Università degli Studi di Milano
Scuola di Dottorato in Medicina Molecolare

Curriculum di Genomics, Proteomics e Tecnologie Correlate

Ciclo XXV

Anno Accademico 2011/2012

Dottorando: Paola CATTANEO

Matricola: R08659

EPIGENETICS DURING HEART DEVELOPMENT: THE ROLE OF THE HISTONE METHYLTRANSFERASE DOT1L IN CARDIAC COMMITMENT

Direttore della Scuola: Ch.mo Prof. Mario Clerici
Tutore: Prof. Mario CLERICI
Correlatore: Prof. Gianluigi CONDORELLI

Choose your own destiny,
Be whoever you want to be.

SOMMARIO

Il cuore è il primo organo a formarsi e a funzionare durante lo sviluppo embrionale, e tutti gli eventi successivi nella vita dell'organismo dipendono dalla sua funzione.

La morfogenesi del cuore e lo sviluppo delle diverse linee cellulari cardiache sono processi molto complessi che si basano sull'interazione di network di fattori di trascrizione che attivano geni cardiaco-specifici. Mutazioni dei fattori di trascrizione cardiaci, dei geni che li regolano e dei geni che questi controllano, risultano in molti difetti cardiaci, evidenziando l'importanza dello studio delle basi molecolari alla base di questi processi.

Le alterazioni chimiche a carico del DNA e degli istoni, note come modifiche epigenetiche, sono sempre più studiate per la loro importanza durante l'organogenesi. Recentemente diversi studi hanno riportato l'importanza delle modifiche istoniche durante il differenziamento cardiaco in vitro, distinte conformazioni della cromatina sono state associate all'espressione di geni importanti per il differenziamento cardiaco. Tuttavia, nonostante il crescente interesse a riguardo, il ruolo degli enzimi che catalizzano queste modifiche rimane ancora poco compreso e in modo particolare durante la cardiomiogenesi in vivo.

In questo studio è stato dimostrato che l'espressione della metiltransferasi DOT1L porta a un cambiamento della distribuzione sul genoma della dimetilazione di H3K79, e che l'azione di questo enzima è fondamentale per la corretta espressione dei geni cardiaci durante il differenziamento dei cardiomiociti.

L'espressione di DOT1L aumenta nei cardiomiociti allo stadio embrionale e neonatale rispetto a cellule staminali embrionali indifferenziate e a cardiomiociti adulti, inoltre, H3K79me2 correla con l'attivazione trascrizionale caratteristica dei cardiomiociti durante il differenziamento. Inoltre i loci dei geni espressi solo in fasi tardive di sviluppo presentano un arricchimento di questo marker attivatorio anche nelle fasi precoci dello sviluppo, suggerendo un ruolo per H3K79me2 nella pre-attivazione di questi geni. I risultati emersi da questo studio dimostrano come la metilazione degli istoni, e in particolare H3K79me2, regola la trascrizione genica nello sviluppo di cardiomiociti e che DOT1L svolge un ruolo centrale in questo processo. Oltre ad aumentare la comprensione dell'epigenetica nel controllo dello sviluppo cardiaco, questa mappa genomica di H3K79me2 potrebbe portare all'identificazione di nuovi geni e a comprendere nuovi meccanismi di regolazione della trascrizione coinvolti nel differenziamento cardiaco.

Nel complesso questo studio illustra l'importanza della regolazione epigenetica nelle prime fasi dello sviluppo al fine di stabilizzare l'espressione genica cardiaca e di delineare il destino di una cellula. È stata aggiunta un'informazione importante al complesso processo di attivazione trascrizionale dei cardiomiociti. Per la prima volta è stata definita una mappa ad alta risoluzione di H3K79me2 nei cardiomiociti, che potrebbe essere il punto di partenza per prevedere nuovi pathway trascrizionali durante il differenziamento cardiaco, nonché potrebbe fornire l'opportunità per identificare nuovi geni utili per comprendere nuovi programmi di sviluppo e per identificare meccanismi alla base dei molti difetti congeniti cardiaci e di malformazioni dello sviluppo ancora sconosciuti.

ABSTRACT

The heart is the first organ to form and function in the embryo, and all subsequent events in the life of the organism depend on its function. Cardiac lineage specification and subsequent morphogenesis of the early developing heart are complex processes that rely on networks of interacting DNA-binding transcription factors and targeted activation of cardiac-specific genes. Mutations in cardiac transcription factors, the genes they regulate and the genes that regulate them, result in many inherited congenital heart defects and point to the importance of understanding the molecular basis behind these processes.

Chemical alterations on DNA and histones, known as epigenetic modifications, are being increasingly studied for their importance in organogenesis, such as that of the heart. Recently, a dynamic landscape of histone modifications has been reported to occur during cardiac differentiation in vitro: distinct chromatin patterns were associated with stage-specific expression of genes functionally relevant to the heart. However, despite the growing number of reports, the role of the enzymes that catalyze these modifications remains poorly understood in cardiac differentiation in vivo. Here, we show that a definite temporal expression pattern of DOT1-like histone H3 methyltransferase (DOT1L) drives a transitional pattern of H3K79 di-methylation in the genome of differentiating cells, and that the function of this enzyme is obligatory for the correct differentiation of cardiomyocytes. In fact, we found that expression of DOT1L was increased in ex vivo embryonic and neonatal cardiomyocytes with respect to undifferentiated embryonic stem cells and adult cardiomyocytes; moreover, H3K79me2 was highly correlated with transcriptional activation in differentiating cardiomyocytes. We also found that the loci of genes expressed only in later stages of development were enriched in this activating mark, suggesting a role for H3K79me2 in the pre-activation of genes. Our results demonstrate how histone methylation, and in particular H3K79me2, regulates the transcription in developing cardiomyocytes and the central role played by DOT1L in this process. Apart from the increase in our understanding of how epigenetics controls development, our genome-wide data on H3K79me2 could lead to the identification of novel genes and transcriptional regulatory networks involved in cardiac differentiation.

Altogether our study illustrates the importance of epigenetic regulation early in development to delineate the fate of a cell and in particular the role of methylation of H3K79 in cardiomyocytes stabilizing the signature for cardiac gene expression. Furthermore, this study add an important information to the intricate process of transcription activation that make a cardiomyocyte; for the first time we built an high resolution map of H3K79me2 in cardiomyocytes that could be the starting point to predict novel transcriptional regulatory networks during cardiomyocyte differentiation, as well provide the opportunity to identify novel genes that might be informative to understand developmental regulatory programs. Indeed we shed light on an additional fundamental enzyme involved in defining the epigenetic code associated with the complex process of heart development and establish a platform useful to identify new mechanisms underlying many congenital heart defects and cardiac developmental malformations.

LIST OF ABBREVIATION

ATP - adenosina trifosfato

BAF – Brg1/Brm associated factor

CBP – Creb Binding Protein

CF – Cardiac Fibroblast

CHIP-SEQ – Chromatin Immunoprecipitation Sequencing

CMa – Adult Cardiomyocyte

CMe – Embryonic Cardiomyocyte

CMp – Postnatal Cardiomyocyte

CRF – Chromatin Remodelling Factor

CTRL - Control

DMEM – Dulbecco's Modified eagle

DNA - Deoxyribonucleic Acid

DNMT - DNA methyl transferase;

DOT1L – Disruptor of telomeric silencing 1 like

E14.5 – Embryonic Day 14.5

EB – Embryoid Bodiy

EMT – Epithelial Mesenchymal Transition

FACS – Fluorescence Activated Cell Sorting

FBS – Fetal Bovine Serum

FDR – False Discovery Rate

FHF – First Heart Field

GATA4 - GATA binding protein 4

GO – Gene Ontology

H3 – Histone 3

H3K27me3 – Trimethylation of Lysin 27 of histone 3

H3K4me3 - Trimethylation of Lysin 4 of histone 3

H3K79me2 - Bimethylation of Lysin 79 of histone 3

H3K9me3 - Trimethylation of Lysin 9 of histone 3

H4 – Histone 4

HAT – Histone Acetyl Transferase

HDAC - Histone Deacetylase

HDM – Histone Demethylase

HMT, Histone Methyl Transferase;

HOXB5 - Homeobox protein Hox-B5

ISL1 – Islet 1 transcription factor,

KO – Knockout

LIF – Leukemia Inhibitor Factor

MEF – Mouse Embryonic Fibroblast

MEF2C – Myocyte Enhancer Factor 2C

mES – Mouse Embryonic Stem Cell

MESP – Mesoderm Posterior 1 Homolog

MLL - Mixed Lineage Leukemia;

Nkx2.5 – NK2 transcription factor related, locus 5 (Drosophila)

OFT - Outflow tract

P/S – Penicillin/Streptomycin solution

PCR – Polimerase Chain Reaction

RNA – Ribonucleic Acid

SET - Su(var), Enhancer of zeste, Trithorax;

SHF – Second Heart Field

shRNA – short herpin RNA

T – Brachyury T

TNNI3 – Troponin I type 3

TSS – Transcription Start Site

WB – Western Blotting

α-MYH – alpha Myosin Heavy Chain

β-MYH – beta Myosin Heavy Chain

TABLE OF CONTENT

1. INTRODUCTION

1.1 Heart Development

The heart is the first organ to form and function in the embryo, and all subsequent events in the life of the organism depend on its function. Abnormalities in heart formation, the most common form of human birth defects, afflict nearly 1% of newborns, and their frequency in spontaneously aborted pregnancies is estimated to be tenfold higher [1, 2]. The proper formation and function of the heart are essential for embryonic survival, and defects in heart formation cause significant morbidity and mortality in prenatal or postnatal life.

During the past decade, our understanding of the development of the embryonic heart has been improved by a number of discoveries. This concerted effort has produced ground breaking achievements, allowing us to have nowadays a better understanding of the genetics of congenital heart disease, innovative genetic tests for diagnostic purposes and a much deeper comprehension of critical events and signal transduction pathways involved in the cardiogenic process [3, 4].

1.1.1   Anatomy of the heart during development

The developing heart arises from lateral plate mesoderm early in embryogenesis and begins to emerge shortly after gastrulation. The mesoderm, one of the three germ layers, can differentiates into many discrete cell types and organs, including bone, blood and the three types of muscle: smooth, skeletal and cardiac.

Following rapid specification and initiation of differentiation into the specific cell types that will form the heart, a series of complex morphogenetic events ensues.

From an anatomical perspective, it is now well established that pools of progenitor cells arising from at least four distinct embryonic populations orchestrate on a series of finely regulated migratory events to give rise to what is the first functional organ in the mammalian embryo [3, 5].

Following gastrulation, cardiac progenitors reside within the anterior lateral mesoderm in a bilaterally symmetrical region to form "the cardiac crescent". During ventral morphogenesis, the most lateral portions of the cardiac crescent are brought together at the midline to form "the linear heart tube" (Figure 1). At this stage the heart begins its slow peristaltic contractions that will later become more rhythmic and regular. The cardiac progenitors that merge in the ventral embryonic midline to form the linear heat tube can be generally divided into two distinct fields: the first heart field (FHF) that will give rise to the atria and to the left ventricular and the second heart field (SHF) that will generate the outflow tract, the right ventricle, the inter-ventricular septum and part of the atria [6]. Consecutive mitotic rounds from the second heart field progenitors generate new cardiac progenitors that are permanently added to the venous and arterial poles of the heart. As a consequence of this permanent cell migration, localized proliferation and stimuli resulting from blood flow, the heart tube starts to undergo a rightward looping process that constitutes the first step for chamber morphogenesis. This looping is concomitant with the establishment of an anterograde unidirectional blood flow and immediately precedes the incorporation of other two sources of cardiac progenitors – the epicardium and the cardiac neural crest [7].

The following processes involve the establishment of the atrial and ventricular septa and the development of bi- and tri-cuspid valves from the endocardial-derived mesenchymal cushions that originated the totally divided four chambers (two atria and two ventricles).
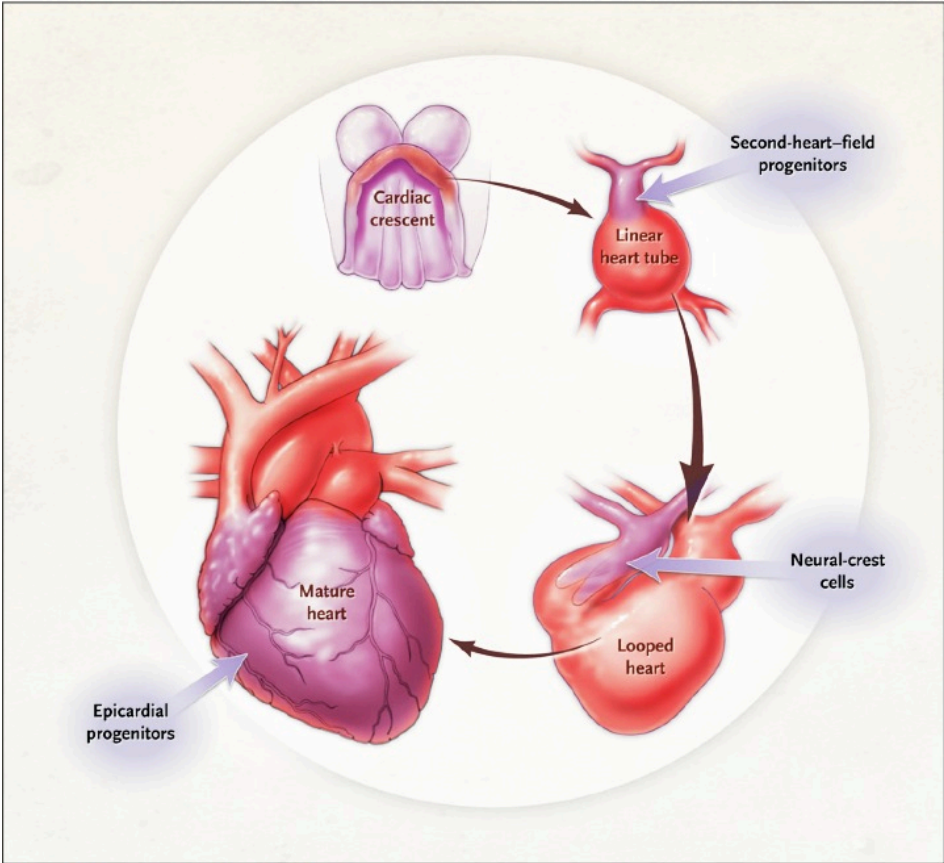
Figure 1 - Heart Development.

At this stage cardiac development exits a cycle dominated by the occurrence of critical morphogenic events to enter a maturation stage in which the coronary circulatory system is formed and in which the cells conclude their differentiation into the lineages they had previously committed to [8].

1.1.2 Lineage specification and cardiac cell types

Heart development is an elaborate process requiring cell specification, cell differentiation, cell migration and interactions among cells from several embryonic origins.

Cellular differentiation involves a complex sequence of events that progressively commits a cell to a defined lineage. It is now clear in fact that precursor cells in the embryo have the potential to differentiate into the various types of cardiac cells, however the potential of differentiate into various types of cells become progressively restricted after lineage decisions. The specification of the cardiovascular lineages involves a transition through a sequence of increasingly restricted progenitor cells, proceeding from a pluripotent state to mesoderm and then to cells committed to cardiovascular fates [9].
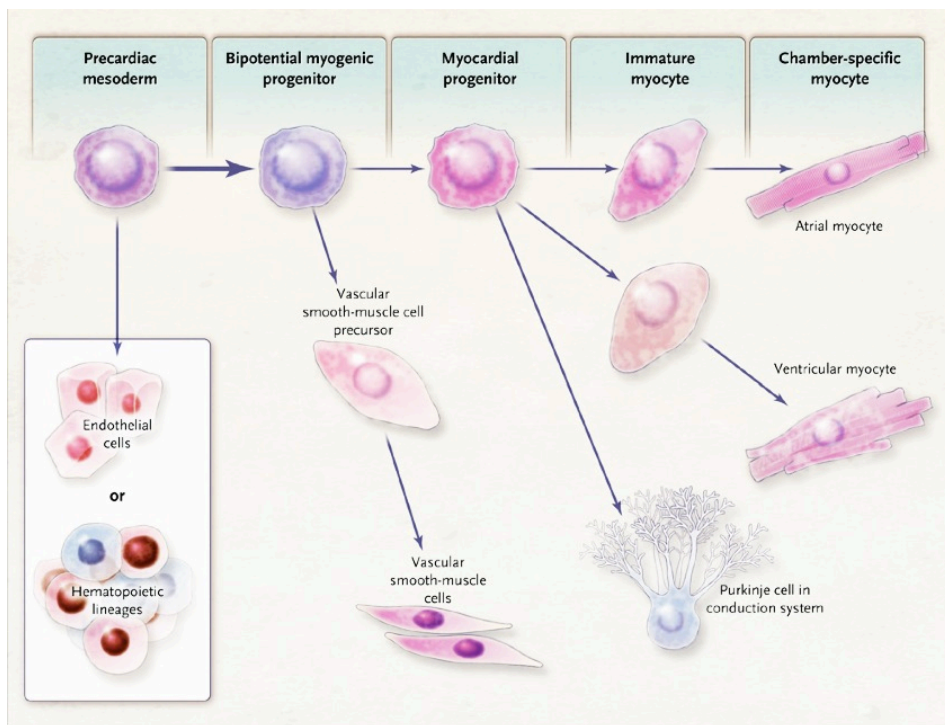
Figure 2 - Cardiac lineage specification

In the mesoderm lineage, decreased activity of pluripotency factors is accompanied by increased activity of lineage-specific transcriptional activators such as Brachyury T and MESP [10].

Mesodermal cells have the potential to differentiate into hematopoietic lineages, endothelial cells, epicardial cells as well as bipotential myogenic progenitors.

Epicardial cells constitute the outermost protective epithelial layer of the heart, dividing the myocardial wall from the fluid filling the pericardial cavity. In embryonic development epicardial cells originate from the mesenchyme of the septum transversum. After creating a continuous monolayer of epithelial nature, a subset of these epicardial cells undergoes epithelial-mesenchymal transition (EMT) and invades the myocardial wall, contributing with several distinct cellular identities to the developing cardiac tissue: endothelial cells, vascular support cells (pericytes and vascular smooth muscle), fibroblasts, and, more controversially, a subset of cardiomyocytes [11-13]. Epicardial cells are readily distinguishable from the adjacent myocardial cells as they exhibit strong expression of epicardial-characteristic genes: TCF21 or Epicardin, WT1 and Tbx18.

However, cardiac progenitor cells segregate into two distinct populations referred to as the primary and secondary heart fields that express unique markers: Isl1 is involved in the differentiation of secondary heart field cells, whereas the Nkx2.5 is a marker of both heart fields. The pool of potential cardiac progenitor cells are involved in continual maintenance of the heart by differentiating into several types of cardiac cell including endothelial cells, vascular smooth muscle cells, conduction cells and ventricular and atrial cardiomyocytes.

Cardiac conduction cells develop and form the cardiac conduction system of the heart. Electrical impulses are propagated through the heart by the cardiac conduction system and by direct cell-cell coupling of cardiac myocytes. Important component of conduction system include Sinoatrial Node, Atrioventricular Node and Purkinje fibers. These cells derive from mycardial progenitor and are specialized myocardial fibers, modified cardiac myocytes, not neurons, which conduct an electrical stimulus that enables the heart to contract [14].

Even if non-myocytes cell type observed in the heart (fibroblasts, endothelial cells, pericyes, vascular muscle cells) make up over 75% of the total number of the cells in the heart, approximately 75% of the total volume of the heart in mammals is make up of cardiac myocytes. Unlike skeletal muscle cells with a analogous, striated appearance, cardiac myocytes have widespread, branching interconnections. Myofilaments are organized into sarcomeres, the fundamental contractile unit of striated muscle cells. During contraction, thick and thin filaments slide past one another resulting in reduced sarcomere length, thus, providing the pumping action of the heart. From immature myocyte, cardiac cells differentiate into chamber specific cardiomyocytes with defined morphology and function; atrial myocytes are striated but much smaller than working ventricular myocytes that are also multinucleated.

In cardiomyocytes two ventricular myosin heavy chains, α-MYH and β-MYH, which exhibit different levels of ATPase activity, are differentially expressed during cardiac development. The β-MYH gene, predominantly expressed in late fetal live, is located 4kb upstream from the α-MYH gene, that is predominantly expressed in the adult. Both types of MYHs, α-MYH and β-MYH, are almost always expressed together in the normal animal, although their respective level varies during development. During heart development in mice and rat the ventricular β-MYH mRNA, dominant in the late fetal life, is almost completely replaces after birth by α-MYH mRNA which is dominant in the adult, leading to parallel changes in the MYH

isozyme distribution [15]. Interestingly, the expression of these genes is not limited to the ventricular cells since α-MYH mRNA is accumulated in the atrial myocardium whereas β-MYH mRNA is present in slow-twitch skeletal muscle fibers [16].

In the myocyte the essential contractile structure that contain the components of the contractile apparatus is the sarcomere. The sarcomere is composed of thick and thin inter-digitating filaments. The fundamental proteins of the contractile apparatus are myosin, actin, tropomyosin, and the troponin complex. In the presence of increased extracellular Ca2+, interactions occur between these proteins, causing the hydrolysis of ATP and changes in physical-chemical dynamics. These processes result in the development of tension within the myocyte. Myosin, the thick filament, is composed of a filamentous tail and a globular head region [17]. This globular head contains the site for actin binding, as well as a catalyzing site for ATPase activity. Actin is the major contractile protein found in the thin filament. Tropomyosin is another protein found in the thin filament. This rigid molecule lies on either side of actin, adding rigidity to the thin filament. Tropomyosin influences actin-myosin cross-bridge formation by physically inter-digitating between the actin-myosin cleft, thus preventing Ca2+ binding [17].

The troponin complex, also present in the thin filament, is formed of three proteins: troponin T, I, and C. Troponin is an key component in that it regulates the extent of cross-bridge formation, as well as contributing to the structural integrity of the sarcomere. Troponin T binds the troponin complex to tropomyosin and anchors the complex to the thin filament. Under normal conditions, phosphorylated troponin I weakens the affinity of Ca2+ for troponin C. Ca2+ binding to troponin C results in a conformational change of the complex, with subsequent actin-myosin interaction, thus initiating cross- bridge formation.

Indeed the cardiomyocyte is one of the most structurally elegant and functionally complex cells in nature. Understanding the molecular mechanism and the transcriptional program that outline lineage commitment and cardiomyocyte specification and function could be important to explain potential defect in cardiac developmental biology, to diagnose potentially lethal conditions but also a potential perspective for tissue engineering and regenerative medicine owing to the extremely reduced regenerative potential of cardiac tissue.

1.2 Transcriptional control of cardiogenesis

1.2.1. Transcription regulation

In the nucleus the DNA-protein complex that forms a chromosome is called chromatin The packaging of chromatin take place in an well-ordered modus in this way the genes present on the DNA are available for the transcription or replication. Hundreds of different cell types play a whole range of specialized functions that depend on the genes expressed and which are activated only in one type of cell. The cells of an organism are not different because they contain different genes, but because they express the genes differently.

The cell can control which proteins produced in various ways: by controlling when and how often a gene is transcribed, checking how occur the process of maturation of the primary transcript RNA at the level of cutting and welding, choosing which mRNA the ribosomes translate, and selectively activating or inactivating proteins already produced. The transcription control is exerted generally in the stage of

initiation of transcription. The promoter region of a gene attracts the enzyme RNA polymerase and directs it properly, so it can make a copy the gene to RNA. The promoters of eukaryotes include an initiation site, and a sequence of about 50 nucleotides that extends upstream from the initiation site [18-20].

This region contains sites necessary for RNA polymerase to bind to the promoter. Almost all the genes have, in addition to the promoter, regulatory sequences of DNA that serve to activate and deactivate a gene. The regulatory sequences don't work independently; to work proteins called regulatory proteins of genes that bind to DNA must recognize them. Within the promoter is located a short DNA sequence that is recognized by a regulatory protein. When the protein binds to this nucleotide sequence blocks access of RNA polymerase to the promoter, which prevents transcription of harmful enzymes [18-20].

Eukaryotic cells have 3 RNA polymerase enzymes that are not capable of initiating transcription without the aid of additional proteins. They require the intervention of a large group of proteins called transcription factors, that be must associate with the polymerase to the promoter site, because the enzyme can start to transcribe. The regulatory proteins may affect the initiation of transcription positioning itself vary distant. This means that a single promoter can be controlled by an almost unlimited number of regulatory sequences scattered along the DNA. Anyway the start of transcription must take account of a DNA organized into nucleosomes and thickened in compact chromatin structures.

Almost all eukaryotic promoters also require activator proteins that facilitate the association between RNA polymerase and transcription factors. The nucleosomes are likely to be present in the promoter regions, when activated gene transcription, these nucleosomes are moved. The option to turn on and off many genes with only one protein is not only in the regulation of cell function: it is also one of the means by which eukaryotic cells differentiate into various types during embryonic development.

1.2.2 Cardiac transcription factors

Intensive studies have revealed numerous genes that control the intricate process of heart development in humans and mice [21, 22]. Most of the inherited forms of congenital heart disease are a result of mutations in cardiovascular transcription factor genes. Cardiac lineage specification and subsequent morphogenesis of the early developing heart are a complex process that relies on networks of interacting DNA-binding transcription factors and targeted activation of cardiac-specific genes. Mutation in cardiac transcription factors, the genes they regulate and the genes that regulate them result in many inherited congenital heart defects and point to the importance of understanding the molecular basis behind these process.

At a molecular level, complementary studies revealed a vast list of genes with critical function in cardiac development. From this indeed transcription factors acquire special importance as components of a regulatory core responsible to make a cardiomyocyte and, simultaneously, to orchestrate the morphogenetic events with the non-myocyte cell fraction of the heart – endothelial cells, vascular support cells, epicardial cells and fibroblasts.

Entry of cells into the cardiac lineage in response to the appropriate signals is coupled to the expression of a set of transcription factors that initiates the program

for cardiac gene expression and drives the morphogenic events involved in formation of the multichambered heart [23].

Members of the GATA, Nkx, T-box (Tbx), MADS box, basic helix-loop-heliz, Iroquois (irk) and Forkhead (Fox) families interact and have critical transcriptional regulatory roles at distinct stages of heart development in multiple cardiogenic processes.

The earliest expressed transcription factors that initiate cardiac fate are the homeobox transcription factor Nkx2-5 and members of the GATA (GATA Binding Protein) family of zinc finger transcription factors, GATA4, GATA5 and GATA6. (GATA Binding Protein-4), Equally important roles in heart development have been shown for members of the T-box (Tbx5, Tbx20), basic helix-loop-helix (DHAND, EHAND), and MADS (MCMI, agamous, deficiens, serum response factor) domain (MEF2) families. LIM (a cysteine-rich motif identified in the homeobox genes lin-11, Isl-1, and mec-3) homeodomain transcription factor, which is expressed in a distinct population of cardiac precursor cells, is essential for the formation of the outflow tract and the right ventricle [24]. Other transcription factors involved in lineage decisions include HF-1b and HOXB5 (Homeobox B5). HOXB5 is necessary to activate the cell-intrinsic events that regulate the differentiation of angioblasts and mature endothelial cells from their mesoderm-derived precursors. HF-1b plays a critical role in conduction system lineage formation and the loss of HF-1b leads to a confused electrophysiological identity in Purkinje and ventricular cell lineages, resulting in cardiac sudden death and marked tachycardia and bradycardia. The transcription factor GATA4 is a critical regulator of cardiac gene expression and plays an essential role in promoting cardiac development and differentiation of the myocardium, as well as in regulating survival and hypertrophic growth of the adult heart [25-33]. In addition to these genes, several members of a family of 18 transcription factors that share a common DNA binding domain – the T-Box - have emerged as critical regulators of multiple cardiogenic events [34]. And their activity is crucial for cardiac morphogenesis [35-41].

Disruption of these genes in model organisms produces extremely severe cardiac phenotypes culminating in embryonic development arrest.

In many cases, systemic loss of individual transcription factors through targeted mutagenesis in mice leads to defective heart tube formation and subsequent embryonic lethality. Loss of GATA4 leads to defective endoderm development and failure of heart primordial fusion, but cardiomyogenic differentiation is initiated, apparent in contractile protein gene expression. Nkx2.5-deficient mice develop a primitive heart tube that differentiates, but these embryos fail to develop a looped heart. Similarly, mice lacking MEF2c or Hand2 (dHand) arrest at the primitive heart tube stage with defects in outflow tract (OFT) and right ventricle (RV) morphogenesis. Loss of Tbx20 prevents heart chamber maturation and leads to reduce cell proliferation with embryonic lethality by embryonic day 10.5. Loss of Tbx5 preferentially affects the caudal region of the primitive heart tube, with loss of venous pole structures and localized gene expression in embryos that do not survive beyond E11. In general, cardiomyogenic differentiation and activation of contractile protein gene expression is apparent in each of these mutant mouse embryos.

Targeted conditional loss-of-function of cardiac transcription factors reveal additional roles in cardiomyocyte differentiation, heart chamber maturation,

morphogenesis, cell proliferation, and conduction system development. Likewise, loss of both GATA4 and GATA6 in embryos after tetraploid rescue of extra-embryonic defects results in complete loss of cardiac differentiation Conditional loss of GATA4 expression later in differentiated cardiomyocytes leads to decreased cell proliferation and thinning of the myocardium. Tbx5 heterozygous mutant mice have congenital heart malformations, including atrial septal defects (ASD), as well as conduction system anomalies [42].

In specialized cardiomyocyte populations, other transcription factors, more locally expressed, confer lineage restricted gene expression. In many cases, localized regulation of transcription factors by signaling pathways in specific regions of the heart drives the specialization of cardiac lineages and morphogenesis.

Although there is apparently no single factor responsible for each cardiomyogenic lineage determination, multiple transcription factors in fact work in concert to regulate the initial differentiation and the subsequent specialization of cardiomyocyte cell lineages in the developing heart. We could say that these multiple transcription factors work combinatorially as a "core" to induce cardiac muscle differentiation.

Importantly, these factors mutually activate each other's transcription to reinforce cardiac differentiation through a series of overlapping feed-forward and positive-feedback loops. For example, in vertebrates, Nkx2-5 and Mef2c are directly activated by GATA factors and GATA genes, such as Gata6, are activated by Nkx2-5. The numerous reinforcing transcriptional pathways among the core cardiac transcription factors provide robustness to the cardiac differentiation program by amplifying the level of the transcription factors responsible for activating genes necessary to make a functional sarcomere. Not surprisingly, the majority of cardiac structural genes and genes involved in energy metabolism in cardiac myocytes contain binding sites for more than one cardiac transcription factor in their promoters and enhancers. For instance, GATA-responsive regulatory elements have been identified in Nkx2.5, Hand2, and GATA6 genes. Nkx2.5-responsive regulatory elements are present in GATA6 and Hand1 genes, while Tbx20 represses expression of Tbx2. In differentiated cardiomyocytes, regulatory elements that control expression of contractile protein genes often contain multiple binding sites enabling combinatorial regulation by multiple cardiac transcription factors [43, 44].

Cell differentiation is associated with a specific gene expression program, where a specific subset of cell-type genes is expressed, leaving the rest of the genome in a repressed state. Gene expression reprogramming is a key feature of cardiac differentiation and heart development and is characterized by activation of specific genes. Changes in the gene transcription program may play a significant role in the cardiac commitment, thus understanding the molecular mechanisms and the signaling events underlying gene transcription in these process could be important. Indeed there are extensive cross-regulatory interactions among cardiogenic transcription factors and the genes that encode them, which have been described as a reinforcing regulatory network, however this tightly controlled expression pattern together with the notion that the gene regulatory networks of coordinately acting and cross-regulating transcription factors control various aspects of cardiogenesis through regulation of target gene expression, was sufficient to activate cardiac gene expression but only in the presence of the cardiac chromatin

remodeling complex that re-modulate the chromatin structure in order to unwrap the DNA and let the promoter accessible to the transcription factor and activate cardiac gene expression [45].

In addition to key signaling and transcriptional events that directs the cardiac lineage, chromatin remodeling and epigenetic modifications are also believed to be necessary in initiating the cardiogenic program. In fact the level of transcriptional control conferred by the structure of chromatin accounts for how cells with the same DNA sequence can form complex organisms

1.3 Epigenetics

An exciting finding of the past 20 years is that transcription regulation in eukaryotic cells does not depend solely on elements of DNA that control gene expression (e.g promoter and enhancer sequences) but also on the state of the chromatin in which a gene is located. In this scenario, the main player is epigenetics – a complex network of mechanisms that control gene expression in a potentially heritable way but without altering the primary nucleotide sequence. These mechanisms regulate gene expression by modulating chromatin structure and DNA-based biological processes such as the binding of transcription factors to promoters and transcription elongation [46, 47].

Regions of the genome that are transcriptionally active have an open chromatin structure (euchromatin) that facilitates binding of transcription factors, whereas inactive regions have a more condensed conformation (heterochromatin) that inhibits the binding of transcription factors.

1.3.1 Origin of Epigenetics

The conceptual origins of epigenetics arise from Aristotele (384-322 BC), who in the fourth century BC in his book "*On the Generation of Animals*" express the theory of epigenesis as the development of organic individual form from the unformed.

Later in the mid-nineteenth century other are the traces of epigenetics in literature. The first believer of this theory was the german physiologist Caspar Friedrich Wolff (1734-1794) who exposed it in 1759 in his work "*Theoria Generationis*". Wolff's theory was in contrast with the preformist theory, supported in previous times, according to which human beings developed from miniscule fully-formed bodies. However, during the late 18th century an extended and controversial debate by biologists finally led epigenesis to eclipse the long-established preformationist view To date, there exists several definitions of epigenetics, and as a result, there are disagreements as to what epigenetics should mean. The term "*Epigenetics*" was coined by C. H. Waddington in 1942, as a portmanteau of the words "*genetics"* and "epigenesis", to describe the differentiation of cells from their initial totipotent state in embryonic development. When Waddington coined the term the physical nature of genes and their role in heredity was not known; he used it as a conceptual model of how genes might interact with their surroundings to produce a phenotype. Robin Holliday defined epigenetics as "the study of the mechanisms of temporal and spatial control of gene activity during the development of complex organisms. Therefore *epigenetic* can be used to describe anything other than DNA sequence that impacts on the development of an organism [48].

The modern usage of the word in science is more narrow; it is, as defined by Arthur Riggs and colleagues, as "the study of mitotically and meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence". The Greek prefix *epi-* in *epigenetics* implies features that are "on top of" or "in addition to" genetics; thus *epigenetic* traits exist on top of or in addition to the traditional molecular basis for inheritance.

In the modern scientific language the term Epigenetic refers to inherited traits that do not match changes in the DNA sequence.

The "epigenome" refers to the overall epigenetic state of a cell. The "epigenetic code" has been used to describe the set of epigenetic features that create different phenotypes in different cells and could represent the total state of the cell, the gene expression, DNA methylation and histone modification status of a particular genomic region. Epigenetic mechanisms can be grouped in three main categories: DNA methylation, histone modifications, nucleosome positioning.

1.3.2 DNA Methylation

DNA methylation was the first epigenetic mechanism discovered. It is always associated with gene silencing. DNA methylation occurs preferentially on the cytosine of CpG dinucleotides, which tend to cluster in regions called CpG islands. In mammals, 50-70% of CpG sites are methylated. The formation of 5-methylcytosine (5meC) residues leads to gene silencing either by directly blocking the binding of transcription factors to DNA or by binding of methyl-binding proteins (MBPs, e.g. MBD1, MECP2, MBD3, and MBD4), which in turn recruit chromatin remodeling co-repressor complexes [49, 50].
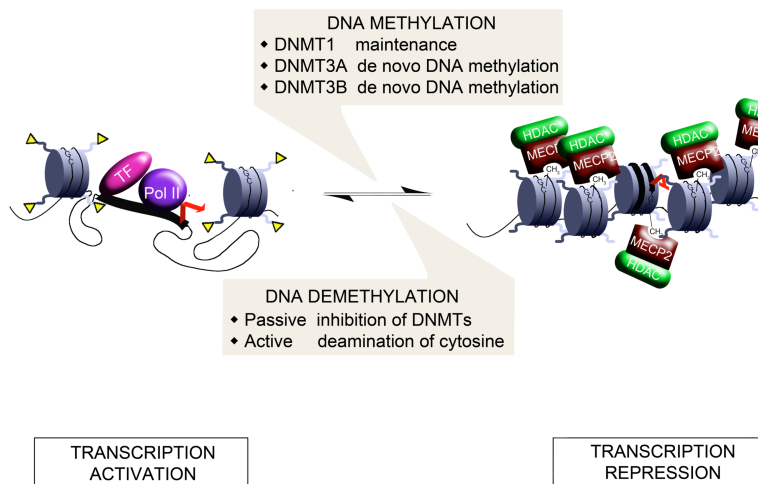
Figure 3 – DNA Methylation

The degree of DNA methylation is regulated by the rate of two events: DNA methylation on one hand – catalyzed by DNA methyl-transferase enzymes such as Dnmt1, Dnmt3a, and Dnmt3b – and DNA demethylation on the other – which can occur either passively, by blocking methylation of newly synthesized DNA during DNA replication, or actively, by the deamination of cytosines following DNA repair [51-53].

1.3.3 Histone Modifications

Most eukaryotic cells contain five main histone proteins: histone H2A, H2B, H3, and H4 – known as the core histones – and H1 – called a linker histone. Histones are small, basic proteins (11-20 kDa in molecular weight) and are formed from a globular domain and charged tails enriched in lysine and arginine residues that are subject to a large number of post-translational modifications such as acetylation, methylation, ubiquitylation, phosphorylation, SUMOylation, and ADP-ribosylation. Among the various histone modifications, acetylation and methylation are the best investigated [51]. Acetylation of the lysine residues of the tails of histone H2B, H3, and H4 is a signal for transcription activation, whereas hypoacetylated histones are found in transcriptionally inactive regions of the genome. Indeed, acetylation, which neutralizes the positive charge of lysine residues, causes an opening up of chromatin, which in this way becomes accessible to transcription factors. Acetylated histones also promote transcription, favoring the binding of transcription factors that have a bromo-domain as a docking site. On the other hand, deacetylation increases the positive charge on histone tails and, thus, strengthens the binding of histones to DNA, resulting in a more compact and less accessible structure for transcription factors [54, 55].
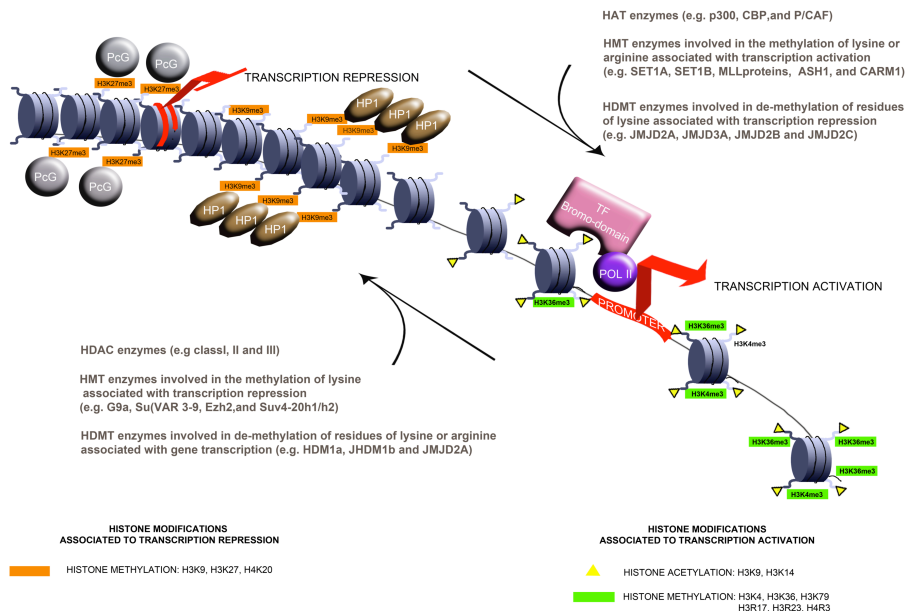
Figure 4 – Histone Modifications

Differently to acetylation, histone methylation can be associated with either activation or repression of transcription depending on the particular residue methylated (lysine or arginine) and on the degree of methylation (mono-, di-, and tri-methylation) [56].

Generally, high levels of acetylation on lysines 9 and 14 and tri-methylation on lysines 4 and 36 of histone H3 are detected in genes that are transcriptionally active, whereas elevated levels of deacetylated histones, histone H3 tri-methylated on lysines 9 and 27, and histone H4 tri-methylated on lysine 20, are associated with transcriptionally inactive regions [57].

1.3.3.1 Epigenetic enzymes encoding histone modifications

An important characteristic of the histone code is dynamicity which depends upon the reversibility of the modifications and on the activities of the many enzymes involved. Histone acetylation is catalyzed by histone acetyl-transferases (HATs), such as E1A-associated protein p300 (p300), CREB-binding protein (CBP), and p300/CBP-associated factor (P/CAF). On the other hand, deacetylation of histones is controlled by histone deacetylases (HDACs), a large family composed of 18 members clustered into three distinct structural classes [58]. Similarly, the methylation of histones is regulated by histone methyl-transferases (HMTs, e.g. G9a and SUV39h1-2) and histone demethylases (HDMs, e.g. LSD1, JmjC, and JARID) [59].
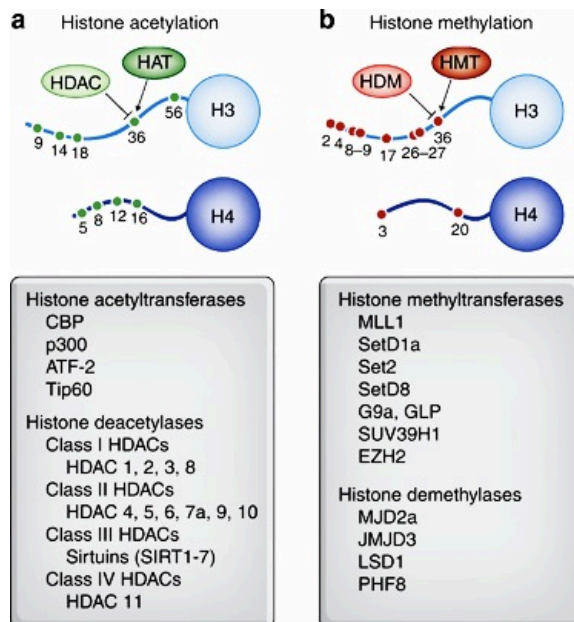
Figure 5 – Epigenetic Enzymes
(a) Histone Acetylation (b) Histone Methylation

1.3.4 Nucleosome positioning and chromatin remodeling factors

The basic unit of chromatin is called the nucleosome. This is formed from a segment of 145-147 base pairs of DNA wrapped around a histone octamer consisting of two molecules of each core histone [60]. The position of the histone octamer on DNA is not random but, rather, is influenced by the rigidity and curvature of the DNA sequence. This phenomenon, known as nucleosome positioning, has an important role in the regulation of several biological processes, including gene transcription. The formation of the nucleosome prevents the binding of proteins, such as transcription factors, to DNA and, thus, acts as a transcription repressor. Indeed, recent studies in yeast suggest that transcriptionally active genes have an "open" promoter, characterized by a nucleosome-depleted region (NDR) directly upstream of the transcription start site (TSS) and, thus, one that favors the binding of transcription factors. In transcriptionally inactive genes, instead, the nucleosome often covers the transcription start site, preventing the binding of transcription activators [61]. The position of the nucleosome is regulated by ATP-dependent chromatin remodeling (RCS) enzymes (e.g. SWI/SNF, ISWI, CHD, and INO80), multiprotein complexes that mobilize or alter the nucleosome structure and, thus, influence gene transcription regulation by using the energy derived from ATP hydrolysis.

1.3.5 Epigenetic technologies

Over the past few years, several studies combining chromatin immunoprecipitation (ChIP) with DNA-microarray analysis (Chip-on-chip) or massive DNA sequencing (ChIP-Seq) techniques have defined the histone modifications occurring within the genome [62]. The general picture emerging from these studies is that a combination of histone modifications creates a "histone code" that influences the transcriptional status of genes.

Determining how proteins interact with DNA to regulate gene expression is fundamental for completely understanding many biological processes and disease states. This epigenetic information is complementary to genotype and expression analysis.

Two are the main techniques used to analyze the histone modification signature and to study how transcription factors and other chromatin-associated proteins affect phenotype-influenced mechanisms. The microarray based ChIP technology (ChIP-chip) includes the hybridizations of the DNA fragments on an array and requires large sets of tiling arrays for lower resolution, however this method is restricted to a fixed number of probes and could introduce some bias and disadvantages.

The modern post-genomic era open the door to the ChIP-seq technology that is currently seen as the best alternative to ChIP-chip; with the ChIP-seq in fact is possible to analyze protein interaction with DNA on a genomic scale in order to map the global binding sites for the protein of interest associated with the DNA.

Specific DNA sites in direct physical interaction with transcription factors and other proteins can be isolated by chromatin immunoprecipitation. ChIP produces a library of target DNA sites bound to a protein of interest *in vivo*. Massively parallel sequence analyses are used in conjunction with whole-genome sequence databases to analyze the interaction pattern of any protein with DNA or the pattern

of any epigenetic chromatin modifications. This can be applied to the set of ChIP-able proteins and modifications, such as transcription factors, polymerases and transcriptional machinery, structural proteins, histone modifications, and DNA modifications. As an alternative to the dependence on specific antibodies, also different methods have been developed to find the superset of all nucleosome-depleted or nucleosome-disrupted active regulatory regions in the genome.

The workflow of ChIP-seq could be divided into three principal processes: ChIP, sequencing, bioinformatics analysis.

1) The Chromatin Immunoprecipitation (ChIP) is a powerful method to selectively enrich for DNA sequences bound by a particular protein in living cells. The crosslinked chromatin is sonicated in order to obtain mono-nucleosome-fragments. The ChIP process enriches specific crosslinked DNA-protein complexes using an antibody against the protein of interest. The enriched DNA-protein complexes are then decrosslinked and the DNA fragment analyzed.

2) Sequencing: after size selection, all the resulting ChIP-DNA fragments are sequenced simultaneously using a genome sequencer. A single sequencing run can scan for genome-wide associations with high resolution, meaning that features can be located precisely on the chromosomes. There are many new sequencing methods used in this sequencing step. Some technologies that analyze the sequences can use cluster amplification of adapter-ligated ChIP DNA fragments on a solid flow cell substrate to create clusters of approximately 1000 clonal copies each. The resulting high density array of template clusters on the flow cell surface is sequenced by a Genome analyzing program. Each template cluster undergoes sequencing-by-synthesis in parallel using novel fluorescently labelled reversible terminator nucleotides. Templates are sequenced base-by-base during each read. Then, the data collection and analysis software aligns sample sequences to a known genomic sequence to identify the ChIP-DNA fragments. Sensitivity of this technology depends on the depth of the sequencing run (i.e. the number of mapped sequence tags), the size of the genome and the distribution of the target factor. The sequencing depth is directly correlated with cost. If abundant binders in large genomes have to be mapped with high sensitivity, costs are high as an enormously high number of sequence tags will be required. By integrating a large number of short reads, highly precise binding site localization is obtained. Compared to ChIP-chip, ChIP-seq data can be used to locate the binding site within few tens of base pairs of the actual protein binding site. Tag densities at the binding sites are a good indicator of protein–DNA binding affinity, which makes it easier to quantify and compare binding affinities of a protein to different DNA sites.

3) Bioinformatic analysis. Because the data are sequence reads, ChIP-seq offers a rapid analysis pipeline, as well as the potential to analyze the enriched sequences with respect to all the genomic information available, is possible to analyze the immunoprecipitated DNA sequence respect the trascription start site (TSS) of the genes, to study the specific distribution of the protein in the region of the gene (upstream TSS, downstream TSS, intra-genic, extra-genic, intronic or exonic, in enhancer or other regolatory regions...), to clusterize the enriched region in functional related pathways, and find differential expressed region within differnet samples.
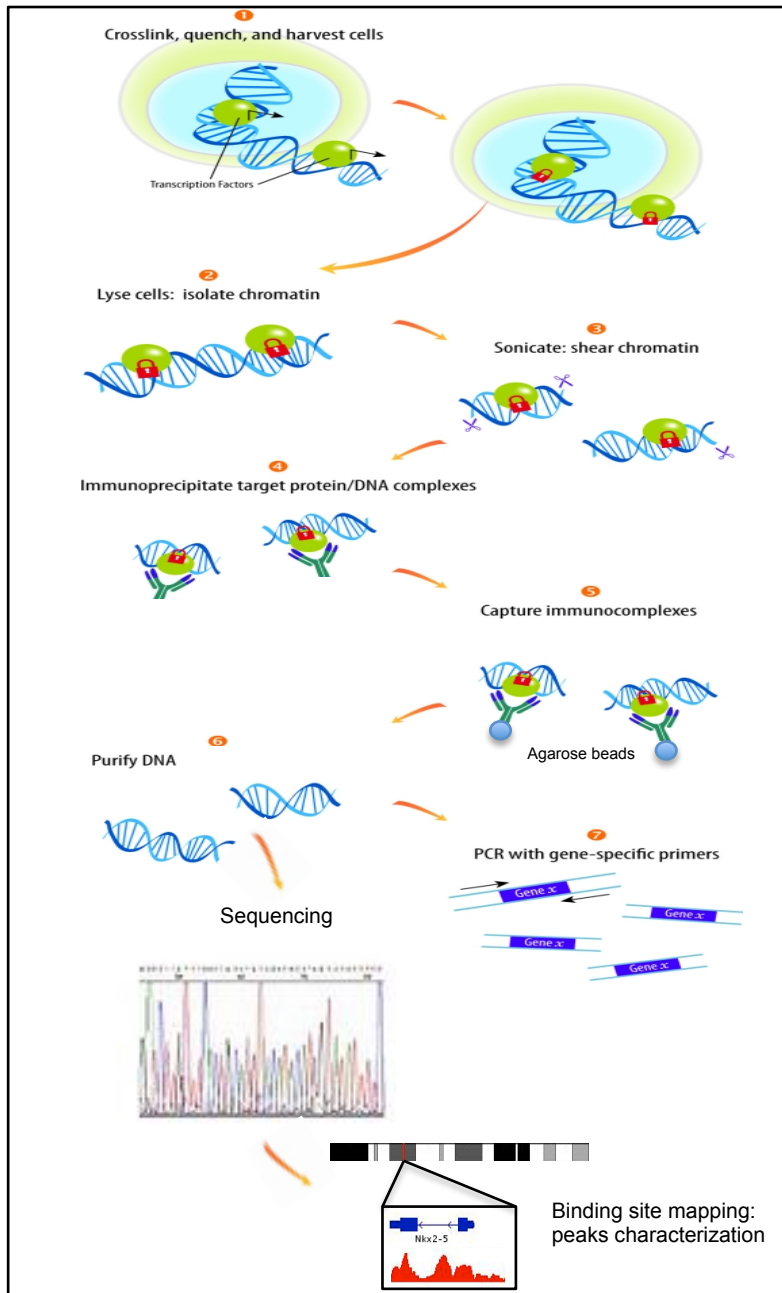
Figure 6 – Epigenetic Technologies: ChIP-seq

1.4 Epigenetics in cardiac differentiation and heart development

Most environmental factors could influence the regulation of gene transcription processes that dictate cell specification during the morphogenic process of heart development and epigenetic mechanisms seems to have a key role in these regulations.

1.4.1 The Waddington's Epigenetic Landscape

The genome sequence is static, but cells differentiate into many different types, which play different functions, and respond differently to the environment and intercellular signaling. Thus, as individuals develop, morphogenesis activates or silences genes in an epigenetically heritable manner, giving cells a "memory". In mammals, most cells terminally differentiate, with only stem cells preserving the capability to differentiate into several cell types ("totipotency" and "multipotency").

Developmental Biology is a relatively young scientific domain resulting from the blending of concepts derived from classic embryology and genetics. For many years these disciplines were mutually exclusive with geneticists disregarding the importance of embryological findings and vice-versa. As described in the paragraph above this hostile environment lasted until the end of the 1930s when critical contributions from two scientists - Salome Gluecksohn-Schoenheimer and Conrad Hal Waddington – opened intercommunication channels between these fields. Their findings provided the first solid proof that mutations in genes can induce abnormalities in early organogenic processes.

Besides the lessons derived from his work on mutations affecting *Drosophila* wing development, in two of his most influential books (*Organizers and Genes*, 1940 and *The Strategy of Genes*, 1957), Conrad coined or emphasized concepts such as competence (the capacity of a cell or tissue to react to an inductive signal), epigenetics (then used to mean external manifestations of genetic activity) and epigenetic landscape (a metaphoric representation cell-fate acquisition processes as a succession of developmental bifurcations controlled by genetic mechanisms) and is, for this reason , regarded by many as the father of modern developmental biology [63-66].

Totipotent

Lineage-committed

Pluripotent (progenitor)
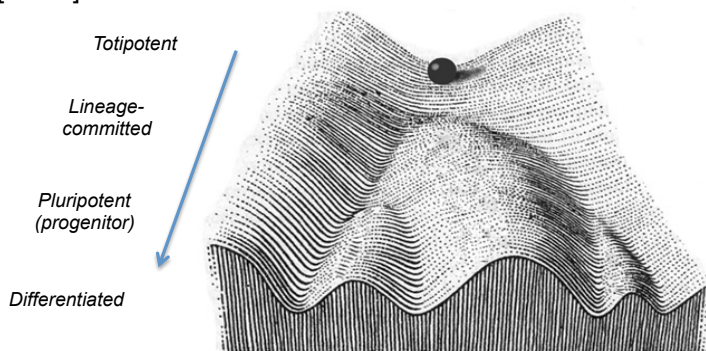
Differentiated

Figure 7 – The Waddington's Epigenetic Landscape

The modern idea of cell differentiation involved the concept of a specific gene expression program associated with a defined cell specification where a specific subset of cell-type genes is expressed, leaving the rest of the genome in a repressed state. But this hypothesis, now milestone of the modern science starts in the fifties with the epigenetic landscape of Waddington and his idea of how gene regulation modulates development and differentiation.

In this metaphor view a totipotent cell is like a marble on the top of a hill. This marble rolling down from the hill can choose different canyons and decide to differentiate to specific cell type, arriving at the bottom of the hill with the gene expression program that characterize a differentiated cell. This is possible thanks to epigenetic decisions that activate or repress specific genes and enable lineage commitment.

In the last decades thanks to the Waddington hypothesis several studies and discoveries have expanded the knowledge in this field and is now a common theory that epigenetics play a central role in the programming of genomes that underlie the establishment and maintenance of differentiated cell states, but how genomic programs are progressively deployed and what are the chromatin regulatory mechanisms that coordinate their deployment still need to be uncovered

1.4.2 Epigenetics in heart development

The heart is a complex organ whose development requires cell specification and differentiation of several cell types (e.g. cardiomyocytes, fibroblasts, and endothelial cells). Gene transcription regulation plays a critical role in orchestrating these processes [1]. Epigenetics, chromatin modifications, and remodeling, are well-known key regulatory elements of gene expression and transcription during development [67, 68]. In fact, the accessibility of the regulatory elements of DNA to transcription factors is controlled by chromatin remodeling complexes that alter chromatin structure in order to activate transcription of cardiac-specific genes and repress the transcription of non-cardiac ones.

1.4.2.1 Chromatin Remodeling During Heart Development

Chromatin remodeling complexes are important players in the modulation of gene transcription in heart development. In particular, Brg1/Brm-associated factor (BAF) complexes play a critical role in regulating cardiac growth, differentiation, and gene expression. BAF complexes act in two opposite ways: they loosen chromatin and facilitate the access of RNA polymerase II to transcriptional initiation sites, thereby activating transcription; at the same time, they associate with proteins involved in transcriptional repression, such as HDACs and methylases [69]. This dual mechanism of action is possible thanks to the several subunits that form the complexes (Figure 8). The ATPase component of the complexes is encoded by one of two genes: either *Brg1* or *Brm*. It appears that during organogenesis, Brg1 may be the primary effector of the function of BAF complexes, whereas Brm is dispensable for embryonic development. All the other subunits of the complexes (e.g. Baf60a, Baf60b, Baf60c) are expressed in a tissue-specific manner and act as a bridge with different DNA binding factors. Baf60c, a cardiac-enriched BAF complex subunit, is expressed very early in pre-cardiac mesoderm and, specifically in the heart and somites of mouse embryo, in particular in the looping heart tube at the poles of the heart, which give rise to the outflow tract and to the atria [70]. Indeed, the deletion or the downregulation of either Baf60c or Brg1 in mouse

models leads to severe defects in heart formation (hypoplastic atrium, single ventricle, abnormal outflow tract, impaired trabeculation, etc.) [70, 71].

The importance of BAF complexes and the transcriptional activation mechanisms driven by Baf60c and Brg1 is highlighted by studies that have demonstrated the role of Baf60c in promoting interactions between BAF complexes and specific cardiac transcription factors involved in heart development (such as Gata4, Nkx2.5 and Tbx5) (Figure 8).
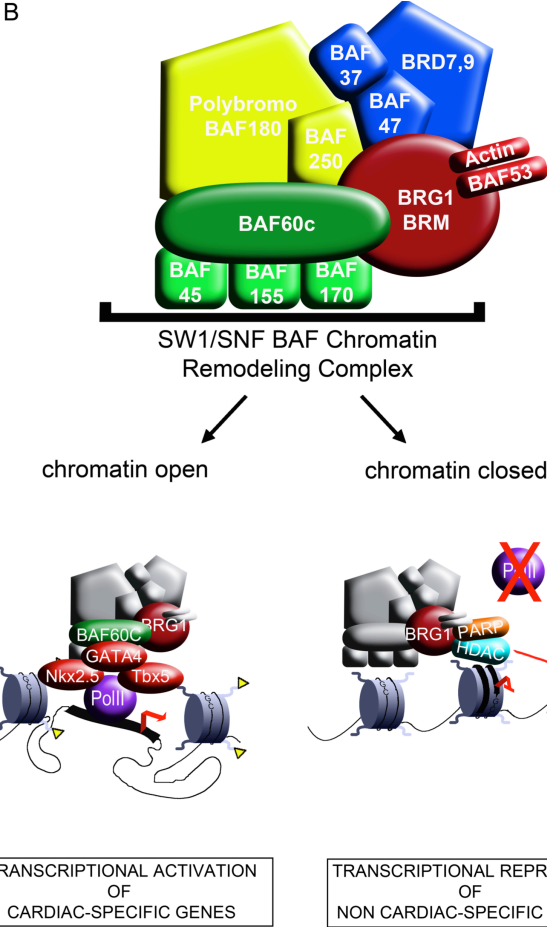
Figure 8 – Chromatin Remodeling and Heart development

In the presence of Brg1, Baf60c induces cardiac differentiation by enhancing the Gata4-dependent activation of Nkx2.5 (the cardiac transcription factor that marks the primary and the secondary heart fields) [70]. The interaction of Baf60c with Gata4 was shown to initiate the cardiac-gene expression program and the differentiation of mesoderm into cardiomyocytes; however, the addition of Tbx5 was required to induce genes responsible for cardiac contraction and to repress

non-cardiac mesodermal genes. Thus, Baf60c potentiates the function of Gata4 and Tbx5, allowing the binding of Gata4 to cardiac loci and, as a result, the induction of a full cardiac-differentiation program [71]. As mentioned above, BAF complexes can also associate with proteins involved in transcriptional repression: in fact, a recent study demonstrated that Brg1 preserves fetal cardiac differentiation by interacting with two other classes of chromatin-modifying enzymes - transcriptional repressor HDACs and poly (ADP-ribose) polymerase (PARP) (Figure 8). Brg1 maintains cardiomyocytes in an embryonic state and regulates the expression of α- and β-myosin heavy chain (MHC) during cardiac growth and differentiation. Brg1, PARP, and HDACs physically form a chromatin-remodeling complex on the α-MHC promoter in order to co-repress α-MHC, whereas Brg1 complexes with PARP on the β-MHC promoter to activate β-MHC. Thus, Brg1 governs two parallel pathways to independently control myocardial growth and differentiation in embryos [72, 73].

1.4.2.2 Histone Acetylation and Heart Development

The function of histone acetylation in heart development has been investigated mainly through the generation of knockout mouse models of genes encoding HDACs and HATs: the study of these models has revealed that both class I and class II HDACs and p300 HAT have a crucial role in heart development. Loss-of-function of HDAC2, a class I HDAC, produced altered myocardial phenotypes and excessive proliferation of cardiomyocytes. Knockout of HDAC1, another class I HDAC, produced severe proliferation defects and general growth retardation but no specific phenotypes. However, the conditional deletion of HDAC1 and HDAC2 together resulted in neonatal lethality, accompanied by cardiac arrhythmias, dilated cardiomyopathy, and upregulation of genes encoding skeletal muscle-specific contractile proteins and calcium channels in the heart, pointing to the redundant roles of HDAC1 and HDAC2 in cardiac growth and development [74-76]. Mice lacking HDAC5 and HDAC9 were also found to have defective heart morphogenesis, with lethal ventricular septal defects and thin-walled myocardium, which typically arise from abnormalities in growth and maturation of cardiomyocytes [72, 77]. Moreover, the targeted deletion of HDAC7 resulted in embryonic lethality due to vascular dilation and rupture. Through the repression of cardiac transcription factors, such as myocyte enhancer factor 2 (MEF2), serum response factor (SRF), nuclear factor of activated T-cell (NFAT), and the zinc-finger protein GATA, HDACs are involved in the coordination of the gene expression programs required for the cardiomyocyte differentiation, proliferation, and morphogenesis mechanisms that underlie heart formation [78]. In particular, a series of studies demonstrated that MEF2 is a critical target of class II HDACs: in fact, these enzymes form a complex with MEF2 on gene regulatory elements, causing the repression of genes harboring MEF2-binding sites [73] (Figure 9). (Indeed, abnormal cardiac growth in HDAC knockout mice correlated with super-activation of MEF2-mediated transcription. In addition to HDACs, the importance of p300 HAT has also been shown for heart development: indeed, p300 knockout mice presented cardiac defects (impaired expression of muscle structural proteins such as b-MHC and a-actinin, as well as reduced ventricular trabeculation) and died between days 9 and 11.5 of gestation. As for class I and class II HDACs, p300 HAT also controls fetal cardiac-gene expression through combinatorial

interactions with the transcription factors involved in heart development [79]. In particular, the interaction of p300 HAT with MEF2D promoted the transcription of cardiac genes required for cardiac development, such as cardiac a-actin. The acetylation of GATA4 was also involved in the differentiation of embryonic stem cells into cardiomyocytes, and Tbx5, critical for early cardiac morphogenesis, mediated the histone acetylation and transactivation of the Nppa promoter. Thus, a mechanism proposed to explain the function of HDACs and p300 HAT in heart development involves the mutually exclusive association of HDACs and HATs with the MEF2 transcription factor, which, consequently, acts either as a transcription repressor or a transcription activator (Figure 9). In this way, p300 HAT could bind to MEF2 during the early stage of heart development in order to activate the transcription of factors critical for cardiac morphogenesis during this period; later on, the interaction of MEF2 with HDACs would represses them.
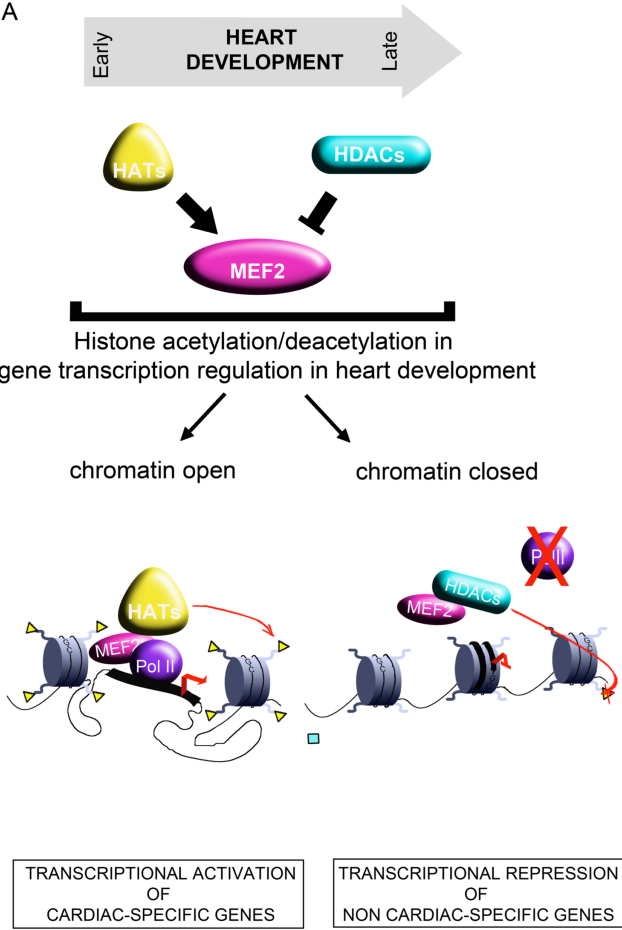
Figure 9 – Histone Acetylation and Heart Development

1.4.2.3 Histone Methylation and Heart Development

Compared to histone acetylation, the role of histone methylation during heart development is less clear. Two classes of enzymes are involved in this process: HDMs and HMTs. HDMs belonging to the jumonji (Jmj) family act mostly as transcriptional repressors and are essential for normal heart development and function. In fact, Jmj is expressed in all stages of normal developing heart from the pre-cardiac mesoderm to the adult mouse heart. The phenotype of Jmj knockout mice was characterized by critical congenital heart malformations (ventricular septal defects, noncompaction of the ventricular wall, double-outlet right ventricle, and dilated atria) and dysregulation of cardiac markers (a-MHC, b-MHC, ANF, MLC2V, MLC2A) and myocardial contractile proteins [80-82]. Therefore, Jmj may interact with transcription factors to modulate target gene expression. Epigenetic regulation via histone modifications stabilizes transcriptional programs in embryonic progenitors and differentiated cells and is likely to be crucial for establishing and maintaining gene expression and stress responses throughout life on the other hand, the functions of HMTs in the embryonic heart are completely unknown.

1.4.2.4 DNA Methylation and Heart Development

Most studies to date have focused on the role of DNA methylation in the differentiation processes (e.g. neuronal differentiation). However, it is currently unknown whether DNA methylation is involved in heart development.

1.5 The H3K79 Histone Methyltransferase DOT1L

Until now histone methylation has not been clearly implicated in cardiovascular development, but mouse genetics findings indicate that histone methylation could be a major regulator of heart development, as it is in other organ systems.

Histone methylation is catalyzed by group of histone methyltransferases (KMTs). The KMTs can be divided into two main classes based on their catalytic domains: with SET domain and without SET domain. The only KMT enzyme belonging to the second class is Dot1/DOT1-Like (disruptor of telometic silencing, also called Kmt4) [83].

Within all the post-transcriptional modifications that occur on the histones, Lysin 79 of the histone H3 is the only globular domain subject to methylation. Knockout models in yeast, flies and mice reveal that Dot1 is the only enzyme that catalyzes H3K79 mono-, di- and tri- methylation, in fact the absence of Dot1 results in the complete lost of methylation at the level of H3K79 [84-86].

While the establishment of H3K79methylation is well described and is highly regulated by multiple processes, the reversibility of this marker is less well studied, few studies suggest mechanisms of H3K79 demethylation [87]. However, several lines of evidence suggest that H3K79methylation might be reversible and subjected to dynamic regulation and dioxygenase may catalyze the removal of H3K79me2 marks.

Dot1 was identified in a genetic screen for genes whose overexpression disrupts telomeric silencing in saccharomyces cerevisiae [88]. The establishment of telomere and telomere-proximal DNA silencing is achieved through the recruitment and binding of Sir proteins and Sir3 inhibits H3K79 methylation by competing with Dot1L.

A part from this function, genome-wide analysis link methylation of H3K79 to active transcription. All H3K79 methylation markers are localized within the body of transcribed gene and the amount of enrichment correlates with expression level, suggesting a role of DOT1L in transcription activation and elongation [57, 89, 90]. DOT1L is recruited to elongating RNA Pol II through its association with a network of proteins that include the elongation factors ELL, AF4, AF10; AF9 and ENL and function as transcription memory to maintain active the transcription machinery.

In mammals besides regulating heterochromatin formation at telomeres [91], indeed several others are the functions that arise from Dot1L's enzymatic activity.

It plays a role in cellular and biological processes such as DNA repair and cell cycle regulation [86, 92]. DOT1L-mediated H3K79 methylation is believed to play two distinct roles in Rad9-mediated DNA damage response: the activation of the G1-S checkpoint and the repair of DNA damage at late G2 phase. Beyond its role in DNA damage response, DOT1L is also important for meiotic checkpoint control and cell cycle regulation in fact H3K79 methylation levels fluctuate with the cell cycle phases [93]. Deficiency of DOT1L caused a G0-G1 cell cycle arrest and induced apoptosis in erythroid progenitor, blocking differentiation. A part from these functions DOT1L is broadly involved in leukemogenesis, particularly those mediated by MLL fusion protein [94, 95], and in cell differentiation/reprogramming and in embryonic development.

Inhibition of Dot1L histone methyltransferase activity significantly increased the efficiency of reprogramming of differentiated cells to induced pluripotent stem cells (iPS) with the down-regulation of mesenchymal regulators indicating a potential role for DOT1L in cell differentiation through this lineage [96].

In addition, during embryogenesis H3K79me2 is absent in the mouse zygotes and is present at very low levels until the blastocyst stage, suggesting that loss of the epigenetic marks is important for early embryogenesis, in fact germline disruption of DOT1L in mice is lethal at embryonic day E10.5 during the organogenesis of the cardiovascular system. Consistently, knockout embryos display growth impairment, cardiovascular defects including heart enlargement, cardiac dilatation, yolk sac angiogenesis defects, decreased vasculature and anemia [86].

In fact, DOT1L is highly expressed in the heart, and a cardiac-specific knockout of mDOT1L (Dot1L$^{f/f}$:αMHC-Cre) caused dilated cardiomyopathy (DCM) with chamber dilatation and systolic dysfunction [97]. The mechanism behind this process consists in the transcriptional regulation of dystrophin mediated by DOT1L-H3K79 methylation.

The cardiac phenotype associated with the loss of DOT1L is complex, besides dystrophin activation many other cardiac genes seems to be regulated be H3K79 methylation, however, studies are needed to shed light on how Dot1L achieved specificity to temporally regulate its catalytic activity at specific chromatin domains.

2. AIM OF THE STUDY

Insights into cardiac development promise to have impact on human disease in many ways, most of which have yet to be realized. The identification of cardiac control genes permits genetic screening for mutations in affected individuals and families. Research on the basic principles of cardiogenesis is believed to have an important impact on the understanding of many cardiac malformations and may give inroads into therapeutic strategies.

A deeper understanding of the regulatory mechanisms controlling cardiovascular gene expression is critical for deciphering the mechanisms underlying congenital heart defects, as well as for helping to design more powerful and specific regenerative therapies for heart disease. Controlling gene expression in specific tissues and developmental windows is essential for cardiovascular development in embryos.

Cardiac lineage specification and subsequent morphogenesis of the early developing heart are complex processes that rely on networks of interacting DNA-binding transcription factors and targeted activation of cardiac-specific genes. Mutations in cardiac transcription factors, the genes they regulate and the genes that regulate them, result in many inherited congenital heart defects and point to the importance of understanding the molecular basis behind these processes. However, mutations on a given transcription factor may result in different types of heart defects, suggesting that other mechanisms related to gene expression are involved in normal and pathologic heart development. It is now clear that epigenetics plays a central role in the programming of genomes that underlie the establishment and maintenance of differentiated cell states, but how genomic programs are progressively deployed and what chromatin regulatory mechanisms coordinate their deployment in cardiomyocytes fate decisions still need to be uncovered. To date, most studies on organ development have focused on the role of remodelling chromatin structure complexes and histone acetylation in defining the transcription program during cardiomyogenesis. Less clear is the importance of other histone modifications and, in particular, the dynamic interactions between histone methyltransferases and demethylases.

To this end the aim of my PhD study was to define the transcriptional regulation profiles in the cardiovascular system and to identify the "epigenetic code" associated with histone methylation in order to identify genes encoding epigenetic enzymes that are regulated during cardiac development and differentiation and to elucidate the signaling events underlying gene transcription during embryonic stem cell cardiac differentiation and heart development.

3. MATERIAL AND METHODS

3.1. Isolation of heart tissues and CMs purification

All the experiments were performed according to the Guide for the Care and Use of Laboratory Animals published by the US National Institutes of Health (NIH Publication no. 85–23, revised 1996) and approved by the local ethical committee. The study was performed on embryonic (E14.5), neonatal (1 day after birth) and adult (2 months old) CD1 mice. Animals were housed in a controlled environment on an illumination schedule of 12h light/dark and fed with chow diet. Mice were sacrificed according to the protocol of the internal ethics committee, the hearts were isolated and primary embryonic, neonatal and adult and cardiomyocytes were isolated and cultured as described elsewhere [98-100].

3.2. mES culture and differentiation protocol

The TBV2 cell line of mouse Embryonic Stem (mES) cells was used throughout this study. Cells were cultured on a feeder layer of mitotically inactivated Mouse Embryonic Fibroblast (MEFs) in order to keep them undifferentiated in pluripotent state. The propagation medium was high glucose DMEM supplemented with sodium pyruvate, L-glutamine, penicillin–streptomycin, 2-mercaptoethanol, 15% ES-screened FBS (Hyclone) and 10^3 U/mL LIF (Millipore). mES were passed twice on 0.1% gelatin-coated tissue culture dishes without MEFs before starting the experiment.

Differentiation of mES into the cardiac lineage was carried in feeder-free conditions using standard techniques [101]. In differentiation medium (high glucose DMEM supplemented with sodium pyruvate, L-glutamine, penicillin–streptomycin, 2-mercaptoethanol, 15% FBS (GIBCO) without LIF) Embryoid Bodies (EBs) were aggregated using the "Hanging Drop" method (300cells/drop) and cultured for 2 days (d0-d2) as hanging drop and for 3 days (d2-d5) in suspension in low attachment petri dish (Falcon). The 5-day-old EBs were plated onto 0.1% gelatin-coated tissue culture dishes in differentiation medium and cultured at different time points for mRNA analysis, ChIP experiment, immunoblotting and FACS.

3.3. Viral production and infection

Different shRNA against DOT1L and non-target shRNA cloned into the lentiviral vector plK0.1 were tested (OpenBiosystems). Viral particles were produced in 293T cells by co-transfection with plasmids pCMV-VSVG AND psPAX2.

3.4. FACS analysis

FACS procedure was carried out using standard protocols. The cells fixed in 4% PFA were stained with the following antibodies: anti-TNNI (Millipore) and anti-SSEA-1 PE (Miltenyi Biotec) and detected with a FACS Canto Cell Analyzer (Becton Dickinson).

3.5. Immunoblotting

Immunoblotting procedure was carried out using standard protocols. The following primary antibodies were used overnight at 4°C: anti-KMT4 DOT1L (abcam), anti-histone H3 dimethyl Lys79 (abcam), anti-histone H3 (abcam), anti-Pou5F1 (abcam), anti-MYH (Chemicon) anti-Lamin B (Santa Cruz).

3.6. ChIP assays

The ChIP assays were carried out according with standard procedure. Briefly 5 x 10^6 cells were used for each immunoprecipitation. Cells were cross-linked for 10 min at RT using 1% formaldehyde. Cross-linking was quenched by adding glycine to a final concentration of 0.125M. The cells were then collected, resuspended in lysis buffer (5mM PIPES pH8, 85mM KCl, 0.5% NP40 and protease inhibitors) and incubated on ice for 15 minutes before proceed to sonication to generate fragment of 200-400bp. The efficiency of sonication was assessed through agarose gel. Chromatin samples were pre-cleared for 1 hours with protein G beads and then IP overnight at 4°C with specific antibodies: anti-histone H3 dimethyl Lys79 (abcam), anti-histone H3 trimethyl Lys4 (Active Motif), anti-histone H3 trimethyl Lys 27 (Millipore-Upstate), anti-histone H3 trimethyl Lys9 (Millipore-Upstate), anti-histone H3 (abcam) and rabbit IgG (Millipore-Upstate). After incubation the immuno-complexes were bound to protein G beads for 2 hours and subsequently washed with low-salt wash buffer (0.1% SDS, 2mM EDTA, 20mM Tris HCl pH8, 1% Triton X-100, 150mM NaCl and protease inhibitors), high-salt wash buffer (0.1% SDS, 2 mM EDTA, 20mM Tris HCl pH8, 1% Triton x-100, 500mM NaCl and protease inhibitors) and TE buffer. Immuno-complexes were eluted in elution buffer (1% SDS, 100mM NaHCO$_3$ and protease inhibitors) and the crosslinking was reverted overnight at 65°C. Samples were treated with proteinase K, extracted with phenol/cloroform and precipitated with ethanol.

Purified DNA was evaluated by qPCR on ABI 7900HT with SYBR green PCR master mix (Applied Biosystem) using specific primer designed close to the promoter region and the TSS of the gene (A = -1000bp/-500bp to TSS, B = +500bp/+1000bp to TSS, C = +3500bp/+4000bp to TSS). Values obtained were normalized to the input and to the H3 content. The sequences of primer used for ChIP-qPCR are available upon request. ChIP DNA fragment were sequenced and libraries were by high throughput sequencing with SoLiD 5500 (Life Technologies).

3.7. ChIP-seq analysis pipeline

The ChIP-seq data analysis was performed using several bioinformatic tools.

For H3K4me3, H3K27me3 and H3K9me3 histone modifications in ES cells, bed files were downloaded from Gene Expression Omnibus (H3K4me3: GSM307618; H3K27me3: GSM307619; H3K9me3: GSM307621). Genome coordinates were converted from mm8 to mm9 mouse reference genome using Batch Coordinate Conversion (liftOver) created by the UCSC Genome Bioinformatics Group.

For H3K79me2 histone modification in ES cells raw sequencing reads were downloaded from Gene Expression Omnibus (GSM307150-GSM307151) and mapped to the mouse genome (version mm9) using BOWTIE (version 0.12.8).

To profile histone modifications in cardiomyocytes, sequencing reads were mapped to the mouse genome (version mm9) using BOWTIE (version 0.12.8). Uniquely mapped reads with no more than 2 mismatches were used for binding peak detection. To identify peaks, two peak calling software's were used: MACS and SICER. Both software detected binding peaks by comparing IP and input control. For MACS and SICER, we used the following parameters.

MACS; effective genome size = 1.87e+09,band width = 300,model fold = 5,30,pvalue cutoff = 1.00e-05. SICER; windows size =200, gap size = 600, redundancy threshold = 1, FDR = 0.05. Occupancy analysis and differential binding affinity analysis was assessed with R Bioconductor package DiffBind. The final set of binding peaks contains those that are called by both software. The averaged levels of epigenetic marks within ±4 kb of the TSS was used to plot the genome-wide distribution of histone modifications with respect to TSS [102-104].

Identification of ChIP enriched genomic regions and genes: Genomic bins with a normalized ChIP-Seq density greater than a defined threshold were considered enriched. The genomic coordinates of the full set of transcripts from the RefSeq database (http://www.ncbi.nlm.nih.gov/RefSeq/) from the July 2007 version of the mouse genome sequence (Mouse NCBIv37, mm9) was downloaded from the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgTables). Genes were associated with H3K4me3 and H3K27me3 occupied genomic regions if the gene transcription start site (TSS) occurred within the region or if the distance from the TSS to the boundary of the region was less than or equal to 2 kb. If multiple regions were associated with a single gene, all of these gene are reported the region with the greatest peak ChIP-Seq density used.

Analysis of Chromatin Marks at TSSs: Chromatin marks at promoters were evaluated by computing the ratio of ChIP to input within a 4kb region centered on the TSS of each Ensembl transcript. ChIP values were evaluated as the sum of the depth of reads of every base in the 4kb window, normalized to the total number of ChIP reads in the given sample.

3.8. Total RNA extraction, cDNA synthesis and qRT-PCR gene expression analysis.

RNA was extracted from mES cells and CMs using TRIzol (Invitrogen). 1μg of RNA was reverse transcribed to cDNA using Super Script VILO cDNA Syntesis Kit (Invitrogen) and amplified by real-time quantitative PCR with SYBR Green PCR master mix (Applied Biosystem) and using specific primers for pluripotency and cardiac markers (Appendix Table II). Each sample was analyzed in triplicated using ABI 7900HT (Applied Biosystems). 18s gene was used as housekeeping for expression normalization.

3.9. Epigenetic enzyme TaqMan assay cards.

RNA was extracted from mES cells and CMs using TRIzol (Invitrogen). RNA was reverse transcribed to cDNA using Super Script VILO cDNA Syntesis Kit (Invitrogen). 1.2μg of cDNA of each sample is used to analyze the expression of 88 genes encoding important epigenetic enzymes. Custom microfluidic gene expression cards were drawn using TaqMan probes (Applied Biosystems).

TaqMan probes information available upon request. Each sample was analyzed in triplicated using ABI 7900HT (Applied Biosystems). Data were analyzed with DataAssist software using the median normalization method.

3.10. Illumina gene expression microarray.

RNA was extracted from mES cells and CMs using TRIzol (Invitrogen). 500ng of RNA is transcribed to cRNA by Illumina TotalPrep RNA Amplification Kit (Ambion). A total of 700ng of cRNA is hybridized at 58°C for 16 hours to the Illumina MOUSEWT-6V BreadChips (Illumina). Bred-Chips are scanned using Illumina BedArray Reader and the Bead Scan software (Illumina). Data is processed with BeadStudio version 3 (Illumina).

3.11. Gene expression microarray analysis pipeline.

The expression levels of transcripts were analyzed by means of genome wide expression analysis with a MouseWG-6 v2.0 Expression BeadChip (Illumina) according to the manufacturer's instructions.
The BeadChips were scanned with the Illumina iScan system. Raw data were background-subtracted and normalized using the quantile normalization method (lumi software package). Normalized data were filtered for genes with significant expression levels compared to negative control beads. Selection for differentially expressed genes was performed on the basis of arbitrary thresholds for fold changes plus statistical significance according to the Illumina t-test error model (limma software). The transcript with the highest median expression was selected to represent the expression of the gene if it was represented with several transcripts.
Functional Annotation and Molecular Network Analysis: Functional annotation of significant genes identified by microarray analysis was searched by the web-accessible program named Database for Annotation, Visualization and Integrated Discovery (DAVID) version 2009, National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH) (david.abcc.ncifcrf.gov).

3.12. Correlation studies between ChIP-seq and gene expression.

For heatmap display expression data genes are ordered by the magnitude of expression analysis between samples. Genes with higher histone modifications occupancy than average expression are shown in red and samples with lower than average expression are shown in green (scale in standard deviations).
Scatter plots were produced to see how the expression level agreed with degree of histone modification. The Pearson correlation coefficients between histone modifications and the expression level of genes was calculated using R.

3.13. Statistical analysis.

Data are presented as mean ± s.d. P values were determined by two-tailed t-test. P<0.05 was considered statistically significant.

4. RESULTS

4.1 High resolution maps of histone methylation markers in embryonic stem cells, neonatal cardiomyocytes and adult cardiomyocytes.

4.1.1 The Post-Genomic Era and the ChIP-seq databases.

The beginning of the post-genomic era, the development of bioinformatics software and the improvement of all the molecular biology techniques generate all around the research world an enormous amount of information about the expression and the regulation of genes within different species and different physiological and pathological conditions. The effort that the research community is now trying to do in order to share important discoveries and information and keep money and time for future research is to produce a big atlas of genomic and epigenomic information This is possible thanks to several databases available online where researchers deposit datasets of gene expression profiles and DNA and RNA sequencing, which allowed people to explore, view and download results from already published studies and to perform metanalysis with new data.

One of these sources is the "NIH Roadmap Epigenomics Project" where genome-wide maps of DNA and histone modification from diverse collections of epigenomic data sets are available (http://www.ncbi.nlm.nih.gov/epigenomics).

In order to investigate the histone methylation profile that regulates gene transcription during cardiomyocytes differentiation we conduct a preliminary bioinformatic analysis browsing all the already deposited ChIP-seq experiments in the GEO profile datasets. ChIP-sequencing (ChIP-seq), is used to analyze protein interactions with DNA. This new method combines chromatin immunoprecipitation (ChIP) with massively parallel DNA sequencing in order to analyze the interaction pattern of any protein with DNA or the pattern of any epigenetic chromatin modifications that regulate gene transcription.

4.1.1.1 The mouse embryonic stem cells ChIP-seq.

Several studies are available on undifferentiated embryonic stem cells and several features have been analyzed with different antibodies. Comparative analysis between several datasets revealed that different mouse embryonic stem cell lines have a characteristic epigenetic signature of histone markers associated at the activation of stem cell genes and the repression of genes involved in late stages of differentiation. Relying on the type of cell line used, the sequencing method applied the type of the antibody used and the different histone modification analyzed we select two different studies to conduct our metanalysis. From the GSE12241 series published from Mikkelsen TS, Ku M, Lander ES and Bernstein BE in 2008 in the study "Genome-wide maps of chromatin state in pluripotent and lineage-commites cells" we select the samples GSM307618 (ES_H3K4me3_ChIPSeq), GSM307619 (ES_H3K27me3_ChIPSeq), GSM307621 (ES_H3K9me3_ChIPSeq) and GSM307625 (ES_WCE_ChIPSeq) and from the GSE11724 series published from Young RA in 2008 in the study "Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells" we select the samples GSM307150 (ES_H3K79me2_ChIPSeq) and GSM307155 (ES_WCE_ChIPSeq) [105, 106].

However no ChIP-seq experiments have been conduct in cardiac progenitor cells, cardiomyocytes or heart tissues.

4.1.2 Chromatin imunoprecipitation and sequencing (ChIP-seq) of neonatal and adult cardiomyocytes.

In order to investigate the genomic regions regulated by histone methylation during cardiomyocyte differentiation, we isolated cardiomyocytes from neonatal and adult mice and performed ChIP-seq experiments to compare the distribution of H3K4me3 and H3K79me2 (associated with transcription activation), and H3K9me3 and H3K27me3 (associated with transcription repression) within the genome of cells at different stages of heart development. Indeed chromatin is not a static structure. The post-translational modification of H3 e H4 affect the structural dynamic of nucleosomes and thus determine the chromatin structure and folding, which in turn leads to gene expression or silencing.

4.1.2.1 Cardiomyocytes isolation from neonatal and adult mice.

Neonatal cardiomyocytes (CMp) have been isolated from 1-day-postnatal CD1 mice. Briefly from 40 hearts CMs have been dissociated at single cell through sequential enzymatic digestions. Mice were sacrificed according to the protocol of the internal ethics committee; the hearts were isolated, washed in HBSS for 1h on ice and kept in agitation over night in trypsin solution in order to dissociate the endothelial cells from the hearts. The hearts were then processed in successive collagenase IV digestions and the obtained suspension of cells plated for the purification process. After 1h in culture at 37°C cardiac fibroblast (CFs) attach to the dish and leave in suspension CMs cells. The efficiency of isolation was tested through FACs analysis staining the cells with TNNI antibodies as marker of CMs; a mean of 84% of purity was obtained from 4 different primary CMs cell preparation experiments, one representative experiment is shown in Figure 10A.

Adult cardiomyocytes (CMa) have been isolated from 2-month-old CD1 male mice. After the incanulation of the heart through the aorta an enzymatic solution of collagenase IV was applied by the Langendorf method using a retrograde flow. This method allowed the matrix and collagen dissociation within the cells. CMa in physiological condition could reach dimension of 100μm and display rod-shape morphology, these allowed us to purify CMa from other cells by gravitometry centrifugation. The efficiency of cell separation was verified throughout the counting of rod-shape cells in the hemocytometer chamber and by the analysis of TNNI positive cells at the cytoflourimetry. CMa where isolated from 8 heart with a mean of $1x10^6$ cells per heart and an efficiency of purification of 90% (Figure 10B).
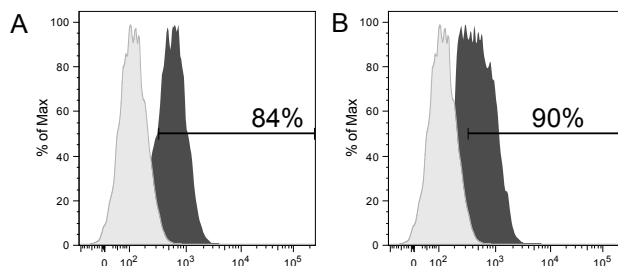
Figure 10 – FACs analysis to access CM purification: A) neonatal CM, B) adult CM

4.1.2.2 ChIP and libraries preparation.

The chromatin obtained after nuclei isolation from neonatal CMs and adult CMs was sonicated to obtain an average chromatin length of 300bp, corresponding to a mono-nucleosome fragment size. IP was performed by incubating the chromatin at 4°C overnight with antibodies anti-tri-methylated H3K4 and anti-di-methylated H3K79, histone modifications associated to transcriptional activation, and anti-tri-methylated H3K9 and anti-tri-methylated H3K27, markers of transcriptional repression. After the evaluation of the quantity and of the quality of the DNA enriched with each antibody in comparison with input, ChIP DNA fragment were sequenced using Next Generation Sequencing techniques and the distribution of these marker on the genomes was studied.

The DNA fragments sequenced, also called "reads" were aligned and mapped to the Mouse Genome (mm9) using two different programs of annotation, MACS (Model-based Analysis) and SICER. The reads could be imagined as breaks that form a peak in the portion of the genome where there is an enrichment of that modification, these peaks were than normalized versus the input (DNA from each samples before immunoprecipitation) and sort with statistical parameters such as Fold Discovery Rate (FDR) and P-value. The most significant enriched regions were then analyzed with respect to the Transcription Start Site (TSS) of each gene and with respect to the different genomic regions. The peaks annotated were than associated with genes, visualized on genome browser and clustered into functional related pathways (Figure 11).
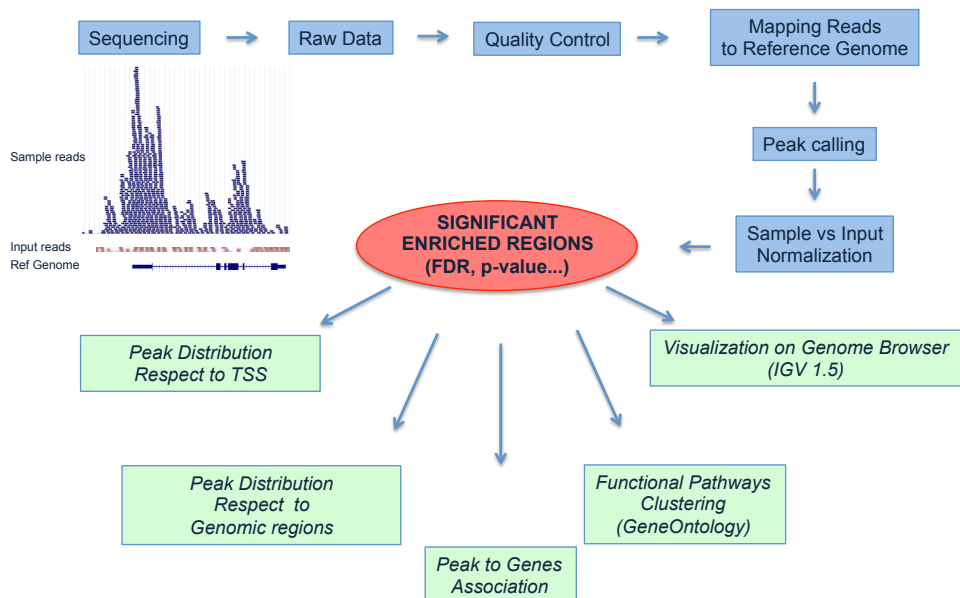
Figure 11 – Workflow of sequencing and bioinformatic analysis.

As ChIP-seq is considered an absolute quantification of histone methylation, this experiment allowed us to generate a high-resolution map of H3K4me3 H3K79me2, H3K27me3 and H3K9me3 histone methylation markers in neonatal and adult CMs that could be analyzed in comparison with the already published datasets of ChIP-seq in mouse embryonic stem cell. This analysis revealed the acquisition or the lost of makers on the locus of differential expressed genes during the cardiac differentiation process.

4.1.3 ChIP-seq analysis and metanalysis.

Comparative analysis was performed to evaluate differences in the distribution of histone markers and in the transcription regulation of genes between mouse embryonic stem cell (mES), neonatal cardiomyocytes (CMp) and adult cardiomyocytes (CMa).

The datasets generated for H3K4me3, H3K79me2, H3K27me3 and H3K9me3 in CMp and CMa from our ChIP-seq experiments have been analyzed in relation with mES ChIP-seq datasets for the same histone methylation markers downloaded from the GSE12241 and the GSE11724 series already published [105, 106].

4.1.3.1 Histone modification distribution respects gene transcription start site (TSS).

To access the distribution of each marker of interest, we first calculate the average coverage around the transcription start site (TSS) across all marked genes.

ChIP-seq libraries from mES cells, CMp, and CMa were compared through SICER software of analysis in order to find differences in marker distribution. The significant-peak lists were identified sorting the peak lists by three parameters: number of tags per peak, fold enrichment, and the False Discovery Rate.

The TSS plot was create by taking a fixed length (in bp) around the TSS, with all genes in the same orientation (5' to 3' open reading frame), and then averaging the signal across all the selected genes. A sampling frequency of 25bp/points was used to cover a region of +/- kb around the TSS of all the genes.

In accordance with previously published data, we found that each histone modification had a characteristic distribution not only within the genome but also with respect to TSSs [57, 105]. When we assessed the distribution of H3K4me3 with a TSS centered plot, we observed a sharp peak of enrichment at the TSS, which extends over a 2kb interval. Additionally the signal was stronger upstream of the TSS, than in the downstream region. By considering the difference between cell types (mES blue line, CMp red line and CMa green line), we observed that H3K4me3 was progressively redistributed during development, becoming more present inside the gene body and less associated with TSSs in CMs with respect to mES. Considering the association in bivalent domains of H3K4me3 and H3K27me3 during differentiation processes we found a similar distribution also for H3K27me3. H3K27me3 had a similar, highly dense and narrow distribution profile close to TSSs; in mES cells there was a strong enrichment of H3K27me3 around the TSS, whereas the signal was much lower in CMp and CMa cells. A common feature of the TSS centered plots is a sharp dip in H3H27me3 around the TSS, corresponding to the position of the nucleosome-depleted zone.

At the level of the TSS we noticed stark differences in the enrichment profile between the cell types, the nucleosome-depleted zone was less evident in the mES cell plots both for H3K4me3 and H3K27me3, suggesting that the distribution

of these markers across the gene could be important for regulating the binding profile of RNApol-II occupancy and transcription during CM cells differentiation.

Analyzing the distribution of H3K79me2 in the region +/-4kb to the TSS, we found a strong signal across the gene body which may indicate the presence of active transcription during termination. Considering the presence of this marker into the gene body we analyzed a wider region of +/- 25kb to the TSS. From this analysis we observed an enrichment of H3K79me2 downstream the TSS and the absence of the marker upstream. The enrichment in the gene body is gradually lost from mES to CMp and CMa indicating a progressive reduction of transcription activation during cell differentiation (Figure 12).
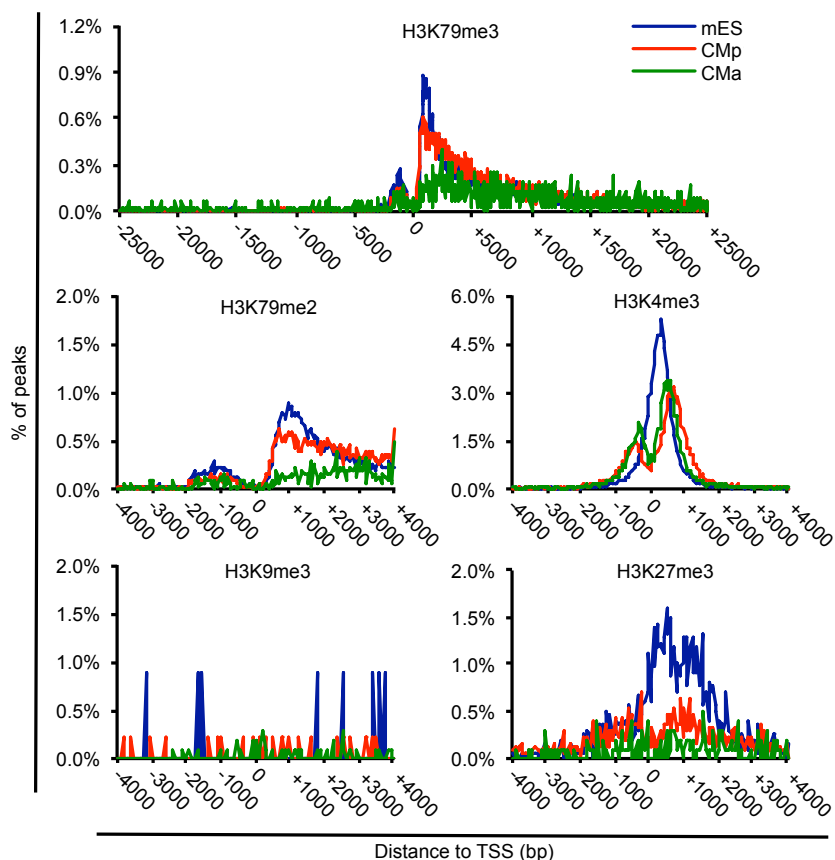
Figure 12 - ChIP-seq analysis of H3K79me2, H3K4me3, H3K27me3 and H3K9me3 in mouse Embryonic Stem cells (mES), cardiomyocytes from post-natal day 1 (CMp) and cardiomyocytes from 2 month-old mice (CMa). The averaged levels of histone modifications are plotted within a region of ± 25Kb around the annotated Transcription Start Sites (TSS) (H3K79me2) and a region ± 4Kb around the annotated TSS (H3K79me2, H3K4me3, H3K27me3 and H3K9me3), each graph shows the distribution of mES (blue line), CMp (red line) and CMa (green line).

4.1.3.2 Histone modification distribution respects genomic regions

Although regulation at TSS is important for gene regulation, understanding the distribution of these markers in the different region of the genome could identify important regulators for tissue specific gene expression patterns during lineage commitment and cardiac development.

We also analyzed the distribution of the histone methylation markers with respect to the different genomic regions, quantify the relative presence of each modification downstream the gene body, upstream the gene body, at the level of the TSS, inside the gene, in the region overlapping the start and in the region overlapping the end.
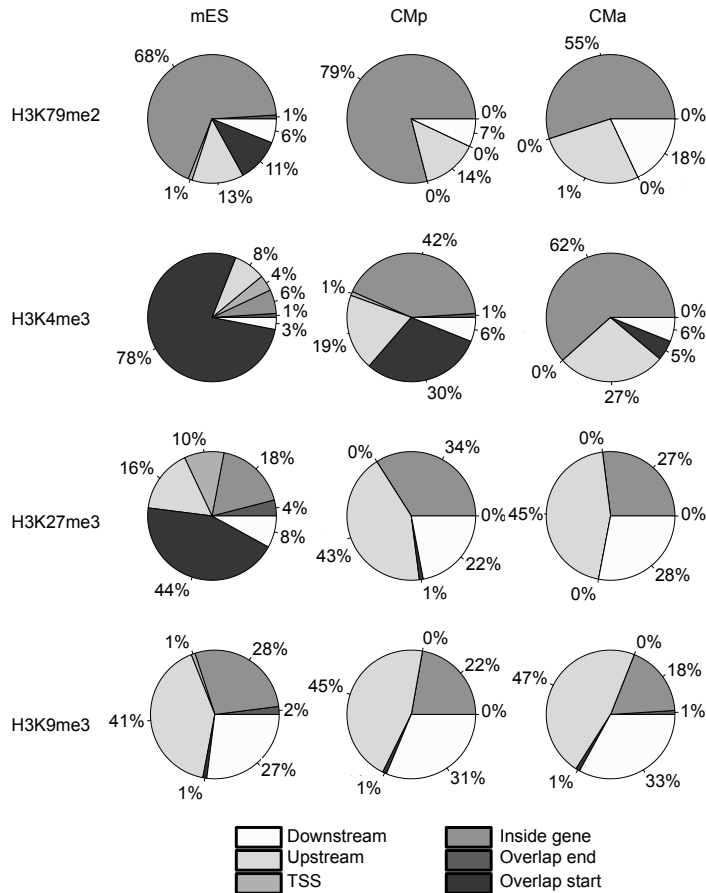
Figure 13 - Distribution of histone marks across the genome. ChIP-seq analysis of H3K79me2, H3K4me3, H3K27me3 and H3K9me3 in mouse Embryonic Stem cells (mES), cardiomyocytes from post-natal day 1 (CMp) and cardiomyocytes from 2 month-old mice (CMa). Results are expressed in percentage and show the average distribution of each marks downstream the gene, upstream the gene, at the level of the Transcription Start Site (TSS), inside the gene and in the region overlapping the end and the start of the gene.

33

According with the profile observed in figure 12 the presence of H3K4me3 significantly decreased into the gene overlapping start region from mES to CMs (78% in mES, 30% in CMp and 5% in CMa) and increased in the region inside the gene in CMs compared to mES (6% in mES, 42 % in CMp and 62% in CMa).

A similar modification in the distribution results also for H3K27me3, where the 44% presence of this marker in mES cells is completely redistributed in CMs. On the other end this marker in both neonatal and adult CMs is enriched in the region upstream the gene body.

The changes in H3K79me2 were only in part observed, both in mES cells and CMs H3K79me2 is manly present inside the gene and it appear to be partially relocated upstream and downstream during differentiation.

This analysis suggest a dynamic redistribution of the epigenetic markers upon differentiation, indicating a role of H3K4me3 and H3K27me3 in the regulation of the transcription of cell specific genes and maintaining the chromatin in a "poised" state, ready to be remodeled (Figure 13).

4.1.3.3 Binding affinity correlation and principal component analysis (PCA).

To identify the genomic loci enriched in the ChIP-seq data where a specific histone methylation is present we perform an analysis with the R Bioconductor package DiffBind using the peak callers and aligned sequence read datasets obtained with the SICER and MACS analysis.

The DiffBind program is designed to work with multiple peak sets simultaneously as well as managing the results of multiple peak callers; this allowed us to identify sites that are differentially bound between the three samples groups, to identify overlapping and merging peak sets, to count sequencing reads overlapping intervals in peak sets, and to identify statistically significantly differentially bound site based on evidence of binding affinity, measured by differences in read densities.

First DiffBind finds all overlapping peaks and originates a single set of unique genomic intervals covering all the supplied peaks from the different peak-sets, this allowed to examine how well similar samples cluster together and to enable overlaps to be examined, counting how many reads overlap each interval for each unique sample.

Figure 14 displays the result of these analyses in which a binding affinity matrix containing a normalized read count for each sample at every potential binding site is shown. With this analysis the samples are clustered using statistically differential binding affinity rather than occupancy data.

In fact, correlation analysis heatmap using affinity data, reveals that the four modifications analyzed (H3K4me3, H3K27me3, H3K79me2 and H3K9me3) in the three samples (mES, CMp and CMa) clustered into two major hierarchical clusters, which according with general epigenetic roles, divided the modifications into an activating group and a repressing group. Using the differentially bound sites we see that, within these two major groups, the same modification in the three different samples create a narrower cluster.
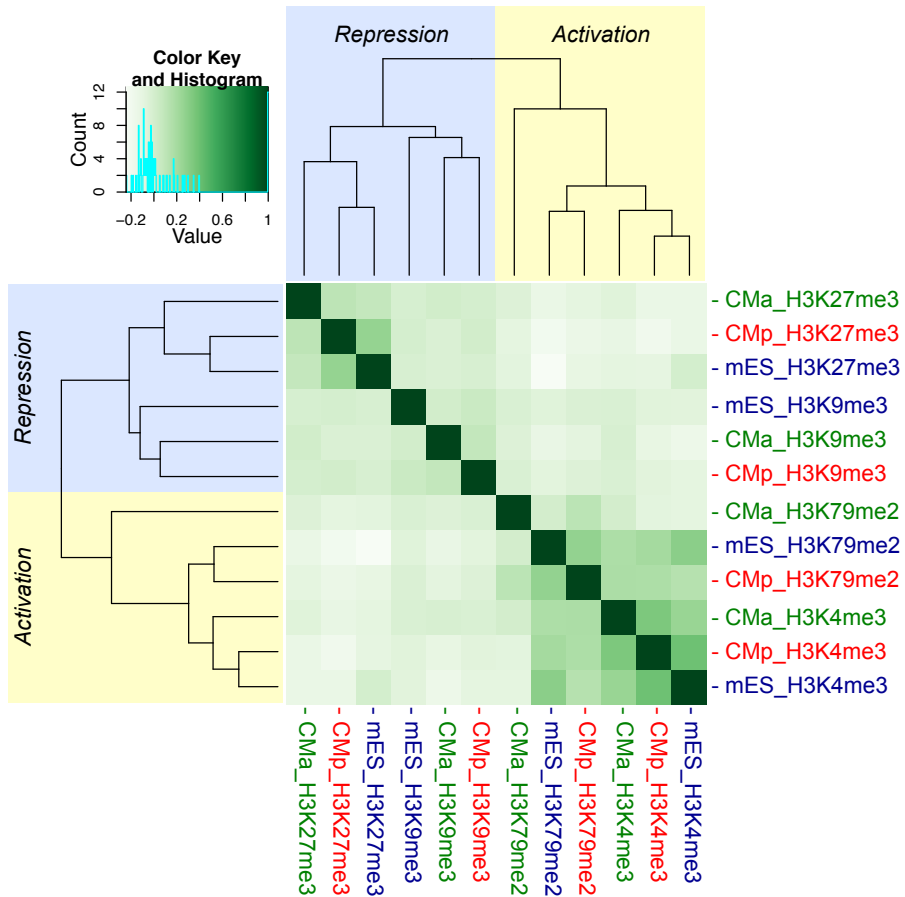
Figure 14 - Correlation heatmap using occupancy (peak caller score) data.

While the correlation heatmaps showed clustering, a Principal Component Analysis (PCA) was used to show how different samples (mES, CMp and CMa) and different histone modifications associate. PCA plot using affinity data for all the sites normalized per read count reveals that the differential analysis identifies in a statistically significant way (FDR < 0.1) several sites that are common for H3K4me3 and H3K27me3 in all the three samples. These results is supported by the fact that H3K4me3 and H3K27me3 in differentiating cells are present in the same moment on the same loci, forming bivalent domains that work together in the regulation of gene expression. On the other hand H3K79me2 seems to be present in other sites respect to H3K4me3 and H3K27me3 suggesting a potentially different and specific role of transcriptional activation associate at this histone

modification. Finally this analysis identify for H3K9me3 sites that separate the three samples associated at this marker in a region opposite to both the H3K79me2 group and the H3K4me3/H3K27me3 group of samples, according with its transcriptional repression activity (Figure 15).

PCA: Condition [49% of total variance]
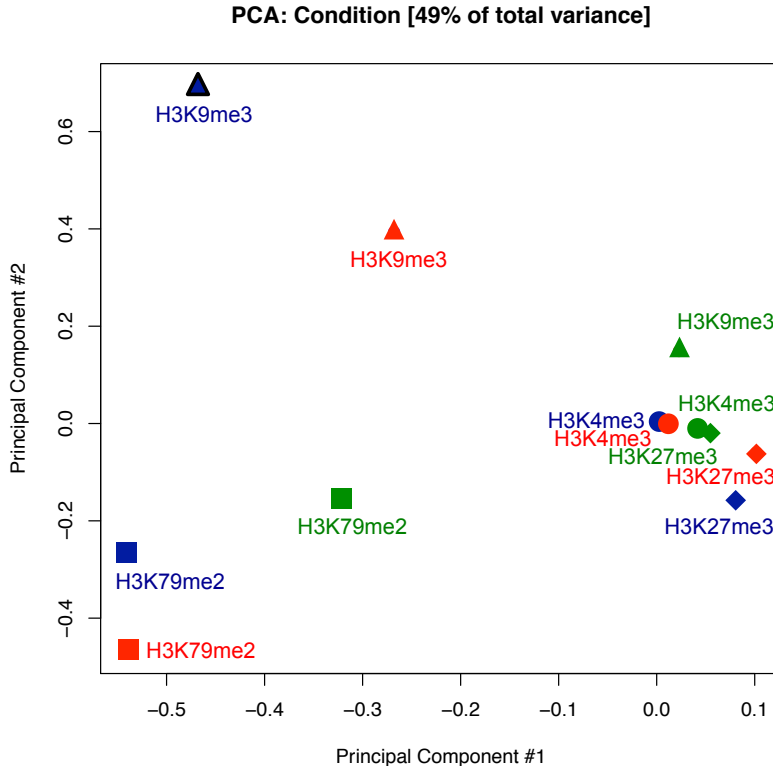
Figure 15 - PCA analysis revealed association within different samples. Affinity data plotted for H3K79me2 (□), H3K4me3 (○), H3K27me3 (◇) and H3K9me3 (△) in mouse Embryonic Stem cells (mES blue), cardiomyocytes from post-natal day 1 (CMp red) and cardiomyocytes from 2 month-old mice (CMa green). The graph displays on x and y axis the first two principal components identified by Principal Component Analysis (PCA) and reveals potential association between samples and modifications.

4.1.3.4 Identification of differential binding sites

In order to identify sites unique to a group of samples we use occupancy-based analysis. The Venn diagrams illustrated in figure 16 shows for H3K79me2 4887 sites unique for mES, whereas 1258 sites specific for CMp and 989 unique to the CMa group, with 610 sites being identified in all the three samples. H3K27me3 was identified in 1065 sites specific for mES, 708 sites in CMp and 362 sites unique for

CMa, whereas less than 10% of the sites identified were common between the samples. On the other hand H3K4me3 surprisingly results more present in shared region between the samples and H3K9me3 instead was present in a few sites in the condition analyzed. Even if some sites belonging to all the three different samples could be useful to investigate the role of the bivalent domains in the cardiac differentiation, our main interest was in finding binding sites differently present between the three groups of samples, this is the reason way we start studing the differential binding sites associates specifically to mES, CMp and CMa.
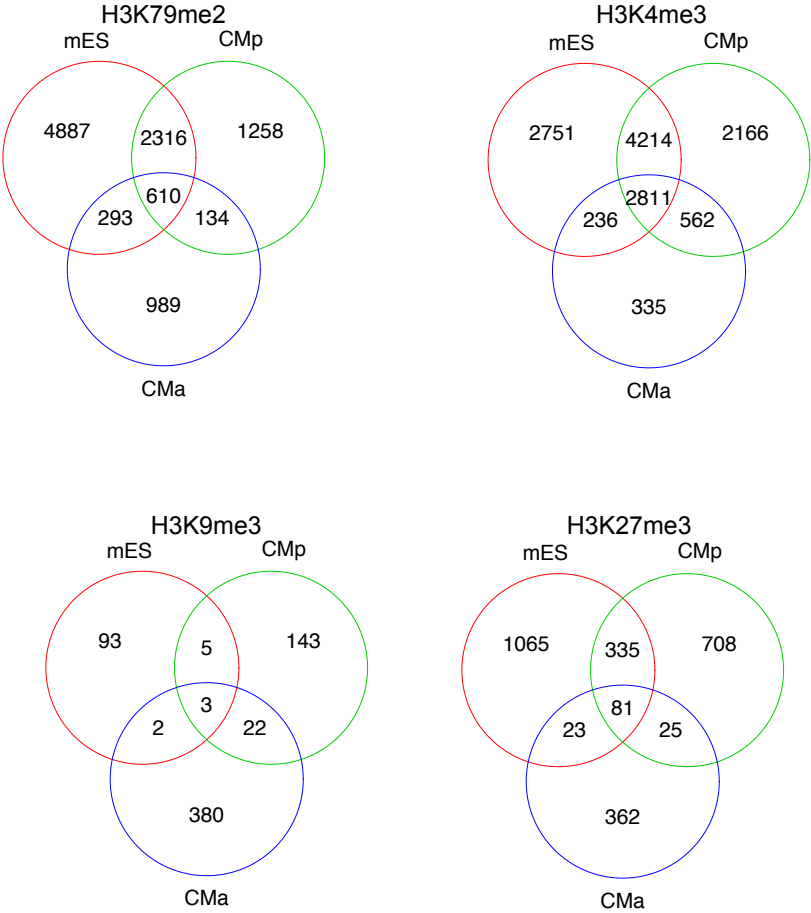
Figure 16 - Histone methylation marks have common and exclusive peaks in cardiomyocytes. Venn diagrams displaying the number of overlapping and mutually exclusive peaks associated with H3K79me2, H3K27me3, H3K4me3 and H3K9me3 in mouse Embryonic Stem cells (mES), cardiomyocytes from post-natal day 1 (CMp) and cardiomyocytes from 2 month-old mice (CMa).

In order to investigate the exclusive binding sites belonging to a specific cell type and a specific differentiation stage we carried out a comparative analysis comparing each one of the three samples within each other: mES with CMp, mES with CMa, and CMp with CMa for each one of the for histone modification analyzed (H3K79me2, H3K4me3, H3K27me3 and H3K9me3).
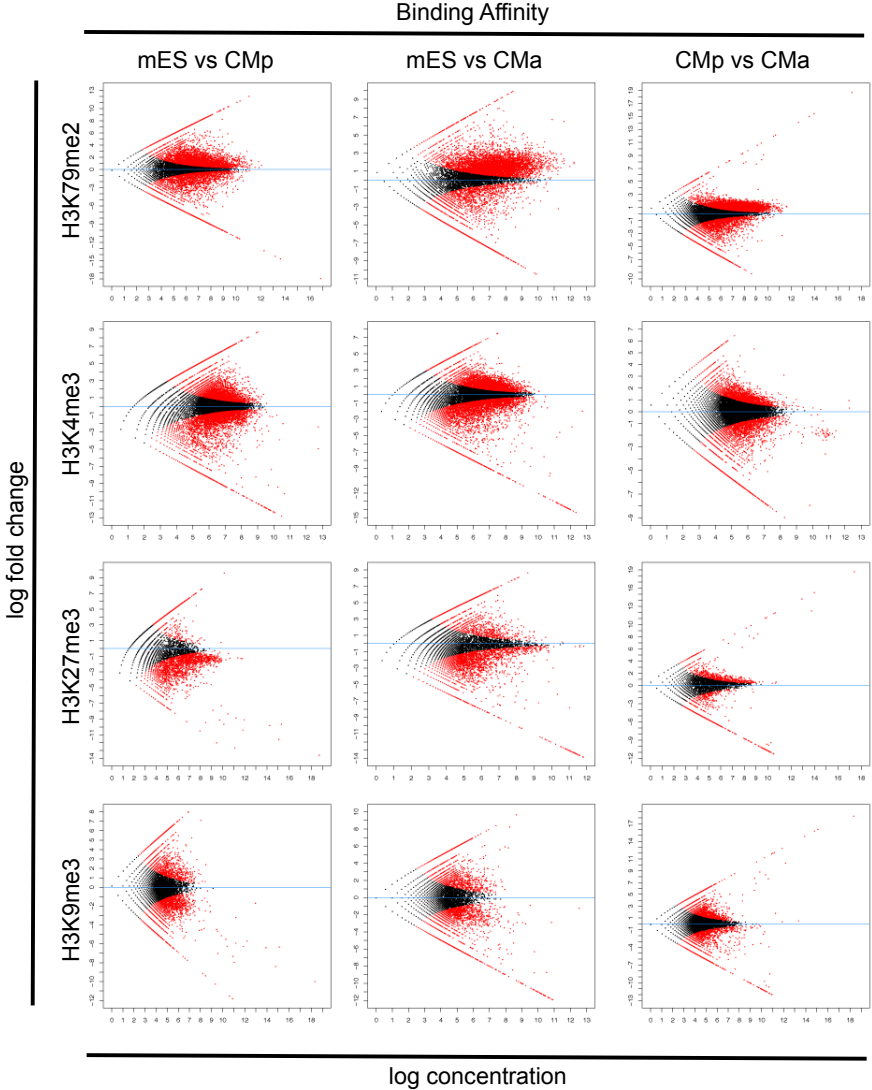
Figure 17 - Differential binding affinity of histone marks. The MA plots show the differentially bound sites for H3K79me2, H3K27me3, H3K4me3 and H3K9me3 that results from three different comparative analyses: mES vs CMp (left), mES vs CMa (middle) and CMp vs CMa (left). Each point represent a binding site, with point in red having an absolute log fold difference of at least 2.

A useful way to visualize which of the data-points are being identified as differentially bound is the analysis with MA plots. Each point represents a binding site, the red points represents the binding sites that have an absolute log fold difference of at least 2 resulting differentially presented in the two samples compared. From this analysis results evident that more sites are differentially presented between mES and CMs both neonatal and adult; then between CMp and CMa is highlighted the intermediate stage of differentiation of the 1day-old CMs compared with the adult CMs, in which the transcription profile should be already defined (Figure 17).

4.1.3.5 Statistical association of the genomic region identified with nearby genes (GREAT) and GO pathways identification

To identify the genes that were proximal and distal to the differential peaks we have analyzed the data set of the most significant ChIP-seq peaks with GREAT: Genomic Regions Enrichment of Annotation Tool software which predict functions of cis-regulatory regions (http://bejerano.stanford.edu/great/public/html/) [107].

GREAT calculates statistics by associating genomic regions with nearby genes and applying the gene annotations to the regions. Association is a two-step process: first, every gene is assigned to a regulatory domain, then, each genomic region is associated with all genes whose regulatory domain it overlaps. Each gene is assigned to a basal regulatory domain of a minimum distance of 5kb upstream and 1kb downstream of the TSS (regardless of other nearby genes). The regulatory domain is extended in both directions to the nearest gene's basal domain but no more than the maximum extension in one direction.

From an input file of genomic regions GREAT associates both proximal and distal input genomic regions with their putative target genes and uses gene annotations from numerous ontologies to associate genomic regions with annotations. Furthermore this analysis allowed us to calculate statistical enrichment for associations between genomic regions and annotations and to create track of the more significant pathways. This permitted us to have information on the specific function of the genomic regions regulated by H3K79me2, H3K4me3 H3K27me3 and H3K9me3 in mES compared with CMp compared with CMa and to clustered the associated genes in GO related pathways. We found relevant changes in the genome-wide distribution of the histone modifications analyzed. This revealed that histone modifications might regulate the expression of genes involved in fundamental cardiac functions: H3K79me2 and H3K4me3 were modulated in putative regulatory regions involved in RNA metabolism in mES cells vs. CMp and CMa, whereas in CMs these modifications were enriched in genes encoding cardiomyocyte structural proteins and in regions involved in cardiac phenotype specification; H3K79me2 underwent a change in distribution between CMp and CMa stages, being more present in regions involved in cellular metabolism in CMa and in genes regulating cytoskeleton assembly in CMp; in contrast, when comparing mES vs. CMp or CMa, the repressor mark H3K27me3 was lost in genomic regions harboring genes involved in actin filament organization and was acquired in regions involved in nervous system formation. GO analysis revealed that H3K79me2 became progressively enriched in cardiomyocyte-specific genes involved in cardiomyocyte structure and function. This finding underscores the importance of these epigenetic modifications in directing cell fate decisions for cardiac differentiation (Figure 18).

Figure 18A - Functional annotation of differentially enriched modifications was performed using GREAT. The top 5 over-represented categories belonging to the Gene Ontology (GO) Biological Process are shown. The x axes values (in logarithmic scale) correspond to the binomial raw (uncorrelated) P-values. For each marks (H3K79me2, H3K4me3, H3K27me3 and H3K9me3), Blue histograms represent mES, red histograms CMp and green histograms CMa

GO Biological Processes

mES vs CMa

-log10(Binomial p value)

H3K79me2

0 5 10 15 20 25 30 35 40 45 50 55 60

- RNA processing
- response to DNA damage stimulus
- DNA metabolic process
- ncRNA metabolic process
- translation

0 1 2 3 4 5 6 7 8 9

- actin filament organization
- cardiac muscle contraction
- myofibril assembly
- actomyosin structure organization
- generation of precursor metabolites and energy

H3K4me3

0 5 10 15 20 25 30 35

- DNA metabolic process
- ncRNA metabolic process
- translation
- M phase
- DNA repair

0 1 2 3 4 5 6 7 8

- face morphogenesis
- head morphogenesis
- central nervous system vasculogenesis
- generation of precursor metabolites and energy
- positive regulation of receptor activity

H3K27me3

0 2 4 6 8 10 12 14

- actin filament-based process
- actin cytoskeleton organization
- cellular response to peptide hormone stimulus
- negative regulation of transferase activity
- negative regulation of kinase activity

0 1 2 3 4 5 6 7 8 9 10 11 12 13

- cerebellar cortex development
- floor plate development
- floor plate morphogenesis
- cerebellar granule cell precursor proliferation
- metencephalon development

H3K9me3

0 2 4 6 8 10 12 14 16

- macromolecule catabolic process
- cellular macromolecule catabolic process
- protein catabolic process
- actin filament-based process
- actin cytoskeleton organization

0 1 2 3 4 5 6 7 8 9

- head morphogenesis
- body morphogenesis
- face morphogenesis
- formation of primary germ layer
- oviduct development

Figure 18A - Continued

41

GO Biological Processes

CMp vs CMa

-log10(Binomial p value)

H3K79me2

0   5   10   15   20   25   30   35

- small protein conjugation or removal
- intracellular transport
- cytoskeleton organization
- protein ubiquitination
- ification by small protein conjugation

0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0

- regulation of cholesterol transport
- midbrain-hindbrain boundary development
- trophoblast giant cell differentiation
- rostrocaudal neural tube patterning
- transcription from RNA polymerase I promoter

H3K4me3

0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0

- sterol biosynthetic process
- cholesterol biosynthetic process
- glial cell migration

0   1   2   3   4   5   6   7   8   9   10

- peptidyl-lysine modification
- protein acetylation
- peptidyl-lysine acetylation
- internal peptidyl-lysine acetylation
- regulation of gene expression, epigenetic

H3K27me3

0   2   4   6   8   10   12   14   16   18

- pattern specification process
- regionalization
- cell fate commitment
- neuron differentiation
- anterior/posterior pattern specification

0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0

- metencephalon development
- bone remodeling
- formation of primary germ layer
- cerebellum development
- tissue remodeling

H3K9me3

0   1   2   3   4   5   6   7   8   9

- covalent chromatin modification
- regulation of translation
- cell-substrate adhesion
- regulation of gene expression, epigenetic
- peptidyl-lysine modification

0   1   2   3   4   5   6   7

- macromolecule methylation
- body morphogenesis
- head morphogenesis
- endoderm development
- face morphogenesis
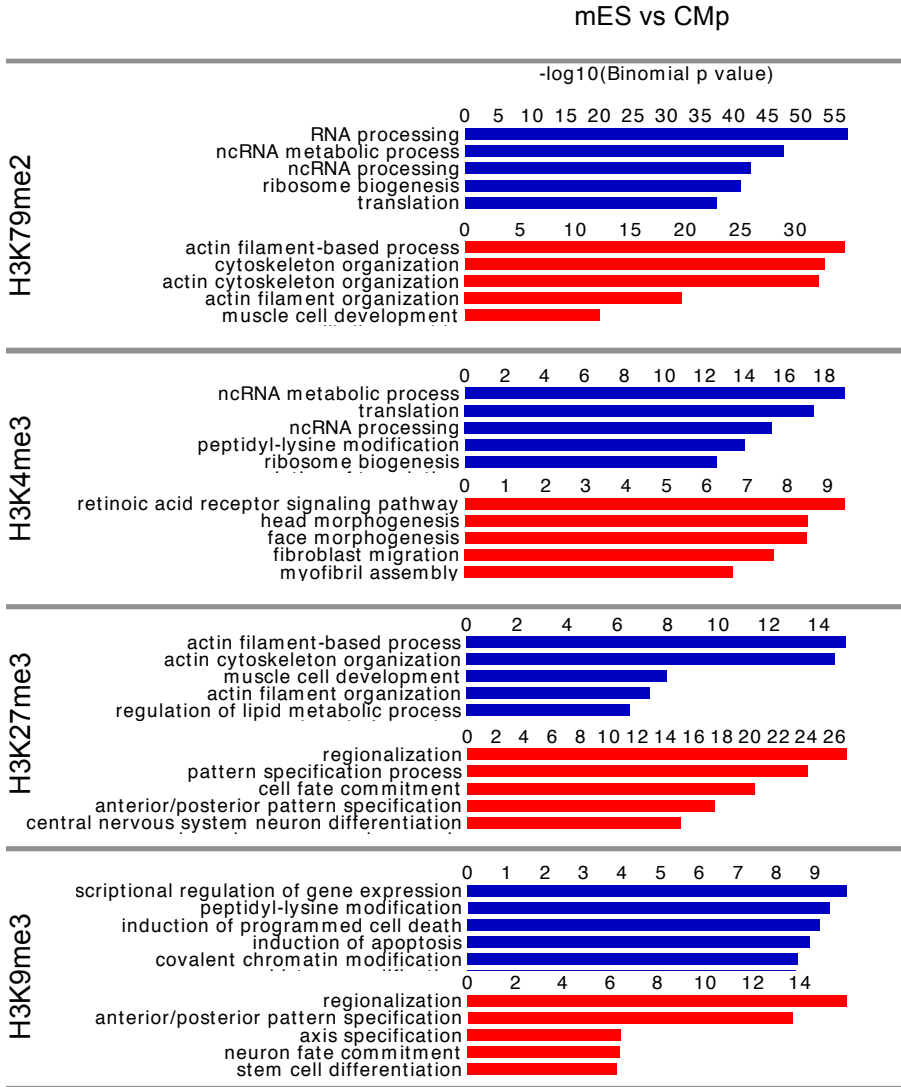
Figure 18A - Continued

42

Figure 18B - Functional annotation of differentially enriched modifications was performed using GREAT. The top 5 over-represented categories belonging to the Gene Ontology (GO) Molecular Functions are shown. The x axes values (in logarithmic scale) correspond to the binomial raw (uncorrelated) P-values. For each marks (H3K79me2, H3K4me3, H3K27me3 and H3K9me3), Blue histograms represent mES, red histograms CMp and green histograms CMa

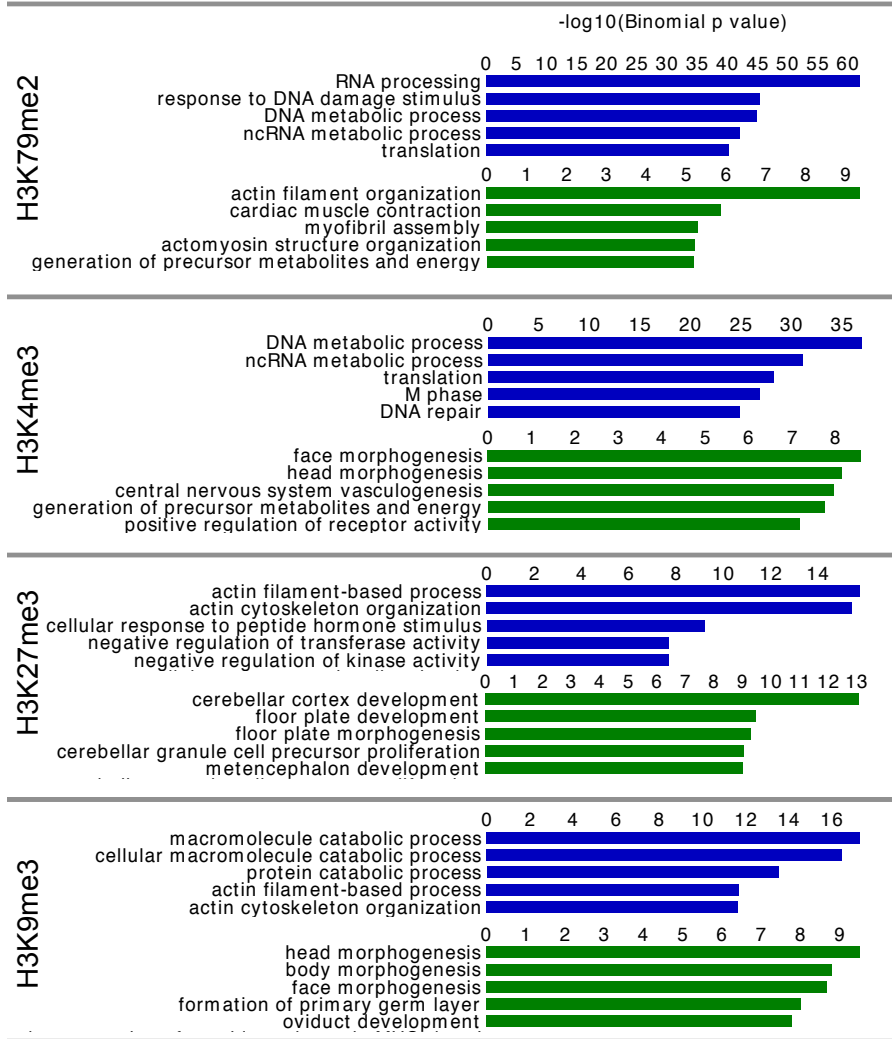# GO Molecular Functions

mES vs CMa

Figure 18B - Continued

GO Molecular Functions

CMp vs CMa

-log10(Binomial p value)

H3K79me2

	0 2 4 6 8 10 12 14 16 18 20 22
small conjugating protein ligase activity	
ubiquitin-protein ligase activity	
acid-amino acid ligase activity	
protein transporter activity	
unfolded protein binding	

0.0 0.5 1.0 1.5 2.0 2.5 3.0

RNA polymerase II distal enhance

H3K4me3

0   1   2   3   4   5   6   7   8

- histone-lysine N-methyltransferase activity
- protein phosphorylated amino acid binding
- phosphotyrosine binding
- transforming growth factor beta-activated receptor activity
- 14-3-3 protein binding

H3K27me3

0   5   10   15   20   25

- sequence-specific DNA binding
- sequence-specific DNA binding transcription factor activity
- nucleic acid binding transcription factor activity
- regulatory region DNA binding
- transcription regulatory region DNA binding

0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4

platelet-derived growth factor binding

H3K9me3

0.0   0.5   1.0   1.5   2.0   2.5   3.0

protein kinase inhibitor activity

0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0

- histone-lysine N-methyltransferase activity
- protein-lysine N-methyltransferase activity
- protein methyltransferase activity

Figure 18B - Continued

GO Cellular Component

mES vs CMp

-log10(Binomial p value)

H3K79me2

| 0 10 20 30 40 50 60 |
| ribonucleoprotein complex |
| nucleolus |
| chromosome |
| chromosomal part |
| nuclear body |

| 0 5 10 15 20 25 30 35 |
| actin cytoskeleton |
| contractile fiber |
| myofibril |
| focal adhesion |
| contractile fiber part |

H3K4me3

| 0 2 4 6 8 10 12 14 16 18 20 22 24 26 |
| ribonucleoprotein complex |
| actomyosin |
| PML body |
| ribonucleoprotein granule |
| histone acetyltransferase complex |

| 0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 |
| endoplasmic reticulum lumen |
| voltage-gated calcium channel complex |
| calcium channel complex |
| integral to membrane of membrane fraction |

H3K27me3

| 0 2 4 6 8 10 12 14 16 |
| actin cytoskeleton |
| focal adhesion |
| cell-substrate adherens junction |
| actin filament bundle |
| cell-substrate junction |

| 0 2 4 6 8 10 12 14 |
| transcription factor complex |
| anchored to plasma membrane |

H3K9me3

| 0 2 4 6 8 10 12 |
| actin cytoskeleton |
| cell leading edge |
| anchoring junction |
| focal adhesion |
| adherens junction |

Figure 18C - Functional annotation of differentially enriched modifications was performed using GREAT. The top 5 over-represented categories belonging to the Gene Ontology (GO) Cellular Component are shown. The x axes values (in logarithmic scale) correspond to the binomial raw (uncorrelated) P-values. For each marks (H3K79me2, H3K4me3, H3K27me3 and H3K9me3), Blue histograms represent mES, red histograms CMp and green histograms CMa
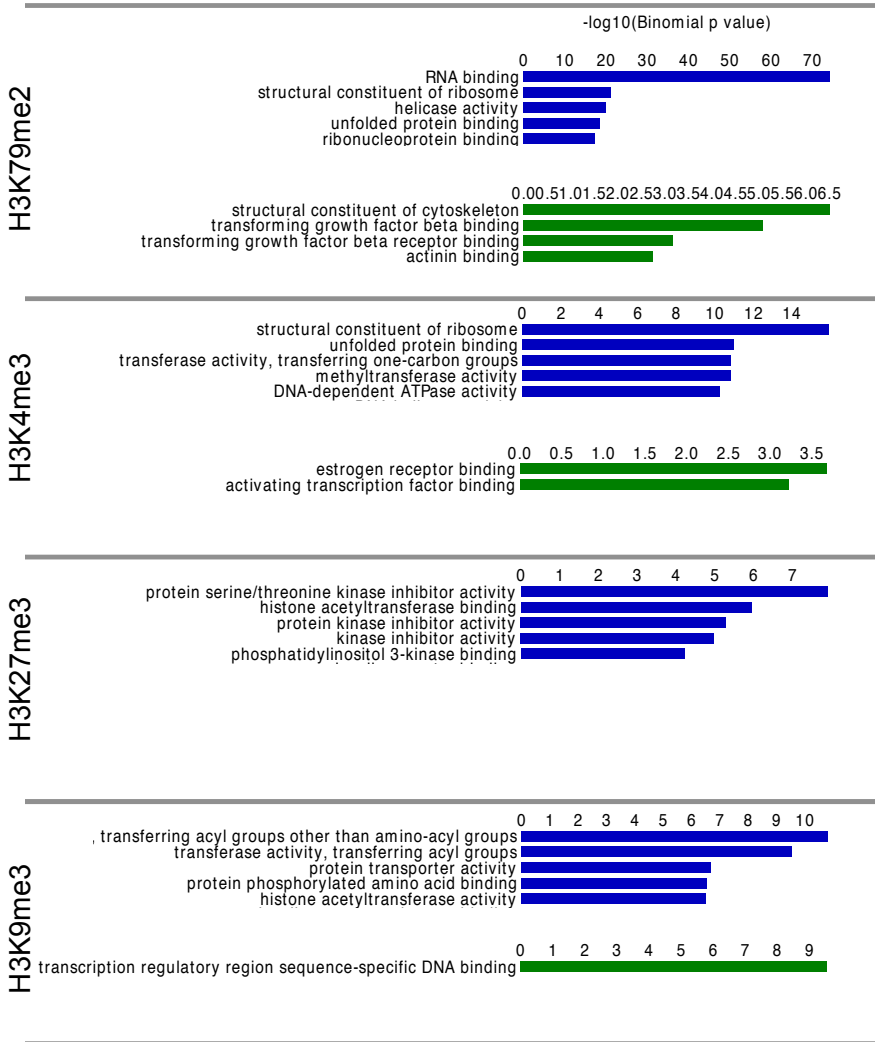
46

GO Cellular Component

mES vs CMa

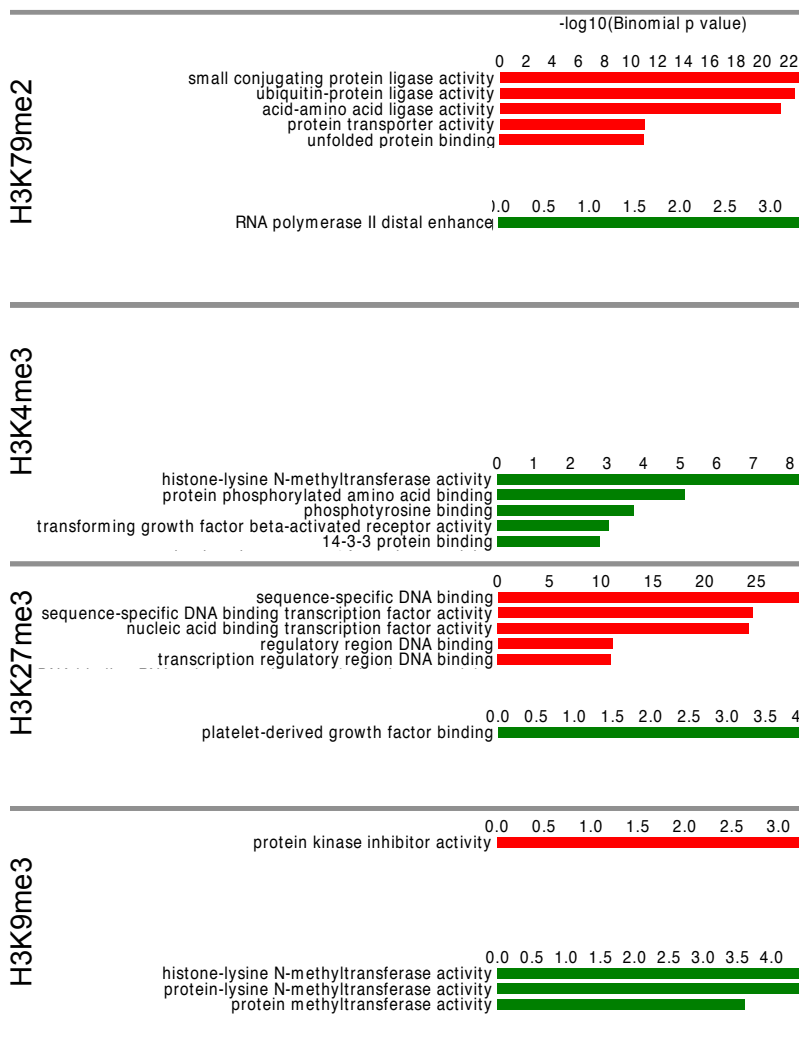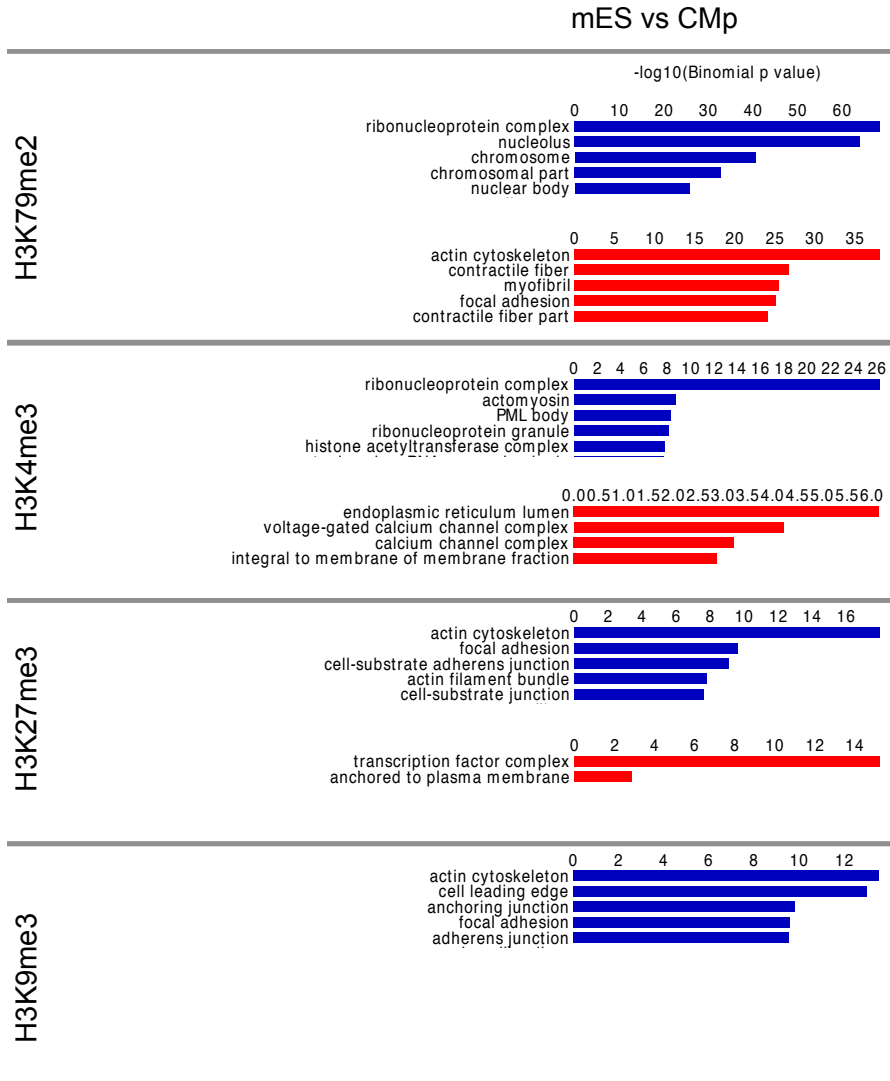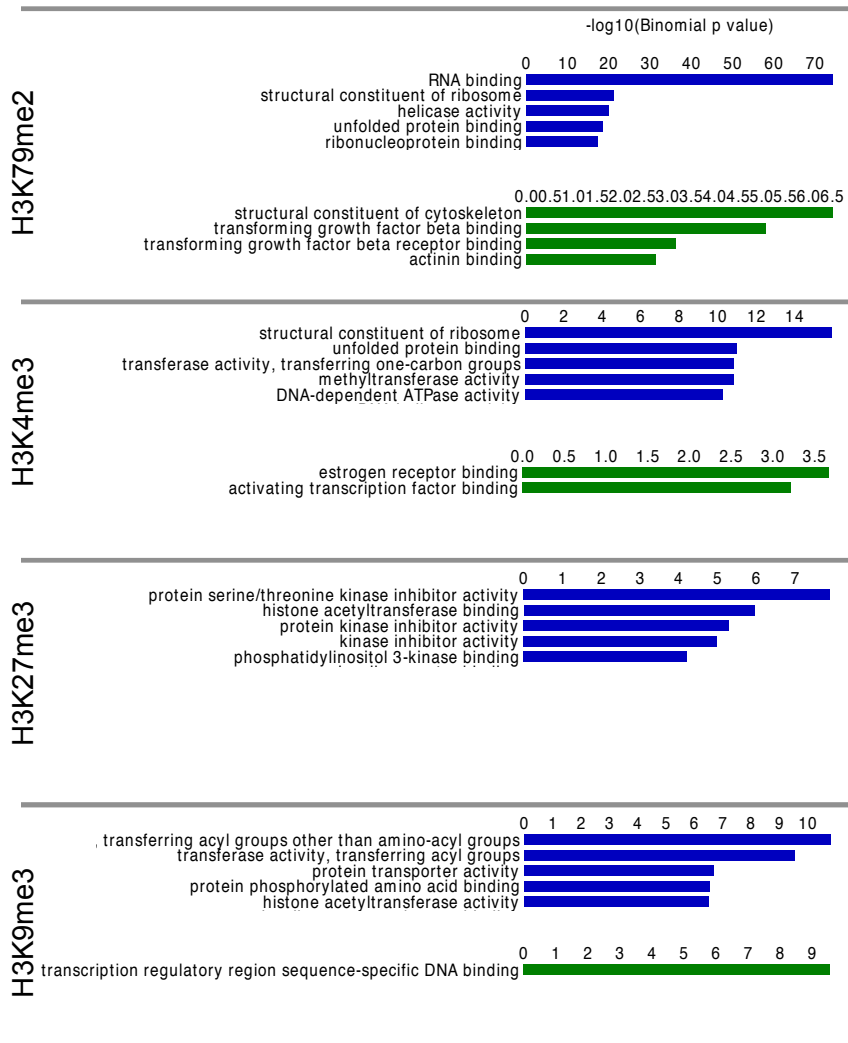Figure 18C - Continued

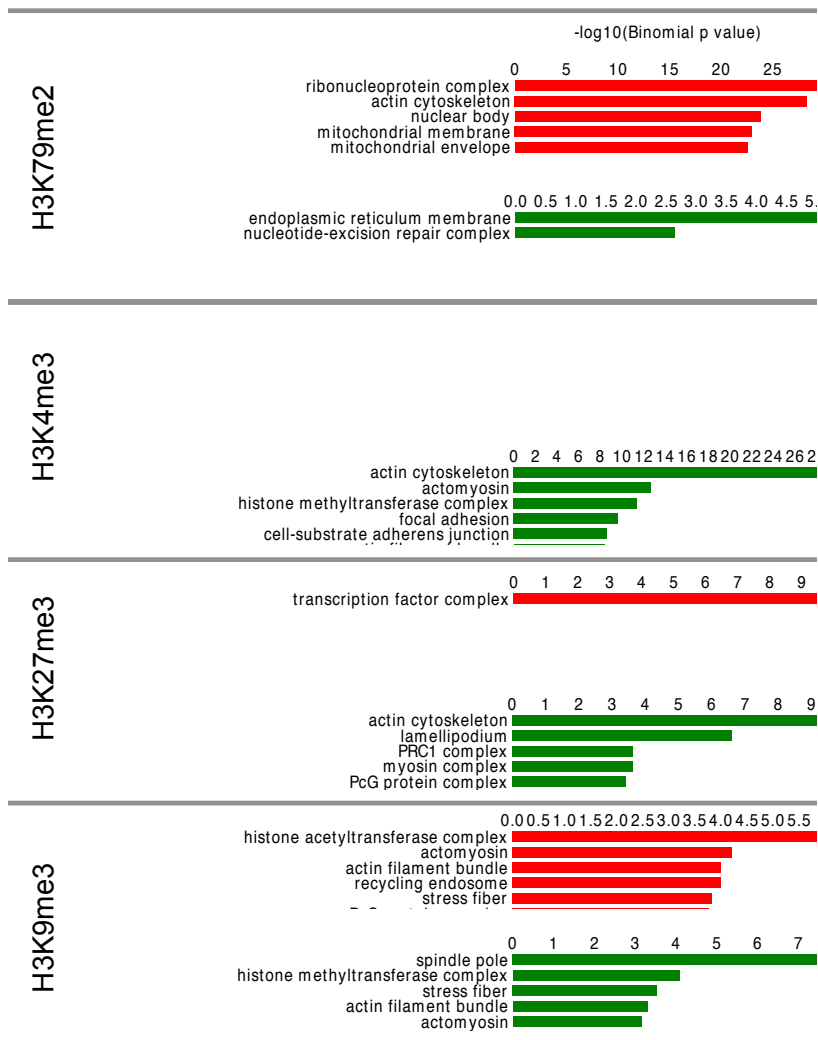# GO Cellular Component

CMp vs CMa

Figure 18C - Continued
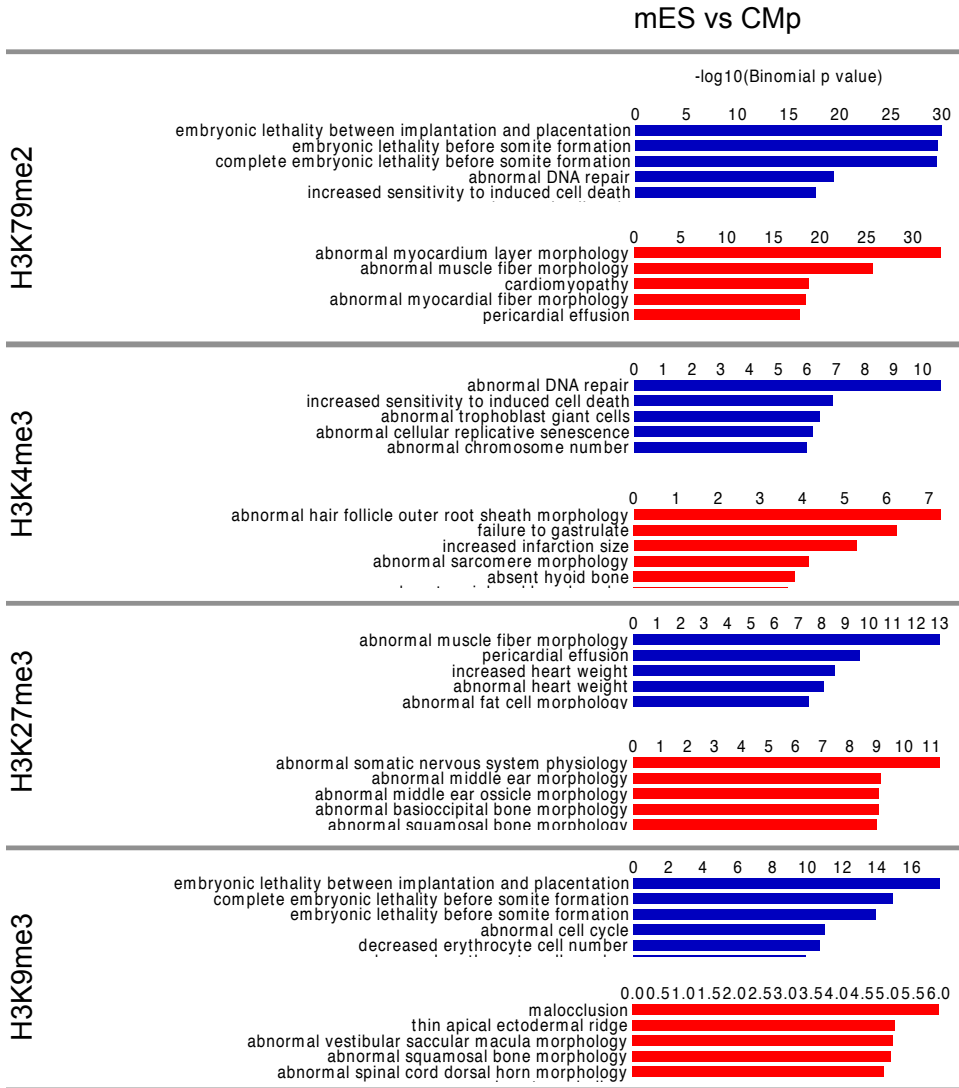
GO Mouse Phenotype

Figure 18D - Functional annotation of differentially enriched modifications was performed using GREAT. The top 5 over-represented categories belonging to the Gene Ontology (GO) Mouse Phenotype are shown. The x axes values (in logarithmic scale) correspond to the binomial raw (uncorrelated) P-values. For each marks (H3K79me2, H3K4me3, H3K27me3 and H3K9me3), Blue histograms represent mES, red histograms CMp and green histograms CMa
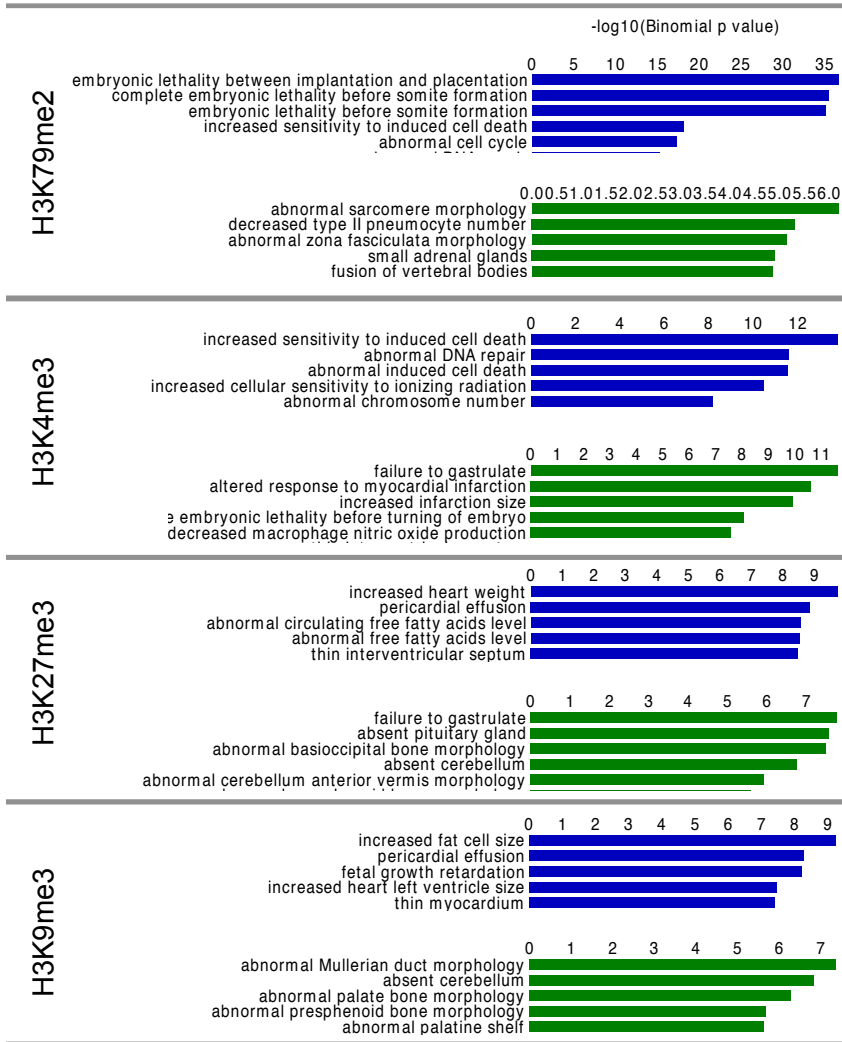
GO Mouse Phenotype

mES vs CMa

Figure 18D - Continued
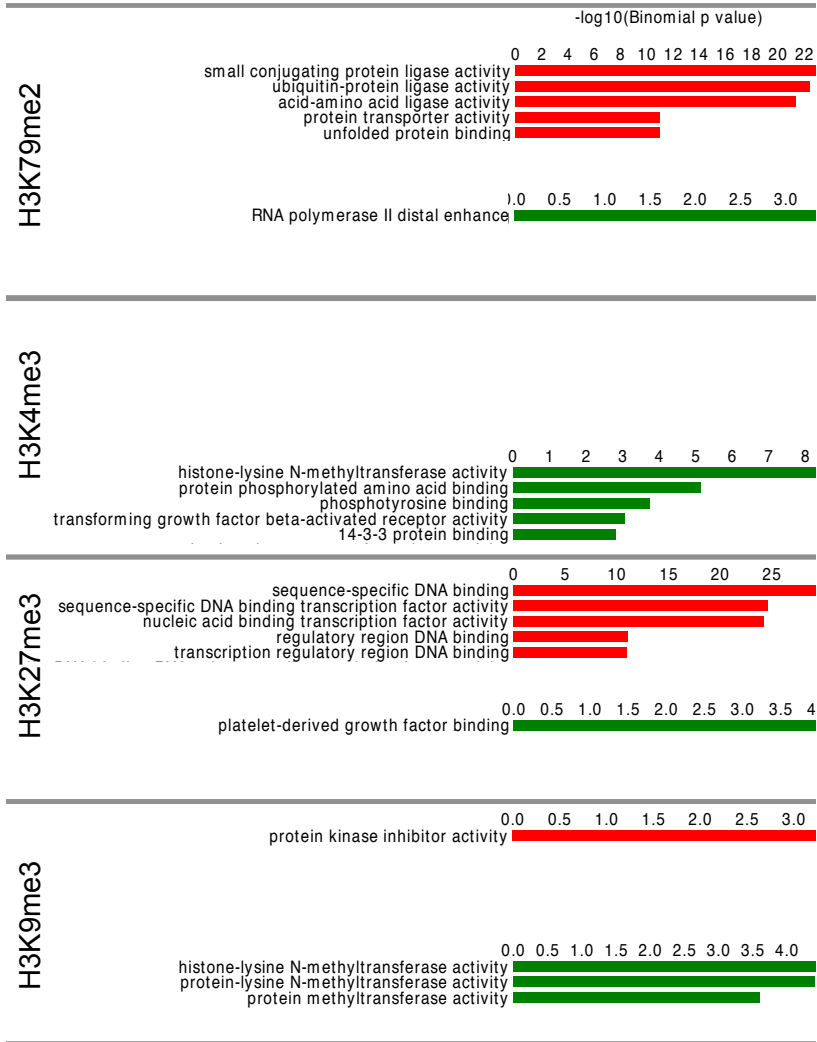
GO Mouse Phenotype

CMp vs CMa

Figure 18D - Continued

4.1.3.6 IGV genomic map visualization

Moreover, ChIP-seq data from single genes modulated during cardiac differentiation (e.g. Nanog, Pou5F1, Gata4, Nkx2-5, Myh6, Myh7, TnnI3) were inspected with either the UCSC Genome Browser (http://genome.ucsc.edu/) or the IGV2.1 bioinformatics program (http://www.broad[...]
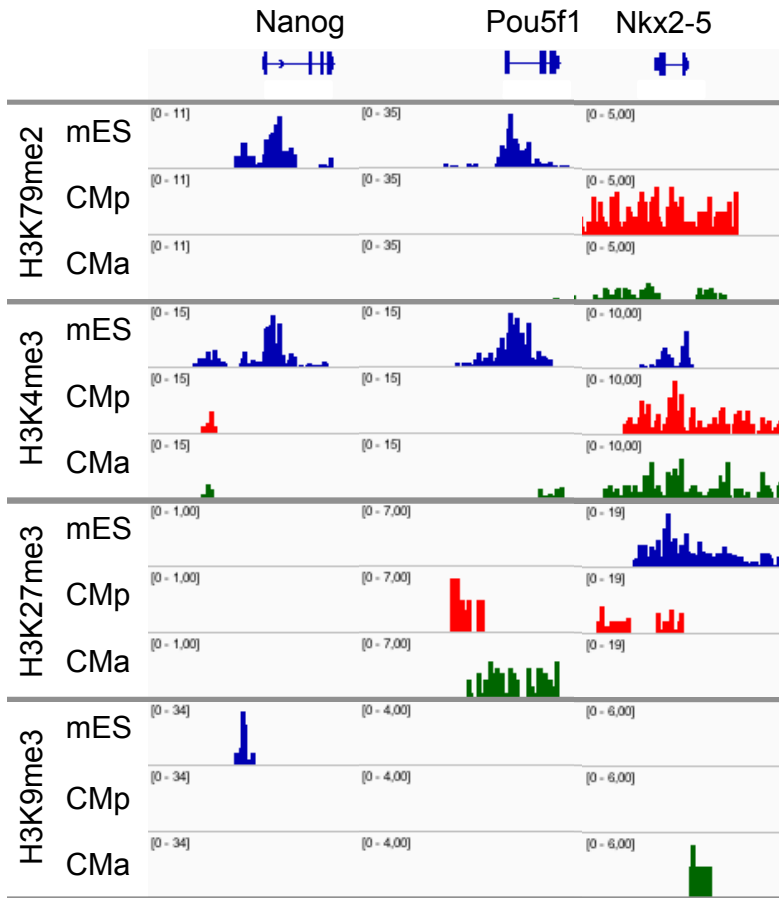
shown for key gene loci of pluripotency (Nanog, Pou5F1), cardiac transcription regulation (Gata4, Nkx2-5) and cardiomyocyte specification (Myh6, Myh7, TnnI3). The peak height is set to correspond to the maximum enriched peak in the region for each mark.

An alteration in the profile of these markers is shown during cardiac differentiation, analyzing key genes of pluripotency and heart development, H3K79me2 and H3K4me3 are present on the locus of Nanog and Pou5F1 in mES and are lost in CMp and CMa and H3K27me3 is acquired in CMp and CMa on the locus of

chr14

qA1   qA2   qA3   qB

63.820 kb   63.840 kb   63.860 k

Gata4

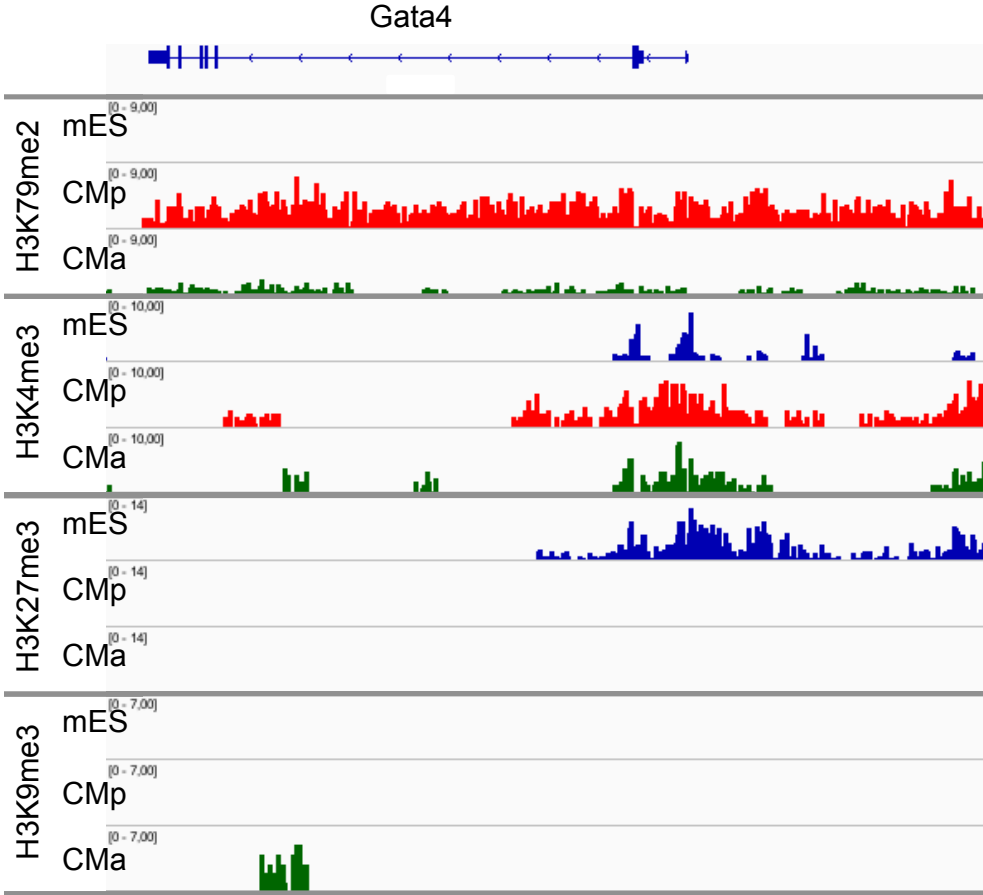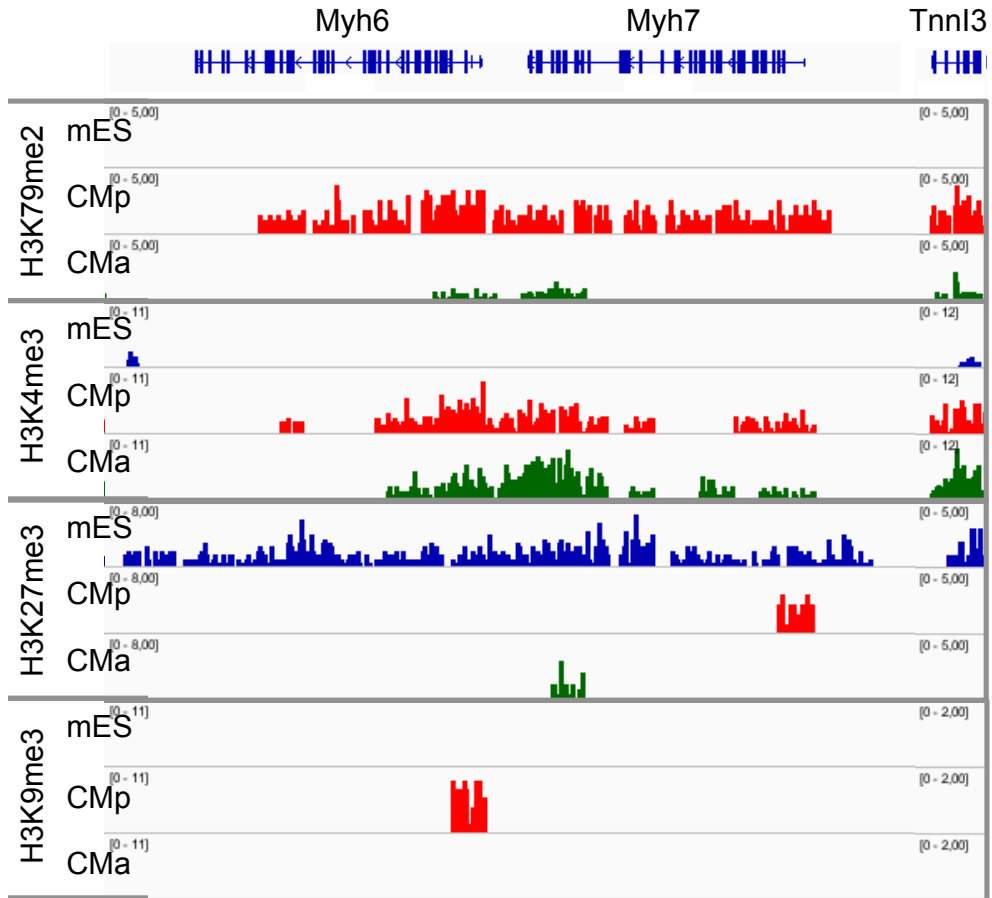| Refseq genes	Refseq genes	Refseq genes
ES_H3K79me2-islandfilter	ES_H3K79me2-islandfiltere	ES_H3K79me2-islandfiltered	[0 - 9,00]
CMN_H3K79me2-islandfilt	CMN_H3K79me2-islandfilte	CMN_H3K79me2-islandfiltered	[0 - 9,00]
Sham_sample_6-islandfilte	Sham_sample_6-islandfilter	Sham_sample_6-islandfiltered	[0 - 9,00]
ES_H3K4me3-islandfiltere	ES_H3K4me3-islandfiltered	ES_H3K4me3-islandfiltered	[0 - 10,00]
CMN_H3K4me3-islandfilter	CMN_H3K4me3-islandfiltere	CMN_H3K4me3-islandfiltered	[0 - 10,00]
Sham_sample_7-islandfilte	Sham_sample_7-islandfilter	Sham_sample_7-islandfiltered	[0 - 10,00]
ES_H3K27me3-islandfilter	ES_H3K27me3-islandfiltere	ES_H3K27me3-islandfiltered	[0 - 14]
CMN_H3K27me3-islandfilt	CMN_H3K27me3-islandfilter	CMN_H3K27me3-islandfiltered	[0 - 14]
Sham_sample_5-islandfilte	Sham_sample_5-islandfilter	Sham_sample_5-islandfiltered	[0 - 14]
ES_H3K9me3-islandfiltere	ES_H3K9me3-islandfiltered	ES_H3K9me3-islandfiltered	[0 - 7,00]
CMN_H3K9me3-islandfilter	CMN_H3K9me3-islandfiltere	CMN_H3K9me3-islandfiltered	[0 - 7,00]
Sham_sample_4-islandfilte	Sham_sample_4-islandfilter	Sham_sample_4-islandfiltered	[0 - 7,00]

Figure 15 – Continued

4.2 The epigenetic code associated with histone methylation correlate with the transcriptional program of cardiomyocytes at different stages of development.

4.2.1 mRNA microarray gene expression analysis

To investigate whether changes in the distribution of the histone modifications analyzed are associated with gene reprogramming accompanying heart development, we analyzed the mRNA gene expression profile of mES, CMp and CMa. Neonatal and adult CMs were purified as described in sections 3.1 and 4.1.2.1. Total RNA was extracted from undifferentiated mES cells (TBV2 line),

54

neonatal and adult CMs using TRIzol method (Invitrogen). The mRNA obtained was quantified with NanoDrop Spectrophotometer and the quality was tested through Agarose gel and Agilent, only samples with RIN number > 8 were used.

Gene expression profiling of mES, CMp and CMa were obtained by Illumina microarray. Briefly, 500ng of RNA were transcribed to cRNA by Illumina TotalPrep RNA Amplification Kit (Ambion) and a total of 500ng of cRNA was hybridized at 58°C for 16 hours on the Illumina MOUSEWT-6V BeadChips (Illumina). Bead-Chips were scanned using Illumina BeadArray Reader and the Bead Scan software. Data were processed with BeadStudio version 3 and successively analyzed with MeV_3 software. Three independent biological replicate were analyzed for each samples. PCA analysis shows a significant correlation within the samples of each group (Figure 20).
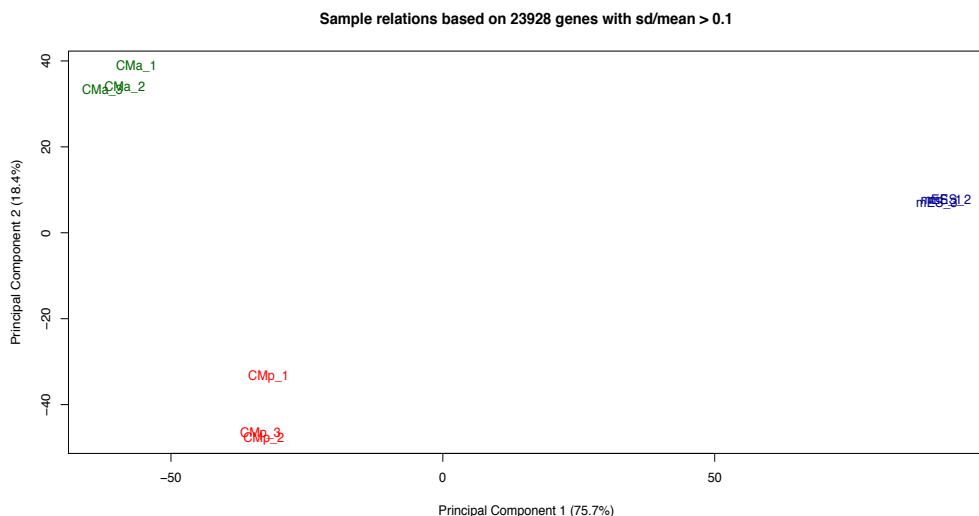
Sample relations based on 23928 genes with sd/mean > 0.1

Figure 20 – PCA plot Illumina analysis

Statistical analysis of raw data revealed that 4607 genes significantly modulated in mES vs. CMp, 5258 genes in mES vs. CMa, and 2163 genes differentially expressed between CMp and CMa. Heararchical clustering shown three main groups, wuth a closer proximity between CMp and CMa compared with mES. In red are presented gene whose expression is up regulated and in green genes down regulated; is evident that genes highly expressed in mES are not expressed in CMp and CMa, within this group of genes no big difference are shown between neonatal and adult CMs (Figure 21).
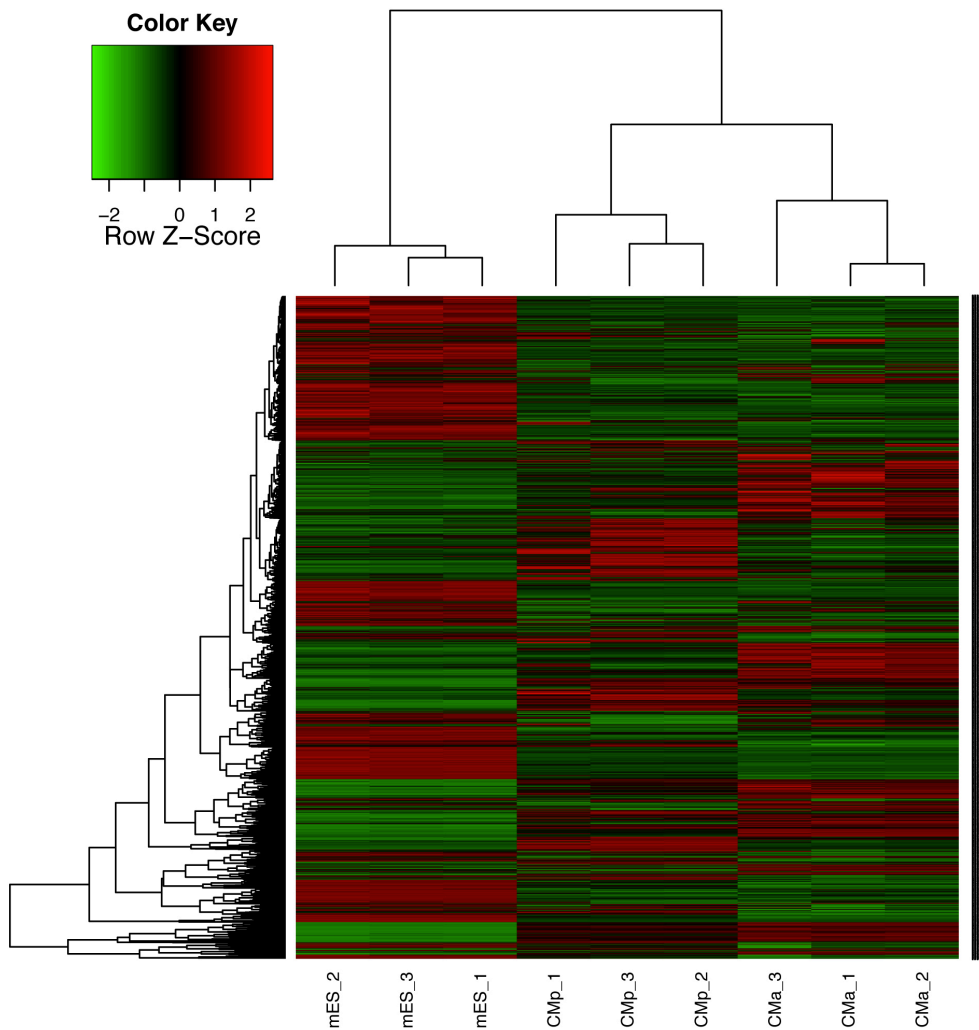
Figure 21 - Gene expression profile of cardiomyotytes at different stages of differentiation. Illumina mRNA gene expression datasets are presented for mouse Embryonic Stem cells (mES), post-natal cardiomyocytes (CMp) an adult cardiomyocytes (CMa). Genes were Z-score normalized and data were subjected to hierarchical clustering (centered correlation distance, centroid linkage). The results from three biological replicates are presented, samples with higher than average expression are shown in red and samples with lower than average expression are shown in green.

As described in the section 4.1.3.4 a comparative analysis was performed within the groups (mES vs CMp; mES vs CMa; CMp vs CMa) in order to identify the up-regulated genes characteristic for each cell type analyzed. Figure 22 revealed that a high percentage of genes resulted differentially expressed within mES and CMs (both neonatal and adult), whereas a lower percentage of genes are differentially expressed within CMp and CMa.
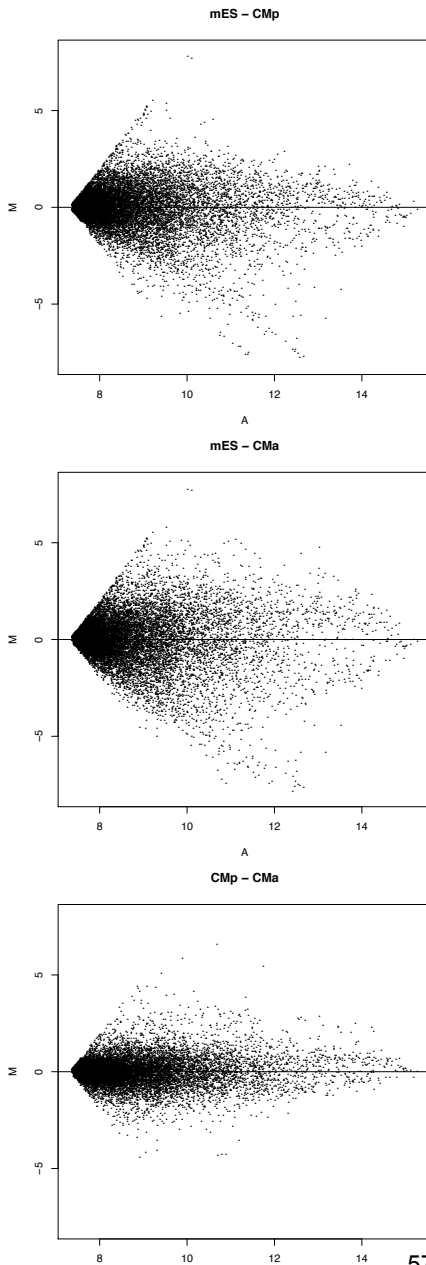
Figure 22 – MA plot resulting from Illumina gene expression analysis. Three comparative analysis are presented: mES vs CMp, mES vs CMa and CMp vs CMa

4.2.2 Functional clustering of differential expressed genes.

Kegg pathway classification, performed with DAVID bioinformatics software (http://david.abcc.ncifcrf.gov/), revealed that the up-regulated genes in CMs are involved in a variety of cellular processes implicated in heart development and cardiac structure and function; down-regulated genes were found to be part of cellular and signal transduction pathways involved in stem cells metabolism (Figure 23).

mES vs CMp

mES vs CMa

GO Biological Process

RNA processing
cell cycle
mRNA processing
mRNA metabolic process
ribonucleoprotein complex biogenesis
cell cycle phase
RNA splicing
cell cycle process
ncRNA metabolic process
DNA metabolic process

1.0E-30 1.0E-50 1.0E-70

generation of precursor metabolites and energy
electron transport chain
oxidation reduction
cofactor metabolic process
energy derivation by oxidation of organic compounds
cellular respiration
coenzyme metabolic process
cofactor biosynthetic process
oxidative phosphorylation
muscle organ development

1.0E-05 1.0E-20 1.0E-35

GO Molecular Function

RNA binding
nucleotide binding
helicase activity
purine ribonucleotide binding
ribonucleotide binding
ATP-dependent helicase activity
purine NTP-dependent helicase activity
DNA binding
purine nucleotide binding
ATP binding

1.0E-10 1.0E-30 1.0E-50

red-ox activity, acting on NADH or NADPH
cofactor binding
red-ox activity, acting on NADH or NADPH
NADH dehydrogenase activity
NADH dehydrogenase (ubiquinone) activity
NADH dehydrogenase (quinone) activity
nt inorganic cation transmb transporter activity
drogen ion transmembrane transporter activity
nucleotide binding
purine nucleotide binding

1.0E+00 1.0E-05 1.0E-10

GO Cellular Component

intracellular non-membrane-bounded organelle
non-membrane-bounded organelle
nuclear lumen
intracellular organelle lumen
organelle lumen
membrane-enclosed lumen
nucleolus
ribonucleoprotein complex
chromosome
chromosomal part

1.0E-40 1.0E-65 1.0E-90

mitochondrion
mitochondrial part
mitochondrial envelope
mitochondrial membrane
mitochondrial inner membrane
organelle inner membrane
organelle membrane
organelle envelope
envelope
respiratory chain

1.0E-10 1.0E-50 1.0E-90

KEGG Pathway

Spliceosome
Ribosome
Cell cycle
DNA replication
Nucleotide excision repair
Mismatch repair
Pyrimidine metabolism
RNA degradation
Homologous recombination
Aminoacyl-tRNA biosynthesis

1.0E+00 1.0E-15 1.0E-30

Oxidative phosphorylation
Alzheimer's disease
Parkinson's disease
Huntington's disease
Cardiac muscle contraction
Valine, leucine and isoleucine degradation
Dilated cardiomyopathy
Hypertrophic cardiomyopathy (HCM)
Citrate cycle (TCA cycle)
hythmogenic right ventricular cardiomyopathy
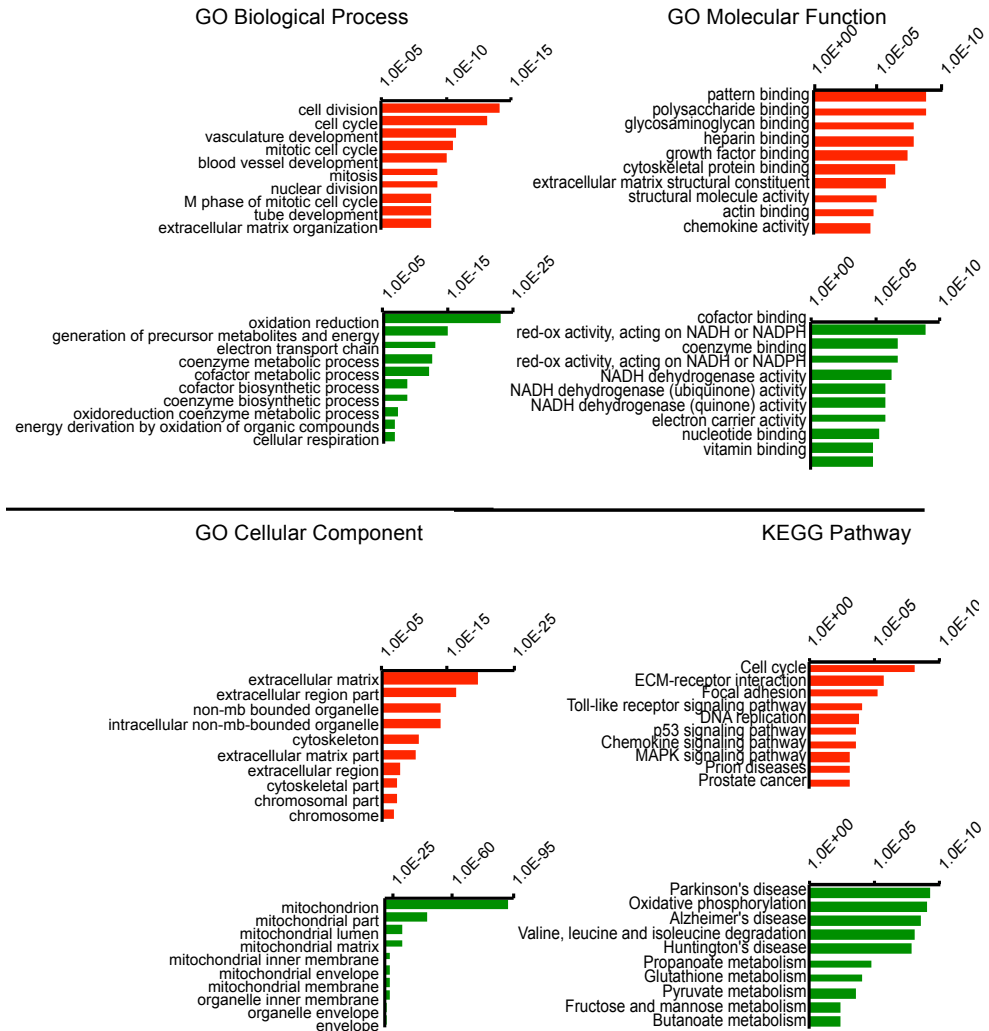
1.0E+00 1.0E-15 1.0E-30

Figure 23 - Functional and molecular characterization of mRNA expression in cardiomyocytes during differentiation was performed using DAVID. The top 10 over-represented categories belonging to the Gene Ontology (GO) Biological Process, GO Cellular Component and GO Molecular Function and to the KEGG Pathway are shown. The x axes values correspond to the P-values. For each category results from three different comparative analyses are shown: mES vs CMp, mES vs CMa and CMp vs CMa. Blue histograms represent mouse Embryonic Stem cells (mES), red histograms post-natal cardiomyocytes (CMp) and green histograms adult cardiomyocytes (CMa).

4.2.3 Statistical correlation between ChIP-seq data and mRNA gene expression profile data.

To investigate whether changes in the distribution of the histone modifications analyzed are associated with gene reprogramming accompanying heart development, we cross-compared ChIP-seq libraries from mES and CMs with Illumina gene expression data and we found that many genes up-regulated in CMs are also enriched in histone methylation markers associated with transcriptional activation.

The heatmaps of figure 24 show a significant correlation (H3K79me2, H3K4me3) and anti-correlation (H3K27me3 and H3K9me3) between the enrichment of a histone methylation modification and the transcriptional profile. Genes are ordered by the magnitude of differential histone methylation occupancy and the relative gene expression is shown, genes up-regulated are shown in red and genes down-regulated are shown in green.

The comparison between ChIP-seq fold change and Illumina gene expression fold change revealed that H3K79me2 is highly correlated with the transcriptional activation profile in CMs. Pearson correlation (R) in mES vs CMp results R=0.698; and in mES vs CMa R=0.607 (R values of correlation and anti-correlation are between +1 and -1) (Figure 25).

This result proves that histone methylation regulates the transcription program in developing cardiac cells and that H3K79me2 play a central role in this regulation. Supporting this idea and according with previously published studies in other differentiating systems H3K4me3 and H3K27me3 correlate and anti-correlate respectively with the expression profile, however no significant correlation was found for H3K9me3, this was expected because a small number of genes are regulated by this marker in our experimental conditions. Taken together this results showed a really significant correlation between histone methylation and transcription regulation.

The bivalent domain H3K4me3 and H3K27me3 defined as "poised" markers characteristic of differentiating cells, also in the cardiac differentiation model plays a key role and of particular attention is the role of H3K79me2, marker of transcriptional activation, that results enriched on the locus of the genes that are activated during heart development.

This result proves that histone methylation regulates the transcription program in developing cardiac cells and that H3K79me2 plays a central role in this regulation; moreover, histone marks distinguish functionally distinct genes with a similar expression pattern.
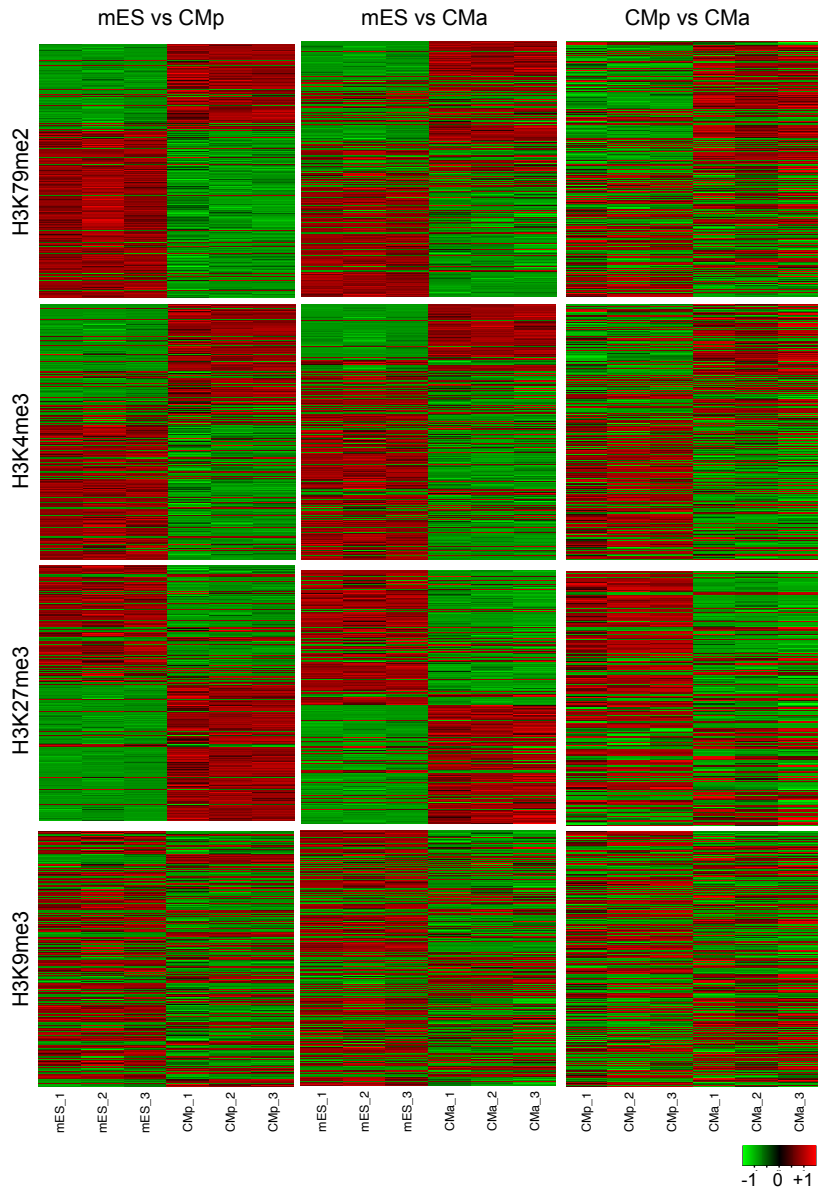
Figure 24 - Expression data for genes differentially occupied by H3K79me2, H3K4me3, H3K27me3 and H3K9me3 in mouse Embryonic Stem cells (mES), cardiomyocytes from post-natal day 1 (CMp) and cardiomyocytes from 2 month-old mice (CMa). Genes are ordered by magnitude of differential marks occupancy and relative gene expression is shown. Samples with higher than average expression are shown in red and samples with lower than average expression are shown in green. Results from three different comparative analyses are shown: mES vs CMp (left), mES vs CMa (middle) and CMp vs CMa (left).
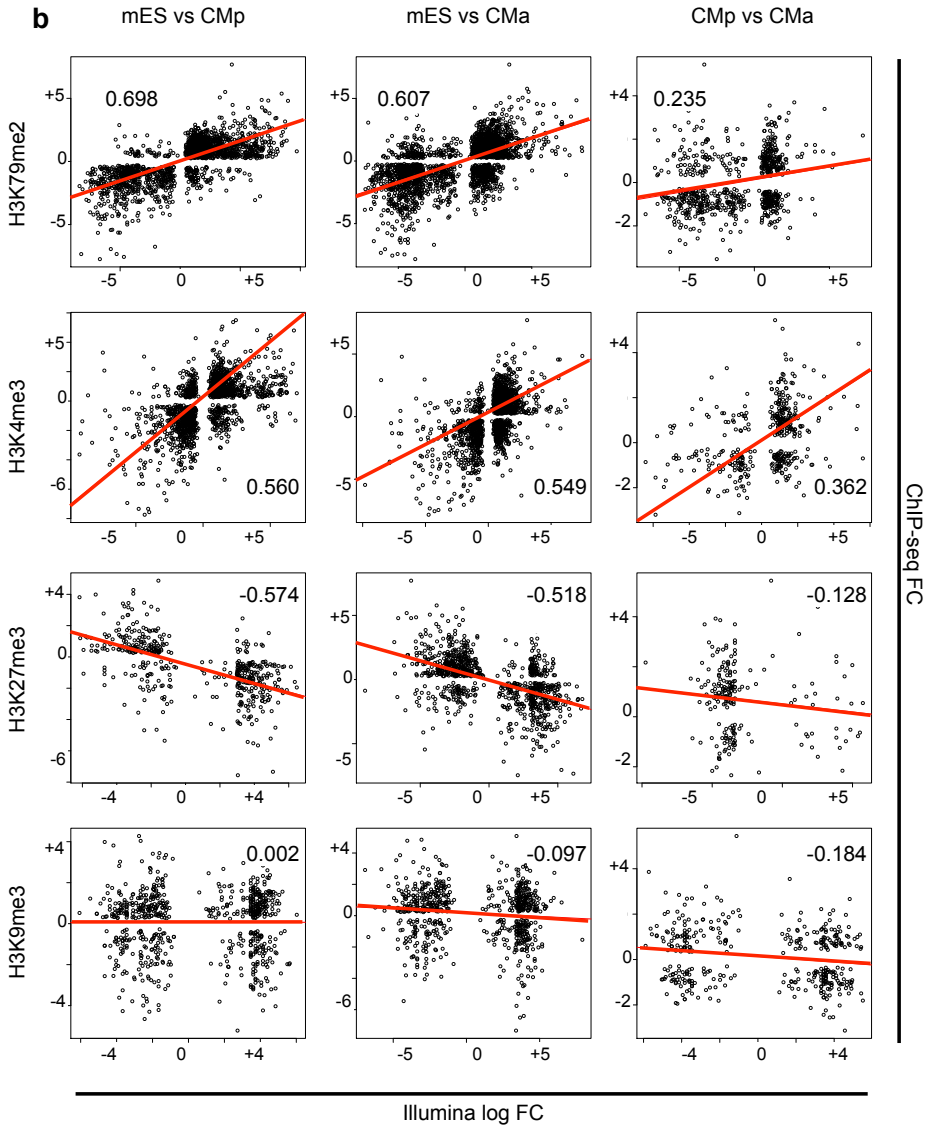
Figure 25 - Correlation analysis between ChIP-seq fold change values (ChIP-seq_FC) and gene expression fold change values in logaritmic scale (Illumina_logFC). The correlation index R for each comparison is indicated inside each graph. Red line indicates correlation or anti-correlation between the histone methylation marks enrichment and the expression level. Results from three different comparative analyses are shown: mES vs CMp (left), mES vs CMa (middle) and CMp vs CMa (left) for H3K79me2, H3K4me3, H3K27me3 and H3K9me3.

4.2.4 RealTime qPCR gene expression and qPCR-ChIP validation.

Real Time qPCR was performed in order to validate the gene expression profile obtained through Illumina microarray. Neonatal and adult CMs were purified as described in sections 3.1 and 4.1.2.1. Total RNA was extracted from undifferentiated mES cells (TBV2 line), neonatal and adult CMs using TRIzol method (Invitrogen). The mRNA obtained was quantified with NanoDrop Spectrophotometer and the quality was tested through agarose gel. 1 µg of total RNA were reverse transcribed to cDNA (SuperScript VILO – Invitrogen) and amplified by real-time qPCR with Syber Green PCR master mix (Applied Biosystem). Each samples was analyzed in triplicated using the 7900HT (Applied Biosystem). In Appendix Table II are indicated the specific primer used for the analysis of the gene expression of cardiac markers of differentiation and markers of pluripotency. Relative mRNA expression profiling in mES, CMp and CMa revealed that Nanog and Pou5f1 are enriched in mES compared with CMp and CMa, the expression of the transcription factors Gata4 and Nkx2-5 increase in CMp compared to mES and remain at high level also in CMa. Cardiac structural genes were also investigated; the fetal isoform of myosin heavy chain (Myh7) is significantly upregulated in CMp, whereas the adult isoform Myh6 is progressively up-regulated in CMp and CMa compared to mES, as well as TnnI3 (Figure 26 A). To validate the ChIP-seq enrichment also qPCR-ChIP was performed. Chromatin isolation and ChIP was performed as previously described in sections 3.6 and 4.1.2.2. Appendix Table I show the list of the antibodies used. After DNA purification the enrichment of H3K79me2, H3K4me3, H3K27me3 and H3K9me3 was evaluated on the promoter region of Nanog, Pou5F1, Gata4, Nkx2-5, Myh7, Myh6 and TnnI3. Three primer sets were analyzed for each gene: -4000/-3500bp to TSS, -1000/-500bp to TSS and +500/+1000bp to TSS (Appendix Table III). The enrichment profile was compared to negative control (IgG) and normalized to the relative presence of unmodified histone H3. According with ChIP-seq data H3K79me2 results enriched on the locus of gene active transcribed during cardiomyocyte differentiation, as well as H3K4me3, whereas H3K27me3 is associated with transcription repression and H3K9me3 results not involved in the differentiation of cardiac cells (Figure 26B).

*Figure 26 - Histone methylation enrichment and mRNA expression on key genes of cardiac differentiation. A, Relative mRNA expression was access through qRT-PCR in mouse Embryonic Stem cells (mES), post-natal cardiomyocytes (CMp) an adult cardiomyocytes (CMa). mRNA levels for Nanog, Pou5F1, Gata4, Nkx2-5, Myh7, Myh6 and TnnI3 are normalized to 18s and expressed as the mean ± s.d. from three independent experiment. ** P<0.01 vs mES, ## P<0.01 vs CMp. B, ChIP assay shows H3K79me2, H3K4me3, H3K27me3 and H3K9me3 binding to the gene locus of Nanog, Pou5F1, Gata4, Nkx2.5, Myh7, Myh6 and TnnI3. Levels were determined by qPCR and are expressed as fold change to the input and relative to H3. Three different regions were analyzed for each gene locus: -1000bp/-500bp to TSS (A), +500bp/+1000bp to TSS (B) and +3500bp/+4000bp to TSS (C). Data are the result of 3 three independent experiments. The enrichment levels for IgG (neg ctrl) and the unmodified H3 relative to input are showed as ctrl.*
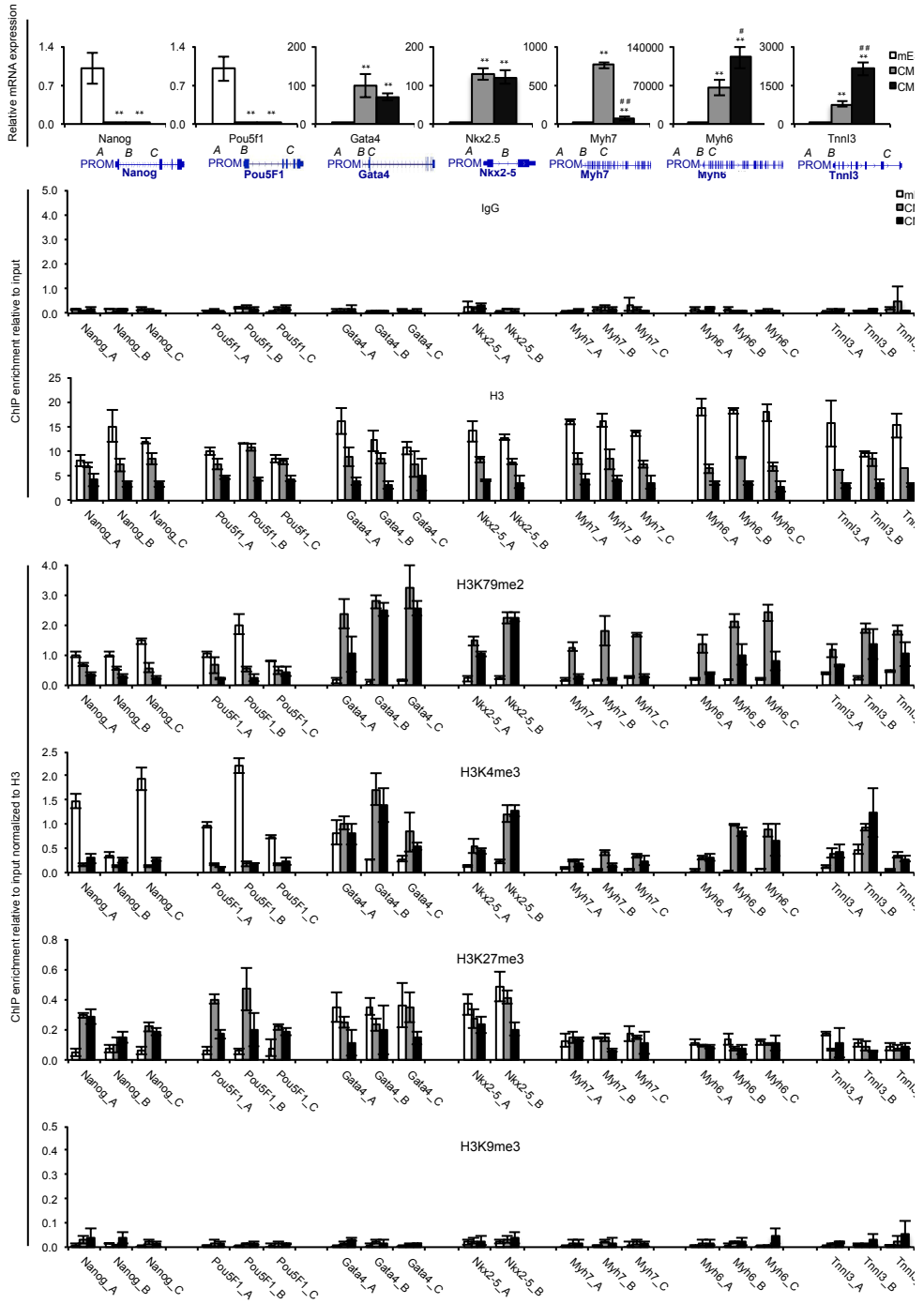
4.3 Wide gene expression analysis of epigenetic enzymes in cardiomyocytes reveals that DOT1L is highly expressed at embryonic and neonatal stages.

4.3.1 Epigenetic enzymes differentially expressed in cardiac development.

The differential RNA expression of several epigenetic enzymes genes and conventional genes involved in cardiac differentiation genes has been analyzed in undifferentiated mES cells and CMs isolated at different stage of heart development (embryos E14.5, neonatal 1 day post natal and adult 2 month old mice. Cardiomyocytes from neonatal and adult mice were isolated as previously described in sections 3.1 and 4.1.2.1; embryonic CMs were isolated from 40 embryos at embryonic day E14.5. Briefly after surgical procedures that allowed us to isolate the heart from the embryos, 40 heart were rinse in HBSS and processed in a collagen IV solution at 37°C for 5 consecutive digestions, the single cells isolated were plated for 1 hour in order to allowed the fibroblast and the endothelial cells to attach to the plate; the suspension of CMs were then collect: 1 million of cells were stained with TNNI and analyzed at FACs to verify the efficiency of isolation and the purity of CMs population, 2 million were collected in Leamli for protein analysis and 2 million were collected in TRIzol for RNA extraction.
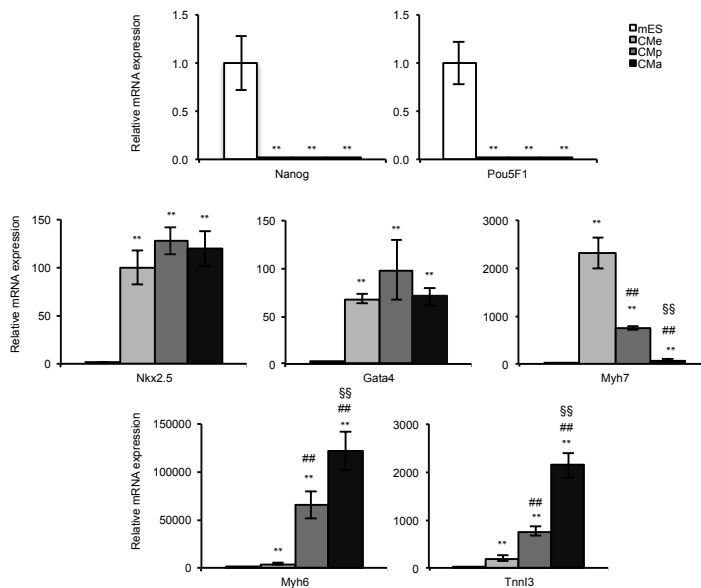

Figure 27 Gene expression profile of marker of CMs during development. qRT-PCR showing relative mRNA expression of pluripotency (Pou5F1, Nanog), and of early (Nkx2-5, GATA4, and Myh7) and late stages of heart development (Myh6 and TnnI3) in mES, CMe, CMp and CMa. Bars represent the mean ± s.d. from three biological replicates. ** P<0.01 vs mES, ## P<0.01 vs CMe , §§ P<0.01 vs CMp.

The RNA from mES, CMe, CMp, CMa was extracted using the TRIzol method (Invitrogen) following the manufacture instruction. The RNA obtained was quantified with NanoDrop Spectrophotometer and the quality tested through agarose gel. 2 µg of RNA was retro-transcribed to cDNA. First realtime qPCR analysis was conduct in order to access the expression of control genes of pluripotency (Pou5F1, Nanog), markers of early stages of heart development (MEF2C, Isl1, Nkx2-5, GATA4 and Myh7) and late stages (Myh6 and TnnI3). Data are the results of three independent experiment and are presented as mean ± SD, 18s gene has been used as housekeeping gene for the normalization levels. (Figure 27).

To interrogate the role of specific enzymes involved in the modulation of the epigenetic code associated with DNA and histone modification pathways and to identify potential molecular targets that drive CMs through heart development, using the Applied Biosystem technology, we design a 384well Taqman microfluidic epigenetic card that allowed us to compare the expression of 88 specific epigenetic enzymes coding for 3 DNA methyltransferases (DNMTs), 16 histone acetyltransferases (HATs), 12 histone deacetylases (HDACs), 31 histone methyltransferases (HMTs), 15 histone demethylases (HDMs) and 11 chromatin remodeling factors involved in polycomb repressive complex (Pcgf) and in DNA binding (Mbd). The cDNA derived from 1.2µg of mRNA was loaded on the cards; quantitative realtime PCR, from three independent experiments, was performed and the results were normalized using the median method of normalization, this method correlate the expression level of a target gene with the median expression of a group of genes, showing in this way the trend of expression within a group. This analysis revealed that, out of 88 genes, the expression of the epigenetic enzymes analyzed clustered in 4 main different groups: in the first one 25 enzymes results up-regulated during heart development compared to mES levels indication a role of this enzymes in the differentiation process and an association with the transcription regulation of genes involved in the structural and functional characterization of adult CMs; in the second group of genes, 17 enzymes results up-regulated in the embryonic/neonatal stage compared to adult CMs and mES, this peak of expression reveal a potential role for these enzymes in the definition of cardiac specification and commitment. An opposite pattern of expression is present for 21 enzymes that results down-regulated in the embryonic/neonatal stage compared to adult CMs and mES, whereas the last group of 25 genes were progressively down-regulated during development indicating a fundamental role for this enzymes in the maintenance of pluripotency (Figure 28). The results of the epigenetic card have been also validated on different samples via conventional realtime qPCR using primer designed intra-spanning for Syber Green chemistry (Appendix Table II).

Figure 28 – Wide expression profiling of epigenetic enzyme genes reveals a potential role for DOT1L in cardiomyocyte differentiation. Custom qRT-PCR card showing the relative expression of 88 epigenetic enzymes in mES, CMe CMp and CMa, the mean of three biological replicates is shown, data are presented as – ΔΔCt relative to mES expression. On the left heatmap showing differential expression levels, on the right graphs highlight 4 groups of differentially expressed genes.
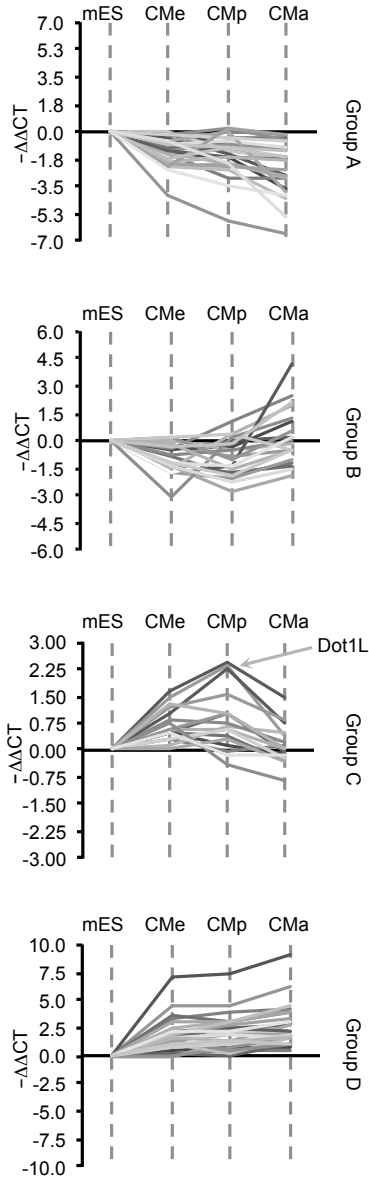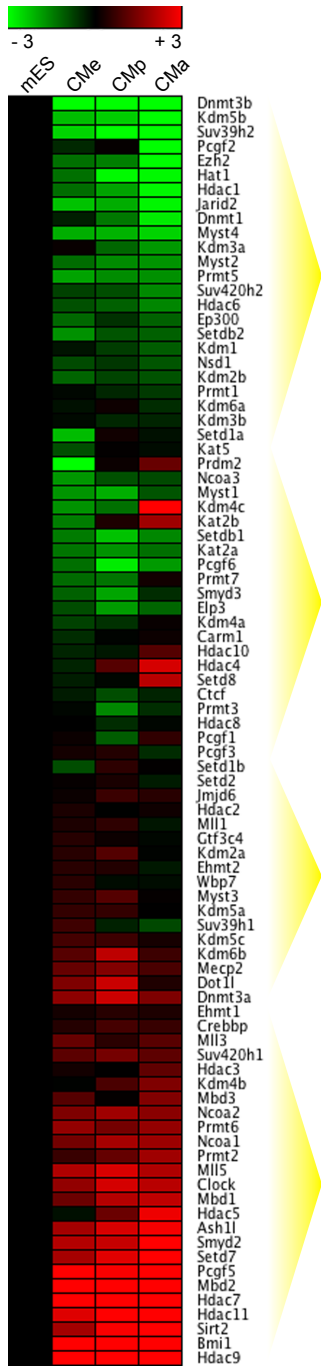
4.3.2 The histone methyltransferase DOT1L espression and function in cardiomyocytes.

Within the third group of genes up-regulated during the embryonic/neonatal stages of heart development we identify the histone methyltransferases DOT1L as one of the most significant enzymes differential expressed.

The RNA level of expression analyzed through conventional realtime qPCR reveals that in CMs DOT1L expression is markedly up-regulated during embryonic and neonatal stages compared to undifferentiated mES cells, the expression level decreased than in a significant way in adult CMs (Figure 29). This result indicates a potential role of DOT1L in the definition of important mechanisms involved in cardiac development.
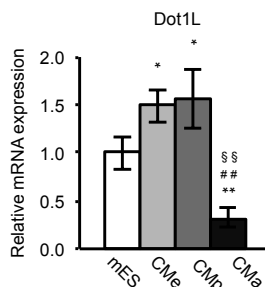

*Figure 29 - qRT-PCR showing relative mRNA expression of DOT1L. Bars represent the mean ± s.d. from three biological replicates. * P < 0.05 vs mES, ** P<0.01 vs mES, ## P<0.01 vs CMe , §§ P<0.01 vs CMp*

The mRNA expression is also confirmed at the protein level where the expression of DOT1L is higher in embryonic and neonatal CMs compered to adult CMs and mES cells (Figure 30 top panel). The expression of DOT1L enzyme correlate also with its histone methyltransferase function; the up-regulation of DOT1L enzyme reflects at the protein levels with the progressive up-regulation of the di-methylation of Lys 79 in embryonic CMs, neonatal CMs and adult CMs (Figure 30 low panel). This indicates the importance of a time dependent expression of the enzyme that results in a consecutive expression of its function.
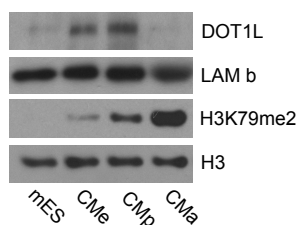

Figure 30 - Western blot protein expression analysis of DOT1L (normalized to LaminB expression level) and H3K79me2 (normalized to unmodified H3). One representative experiment out of three is shown.

4.4 Dimethylation of H3K79 mediate by DOT1L is increased during embryonic stem cells differentiation.

4.4.1 In vitro model of mES differentiation into cardiomyocytes

To investigate the mechanism involved in the activation of the specific cardiac signature activated during cardiac development and differentiation, following our data on histone methylation profile, we focused our studies on DOT1L, the only known methyltrasferase of lysine 79 of histone H3 associated with transcriptional activation.

We used an in vitro model of embryonic stem cell differentiation to further investigate the expression of DOT1L and the methylation of H3K79.

The TBV2 cell line of murine embryonic stem cells (mES) was used throughout this study. mES Cells were cultured on a feeder layer of mitotically inactivated mouse embryonic fibroblast (MEFs) in order to keep the pluripotency. The propagation medium was composed of high glucose DMEM supplemented with sodium pyruvate, L-glutamine, penicillin-streptomycin, 2-mercaptoethanole, 15% ES-screened FBS (Hyclone) and 10^3 U/mL LIF (Millipore). The time course of the differentiation experiment is presented in figure 31.
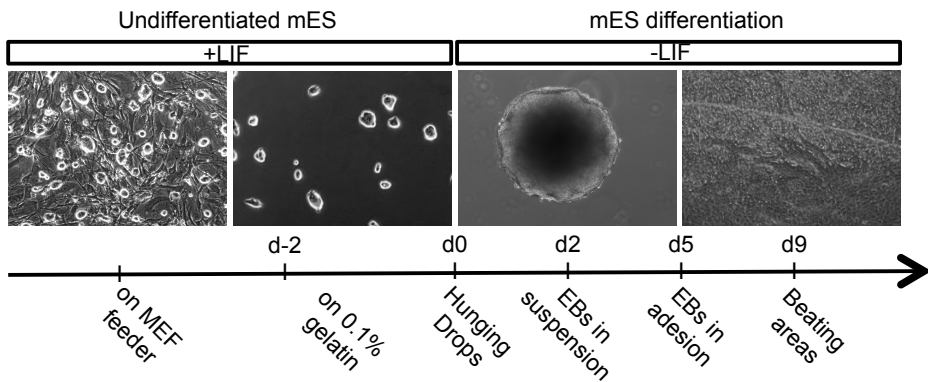

Figure 31 - Time course scheme and phase-contrast microscopy images of mES differentiation in vitro using the hanging drop method.

mES cells were passed twice on gelatin-coated (0.1%) tissue culture dishes in order to deplete the contaminant of the MEFs feeder before starting the experiment In order to generate cardiomyocytes from mES cell, the "hanging drop" methodology and embryoid bodies (EBs) aggregation approach was used to differentiate mES cells in vitro. At d0 a suspension of mES cells in differentiation

media (15% FBS without LIF) was used to form hanging drops at the concentration of 300cells/drop in which EBs were formed. EBs were cultured for 2 days in hanging drops (d0-2) and subsequently in suspension for 3 days (d2-d5) in bacterial Petri dish to avoid attachment. The 5-day-old EBs were plated onto 0.1% gelatin-coated tissue culture dishes with fresh medium every two days.

During the differentiation process mES start to show spontaneous beating activity between day 6 and day9. The efficiency of differentiation into the cardiac lineage was monitored using FACs analysis and realtime qPCR analysis at different time points of differentiation. Citofluorimetry results showed a drastic reduction of the number of cells positive for SSEA-1 (marker of staminality) from 98% at day0 to 2% at day9 and an enrichment of TNNI positive cells at day6 (17%) and day9 (27%) (Figure 32).
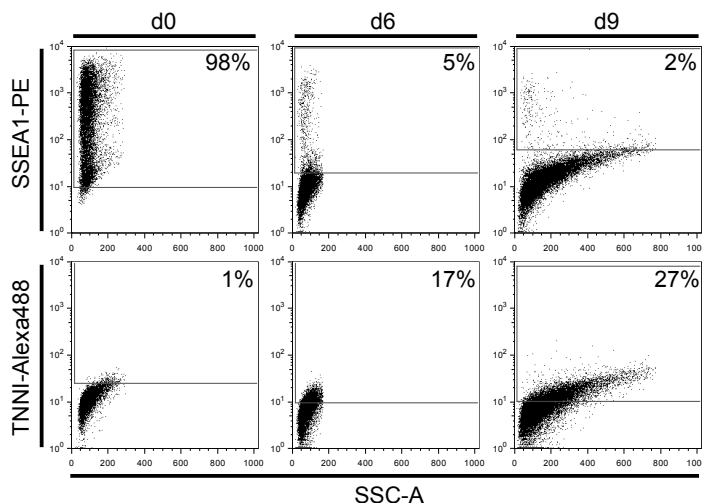

Figure 32 - FACS analysis to determine the percentage of cells expressing SSEA1 and TNNI at day 0 of differentiation, day 6 and day 9. Three independent experiment analyses were carried out; one representative analysis is shown.

Gene expression analysis revealed that differentiating cultures lose, after the first two days of differentiation, the pluripotency markers Nanog and Pou5f1; marker of cardiac commitment (Brachyury) and cardiac transcription factors (Nkx2-5, Gata4) are enriched at early stages (day 5), whereas marker of functional cardiomyocytes (Myh7, Myh6, TnnI3) become highly expressed at day 9, with evident beating areas (Figure 33).
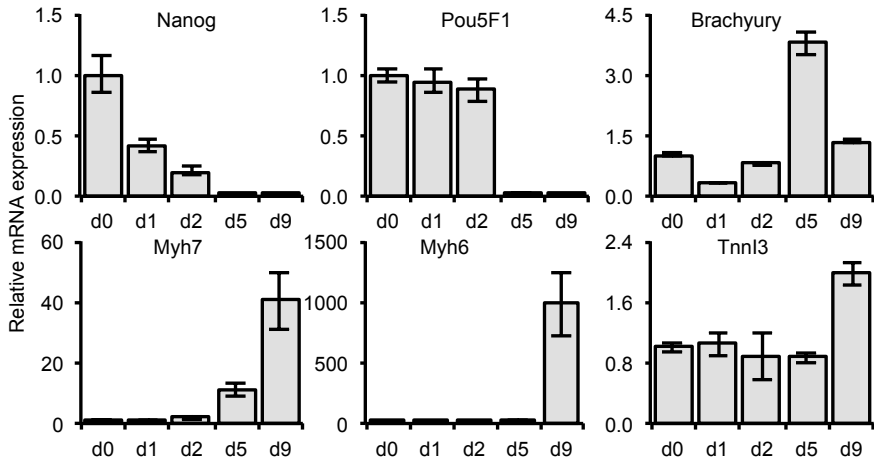
Figure 33 - Relative qRT-PCR of genes differential expressed during mES differentiation through the cardiac lineage; mRNA levels are normalized to 18s and expressed as the mean ± s.d. of three independent experiments.

POU5F1 and α/βMYH expression is also confirmed at the protein level with Western Blot experiments, with the complete lost of the protein expression of POU5F1 after day 2 and the acquisition of α/βMYH at day 9 (Figure 34).

Figure 34 - Protein levels detected by western blotting from mES cell at 0, 1, 2, 5, 9 days of differentiation. One representative experiment out of three is shown; the expression levels for POU5F1, α/βMYH are normalized to that of Lamin B (LAM b).

4.4.2 DOT1L and H3K79me2 expression during mES differentiation into CMs.

To test the role of Dot1L during in vitro differentiation of mES into CMs the level of expression of mRNA was evaluated. This analysis revealed that Dot1L results increased at day 1 and day 2 compared to day 0 and than decreased at day 5 and day 9 (Figure 35).

Figure 35 - Relative qRT-PCR of DOT1L expression at day 0, 1, 2, 5 and 9 of differentiation; mRNA levels are normalized to 18s and expressed as the mean ± s.d. from three independent experiment.

The mRNA expression is translated also at the protein level where the level of Dot1L, results increased at day 1 and day 2, resulting in the progressive enrichment of H3K79 di-methylated at day 2, day 5 and day 9. The expression levels are normalized with respect to the nuclear marker Lamin B and the unmodified H3 (Figure 36).
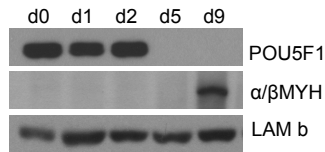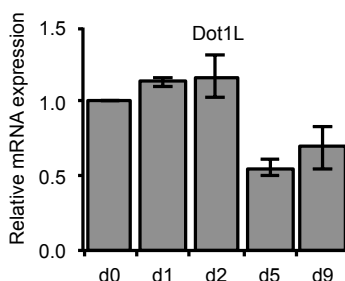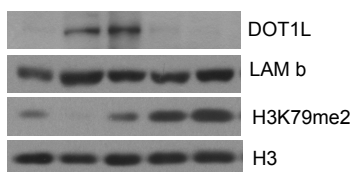

Figure 36 - Protein levels detected by western blotting from mES cell at 0, 1, 2, 5, 9 days of differentiation. One representative experiment out of three is shown; the expression levels DOT1L are normalized to that of Lamin B (LAM b) and the expression levels of H3K79me2 to that of unmodified histone 3 (H3).

4.4.3 During mES differentiation into CMs DOT1L activates cardiac specific loci through di-methylation of H3K79.

To test if the expression of Dot1L and the di-methylation of H3K79 occur at gene loci involved in cardiac differentiation we performed chromatin immunoprecipitation experiments. qPCR-ChIP was performed as previously described in section 4.1.2.2, the immunoprecipitation was performed over night using IgG, H3K79me2 and H3 antibodies (Appendix Table I). Through real time-qPCR the enrichment of H3K79me2 was analyzed in two different regions of the gene locus of Nanog, Pou5F1, Brachyury, Myh7, Myh6 and TnnI3, specific primers were designed in the region -1000bp / -500bp to TSS and in the region +500bp / +1000bp to TSS (Appendix Table III). According with gene expression analysis, H3K79 results di-methylated from day 2 at the locus of Brachyury and at day 9 at the locus of Myh7, Myh6 and TnnI3 genes in all the regions analyzed (Figure 37).

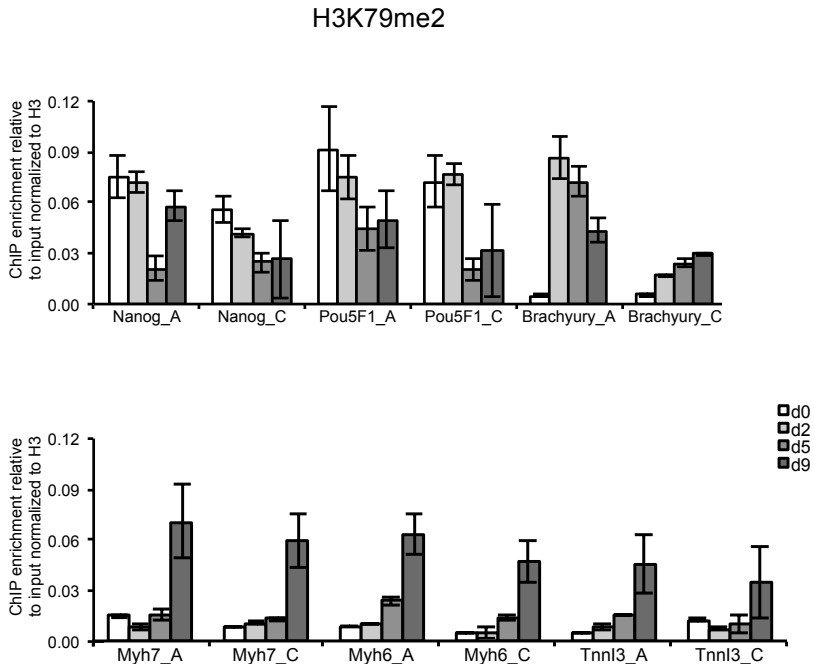*Figure 37 - ChIP assay showing H3K79me2 binding to the gene locus of Nanog, Pou5F1, Brachyury, Myh7, Myh6 and TnnI3. Levels were determined by qPCR and are expressed as fold change to the input and relative to H3. Three different regions were analyzed for each gene locus: -1000bp/-500bp to TSS (A), +500bp/+1000bp to TSS (B) and +3500bp/+4000bp to TSS (C). Data are the result of three independent experiments.*

4.5 DOT1L methyltransferase activity is essential to activate cardiac genes commitment.

4.5.1 DOT1L silencing in mES differentiation into CMs.

To test if the activity of DOT1L is mandatory for the transcriptional activation of genes involved in cardiac commitment we use a lentiviral-mediated knock-out approach during mES cells differentiation in vitro. The variation of the efficiency of differentiation elucidates the function of DOT1L in the specific gene reprogramming occurring during cardiac commitment. mES cells were cultured and differentiated as described in sections 3.2. and 4.4.1. At day 0 we infected the cells with a lentiviral vector carrying an shRNA that silence DOT1L for 2 hours before the hanging drops procedure and then the differentiation protocol has been carried out as previously described. In parallel we performed an infection with an empty vector as control. Comparing the mRNA level of Dot1L during differentiation, we observed that at day 1 and day 2 the expression of Dot1L was efficiently silenced with the shRNA compared to the shCtrl (Figure 38)

Figure 38 - mRNA levels of DOT1L gene in undifferentiated day 0 mES cells and at day 1, 2 5 and 9 after differentiation in control (sh-Ctrl grey line) and DOT1L-knowkdown cells (sh-Dot1L black line). Levels are normalized to 18s and plotted as relative mRNA expression. Three independent experiments were carried.

At the protein level Western Blot experiments showed that the down-regulation of DOT1L in the shRNA differentiation (shDOT1L) is evident at day 1 and day 2 and therefore the cells don't gain the activating marker H3K79me2 at day 2, day 5 and day 9 (Figure 39).
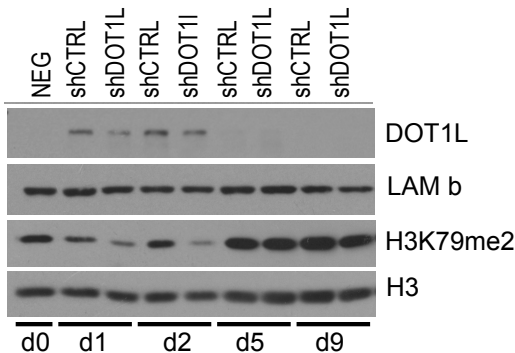
Figure 39 – Protein levels detected by western blotting from mES cell at day 0, 1, 2, 5, 9 days of differentiation in control (sh-Ctrl) and DOT1L-knowkdown cells (sh-Dot1L). One representative experiment out of three is shown; the expression levels DOT1L are normalized to that of Lamin B (LAM b) and the expression levels of H3K79me2 to that of unmodified histone 3 (H3).

4.5.2 The absence of DOT1L results in less efficiency of cardiac differentiation.

Differentiation of mES cells in the absence of DOT1L (shDOT1L) results, at the protein level through Western Blot experiments, in the down-regulation of the expression of α/βMYH at day 9 whereas no effects are observed on the expression of pluripotency markers such as POU5F1 (Figure 40).

Figure 40 - Protein leve......................................ES cell at day 0, 1, 2, 5, 9 days of differentiation in control (sh-Ctrl) and DOT1L-knowkdown cells (sh-Dot1L). One representative experiment out of three is shown; the expression levels for POU5F1, α/βMYH are normalized to that of Lamin B (LAM b).

At the mRNA level we analyzed the expression of several genes important in each step of cardiac differentiation in order to individuate the specific moment in which DOT1L plays its role.

Comparing the level of expression of Pou5F1 and Nanog important for maintain pluripotency no difference were observed in sh-Dot1L differentiation compared to sh-Ctrl differentiation. We analyzed then the expression of markers of the three germ layers: Brachyury as marker of mesoderm, Sox17 as marker of ectoderm and Nestin as marker of endoderm. Interestingly we found that the expression of Brachyury is significantly down-regulated in the absence of Dot1L whereas the expression of Sox17 and Nestin are not affected, this suggest a specific role of this epigenetic enzyme in mesodermal commitment.

We than examined the expression of transcription factors involved in the activation of cardiac genes but we didn't observed any significant change of expression for Mef2C, Gata4, Isl1 and Nkx2-5 at any time of differentiation in the absence of Dot1L, this results is explained by the fact that at the moment of the methyltransferase activity of Dot1L the transcription factors are already synthetized and present on the locus of the gene that is going to be activated. But the absence of Dot1L was observed to modulate the expression of Myh7, Myh6 and TnnI3 genes, which in the shDo1IL differentiation resulted down-regulated respect the sh-Ctrl (Figure 41).

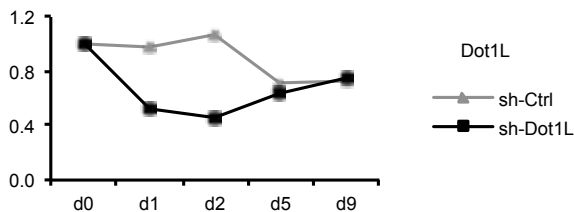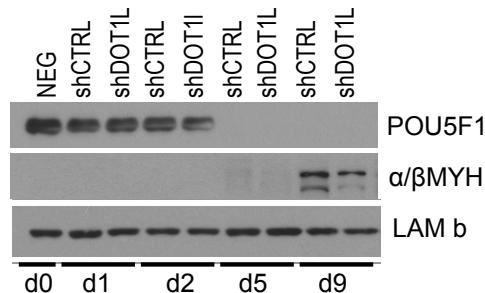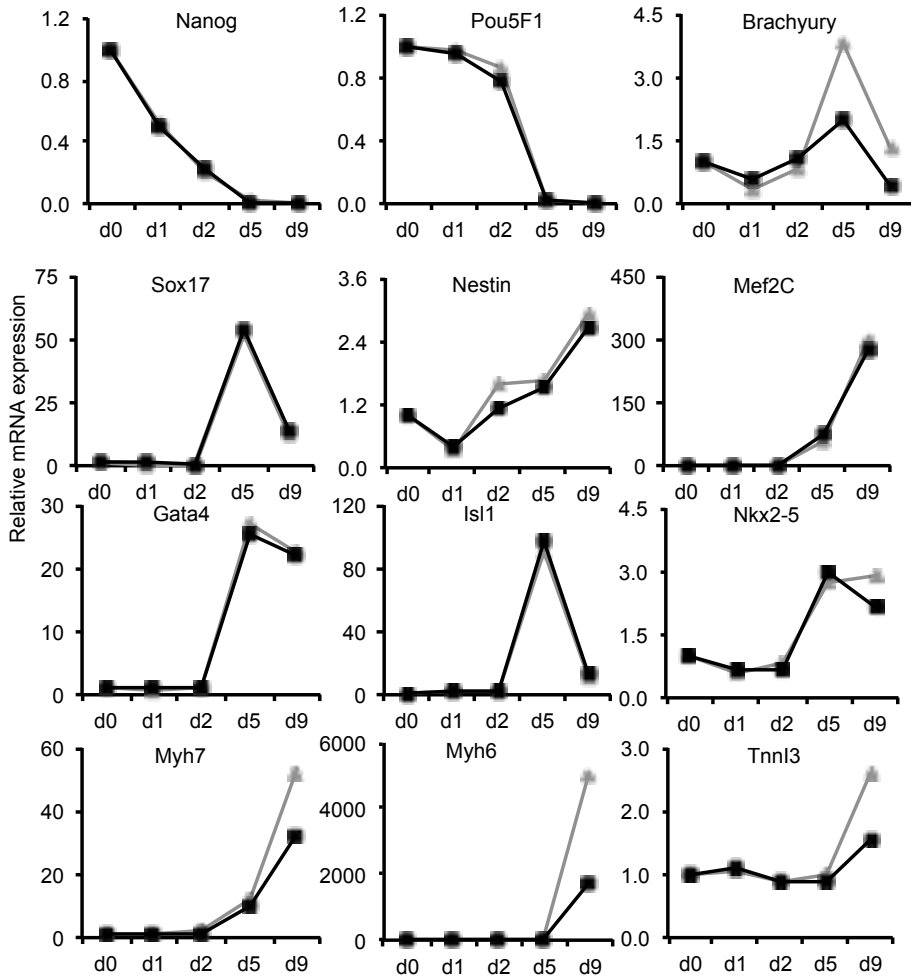*Figure 41 - mRNA levels of pluripotency and developmental genes in undifferentiated day 0 mES cells and at day 1, 2, 5 and 9 after differentiation in control (sh-Ctrl grey line) and DOT1L-knowkdown cells (sh-Dot1L black line). Levels are normalized to 18s and plotted as relative mRNA expression. Three independent experiments were carried out and one representative experiment is shown.*

To test is the lower expression of cardiac genes results also in a less efficiency of differentiation in term of number cardiomyocytes we perform a FACs experiment staining the cells with TNNI antibody as marker of CMs and we found that in the absence of Dot1L during differentiation at day 6 and day 9 less cells are positive in the sh-Dot1L compared to sh-Ctrl. In control condition, at day 6, 20% of cells are TNNI positive and a percentage of 25% of TNNI positive cell is reached at day 9.

The absence of Dot1L (sh-Dot1L) result in the lost of 25% of the efficiency of differentiation: at day 9 only 15% of the cells are TNNI positive and at day 9 only 18% (Figure 42).
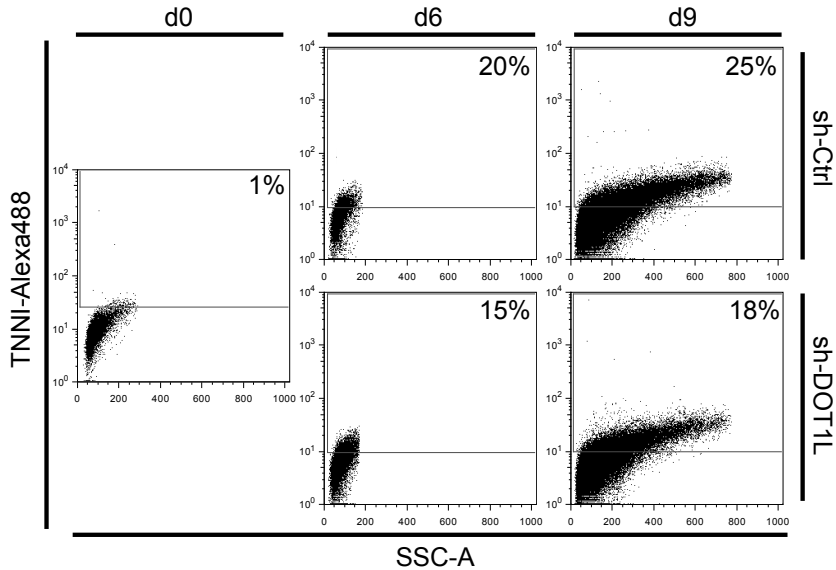
Figure 42 - FACS analysis shows the percentage of cells expressing TNNI at d0, d6 and d9 of differentiation in in control (sh-Ctrl) and DOT1L-knowkdown cells (sh-Dot1L). Three independent experiments were carried out; one representative analysis is shown.

The possibility to silence Dot1L during differnetiation helped us to indientify the moment in which this enzyme explicate its methyltransferase activity, di-methylate H3K79 resulting in transcriptional activation. From our results it seem that it plays a central role in the activation of mesodermal and cardiac specific commitment.

5. CONCLUSION

During development, cells proceed from totipotency through lineage commitment to terminal differentiation. From essentially identical genomes cells choose their fate, establishing developmental commitments that can be transmitted to daughter cells. This is possible thanks to epigenetic decision that control gene expression through the modification of chromatin conformation but without modifying the actual DNA sequence.

Cell differentiation is associated with the activation of a specific gene expression program in which a defined subset of lineage-restricted genes is expressed, leaving the rest of the genome in a repressed state. Within this context, epigenetic decisions either activate or repress specific genes to enable lineage commitment.

Histone post-traslational modifications are implicated in influencing gene expression and genome function by establishing global chromatin environment and orchestrating DNA-based biological processes [109]. The general picture emerging is that differentially expressed genes are also epigenetically distinguishable.

Remarkable progress has been made during the past few years in the characterization of histone modifications on a genome-wide scale [46, 110]. Several large-scale chromatin immunoprecipitation (ChIP) combined with massive DNA sequencing studies (ChIP-seq) have revealed interesting insights into the complex relationship between gene expression and histone modifications. The general picture emerging from these studies is that high level of histone acetylation, methylation of H3K4, H3K36 and H3K79 are detected on active genes, whereas elevated methylation of H3K9 and H3K27 correlates with gene repression [111-113]. Generally the apparently opposite modifications H3K4me3 and H3K27me3, colocalize in regions termed "bivalent domains" in embryonic stem cells, which have been suggested to function in the differentiation of these cells [113].

Chromatin-state maps for a variety of cell types, showing the genome wide distribution of important chromatin modifications, have been done, but no whole-genome ChIP studies in cardiac cells have been reported.

Hitherto, some studies on cell differentiation during cardiomyogenesis have focused on the roles of remodeling chromatin structure complexes and histone acetylation in defining the transcription program [114, 115], less clear in this process is the importance of other histone modifications and, in particular, the dynamic interaction between histone methyltransferases and demethylases. These mechanisms create a specific and complex "epigenetic code", an important characteristic of which is its dynamicity, which depends upon the reversibility of the modifications and on the activities of the enzymes involved.

Histone methylation is catalyzed by group of histone methyltransferases (KMTs). The KMTs can be divided into two main classes based on their catalytic domains: with SET domain and without SET domain. The only KMT enzyme belonging to the second class is Dot1/DOT1-Like (disruptor of telometic silencing, also called Kmt4) [83].

Knockout models in yeast, flies and mice reveal that Dot1 is the only enzyme that catalyzes H3K79 mono-, di- and tri- methylation, in fact the absence of Dot1 results in the complete lost of methylation at the level of H3K79 [84-86]. The reversibility of

this marker is less well studied; few studies suggest mechanisms of H3K79 demethylation [87].

Indeed genome-wide analysis link methylation of H3K79 to active transcription. All H3K79 methylation markers are localized within the body of transcribed gene and the amount of enrichment correlates with expression level, suggesting a role of DOT1L in transcription activation and elongation [57, 89, 90].

Dot1 was identified in a genetic screen for genes whose overexpression disrupts telomeric silencing in saccharomyces cerevisiae [88], in mammals besides regulating heterochromatin formation at telomeres [91] indeed several others are the functions that arise from Dot1L's enzymatic activity: it play a role in cellular and biological processes such as DNA repair and cell cycle regulation [86, 92], embryonic development [116] [86], MLL-associated leukemogenesis [94, 95], hematopoiesis [93] and cardiac function [97]. In fact, DOT1L is highly expressed in the heart, and germline disruption of DOT1L in mice is lethal at embryonic day E10.5, generating growth impairment, yolk sac angiogenesis defects, and cardiac dilatation [86]. A cardiac-specific knockout of mDOT1L (Dot1L$^{f/f}$: αMHC Cre) caused dilated cardiomyopathy (DCM) with chamber dilatation and systolic dysfunction [97]. However, studies were needed to shed light on how Dot1L achieved specificity to temporally regulate its catalytic activity at specific chromatin domains.

While it is now clear that epigenetics plays a central role in the programming of genomes that underlie the establishment and maintenance of differentiated cell states, how genomic programs are progressively deployed and what chromatin regulatory mechanisms coordinate their deployment in cardiomyocytes fate decisions still need to be uncovered. Here, we have generated high-resolution maps of histone methylation in neonatal and adult cardiomyocytes, clarifying the signaling events underlying gene transcription associated with histone methylation in cardiomyocyte differentiation. Indeed we have identified a specific role of H3K79 methylation mediated by Dot1L in the activation of specific genes involved in cardiac commitment.

Our results prove that histone methylation regulates the transcription program in developing cardiac cells and that H3K79me2 plays a central role in this regulation; moreover, DOT1L is fundamental to establish the epigenetic signature needed for cardiac commitment.

The genomic profile of histone methylation and the epigenetic enzymes involved have fundamental roles in defining the characteristics of cardiac cells during differentiation. Epigenetic mechanisms are intimately associated with transcriptional networks, adding a layer of complexity and fine-tuning to the regulation of cardiac commitment and differentiation.

In a differentiating cell, the maintenance of the proper balance between H3K4 and H3K27 methylation levels at the promoter regions of poised developmental genes is important to govern the highly dynamic regulation of gene expression needed to define a differentiated cell. In addition to the methylation of the bivalent domain that govern a differentiating cell, the whole program of gene expression can be driven by the methylation of H3K79 as a signature needed to define cardiac lineage commitment and the subsequent activation of key cardiac genes.

We have identified a large number of genes that show H3K79 methylation enrichment in cardiomyocytes; the tiny correlation with the transcription profile

prove that histone methylation regulates the transcription program in developing cardiac cells and that H3K79me2 plays a central role in this regulation. Indeed the genes associated with H3K79me2 functionally clustered in pathways involved in the definition of different aspect of cardiac structure and function. The rich dataset generated in our study has allowed us to identify DOT1L as a key actor that play an important role in defining cardiac commitment through the control of the levels of H3K79 methylation at the locus of cardiac-specific genes. This pre-activation pattern defined by DOT1L is consistent with the idea that molecular events in early lineage commitment mark gene for subsequent activation. This pre-activation pattern is likely important for genes that are not regulated by polycomb complexes. We propose that early deposition of H3K79me2 at specific cardiac genes is a regulatory step that facilitates later activation of these genes and that DOT1L is fundamental and absolutely required for defining the epigenetic signature that drives a cell to cardiac commitment. Our results demonstrate that high levels of DOT1L are needed to activate cardiac specific genes and that this is achieved through tight control of the level of H3K79 methylation at the regulatory regions of cardiac differentiation genes. Thus, the downregulation of DOT1L during differentiation constitutes a critical mechanism that lacks a fundamental part for the transcriptional program of activation of cardiac gene. In fact our idea is supported by the fact that mice lacking DOT1L in cardiomyocytes during development show sever cardiac phenotype [86] [97].

We propose the following mechanism:

Altogether our study illustrates the importance of epigenetic regulation early in development to delineate the fate of a cell and in particular the role of methylation of H3K79 in cardiomyocytes stabilizing the signature for cardiac gene expression. Furthermore, this study add an important information to the intricate process of

transcription activation that make a cardiomyocyte; for the first time we built an high resolution map of H3K79me2 in cardiomyocytes that could be the starting point to predict novel transcriptional regulatory networks during cardiomyocyte differentiation, as well provide the opportunity to identify novel genes that might be informative to understand developmental regulatory programs. Indeed we shed light on an additional fundamental enzyme involved in defining the epigenetic code associated with the complex process of heart development and establish a platform useful to identify new mechanisms underlying many congenital heart defects and cardiac developmental malformations.

Pharmacological modulation of epigenetic mechanisms has a great therapeutic potential. Epigenetic therapy, feasible on account of the reversibility of the epigenetic profile of cells, consists in restoring correct gene expression in affected cells through the use of drugs that inhibit the activity of epigenetic enzymes [117]. Recently, it was shown that HDACi prevent cardiac hypertrophy and heart failure, suggesting the possibility of epigenetic therapy also for the management of heart disease [118, 119].

However, the development of epigenetic therapy for cardiovascular disease is limited, at the moment, by the pleiotropic effect of these drugs and by a lack of knowledge on the epigenetic mechanisms involved. A better definition of the epigenetic mechanisms implicated in cardiovascular disease could help to overcame these problems and to develop more specific therapeutic strategies for heart diseases based on the identification of new molecular targets (e.g. DNA and histone methyltransferases and demethylases).

Epigenetics could also play an important role in the diagnosis of cardiovascular defects. Indeed, a distinguishing feature of epigenetic changes with respect to those of a genetic nature is that they tend to be acquired in a gradual, rather than an abrupt, manner. This makes epigenetic alterations a good target for prevention strategies.

An important application of epigenetics could be in modulating the efficiency and specificity of cardiac differentiation and to explain defects underlying cardiac developmental malformation. However, due to a lack of knowledge on the mechanisms involved, at present we are far from a clinical application of epigenetic mechanisms to control heart development. Therefore, further studies are necessary in order to shed light on the signaling events underlying transcription in cardiac differentiation and development. To this end high-resolution maps of histone methylation in neonatal and adult cardiomyocytes were generated, clarifying the signaling events underlying gene transcription associated with histone methylation in cardiomyocyte differentiation and development.

Indeed we have identified a specific role of H3K79 methylation mediated by Dot1L in the activation of specific genes involved in cardiac commitment.

6. REFERENCES

1.   Epstein, J.A., *Franklin H. Epstein Lecture. Cardiac development and implications for heart disease.* The New England journal of medicine, 2010 363(17): p. 1638-47.
2.   Olson, E.N., *A decade of discoveries in cardiac biology.* Nature medicine, 2004. 10(5): p. 467-74.
3.   Bruneau, B.G., *The developmental genetics of congenital heart disease.* Nature, 2008. 451(7181): p. 943-8.
4.   Pierpont, M.E., et al., *Genetic basis for congenital heart defects: current knowledge: a scientific statement from the American Heart Association Congenital Cardiac Defects Committee, Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics.* Circulation, 2007. 115(23): p. 3015-38.
5.   Evans, S.M., et al., *Myocardial lineage development.* Circulation research, 2010. 107(12): p. 1428-44.
6.   Olson, E.N., *Gene regulatory networks in the evolution and development of the heart.* Science, 2006. 313(5795): p. 1922-7.
7.   Jiang, X., et al., *Fate of the mammalian cardiac neural crest.* Development, 2000. 127(8): p. 1607-16.
8.   Gu, W., et al., *Interaction of myogenic factors and the retinoblastoma protein mediates muscle cell commitment and differentiation.* Cell, 1993. 72(3): p. 309-24.
9.   Laflamme, M.A. and C.E. Murry, *Heart regeneration.* Nature, 2011. 473(7347): p. 326-35.
10.   Wilkinson, D.G., S. Bhatt, and B.G. Herrmann, *Expression pattern of the mouse T gene and its role in mesoderm formation.* Nature, 1990. 343(6259): p. 657-9.
11.   Gourdie, R.G., et al., *Endothelin-induced conversion of embryonic heart muscle cells into impulse-conducting Purkinje fibers.* Proceedings of the National Academy of Sciences of the United States of America, 1998. 95(12): p. 6815-8.
12.   Christoffels, V.M., et al., *Tbx18 and the fate of epicardial progenitors.* Nature, 2009. 458(7240): p. E8-9; discussion E9-10.
13.   Cai, C.L., et al., *A myocardial lineage derives from Tbx18 epicardial cells.* Nature, 2008. 454(7200): p. 104-8.
14.   Moorman, A.F. and V.M. Christoffels, *Development of the cardiac conduction system: a matter of chamber development.* Novartis Foundation symposium, 2003. 250: p. 25-34; discussion 34-43, 276-9.
15.   Lompre, A.M., B. Nadal-Ginard, and V. Mahdavi, *Expression of the cardiac ventricular alpha- and beta-myosin heavy chain genes is developmentally and hormonally regulated.* The Journal of biological chemistry, 1984. 259(10): p. 6437-46.
16.   Subramaniam, A., et al., *Tissue-specific regulation of the alpha-myosin heavy chain gene promoter in transgenic mice.* The Journal of biological chemistry, 1991. 266(36): p. 24613-20.

17. Walker, C.A. and F.G. Spinale, *The structure and function of the cardiac myocyte: a review of fundamental concepts.* The Journal of thoracic and cardiovascular surgery, 1999. 118(2): p. 375-82.

18. Babu, M.M., et al., *Structure and evolution of transcriptional regulatory networks.* Current opinion in structural biology, 2004. 14(3): p. 283-91.

19. Brivanlou, A.H. and J.E. Darnell, Jr., *Signal transduction and the control of gene expression.* Science, 2002. 295(5556): p. 813-8.

20. Nikolov, D.B. and S.K. Burley, *RNA polymerase II transcription initiation: a structural view.* Proceedings of the National Academy of Sciences of the United States of America, 1997. 94(1): p. 15-22.

21. Lobe, C.G., *Transcription factors and mammalian development.* Current topics in developmental biology, 1992. 27: p. 351-83.

22. Kaneko, K.J. and M.L. DePamphilis, *Regulation of gene expression at the beginning of mammalian development and the TEAD family of transcription factors.* Developmental genetics, 1998. 22(1): p. 43-55.

23. Pandur, P., *What does it take to make a heart?* Biology of the cell / under the auspices of the European Cell Biology Organization, 2005. 97(3): p. 197-210.

24. Moorman, A.F., et al., *Development of the building plan of the heart.* Annals of the New York Academy of Sciences, 2004. 1015: p. 171-81.

25. Cai, C.-L., et al., *Isl1 Identifies a Cardiac Progenitor Population that Proliferates Prior to Differentiation and Contributes a Majority of Cells to the Heart.* Developmental cell, 2003. 5(6): p. 877-889.

26. Elliott, D.A., et al., *Cardiac homeobox gene NKX2-5 mutations and congenital heart disease: associations with atrial septal defect and hypoplastic left heart syndrome.* J Am Coll Cardiol, 2003. 41(12798584): p. 2072-2076.

27. Garg, V., et al., *GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5.* Nature, 2003. 424(12845333): p. 443-447.

28. Goldmuntz, E., E. Geiger, and D.W. Benson, *NKX2.5 mutations in patients with tetralogy of fallot.* Circulation, 2001. 104(11714651): p. 2565-2568.

29. Kuo, C.T., et al., *GATA4 transcription factor is required for ventral morphogenesis and heart tube formation.* Genes & development, 1997. 11(9136932): p. 1048-1060.

30. Lyons, I., et al., *Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5.* Genes & development, 1995. 9(7628699): p. 1654-1666.

31. McElhinney, D.B., et al., *NKX2.5 mutations in patients with congenital heart disease.* J Am Coll Cardiol, 2003. 42(14607454): p. 1650-1655.

32. Molkentin, J.D., et al., *Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis.* Genes & development, 1997. 11(9136933): p. 1061-1072.

33. Schott, J.J., et al., *Congenital heart disease caused by mutations in the transcription factor NKX2-5.* Science, 1998. 281(9651244): p. 108-111.

34. Plageman, T.F. and K.E. Yutzey, *T-box genes and heart development: putting the "T" in heart.* Dev Dyn, 2005. 232(15580613): p. 11-20.

35. Basson, C.T., et al., *Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome.* Nat Genet, 1997. 15(8988165): p. 30-35.

36. Kirk, E.P., et al., *Mutations in cardiac T-box factor gene TBX20 are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy.* Am J Hum Genet, 2007. 81(17668378): p. 280-291.

37. Li, Q.Y., et al., *Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family.* Nat Genet, 1997. 15(8988164): p. 21-29.

38. Lindsay, E.A., et al., *Tbx1 haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice.* Nature, 2001. 410(11242049): p 97-9101.

39. Singh, M.K., et al., *Tbx20 is essential for cardiac chamber differentiation and repression of Tbx2.* Development, 2005. 132(15901664): p. 2697-2707.

40. Stennard, F.A., et al., *Murine T-box transcription factor Tbx20 acts as a repressor during heart development, and is essential for adult heart integrity, function and adaptation.* Development, 2005. 132(15843414): p. 2451-2462.

41. Takeuchi, J.K., et al., *Tbx20 dose-dependently regulates transcription factor networks required for mouse heart and motoneuron development.* Development, 2005. 132(15843409): p. 2463-2474.

42. Peterkin, T., A. Gibson, and R. Patient, *GATA-6 maintains BMP-4 and Nkx2 expression during cardiomyocyte precursor maturation.* The EMBO journal, 2003. 22(16): p. 4260-73.

43. Schwartz, R.J. and E.N. Olson, *Building the heart piece by piece: modularity of cis-elements regulating Nkx2-5 transcription.* Development, 1999. 126(19): p. 4187-92.

44. Molkentin, J.D., et al., *Direct activation of a GATA6 cardiac enhancer by Nkx2.5: evidence for a reinforcing regulatory network of Nkx2.5 and GATA transcription factors in the developing heart.* Developmental biology, 2000. 217(2): p. 301-9.

45. Schlesinger, J., et al., *The cardiac transcription network modulated by Gata4, Mef2a, Nkx2.5, Srf, histone modifications, and microRNAs.* PLoS genetics, 2011. 7(2): p. e1001313.

46. Bernstein, B.E., A. Meissner, and E.S. Lander, *The mammalian epigenome.* Cell, 2007. 128(4): p. 669-81.

47. Wu, J.I., J. Lessard, and G.R. Crabtree, *Understanding the words of chromatin regulation.* Cell, 2009. 136(2): p. 200-6.

48. Waddington, C., *"The epigenotype".* (1942).

49. Hendrich, B. and S. Tweedie, *The methyl-CpG binding domain and the evolving role of DNA methylation in animals.* Trends in genetics : TIG, 2003. 19(5): p. 269-77.

50. Suzuki, M.M. and A. Bird, *DNA methylation landscapes: provocative insights from epigenomics.* Nature reviews. Genetics, 2008. 9(6): p. 465-76.

51.   Ooi, S.K. and T.H. Bestor, *The colorful history of active DNA demethylation.* Cell, 2008. 133(7): p. 1145-8.

52.   Bhutani, N., et al., *Reprogramming towards pluripotency requires AID-dependent DNA demethylation.* Nature, 2010. 463(7284): p. 1042-7.

53.   Popp, C., et al., *Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency.* Nature, 2010. 463(7284): p. 1101-5.

54.   Grunstein, M., *Histone acetylation in chromatin structure and transcription.* Nature, 1997. 389(6649): p. 349-52.

55.   Wang, X. and J.J. Hayes, *Acetylation mimics within individual core histone tail domains indicate distinct roles in regulating the stability of higher-order chromatin structure.* Molecular and cellular biology, 2008. 28(1): p. 227-36.

56.   Kouzarides, T., *Histone methylation in transcriptional control.* Current opinion in genetics & development, 2002. 12(2): p. 198-209.

57.   Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. 129(4): p. 823-37.

58.   Haberland, M., R.L. Montgomery, and E.N. Olson, *The many roles of histone deacetylases in development and physiology: implications for disease and therapy.* Nature reviews. Genetics, 2009. 10(1): p. 32-42.

59.   Cloos, P.A., et al., *Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease.* Genes & development, 2008. 22(9): p. 1115-40.

60.   Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 A resolution.* Nature, 1997. 389(6648): p. 251-60.

61.   Cairns, B.R., *The logic of chromatin architecture and remodelling at promoters.* Nature, 2009. 461(7261): p. 193-8.

62.   Schones, D.E. and K. Zhao, *Genome-wide approaches to studying chromatin modifications.* Nature reviews. Genetics, 2008. 9(3): p. 179-91.

63.   Slack, J.M., *Conrad Hal Waddington: the last Renaissance biologist?* Nature reviews. Genetics, 2002. 3(11): p. 889-95.

64.   Goldberg, A.D., C.D. Allis, and E. Bernstein, *Epigenetics: A Landscape Takes Shape.* Cell, 2007. 128(4): p. 635-638.

65.   Waddington, C., *The Strategy of the Genes.* 1957, London George Allen & Unwin.

66.   Waddington, C., *The Scientific Attitude* 1941: Pelican Books.

67.   Pawlikowska, L., et al., *Association of common genetic variation in the insulin/IGF1 signaling pathway with human longevity.* Aging Cell, 2009. 8(19489743): p. 460-472.

68.   Tachibana, M., et al., *G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis.* Genes & development, 2002. 16(12130538): p. 1779-1791.

69.   de la Serna, I.L., Y. Ohkawa, and A.N. Imbalzano, *Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers.* Nature reviews. Genetics, 2006. 7(6): p. 461-73.

70.   Lickert, H., et al., *Baf60c is essential for function of BAF chromatin remodelling complexes in heart development.* Nature, 2004. 432(7013): p. 107-12.

71.   Takeuchi, J.K. and B.G. Bruneau, *Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors.* Nature, 2009. 459(7247): p. 708-11.

72.   Zhang, C.L., et al., *Class II histone deacetylases act as signal-responsive repressors of cardiac hypertrophy.* Cell, 2002. 110(12202037): p. 479-488.

73.   Han, A., et al., *Mechanism of recruitment of class II histone deacetylases by myocyte enhancer factor-2.* J Mol Biol, 2005. 345(15567413): p. 91-9102.

74.   Lagger, G., et al., *Essential function of histone deacetylase 1 in proliferation control and CDK inhibitor repression.* EMBO J, 2002. 21(12032080): p. 2672-2681.

75.   Montgomery, R.L., et al., *Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility.* Genes & development, 2007. 21(17639084): p. 1790-1802.

76.   Zupkovitz, G., et al., *Negative and positive regulation of gene expression by mouse histone deacetylase 1.* Molecular and cellular biology, 2006. 26(16940178): p. 7913-7928.

77.   Chang, S., et al., *Histone deacetylases 5 and 9 govern responsiveness of the heart to a subset of stress signals and play redundant roles in heart development.* Molecular and cellular biology, 2004. 24(15367668): p. 8467-8476.

78.   Naya, F.J., et al., *Transcriptional activity of MEF2 during mouse embryogenesis monitored with a MEF2-dependent transgene.* Development, 1999. 126(10207130): p. 2045-2052.

79.   Yao, T.P., et al., *Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300.* Cell 1998. 93(3): p. 361-72.

80.   Kim, T.-G., et al., *Jumonji represses atrial natriuretic factor gene expression by inhibiting transcriptional activities of cardiac transcription factors.* Molecular and cellular biology, 2004. 24(15542826): p. 10151-10160.

81.   Mysliwiec, M.R., et al., *Generation of a conditional null allele of jumonji.* Genesis, 2006. 44(16900512): p. 407-411.

82.   Lee, Y., et al., *Jumonji, a nuclear protein that is necessary for normal heart development.* Circulation research, 2000. 86(9): p. 932-8.

83.   Feng, Q., et al., *Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain.* Current biology : CB, 2002. 12(12): p. 1052-8.

84.   van Leeuwen, F., P.R. Gafken, and D.E. Gottschling, *Dot1p modulates silencing in yeast by methylation of the nucleosome core.* Cell, 2002. 109(6): p. 745-56.

85.   Shanower, G.A., et al., *Characterization of the grappa gene, the Drosophila histone H3 lysine 79 methyltransferase.* Genetics, 2005. 169(1): p. 173-84.

86.   Jones, B., et al., *The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure.* PLoS genetics, 2008. 4(9): p. e1000190.

87. Xu, W., et al., *Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases.* Cancer cell, 2011. 19(1): p. 17-30.

88. Singer, M.S., et al., *Identification of high-copy disruptors of telomeric silencing in Saccharomyces cerevisiae.* Genetics, 1998. 150(2): p. 613-32.

89. Steger, D.J., et al., *DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells.* Molecular and cellular biology, 2008. 28(8): p. 2825-39.

90. Wang, Z., et al., *Combinatorial patterns of histone acetylations and methylations in the human genome.* Nature genetics, 2008. 40(7): p. 897-903.

91. Lacoste, N., et al., *Disruptor of telomeric silencing-1 is a chromatin-specific histone H3 methyltransferase.* The Journal of biological chemistry, 2002. 277(34): p. 30421-4.

92. Barry, E.R., et al., *ES cell cycle progression and differentiation require the action of the histone methyltransferase Dot1L.* Stem Cells, 2009. 27(7): p. 1538-47.

93. Feng, Y., et al., *Early mammalian erythropoiesis requires the Dot1L methyltransferase.* Blood, 2010. 116(22): p. 4483-91.

94. Nguyen, A.T., et al., *DOT1L, the H3K79 methyltransferase, is required for MLL-AF9-mediated leukemogenesis.* Blood, 2011. 117(25): p. 6912-22.

95. Daigle, S.R., et al., *Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor.* Cancer cell, 2011. 20(1): p. 53-65.

96. Onder, T.T., et al., *Chromatin-modifying enzymes as modulators of reprogramming.* Nature, 2012. 483(7391): p. 598-602.

97. Nguyen, A.T., et al., *DOT1L regulates dystrophin expression and is critical for cardiac function.* Genes & development, 2011. 25(3): p. 263-74.

98. Rodgers, L.S., et al., *An improved protocol for the isolation and cultivation of embryonic mouse myocytes.* Cytotechnology, 2009. 59(2): p. 93-102.

99. Rizzi, R., et al., *Post-natal cardiomyocytes can generate iPS cells with an enhanced capacity toward cardiomyogenic re-differentiation.* Cell death and differentiation, 2012. 19(7): p. 1162-74.

100. Catalucci, D., et al., *Akt regulates L-type Ca2+ channel activity by modulating Cavalpha1 protein stability.* The Journal of cell biology, 2009. 184(6): p. 923-33.

101. Zeineddine, D., et al., *Cardiac commitment of embryonic stem cells for myocardial repair.* Methods in molecular medicine, 2005. 112: p. 175-82.

102. Feng, J., et al., *Identifying ChIP-seq enrichment using MACS.* Nature protocols, 2012. 7(9): p. 1728-40.

103. Zang, C., et al., *A clustering approach for identification of enriched domains from histone modification ChIP-Seq data.* Bioinformatics, 2009. 25(15): p. 1952-8.

104. Ross-Innes, C.S., et al., *Differential oestrogen receptor binding is associated with clinical outcome in breast cancer.* Nature, 2012. 481(7381): p. 389-93.

105. Marson, A., et al., *Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells.* Cell, 2008. 134(3): p. 521-33.

106.   Mikkelsen, T.S., et al., *Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.* Nature, 2007. 448(7153): p. 553-60.

107.   McLean, C.Y., et al., *GREAT improves functional interpretation of cis-regulatory regions.* Nature biotechnology, 2010. 28(5): p. 495-501.

108.   Robinson, J.T., et al., *Integrative genomics viewer.* Nature biotechnology, 2011. 29(1): p. 24-6.

109.   Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. 128(4): p. 693-705.

110.   Ang, Y.S., et al., *Stem cells and reprogramming: breaking the epigenetic barrier?* Trends in pharmacological sciences, 2011. 32(7): p. 394-401.

111.   Boyer, L.A., et al., *Polycomb complexes repress developmental regulators in murine embryonic stem cells.* Nature, 2006. 441(7091): p. 349-53.

112.   Lee, T.I., et al., *Control of developmental regulators by Polycomb in human embryonic stem cells.* Cell, 2006. 125(2): p. 301-13.

113.   Bernstein, B.E., et al., *A bivalent chromatin structure marks key developmental genes in embryonic stem cells.* Cell, 2006. 125(2): p. 315-26.

114.   Meshorer, E. and T. Misteli, *Chromatin in pluripotent embryonic stem cells and differentiation.* Nature reviews. Molecular cell biology, 2006. 7(7): p. 540-6.

115.   Meshorer, E., et al., *Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells.* Developmental cell, 2006. 10(1): p. 105-16.

116.   Ooga, M., et al., *Changes in H3K79 methylation during preimplantation development in mice.* Biology of reproduction, 2008. 78(3): p. 413-24.

117.   Karberg, S., *Switching on epigenetic therapy.* Cell, 2009. 139(20005793): p. 1029-1031.

118.   Kong, Y., et al., *Suppression of class I and II histone deacetylases blunts pressure-overload cardiac hypertrophy.* Circulation, 2006. 113(16735673): p. 2579-2588.

119.   Gallo, P., et al., *Inhibition of class I histone deacetylase with an apicidin derivative prevents cardiac hypertrophy and failure.* Cardiovasc Res, 2008. 80(18697792): p. 416-424.

120.   Castaldi, A et al., *December and April: two deadlines to start your future.* Thanks for your friendship.

7. APPENDIX

Table I – List of antibodies used for Western Blot, FACs and ChIP

Antibody	Specie	Brand	Catalog #	WB	FACs	ChIP
a/b MHC	Ms	Chemicon	MAB 1552	1:200		
cTNNI	Ms	Millipore-Upstate	MAB3150		1:100	
DOT1L	Rb	Abcam	ab64077	1:1000		
H3	Rb	Abcam	ab1791	1:5000		
H3K27me3	Rb	Millipore-Upstate	07-449	1:2000		3ug
H3K4me3	Rb	Active Motif	39159	1:1000		3ug
H3K79me2	Rb	Abcam	ab3594	1:1000		3ug
H3K9me3	Rb	Millipore-Upstate	07-442	1:1000		3ug
Oct4	Rb	Abcam	ab19857	1:2000		
Rb IgG	Rb	Millipore-Upstate	12-370			3ug
SSEA1	Ms	Millipore-Upstate	MAB4301		1:100	
Lamin B	Gt	Santa Cruz	sc6216	1:500		

Table II – List of primer used for Gene expression realtime qPCR (Sybr Green)

Primer	5'-3'	Category
qRT_m18S_Fw	AAATCAGTTATGGTTCCTTTGGTC	house keeping
qRT_m18S_Rev	GCTCTAGAATTACCACAGTTATCCAA	house keeping
qRT_mGAPDH_Fw	GGCAAATTCAACGGCACA	house keeping
qRT_mGAPDH_Rev	GTTAGTGGGGTCTCGCTCTG	house keeping
qRT_mGAPDH_Fw	AATGTCAGCAATGCATCCTG	house keeping
qRT_mGAPDH_Rev	ATGGACTGTGGTCATGAGCC	house keeping
qRT_mDot1L Fw	GCGGCTGTGTGACAAATACA	Epigenetic Enzyme
qRT_mDot1LRev	GGCTGTGTAGTGCCCTTCC	Epigenetic Enzyme
qRT_mSOX2_Fw	GCACATGAACGGCTGGAGCAACG	Stem Marker
qRT_mSOX2_Rev	TGCTGCGAGTAGGACATGCTGTAGG	Stem Marker
qRT_mSOX2_Fw	TCAAGCATGTCCTACTCGCAG	Stem Marker
qRT_mSOX2_Rev	GAGGAAGAGGTAACCACGGG	Stem Marker
qRT_mOCT3/4_Fw	CCCTCTGTTCCCGTCACTG	Stem Marker
qRT_mOCT3/4_Rev	ACCTCCCTTGCCTTGGCT	Stem Marker
qRT_mNANOG_Fw	CAGGTGTTTGAGGGTAGCTC	Stem Marker
qRT_mNANOG_Rev	CGGTTCATCATGGTACAGTC	Stem Marker
qRT_mBrachyury_Fw	CAGCCCACCTACTGGCTCTA	Mesoderm Marker
qRT_mBrachyury_Rev	GAGCCTGGGGTGATGGTA	Mesoderm Marker
qRT_mNestin_Fw	CTGCAGGCCACTGAAAAGTT	Endoderm Marker
qRT_mNestin_Rev	AGGTGTCTGCAAGCGAGAGT	Endoderm Marker
qRT_mSox17_Fw	CTTTATGGTGTGGGCCAAAG	Esoderm Marker
qRT_mSox17_Rev	GGTCAACGCCTTCCAAGACT	Esoderm Marker
qRT_mMEF2C_Fw	ACTGGGAAACCCCAATCTTC	Cardiac Marker
qRT_mMEF2C_Rev	ATCAGACCGCCTGTGTTACC	Cardiac Marker
qRT_mGATA4_Fw	TCTCACTATGGGCACAGCAG	Cardiac Marker
qRT_mGATA4_Rev	GCGATGTCTGAGTGACAGGA	Cardiac Marker
qRT_mISL1_Fw	GCAACCCAACGACAAAACTAAT	Cardiac Marker
qRT_mISL1_Rev	CCATCATGTCTCTCCGGACT	Cardiac Marker
qRT_mNkx2.5_Fw	CAAGTGCTCTCCTGCTTTCC	Cardiac Marker
qRT_mNkx2.5_Rev	GGCTTTGTCCAGCTCCACT	Cardiac Marker
qRT_m_cTnI3_Fw	GCAGGTGAAGAAGGAGGACA	Cardiac Marker
qRT_m_cTnI3_Rev	CGATATTCTTGCGCCAGTC	Cardiac Marker
qRT_m_aMYHC6_Fw	CGAAACTGAAAACGGCAAG	Cardiac Marker
qRT_m_aMYHC6_Rev	TGGCCATGTCCTCGATCT	Cardiac Marker
qRT_m_bMYHC7_ Fw	GAGGAGAGGGCGGACATT	Cardiac Marker
qRT_m_bMYHC7_ Rev	ACTCTTCATTCAGGCCCTTG	Cardiac Marker
qRT_m_aMYOSIN_Fw	CGCATCAAGGAGCTCACC	Cardiac Marker
qRT_m_aMYOSIN_Rev	CCTGCAGCCGCATTAAGT	Cardiac Marker
qRT_m_bMYOSIN_Fw	CGCATCAAGGAGCTCACC	Cardiac Marker
qRT_m_bMYOSIN_Rev	CTGCAGCCGCAGTAGGTT	Cardiac Marker

Table III – List of primer used for ChIP-realtime qPCR (Sybr Green)

Primer	5'-3'	Category
mPOU5F1_A_Fw	CCTGGACAACACAAGATGGA	Stem Marker
mPOU5F1_A_Rev	CCCAGAGACCCCAGAGAAGT	Stem Marker
mPOU5F1_B_Fw	AGTGAAGGGAATTGGGAACA	Stem Marker
mPOU5F1_B_Rev	ACACTGCGTCGTGCTTCTC	Stem Marker
mPOU5F1_C_Fw	GCAGGATGTGTGCAATGTCT	Stem Marker
mPOU5F1_C_Rev	GAGTTTCAAGCCAGGAATGG	Stem Marker
mNANOG_A1_Fw	AGCCGACTTAAGCTGGGTTAG	Stem Marker
mNANOG_A1_Rev	AAAGTTTGCCGATCAGTCCTT	Stem Marker
mNANOG_A2_Fw	GGGTAGGGTAGGAGGCTTGA	Stem Marker
mNANOG_A2_Rev	CGGCTCAAGGCGATAGATT	Stem Marker
mNANOG_B_Fw	TTGGTGAGGTTATACAGTTAGTTTGC	Stem Marker
mNANOG_B_Rev	CCCAAAGGTTGAGAGAAATGC	Stem Marker
mNANOG_C_Fw	GCTGGGACTAAAGGTATGTACCAC	Stem Marker
mNANOG_C_Rev	GCGAGGAGAGGCTGTTAGAA	Stem Marker
Brachyury_A_Fw	GGACGTGTCCCAAAGCTG	Mesoderm Marker
Brachyury_A_Rev	GGTCTCCTTGACCTCTCCAA	Mesoderm Marker
Brachyury_B_Fw	GGAATGACCAGGTTTGCCTA	Mesoderm Marker
Brachyury_B_Rev	CACTAAGCCCACGGGTTCT	Mesoderm Marker
mBrachyury_C_Fw	GGAAATGGACCCATTAGCTATTCT	Mesoderm Marker
mBrachyury_C_Rev	TTCCCACTCCCCAGCTACT	Mesoderm Marker
mNkx2-5_A_Fw	GAGAATCCAGGCAGACAACC	Cardiac Marker
mNkx2-5_A_Rev	GCATTGTTGGGGAATTGACT	Cardiac Marker
mNkx2-5_B_Fw	GCGGGGAGTTTGGAGTATAA	Cardiac Marker
mNkx2-5_B_Rev	AAATTCGGCGTTCCCTCTAA	Cardiac Marker
mNkx2-5_C_Fw	TACCCCTACCCCAGCTACG	Cardiac Marker
mNkx2-5_C_Rev	GCCAAAGTTCACGAAGTTGC	Cardiac Marker
mGATA4_A_Fw	GGGCTGGTGGAGGTTCTC	Cardiac Marker
mGATA4_A_Rev	TCAGTGCCTAGAGACGCAAG	Cardiac Marker
mGATA4_B_Fw	CTACCCACATACACCGCTTTC	Cardiac Marker
mGATA4_B_Rev	GCCGACTACCCAAGACTATCC	Cardiac Marker
mGATA4_C_Fw	CTCAGGGTGTTCGAGACCAG	Cardiac Marker
mGATA4_C_Rev	GGTTGCTCCAGAAATCGTG	Cardiac Marker
mMyh7_A_Fw	CCCCCTAAAGCAAAGCACTTA	Cardiac Marker
mMyh7_A_Rev	GTCCCTGGGATCATGGTG	Cardiac Marker
mMyh7_B_Fw	TGTAGGTGGCTCCGAGAAAG	Cardiac Marker
mMyh7_B_Rev	GGGGCTGTTCTACCCTTACC	Cardiac Marker
mMyh7_C_Fw	TGTCACAACAGCGGAGAATC	Cardiac Marker
mMyh7_C_Rev	TTGGATGACCCTCTTAGTGTTG	Cardiac Marker
mMyh6_A_Fw	GGGATGGGAGCTTGTGTG	Cardiac Marker
mMyh6_A_Rev	TGGGTAAGGGTCACCTTCTCT	Cardiac Marker
mMyh6_B_Fw	GGTCAGGATCTCTGGATTGG	Cardiac Marker
mMyh6_B_Rev	GCTGGACGGAGAGAGGAAC	Cardiac Marker
mMyh6_C_Fw	ATGGAAGATAAACCCCCACA	Cardiac Marker

mMyh6_C_Rev	AGCCTGGATAATCTGGTCCTC	Cardiac Marker
mTNNI3_A_Fw	TGAATTTCACAGGAGTGAAGGA	Cardiac Marker
mTNNI3_A_Rev	CCACGTTTCTGTCGGTTTC	Cardiac Marker
mTNNI3_B_Fw	GGGCTTCTGGGTTCAAGAAT	Cardiac Marker
mTNNI3_B_Rev	CTCTGCCCATCACCCTACC	Cardiac Marker
mTNNI3_C_Fw	GAGACAGGAAGTGCTCTAGAATCAT	Cardiac Marker
mTNNI3_C_Rev	CCCTTTCTGGTCTCTATCTACCC	Cardiac Marker