UNIVERSITÀ DEGLI STUDI DI TORINO

Department of Life Sciences and Systems Biology

Doctoral School in Life and Health Sciences

PhD Program in Complex Systems for Life Sciences

XXX Cycle

Novel immediate-early long non-coding RNAs control cell-cycle progression with distinct molecular mechanisms

TUTOR:
Prof. Salvatore Oliviero

CYCLE COORDINATOR:
Prof. Michele De Bortoli

PhD CANDIDATE:
Mara Maldotti

Academic years: 2014-2017

SSD: BIO/11

A mamma Lucia, papà Ivano e Giulia

Lo strumento più raffinato e straordinario a nostra disposizione
resterà sempre la nostra mente.
- Rita Levi-Montalcini

Contents

Abbreviations

CCR	Cell-Cycle Regulated
ceRNA	competing endogenous RNA
ChIP	Chromatin Immunoprecipitation
c-Myc	avian myelocytomatosis viral oncogene homolog
ESCs, mES	E14 mouse Embryonic Stem Cells
FACS	Fluorescence-Activated Sell Sorting
H3K4me1	Histone 3 lysine 4 monomethylation
H3K4me3	Histone 3 lysine 4 trimethylation
H3K27ac	Histone 3 lysine 27 acetylation
H3K36me3	Histone 3 lysine 36 trimethylation
I-SMAD	Inhibitory SMAD
let-7	lethal-7
lncRNA	long non-coding RNA
lincRNA	long intergenic non-coding RNA
Med	Mediator
MEFs	Mouse Embryonic Fibroblasts
miRNAs	microRNAs
ncRNAs	non-coding RNAs
RNA-FISH	RNA Fluorescent in Situ Hybridization
RNA-seq	RNA sequencing
RT-qPCR	**Q**uantitative Real-Time PCR
TGF-β	Transforming growth factor-β
TSS	Transcription Start Site

List of Figures

List of Tables

Abstract

In mammals the cell-cycle progression through the G1 phase is a tightly regulated process mediated by the transcriptional activation of coding genes, whose dysregulation often leads to tumorigenesis.

We report the discovery of two novel cell-cycle regulated (CCR) long non-coding RNAs (lncRNAs), expressed in response to serum treatment in mouse embryonic fibroblasts (MEFs), comparably to immediate-early genes.

The localization in specific subcellular compartments and the function of the non-coding RNAs are closely related. By loss-of-function experiments we find that CCR492 and CCR102 lncRNAs are required for G1/S phase progression. In the cell they show a cytoplasmic and chromatin localization, respectively.

Mechanistically, CCR492 contains 4 let-7 MREs and acts as a competing endogenous RNA (ceRNA) to antagonize the function of let-7 microRNAs, leading to the derepression of c-Myc. Moreover, we show that the ectopic expression of CCR492 along with a constitutively active H-Ras promotes cell transformation.

In contrast to coding genes, most lncRNAs do not exhibit high levels of sequence conservation across species. Interestingly, murine CCR102 and its identified human counterpart are localized in a syntenic region and are associated with orthologous genes. Differently from the majority of known chromatin-associated lncRNAs, which are involved in gene silencing, we observe that CCR102 promotes an active chromatin state activating the expression of its neighboring protein-coding gene Smad7 *in cis*. This transcriptional regulation is conserved also for its human counterpart indicating the analogous biological role. CCR102 is ubiquitously expressed and it is upregulated in a subset of human colorectal cancers in comparison with normal colon tissues.

Thus, our findings identified two novel lncRNAs and demonstrated that CCR492 and CCR102, acting through distinct mechanisms, play a key role in cell proliferation in physiological and pathological contexts.

Chapter 1

Introduction

1.1 The central dogma of molecular biology

The central dogma of molecular biology was established by Francis Crick in 1958 and describes the flow of genetic information to be unidirectional from DNA into RNA and into protein as the final functional product, but never from protein to nucleic acids. According to the central dogma, a gene is therefore defined as a part of DNA essential for the generation of RNA and protein (Crick, 1970).

The human genome consists of 3.2 billion base pairs (bp) organized into 23 chromosomes that encode for a number of 60,483 genes including protein-coding genes, non-coding genes and pseudogenes (Siggens & Ekwall, 2014). While the number of protein-coding genes stays under broad consensus, recent studies of the human transcriptome have revealed a huge number of non-coding RNAs (ncRNAs). These transcribed elements, which lack the capacity to code for a protein, represent a break with the central dogma of biology because they are not translated. The role of the non-coding genome became a new focus of RNA biology research, which has been pivotal for the development of modern molecular biology (Figure 1.1).

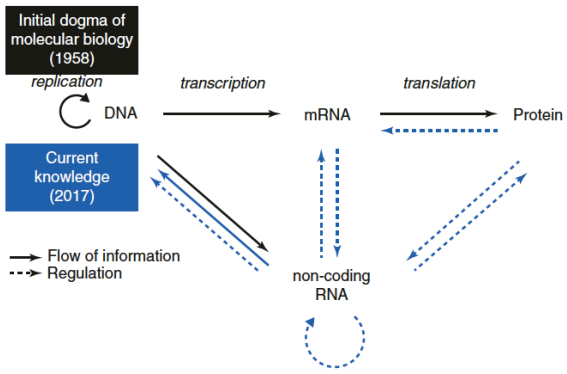


Figure 1.1 Initial and current dogma of molecular biology
(M.R.S. Rao et al., 2017)

1.2 From non-coding genome to non-coding transcriptome

Despite its status as "junk", non-coding genome received sustained interest since the 1970s. Indeed, technical innovation in eukaryotic transcriptome landscape analyses has revealed that the majority of the genome is transcribed, producing large numbers of non-coding RNAs.

The diversity of the non-coding transcriptome is considered an argument to explain the phenotypic differences observed across species given a relatively similar number of protein-coding genes among fruit fly (13,985; BDGP release 4), nematode worm (21,009; Worm base release 150), and human (23,341; NCBI release 36) (Siggens & Ekwall, 2014). Although transcriptional noise or pervasive transcription of the genome have only recently been recognized due to the development of sensitive high-throughput sequencing technologies, the existence of functional non-coding transcripts is established for several decades (Morris & Mattick, 2014). To date, the catalog of the functional non-coding transcriptome is growing continuously (Cech & Steitz, 2014). General conventions divide non-coding RNAs into two main categories, based on the size: small (< 200 nt) and long (> 200 nt) non-coding RNAs (Gibb, Brown, & Lam, 2011). Within these two categories, there are also many subclasses of ncRNAs.

1.2.1 Small non-coding RNAs

Small ncRNAs are very heterogeneous and could be classified in several dozens of classes. These include well-characterized housekeeping ncRNAs: transfer RNA (tRNA) and some ribosomal RNA (rRNA) essential for fundamental aspects of cell biology, small nuclear RNAs (snRNAs) involved in splicing and small nucleolar RNAs (snoRNAs). Another class of small RNAs encoded in the human genome and functional in regulation of RNA stability are piwi-interacting RNAs (piRNAs) of ~27 nt in length. This species interacts with PIWI proteins and degrades expressed transposable elements in the germline (Luteijn & Ketting, 2013).

The most extensively studied small RNAs with a regulatory role are microRNAs (miRNAs) of ~22 nt in length. Elegant studies have defined an intricate mechanistic basis for miRNA-mediated silencing of target gene expression through the RNA-induced silencing complex (RISC), which employs Argonaute family proteins (such as AGO2) to cleave target mRNA transcripts or to inhibit the translation of that mRNA (Ameres & Zamore, 2013) miRNAs modulate the expression levels of genes involved in many biological processes, for instance cell proliferation, and could be aberrantly expressed in cancer (Gu & Iyer, 2006),(Garzon, Calin, & Croce, 2009),(Bueno & Malumbres, 2011). In particular, the let-7 miRNA family was initially discovered in *Caenorhabditis elegans* (Reinhart et al., 2000), and has been shown to be evolutionarily highly conserved (Pasquinelli et al., 2000). Let-7 is undetectable in embryonic stem cells, but it is expressed at high levels in adult tissues and its expression is deregulated or lost in many cancers (Boyerinas, Park, Hau, Murmann, & Peter, 2010),(Johnson et al., 2007).

1.2.2 Long non-coding RNAs

Although the discovery and functional characterization of small ncRNAs have dominated the field of RNA biology in the past, long non-coding RNAs (lncRNAs) are the least explored emerging regulatory molecules.

Transcripts >200 nt in size that lack coding potential, as determined by the absence of large open reading frames and codon conservation, have been classified as lncRNAs (Morris & Mattick, 2014).

Many lncRNA genes have a structure similar to genomic protein-coding genes. The majority of eukaryotic lncRNAs are produced by RNA polymerase II and their transcriptional start sites

(TSS) reside within DNase I hypersensitive sites suggesting nucleosome depletion from these regionss are further associated with epigenetic signatures common to protein-coding genes, such as trimethylation of histone 3 lysine 4 (H3K4me3) at the TSS and trimethylation of histone 3 lysine 36 (H3K36me3) along the gene body (Guttman et al., 2009).

LncRNA transcripts also share characteristics with mRNAs, such as polyadenylation, capping and splicing. They may or may not be 3' end polyadenylated and they may also be present in both forms, such as bimorphic transcripts. Many lncRNAs are capped at the 5' end, except those processed from longer precursors, for instance circRNAs. Early studies did not consider single-exon lncRNAs in order to distinguish lowly expressed lncRNAs from transcriptional noise. However, many lncRNAs are unspliced single-exon transcripts or are characterized by the presence of few exons. Despite having few exons, lncRNAs show high tendencies for alternative splicing (Derrien et al., 2012).

In addition, lncRNAs differ from other classes of RNAs in their lower expression level and higher tissue-specificity. A large number of lncRNAs are developmentally and temporally regulated or restricted to particular tissues or organs indicating their role in specific cellular processes, rendering their discovery more difficult (Derrien et al., 2012).

Most importantly, lncRNAs generally show low sequence conservation across different species unlike many protein-coding genes (Derrien et al., 2012). Recently, a study examining the non-coding transcriptome of 17 different species showed that although the body of non-coding genes tends not to be conserved, short patches of conserved sequences could be found at their 5' ends. This confirmed a higher conservation of TSS and synteny, as well as expression patterns in different tissues, especially in those involved in development (Hezroni et al., 2015). In general, promoters of lncRNA genes exhibit higher conservation rates than their exonic (Derrien et al., 2012),(M. K. Iyer et al., 2015), have only a moderate sequence identity and alignment to their orthologs and this phenomenon may indicate the importance of transcription from a specific genomic location and the reduced selective pressure on the primary sequence of lncRNAs (J. Zhao, Sun, Erwin, Song, & Lee, 2008). Interestingly, some lncRNAs are at conserved genomic locations, with conserved exon-intron structures yet no detectable sequence conservation. For example, protein-coding genes adjacent to a lncRNA gene in zebrafish are more likely to have orthologs adjacent to lncRNA genes in human or

mouse, even when all lncRNAs with sequence homology are excluded from the analysis (Ulitsky, Shkumatava, Jan, Sive, & Bartel, 2011). Because of the low expression and the low evolutionary conservation levels of lncRNAs, many questions regarding their functionality have been raised. However, various studies have revealed an active role of lncRNAs in controlling multiple cell regulatory layers including chromosome architecture, chromatin modulation and epigenetic modification, transcription, RNA maturation, splicing and translation (Rinn & Chang, 2012).

1.3 Techniques to identify lncRNAs

Non-coding RNAs have historically been more difficult to detect than protein-coding genes. This is because lncRNAs exhibit lower levels of expression (Rinn & Chang, 2012) tend to be more tissue-specific in their expression and cannot be predicted from open reading frames (ORFs). In the past decade, however, large-scale analyses have focused on identifying lncRNAs in a comprehensive manner. The first large-scale catalog of putative non-coding transcripts came from the FANTOM project (The FANTOM Consortium, 2005) which used cDNA cloning followed by Sanger sequencing. Subsequent studies refined EST- and cDNA- based lncRNA catalogs in mouse and human, which include the current RefSeq and Ensembl lncRNA annotations (Derrien et al., 2012),(Pruitt, Tatusova, Brown, & Maglott, 2012). A major advance was presented by Guttman et al. who combined the methodology of DNA tiling arrays with the logic of epigenetics. Since non-coding RNAs could resemble protein-coding genes in their fundamental structure and epigenetic characteristics, the authors considered that ncRNAs could be polyadenylated, spliced and they could have a gene promoter marked by H3K4me3 and a gene body marked by H3K36me3. This strategy identified several thousand regions of unannotated transcription, defined as lncRNAs, based on their lengths and characteristics (Guttman et al., 2009) (Figure1.2).

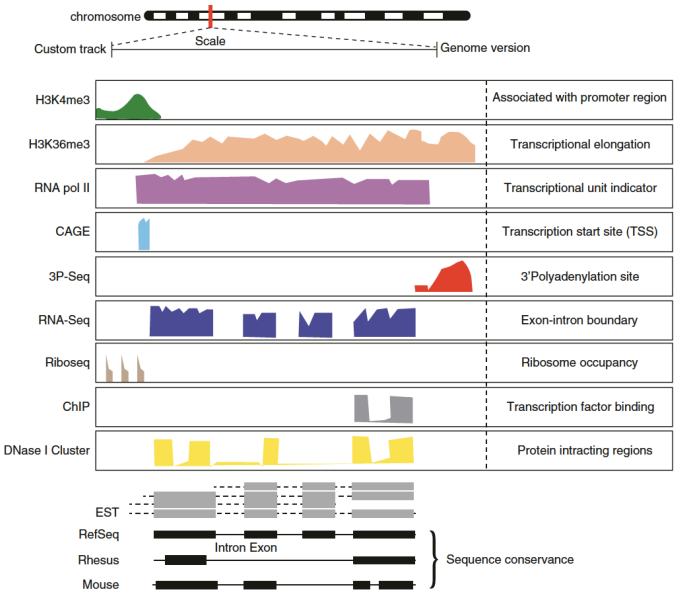


Figure 1.2 Prediction of lncRNA based on NGS technology
(M.R.S. Rao et al., 2017)

.

The introduction of RNA-sequencing (RNA-seq) subsequently enabled the direct detection of all RNA species in a cell. Indeed, RNA-seq is a quantitative and exceptionally sensitive technique and is now the gold standard method to discover novel lncRNAs. The library is prepared by RNA fragmentation, adapter ligation, cDNA synthesis, size selection and limited cycles of amplification. The libraries can be prepared from poly (A)+ enriched RNA by selectively depleting rRNA. Recently, Carninci's group in conjunction with the FANTOM consortium performed a complete profiling of nucleus- and cytoplasmic- enriched RNA fractions from a representative set of human and mouse stem cell and differentiated cell lines using four complementary sequencing technologies (deep sequencing, CAGE, nanoCAGE and CAGEscan) and identified several thousand novel lncRNAs, which were principally enriched in nucleus. They also discovered that a large number of these novel transcripts originated from long terminal repeats (LTR) elements (Fort et al., 2014).

1.4 Genomic contexts of lncRNAs

The GENCODE gene annotation classifies lncRNAs depending on their location and orientation with respect to the nearest protein-coding genes (Derrien et al., 2012) (Figure 1.3):

(1) Exonic antisense lncRNAs: intersect an exon of a protein-coding gene locus on the opposite DNA strand;

(2) Divergent lncRNAs: when its expression and the expression of a neighboring coding transcript on the opposite strand are initiated in close genomic proximity;

(3) Intronic lncRNAs: reside within an intron of a protein-coding gene on the same or opposite strand, but without any overlap with protein-coding exons;

(4) Intergenic lncRNAs: do not intersect any protein-coding gene locus.

However, no intrinsic functional differences between these lncRNA categories have yet been demonstrated (Morris & Mattick, 2014).

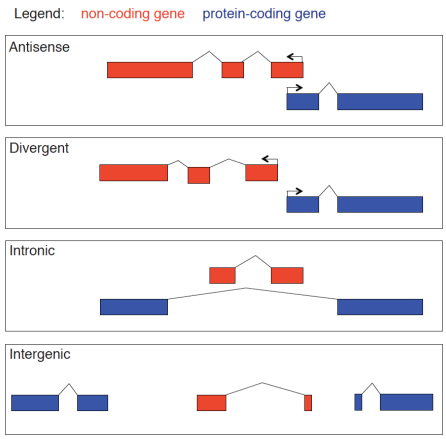


Figure 1.3 lncRNAs classification on the bases of their localization in genome architecture. Antisense, Divergent, Intronic, Intergenic

1.5 Subcellular localization of lncRNAs

LncRNAs can be located predominantly in the nucleus or in the cytoplasm, or they can have a dual localization (e.g. HOTAIR) (Ayupe et al., 2015). Depending on the transcript, however, compared to mRNAs, lncRNAs show higher enrichment in the nucleus (Ulitsky & Bartel, 2013). Some of the best-studied lncRNAs, such as Xist, Malat1, Neat1, and Miat, are almost exclusively located in the nucleus and even define specific nuclear domains (C. J. Brown et al.,

1992),(Hutchinson et al., 2007).Within the nucleus, these transcripts are particularly found in the chromatin-associated fraction suggesting their important role in epigenetic modification and gene regulation (Derrien et al., 2012). Indeed, several subgroups of lncRNAs with a precise subcellular localization have been defined, such as chromatin-enriched (che)RNAs (Werner & Ruthenburg, 2015) and chromatin-associated lncRNAs, CARs (Mondal, Rasmussen, Pandey, Isaksson, & Kanduri, 2010). However, other studied lncRNAs are found mostly in the cytoplasm (Cesana et al., 2011),(K. C. Wang et al., 2011),(Huarte et al., 2010) they are responsible for post-transcriptional regulations, such as mRNA stability and translational control. The more specific localization of lncRNAs within either the cytoplasm or the nucleus, as well as the factors and sequence elements that dictate this localization, remain largely unexplored.

1.6 Molecular models of lncRNA–protein interaction

Molecular mechanisms for lncRNA action involve their specific interactions with proteins, chromatin and RNA (Rinn & Chang, 2012). In this way, lncRNAs, independent from their subcellular localization, could act as decoys, scaffolds and guides facilitating various biological processes (Figure 1.4).

In particular, lncRNAs can act as decoys for proteins, such as transcription factors (TF), modulating their binding to DNA. Indeed, they titrate the protein away from its potential targets, as has been reported for lincRNA Gas5 and glucocorticoid receptor (Kino, Hurt, Ichijo, Nader, & Chrousos, 2010), PANDA and NF-Y (Hung et al., 2011), sno-lncRNAs and Fox2 (Yin et al., 2012) and Gadd7 and TDP-43 (Liu, Li, Zhang, Guo, & Zhan, 2012).

Moreover, lncRNAs function as scaffolds for protein-complex formation. They are known to associate with a multitude of histone- or DNA-modifying and nucleosome remodeling complexes (Han & Chang, 2015),(Davidovich & Cech, 2015) and are essential for the assembly of particular nuclear substructures (Chujo, Yamazaki, & Hirose, 2016). Given that many lncRNAs are mostly cytoplasmic, this scaffolding mechanism might also play important roles in the cytosol. LncRNAs have additionally been identified to play a role in the establishment of chromatin structure, more specifically in enhancer-promoter interactions for example by binding to the Mediator complex, facilitating transcription initiation (Lai et al., 2013).

A guide function to recruit proteins to specific genomic locations depends on the biological context (cell- or tissue-type, developmental stage, pathology) and often cannot be explained by a simple RNA-DNA sequence complementarity. For some lncRNA guides the formation of a triple-helix structure between lncRNA and DNA was experimentally proven, as in the case of Khps1 which anchors the CBP/p300 complex to the proto-oncogene SPHK1 (Postepska-Igielska et al., 2015). Another example is MEG3 which guides the EZH2 subunit of PRC2 to TGF-β-regulated genes (Mondal et al., 2010).

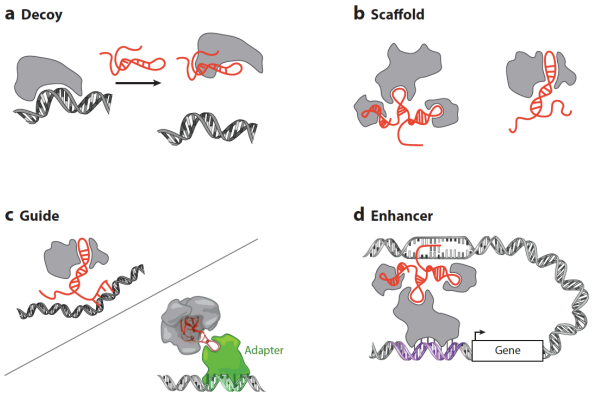


Figure 1.4 Molecular interactions. lncRNAs can act as decoys, scaffolds and guides (Rinn & Chang, 2012)

In summary, lncRNAs present high functional versatility, which likely relies on their ability as RNA molecules to acquire different structures and molecular interactions with different target molecules.

1.7 *cis*- versus *trans*-regulatory mechanisms

The increasing number of lncRNAs reported is in opposition to the relatively few examples of functionally characterized lncRNAs.

The potential mechanisms of lncRNA function can be divided into three groups:

(1) those that rely solely on the act of transcription or on the nascent RNA;

(2) those that require the processed RNA yet depend on the site of transcription;

(3) those that are independent of the site of transcription.

A major difference between the first two groups and the last one is in whether the direct targets of the lncRNA activity are found only in proximity to the lncRNA gene (*cis* targets, groups 1 and 2), or anywhere in the cell (*trans* targets, group 3) (Ulitsky & Bartel, 2013). A *cis* regulation by lncRNAs contributes to local control of gene expression of neighboring genes via direct or indirect binding in order to modify the chromatin state on the same allele from which it is transcribed (Guttman & Rinn, 2012). The most intensively studied and best understood *cis* mechanism of regulation by lncRNAs is the mammalian X inactivation (Xic), a genetic locus that specifies a number of ncRNAs, including Xist (Lee, Davidow, & Warshawsky, 1999), HOTTIP lncRNA from the human HoxA cluster adds a further dimension to *cis* regulation by lncRNAs defining a central role for chromosome looping in delivering a lncRNA to its site of action. In particular, it activates the expression of HoxA genes by interacting with the WDR5/MLL complex through chromatin loop formation (K. C. Wang et al., 2011). By serving as key intermediates that transmit information from higher order chromosomal looping into chromatin modifications, lncRNAs may organize chromatin domains to coordinate long-range gene activation. Similar mechanisms of action appear to be in play for other lncRNAs with transcriptional repression activities. The lncRNA Air silences transcription of its target gene on the paternal chromosome via a specific interaction between the ncRNA and chromatin at its promoter by recruiting G9a (Nagano et al., 2008).

trans-acting lncRNAs typically facilitate epigenetic regulation of gene expression in geographically distant locations of the genome. The interaction between lncRNAs and the chromatin complexes is required for the repression of specific gene loci as described for lncRNA HOX antisense RNA (HOTAIR) (Rinn et al., 2007). This lncRNA is expressed from the HOXC gene cluster and regulates HOXD1 locus and additional genomic loci expression, as recently demonstrated, through repressive effects independent of Polycomb Repressive Complex 2 (PRC2), whose recruitment may rather be a downstream consequence of gene silencing (Portoso et al., 2017). Another well-known lncRNA, lincRNA-p21, a repressor in p53-dependent transcriptional responses, is able to exert its effects *in trans* on chromatin structure and gene expression across multiple sites in the genome through the physical association with hnRNP-K (Huarte et al., 2010). Importantly, the wild-type allele of lincRNA-p21 located *in trans* failed to rescue p21 levels in mouse embryonic fibroblasts (MEFs),

demonstrating that lincRNA-p21 acts *in cis* on its neighboring gene p21 to enforce the G1/S checkpoint (Dimitrova et al., 2014).

1.8 Molecular mechanisms of action of lncRNAs

LncRNAs have emerged as key regulators in a wide range of biological processes such as cell-cycle, metabolism, apoptosis, differentiation, maintenance of pluripotency, etc. (Geisler & Coller, 2013). Since the spectrum of lncRNA functional involvement has extended to various biological processes, a common theme is the regulation of gene expression by lncRNAs resulting in activation or repression of target genes involved in specific processes (Rinn & Chang, 2012). Mechanistically, most well-characterized lncRNAs show a functional role in gene expression regulation, typically transcriptional rather than post-transcriptional.

1.8.1 lncRNAs in transcriptional regulation

Nuclear lncRNAs are crucial players in the regulation of transcription via modulation of chromatin. Indeed, a significant number of chromatin-associated lncRNAs interact with chromatin-remodeling complexes, driving them to specific genomic loci and others are implicated in the architectural conformation and activity of transcriptional enhancers (Lai et al., 2013).

Promoters and enhancers constitute fundamental *cis*-regulatory elements for the control of protein-coding gene expression, serving as platforms for the recruitment of transcription factors and transcription machinery and the establishment of a particular chromatin organization (Lenhard, Sandelin, & Carninci, 2012) (Figure 1.5).

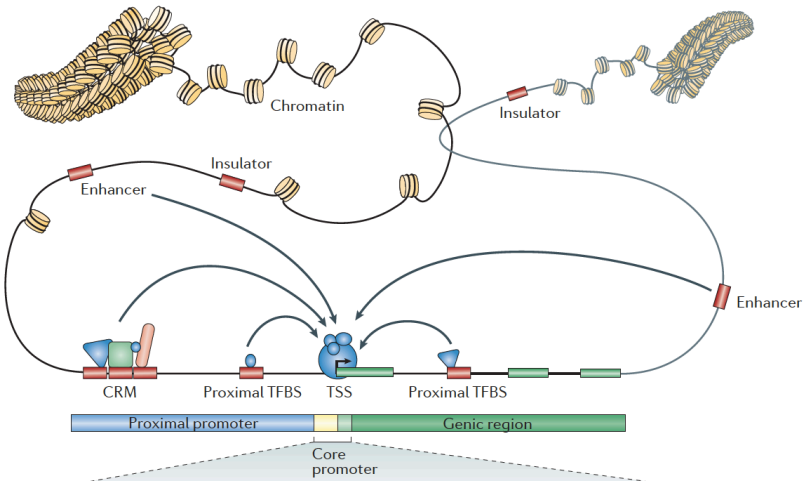


Figure 1.5 Regulatory elements controlling gene expression
(Lenhard et al., 2012 – Modified)

In general promoters can be subdivided into proximal and core promoters, with proximal promoters being similar to enhancers in serving as transcription factor binding sites and with core promoters mediating the assembly of the pre-initiation complex.

Enhancers have been classically defined as DNA regions that positively influence the expression of target genes in an orientation-independent manner. Active enhancer elements are characterized by the presence of high levels of the activating H3K4me1 and H3K27ac histone modifications (Creyghton et al., 2010). At the same time, the H3K4me3 histone mark is mostly absent from enhancer sequences (ENCODE Project Consortium et al., 2007). The gene activating function of enhancers is completed by clustering of transcription factor binding sites, thereby serving as platforms for a cooperative binding of transcription factors. As a consequence, nucleosomes are excluded from active enhancer regions.

Promoters and enhancers share many similar features in mammalian genomes such as the binding of transcription factors to both elements and the presence of core promoter elements at similar frequencies and the initiation of divergent transcription (Venters & Pugh, 2013),(Andersson et al., 2014)(Figure 1.6).

Regulatory element	Activation marks						
	DHS	Methyl	H3K4me1	H3K4me3	H3K27ac	Med	CTCF
Promoter	+	Low	–	+	++	+	+/–
Enhancer	+	Low	+	–	++	+	+/–
Super-enhancer	+	Low	+	–	+++	+++	+/–
Insulator	+	Low	–	+/–	–	–	+

Figure 1.6 Markers used to identify regulatory elements controlling gene expression
(Sun and Taipale, 2016 – Modified)

Chromatin looping is mediated by the interaction of transcriptional co-activator complexes such as Mediator and cohesion with enhancer-bound transcription factors and brings enhancer sequences and gene promoters in physical proximity. This positively influences the pre-initiation complex assembly at promoters (Kagey et al., 2010). Mechanistically, the Mediator complex, which is comprised of 26 subunits in mammals, is crucial for the organization of genomic DNA into topological domains. More recently, it has been suggested that lncRNAs favor these chromatin loops by interacting with the Mediator complex (Lai et al., 2013) (Figure 1.7).

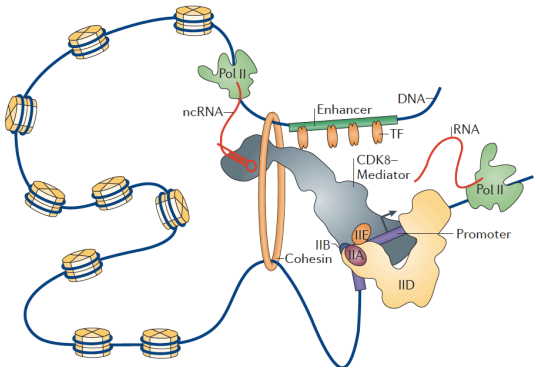


Figure 1.7 Regulation of chromatin architecture by Mediator
(Allen & Taatjes, 2015)

Instead, insulator prevents unsuitable interactions of *cis*-regulatory elements with promoters when it is located between an enhancer and a promoter. In vertebrates, this function is mediated by binding of CTCF (CCCTC-binding factor). Also, insulators function as barriers to the spread of heterochromatin and reside at the border of eu- and heterochromatin (Recillas-Targa et al., 2002).

Basically, lncRNAs modulate local gene expression either by recruiting transcription factors, epigenetic modifiers, or chromatin organizers, or by forming a RNA-DNA triplex that anchors them and associated effector proteins to specific genomic loci. One of the most common lncRNA, Xist, mediates the recruitment of PRC2 to the Xi to silence gene expression (J. Zhao et al., 2008). Moreover, ANRIL was shown to be involved in epigenetic regulation of the INK4b-ARF-INK4A locus by recruiting PRC1 and PRC2 to form heterochromatin surrounding the INK4b-ARF-INK4a locus, leading to repression of gene expression (Bracken et al., 2007). KCNQ1OT lncRNA regulates imprinting-associated chromatin interactions with EZH2. Silencing of KCNQ1OT1 leads to depletion of the loop and loss of imprinting (Redrup et al., 2009). It has also been described that MEG3 and EZH2 share common target genes. MEG3 binding sites have GA-rich sequences that guide MEG3 to the chromatin through RNA-DNA triplex formation. These structures are widespread and are present over the MEG3 binding sites associated with the TGF-β pathway genes (Mondal et al., 2015). Some lncRNAs associate with chromatin-remodeling complexes: lncRNA SChLAP1, Mhrt and Evf2 lncRNA interact with SWI/SNF-related chromatin remodelers Brg1 and Baf170 (Tang et al., 2017) to promote gene repression. Instead, others lncRNAs recruit the SWI/SNF complex to mediate gene activation, such as lncTCF7 (K. C. Wang et al., 2011) and Linc-Cox2 (Hu et al., 2016). Meanwhile, lncRNAs such as NEAT1 and MALAT1 are related to nuclear architecture. NEAT1 contributes to nuclear paraspeckle formation, while MALAT1 is deregulated in many human cancers and localized in nuclear speckles domains, which are involved in the assembly, modification, and storage of pre-mRNA processing machinery. Recently it has been reported that NEAT1 and MALAT1 are coregulated, and associated with histone modifications of active chromatin (Clemson et al., 2009),

In addition, to facilitate epigenetic changes that impact gene transcription, emerging evidence suggests that lncRNAs can contribute to gene regulation by influencing the activity of enhancers. For instance, HOTTIP, as described in detail before, is implicated in chromatin looping of active enhancers to the distal HoxA locus, though knockdown and overexpression of HOTTIP is not sufficient to alter chromosomal conformations (K. C. Wang et al., 2011).

1.8.2 lncRNAs deriving from particular promoter and enhancer transcriptional contexts

The discovery of promoter divergent (bidirectional or head-to-head) transcription as an intrinsic feature of the eukaryotic transcriptional machinery has given rise to the identification of divergent lncRNAs (Seila et al., 2008). Transcription at divergent TSSs in promoters and enhancers initiates within a comparable distance of 100-200 bp and at nucleosomes flanking both TSSs (Duttke et al., 2015),(Core et al., 2014).

A major recent development has been the discovery of enhancer RNAs (eRNAs), which are critical for the proper coordination of enhancer genomic loci with gene expression regulation. Production of these transcripts correlates well with characteristic enhancer modifications at chromatin such as H3K4me1, H3K27ac and p300 binding (Core et al., 2014). The expression levels of eRNAs also correlate with levels of mRNA synthesis of neighboring genes implying eRNA involvement in transcriptional activation. In accordance with this, gene activation can be lost when eRNAs are specifically depleted from cells (T.-K. Kim et al., 2010). The functionality of eRNAs is still controversial but could be due to the involvement of these transcripts in enhancer-promoter looping, and/or they could facilitate the release of paused Pol II into productive elongation, both of which contribute to a robust transcription activation of nearby and distant genes (W. Li et al., 2013),(Schaukowitch & Kim, 2014).

The existence of bidirectional promoter architecture is a common and conserved feature across many species, indicating a significant ancestral origin of this type of genomic organization that became more heavily utilized by higher organisms (Y.-Y. Li et al., 2006). The chromatin of bidirectional promoters shows many features commonly associated with more active promoters, including higher density Pol II binding, increased density of Histone H3 acetylation and increased promoter proximal markings with H3K4Me2/3. Meanwhile, Histone H4 acetylation appeared to be under-represented at bidirectional promoters (J. M. Lin et al., 2007). Promoter bidirectional transcription was attributed to the presence of reverse oriented core promoter elements that are similar to their counterparts in the protein-coding gene direction. This suggests that divergent antisense transcription initiation is not due to promiscuous transcription initiation of Pol II, but due to the presence of core promoter-like elements that favor pre-initiation complex assembly (Duttke et al., 2015). Originating from the opposite

strand of a protein-coding gene strand, these lncRNA transcripts do not overlap or only partially overlap with the 5' region of paired protein-coding genes (Figure 1.8). One study has catalogued lncRNAs expressed in human embryonic stem cells (hESCs) and suggested that >60% of lncRNAs are divergently transcribed from active protein-coding genes (within a window of ±2 kb around the protein-coding gene TSSs) (Sigova et al., 2013).

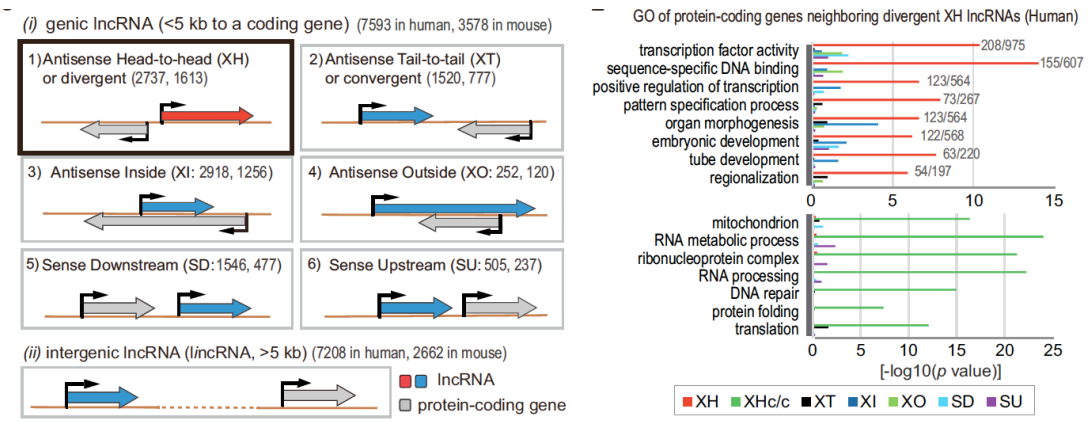


Figure 1.8 GO analysis of coding genes neighboring particular biotypes of lncRNA
(divergent lncRNA organization is highlighted by a black box)
(Luo et al., 2016)

The correlations of expressions between neighboring lncRNAs and protein-coding genes show contradictory results, on the one hand high positive correlation was suggested (Derrien et al., 2012),(Ørom et al., 2010), on the other hand no such relationship was found (Cabili et al., 2011; Hung et al., 2011). Orom's group used a GENCODE annotation of the human genome by which they revealed *cis*-regulatory functions of lncRNAs on nearby transcription. They focused on lncRNAs distant at least 1 kb of the first and last exons as to avoid promoter- and 3'-associated transcripts. These lncRNAs function as enhancers to activate expression in their chromosomal neighborhood *in cis* (Ørom et al., 2010). Another study focused on the expression of lncRNAs around the promoters of 56 cell-cycle genes under diverse perturbation conditions, such as DNA damage or oncogenic stimuli. The expression of these lncRNAs and their closest protein-coding genes did not correlate (Hung et al., 2011).

However, genic lncRNAs (< 5 kb from a protein-coding gene) exhibit significantly greater expression correlation with their nearest coding genes than intergenic lncRNAs (> 5 kb from a

protein-coding gene) (Luo et al., 2016) and *cis* regulation of nearby transcription by divergent lncRNAs represents a general theme in mammalian gene regulation.

Some of the biological functions of lncRNAs overlap with known functions of Mediator complex and functionally cooperate with it, to associate with enhancer regions and regulate chromatin architecture. Indeed, a type of lncRNA called activator RNA (lncRNA-a) interacts with Mediator, by favoring the chromatin looping between the lncRNA-a loci and their targets, to increase the transcription of neighboring genes with a *cis*-mediated mechanism (Lai et al., 2013). In a recent elegant report of allele-specific engineering of five lncRNA loci that act *in cis* to enhance proximal gene expression, it has been reported that the promoters of lncRNAs may act as enhancer elements (Engreitz et al., 2016). Highly chromatin enriched-RNAs (cheRNAs) are largely cell type specific and their depletion produces significant decreases in neighboring gene expression, establishing cheRNAs as transcriptional activators. The cheRNA product could facilitate looping from the newly activated enhancer to the gene promoter, without interaction with Mediator as is the case for HIDALGO lncRNA (Werner et al., 2017). The antisense E2F1-regulated lncRNA named Khps1 activates SPHK1 expression by recruiting the histone acetyltransferase p300/CBP to the SPHK1 promoter, which leads to local changes of the chromatin structure that ensures E2F1 binding and enhances transcription. Mechanistically, Khps1 forms a DNA-RNA triplex, upstream of the TSS of SPHK1, serving as anchor points to guide chromatin modifiers to favor a E2F1-driven regulatory loop (Postepska-Igielska et al., 2015). A rigorous study of the Evx1as/EVX1 locus, by complementary genetic approaches together with single-cell time-course analysis of gene expression, reveals an essential role of the lncRNA Evx1as in promoting EVX1 activation and ESC differentiation. In detail, Evx1as RNA binds to chromatin at an Evx1as/EVX1 enhancer site, facilitates Mediator binding, and promotes chromatin looping between the Evx1as/EVX1 promoter and enhancer. Thus, divergent lncRNAs, or at least a subset of them, can positively modulate nearby transcription and participate in biological processes similar to those controlled by the nearby protein-coding genes (Luo et al., 2016). Therefore, lncRNAs may fine-tune the nuclear architecture and expression of specific genomic loci, contributing in part to the increased phenotypic complexity of higher eukaryotes (Yan, Luo, Lu, & Shen, 2017).

1.8.3 lncRNAs in post-transcriptional regulation

Several works described a functional role for lncRNAs localized within the cytoplasm to regulate gene expression at the post-transcriptional level. Some of them have been reported to work through RNA-RNA interactions. Therefore, sequence complementarity-mediated interactions could regulate mRNA stability and translation. For instance, half-STAU1-binding site RNA (1/2-sbsRNA) contains Alu short interspersed elements (SINE) and its mechanism of action is mediated by the repeat element, which forms imperfect complementary RNA duplexes with Alu elements of 3' UTRs of target mRNAs and recruits double-stranded RNA binding protein Staufen1, which targets mRNA for degradation. Thus, the cells use this strategy to recruit proteins to mRNAs and mediate the decay of these mRNAs (Gong & Maquat, 2011). Some natural-antisense transcripts (NATs) may bind to mRNA to mask the binding sites of miRNAs and thus stabilize mRNA. BACE1AS (BACE1 antisense RNA), a ~2 kb lncRNA, has been reported to be activated in Alzheimer disease to form an RNA duplex with BACE1 (β-secretase 1) mRNA, which may mask the binding site for miR-485-5p and prevent translational repression of BACE1 mRNA by miRNA (Faghihi & Wahlestedt, 2009) is also get involved in regulation at the translational level, which can either repress, as exemplified by lincRNA-p21 on CTNNB1 and JUNB mRNAs (Yoon et al., 2012) or promote, as exemplified by AS Uchl1 on Uchl1 mRNA translation (Carrieri et al., 2012).

However, a peculiar mode of action is that lncRNAs can function as competing endogenous RNAs (ceRNAs) (Salmena, Poliseno, Tay, Kats, & Pandolfi, 2011) by binding to and sequestering specific miRNAs through their miRNA binding sites (also referred to as miRNA response elements, MRE). Mechanistically, ceRNAs function as 'miRNA sponges' to protect the target mRNAs from repression (Fatica & Bozzoni, 2013). Therefore, coding and non-coding RNAs can regulate crosstalk between each other by competing for shared miRNAs (Fatica & Bozzoni, 2013). This activity, preserved during evolution, was first discovered in plants (Franco-Zorrilla et al., 2007) and, subsequently, in mammals (Poliseno et al., 2010), in which it was shown to be relevant in many processes, including tumorigenesis (Karreth et al., 2011),(Sumazin et al., 2011), cell differentiation (Deng, 2015) and pluripotency. A clear example of this mechanism is represented by linc-MD1, a muscle-specific lncRNA, which is indispensable for muscle differentiation and plays an important role in myogenesis. linc-MD1

acts as a natural decoy for two muscle-specific miRNAs, miR-133 and miR-135. Expression of mastermind-like-1 (MAML1) is controlled by miR-133, and myocyte-specific enhancer factor 2C (MEF2C) is the target of miR-135 (Cesana et al., 2011). MAML1 and MEF2C are important myogenic factors required for activation of muscle-specific genes. MEF2C binds to the promoter region of cardiac muscle genes and positively regulates the differentiation of muscle cells, while MAML1 acts as a transcription coactivator in some signal transduction pathways (such as Notch signaling) related to muscle differentiation. With the depletion of linc-MD1, expression of both MAML1 and MEF2C is repressed, whereas overexpression of linc-MD1 resulted in high levels of MAML1 and MEF2C. These observations argue for a direct competition between linc-MD1 and the respective mRNAs for miRNA binding (Cesana et al., 2011) (Figure 1.9).

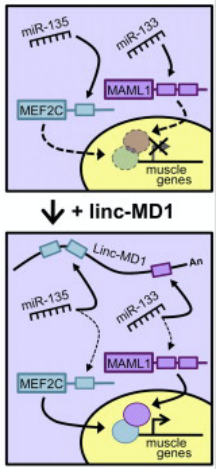


Figure 1.9 Schematic representation of ceRNA model of action
(Cesana, Cacchairelli et al., 2011)

Similarly, other lncRNAs function as miRNA sponges (e.g HULC to miR-372 (J. Wang et al., 2010), lncRNA-RoR to miR-145 (Y. Wang et al., 2013), H19 to miR-141 and miR-22 (Liang et al., 2015) by competing for miRNA binding and thus exerting post-transcriptional control.

Some pseudogenes represent an important source of sponge/target mimics for miRNAs to stabilize their homologous mRNAs, such as PTENP1, which is pseudogene of PTEN (phosphatase and tensin homolog) (Poliseno et al., 2010). An additional example of ceRNA was found in the newly identified class of circular RNAs (circRNAs) (Memczak et al., 2013).

The subcellular balance between ceRNA, one or multiple miRNAs, and mRNA targets constitutes a complex network allowing a fine-tuning of the regulation of gene expression in many biological processes.

1.8.4 Other classes of lncRNAs

UTR-Associated (ua) RNAs: the 3' UTR regions of eukaryotic genes can be transcribed into uaRNAs. They are regulated in a developmental stage- and tissue-specific manner and are evolutionarily conserved (T. R. Mercer & Mattick, 2013).

Promoter upstream transcripts (PROMPTs): are transcribed in sense and antisense directions at promoter regions and are highly unstable. Polyadenylation-dependent degradation of PROMPTs was proposed to ensure directional RNA production from otherwise bidirectional promoters (Preker et al., 2011).

Transcribed Ultra-conserved Region (T-UCR) lncRNAs: ultra-conserved regions (UCRs) are genome segments that exhibit 100% DNA sequence conservation between human, mouse, and rat. The human genome contains 481 UCRs within intragenic (39%), intronic (43%), and exonic (15%) sequences. These regions are transcribed into T-UCR lncRNAs, which are aberrantly expressed in different cancers and some are associated with poor prognosis (Bejerano et al., 2004).

Telomeric repeat-containing RNAs (TERRA): telomeres, which are protective nucleoprotein structures at the ends of chromosomes, are transcribed into non-coding RNAs. This family of transcripts is generated from both Watson and Crick strands in a cell cycle-dependent manner. Formation of RNA-DNA hybrids by TERRA at chromosome ends promotes recombination and delays senescence (Balk et al., 2013).

Centromeric lncRNAs: centromeric repeats are actively transcribed into lncRNAs during the progression from late mitosis to early G1 phase and they are required for correct kinetochore assembly and the maintenance of centromere integrity (Blower, 2016).

Promoter and pre-rRNA antisense (PAPAS): ribosomal (r)DNA loci were shown to be transcribed by RNA polymerase II. Their expression is induced in quiescent cells and triggers the recruitment of histone H4K20 methyltransferase Suv4-20h2 to ribosomal RNA genes for histone modification and transcriptional silencing (Bierhoff et al., 2014).

1.9 Techniques to probe lncRNA molecular interactions

Once the annotation of an uncharacterized lncRNA is done, it is necessary to determine if it actually possesses biological relevance. The major step towards understanding this is to determine if a non-coding transcript has any functional consequence in a particular context by gain- or loss-of-function analysis and the observation of the resulting phenotypes.

Two main approaches are followed to experimentally achieve successful knockdown of a particular lncRNA: transient methods like RNA interference (RNAi) or antisense oligonucleotide (ASO)-mediated depletion and CRISPR interference (CRISPRi) that involves Cas9-mediated modulation of gene transcription levels (Gilbert et al., 2014). In one of the first large-scale screens, Guttman's group used lentiviral shRNA-mediated analysis to analyze the depletion of 147 lncRNAs in mESCs (Guttman et al., 2011). A successful rescue of the phenotype is extremely important to prove that the phenotype is caused by the alteration of the lncRNA transcript and not by off-target effects.

To gain insight into the relationship between the function of lncRNAs and the molecules with which they interact, several *in vitro* and *in vivo* approaches are available. (1) RNA-Protein interaction: lncRNAs perform their regulatory functions in association with proteins, like chromatin-modifying complexes, transcription factors and RNP complexes. The first approach is protein-centered, using an antibody specific to a particular protein under native conditions, followed by immunoprecipitation and RNA purification, is named RNA immunoprecipitation (RIP) (J. Zhao et al., 2008), or Cross-Linking Immunoprecipitation (CLIP) using UV light (Spitzer et al., 2014). CLIP, HITS-CLIP or CLIP-seq can be considered to be powerful tools for studying protein-RNA interactions on a genome-wide scale when combined with high-throughput sequencing technologies. However, despite the high specificity, due to covalent cross-linking that permits stringent washing and RNase treatment, CLIP experiments often generate cDNA libraries of limited sequence complexity due to the restricted amount of co-purified RNA. The second approach is RNA-centered using biotinylated antisense probes to capture RNA-protein complexes but also to identify (2) RNA-Chromatin and (3) RNA-RNA domains of interaction. This approach includes RNA pull-down assay, chromatin isolation by RNA purification (ChIRP) (Chu, Quinn, & Chang, 2012), capture hybridization analysis of RNA targets (CHART) (Davis & West, 2015) and RNA antisense purification (RAP) (Engreitz

et al., 2014). Chromatin Isolation by RNA Purification (ChIRP) involves extensive cross-linking of cells with formaldehyde followed by capture of target RNA with short biotinylated antisense DNA oligos. The RNA-protein complexes are then captured on streptavidin beads and the enriched proteins are finally subjected to MS, while purified RNA and DNA fragments are sequenced following the same procedures as RNA-seq or DNA-seq, respectively. CHART is more similar to ChIRP: short antisense DNA oligonucleotides cover the entire target lncRNA but the candidate hybridization regions are empirically determined after an RNase H assay. Compared to ChIRP and CHART, the most distinctive feature of RAP is the use of long-RNA antisense capture probes (>60 nt), which form very stable RNA-DNA hybrids. In addition, RAP can be performed in various cross-linking conditions and under stringent hybridization and washing conditions that dramatically reduce off-target nucleic acids or proteins (Figure 1.10).

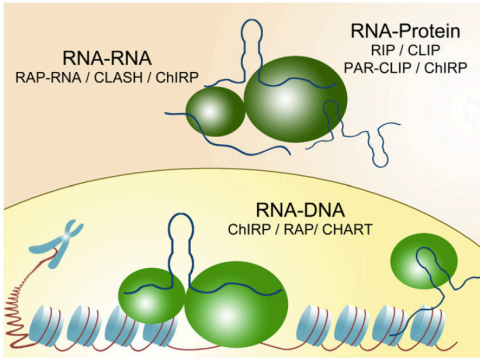


Figure 1.10 Schematic representation of methodologies to study lncRNA molecular interactions

1.10 Cell cycle and proliferation

The cell cycle is the sequence of events that occur in cells stimulated to grow. A complex network of signaling pathways regulates the entry into the cell cycle and is crucial for cell homeostasis.

Growth factors induce temporary resting cells, in G0 phase, to re-enter the cell cycle, which begins with a first period of growth (G1) during which it prepares for a period of DNA synthesis (S) (Blomen & Boonstra, 2007). Towards the end of G1 phase, there is an important restriction point (R) (Blagosklonny & Pardee, 2002), which marks the point where the cell becomes irreversibly committed to traverse the rest of the cell cycle. Growth factors act early in G1 phase to engage the cell-cycle signaling system, whose interaction with the growth-factor signaling pathways, represent a critical event in the control of cell proliferation. To understand how proliferation is controlled, it is therefore necessary to follow the sequence of events during the G1 phase that lead up to R. Once R is overcome, cells become independent of growth factors in that further progression through the cell cycle is now in the hands of the cell-cycle control system. Once duplicated the chromosomes, the cell enters a second period of growth (G2) where it prepares to divide into two daughter cells during the period of mitosis (M) (Berridge, 2012) (Figure 1.11).

Cyclins and CDKs are key players in cell-cycle regulation. In general, the CDKs, particularly CDK1, CDK2, and CDK4/6, are activated via binding to their selected cyclins, including cyclins A, B, D, and E, in specific phases of the cell cycle, followed by phosphorylation of their target proteins to enable cell-cycle progression. The activities of the CDKs are controlled not only by cyclins but also by phosphorylation or dephosphorylation by Cdc25 phosphatase (Nurse, 2002). Moreover, CDK inhibitors including p15, p16, p18, p21, p27, and p57 specifically bind to their target cyclin–CDK complexes and inhibit their activity to negatively regulate the cell cycle (Vidal & Koff, 2000). CDKs and their related pathways control the cell proliferation by maintaining exit from and entry to the different phases of the cell cycle.

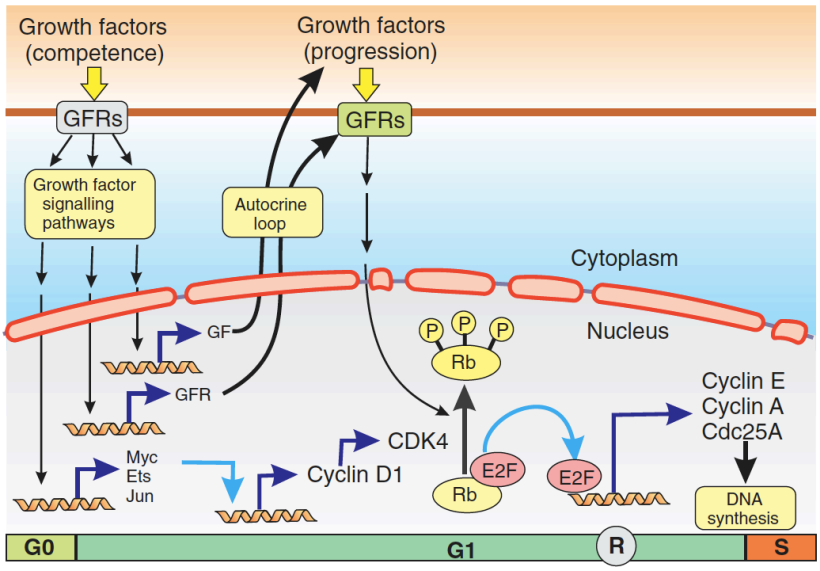


Figure 1.11 Spatiotemporal aspects of proliferative signaling
(Berridge, 2012)

Growth factors initiate a number of signaling pathways that carry information into the nucleus to activate immediate-early genes (IEGs). Some of these early protein-coding genes are involved in an autocrine loop to drive later events. Other early genes encode transcription factors, such as c-Myc, c-Jun and c-Fos, to initiate transcription of later genes, such as cyclin D, which combines with the cyclin-dependent kinases (CDKs), CDK4/6, to phosphorylate the retinoblastoma protein (Rb) which is one of the key factors guarding passage through the restriction point (R). A key event at R is the phosphorylation of Rb, which removes its inhibition of E2F and enables this transcription factor to increase the expression of cell-cycle components (e.g. cyclin E, cyclin A and Cdc25A) responsible for initiating DNA synthesis.

1.10.1 lncRNAs in cell proliferation

Advancements in genomics technologies over the last decade allowed for the identification of tens of thousands of lncRNAs previously unthought, of thereby revolutionizing the way we see the complex eukaryotic transcriptome and the consequent regulation of many biological processes.

Although the biological relevance for the majority of lncRNAs has not yet been assessed, those lncRNAs which have been functionally characterized are reported to play important roles in a variety of physiological processes, such as regulating cell proliferation and differentiation, apoptosis, imprinting, promoting pluripotency, and controlling gene expression.

Many lncRNAs are involved in the modulation of critical cell-cycle regulators such as cyclins, CDKs, CDK inhibitors, pRb and p53 via various mechanisms. These lncRNAs principally act as epigenetic regulators of transcription, post-transcriptional regulators, and protein scaffolds.

Gadd7 is a lncRNA involved in regulating CDK6 expression in a post-transcriptional manner (Liu et al., 2012). It is transcriptionally induced via DNA damage mediated by UV and cisplatin, binds to TDP-43 and dissociates from CDK6 mRNA. The CDK6 mRNA is then degraded, resulting in inhibition of the G1/S transition. MALAT1, a mediator of mRNA splicing (Tripathi et al., 2010), is upregulated in several human cancers and contributes to cancer cell proliferation (Gutschner & Diederichs, 2014). MALAT1 depletion results in arrest in the G1 phase and promotes expression of p53 as well as p16, p21, and p27 in human fibroblasts. On the contrary, it suppresses various genes involved in cell-cycle progression such as cyclin A2 and Cdc25A. Moreover, in G2/M progression, MALAT1 is required for expression of B-Myb, thereby contributing to cell-cycle progression in each phase (Tripathi et al., 2013). ANRIL is an anti-sense transcript of the p15 gene in the INK4 locus induced by DNA-damaging agents via the ATM-E2F1 pathway and it is involved in epigenetic repression of the transcription of the INK4 locus via the recruitment of PRC1 and PRC2 (Wan et al., 2013).

The integrity of the genome is maintained by a process of checkpoint signaling that operates to arrest the cell cycle in response to DNA damage. Tumor protein p53, described as “the guardian of the genome” (Lane, 1992), is crucial for the induction of growth arrest by numerous stress signals. Moreover, the importance of the tumor suppressing functions of p53 is shown by its high mutation frequency in cancers and by the highly tumorigenic phenotype of p53 null mice-RoR negatively regulates p53 expression, thereby suppressing doxorubicin-induced G2/M phase arrest and apoptosis. Conversely, depletion of lncRNA-RoR leads to p53 accumulation (Y. Wang et al., 2013). Pint lncRNA (p53 induced non-coding transcript), a tumor suppressor candidate lncRNA, is finely regulated by p53 and connects p53 activation with epigenetic silencing by PRC2 in a p53 auto-regulatory negative mechanism. In addition, p53 directly binds to lncRNA-p21 promoter-p21 transcription is up-regulated upon DNA damage in MEFs (Huarte et al., 2010).

PANDA is induced by DNA damage in a p53-dependent manner. It binds to and inhibits the NF-YA transcription factor, which limits the expression of pro-apoptotic genes such as FAS and BIK and thus results in the repression of apoptosis (Hung et al., 2011).

c-Myc is a basic helix-loop-helix leucine zipper (bHLHZ) transcription factor, induced in early G1 phase by serum and a variety of mitogens, which bind the consensus DNA sequence known as the E-Box (CACGTG) when dimerized with Max (Kato & Dang, 1992). c-Myc is a master regulatory transcription factor that modulates a large number of cellular processes, such as proliferation, growth, differentiation, metabolism, and even apoptosis (Eilers & Eisenman, 2008). Several reports have identified lncRNAs as components of c-Myc transcriptional network and many lncRNAs can be found differentially regulated by c-Myc in different cancer types. Some of them have been shown to be able to control the expression of c-Myc itself, both at transcriptional and post-transcriptional levels (Figure 1.12).

Name	Alternative Name	Ensembl Gene ID	Regulation
lncRNAs regulated by MYC			
H19		ENSG00000130600	positive
DANCR	ANCR	ENSG00000226950.2	positive
SNHG16	NcRAN	ENSG00000163597	positive
MINCR	LINC01604	ENSG00000253716.1	positive
AK021907	MYCLo-1	None	positive
ELFN1-AS1	MYCLo-2	ENSG00000236081	positive
KTN1-AS1	MYCLo-3	ENSG00000186615	positive
AK098037	MYCLo-4	None	negative
LPP-AS2	MYCLo-5,6	ENSG00000270959	negative
VPS9D1-AS1	MYU	ENSG00000261373	positive
ENST00000553181.5	lncRNA-MIF	ENSG00000257135	positive
BCYRN1	BC200a	None	positive
lncRNAs affecting MYC Expression			
PVT1	LINC00079	ENSG00000249859	positive
CCAT1	CARLo-5	ENSG00000247844	unknown
CCAT1-L		ENSG00000247844	positive
CCAT2	LINC00873	ENSG00000280997.1	unknown
PCAT1		ENSG00000253438	positive
MYCNUT	lncUSMycN	ENSG00000223850	positive
MYCNOS		ENSG00000233718	positive
lncRNAs affecting MYC stability/translation			
GHET1		ENSG00000281189.1	positive
LINC-ROR	CTD-2526M8.1	ENSG00000258609	positive
ENST00000553181.5	lncRNA-MIF	ENSG00000257135	negative
GAS5	SNHG2	ENSG00000234741.7	negative
lncRNAs affecting MYC transcriptional activity			
PCGEM1	PCAT9	ENSG00000227418.6	positive
MINCR	LINC01604	ENSG00000253716.1	positive
LPP-AS2	MYCLo-6	ENSG00000270959	negative

Figure 1.12 lncRNAs that have been described as being part of the c-Myc network (Ingram et al., 2017)

Moreover, well-characterized miRNAs participate in cell-cycle regulation through post-transcriptional regulation (Bueno & Malumbres, 2011). One of the common miRNA families is the *let-7* family which has been shown to repress cell-cycle regulators (e.g., cyclin A, cyclin D and CDK4) and to block cell-cycle progression and anchorage-independent growth in cancer cells (Johnson et al., 2007). Additionally, *let-7a* inhibits c-Myc-induced cell growth in Burkitt lymphoma cells by blocking c-Myc expression (Sampson et al., 2007). Further, HuR RNA-binding protein (RBP) binds and represses c-Myc mRNA by recruiting the *let-7*/RISC complex to the 3' UTR region of c-Myc (H. H. Kim et al., 2009). Particularly, recruitment of HuR and *let-7* to the transcript of c-Myc is interdependent (Gunzburg et al., 2015).

1.11 Growth-factor signaling pathways

Growth factors use a variety of signaling pathways that co-operate with each other to transmit information into the nucleus. The interchange between the cell-cycle signaling system and growth-factor signaling pathways is a critical event in the control of cell proliferation. In general, cyclins are the primary targets of the signals coming into the cell from growth-factor receptors on the cell surface. Proliferative signaling pathways drive cells into the cell cycle by activating events early in G1. One of these pathways depends on Ras and the mitogen-activated protein kinase (MAPK) signaling pathway. Another very common example is the Wnt signaling pathway. In contrast to the positive effects of certain growth factors in initiating the cell cycle, other growth factors such as transforming growth factor-β (TGF-β) act by inhibiting the cell-cycle signaling pathway in a particular context (Figure 1.13).

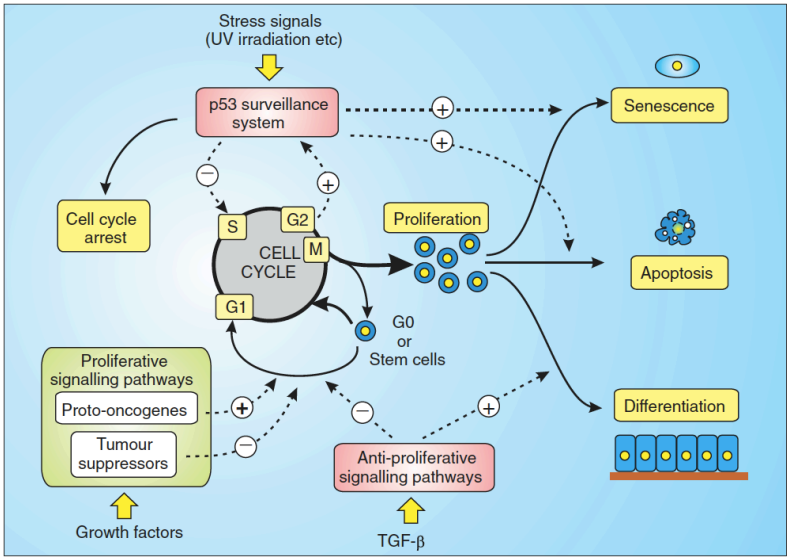


Figure 1.13 Cell-cycle network of signaling systems of cell proliferation and related cellular processes
(Berridge, 2012)

The interest in signaling pathways regulating cell proliferation increased with growing evidence of their alteration in many forms of cancer. Very often key components of these signaling pathways are proto-oncogenes that become constitutively active to function as oncogenes. Conversely, there are negative elements, and many of these are tumor suppressors, which are inactivated in many tumor cells. In addition to the protein-coding genes, also lncRNAs are deregulated in a number of cancers, suggesting that their aberrant expression may be a substantial contributor in cancer development (Gutschner & Diederichs, 2014).

1.11.1 Transforming growth factor-β signaling pathway: the inhibitory SMAD, SMAD7, in colorectal cancer

Transforming growth factor-β (TGF-β) signaling pathway plays crucial roles during development and homeostasis and exerts strong anti-proliferative effects on normal and premalignant cells. However, advanced stage cancers often become insensitive to the tumor-suppressive actions of TGF-β signaling. Indeed, TGF-β signaling is involved in epithelial-to-mesenchymal transition (EMT) induction, angiogenesis promotion, altered extracellular matrix deposition, immune suppression, and increased metastatic colonization (Ikushima &

Miyazono, 2010).

SMAD family members have been identified as essential genes for the intracellular mediation of TGF-β signaling (Massagué, 2008). In the canonical pathway, the activation of TGF-β and BMP branches leads to the recruitment of R-SMADs, SMAD2/3 in the case of TGF-β or SMAD1/5/8 in the case of BMP, which form a complex with SMAD4. This complex then directs the transcription of many target genes, including SMAD7 (Figure 1.14).

SMAD7, a I-SMAD like SMAD6, in turn serves as a negative feedback regulator of TGF-β signaling (Wrighton, Lin, & Feng, 2009). Indeed, it forms stable complexes with activated type I receptors and thereby blocks the phosphorylation of R-SMADs, or recruits ubiquitin E3 ligases, such as SMURF1/2, resulting in the ubiquitination and degradation of the activated type I receptors. Besides, SMAD7 also inhibits TGF-β/BMP signaling in the nucleus by interacting with transcriptional repressors, such as histone deacetylases, HOXC-8 and CtBP, or disrupting the formation of TGF-β induced functional SMAD-DNA complexes (S. Zhang et al., 2007). The level of SMAD7 transcription is also upregulated by other stimuli, including IFN-γ, TNF-α as well as ultraviolet light and TPA, by which SMAD7 mediates the crosstalk between TGF-β and other signaling pathways.

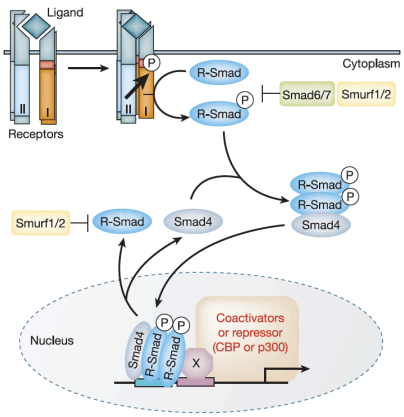


Figure 1.14 General mechanism of TGF-β signaling
(Rik Derynck & Ying E. Zhang, 2008)

TGF-β superfamily signaling is established as a crucial signaling family in colorectal cancer (CRC). CRC is the third most common malignant tumor and the fourth most common cause of cancer mortality worldwide (Center, Jemal, & Ward, 2009). The onset of CRC involves

multi-factorial and complex steps in which abnormal gene expression, due to mutational events, plays an important role. Also SMAD7 is involved in the etiology of CRC, indeed CRC patients with deletion of SMAD7 have a favorable outcome compared to patients with SMAD7 amplification (Boulay et al., 2003). SMAD7 overexpression has been observed in some CRC cells, and reduction of SMAD7 expression using antisense RNA leads to decreased proliferation in the HCT-116 CRC cell line and human CRC neoplastic explants, and it further leads to reduced tumorigenesis in $APC^{min/-}$ mice (Stolfi et al., 2014). Moreover, stable expression of SMAD7 promotes tumor progression and enhances liver metastasis of CRC (Halder, Rachakonda, Deane, & Datta, 2008).

Aim of the thesis

Although increasing numbers of lncRNAs are reported, their functionality remains mostly elusive.

Essentially, an accurate transition from G1 phase of the cell cycle to S phase is crucial for the control of eukaryotic cell proliferation, and its deregulation promotes oncogenesis. Several studies used cells that re-enter the cell cycle, after serum removal and subsequent stimulation, to analyze transcriptional activation in the G1 phase. Many protein-coding genes control this process but little is known about the regulatory role of immediate-early lncRNAs.

This thesis aims to identify, by a high-throughput RNA-seq approach, novel cell-cycle regulated lncRNAs (CCR lncRNAs) and characterize their molecular mechanisms of action.

A common theme in the regulation of gene expression by lncRNAs is the activation or repression of their target genes. Indeed, lncRNAs can accumulate *in cis*, localize in the nucleus *in trans*, or in the cytoplasm to execute their function via distinct regulatory mechanisms.

Thus, this thesis intends to highlight the regulatory circuits that involve immediate-early lncRNA action at transcriptional or post-transcriptional level in the context of cell proliferation and decipher whether their deregulation might have an impact in a pathological condition such as tumorigenesis.

Chapter 2

High-throughput RNA-seq analysis identifies
novel immediate-early lncRNAs involved in cell proliferation

Results

2.1 Cell-cycle re-entry by serum promotes the expression of immediate-early lncRNAs

Normal cells, like fibroblasts, require growth factors to proliferate, which are provided *in vitro* by the addition of fetal bovine serum (FBS). In serum deprivation conditions, the cells enter in a quiescent state, termed G0, while the addition of serum triggers a proliferative response. We used this system that represent a traditional model for studying cell-cycle progression and the identification of the early genes which are induced by mitogens to progress into the in G1 phase (Figure 2.1A). We first performed a fluorescence-activated cell sorting (FACS) analysis with propidium iodide staining to determine the percentage of cells in G0/G1, S and G2/M phase. We found that serum starved mouse embryonic fibroblasts (MEFs) contained higher percentages of cells in G0/G1 (Figure 2.1B, left panel). Moreover, we observed, by RT-qPCR analysis, an increase in the expression levels of some well-known early protein-coding genes, such as c-Jun, c-Fos and c-Myc (Figure 2.1B, right panel), confirming the accuracy of the proliferative model.

To identify novel long non-coding RNAs (lncRNAs) expressed as immediate-early genes upon mitogenic stimulation, we extracted the RNA from MEFs harvested at different time points after serum treatment (0.5h, 1h and 2h) and we performed a high-throughput RNA sequencing (RNA-seq) analysis. From the sequencing, we obtained short-read gapped alignment yielded approximately 2.4 billion mappings on the mouse genome. Reference annotation based transcript (RABT) assembly of mapped reads generated about $3x10^5$ potential transcripts. We then designed a filtering pipeline to select for previously unannotated transcripts lacking coding potential. All single exon transcripts, transcripts shorter than 200 bases, transcripts with low coverage (<3 reads per base on average), transcripts that have been previously annotated (in RefSeq, UCSC, Ensembl, and Vega annotation datasets) and transcripts with putative coding potential were excluded (Figure 2.2A). We selected only those transcripts whose expression was significantly upregulated at any time point following serum induction. This analysis produced a set of novel 19 high-fidelity long non-coding transcripts upregulated by serum treatment in starved MEFs (Figure 2.2B). The validation by RT-qPCR of the novel lncRNA revealed that only 10 of them showed a good expression profiles with specific peaks of expression, under serum stimulation, in BALB/c fibroblasts (Figure 2.3). We name these novel immediate-early transcripts cell-cycle regulated (CCR) lncRNAs.

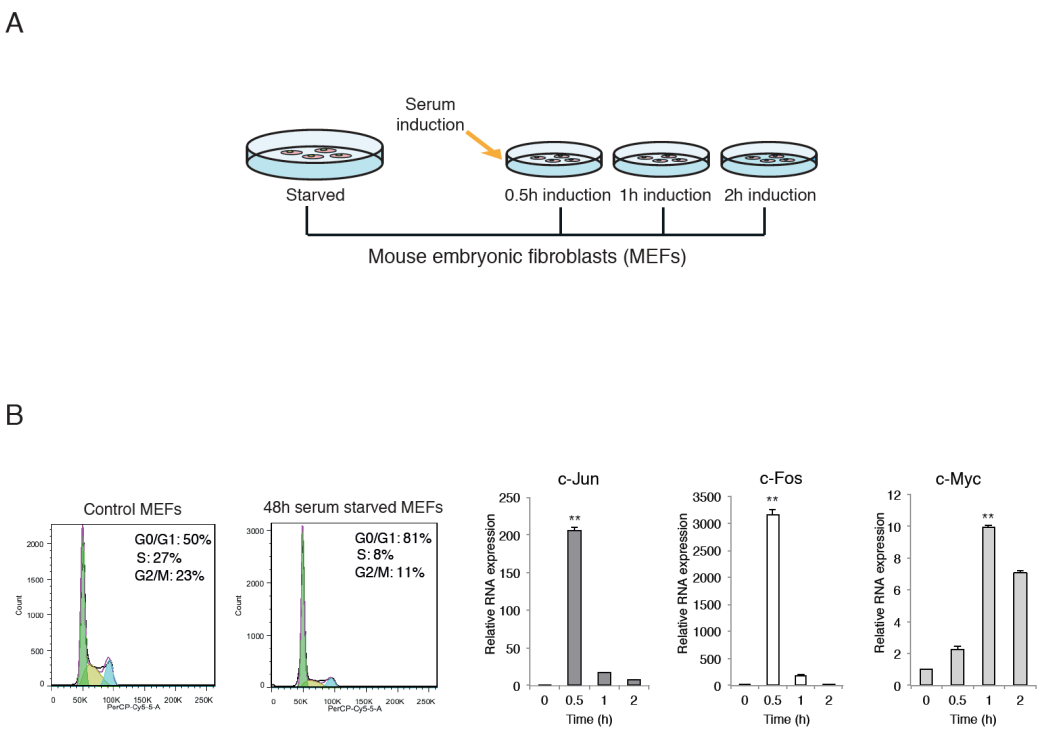
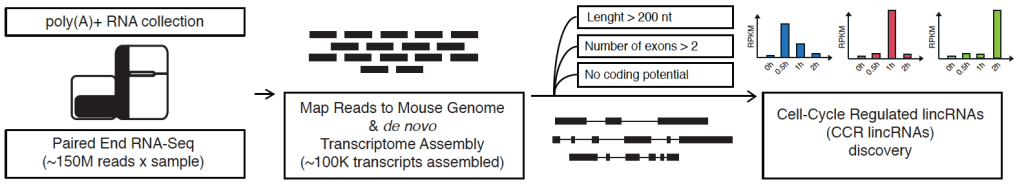


Figure 2.1 Time course of serum induction in MEFs.

A. Experiment layout to identify serum-dependent transcription. Starved MEFs were treated with serum for the times indicated.

B. Cell-cycle FACS analysis of starved MEFs compared to serum-induced MEFs (left panel). RT-qPCR analysis of: c-Jun, c-Fos and c-Myc transcripts in serum time course, as controls of serum induction (right panel). Data are presented as mean values ± SD of 3 independent experiments; **P < 0.01, *P < 0.05 by the Student's t-test.

A



B

CCR_lincRNAs Symbol	Genomic Coordinates	Time after serum induction (RPKM) 0	0.5	1	2
CCR060	chr1:188410855-188439211	0,14160	0,24083	1,10166	1,05795
CCR065	chr10:21478232-21505858	0	0,01921	0,73612	0,57483
CCR071	chr15:11748729-11758896	0,04945	0,21453	0,19886	0,13986
CCR073	chrX:65783166-65800793	0,20826	0,14290	0,45305	0
CCR102	chr18:75518303-75524655	0,67972	0,68417	1,07602	1,70354
CCR115	chr2:160430210-160445851	0,34222	0,24436	0,34764	0,70412
CCR188	chr8:105762496-105812526	0,08039	0,09097	0,09252	0,03570
CCR212	chr13:46060501-46061828	2,25441	2,37400	4,40398	0
CCR250	chr15:38380574-38385000	0	0,02181	0,37934	0
CCR301	chr10:120125496-120139854	0,28095	0,37205	1,51241	2,08313
CCR306	chr17:73419743-73422205	0,13933	0,15091	0,14424	0,35251
CCR327	chr17:71220042-71221981	0,03979	0,05998	0,08477	0,02067
CCR339	chr18:46377975-46385421	0,03026	0,15564	0,04387	0,03175
CCR374	chr2:94000791-94026128	0,01188	0,01286	0,14002	1,06417
CCR363	chr19:53496774-53498840	0	0	0,11442	0,17511
CCR492	chr4:84387297-84390980	0	0,01953	0,06285	0,11043
CCR548	chr6:31014639-31038730	0,06948	0,09374	3,01249	0
CCR559	chr6:115476690-115490587	0	0,00648	0,00411	0
CCR576	chr6:42338402-42342246	0,00975	0,03279	0,05759	0

Figure 2.2 Discovery of novel CCR lncRNAs

A. A schematic overview of the lncRNAs discovery pipeline employed: (1) paired-end reads from Illumina sequencing are mapped to the mouse genome, (2) a parsimonious set of transcripts justifying the observed read is built using Cufflinks reference annotation based transcript (RABT) assembly, (3) transcripts are filtered by length, number of exons, and coding potential, (4) transcripts induced by serum are selected at each time point.

B. List of the novel 19 CCR lncRNAs discovered by our RNA-seq approach in term of genomic coordinates and gene expression levels (RPKM) during serum stimulation in MEFs.

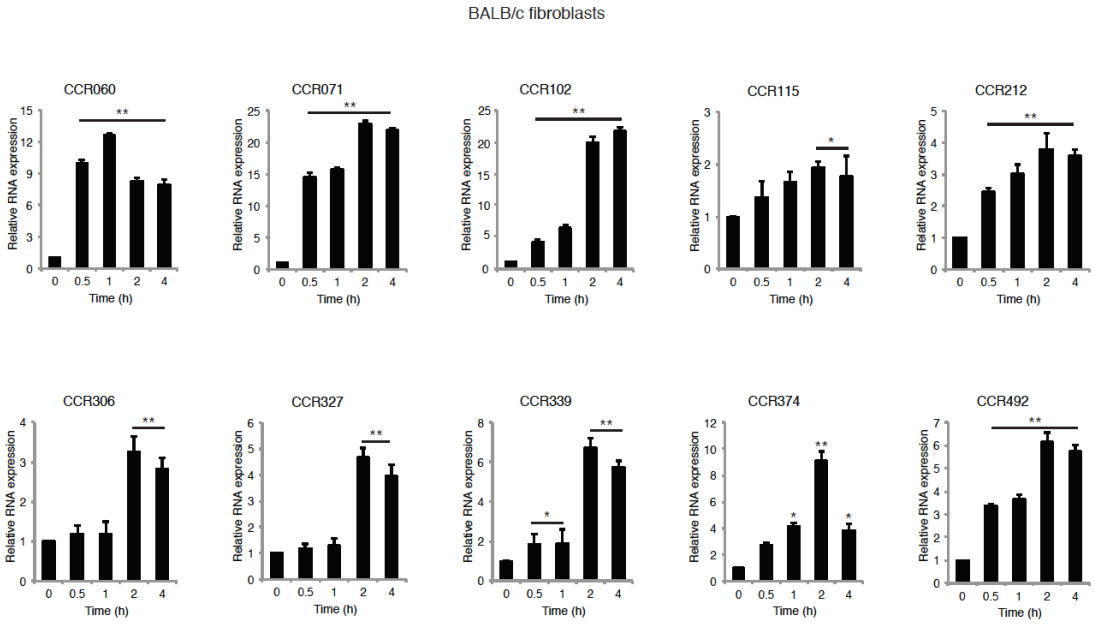


Figure 2.3 CCR lncRNAs validation by RT-qPCR in BALB/c fibroblasts

Validation of CCR lncRNA expression levels by RT-qPCR, with specific oligonucleotides, in serum-induced BALB/c fibroblasts. The gene expression values were normalized to β-actin. Data are presented as mean values ± SD of 3 independent experiments; **P < 0.01, *P < 0.05 by the Student's t-test.

2.2 Two novel CCR lncRNAs show a significant phenotype in cell proliferation

To explore the biological role of the identified CCR lncRNAs, we performed loss-of-function experiments by transfecting BALB/c fibroblasts with short independent hairpin RNAs (shRNAs) against each lncRNA (Figure 2.4A). Among the 10 CCR lncRNAs, only the silencing of CCR492 and CCR102 resulted in a significant decrease in cell growth with respect to control cells (Figure 2.4B).

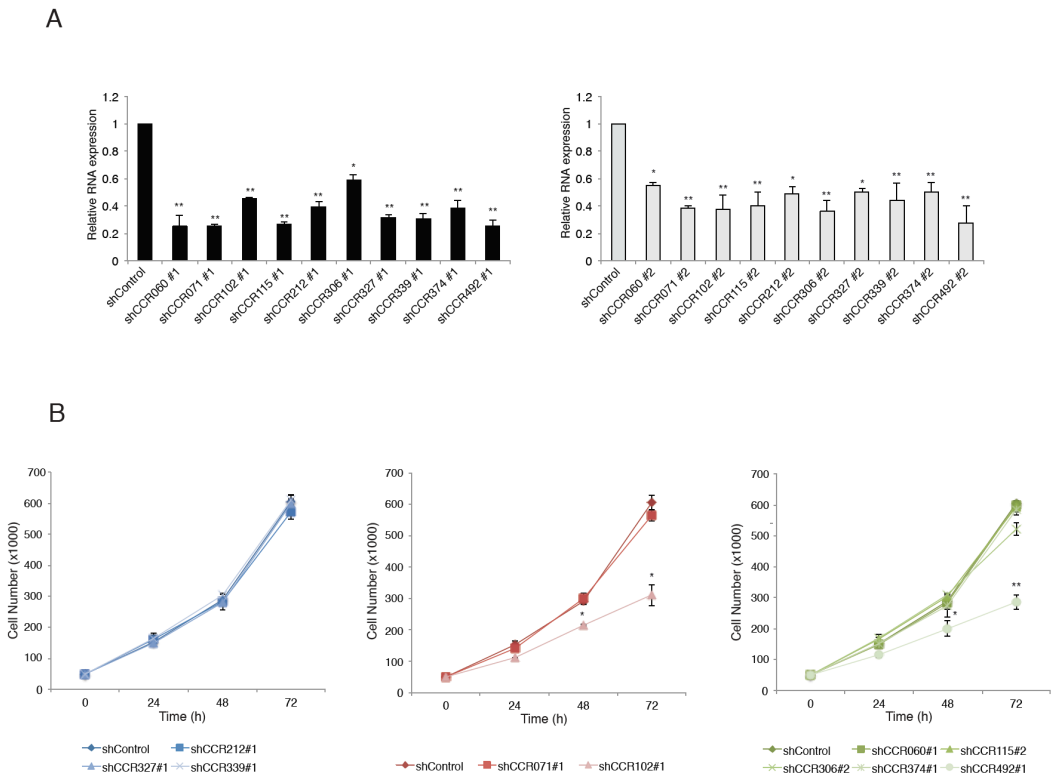


Figure 2.4 Selection of CCR lncRNAs with a phenotype in cell proliferation

A. BALB/c fibroblasts were transfected with two independent shRNAs for each lncRNA. LncRNA expression levels were evaluated by RT-qPCR.

B. Screening of the biological phenotype in loss-of-function experiment. BALB/c fibroblasts were transfected with the best shRNA and analyzed in cell growth, compared to shControl. Data are presented as mean values ± SD of 3 independent experiments; **P < 0.01, *P < 0.05 by the Student's t-test.

Chapter 3

CCR492 lncRNA acts as a ceRNA
reducing the let-7 repressing activity on c-Myc
and contributes to cell transformation

Results

3.1 Characterization of CCR492 lncRNA

The CCR492 gene is localized on chromosome 4, in a 200 kb-long intergenic region between the Bnc1 and Cntln protein-coding genes. The promoter specific H3K4me3 modification and the presence of the H3K36me3, along the gene body, indicated its active transcription in fibroblasts. It is composed of 4 exons, as shown by the RNA-seq profile (Figure 3.1A). Importantly, northern blot analysis and rapid amplification of cDNA ends (RACE) revealed that CCR492 is a transcript of 953 nucleotides (Figure 3.1B-C).

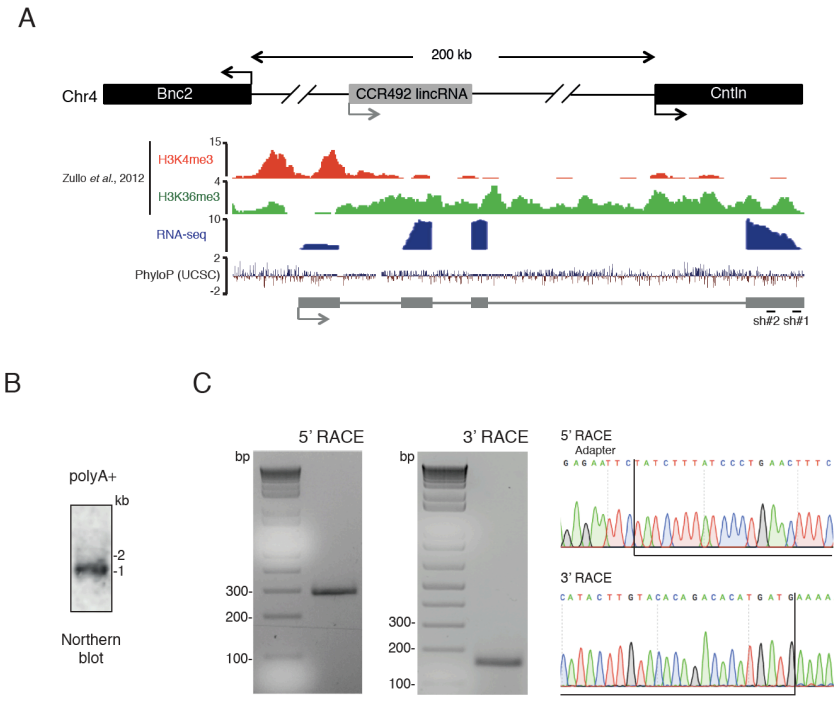


Figure 3.1 Characterization of CCR492 lncRNA

A. Schematic representation of the chromosomal location of the CCR492 lncRNA. The arrowhead indicates the transcription direction. Chromatin structure is shown from fibroblasts ChIP-seq data for histone modifications (Zullo et al, 2002). Positions of shRNAs directed against exon 4 are indicated.

B. Northern blot analysis for CCR492 in polyA$^+$ fraction from serum-induced fibroblasts confirms the predicted size of transcript (~1 kb).

C. CCR492 mature RNA start and stop sites (mapped by 5' and 3' RACE, respectively).

D. Expression of CCR492 in BALB/c fibroblasts transfected with either a non-targeting control (shControl) or shRNAs targeting lncRNA CCR492 (shCCR492 #1, #2). Data are presented as mean values ± SD of 3 independent experiments; **P < 0.01, *P < 0.05 by the Student's t-test.

3.2 CCR492 controls cell-cycle progression in fibroblasts

To investigate the potential biological role of CCR492 in cell-cycle progression, we transfected two shRNAs targeting CCR492 (shCCR492 #1, #2) or shControl in fibroblasts. Cell-growth curve assay showed a significant reduction in cell proliferation of CCR492 silenced cells, with respect to control cells (Figure 3.2 A, Figure 3.2B top panel). Analysis of the cell cycle by FACS revealed an increase of G1-G0-phase cells, together with a reduction of S-phase cells upon CCR492 knockdown (Figure 3.2B bottom panel). To verify that the effect on cell proliferation was not due to off-target effects of the shRNAs, we performed a complementation experiment by expressing two different CCR492 transcripts, a site-specific mutant and a deleted mutant, which are not affected by silencing (Figure 3.2C). Analysis of cell proliferation showed that ectopic expression of the lncRNA was able to rescue the wild-type proliferation phenotype, supporting the involvement of immediate-early CCR492 lncRNA in regulating the cell cycle in fibroblasts (Figure 3.2D-E).

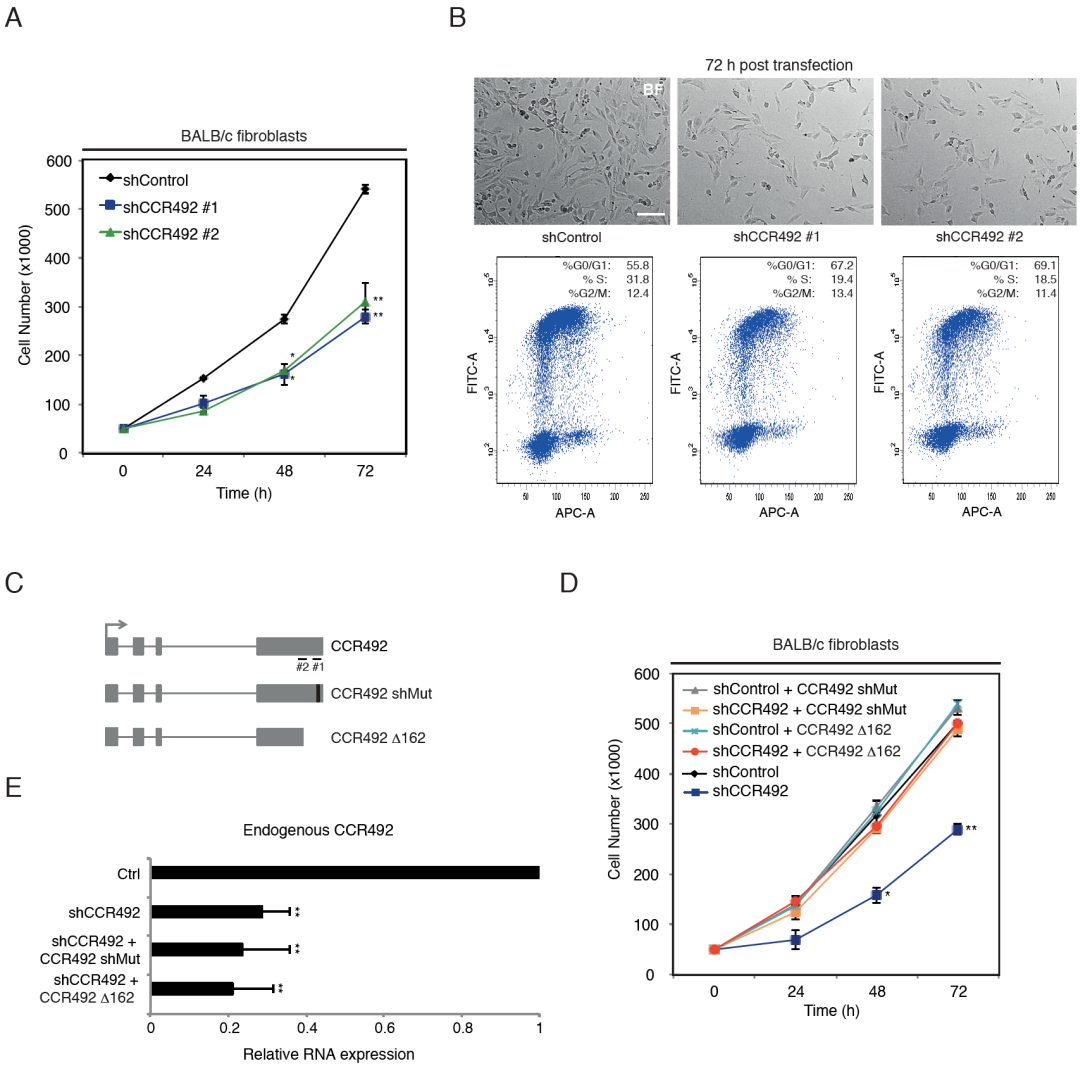


Figure 3.2 CCR492 depletion suppresses the G1/S progression

A. Cell-growth curve assay in control (shControl) and CCR492 silenced fibroblasts (shCCR492 #1, #2)

B. Bright field of control and CCR492 silenced fibroblasts (Top panel). EdU-CellCycle 633-red flow cytometry analysis. Cells were pulsed with EdU for 1 h following 72 h transfected with shControl or shCCR492 #1, #2. Scatter plot histograms of EdU-labeled cells were stained for DNA content (X-axis) and EdU (Y-axis). The following populations are shown: G0/G1, S, and G2/M (Bottom panel).

C. Schematic representation of the CCR492 WT, CCR492 shMut and CCR492 Δ162 used for the rescue experiment.

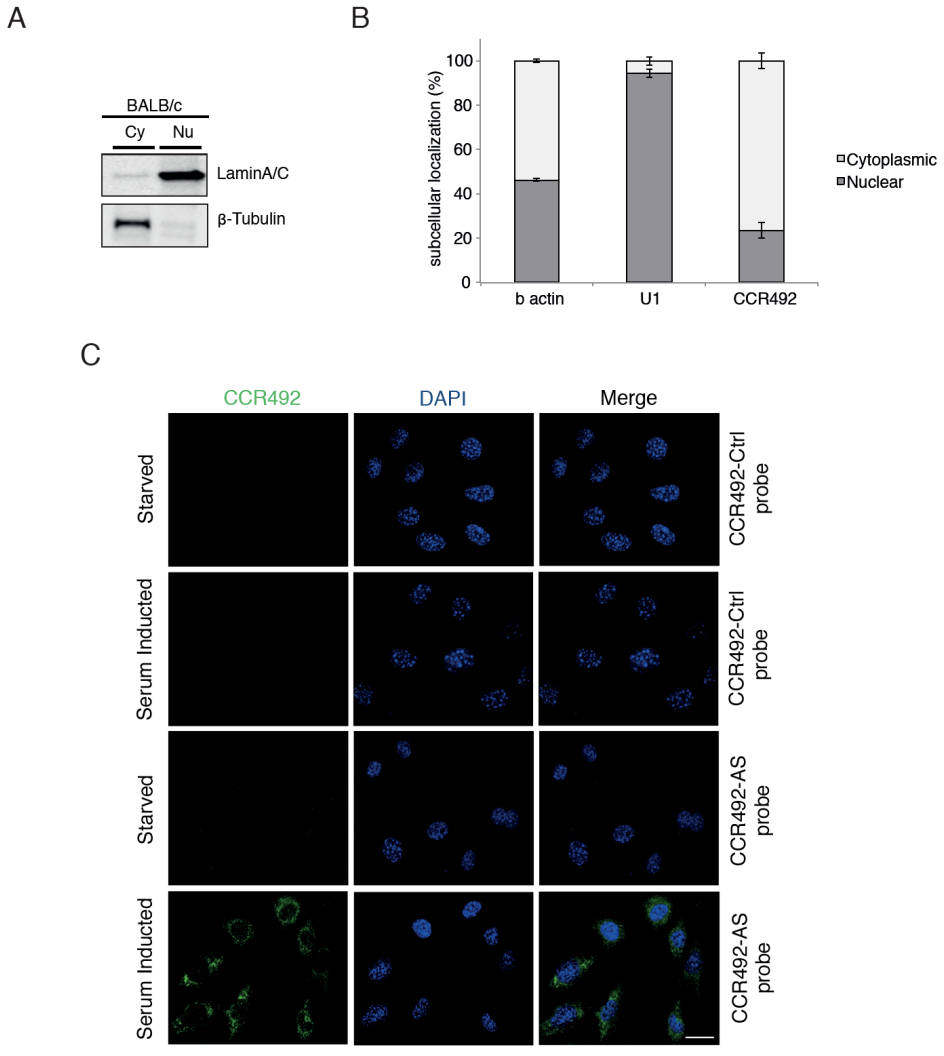
D. Expression levels of the endogenous CCR492 were determined by RT-qPCR.

E. Cell-growth curve assay of CCR492 silenced fibroblasts in the absence or presence of the CCR492 Mutants. Data are presented as mean values ± SD of 3 independent experiments; **P < 0.01, *P < 0.05 by the Student's t-test.

3.3 CCR492 is a cytosolic lncRNA

Next we examined the localization of CCR492 in BALB/c fibroblasts by nuclear and cytoplasmic subcellular fractionation (Figure 3.3A). The RT-qPCR showed that CCR492 RNA was prevalently present in the cell cytoplasm (Figure 3.3B).

We further analysed the cellular distribution of CCR492 by RNA Fluorescent in Situ Hybridization (RNA-FISH). CCR492 signal was visible in the cytoplasm while no signal was evident in the DAPI-stained nuclei (Figure 3.3C).



(Figure 3.3 – legend on the next page)

Figure 3.3 CCR492 is a cytoplasmic lncRNA

A. Western blot of subcellular fractionation in BALB/c fibroblasts. Controls of subcellular fractionation are β-Tubulin for cytoplasm, and Lamin A/C for nucleus.

B. RT-qPCR of subcellular fractionation in serum-induced BALB/c fibroblasts. Percentage ratio of CCR492 nuclear and CCR492 cytoplasmic over the whole total CCR492 expression levels represents the distribution of CC492 in fibroblasts. Controls of subcellular fractionation are β-actin for cytoplasm, and U1 for nucleus.

C. RNA Fluorescent In Situ Hybridization (RNA-FISH) of CCR492 in starved or serum-induced BALB/c fibroblasts. Top shows the negative control obtained with sense 5'-biotinylated CCR492 probes. Bottom shows single-molecule CCR492 localization obtained with 5'-biotinylated antisense (CCR492-AS) probes. Data are presented as mean values ± SD of 3 independent experiments; **P < 0.01, *P < 0.05 by the Student's t-test.

3.4 CCR492 acts as a molecular sponge modulating let-7 activity

Considering that CCR492 is expressed in the cytoplasm in response to growth-factor treatment to favor the cell-cycle progression, we hypothesized that it could act as a competing endogenous RNA (ceRNA) for miRNAs targeting genes involved in cell growth.

By performing small RNA-seq we identified the top 20 miRNAs expressed in MEFs, and bioinformatically evaluated the presence of the corresponding MREs on CCR492. Our analysis revealed that CCR492 contains 2 or more putative MREs for 13 of them (Table 3). We focused our attention to the let-7 miRNAs, which showed 4 high-affinity MREs to CCR492 (Figure 3.4A-B-C) and are known to play a role in cell growth (Johnson et al., 2007).

To evaluate the relationship between CCR492 and let-7 we performed a cell growth assay in the presence of let-7 inhibitors. The silencing of CCR492 does not affect cell growth in the presence of let-7 inhibitors indicating that CCR492-dependent phenotype requires the presence of let-7 (Figure 3.4D-E). Taken together these experiments show that CCR492 acts by titrating let-7 to limit its activity.

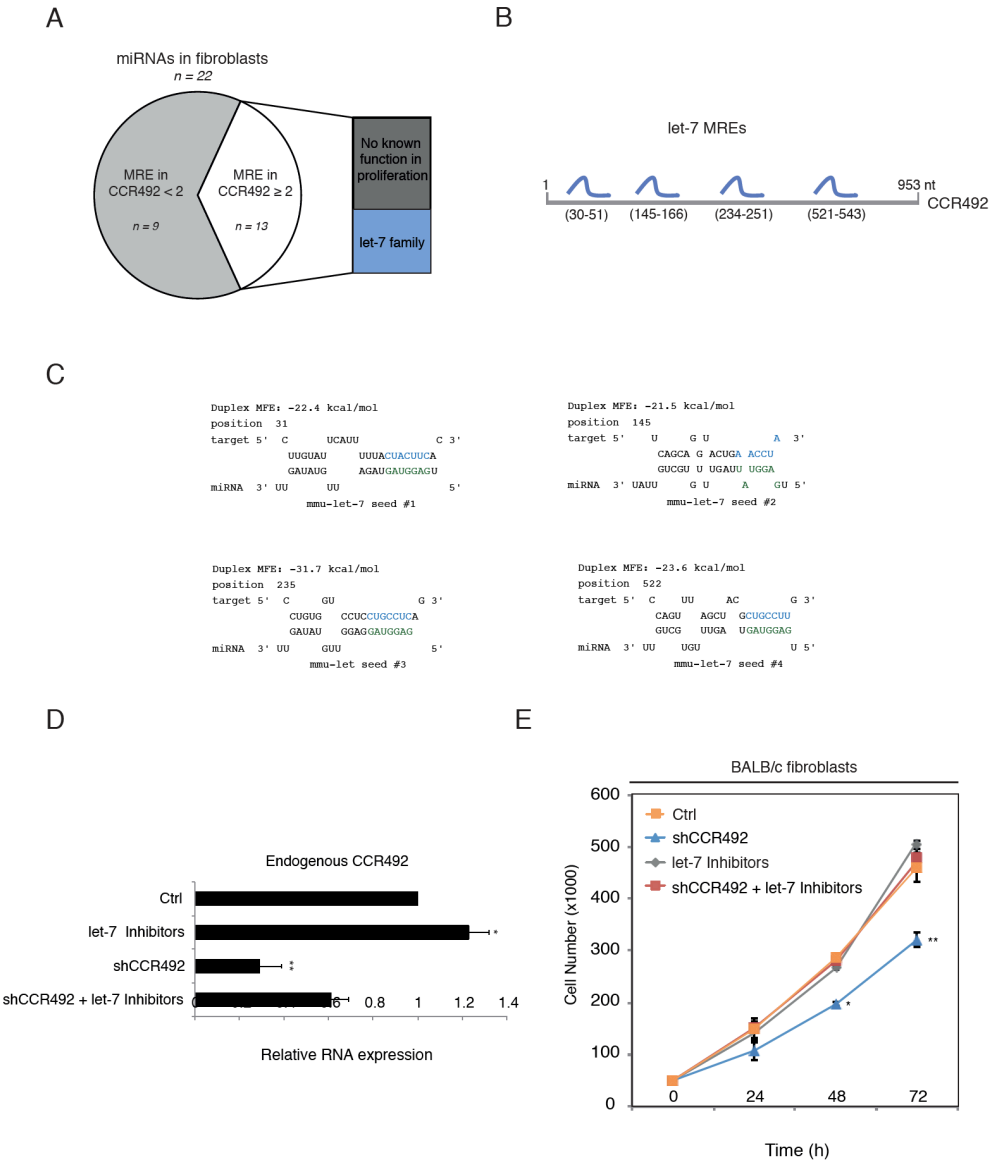


Figure 3.4 CCR492 acts as a natural decoy for let-7

A. Analysis of potential MREs in CCR492.

B. Schematic representation of the 4 bioinformatics predicted let-7 binding sites in CCR492.

C. Bioinformatic prediction of target-miRNA seed regions in CCR492 for let-7. Duplex MEF (ΔG seed duplex, Minimum Free Energy) values were obtained from RNAhybrid program (Kruger & Rehmsmeier, 2006)

D. Expression levels of CCR492 in control and CCR492 silenced fibroblasts, in the absence or presence of let-7 Inhibitors.

E. Cell-growth assay with let-7 Inhibitors in CCR492 depleted BALB/c fibroblasts. Data are presented as mean values ± SD of 3 independent experiments; **P < 0.01, *P < 0.05 by the Student's t-test.

.

3.5 CCR492 interacts directly with let-7 to regulate c-Myc expression

The intersection of the serum-induced genes and those predicted to be let-7 target genes identified 14 putative target RNAs (Figure 3.5A). Most of these are not related with the cell cycle and are not affected by CCR492 silencing (Figure 3.5B). Interestingly, c-Myc, which is known to play a central role in cell proliferation, showed a significant decrease of expression in CCR492 silenced cells both at the mRNA and protein levels, but not at the c-Myc hnRNA levels (Figure 3.5B-C-D). Demonstrating that this regulation is at the post-transcriptional level.

To verify whether the effect of CCR492-dependent depletion of let-7 acted directly on the c-Myc RNA, we cloned the 3'UTR of c-Myc containing the let-7 binding site downstream of a firefly luciferase reporter gene (pRLuc-c-Myc 3'UTR), and analysed the effect of CCR492 knockdown on the c-Myc 3'UTR. The inhibition of CCR492 expression resulted in a significant reduction of the luciferase expression with respect to the control plasmid. Importantly, the inhibitory effect on the luciferase construct by CCR492 silencing was abolished either by the co-expression of the CCR492 mutant or by let-7 inhibitors (Figure 3.6A). To determine the functional interaction between CCR492 and let-7, we performed an RNA pull-down assay. CCR492 pull-down strongly is enriched in let-7 and Ago2, while we could not detect c-Myc RNA (Figure 3.6B). Thus, demonstrating that CCR492 and let-7 are associated together within a RISC complex (Figure 3.6C). Importantly, CCR492 mutated all its let-7 binding sites was no longer able to rescue the wild-type proliferation phenotype (Figure 3.6D). Finally, the overexpression of CCR492 in fibroblasts produces a decrease in let-7 activity that is in agreement with enhancement of expression levels of some target genes (Suppl. Figure S3.3E). Taken together these results demonstrate that, CCR492 by binding directly to let-7 family miRNAs, acts as ceRNA reducing, in a post-transcriptional manner, the let-7 repressing activity on c-Myc.

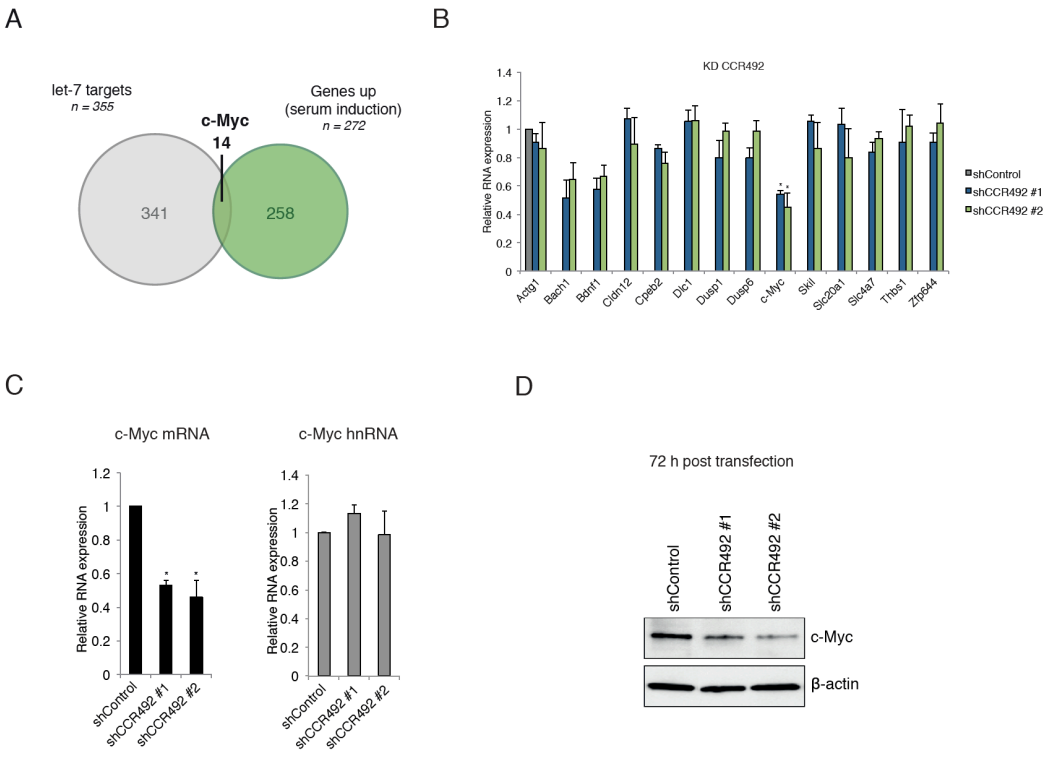


Figure 3.5 CCR492 is a post-transcriptional regulator of c-Myc

A. Venn diagram of serum-induced genes versus let-7 target genes in BALB/c fibroblasts.

B. RT-qPCR of 14 putative let-7 target RNAs and induced by serum treatment in CCR492 depleted BALB/c fibroblasts.

C. RT-qPCR of c-Myc mRNA and c-Myc hnRNA expression levels in control or CCR492 silenced BALB/c fibroblasts.

D. Western blot of c-Myc protein expression in control or CCR492 silenced BALB/c fibroblasts.

Data are presented as mean values ± SD of 3 independent experiments; **P < 0.01, *P < 0.05 by the Student's t-test.

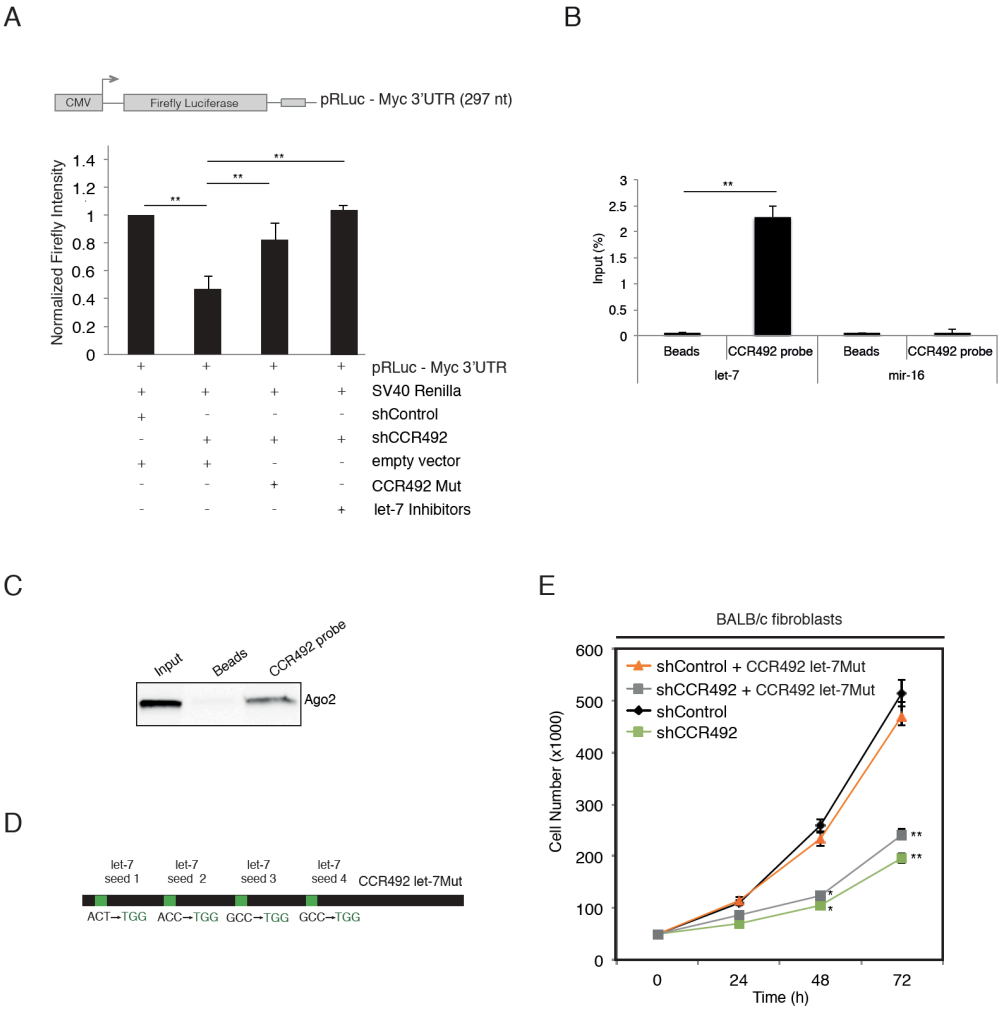


Figure 3.6 CCR492 directly interacts with let-7

A. Dual luciferase assay. Schematic representation of the c-Myc 3'UTR cloned downstream to the Firefly luciferase open reading frame (pRLuc-Myc 3'UTR) (Top panel). Luciferase level obtained by RLuc-c-Myc 3'UTR transfected in control or CCR492 silenced BALB/c fibroblasts. In addition, complementation assay using the CCR492 mutant insensitive to the silencing or with let-7 Inhibitors is shown (Bottom panel).

B. RT-qPCR detection of let-7 in RNA pull-down assay. Mir-16 is the negative control. The results are shown as the percentage (1/100) of the input.

C. Pull-down of Ago2 by biotin-labeled CCR492 probes, as detected by Western Blot.

D. Schematic representation of CCR492 mutant in let-7 seeds.

E. Cell-growth curve assay of CCR492 silenced fibroblasts in the absence or presence of the CCR492 let-7 Mut. Data are presented as mean values ± SD of 3 independent experiments; **P < 0.01, *P < 0.05 by the Student's t-test.

3.6 CCR492 contributes to cell transformation

The above results demonstrate that the expression of CCR492 in response to mitogenic stimuli antagonizes the function of let-7. The finding that CCR492 protects c-Myc mRNA by sequestering let-7 miRNAs prompted us to verify whether CCR492 can cooperate to cell transformation.

To this end, we transfected BALB/c fibroblasts with a constitutively active mutant of the H-Ras proto-oncogene (H-Ras V12), with and without CCR492. As expected, significantly increased c-Myc expression (Figure 3.7A). Importantly, the ectopic expression of CCR492 strongly induced the Ras-dependent transformed foci and soft agar colonies compared to either control cells and cells overexpressing H-Ras alone (Figure 3.7B-C). Taken together these data show that CCR492 plays a role in the control of the cell cycle and when overexpressed contributes to cell transformation.

,

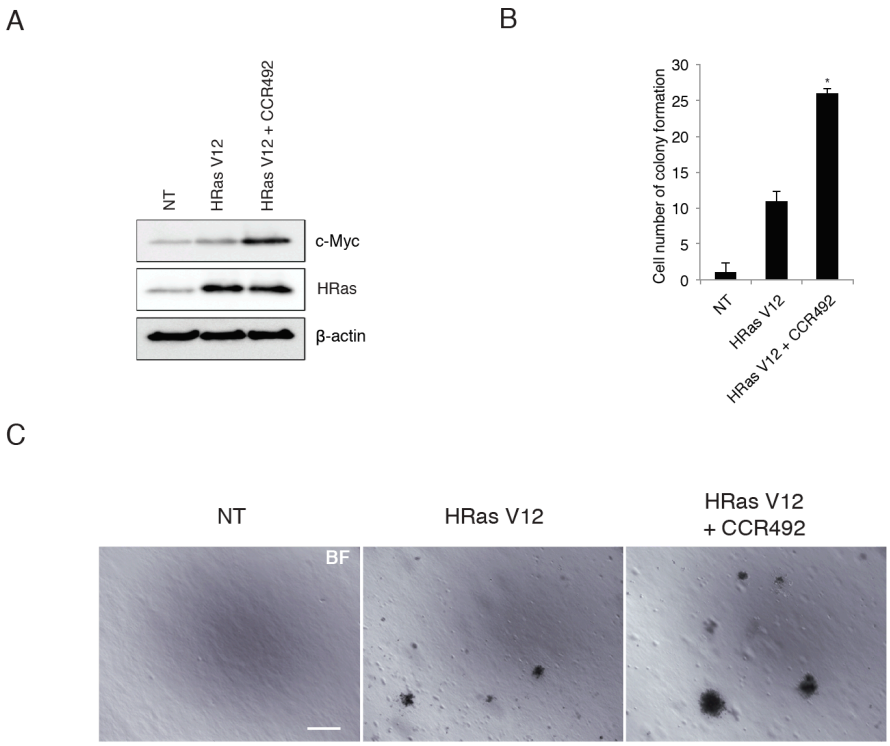


Figure 3.7 CCR492 cooperates with H-Ras V12 for cell transformation

A. Western Blot analysis of c-Myc and H-Ras expression levels in BALB/c transfected either with a construct expressing H-RAs V12 alone or together with CCR492.

B. Soft agar quantification colony assay of BALB/c fibroblasts transfected as in A.

C. A representative bright field image of soft agar colonies.

Data are presented as mean values ± SD of 3 independent experiments; **P < 0.01, *P < 0.05 by the Student's t-test.

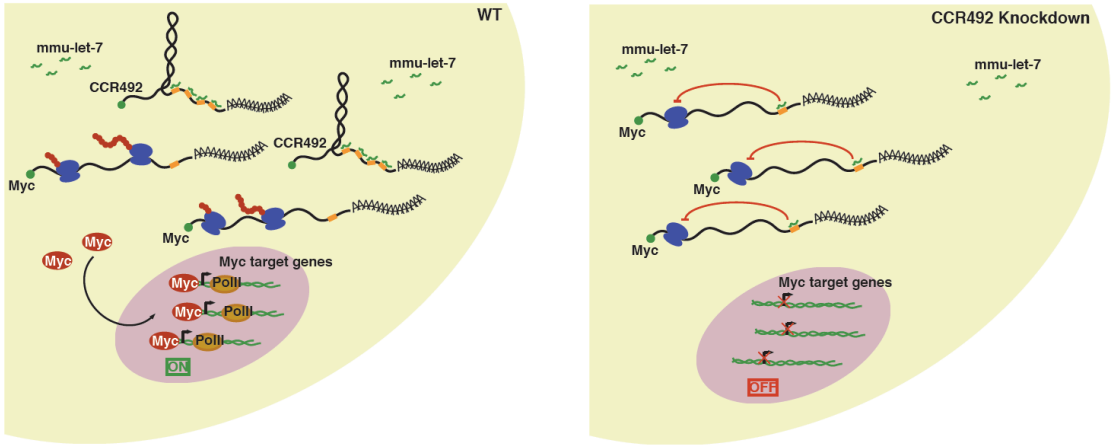


Figure 3.8 Molecular mechanism of action of CCR492

Model of the interactions between CCR492, let-7, and c-Myc. Left panel shows cells expressing CCR492. Right panel shows the effect of the let-7 miRNAs on the c-Myc transcripts.

Table 1. Oligonucleotides

Gene	Sequence	Strand
CCR301	CTGTTCTTTCCCCAACGAAA	FW
CCR071	AGGGCCACCAGAGACTGAG	FW
CCR102	CATAGGGCTTGGGGTTCTAA	FW
CCR115	CCAGGAAGCTGAGAATGAGC	FW
CCR188	CAGAGATCCACAGCCAAACA	FW
CCR339	ATAGTCGTGAGCTACCATGT	FW
CCR327	CAGAGATCCTGACACTGACT	FW
CCR374	ACAGAAGGAACTGGTACACA	FW
CCR363	TGCAGATACACAAGTGTTGG	FW
CCR065	GTGAGACAGAATGTTTGCCT	FW
CCR492	CCAGCTAGAGACAACAGAGA	FW
CCR306	CACACCAGATTGCTAAAGCT	FW
CCR060	ATCGGGAGAACTAGGAAAGG	FW
CCR576	TAGACAGCAGGGAATACAGG	FW
CCR073	GTACCTGCTCCACTTTTCTG	FW
CCR559	CTTTGTGGACAGATCCCATG	FW
CCR250	GTTTGAAGGTGTTTTGCAGG	FW
CCR212	ACTCTAAGCCGGATTTGAGA	FW
CCR548	TACTTTCTGTTCTGCACTGC	FW
CCR301	AGGGAGGACTTGAACCTATG	REV
CCR071	AGGACCCACATGAAGACCAG	REV
CCR102	CTACAGTGGTAGCTCACAGA	REV
CCR115	AAACGGTGGGAAACACTCTG	REV
CCR188	TTGGTTGGTGATTCCCTTTC	REV
CCR339	AGAGATGTCTGTCAGCGTTA	REV
CCR327	TATTCAGCTTGTGTGCTTCC	REV
CCR374	ACCCCTTATACCTGAGCTTC	REV
CCR363	GCCTGGAGTCACATCTTTTT	REV
CCR065	CCCCTAAGTGTCATGCTCTA	REV
CCR492	GTGATATGAGTGCAAGGACC	REV
CCR306	CTGCTTTTGTTACCCTCCTC	REV
CCR060	ATTCACTCAGTTCTCGGCTA	REV
CCR576	GTCTGTGGTTCCACTTTGTT	REV
CCR073	CTGTGAGGTTACTCTTTGCC	REV
CCR559	TCCAGAACCTAATCCACCTC	REV
CCR250	GTTTGAAGGTGTTTTGCAGG	REV
CCR212	CACACGTTTCTCTTGGTGAT	REV
CCR548	TGTTCACCTTTGCTCTATGC	REV

Gene	Sequence	Strand
c-Jun	GGGAAGCACTGCCGTCTGGA	FW
c-Fos	TCTACCCCTGGACCCCTTGCC	FW
Actb	TCTTTGCAGCTCCTTCGTTG	FW
c-Myc	CAGAGGAGGAACGAGCTGAAGCGC	FW
c-Myc hnRNA	ATTCAGGAGGCGTGGCTAAC	FW
Exogenous CCR492	GAAGCCGCACGTCTCACTA	FW
c-Jun	ACTGTAGCCGTAGGCACCGC	REV
c-Fos	GGAGATGGCTGTCACCGTGGG	REV
Actb	ACGATGGAGGGGAATACAGC	REV
c-Myc	TTATGCACCAGAGTTTCGAAGCTGTTCG	REV
c-Myc hnRNA	TTGTGCTGGTGAGTGGAGAC	REV
Exogenous CCR492	CTGGGGCCAACAAGAATAGA	REV

Table 2. Oligonucleotides

Target Gene	Target sequence	Strand	Name
CCR060	CCGGCCCAATACAAGGGTGAATTATCTCGAGATAATTCACCCTTGTATTGGGTTTTTG	FW	CCR060 #1
CCR071	CCGGGGACCTAGGCCCTCTACATATCTCGAGATATGTAGAGGGCCTAGGTCCTTTTTG	FW	CCR071 #1
CCR102	CCGGATTTCTCAAGCCATGTTTATTCTCGAGAATAAACATGGCTTGAGAAATTTTTTG	FW	CCR102 #1
CCR115	CCGGGGCTGGCAACCATGTGTTTAAACTCGAGTTTAAACACATGGTTGCCAGCTTTTTG	FW	CCR115 #1
CCR212	CCGGGCTGGAAGCCACTACTATTTACTCGAGTAAATAGTAGTGGCTTCCAGCTTTTTG	FW	CCR212 #1
CCR306	CCGGGAAGGTTCACTCTGGATTAAACTCGAGTTTAATCCAGAGTGAACCTTCTTTTTG	FW	CCR306 #1
CCR327	CCGGGAAGCACACAAGCTGAATAAACTCGAGTTTATTCAGCTTGTGTGCTTCTTTTTG	FW	CCR327 #1
CCR339	CCGGTCAAGGGAGTAAAGGTAATTTCTCGAGAAATTACCTTTACTCCCTTGATTTTTG	FW	CCR339 #1
CCR374	CCGGCTCAACGTGCTGAACTATATTCTCGAGAATATAGTTCAGCACGTTGAGTTTTTG	FW	CCR374 #1
CCR492	CCGGCTTTGGTCCTTGCACTCATATCTCGAGATATGAGTGCAAGGACCAAAGTTTTTG	FW	CCR492 #1
CCR060	AATTCAAAAACCCAATACAAGGGTGAATTATCTCGAGATAATTCACCCTTGTATTGGG	REV	CCR060 #1
CCR071	AATTCAAAAAGGACCTAGGCCCTCTACATATCTCGAGATATGTAGAGGGCCTAGGTCC	REV	CCR071 #1
CCR102	AATTCAAAAAATTTCTCAAGCCATGTTTATTCTCGAGAATAAACATGGCTTGAGAAAT	REV	CCR102 #1
CCR115	AATTCAAAAAGCTGGCAACCATGTGTTTAAACTCGAGTTTAAACACATGGTTGCCAGC	REV	CCR115 #1
CCR212	AATTCAAAAAGCTGGAAGCCACTACTATTTACTCGAGTAAATAGTAGTGGCTTCCAGC	REV	CCR212 #1
CCR306	AATTCAAAAAGAAGGTTCACTCTGGATTAAACTCGAGTTTAATCCAGAGTGAACCTTC	REV	CCR306 #1
CCR327	AATTCAAAAAGAAGCACACAAGCTGAATAAACTCGAGTTTATTCAGCTTGTGTGCTTC	REV	CCR327 #1
CCR339	AATTCAAAAATCAAGGGAGTAAAGGTAATTTCTCGAGAAATTACCTTTACTCCCTTGA	REV	CCR339 #1
CCR374	AATTCAAAAACTCAACGTGCTGAACTATATTCTCGAGAATATAGTTCAGCACGTTGAG	REV	CCR374 #1
CCR492	AATTCAAAAACTTTGGTCCTTGCACTCATATCTCGAGATATGAGTGCAAGGACCAAAG	REV	CCR492 #1
CCR060	CCGGTCATCGGGAGAACTAGGAAAGCTCGAGCTTTCCTAGTTCTCCCGATGATTTTTG	FW	CCR060 #2
CCR071	CCGGGACTGAGACAGAGATGGATTACTCGAGTAATCCATCTCTGTCTCAGTCTTTTTG	FW	CCR071 #2
CCR102	CCGGTCTCAAGTGTCAGGGATTAATCTCGAGATTAATCCCTGACACTTGAGATTTTTG	FW	CCR102 #2
CCR115	CCGGCATGGGCTTCAGACCATTATACTCGAGTATAATGGTCTGAAGCCCATGTTTTTG	FW	CCR115 #2
CCR212	CCGGTTTCATCGCAAGCGAGGTAACCTCGAGGTTACCTCGCTTGCGATGAAATTTTTG	FW	CCR212 #2
CCR306	CCGGTATCTTGATACCACCTATATTCTCGAGAATATAGGTGGTATCAAGATATTTTTG	FW	CCR306 #2
CCR327	CCGGGGCCAAAGGTGGTTGAAATATCTCGAGATATTTCAACCACCTTTGGCCTTTTTG	FW	CCR327 #2
CCR339	CCGGAGGCATCAAGAACAGTGAAATCTCGAGATTTCACTGTTCTTGATGCCTTTTTTG	FW	CCR339 #2
CCR374	CCGGTTCTATAGAGTTCCCTAATTACTCGAGTAATTAGGGAACTCTATAGAATTTTTG	FW	CCR374 #2
CCR492	CCGGATCAGCGGTTAGCAGCTTATACTCGAGTATAAGCTGCTAACCGCTGATTTTTTG	FW	CCR492 #2
CCR060	AATTCAAAAATCATCGGGAGAACTAGGAAAGCTCGAGCTTTCCTAGTTCTCCCGATGA	REV	CCR060 #2
CCR071	AATTCAAAAAGACTGAGACAGAGATGGATTACTCGAGTAATCCATCTCTGTCTCAGTC	REV	CCR071 #2
CCR102	AATTCAAAAATCTCAAGTGTCAGGGATTAATCTCGAGATTAATCCCTGACACTTGAGA	REV	CCR102 #2
CCR115	AATTCAAAAACATGGGCTTCAGACCATTATACTCGAGTATAATGGTCTGAAGCCCATG	REV	CCR115 #2
CCR212	AATTCAAAAATTTCATCGCAAGCGAGGTAACCTCGAGGTTACCTCGCTTGCGATGAAA	REV	CCR212 #2
CCR306	AATTCAAAAATATCTTGATACCACCTATATTCTCGAGAATATAGGTGGTATCAAGATA	REV	CCR306 #2
CCR327	AATTCAAAAAGGCCAAAGGTGGTTGAAATATCTCGAGATATTTCAACCACCTTTGGCC	REV	CCR327 #2
CCR339	AATTCAAAAAAGGCATCAAGAACAGTGAAATCTCGAGATTTCACTGTTCTTGATGCCT	REV	CCR339 #2
CCR374	AATTCAAAAATTCTATAGAGTTCCCTAATTACTCGAGTAATTAGGGAACTCTATAGAA	REV	CCR374 #2
CCR492	AATTCAAAAAATCAGCGGTTAGCAGCTTATACTCGAGTATAAGCTGCTAACCGCTGAT	REV	CCR492 #2

Table 3. Summary of putative miRNA binding sites in the CCR492 sequence in fibroblasts

Mature miRNA	RPM	Number of MREs
mmu-miR-143-3p	148.252	2
mmu-miR-10-5p	116.442	2
mmu-miR-21a-5p	120.239	none
mmu-miR-99b-5p	66.174	2
mmu-let-7-5p	60.900	4
mmu-miR-30-5p	46.298	2
mmu-miR-26a-5p	31.122	2
mmu-miR-411-5p	27.805	1
mmu-miR-181a-5p	25.281	1
mmu-miR-34c-5p	22.133	2
mmu-miR-125b-5p	20.918	1
mmu-miR-191-5p	15.832	1
mmu-miR-125a-5p	16.825	none
mmu-miR-199a-5p	13.595	1
mmu-miR-541-5p	13.438	3
mmu-miR-100-5p	9.169	2
mmu-miR-351-5p	9.641	3
mmu-miR-16-5p	7.624	none
mmu-miR-182-5p	7.321	3
mmu-miR-434-5p	5.905	none
mmu-miR-186-5p	3.523	2
mmu-miR-26b-5p	3.629	2

Chapter 4

CCR102 lncRNA positively affects Smad7 expression *in cis* taking part in TGF-β signaling

Results

4.1 Characterization of CCR102 lncRNA

The CCR102 gene is localized on chromosome 18 and the presence of active transcription marks, H3K4me3, on the promoter, and H3K36me3, on the gene body, indicate its transcription in fibroblasts. CCR102 transcript is long 2700 bp and is composed by 8 exons (Figure 4.1A). Moreover, CCR102 is ubiquitously expressed in other cell types and tissues according to bioinformatics analysis obtained from our RNA-seq approach and CSHL Resource (Figure 4.1B). Importantly, CCR102 is transcribed on the opposite strand from its neighboring protein-coding gene Smad7 (260 bp away), indicating that CCR102 is a divergent lncRNA (Figure 4.1C).

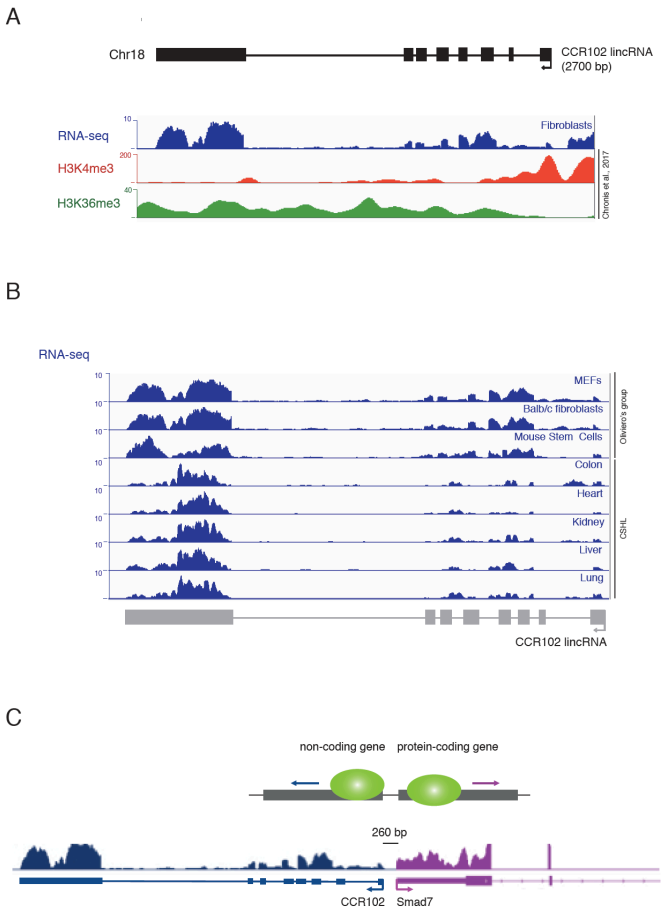


Figure 4.1 Characterization of CCR102 lncRNA

A. Schematic representation of CCR102 lncRNA in term of number of exons (8 exons) and introns. The arrowhead indicates the transcription direction. Chromatin structure is shown from ChIP-seq data for histone modifications (H3K4me3, H3K36me3) in fibroblasts (Chronis et al., 2017).

B. Profile of CCR102 transcript in different cell types and tissues.

C. Schematic representation of a divergent transcription. CCR102 non-coding gene and Smad7 protein-coding gene are transcribed from a bidirectional promoter.

4.2 CCR102 favors G1/S phase progression

To elucidate the biological role of CCR102, we performed loss-of-function studies in proliferation assays. The CCR102 silencing, by transfection of two independent shRNAs (shCCR102 #1, #2) (Figure 4.2A), led to a significant decrease in cell growth, compared to the control fibroblasts (Figure 4.2B). Importantly, FACS analysis revealed an increase of G0/G1-phase cells, together with a reduction of S-phase cells upon CCR102 knockdown (Figure 4.2C). To verify whether CCR102 play a key role in cell proliferation, we performed a deeply analysis by RNA-seq of CCR102 depleted cells, with respect to the control. The silencing of CCR102 resulted in a modulation of some cell-cycle-associated genes (Figure 4.2D). To gain more detailed functional insights into the differentially expressed gene list, a Gene Ontology enrichment analysis was conducted. The categories related to cell cycle were strongly represented (Figure 4.2E). We thus concluded that immediate-early CCR102 lncRNA is directly involved in the regulation of cell-cycle progression.

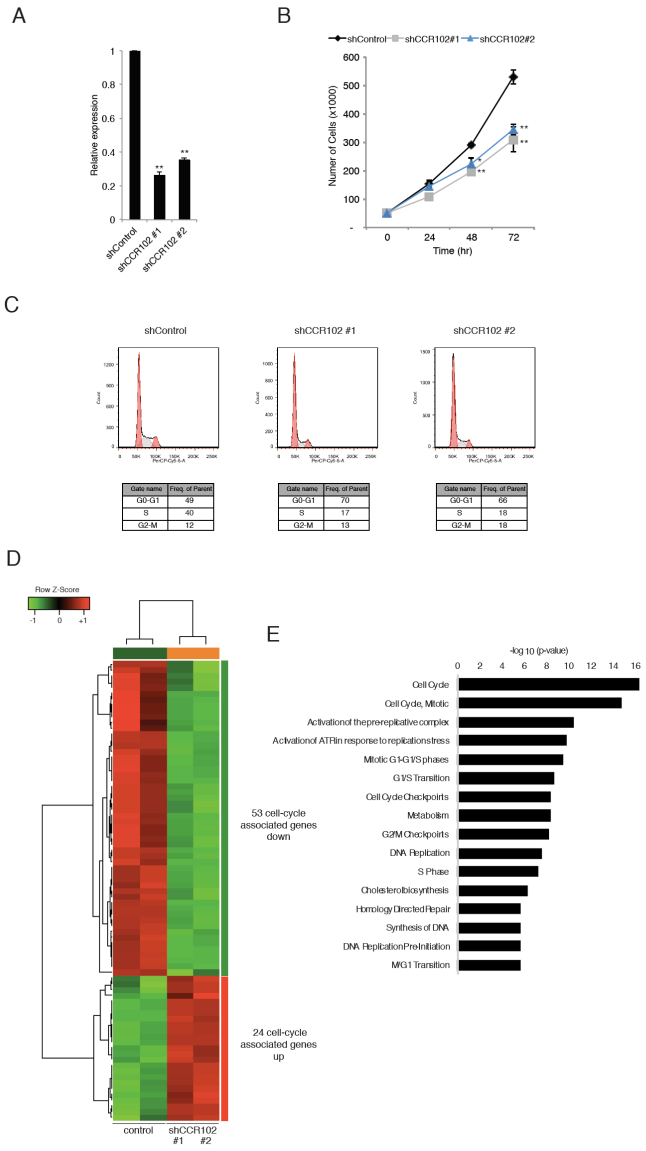


Figure 4.2 CCR102 favors cell-cycle progression

A. Expression of CCR102 in BALB/c fibroblasts transfected with either a non-targeting control (shControl) or two shRNAs targeting CCR102 (shCCR102 #1, #2).

B. Cell-growth curve assay in control and CCR102 silenced fibroblasts.

C. Cell-cycle profile by FACS of CCR102 depleted BALB/c fibroblasts compared to the control shows each phase.

D.Heatmap representation of gene commonly affected by CCR102 silencing (shCCR102 #1, #2). Colors represent transcripts related to cell cycle with a Row Z-Score positive (red) or negative (green).

E. Gene Ontology term enrichment analysis of differentially expressed genes, in CCR102 depleted cells compared to the control. A significant enrichment in pathways relevant for cell cycle and G1/S transition is shown. Data are presented as mean values ± SD of 2 independent experiments; **P < 0.01, *P < 0.05 by the Student's t-test.

4.3 CCR102 is a chromatin-associated lncRNA

To obtain information about the molecular mechanism of action of CCR102, we first defined its subcellular localization. The analysis of biochemical subcellular fractionation by RT-qPCR showed that CCR102 RNA was, for the most part, present in the chromatin fraction of different cell types (Figure 4.3A). Moreover, RNA-FISH assay revealed that CCR102 signal was detected in the euchromatin, while no signal was evident in the DAPI-stained heterochromatin (Figure 4.3B). Taken together, these results suggest that CCR102 is a euchromatin-associated lncRNA.

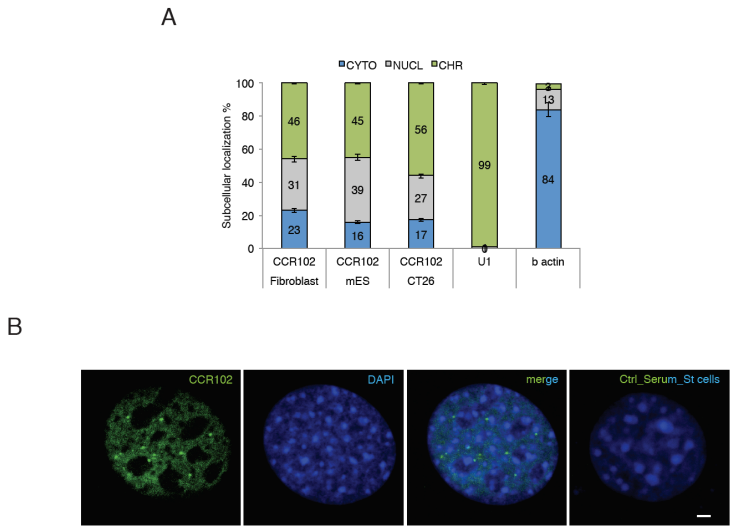


Figure 4.3 CCR102 is a chromatin-associated lncRNA

A. RT-qPCR of subcellular fractionation in fibroblasts, mouse embryonic stem cells and CT26. Percentage ratio of CCR102 in chromatin, nuclear and cytoplasmic over the whole total CCR102 expression levels represents the distribution of CCR102 in cells. Controls of subcellular fractionation are U1 for nucleus and β-actin for cytoplasm.

B. RNA-FISH assay of CCR102 in serum-induced or starved BALB/c fibroblasts as a control. CCR102 localization was obtained with 5'-biotinylated antisense (CCR102-AS) probes. Data are presented as mean values ± SD of 2 independent experiments; **P < 0.01, *P < 0.05 by the Student's t-test.

4.4 CCR102 genomic locus harbors an enhancer region by which positively affects Smad7 expression

An interesting finding is that divergent gene organization may allow chromatin-associated lncRNAs to regulate their adjacent protein-coding genes *in cis* (Sigova et al., 2013; Ørom et al., 2010),(Luo et al., 2016). To gain insight into this, we performed a ChIP-seq analysis in fibroblasts that revealed the presence of an enhancer region, within the CCR102 locus (Figure 4.4A top panel). The high level of H3K4m1 and H3K27ac occupancy, together with the enrichment in p300, as obtained by RT-qPCR analysis (Figure 4.4A bottom panel), indicated that it was an active enhancer.

Our working hypothesis was that CCR102 could promote the expression of its neighboring protein-coding gene Smad7, via the activation of its internal enhancer. Smad7, an Inhibitory Smad (I-Smad), regulates TGF-β signaling by negative feedback loop (Wrighton, Lin, & Feng, 2009). The activity of Smad7 is modulated at many levels, suggesting that it serves as key regulators for fine-tuning the responses to TGF-β signaling. Since CCR102 and Smad7 are constitutively expressed in many different cell types, to investigate their potential crosstalk, we considered the changes in gene expression due to CCR102 silencing, in CCR102 locus (2 Mb). This accurate analysis of the expression levels of the CCR102 neighboring protein-coding genes, represented in the heatmap by a color on a continuous scale according to normalized counts, revealed that in CCR102 depleted cells, only Smad7 mRNA was significantly affected, while Acaa2, Dym, Ctif and Smad2 genes are not affected with respect to the control. Moreover, the genes that are not expressed remain off (Figure 4.4B). These results demonstrate that CCR102 positively regulates the expression of Smad7 with a novel *cis*-regulatory mechanism. Indeed, silencing of CCR102 also produces a significant reduction of active chromatin signatures, H3K27ac and p300 markers, in the enhancer localized within CCR102 coding, which from now on we will name Smad7 enhancer (Figure 4.4C).

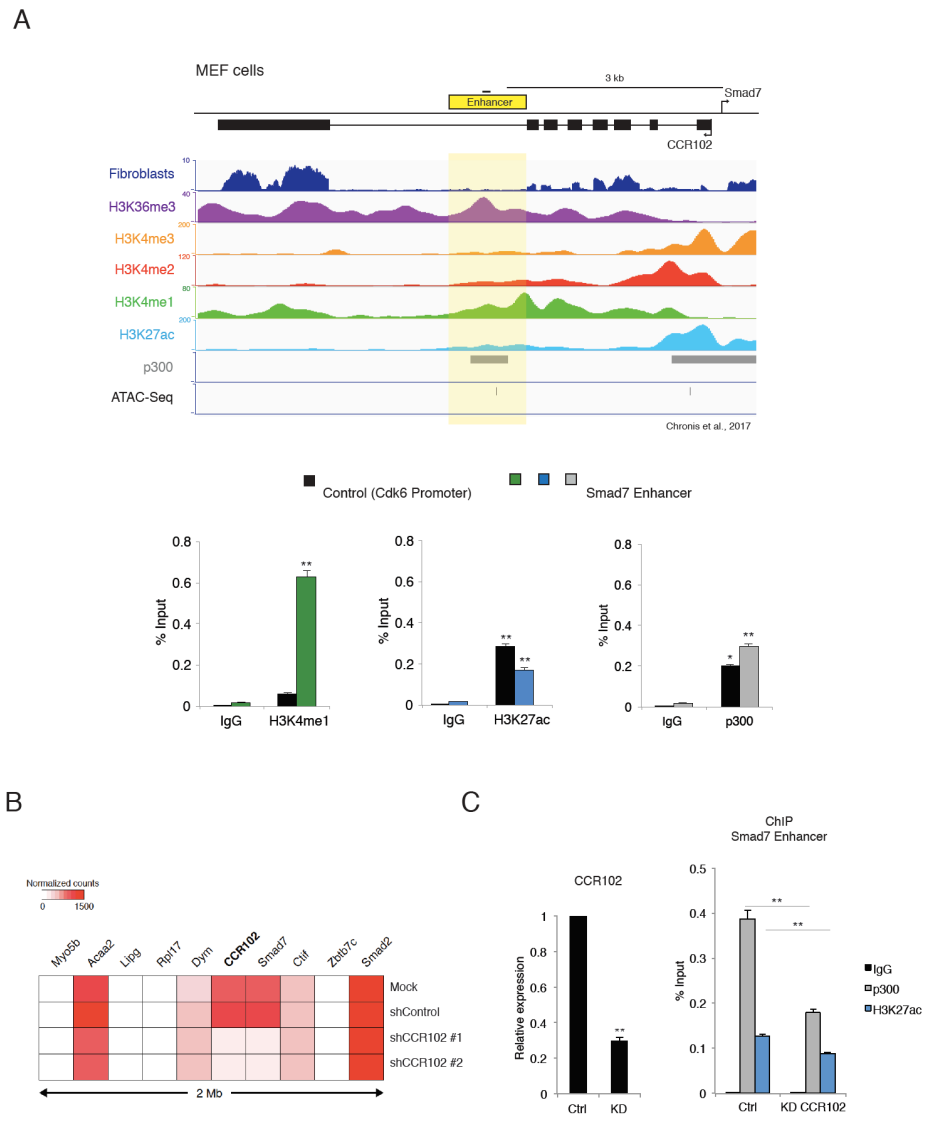


Figure 4.4 CCR102 locus shows an enhancer region and positively affects Smad7 expression

A. Chromatin signatures from ChIP-seq and ATAC-seq measurements in fibroblasts (Chronis et al., 2017) indicate the presence of an enhancer in an intronic region of CCR102.

B. Heatmap representation of gene expression levels (Normalized count) of the CCR102 genomic locus (2Mb). The represented values derived from CCR102 silenced fibroblasts (shCCR102 #1, #2) and control cells (Mock, shControl).

C. ChIP analysis of Smad7 enhancer in CCR102 depleted fibroblasts compared to the control. The antibodies used are specified in the legend. Data are presented as mean values ± SD of 2 independent experiments; **P<0.01, *P<0.05 by the Student's t-test.

4.5 CCR102 interacts with Mediator proving a mechanistic evidence for lncRNA-mediated *cis* regulation of Smad7

CCR102 is highly expressed also in mouse embryonic stem cells (ESCs). ChIP-seq data revealed in ESCs chromatin features of an enhancer, within the CCR102 locus, with overlapping peaks for H3K4me1, H3K27ac and transcription factor binding sites, although with a different genomic localization (Figure 4.5A). Importantly, Smad7 was significantly downregulated in CCR012 silenced cells suggesting a similar molecular mechanism of action modulation than in fibroblasts (Figure 4.5B).

The Mediator complex contributes to the formation and maintenance of chromatin loops between enhancer sequences and the general transcription machinery present at the promoter. Recently, it was proposed that lncRNAs can interact with Mediator, by favoring the chromatin looping between the lncRNA loci and their targets, to increase the transcription of neighboring genes with a *cis*-mediated mechanism (Lai et al., 2013). To confirm that CCR102 acts as an activator of Smad7 enhancer, we performed a RNA pull-down assay with specific biotinylated antisense probes (CCR102-AS), by which we observed that CCR102 directly interacts with Med12 (Figure 4.5C). In agreement with ChIP-seq experiments, which highlighted a strong enrichment of Med12 in Smad7 enhancer and promoter regions, we also demonstrated that CCR102 depletion produces a decrease of Med12 occupancy and H3K27ac active enhancer marker (Figure 4.5D).

Taken together, these experiments suggest that CCR102 transcription positively affects Smad7 expression favoring chromatin accessibility, by the recruitment of the Mediator to the enhancer.

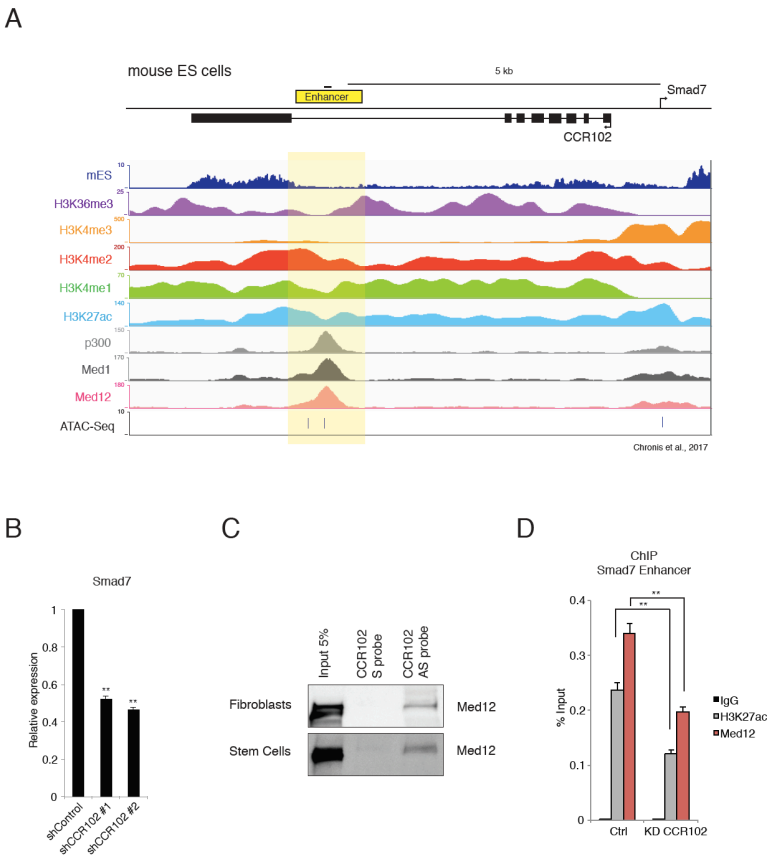


Figure 4.5 CCR102 modulates local chromatin state

A. Chromatin signatures from ChIP-seq and ATAC-seq measurements in mouse embryonic stem cells (Chronis et al., 2017) indicate the presence of an enhancer, in a CCR102 intronic region, with a strongly enrichment of Mediator.

B. RT-qPCR of Smad7 expression levels in control and CCR102 silenced mouse embryonic stem cells.

C. Pull-down of Med12 obtained with 5'-biotinylated antisense (CCR102-AS) probes, as detected by Western blot. Sense (CCR102-S) 5'-biotinylated probes as negative control.

D. ChIP analysis of Smad7 enhancer in CCR102 depleted mouse embryonic stem cells compared to the control. The antibodies used are specified in the legend. Data are presented as mean values ± SD of 2 independent experiments; **P<0.01, *P<0.05 by the Student's t-test.

4.6 CCR102 is upregulated by TGF-β signaling

Ligands of the TGF-β superfamily regulate many cellular functions including cell growth, adhesion, migration, cell-fate determination, differentiation and apoptosis. Failures in signaling downstream of TGF-β are implicated both in the development of tumors and metastasis. Immediate-early CCR102 was expressed rapidly upon serum treatment and Smad7 bidirectional promoter contains a recognized Smad-Binding Element (SBE). Indeed, the expression of Smad7 gene is induced by TGF-β itself (Nakao et al., 1997). To test whether the presence of SBE also regulates CCR102 transcription, we treated cells with TGF-β1 cytokine. Western blot analysis showed a significant increase of pSmad3 proteins, under TGF-β1 and Activin A stimulation, confirming the activation of TGF-β signaling (Figure 4.6A top panel). Moreover, we observed, by RT-qPCR analysis, an upregulation of CCR102 and, as expected, of Smad7 expression levels in primary cells, like fibroblasts and mouse stem cells, but also in CT26 colorectal cancer cells (Figure 4.6A bottom panel).

To examine whether CCR102 takes part in TGF-β signaling, we performed a Gene Ontology analysis considering the list of all downregulated genes by CCR102 silencing, compared to the control. Importantly, we found a significant enrichment in the DNA replication, cell-cycle categories and in TGF-β signaling (Figure 4.6B).

This analysis strongly confirms that the TGF-β1 early-response Smad7 represents a relevant CCR102 target gene.

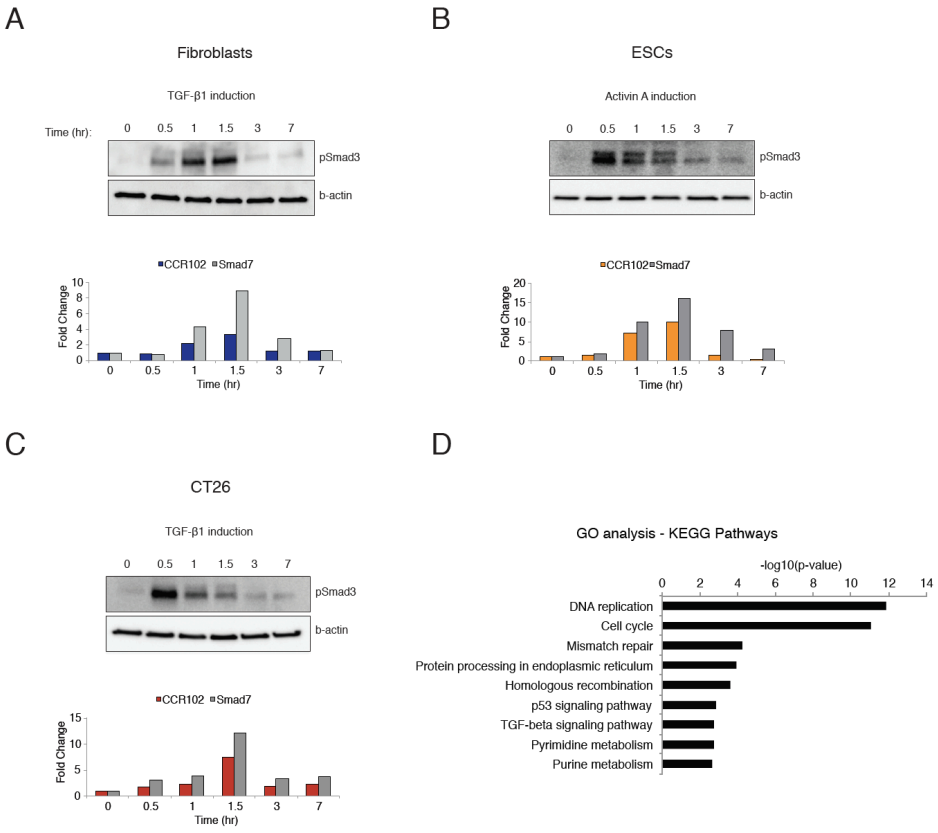


Figure 4.6 CCR102 takes part in TGF-β signaling

A, B, C Western blot of Smad3 phosphorylation following TGF-β1 (in fibroblasts and CT26) and Activin A (in ESCs) stimulation, as a control of TGF-β signaling activation (Top panel). RT-qPCR analysis of expression levels of CCR102 and Smad7 genes (Bottom panel).

D. GO enrichment - KEGG pathway analysis of all downregulated genes by CCR102 silencing. The enrichment score (−Log10 p-value) of the GO terms enriched among differentially expression genes. Data are presented as mean values ± SD of 2 independent experiments; **P<0.01, *P<0.05 by the Student's t-test.

4.7 Human counterpart of CCR102, lncSMAD7, is involved in cell proliferation

Given the importance of *cis*-regulatory role on Smad7 expression and the effects on TGF-β signaling in mouse, we decided to explore whether a human counterpart of CCR102 exists. Despite the majority of lncRNAs do not exhibit high levels of sequence conservation across species and CCR102 sequence is not conserved, the potential human counterpart that we considered, named as lncSMAD7, is localized in a syntenic region, in chromosome 18, associated with the same CTIF, SMAD7 and DYM orthologous genes (Figure 4.7A). lncSMAD7 is already annotated (other names: RP11-1058N17.1 or AC114684.1) and is encoded in three isoforms (Figure 4.7B left panel) with two independent TSSs characterized by the presence of H3K4me3 (Figure 4.7B middle panel) and determined by 5' RACE (Figure 4.7B right panel). The analysis of expression levels by RT-qPCR, under serum-induction condition in starved-fibroblasts and HEK293 cells, revealed an upregulation of lncSMAD7 during the time course (Figure 4.7C), comparably to CCR102 transcript in mouse. Next, we examined the possible biological role of lncSMAD7 in loss-of-function experiments. Fibroblasts silenced for the expression of lncSMAD7, by transfecting two independent shRNAs (shlncSMAD7 #1, #2) (Figure 4.7D), showed a reduction of the rate of proliferation in growth curve, with respect to the control cells (Figure 4.7E).

These experiments demonstrate that lncSMAD7 favors the cell-cycle progression, as the murine CCR102, suggesting that the conservation of the genomic locus is an indicator of analogous biological role.

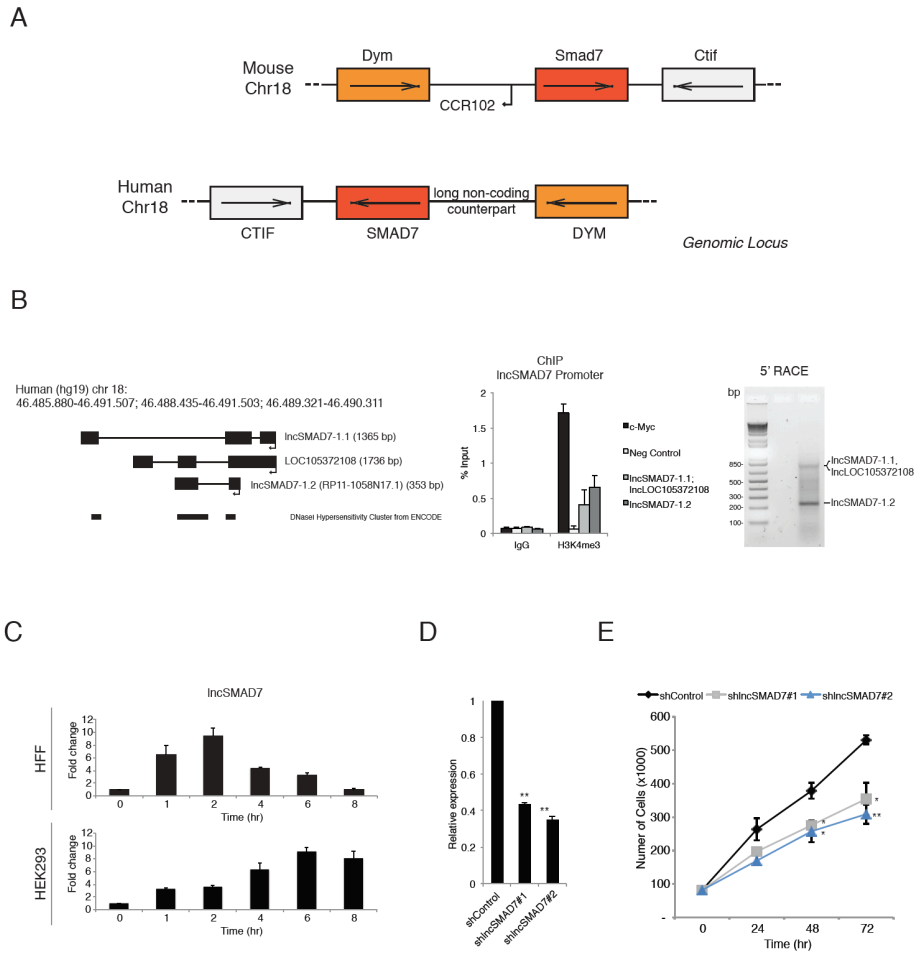


Figure 4.7 Characteziation of Human counterpart of CCR102 (lncSMAD7)

A. Schematic representation of the CCR102 non-conding gene and its neighboring protein-coding genes positioning in the Human and Mouse genomic loci.

B. Human counterpart of CCR102 (named lncSMAD7) isoform representation (left panel); ChIP analysis of two promoter regions of the lncSMAD7 gene (middle panel); lncSMAD7 isoforms start sites (mapped by 5' RACE) (right panel).

C. RT-qPCR validation of lncSMAD7 in HFF and HEK293 with specific oligos. The expression levels are indicated as fold change on starved cells in a serum treatment time course.

D. Expression of lncSMAD7 in HEK293 transfected with either a non-targeting control (shControl) or two shRNAs targeting lncSMAD7 (lncSMAD7 #1, #2).

E. Cell-growth curve assay in control and lncSMAD7 silenced cells. Data are presented as mean values ± SD of 2 independent experiments; **P < 0.01, *P < 0.05 by the Student's t-test.

4.8 lncSMAD7 is a chromatin lncRNA that positively impacts on SMAD7 expression

Like CCR102, lncSMAD7 is predominantly expressed in the chromatin fraction, as confirmed by biochemical subcellular fractionation (Figure 4.8A). RNA-FISH analysis showed a strong enrichment of lncSMAD7 transcripts in the nuclear compartment (Figure 4.8B).

The analysis of the lncSMAD7 locus revealed the presence of two active enhancer regions, well described in EBI-EMBL source, supported by the comparable features to CCR102 enhancer (Figure 4.9A). Importantly, lncSMAD7 is localized downstream of the SMAD7 TSS, transcribed in the same direction as the nearest gene (Figure 4.9B). Moreover, lncSMAD7 contains an enhancer region that we named distal enhancer. Proximal (in green) and distal (in yellow) SMAD7 enhancers were characterized by ChIP experiments for the presence of H3K4me, H3K27ac, p300 markers and CTFC specifically for distal enhancer, in agreement with ENCODE data analysis (Figure 4.8B). To verify whether lncSMAD7 affects the expression of SMAD7 like in mouse, we performed a knockdown of lncSMAD7. The silencing of lncSMAD7 resulted in a significant reduction of SMAD7 expression levels (Figure 4.9C).

Our data indicate that the targeted depletion of lncSMAD7 in the downstream sense orientation led to a decrease in expression of its uptstream neighboring protein-coding SMAD7.

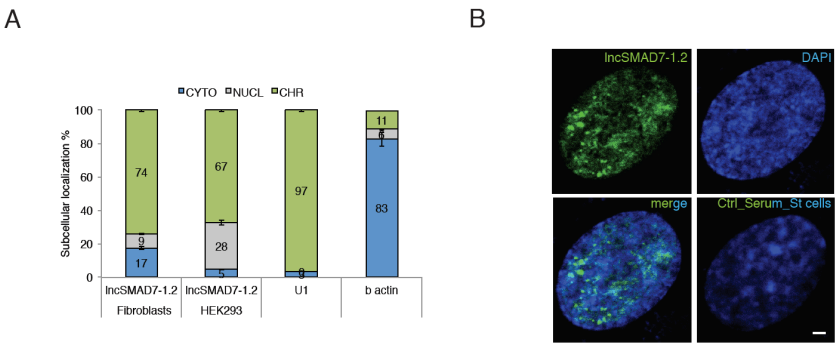


Figure 4.8 lncSMAD7 is a chromatin-associated lncRNA

A. RT-qPCR of subcellular fractionation. Percentage ratio of lncSMAD7 in chromatin, nuclear and cytoplasmic over the whole total lncSMAD7 expression levels represents the distribution of lncSMAD7 in HFF and HEK293. Controls of subcellular fractionation are U1 for nucleus and β-actin for cytoplasm.

B. RNA-FISH assay of lncSMAD7 in serum-induced and starved cells. lncSMAD7 localization was obtained with 5'-biotinylated antisense (lncSMAD7-AS) probes. Data are presented as mean values ± SD of 2 independent experiments; **P < 0.01, *P < 0.05 by the Student's t-test.

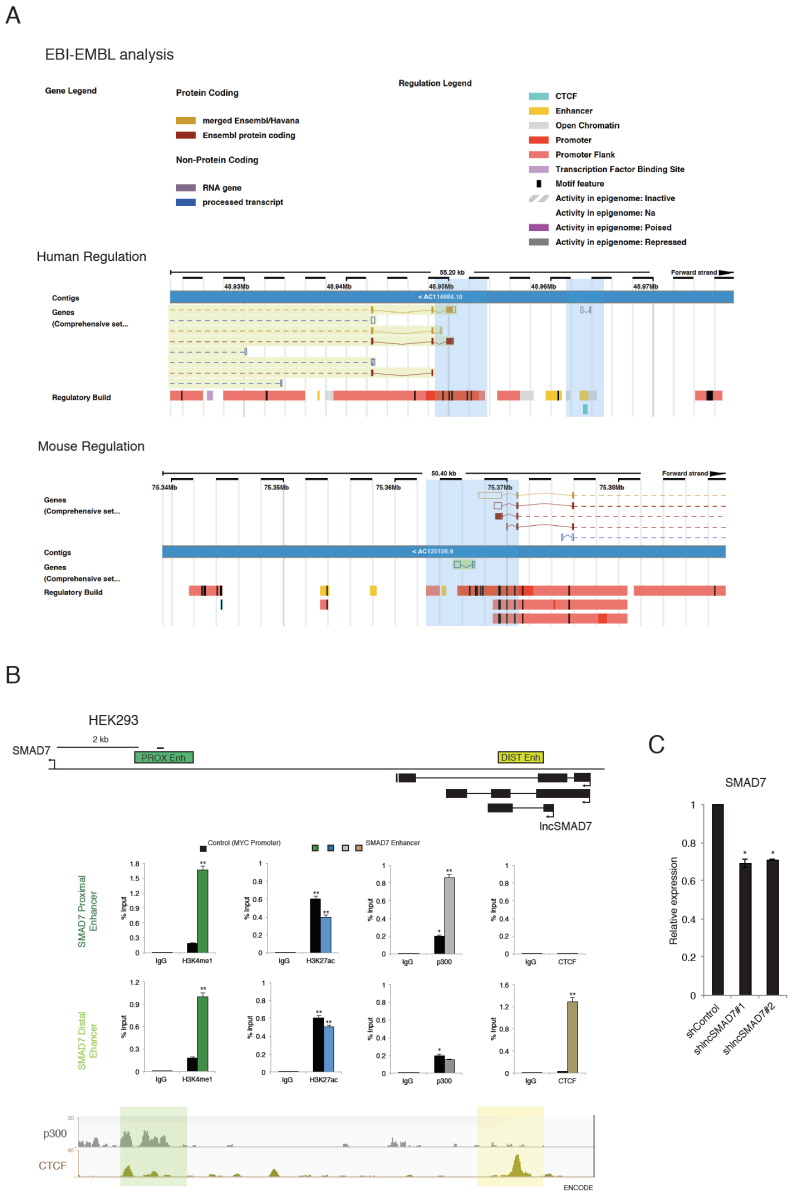


Figure 4.9 lncSMAD7 locus exhibits enhancer region as CCR102 locus

A. EBI-EMBL analysis of lncSMAD7 locus and CCR102 locus.

B. Schematic representation of lncSMAD7 genomic locus. Arrows indicate the transcription start sites (TSSs) of SMAD7 and lncSMAD7 isoforms. Squares represent the enhancer regions (Top panel). ChIP analysis of proximal and distal SMAD7 enhancer regions, the antibodies used are specified in each panel (middle panel). Chromatin signatures from ChIP-seq measurements (ENCODE) in HEK293 (Bottom panel).

C. Expression of SMAD7 in HEK293 transfected with either a non-targeting control (shControl) or two shRNAs targeting lncSMAD7 (lncSMAD7 #1, #2). Data are presented as mean values ± SD of 2 independent experiments; **P < 0.01, *P < 0.05 by the Student's t-test.

4.9 lncSMAD7 is upregulated in colorectal cancer

Considering that TGF-β signaling has a crucial role in colorectal cancer (Pritchard & Grady, 2011) and murine CCR102 takes part in this pathway by modulating the expression levels of Smad7, we speculated that lncSMAD7 might be altered in tumors. To confirm this hypothesis, we analyzed RNA-seq data obtained from The Cancer Genome Atlas (TCGA). Interestingly, we found that lncSMAD7 is upregulated in a cohort of 10 patients with colorectal cancer (stages II, III) in comparison with normal tissue samples (Figure 4.10A). Futhermore, the upregulation of SMAD7 in colorectal cancer has been correlated with poor survival (Boulay et al., 2003). A positive correlation between the expression levels of lncSMAD7 and SMAD7 mRNAs was observed in colorectal tumors, indicating that these transcripts are functionally linked to the tumor progression (Figure 4.10B).

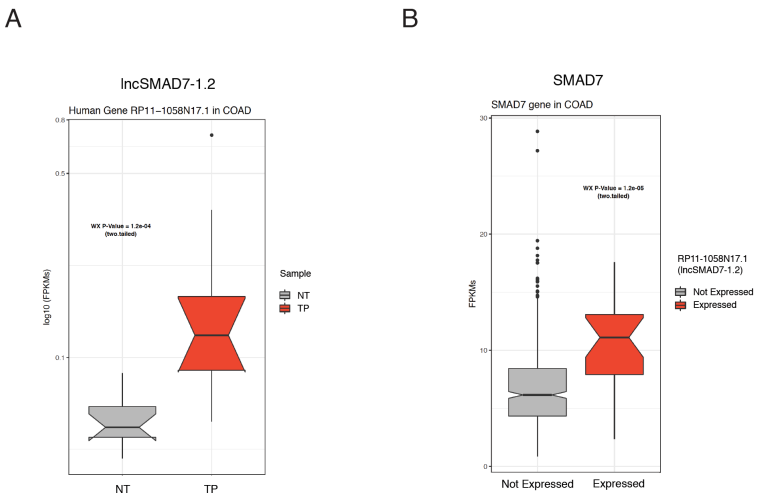


Figure 4.10 lncSMAD7-1.2 is upregulated in a subset of colorectal cancers and correlates with SMAD7

A: Expression analysis of lncSMAD7-1.2 in a cohort of 10 colon cancer tissues in comparison with the corresponding normal tissue in RNA-seq datasets obtained from TCGA.

B: SMAD7 levels in samples expressing lncSMAD7-1.2 compared to samples that do not express lncSMAD7-1.2.

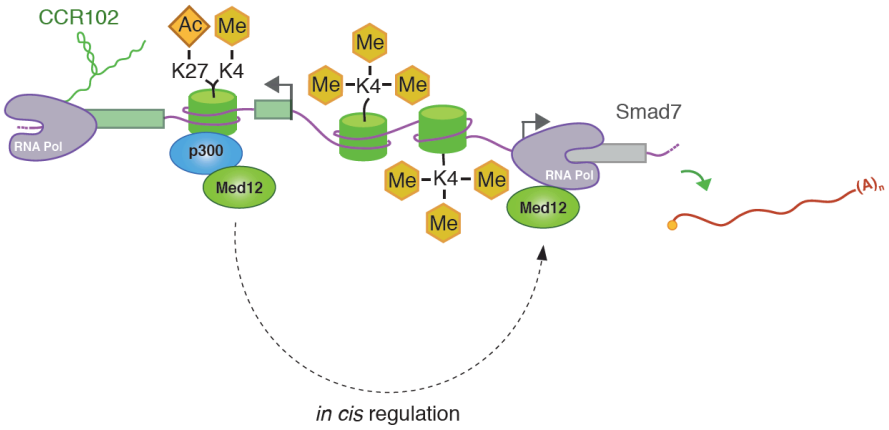


Figure 4.11 Molecular mechanism of action of CCR102

Model of *cis*-regulatory mechanism by which CCR102 positively affects Smad7 expression.

Table 4.

UCSC Genome Browser analysis

					HchrStart (bp)	HchrEnd (bp)		
Ensembl Gene ID	Ensembl Transcript ID	Human Ensembl Gene ID	HaGeneName	HchrName			%id.targetHgeneIdtoQG	LastCancwithHuman
ENSMUSG00000025880	ENSMUST00000026999	ENSG00000101665	SMAD7	18	48919853	48950711	98	Euarchontoglires
ENSMUSG00000025880	ENSMUST00000174843	ENSG00000101665	SMAD7	18	48919853	48950711	98	Euarchontoglires
ENSMUSG00000025880	ENSMUST00000174411	ENSG00000101665	SMAD7	18	48919853	48950711	98	Euarchontoglires
ENSMUSG00000025880	ENSMUST00000172718	ENSG00000101665	SMAD7	18	48919853	48950711	98	Euarchontoglires
ENSMUSG00000025880	ENSMUST00000168918	ENSG00000101665	SMAD7	18	48919853	48950711	98	Euarchontoglires
ENSMUSG00000035765	ENSMUST00000039608	ENSG00000141627	DYM	18	49041474	49461347	95	Eutheria
ENSMUSG00000052928	ENSMUST00000165559	ENSG00000134030	CTIF	18	48539046	48863217	90	Eutheria

Legend:
HaGeneName: Human Associated Gene Name
HchrName: Human Chromosome Name
HchrStart (bp): Human Chromosome Start (bp)
HchrEnd (bp): Human Chromosome End (bp)
%id.targetHgeneIdtoQG: %id. target Human Gene Identical to Query gene
LastCancwithHuman: Last common ancestor with Human

Table 5. Oligonucleotides

Gene	Sequence	Strand
CCR102	CATAGGGCTTGGGGTTCTAA	FW
CCR102	CATAGGGCTTGGGGTTCTAA	REV
Smad7	GGGGCTTTCAGATTCCCAAC	FW
Smad7	AGCTGACTCTTGTTGTCCGA	REV
Gapdh	GTGTTCCTACCCCCAATGTGT	FW
Gapdh	ATTGTCATACCAGGAAATGAGCTT	REV
MEG3	ACTCTTGCCACATTAGCCCC	FW
MEG3	AATGCCAGCTGCTCTTCTGT	REV
lncSMAD71.2	ACATGCTCTGCCCATTCCAA	FW
lncSMAD71.2	ACTGAGCGCGTTCACTTACA	REV
SMAD7	CTCAAACCAACTGACTGTCCA	FW
SMAD7	ATTGAGCTGTCCGAGGCAAA	REV
ACTIN	AGAAAATCTGGCACCACACC	FW
ACTIN	GGGGTGTTGAAGGTCTCAAA	REV
mm: Smad7_TSS_ChIP	CTGATCGGCGCACAGAGGAT	FW
mm: Smad7_TSS_ChIP	GTCTCAGGCAGCTCTCTCCG	REV
mm: Enhancer mouse_Stem cells_ChIP	ATGCATTTCCTAGGCTCCGC	FW
mm: Enhancer mouse_Stem cells_ChIP	ACTCTGAGCATCCCTCCGAA	REV
mm: Enhancer mouse_Fibroblasts_ChIP	CCAAAAAGCTACGTGCCTGG	FW
mm: Enhancer mouse_Fibroblasts_ChIP	TGAGGACACGCGAAATGGAT	REV
mm: ckd6_TSS_ChIP	GAGTAGAGGGAGGAGGAGA	FW
mm: ckd6_TSS_ChIP	CGATTACATAGTCTGCCCAC	REV
HS: Distal Enhancer_ChIP	TGCAATACCTCCCAGTGCAA	FW
HS: Distal Enhancer_ChIP	GAGGGCTGCTACCCAGTAGA	REV
HS: Proximal Enhancer_ChIP	CTCGGTCCAGTCCGGTATAA	FW
HS: Proximal Enhancer_ChIP	CGCCCTAGCGCTTCATTCAT	REV
HS: c-MycChIP+0.4kbpos	TCCGCAACCCTTGCCGCATC	FW
HS:c-MycChIP+0.4kbpos	CGCGTCCTTGCTCGGGTGTT	REV

Chapter 5

Discussion

In this thesis work we reported the identification of novel immediate-early lncRNAs whose expression is induced by mitogens in mouse fibroblasts and the characterization of the molecular function of two of these lncRNAs. We found that the lncRNAs CCR492 and CCR102 play a role in cell proliferation with post-transcriptional and transcriptional regulatory mechanisms, respectively: CCR492 acts as a ceRNA by binding let-7 to upregulate c-Myc, and CCR102 acts at the transcriptional level regulating *in cis* the transcription of Smad7.

In response to serum treatments quiescent fibroblasts trigger a cellular response to upregulate over one hundred coding genes including the nuclear factors c-Jun, c-Fos, c-Myc and many genes, coding for other cellular components, required to proceed toward the G1 phase. Only recently, the importance of non-coding RNAs in cell regulatory circuits has become clear. We therefore performed a screening to identify non-coding RNAs involved in the regulation of the cell growth. Since it is impossible to predict RNA function from its primary sequence alone, we performed a selection based on their regulation in response to proliferative signals. Thus, we took advantage of the knowledge accumulated from the studies of protein-coding genes, which revealed the importance of the immediate-early genes in the cell-cycle regulatory circuit.

Our screening identified a number of transcripts induced by serum in starved MEFs. We further characterized the functional role of two of such cell-cycle regulated lncRNAs: CCR492 and CCR102. CCR492 is a novel lncRNA, of 953 bp, whose gene is organized in 4 exons, it maps in a gene devoid region of chromosome 4, distant about 65 kb from the closest gene. CCR102 is a divergent lncRNA, transcribed from the same bidirectional promoter of the gene

Smad7 and it is localized in the mouse chromosome 18. Both lncRNAs are strongly induced by serum (Maldotti et al., 2016). Similarly to Smad7, CCR102 is also induced by TGF-β (Arase et al., 2014). Moreover, despite the majority of lncRNAs, which are known to show more cell-type specificity than protein-coding transcripts, CCR102 is expressed in many different cell types and tissues.

By functional experiments we found that both CCR492 and CCR102 contribute to the progression of the cell cycle, since their depletion significantly reduces the growth rate affecting the cell entry into the S phase. By analysis of the subcellular localization we found that CCR492 is cytoplasmic while CCR102 is chromatin-associated.

LncRNAs have been shown to function in several contexts to regulate gene expression (Geisler & Coller, 2013). In the cytoplasm, a number of lncRNAs has been proposed to function as ceRNAs to modulate the availability of miRNAs, sequestering them from their protein-coding target RNAs (Poliseno et al., 2010),(Salmena et al., 2011),(Cesana et al., 2011),(Kallen et al., 2013). Inspection of CCR492 sequence revealed the presence of 4 high-affinity recognition elements for the let-7 family of miRNAs, which are known to be involved in the control of the cell cycle (Johnson et al., 2007) suggesting that CCR492 could function as a sponge for let-7 miRNAs. Indeed, in the presence of let-7 inhibitors the effect of CCR492 depletion was abolished suggesting that CCR492 acts as a ceRNA to the let-7 miRNAs. In addition, CCR492 directly interacts with let-7 and the co-immunoprecipitation with Ago2 supports that this reciprocal repression is through RISC complex. One important target of let-7 is c-Myc (Kumar, Lu, Mercer, Golub, & Jacks, 2007) that contains a let-7 MRE on its 3'UTR region and has been recently demonstrated that even modest variations of let-7 levels significantly alter the c-Myc expression (Wu et al., 2015). We found that the silencing of CCR492 results in a significant decrease of c-Myc expression in fibroblasts, while its overexpression restored elevated c-Myc levels. Importantly, we could demonstrate that this regulation is via the c-Myc 3'UTR and it is abolished by let-7 inhibitors. Thus, our results are compatible with a model (Figure 3.8) in which CCR492, by binding directly let-7 miRNAs, releases c-Myc from let-7 dependent inhibition. In agreement with this result we found that the constitutive expression of CCR492 in fibroblasts cooperates with mutated H-Ras to promote cell transformation. Let-7 family of miRNAs must be tightly regulated in the cells as their deregulation leads to developmental

alterations and cancer (Johnson et al., 2007). However, their regulation is quite complex because in mammals the family of let-7 miRNAs is present in multiple copies in the genome. Let-7 expression is regulated at the transcriptional and post-transcriptional level. In spite of their different genome organization, let-7 precursors share conserved structural elements and the sequence of the mature miRNA is highly conserved. One way to regulate let-7 expression is by Lin28 proteins, which inhibit the mature form by their specific interaction with let-7 precursors (Nam, Chen, Gregory, Chou, & Sliz, 2011),(Stefani, Chen, Zhao, & Slack, 2015). Another simple and effective way to regulate let-7 function is to inhibit its activity by direct interaction with the mature miRNAs as all let-7 genes share the same seed to bind to their targets. This is accomplished by the expression of non-coding RNAs containing multiple MREs. In humans, the lncRNA H19 functions as a sponge to antagonize let-7 to regulate muscle differentiation (Kallen et al., 2013). Our work unveiled a novel lncRNA that is not conserved hence performs the same function of inhibiting let-7 activity in mouse, in a different cellular context. Thus, demonstrating the parallel evolution of different molecules to achieve the same regulatory mechanism, we establish the importance of this mechanism for the control of let-7 miRNAs.

When compared with protein-coding mRNAs, a significant number of lncRNAs is enriched in the nucleus (Derrien et al., 2012) and many of them complex with chromatin modifiers to regulate gene expression in a transcriptional manner. Analogous to the enhancer-derived lncRNA HOTTIP (K. C. Wang et al., 2011) and che-RNA (Werner et al., 2017), euchromatin-associated CCR102 promotes an active chromatin state. Some lncRNAs functionally cooperate with Mediator, a crucial complex for enhancer–promoter looping formation, to activate the transcription (Lai et al., 2013). Consistent with this, inspection of CCR102 genomic region, in primary cells, revealed the presence of an active Smad7 enhancer (H3K4me1 and H3K27ac), strongly enriched for, the kinase component of Mediator, Med12. CCR102 is localized adjacent to the protein-coding gene Smad7, a potent inhibitory of TGF-β signaling that also mediates the crosstalk with other signaling pathways to transmit information into the nucleus (Nakao et al., 1997). A recent study proposes that lncRNAs transcribed very close to their protein-coding genes function as enhancers to activate expression in their chromosomal neighborhood, *in cis* (Ørom et al., 2010). Interestingly, knockdown of CCR102, in primary cells, led to decreased

expression of Smad7 together with a significant reduction of Med12 binding to Smad7 enhancer. Moreover, CCR102 physically interacts with Med12, providing a mechanistic evidence for a novel lncRNA-mediated *cis* regulation. Thus, these experiments highlight an important mechanism that may be the key to regulate Smad7 transcription. We propose a model in which the transcription of CCR102 favors chromatin accessibility to precisely and rapidly regulate Smad7 gene expression in response to growth factors. Above described regulation of the TGF-β pathway might be relevant in pathological condition, such as colorectal cancer. We therefore also analyzed the human counterpart of CCR102, lncSMAD7, which, despite undetectable sequence identities, shows several analogies with CCR102: it is a chromatin-associated lncRNA and it is localized close to the TSS of SMAD7 gene, in a human-mouse syntenic region on chromosome 18. Moreover, the lncSMAD7 locus also appears to contain a transcriptional enhancer. Indeed, ENCODE data analysis indicate lncSMAD7 locus to be marked by high levels of H3K4me1 and H3K27ac in two active SMAD7 enhancer regions and in human cells, silencing of lncSMAD7 results in a significant decrease in SMAD7 expression levels. Thus, showing that similarly to CCR102 lncSMAD7 positively regulate the SMAD7 transcription. Interestingly, SMAD7 is upregulated in a subset of colorectal cancer in comparison with normal tissue samples (Halder et al., 2008), (Stolfi et al., 2014),(De Simone et al., 2017) and analysis of lncSMAD7 in cancers shows its high correlation with SMAD7. Thus, lncSMAD7 may represent a potential target for therapeutic intervention against cancer.

Further studies will be necessary to fully characterize the CCR102/lncSMAD7-mediated regulatory mechanism. One of the research priorities will be to use of a knockout (KO) strategy, based on CRISPR/Cas9 system, to insert a premature polyadenylation (polyA) signal allele-specific into the CCR102 locus to formally demonstrate that CCR102 and lncSMAD7 act *in cis* to enhance gene expression as recently shown (Engreitz et al., 2016). In addition, our results provide strong evidence that Mediator complex occupancy in CCR102 locus facilitates the contact between promoter and enhancer elements through chromatin looping. Nevertheless, a Chromatin Conformation Capture (3C) (J. Dekker, Rippe, Dekker, & Kleckner, 2002) analysis will allow to confirm the presence of a DNA looping, in comparison with CCR102-depleted cells.

Chapter 6

Experimental Procedures

6.1 Cell culture

Primary mouse embryonic fibroblasts (MEFs) derived from 13.5d pregnant female mice, BALB/c fibroblasts (ATCC, Manassas, VA, USA) and CT26 (Mus musculus colon carcinoma) were cultured in DMEM medium with 10% fetal bovine serum (FBS) (Sigma), 1 mM sodium pyruvate (Invitrogen), 50 U/ml of penicillin/ml, and 50 µg/ml of streptomycin/ml. Cell synchronization was performed starving MEFs in 0.2% serum for 48 h and BALB/c fibroblasts in 0% serum for 48 h then released into cell cycle by adding 20% serum for the times indicated (30min, 1 h and 2 h). 60 nM pool of let-7miRNA Inhibitors (Ambion) and 60 nM pool of let-7 miRNA Mimics (Ambion) were respectively used for let-7 inhibition and overexpression experiments. For 5,6-dichloro-1-b-D-ribofuranosylbenzimidazole (DRB) treatment, 40 µM DRB (diluted in DMSO) were added directly to the cell culture medium and incubated for the indicated times.
E14 mouse Embryonic Stem Cells (ESCs) were cultured in high-glucose DMEM (Invitrogen) supplemented with 15% FBS (Millipore), 0.1 mM nonessential amino acids (Invitrogen), 1 mM sodium pyruvate (Invitrogen), 0.1 mM 2-mercaptoethanol, 1500 U/ml LIF (Millipore), 25 U/ml penicillin, and 25 µg/ml streptomycin. Cytokines treatments were conducted in serum-free media supplemented with 10 ng/ml TGF-β1 (R&D) or 40 ng/ml Activin A (ThermoFisher Scientific) in a rapid time course.

6.2 RNA extraction and quantitative Real-Time PCR (RT-qPCR)

RNA was extracted using TRIzol reagent (Invitrogen), following manufacturer instructions. RNA integrity measurements were performed using Fragment Analyzer™ (Advanced Analytical). All samples had RNA Quality Number (RQN) N 9.8. RT-qPCR was performed

using the SuperScript III PlatinumOne-Step Quantitative RT-PCR System (Invitrogen) following the manufacturer's instructions.
Primers sequences are shown in Table 1 and 5. Expression levels of mature let-7 were quantified by RT-qPCR using TaqMan MicroRNA Assays Kit (Applied Biosystems, Carlsbad, CA). In addition, the expression levels of miR-16 were evaluated as a negative control of RNA precipitation. The reverse transcription reaction was carried out with TaqMan MicroRNA Reverse Transcription Kit (Applied Biosystems, Carlsbad, CA) following the manufacturer's instructions.

6.3 RNA-seq library preparation

For RNA-seq library preparations, 2 µg of total RNA were used as input for the TruSeq RNA Library Prep Kit v2 (Illumina), and libraries were prepared following manufacturer instructions. Gene ontology was performed using the DAVID Software (Frederick National Laboratory, Rockville, MD, USA). Heatmap and clustered heatmap were prepared using publically available Morpheus software https://software.broadinstitute.org/morpheus.

6.4 RNA-seq for novel lncRNAs discovery

Next-generation RN-seq was performed on Illumina HiScanSQ Platform, producing approximately 180.000.000 paired-end reads per lane. Reads were preprocessed using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), in order to remove poor-quality reads, and clip adapter sequences. After preprocessing, we obtained approximately 150.000.000 high quality paired-end reads.
Reads were mapped to mouse genome (Mus musculus, assembly mm9) using the TopHat v2.0.0 spliced-mapper, a gapped aligner able to discover new splice-junctions ab initio. Briefly, TopHat maps reads to the genome using the Bowtie aligner to build a set of putative exons (or contigs). The unmapped reads, which may cross splice-junctions, are then mapped allowing for large gaps, against a reference built onto GT-AG genomic splicing motifs. Reads mapped from TopHat were then assembled into a parsimonious set of transcripts using Cufflinks v2.0.2. Cufflinks uses TopHat output to build the minimal set of transcripts that can explain the majority of mapped reads. To increase Cufflinks accuracy, we provided it with an annotation of both annotated and predicted transcripts from Ensembl release 68 (July 2012).
To create a comprehensive and combined annotation of any discovered transcript in all sequenced lanes, we used the Cuffcompare utility. Cuffcompare performs a union of all transcripts by merging transcripts that share all introns and exons, with some flexibility for the first and last exons.
At this stage, Cufflinks produced approximately 300.000 transcripts, many of which are single exon transcripts deriving from transcriptional noise, residuals of pre-mRNAs, and exons that

Cufflinks was unable to connect due to the lack of spliced-read information. To clean our annotation from background noise, we designed a filtering pipeline, which allows us to simultaneously remove low-fidelity transcripts, and to select putative previously unannotated transcripts lacking coding potential:
(1) Size selection. All single exon transcripts, and multi-exon transcripts shorter than 200 bases were excluded; (2) Minimal read coverage. Using Cufflinks estimation read coverage for each transcript, we excluded those transcripts with a maximal coverage below 3 reads per base; (3) Filtering of known annotations. We removed all transcripts that have at least an exon overalapping a transcript from RefSeq, UCSC, Ensembl, and Vega annotation sets, and any transcript overlapping known microRNAs, tRNAs, snoRNAs, and rRNAs; (4) Coding potential estimation. To estimate the coding potential of novel transcripts, we used two different Support Vector Machines, named LNSeeRNA and Coding Potential Calculator, which have been trained onto different dataset. Transcripts identified by at least one SVM as coding, were excluded. The final annotation yields approximately high-fidelity 750 transcripts. To determine which transcripts may have a role in cell proliferation, we then performed differential expression analysis across time-course samples, using Cuffdiff, and selected only those showing significant upregulation upon serum induction respect to serum-starved cells (point 0 of our time-course). Cuffdiff estimates the abbundance of each transcript, by calculating the number of reads which map on it, normalized over the lenght of the transcript and the total number of reads mapped in the experiment (RPKM, Reads Per Kilobase per Million reads).
miRNA target prediction on lncRNA CCR492 was performed using three algorithms: MREdictor (Incarnato, Neri, Diamanti, & Oliviero, 2013), RNAhybrid (Kruger & Rehmsmeier, 2006), and Pita (Kertesz, Iovino, Unnerstall, Gaul, & Segal, 2007) .
The RNA-seq of the present manuscript are deposited in:
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=mhgdymwgxloxvij&acc=GSE77324

6.5 Chromatin Immunoprecipitation assay (ChIP)

Approximately $7*10^6$ cells were cross-linked by addition of formaldehyde to 1% for 10 min at RT, quenched with 0.125 M glycine for 5 min at RT, and then washed twice in cold PBS. The cells were resuspended in Lysis Buffer 1 (50 mM Hepes-KOH pH 7.5, 140mM NaCl, 1mM EDTA, 10% Glycerol, 0.5% NP-40, 0.25% Triton X-100 and protease inhibitor) to disrupt the cell membrane and in Lysis Buffer 2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1mM EDTA, 0.5 mM EGTA and protease inhibitor) to isolate nuclei. The isolated nuclei were then resuspended in SDS ChIP Buffer (20 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS and protease inhibitors). Extracts were sonicated using the Bioruptor H Twin (Diagenode) for 2 runs of 10 cycles [30 sec ''ON'', 30 sec ''OFF''] at high power setting. Cell lysate was centrifuged at 12,000 g for 10 min at 4°C. The supernatant was diluted with ChIP Dilution Buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton) before immunoprecipitation step. Streptavidin beads

(Dynabeads®Protein G, Life Technologies) were saturated with PBS/1% BSA and the samples were incubated with 2 mg of antibody overnight at 4°C on a rotator. Next day samples were incubated with saturated beads for two hours at 4°C on a rotator. Successively immunoprecipitated complexes were washed five times with RIPA buffer (50 mM Hepes-KOH pH7.6, 500mM LiCl, 1mM EDTA, 1% NP-40, 0,7% Na-Deoxycholate) at 4°C for 5 minutes each on a rotator. Elution Buffer was added and incubated at 65°C for 15 minutes. The decrosslinking was performed at 65°C overnight. Decrosslinked DNA was purified using QIAQuick PCR Purification Kit (QIAGEN) according to the manufacture's instruction.

6.6 ChIP-seq library preparation

Approximately 10 ng of purified ChIP DNA were end-repaired, dA-tailed, and adapter-ligated using the NEBNext ChIP-seq Library Prep Master Mix Set (NEB), following manufacturer instructions.

6.7 Reads mapping and data analysis

Samples were sequenced on the HiScanSQ or Next500 platforms (Illumina). All of the analysed data sets, were mapped to a recently published variant of the Mus musculus mm9 genome assembly. Prior mapping, sequencing reads were trimmed from low quality score basis and clipped from the adapter sequence by using FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). For RNA-seq data analysis, reads were mapped using TopHat v2.0.6 and mRNA quantification was performed using Cuffdiff v2.0.2. For ChIP-seq data analysis, reads were mapped Bowtie version 0.12.7, reporting only unique hits with up to two mismatches (parameters: -m 1 –v 2).

6.8 Constructs and transfection

shRNAs were constructed using the TRC hairpin design tool (http://www.broadinstitute.org/rnai/public/seq/search), choosing the hairpin sequences provided in Table 2. Annealed oligonucleotides were cloned into pLKO.1 vector (Addgene: 10,878) and each construct was verified by sequencing. For shRNA transient transfection, 5 µg of PLKO.1 vectors against lncRNAs and control hairpins were incubated with cells for 24 h. CCR102 silencing in chromatin-related experiments was obtained by using GapmeRs (IDT Custum Oligos) targeting CCR102 or random scrambled sequence.

To generate the CCR492 expression plasmid the full length of CCR492 was PCR amplified from MEF-cDNA and inserted into pCCLsin.PPT.hPGK.GFPpre using AscI/SalI restriction sites as previously described [16]. The primer sequences are as follow: CCR492-FW: 5'-TTGGCGCGCCTATCTTTATCCCTGAACTTTCTTTC-3' and CCR492-REV: 5'-

GCGTCGACCATCATGTGTCTGTGTACAAG-3'. A deleted form of CCR492 was produced by PCR to create CCR492 Δ162 with a new reverse primer: CCR492 Δ162-REV: 5'-GCGTCGACTAACCGCTGATATCTCTCCAG-3'. A mutant form of CCR492 with within the shRNA #1 target sequence, CCR492 shMut, was produced with two-step PCR procedure and these primers: CCR492 shMut-FW: 5'-CTAGATTCTAGATGGAAAAATAAAGGAT GAAGCAAC-3'; CCR492 shMut-REV: 5'-CATCCTTTATTTTTCCATCTAGAA TCTAGTTCCTC-3'. A 297-bp fragment of c-Myc 3'UTR containing the let- 7 target-miRNA seed region was PCR amplified from c-Myc cells cDNA, and cloned into MluI/SpeI sites of pMIR-Report (Invitrogen) to generate pMIR-RLuc-Myc-3'UTR. Primer sequences are as follows: c-Myc-3'UTRFW: 5'-ACGCGTACTGACCTAACTCGAGGAGGA-3' and c-Myc-3'UTR-REV: 5'-CCCTATTTACATGGGAAAATTGGATACTAGT-3'. The CCR492 mutant in the 4 let-7 binding sites was generate by QuikChange II XL Site-Directed Mutagenesis kit. Transient transfections of the constructs were performed using Lipofectamine 2000 Transfection Reagent (Invitrogen) according to the manufacturer's protocol.

6.9 Luciferase assay

For the luciferase assays, $3x10^3$ BALB/c fibroblasts were seeded per well in a 96-well plate the day before transfection. The cells were transfected with: 25 ng of pMIR-RLuc-Myc-3'UTR vector bearing the tested MRE, 2.25 ng of SV40-Renilla and either 86.9 ng of shControl or 86.9 ng of shCCR492, plus 86.9 ng of empty vector or 86.9 ng of CCR492 Mutant. For complementation assay cells were also transfected with 30 nM miRNA Inhibitor negative control #1 (Ambion) or a 30 nM pool of let-7 miRNA Inhibitors (Ambion). Firefly luciferase activity was assayed after 48 h using the Dual Luciferase Reporter Assay System (Promega, cat. E1910) and was normalized over the Renilla intensity.

6.10 Northern blot and rapid amplification of cDNA ends (RACE)

Northern blot analysis was performed with NorthernMax kit (Ambion) following manufacturer instructions. In brief, 4 μg of polyA+ RNAs was purified from total RNA of serum-induced fibroblasts and run on formaldehyde 1% agarose gel, transferred to BrightStar-Plus Membrane (Ambion) followed by cross-linking through UV irradiation. The membrane was subjected to hybridization with 0.1 nM of biotinylated sntisense probe overnight at 68 °C.
The probe complementary to CCR492 was transcribed in vitro with T7-FlashScribe Transcription kit (CellScript) and labeled with biotin-16- UTP (Roche) for a length of 351 nt (561–912), using the following primers: AS-PROBE-FW:
5'-GAGATAATACGACTCACTATAGGGAGAAGCATCATGTGTCTGTGTACAAG-3' and
AS-PROBE-REV: 5'-TATCTTTATCCCTGAACTTTCTTTC-3'. After washing, membranes were exposed for 2 h. 5' and 3' RACE were performed using SMARTer RACE cDNA Amplification

kit (Clontech) and followed by DNA sequencing. The Gene-Specific Primers used for PCR of RACE analysis were as follows: CCR492-REV: 5'-TTCCAGCTCCTAAGGTCACTG-3' (5'RACE) and CCR492-FW: 5'-CTACTCTTATAAAGGACCCGACT-3' (3'RACE).

6.11 RNA/protein extracts and Western blotting

RNA and protein extracts of cytoplasmic and nucleoplasmic fractions were isolated using the PARIS kit (Ambion) according to the manufacturer's specifications.
Chromatin fraction was obtained with a further step on the nucleoplasmic fraction: the nuclei pellet was resuspended in a prechilled glycerol buffer (20 mM Tris-HCl pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 0.85 mM DTT, 0.125 mM PMSF, 50% glycerol) by gentle flicking of the tube. An equal volume of cold nuclei lysis buffer (10 mM HEPES pH 7.6, 1mM DTT, 7,5 mM $MgCl_{2,}$ 0.2 mM EDTA, 0.3 M NaCl, 1M UREA, 1% NP-40) was added, the tube was gently vortexed for 2 x 2 sec, incubated for 2 min on ice, and then centrifuged for 2 min, 4°C, 14000 rpm. The supernatant correspond to the nuclear fraction while the pellet the chromatin fraction. The distribution of lncRNA in cellular compartment was calculated as percentage ratio of lncRNA nuclear, cytoplasmic and chromatin over the whole total lncRNA expression levels in all compartments. For total protein extracts, cells were resuspended in F-buffer (10 mM Tris-HCl pH 7.0, 50 mM NaCl, 30 mM Napyrophosphate, 50 mM NaF, anti-proteases) and sonicated for 3 pulses. Extracts were quantified using bicinchoninic acid (BCA) assay (Pierce) and were run in SDS-polyacrylamide gels, transferred to nitrocellulose membranes and incubated 16 h with specific primary antibodies.

6.12 RNA Fluorescent in Situ Hybridization (RNA-FISH)

5'-biotinylated sense (negative Ctrl) and antisense (CCR492-AS and CCR102-AS) probes were purchased from Eurofins MWG Operon. RNA-FISH analysis was performed with both probes in starved and serum-induced cells. $2x10^4$ BALB/c fibroblasts were seeded onto 0.1% gelatin-coated glass slide chambers. Cells were rinsed once in PBS buffer, fixed in 4% paraformaldehyde and permeabilized with 0.5% TRITON X-100. Cells were saturated in saturation buffer (SSC Buffer, 1% BSA, 1 µg yeast tRNA) for 2 h at room temperature. Then, 30 ng of 5'-biotinylated probes were denatured at 80 °C for 5 min and added to hybridization buffer (SSC Buffer, 10% formamide, 100 mg/ml dextran sulfate). Cells were incubated at 37 °C for 3 h in a humidified chamber protected from light for the hybridization. After incubation, RNA probes were revealed using Streptavidin, Alexa Fluor 488 conjugate (Life Technologies). DAPI (0.5mg/ml) was used to visualize cell nuclei and the images were performed with Leica TCS SP5 confocal microscope.

6.13 RNA pull-down assay

Whole-cell extracts were prepared from 3.5x 10^6 BALB/c fibroblasts with RIP buffer (150 mM KCl, 25 mM Tris-HCl pH 7.4, 0.5 mM DTT, 0.5% NP40, 1 mM PMSF, protease inhibitor and 20 U/ml RNaseout (Invitrogen)). The same number of cells was used for each condition. For precipitation assays, 1 µg of lncRNA-antisense probes and whole-cell extracts were incubated for 2 h at 4 °C; complexes were isolated with Streptavidin-coupled Dynabeads (Invitrogen). Beads were washed 5 times for 5 min at 4 °C with RIP buffer. The co-precipitated RNAs were extracted with phenol:chloroform:isoamyl alcohol and detected by RT-qPCR. Proteins isolated from the beads were detected by Western blot.

6.14 Proliferation assay and fluorescence-activated cell sorting (FACS) analysis

For cell-growth assay, 5x10^4 cells were plated in 35mm wells and counted at the indicated time point using Scepter™ Automated Cell Counter (Millipore). A growth curve was plotted to examine the effects of the shRNAs on cell proliferation. 72 h post-transfection, the cells were harvested, fixed in 70% ethanol, and stained for flow cytometry analysis. For the One-dimensional cell-cycle analysis, the cells were stained with propidium iodide (PI) solution (0.1% TRITON X-100, 200 mg/ml RNase, 20 mg/ml PI in PBS buffer) for 30 min at room temperature. Two dimensional cell-cycle analysis was performed using Click-iT EdU-Cell Cycle 633-red assay from Invitrogen according to the manufacturer's protocol with 1 h of EdU pulse. Acquisition was performed using Becton Dickinson FACS Canto and analysis was done with FACS FlowJo Software.

6.15 Monolayer and soft agar colony-formation assays

For monolayer colony-formation assay, 1000 BALB/c fibroblasts were transduced with pCCL-CCR492 and H-RasV12 (Addgene Plasmid 9051: pBABEpuro H-RasV12) expression vectors and plated in 100 mm2 plates and allowed to grow in appropriate culture medium for 10 days. Fresh media were supplied every 3 days. Colonies were stained with crystal violet dye after formaldehyde fixation. Soft agar colony-formation assays, 3x10^4 cells/ml were transduced in the same conditions and seeded on top of a solidified layer in a volume of 2ml of 0.5% Bacto Agar (Sigma-Aldrich) over 2 ml 0.4% agar base layers in each six-well plate.

6.16 Bioinformatics analysis of colorectal cancer tissues

Gene expression analysis of RP11-1058N17.1 (other names: lncSMAD7, AC114684.1) was performed using The Cancer Genome Atlas TCGA database (http://cancergenome.nih.gov/). Statistical analyses were performed using two-tailed Student's t-test.

6.17 Statistical data analysis

Data are presented as mean values ± SD of minimally 2 independent experiments; **P<0.01, *P<0.05 by the Student's t-test.

6.18 Antibodies

The antibodies were purchased from
Abcam (anti-c-Myc-ab11917, anti-Ago2-ab186733, anti-H3K4me1-ab8895, anti-Med12-ab70842, anti-pSmad3-ab28379), Sigma-Aldrich (anti-b-tubulin- T8328), SantaCruz (anti-Ras-sc520, anti-LaminA-sc20680, anti-p300-sc48343), Millipore (anti-H3K4me3-07473, anti-H3K27ac-07360, anti-CTCF-07729) Cell Signaling (anti-Ezh2-4905).

References

Ameres, S. L., & Zamore, P. D. (2013). Diversifying microRNA sequence and function. *Nature Reviews Molecular Cell Biology*, *14*(8), 475–488. http://doi.org/10.1038/nrm3611

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., et al. (2014). An atlas of active enhancers across human cell types and tissues. *Nature*, *507*(7493), 455–461. http://doi.org/10.1038/nature12787

Arase, M., Horiguchi, K., Ehata, S., Morikawa, M., Tsutsumi, S., Aburatani, H., et al. (2014). Transforming growth factor-β-induced lncRNA-Smad7 inhibits apoptosis of mouse breast cancer JygMC(A) cells. *Cancer Science*, *105*(8), 974–982. http://doi.org/10.1111/cas.12454

Ayupe, A. C., Tahira, A. C., Camargo, L., Beckedorff, F. C., Verjovski-Almeida, S., & Reis, E. M. (2015). Global analysis of biogenesis, stability and sub-cellular localization of lncRNAs mapping to intragenic regions of the human genome. *RNA Biology*, *12*(8), 877–892. http://doi.org/10.1080/15476286.2015.1062960

Balk, B., Maicher, A., Dees, M., Klermund, J., Luke-Glaser, S., Bender, K., & Luke, B. (2013). Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. *Nature Structural & Molecular Biology*, *20*(10), 1199–1205. http://doi.org/10.1038/nsmb.2662

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., & Haussler, D. (2004). Ultraconserved elements in the human genome. *Science*, *304*(5675), 1321–1325. http://doi.org/10.1126/science.1098119

Berridge, M. J. (2012). Cell Signalling Biology: Module 9 - Cell Cycle and Proliferation. *Biochemical Journal*, *6*, csb0001009. http://doi.org/10.1042/csb0001009

Bierhoff, H., Dammert, M. A., Brocks, D., Dambacher, S., Schotta, G., & Grummt, I. (2014). Quiescence-induced LncRNAs trigger H4K20 trimethylation and transcriptional silencing. *Molecular Cell*, *54*(4), 675–682. http://doi.org/10.1016/j.molcel.2014.03.032

Blagosklonny, M. V., & Pardee, A. B. (2002). The restriction point of the cell cycle. *Cell Cycle*, *1*(2), 103–110.

Blomen, V. A., & Boonstra, J. (2007). Cell fate determination during G1 phase progression. *Cellular and Molecular Life Sciences*, *64*(23), 3084–3104. http://doi.org/10.1007/s00018-007-7271-z

Blower, M. D. (2016). Centromeric Transcription Regulates Aurora-B Localization and Activation. *CellReports*, *15*(8), 1624–1633. http://doi.org/10.1016/j.celrep.2016.04.054

Boulay, J.-L., Mild, G., Lowy, A., Reuter, J., Lagrange, M., Terracciano, L., et al. (2003). SMAD7 is a prognostic marker in patients with colorectal cancer. *International Journal of Cancer*, *104*(4), 446–449. http://doi.org/10.1002/ijc.10908

Boyerinas, B., Park, S. M., Hau, A., Murmann, A. E., & Peter, M. E. (2010). The role of let-7 in cell differentiation and cancer. *Endocrine Related Cancer*, *17*(1), F19–F36. http://doi.org/10.1677/ERC-09-0184

Bracken, A. P., Kleine-Kohlbrecher, D., Dietrich, N., Pasini, D., Gargiulo, G., Beekman, C., et al. (2007). The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. *Genes & Development*, *21*(5), 525–530. http://doi.org/10.1101/gad.415507

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafrenière, R. G., Xing, Y., Lawrence, J., & Willard, H. F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*, *71*(3), 527–542.

Bueno, M. J., & Malumbres, M. (2011). Biochimica et Biophysica Acta. *BBA - Molecular Basis of Disease*, *1812*(5), 592–601. http://doi.org/10.1016/j.bbadis.2011.02.002

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., & Rinn, J. L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & Development*, *25*(18), 1915–1927. http://doi.org/10.1101/gad.17446611

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., et al. (2012). Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature*, *491*(7424), 454–457. http://doi.org/10.1038/nature11508

Cech, T. R., & Steitz, J. A. (2014). The Noncoding RNA Revolution— Trashing Old Rules to Forge New Ones. *Cell*, *157*(1), 77–94. http://doi.org/10.1016/j.cell.2014.03.008

Center, M. M., Jemal, A., & Ward, E. (2009). International trends in colorectal cancer incidence rates. *Cancer Epidemiology, Biomarkers & Prevention : a Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *18*(6), 1688–1694. http://doi.org/10.1158/1055-9965.EPI-09-0090

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., et al. (2011). A Long Noncoding RNA Controls Muscle Differentiation by Functioning as a Competing Endogenous RNA. *Cell*, *147*(2), 358–369. http://doi.org/10.1016/j.cell.2011.09.028

Chu, C., Quinn, J., & Chang, H. Y. (2012). Chromatin Isolation by RNA Purification (ChIRP). *Journal of Visualized Experiments*, (61). http://doi.org/10.3791/3912

Chujo, T., Yamazaki, T., & Hirose, T. (2016). Architectural RNAs (arcRNAs): A class of long noncoding RNAs that function as the scaffold of nuclear bodies. *Biochimica Et Biophysica Acta*, *1859*(1), 139–146. http://doi.org/10.1016/j.bbagrm.2015.05.007

Clemson, C. M., Hutchinson, J. N., Sara, S. A., Ensminger, A. W., Fox, A. H., Chess, A., & Lawrence, J. B. (2009). An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Molecular Cell*, *33*(6), 717–726. http://doi.org/10.1016/j.molcel.2009.01.026

Core, L. J., Martins, A. L., Danko, C. G., Waters, C. T., Siepel, A., & Lis, J. T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature Genetics*, *46*(12), 1311–1320. http://doi.org/10.1038/ng.3142

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences*, *107*(50), 21931–21936. http://doi.org/10.1073/pnas.1016071107

Crick, F. (1970). Central dogma of molecular biology. *Nature*, *227*(5258), 561–563.

Davidovich, C., & Cech, T. R. (2015). The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. *RNA (New York, N.Y.)*, *21*(12), 2007–2022. http://doi.org/10.1261/rna.053918.115

Davis, C. P., & West, J. A. (2015). Purification of specific chromatin regions using oligonucleotides: capture hybridization analysis of RNA targets (CHART). *Methods in Molecular Biology (Clifton, N.J.)*, *1262*, 167–182. http://doi.org/10.1007/978-1-4939-2253-6_10

De Simone, V., Bevivino, G., Sedda, S., Izzo, R., Laudisi, F., Dinallo, V., et al. (2017). Smad7 knockdown activates protein kinase RNA-associated eIF2α pathway leading to colon cancer cell death. *Cell Death and Disease*, *8*(3), e2681. http://doi.org/10.1038/cddis.2017.103

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science*, *295*(5558), 1306–1311. http://doi.org/10.1126/science.1067799

Deng, L. (2015). Long noncoding RNA CCAT1 promotes hepatocellular carcinoma progression by functioning as let-7 sponge, 1–10. http://doi.org/10.1186/s13046-015-0136-7

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Research*, *22*(9), 1775–1789. http://doi.org/10.1101/gr.132159.111

Dimitrova, N., Zamudio, J. R., Jong, R. M., Soukup, D., Resnick, R., Sarma, K., et al. (2014). LincRNA-p21 Activates p21 In cisto Promote Polycomb Target Gene Expression and to Enforce the G1/S Checkpoint. *Molcel*, *54*(5), 777–790. http://doi.org/10.1016/j.molcel.2014.04.025

Duttke, S. H. C., Lacadie, S. A., Ibrahim, M. M., Glass, C. K., Corcoran, D. L., Benner, C., et al. (2015). Human promoters are intrinsically directional. *Molecular Cell*, *57*(4), 674–684. http://doi.org/10.1016/j.molcel.2014.12.029

Ebert, M. S., & Sharp, P. A. (2010). Emerging roles for natural microRNA sponges. *Current Biology : CB*, *20*(19), R858–61. http://doi.org/10.1016/j.cub.2010.08.052

Eilers, M., & Eisenman, R. N. (2008). Myc's broad reach. *Genes & Development*, *22*(20), 2755–2766. http://doi.org/10.1101/gad.1712408

ENCODE Project Consortium, Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, *447*(7146), 799–816. http://doi.org/10.1038/nature05874

Engreitz, J. M., Haines, J. E., Perez, E. M., Munson, G., Chen, J., Kane, M., et al. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*, *539*(7629), 452–455. http://doi.org/10.1038/nature20149

Engreitz, J. M., Sirokman, K., McDonel, P., Shishkin, A. A., Surka, C., Russell, P., et al. (2014). RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell*, *159*(1), 188–199. http://doi.org/10.1016/j.cell.2014.08.018

Faghihi, M. A., & Wahlestedt, C. (2009). 2009 Faghihi, 1–7. http://doi.org/10.1038/nrm2738

Fatica, A., & Bozzoni, I. (2013). Long non-coding RNAs: new players in cell differentiation and development. *Nature Publishing Group*, *15*(1), 7–21. http://doi.org/10.1038/nrg3606

Fort, A., Hashimoto, K., Yamada, D., Salimullah, M., Keya, C. A., Saxena, A., et al. (2014). Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nature Genetics*, *46*(6), 558–566. http://doi.org/10.1038/ng.2965

Franco-Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., et al. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. *Nature Genetics*, *39*(8), 1033–1037. http://doi.org/10.1038/ng2079

Garzon, R., Calin, G. A., & Croce, C. M. (2009). MicroRNAs in Cancer. *Annual Review of Medicine*, *60*, 167–179. http://doi.org/10.1146/annurev.med.59.053006.104707

Geisler, S., & Coller, J. (2013). RNA in unexpected places:long non-coding RNA functionsin diverse cellular contexts. *Nature Reviews Molecular Cell Biology*, *14*(11), 699–712. http://doi.org/10.1038/nrm3679

Gibb, E. A., Brown, C. J., & Lam, W. L. (2011). The functional role of long non-coding RNA in human carcinomas. *Molecular Cancer*, *10*(1), 38. http://doi.org/10.1186/1476-4598-10-38

Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*, *159*(3), 647–661. http://doi.org/10.1016/j.cell.2014.09.029

Gong, C., & Maquat, L. E. (2011). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature*, *470*(7333), 284–288. http://doi.org/10.1038/nature09701

Gu, J., & Iyer, V. R. (2006). PI3K signaling and miRNA expression during the response of quiescent human fibroblasts to distinct proliferative stimuli. *Genome Biology*, *7*(5), R42. http://doi.org/10.1186/gb-2006-7-5-r42

Gunzburg, M. J., Sivakumaran, A., Pendini, N. R., Yoon, J.-H., Gorospe, M., Wilce, M. C. J., & Wilce, J. A. (2015). Cooperative interplay of let-7 mimic and HuR with MYC RNA. *Cell Cycle*, *14*(17), 2729–2733. http://doi.org/10.1080/15384101.2015.1069930

Gutschner, T., & Diederichs, S. (2014). The hallmarks of cancer. *RNA Biology*, *9*(6), 703–719. http://doi.org/10.4161/rna.20481

Guttman, M., & Rinn, J. L. (2012). Modular regulatory principles of large non-coding RNAs. *Nature*, *482*(7385), 339–346. http://doi.org/10.1038/nature10887

Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, *458*(7235), 223–227. http://doi.org/10.1038/nature07672

Guttman, M., Donaghey, J., Carey, B. W., Garber, M., Grenier, J. K., Munson, G., et al. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*, *477*(7364), 295–300. http://doi.org/10.1038/nature10398

Halder, S. K., Rachakonda, G., Deane, N. G., & Datta, P. K. (2008). Smad7 induces hepatic metastasis in colorectal cancer. *British Journal of Cancer*, *99*(6), 957–965. http://doi.org/10.1038/sj.bjc.6604562

Han, P., & Chang, C.-P. (2015). Long non-coding RNA and chromatin remodeling. *RNA Biology*, *12*(10), 1094–1098. http://doi.org/10.1080/15476286.2015.1063770

Hezroni, H., Koppstein, D., Schwartz, M. G., Avrutin, A., Bartel, D. P., & Ulitsky, I. (2015). Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *CellReports*, *11*(7), 1110–1122. http://doi.org/10.1016/j.celrep.2015.04.023

Hu, G., Gong, A.-Y., Wang, Y., Ma, S., Chen, X., Chen, J., et al. (2016). LincRNA-Cox2 Promotes Late Inflammatory Gene Transcription in Macrophages through Modulating SWI/SNF-Mediated Chromatin Remodeling. *The Journal of Immunology*, *196*(6), 2799–2808. http://doi.org/10.4049/jimmunol.1502146

Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M. J., Kenzelmann-Broz, D., et al. (2010). A Large Intergenic Noncoding RNA Induced by p53 Mediates Global Gene Repression in the p53 Response. *Cell*, *142*(3), 409–419. http://doi.org/10.1016/j.cell.2010.06.040

Hung, T., Wang, Y., Lin, M. F., Koegel, A. K., Kotake, Y., Grant, G. D., et al. (2011). Nature Publishing Group 2011 Hung. *Nature Genetics*, *43*(7), 621–629. http://doi.org/10.1038/ng.848

Hutchinson, J. N., Ensminger, A. W., Clemson, C. M., Lynch, C. R., Lawrence, J. B., & Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics*, *8*, 39. http://doi.org/10.1186/1471-2164-8-39

Ikushima, H., & Miyazono, K. (2010). TGFbeta signalling: a complex web in cancer progression. *Nature Publishing Group*, *10*(6), 415–424. http://doi.org/10.1038/nrc2853

Incarnato, D., Neri, F., Diamanti, D., & Oliviero, S. (2013). MREdictor: a two-step dynamic interaction model that accounts for mRNA accessibility and Pumilio binding accurately predicts microRNA targets. *Nucleic Acids Research*, *41*(18), 8421–8433. http://doi.org/10.1093/nar/gkt629

Iyer, M. K., Niknafs, Y. S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nature Genetics*, *47*(3), 199–208. http://doi.org/10.1038/ng.3192

Iyer, V. R., Eisen, M. B., Ross, D. T., Schuler, G., Moore, T., Lee, J. C., et al. (1999). The transcriptional program in the response of human fibroblasts to serum. *Science*, *283*(5398), 83–87.

Johnson, C. D., Esquela-Kerscher, A., Stefani, G., Byrom, M., Kelnar, K., Ovcharenko, D., et al. (2007). The let-7 MicroRNA Represses Cell Proliferation Pathways in Human Cells. *Cancer Research*, *67*(16), 7713–7722. http://doi.org/10.1158/0008-5472.CAN-07-1083

Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, *467*(7314), 430–435. http://doi.org/10.1038/nature09380

Kallen, A. N., Zhou, X.-B., Xu, J., Qiao, C., Ma, J., Yan, L., et al. (2013). The Imprinted H19 LncRNA Antagonizes Let-7 MicroRNAs. *Molcel*, *52*(1), 101–112. http://doi.org/10.1016/j.molcel.2013.08.027

Karreth, F. A., Tay, Y., Perna, D., Ala, U., Tan, S. M., Rust, A. G., et al. (2011). In Vivo Identification of Tumor-Suppressive PTEN ceRNAs in an Oncogenic BRAF-Induced Mouse Model of Melanoma. *Cell*, *147*(2), 382–395. http://doi.org/10.1016/j.cell.2011.09.032

Kato, G. J., & Dang, C. V. (1992). Function of the c-Myc oncoprotein. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *6*(12), 3065–3072.

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., & Segal, E. (2007). The role of site accessibility in microRNA target recognition. *Nature Genetics*, *39*(10), 1278–1284. http://doi.org/10.1038/ng2135

Kim, H. H., Kuwano, Y., Srikantan, S., Lee, E. K., Martindale, J. L., & Gorospe, M. (2009). HuR recruits let-7/RISC to repress c-Myc expression. *Genes & Development*, *23*(15), 1743–1748. http://doi.org/10.1101/gad.1812509

Kim, T.-K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature*, *465*(7295), 182–187. http://doi.org/10.1038/nature09033

Kino, T., Hurt, D. E., Ichijo, T., Nader, N., & Chrousos, G. P. (2010). Noncoding RNA Gas5 Is a Growth Arrest- and Starvation-Associated Repressor of the Glucocorticoid Receptor. *Science Signaling*, *3*(107), ra8–ra8. http://doi.org/10.1126/scisignal.2000568

Kruger, J., & Rehmsmeier, M. (2006). RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Research*, *34*(Web Server), W451–W454. http://doi.org/10.1093/nar/gkl243

Kumar, M. S., Lu, J., Mercer, K. L., Golub, T. R., & Jacks, T. (2007). Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nature Genetics*, *39*(5), 673–677. http://doi.org/10.1038/ng2003

Lai, F., Orom, U. A., Cesaroni, M., Beringer, M., Taatjes, D. J., Blobel, G. A., & Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature*, *494*(7438), 497–501. http://doi.org/10.1038/nature11884

Lane, D. P. (1992). Cancer. p53, guardian of the genome. *Nature*, *358*(6381), 15–16. http://doi.org/10.1038/358015a0

Lee, J. T., Davidow, L. S., & Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. *Nature Genetics*, *21*(4), 400–404. http://doi.org/10.1038/7734

Lenhard, B., Sandelin, A., & Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature Reviews Genetics*, *13*(4), 233–245. http://doi.org/10.1038/nrg3163

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A. Y., et al. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, *498*(7455), 516–520. http://doi.org/10.1038/nature12210

Li, Y.-Y., Yu, H., Guo, Z.-M., Guo, T.-Q., Tu, K., & Li, Y.-X. (2006). Systematic analysis of head-to-head gene organization: evolutionary conservation and potential biological relevance. *PLoS Computational Biology*, *2*(7), e74. http://doi.org/10.1371/journal.pcbi.0020074

Liang, W.-C., Fu, W.-M., Wong, C.-W., Wang, Y., Wang, W.-M., Hu, G.-X., et al. (2015). The lncRNA H19 promotes epithelial to mesenchymal transition by functioning as miRNA sponges in colorectal cancer. *Oncotarget*, *6*(26), 22513–22525. http://doi.org/10.18632/oncotarget.4154

Lin, J. M., Collins, P. J., Trinklein, N. D., Fu, Y., Xi, H., Myers, R. M., & Weng, Z. (2007). Transcription factor binding and modified histones in human bidirectional promoters. *Genome Research*, *17*(6), 818–827. http://doi.org/10.1101/gr.5623407

Liu, X., Li, D., Zhang, W., Guo, M., & Zhan, Q. (2012). Long non-coding RNA gadd7 interacts with TDP-43 and regulates Cdk6 mRNA decay. *The EMBO Journal*, *31*(23), 4415–4427. http://doi.org/10.1038/emboj.2012.292

Luo, S., Lu, J. Y., Liu, L., Yin, Y., Chen, C., Han, X., et al. (2016). Divergent lncRNAs Regulate Gene Expression and Lineage Differentiation in Pluripotent Cells. *Cell Stem Cell*, *18*(5), 637–652. http://doi.org/10.1016/j.stem.2016.01.024

Luteijn, M. J., & Ketting, R. F. (2013). PIWI-interacting RNAs: from generation to transgenerational epigenetics. *Nature Reviews Genetics*, *14*(8), 523–534. http://doi.org/10.1038/nrg3495

Maldotti, M., Incarnato, D., Neri, F., Krepelova, A., Rapelli, S., Anselmi, F., et al. (2016). The long intergenic non-coding RNA CCR492 functions as a let-7 competitive endogenous RNA to regulate c-Myc expression. *Biochimica Et Biophysica Acta*, *1859*(10), 1322–1332. http://doi.org/10.1016/j.bbagrm.2016.06.010

Massagué, J. (2008). TGFβ in Cancer. *Cell*, *134*(2), 215–230. http://doi.org/10.1016/j.cell.2008.07.001

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, *495*(7441), 333–338. http://doi.org/10.1038/nature11928

Mercer, T. R., & Mattick, J. S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. *Nature Publishing Group*, *20*(3), 300–307. http://doi.org/10.1038/nsmb.2480

Mondal, T., Rasmussen, M., Pandey, G. K., Isaksson, A., & Kanduri, C. (2010). Characterization of the RNA content of chromatin. *Genome Research*, *20*(7), 899–907. http://doi.org/10.1101/gr.103473.109

Mondal, T., Subhash, S., Vaid, R., Enroth, S., Uday, S., Reinius, B., et al. (2015). MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA-DNA triplex structures. *Nature Communications*, *6*, 7743. http://doi.org/10.1038/ncomms8743

Morris, K. V., & Mattick, J. S. (2014). The rise of regulatory RNA. *Nature Reviews Genetics*, *15*(6), 423–437. http://doi.org/10.1038/nrg3722

Nagano, T., Mitchell, J. A., Sanz, L. A., Pauler, F. M., Ferguson-Smith, A. C., Feil, R., & Fraser, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science*, *322*(5908), 1717–1720. http://doi.org/10.1126/science.1163802

Nakao, A., Afrakhte, M., Morén, A., Nakayama, T., Christian, J. L., Heuchel, R., et al. (1997). Identification of Smad7, a TGFbeta-inducible antagonist of TGF-beta signalling. *Nature*, *389*(6651), 631–635. http://doi.org/10.1038/39369

Nam, Y., Chen, C., Gregory, R. I., Chou, J. J., & Sliz, P. (2011). Molecular Basis for Interaction of let-7 MicroRNAs with Lin28. *Cell*, 1–21. http://doi.org/10.1016/j.cell.2011.10.020

Nurse, P. (2002). Cyclin dependent kinases and cell cycle control (nobel lecture). *Chembiochem : a European Journal of Chemical Biology*, *3*(7), 596–603. http://doi.org/10.1002/1439-7633(20020703)3:7<596::AID-CBIC596>3.0.CO;2-U

Ong, C.-T., & Corces, V. G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Reviews Genetics*, *12*(4), 283–293. http://doi.org/10.1038/nrg2957

Pardee, A. B. (1989). G1 events and regulation of cell proliferation. *Science*, *246*(4930), 603–608.

Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., et al. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, *408*(6808), 86–89. http://doi.org/10.1038/35040556

Plasari, G., Calabrese, A., Dusserre, Y., Gronostajski, R. M., McNair, A., Michalik, L., & Mermod, N. (2009). Nuclear factor I-C links platelet-derived growth factor and transforming growth factor beta1 signaling to skin wound healing progression. *Molecular and Cellular Biology*, *29*(22), 6006–6017. http://doi.org/10.1128/MCB.01921-08

Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W. J., & Pandolfi, P. P. (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*, *465*(7301), 1033–1038. http://doi.org/10.1038/nature09144

Portoso, M., Ragazzini, R., Brenčič, Ž., Moiani, A., Michaud, A., Vassilev, I., et al. (2017). PRC2 is dispensable for HOTAIR-mediated transcriptional repression. *The EMBO Journal*, *36*(8), 981–994. http://doi.org/10.15252/embj.201695335

Postepska-Igielska, A., Giwojna, A., Gasri-Plotnitsky, L., Schmitt, N., Dold, A., Ginsberg, D., & Grummt, I. (2015). LncRNA Khps1 Regulates Expression of the Proto-oncogene SPHK1 via Triplex-Mediated Changes in Chromatin Structure. *Molecular Cell*, *60*(4), 626–636. http://doi.org/10.1016/j.molcel.2015.10.001

Preker, P., Almvig, K., Christensen, M. S., Valen, E., Mapendano, C. K., Sandelin, A., & Jensen, T. H. (2011). PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. *Nucleic Acids Research*, *39*(16), 7179–7193. http://doi.org/10.1093/nar/gkr370

Pritchard, C. C., & Grady, W. M. (2011). Colorectal cancer molecular biology moves into clinical practice. *Gut*, *60*(1), 116–129. http://doi.org/10.1136/gut.2009.206250

Pruitt, K. D., Tatusova, T., Brown, G. R., & Maglott, D. R. (2012). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Research*, *40*(Database issue), D130–5. http://doi.org/10.1093/nar/gkr1079

Recillas-Targa, F., Pikaart, M. J., Burgess-Beusse, B., Bell, A. C., Litt, M. D., West, A. G., et al. (2002). Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(10), 6883–6888. http://doi.org/10.1073/pnas.102179399

Redrup, L., Branco, M. R., Perdeaux, E. R., Krueger, C., Lewis, A., Santos, F., et al. (2009). The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing. *Development (Cambridge, England)*, *136*(4), 525–530. http://doi.org/10.1242/dev.031328

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., et al. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*, *403*(6772), 901–906. http://doi.org/10.1038/35002607

Rinn, J. L., & Chang, H. Y. (2012). Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry*, *81*(1), 145–166. http://doi.org/10.1146/annurev-biochem-051410-092902

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, *129*(7), 1311–1323. http://doi.org/10.1016/j.cell.2007.05.022

Salmena, L., Poliseno, L., Tay, Y., Kats, L., & Pandolfi, P. P. (2011). A ceRNA Hypothesis: The Rosetta Stone of a Hidden RNA Language? *Cell*, *146*(3), 353–358. http://doi.org/10.1016/j.cell.2011.07.014

Sampson, V. B., Rong, N. H., Han, J., Yang, Q., Aris, V., Soteropoulos, P., et al. (2007). MicroRNA let-7a down-regulates MYC and reverts MYC-induced growth in Burkitt lymphoma cells. *Cancer Research*, *67*(20), 9762–9770. http://doi.org/10.1158/0008-5472.CAN-07-2462

Schaukowitch, K., & Kim, T. K. (2014). REVIEWEMERGING EPIGENETIC MECHANISMS OF LONG NON-CODING RNAS. *Neuroscience*, *264*(C), 25–38. http://doi.org/10.1016/j.neuroscience.2013.12.009

Seila, A. C., Calabrese, J. M., Levine, S. S., Yeo, G. W., Rahl, P. B., Flynn, R. A., et al. (2008). Divergent transcription from active promoters. *Science*, *322*(5909), 1849–1851. http://doi.org/10.1126/science.1162253

Siggens, L., & Ekwall, K. (2014). Epigenetics, chromatin and genome organization: recent advances from the ENCODE project. *Journal of Internal Medicine*, *276*(3), 201–214. http://doi.org/10.1111/joim.12231

Sigova, A. A., Mullen, A. C., Molinie, B., Gupta, S., Orlando, D. A., Guenther, M. G., et al. (2013). Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proceedings of the National Academy of Sciences*, *110*(8), 2876–2881. http://doi.org/10.1073/pnas.1221904110

Spitzer, J., Hafner, M., Landthaler, M., Ascano, M., Farazi, T., Wardle, G., et al. (2014). PAR-CLIP (Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation): a step-by-step protocol to the transcriptome-wide identification of binding sites of RNA-binding proteins. *Methods in Enzymology*, *539*, 113–161. http://doi.org/10.1016/B978-0-12-420120-0.00008-6

Stefani, G., Chen, X., Zhao, H., & Slack, F. J. (2015). A novel mechanism of LIN-28 regulation of let-7 microRNA expression revealed by in vivo HITS-CLIP in C. elegans. *RNA (New York, N.Y.)*, *21*(5), 985–996. http://doi.org/10.1261/rna.045542.114

Stolfi, C., De Simone, V., Colantoni, A., Franzè, E., Ribichini, E., Fantini, M. C., et al. (2014). A functional role for Smad7 in sustaining colon cancer cell growth and survival. *Cell Death and Disease*, *5*, e1073. http://doi.org/10.1038/cddis.2014.49

Sumazin, P., Yang, X., Chiu, H.-S., Chung, W.-J., Iyer, A., Llobet-Navas, D., et al. (2011). An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell*, *147*(2), 370–381. http://doi.org/10.1016/j.cell.2011.09.041

Tang, Y., Wang, J., Lian, Y., Fan, C., Zhang, P., Wu, Y., et al. (2017). Linking long non-coding RNAs and SWI/SNF complexes to chromatin remodeling in cancer. *Molecular Cancer*, *16*(1), 42. http://doi.org/10.1186/s12943-017-0612-0

The FANTOM Consortium. (2005). The Transcriptional Landscape of the Mammalian Genome. *Science*, *309*(5740), 1559–1563. http://doi.org/10.1126/science.1112014

Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., et al. (2010). The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation. *Molecular Cell*, *39*(6), 925–938. http://doi.org/10.1016/j.molcel.2010.08.011

Tripathi, V., Shen, Z., Chakraborty, A., Giri, S., Freier, S. M., Wu, X., et al. (2013). Long Noncoding RNA MALAT1 Controls Cell Cycle Progression by Regulating the Expression of Oncogenic Transcription Factor B-MYB. *PLoS Genetics*, *9*(3), e1003368. http://doi.org/10.1371/journal.pgen.1003368.s014

Ulitsky, I., & Bartel, D. P. (2013). lincRNAs: Genomics, Evolution, and Mechanisms. *Cell*, *154*(1), 26–46. http://doi.org/10.1016/j.cell.2013.06.020

Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H., & Bartel, D. P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell*, *147*(7), 1537–1550. http://doi.org/10.1016/j.cell.2011.11.055

Venters, B. J., & Pugh, B. F. (2013). Genomic organization of human transcription initiation complexes. *Nature*, *502*(7469), 53–58. http://doi.org/10.1038/nature12535

Vidal, A., & Koff, A. (2000). Cell-cycle inhibitors: three families united by a common cause. *Gene*, *247*(1-2), 1–15.

Wan, G., Mathur, R., Hu, X., Liu, Y., Zhang, X., Peng, G., & Lu, X. (2013). Long non-coding RNA ANRIL (CDKN2B-AS) is induced by the ATM-E2F1 signaling pathway. *Cellular Signalling*, *25*(5), 1086–1095. http://doi.org/10.1016/j.cellsig.2013.02.006

Wang, J., Liu, X., Wu, H., Ni, P., Gu, Z., Qiao, Y., et al. (2010). CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. *Nucleic Acids Research*, *38*(16), 5366–5383. http://doi.org/10.1093/nar/gkq285

Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*, 1–7. http://doi.org/10.1038/nature09819

Wang, Y., Xu, Z., Jiang, J., Xu, C., Kang, J., Xiao, L., et al. (2013). Endogenous miRNA Sponge lincRNA-RoR Regulates Oct4, Nanog, and Sox2in Human Embryonic Stem Cell Self-Renewal. *Developmental Cell*, *25*(1), 69–80. http://doi.org/10.1016/j.devcel.2013.03.002

Werner, M. S., & Ruthenburg, A. J. (2015). Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes. *CellReports*, *12*(7), 1089–1098. http://doi.org/10.1016/j.celrep.2015.07.033

Werner, M. S., Sullivan, M. A., Shah, R. N., Nadadur, R. D., Grzybowski, A. T., Galat, V., et al. (2017). Chromatin-enriched lncRNAs can act as cell-type specific activators of proximal gene transcription. *Nature Structural & Molecular Biology*, *24*(7), 596–603. http://doi.org/10.1038/nsmb.3424

Wrighton, K. H., Lin, X., & Feng, X.-H. (2009). Phospho-control of TGF-beta superfamily signaling. *Cell Research*, *19*(1), 8–20. http://doi.org/10.1038/cr.2008.327

Wu, L., Nguyen, L. H., Zhou, K., de Soysa, T. Y., Li, L., Miller, J. B., et al. (2015). Precise let-7 expression levels balance organ regeneration against tumor suppression. *eLife*, *4*, e09431. http://doi.org/10.7554/eLife.09431

Yan, P., Luo, S., Lu, J. Y., & Shen, X. (2017). Cis- and trans-acting lncRNAs in pluripotency and reprogramming. *Current Opinion in Genetics & Development*, *46*, 170–178. http://doi.org/10.1016/j.gde.2017.07.009

Yin, Q.-F., Yang, L., Zhang, Y., Xiang, J.-F., Wu, Y.-W., Carmichael, G. G., & Chen, L.-L. (2012). Long Noncoding RNAs with snoRNA Ends. *Molecular Cell*, *48*(2), 219–230. http://doi.org/10.1016/j.molcel.2012.07.033

Yoon, J.-H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J. L., De, S., et al. (2012). Short Article. *Molecular Cell*, *47*(4), 648–655. http://doi.org/10.1016/j.molcel.2012.06.027

Zhang, S., Fei, T., Zhang, L., Zhang, R., Chen, F., Ning, Y., et al. (2007). Smad7 antagonizes transforming growth factor beta signaling in the nucleus by interfering with functional Smad-DNA complex formation. *Molecular and Cellular Biology*, *27*(12), 4488–4499. http://doi.org/10.1128/MCB.01636-06

Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J., & Lee, J. T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*, *322*(5902), 750–756. http://doi.org/10.1126/science.1163045

Ørom, U. A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., et al. (2010). Long noncoding RNAs with enhancer-like function in human cells. *Cell*, *143*(1), 46–58. http://doi.org/10.1016/j.cell.2010.09.001

Appendix

Publications

1. Neri F, Rapelli S, Krepelova A, Incarnato D, Parlato C, Basile G, **Maldotti M**, Anselmi F, Oliviero S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. NATURE, vol. 543, p. 72-77, ISSN: 0028-0836, doi: 10.1038/nature21373
IRIS AperTo ID (handle): http://hdl.handle.net/2318/1634241

2. **Maldotti M**, Incarnato D, Neri F, Krepelova A, Rapelli S, Anselmi F, Parlato C, Basile G, Dettori D, Calogero R, Oliviero S. (2016). The long intergenic non-coding RNA CCR492 functions as a let-7 competitive endogenous RNA to regulate c-Myc expression. BIOCHIMICA ET BIOPHYSICA ACTA, vol. 1859, p. 1322-1332, ISSN: 0006-3002, doi: 10.1016/j.bbagrm.2016.06.010
IRIS AperTo ID (handle): http://hdl.handle.net/2318/1588990

3. Incarnato D, Anselmi F, Morandi E, Neri F, **Maldotti M**, Rapelli S, Parlato C, Basile G, Oliviero S. (2016). High-throughput single-base resolution mapping of RNA 2'-O-methylated residues. NUCLEIC ACIDS RESEARCH, vol. 45, p. 1433-1441, ISSN: 0305-1048, doi: 10.1093/nar/gkv392
IRIS AperTo ID (handle): http://hdl.handle.net/2318/1558946

4. Neri F, Dettori D, Incarnato D, Krepelova A, Rapelli S, **Maldotti M**, Parlato C, Paliogiannis P, Oliviero S (2015). TET1 is a tumour suppressor that inhibits colon cancer growth by derepressing inhibitors of the WNT pathway. ONCOGENE, vol. 34, p. 4168-4176, ISSN: 0950-9232, doi: 10.1038/onc.2014.356
IRIS AperTo ID (handle): http://hdl.handle.net/2318/150019

5. Neri F, Incarnato D, Krepelova A, Dettori D, Rapelli S, **Maldotti M**, Parlato C, Anselmi F, Galvagni F, Oliviero S. (2015). TET1 is controlled by pluripotency-associated factors in ESCs and downmodulated by PRC2 in differentiated cells and tissues. NUCLEIC ACIDS RESEARCH, vol. 43, p. 6814-6826, ISSN: 0305-1048, doi: 10.1093/nar/gkv392
IRIS AperTo ID (handle): http://hdl.handle.net/2318/141912

Acknowledgments

I would like to thank my supervisors, Professor Salvatore Oliviero for giving me the great opportunity to work in his lab and for his considerable scientific mentor in this important PhD experience. Six years have passed, but I often remember the day when I arrived in Turin and you welcomed me into your lab. For all of these reasons, I just want to say thank you Prof.

I am also very grateful to all members of the Oliviero's lab, past and present, for their precious scientific advices and the time spent together: Danny, Francesca, Lisa, Caterina, Giulia, Francesco, Anna, Edoardo, Matteo, Valentina and Andrea. Thank you.

A special thanks to Stefania, for the support and the friendship that she gave me in these years, finally, I can really say thank you Ste!

In addition, I would like to acknowledge the reviewers of my PhD Thesis: Professor Irene Bozzoni and Professor Giannino Del Sal.

Last but not the least, I want to thank my family, my parents and my little sister Giulia, for their support in my studies, the encouragement to follow my dreams and the fantastic environment in which I grew up.

Thank you grandma, your memory is always alive.

And the biggest thank to Filippo, who has been at my side in these years and who has always given me unconditional love and support, especially in this last period of the Thesis.

December 2017,

Mara