TBX1 TRANSCRIPTION FACTOR: MECHANISMS OF GENE REGULATION

Tutor
Prof. Antonio Baldini

Candidate
Andrea Cirino

COORDINATOR
Prof. Vittorio Enrico Avvedimento

Academic Year 2017/2018
# TABLE OF CONTENT

**ABSTRACT** .......................................................................................................................... 4

**ABBREVIATIONS** ................................................................................................................ 5

**LIST OF FIGURES AND TABLES** ........................................................................................ 7

**INTRODUCTION** .................................................................................................................. 9

- DiGeorge Syndrome and the candidate gene ........................................................................... 9
- *Tbx1* encodes a Transcription factor of the T-box family. ...................................................... 11
- *Tbx1* can regulate target genes without binding the DNA. .................................................... 15
- Chromatin remodeling and Transcription factor activity. ......................................................... 16
- *Tbx1* and the Second Heart Field (SHF) .................................................................................. 19

**AIMS** .................................................................................................................................... 22

**MATERIALS AND METHODS** ............................................................................................. 23

1. Experimental Models ............................................................................................................. 23
   - Mouse lines ......................................................................................................................... 23
   - Cell lines (P19Cl6, C2C12, mESC) ...................................................................................... 25
   - In vitro differentiation .......................................................................................................... 27

2. Molecular Biology procedures ............................................................................................... 29
   - RNA-seq ............................................................................................................................. 29
   - Assay for Transposase-Accessible Chromatin (ATAC-seq) and quantitative ATAC (Q-ATAC) .................................................................................................................. 30
   - Chromatin Immunoprecipitation (ChIP) and quantitative ChIP ....................................... 33
   - Antibodies .......................................................................................................................... 34

3. Bioinformatics analyses ......................................................................................................... 35
   - RNA-seq ............................................................................................................................. 35
   - ATAC-seq .............................................................................................................................. 36
   - Real-time PCR, statistical analysis ....................................................................................... 40

**RESULTS** ............................................................................................................................... 41

Chapter 1 ..................................................................................................................................... 41
TBX1 controls histone acetylation at the Mef2c-Anterior Heart Field (AHF) locus. ..............................................................................................................................................41

Chapter 2 .........................................................................................................................................................................................43

Generation of chromatin accessibility maps of P19Cl6 cells with and without Tbx1..................................................................................................................................................................................43

Intersection of ATAC-seq and ChIP-seq data: many TBX1 binding sites are located within closed chromatin.........................................................50

Some loci show delayed response to loss of Tbx1 dosage in time-course experiments.................................................................................................................................55

Chapter 3 .........................................................................................................................................................................................59

mESC differentiation and Tbx1 expression ........................................................................................................................................59

Chromatin remodelling and gene expression in differentiated mESCs WT and Tbx1−/−. .................................................................................................................................................................65

Chromatin remodelling studies in vivo: Initial experiments and future perspectives ..............................................................................................71

DISCUSSION ....................................................................................................................................................................................73

CONCLUSIONS .............................................................................................................................................................................77

ACKNOWLEDGMENTS ....................................................................................................................................................................78

REFERENCES ...............................................................................................................................................................................79

LIST OF PUBLICATIONS .................................................................................................................................................................90
**ABSTRACT**

The *Tbx1* gene encodes a transcription factor, TBX1, critical for heart development in several species, including humans. The haploinsufficiency of this gene is associated with DiGeorge Syndrome (DGS) named also 22q11.2 Deletion Syndrome (22q11.2DS) which is characterized by multiple congenital anomalies, including heart disease (CHD). The molecular mechanisms by which TBX1 regulates its targets are unclear. In my thesis work I have focused on chromatin interactions mechanisms. In a first set of experiments using as a model a specific target gene, I have demonstrated that loss of TBX1 correlate with acetylation of a specific enhancer named anterior Heart Field (AHF) of the *Mef2c* gene, a gene critical for cardiogenesis. The mechanisms by which TBX1 affects histone acetylation need to be clarified, but we could not demonstrate a direct interaction with HDAC1 and HDAC2. Most of my thesis work has been dedicated to understanding the role of TBX1 in chromatin remodelling. By manipulation of two different model systems, I have generated maps of the accessible regions in different dosages of *Tbx1*. In P19Cl6 cells, I found that 86% of TBX1 binding sites are in closed chromatin. After *Tbx1* knock down, I found that differentially accessible regions (DARs) are not localized in regions bound by TBX1. Consistent with this finding, I did not find T-box motifs in DARs. However, a limited study using time-course experiments identified a delayed chromatin remodelling in selected loci bound by TBX1. In contrast, a study of chromatin remodelling in differentiated murine embryonic stem cells (mESCs), WT and *Tbx1*−/− revealed that DARs do have T-box motifs suggesting that a good portion of chromatin changes may be located in TBX1-binding regions. Comparison between P19Cl6 and mESCs reveals differences about DARs binding motifs and communalities about the increase numbers of accessible regions after *Tbx1* loss of function. In conclusion, my studies revealed new insight into the mechanisms by which TBX1 affects the chromatin landscape and indicate that mechanisms may be different depending on the cellular context.
ABBREVIATIONS

- 22q11.2 Deletion Syndrome → 22q11.2DS
- 5-Azacytidine → 5-Aza
- acetylation of lysine 27 of histone 3 → H3K27Ac
- Adrenocorticotropic Hormone → ACTH
- Anterior Heart Field → AHF
- Assay for Transposase-Accessible Chromatin → ATAC
- Atrial Siphon Muscle → ASM
- Bone Morphogenetic Factor → BMF
- Cardiac Progenitors → CP
- Chromatin Immunoprecipitation → ChIP
- Clustered Regularly Interspaced
- Coloboma, Heart defects, Atresia chonae, Retarded growth, Genital hypoplasia, Ear anomalies → CHARGE
- complementary Deoxyribonucleic acid → cDNA
- Congenital Heart Defects → CHD
- Deoxyribonucleic acid → DNA
- Differentially Accessible Regions → DARs
- Differentially Expressed → DE
- DiGeorge Syndrome → DGS
- Dimethyl sulfoxide → DMSO
- epithelial-like layer of the SHF → eSHF
- First Heart Field → FHF
- Fluorescent Activated Cell Sorter → FACS
- Gene Ontology → GO
- Holt-Oram Syndrome → HOS
- Hypergeometric Optimization of Motif Enrichment → HOMER
- Knock-down → KD
- Low Copy Repeat → LCR
- Mitochondrial reads → M
- monomethylation of lysine 4 of histone 3 → H3K4me1
- murine Embryonic Stem cells → mESC
- Polymerase Chain Reaction → PCR
- Prepulse Inhibition → PPI
- Principal Component Analysis → PCA
- Ribonucleic acid → RA
- Second Heart Field → SHF
- Serum Response Factor → SRF
- small interfering RNA → siRNA
- Small Patella Syndrome → SPS
- Spatial Clustered Identification of ChIP-Enriched Regions → SICER
- Systematic Evolution of Ligands by Exponential Enrichment → SELEX
- Transcriptional Start Site → TSS
- Ulnar-Mammary Syndrome → UMS
- Vascular Endothelial Grown Factor → VEGF
- Velocardiofacial syndrome → VCFS
- Ventricular Septal Defects → VSD
LIST OF FIGURES AND TABLES

Figure 1: Structure of the TBX3/DNA Complex.
Figure 2: Selection of specific oligonucleotides bound to TBX1.
Figure 3: TBX1 de novo motif.
Figure 4: Structure of the TBX1/DNA Complex.
Figure 5: An evolutionary conserved cardiopharyngeal ontogenic motif.
Figure 6: murine Embryonic Stem Cells (mESCs) differentiation.
Figure 7: ATAC-seq schematic pipeline.
Figure 8: Custom Nextera PCR Primers.
Figure 9: Sam output structure.
Figure 10: Adapter contamination in raw reads.
Figure 11: Contamination free in raw reads.
Figure 12: Tbx1-KD in C2C12 cells.
Figure 13: TBX1 and H3K27Ac enrichment at Mef2c-AHF enhancer in C2C12 cells.
Figure 14: P19Cl6 cells differentiation until D1.
Figure 15: P19Cl6 RNA quality.
Figure 16: Tbx1-KD in P19Cl6 cells.
Figure 17: DNA concentration and size distribution in P19Cl6 cells.
Figure 18: Total accessible regions distribution in Control and Tbx1-KD P19Cl6 cells.
Figure 19: Intersection between Control and Tbx1-KD accessible regions and TBX1-binding regions in P19Cl6 cells.
Figure 20: Intersection between Control and Tbx1-KD accessible regions and TBX1-binding regions based on genome coverage in P19Cl6 cells.
Figure 21: Intersection between differential accessible regions (DARs)/ Total Regions and TBX1-binding sites in P19Cl6 cells.
Figure 22: Motifs discovery in differential accessible regions (DARs) which increase accessibility in P19Cl6 Tbx1-KD cells.
Figure 23: Putative responsive loci associated to differential accessible regions (DARs) which increase accessibility in P19Cl6 Tbx1-KD cells.
Figure 24: P19Cl6 cells differentiation until D2.
Figure 25: Tbx1-KD in P19Cl6 cells during differentiation.
Figure 26: Putative responsive loci to Tbx1-KD in P19Cl6 cells.
Figure 27: Tbx1-overexpression in P19Cl6 cells during differentiation.
Figure 28: Putative responsive loci associated to Tbx1 overexpression in P19Cl6 cells.
Figure 29: Tbx1 and cTnt2 expression during murine embryonic stem cells (mESCs) differentiation.
Figure 30: Chromatin accessibility variations in genes involved in cardiac differentiation.
Figure 31: FACS-sorted murine embryonic stem cells (mESCs) differentiated at d4.
Figure 32: Tbx1 expression in murine embryonic stem cells (mESCs).
Figure 33: Total accessible regions distribution in d4-PDGFRa+, Tbx1 WT and KO murine embryonic stem cells (mESC).
Figure 34: Motifs discovery in d4-PDGFRa+ differential accessible regions (DARs) which increase accessibility murine embryonic stem cells (mESCs) Tbx1-KO cells.
Figure 35: Motifs discovery in regions which gain accessibility in d4-PDGFRa+ murine embryonic stem cells (mESCs) Tbx1-KO cells.
Figure 36: T-box motif in regions which gain accessibility in d4-PDGFRa+ murine embryonic stem cells (mESCs) Tbx1-KO cells.
Figure 37: RNA-seq data on d4-PDGFRa+, Tbx1 WT and KO murine embryonic stem cells (mESCs).
Figure 38: Tbx1 expression in Tbx1cre/+; R26mt/mG embryo.
Figure 39: DNA size distribution in Tbx1cre/flox; R26mt/mG embryos.
Figure 40: Cartoon showing a working model by which TBX1 may regulate target genes.

Table 1: DNA binding Motifs for T-box Transcription Factors identified with different techniques.
Table 2: Antibody list used in this work.
Table 3: Alignment rate (%) before and after removing of Nextera transposase Sequence in P19Cl6 cells.
Table 4: Comparison between the number of total reads with and without Mitochondrial reads in P19Cl6 cells.
Table 5: Total Peaks number in P19Cl6 cells.
Table 6: Genes involved in cardiopharyngeal mesoderm during murine embryonic stem cells (mESCs) differentiation.
Table 7: Total Peaks number in d4-PDGFRa+, Tbx1 WT and KO murine embryonic stem cells (mESCs).
Table 8: Total pooled Peaks number in d4-PDGFRa+, Tbx1 WT and KO murine embryonic stem cells (mESCs).
Table 9: Common genes associated to differentially accessible regions (DARs) and differentially expressed (DE) genes in d4-PDGFRa+, Tbx1 WT and KO murine embryonic stem cells (mESCs).
Table 10: Differentially expressed (DE) genes between d4-PDGFRa+, Tbx1 WT and KO murine embryonic stem cells (mESCs).
**INTRODUCTION**

*DiGeorge Syndrome and the candidate gene.*

The gene studied in the thesis work was identified in an effort to isolate the gene/s involved in DiGeorge syndrome (DGS) or 22q11.2 deletion syndrome (22q11.2DS) (Lindsay EA et al., 2001; Merscher S et al., 2001). Congenital heart disease (CHD) affects 8/1000 live births. A common genetic cause of CHD is the 22q11.2 deletion, also known as DiGeorge syndrome (DGS) (McDonald-McGinn DM et al., 2015). It has been estimated that a substantial portion of patients with some specific heart defects have 22q11.2 deletions: 52% of those with interrupted aortic arch type B, 34% with truncus arteriosus, 16% with tetralogy of Fallot and 5-10% with Ventricular Septal Defects (VSD). Besides CHD, patients have a number of other phenotypic features, for example cleft palate, development disabilities and schizophrenia. At the genetic level, the deletion could result from aberrant homologous recombination between low copy repeat (LCR) sequences, which flank the deleted regions (Edelman L et al., 1999). The name of this genetic disorders derives from Angelo DiGeorge who, in the late ‘60s, described this syndrome characterized by aberrant development of the thymus and parathyroid. Some of the development anomalies of DGS were also reproduced on chick models of neural crest ablation (Farrel MJ et al., 1999) leading investigators to hypothesize that DGS may derive from abnormal development of neural crest cell-contributed organs. However, more recent studies have implicated other cell lineages, specifically the cardiopharyngeal mesoderm, that will be discussed later. DGS is caused by chromosomal microdeletion of chromosome 22, at q11.2 locus; therefore, the disease is now commonly referred to as 22q11.2 deletion syndrome. Most patients with this syndrome have a large (3Mb) genomic deletion. About 10% of patients have a smaller deletion of about 1.5 Mb. Most genes localized in the region are conserved in the mouse on chromosome 16 (Gong W. et al., 1996; Botta A et al., 1997; Sutherland HF et al., 1998; Puech A et al., 1997; Lund J et al., 1999). This system allowed for the engineering of the first model, named Df1, which carries a deletion that encompasses mouse homologues of 18 genes that are deleted in patients with 1.5 Mb deletion whose phenotype is characterized by heart defects, thymus and parathyroid defects (Lindsay EA et al., 1999). In particular, the cardiac defects may be rescued in Df1 mice on chromosome 16 by reciprocal duplication (Dp1) on the homolog, restoring normal dosage of the Df1 region. Further mouse studies discovered that mutation of *Tbx1*, included in the Df1 region, is sufficient to recapitulate the phenotypic spectrum of the disease.
(Jerome LA et al., 2001; Linsday EA et al., 2001; Merscher S et al., 2001). The role of \(TBX1\) in the human disease has been subsequently confirmed in human patients with clinical phenotypes consistent with DGS but lacking chromosomal deletions, but had mutations of \(TBX1\) gene (Yagi H et al., 2003). These studies demonstrated that heterozygous mutation of \(TBX1\) is sufficient to cause most of the clinical features of 22q11.2DS patients. \(Tbx1\) mouse mutants show different phenotypes also seen in patients, for example: \(Tbx1\)-null mice have a small otocyst that fails to grow and does not give rise to the vestibular and cochlear apparata (Vitelli F et al., 2003). Auditory problems are common in 22q11.2DS patients (McDonald-McGinn DM et al., 2015). In addition, loss of \(Tbx1\) causes abnormalities of lymphatic vessel development as it regulates \(Vegfr3\), a gene essential for lymphangiogenesis (Chen L et al., 2010). Lymphatic vessel defects have been found in 22q11.2DS patients (Unolt M et al., 2018).
**Tbx1 encodes a Transcription factor of the T-box family.**

The T-box family of transcription factors includes TBX2, TBX3, TBX4, TBX5, TBX6, TBX10, TBX13, TBX14, TBX15, TBX18, TBX19, TBX20, TBX21, TBX22, BRACHYURY, T-BRAIN1, EOMESODERMIN, and of course TBX1. These T-box proteins have a common DNA-binding motif called “T-domain” that binds DNA in a sequence-specific manner (Papaioannou VE, 2014). Genes encoding them have a vital role in embryogenesis and more recently they have also been implicated in cancer biology. The T-box binding element is a palindromic DNA sequence with strong affinity for BRACHYURY, also called as “T”, which interacts with that sequence in a dimeric form, each monomer binds a half site called T-half site (5’-AGGTGTGAAATT-3’) (Kispert A et al., 1993). The T-gene was discovered in 1927 because of a spontaneous mutation, which caused truncated tails in mice (Dobrovolskiaia-Zavadskiaia N et al., 1927).

The T-box domain is about 180 amino acids long. The T-box domain, mostly situated in the middle portion of the protein, is a stretch of 180-190 aa residues and is defined as the minimal region that is necessary for the sequence-specific binding to DNA (Papaioannou VE et al., 1998). Different human genetic diseases are associated with mutations of T-box genes, for example mutation in Tbx3 causes an autosomal dominant disorder characterized by mammary gland hypoplasia, dental and genital abnormalities (Bamshad M et al., 1997; 1999) and upper limb malformations called Ultrasound Mammary Syndrome (UMS). Mutation in Tbx4 are associated with another genetic disorder named small patella syndrome (SPS) related to problem in skeletal development of patella (Bongers EM et al., 2004). Alteration in Tbx5 lead to an autosomal dominant disorder, Holt-Oram syndrome (HOS) in which most affected patients exhibit cardiac and limb malformations (Li QY et al., 1997). Finally, Yi et al., 1999 demonstrated that mutations of Tbx19 are associated with loss of adrenocorticotrophic hormone (ACTH) and melanocyte stimulating hormone in the corticotroph and melanotroph cell lineages in the pituitary resulting in adrenal insufficiency.

T-box proteins have a variable weight: the range is from 50 to 78 kDa. Although, as described above, BRACHYURY binds DNA as a dimer (Papapetrou C et al., 1997) and exhibits a common human polymorphism Gly-177-Asp in the conserved DNA-binding domain, with each monomer binding half of the sequence, or T-half site (5’-AGGTGTGAAATT-3’). The T-box domains change between the different T-box proteins, although some specific residues are 100% conserved between the different T-box domains whose preference for different combinations of orientations, number and spacing of T-half sites may help to create binding specificity for target genes.
(Conlon FL et al., 2001). The figure 1 shows a ribbon diagram of a human TBX3 monomer bound to its DNA target site.

Fig. 1

![Ribbon diagram of a human TBX3 monomer bound to its DNA target site. Secondary structure elements are labelled, and helices, strands, and loops are depicted in turquoise, red, and grey, respectively (Coll M et al., 2002).](image)

After the discovery of the first Tbx1 mutations (Yagi H et al., 2003), additional patients have been described and two of them had a truncating mutation that resulted in loss of function due to the deletion of a C-terminal nuclear localization (Stoller JZ et al., 2005). Another missense mutation has been described in a familial case of Shprintzen syndrome. This missense mutation results in gain of function, possibly through stabilization of the protein dimer DNA complex (Zweier C et al., 2007). Prepulse inhibition (PPI) deficits in Df1/+ mice are caused by haploinsufficiency of Tbx1 and mutation in this gene is sufficient to cause reduced PPI (Paylor R et al., 2006). Screening of TBX1 coding sequence identified a frameshift deletion (1320-1342del23bp) in patient with characteristic facial appearance of velocardiofacial (VCFS) and hypernasal speech (Paylor R et al., 2006). Overall very few Tbx1 mutations have been reported suggesting that they are rare. TBX1 binds to a consensus sequence that has been identified recently. Two studies have identified the consensus using different technologies: Systematic Evolution of Ligands by Exponential Enrichment (SELEX) (Castellanos R et al., 2014) and ChIP-seq (Fulcoli FG et al., 2016). TBX1 preferentially binds to a tandem repeat of 5’-AGGTGTGAAGGTGTG-3’
but it can also interact with Half Sites. The figure 2 shows the motifs identified with SELEX (Tandem and Half site).

Fig. 2

![Sequence alignment showing the optimal DNA binding motif for TBX1](image)

Figure 2. **Selection of specific oligonucleotides bound to TBX1.** Sequence alignment shows that the optimal DNA binding motif for TBX1 is AGGTGT(G/T) (A/T) followed by two repeated similar motifs termed the Tandem Repeat (TR) and Half Site Partial Site as shown (½SPS) (Castellano R et al., 2014)

De novo motif discovery, using ChIP-seq with a Tbx1 antibody, has uncovered an 8-bp consensus sequence (Fig. 3).

Fig. 3

![Sequence logo representing the enriched Tbx1 motif identified by de novo motif discovery](image)

Figure 3. **TBX1 de novo motif.** Sequence logo representing the enriched Tbx1 motif identified by de novo motif discovery (Fulcoli FG et al., 2016).
Table 1 lists the consensus sequences for a group of T-box proteins.

### Table 1

<table>
<thead>
<tr>
<th>Gene</th>
<th>Publication</th>
<th>Sequence</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tbx1</td>
<td>Fulcoli FG et al., 2016</td>
<td>GG A G G/C T G G</td>
<td>ChIP-seq</td>
</tr>
<tr>
<td>Tbx1</td>
<td>Castellanos R et al., 2014</td>
<td>A G G T G T G/T A/T</td>
<td>SELEX</td>
</tr>
<tr>
<td>Tbx5</td>
<td>Narlikar L et al., 2013</td>
<td>GG/AA G G/A T G G/A</td>
<td>ChIP-seq</td>
</tr>
<tr>
<td>Tbx5</td>
<td>Mori AD et al., 2006</td>
<td>A/G/T G/A G T G N N A</td>
<td>ChIP-seq</td>
</tr>
<tr>
<td>Tbx5</td>
<td>Luna-Zurita L et al., 2016</td>
<td>G A G G T G</td>
<td>ChIP-EXO</td>
</tr>
</tbody>
</table>

Table 1. DNA binding Motifs for T-box Transcription Factors identified with different techniques. The table is divided in Gene, Publication, Sequence and Method used (Baldini A et al., 2017).

How does TBX1 bind DNA? The T-box domain of TBX1 is composed of a seven beta-barrel domain core related to an s-type immunoglobulin fold (Bork P et al., 1994), and closed by a smaller beta-pleated sheet as illustrated in the Fig. 4 (El Omari K et al., 2012). It has been proposed that TBX1 binds the DNA as monomer based on the observation that the molecular surface for potential dimerization is too small to be of biological significance.

Fig. 4

![Figure 4. Structure of the TBX1/DNA Complex.](image)

Figure 4. Structure of the TBX1/DNA Complex. Ribbon diagram of a TBX1 bound to its DNA target site. TBX1 is in yellow (monomer “a” light yellow, monomer “b” dark yellow), DNA is in blue) (El Omari K et al., 2012).
Tbx1 can regulate target genes without binding the DNA.

DNA binding is not the unique mechanism by which Tbx1 can function. Two papers published in 2009 reported that TBX1 interacts with specific proteins and its biological function is independent from its binding to DNA. Fulcoli FG et al., 2009 have demonstrated an interaction between TBX1 and SMAD1. SMAD1 is a member of SMADS family of proteins that are key signal transducers downstream of the TGF-Beta superfamily type I receptors (Mehra A et al., 2000). SMAD1 has a central role in BMP signal transduction essential for mesoderm formation, cartilage development, postnatal bone formation and heart development. In addition, BMP4 plays an important role during gastrulation stage, in particular conditional deletion of BMPR1a in embryos showed a shortened cardiac outflow tract with septation defects, a process known to require neural crest and is necessary for perinatal viability (Stottmann RW et al., 2004). The interaction TBX1-SMAD1 suppresses SMAD1 binding to SMAD4, and a Tbx1 mutation that prevents binding to DNA does not affect binding to SMAD1 nor does it affect the ability to suppress SMAD1 activity (Fulcoli FG et al., 2009). Thus, TBX1 can affect the BMP signal transduction mechanism without binding the DNA. The TBX1-SMAD1 interaction is not the only mechanism by which TBX1 acts without binding to DNA. In another paper it has been shown that Tbx1 plays an important role in regulating proliferation and differentiation of multipotent heart progenitors by interacting with another transcription factor, Serum Response Factor (SRF), a master regulator of muscle differentiation, and it regulates negatively its level. The TBX1 effect on SRF dosage was not due to transcriptional downregulation but was proteasome-dependent (Chen L et al., 2009). The TBX1-SRF interaction is potentially important for regulating the differentiation of cardiac progenitors (CP). Indeed, it was demonstrated that TBX1 promotes CP proliferation and inhibits their differentiation (Chen L et al., 2009).
Chromatin remodeling and Transcription factor activity.

Although a systematic, unbiased screening of TBX1 protein interactions has not been done, an initial picture of TBX1-chromatin interactions is emerging. Many biological processes, in particular developmental programs, are controlled, by transcription factors and chromatin regulators. The cooperation between transcription factor and chromatin remodelers is necessary to maintain specific gene expression programs through epigenetic modification of the genome (Zaret KS et al., 2011). Thanks to new technologies it is becoming clear that there is a strong correlation between genome and epigenome. The epigenetic landscape of enhancer elements is important for commitment of embryonic stem cells and their self-renewal. There are at least two post-translational histone modifications correlated with TBX1 known to date: H3K27Ac (the acetylation of lysine 27 of histone 3) and H3K4me1 (the monomethylation of lysine 4 of histone 3). H3K27Ac distinguishes active enhancers from primed ones, which are characterized only by H3K4me1 (Creighton MP et al., 2010). Chen et al., 2012, have demonstrated an interaction between TBX1 and a methyltransferase SETD7, but it has later been shown that this occurs outside from the chromatin context, so the significance of this interaction is unclear (Fulcoli FG et al., 2016). In contrast, in the chromatin context, TBX1 interacts with KMT2C (MLL3). KMT2C and LSD1, a histone demethylase, maintain the methylation level of H3K4. ChIP-seq using anti TBX1 and H3K4me1 antibodies revealed a highly significant overlap between TBX1 binding sites and H3K4me1-enriched regions. Furthermore, ChIP-seq with an antibody anti H3K27Ac demonstrated that TBX1 binds to H3K27Ac-poor regions (Fulcoli FG et al., 2016). Whether TBX1 actively maintains histone hypoacetylation is still unclear. ChIP-seq and RNA-seq correlations using P19Cl6 with and without Tbx1 knock down suggested that Tbx1 is neither a strong activator nor a strong repressor. In general, T-box proteins can be activators or repressors (Kawamura A et al., 2008), depending on biological context. For example, Tbx1 is “activator” for mechanisms correlated to proliferation and “repressor” for mechanisms correlated to differentiation of muscle cells (Chen et al., 2012). It is possible that the TBX1 function is to prime target enhancers and make them accessible to other regulators which can act as activator or repressor (Baldini A et al., 2017). It is possible that the Tbx1 haploinsufficiency phenotype could be a direct consequence of reduced H3K4me1, resulting in reduced accessibility of specific enhancers (Baldini A et al., 2017).

TBX1 interacts also with ASH2l, the mammalian homolog of Drosophila ASH2 (absent small homeotic 2), a core component of multimeric histone
methyltransferase necessary for methylation of histone lysine residues. The interaction could be seen with a yeast 2-hybrid system and in mammalian cells, although its biological significance is unknown. (Stoller JZ et al., 2010). TBX1 interacts also with Nkx2-5 which encodes a homeobox-containing transcription factor required for heart development (Nowotschin S et al., 2006). Experiments in Ciona intestinalis have revealed an antagonism mechanism between TBX1 and NKX2-5. Specifically, in cells undergoing asymmetric cell division of common progenitors, Nkx2-5 promotes GATA expression and cardiac specification in the second heart precursor by antagonizing Tbx1-mediated inhibition of GATA and activation of determinants of atrial siphon muscle (ASM) specification, the homologous regions of mastication muscle in mammalian (Wang W et al., 2013).

Furthermore, it has been reported a genetic interaction between Chd7 and Tbx1, in particular it was found that these two genes were in epistasis and that the correct level of their expression was required in pharyngeal ectoderm for proper morphogenesis of the great arteries (Randall V et al., 2009). The Chd7 gene encodes a chromo-domain containing, chromatin remodeling protein, involved in CHARGE syndrome (an acronym for Coloboma, Heart defect, Atresia choanae, Retarded growth, Genital hypoplasia, and Ear anomalies) which has some similarities with DGS. The mechanism of genetic interaction is unclear but CHD7 binds approximately 10000 regions in the mouse genome and most of them localize in H3K4me1 rich regions (Schnetz MP et al., 2010). This parallelism suggests that TBX1 and CHD7 may co-localize in some enhancers.

Chen and co-workers (Chen L. et al., 2012) found an interaction between TBX1 and BAF60A also known as SMARCD1. This protein belongs to a SWI/SNF complex, whose members have helicase and ATPase activities and can regulate gene expression by remodelling chromatin (Hsiao PW et al., 2003). TBX1 co-immunoprecipitates with BAF60A and is able to recruit BAF60A onto the target gene Wnt5a. Knock down of Baf60a affects the ability of TBX1 to regulate different target genes in vitro, including Wnt5a. Tbx1 and Wnt5a interact genetically because the loss of both genes produces a more severe phenotype than loss of one of them in vivo. In cultured cells, TBX1 binds the T-box binding elements of the Wnt5a gene (Chen et al., 2012). It may be that TBX1 is able to interact with BAF60A and recruit the member of the complex in proximity to its target sites where it remodels chromatin and active target enhancers through recruiting of histone methyltransferase. Time-course experiments in vitro and expression analyses in vivo suggest that Baf60a is expressed in most tissues during mouse embryonic development and it is downregulated during cardiac differentiation. In contrast the expression of Baf60c (Smarcd3), an alternative member of the same complex, has an inverse correlation with cardiac
differentiation. Overall, the review of the literature presented here suggests that $Tbx1$ functions may be related to epigenetic modifications operated through interactions with other transcription factors, histone modifiers and chromatin remodelers.
Tbx1 and the Second Heart Field (SHF).

Tbx1 expression is variable both spatially and temporally across tissues, it is expressed in mice during embryogenesis and is finely regulated (Xu H et al., 2005). Specifically, it is expressed in the pharyngeal apparatus which includes pharyngeal endoderm, ectoderm, mesodermal core of pharyngeal arches, head mesenchyme and the second heart field (SHF) (Xu H et al., 2005). The SHF is defined as a reservoir of cardiac progenitors, which gradually migrate into the heart and contribute to the growth of the outflow tract, the right ventricle and atria (Mjaatvedt CH et al., 2001; Kelly RG et al., 2001; Waldo KL et al., 2001). The amniote heart is made up of cardiomyocytes arisen from two different adjacent progenitor cell populations in the early embryo (Meilhac SM et al., 2004). Early differentiating cardiac progenitor cells of the first heart field (FHF) give rise to the linear heart tube, that eventually becomes the left ventricle, and to part of the atria (Kelly RG 2012, Tzahor E et al., 2011). Progenitors of the SHF derive from the primitive streak and later localize to the pharyngeal and splanchnic mesoderm. These progenitors are multipotent and will give rise to cardiomyocytes and endothelial cells of the outflow tract, right ventricle and most of the atria (Kelly RG et al., 2001; Mjaatvedt CH et al., 2001). The SHF is divided into two subpopulations, the anterior and posterior SHF that contribute to the arterial and venous poles of the heart, respectively (De Bono C et al., 2018). SHF-derived segments of the heart share a lineage relationship with craniofacial skeletal muscle (Lescroart F et al., 2014) revealing that there are common progenitors that give rise to both cardiac and craniofacial muscle cells. This newly discovered lineage is conserved across species and has been named the cardiopharyngeal mesoderm (Diogo R et al., 2015) (Fig. 5).
Genetic experiments in mouse and other organisms have shown that Tbx1 is necessary for outflow tract development and it also affects the development of craniofacial muscle as well as other organs derived from the pharyngeal apparatus. Therefore, Tbx1 is hypothesized to be a major regulator of the development of the cardiopharyngeal mesoderm lineage. Genome-wide gene expression analysis in Tbx1 mutant embryos have highlighted a number of pathways that are affected by the loss of the gene. For example, in one study Hod and Nkx2-6 were downregulated in Tbx1 null embryos, in contrast some of the genes necessary for cardiac morphogenesis, such as Gata factors (Gata4), Raldh2 and Tbx5, and a subset of muscle genes were ectopically expressed (Liao J et al., 2008). Tbx1 also downregulates the expression of Vegfr2 in the posterior SHF and the expression of Vegfr2 is up regulated and extended in the absence of Tbx1 (Lania G et al., 2015). Tbx1 plays an important role for tissue architecture. Pathway analyses of Tbx1 target genes identified genes involved in focal adhesion and tissue architecture (Fulcoli FG et al., 2016). Recent data from the lab have confirmed that Tbx1 is required for the integrity of the axis ECM-Integrin-Focal adhesion in the SHF (Alfano D et al., 2018 Biorxiv). The communication between extracellular...
matrix, integrin and focal adhesion, is altered in SHF of Tbx1 null mice, suggesting that Tbx1 is required for correct ECM-cell interactions. In particular, in the epithelial-like layer of the SHF (eSHF), Paxillin, Vinculin, E-cadherin, F-actin and NMIIB are mis-localized, possibly compromising the cohesiveness of the epithelial cell layer. Discoveries concerning the exact relationship between Tbx1 and SHF development will lead new insights into the pathogenesis of congenital heart defects.

The developmental roles of Tbx1 has been under study for years. In Tbx1-null mice, outflow tract progenitor cells fail to expand and contribute to the dorsal pericardial wall and then to the heart poles, leading to abnormalities of the outflow and inflow tracts (Rana MS et al., 2014). The developmental roles of Tbx1 are at least partially conserved during evolution. For example, in zebrafish the loss of Tbx1 is associated with ventricular and outflow tract defects consistent with a conserved role in SHF-mediated cardiogenesis. More in details, in Tbx1 null animals there were 25% less cardiomyocytes that contributes to the heart tube suggesting a defective proliferation of cardiac precursor (Nevis K et al., 2013).
AIMS

The overall aim of my thesis work is to explore mechanisms by which the transcription factor TBX1 regulates its targets through interaction with chromatin. In particular, I will address the following specific aims:

1) To understand the relationship between TBX1 and histone acetylation.

TBX1 binds H3K4me1-rich regions and H3K27Ac-poor regions. I will ask the question if TBX1 has an active role in maintaining the histone acetylation status on chromatin. To do this, I will use a cell culture model to understand if TBX1 dosage affects the acetylation status of the Mef2c “anterior heart field” enhancer.

2) To map chromatin accessibility genome-wide with and without TBX1.

I will use a recently developed technology, known as “Assay for Transposase-Accessible Chromatin followed by deep sequencing” (ATAC-seq) to quantitatively measure chromatin accessibility genome-wide. I will use two cell models: mouse P19Cl6 cells and mouse embryonic stem cells (mESC). Both models can be differentiated in vitro. Manipulation of TBX1 dosage will be obtained by small interfering RNA (for P19Cl6 cells) or by gene deletion (mESC). Finally, I will extend the experiments to genetically-labelled cells purified from heterozygous and homozygous mutant mouse embryos.

3) To determine the correlation between changes in chromatin accessibility and TBX1 binding to chromatin.

I will integrate chromatin accessibility data with available map of TBX1 binding sites and gene expression data. I will ask whether TBX1 has a local impact on chromatin remodelling and, if so, which genes respond to TBX1 dosage changes.
MATERIALS AND METHODS

1. Experimental Models

- **Mouse lines**

  \( Tbx1^{\text{Cre/+}} \)
  Cre-recombinase is knocked-into exon 5 of the \( Tbx1 \) gene. The insertion caused inactivation of the gene, so \( Tbx1^{\text{Cre/+}} \) animals are functionally heterozygous mutants (Huynh T et al., 2007). In this line, Cre is faithfully expressed in the \( Tbx1 \) expression domain. The Cre/loxP system has been used for many years in conditional mutagenesis in mice.

  \( R26R^{mT-mG} \)
  This is a Cre recombination reporter line. Reporters of Cre enzyme are important for defining the spatial and temporal extent of Cre-mediated recombination. \( R26R^{mT-mG} \) is a double-fluorescent Cre reporter mouse that expresses membrane-targeted tandem dimer Tomato (mT) in every tissue. Upon Cre-mediated recombination, it switches to expression of a membrane-targeted green fluorescent protein (mG) (Muzumdar MD et al., 2007).

  \( Tbx1^{\text{flox/+}} \)
  This line carries a loxP-flanked (flox) exon 5 of the \( Tbx1 \) gene (Xu et al., 2004). Upon Cre recombination, the floxed allele is excised and the gene inactivated.
  Embryos with genotype \( Tbx1^{\text{Cre/flox}} \) are functionally homozygous mutants but only in cells in which \( Tbx1 \) is expressed (and thus Cre is expressed). In embryos \( Tbx1^{\text{Cre/flox}}, R26^{mT-mG} \), homozygous (recombined) cells are identifiable because they express green fluorescence. Mice were genotyped by PCR using DNA extracted from tail biopsies (or embryo yolk sacs) using the following primers pairs:

  \[ \begin{align*}
  Tbx1\text{Cre}: & \quad Tbx1\text{Cre-F} (5'-\text{TGATGAGGTTCGCAAGAACC-3'}) \\
  & \quad Tbx1\text{Cre-R} (5'-\text{CCATGAGTGAACGAACCTGG-3'})
  \end{align*} \]

  \[ \begin{align*}
  Tbx1\text{flox}: & \quad Tbx1\text{flox-F} (5'-\text{CGACCCCTTCTCTGGCTATG-3'})
  \end{align*} \]
Tbx1flox-R (5’-AAAGACTCCTGCCCTTTTCC-3’)

R26R mT-mG:

TOM R1 AAAGTCGCTCTGAGTTGTAT
TOM R3 GGAGCGGGAGAAATGGATATG
pCAG GTCGTTGGGCGGTCAG

PCR products were separated on 2% agarose gel.
• **Cell lines (P19Cl6, C2C12, mESC)**

P19Cl6.
P19Cl6 cells are a clonal derivative isolated from murine P19 embryonic carcinoma cells (Mueller I et al., 2010). This Cl6 subline efficiently differentiates into beating cardiomyocytes with adherent conditions when treated with 1% DMSO. Cells are maintained a minimal essential medium supplemented with 10% fetal bovine serum, penicillin (100 U/mL), streptomycin (100 µg/mL) and L-glutamine (300 µg/mL). Cells were maintained at 37°C in 5% CO2.

C2C12.
C2C12 cells are mouse undifferentiated myoblast cells (ATCC CRL-1772) which express high levels of Tbx1 and can be manipulated through siRNA transfection. C2C12 were cultured in Dulbecco’s modified Eagle’s medium (DMEM, Invitrogen) supplemented with 10% FBS and L-glutamine (300 µg/mL). Cells were maintained at 37°C in 5% CO2.

Mouse embryonic stem cells (mESC).
Mouse embryonic stem cells (mESC) derive from the inner cell mass of blastocyst. Due to their ability to differentiate in many cell types, mESCs are largely used in the research lab for “in vitro” studies and it is a very powerful tool for genetic disease and obviously are extremely used in development and stemness field. We have used the line ES-E14TG2a (ATCC CRL-1821) that is grown in feeder free conditions. Cells were cultured in Glasgow medium (GMEM, Sigma), 15% FBS, ESGRO Chemicon (ESG1106), 100x β-mercaptoethanol (BME), 100x Glutammine, 100x Non-Essential Amino Acid (NEAA), 100x Sodium pyruvate (NA-Pyr), 100x PenStrep. Cells were maintained at 37°C in 5% CO2.

Gene knock down using small interfering RNA (siRNA).
In P19Cl6 siRNA transfection, cells were plated at 5.0 x 10⁵ per well in six-well plates and transfected with a pool of Silencer Select Pre-Designed Tbx1 siRNA (pool of s74767, s74768 and s74769, Life Technology) in antibiotic-free medium using Lipofectamine RNA iMAX Reagent (Life Technology) according to the instructions. In ATAC-seq experiment, 42 hours after siRNA, using in the same time a siRNA control, cells were harvested and processed. In time course experiments, 13 hours after siRNA, using in the same time a siRNA control, cells were harvested and processed.
In C2C12 siRNA transfection, cells were plated at $1.2 \times 10^5$ per well in six-well plates and transfected with a pool of Silencer Select Pre-Designed Tbx1 siRNA (pool of s74767 and s74769, Life Technology) in antibiotic-free medium using Lipofectamine RNA iMAX Reagent (Life Technology) according to the manufacturer’s instructions. 48 hours after siRNA, using in the same time a siRNA control, cells were harvested and processed for further analyses.
• **In vitro differentiation**

P19Cl6

P19Cl6 cells differentiate into beating cardiomyocytes in adherent conditions when treated with 1% DMSO. The protocol used for this cells was established in the lab, and consists in plating about 500,000 cells in a 35-mm dish. Next day, when the cells are confluent, it was added 10uM 5-Azacytidine (5-Aza) to induce differentiation. After 24h, it was added a fresh medium containing DMSO 1%. For time course experiment for both loss and gain of function, cells were collected at T1 (13 hr after Transfection), at D1 (24 hr after 5-Aza induction) and at Day2 (24 hr after DMSO adding).

mESCs

mESCs differentiate into the cardiac lineage using a protocol established in Dr. Keller’s laboratory (Keller G 2005). This protocol generates, starting from undifferentiated embryonic stem cells, cardiac precursor (CP) and cardiomyocytes (CM). mESCs were maintained at Day 0 using leukemia inhibitory factor (LIF, 10⁶ units/ml) when the cells were stimulated to proliferate as embryonic bodies by adding ascorbic acid (5 mg/mL) and monotioglycerol (MTG, dilution from 1.25 g/mL). At day 2, the differentiation can start by adding general growth factor: activin A (10 ng/µL), bone morphogenetic factor (BMP4, 10 ng/µL), vascular endothelial growth factor (VEGF 5 ng/µL) while at day 4 starts the specification and maturation of cardiomyocytes by plating cells on gelatin coat with fibroblast growth factors (FGFb, 10 ng/µL; FGF10, 50 ng/µL) and VEGF (5 ng/µL), therefore day 4 cells encompass cardiac precursor markers. At day 10, cells become beating cardiomyocytes. The Fig. 6 illustrates the differentiation scheme.
Figure 6. **murine Embryonic Stem Cells (mESCs) differentiation.** Three different phases that mESCs need to differentiate into beating cardiomyocytes: Proliferation, Mesoderm induction, Endothelial/Cardiac specification and maturation (Keller G 2005).

mESCs, differentiated at day 4, sorted for PDGFRa were isolated through flow cytometry. *Tbx1*-KO cells were generated in the lab using CRISPR-Cas9 technology (Clustered regularly interspaced short palindromic repeats).
2. Molecular Biology procedures

- RNA-seq

Cells in dishes were washed with PBS cold and then 1 mL of Trizol was added directly on a single dish. Lysate was then harvested and vortexed in order to promote lysis of cells. Then, 200 µl of chloroform was added to 1 mL in order to separate three distinct phases: upper phases (where we can find RNA), intermediated phases (where we can find DNA) and lower phases (where we can find proteins and other cells-derived). The mixture was centrifuged at 12000g for 15 min. The top (aqueous) phase was removed and transferred into a new tube in which was added 500 µl of isopropanol and the solution was incubated for 20 min at room temperature (RT). After 20 min, solution was centrifuged for 10 min at 12000g. Pellet (RNA) was washed two times with Ethanol 80% and centrifuged for 5 min at 7500 g. Pellet was resuspended in a fresh water and then processed for further analyses. The concentration was estimated with Nanodrop. All samples were run on agarose gel and all three different RNA fragments (28S, 18S and 5S) were distinguished. After quality/quantity check RNA samples were used for libraries preparation with the Illumina’s strand specific RNA seq protocol, barcoded and pooled in one lane. The raw data for sequencing of cDNA were generated with Illumina platform (NextSeq 500) for paired-end reads of length 75bp.
• Assay for Transposase-Accessible Chromatin (ATAC-seq) and quantitative ATAC (Q-ATAC)

This new method, developed by Buenrostro et al., 2015 is now widely used for mapping chromatin accessibility genome-wide. This technique uses Tn5 transposase which is able, when the chromatin is open, to insert sequencing adapters into accessible regions of chromatin. More in details, Tn5 is a prokaryotic transposase, which endogenously functions through the “cut and paste” mechanism, requiring sequence-specific excision of a locus containing 19 bp inverted repeats. The Fig. 7 is a schematic representation of the technique.

Figure 7. ATAC-seq schematic pipeline. Tn5 transposase (orange) insert sequencing adapters (green and red) in chromatin accessible regions. Open chromatin (violet) can be isolated, amplified and then sequenced (Buenrostro J et al., 2015).

Sequencing adapters, associated to open chromatin regions, can be amplified and then sequenced. Sequencing adapters associated with regions of increased accessibility were recognized by Customer Nextera PCR Primer 1 and Custom Nextera PCR primer 2 which contains barcode necessary for sample pooling. A complete list of primers is illustrated in the figure 8.
Figure 8. Custom Nextera PCR Primers. PCR primers are used to amplify fragment of opened chromatin. Primer 1 does not contain barcode while Primer 2 contains barcode necessary for sample pooling (Buenrostro J et al., 2015).

It is possible to map regions of transcription factor binding and nucleosome position. ATAC-seq has replaced DNase-seq for open chromatin regions and MNase-seq for assaying nucleosome position because it is fast and sensitive. A critical point of this new technique is cells number: in general, too few cells causes over-digestion of chromatin and appears to create a larger fraction of reads that map to inaccessible regions of the genome; using too many cells may cause under-digestion and creates high molecular weight fragments, which certainly is difficult to sequence. Cells were harvested and they absolutely must not be fixed, intact cells in a homogenous suspension tend to give the best results. In order to determine the quantification of DNA obtained is not recommended Qubit analysis but we have used Tapestation, automated sample processing for quality control of Next Generation Sequencing and microarray data.

For sequencing was used NextSeq 500, based on Illumina protocol, 60bp each reads in paired-end, MID flowcell, 240.000.000 total reads.

ATAC-seq was optimized in about three hours: 30 minutes for nuclei extraction, 45 minutes for transposition and purification, 1 hour and 45 minutes for PCR and purification and the remaining time for quantification analysis. More in details there are 4 main areas in which the protocol can be divided. The first one is cell preparation: we have used 50.000 cells for P19Cl6 cells and 12500 cells for mESCs. After washes in PBS, cells were suspended in 50 µL of cold lysis buffer (10mM Tris-HCl, pH 7.4, 10mM
NaCl, 3mM MgCl2, 0.1% IGEPAL CA-630) and immediately spin down at 500 x g for 10 min at 4 °C. The second part consists of Transposition mix and Purification, the nuclei were incubated at 37 °C in Transposition Reaction Mix (25 μL reaction buffer, 2.5 μL of Transposase, 22.5 μL Nuclease free water), purified using Qiagen MinElute PCR Purification Kit and eluted in 10 μL of Nuclease Free water. The Third part consists of PCR amplification of 10 μL Tagmented DNA with 2.5 μL of PCR primer without barcode, 2.5 μL of specific barcode PCR primer, 10 μL Nuclease free water and 25 μL NEBNext High-Fidelity 2x PCR Master Mix. The PCR cycles are: 1 cycle of 72°C for 5 min (critical to allow extension of both ends of the primer after transposition), 4 cycle of 98 °C for 10 secs, 63 °C for 30 secs and 72°C for 1 min. In order to reduce GC and size bias in PCR, the appropriate number of PCR was calculated through real-time to stop amplification prior to saturation (Buenrostro J et al., 2013). The last part consists of assessing the quality/quantity of DNA purified which represent the regions of accessible chromatin.

ATAC can be also associated to a screening, through real-time, of different loci in order to estimate how chromatin accessibility changes. This technique is called quantitative ATAC (Q-ATAC). In this work, Q-ATAC has done on a set of loci during loss and gain of function time-course experiments.
**Chromatin Immunoprecipitation (ChIP) and quantitative ChIP**

C2C12 cells were cross-linked with 1% formaldehyde for 15 min at room temperature and glycine was added to stop the reaction to a final concentration of 0.125 M for 5 min. The cell pellet was suspended in 6 x volumes of cell lysis buffer (10 mM HEPES, 60 mM KCl, 1 mM EDTA, 0.075% v/v NP40, 1 mM DTT and 1X protease inhibitors, adjusted to pH 7.6) in a 1.5 mL tube incubating on ice for 15 min. Isolated nuclei were suspended in Buffer B of LowCell ChIP Kit reagent and chromatin was sonicated into 200-500 bp long fragment using the Covaris S2 Sample Preparation System (Duty Cycle: 5%, Cycles: 5, Intensity: 3, Bath temperature: 4 °C. Cycles per Burst: 200, Power mode: Frequency Sweeping, Cycle Time: 60 seconds, Degassing mode: Continuous). After sonication, chromatin was diluted in Buffer A, according with LowCell Kit in order to make a SDS-dilution (SDS-high concentration may interfere between antibody and protein interaction). Chromatin was incubated with anti H3K27Ac (Abcam, ab4729), or normal rabbit IgG (Santa Cruz Biotechnology, 2027). Next, steps included extensive washes and reverse crosslinking following the same kit previously described. For quantitative ChIP, I performed real-time PCR of the immunoprecipitated DNA and inputs, using the FastStart Universal SYBR Green Master kit (Roche) on System or Step-one plus (Applied Biosystems) using primers specific for the AHF-enhancer region of Mef2c.

Forward primer: TGAGGAGGGAGCTGCAGTAT
Reverse primer: CCGTTTCTCTATCCCAACCA

The experiment was performed two different Tbx1 dosages (Control and Tbx1-KD). The H3K27Ac around AHF-enhancer region was calculated on 1% of input (sonicated and purified chromatin). Results are the mean of two biological replicates (error bars indicate s.e.m.).
- **Antibodies**

All antibodies used for this Ph.D. project are listed in the Table 2.

Table 2

<table>
<thead>
<tr>
<th>Group</th>
<th>Antibody</th>
<th>Code</th>
<th>Secondary detecting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>Anti-Histone H3 (acetyl K27)</td>
<td>ab4729</td>
<td>VeriBlot (ab131366)</td>
</tr>
<tr>
<td>Primary</td>
<td>NF-YA</td>
<td>sc-17753</td>
<td>Mouse monoclonal</td>
</tr>
<tr>
<td>Primary</td>
<td>HA</td>
<td>12CA5</td>
<td>Mouse monoclonal</td>
</tr>
<tr>
<td>Primary</td>
<td>Lamin B</td>
<td>sc-6216</td>
<td>Goat polyclonal</td>
</tr>
<tr>
<td>Primary</td>
<td>B-actin</td>
<td>A5316</td>
<td>Mouse monoclonal</td>
</tr>
<tr>
<td>Primary</td>
<td>IgG</td>
<td>sc-2027</td>
<td>VeriBlot (ab131366)</td>
</tr>
<tr>
<td>Secondary</td>
<td>mouse IgG HRP linked</td>
<td>NA931V</td>
<td>--</td>
</tr>
<tr>
<td>Secondary</td>
<td>rabbit IgG HRP linked</td>
<td>NA934V</td>
<td>--</td>
</tr>
<tr>
<td>Secondary</td>
<td>goat IgG HRP linked</td>
<td>SC-2020</td>
<td>--</td>
</tr>
</tbody>
</table>

Table 2. **Antibody list used in this work.** The table is divided in Group, Antibody, Code and Secondary detecting.
3. Bioinformatics analyses

- **RNA-seq**

Two biological replicates were sequenced for each sample. For each replicate, we have used about $2.0 \times 10^7$, each read about 150bp long. Reads obtained were mapped to the mouse genome (mm9) using TopHAt2 (Trapnell C et al., 2009). All other parameters were used as default. The reference annotation, Mus_musculus.NCBIM37.67.gtf, was downloaded from ensemble database (http://www.ensemble.org). Gene expression levels were estimated for each sample in term of Fragment Per Kilo base of exon model per Million mapped reads (FPKM) using Cufflinks (Trapnell C et al., 2012). I selected only protein coding genes, all other proteins were masked from our analysis and for each gene we tested the significance of 95% confidence interval. For counting reads, I have used HT-seq (Anders S et al., 2015) for gene count matrix and in order to pre-process RNA-seq data for differential expression analysis by counting the overlap of reads with genes. Then from this point I moved for further analysis with a graphical user interface (GUI) for the identification of differentially expressed genes across multiple biological condition, RNA-seqGUI (Russo F et al., 2014). This R package includes some well-known tools always used in the RNA-seq pipeline. All counted reads were normalized and for all data were plotted HeatMap Profiles and Principal Component Analysis (PCA) to establish the reproducibility in all experiments. At least, differentially expressed (DE) genes were estimated using DeSeq2 (Love MI et al., 2014) using default parameters. Gene with adjusted P-values < 0.05 were considered DE.
- **ATAC-seq**

Although ATAC-seq is being adopted in many laboratories because is fast and sensitive, in literature there are no available pipelines. A new pipeline was developed in the lab in order to analyse these kind of data. First of all, each sample was sequenced in paired-end, 60 bp reads long. Quality check control on the raw sequencing data was made using FastaQC with different parameters: 1) Per base quality means how is it correct the position of each nucleotide in a read. This parameter is measured through a scale (0 is worst, 36 is excellent); 2) GC content means how results, in the GC context, overlap with hypothetical distribution; 3) Per base N content means how many undefined nucleotides are calling; 4) Duplication levels means how many redundant reads are generated by PCR; 5) Adapter content means which is the percentage of adapters in selected reads.

Reads were aligned using Bowtie2 (Langmead B et al., 2012). Bowtie2 is a tool for aligning sequencing reads to long reference sequences and it is used for reads of about 50 up to 100s or 1000s of characters. Analysis were performed using bash command from LINUX operative system and the output file was in Sam format.

The Sam file stands for Sequence Alignment/Map format and it is a Tab-delimited text format divided into two parts: the first one includes header and second one which has details about alignment, for example mapping position. A detailed example of Sam field output alignment is illustrated in the figure 9 (Li H et al., 2009).

**Fig.9**

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>QNAME</td>
<td>Query NAME of the read or the read pair</td>
</tr>
<tr>
<td>2</td>
<td>FLAG</td>
<td>Bitwise FLAG (pairing, strand, mate strand, etc.)</td>
</tr>
<tr>
<td>3</td>
<td>RNAME</td>
<td>Reference sequence NAME</td>
</tr>
<tr>
<td>4</td>
<td>POS</td>
<td>1-Based leftmost POSition of clipped alignment</td>
</tr>
<tr>
<td>5</td>
<td>MAPQ</td>
<td>MAPping Quality (Phred-scaled)</td>
</tr>
<tr>
<td>6</td>
<td>CIGAR</td>
<td>Extended CIGAR string (operations: MIDNESH)</td>
</tr>
<tr>
<td>7</td>
<td>MRNM</td>
<td>Mate Reference NaMe (&quot;=&quot; if same as RNAME)</td>
</tr>
<tr>
<td>8</td>
<td>MPOS</td>
<td>1-Based leftmost Mate POSition</td>
</tr>
<tr>
<td>9</td>
<td>ISIZE</td>
<td>Inferred Insert SIZE</td>
</tr>
<tr>
<td>10</td>
<td>SEQ</td>
<td>Query SEQuence on the same strand as the reference</td>
</tr>
<tr>
<td>11</td>
<td>QUAL</td>
<td>Query QUALity (ASCII-33=Phred base quality)</td>
</tr>
</tbody>
</table>

Figure 9. **Sam output structure.** Detailed field in Sam file are QNAME, FLAG, RNAME, POS, MAPQ, CIGAR, MRNM, MPOS, ISIZE, SEQ, QUAL (Li H et al., 2009).
FASTQC report, for all samples in our sequencing has found a short contamination around 33-35 bp to 47 bp of Nextera transposase sequence (black curve), figure 10.

**Fig. 10**

![Figure 10. Adapter contamination in raw reads.](image)

On X-axis there is the position (bp), on Y-axis the reads percentage. It is clearly visible Nextera Transposase Sequence contamination between 33-35 and 47 bp (black curve).

Nextera Transposase Sequence were removed in order to increase the alignment rate output. The “contaminations” sequences were CTGTCTCTTATACACATCTCCGAGCCCACGAGAC which is the reverse complement of the Nextera transposase sequence attached to read 2 and CTGTCTCTTATACACATCTGACGCTGCCGACGA which is the reverse complement of the Nextera transposase sequence attached to read 1 (Turner FS et al., 2014). Cutadapt was used to remove the Nextera transposase sequences. This algorithm is largely used in bioinformatics analysis, it searches and cut off specific sequence independent from its localization (Martin M 2011). Reads without Nextera transposase sequences are illustrated in the figure 11.
Once files in Sam format with uncontaminated reads were obtained, they were converted into Bam files and then in Bed files. Bam file is the compressed binary version of a Sam file that is used to represent aligned sequences. It is divided into Header which contains general information: sample name and length, the alignment method, start/end position, alignment quality and the match descriptor string. Furthermore, the alignment section includes the following information: RG (number of reads for specific sample), BC (indicate demultiplexed sample ID associated with the read), SM (single-end alignment quality), AS (paired end alignment quality), NM (Edit distance tag), XN (Amplicon name tag) (Li H. et al., 2009).

Bed format stand for Browser Extensible Data and it can be defined as a flexible way to represent genomic coordinates in a very sample manner. In Bed file there are different parameters: chromosome (the name of the chromosome on which the genomic feature exists), start (the zero-based starting position of the features in the chromosome), end (the one-based ending position of the feature in the chromosome). Then there are other optional columns, for example: name (the name of the Bed feature), score (Bed score range between 0 to 1000), strand (defines the strand “+” or “−”) (Quinlan AR et al., 2010). From file Bed, R1 and R2 reads (sequencing was made in paired-end mode) were isolated using “grep” as bash command and then sorted using bedtools. Starting from Bed files, it was calculated files which contain all genomic coordinates without PCR duplicates. Last problem in ATAC-seq technique was linked to mitochondrial reads (M) that can strongly reduce number of reads. There is one literature study which explain
a possible method to reduce mitochondrial reads using CRISPR-CAS9. Researchers, through CRISPR, were able to reduce mitochondrial noise from samples by 1.5 to 3 fold (Montefiori L et al., 2017). Mitochondrial reads were counted and removed from files and used for genome coverage and at least they were uploaded on UCSC Genome Browser. The last bioinformatics analysis was peak calling, which quantify how many open regions (peaks) were there in a specific sample. The algorithm used is MACS2 (Model-Based Analysis of ChIP-seq), has been optimized for ChIP-seq and DNase-seq data but it is largely used in the ATAC-seq work (Feng J et al., 2012). This tool identifies statistically enriched genomic regions. By default, it is able to calculate also PCR duplicates by removing them and calculates also a p-value for each peak using a dynamic Poisson distribution to capture local bias in read background levels. In ATAC-seq, due to absence of input sample, MACS2 measures the total genome-wide coverage background in order to estimate enriched accessible regions. The bash command used was: /share/apps/MACS2-2.1.1/bin/macs2 callpeak -t file_input.bed -f BED -g mm -n file_output.bed -nomodel --shif100 --extsize 200. /share/apps/MACS2-2.1.1/bin/ represents the directory path in which MACS is located; macs2 was the version used; -t recalls the input file; -f specifies the file input type; -g is the animal genome (mm9 in our analyses); shift and extsize represent some advanced parameters which consider the Tn5 transposase cut site.
• **Real-time PCR, statistical analysis**

Real-time polymerase chain reaction, also known as quantitative real time polymerase chain reaction (qPCR) or kinetic polymerase chain reaction, is a molecular technique based on PCR, which is used to amplify and quantify a targeted DNA molecule. It enables both detection and quantification (as absolute number of copies or relative amount when normalized to DNA input or additional normalizing genes) of a specific sequence in a DNA sample. RNA was extracted from cells using Trizol reagent (explained in more details in the session “RNA-seq”). Before reverse transcription, RNA samples were treated with DNase I. Contamination with genomic DNA (located in the interphase during trizol extraction) was identified by including “no RT/RT-controls”. cDNA was retro-transcribed from 1 to 2 µg of RNA. All samples were run in duplicates in 15 µL reaction volume. The run used was similar to PCR default condition but the number of cycles is increases up to 45 cycles. The cycle threshold (Ct) was determined during geometric phase of the PCR amplification plots, as illustrated in the manufacturer. Relative differences in transcript levels were quantified using the ΔΔCt method and normalized to Gapdh and Rpl13a expression references. The Tbx1 primers were:

Forward: 5'-CTGACCAATAACCTGCTGGATGA-3'
Reverse: 5'-GGCTGATATCTGtGCATGGAGT-3'

For time-course experiments with loss of function (LoF) and gain of function (GoF) they were set different statistical parameter in order to understand if there are chromatin accessible variations between control and treated samples on some specific loci (Q-ATAC). It was calculated 2^-ΔΔCt by considering the differences between Ct of sample and Ct of endogenous control. It was added, as internal control, the DNA extracted from the same undifferentiated cell lines (P19Cl6) in order to minimize the differences in primers annealing efficiency. Biological duplicates were considered and for each sample it was measured the standard deviations and error standard in order to assess whether each locus has or not variations in chromatin accessibility between samples. Negative control was used to establish the threshold of opened/closed chromatin. Primers were designed in a region that does not contain genes for about 80 kb. The Ct signal of this genomic window was set as the threshold: trend above threshold has indicated as “open chromatin”; in contrast, trend below threshold has indicated as “closed chromatin”.


RESULTS

Chapter 1

TBX1 controls histone acetylation at the Mef2c-Anterior Heart Field (AHF) locus.

TBX1, in P19Cl6 cells, tends to coincide with H3K27Ac-poor regions (Fulcoli FG et al., 2016). In order to explain why TBX1 tends to bind hypo-acetylated regions, the mechanism remains unknown. TBX1 binds to and represses Mef2c expression (Pane LS et al., 2012), therefore, we used this gene as a model to understand mechanisms of gene regulation. TBX1 binds the so called “anterior heart field” (AHF) enhancer of Mef2c gene. To understand if this enhancer could be differentially acetylated following by Tbx1 dosage variation, we performed a Chromatin Immunoprecipitation (ChIP) with anti H3K27Ac antibody on undifferentiated C2C12 cells (see Materials and Methods paragraph for further informations) with and without Tbx1 knock down using transfection with siRNA pool (Fig. 12).

Fig. 12

![Figure 12](image)

I have chosen the first and second replicate because the transfection was better than third biological replicate and then it was performed ChIP with H3K27Ac followed by real-time PCR on AHF enhancer. I found that C2C12 with low level of Tbx1 have a significant increase of H3K27Ac enrichment at the Mef2c-AHF enhancer compared to Control (Fig. 13) suggesting that TBX1 is able to maintain H3K27Ac enrichment low. The H3K27Ac
enrichment of *Mef2c-AHF* has been calculated on two biological duplicates and immunoprecipitation (expressed in percentage) was based on 1% DNA input from the same cell line. Colleagues in the lab have tested whether TBX1 interacts with HDAC1 or HDAC2, but with negative results. Thus, decreased dosage of TBX1 is associated with increased H3K27Ac, but the mechanism by which low acetylation is maintained by TBX1 is probably indirect. These results have been included in a publication (Pane LS et al., 2018). Future experiments will explore the relationship between TBX1 dosage and H3K27Ac genome-wide using ChIP-seq.

Fig. 13

Figure 13. **TBX1 and H3K27Ac enrichment at *Mef2c-AHF* enhancer in C2C12 cells.** Histogram showing the results of Q-ChIP analyses using anti H3K27Ac antibodies on C2C12 cells treated with non-targeting siRNA (Control) or with Tbx1-targeted siRNA (*Tbx1-KD*) on *Mef2c-AHF* locus. Enrichment is shown as percentage of input. Results are the mean of two biological replicates (error bars indicate s.e.m.).
Chapter 2

*Generation of chromatin accessibility maps of P19Cl6 cells with and without Tbx1.*

To generate maps of chromatin accessibility, we started with P19Cl6 cells. The choice of P19Cl6 model system was due to the availability of TBX1 binding sites map previously generated on the same model. Starting from published TBX1 ChIP-seq data (Fulcoli FG et al., 2016) and then compared them to new ATAC-seq data, we have tried to comprehend how chromatin variations change where TBX1 binds chromatin. Principal Component Analysis (PCA) of P19Cl6 transcriptional profile and differentiated mouse embryonic stem cells (mESCs) established that P19Cl6 differentiated at Day 1, with highest Tbx1 expression, have a transcriptional profile intermediate between ESC and mesodermal differentiation states (Fulcoli FG et al., 2016) therefore this time-point mimics what happen in mouse model during Tbx1 expression. I started experiments through manipulation and differentiation of P19Cl6 cells by culturing cells and after one passage, cells were transfected using a pool of Tbx1 small interference RNA (siRNA). About 5.0 x 10^5 cells were seeded in the 35mm dishes and it was used 25 pmol of siRNA conjugated with Lipofectamine (RNAmax). 18hr after transfection, cells were induced with 5-Aza (5-Aza, is an analogue of cytidine that cannot be methylated). Once added 5-Aza to cell culture medium, cells start to differentiate in the cardiac cells lineage. Under 5-Aza addition, P19Cl6 showed the highest Tbx1 expression. After 24hr from 5-Aza induction, cells were collected and processed for analysis. The protocol of P19Cl6 differentiation is illustrated in the figure 14.

Fig. 14

![Figure 14. P19Cl6 cells differentiation until D1. The differentiation scheme starts from transfection point (siRNA). 18 hours after transfection, cells were induced with 5-Aza and collected 24 hours later at D1.](image)

Cells were collected (two biological duplicates for each conditions: Control and Tbx1-KD) exactly after 24hr from 5-Aza induction and 50.000 cells were processed for ATAC while all the others were used for RNA extraction. The
DNA tagmented with Tn5 enzyme was stored at -20 °C until transfection evaluation. The RNA was extracted and checked for quality was verified using a 1.5% agarose gel. About 1.5 microgram of RNA was uploaded on gel. The results illustrated in the figure 15 suggested that quality was good because bands of 28S, 18S and 5S were visible.

Fig. 15

![Agarose gel image](image_url)

**Control  Tbx1-KD**

Figure 15. **P19Cl6 RNA quality.** RNA extracted from P19Cl6 was uploaded on 1.5% agarose gel. Bands of 28S, 18S and 5S were detectable.

Once assessed the quality and quantity with NanoDrop fluorimeter of RNA, 1µg was retrotranscribed into cDNA and *Tbx1* expression levels were measured using real-time PCR using technical triplicates (Fig. 16). As normalizer we used *Gapdh* gene expression.
Figure 16. **Tbx1-KD in P19Cl6 cells.** *Tbx1* was reduced of 88% in KD sample compared to Control. All samples were in technical triplicates. RQ expresses the normalized expression value.

I obtained a reduction of *Tbx1* expression level of about 90% compared to Control. After KD evaluation, tagmented samples, previously stored at -20 °C, were used for PCR amplification. Amplified DNA was then purified, the fragment size and concentration were estimated using a Tapestation instrument which is able to quantify the double strand DNA concentration and size distribution during electrophoresis assay. The figure 17 shows details of the tapestation output. On the X-axis is indicated the sample fragment size while on Y-axis there is a relative measure of DNA quantity.

**Fig. 17**

Control-1
**Figure 17.** **DNA concentration and size distribution in P19Cl6 cells.** On the X-axis is indicated the sample fragment size while on Y-axis there is a relative measure of DNA quantity. Electrophoresis assay were estimated for Control-1, Control-2, *Tbx1*-KD-1 and *Tbx1*-KD-2.
The two biological replicates for both conditions were sequenced based on fragment distribution reported in figure 17. The sequencing was done using NextSeq 500, based on Illumina protocol, 60bp each reads in paired-end, MID flowcell, 240.000.000 total reads. The ATAC-seq pipeline has been described in the Materials and Methods section. I describe some additional details concerning the P19Cl6 experiment. First of all, I have examined the number of raw reads from each biological replicate. All 4 samples (Control and Tbx1-KD in biological duplicates) were sequenced in a single flow cell and the number of reads for each sample was: 25373112 for Control (first replicate), 63880623 for Control (second replicate), 33998199 for Tbx1-KD (first replicate) and 65812087 for Tbx1-KD (second replicate). Before mapping with reference genome, I evaluated the sequencing quality of all samples. The first replicate for both conditions was bad because more than half reads were PCR duplication artefacts. However, I decided to continue analyses. The second problem was removing the Nextera transposase sequence because the presence of adapters contamination could alter alignment between reads and reference genome. We aligned before and after Nextera transposase sequence and we found an improvement of alignment rate in the reads without Nextera transposase sequence (Table 3).

<table>
<thead>
<tr>
<th>Sample</th>
<th>Reads with Nextera transposase sequence</th>
<th>Reads without Nextera transposase sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control-1</td>
<td>71.68%</td>
<td>89.59%</td>
</tr>
<tr>
<td>Tbx1-KD-1</td>
<td>44.24%</td>
<td>49.52%</td>
</tr>
<tr>
<td>Control-2</td>
<td>84.29%</td>
<td>96.26%</td>
</tr>
<tr>
<td>Tbx1-KD-2</td>
<td>41.62%</td>
<td>44.72%</td>
</tr>
</tbody>
</table>

Table 3. Alignment rate (%) before and after removing of Nextera transposase sequence in P19Cl6 cells. The table is divided in Sample, Reads with Nextera transposase Sequence and Reads without Nextera transposase Sequence).

By removing Nextera transposase Sequence, I found an alignment improvement of 17.88% for Control and 11.97% for Tbx1-KD for the first replicate. For second replicate there was an improvement of 5.28% for Control and 3.1% for Tbx1-KD. Then I moved to mitochondrial reads (M) calculation and removing them from samples. The table 4 illustrates for each sample how many reads were found and the effective reads used for genome coverage and peak calling (Reads after M depletion).
Table 4

<table>
<thead>
<tr>
<th>Sample</th>
<th>Reads before M depletion</th>
<th>Mitochondrial Reads</th>
<th>Reads after M depletion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control-1</td>
<td>3289151</td>
<td>56757</td>
<td>3226394</td>
</tr>
<tr>
<td>Tbx1-KD-1</td>
<td>16022976</td>
<td>1040265</td>
<td>14982711</td>
</tr>
<tr>
<td>Control -2</td>
<td>12394516</td>
<td>959542</td>
<td>11434971</td>
</tr>
<tr>
<td>Tbx1-KD-2</td>
<td>12299146</td>
<td>1090975</td>
<td>11208171</td>
</tr>
</tbody>
</table>

Table 4. **Comparison between the number of total reads with and without Mitochondrial reads in P19C16 cells.** The table is divided in Sample, Reads before M depletion, Mitochondrial Reads and Reads after M depletion.

As demonstrated in the table 4, the mitochondrial reads were less than 10% of the total, so I decided to continue removing all mitochondrial reads and, once obtained the final reads we had called the number of peaks which refer to effective number of accessible regions (Table 5).

Table 5

<table>
<thead>
<tr>
<th>Sample</th>
<th>Peaks number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control-1</td>
<td>3113</td>
</tr>
<tr>
<td>Tbx1-KD-1</td>
<td>26833</td>
</tr>
<tr>
<td>Control -2</td>
<td>36615</td>
</tr>
<tr>
<td>Tbx1-KD-2</td>
<td>50920</td>
</tr>
</tbody>
</table>

Table 5. **Total Peaks number in P19C16 cells.** The table contain Sample and Peaks number (the overall chromatin accessible regions).

Because of the large discrepancy between two different replicates, I decided to eliminate samples with low peak numbers. Peaks annotation and comparison with previously published ChIP-seq and RNA-seq data on the same cell line were performed using only one replicate for each condition: Control-2 and Tbx1-KD-2. Overall, I found 36615 chromatin accessible peaks in Control and 50916 accessible regions in Tbx1-KD cells. We evaluated the distribution of peaks, using ChIP-seeker (Yu G et al., 2015) relative to gene features defined as Promoter 1 to 3 kb (from 1 kb to 3kb from the transcription start site), 5’-UTR, 3’-UTR, First Exon, First Intron, Other Exon, Other Intron, Downstream, Distal Intergenic. In control sample 33.45% of regions were localized around promoter, 49.61% were intragenic and 16.94% were distal intergenic while in the KD sample 31.48% of regions were localized around promoter, 49.14% were intragenic and 19.38% were distal intergenic. In both cases, peaks were distributed mostly at the promoter...
regions of the genes and in the intragenic regions, as expected. Overall distribution suggested no differences in chromatin accessibility regions between the two different conditions. This first analysis revealed that technique was working well because a good percentage of chromatin accessibility were located around 3000 bp of TSS of genes. We have plotted two different graphs: Annotation Pie and Heatmap Profile (Fig. 18).

Fig. 18

![Annotation Pie and Heatmap Profile](image)

Figure 18. **Total accessible regions distribution in Control and Tbx1-KD P19Cl6 cells.** Total accessible regions of Control (left), KD (right). On the top: Annotation pie shows accessible regions distribution around gene features. On the bottom: Heatmap shows accessible regions distribution around TSS (from 1kb to 3kb).
Intersection of ATAC-seq and ChIP-seq data: many TBX1 binding sites are located within closed chromatin.

Overall, considering the all peaks there were no obvious differences between Control and Tbx1-KD sample. Next question that we want to answer was: Does Tbx1 bind opened or closed chromatin? In order to answer this question, I compared ATAC peaks with TBX1 binding sites obtained with the same cell line under the same conditions (Fulcoli FG et al., 2016). Specifically, I considered the highly statistically significant 2388 sites (“golden peaks”). Results indicate that TBX1 peaks are mostly located in ATAC-negative regions (closed chromatin), as only 335 of them (14%) overlap with ATAC peaks (Fig. 19). In addition, Tbx1-KD did not affect significantly chromatin accessibility in TBX1 binding sites: 450 regions (18.8% of TBX1-binding sites) were localized in open chromatin (Fig. 19). After KD, 28 TBX1 binding sites (1.4%) lose accessibility while 143 (6%) gain accessibility.

Fig. 19

To confirm these results, I repeated the same comparison analysis between chromatin accessibility and TBX1-binding sites without using chromatin regions as “peaks” but by using the genome coverage for both conditions (Quinlan AR et al., 2010)
I have plotted the genome coverage distribution of accessible regions in Control and KD around the TBX1 binding sites. The figure 20 shows that there is no overlap or vicinity between TBX1 binding sites and chromatin accessible regions.

Figure 20. Intersection between Control and *Tbx1*-KD accessible regions and TBX1-binding regions based on genome coverage in P19Cl6 cells. Coverage distribution of accessible regions in Control (violet) and *Tbx1*-KD (light blue) around the TBX1 binding sites. The figure shows that there is no overlap or vicinity between TBX1 binding sites and accessible regions. Y-axis (Read count Per Million mapped reads), X-axis (Genomic Regions).
Many genomic regions which are not bound by TBX1 are affected by \textit{Tbx1} loss of function. Although I found no effect of loss of \textit{Tbx1} at \textit{TBX1}-binding regions, I noted that the number of peaks in the \textit{Tbx1}-KD condition are higher than in the Control sample. Therefore, I performed a search of differentially accessible regions (DARs) using a method that takes into account the differences in the number of reads in different sample by performing normalization which, in ATAC experiments, because of the lack of input sample, taking into account a large window of background genome-wide signal. DARs between Control and \textit{Tbx1}-KD conditions was carried out using a tool: Spatial Clustering for Identification of ChIP-Enriched Regions (SICER) (Shiliyang X et al., 2014). SICER-df was used in order to evaluate common regions (Control and \textit{Tbx1}-KD) which increase/decrease accessibility in KD cells and it works by calculating how many reads mapped within specific regions. Using this algorithm, I found 2401 regions that increase accessibility in KD cells, while I found only 16 regions that decrease accessibility. About 72\% of these DARs were located in promoter regions (1 kb from the transcription start site). In contrast, only 27.7\% of the non DARs regions were located in promoters. In addition, only 1.4\% (34) of TBX1 binding sites overlap with DARs suggesting that TBX1 might not remodel these 2401 regions through DNA binding (Fig. 21).

Figure 21. \textbf{Intersection between differential accessible regions (DARs)/ Total Regions and TBX1-binding sites in P19Cl6 cells.} On the left: intersection between Total Regions (Ctrl) localized for about 30\% around promoter and TBX1-binding regions (only 14\%). On the right: intersection between DARs localized for about 72\% around promoter and TBX1-binding regions (only 1.4\%, less than Total Regions).

Next, I asked whether DARs are characterized by the presence of specific transcription factors binding motifs. To this end, I have used Hypergeometric Optimization of Motif EnRichment (HOMER), a tool for motifs discovery (Heinz S et al., 2010). Results showed that there is no enrichment of T-box
binding motifs, confirming that does not bind these regions (Fig. 22). Instead, I found enrichment for other transcription factors listed in the figure 22.

**Fig. 22**

<table>
<thead>
<tr>
<th>Known Motif</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
<td>Motif</td>
<td>Transcription factor (TF)</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>NFY(CCAAT)</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>KLF9</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>SP1</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>SP5</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>KLF3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>De novo Motif</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
<td>Motif</td>
<td>P-value</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>$10^{-23}$</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>$10^{-22}$</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>$10^{-22}$</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>$10^{-22}$</td>
</tr>
</tbody>
</table>

Figure 22. **Motifs discovery in differential accessible regions (DARs) which increase accessibility in P19Cl6 Tbx1-KD cells.** Known motifs (top) and *De novo* motifs (bottom) found no T-box sequences. The top scorer is NFY for both motifs categories. Figure is subdivided for Rank, Motif, P-value, Transcription factor (TF).

Next, I asked how TBX1 may remodel these 2401 regions without binding to them. There are at least two possibilities: 1) Tbx1 may regulate transcription factor-encoding genes, and their product, in turn, may remodel chromatin; 2) Remodelling may be due to a DNA-independent effect of *Tbx1*. To address this possibility, we will test whether TBX1 immunoprecipitated with some of the transcription factors listed in the Fig. 22. I will also determine if loss of *Tbx1* increases the occupation of TFs like NF-Y to promoters. I have selected some putative responsive loci associated to DARs and I found no differences in chromatin accessibility associated to these loci following *Tbx1* overexpression (Fig. 23). In future it could be interesting to select other putative responsive loci or maybe it could be necessary another approach to address this point.
Fig. 23

Figure 23. *Putative responsive loci associated to differential accessible regions (DARs) which increase accessibility in P19Cl6 Tbx1-KD cells.* Enrichment of chromatin accessibility associated to *Rtn3, Atp5e, Eef2k, Sesn2* found no differences following *Tbx1* overexpression. On the Y-axis there is a normalized expression value (2^ΔCt), on the X-axis there is EV (Empty Vector), WT (*Tbx1* exogenous copy) and MUT (*Tbx1* exogenous copy unable to bind DNA). *D1* refers to differentiation point while *Neg Ctrl* refers to the threshold of closed chromatin: primers were built in a region that does not contain any genes in a range of about 80 kb.
Some loci show delayed response to loss of Tbx1 dosage in time-course experiments.

The lack of a strong effect of TBX1 on chromatin remodelling is surprising, given the interactions with chromatin remodelling proteins reviewed in the introduction. Therefore, I have asked the question: Why TBX1 does not induce significant chromatin remodelling at its binding regions? I have considered two hypotheses to explain my results: 1) P19Cl6 cells may lack required cofactors that could help TBX1 in chromatin remodeling. Therefore, I have tested other model systems (see Chapter 3). 2) The chromatin response may not be evident at this time point of differentiation but may appear at a later stage because cofactors may become biologically available later. To address this second point, I have performed time-course experiment using Q-ATAC instead of ATAC-seq, and tested chromatin access at specific loci. I have collected cells at three different time points during the differentiation protocol, T1 (13 hours after siRNA transfection), at D1 (24 hours after 5-Aza induction) and at D2 (24 hours after DMSO addition) and for each point two biological replicates were considered (Fig. 24).

Fig. 24
Transfection

Figure 24. P19Cl6 cells differentiation until D2. The differentiation scheme starts from transfection point. 13 hours after transfection, cells were collected at T1. After 5 hours, cells were induced with 5-Aza and collected 24 hours later at D1. DMSO was added and after 24 hours cells were collected (D2) (Fulcoli FG et al., 2016).

P19Cl6 cells were treated with siRNA pools (targeted to Tbx1 or non-targeted control) After 13 hours (hr) the first sample was collected T1, then after 5hr 5-Aza was added to the remaining cells and other sample was collected at after 24hr D1, and DMSO was added to the remaining cells. The third and last sample was collected after 24hr from DMSO exposure (D2) (Fulcoli FG et al., 2016). As we expected, Tbx1 was knocked down at all three time points, compared to controls (Fig. 25).
Figure 25. *Tbx1*-KD in P19Cl6 cells during differentiation. *Tbx1* was reduced in KD sample (blue) compared to Control (orange) at all three time points. RQ expresses the normalized expression value.

Once verified the *Tbx1* expression levels in Control and KD conditions, tagmented chromatin for both replicates on 50000 cells were amplified, purified and then used for quantitative ATAC (Q-ATAC). For chromatin enrichment analyses were chosen loci bound by TBX1 that contain accessible regions. The loci selected were: *Bai2* (encoding brain inhibitor of angiogenesis, a GPCR receptor of secretin), *Cdc42bpg* (protein tyrosine kinase), *Brd4* (encoding serine/threonine kinase involved in chromatin remodelling), *Pxn* (encoding protein involved in focal adhesion), *Dusp7* (encoding phosphatases of threonine, tyrosine and serine). For real-time PCR, we have used both biological duplicates for each time point and each duplicate was divided in two technical replicates. We used two different controls: *Gapdh* promoter (positive control) representing the open chromatin, and a desert island locus (negative control) which does not contain any genes in a range of about 80 kb. The figure 26 illustrated the time-course enrichment ($2^{- \Delta Ct}$). The dotted lines indicate the values of negative control regions Control (orange) and KD (blue) samples.
Figure 26. **Putative responsive loci to Tbx1-KD in P19C16 cells.** Enrichment of chromatin accessibility associated to Bai2, Cdc42bpg, Brd4, Pxn, Dusp7 found that the first three loci became accessible at D2, after 66 hours from Tbx1-KD. On the Y-axis there is a normalized expression value ($2^{-\Delta Ct}$), on the X-axis there are three different time points (T1, D1, D2). Colours legend are illustrated in the figure.

Results showed that 3 out 5 loci tested (Bai2, Cdc42bpg and Brd4) became accessible after 66 hours from Tbx1-KD suggesting a possible role in enhancer priming (Wang C et al., 2016). Next, I have tested the hypothesis that overexpression of Tbx1 may be associated with chromatin changes by displacing the nucleosome core (Luger K et al., 2012). Thus, I have designed a gain of function experiment to test chromatin response. The time-course gain of function experiment was performed under the same conditions described for the loss of function experiment but, I transfected cells with an expressing vector containing a Tbx1-3xHA cDNA or with a control, empty vector. Chromatin from harvested cells was tagmented and stored while I checked the transfection efficiency, as normalizer was used beta actin (Fig. 27).
Fig. 27

Figure 27. **Tbx1-overexpression in P19Cl6 cells during differentiation.** Tbx1 was overexpressed in Tbx1-3xHA at T1, D1, D2 (detected with Anti-HA tag). Empty Vector was used as negative control. Beta actin was used as normalizer.

Subsequently, I have carried out Q-ATAC for 5 selected loci (Fig. 28)

Fig. 28

Figure 28. **Putative responsive loci associated to Tbx1 overexpression in P19Cl6 cells.** Enrichment of chromatin accessibility associated to Bai2, Cdc42bpg, Brd4, Pxn, Dusp7 found no differences following Tbx1 overexpression. On the Y-axis there is a normalized expression value (2^{ΔCt}), on the X-axis there is T1, D1 and D2. Colours legend are illustrated in the figure.

Results showed that none of the loci tested is affected by increased dosage of Tbx1. In the future, it would be of interest to perform ATAC-seq experiments, rather than Q-ATAC to obtain a global view of chromatin accessibility.
Chapter 3

mESC differentiation and Tbx1 expression.

I wanted to confirm the data obtained in P19Cl6 cells a using different model system. We selected murine embryonic stem cells (mESCs) because of their potential differentiation skills. These undifferentiated cells can be induced, using specific factors, to differentiate into cardiomyocytes (Keller G et al., 2005). This protocol (described in details in the Materials and Methods section) can be divided in three parts: the first one is cell adhesion and proliferation, the second one is mesoderm induction and the last one is cardiac specification and maturation.

We have analysed the expression markers involved in cardiomyocytes fate and differentiation and we found that \( cTnt2 \) was progressively expressed during differentiation, starting at day 4 and increasing up until day 10. \( Tbx1 \) started to be expressed at day 4 (highest \( Tbx1 \) expression during differentiation) and was also expressed at day 10 (Fig. 29).

I have performed RNA-seq and ATAC-seq at day 2 and day 4. Results confirmed that \( Tbx1 \) is activated at day 4 but it is expressed at low level in this system. The data also showed that between day 2 and day 4 most genes encoding cardiogenic transcription factors are activated at day 4 (Table 6).

Figure 29. \( Tbx1 \) and \( cTnt2 \) expression during murine embryonic stem cells (mESCs) differentiation. On the left: \( Tbx1 \) (239bp) expression at d0, d2, d4 (expressed), d6, d10 (low expressed). On the middle: \( cTnt2 \) (102bp) expression at d0, d2, d4 (low expressed), d6 (expressed) d10 (high expressed). On the right: \( Gapdh \) (106bp) during the same differentiation used as normalizer.
Table 6

<table>
<thead>
<tr>
<th>Gene</th>
<th>d2_FPKM</th>
<th>d4_FPKM</th>
<th>d2 -&gt; d4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mesp1</td>
<td>0.13</td>
<td>1.23</td>
<td>+</td>
</tr>
<tr>
<td>T/Brach</td>
<td>0.37</td>
<td>10.85</td>
<td>+</td>
</tr>
<tr>
<td>Smarcd1</td>
<td>75.95</td>
<td>37.61</td>
<td>-</td>
</tr>
<tr>
<td>Kdr</td>
<td>0.10</td>
<td>74.82</td>
<td>+</td>
</tr>
<tr>
<td>Gata4</td>
<td>0.13</td>
<td>43.78</td>
<td>+</td>
</tr>
<tr>
<td>Isl1</td>
<td>0.27</td>
<td>4.04</td>
<td>+</td>
</tr>
<tr>
<td>Nkx2-5</td>
<td>0.05</td>
<td>0.22</td>
<td>+</td>
</tr>
<tr>
<td>Tbx1</td>
<td>0.00</td>
<td>0.40</td>
<td>+</td>
</tr>
<tr>
<td>Pecam1</td>
<td>0.71</td>
<td>24.54</td>
<td>+</td>
</tr>
<tr>
<td>Smarcd3</td>
<td>1.41</td>
<td>14.01</td>
<td>+</td>
</tr>
<tr>
<td>Tnnt2</td>
<td>0.02</td>
<td>4.94</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 6. **Genes involved in cardiopharyngeal mesoderm during murine embryonic stem cells (mESCs) differentiation.** The table is divided in Gene, d2_FPKM (the FPKM average of two biological duplicates), d4_FPKM (the FPKM average of two biological duplicates), d2 -> d4 (gene expression variations during d2-d4 differentiation; + refers to gene whose expression increases at day 4 while - refers to gene whose expression decreases at day 4).

ATAC-seq signal between day 2 (d2) and day 4 (d4) has identified genes involved in cardiac differentiation: **Smarcd3** essential for function of BAF chromatin remodelling complexes in heart development (Lickert H et al., 2004), **Gata4** important for cardiomyocytes stem cells differentiation and **Mesp1**, a key regulator of cardiovascular lineage commitment (Bondue A et al., 2010) (Fig. 30).
Fig. 30

Smarcd3

[Graph showing coverage and other details]

Gata4

[Graph showing coverage and other details]
Chromatin accessibility variations in genes involved in cardiac differentiation. Differential Enrichment of chromatin accessibility associated to Smarcd3 (Top), Gata4 (middle), Mesp1 (bottom). On vertical axis there are the genome coverage of day 2 first replicate, day 2 second replicate, day 4 first replicate and day 4 second replicate. Red arrow indicates the open chromatin open in day 4 compared to day 2.

Some of the regions which show different chromatin accessibility at these two stages of differentiation are being validated. Although results with these experiments should allow the identification of enhancers involved in differentiation, the low expression levels of Tbx1 makes it difficult to attribute any chromatin change to Tbx1 gene activation. To enrich the cell population with Tbx1-expressing cells, the lab has carried out flow cytometry analyses followed by cell sorting, with standard surface markers: FLK1 and PDGFRa. We were able to isolate at day 4 (highest Tbx1 expression) three different subpopulations which are separated because of their surface markers. The first one was positive for PDGFRa, the second double positive for PDGFRa and FLK1 (also known as VEGFR2) and third one which was positive for FLK1. Cell populations of three subgroups were respectively 9%, 68.2% and 21 % of the total (Fig. 31).
Figure 31. FACS-sorted murine embryonic stem cells (mESCs) differentiated at d4. FACS-sorted d4 identifies three distinct subpopulations: PDGFRa+, PDGFRa+/FLK1+, FLK1+. Cell populations were respectively 9%, 68.2% and 21% of the total. RNA was extracted from sorted cells of these three populations and we tested the expression of Tbx1. We found that at day 4 the highest expression was in the PDGFRa+; FLK1- subpopulation (Fig. 32).

Figure 32. Tbx1 expression in murine embryonic stem cells (mESCs). Histogram showing Tbx1 expression (mRNA relative expression levels) in d4 total cells (pink), Pdgfra+ (red), Pdgfra+/Flk1+ (yellow), Flk1+ (blue) cells. Tbx1 is highly expressed in Pdgfra+ cells.
Therefore, we have chosen this specific subpopulation to understand how and if TBX1 is able to have an impact on chromatin accessibility. The laboratory has generated Tbx1−/− mESC using CRISPR-Cas9 gene targeting. I have used these cells to perform ATAC-seq and RNA-seq experiments along with WT cells at day 4, after sorting Pdgfra+; Flk1- cells.
**Chromatin remodelling and gene expression in differentiated mESCs WT and Tbx1<sup>-/-</sup>.**

ATAC-seq and RNA-seq experiments were done on two biological replicates of mESCs (*Tbx1* WT and KO) differentiated at day 4 and sorted for PDGFRα and FLK1. The overall number of accessible chromatin regions is illustrated in the table 7.

Table 7:

<table>
<thead>
<tr>
<th>Sample</th>
<th>Peaks number</th>
</tr>
</thead>
<tbody>
<tr>
<td>d4-PDGFRα+_{WT}-1</td>
<td>4101</td>
</tr>
<tr>
<td>d4-PDGFRα+-_{KO}-1</td>
<td>10305</td>
</tr>
<tr>
<td>d4-PDGFRα+_{WT}-2</td>
<td>8228</td>
</tr>
<tr>
<td>d4-PDGFRα+-_{KO}-2</td>
<td>4115</td>
</tr>
</tbody>
</table>

Table 7: **Total Peaks number in d4-PDGFRα+, Tbx1 WT and KO murine embryonic stem cells (mESCs).** Table is divided in Sample and Peaks number (total accessible chromatin regions).

The number of peaks in d4-PDGFRα+WT-1 and d4-PDGFRα+KO-2 was very low, so we decided to pool replicates for both conditions. Table 8 shows number of pooled peaks.

Table 8:

<table>
<thead>
<tr>
<th>Sample</th>
<th>Pooled Peaks number</th>
</tr>
</thead>
<tbody>
<tr>
<td>d4-PDGFRα+_WT_pooled</td>
<td>12109</td>
</tr>
<tr>
<td>d4-PDGFRα+_KO_pooled</td>
<td>14443</td>
</tr>
</tbody>
</table>

Table 8: **Total pooled Peaks number in d4-PDGFRα+, Tbx1 WT and KO murine embryonic stem cells (mESCs).** Table is divided in Sample and Pooled Peaks number (total accessible chromatin regions).

Next, we annotated peaks using ChIP-seeker and we found that overall peaks distribution was similar between WT and mutant cells. ATAC-seq peaks were located around the transcription start site (TSS) of genes (Fig. 33).
Figure 33. **Total accessible regions distribution in d4-PDGFRα+, Tbx1 WT and KO murine embryonic stem cells (mESC).** Heat map of chromatin accessible regions distribution around TSS (from 1kb to 3kb) of WT (Left) and KO (Right).

There is no map of TBX1 binding sites in mESCs, and there are no longer antibodies suitable for TBX1 ChIP experiments. Therefore, we cannot establish whether TBX1 remodels chromatin at its binding sites. Thus, I have limited my studies to the identification of differentially accessible regions (DARs) following by Tbx1 dosage variations and analysis of RNA-seq data. I analysed DARs which taking into account all common regions between two different conditions with a differentially enrichment. I found 48 regions that decrease accessibility in KO and 117 regions that increase accessibility in the same condition. Suddenly I noted that, also for this model system when Tbx1 was absent, the number of differential accessible regions increased and were distributed around TSS and promotor of genes as expected. Started from these 117 regions, I obtained a list of associated genes that we compared with differentially expressed genes between Tbx1 WT and KO. I generated the table 9 divided in different fields: log2FC which measures how many genes were differentially expressed between both conditions, P-value, the Gene name, the ATAC ratio which expresses the ratio between normalized reads of both conditions, and gene annotation. Two accessible regions, associated to
Myc1 and Robo2 were localized around promoters while other six regions were localized in intra/intergenic features. We are going to validate some of these genes “in vivo” using a model system in which Tbx1 expression could be monitored and studied; one is very interesting: Cyp26a1 a known Tbx1 marker (Caterino M et al., 2009). Cyp26a1, gene required for retinoic acid inactivation during embryogenesis, is a potential Tbx1 target from a microarray screen comparing wild-type and null Tbx1 mouse embryonic pharyngeal arches (PA) at E9.5 (Roberts C et al., 2006).

Table 9:

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>log2FC</th>
<th>P-value</th>
<th>Gene Name</th>
<th>ATAC ratio KO/WT</th>
<th>annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cemip</td>
<td>1.27</td>
<td>0.005</td>
<td>cell migration inducing protein, hyaluronan</td>
<td>2.46 Intron</td>
<td></td>
</tr>
<tr>
<td>Peli3</td>
<td>0.94</td>
<td>0.003</td>
<td>pellino 3</td>
<td>2.22 Exon</td>
<td></td>
</tr>
<tr>
<td>Scl6a6</td>
<td>0.27</td>
<td>0.016</td>
<td>solute carrier family 6 (neurotransmitter, taurine), member 6</td>
<td>1.94 Distal Intergenic</td>
<td></td>
</tr>
<tr>
<td>Cyp26a1</td>
<td>0.49</td>
<td>0.004</td>
<td>cytochrome P450, family 26, subfamily a, polypeptide 1</td>
<td>1.93 Distal Intergenic</td>
<td></td>
</tr>
<tr>
<td>Myc1</td>
<td>0.49</td>
<td>0.013</td>
<td>v-myc avian myelocytomatosis viral oncogene</td>
<td>1.48 Promoter</td>
<td></td>
</tr>
<tr>
<td>Robo2</td>
<td>-1.12</td>
<td>0.015</td>
<td>roundabout guidance receptor 2</td>
<td>1.81 Promoter</td>
<td></td>
</tr>
<tr>
<td>Zfp516</td>
<td>-0.58</td>
<td>1.65E-05</td>
<td>zinc finger protein 516</td>
<td>1.75 Distal Intergenic</td>
<td></td>
</tr>
<tr>
<td>Epas1</td>
<td>-1.35</td>
<td>0.009</td>
<td>endothelial PAS domain protein 1</td>
<td>1.59 Intron</td>
<td></td>
</tr>
</tbody>
</table>

Table 9: Common genes associated to differentially accessible regions (DARs) and differentially expressed (DE) genes in d4-PDGFRα+, Tbx1 WT and KO murine embryonic stem cells (mESCs). The table is divided in Gene Symbol, log2FC (measures how many genes were differentially expressed between both conditions), P-value, Gene Name, ATAC ratio KO/WT (expresses the ratio between normalized reads of both conditions), and gene annotation.

Although P19Cl6 showed that there was no chromatin accessibility where TBX1 binds chromatin, in this cells we found a list of Motifs statistically significant and most of them were bound by T-box transcription factors. (Fig. 34).
**Fig. 34**

**De novo Motif**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Motif</th>
<th>P-value</th>
<th>Transcription factor (TF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="motif1.png" alt="" /></td>
<td>$10^{-7}$</td>
<td>TBR1 (T-BOX)</td>
</tr>
<tr>
<td>2</td>
<td><img src="motif2.png" alt="" /></td>
<td>$10^{-7}$</td>
<td>TBX6 (T-BOX)</td>
</tr>
<tr>
<td>3</td>
<td><img src="motif3.png" alt="" /></td>
<td>$10^{-5}$</td>
<td>TBET (T-BOX)</td>
</tr>
<tr>
<td>4</td>
<td><img src="motif4.png" alt="" /></td>
<td>$10^{-4}$</td>
<td>EOMES (T-BOX)</td>
</tr>
<tr>
<td>5</td>
<td><img src="motif5.png" alt="" /></td>
<td>$10^{-3}$</td>
<td>NEUROD1 (bHLH)</td>
</tr>
</tbody>
</table>

Figure 34. Motifs discovery in d4-PDGFRα+ differential accessible regions (DARs) which increase accessibility in murine embryonic stem cells (mESCs) *Tbx1*-KO cells. Motifs found many T-box sequence. Figure is subdivided for Rank, Motif, Name of transcription factors and P-value.

Next I have focused on regions which gain or lose accessibility following by *Tbx1* dosage variations, they showed a distinct pattern around gene features compared to total accessible regions. The regions which gain or lose accessibility refers to regions which are unique in one specific condition different from DARs which refers to all common regions between two conditions with differentially enrichment. This analysis may take with caution because data were not normalized. Also in this context, in the accessible regions where *Tbx1* was absent I found significant T-box motif including EOMES, TBX21, TBX2, TBET, TBX20, TBX1, TBX4 (Fig. 35).

**Fig. 35**

**De novo Motif**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Motif</th>
<th>P-value</th>
<th>Transcription factor (TF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="motif6.png" alt="" /></td>
<td>$10^{-64}$</td>
<td>NEUROD1, OUS2, NEUROD2, TWIST2, BHLM1, ATG1, NEUROD2, SOX2, NEUROG1</td>
</tr>
<tr>
<td>2</td>
<td><img src="motif7.png" alt="" /></td>
<td>$10^{-57}$</td>
<td>MED-1, OUS1, NR2E3, NR8-F45, MT2, OUS2</td>
</tr>
<tr>
<td>3</td>
<td><img src="motif8.png" alt="" /></td>
<td>$10^{-57}$</td>
<td>EOMES, TBX5, TBX1, TBX2, TBET, TBX20, TBX1, TBX4, TBX21</td>
</tr>
<tr>
<td>4</td>
<td><img src="motif9.png" alt="" /></td>
<td>$10^{-56}$</td>
<td>ZIC, ZIC2, ZIC3, ZIC1, TFX2</td>
</tr>
</tbody>
</table>

Figure 35. Motifs discovery in regions which gain accessibility in d4-PDGFRα+ murine embryonic stem cells (mESCs) *Tbx1*-KO cells. Motifs analyses found a T-box sequence highly significant. Figure is subdivided for Rank, Motif, and P-value. In red arrow are highlighted T-box Transcription Factors: EOMES, TBX5, TBX2, TBET, TBX20, TBX1, TBX4, TBX21.
I uploaded on UCSC genome browser the T-box bound motifs and I noted a peak upstream \textit{Rpf1} gene (Fig. 36). This region, as an example, was opened in \textit{Tbx1} KO cells and closed in WT cells, suggesting that TBX1, or other T-box transcription factors, may limit access to the chromatin.

![Figure 36](image)

Figure 36. \textbf{T-box motif in regions which gain accessibility in d4-PDGFRa+ murine embryonic stem cells (mESCs) \textit{Tbx1}-KO cells.} Genome coverage identifies a T-box motif (Motif_Tbox), upstream to \textit{Rpf1}, in \textit{Tbx1} KO cells compared to WT cells suggesting that T-box Transcription Factor may limit access to the chromatin.

Gene expression analyses of WT vs KO samples resulted in good quality data and good reproducibility among replicates (Fig. 37).

![Figure 37](image)

Figure 37. \textbf{RNA-seq data on d4-PDGFRa+, \textit{Tbx1} WT and KO murine embryonic stem cells (mESCs).} Heat Map showing a good reproducibility between independent biological duplicates WT and KO. Colour Key indicates the gene expression, red (high), green (low).
I have determined differentially expressed (DE) genes using DeSeq2 algorithm (Love Mi et al., 2014). I found 642 DE genes of which 230 are up regulated and 412 are down regulated. DE genes were analysed for Gene Ontology (GO) terms using DAVID (Huang DW et al., 2009). Results (Table 10) showed terms “enrichment” relative to pathways in which Tbx1 has already been shown to play a role; for example, “heart development”, “heart morphogenesis” etc.

Table 10

<table>
<thead>
<tr>
<th>TERM</th>
<th>PValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0007275~multicellular organism development</td>
<td>1.87E-22</td>
</tr>
<tr>
<td>GO:0007507~heart development</td>
<td>1.00E-11</td>
</tr>
<tr>
<td>GO:0045944~positive regulation of transcription from RNA polymerase II promoter</td>
<td>1.01E-11</td>
</tr>
<tr>
<td>GO:0007411~axon guidance</td>
<td>4.40E-11</td>
</tr>
<tr>
<td>GO:0007155~cell adhesion</td>
<td>2.92E-10</td>
</tr>
<tr>
<td>GO:0001525~angiogenesis</td>
<td>4.93E-10</td>
</tr>
<tr>
<td>GO:0000122~negative regulation of transcription from RNA polymerase II promoter</td>
<td>6.07E-10</td>
</tr>
<tr>
<td>GO:0030335~positive regulation of cell migration</td>
<td>1.22E-09</td>
</tr>
<tr>
<td>GO:0001822~kidney development</td>
<td>4.21E-09</td>
</tr>
<tr>
<td>GO:0003007~heart morphogenesis</td>
<td>8.76E-09</td>
</tr>
</tbody>
</table>

Table 10: Differentially expressed (DE) genes between d4-PDGFRα+, Tbx1 WT and KO murine embryonic stem cells (mESCs). Term indicates the biological process; P-value indicates the significant TERM. In red there are some interesting biological processes: heart development, angiogenesis, heart morphogenesis.

Overall, the experiments performed in Tbx1 WT and mutant mESCs provide a good level of confidence that this is a useful model for further studies. In addition, chromatin remodelling data essentially confirm the P19Cl6 data showing that loss of Tbx1 is associated with gain of accessibility in discrete loci. However, in contrast to P19Cl6 data, mESC DARs were enriched in T-box binding sites, leaving open the possibility that TBX1 may be directly responsible for chromatin remodelling. A map of TBX1 binding sites in this model will be necessary to confirm this possibility.
Chromatin remodelling studies in vivo: Initial experiments and future perspectives.

It is desirable to replicate what we have obtained in the cell system using an “in vivo” approach. Thanks to the use of different cell types, we were able to demonstrate that there was chromatin remodelling due to Tbx1 dosage variations. In the last period of Ph.D. program, I moved from cells to embryos in order to demonstrate that there are chromatin changes following loss of Tbx1 expression also in vivo. We initiated ATAC-seq on FACS-purified cells from Tbx1 heterozygous and homozygous embryos. To label Tbx1-expressing cells we are crossing Tbx1cre/+ mice with Tbx1flox/++; R26mT/mG mice to obtain Tbx1cre/+ and Tbx1cre/flox embryos. The results are GFP+ cells which can be Tbx1cre/+ (Tbx1 heterozygous) or Tbx1cre/flox (Tbx1 homozygous). Figure 38 shows an example of GFP+ cells in a Tbx1cre/+, R26mT/mG E9.5 mouse embryo section.

Fig. 38

Figure 38. Tbx1 expression in Tbx1cre/+; R26mT/mG embryo. GFP+ cells mark the Tbx1 expression domain in E9.5 embryo.

Embryos are harvested at E9.5 (stage selected because this is a critical point for Tbx1 function), disaggregated and subjected to FACS purification of GFP+ cells. At the moment, we obtained two good libraries for two different Tbx1cre/flox; R26mT/mG embryos: profiles of DNA fragment are distributed between 200bp to 600 bp and we have selected, at moment, both embryos with 19 somites (Fig. 39). At this point we have two biological replicates for Tbx1 homozygous condition, we are waiting for two heterozygous mice for sequencing.
Figure 39. **DNA size distribution in $Tbx1^{cre/flox}$; $R26^{mT/mG}$ embryos.** The figure shows the DNA size distribution between 100bp to 600bp. 1 and 2 refers to first and second replicate and blue arrows indicate the fragments more abundant than others.

These two libraries will be paired with other two libraries of $Tbx1$ heterozygous mice to complete one run sequencing. ATAC-seq results will give us information regarding the chromatin accessible distribution following by $Tbx1$ dosage variations. These data will be combined with data from Morrow lab that is collaborating with us and generating TBX1 ChIP-seq data using a new mouse line carrying a tagged isoform of the $Tbx1$ gene.
DISCUSSION

Tbx1 is a candidate gene of DiGeorge syndrome and is necessary for outflow tract and craniofacial muscles development. In the last 20 years, different scientists have been focusing on this gene and many genetic experiments in mouse models have established that Tbx1 is a major regulator of the development of the cardiopharyngeal mesoderm. Although many epigenetic approaches have been pursued, the many details of the molecular mechanisms by which Tbx1 functions remain to be elucidated. The principal aim of this work thesis is to determine mechanisms of chromatin interactions by which Tbx1 regulates target genes. Tbx1 encodes a protein that belongs to T-box transcription factor family, which is essential for many development processes. Indeed, several T-box genes are haploinsufficient in a large number of birth defects (Naiche LA et al., 2005). Research in molecular mechanisms have been hampered by the embryonic lethality associated with T-box deficiency in mice, leading to difficulties in finding target genes (Miller S et al., 2009). Recently, many studies have been focusing on the interactions of T-box proteins with the histone modifying machinery. It seems that T-box factors cooperate with histone modifying enzymes such as, histone acetyltransferases or deacetylases and methyltransferases or demethylases in order to modulate the expression of target genes. For example, T-BET interacts with both H3K27-demethylase and H3K4-methyltransferase activities. It was found that this correlation between T-box and histone modifiers activities is associated with conserved residues in the T-box DNA binding domain provided the potential that this is a common mechanism utilized by the T-box family to regulate epigenetic states at development transition (Miller S et al., 2009). Fulcoli and co-workers (Fulcoli FG et al., 2016) found that TBX1, through interaction with a histone methyltransferase MLL3, promotes H3K4me1-regions and it binds to H3K27Ac-poor regions (Fulcoli FG et al., 2016) but the mechanism is still unclear. In the first period of my PhD program, I have tried to understand better this inverse correlation between TBX1 and histone acetylation. Through cells manipulation, I have done a ChIP-seq experiment using H3K27Ac antibody, following by Tbx1 dosage variations in order to understand if it could be differences in the genome-wide acetylation levels. ChIP experiment was abandoned because we had many technical problems related to sequencing. Although genome-wide acetylation maps following by Tbx1 dosage variations were not available, I demonstrated, using another cell system, in one specific case, that Tbx1 repression of Mef2c gene expression is correlated with the de-acetylation of a specific enhancer named Anterior Heart Field (AHF) of the Mef2c gene (Pane LS et al., 2018). The mechanism has not been elucidated yet. We have done many co-immunoprecipitation
experiments between TBX1 and HDAC proteins (HDAC1 and HDAC2) and it seems that TBX1 does not interact with any of them. Of course, there are other HDACs that may interact with TBX1, and it is also possible that the interaction is indirect so that it may not be revealed by a standard immunoprecipitation experiment. Furthermore, TBX1 might cause hypoacetylation by interfering with acetyltransferases rather than recruiting HDACs. Thus, further experiments are required to establish the mechanisms by which TBX1 binds H3K27Ac-poor regions. Histone post-translational modifications are not the only mechanism that TBX1 uses to regulate target genes because interaction with chromatin remodelers may be also important. TBX1 recruits BAF60a and activate or enhance transcription of a specific target (Chen L et al., 2012). Because the main function of chromatin remodelers is to make DNA more or less accessible to transcription factors and to the transcriptional machinery, I have focused on a newly developed and powerful technology to map chromatin accessibility genome-wide ATAC-seq. By manipulating the cell systems used in this work, I was able to generate maps of accessible regions with different dosages of Tbx1. I found that 86% of TBX1 binding sites are in closed chromatin. These regions do not remodel after 42 hours of TBX1 knock down. Differentially accessible regions (DARs) were mostly localized at the promoter of genes, and in regions that do not bind TBX1. Consistently, motif discovery analyses found no enrichment of T-box binding motifs in DARs. Therefore, TBX1 may remodel these regions, a) Indirectly or b) through a DNA-binding independent function.

Concerning the possible indirect effect, Tbx1 may regulate transcription factor-encoding genes, and their product, in turn, may remodel chromatin. A DNA-binding independent function is also possible as there are precedents in the literature for TBX1 and for at least another T-box protein (Messenger NJ et al., 2005). To demonstrate that TBX1 can remodel chromatin without binding DNA it would be necessary to test a mutant isoform unable to bind DNA. I have attempted to perform this experiment, but only using transfection of the mutant isoform compared to transfection of a WT expression vector. Unfortunately, the latter did not cause any chromatin remodelling at selected loci. There are different reasons that could explain why in these loci I did not find any differences in chromatin accessibility. DARs were detected after Tbx1 loss of function and these loci may not be responsive to overexpression. Furthermore, DARs have not been validated yet with additional experiments, therefore, it will be necessary to select validate loci for gain of function experiment. HOMER analyses of DARs in P19Cl6 cells returned strong enrichment of a particular transcription factor, NF-Y.

NF-Y is a pioneer factor which promotes chromatin accessibility by displacement of nucleosome core (Oldfield AJ et al., 2014). We will test
whether TBX1 knock down increases NF-Y occupancy at selected loci. If this is the case, we will test whether TBX1 may directly interact with this pioneer factor and perhaps make it less biologically available.

Next, we asked why TBX1 does not induce significant chromatin remodelling at its binding regions?

The results of my time course-experiment suggest that chromatin remodelling may occur at a later time point, perhaps because additional, required co-factors may only become available later during differentiation. Another, but related reason as to why we do not see remodelling at TBX1 binding sites, may be that P19Cl6 cells may not express co-factors required for chromatin remodelling. For this reason, I have also used mESC cells to study chromatin remodelling, and I am planning to use also an in vivo model, with the caveat that a map of TBX1 binding sites in these systems is not yet available.

An overall comparison of P19Cl6 and mES cell-derived data reveals differences and communalities. The main difference is that HOMER analyses of DARs produced substantially different lists of binding motifs, it seems that in P19Cl6 cells T-box factors do not bind any DARs suggesting an indirectly or DNA-binding independent mechanism. The finding of T-Box motifs enrichment in DARs and in regions which increase accessibility in KO condition of mESC model suggest that a good portion of chromatin changes may be located in TBX1-binding regions. Unfortunately, at the moment, we lack a good Tbx1 antibody necessary to depict the TBX1-binding regions in this cell system, therefore we cannot compare the accessibility regions with TBX1-binding sites. The hypothetical relation between T-box and regions which increase accessibility in KO condition remains a bioinformatics approach that certainly needs further investigations. The communality is that reduction or loss of TBX1 is associated with increased chromatin accessibility in both systems, suggesting that TBX1 suppresses chromatin accessibility at many loci. Figure 40 shows a working model by which TBX1 may regulate target genes. This could be a mechanism by which TBX1 inhibits cell differentiation in certain context such as cardiomyocyte differentiation.
Figure 40. Cartoon showing a working model by which TBX1 may regulate target genes. On the top, there is a schematic view of TBX1 working model: TBX1 (blue dot) can interact with Histone methyltransferases (HMTs – dark yellow dot) and tissue specific transcription factor (tsTF – green dot) and promotes the monomethylation of lysine 4 (H3K4me1) TBX1 binds closed chromatin. Without TBX1, the chromatin may change conformation (closed-opened) and gene target transcription can start: RNA pol II (yellow dot) binds promoter (blue rectangle). H3K27Ac (violet dot) is a marker of active enhancer. At the bottom, there are three putative mechanisms. 1) Indirect effect: TBX1 (blue dot) may regulate transcription factors-encoding genes (red rectangle), and their product (TFx – orange dot), in turn, may remodel chromatin. 2) DNA-binding independent effect: TBX1 (blue dot) interact with TFx (orange dot) or other proteins outside from chromatin context and, with a mechanism not yet demonstrated, the chromatin may change accessibility. 3) Repressor effect: Without TBX1 (blue dot), enhancers (black rectangle) may be active (H3K27Ac in violet) and RNA pol II (yellow dot) promotes transcription. By the contrast, TBX1 interacts with another TBX1 (blue dots) and repress transcription by masking enhancer (black rectangle) region.
CONCLUSIONS

Altogether, my results indicate that TBX1, a candidate gene of DiGeorge syndrome necessary plays important roles in gene regulation. Although TBX1 was identified about twenty years ago, its molecular functions are still to be clarified. In this work thesis I have demonstrated that TBX1 function may be cell type and context dependent. In P19Cl6 cells, TBX1 binds closed chromatin, and its loss of function increases chromatin accessibility in regions that are not bound by TBX1. This effect may occur through an indirect mechanism or through a non-DNA binding mechanism. Data that I have obtained using differentiating murine embryonic stem cells (mESCs) confirm that loss TBX1 opens chromatin. However, in these cells, I did find in opened regions, T-box binding motifs, albeit by bioinformatics analysis. These data suggest that TBX1 works differently in different cell types, perhaps because of the availability of different cofactors. However, bioinformatics finding need to be validated, therefore in the last period of my PhD program, I am using mESCs in which a specific tag (V5) has been knocked-into the Tbx1 locus and use anti-tag antibodies for ChIP-seq analysis. The epigenomic approach used in this work revealed new, unexpected findings concerning the chromatin response to TBX1 dosage, and open a new perspective onto the molecular functions of this transcription factor.
ACKNOWLEDGMENTS

First, I would like to thank my tutor, Prof. Antonio Baldini at the Institute of Genetic and Biophysics, CNR, Naples for the opportunity to work in his laboratory. I am very grateful for his constant support, as well as for the opportunity to work independently. I also want to thank my lab members: Prof. Elizabeth Illingworth, Rosa Ferrentino, Marca Bilio, Daniela Alfano, Sara Cioffi, Cinzia Caprio, Gemma Flore, Gabriella Lania, Ilaria Aurigemma, Ilaria Favicchia, Annalaura Russo for their assistance and helpful discussions. Next I would like to thanks Robert Kelly at the Marseille University for his wonderful revision about this work, Gabriella Fulcoli for technical support, Claudia Angelini and Monica Franzese for bioinformatics support and of course, my family. At the last but not the least, I would like to thanks Alessandra for her support and encouragement during these years.
REFERENCES

Alfano D, Altomonte A, Cortes C, Bilio M, Kelly RG, Baldini A. Tbx1 regulates extracellular matrix-cell interactions in the second heart field. bioRxiv 267906.


Botta A, Lindsay EA, Jurecic V, Baldini A. Comparative mapping of the DiGeorge syndrome region in mouse shows inconsistent gene order and


Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. Histone H3K27ac separates active from poised enhancers and


Xu H, Morishima M, Wylie JN, Schwartz RJ, Bruneau BG, Lindsay EA, Baldini A. **Tbx1 has a dual role in the morphogenesis of the cardiac outflow tract.** Development. 2004 Jul;131(13):3217-27.


LIST OF PUBLICATIONS
