UNIVERSITÀ DEGLI STUDI DI MILANO

DOCTORAL SCHOOL IN BIOCHEMICAL, NUTRITIONAL AND METABOLIC SCIENCES

DIPARTIMENTO DI SCIENZE CLINICHE E DI COMUNITÀ

PHD COURSE IN BIOCHEMISTRY

XXVII Cycle

PhD THESIS

Definition of transcriptional landscape in cardiac maturation and cardiac hypertrophy

BIO/12

Giuliano Giuseppe Stirparo

Matricola R09748

TUTOR: Professor Viviana Cavalca

COORDINATORE DEL DOTTORATO: Professor Francesco Bonomi

ANNO ACCADEMICO 2013/2014

Index

ABSTRACT

Heart failure (HF) is a syndrome resulting from a complex genetic predisposition and multiple environmental factors: it is a leading cause of morbidity and mortality. Specific gene expression patterns are activated in the hypertrophic and failing heart and are thought to contribute to the development of HF. Many regulatory molecules are involved in the control of gene expression: among these, long non-coding RNA (lncRNA) is gaining importance for several cellular process and diseases. However, little is still known about its involvement in HF. Many functions have been attributed to lncRNAs, such as cell proliferation, apoptosis and cell invasion, indicating that they may represent a major regulatory component of the eukaryotic genome. Not surprisingly, lncRNAs have been found implicated in several aspects of cancer, and in many neuronal diseases. Despite this, and the known role of other ncRNAs, such as miRNA, in HF, the function of lncRNAs in this pathologic state has been not studied. Thus, the general hypothesis behind this project is that lncRNAs have an important role in defining gene expression re-programming in HF. Consequently, the overall scientific objective of this proposal is to study the role of lncRNAs in gene transcription regulation accompanying heart failure. To this end, we propose to use high-throughput RNA sequencing (RNA-seq) to identify lncRNAs that are modulated in cardiomyocytes during HF. In order do to this, we performed RNA-seq on cardiomyocytes isolated from mice after 1, 2, 4 and 7 days of transverse aortic constriction (TAC) and from sham-operated mice. The importance of this study lies not only in the furthering of our understanding of the pathological mechanisms leading HF, but aims to generate – in the light of recent progress in RNA-based therapeutic strategies – data that may be instrumental to the development of improved therapeutic strategies for this increasingly frequent pathology.

INTRODUCTION

1.1- Heart Failure

Heart failure (HF) is not only a leading cause of morbidity and mortality (Ahmad et al., 2012), it is the ultimate outcome of many cardiovascular pathologies. It is underlain by gene expression and epigenetic reprogramming, wherein certain fetal genes are upregulated while adult genes become downregulated (Izumo et al., 1988, Lompre et al., 1979).

Heart failure may be due to cardiac hypertrophy, an adaptive response to pressure or volume stress. Hypertrophic growth is the cause of many forms of heart disease, such as hypertension, ischemic disease, valvular disease, and heart failure. The hypertrophic mechanism is the way in which cardiomyocytes respond to mechanical stress, enabling myocytes to increase their work output and improve cardiac pump function (Figure 1). Ventricular hypertrophy is therefore associated with a high risk of heart failure (Levy et al., 1990) (Koren et al., 1991).

The hypertrophic transformation of heart has been divided in the past (Meerson, 1961) in 3 steps:

1) Stage in which load exceeds output, developing hypertrophy
2) Stage which the workload/mass ratio is normalized, compensatory hypertrophy
3) Stage characterized by ventricular dilation and subsequent heart failure

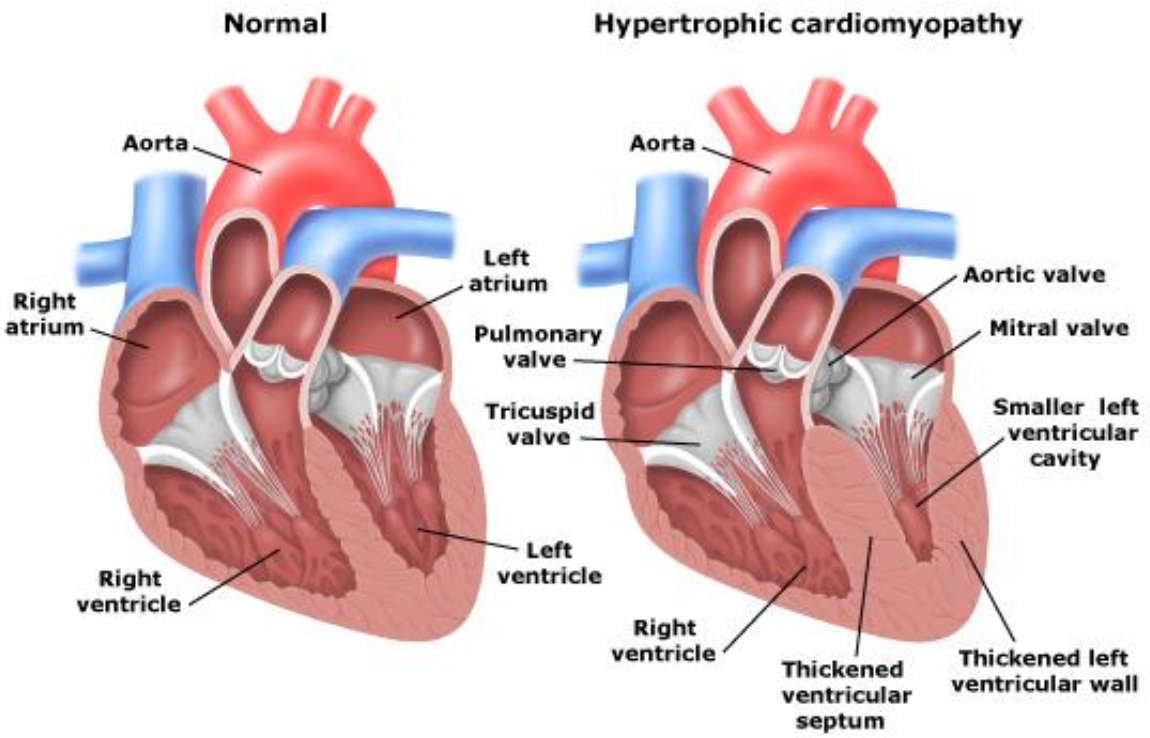


Figure 1. The picture shows the difference in size and structure between normal and hypertrophic heart.

Furthermore, cardiac hypertrophy, a pathological event that is ultimately responsible for a deterioration in heart function, is principally due to gene reprogramming. It has been discovered that mechanisms similar to those that govern gene expression in heart development (and maturation) are at the basis of those that occur in HF (Hunter & Chien, 1999). The study of the mechanisms that control cardiac gene expression in HF and heart development provides information conducive to the better understanding of HF pathogenesis, and could lead to new diagnostic and therapeutic tools (Han et al., 2011).

1.2- ncRNAs

One of the greatest surprises obtained during the last years has been the discovery of noncoding RNAs (ncRNAs) with high-throughput analysis of the transcriptome. ncRNA are a variety of RNA that are not translated into a protein and act at the level of transcription or translation and influence gene expression through various mechanisms—such as RNA interference. Based on size, ncRNAs can be divided into two groups (Table1): (1) small ncRNAs (<200nt), which include microRNA (miRNA), PIWI-interacting RNA (piRNA), and endogenous small interfering RNA (siRNA), and (2) long ncRNAs (lncRNAs), which have a length between 0.2 and 2 kb and, thus, a coding potential of less than 100 amino acids (Mitra et al., 2012, Qureshi & Mehler, 2012, Saxena & Carninci, 2011, Wang & Chang, 2011).

tRNA/mt-tRNA	Transfer RNA & mitochondrial transfer RNA	It is an adaptor molecule composed of RNA, typically 73 to 94 nucleotides in length, that serves as the physical link between the nucleotide sequence of nucleic acids (DNA and RNA) and the amino acid sequence of proteins.
rRNA/mt-rRNA	Ribosomal RNA & mitochondrial ribosomal RNA	It is the RNA component of the ribosome, and is essential for protein synthesis in all living organisms
scRNA	Small cytoplasmic RNA	Small (7S; 129 nucleotides) RNA molecules found in the cytosol and rough endoplasmic reticulum associated with proteins that are involved in specific selection and transport of other proteins.

snRNA	Small nuclear	A general term for many different kinds of small RNA molecules found in the nucleus of a cell that include as examples species involved in splicing of introns from mRNA and in RNA interference.
snoRNA	Small nucleolar	They are a class of small RNA molecules that primarily guide chemical modifications of other RNAs, mainly ribosomal RNAs, transfer RNAs and small nuclear RNAs.
piRNA	piwi RNA	piRNAs guide PIWI proteins to complementary RNAs derived from transposable elements. Similarly to in RNA interference, PIWI proteins cleave the transposon RNA, leading to silencing.
miRNA	microRNA	Small RNAs containing 21 to 33 nucleotides that associated with multiple proteins in a RNA-induced silencing complex (RISC) that repress transcription of specific target mRNA by hybridizing to its 3' untranslated region.
siRNA	small interfering	It is a class of double-stranded RNA molecules, 20-25 base pairs in length. siRNA plays many roles, but it is most notable in the RNA interference (RNAi) pathway
lncRNA	long non coding RNA	Large intergenic non-coding RNAs, usually associated with open chromatin signatures such as histone modification sites. Could be divided in different categories according to their genomic position and have multiple functions.

Table 1. The table shows the different classes of small non coding RNA and long ncRNA.

So far, a large range of functions have been attributed to ncRNA, such as cell proliferation, apoptosis, and cell invasion and imprinting, indicating that these molecules may represent a major regulatory component of the eukaryotic genome. Not surprisingly, ncRNAs are emerging as important players in several human pathologies, including cardiovascular diseases (Mitra et al., 2012, Qureshi & Mehler, 2012, Saxena & Carninci, 2011, Wang & Chang, 2011). Indeed, miRNAs play a key role in driving the gene expression changes of HF, atherosclerosis, and cardiac ischemic/reperfusion injury (Latronico & Condorelli, 2009). However, we are just beginning to understand the biology of piRNA, siRNA, and lncRNA in the cardiovascular system.

1.3- Biological Roles of lncRNAs

lncRNAs are currently defined as transcripts greater than 200 nt without known protein-coding function (Batista & Chang, 2013). Though thousands of lncRNAs have been identified across eukaryotes, many appear less conserved than protein-coding genes and therefore seems to be species specific (Cabili et al., 2011, Derrien et al., 2012). lncRNAs can be classified into one or more of the following five categories, based on their genomic distribution, (Figure 2): (A) sense, the lncRNA sequence overlaps with the sense strand of a protein coding gene; (B) intronic, the lncRNA sequence is derived entirely from within an intron of another transcript; (C) bidirectional, the lncRNA is located on the opposite strand from a protein coding gene whose transcription is initiated less than 1000 base pairs away; (D-E) antisense, the lncRNA sequence overlaps with the antisense strand of a protein coding gene; and/or (F) intergenic, the lncRNA sequence is not

located near any other protein coding loci. Although only a very small percentage of identified lncRNAs have been explored experimentally, they are known to be implicated in many biological processes such as cell fate specification, X chromosome inactivation, nuclear compartmentalization and architecture, genomic imprinting, translational control, RNA splicing and chromatin modification (Wutz & Gribnau, 2007).

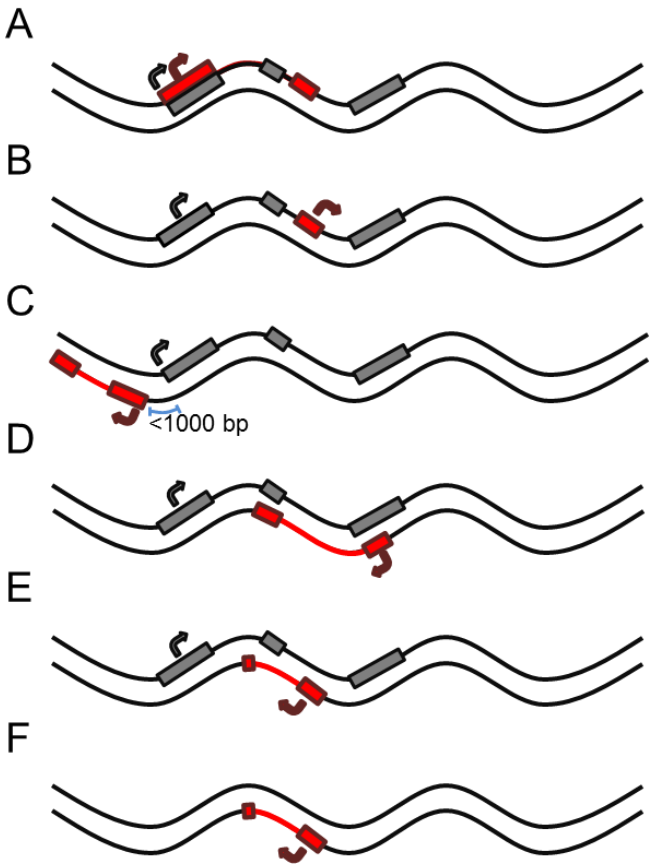


Figure 2. The picture shows the genomic position of lncRNA, (A)Sense, (B) Intronic, (C) Bidirectional, (D,E) Exonic/Intronic Antisense, (F) Intergenic.

1.3.1- X chromosome Inactivation and Genomic Imprinting

Xist is one of the best-studied lncRNAs. It was discovered in 1991 and now it is widely accepted to be required for the silencing of hundreds of genes on the inactive X chromosome in female somatic cells by favouring the formation of a chromatin structure with an epigenetic profile linked to transcription repression. This process is known as X chromosome inactivation (XCI) (Figure 3) (Heard et al., 2001). However the exact mechanism of Xist-mediated XCI is yet to be fully elucidated, despite having been studied extensively (Wutz & Gribnau, 2007). lncRNAs are also implicated in genomic imprinting, a process that inactivates either the maternally or the paternally inherited allele. Imprinted genes play crucial roles in mammalian development, and so their expression must be highly regulated. Intriguingly, many imprinted gene loci express, in addition to mRNAs, a significant number of lncRNAs (e.g., Air, H19, and Kcnq1ot1) that appear to play major roles in regulating the expression of neighbouring imprinted protein coding genes in *cis* (Reik & Murrell, 2000).

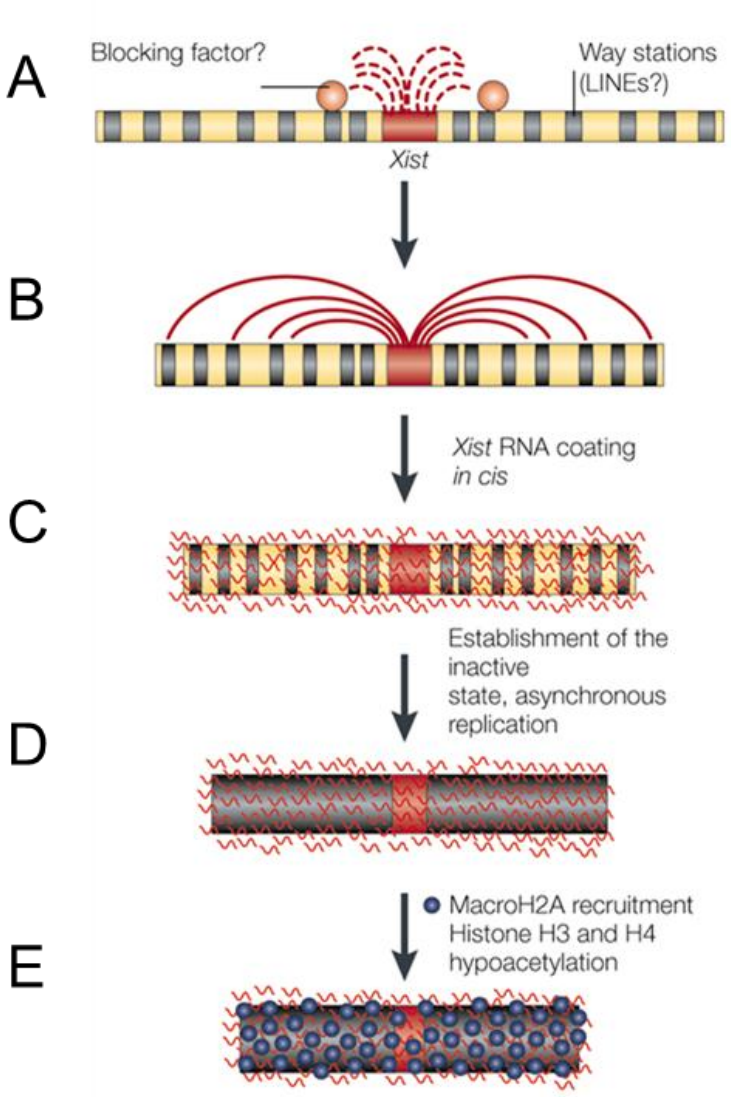


Figure 3. Model of Xist Inactivation: A, blocking factors prevent Xist RNA, expressed in an unstable form (dotted red lines), the association with the chromosome in cis. B, through stabilization Xist RNA becomes upregulated and release of the blocking factor. C, Xist RNA coats the X chromosome. D, Xist RNA coating cause transcriptional silencing of genes on the X chromosome. E, chromatin modifications, transform the Xist RNA-coated chromosome into a stably inactive and condensed chromatin state (Avner & Heard, 2001).

1.3.2- Nuclear Compartmentalization and Architecture

The contribution of lncRNAs in the building of a cell's architecture is well known. A clear example comes from paraspeckles—ribonucleoprotein bodies found in the interchromatin space of mammalian cell nuclei—which play a central role in regulating gene expression during cell differentiation. These structures control many aspects of transcription and RNA processing, such as transcription initiation, coactivation and corepression, RNA splicing, and transcription termination, through retention of RNA in the nucleus (Bond & Fox, 2009). Given the large numbers of long noncoding transcripts currently being discovered through whole-transcriptome analysis, paraspeckles may be a paradigm for a class of subnuclear bodies formed around lncRNA.

1.3.3- Cell Fate Specification

It is well accepted that many lncRNAs provide an additional layer of regulation in the specification of cellular identities through the regulation of gene expression at transcriptional and translational levels (Hu et al., 2012). Recently, the muscle-specific lncRNA linc-MD1 was identified as a regulator of muscle cell differentiation (Figure 4), by acting as a competing endogenous RNA in mouse and human myoblasts. Downregulation and over-expression of linc-MD1 correlate with retardation and anticipation of the muscle differentiation program, respectively (Cesana et al., 2011).

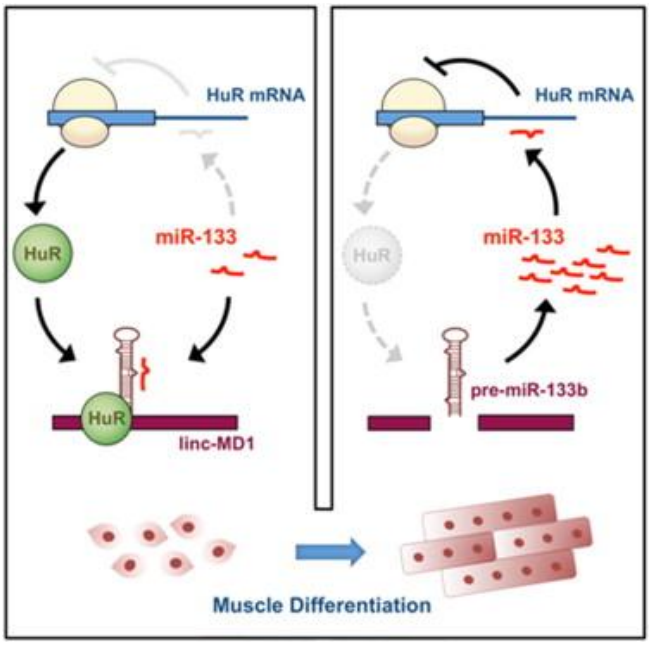


Figure 4. Muscle-specific long noncoding RNA linc-MD1 as a sponge for miR-133 and miR-135. Legnini I. *et al.* describe alternative synthesis between linc-MD1 and miR-133 is controlled by the HuR protein, which favours linc-MD1 accumulation through its ability to bind linc-MD1. The author shows that HuR is under repressive control of miR-133 and that the sponging activity of linc-MD1 consolidates HuR expression in a feedforward positive loop (Legnini et al., 2014).

1.3.4- RNA Splicing and Translational Control

Alternative splicing of pre-mRNAs increases the diversity of the proteome by generating different protein products having non-overlapping functions from a single mRNA. MALAT1 (metastasis-associated lung adenocarcinoma transcript 1) was originally identified as a gene that was specifically upregulated in metastatic non-small cell lung cancer cells, but has subsequently been recharacterized as an lncRNA that accumulates in the nucleus and is now referred to as NEAT2 (nuclear-enriched noncoding transcript 2) (Hutchinson et al., 2007). This lncRNA regulates alternative splicing through its

interaction with serine-/arginine-rich (SR) nuclear phosphoproteins involved in the splicing machinery (Lin et al., 2011b).

1.3.5- Chromatin Modification

lncRNAs have been proposed to regulate transcription by recruiting chromatin-remodelling complexes, which in turn mediate epigenetic changes. For example, Kcnq1ot1, Airn, Xist and HOTAIR are four lncRNAs that act to promote the formation of repressive chromatin structure across large genomic regions and even entire chromosomes (Figure 5). This is achieved through recruitment of epigenetic enzymes, such as polycomb repressive complex 1 (PRC1) and polycomb repressive complex 2 (PRC2), which mediate mono-ubiquitinylation of lysine 119 on histone 2A and trimethylation of lysine 27 on histone H3, respectively (Pandey et al., 2008, Terranova et al., 2008). Another class of ncRNA that has received much emphasis is found expressed at enhancer regions: this type of RNA has been named enhancer RNA (eRNA). The size of eRNA ranges from 0.1 to 9 kb, comprising an average of 800 nt. Most promote transcription of neighbouring genes. An example is ncRNA-a7, which acts in *cis* to induce transcriptional activation of the nearby gene *Snai1* (Orom et al., 2010). There are probably many other functions of lncRNAs awaiting discovery. For example, the lncRNA NRON has been shown to regulate nuclear trafficking of the transcription factor nuclear factor of activated T cells, and the observation that many lncRNAs are located in the cytoplasm suggests that they might have undiscovered roles in cell biology (Kapranov et al., 2007).

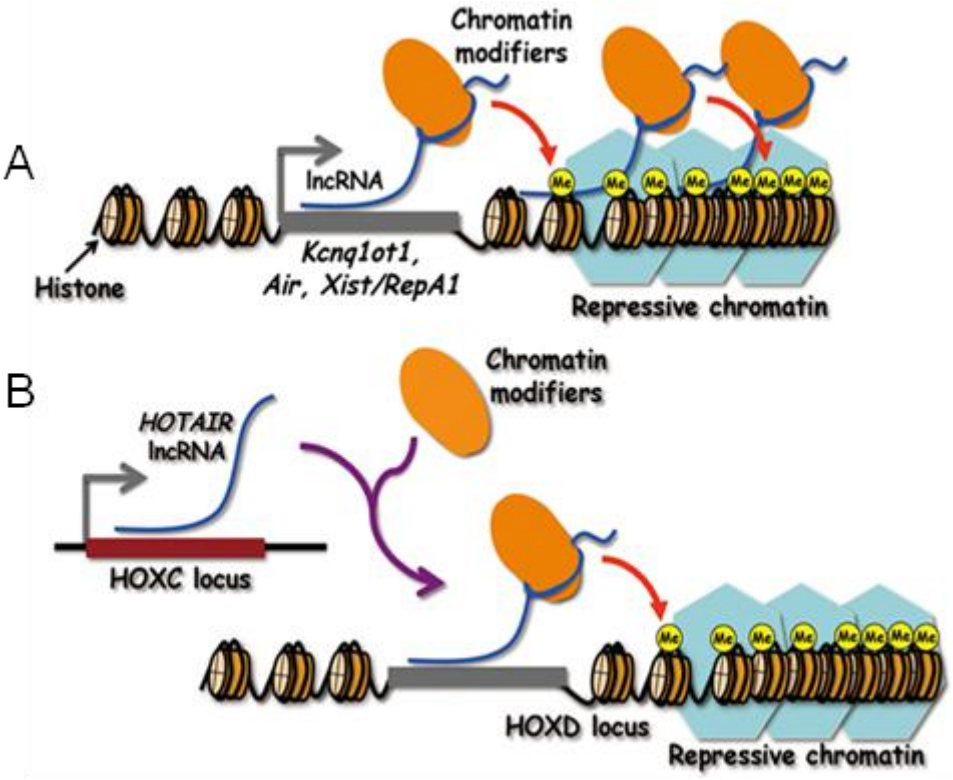


Figure 5. Chromatin modifiers in *cis*. (A) lncRNAs Kcnq1ot1, Air and Xist/Rep1 that are transcribed from the coding loci are accumulated on a chromatin in cis to recruit chromatin modifies. By modification of histones and DNA the modifiers facilitate the formation of the repressive chromatin. "Me" indicates methylation of a histone. (B) Chromatin modifiers in trans. HOXC locus express HOTAIR lncRNA that binds to the HOXD locus in trans and inactivates transcription by recruiting the chromatin modifiers to form the repressive chromatin.

1.4- Molecular Mechanisms of lncRNAs

The precise mechanism of how lncRNAs function still remains largely unknown and is the subject of intense investigation. Wang and Chang have proposed to classify the molecular mechanisms of lncRNAs into four main categories (Figure 6) (Wang & Chang, 2011):

1.4.1- Signal lncRNAs

These lncRNAs can serve as molecular signals because transcription of individual lncRNAs occurs at very specific times and places in order to interpret cellular context or to respond to diverse stimuli. For instance, Air and Kcnq1ot1 mediate the transcriptional silencing of multiple genes by interacting with chromatin and recruiting the chromatin-modifying machinery (Mohammad et al., 2009, Wang & Chang, 2011).

1.4.2- Decoy lncRNAs

The molecular decoy type of activity takes place when specific lncRNAs are transcribed and then bind to and titrate away protein factors. Decoy lncRNAs can “sponge” protein factors such as transcription factors and chromatin modifiers. This leads to broad changes in the cell’s transcriptome. One of the most abundant nuclear lncRNAs in mammalian cells is NEAT2, which is localized in nuclear speckles. NEAT2 binds to and sequesters several SR splicing factors to nuclear speckles. Depletion of NEAT2 alters splicing-factor localization and activity, leading to an altered pattern of alternative splicing for a set of pre-mRNAs (Tripathi et al., 2010). In addition, lncRNAs can titrate microRNAs out from their target, functioning as competitive endogenous RNA (Cesana et al., 2011).

1.4.3- Guide lncRNAs

These can act as molecular chaperons by localizing particular ribonucleoprotein complexes to specific chromatin targets. This activity can cause changes in the gene expression of neighbouring (cis) or distantly located (trans) genes that cannot be easily predicted by just the lncRNA sequence itself. Examples are HOTAIR and HOTTIP, both transcribed within the human HOX clusters. HOTTIP may organize chromatin domains to coordinate long-range gene activation by serving as key intermediates that transmit information from higher order chromosomal looping into chromatin modifications. In contrast, lncRNAs such as HOTAIR are able to alter and regulate epigenetic states in cells through their targeting of chromatin-modifying complex occupancy-localization-enzymatic activity in trans (Wang et al., 2011).

1.4.4- Scaffold lncRNAs

Some lncRNAs possess multiple domains that bind distinct proteins. The complexes formed perform functions such as transcriptional activation or repression. Thus, the lncRNA serves as an adaptor to form the functional protein complexes. The telomerase RNA TERC (TERRA), which is part of telomeric heterochromatin in addition to being present in the nucleoplasm, is a classic example of an RNA scaffold and is essential for telomerase function (Collins, 2008).

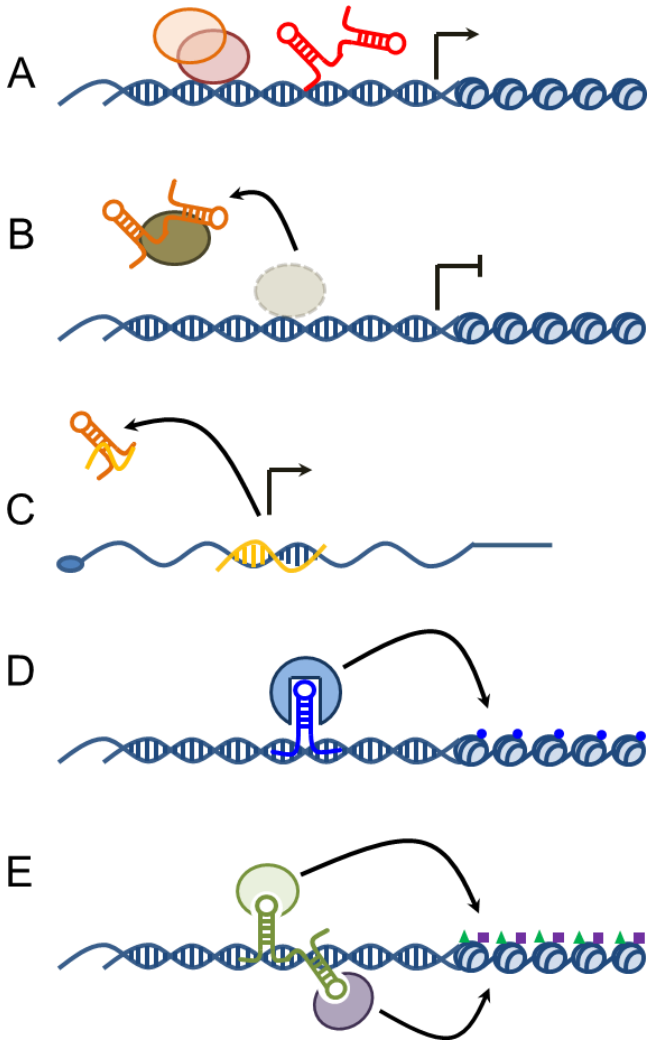


Figure 6. Molecular mechanism of long noncoding RNAs. (A) Signalling lncRNA: combination of lncRNAs and transcription factors may affect gene regulation; (B)decoy lncRNA: lncRNAs can displace transcription factors away from the chromatin, leading to the silencing of a nearby gene or (C) lncRNAs can titrate microRNAs out from their target; (D) guide lncRNA: lncRNAs may guide chromatin-modifying enzymes to their site of action, either in *cis* or in *trans* ; (E) scaffold lncRNA: lncRNAs can induce the assembly of multiple proteins affecting histone modifications. TF transcription factor, CME chromatin-modifying enzymes, RbNC ribonucleoprotein complexes.

1.5- Are lncRNAs Important Players of Heart Development?

Over recent years, next-generation sequencing technologies have enabled the study of transcriptomes in various cell types at different developmental stages. These studies have revealed that the expression of many lncRNAs is regulated during development and that their role ranges from the control of pluripotency to lineage specification (Collins, 2008, Kung et al., 2013, Lin et al., 2011a). For instance, XCI is tightly correlated with both early embryonic development and pluripotency in embryonic stem cells and induced pluripotent stem cells (Deuve & Avner, 2011). Knockdown of dozens of lncRNAs causes either exit from the pluripotent state or upregulation of lineage commitment programs, demonstrating that lncRNAs play key roles in the circuitry controlling the embryonic stem cell state (Guttman et al., 2011). The role of lncRNA in heart development is only starting to emerge. Indeed, two recent reports demonstrated that two lncRNAs, Fendrr and Braveheart (Bvht), are involved in defining the gene transcription program underlying the development of lateral mesoderm in heart and the differentiation of cardiomyocytes, respectively (Grote et al., 2013, Klattenhoff et al., 2013). Fendrr is specifically expressed in nascent lateral plate mesoderm and is required for the correct development of this tissue in heart and the body wall. Indeed, Fendrr-deficient mice die around E13.75 due to abnormal functioning of the heart. The heart of these mice presents with a thin ventricular wall, caused by cardiac hypoplasia linked to an altered proliferation of cardiac myocytes at later stages of heart development. Fendrr regulates the expression of important transcription factors, such GATA-6, NKX2-5, FOXF1, TBX3, IRX3 and PITX2, by controlling the epigenetic profile of the promoters of these genes. In

fact, Fendrr can bind either to PRC2 or to the Trithorax group/MLL protein complex (TrxG/MLL), which induces trimethylation of lysine 27—an H3 histone marker associated with transcription activation—and trimethylation of lysine 4 on histone H3—a marker associated with transcription activation—respectively (Grote et al., 2013). The lncRNA Braveheart (Bvht) is required for cardiomyocytes differentiation in vitro and for maintaining the cardiac phenotype in neonatal cardiomyocytes. Bvht is required for expression of core gene regulatory networks involved in defining cardiovascular cell fate and acts upstream of mesoderm posterior (MesP1), a master gene of cardiovascular lineage commitment. In a similar way to Fendrr, Bvht regulates cardiomyocytes differentiation by modulating the epigenetic profile of cells through interaction with SUZ12, a component of PRC2 (Klattenhoff et al., 2013).

1.6- lncRNA: a Possible Player in Heart Failure

If we are only at the beginning of understanding the role of lncRNA in heart development, we are even further away from comprehending the function of these molecules in cardiovascular disease. However, recent studies have started to investigate their role in this area. A genome-wide association study (GWAS) found six single-nucleotide polymorphisms (SNP) in the lncRNA MIAT (myocardial infarction associated transcript) associated with myocardial infarction; one SNP (A11741G) caused a 1.3-fold increase in MIAT transcription in vitro (Ishii et al., 2006). Another study revealed 15 lncRNAs modulated in the heart of mice subjected to aortic constriction—a procedure that, through the creation of pressure overload, induces first compensated hypertrophy

and then HF (Lee et al., 2011). Moreover, transcription in the heart of Kcnq1—a gene encoding a potassium channel—depends on the expression of the lncRNA Kcnq1ot1. This is an unspliced, ~60 kb lncRNA whose transcription starts at intron 10 of Kcnq1 and in the opposite direction to that of the host gene. Mice that express a truncated Kcnq1ot1 have an increased expression of Kcnq1. Because correct potassium channel activity is required for normal cardiac functioning, any alteration of the Kcnq1ot1-mediated control of Kcnq1 could be responsible for abnormal heart function (Korostowski et al., 2012). The involvement of lncRNA in HF is also suggested by the role of ANRIL (antisense noncoding RNA in the INK4 locus) in coronary heart disease (McPherson et al., 2007). This lncRNA is an antisense noncoding RNA transcribed with the INK4b– ARF–INK4a gene cluster, which encodes oncosuppressor proteins p15INK4b, p14 ARF and p16INK4a. GWAS revealed that many SNPs mapped to this genomic locus and, in particular, to the gene encoding ANRIL and were associated with increased susceptibility to coronary artery disease and several other diseases, including cancer. ANRIL regulates the expression of the INK4b–ARF–INK4a gene cluster through recruitment of both PRC1 and PRC2, promoting the formation of a repressive chromatin structure (Yap et al., 2010). In the light of these findings, it was proposed that loss-of-function mutations in this lncRNA lead to altered expression of p15INK4b, p14 ARF, and p16INK4a, influencing thus the susceptibility to coronary heart disease and cancer.

1.7- lncRNA: a Possible Therapeutic Target for Heart Failure?

A therapeutic strategy for HF is to cure the sick heart cells by interfering with the gene expression program that underlies HF. To this end, novel therapeutic opportunities might arise from RNA-based strategies that consist in the use of synthetic RNAs capable of programming biological function. The potential of this strategy for HF is supported by the discovery of the role of miRNAs in triggering the gene expression program of HF and by the possibility of manipulating ncRNAs in vivo. Indeed, anti-miRNAs and antisense oligonucleotides are being employed to inhibit the expression of specific miRNAs in vitro and in vivo for investigative and clinical purposes (Care et al., 2007, van Rooij et al., 2008). The use of the same strategies for lncRNA could open new therapeutic avenues for HF, overcoming problems linked to the use of miRNA-based molecules, such as low specificity and poor pharmacokinetics. Indeed, future studies on the role of lncRNA in HF and heart development will improve our understanding of the ncRNA network involved in regulating gene expression changes underlying HF and, thus, allow the development of specific therapeutic strategies based on the interference not only of miRNAs but also of lncRNA important for HF. These studies will greatly benefit from the combination of next-generation sequencing technologies applied to RNA (RNA-seq) with bioinformatics tools developed to identify lncRNAs that are differentially expressed in different biological conditions and for the prediction of their mechanism of action. Table 2 lists some bioinformatics tools used today to study lncRNAs.

1.8- Epigenetic

The regulation of gene transcription in multicellular eukaryotic cells not only is driven by DNA elements that control gene expression (for example promoters and enhancer) but also by the state of chromatin in which the gene is localized (Birney et al., 2007). In this context, a mechanism that plays a crucial role is epigenetics, a complex network of mechanisms that control heritable gene expression without altering the primary nucleotide sequence. These mechanisms regulate gene expression by modulating chromatin structure and adjusting the biological processes that are the basis of gene expression (Care et al.). Epigenetic mechanisms are grouped mainly in four categories: DNA methylation, post-translational modifications of proteins histone, chromatin remodelling and small non-coding RNAs function as regulatory micro RNA (miRNA) and long non-coding RNAs (long non-coding RNA) (Figure 7) (Relton & Davey Smith, 2010). Many functional studies have shown how the epigenetic mechanisms play an important role in regulating the expression of genes involved in important biological processes such as cell differentiation, cell cycle and the DNA damage repair. So it is not surprising as epigenetic alterations are responsible for several human diseases such as cancer and some neuronal disorders such as Alzheimer's disease and Huntington's disease (Urdinguio et al., 2009).

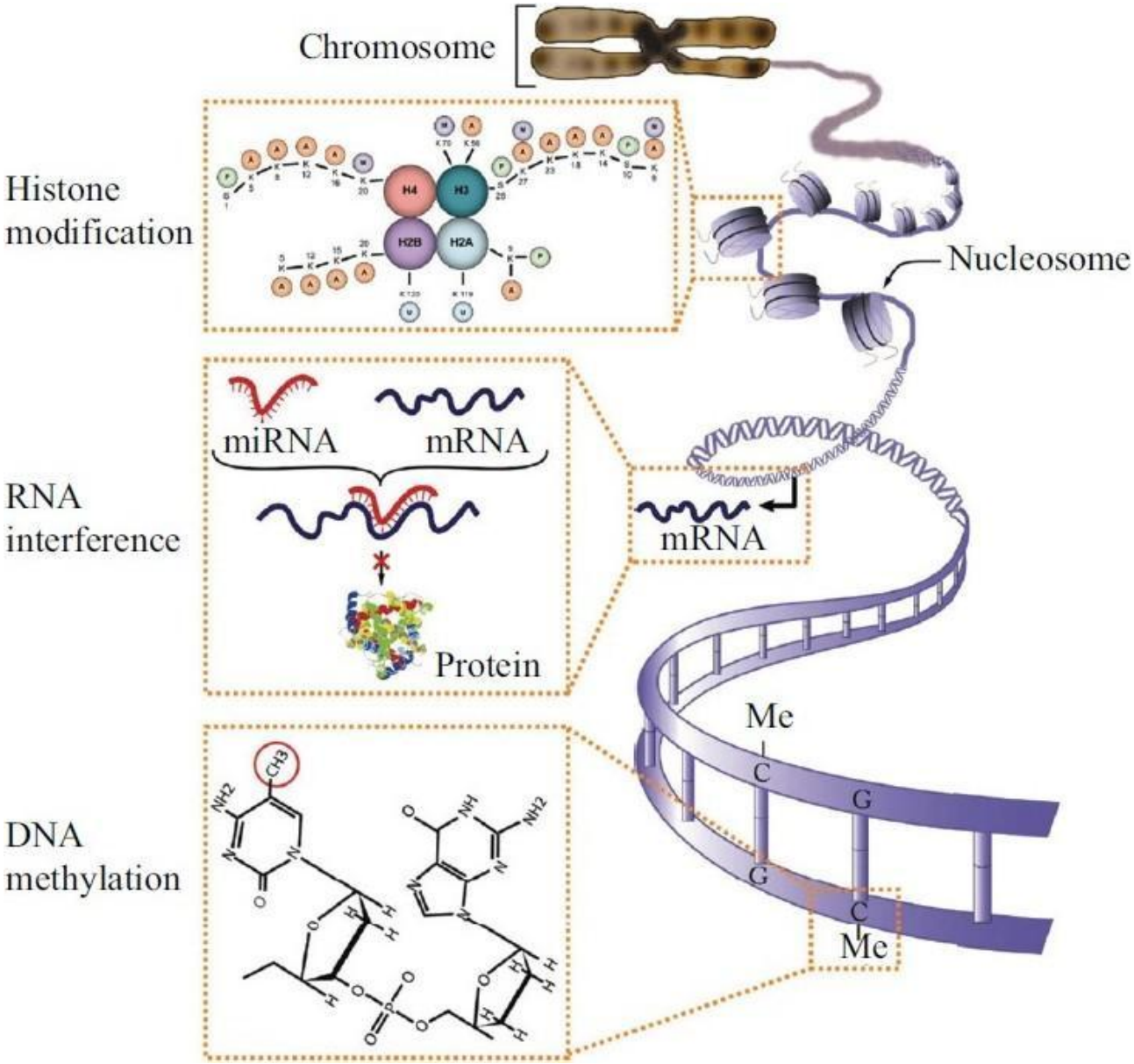


Figure 7. Schema of principal mechanism of epigenetic regulation.

1.8.1- Chromatin Structure

The entire genomic DNA of eukaryotic nuclei is organized into a highly compact structure called chromatin. It is made from DNA, small amount of RNA, and by two groups of protein molecules: histones and not histone proteins such as transcription factors, RNA polymerases, DNA polymerase and the topoisomerase. There are two types of chromatin: one that consists of thin filaments, called euchromatin and another one which is called heterochromatin. The euchromatin is characterized by an open conformation due to the relaxed nucleosomic state arrangement and a high flexibility, features that make the DNA accessible to proteins that regulate gene expression (Birney et al., 2007). The heterochromatin, however, is made up of portions temporarily (optional heterochromatin) or permanently (constitutive heterochromatin) very condensed. This feature makes the DNA inactive from transcriptional point of view being inaccessible to transcription factors. The heterochromatin in addition to having a role in the regulation of transcription also has a function to control the stability and protection from chromosomal mutations and translocations (Huang et al., 2004). This dynamic composition is regulated by multiple epigenetic mechanisms (Groth et al., 2007). The fundamental unit of chromatin is the nucleosome that consists of approximately 146 base pairs of genomic DNA that wrap around an octamer of basic proteins, histones, called the "core histone". The nucleosome is the same type of organization in all eukaryotes. There are five types of histones: H1, H2A, H2B, H3 and H4. These, except histone H1, are present in equimolar amounts in the nucleosome (two histones of each class) and are arranged in an octamer composed of a tetramer of histones H3 and H4

histones of two dimers H2A-H2B. Histones are small basic proteins (the molecular weight is comprised between 11 and 20kDa), being rich in basic amino acids such as arginine and lysine. They are highly conserved from the evolutionary point of view and have a structure consisting of a carboxyl-terminal globular domain and amino-terminal end that protrudes from the nucleosome (called N-terminal tail). H1 is not part of the core histone, but seems to play an important role in inducing the chromatin to assume higher order structures.

1.8.2- The Regulation of Chromatin

The reorganization of chromatin superstructure allows the DNA to be suitably condensed, carefully replicated and ordered (Luger et al., 1997). Chromatin is a dynamic structure that can undergo strong conformational changes during DNA replication, gene expression, DNA repair and recombination events. The mechanisms that regulate the activity of chromatin are:

• The reorganization of the nucleosome remodelling factors ATP-dependent.

• The post-translational modifications of histones, creating a complex combination of signals that can activate or repress transcription, depending on the chemical nature of the molecule that makes the modification.

1.8.3- The Post-Translational Modifications of Histones

Many studies have shown that combinations of post-translational modifications to level of the same histone or between different histones present in the same chromosome region constitute a "code", called epigenetic code that regulates many functions phones. The post-translational modifications of histones comprise methylation, acetylation, phosphorylation, ADP-ribosylation and ubiquitination (Fischle et al., 2003). Among these, the best known are acetylation and methylation.

1.8.4- Acetylation of Histones H3 (H3ac)

Acetylation is the addition of an acetyl group (-COCH3) group amino (NH2) present in the lysine residues of histones H3, H4, H2A and H2B. The acetylation is catalysed by histone acetyl transferase (Histone Acetyl Transferase, HAT), while the removal of the acetyl group is carried out by histone deacetylase (Histone deacetylases, HDAC) (Conerly & Grady, 2010). Genetic studies in combination with analysis of ChIP (chromatin immunoprecipitation) suggested that the opposing action of these two groups of proteins helps to maintain the equilibrium state of the level acetylation. Depending on the recruitment of HAT or HDAC on promoters determines a state of hyper or hypo-acetylation correlated with activation or transcriptional repression, respectively. Acetylation, indeed, is a modification that promotes histone transcription through by promoting formation of a more open chromatin structure and therefore facilitating the binding of transcription factors on chromatin. Furthermore, the acetylation of lysines

determines a decrease of the positive charge on histone tails, which has the effect of reducing the affinity between adjacent nucleosomes. This event causes chromatin to be very dynamic and less condensed. (Ruthenburg et al., 2007).

1.8.5- Methylation of Histone H3 (H3me)

The methylation of histones involves the inclusion of one or more methyl groups (-CH3) to arginine and lysine on histone H3 and H4. Histone methylation, as acetylation, is a reversible process. In recent years, it has been identified several histone lysine methyltransferase (Lysine methyltransferase, KMT) and histone lysine demethylases (Lysine DeMethylases, KDM) (Black et al., 2012). From the functional point of view, if the acetylation of histones is always associated with transcriptional activation, the situation is different for methylation, which according to the methylated residue and the degree of methylation can promote transcriptional activation or repression (Figure 8). The methylation of lysine residues 4 and 79 of histone H3 are associated with transcriptional activation and lysine 9 and 27 residues on histone H3 promote transcriptional repression. Methylation also has different roles in regulation of transcription because each methylated residue is able to recruit specific proteins that can activate or repress transcription (Kouzarides, 2007). For example, the trimethylation of lysine 4 (K4) on histone H3 (H3K4me3) is a modification that promotes transcription through the binding of CHD1 (chromatin organization modifier), a transcription factor involved in transcription elongation and the processing of pre-mRNA. This histone modification is present on the promoter of transcriptionally active genes. An important aspect of the

regulation of gene expression by histone modifications is their interdependence that arises from the impossibility of two histone modifications to co-exist on the same amino acid residue and the ability to trigger different cellular signals according to the chemical group conjugated to the residue. For example, phosphorylation of serine 10 on histone H3 favours acetylation on lysine 9, preventing methylation of this residue. Lysine 9 acetylation and methylation are associated with transcription activation and repression (Fischle et al., 2003). Many of the enzymes that modify histones post-translationally show a degree of specificity not only to a particular amino acid residue, but also to the state of histone modification of the existing one.

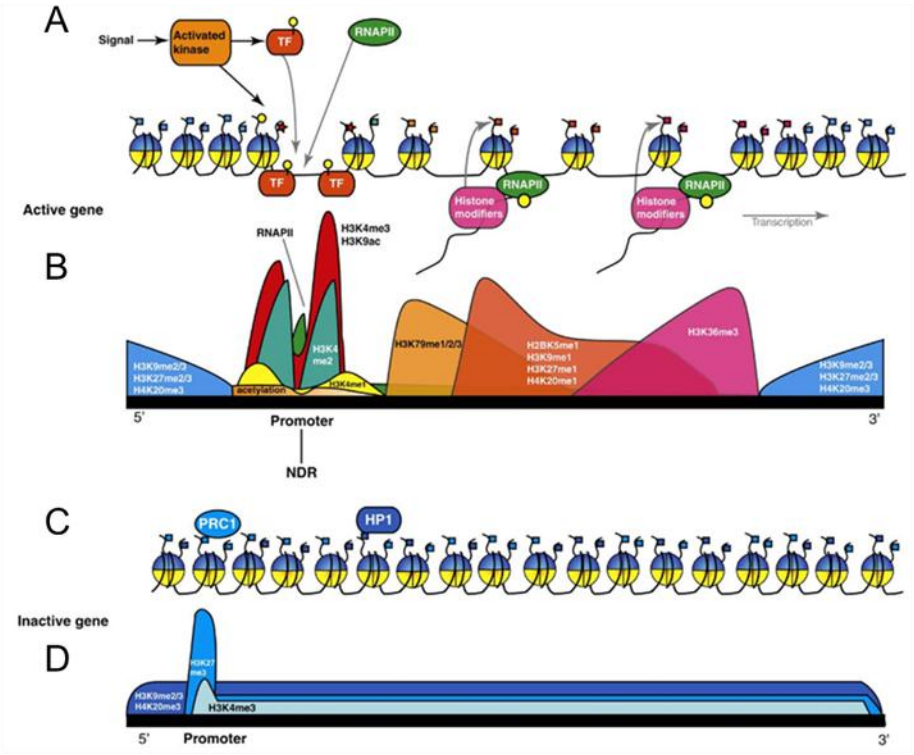


Figure 8. Distribution of histone modifications in active genes or silenced. In theFigure are schematized the nucleosome (A and C) and the distribution of changes in the gene (B e D). An active gene is displayed in (A) and (B). An external signal can lead to the activation of kinases (orange) in the nucleus, which phosphorylate (yellow circles) both histone and transcription factors (TF, red) to cause an appropriate physiological response. The transcription factor can bind to the promoter DNA to facilitate the docking of the RNA polymerase II (RNAPII, green) giving rise to the transcript. The elongation of RNA polymerase II to highly phosphorylated in the C-terminal domain, can interact, directly or indirectly, with histone modifiers (pink) (Kim et al., 2012). In particular, nucleosomes in promoter of transcriptionally active genes contain high levels of activators, such as acetylation and methylation of H3K4, creating a region of nucleosome depletion (NDR) in which bind the RNAPII. Inactive genes, shown in (C) and (D), have a regular distribution of changes in silencing, such as H3K9me2/3, while H3K27me3 is enriched in the promoter. These modifications constitute the binding site for heterochromatic proteins (blue) and which make the condensed chromatin and therefore transcriptionally inactive.

1.9- Enhancers

The eukaryotic promoters immediately upstream of the transcription start site (TSS, transcription start site) are needed to establish the appropriate level of expression of genes. To express their function, the promoter sequence is assembled with a class of proteins known as transcription factors of polymerase II (TFII). The function of the trans-agent of transcription factors is not exclusively related to the promoters, but often relates to other sites or distal *cis*-regulatory elements, which contribute to the modulation of transcription of genes located on the same DNA molecule. These elements are called "enhancers" or "silencers" depending on their function. They are able to operate regardless of orientation, position and distance, and contain sequences recognized by tissue-specific and/or inducible transcription factors. The enhancer and silencer may act on the promoter, increasing or inhibiting its activity through interactions with the basal transcription complex, or they can modify the chromatin structure so as to facilitate access or disadvantage of the complex transcription start on the DNA.

Recently, genome-wide studies of the distribution of different histone modifications have shown that the enhancers have a specific epigenetic profile which allows not only the identification but also the prediction of their state of activities. The mono-methylation of lysine 4 of histone H3 (H3K4me1) is considered a marker of enhancers, as it distributes much more at the level of these genetic elements than on the promoters. In contrast, H3K4me3 is largely absent at the enhancer, but is predominant at the promoters of active genes (Birney et al., 2007). It is possible to identify the putative enhancer through the analysis of the genomic distribution of these two types of histone

H3 methylation. This prediction becomes much more significant when the histone methylation markers are combined with other indicators of the activity of distal elements in specific cell types, such as the binding of the transcriptional cofactor p300, acetyl-transferase (Blow et al., 2010). In addition, other studies have shown that acetylation of histone H3K27 in combination with histone H3K4 mono methylation marks enhancers that distribute near active genes. Inactive enhancers called "poised", however, are associated with the presence of H3K4me1, combined with trimethylation of H3K27 and the absence of H3K27 acetylation (Rada-Iglesias et al., 2011).

AIM OF THE PROJECT

The aim of this project was to identify modulated protein coding genes and long-intergenic noncoding RNA in order to study the development and progression of cardiac hypertrophy. To achieve this, we sequenced the transcriptome of isolated cardiomyocytes after 1, 2, 4 and 7 days of transverse aortic constriction and Sham-operated mice as control. These data allowed us to determine networks and biological processes, important key genes, lncRNAs or novel lncRNAs involved in the pathology. We also analysed how promoter regions of protein coding genes and lncRNAs are regulated by transcription factors. Furthermore, to better understand the progression of the disease, we studied the fetal-like induced genes and lncRNAs involved in heart failure. The importance of this study lies not only in the furthering of our understanding of the pathological mechanisms leading HF, but aims to generate – in the light of recent progress in RNA-based therapeutic strategies – data that may be instrumental to the development of improved therapeutic strategies for this increasingly frequent pathology.

MATERIAL AND METHODS

2.1- Transverse Aortic Constriction

To induce cardiac hypertrophy in mice, with subsequent heart failure, transverse aortic constriction is often used for its reproducibility and gradual time course. Initially, compensatory hypertrophy is associated with an enhanced cardiac contractility. Later, the hemodynamic overload became detrimental, causing cardiac dilatation and heart failure (deAlmeida et al., 2010)(Figure 9).

According to the New Directive 2010/63/EU, special attention was paid to minimize animal suffering and welfare. During the procedure, anesthetized mice are closely monitored to assure that they were maintained in the proper anaesthetic plane. By pinching the tail of the animal while monitoring body temperature and respiratory rate is assessed the anaesthetic efficacy. Any reaction of the animal, hypothermia or increase in respiratory rate indicate that anaesthesia is too light and that should be given additional anaesthetic. All experiments were conducted in conformity with European Communities Council Directive. TAC surgery was adapted from Rockman HA *et al* (Rockman et al., 1991). Briefly, C57Bl6/J mice were anesthetized by intraperitoneal injection of a mixture ketamine (100mg/kg) and xylazine (10 mg/kg). The mice were then placed in a supine position, and the chest was shaved using a chemical hair remover. The aortic arch was exposed through the 1st intercostal space. A 8-0 Prolene suture was passed between the truncus anonymous and the left carotid artery, a blunted 27-gauge needle was placed against the aorta and the knot tightened along with the needle and secured with a second knot. The needle was removed to create a lumen with a fixed stenotic diameter. The chest cavity was then closed with a 6-0 silk suture. A separate group of mice

underwent the same surgical procedure but without any tightening of the knot (sham group). The pressure load caused by the knot was verified through the measurement of the pressure gradient across the aortic constriction with echocardiography. The mean pressure gradient in TAC mice made to be approximately 70 mmHg in order to induce a rapid deterioration of heart function.

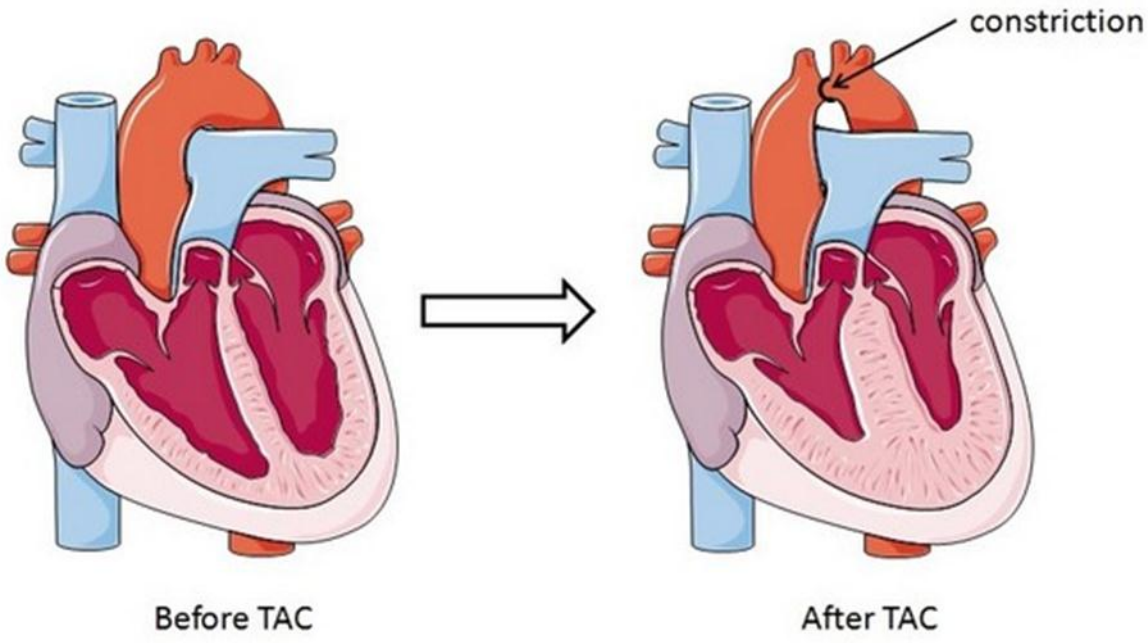


Figure 9. The picture show the enlargement of ventricular walls caused by the transverse aortic constriction.

2.2- Cardiomyocytes Isolation

The procedures used to obtain adult cardiomyocytes from sham- and TAC-operated mice, and for the culture of cardiomyocytes from neonatal mice, have been described elsewhere (Louch et al., 2011).

2.3- RNA Extraction

RNA was obtained from cells using TRIzol® reagent (Life Technologies®) following the manufacturer's instruction. The process is divided in 4 steps.

2.3.1- Sample Homogenization

Cells were harvest by centrifugation and add to 0.75 mL TRIzol® Reagent per 0.25 mL of sample. Cells were lysed by pipetting up and down several times.

2.3.2- Phase Separation

Add 0.2 mL of chloroform to 1 mL of TRIzol® reagent and shake. The samples are then incubated for 2-3 minutes and centrifuged at 12.000 rpm for 15 minutes at 4°C degree. The aqueous phase, containing the RNA, were removed and dispensed in new tube. Then, add 0.5 mL of isopropanol to the aqueous phase, incubates for 10 minutes at room temperature centrifuge at 12.000 rpm for 15 minutes at 4°.

2.3.3- RNA Precipitation

Add 0.5 mL of isopropanol 100% to the aqueous phase and incubate for 10 minute sat room temperature. Centrifuge at 12000 rpm for 10 minutes at 4°C.

2.3.4- RNA Wash

After the removal of supernatant, wash pellet with 1 mL of 75% ethanol, vortex and centrifuge for 5 minutes at 4°C. Repeat this procedure two times. At the end, resuspend the RNA in 20-25 µl of RNase-free water.

2.4- Library Preparation and RNA-Sequencing

The library was prepared following TruSeq Stranded Total RNA Sample Prep Guide (Illumina®). Briefly, after quality assessment of RNA integrity and purity with Bioanalyzer 2100 (Agilent Technology, Inc), the RNA were depleted from ribosomal RNA fraction. Subsequently are synthetized the first and second strand of cDNA respectively. To ensure a low rate of chimera transcripts the 3'end are adenylated. At the end, multiple indexed adapter were ligated at the ends of ds-cDNA template and these molecules were then amplify through PCR. Finally were checked quality of the library with Bioanalyzer 2100.

The RNA was sequenced with Illumina HiSeq 2000. Briefly, each fragment of RNA is bound randomly inside the surface of a flowcell and then amplified forming cluster of dsRNA. Then, for each cluster is read the incorporation of labelled nucleotides. This process generates an amount of sequences of nucleotide corresponding to the original

cluster; these sequence are then aligned to the genome to quantify the expression of the relative molecules.

Below are show the number of reads generates for each samples (Table 2).

Samples	Total Reads	Mapped Reads
Sham Day1 Replicate1	52947412	47955356
Sham Day1 Replicate2	59397336	52902223
Sham Day2 Replicate1	56361481	50590870
Sham Day2 Replicate2	60133321	53772985
Sham Day4 Replicate1	55819479	50554195
Sham Day4 Replicate2	57352629	50888211
Sham Day7 Replicate1	56743781	48811810
Sham Day7 Replicate2	52773918	46918196
Tac Day1 Replicate1	65731893	59198328
Tac Day1 Replicate2	59588156	53345326
Tac Day2 Replicate1	53383412	48073849
Tac Day2 Replicate2	51068134	45814014
Tac Day4 Replicate1	57711882	51800046
Tac Day4 Replicate2	58465497	52441887
Tac Day7 Replicate1	51185723	45069530
Tac Day7 Replicate2	55496080	48734104
Embryonic Replicate 1	59677804	52744873
Embryonic Replicate 2	62122420	55201554
Pups Replicate 1	58629137	52636914
Pups Replicate 2	61193627	55333980

Table2. Table indicating the number of total sequenced reads and mapped read.

2.5- Alignment and Filtering

The reads obtained from sequencing were aligned against the last Ensembl version of *Mus Musculus* genome (GRCm38_76) with Tophat (Trapnell et al., 2009, Langmead et al., 2009). Tophat is a program that allow to aligns sequenced RNA reads in order to identify exon-exon splice junctions and to assign univocally the reads to a region of genome. It is built on the Bowtie (Langmead et al., 2009), ultrafast short read mapping algorithm.

Briefly, the algorithm creates an index of the genome, by using Burrows-Wheeler transform (BWT) that allow to increase sensitivity and reducing time and memory calculation, and FM index with two novel extension for short reads alignment: a quality backtracking, to favours alignment with high quality and allow mismatch, and double indexing strategy to prevent excessive mismatching. Furthermore, as the transcripts coordinates are not continuative on the genome, the software is supply a reference annotation, that contain the relative coordinates of genes on chromosome, to help the reconstruction of the transcriptome (Langmead et al., 2009).

Below, is the command used to align all the FASTA file, and list and the description of Tophat parameters used for the analysis (Table3).

```
tophat -g 10 -r 200 --mate-std-dev 150 -p 3 --library-type fr-firststrand
       -G $reference/Mus_musculus_GRCm38_76.gtf -o $sham_day1_rep1 $bow/bowtie_index
       $fasta/CM_sham_1gg1_R1 $fasta/CM_sham_1gg1_R2
```

Command	Description
-g	Instructs TopHat to allow up to this many alignments to the reference for a given read, and choose the alignments based on their alignment scores if there are more than this number. The default is 20 for read mapping. Unless you use --report-secondary-alignments, TopHat will report the alignments with the best alignment score. If there are more alignments with the same score than this number, TopHat will randomly report only this many alignments. In case of using --report-secondary-alignments, TopHat will try to report alignments up to this option value, and TopHat may randomly output some of the alignments with the same score to meet this number.
-r	This is the expected (mean) inner distance between mate pairs. For, example, for paired end runs with fragments selected at 300bp, where each end is 50bp, you should set -r to be 200. The default is 50bp.
--mate-std-dev	The standard deviation for the distribution on inner distances between mate pairs. The default is 20bp.
-p	Use this many threads to align reads. The default is 1.
--library-type	The default is unstranded (fr-unstranded). If either fr-firststrand or fr-second strand is specified, every read alignment will have an XS attribute tag as explained below. Consider supplying library type options below to select the correct RNA-seq protocol. Fr-Firststrand: we enforce the rule that the right-most end of the fragment (in transcript coordinates) is the first sequenced (or only sequenced for single-end reads). Equivalently, it is assumed that only the strand generated during first strand synthesis is sequenced.
-G	Supply TopHat with a set of gene model annotations and/or known transcripts, as a GTF 2.2 or GFF3 formatted file. If this option is provided, TopHat will first extract the transcript sequences and use

	Bowtie to align reads to this virtual transcriptome first. Only the reads that do not fully map to the transcriptome will then be mapped on the genome. The reads that did map on the transcriptome will be converted to genomic mappings (spliced as needed) and merged with the novel mappings and junctions in the final tophat output.
-o	Sets the name of the directory in which TopHat will write all of its output. The default is "./tophat_out".

Table3. List and description of parameters used for alignment.

After alignment, the reads were filtered according to Phred quality score of 20. The filtering was performed with Samtools (Li et al., 2009) software for the manipulation of the alignment files. In the details the alignment files could be indeed in SAM format (sequence alignment map). These files contain not only the base composition of the reads and the chromosome coordinate at which it aligns, but also the quality score for each base called by sequencer and aligned. In this way is possible to extrapolate the desired information, or to filter out the low quality reads from the SAM file.

Subsequently, to retrieve quantitative information, HTseq count software (Anders et al., 2014) was used to count the aligned reads in the exonic region according to a list of genomic features. In details, the feature is an interval of coordinates on a particular chromosome. In this case the feature are the annotation of genes. With this software. the user has to pay attention in deciding how to act with reads that overlap more than one feature. The software allow 3 types of resolution mode (Figure 10):

-union

-intersect-strict

-intersection-nonempty

Figure 10. Different possibility of association between the sequenced reads and the annotation. For each possibility is shown if gene reads is assigned to gene or not.

To reduce the number of false positive assignments and increase sensitivity and sensibility, we decide to adopt the intersection-strict resolution.

HT-seq count generates, as output, a table with the list of the feature (in this case the gene name) and the relative number of reads associated to them.

2.6- Differential Analysis

The differential analysis was performed in R, a software environment for statistical computing and graphics (http://www.r-project.org/). EdgeR (Empirical analysis of digital gene expression data in R) (Robinson et al., 2010) package was used to analyse the modulated gene. Briefly, EdgeR uses empirical Bayes estimation and exact tests based on the negative binomial distribution; the row count are normalized on the total number of reads and log2 transformed. The significance threshold were set at pval<0.01 (pval<0.001 for novel lincRNA).

2.7- Novel lincRNA Discovery

Cufflinks (Roberts et al., 2011, Trapnell et al., 2012) software was used to identify novel lincRNA. In particular, cufflinks is able to estimate the abundance of annotated transcripts but more important, is able to assemble novel islands of transcription in unannotated regions. To achieve this, Cuffcompare command was used to compare the transcribed region found on the experiment (based on RNA-seq evidence) to the annotation file and output the unannotated transcribed region. Subsequently on these regions, the number of reads were counted, normalized, transformed and analysed with EdgeR package to find modulated lincRNA across time and condition. Furthermore, was possible to reduce the number of false positives by intersecting the region actively transcribed (obtained from RNA Sequencing) with regions epigenetically regulated (obtained from Chip Sequencing) (Figure 11).

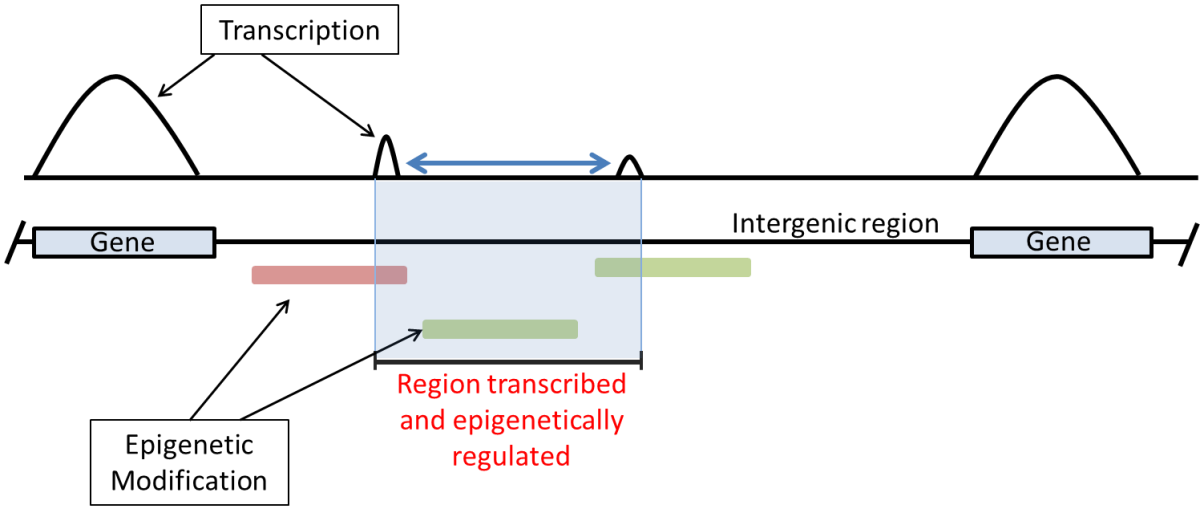


Figure 11. The picture shown how the region actively transcribed could be also epigenetically regulated. This combination could improve the discover and identification of novel lncRNAs.

2.8- Encode

The Encyclopaedia of DNA Elements (ENCODE; http://genome.ucsc.edu/ENCODE/) (ENCODE, 2012) Consortium is an international collaboration of research groups aimed to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels. From ENCODE, we downloaded from the University of Washington RNA-seq experiments performed on C57BL/6 mice at 8 weeks (http://www.washington.edu/). Each alignment file was convert into the most recent version of genome and annotation with CrossMap.py script (Zhao et al., 2014) and analysed with HT-seq count.

2.9- PhastCons

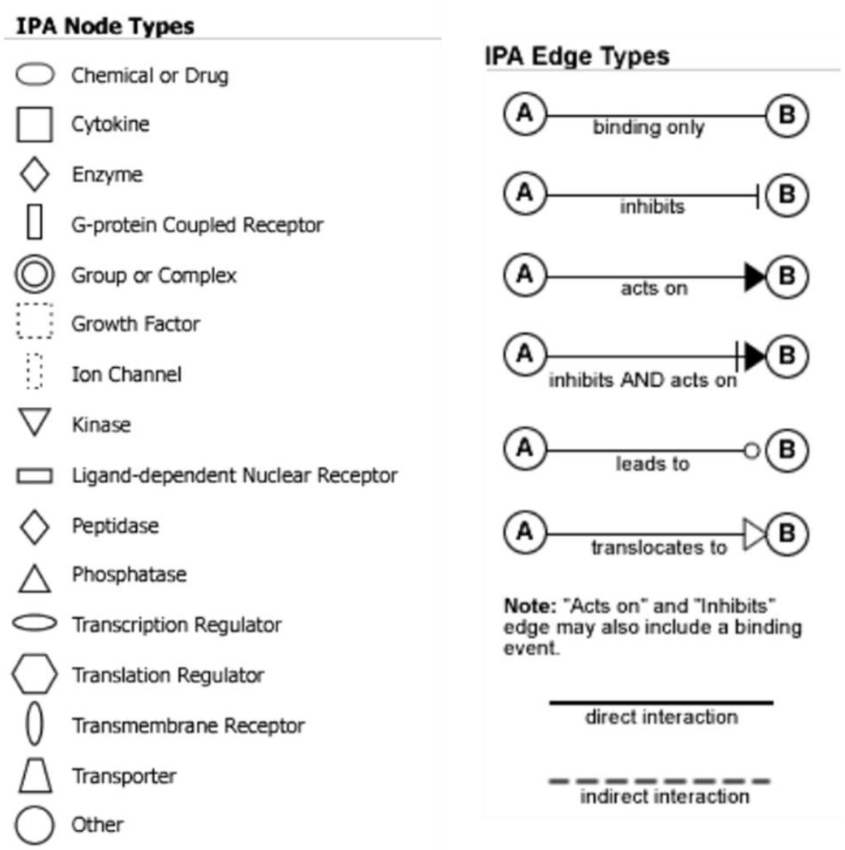
Phast score of "phastCons60wayEuarchontoGlire" were downloaded from UCSC (http://genome.ucsc.edu/) and for each base-position of genes were calculated the relative conservation score.

2.10- Coding Potential Calculation

CPAT (Wang et al., 2013) software was used to calculate the coding potential for protein coding genes and lncRNA in order to assess their capacity to codify peptide and protein. Coding Potential Assessment Tool (CPAT) is able to recognize coding and noncoding transcripts from a large pool of candidates. To do this, CPAT uses a logistic regression model with 4 features: hexamer usage bias, open reading frame size, Fickett TESTCODE statistic and open reading frame coverage.

2.11- IPA Analysis

The networks, biological process, TOX and BIO function were generated through the use of QIAGEN's Ingenuity Pathway Analysis (IPA®,QIAGEN Redwood City, www.qiagen.com/ingenuity). To perform the analysis we consider direct and indirect relationship among genes, by taking into account only experimentally observed interaction in mouse. Below are show the nodes and edges types legend for the comprehension of IPA networks.



2.12- Distribution of Pearson Correlation

Person correlation is defined as the measure of correlation between two variables. In particular, the variable used in the experiments are the fold-change of lncRNA, and the fold change of protein coding genes. Due to the small dataset we decide to use the bootstrap methods, that is a statistical technique of resampling.

```
X<-as.matrix(read.table("Table_of_FC.txt",as.is=T))
set.seed(1)
rows      = c(T)                                  # Total number of rows for each table(T)
for(i in 1:N) {                                   # used a for loop, Number of Iterations(N)
  bootRows = sample(rows, size=r, replace=F)      # How many rows(r) use?
  bootDist[i] = cor(X[bootRows,1], X[bootRows,2]) # Calculation of Pearson Correlation for N times on r Random rows
}
```

2.13- Heat Map

Two-way hierarchical clustering of significantly modulated genes was performed with Cluster 3.0 software using gene-wise median-centred normalized intensities. Centred correlation and complete linkage clustering were used as the distance metric. Heat maps were visualized using Java TreeView software.

2.14- cDNA Amplification and RT-PCR Analysis

From total RNA we synthetized complementary DNA (cDNA) with High capacity cDNA reverse transcription kit (LifeTechnologies) following the manufacturer's instructions. Briefly, total RNA (1µg) was mixed to a 10X RT Buffer (2µl), 25X dNTPS (0.8µl), 10X random primer (2µl), reverse transiptase (1µl), RNase inhibitor (1µl) and RNase-free water up to 20 µl. The mix was incubated in thermalcycler for:

- 10 minutes 25°
- 120 minutes 120°
- 5 minutes 85°

Then samples were stored at -80°. Validation were performed with RT-PCR using SYBR Master Mix on Viia7 Fast Real-Time PCR instrument (Applied Biosystem)Reaction contained 20 ng of cDNA, 1 µl of forward and reverse primer and 10ul of SYBR green master Mix. List of specific primer are provided below.

2.15- Chip-Seq Sequencing and Analysis

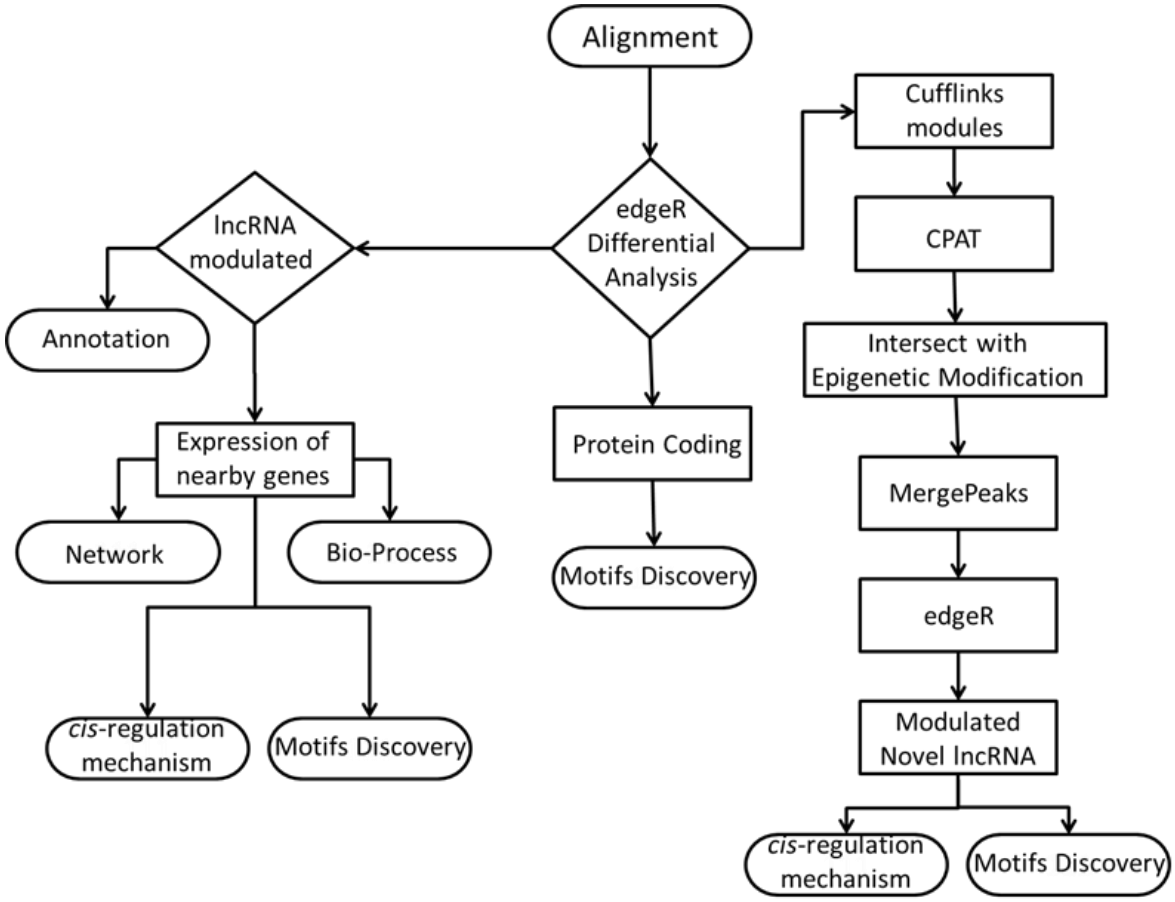
ChIP was performed as described elsewhere (Peters et al., 2003). Briefly, the chromatin obtained from cardiomyocytes isolated from mice after 1 week of transverse aortic constriction (TAC) and from mice that were sham operated was sonicated to obtain an average chromatin length of 300bp. Immunoprecipitation was performed by incubating the chromatin at 4 °C overnight with antibodies against H3K9ac (ac, acetylation; K, lysine; H3, histone H3; Millipore-Upstate code: 07–352), H3K27ac (Abcamcode, ab4729), H3K4me3 (me, methylation; Active Motif code,39159), H3K79me2 (Abcam code, ab3594), H3K9me2 (Abcamcode, ab1220), H3K9me3 (Millipore-Upstate code, 07442), and H3K27me3 (Millipore-Upstate code, 07449). Then, ChIP DNA fragments were sequenced with the SOLiD system (Life Technologies) at Genomnia srl. ChIP DNA samples (5 ng) eluted in 50µL Tris-EDTA buffer (Tris·HCl 10mM, EDTA 1mM at pH 8) were sheared using the Covaris S2 System to reach a median size of 50-300bp. ChIP sequencing (ChIP-seq) library preparation was performed according to the manufacturer's protocol for 5500 SOLiD ALPHA ChIP-seq Library Preparation (Life Technologies). Briefly, sheared samples were first end-repaired, modified by adding a poly-deoxy-adenylic acid-tail to their 3'ends, and ligated to specific SOLiD adaptors containing P1 and different barcode sequences. The nick-translated adaptor-ligated DNA was amplified using P1 and barcode primers, and after two Agencourt AMPure XP (Agencourt Bioscience Corporation) bead purifications steps, the yield and size distribution of the libraries were checked using the Agilent DNA 1000 Kit (Agilent Technologies, Inc.). Equal-molar barcoded libraries were pooled, and emulsion PCR was performed using the SOLiD EZ Bead system, according to

the manufacturer's instructions. Sequencing was performed using standard settings on the 5500xl SOLiD Systems according to the 5500 Genetic Analysis System, Run sequencer protocol. At least 20 million tags for each library, 50 bp long, were sequenced. To profile histone modifications in cardiomyocytes, sequencing reads were mapped to the mouse genome(version mm9), using BOWTIE. Uniquely mapped reads with no more than 2 mismatches were used for binding peak detection. To identify peaks, two software sets for peak calling were used: MACS (model-based analysis of ChIP-seq) and SICER (spatial clustering approach for the identification of ChIP-enriched regions). Both programs detected binding peaks by comparing IP with the input control. MACS and SICER were used with default parameters (Zang et al., 2009, Feng et al., 2012). Occupancy analysis and differential binding affinity analysis were assessed with the R Bioconductor package DiffBind (differential binding analysis of ChIP-seq peak data). The final set of binding peaks contained those that were identified by both software (Ross-Innes et al., 2012). Then the peaks were converted with UCSC liftover (http://genome.ucsc.edu/) in last version of genome coordinates.

2.16- Project Pipeline

Below is shown the pipeline used for the analysis. Briefly, starting from the alignment and differential analysis we were able to identify annotated lncRNA modulated. Starting from here, we determined the expression of nearby gene, that allowed us to find the modulated and enriched biological processes and networks, motifs enriched in promoter regions of lncRNAs and protein coding genes. Furthermore, Cufflinks modules, with the

application of different filter, and the intersection with epigenetic modification, allowed us to identify a set of novel modulated lncRNAs.



RESULTS

3.1- Hypertrophic Outcome

To identify the lncRNAs modulated in cardiac hypertrophy, we carried out RNA-seq experiments on RNA purified from cardiomyocytes isolated from the left ventricle of mice that had been subjected to transverse aortic constriction (TAC), a surgical procedure that induces cardiac hypertrophy and then heart failure. Cardiomyocytes were isolated after 1, 2, 4, 7 days of TAC. The hypertrophy state was evaluated analysing the level of expression of several hypertrophy markers. Hypertrophic cardiomyocytes (Figure 12 Panel a; *<0.05; **<0.01) had an increase (*NPPA, NPPB, MYH7*) or decrease (*MYH6*) according to hypertrophic outcome (Figure 12, Panel b). Furthermore, principal component analysis (PCA, Figure 12 Panel c), a statistical technique used to reduce the dimension of large datasets, performed on total RNA-sequenced, showed a high reproducibility between replicates and increased shifting of principal components after 4 and 7 days of TAC surgery.

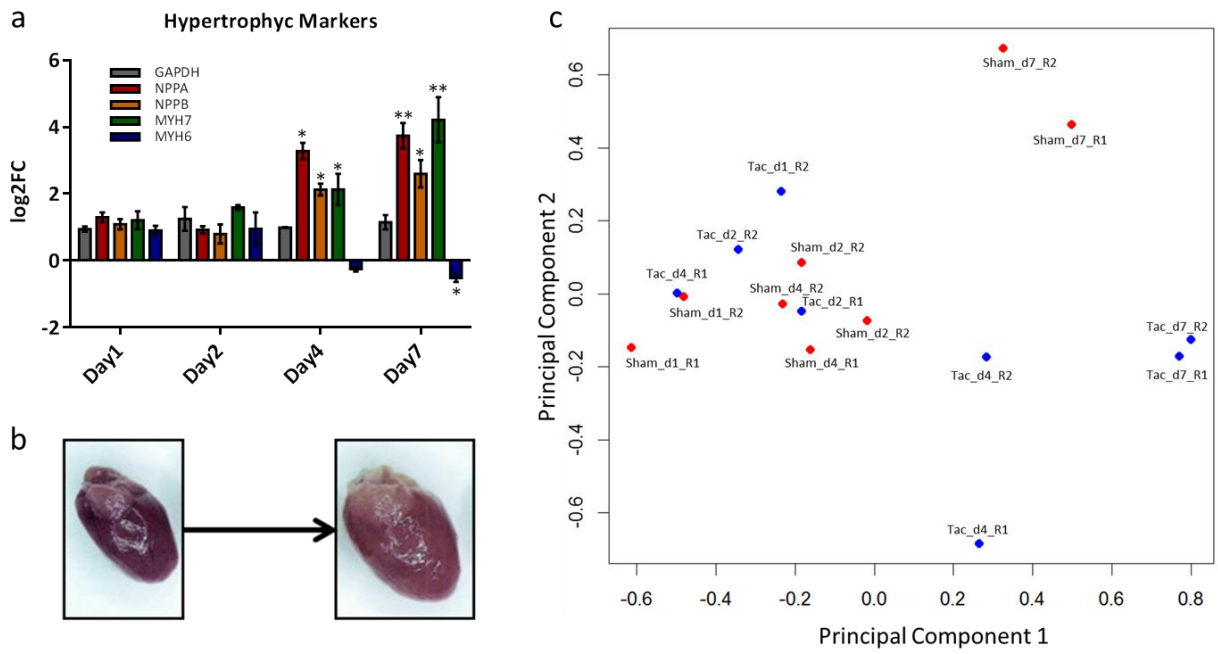


Figure 12. Panel a; RT-PCR analysis performed on hypertrophy markers (Nppa, Nppb, Myh7,Myh6). The picture shows the relative fold change ± SD, *<0.05; **<0.01. Panel b; phenotype of heart after 1 week of constriction. Panel c; PCA plot.

Then we identified lncRNA and protein coding genes expressed and modulated in TAC mice via RNA sequencing and row counts analysis. Massive parallel sequencing was used to retrieve paired-end and strand-specific reads of total RNA libraries of protein coding and lncRNA genes identified with the latest version of genome and annotation from Ensembl, together with edgeR analysis (R Bioconductor) to quantify the gene expression. The differential analysis identified at day1, 230 (156 up, 74 down); day2, 848 (334 up, 514 down); day4, 1348 (968 up, 380 down); day7, 2259(1268 up, 991 down) modulated protein coding genes (Figure 13, Panel a). Regarding lncRNA, the analysis identified at day1, 12 (10 up, 2 down); day2, 28 (6 up, 22 down); day4, 35 (21 up, 14 down); day7, 64 (18 up, 46 down) modulated lncRNAs (Figure 13, Panel b). These data underline a

correlation between the hypertrophic outcome and gene expression reprogramming not only for protein coding genes but also for lncRNA, suggesting that lncRNA could have an important role in the development and progression of cardiac hypertrophy.

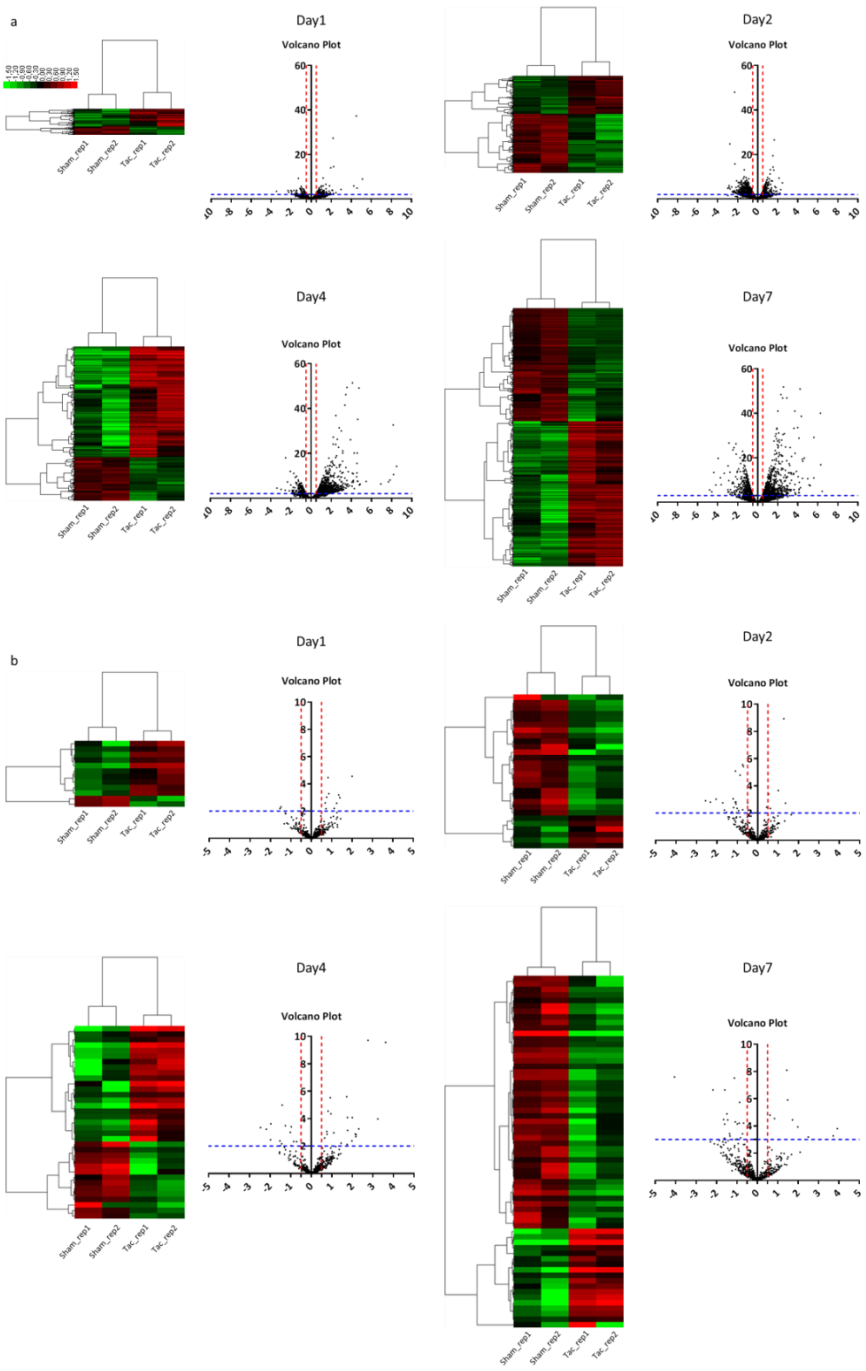


Figure 13. Panel a; Heatmap and Volcano Plot showing the number and level of modulated protein coding genes and, Panel b,. lncRNA.

In order to have information on the biological role of genes found modulated, we performed gene-gene network interaction analysis with IPA (Ingenuity Pathway Analysis). The first two days were characterized by networks for Energy Production and Metabolism (Day 1), as evidenced by up-regulation of Pdk4, a serine/threonine kinase that plays a key role in regulation of glucose and fatty acid metabolism, and thus involved in proliferation via its role in regulating carbohydrate and fatty acid metabolism (Zhao et al., 2008), and Acot2, an acyl-CoA thioesterase regulating intracellular levels of acyl-CoAs, free fatty acids and CoASH (Rame et al., 2011); on the other hand, Cell Death and Survival (Day2) processes were strictly related to cell cycle as indicated by the modulation of Lims2, involved in modulating cell spreading and migration, and Fstl1, involved in the modulation of the action of some growth factors on cell proliferation and differentiation. Next we found an enrichment of cardiac-specific networks (Cardiovascular Disease, Cardiac System Development and Cardiac Hypertrophy) starting from day 4, when the stress and the induced overload became deleterious. The up-regulation of Nppa and Nppb, well-known hypertrophic markers, starts processes related to Cardiovascular Disease. Furthermore we observed a strong down-regulation of Angpt1, a regulator of endothelial cell survival, angiogenesis, migration, proliferation and maintenance of vascular quiescence, but also reorganization of the actin cytoskeleton. It is required for normal heart development and angiogenesis during embryogenesis. Day 7 evidenced a stronger modulation of Nppa, Nppb and Myh7, involved in muscle contraction through interactions with actin-rich filaments that create a contractile force.

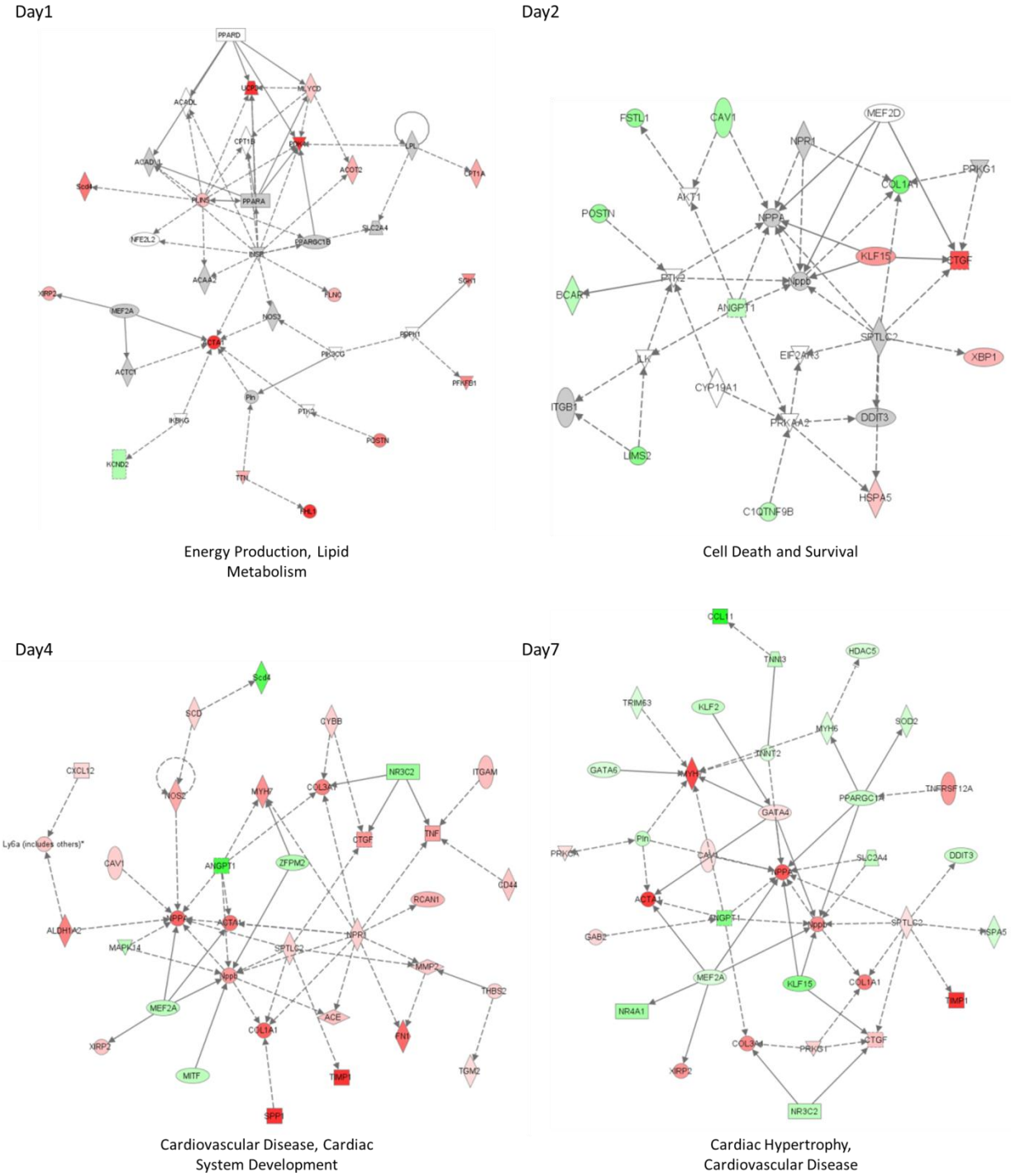


Figure 14. Gene-gene enriched network interaction obtained with IPA software. The colour gradient represent the level of modulation relative to the control. The red gradient is indicative if upregulation whereas the green gradient is indicative of downregulation. For more details about the legend see Material and Methods 2.11.

3.2- lncRNA in Cardiac Hypertrophy

Next we characterized the nature of non-coding RNAs by assessing the mean level of transcription compared to protein coding genes, the conservation score (phastCons) and the coding probability. Figure 15, Panel a top, shows that lncRNAs and novel lncRNAs have a low level of transcription if compared to those of protein coding genes, suggesting that they may be involved in fine mechanisms of regulation. Furthermore, Figure 15, Panel a middle, shows that annotated and novel sequences of lncRNAs are not preserved between species if compared to exons of protein coding genes. An explanation for the low sequence conservation of lncRNAs and novel lncRNAs is that, to maintain their functionality, they do not require much nucleotide sequence conservation. Indeed, while protein coding genes are under very intense selection restraints to maintain correct amino acid coding, and thus their functionality, RNA molecules, that have much less rigid sequence requirements to maintain their secondary structures, only need short stretches of conserved sequences to maintain normal function (Louro et al., 2009, Mercer et al., 2009). Finally, Figure 15, Panel a bottom, shows that, according to their classification, lncRNAs and novel lncRNAs identified have a very low level of coding probability.

Differential analysis identified a total of 108 time-dependent modulated lncRNAs (Figure 15, Panel b and c); the Venn diagram shows that most lncRNAs are time-specific, suggesting that they could have specific roles at each time-point in regulating the change in gene expression.

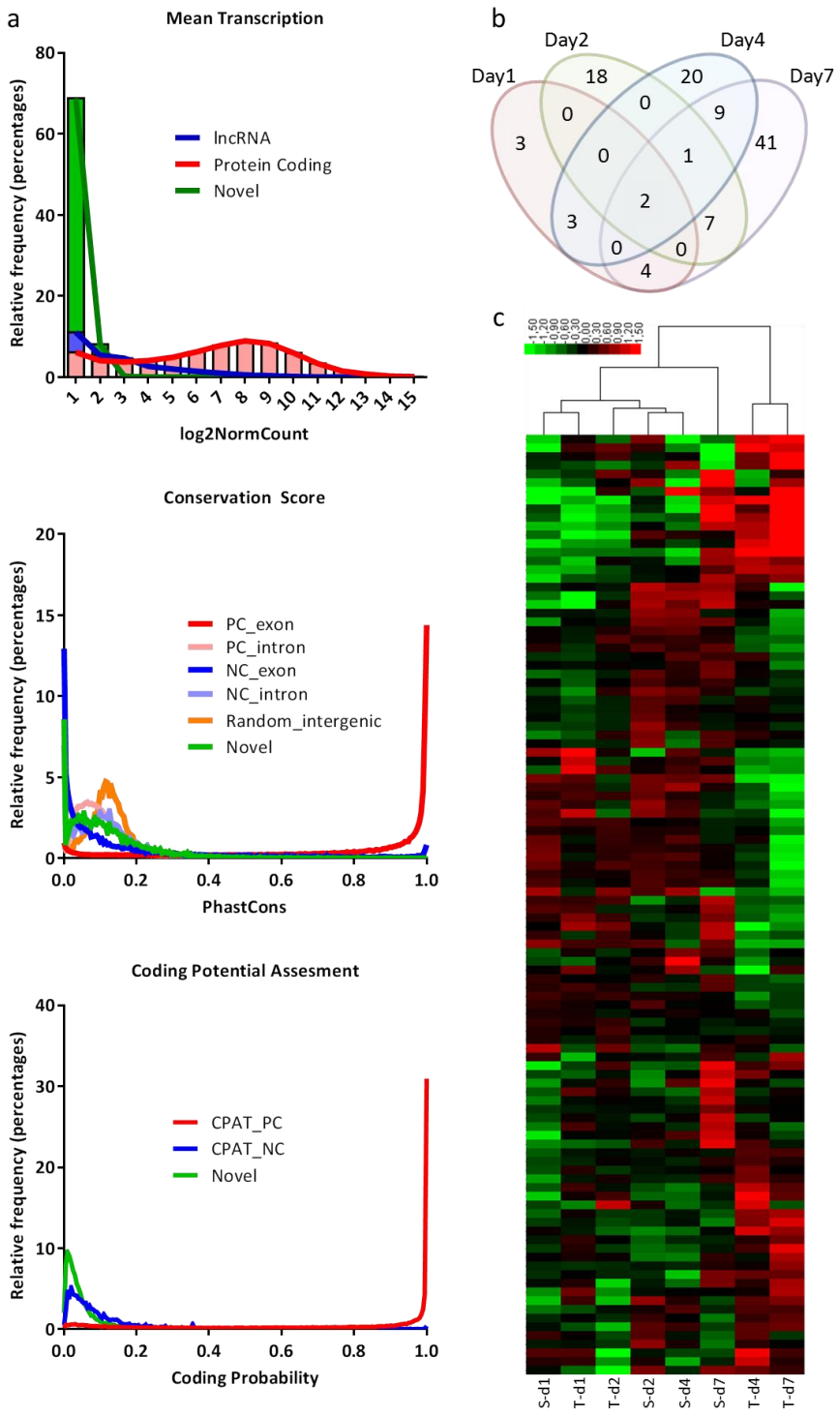


Figure 15. Panel a *top*, Mean transcription of protein coding genes, lncRNAs and novel lncRNAs; middle Phast conservation score with Euarchontoglires; *bottom*, coding probability obtained with CPAT software. Panel b, Venn-Diagram of modulated lncRNAs at each time-point. Panel c, one-way hierarchical cluster analysis of modulated lncRNAs.

According to their genomic position, lncRNAs can regulate the expression of protein coding genes in *cis* (Figure 16, Panel a, b). Different mechanisms of regulation and interaction have been proposed (Figure 2 and Figure 6), but still no global properties have been identified for lncRNAs modulated in cardiac hypertrophy. To do this we correlated the level of modulation of 108 modulated lncRNA with those of the first five protein coding genes identified according to distance from the transcription start site (TSS). Moreover, to discriminate casual transcriptional clusters from a consistent association, we also checked the correlation between modulated protein coding genes and the nearby protein coding genes. Indeed, the maximum area included below the curves of Figure 16 panel d left, the x axis and the value of Pearson correlation of 0.5 is 6.4% of the total. We set this percentage as the cut-off to determine possible positive correlation. Thus, by adopting this threshold, we determined a positive correlation between lncRNA and the first and second protein coding genes (Figure 16, Panel c). Figure 17 recapitulates all the annotated lncRNA modulated across time with a red or green bar indicating the value of fold change relative to the control. For each lncRNA, we also show the first two protein coding genes with their fold change. lncRNA were divided into clusters of regulation to better define and discover functionally associated lncRNA and protein coding genes (Figure 18).

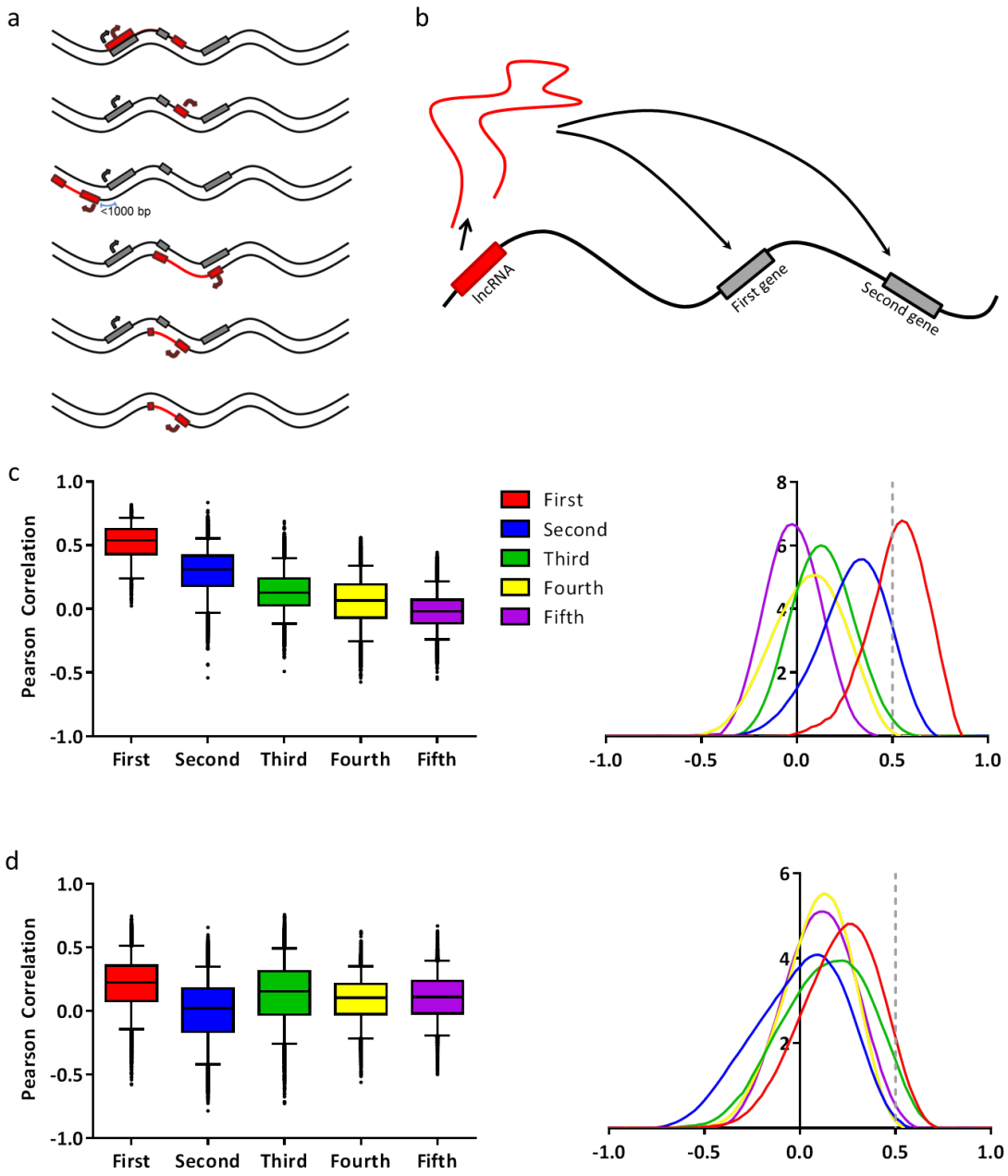


Figure 16. Panel a, genomic position of lncRNAs, in ascending order, sense, intronic, bidirectional, exonic/intronic antisense, intergenic. Panel b, hypothetical way of mechanism of *cis* regulation. Panel c and d, *left*, whisker plot of distribution of Pearson correlation between lncRNA/protein coding genes and first five protein coding genes according to the TSS distance and *right,* relative frequency of Pearson correlation distribution. The percentage of areas included between the curves, x axis and 0.5 cut-off of Pearson correlation are the following: for lncRNA/protein coding association, starting from the first genes, 60%, 12%, 0.8%, 0.08% and 0%. For protein coding/protein coding association are: 6.4%, 0.32%, 4.8%,0.12%,0.8%.

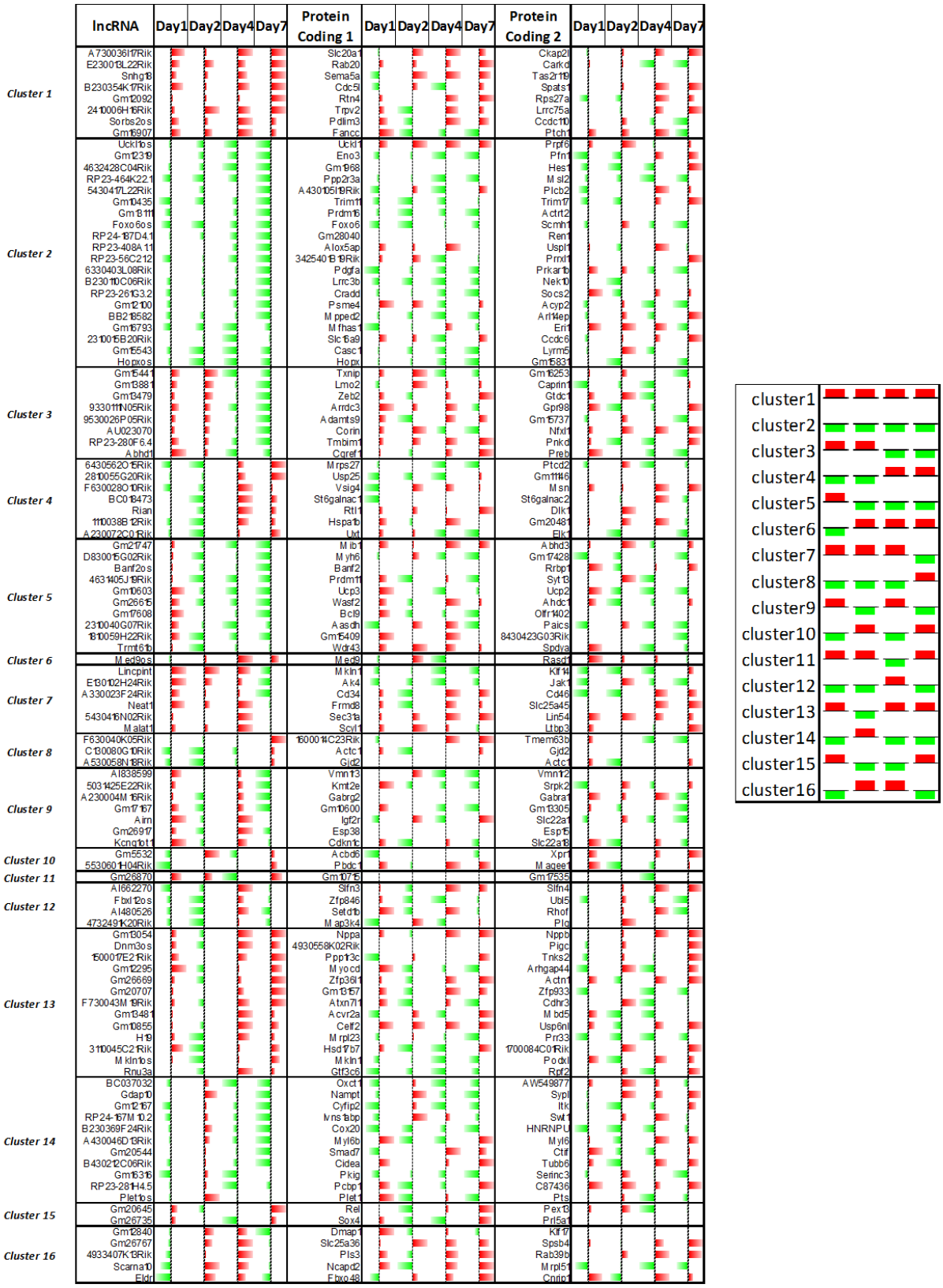


Figure 17. Summary of modulated lncRNA at each time point and first two nearby protein coding genes. Red and green bar represent the level of upregulation or downregulation respectively.

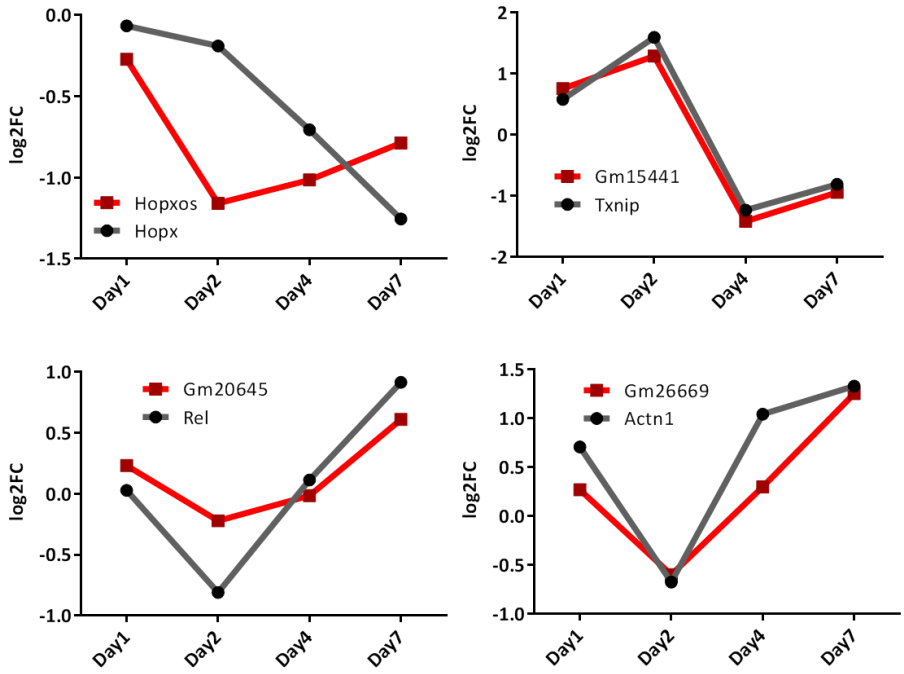


Figure 18.Cardiac related protein coding genes and lncRNAs.

3.2.1-Enriched lncRNA Networks in Cardiac Hypertrophy

The next step was to identify gene-gene network associations to better clarify the role of modulated lncRNAs and the associated protein coding genes.

Ingenuity pathway analysis (Figure 19) was used to determine principal networks at each time-point. Day1 had networks related to Cell Function and Maintenance: the network identified modulation of Ak1, that catalyses the transfer of the terminal phosphate group between ATP and AMP, and thus is involved in cellular energy homeostasis; and Ucp3, a mitochondrial transporter protein that creates proton leaks across the inner mitochondrial, also involved in glucose homeostasis (Schrauwen et al., 2004). Day2 is enriched for Molecular Transport and Lipid Metabolism networks, as suggested by the

modulation of Abdh3, recently found to be a physiologic regulator of medium-chain phospholipids, and thus leading to greater energy expenditure (St-Onge & Jones, 2002, Long et al., 2011).

As for the protein coding genes enriched network, Cardiac-related network started from day 4. Indeed, Corin, a serine-type endopeptidase involved in atrial natriuretic peptide hormone (NPPA) processing, was down-regulated. It has been demonstrated that Corin is a biomarker of hypertrophy (Gladysheva et al., 2013) and its regulation by lncRNA could help the understanding of how this gene is regulated in cardiac hypertrophy. Furthermore day 4 had down-regulation of Txnip, that may act as an oxidative stress mediator. Studies showed that cells that overexpress Txnip develop less hypertrophy after aortic constriction (Yoshioka et al., 2004) and thus the demonstration of a lncRNA-dependent downregulation could improve our understanding of the involved network. Day 7 showed cardiac hypertrophy enrichment but also a re-expression of embryonic development-associated genes. For instance the up-regulation of Igf2r is involved both in embryonic maturation and cardiac hypertrophy. Moreover, Myocd, that plays a crucial role in cardiogenesis and differentiation of the smooth muscle cell lineage, is also an important modulator of cardiac hypertrophy (Kuwahara et al., 2010).

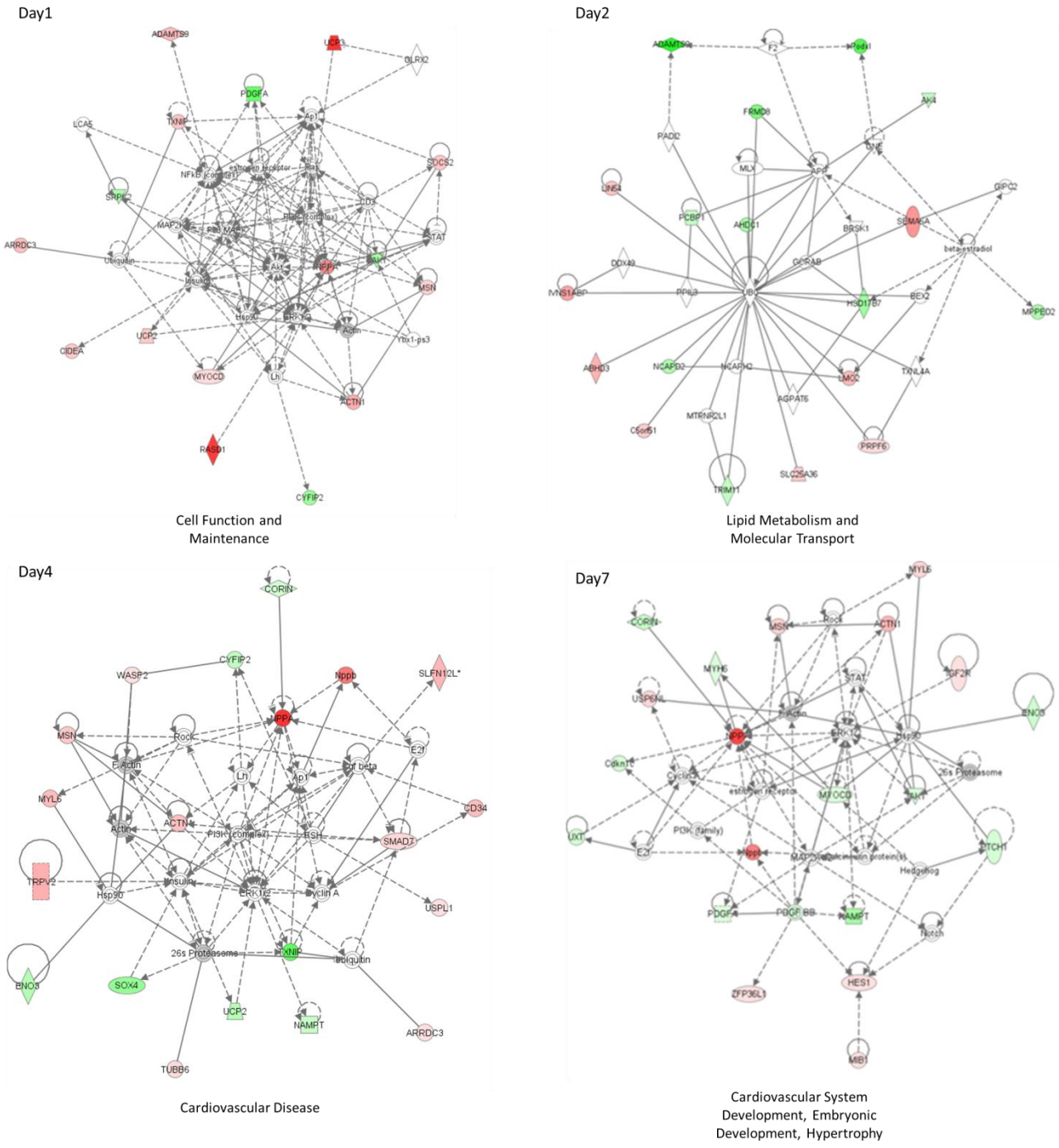


Figure 19. Gene-gene enriched network interaction obtained with IPA software. The colour gradient represents the level of modulation relative to control. The red gradient is indicative of upregulation, whereas the green gradient is indicative of downregulation. For more details on the legend see Material and Methods 2.11.

3.2.2-Enriched lncRNA Biological Processes in Cardiac Hypertrophy

Furthermore we decided to elucidate the involvement of enriched biological process to better define the role of lncRNA and nearby protein coding genes. Analysis of Biological processes (Figure 20) showed a constant increasing of energy and morphological related process such as cardiogenesis, vasculogenesis and contraction of actin cytoskeleton. This is coherent with an increased energy requirement and the remodelling of the heart during cardiac hypertrophy. Furthermore specific biological processes are enriched only at days 4 and 7, such as fibrosis or cardiac hypertrophy. By contrast, a decrease in cardioprotection processes was observed during the development of cardiac hypertrophy, probably due to the increase of stress and hemodynamic overload.

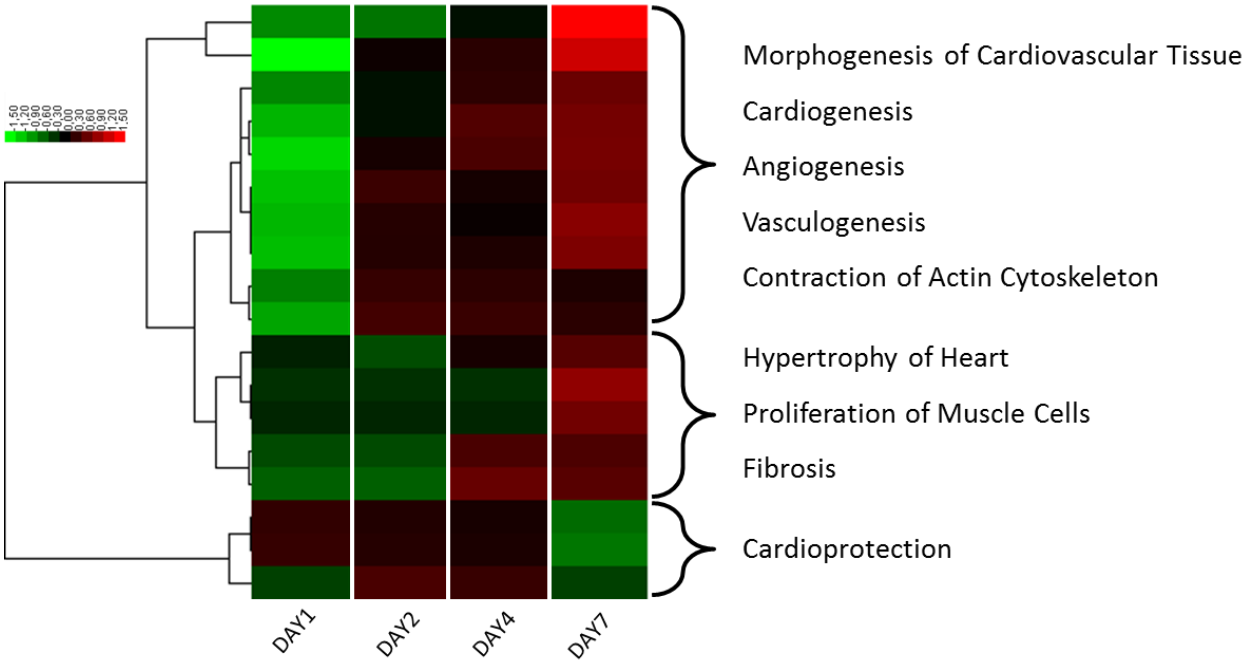


Figure 20. Heatmap of enriched biological processes for each day. Colour gradient represent the increase in significance, from green (less significant) to red (more significant) of the processes.

3.2.3- Cardiac Specific lncRNA

Then we asked which modulated lncRNAs could have a heart- or cardiomyocyte-specific expression pattern. We downloaded RNA-seq expression data from the ENCODE project for different tissues (Figure 21, Panel a). By comparing the expression levels of modulated lncRNAs, it was possible to identify 6 cardiomyocyte-specific lncRNA (for instance Gm13054 and AU023070) and 25 heart-specific lncRNA (for instance Gm12167). These lncRNAs showed time-dependent correlation with the nearby (strongly heart related) genes (Figure 21, Panel b). For example Gm13054 correlated with the expression of NPPA and NPPB, natriuretic peptide (NP) receptors, playing a key role in cardiovascular homeostasis through regulation of natriuresis, diuresis, and vasodilation and inhibition of renin and aldosterone secretion. Moreover AU023070 showed a correlation with Corin, a serine-type endopeptidase involved in atrial natriuretic peptide hormone (NPPA) processing. It converts, through proteolytic cleavage, the non-functional propeptide NPPA into the active hormone, thereby regulating blood pressure in heart and promoting natriuresis, diuresis and vasodilation. This suggests a fine regulatory mechanism that involves gene-gene and non-coding-gene interaction.

Cyfip, nearby Gm12167, is involved in T-cell adhesion and p53/TP53-dependent induction of apoptosis, but its function remains largely unknown and in need of clarification.

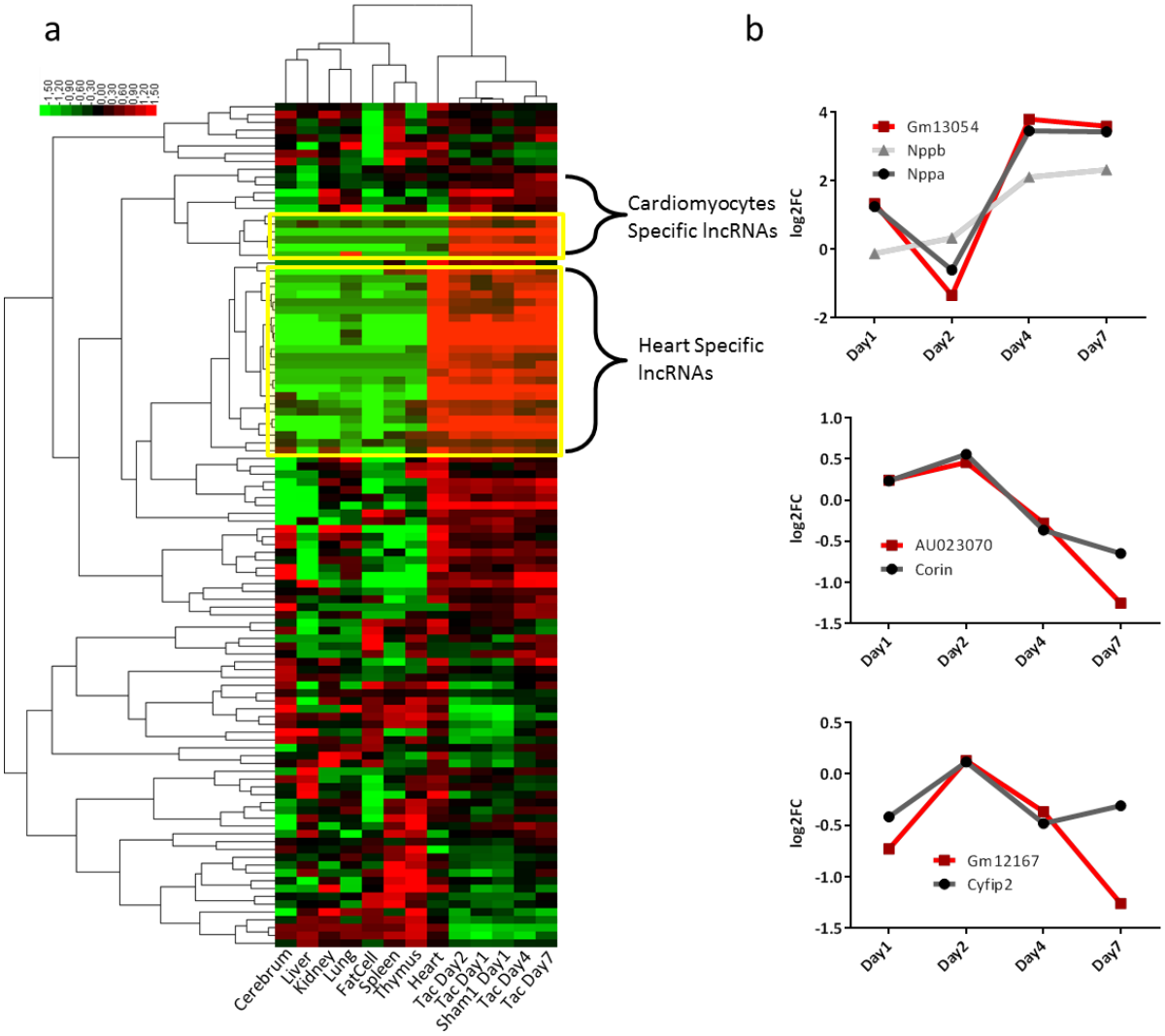


Figure 21. Panel a: Two-way hierarchical cluster of expression values (normalized row counts) of 108 modulated lncRNAs and various tissues downloaded from ENCODE. The yellow box evidence cardiomyocytes and heart specific clusters. Panel b, *top, middle* show two important cardiomyocyte specific lncRNAs and associated protein coding genes with consistent modulation through the time. *bottom*, heart specific lncRNA and associated gene.

3.3- lncRNA Modulated in Embryonic Maturation

Cardiac hypertrophy is characterized by different mechanisms and processes that are activated by overload: for instance, signalling pathways, metabolic network that increase cell size, accumulation and assembly of contractile proteins and, in particular, an increase in the expression of embryonic genes (Hunter & Chien, 1999). Thus, we tried to quantify the disease-induced fetal-like changes in protein coding and lncRNA expression signatures. To do this, we sequenced RNA extracted from embryos after 14.5 days of maturation and performed differential analysis against adult mice (Sham 7 day). Differential analysis showed a huge number (545) of modulated lncRNA, suggesting that they play important roles not only in disease progression but also in specific processes and networks involved in cardiomyocyte maturation (Figure 22, Panel b). Furthermore biological enriched processes were significant with the nearby genes of lncRNAs partially resembling those found enriched for protein coding genes, such as Heart Development and Wnt Receptor Pathway, that regulate important aspects of organogenesis and cell fate determination during embryonic development (Komiya & Habas, 2008)(Figure 22, Panel a, b). The presence of lncRNA in this pathway strongly suggests that they may be fully involved in maturation of embryonic cardiomyocytes and thus plays an important role in these processes.

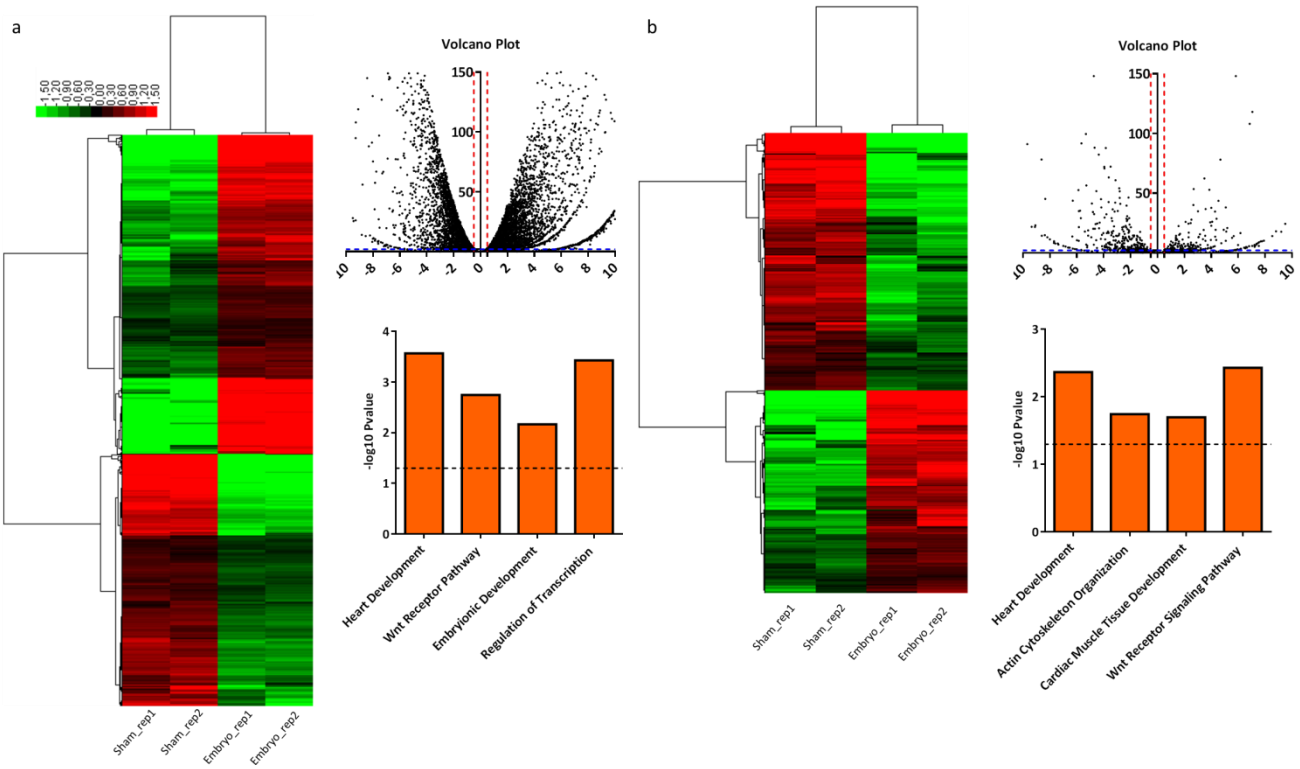


Figure 22.Panel a; two-way hierarchical cluster of expression values (normalized row counts) of modulated protein coding genes, volcano plot and barplot of enriched biological processes. A total of 10293 (5757 upregulated and 4536 downregulated) protein coding genes were found modulated. Panel b; two-way hierarchical cluster of expression values (normalized row counts) of modulated lncRNAs, volcano plot and barplot of enriched biological processes. A total of 545 (239 upregulated and 305 downregulated) protein coding genes were found modulated.

3.3.1-TAC-Induced Fetal-Like Changes

Next, to better define the role of lncRNA induced fetal like changes, we asked which of the upregulated and downregulated protein coding genes and lncRNAs were shared between embryonic and TAC-induced modulation, and which biological process were specific for each condition or shared between them. We found very specific enriched biological processes for embryonic and TAC conditions: Embryogenesis, Wnt pathway, Genomic imprinting and Muscle Morphology, Cardiac Tissue Morphogenesis (Figure 23). On the other hand, modulation of shared lncRNA enriched process involved both in embryonic maturation and in cardiac hypertrophy, such as Cell Proliferation and Cardiovascular Development. The same phenomena was observed for protein coding genes (Figure 23).

Figure 24 summarizes lncRNA modulated both in embryonic and TAC condition with the nearby protein coding genes and the level of modulation. Furthermore panel b shows lncRNAs and probable nearby associated protein coding genes that resemble embryonic modulation in TAC-mice. Sema5a, which increases endothelial cell proliferation and migration, may promote angiogenesis and inactivation of this gene and results in embryonic lethality (Fiore et al., 2005); Eno3 has a function in striated muscle development and regeneration (Shi & Garry, 2006).

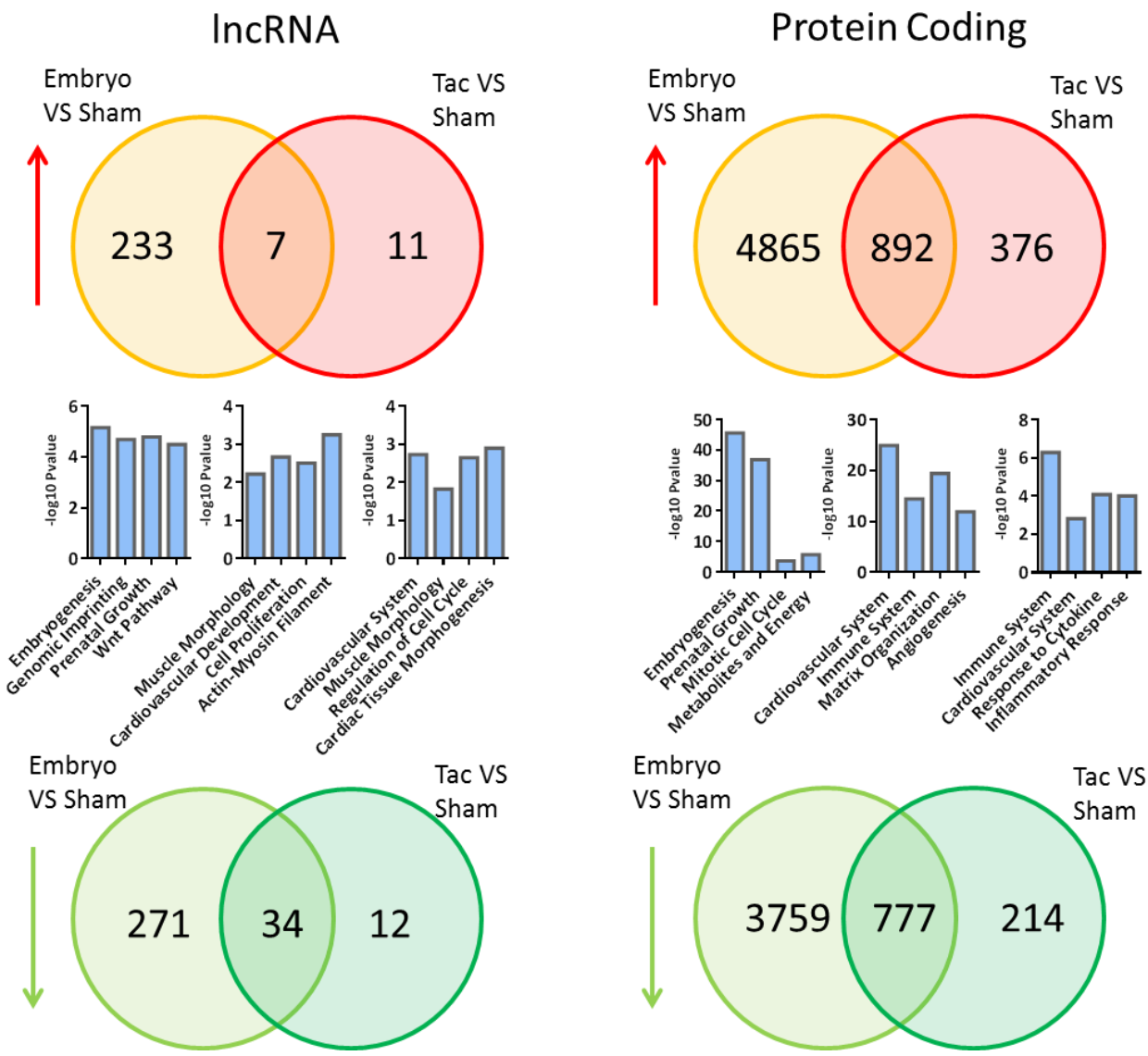
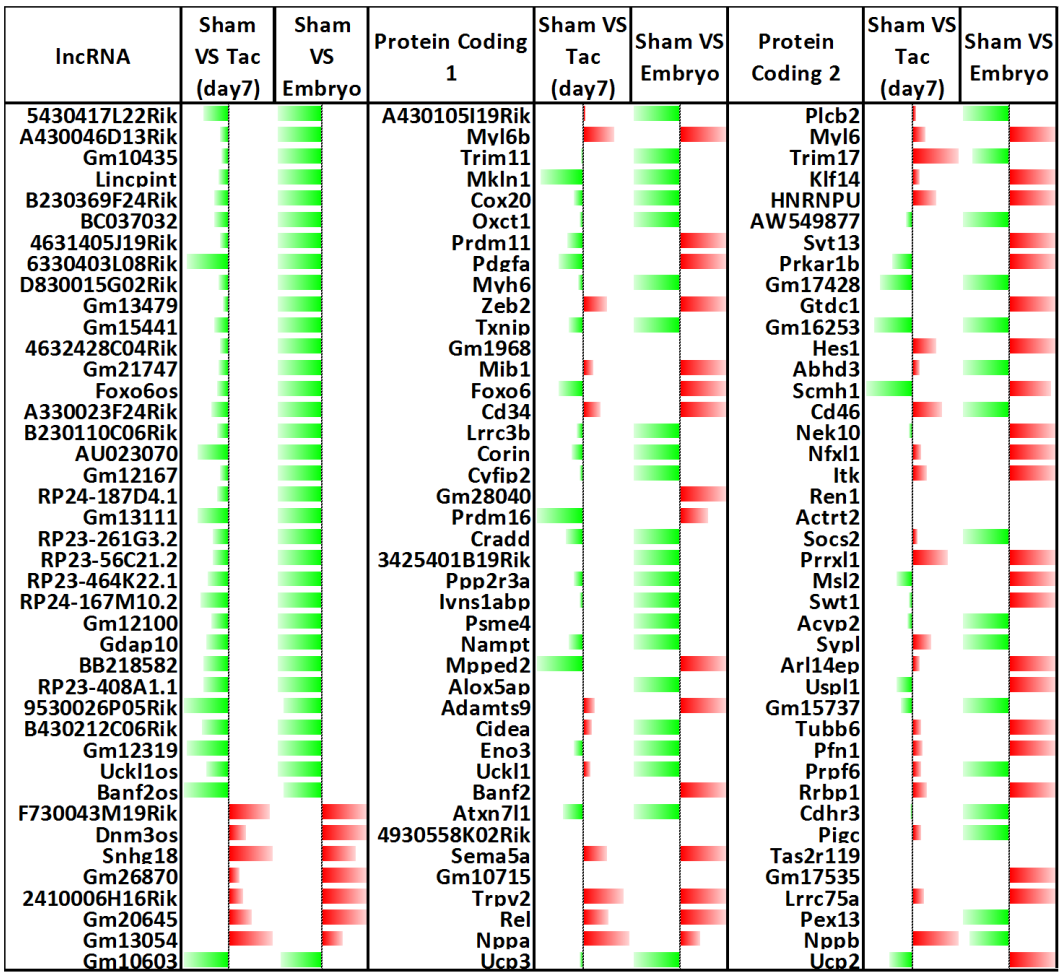


Figure 23. *left,* Venn diagram of upregulated (up) or downregulated (down) lncRNAs in Embryonic/Adult and Tac/Sham comparison. Bar plot shows enriched biological processes for specific Embryonic lncRNAs, common lncRNAs and TAC specific lncRNAs respectively. *Right*, Venn diagram of upregulated or downregulated protein coding genes in Embryonic/Adult and Tac/Sham comparison. Bar plot show enriched biological process for specific Embryonic, common and TAC specific protein coding genes, respectively.

a



b

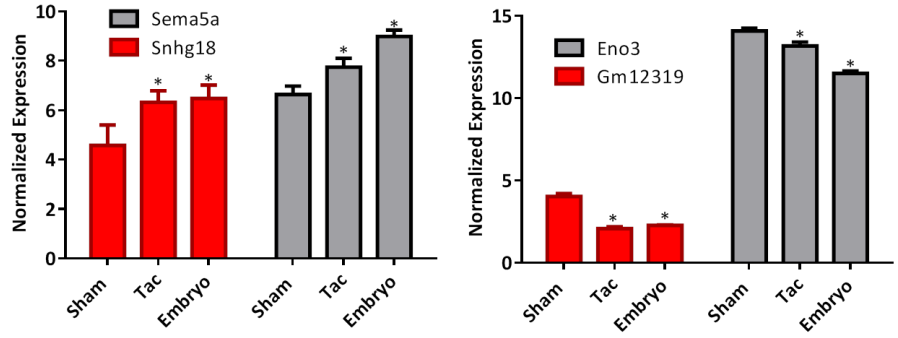


Figure 24. Panel a: summary of TAC induced fetal like changes lncRNA and first two nearby protein coding genes. Red and green bar represent the level of upregulation or downregulation respectively. Panel b: lncRNA and nearby protein coding genes that resemble embryonic modulation in TAC mice.

3.4- Motifs Discovery

To identify the transcription network within which the regulated promoters act to govern gene expression in TAC cardiomyocytes, we searched for the presence of transcription factor-binding motifs within -2000 kb from the TSS. RSAT (Regulatory Sequence Analysis Tools) identified conservative stretches of nucleotides shared between sequences recognized by transcription factors. The analysis revealed that promoter regions of protein coding genes are enriched for SP1 and KLF family transcription factors, both for TAC- and maturation-induced genes. Sp1 is a transcription factor that can activate or repress transcription in response to pathological stimuli and is involved in many processes such as apoptosis, cell growth, differentiation and immune responses. These data suggest that TAC-induced fetal-like reprogramming involves also transcription factor networks. On the other hand, specific transcription factors are enriched in promoter regions for modulated TAC or embryonic protein coding genes. Indeed, Mef2a and Mef2c transcription factors are induced only in TAC-modulated genes, suggesting that they may have a specific role in the development of hypertrophy (Passier et al., 2000)(Figure 26, Panel b). Finally we validated the binding of SP1 to promoters of hypertrophy-modulated genes after 1 week of TAC surgery. (Figure 26, Panel c)

Next we checked which transcription factors were present at the promoter region of lncRNAs. The analysis showed that lncRNAs are mostly regulated by their own transcription factor (in particular FOX related), that plays important roles in regulating the expression of genes involved in cell growth, proliferation, differentiation, and longevity. Fox related transcription factor seems to be specific for TAC-induce lncRNA.

Common transcription factor belong to the family of Arnt, that may form heterodimers with Ahr and Hif1a. Arnt::Ahr complex is involved in cell-cycle regulation and plays an important role in the development and maturation of many tissues. Arnt::Hif1a complex regulates the adaptive response to hypoxia and maintains biological homeostasis, influences cell metabolism, survival and angiogenesis (Alkon et al., 2012). Finally, embryonic-specific transcription factors belonged to the E2F family: these transcription factors are involved in cell cycle regulation and synthesis of DNA in mammalian cells (Figure 26, Panel a).

a

Transcription Factor	Sham Vs Tac	Sham Vs Embryo
FOXD3	√	X
FOXP1	√	X
FOXO3	√	X
SP1	√	√
KLF5	√	X
ARNT::AHR	√	√
NRF1	√	X
HIF1A::ARNT	√	√
E2F1	X	√
E2F4	X	√
NFATC2	X	√
SOX5	X	√

b

Transcription Factor	Sham Vs Tac	Sham Vs Embryo
SP1	√	√
KLF4	√	X
KLF5	√	√
SRY	√	√
SOX5	√	X
ARID3A	√	X
MEF2C	√	X
MEF2A	√	X
NFATC2	X	√
SOX3	X	√
SOX6	X	√

c

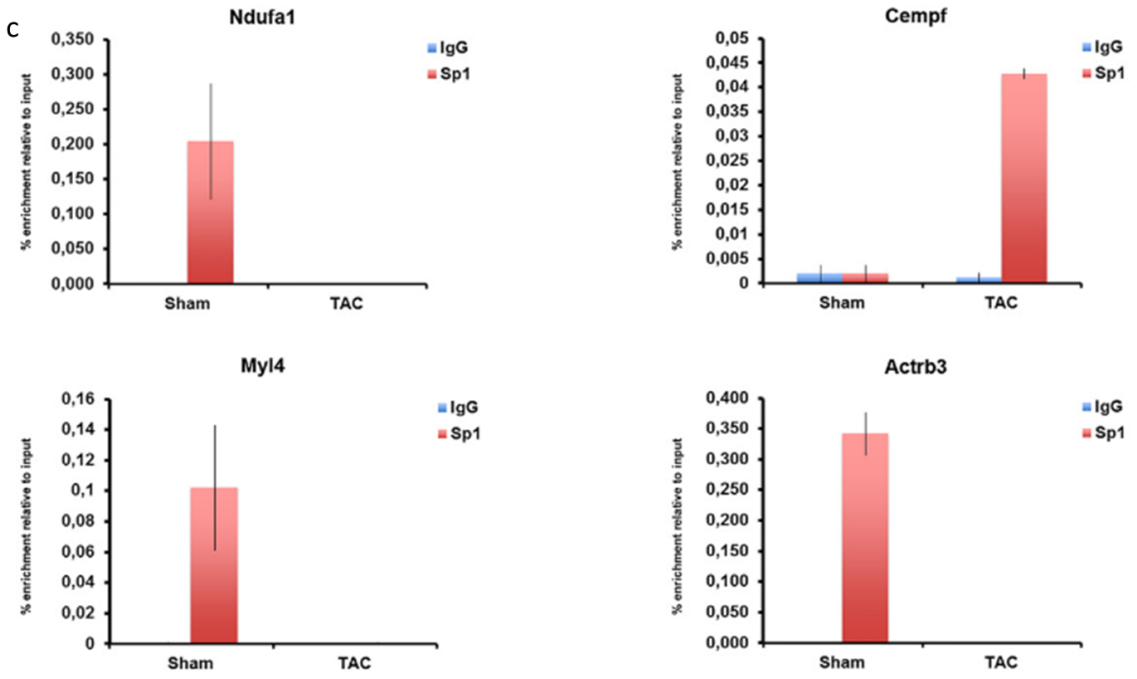


Figure 25. Panel a: enriched transcription factor binding motifs in modulated lncRNA promoters. Panel b: enriched transcription factor binding motifs in modulated protein coding genes promoters. Panel c. validation of Sp1 binding site in promoters of selected genes (Ndufa1, Cempf, Actrb3, Myl4).

3.5- Novel lncRNA

Due to the difficulty in the discovery and identification of lncRNA, also the most recent version of genome and annotation may lack a complete definition of the lncRNA landscape. In particular this gap could be due to time-specific or tissue-specific expression of non-coding RNA. Thus, novel lncRNAs could play important and tissue/time-specific roles in a wide range of biological processes. To identify novel lncRNA we looked for islands of transcription in genomic regions lacking annotation. Then, different filters were applied to discriminate potential novel lncRNA from transcriptional noise. Each region was filtered according to their coding potential, number of exon, and to drastically reduce the number of false positive, we called enriched peaks of histone modification (see Material and Methods) between the condition and determined in this way which transcribed regions were also epigenetically regulated. We performed this analysis only at day 7, when we found the most gene expression reprogramming according to the hypertrophic response. Our analysis revealed 155 modulated novel lncRNA (68 upregulated and 87 downregulated). Gene-gene interaction analysis showed an enrichment of Cardiac Hypertrophy and Cardiac Disease networks, consistent with day 7 hypertrophic outcome. Interestingly, we found new lncRNAs nearby Ankrd1 - a nuclear transcription factor that negatively regulates the expression of cardiac genes - and HOPX, required to modulate cardiac growth and development. The enriched biological processes were perfectly coherent with the outcome of hypertrophy after seven days of constriction. Indeed we found process related to energy and metabolism, to morphology and, in particular, hypertrophy-related

processes. Furthermore, motifs analysis revealed some binding sites of the annotated lncRNAs. Indeed Fox family binding sites were found enriched in the promoter regions of novel lncRNA.

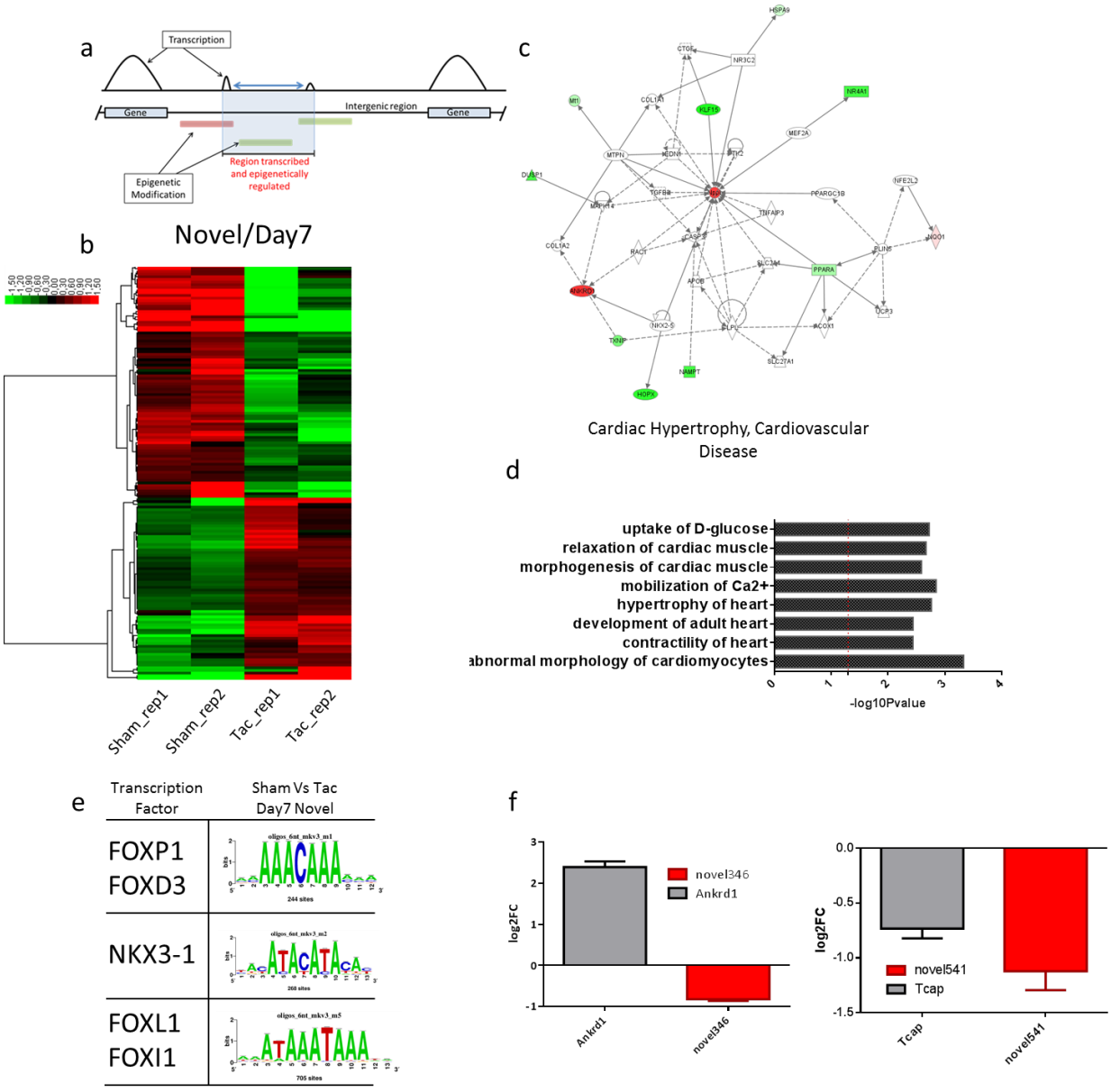


Figure 26. Panel a. Principle with it has been possible identify novel lncRNA transcribed and epigenetically regulated. Panel b: One-way hierarchical cluster of 158 row counts values of modulated novel lncRNA. Panel c: Network analysis. Panel d: enriched biological processes. Panel e: enriched binding motifs for promoter regions of novel lncRNAs. Panel f: novel lncRNAs that may regulates the expression of important cardiac related genes.

DISCUSSION AND CONCLUSION

Massive parallel, strand-specific and paired-end sequencing was used to identify modulated protein coding genes and modulated lncRNA after different time-points of hypertrophic stimuli. A number of modulated protein coding genes were strongly enhanced through time in response to the induced overload. We observed an enrichment of Energy Production and Metabolism and Cell Death and Survival networks in the first and second days, respectively. Important and specific gene networks were regulated at each day: for instance Pdk4, a regulator of glucose and fatty acid metabolism, and Acot2, an acyl-CoA thioesterase that regulates intracellular levels of acyl-CoAs, free fatty acids and CoASH, are modulated at day1; while Lims2, involved in modulating cell spreading and migration, and Fstl1, involved in cell proliferation and differentiation, are modulated at day 2. Cardiac specific networks become significant starting from day 4: Cardiac System Development, evidenced by upregulation of hypertrophic markers (Nppa, Nppb) and downregulation of Angpt1, a regulator of endothelial cell survival, angiogenesis, migration, proliferation and maintenance of vascular quiescence. TAC induced mice after 7 days showed a stronger upregulation of Nppa, Nppb and Myh7 that determine the hypertrophic response.

Next we defined the signature of lncRNA modulated during the development of cardiac hypertrophy to better define the transcriptional landscape of this pathology. By assessing the mean level of transcription of lncRNA compared to protein coding genes, the conservation score (phastCons) and the coding probability, we characterized the nature of non-coding RNAs. Low level transcription of lncRNA relative to protein coding genes suggests that they may be involved in fine mechanisms of regulation. Furthermore both annotated and novel sequences of lncRNAs are not preserved between species as

they do not require much nucleotide sequence conservation to maintain their functionality. We show that most of lncRNAs are time specific, suggesting that they could have specific roles at each time point in regulating the change in gene expression.

Furthermore we demonstrate a global correlation between the expression of lncRNA and the first two nearby genes. These genes were then used to perform gene-gene network analysis and biological process enrichment. As for protein coding genes, lncRNA are enriched for cardiac specific networks, starting from day4 with down-regulation of Corin, a serine-type endopeptidase involved in atrial natriuretic peptide hormone (NPPA) processing. The regulation of Corin by lncRNA could help to understanding how this gene is regulated in cardiac hypertrophy. After 4 days from TAC is also showed downregulation of Txnip. Cells that overexpress Txnip develop less hypertrophy after aortic constriction and thus the demonstration of a lncRNA-dependent downregulation could improve our understanding of the involved network. We observed both cardiac hypertrophy enrichment and re-expression of embryonic development associated genes after 7 days from TAC. Upregulation of Igf2r is involved both in embryonic maturation and cardiac hypertrophy. Moreover, Myocd, which plays a crucial role in cardiogenesis and differentiation of the smooth muscle cell lineage, is also an important modulator of cardiac hypertrophy.

To better define the role of lncRNA and nearby protein coding genes, we performed gene ontology analysis to find out which biological processes were enriched. There was a constant increase of energy and morphological related processes, such as cardiogenesis, vasculogenesis and contraction of actin cytoskeleton. This suggests an important role of lncRNA in increased energy requirement and in the remodelling of the heart during

hypertrophy. Furthermore specific biological processes are enriched only at days 4 and 7, such as fibrosis or cardiac hypertrophy.

By comparing the expression levels of modulated lncRNA in TAC mice with those obtained from different tissues downloaded from the ENCODE project, we identified 6 cardiomyocyte-specific lncRNA (for instance Gm13054 and AU023070) and 25 heart-specific lncRNA (for instance Gm12167). Moreover some of these lncRNAs show time-dependent correlation with (strongly heart related) nearby genes: for instance Gm13054 correlates with the expression of NPPA and NPPB, natriuretic peptide (NP) receptors playing a key role in cardiovascular homeostasis through regulation of natriuresis, diuresis, and vasodilation and inhibition of renin and aldosterone secretion respectively; AU023070 shows a correlation with Corin, a serine-type endopeptidase involved in atrial natriuretic peptide hormone (NPPA) processing. It converts, through proteolytic cleavage, the non-functional propeptide NPPA into the active hormone, thereby regulating blood pressure in heart and promoting natriuresis, diuresis and vasodilation. These data suggest that lncRNA may be expressed in specific tissues and time-points to regulate the stress response.

As cardiac hypertrophy is characterized by the re-expression of embryonic genes first we decided to sequence RNA extracted from embryos after 14.5 days of maturation and performed differential analysis against adult mice (Sham 7-day) and compared the modulated lncRNA and protein coding genes with those modulated in TAC mice to find out which of these elements could play a role in maturation and disease.

As we found a huge number of modulated lncRNA (545), we hypothesize that they play important roles not only in disease progression but also in specific processes and

networks involved in cardiomyocyte maturation. Furthermore, Heart Development, which regulates important aspects of organogenesis and cell fate determination during embryonic development, is biological enriched process both for protein coding genes and lncRNA. The presence of lncRNA in the Wnt Receptor pathway strongly suggests that they may be fully involved in maturation of embryonic cardiomyocytes and thus play and important role in this process.

Next, we defined the role of lncRNA inducing fetal-like changes. We found specific enriched biological processes in embryonic and TAC conditions such as Embryogenesis, Wnt pathway, Genomic imprinting and Muscle Morphology, Cardiac Tissue Morphogenesis,. On the other hand, modulation of shared lncRNA is enriched in processes involved both in embryonic maturation and in cardiac hypertrophy, such as Cell Proliferation and Cardiovascular Development. Sema5a, that may promote angiogenesis, and Eno3, that has a function in striated muscle development and regeneration, are two important genes nearby lncRNA modulated both in embryos and the TAC condition.

To identify the transcription network within which the regulated promoters acted to govern gene expression in TAC cardiomyocytes, we searched for the presence of transcription factor-binding motifs within -2000 kb from the TSS. We found specific transcription factors for promoters of protein coding genes and lncRNA.

Promoter regions of protein coding genes are enriched for SP1, a factor that can activate or repress transcription in response to pathological stimuli and for KLF, involved in many processes such as apoptosis, cell growth, differentiation and immune responses, both for TAC- and maturation-induced genes. On the other hand, Mef2a and Mef2c transcription

factors are induced only in TAC-modulated genes, suggesting that they may have a specific role in the development of hypertrophy.

lncRNA promoter regions are mostly regulated by their own transcription factor (in particular FOX related) that plays important roles in regulating the expression of genes involved in cell growth, proliferation, differentiation and longevity. Fox related transcription factor seems to be specific for TAC-induced lncRNA.

Transcription factors shared between maturation and TAC mice belonged to the Arnt family, that may form heterodimers with Ahr and Hif1a. Arnt::Ahr complex is involved in cell-cycle regulation and plays an important role in the development and maturation of many tissues. Arnt::Hif1a complex regulates the adaptive response to hypoxia and maintains biological homeostasis, and influences cell metabolism, survival and angiogenesis.

Novel lncRNA could play important and tissue/time-specific roles in a wide range of biological processes. We identify novel lncRNA by intersecting transcribed regions with those that were also epigenetically regulated.

We found 155 modulated novel lncRNA (68 upregulated and 87 downregulated). Gene-gene interaction analysis showed an enrichment of Cardiac hypertrophy and Cardiac Disease networks, consistent with day 7 hypertrophic outcome. Interestingly, we found new lncRNAs nearby two important genes: Ankrd1, a nuclear transcription factor that negatively regulates the expression of cardiac genes; and HOPX, required to modulate cardiac growth and development.

Furthermore, motifs analysis revealed some binding sites of the annotated lncRNAs. Indeed Fox family binding sites were found enriched in the promoter regions of novel lncRNA.

Although these results do not elucidate the molecular mechanisms by which lncRNAs are involved in cardiac hypertrophy and heart development, the identification of a lot lncRNAs modulated in these biological processes suggests an important role of these regulatory molecules in heart physiology and pathophysiology.

REFERENCES

Ahmad T, Fiuzat M, Felker GM, O'connor C, 2012. Novel biomarkers in chronic heart failure. *Nat Rev Cardiol* **9**, 347-59.

Alkon J, Friedberg MK, Manlhiot C*, et al.*, 2012. Genetic variations in hypoxia response genes influence hypertrophic cardiomyopathy phenotype. *Pediatr Res* **72**, 583-92.

Anders S, Pyl PT, Huber W, 2014. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics*.

Avner P, Heard E, 2001. X-chromosome inactivation: counting, choice and initiation. *Nat Rev Genet* **2**, 59-67.

Batista PJ, Chang HY, 2013. Long noncoding RNAs: cellular address codes in development and disease. *Cell* **152**, 1298-307.

Birney E, Stamatoyannopoulos JA, Dutta A*, et al.*, 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816.

Black JC, Van Rechem C, Whetstine JR, 2012. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol Cell* **48**, 491-507.

Blow MJ, Mcculley DJ, Li Z*, et al.*, 2010. ChIP-Seq identification of weakly conserved heart enhancers. *Nat Genet* **42**, 806-10.

Bond CS, Fox AH, 2009. Paraspeckles: nuclear bodies built on long noncoding RNA. *J Cell Biol* **186**, 637-44.

Cabili MN, Trapnell C, Goff L*, et al.*, 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **25**, 1915-27.

Care A, Catalucci D, Felicetti F*, et al.*, 2007. MicroRNA-133 controls cardiac hypertrophy. *Nat Med* **13**, 613-8.

Cesana M, Cacchiarelli D, Legnini I*, et al.*, 2011. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147**, 358-69.

Collins K, 2008. Physiological assembly and activity of human telomerase complexes. *Mech Ageing Dev* **129**, 91-8.

Conerly M, Grady WM, 2010. Insights into the role of DNA methylation in disease through the use of mouse models. *Dis Model Mech* **3**, 290-7.

Dealmeida AC, Van Oort RJ, Wehrens XH, 2010. Transverse aortic constriction in mice. *J Vis Exp*.

Derrien T, Johnson R, Bussotti G*, et al.*, 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* **22**, 1775-89.

Deuve JL, Avner P, 2011. The coupling of X-chromosome inactivation to pluripotency. *Annu Rev Cell Dev Biol* **27**, 611-29.

Encode, 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74.

Feng J, Liu T, Qin B, Zhang Y, Liu XS, 2012. Identifying ChIP-seq enrichment using MACS. *Nat Protoc* **7**, 1728-40.

Fiore R, Rahim B, Christoffels VM, Moorman AF, Puschel AW, 2005. Inactivation of the Sema5a gene results in embryonic lethality and defective remodeling of the cranial vascular system. *Mol Cell Biol* **25**, 2310-9.

Fischle W, Wang Y, Allis CD, 2003. Histone and chromatin cross-talk. *Curr Opin Cell Biol* **15**, 172-83.

Gladysheva IP, Wang D, Mcnamee RA*, et al.*, 2013. Corin overexpression improves cardiac function, heart failure, and survival in mice with dilated cardiomyopathy. *Hypertension* **61**, 327-32.

Grote P, Wittler L, Hendrix D*, et al.*, 2013. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev Cell* **24**, 206-14.

Groth A, Corpet A, Cook AJ*, et al.*, 2007. Regulation of replication fork progression through histone supply and demand. *Science* **318**, 1928-31.

Guttman M, Donaghey J, Carey BW*, et al.*, 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295-300.

Han P, Hang CT, Yang J, Chang CP, 2011. Chromatin remodeling in cardiovascular development and physiology. *Circ Res* **108**, 378-96.

Heard E, Rougeulle C, Arnaud D, Avner P, Allis CD, Spector DL, 2001. Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell* **107**, 727-38.

Hu W, Alvarez-Dominguez JR, Lodish HF, 2012. Regulation of mammalian cell differentiation by long non-coding RNAs. *EMBO Rep* **13**, 971-83.

Huang J, Fan T, Yan Q*, et al.*, 2004. Lsh, an epigenetic guardian of repetitive elements. *Nucleic Acids Res* **32**, 5019-28.

Hunter JJ, Chien KR, 1999. Signaling pathways for cardiac hypertrophy and failure. *N Engl J Med* **341**, 1276-83.

Hutchinson JN, Ensminger AW, Clemson CM, Lynch CR, Lawrence JB, Chess A, 2007. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* **8**, 39.

Ishii N, Ozaki K, Sato H*, et al.*, 2006. Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction. *J Hum Genet* **51**, 1087-99.

Izumo S, Nadal-Ginard B, Mahdavi V, 1988. Protooncogene induction and reprogramming of cardiac gene expression produced by pressure overload. *Proc Natl Acad Sci U S A* **85**, 339-43.

Kapranov P, Cheng J, Dike S*, et al.*, 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484-8.

Kim SK, Jung I, Lee H*, et al.*, 2012. Human histone H3K79 methyltransferase DOT1L protein [corrected] binds actively transcribing RNA polymerase II to regulate gene expression. *J Biol Chem* **287**, 39698-709.

Klattenhoff CA, Scheuermann JC, Surface LE*, et al.*, 2013. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* **152**, 570-83.

Komiya Y, Habas R, 2008. Wnt signal transduction pathways. *Organogenesis* **4**, 68-75.

Koren MJ, Devereux RB, Casale PN, Savage DD, Laragh JH, 1991. Relation of left ventricular mass and geometry to morbidity and mortality in uncomplicated essential hypertension. *Ann Intern Med* **114**, 345-52.

Korostowski L, Sedlak N, Engel N, 2012. The Kcnq1ot1 long non-coding RNA affects chromatin conformation and expression of Kcnq1, but does not regulate its imprinting in the developing heart. *PLoS Genet* **8**, e1002956.

Kouzarides T, 2007. Chromatin modifications and their function. *Cell* **128**, 693-705.

Kung JT, Colognori D, Lee JT, 2013. Long noncoding RNAs: past, present, and future. *Genetics* **193**, 651-69.

Kuwahara K, Kinoshita H, Kuwabara Y*, et al.*, 2010. Myocardin-related transcription factor A is a common mediator of mechanical stress- and neurohumoral stimulation-induced cardiac hypertrophic signaling leading to activation of brain natriuretic peptide gene expression. *Mol Cell Biol* **30**, 4134-48.

Langmead B, Trapnell C, Pop M, Salzberg SL, 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25.

Latronico MV, Condorelli G, 2009. MicroRNAs and cardiac pathology. *Nat Rev Cardiol* **6**, 419-29.

Lee JH, Gao C, Peng G*, et al.*, 2011. Analysis of transcriptome complexity through RNA sequencing in normal and failing murine hearts. *Circ Res* **109**, 1332-41.

Legnini I, Morlando M, Mangiavacchi A, Fatica A, Bozzoni I, 2014. A feedforward regulatory loop between HuR and the long noncoding RNA linc-MD1 controls early phases of myogenesis. *Mol Cell* **53**, 506-14.

Levy D, Garrison RJ, Savage DD, Kannel WB, Castelli WP, 1990. Prognostic implications of echocardiographically determined left ventricular mass in the Framingham Heart Study. *N Engl J Med* **322**, 1561-6.

Li H, Handsaker B, Wysoker A*, et al.*, 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9.

Lin M, Pedrosa E, Shah A*, et al.*, 2011a. RNA-Seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. *PLoS One* **6**, e23356.

Lin R, Roychowdhury-Saha M, Black C*, et al.*, 2011b. Control of RNA processing by a large non-coding RNA over-expressed in carcinomas. *FEBS Lett* **585**, 671-6.

Lompre AM, Schwartz K, D'albis A, Lacombe G, Van Thiem N, Swynghedauw B, 1979. Myosin isoenzyme redistribution in chronic heart overload. *Nature* **282**, 105-7.

Long JZ, Cisar JS, Milliken D*, et al.*, 2011. Metabolomics annotates ABHD3 as a physiologic regulator of medium-chain phospholipids. *Nat Chem Biol* **7**, 763-5.

Louch WE, Sheehan KA, Wolska BM, 2011. Methods in cardiomyocyte isolation, culture, and gene transfer. *J Mol Cell Cardiol* **51**, 288-98.

Louro R, Smirnova AS, Verjovski-Almeida S, 2009. Long intronic noncoding RNA transcription: expression noise or expression choice? *Genomics* **93**, 291-8.

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ, 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-60.

Mcpherson R, Pertsemlidis A, Kavaslar N*, et al.*, 2007. A common allele on chromosome 9 associated with coronary heart disease. *Science* **316**, 1488-91.

Meerson FZ, 1961. On the mechanism of compensatory hyperfunction and insufficiency of the heart. *Cor Vasa* **3**, 161-77.

Mercer TR, Dinger ME, Mattick JS, 2009. Long non-coding RNAs: insights into functions. *Nat Rev Genet* **10**, 155-9.

Mitra SA, Mitra AP, Triche TJ, 2012. A central role for long non-coding RNA in cancer. *Front Genet* **3**, 17.

Mohammad F, Mondal T, Kanduri C, 2009. Epigenetics of imprinted long noncoding RNAs. *Epigenetics* **4**, 277-86.

Orom UA, Derrien T, Beringer M*, et al.*, 2010. Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46-58.

Pandey RR, Mondal T, Mohammad F*, et al.*, 2008. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* **32**, 232-46.

Passier R, Zeng H, Frey N*, et al.*, 2000. CaM kinase signaling induces cardiac hypertrophy and activates the MEF2 transcription factor in vivo. *J Clin Invest* **105**, 1395-406.

Peters AH, Kubicek S, Mechtler K*, et al.*, 2003. Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol Cell* **12**, 1577-89.

Qureshi IA, Mehler MF, 2012. Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. *Nat Rev Neurosci* **13**, 528-41.

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J, 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-83.

Rame JE, Barouch LA, Sack MN*, et al.*, 2011. Caloric restriction in leptin deficiency does not correct myocardial steatosis: failure to normalize PPAR{alpha}/PGC1{alpha} and thermogenic glycerolipid/fatty acid cycling. *Physiol Genomics* **43**, 726-38.

Reik W, Murrell A, 2000. Genomic imprinting. Silence across the border. *Nature* **405**, 408-9.

Relton CL, Davey Smith G, 2010. Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment. *PLoS Med* **7**, e1000356.

Roberts A, Pimentel H, Trapnell C, Pachter L, 2011. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **27**, 2325-9.

Robinson MD, Mccarthy DJ, Smyth GK, 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-40.

Rockman HA, Ross RS, Harris AN*, et al.*, 1991. Segregation of atrial-specific and inducible expression of an atrial natriuretic factor transgene in an in vivo murine model of cardiac hypertrophy. *Proc Natl Acad Sci U S A* **88**, 8277-81.

Ross-Innes CS, Stark R, Teschendorff AE*, et al.*, 2012. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389-93.

Ruthenburg AJ, Li H, Patel DJ, Allis CD, 2007. Multivalent engagement of chromatin modifications by linked binding modules. *Nat Rev Mol Cell Biol* **8**, 983-94.

Saxena A, Carninci P, 2011. Long non-coding RNA modifies chromatin: epigenetic silencing by long non-coding RNAs. *Bioessays* **33**, 830-9.

Schrauwen P, Hardie DG, Roorda B*, et al.*, 2004. Improved glucose homeostasis in mice overexpressing human UCP3: a role for AMP-kinase? *Int J Obes Relat Metab Disord* **28**, 824-8.

Shi X, Garry DJ, 2006. Muscle stem cells in development, regeneration, and disease. *Genes Dev* **20**, 1692-708.

St-Onge MP, Jones PJ, 2002. Physiological effects of medium-chain triglycerides: potential agents in the prevention of obesity. *J Nutr* **132**, 329-32.

Terranova R, Yokobayashi S, Stadler MB*, et al.*, 2008. Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev Cell* **15**, 668-79.

Trapnell C, Pachter L, Salzberg SL, 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-11.

Trapnell C, Roberts A, Goff L*, et al.*, 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562-78.

Tripathi V, Ellis JD, Shen Z*, et al.*, 2010. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* **39**, 925-38.

Urdinguio RG, Sanchez-Mut JV, Esteller M, 2009. Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. *Lancet Neurol* **8**, 1056-72.

Van Rooij E, Marshall WS, Olson EN, 2008. Toward microRNA-based therapeutics for heart disease: the sense in antisense. *Circ Res* **103**, 919-28.

Wang KC, Chang HY, 2011. Molecular mechanisms of long noncoding RNAs. *Mol Cell* **43**, 904-14.

Wang KC, Yang YW, Liu B*, et al.*, 2011. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120-4.

Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W, 2013. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res* **41**, e74.

Wutz A, Gribnau J, 2007. X inactivation Xplained. *Curr Opin Genet Dev* **17**, 387-93.

Yap KL, Li S, Munoz-Cabello AM*, et al.*, 2010. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell* **38**, 662-74.

Yoshioka J, Schulze PC, Cupesi M*, et al.*, 2004. Thioredoxin-interacting protein controls cardiac hypertrophy through regulation of thioredoxin activity. *Circulation* **109**, 2581-6.

Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W, 2009. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952-8.

Zhao G, Jeoung NH, Burgess SC*, et al.*, 2008. Overexpression of pyruvate dehydrogenase kinase 4 in heart perturbs metabolism and exacerbates calcineurin-induced cardiomyopathy. *Am J Physiol Heart Circ Physiol* **294**, H936-43.

Zhao H, Sun Z, Wang J, Huang H, Kocher JP, Wang L, 2014. CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006-7.