Kulbokas, D. Kulp, T. Landers, J. P. Leger, S. Leonard, I. Letunic, R. Levine, J. Li, M. Li, C. Lloyd, S. Lucas, B. Ma, D. R. Maglott, E. R. Mardis, L. Matthews, E. Mauceli, J. H. Mayer, M. McCarthy, W. R. McCombie, S. McLaren, K. McLay, J. D. McPherson, J. Meldrim, B. Meredith, J. P. Mesirov, W. Miller, T. L. Miner, E. Mongin, K. T. Montgomery, M. Morgan, R. Mott, J. C. Mullikin, D. M. Muzny, W. E. Nash, J. O. Nelson, M. N. Nhan, R. Nicol, Z. Ning, C. Nusbaum, M. J. O'Connor, Y. OkAzaki, K. Oliver, E. Overton-Larty, L. Pachter, G. Parra, K. H. Pepin, J. Peterson, P. Pevzner, R. Plumb, C. S. Pohl, A. Poliakov, T. C. Ponce, C. P. Ponting, S. Potter, M. Quail, A. Reymond, B. A. Roe, K. M. Roskin, E. M. Rubin, A. G. Rust, R. Santos, V. Sapojnikov, B. Schultz, J. Schultz, M. S. Schwartz, S. Schwartz, C. Scott, S. Seaman, S. Searle, T. Sharpe,

A. Sheridan, R. Shownkeen, S. Sims, J. B. Singer, G. Slater, A. Smit, D. R. Smith, B. Spencer, A. Stabenau, N. Stange-Thomann, C. Sugnet, M. Suyama, G. Tesler, J. Thompson, D. Torrents, E. Trevaskis, J. Tromp, C. Ucla, A. Ureta-Vidal, J. P. Vinson, A. C. Von Niederhausern, C. M. Wade, M. Wall, R. J. Weber, R. B. Weiss, M. C. Wendl, A. P. West, K. Wetterstrand, R. Wheeler, S. Whelan, J. Wierzbowski, D. Willey, S. Williams, R. K. Wilson, E. Winter, K. C. Worley, D. Wyman, S. Yang, S. P. Yang, E. M. Zdobnov, M. C. Zody, E. S. Lander, and Mouse Genome Sequencing Consortium. 2002. "Initial sequencing and comparative analysis of the mouse genome." *Nature* 420 (6915):520-62. doi: 10.1038/nature01262.

Watt, F., and P. L. Molloy. 1988. "Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter." *Genes Dev* 2 (9):1136-43.

Weber, M., I. Hellmann, M. B. Stadler, L. Ramos, S. Pääbo, M. Rebhan, and D. Schübeler. 2007. "Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome." *Nat Genet* 39 (4):457-66. doi: 10.1038/ng1990.

Weintraub, H., S. J. Tapscott, R. L. Davis, M. J. Thayer, M. A. Adam, A. B. Lassar, and A. D. Miller. 1989. "Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD." *Proc Natl Acad Sci U S A* 86 (14):5434-8.

Wen, Y., P. Bi, W. Liu, A. Asakura, C. Keller, and S. Kuang. 2012. "Constitutive Notch activation upregulates Pax7 and promotes the self-renewal of skeletal muscle satellite cells." *Mol Cell Biol* 32 (12):2300-11. doi: 10.1128/MCB.06753-11.

Wheeler, M. T., E. C. Snyder, M. N. Patterson, and S. J. Swoap. 1999. "An E-box within the MHC IIB gene is bound by MyoD and is required for gene expression in fast muscle." *Am J Physiol* 276 (5 Pt 1):C1069-78.

Whetstine, J. R., A. Nottke, F. Lan, M. Huarte, S. Smolikov, Z. Chen, E. Spooner, E. Li, G. Zhang, M. Colaiacovo, and Y. Shi. 2006. "Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases." *Cell* 125 (3):467-81. doi: 10.1016/j.cell.2006.03.028.

Whitcomb, S. J., A. Basu, C. D. Allis, and E. Bernstein. 2007. "Polycomb Group proteins: an evolutionary perspective." *Trends Genet* 23 (10):494-502. doi: 10.1016/j.tig.2007.08.006.

Whyte, W. A., D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee, and R. A. Young. 2013. "Master transcription factors and mediator establish super-enhancers at key cell identity genes." *Cell* 153 (2):307-19. doi: 10.1016/j.cell.2013.03.035.

Wiles, M. V., and G. Keller. 1991. "Multiple hematopoietic lineages develop from embryonic stem (ES) cells in culture." *Development* 111 (2):259-67.

Williams, K., J. Christensen, M. T. Pedersen, J. V. Johansen, P. A. Cloos, J. Rappsilber, and K. Helin. 2011. "Tet1 and hydroxymethylcytosine in transcription and DNA methylation fidelity." *Nature* 473 (7347):343-8. doi: 10.1038/nature10066.

Workman, J. L., and R. E. Kingston. 1998. "Alteration of nucleosome structure as a mechanism of transcriptional regulation." *Annu Rev Biochem* 67:545-79. doi: 10.1146/annurev.biochem.67.1.545.

Wossidlo, M., J. Arand, V. Sebastiano, K. Lepikhov, M. Boiani, R. Reinhardt, H. Schöler, and J. Walter. 2010. "Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes." *EMBO J* 29 (11):1877-88. doi: 10.1038/emboj.2010.80.

Wossidlo, M., T. Nakamura, K. Lepikhov, C. J. Marques, V. Zakhartchenko, M. Boiani, J. Arand, T. Nakano, W. Reik, and J. Walter. 2011. "5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming." *Nat Commun* 2:241. doi: 10.1038/ncomms1240.

Wu, H., V. Coskun, J. Tao, W. Xie, W. Ge, K. Yoshikawa, E. Li, Y. Zhang, and Y. E. Sun. 2010. "Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes." *Science* 329 (5990):444-8. doi: 10.1126/science.1190485.

Wu, H., A. C. D'Alessio, S. Ito, Z. Wang, K. Cui, K. Zhao, Y. E. Sun, and Y. Zhang. 2011. "Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells." *Genes Dev* 25 (7):679-84. doi: 10.1101/gad.2036011.

Wu, H., A. C. D'Alessio, S. Ito, K. Xia, Z. Wang, K. Cui, K. Zhao, Y. E. Sun, and Y. Zhang. 2011. "Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells." *Nature* 473 (7347):389-93. doi: 10.1038/nature09934.

Wu, H., and Y. Zhang. 2011. "Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation." *Genes Dev* 25 (23):2436-52. doi: 10.1101/gad.179184.111.

Wu, L., and Q. Zheng. 2014. "Active demethylation of the IL-2 Promoter in CD4+ T cells is mediated by an inducible DNA glycosylase, Myh." *Mol Immunol* 58 (1):38-49. doi: 10.1016/j.molimm.2013.10.016.

Xia, X., Y. Zhang, C. R. Zieth, and S. C. Zhang. 2007. "Transgenes delivered by lentiviral vector are suppressed in human embryonic stem cells in a promoter-dependent manner." *Stem Cells Dev* 16 (1):167-76. doi: 10.1089/scd. 2006.0057.

Xie, W., C. L. Barr, A. Kim, F. Yue, A. Y. Lee, J. Eubanks, E. L. Dempster, and B. Ren. 2012. "Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome." *Cell* 148 (4):816-31. doi: 10.1016/j.cell.2011.12.035.

Xie, W., M. D. Schultz, R. Lister, Z. Hou, N. Rajagopal, P. Ray, J. W. Whitaker, S. Tian, R. D. Hawkins, D. Leung, H. Yang, T. Wang, A. Y. Lee, S. A. Swanson, J. Zhang, Y. Zhu, A. Kim, J. R. Nery, M. A. Urich, S. Kuan, C. A. Yen, S. Klugman, P. Yu, K. Suknuntha, N. E. Propson, H. Chen, L. E. Edsall, U. Wagner, Y. Li, Z. Ye, A. Kulkarni, Z. Xuan, W. Y. Chung, N. C. Chi, J. E. Antosiewicz-Bourget, I. Slukvin, R. Stewart, M. Q. Zhang, W. Wang, J. A. Thomson, J. R. Ecker, and B. Ren. 2013. "Epigenomic analysis of multilineage differentiation of human embryonic stem cells." *Cell* 153 (5):1134-48. doi: 10.1016/j.cell.2013.04.022.

Xu, J., S. D. Pope, A. R. Jazirehi, J. L. Attema, P. Papathanasiou, J. A. Watts, K. S. Zaret, I. L. Weissman, and S. T. Smale. 2007. "Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells." *Proc Natl Acad Sci U S A* 104 (30):12377-82. doi: 10.1073/pnas.0704579104.

Xu, J., J. A. Watts, S. D. Pope, P. Gadue, M. Kamps, K. Plath, K. S. Zaret, and S. T. Smale. 2009. "Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells." *Genes Dev* 23 (24):2824-38. doi: 10.1101/gad.1861209.

Xu, K., D. C. Chong, S. A. Rankin, A. M. Zorn, and O. Cleaver. 2009. "Rasip1 is required for endothelial cell motility, angiogenesis and vessel formation." *Dev Biol* 329 (2):269-79. doi: 10.1016/j.ydbio.2009.02.033.

Xu, Y., C. Xu, A. Kato, W. Tempel, J. G. Abreu, C. Bian, Y. Hu, D. Hu, B. Zhao, T. Cerovina, J. Diao, F. Wu, H. H. He, Q. Cui, E. Clark, C. Ma, A. Barbara, G. J. Veenstra, G. Xu, U. B. Kaiser, X. S. Liu, S. P. Sugrue, X. He, J. Min, Y. Kato, and Y. G. Shi. 2012. "Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development." *Cell* 151 (6):1200-13. doi: 10.1016/j.cell.2012.11.014.

Yamaguchi, S., K. Hong, R. Liu, A. Inoue, L. Shen, K. Zhang, and Y. Zhang. 2013. "Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming." *Cell Res* 23 (3):329-39. doi: 10.1038/cr.2013.22.

Yamaji, M., Y. Seki, K. Kurimoto, Y. Yabuta, M. Yuasa, M. Shigeta, K. Yamanaka, Y. Ohinata, and M. Saitou. 2008. "Critical function of Prdm14 for the establishment of the germ cell lineage in mice." *Nat Genet* 40 (8):1016-22. doi: 10.1038/ng.186.

Yee, S. P., and P. W. Rigby. 1993. "The regulation of myogenin gene expression during the embryonic development of the mouse." *Genes Dev* 7 (7A):1277-89.

Yoder, J. A., C. P. Walsh, and T. H. Bestor. 1997. "Cytosine methylation and the ecology of intragenomic parasites." *Trends Genet* 13 (8):335-40.

Yuen, R. K., S. M. Neumann, A. K. Fok, M. S. Peñaherrera, D. E. McFadden, W. P. Robinson, and M. S. Kobor. 2011. "Extensive epigenetic reprogramming in human somatic tissues between fetus and adult." *Epigenetics Chromatin* 4:7. doi: 10.1186/1756-8935-4-7.

Zammit, P. S., J. J. Carvajal, J. P. Golding, J. E. Morgan, D. Summerbell, J. Zolnerciks, T. A. Partridge, P. W. Rigby, and J. R. Beauchamp. 2004. "Myf5 expression in satellite cells and spindles in adult muscle is controlled by separate genetic elements." *Dev Biol* 273 (2):454-65. doi: 10.1016/j.ydbio.2004.05.038.

Zhan, M., D. R. Riordon, B. Yan, Y. S. Tarasova, S. Bruweleit, K. V. Tarasov, R. A. Li, R. P. Wersto, and K. R. Boheler. 2012. "The B-MYB transcriptional network guides cell cycle progression and fate decisions to sustain self-renewal and the identity of pluripotent stem cells." *PLoS One* 7 (8):e42350. doi: 10.1371/journal.pone.0042350.

Zhang, F., J. H. Pomerantz, G. Sen, A. T. Palermo, and H. M. Blau. 2007. "Active tissue-specific DNA demethylation conferred by somatic cell nuclei in stable heterokaryons." *Proc Natl Acad Sci U S A* 104 (11):4395-400. doi: 10.1073/pnas.0700181104.

Zhang, H., X. Zhang, E. Clark, M. Mulcahey, S. Huang, and Y. G. Shi. 2010. "Tet1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine." *Cell Res* 20 (12):1390-3. doi: 10.1038/cr.2010.156.

Zhang, L., X. Lu, J. Lu, H. Liang, Q. Dai, G. L. Xu, C. Luo, H. Jiang, and C. He. 2012. "Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA." *Nat Chem Biol* 8 (4):328-30. doi: 10.1038/nchembio.914.

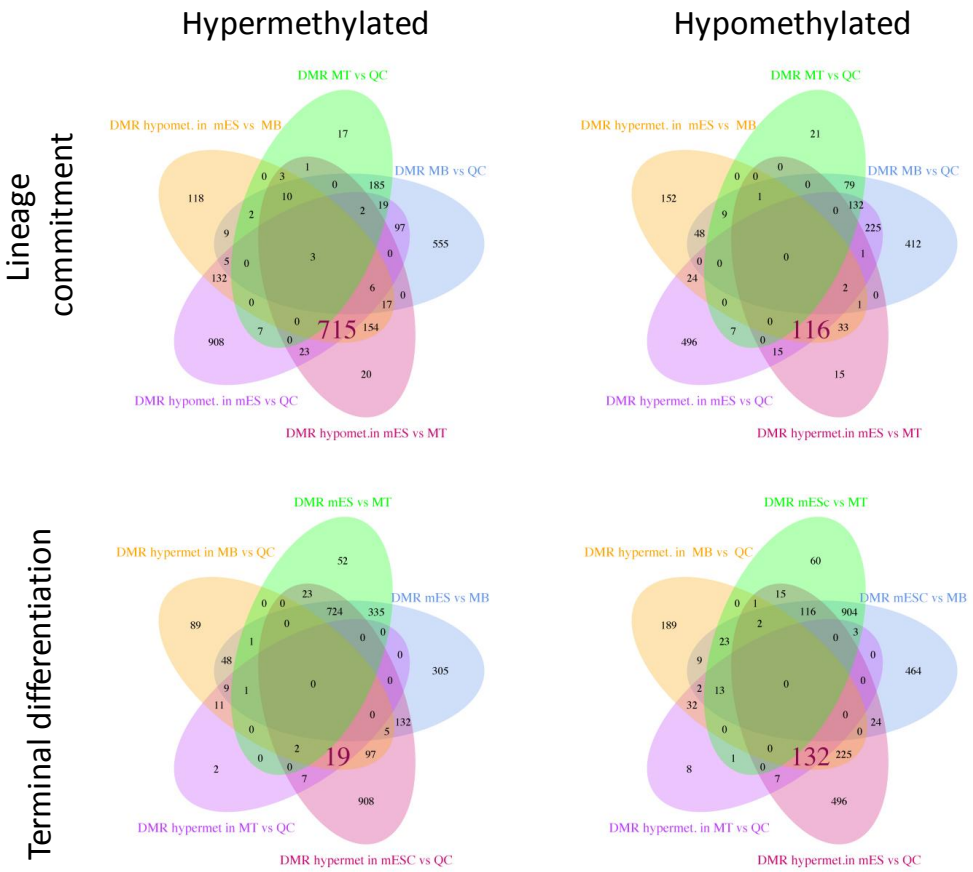
Zhang, R. P., J. Z. Shao, and L. X. Xiang. 2011. "GADD45A protein plays an essential role in active DNA demethylation during terminal osteogenic differentiation of adipose-derived mesenchymal stem cells." *J Biol Chem* 286 (47):41083-94. doi: 10.1074/jbc.M111.258715.

Zhang, R. R., Q. Y. Cui, K. Murai, Y. C. Lim, Z. D. Smith, S. Jin, P. Ye, L. Rosa, Y. K. Lee, H. P. Wu, W. Liu, Z. M. Xu, L. Yang, Y. Q. Ding, F. Tang, A. Meissner, C. Ding, Y. Shi, and G. L. Xu. 2013. "Tet1 regulates adult hippocampal neurogenesis and cognition." *Cell Stem Cell* 13 (2):237-45. doi: 10.1016/j.stem.2013.05.006.

Ziller, M. J., H. Gu, F. Müller, J. Donaghey, L. T. Tsai, O. Kohlbacher, P. L. De Jager, E. D. Rosen, D. A. Bennett, B. E. Bernstein, A. Gnirke, and A. Meissner. 2013. "Charting a dynamic DNA methylation landscape of the human genome." *Nature* 500 (7463):477-81. doi: 10.1038/nature12433.

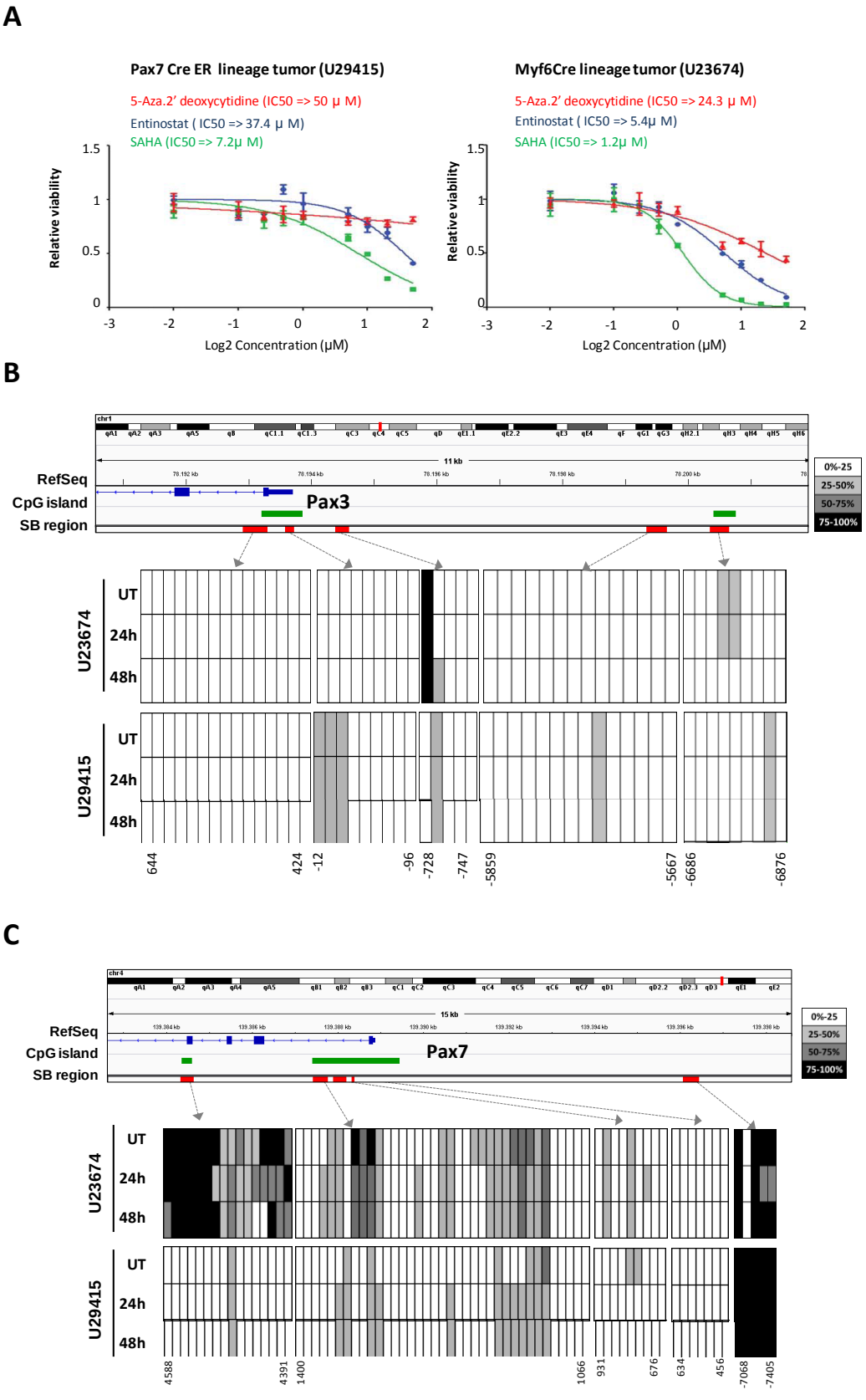
Ziller, M. J., F. Müller, J. Liao, Y. Zhang, H. Gu, C. Bock, P. Boyle, C. B. Epstein, B. E. Bernstein, T. Lengauer, A. Gnirke, and A. Meissner. 2011. "Genomic distribution and inter-sample variation of non-CpG methylation across human cell types." *PLoS Genet* 7 (12):e1002389. doi: 10.1371/journal.pgen.1002389.

APPENDIX I

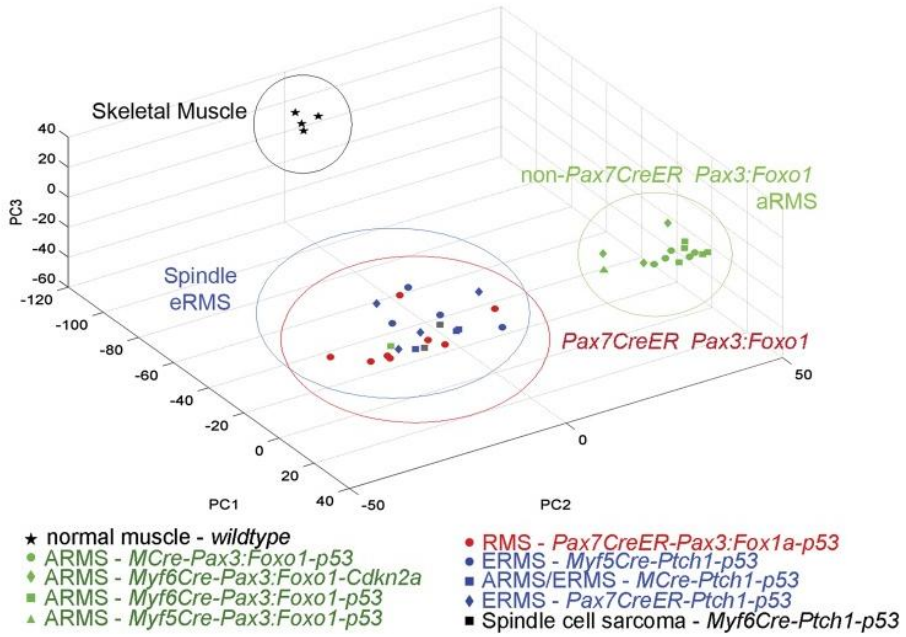


Appendix I Figure 1. Venn diagrams of the DMRs identified in all the comparison.

The Venn diagrams allowed the identification of non fluctuant hypo- and hyper- methylation changes at lineage commitment and terminal differentiation. The number of selected DMRs is indicated in bold letters. For example, to identify the hypermethylated DMRs during lineage commitment we selected the amplicons that showed increased number of reads in MB, MT and MF compared to ESC, and were stable maintained between MB, MT and MF. This restrictive selection ensured that the selected DMRs represented qualitative changes specific to each differentiation stage.



Appendix I Figure 2. Rhabdomyosarcoma cell lines treated with 5Aza 2' Deoycytidine. A Cell viability assay showing the effect of DNA methyltransferase inhibitor 5-Aza-2` deoxycytidine on U29415 and U23674. Data obtained from our collaborators (Abraham et al. 2014). DNA methylation analysis of Pax3 (**B**) and Pax7 (**C**) locus at U23674 and U29415cell lines after 5-Aza-2` deoxycytidine treatment.



Appendix I Figure 3. RMS cell lines and physiological skeletal muscle expression profiling. PCA showing the proximity of embyronal Rambdomyosarcoma cells bearing Ptch1 and p53 mutations (U37125) with Pax7CreER aRMS cell line (Pax7 Cre ER Pax3:FKHL-p53) (U23674) (Abraham et al. 2014).

Appendix I Table 1: AIMS-seq amplicon distribution by chromosomes (mm9 mouse genome)

chr.	Number of amplicons per chr.	% of amplicons per chr.
chr1	3107	6.123
chr2	3580	7.055
chr3	2760	5.439
chr4	3808	7.504
chr5	3524	6.945
chr6	2594	5.112
chr7	3224	6.353
chr8	2686	5.293
chr9	2308	4.548
chr10	2536	4.998
chr11	3330	6.562
chr12	2092	4.123
chr13	2050	4.040
chr14	1918	3.780
chr15	2217	4.369
chr16	1603	3.159
chr17	2181	4.298
chr18	1383	2.725
chr19	1301	2.564
chrX	2512	4.950
chrY	31	0.061

Appendix I Table 2: Positive and Negative amplicons

Classification across the gene-related distribution: number of amplicons, Fisher test p. adjusted values and percentages

Number of amplicons	Positive	Negative	Virtual
Exon	556	2098	2654
E/I junction	3306	3462	6768
Intron	6491	8176	14667
Promoter	1211	5503	6714
Proximal	913	839	1752
Distal	1552	1952	3504
Intergenic	6162	8524	14686
Total	20191	30554	50745

Positive vs Negative	Fisher test p-adjustet value		%		
Classification	Positive	Negative	Positive	Negative	Virtual
Exon	1.00E+00	1.79E-99	2.8	6.9	5.2
E/I junction	6.57E-01	1.00E+00	16.4	11.3	13.3
Intron	2.35E-38	1.00E+00	32.1	26.8	28.9
Promoter	1.00E+00	0.00E+00	6.0	18.0	13.2
Proximal	7.52E-26	1.00E+00	4.5	2.7	3.5
Distal	2.17E-08	1.00E+00	7.7	6.4	6.9
Intergenic	2.86E-10	1.00E+00	30.5	27.9	28.9

Classification across the CpG island distribution: number of amplicons , Fisher test p. adjusted values and percentages

Number of amplicons	Positive	Negative	Virtual
island	1432	13352	14784
shore	1544	1129	2673
shell	734	314	1048
Out side	16481	15759	32240

Positive vs Negative	Fisher test p-adjustet value		%		
Classification	Positive	Negative	Positive	Negative	Virtual
island	1	2.2E-16	7.1	43.7	29.1
shore	1.35E-82	1	7.6	3.7	5.3
shell	9.25E-89	1	3.6	1.0	2.1
Out side	0	1	81.6	51.6	63.5

Classification across the repetitive element distribution: number of amplicons , Fisher test p. adusted

values and percentages

Number of amplicons	Positive	Negative	Virtual
LINE	3963	11360	15323
LTR	2673	1582	4255
SINE	2580	607	3187
No repetitive	10975	17005	27980
Total	20191	30554	50745

Positive vs Negative	Fisher test p-adjustet value		%		
Classification	Positive	Negative	Positive	Negative	Virtual
LINE	0	1	19.6	37.2	30.2
LTR	1	7.04E-220	13.2	5.2	8.4
SINE	1	0	12.8	2.0	6.3
No repetitive	4.08E-3	1	54.4	55.7	55.1

Appendix I Table 3: Hypo- and Hyperylated DMRs

Classification across the gene-related distribution : counts of amplicons Fisher test p. adjusted values and percentages :

Counts of amplicons in in the gene-related distribution	Hypo.	Hyper.	Virtual
Exon	2	6	2654
Junction	39	86	6768
Intron	76	215	14667
Promoter	21	18	6714
Proximal	13	38	1752
Distal	23	68	3504
Intergenic	74	303	14686
Total	248	734	50745

DMRs	Fisher test p-adjustet value		%		
Classification	Hypo.	Hyper.	Hypo.	Hyper.	Virtual
Exon	0.85	1	0.806	0.817	5.23
Junction	0.0123	1	15.7	11.7	13.3
Intron	0.854	1	30.6	29.3	28.9
Promoter	0.000724	1	8.47	2.45	13.2
Proximal	0.854	1	5.24	5.18	3.45
Distal	0.854	1	9.27	9.26	6.91
Intergenic	1	0.00631	29.8	41.3	28.9

Classification across the CpG island distribution : counts of amplicons , Fisher test p. adjusted values and percentages:

Count of amplicons in in the CpG island-related distribution	Hypo.	Hyper.	Virtual
island	12	8	14784
shore	14	23	2673
shell	9	30	1048
Out side	213	673	32240
Total	248	734	50745

DMRs	Fisher test p-adjustet value		%		
Classification	Hypo.	Hyper.	Hypo.	Hyper.	Virtual
island	0.00365	1	3.74	1.26	0.136
shore	0.117	1	12	5.62	0.057
shell	0.915	0.918	0	0	0
Out side	0.996	0.0274	0	0	0

Classification across the repetitive element distribution : counts of amplicons , Fisher test p. adjusted values and percentages:

Count of amplicons in repetitve elements	Hypo.	Hyper.	Virtual
LINE	54	155	15323
LTR	46	179	4255
SINE	41	112	3187
No repetitive	107	288	27980
Total	248	734	50745

DMRs	Fisher test p-adjustet value		%		
DMRs	Hypo.	Hyper.	Hypo.	Hyper.	Virtual
LINE	0.59	0.88	21.77	21.12	30.20
LTR	0.98	0.14	18.55	24.39	8.39
SINE	0.59	0.88	16.53	15.26	6.28
No repetitive	0.59	0.88	43.15	39.24	55.14

APPENDIX II

Appendix II Table 1: Sodium bisulphite sequencing primers

Gene	Region	First PCR		Nested PCR	
		Forward	Reverse	Forward	Reverse
Ckm	Promoter	GGTTTTTTTAGTTGTTAATGT	ACCCCAATAATTTCTCTCAA	TTTGGGGGTTAGGGTTTA	TATCTATACTATAAAAATAATAA
Dppa4	Promoter	GTTTGGAGAAGAGAAGAAT	AATATTTAATATTTTATTT	GGAGAAGAGAAGAATGAG	AATTAATAAATTTATTATAT
mir302	Promoter	TTTAATATGTAGATGTATTT	CACCCACAAAACAACTAA	AGAAGTATAATGTATTTTTTGG	ATAAACTTTACCTCCTTTAC
Myf5	-17 kb enhancer	TAAATGTGGTTAATGGTTAT	ACCTCTTAAAAAAATAATATA	TTAAATAGTAAGTTTATGTAATA	CTCTCTCTTTTTTCCTAAT
	-111 kb enhancer	GTATGTTAGTATTATTTTT	CAAACTCACCTTTCAAC	TATAGTGTGTGATAAGTTAT	ATAACCCTAACCCACATAA
	Control -118 kb	AGAATAAGAGGATTAAGATTTT	AACTAAAACACACACTCAAA	GTTTTTTAAGAAAGATTTTTAT	AATTTTTTTTTACATCAACAC
	Controll Down stream	TA GGA AGT AAT AGA AAG TAA	TTCAAAAAATAATACCACTAA	AAGTAATAGAAAGTAATGATTAA	CCTAAAAAAAATATACTACCAAA
	Control -4.4 kb	TTGTAAAGATTGAATAAAAT	TACTATACATCCAAAAAAAT	TAAGGAAAGATTAGGAGTTG	AAAAATACAAACATCTATAAT
	Intergenic enhancer	TATTTTGTTTAGTTTGTTTTT	AATCTTCCACACCTCTAC	TTTTTTTAGTAGGAGTGATTAT	TTTACATCAATCCCTCTATA
	Promoter	TAGGAGTTTTTATAGAAGTATTT	AACCAACCCCAACCCCTT	AGTATTTAGAAGGGGAGAA	TTTATCCAAAAAACCACCA
	Proximal Arch enhancer	GATAAAGATTAAAGAAGTAAAG	AACAACTAAAACCCATCTATAA	GATGAGAGAGATGGTTATTT	AAACTTTCTATAAAAAACTAAAC
	Early epaxial enhancer 3'	TGGTAAAATAGAGTTGTTTGT	ACTACTTCAATAAATAAAAAAAT	TTTGTAGAGTGTAGTGTTTGT	ACATAAAACCTCTTAACTAAAC
	Early epaxial enhancer 5'	TTAGGATTTTTATATATTTTAT	CTCACAAATACCAACTTTT	GTTTTAAGATTATTTATA	AACTTCTCACCACCTCAA
Myf6	Promoter	AGTAATGGTTATTGTTTATGA	TAAAACTCAATCCAACTTCTAA	TATTGGAAATATTAATGAGG	TAACTAATCTAATCTAAATAAT
Myh1	Promoter	GGAGGAAGTAATAGTTGTT	ACACTCTTACCTTAAAACTT	GTTATATTGAGGAGTAGAAAG	TAAAAAAAAACAAATCACTCTC
Myh4	Promoter	TTAGTTATAGTGTTAGATTTA	CACCCCAACTTCACTTTT	AATGGTTTTAAGTATTAGTAGA	CCTCAACTATCCTAACTAC
Myh8	Promoter	TTTAAGGGAATGTAGTGTGG	ATCACTTACCTCTAAACTCTT	GAATAAATAAGGAAAGTGAGTT	AATAAAATATAAAAAACAACCCTA
MyoD	Distal	ATTGAGAGTTAGGTAGGG	AACCAACTCACTTTCTCC	TTATAATATAGTTAGTTGGGG	CTAACCTCTCATACCTAATA
Myog	Promoter	AAGAGAAGGTTAAGTGGA	AAACCCTAAAAATAAACAAAAAA	TTGGATTATGGAGGAGAGA	AATATCTCATACAACTCC
Nfix	+55637	GGAATAAAGGTTAAAGGATTA	CTAATACACCAATACATATCA	GTAAGGTAGTGAAAGTAGG	CTAATACACCAATACATATCA
	+7231 +7510	GTGATAGGTTTTTGTTGG	TCTCTCTTCCTTCTTTCTA	AGGGTATTTTTGGGGTTA	TCCTTCTTTCTAACTCTAA
Pax3	+7587 +7454	TTTAGATGAAAGTGAGAGG	TTCCTACTCCAACACCCTAA	GATATAAAGGAGGTAATTGAGT	ACACCCTAATAAAAAAAAAAAATA
	+644 +424	AGGTGAATAGAAAGAGAAT	ACTACCCCCAAAATAAC	AGGTGAATAGAAAGAGAAT	ACTATACCCAAAATAATACAA
	-96 -12	TTAAAAAAGGTTAGAGAGGG	AACTCAAAACTCCTAATCAA	TAG TTG GGG TTA TTG GTA AA	TAAAAAAACTAATAAATACTCC
	-728 -747	TGGTATAGTTATGATAGAGA	AATCTTTACAACACTCCTAA	TAAGTTGTAAGTAATGGGGA	AACACTTCAACTCCTAACC
	-5859 -5667	TATAGTGGAATAAGTTTATATT	CTTACCTCTTCTCCTAA	GAGAAGTAGGGTAGAGTT	CCAAAAAAAAACCAAAAATCA
	-6686 -6876	GAAAGAATGAAGGTATAT	TCTCCCACAATAAAAACTA	GTATTGTTTAGAGAATATTTT	AATAAAAAAACCTTAAAAC
	-18774 -19025	GGGGTAATAGTTAGTTAGT	ATAACCTAAAAATATACAAAAAA	AGAAGATGAAAATGATTGTTTG	ATTCCTAAAACCTAAAACTTAAA
	-20296 -	GAGTTGTAGTTGGTTTAA	ATACCTAAATAAAAAAAATTAA	GTAGAAGGTGTGTAGGTA	AAATAAATAAACCAAAAAACC
Pax7	+1400 + 1066	TAAAAGTAGGTATTTGAAATTA	CCCCTCTTATCATTATAT	TAAAAGTAGGTATTTGAAATTA	CTAAAAAAAAAACCCCCTC
	+931 +676	ATATAATGATAAGAGGGG	TTCCACTCCACCAAAC	GTTTTTAATTGTTTTTGAGATAT	AATACATAATACCTTATTTCCc
	+634 +456	GGAAATAAGGTATTATGTAT	TCTACTAAATCCCAATCTC	GGAAATAAGGTATTATGTAT	CTCTTCTCTAACACAACA
	+4588 +4391	GAGGTATAGGATTGTGTTA	CAACTTATACTCTCCCTTT	GTAGTTTTATTTTTAGGATTTG	CCTTTTACCTTTCATTTCTAA
	-7068 -1066	TTTAAGTTTTTTGTAAGAG	ATTCTCTCCATACCCATTAA	GTAAGTTATTAAAGATAAAAATA	TTAAAACAAATCAAAAAATACC
Pou5f1	Promoter	AAGTAAGAATTGAGGAGTGGT	CAATCCCACCCTCTAACCT	GGTTTTAGAAATAATTGGTATA	AAACCTTCTATATCCATCTAT
Myf5 (human)	Early epaxial enhancer	TAATTTTTTAGATTAGGTTATA	TTATATACATCATTCCACAT	AAG TTG TTG AAA GGG AAG GA	CAAAAATACATTAAATACCC

Appendix II Table 2: Gene expression primers

	Forward	Reverse
18S	TTGACGGAAGGGCACCACCAG	GCACCACCACCCACGGAATCG
Apobec2	GACCCTGAGAAGCTGAAAGAG	CGACCACATAGCAGAGAAAGG
Ckm	AGGCATGGCCCGAGAC	AGATCACGCGAAGGTGGTC
Dnmt1	CTCAGGGACCATATCTGCAAG	GGTGTACTGTAGCTTATGGGC
Dnmt3a	GGTCATGTGGTTCGGAGATG	AGGACTTCGTAGATGGCTTTG
Dnmt3b	GTACCCCATCAGTTGACTTGAG	TTGATCTTTCCCCACACGAG
Gadph	ACTCCCACTCTTCCACCTTC	TCTTGCTCAGTGTCCTTGC
Myf5	GCCATCCGCTACATTGAGAG	ACAGGGCTGTTACATTCAGG
Myf6	ATGGACCTTTTTGAAACTGGCTCC	CTGGCCAGGGCAGTGGGGAGGCTG
Myh1	CTCCAGGCTGCTTTAGAGGAA	CCTGCTCCTAATCTCAGCATCC
Myh4	AAACCACCTCAGAGTTGTGGA	CTTCCGAAGGTTCCTGATTGC
Myh8	AACAGAAACGCAATGCTGAGG	TCGCCTGTAATTTGTCCACCA
MyoD	GCCGCCTGAGCAAAGTGAATG	CAGCGGTCCAGGTGCGTAGAAG
Myogenin	GGTGTGTAAGAGGAAGTCTGTG	TAGGCGCTCAATGTACTGGAT
Nfix	ACATCAAACCACTGCCCA	GTCCGCCAGTGAGAAGTTG
Nfix (isoforms discriminative)	ATGGCCCAATGACGTGG	TCCGATGCTGACAAACCG
Pax3	GAGCGAAGCTGCCCCCAG	GCCGTTGATAAATACTCCTCC
Pax3-Fkhr	AGACAGCTTTGTGCCTCCAT	CTCTTGCCTCCCTCTGGATT
Pax7	CAGGAGACTGCGTCCATCCG	CCGAACTTGATTCTGAGCAC
Pou5f1	GAGGAGTCCCAGGACATGAA	AGATGGTGGTCTGGCTGAAC
Tdg	TTCCTAACATGGCAGTCACG	TCTGGGTTTCCTTTTCCTTCTC
Tet1	GACCGAAGATGTACCCTCAAC	CCTCCCAAACTTACAGCCG
Tet2	AACCTGGCTACTGTCATTGCTCCA	ATGTTCTGCTGGTCTCTGTGGGAA
Tet3	TCCGGATTGAGAAGGTCATC	CCAGGCCAGGATCAAGATAA
Usf1	GTCTTCCGAACTGAGAATGGG	GAAAGCTCCCTGGATCACTG

Appendix II Table 3: ChIP assay primers

Gen	Region	Forward	Reverse
Myf5	**Early epaxial enhancer**	ACAGCCCAAGGAGAAAAGTG	GCAAAGTTTTACATCAGTCCCTC
	17 kb enhancer	TGTTAACTGCGGTCTGATCAC	GAAGGTACTGTCTGAAAGGGAAG
Nfix	**DMR**	GGCATCTGGAAAAGTTTGGG	TTGCTTTCTGTCCCCACTC
	US	CCAGAAAACCAGCAAAACCC	CAAATGAGAGTACCCTTCCTGG
Control region	**118 kb**	ACAGTAAACAGAGTCAGTGTGTG	TCGAGGGTCCTGTCTGAC

Appendix II Table 4: GEO Accession numbers

Samples	GEO Accesion number
ESC ChIP seq	
H3K4me1	GSM769009
H3K4me3	GSM769008
H3K27me3	GSM1000089
H3K27Ac	GSM1000099
p300	GSM918750
Pol II	GSM918749

MB/MT ChIP seq	
Asp et al, 2011 (H3K4me1)	GSE25308
Blum et al, 2012 (H3K27Ac, p300)	GSE37525
H3K4me3 MB	GSM918415
H3K4me3 MT	GSM918416
H3K27me3 MB	GSM918408
H3K27me3 MT	GSM918414
Usf1 C2C12 MB	GSM915162
Usf1 C2C12 MT	GSM915161
Pol II C2C12 MT	GSM915176