Cristina Barbagallo

LncRNA and circRNA expression profiles in tissues and serum exosomes of colorectal cancer patients and cell lines

TESI DI DOTTORATO

Coordinatore del Dottorato: Prof. Lorenzo Malatino
Coordinatore del Curriculum: Prof. Michele Purrello
Tutor: Prof. Marco Ragusa

Anno Accademico 2015-2016
# Table of contents

1. Abstract ................................................................................................................................ 1

2. Introduction .......................................................................................................................... 3
   2.1 Long non-coding RNAs ................................................................................................. 8
      2.1.1 LncRNA biogenesis .............................................................................................. 9
      2.1.2 LncRNA functions ............................................................................................... 10
   2.2 Circular RNAs .............................................................................................................. 16
      2.2.1 CircRNA biogenesis ............................................................................................ 17
      2.2.2 CircRNA functions ............................................................................................. 22
   2.3 Small non-coding RNAs .............................................................................................. 25
      2.3.1 MiRNA biogenesis .............................................................................................. 25
      2.3.2 MiRNA functions ................................................................................................. 28
   2.4 Exosomes ..................................................................................................................... 30
      2.4.1 Exosomes as mediators of intercellular communication ....................................... 36
   2.5 Colorectal cancer ........................................................................................................ 39
      2.5.1 CRC genetic alterations ....................................................................................... 40
      2.5.2 CRC diagnosis and therapy ................................................................................ 47
   2.6 NcRNAs as mediators of tumor progression ............................................................... 52
      2.6.1 LncRNA role in CRC pathogenesis ..................................................................... 52
      2.6.2 CircRNAs and colorectal cancer ......................................................................... 56
2.6.3 MiRNAs involved in CRC carcinogenesis ........................................................... 57

2.7 Exosomes in tumor progression ........................................................................... 58

2.8 NcRNAs in biological fluids ................................................................................. 59

3. Materials and methods ............................................................................................ 61

3.1 Cell lines .............................................................................................................. 61

3.2 Isolation of exosomes secreted by CRC cell lines .............................................. 61

3.3 RNA isolation from cell lines and exosomes ...................................................... 62

3.4 RNA isolation from CRC formalin-fixed, paraffin-embedded samples .......... 62

3.5 Exosome and RNA isolation from CRC patient serum ..................................... 62

3.6 PCR primer design .............................................................................................. 63

3.7 LncRNA and circRNA expression analysis by Real-Time PCR ..................... 65

3.8 Computational analysis ....................................................................................... 65

3.9 Biological network construction and Gene ontology analysis ......................... 69

3.10 ROC curve analysis ........................................................................................... 69

3.11 CRC cell lines treatment with MAPK inhibitors ............................................. 70

3.12 Protein extraction and Western Blot analysis ................................................ 70

3.13 In vitro silencing of UCA1 ........................................................................... 71

4. Results .................................................................................................................... 72

4.1 NcRNA expression in CRC cell lines and exosomes ....................................... 72

4.2 NcRNA expression in FFPE CRC biopsies ..................................................... 73

4.3 NcRNA expression in serum exosomes of CRC patients ............................ 76
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.4 Analysis of GEO DataSets and identification of mRNA and miRNA targets of DE ncRNAs</td>
<td>78</td>
</tr>
<tr>
<td>4.5 Analysis of gene ontologies</td>
<td>82</td>
</tr>
<tr>
<td>4.6 Evaluation of the diagnostic accuracy of DE ncRNAs in serum exosomes</td>
<td>84</td>
</tr>
<tr>
<td>4.7 <em>In vitro</em> inhibition of MAPKs</td>
<td>86</td>
</tr>
<tr>
<td>4.7.1 Cell viability assay</td>
<td>86</td>
</tr>
<tr>
<td>4.7.2 Confirmation of MAPK inhibition through Western Blot analysis</td>
<td>86</td>
</tr>
<tr>
<td>4.7.3 MAPK inhibition affects ncRNA expression</td>
<td>87</td>
</tr>
<tr>
<td>4.7.4 Evaluation of correlation between ncRNA expression and ERK activation</td>
<td>88</td>
</tr>
<tr>
<td>4.8 UCA1 target expression after <em>in vitro</em> silencing</td>
<td>89</td>
</tr>
<tr>
<td>5. Discussion</td>
<td>91</td>
</tr>
<tr>
<td>5.1 l NcRNAs contribute to CRC progression</td>
<td>91</td>
</tr>
<tr>
<td>5.2 NcRNAs in serum exosomes of CRC patients: new diagnostic biomarkers?</td>
<td>96</td>
</tr>
<tr>
<td>5.3 Identification of miRNA and mRNA targets of DE ncRNAs: future perspectives for the understanding of miRNA-ncRNA and mRNA-ncRNA interaction effects</td>
<td>98</td>
</tr>
<tr>
<td>5.4 Altered expression of ncRNAs after treatment with MAPK inhibitors</td>
<td>100</td>
</tr>
<tr>
<td>5.5 <em>In vitro</em> silencing of UCA1 affects the expression of its targets</td>
<td>102</td>
</tr>
<tr>
<td>6. Conclusions and future perspectives</td>
<td>105</td>
</tr>
<tr>
<td>7. Bibliography</td>
<td>106</td>
</tr>
</tbody>
</table>
1. Abstract

In the last few years several studies demonstrated the fundamental role of non-coding RNAs (ncRNAs) in tumor onset and progression. While the involvement and the mechanism of action of microRNAs (miRNAs) have been widely investigated, little is known about long non-coding RNAs (lncRNAs) and circular RNAs (circRNAs), which seem to act through a plethora of molecular mechanisms regulating essential biological processes, such as cell cycle, splicing, chromatin remodeling, apoptosis, adhesion and migration. LncRNAs are non-coding RNAs longer than 200 nucleotides, with or without 5’-cap and poly(A) tail; circRNAs are circular molecules lacking free ends and thus resistant to exonucleasic degradation, transcribed from protein-coding genes. Several evidences showed that aberrant expression of lncRNAs and circRNAs is associated with various tumors, including CRC. In our study we analysed through Real-Time PCR the expression of a set of 17 lncRNAs and 31 circRNAs selected from literature in 20 CRC tissues compared to normal adjacent tissues (NATs), and in serum exosomes of 20 CRC patients compared to 20 healthy individuals. We identified 8 ncRNAs (CCAT1, CCAT2, CDR1AS, HOTAIR, MALAT1, TUG1, UCA1 and ZEB2AS1) differentially expressed (DE) in tissues and 3 ncRNAs (circ16, TUG1 and UCA1) DE in serum exosomes. Through gene ontology analysis we verified the involvement of DE ncRNAs in pathways associated with tumor progression; we also evaluated the diagnostic accuracy of ncRNAs deregulated in serum exosomes through ROC curves, suggesting their possible application in CRC non-invasive diagnosis. We observed that inhibition of MAPK pathway, associated with CRC, altered the expression of HOTAIR, MALAT1, TUG1 and UCA1 in HCT-116 colon cancer cells treated with U0126 (MEK1/2 specific inhibitor) compared to untreated cells, suggesting a possible connection between ncRNAs and MAPKs. Also, we performed a computational analysis to identify mRNAs and miRNAs involved in CRC and characterized by correlation of expression and sequence complementarity with ncRNAs deregulated in CRC tissues.
Finally, we performed *in vitro* functional assays by silencing UCA1 expression through ASOs in HCT-116 cells and analysed the expression of its mRNA and miRNA targets, aiming to investigate the eventual degradation triggered by the interaction between ncRNAs and their miRNA/mRNA targets. Our study provide new data about aberrant expression of ncRNAs in CRC and their involvement in carcinogenesis; further analyses will be necessary to fully understand ncRNA molecular mechanisms and to evaluate their possible application in CRC diagnosis and therapy.
2. Introduction

A few years after the discovery of DNA structure, Francis Crick stated the central dogma of molecular biology: “Once information has got into a protein it can't get out again” (1); with this simple sentence he summarized the flow of genetic information through DNA, RNA and proteins (Figure 2.1).

![Figure 2.1: the original concept diagram published by Francis Crick in 1956.](image)

Shortly after, François Jacob and Jacques Monod confirmed the central role of RNA in genetic information transfer from DNA to proteins (2). At that time, with few evidences, it was assumed that all biological processes in cells were regulated by proteins, characterized by more complex structures, while nucleic acids were considered genetic information reservoirs. To date, a
variety of new technologies and skills have been improved and a huge amount of sequencing data has been produced, leading researchers all over the world to go beyond these models and reconsider RNA role in cellular biology.

In 1990 an international consortium (including USA, United Kingdom, France, Australia, and China) launched the Human Genome Project, aiming to determine the entire sequence and to detect all genes of human genome (3). In 2001 a first draft of the human genome was published (4), followed by the final version in 2004 (5). Until then, scientists tried to indirectly estimate the total number of human protein-coding genes using many approaches, such as the typical vertebrate tissue mRNA complexity, the approximate ratio of typical gene size to the genome size, the number of CpG islands and the frequency of their association with known genes; finally it was supposed a total of 30,000-35,000 protein-coding genes in human genome (3). The final results of Human Genome Project revealed that our genome includes approximately 20,000-25,000 protein-coding genes (5), but this number has been recently corrected in about 19,000 genes (6). Surprisingly, this number is lower than expected, and very close to total protein-coding gene number of less complex organisms (Caenorhabditis elegans, 19,735 genes, Gallus gallus, 16,736 genes, Drosophila melanogaster, 14,889 genes, Mus musculus (21,839 genes) (7) (Figure 2.2).
This discovery surprised scientists, which believed that *Homo sapiens* higher complexity was due to higher gene number compared to inferior organisms. Also, it was observed that all genes identified during the Human Genome Project represent less than 2% of the entire genome, while in other organisms, eukaryotic or prokaryotic, the coding fraction of genome is considerably higher (8) (Figure 2.3).

**Figure 2.2:** total number of protein-coding genes comparison between *Homo sapiens* and inferior organisms (from https://www.scienecnews.org/article/more-chicken-fewer-grape).

**Figure 2.3:** coding genome percentage comparison between *Homo sapiens* and other organisms, both prokaryotic and eukaryotic (from Sana et al., *J Transl Med.* 2012).
This illusory paradox was partially explained by alternative splicing events and post-translational modifications, which allow cells to produce more proteins with different functions from a single gene. However, scientists concluded that a large fraction of the human genome, even 98%, was represented by “junk DNA”, that is DNA with no function.

In order to understand the molecular bases of Homo sapiens higher complexity and the “junk DNA” function, the ENCODE (Encyclopedia of DNA Elements) Project was launched with the purpose to identify all functional elements in the human genome (9). Final results, confirmed by following studies, showed that the 85-90% of human genome is pervasively transcribed: numerous new transcription start sites were identified, as well as many new transcripts, originated from genes located at non-coding regions or overlapping protein-coding loci (10, 11, 12). Since these new transcripts were not translated into proteins, they were defined non-coding RNAs (ncRNAs). It was recently estimated that non-coding gene number tends to exponentially increase ascending the evolutionary scale, while protein-coding gene number reaches a plateau: therefore, Homo sapiens higher complexity in comparison to inferior organisms with similar protein-coding gene number may be due to this new class of RNAs (13) (Figure 2.4).
Next Generation Sequencing (NGS) new technologies led to the identification of ncRNA sequences, stressing their great structural heterogeneity. The same heterogeneity can also be observed in the structure of ncRNA genes: i) their length is extremely variable, from few tens of nucleotides to several hundreds of kilobases; ii) they lack ORFs (Open Reading Frames); iii) evolutionary conservation is often low or absent; iv) there is no preferential localization within the genome. Also, a relative tolerance to point mutations has been observed (14, 15). These features make ncRNA genes difficult to identify, which is why their exact number is to date unknown; however, it has been estimated that the human genome includes from 10,000 to 20,000 non-protein coding genes (16). Classifying these molecules is also difficult: two different classification are currently used, based on ncRNA length or function. According to length, ncRNAs are divided into i) long non-coding RNAs (IncRNAs), which are longer than 200 nucleotides, and ii) small non-coding RNAs, whose length is equal or inferior to 200 nucleotides; the latter includes ncRNA molecules deeply analysed, such as microRNAs (miRNAs), small interfering RNAs (siRNAs), and PIWI-interacting RNAs (piRNAs). The second classification, based on function, consists of i)
Introduction

housekeeping ncRNAs, constitutively expressed and responsible of fundamental physiological processes; ii) regulatory ncRNAs, whose expression can be induced in different cell histotypes, in defined development and differentiation stages or following external stimuli. Housekeeping ncRNAs include transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), and telomerase RNAs, while regulatory ncRNAs comprise miRNAs, siRNAs, piRNAs and lncRNAs (14, 15).

2.1 Long non-coding RNAs

LncRNAs represent a highly heterogeneous class of ncRNAs, characterized by length greater than 200 nucleotides (cut-off value derived from biochemical techniques that allow to exclude all known classes of small RNAs) frequently reaching 100 kb (14). These transcripts are considered non-protein coding because they lack long ORF (more than 100 codons) and/or codon conservation; however, recent studies showed that some lncRNAs can code for small proteins (17). LncRNA genes are difficult to identify because of their variable features; indeed, they can:

i. Be located within intergenic regions, generating long intergenic non-coding RNAs (lincRNAs) (18);

ii. Be located inside introns of protein-coding genes (19);

iii. Partially overlap untranslated regions (UTRs) or protein-coding gene promoters (20, 21);

iv. Be pseudogenes whose transcripts (sense or antisense) control the expression of their protein-coding paralog genes (22, 23);

v. Lie within transcribed ultra-conserved regions (tUCRs), characterized by high evolutionary conservation and located at intra- or intergenic regions (24);

vi. Be located within the mitochondrial genome and regulated by proteins coded by nuclear genes (25).
The discovery of ncRNAs amazed researchers and started off a debate about the effective functions of these molecules, initially considered “transcriptional noise” (26, 27). Nevertheless, numerous studies have highlighted that ncRNA expression is tissue-specific, and it varies in different development or differentiation stages or in response to external stimuli (28, 29, 30, 31). These data allowed to go beyond the “transcriptional noise” hypothesis and sparked scientist interest in IncRNAs: recent studies showed that these molecules play an important role in biological processes, both physiological, such as cell cycle (32, 33), apoptosis (34, 35), differentiation (36), stemness (37), aging (38), and pathological (39, 40, 41). In particular, many researchers focused their attention on the possible involvement of IncRNAs in neoplastic processes, which by now seems to be demonstrated (42, 43, 44, 45, 46).

2.1.1 LncRNA biogenesis

Most of IncRNAs share with mRNAs the transcription by RNA polymerase II, the epigenetic regulation (histone modifications) and splicing signals (47); only a small part of them is transcribed by RNA polymerase I and III (48, 49). About half of IncRNAs are post-transcriptionally processed undergoing the same steps of mRNAs (5’-cap addition, polyadenylation, splicing), although some of them are monoexonic and non-polyadenylated (47, 50). Poly(A) tail protects 3’-ends from enzymatic degradation; in its absence, the presence of alternative structures with similar function, such as triple helix, has been shown. This kind of structure has been identified at 3’-end of MALAT1 (metastasis-associated lung adenocarcinoma transcript 1), a 8.7-kb IncRNA transcribed from a gene located at chromosome 11. During MALAT1 processing, the primary transcript loses from its 3’-end a 61-nucleotide fragment, called mascRNA (MALAT1-associated small cytoplasmic RNA); hence, the new 3’-end of mature transcript presents a stem-loop structure and three conserved motifs, composed of two U-rich and one A-rich tracts; this A-rich tract is similar to a poly(A) tail, but it’s too short to be recognized by poly(A)-binding proteins. These three motifs
interact with each other and form a triple helix structure that stabilizes MALAT1 3’-end and protects it from enzymatic degradation (51) (Figure 2.5).

**Figure 2.5:** triple helix structure stabilizing MALAT1 3’-end in absence of poly(A) tail. Dashed lines show base triplets (from Wilusz, Biochim Biophys Acta. 2016).

### 2.1.2 LncRNA functions

LncRNAs play a crucial role in cellular physiology through a plethora of functions and molecular mechanisms. It is now clear that they can regulate gene expression at different levels (transcriptional, post-transcriptional, translational, post-translational, epigenetic) (52) (Figure 2.6); however, the exact molecular mechanism has been elucidated only for few lncRNAs.
Figure 2.6: lncRNAs regulate cell physiology performing several functions (from Morris and Mattick, Nat. Rev. Genet. 2014).

Xist (X inactive specific transcript) is one of the first IncRNAs identified (53, 54). Xist is a 19-kb IncRNA transcribed by the X chromosome; it is involved in the X-chromosome inactivation process, the epigenetic mechanism leading to the random inactivation of one of the two X
chromosomes in female individual cells. X inactivation depends on histone modifications (in particular lysine methylation) performed by PRC1 and PRC2 (Polycomb-repressing complex 1 and 2), which in turn are recruited on the inactive X chromosome by Xist (55). Similarly, other lncRNAs can modulate gene expression by recruiting chromatin remodeling complexes: one of the best known is HOTAIR (Hox transcript antisense intergenic RNA), a 2.2-kb lncRNA transcribed from the antisense strand of HOXC (Homeobox C) cluster on chromosome 12 (56, 57). HOTAIR also interacts with PRC2, inducing the trimethylation of histone H3 lysine 27, and the consequent transcription inhibition, of HOXD (Homeobox D) cluster on chromosome 2 (58). Moreover, HOTAIR interacts not only with PRC2 through a 5’-domain, but also with LSD1 (lysine demethylase 1A) through a 3’-domain, thus modulating demethylation and transcriptional activation of the same cluster (57, 59). TUG1 (taurine up-regulated 1) is a 7.5-kb lncRNA located at chromosome 22; its expression is p53-mediated thanks to many highly conserved p53-binding sites included in its promoter. After DNA damage, p53 binds TUG1 promoter and activates TUG1 expression; the lncRNA in turn recruits PRC2 and inhibits expression of cell cycle regulator genes (60).

LncRNAs can also bind transcriptional factors, recruiting them on DNA or sequestering them, therefore regulating the expression of specific genes. PANDAR (promoter of CDKN1A antisense DNA damage activated RNA) is a 1.5-kb polyadenylated and capped lncRNA, antisense transcript of CDKN1A (cyclin-dependent kinase inhibitor 1A) on chromosome 6. Its expression is regulated by p53 after DNA damage; PANDAR binds the transcription factor NFYA (nuclear transcription factor Y subunit alpha), impeding it to bind the promoters of pro-apoptotic genes (61).

Some lncRNAs can interact with DNA inducing loop formation, thus moving closer two otherwise very distant regions; these lncRNAs act as “bridges” between promoters and enhancers or silencers. Their mechanism of action is not known yet, although it has been proposed the formation of a RNA-DNA-DNA triplex or a RNA-protein-DNA complex (62). The DHFR (dihydrofolate
reductase) locus on chromosome 5 contains two promoters: the major promoter codes for the DHFR mRNA, while the upstream minor promoter codes for a ncRNA which negatively regulates mRNA levels in cis and in trans: the DHFR ncRNA interacts with the major promoter, forming a stable complex, and with TFIIB (transcription factor IIB), causing PIC (pre-initiation complex) dissociation and transcription inhibition (63) (Figure 2.7).

![Figure 2.7: DHFR ncRNA inhibits the transcription of DHFR mRNA by interacting with the major promoter and TFIIB (from Martianov et al., Nature 2007).](image)

LncRNAs can act at post-transcriptional level modulating mRNA stability, translation and splicing. BACE1AS (BACE1 antisense RNA) is a 825-nucleotide lncRNA transcribed from a gene located at chromosome 11, apparently involved in Alzheimer’s disease pathogenesis. BACE1AS is the natural antisense of BACE1 (beta-secretase 1), a transmembrane protease that catalyzes the first step in amyloid beta peptide synthesis; amyloid beta peptide accumulates in Alzheimer’s patient cerebral neurons, generating the typical amyloid plaques. Thanks to sequence complementarity, BACE1AS binds BACE1 mRNA, thus stabilizing it and inducing increased amyloid beta production (64). ZEB2AS1 (ZEB2 antisense RNA 1), antisense transcript of ZEB2 (zinc finger E-box binding homeobox 2), is a 430-nucleotide lncRNA transcribed from chromosome 2. ZEB2 plays a key role in neoplastic processes by repressing E-cadherin expression, thus promoting epithelial-mesenchymal transition (EMT). ZEB2AS1 is abundantly expressed in mesenchymal cells through the activation of an alternative promoter; the lncRNA binds the 5’-UTR of ZEB2 mRNA,
masking splicing sites and preventing the removal of a long intron (about 2.5 kb) including an internal ribosome entry site (IRES). So, in presence of its antisense, ZEB2 mRNA is recognised by ribosomes and translated, producing a transcription factor that induces E-cadherin and EMT repression (20) (Figure 2.8).

Another function of lncRNAs is to regulate cellular localization of proteins, thus modulating their activity. NRON (non-protein coding RNA, repressor of NFAT), a 2.7-kb transcript of chromosome 9, controls the trafficking of NFAT (nuclear factor of activated T cells) transcription factor: NFAT is normally contained in the cytoplasm, but it is imported to the nucleus in response to calcium-dependent signals. NRON interacts with several proteins, including members of the importin-beta superfamily, thus inhibiting NFAT nuclear accumulation (52, 65).

Some lncRNAs play structural roles, allowing correct formation of specific cellular structures. NEAT1 (nuclear paraspeckle assembly transcript 1) is a 4-kb monoexonic polyadenilated lncRNA transcribed from chromosome 11; it is contained in the nucleus overlapping SC35 splicing

---

**Figure 2.8:** ZEB2 mRNA alternative splicing mediated by its antisense ZEB2AS1 in mesenchymal cells; ZEB2 main promoter is shown in black, IRES-containing 3-kb 5'-UTR in gray, ZEB2 ORF in white (from Beltran et al., Genes Dev. 2008).
factor enriched regions; NEAT1 presence in the nucleus is essential for the formation of paraspeckles, nuclear domains involved in nuclear retention of specific mRNAs (66).

Recent studies revealed a connection between IncRNAs and small non-coding RNAs, such as miRNAs: some IncRNAs can be miRNA precursors or can interact with them, sequestering them and inhibiting their transcriptional repression activity; this function has been defined “miRNA sponge”. H19, the first IncRNA identified (67), is transcribed from a gene localized on chromosome 11, near IGF2 (insulin-like growth factor 2) locus; both genes are epigenetically silenced: only the maternal allele of H19 and the paternal allele of IGF2 are expressed, while the other ones are silenced by methylation (68). H19 gene produces a 2.3-kb capped, polyadenilated and spliced IncRNA which is expressed from the earliest stages of embryonic development in several organs and quite completely silenced after birth. H19 is the precursor of miR-675, which targets, among others, RB1 (retinoblastoma 1), a key cell cycle regulator (69). Plus, H19 contains many binding sites for let-7 miRNA family, well known tumor suppressor miRNAs; H19 can bind and sequester let-7 miRNAs, acting as “miRNA sponge” (70). TUSC7 (tumor suppressor candidate 7), a 2.1-kb IncRNA transcribed from chromosome 3, also acts as “sponge” for miR-211 thanks to two binding sites lying within its fourth exon (71). MALAT1 is highly expressed and retained inside the nucleus, where it acts as “riboregulator” modulating the expression of its target genes, involved in cell cycle regulation (51, 72). MALAT1 lacks poly(A) tail, and its 3’-end folds to form a tRNA-like secondary structure, recognised and cleaved by RNAs e P; this cleavage creates a short fragment, subsequently processed by enzymes involved in tRNA biogenesis; at the end of the process, a 61-nucleotide mature transcript called mascRNA is produced (Figure 2.9). MascRNA is exported to the cytoplasm but, despite its structure, it doesn’t contribute to protein synthesis, and its function is still unknown (51). A recent study suggested a mascRNA involvement in immune system regulation: high levels of mascRNA (but not of MALAT1) have been observed in human immune system circulating cells; furthermore, mascRNA downregulation in monocytes-macrophages...
Introduction

Strongly affects the expression of fundamental genes, including FASLG (Fas ligand), FAS (Fas cell surface death receptor), TNF-α (tumor necrosis factor alpha), and IL6 (interleukin 6) (73).

Figure 2.9: mascRNA biogenesis mediated by RNase P cleavage of 3’-end of MALAT1 (from Wilusz, Genes Dev. 2012).

Therefore, it is clear that IncRNAs perform multiple functions through very heterogeneous molecular mechanisms, escaping any attempt at classification.

2.2 Circular RNAs

Next Generation Sequencing techniques led to the identification of a unique class of IncRNAs represented by circular non-coding RNA molecules, called circular RNAs (circRNAs). The first single-stranded circular RNA molecule was identified in 1976 within viroids, infectious agents similar to viruses but lacking capsids (74); since then other circular RNAs have been
observed in other organisms, including animals, but they have been considered by-products of mRNA processing derived from splicing errors (75). Recently, several new circRNAs previously unknown have been identified, characterized by high expression levels in mammalian cells and, sometimes, by high evolutionary and expression conservation, suggesting a specific function in cellular physiology (76, 77).

2.2.1 CircRNA biogenesis

As it is known, one of the characteristics distinguishing prokaryotes and eukaryotes is the alternation of introns (non-coding segments) and exons (coding segments) in eukaryotic gene structure; after transcription, pre-mRNA undergoes a processing which includes splicing, that is intron removal by a ribonucleoprotein complex called spliceosome. Splicing has been widely studied; it consists of two phases, both regulated by the spliceosome: i) the 2′-OH of an intronic base, called branchpoint (usually an adenosine), performs a nucleophilic attack on the 5′-splicing site, localized at exon-intron boundary and identified by consensus dinucleotide GU, thus forming a phosphodiester 2′,5′-bond and releasing the 3′-OH end of the exon; ii) this 3′-OH in turn attacks on the 3′-splicing site, localized at intron-exon boundary and identified by consensus dinucleotide AG: hence, the two exons are bound through a phosphodiester 3′,5′-bond, while the intron is removed as a lariat, because of a 2′,5′-bond (Figure 2.10). This is an essential event finely regulated but also flexible, that allows cells to produce several protein variants modulating splicing sites choice (alternative splicing) (78).
Many circRNAs are transcribed from protein-coding gene exons, so they are called exonic circRNAs (ecircRNAs); it has been proposed that ecircRNA biogenesis biogenesis is catalyzed by the spliceosome following the same splicing mechanism. This hypothesis is supported by several evidences: both splicing inhibitors and mutagenesis of splicing sites (GU and AG consensus sequences) significantly reduce circular RNA levels (79). The exact mechanism of ecircRNA biogenesis is still unknown: however, it seems that spliceosome regulates backsplicing (or inverse splicing) events, where a downstream 5' splicing site (donor) is bound to an upstream 3' splicing site (acceptor); this mechanism produces a head-to-tail junction, typical of backsplicing events (80) (Figure 2.11).
Three models for ecircRNA biogenesis have been proposed:

i. Lariat-driven circularization (Figure 2.12a): during splicing events, some exons can be skipped and removed together with introns (exon skipping); ecircRNAs can be produced by a second splicing on exon-containing lariats, before they are unraveled by debranching enzymes (77);

ii. Intron pairing-driven circularization (Figure 2.12b): intronic complementary motifs flanking circularized exons can anneal and promote circularization (77);

iii. Resplicing-driven circularization (Figure 2.12c): sometimes, a second splicing event can happen on a mature mRNA in presence of specific splicing signals. The first event is regulated by canonical splicing sites and removes introns from pre-mRNA; the second one requires cryptic splicing sites, which are recognised only in specific conditions, and induces exon circularization. Re-splicing is likely an aberrant phenomenon, often observed in carcinogenesis (81).
As suggested by the intron pairing-driven circularization model, the presence of reverse complementary sequences (RCSs) in flanking introns can move exons closer and promote their circularization; indeed, several studies reported a positive correlation between flanking intron RCSs (often represented by Alu repeats in primates) and exon circularization (77, 82, 83). However, RCSs are not essential, and sometimes longer RCSs can even inhibit circularization through increased base-pairing stability; moreover, a gene can include many RCS copies competing with each other, so alternative circularization events can occur, thus raising ecircRNA number produced by a single gene (79, 84). Another factor promoting ecircRNA biogenesis is flanking intron length, that is...
three- to fivefold longer than in randomly selected introns both in human genome, (77, 82), and in other species (79). Again, long flanking introns are not essential for ecircRNA biogenesis, but longer introns are likely to include more cis-acting regulatory element (77, 85).

In addition to ecircRNAs, circular transcripts originated from spliced introns, called circular intronic RNAs (ciRNAs), have been identified. They are originated from the lariat, which is not debranched but circularized thanks to opportune signals, that are: i) a 7-nucleotide GU-rich sequence sited near to 5’ splicing site; ii) an 11-nucleotide C-rich sequence near to branchpoint; molecules participating in this process are still unknown (86) (Figure 2.13).

![Figure 2.13: consensus sequences needed for ciRNA biogenesis: a 7-nucleotide GU-rich sequence sited near to 5’ splicing site (pink) and an 11-nucleotide C-rich sequence near to branchpoint (blue) (from Zhang et al., Mol Cell. 2013).](image)

It has been hypothesized that the same factors regulating canonical (or linear) splicing also modulate backsplicing events; therefore, the two phenomena compete with each other to bind these molecules. A recent study showed that Quaking (QKI), an RNA-binding protein regulating alternative splicing, controls the expression of several circRNAs during EMT. QKI binds RCSs, moving circularized exons closer; the addition of QKI-binding sites in flanking introns highly increases circRNA production (87). ADAR (adenosine deaminase acting on RNA) protein family includes enzymes involved in RNA editing, a process modifying some nucleotides during pre-mRNA maturation. In particular, ADAR enzymes bind the adenosines of double-stranded RNA molecules, converting them into inosines through deamination. Recently, it has been observed that
the downregulation of ADAR enzymes induces increased circRNA expression, suggesting that these enzymes could negatively regulate circRNA biogenesis (83).

### 2.2.2 CircRNA functions

CircRNAs have been initially considered as splicing errors. Successively, some circRNAs showed evolutionary conservation between human and mouse and among species, probably because of the conservation of splicing sites; this observation led to the hypothesis of a molecular function performed by this class of ncRNAs (79). Another evidence supporting this hypothesis is the high number of different circRNAs produced by cells: according to the database circBase (http://www.circbase.org/), approximately 100,000 circRNAs have been identified in human cells (89). It has been estimated that circRNAs represent 1-10% of the entire cell transcriptome (76, 85, 90); most of them have very low expression levels (0.1-1%) compared to co-linear mRNAs, although sometimes the circular transcript exceeds the linear one (77); moreover, circRNAs are often tissue-specific, so they are expressed only in some cell histotypes, even if they tend to be more abundant in brain tissues (79).

Several studies reported the presence of ecircRNAs in the cytoplasm, while their biogenesis occurs within the nucleus: nucleus export systems likely regulate the transport of these molecules (76, 77, 90). One of ecircRNA functions, also performed by IncRNAs, is to bind miRNAs and modulate their action. Some ecircRNAs are strongly associated with Argonaute (Ago) proteins, key factors of miRNA-induced gene silencing (RNA interference, RNAi), and include many miRNA-binding sites. The most known example is represented by a circRNA called CDR1AS/ciRs-7 (CDR1 antisense) and its 80 binding sites for miR-7. CircRNA-miRNA interaction causes a reduced miRNA availability within cell cytoplasm, and consequently a reduced translational repression of specific mRNA targets: increased CDR1AS expression induces an incremented expression of proteins coded by mRNA targets of miR-7. On the contrary, CDR1AS
downregulation provokes reduced protein levels of miR-7 targets, because miR-7 is not sequestered by the circRNA (80, 91) (Figure 2.14). A bioinformatic study showed that circRNAs present miRNA-binding sites lacking polymorphisms, strengthening their miRNA sponge or miRNA reservoir function (92); nevertheless, most of circRNAs include a number of miRNA-binding sites comparable to co-linear mRNAs, suggesting that only some circRNAs act as miRNA sponge (90, 93).

![Figure 2.14: CDR1AS/ciRS-7 acts as miR-7 “sponge” (from Hansen et al., Cancer Res. 2013).](image)

Moreover, ecircRNAs can modulate parental gene expression through transcription regulation. Recent studies showed that some ecircRNAs are retained within the nucleus and interact with RNA polymerase II, suggesting an involvement in gene expression regulation; these ecircRNAs contain, in addition to exons, non-spliced introns, and therefore they have been called exon-intron circRNAs (EIcircRNAs) (94, 95, 96). EIcircRNAs are prevalently expressed in the nucleus; they promote parental gene transcription by binding small nuclear ribonucleoprotein U1 (snRNU1), a splicing regulator also involved in the initial stages of transcription; the EIcircRNA-snRNU1 complex interacts with RNA polymerase II on the promoter of the parental gene, promoting its expression (94) (Figure 2.15).
Figure 2.15: positive regulation of parental gene transcription mediated by ECircRNAs: they bind snRNP-U1 forming a complex which interacts with RNA polymerase II, promoting transcription (from Li et al., Nat Struct Mol Biol. 2015).

Another study suggested that circRNA biogenesis itself reduces co-linear mRNA expression. MBNL1 (muscleblind like splicing regulator 1) gene codes for a splicing regulating protein, but also for a circRNA called circMbl, produced co-transcriptionally. circMbl flanking introns include several binding sites for MBNL protein: when highly expressed, MBNL protein binds circMbl, promoting its circularization to the detriment of splicing. Thus, the cell reacts to MBNL accumulation in two ways: i) splicing is inhibited, reducing mature mRNA production; ii) protein availability is decreased through its circMbl-mediated sequestering (97).

Despite many ecircRNAs include ORFs, to date contrasting evidences suggesting peptide production has been reported; therefore, the possibility that ecircRNAs act as mRNAs needs further investigations (98).

CiRNA functions are not fully understood. A recent study showed that ciRNAs present reduced evolutionary conservation between human and mouse, accumulate in the nucleus and contain few miRNA-binding sites; these observations suggest different functions and mechanisms compared to ecircRNAs. CiRNAs interact with phosphorylated RNA polymerase II, suggesting that they act as positive regulator of gene expression, consistently with their nuclear localization. Interestingly, ciRNA expression is positively correlated with parental gene levels: it has been hypothesized that ciRNAs promote co-linear mRNA expression, as suggested by their accumulation
in the sites of synthesis. Low evolutionary conservation can be due to ciRNA origin: since introns
don’t code for functional proteins, they undergo no or very low evolutive pressure and accumulate
more mutations compared to exons (86).

2.3 Small non-coding RNAs

Small non-coding RNAs, characterized by length equal or inferior to 200 nucleotides, have
been widely studied for many years. By now, it is well known that they regulate many biological
processes: i) RNA synthesis, processing and translation and transcriptional initiation [piRNAs,
PASRs (promoter-associated small RNAs)]; ii) RNA maturation (snoRNAs); iii) RNA degradation
or translation inhibition (miRNA, siRNA) (15). Among all small non-coding RNAs, the most
known and studied are definitely miRNAs, whose biogenesis and function as post-transcriptional
repressors have been well characterized.

2.3.1 MiRNA biogenesis

MicroRNAs are a class of endogenous single-stranded RNAs consisting of approximately
18-25 nucleotides and expressed both in animals and in plants. MiRNA-coding genes are located at
all chromosomes in humans, except for Y chromosome; frequently, genes coding for different
miRNAs are located at adjacent loci on the same chromosome, forming clusters (e.g. let-7 cluster
on chromosome 9, including hsa-let-7a-1, hsa-let-7f-1 e hsa-let-7d). Transcription of a cluster is
simultaneous and generates a polycistronic primary transcript, successively processed into single
miRNAs. MiRNAs belonging to the same cluster are often related to each other, suggesting that
clusters derive from gene duplication; consequently, miRNAs from the same cluster are often
functionally correlated, since function depends on sequence. Genomic localization of miRNA-
coding genes is highly heterogeneous; indeed, they can be located at: i) intergenic regions, ii)
protein-coding gene introns, iii) non-coding gene introns, iv) non-coding gene exons; moreover,
some miRNA-coding genes have an independent promoter (99). According to miRBase database (http://www.mirbase.org/), to date 1881 precursors and 2588 mature miRNAs have been identified, even if a recent study hypothesized a higher number of miRNA-coding genes in human genome (100).

MiRNA biogenesis starts in the nucleus; transcription is prevalently performed by RNA polymerase II, rarely by RNA polymerase III; transcription produces a long (many kb) primary transcript called pri-miRNA, which folds into a double-stranded hairpin structure; the hairpin includes a loop and a double-stranded stem, made up of 33 complementary base pairs (bp), and ends with two long single-stranded traits (101, 102, 103). Successively, pri-miRNA undergoes two sequential cleavages: the first cleavage occurs in the nucleus by Drosha, the second one in the cytoplasm by Dicer, two endonucleases including conserved RNAse III catalytic domains (RRIIDa, RRIIDb); these enzymes act in association with proteins containing double-stranded RNA-binding domains (dsRBDs) (99). The first cleavage performed by a complex including Drosha and Pasha (DGCR8) converts the pri-miRNA into a 700-nucleotide long molecule with a hairpin structure called pre-miRNA, which is exported to the cytoplasm by exportin-5 (Exp5), a GTP-dependent nuclear/cytoplasmic transporter (104). In the cytoplasm, pre-miRNA interacts with a big multiprotein complex called RISC loading complex (RLC), consisting of the endonuclease Dicer, TRBP (Tar RNA Binding Protein), which contains three dsRBDs, PACT (protein activator of PKR), and Argonaute protein Ago-2, an RNAse with catalytic function (105, 106, 107, 108). Argonaute (Ago) proteins are expressed in all eukaryotes and contain specific domains: PAZ and MID domains to bind RNA targets at 3’-end, and PIWI domain at 5’-end to catalyze RNA cleavage (109). Dicer catalyzes the cleavage of the hairpin loop, producing a 22-nucleotide long double-stranded miRNA characterized by mismatches and two protrusive nucleotides at both 3’-ends. This RNA duplex is released from RLC and splits into two strands, defined: i) guide strand, complementary to the targets, and therefore functional; ii) passenger strand (miRNA*), initially
considered non-functional and degraded (110) (Figure 2.16). Recent studies showed that passenger strand also is functional and accumulates in the cytoplasm (111, 112, 113): by now, it is known that every miRNA-coding gene originates two mature molecules, defined -5p and -3p depending on the duplex strand from which they derive.

![Figure 2.16: miRNA biogenesis starts in the nucleus with the transcription of a long pri-miRNA performed by RNA polymerase II or III; the pri-miRNA folds into a hairpin structure and undergoes an enzymatic cleavage performed by Drosha, generating the pre-miRNA, characterized by a stem-loop structure. The pre-miRNA is exported to the cytoplasm; the hairpin structure undergoes a cleavage performed by Dicer, generating a 22-nucleotide long double-stranded miRNA (from Winter et al., Nat Cell Biol. 2009).](image)

A particular miRNA subfamily includes genes located within protein-coding gene introns, which code for miRNA called mirtrons. Mirtron biogenesis is Drosha-independent: after splicing, they are processed into a hairpin RNA, which is exported to the cytoplasm by exportin-5 and directly cleaved by Dicer (114) (Figure 2.17).
Figure 2.17: mirtron biogenesis is Drosha-independent: after removal, the intron lariat is exported to the cytoplasm by exportin-5 and cleaved by Dicer (from Ruby et al., Nature. 2007).

2.3.2 MiRNA functions

All miRNAs perform the same function of post-transcriptional repressors of gene expression through the same molecular mechanism.

Mature miRNAs bind Ago-2 protein and are led to the RNA-induced silencing complex (RISC), the true effector of miRNA-mediated silencing; RISC consists of several proteins, including Ago. Once associated to RISC, mature miRNA becomes active and recognizes its mRNA target: in animals, interaction occurs between the 3’-UTR of mRNAs, where many miRNA-binding sites can be located, and the seed region of miRNAs, included between the second and the eighth nucleotide of the 5’-end of mature miRNAs. MiRNA-mRNA complementarity determines mRNA destiny: i) a perfect match causes the degradation of the mRNA, which is first deprived of its poly(A) tail; ii) a partial complementarity induces the inhibition of mRNA translation (115) (Figure 2.18).
**Figure 2.18:** miRNA-mediated post-transcriptional silencing occurs through mRNA degradation or through inhibition of mRNA translation (from Vidigal and Ventura, Trends Cell Biol. 2015).

MiRNA-mediated mRNA degradation mechanism is still not clear; it seems that RISC recruits on mRNA target the effectors of deadenylation and degradation (exonucleases). This process starts in the cytoplasm and continues in specialized organelles called P-bodies (processing bodies), subcellular structures representing transcript storage and decay areas (109, 116). Translation inhibition seems to be due to RISC-included Ago proteins, which compete with eukaryotic translation initiation factors (eIFs) to bind mRNA 5’-cap: during the first stages of protein synthesis, eIF factors bind the 5’-cap, while PABPC1 [poly(A) binding protein cytoplasmic 1] binds poly(A) tail; eIFs-PABPC1 interaction induces the approach of the two ends of mRNA, promoting its recognition by ribosomes. When mRNA is recognized by miRNA loaded on RISC, Ago proteins bind its 5’-cap and prevent translation start (109).

It has been demonstrated that every miRNA can bind several mRNA targets, and, similarly, a single mRNA can be targeted by many miRNAs (99); therefore, miRNAs can be considered small effectors of wider regulatory pathways, controlling fundamental processes, such as cell cycle, differentiation, apoptosis, and also more complex processes regarding the entire organism, such as embryonic development, immune response, and many others.
2.4 Exosomes

In the last decade, the scientific community welcomed with great interest the discovery of a new mechanism of intercellular communication mediated by membranous vesicles, structures delimited by a phospholipid bilayer with transmembrane proteins; these vesicles carry the same molecules included within cell cytoplasm, such as hydrophilic solutes and biological macromolecules (proteins, lipids, RNAs), so they have been initially considered miniature cells. Further studies showed that cells secrete several types of vesicles, identified with different names (exosomes, exosome-like vesicles, microparticles, microvesicles, membrane-bound particles, apoptotic bodies, apoptotic microparticles), conferred in relation to vesicle dimension, density, origin and biogenesis. It is possible to classify vesicles in: i) exosomes, vesicles of endocytic origin secreted after multivesicular bodies and plasma membrane fusion; ii) shedding microvesicles, which directly origin from plasma membrane; iii) apoptotic bodies, derived from membrane blebbing during apoptosis (117, 118) (Figure 2.19).

![Figure 2.19: different biogenesis of shedding microvesicles, exosomes and apoptotic bodies (from Akers et al., J Neurooncol. 2013).](image-url)
Exosomes are small membrane vesicles characteristically cup-shaped, with a diameter of 30-100 nm \((119, 120)\) (Figure 2.20). Similarly to cells, exosomes present lipid rafts on their membrane and contain proteins and RNAs; typically, it is believed that exosomes lack DNA, even if some studies recently showed the contrary \((121, 122)\).

**Figure 2.20:** exosomes observed by transmission electron microscopy. Several stages of biogenesis are shown: i) intraluminal vesicles within multivesicular bodies; ii) intraluminal vesicles formation through inward invagination of multivesicular body membrane; iii) multivesicular bodies and plasma membrane fusion; iv) exosome secretion \((from Sahoo et al., Circ Res. 2011)\).

Exosomes have been identified for the first time in 1987 as vesicles secreted during maturation of reticulocytes, red blood cells precursors \((123)\); successively, it has been observed that similar structures are secreted by several cell histotypes, such as immune system cells \((124, 125, 126)\), platelets \((127)\), intestinal epithelial cells \((128)\), cancer cells \((129, 130)\). By now it is clear that many cell types, if not all, can secrete exosomes \((131, 132, 133)\), which are released in extracellular space and biological fluids: indeed, exosomes have been identified in all body fluids analysed to
date, such as urine (134), seminal fluid (135), amniotic fluid (136), saliva (137), breast milk and plasma (138).

Exosome biogenesis begins on plasma membrane with an endocytosis event that generates a vesicle called early endosome, typically characterized by tube-like shape and cortical localization within the cytoplasm; early endosome undergoes a maturation process during which internal pH decreases because of the activation of membrane proton pumps, converting into a late endosome, spherical and closer to the nucleus. Late endosome evolves into a multivesicular body (MVB) when its membrane, through inward invaginations, originates intraluminal vesicles (ILVs). MVB fate is variable: indeed, MVBs can: i) fuse together with lysosomes, causing the degradation of their cargo by lysosome hydrolytic enzymes; ii) move toward plasma membrane, fuse with it and release ILVs in the extracellular space; secreted ILVs are called exosomes (139) (Figure 2.21)

![Figure 2.21](image1.png)

**Figure 2.21:** exosome biogenesis starts with an endocytosis event and the formation of an early endosome; after a maturation process, it becomes a late endosome, which converts into a multivesicular body after inward invaginations of its membrane, leading to the formation of intraluminal vesicles. Once secreted through MVBs and plasma membrane fusion, ILVs are called exosomes (from Bellingham et al., Front. Physiol. 2012).

Exosome biogenesis regulation is still under investigation. A calcium-mediated mechanism modulating vesicles secretion has been proposed; calcium ions control similar events, such as neurotransmitter release through synaptic vesicle secretion, or lysosome fusion with the plasma
membrane; in particular, exosome secretion is induced by increased cytoplasmic levels of calcium ions in haematopoietic cells and oligodendrocytes (140, 141, 142). Similarly, plasma membrane depolarization due to the increase of potassium ions promotes exosome release in cortical neurons (143). Moreover, exosome biogenesis is impaired by inhibition of sphingomyelinases, ubiquitous enzymes controlling ceramide production through hydrolysis of sphingomyelin molecules included in plasma membranes: this evidence suggests that ceramide, of which exosome membranes are enriched, can act as exosome biogenesis regulator (144). Rab protein family includes small GTPases involved in trafficking and fusion of intracellular membranes; many members of this family (in particular Rab27a and Rab27b) regulate exosome biogenesis, even if it’s not clear if different members act in different cell types or at different stages of the process (145, 146, 147, 148, 149). Other proteins involved in exosome production are SNAREs (soluble N-ethylmaleimide-sensitive factor attachment protein receptors), a family of membrane fusion regulators; SNAREs modulate MVBs fusion with the plasma membrane. Vesicular SNAREs (v-SNAREs), included within MVB membrane, interact with target SNAREs (t-SNARE), located on the cytoplasmic side of the plasma membrane; this interaction causes the formation of a membrane-bridging SNARE complex, which controls the fusion of the membranes (150, 151) (Figure 2.22).

**Figure 2.22:** membrane fusion mediated by SNARE proteins: vesicular SNAREs interact with target SNAREs, moving the two membranes closer and triggering their fusion (from David Tareste, Biophysics of membrane fusion).

Also the exosome cargo is still under investigation; ExoCarta database (http://www.exocarta.org/) lists all molecules which have been indentified to date within exosomes
secreted by several cell histotypes. However, several studies underline a specific cargo including lipids, proteins and RNAs:

i. Exosomal membranes are particularly enriched in cholesterol, ceramide, sphingomyelin, gangliosides and polyunsaturated fatty acids, while phosphatidylcholine and diacylglycerol levels are lower compared to donor cell membranes (152, 153); moreover, compared to cell membranes, exosomal membranes contain more phosphatidylserine on outer side, which could promote their internalization by recipient cells (154). In its entirety, exosomal membrane is more rigid than cell membrane, and this rigidity could be pH-dependent. This hypothesis isn’t surprising considering exosome biogenesis, since MVB pH is lower than cytosolic one; plus, it has been observed that exosome uptake increases in acid microenvironment, such as the cancer niche, where decreased pH makes exosome membrane more fluid, promoting its fusion with recipient cells (153, 155). In addition, exosome membrane rigidity confers on them greater resistance to degradation, making exosomes excellent vehicles for several different molecules (156, 157). Exosomes also contain lipids acting as signal mediators, such as prostaglandins, and enzymes essential for their production, among which phospholipases A and D (158). It has been demonstrated that, overall, exosomes present an 8.4-fold enrichment of lipids per mg of protein (152).

ii. Proteins identified within exosomes always come from donor cell cytoplasm, while organelle characteristic proteins have never been found within exosomes (159, 160); in particular, exosomes transport heat shock proteins, such as Hsc73 and Hsc90 chaperons, proteins constituting ESCRT (endosomal sorting complex required for transport) complex, such as Tsg101 (tumor susceptibility gene 101) and Alix; tetraspanins, among which CD63 and CD81; cell cycle regulators, such as
phosphatidylinositol-3-kinase (PI3K), mitogen-activated protein kinases (MAPKs) and ErbB family members; integrins; cytoskeleton proteins, such as actin, myosin and tubulins; metabolism enzymes, among which glyceraldehyde 3-phosphate dehydrogenase (GAPDH); proteins involved in signal transduction, such as beta-catenin (120, 161, 162, 163). Among all these proteins, tetraspanins are particularly abundant in exosomes, so they are considered exosome specific markers.

iii. Exosomes are enriched in RNA molecules long less than 200 nucleotides, typically mRNAs and miRNAs, while they lack ribosomal RNAs (164); recently, lncRNAs and circRNAs have been identified within exosomes (165, 166, 167, 168).

It has been hypothesized that exosome cargo is casually determined during ILV formation, because late endosome membrane invaginations could lead to random incorporation of surrounding portions of the cytoplasm (139); recently, a second theory is emerging, according to which cells strictly control molecular trafficking toward exosomes. This theory is supported by several studies demonstrating that exosome cargo depends on donor cell histotype and physiological state (169, 170, 171, 172, 173). Nevertheless, it is not clear yet how cells recognize those molecules that have to be directed toward exosomes: protein sequence or structure motifs (e.g., acylation sites or phospholipid-binding domains) and membrane myristoylation and palmitoylation sites which may mediate protein recognition have been identified (174); also RNA molecules seem to be identified through specific sequence motifs discriminating cellular and exosomal miRNAs; these consensus sequences seem to be recognized by ribonucleoproteins, such as hnRNPA2B1 (heterogeneous nuclear ribonucleoprotein A2B1) (173). ESCRT machinery involvement in exosomal cargo sorting seems to be certain. ESCRT consists of four multiprotein complexes, called ESCRT-0, ESCRT-I, ESCRT-II and ESCRT-III, usually recruited on the cytoplasmic side of endosome membrane to sort specific proteins within ILVs. This process requires ubiquitination of target proteins, recognized by Tsg101, included in ESCRT-I; the interaction between Tsg101 and target proteins activates
ESCRT-II: ESCRT-I and ESCRT-II together regulate the inward invaginations of MVB membrane, allowing cytoplasmic mRNAs, miRNAs and proteins to enter into the nascent ILVs. ESCRT-III is recruited by ESCRT-II to cleave the nascent ILVs into free vesicles within the MVB. All ESCRT subunits and free ubiquitin molecules are released within the cytoplasm to be recycled (Figure 2.23). However, ESCRT-independent mechanisms regulating exosomal cargo sorting have been proposed (175).

![Figure 2.23](image_url)

**Figure 2.23:** ESCRT complex involvement in exosomal cargo sorting: the ESCRT-0 complex recognizes ubiquitinated target proteins in the cytoplasm and binds the ESCRT-I complex, which in turn recruits ESCRT-II subunits. ESCRT-I and ESCRT-II regulates the inward invaginations of MVB membrane, allowing cytoplasmic mRNAs, miRNAs and proteins to enter into the nascent ILVs. ESCRT-II recruits ESCRT-III subunits, which cleave the nascent ILVs into free vesicles. All ESCRT subunits and ubiquitin molecules are released in the cytoplasm to be recycled (from Robbins and Morelli, Nat Rev Immunol. 2014).

### 2.4.1 Exosomes as mediators of intercellular communication

Despite exosomes were initially considered cell “waste bins” secreted to remove dangerous or useless compounds, it is by now clear that their main function is to mediate intercellular communication. Once secreted, exosomes interact with recipient cells influencing their physiology.
Exosomes can act on the same donor cell in autocrine manner, on recipient cells near to donor cell in paracrine manner, or they can reach distant target cells through bloodstream and act in endocrine manner (176, 177, 178). The exact molecular mechanism of this interaction is still under investigation, even if four models have been proposed: i) a protein of exosome membrane interact with a surface receptor of the recipient cell, acting as a ligand (Figure 2.24b); ii) exosome membrane fuses with target cell membrane, transferring in its cytoplasm their entire cargo (Figure 2.24c); iii) a protein of exosome membrane undergoes a proteolytic cleavage in the extracellular space, generating a soluble ligand that interacts with its receptor on the surface of the target cell (Figure 2.24d); iv) a recipient cell with phagocytic activity absorbs the entire vesicle through phagocytosis (Figure 2.24e) (155, 179, 180, 181).

Figure 2.24: proposed models explaining the molecular mechanism of interactions between exosomes and recipient cells. A) exosomes are released by donor cells; B) a protein of exosome membrane interacts with a surface receptor on recipient cell membrane; C) exosome membrane fuses with target cell membrane; D) through a proteolytic cleavage in the extracellular space, a protein of exosome membrane generates a soluble ligand, which binds its receptor on recipient cell surface; E) a phagocytic cell absorbs the entire exosome through phagocytosis (from Mathivanan, J Biotechnol Biomater. 2012).
Whatever the mechanism, the interaction between exosomes and recipient cells induces a response modulating cell phenotype, because of a signal transduction event or the entrance of specific molecules carried by vesicles. Once entered into target cells, exosome mRNAs are actively translated into functional proteins (164); in particular, exosomes secreted by cells cultured in oxidative stress conditions, induced by hydrogen peroxide administration, confer on recipient cells higher resistance to the same treatment through RNA molecules transfer (169).

Among the first exosome-secreting cells identified there are several immune system cells; indeed, exosomes contribute to immune response modulation both mediating intercellular communication and acting as antigen-presenting vesicles thanks to their enrichment in class II MHC (major histocompatibility complex) molecules (124). Particularly, MHC-antigen complexes on exosome surface can directly interact with T-cell receptor; otherwise, the whole microvesicle can be captured by antigen-presenting cells (APCs) and indirectly activate T lymphocytes (182, 183, 184, 185); in addition, exosomes secreted by mature dendritic cells are more efficient in the activation of T lymphocytes than exosomes secreted by immature dendritic cells, suggesting that donor cell type influences exosome function (184, 186, 187).

As suggested by the presence of several molecules involved in signal transduction pathways within exosomes, these microvesicles mediate the transfer of pro-proliferative stimuli. First proteomic studies showed that beta-catenin, the main cytosolic regulator of Wnt pathway, is contained into exosomes; in addition, beta-catenin packaging into exosomes increases in presence of CD9 and CD82, two tetraspanins inhibiting Wnt pathway, and cytoplasmic protein levels decrease through its expulsion via exosome secretion (188).

Recently, the role played by exosomes in nervous system physiology has attracted increasing attention, since exosome secretion has been observed in neurons, Schwann cells, oligodendrocytes, astrocytes and microglia (133, 142, 143, 189). Exosomes produced by oligodendrocytes transfer enzymes, such as catalase and superoxide dismutase I, to neurons, helping
them fighting oxidative stress (190). Moreover, exosomes regulate synaptic plasticity by transporting AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) glutamatergic receptor, thus modulating their number on post-synaptic membrane (189).

### 2.5 Colorectal cancer

Colorectal cancer (CRC) is one of the most common diseases in industrial society. About 875,000 new cases are reported every year, among which about 500,000 are fatal. In Western society, CRC represents the second leading cause of cancer death and the third most frequent cancer type (191).

CRC has the highest incidence rate among tumors in the Italian population: among men it represents 14% of all new tumor cases, being preceded only by prostate and lung cancer, while among women it represents 13% of new cases, being preceded only by breast cancer. The causes of CRC increased incidence observed in the last decades are population aging and higher spread of risk factors, such as the consumption of red meat, sausages, refined flours and sugars, obesity and reduced physical exercise, smoking and alcohol excess; however, prevention campaigns led in recent years to reduced incidence and mortality (192). Average onset age is 62 years, with an increased risk after 40 for both men and women (191). Prognosis is generally favorable and tends to get better, being 5-years survival rate increased from 50% in early 90s to 64% in men in 2006-2007, and from 51% to 63% in women (192).

The disease begins as a benign adenomatous polyp, which successively evolves into advanced adenoma with high-grade dysplasia and finally in invasive cancer. TNM (tumor-node-metastasis) classification allows to classify tumor stage according on three parameters: i) tumor extension (T), ii) lymph node involvement (N), iii) distant metastasis (M). Invasive cancers confined within colon wall (TNM stage I and II) are curable, but if not treated they spread to regional lymph nodes (TNM stage III) and finally cause distant metastasis formation (TNM stage
IV). Stage I and II cancers are curable through surgical excision, and up to 73% of stage III tumors are curable by surgical excision associated with chemotherapy. Even if recent progresses in chemotherapy led to increased survival, stage IV cancers are usually incurable (191, 193).

Clearly, CRC prognosis depends largely on pathological stage at diagnosis and is influenced by several factors, among which the most important is tumor stage; bowel wall invasion degree, lymph node and distant metastases negatively influence prognosis (191).

2.5.1 CRC genetic alterations

CRC can be classified into two different forms: i) sporadic CRC is more common, representing about 75-80% of all cases; ii) familial (hereditary) CRC is rarer, even if it has been observed that a high percentage of CRC patients has relatives affected, with a familial degree comprised between the first and the third; this finding suggests that even in sporadic cancers familiarity leads to a higher risk of disease onset. According to genetic alterations involved in carcinogenesis, sporadic CRC can be further classified into two subtypes, associated with two main pathways involved in CRC: i) the “suppressor” or “canonical” pathway, which causes chromosomal instability (CIN); ii) the “mutator” pathway, associated with microsatellite instability (MSI) (194).

It is believed that the canonical pathway (about 80-85% of sporadic CRC cases) follows the sequence of molecular events described in Fearon and Vogelstein model (195), according to which specific genetic events correlate with tissue morphology during CRC carcinogenesis (Figure 2.25): the first genetic alteration is represented by tumor suppressor APC gene mutation during the adenoma stage, followed by KRAS and p53 mutations and chromosome 18q deletions. This model, although considered linear and simplistic, shows that mutations of tumor suppressor genes, such as APC and p53, and of oncogenes, such as KRAS, are fundamental events in the canonical pathway. Cancers characterized by CIN also present allelic imbalance, with frequent allelic loss on chromosomes 5q, 8p, 17p and 18q, and chromosomal duplications and translocations (194).
Introduction

Figure 2.25: molecular alterations during CRC carcinogenesis; the upper part shows mutations associated with the “suppressor” pathway, the bottom those associated with the “mutator” pathway (from Walther et al., Nat Rev Cancer. 2009).

On the contrary, the “mutator” pathway is rarer, representing 15-20% of sporadic CRC cases. MSI cancers are characterized by high microsatellite instability (MSI-H), that is the accumulation of mutations in microsatellites, short non-coding DNA sequences repeated in tandem throughout the genome. This accumulation of mutations is due to other mutations occurring in genes coding for the mismatch repair (MMR) system, which normally corrects mistakes made by DNA polymerase during DNA replication; mutation-induced MMR system inactivation causes a further accumulation of mutations in microsatellite and in the entire genome. MSI-H sporadic cancers generally don’t show cytogenetic abnormalities or aneuploidy; plus, they present rare or no mutations typical of “suppressor” pathway, while other mutations have been described in microsatellite sequences of genes involved in carcinogenesis, such as TGFβRII, BAX, MSH3, MSH6, caspase 5, APC, beta-catenin. In addition, most of MSI-H sporadic cancers presents the CpG island methylator phenotype (CIMP), characterized by DNA aberrant methylation which causes silencing of functional genes (194).
At molecular level, it is by now known that some signal transduction pathways play a crucial role in CRC neoplastic transformation: among them there are Wnt, MAPK, PI3K and p53 pathways.

The first event in CRC pathogenesis is represented by the activation of Wnt signaling pathway, which regulates proliferation, differentiation and motility of intestinal epithelial cells. In physiological conditions, the APC (adenomatous polyposis coli) protein interacts with Axin in the cytoplasm; the APC-Axin complex recruits the serine/threonine protein kinase GSK3β (glicogen synthase kinase 3β) and the kinase CK-1 (casein kinase 1), forming the beta-catenin “destruction complex”; this complex binds and phosphorylates cytoplasmic beta-catenin, earmarking it for the ubiquitin-proteasome-mediated degradation (Figure 2.26). Beta-catenin plays a dual role within cells: it acts as transcriptional factor and as cellular adhesion stabilizer. In the cytoplasm, beta-catenin binds E-cadherin, a transmembrane glycoprotein involved in formation and maintenance of adherens junctions between epithelial cells. In late carcinogenesis, E-cadherin loses its function; consequently, adherens junctions are disrupted and cell motility and migration increase. Normally, E-cadherin and APC compete for beta-catenin binding, controlling its cytoplasmic levels by inducing its degradation; thus, beta-catenin doesn’t accumulate within the cytoplasm and is not imported to the nucleus. On the contrary, when APC is mutated and non-functional, E-cadherin binds beta-catenin until saturation, inducing a cytoplasmic accumulation of beta-catenin; therefore, it is imported to the nucleus to perform its function of transcriptional co-factor, inducing the expression of genes regulating proliferation, differentiation and cell migration, such as c-Myc, MMP-7 (matrix metalloproteinase-7), and cyclin D1. Since APC is a key regulator of Wnt pathway, it is believed that its mutation with loss of function is the initial event in CRC pathogenesis: mutated APC has been observed in about 80% of CRC patients, while beta-catenin mutations are rarer, being reported in about 5% of cases. APC mutations can be due to hypermethylation or
deregulation of APC gene; also, they can be inherited, as in FAP (familial adenomatous polyposis), an inherited syndrome causing predisposition to cancer (196).

Another pathway classically associated with CRC is MAPK pathway, involved in both sporadic CRC subtypes: KRAS (KRAS proto-oncogene, GTPase) mutations have been observed in 30-35% of patients and are typically associated with chromosomal instability, while BRAF (B-Raf proto-oncogene, serine/threonine kinase) mutations are less frequent (9-11%) and are often associated with microsatellite instability (197). KRAS is a small monomeric G-protein involved in the transduction of mitogenic signals activated by growth factor receptors located on cell membranes. KRAS activation induces a sequence of downstream phosphorylations performed by several kinases, such as MAPK1/ERK (mitogen-activated protein kinase 1) and PI3K/AKT. In physiological conditions, the growth factor binds its receptor tyrosine kinase (RTK), causing its dimerization and autophosphorylation; the phosphorylated dimeric RTK, thus activated, recruits in the cortical cytoplasm GRB2/SOS (growth factor receptor/bound protein2/son of sevenless), which

Figure 2.26: schematic representation of Wnt signaling pathway (from Roper and Hung, Molecular Pathogenesis of Colorectal Cancer, Chapter 2. Springer New York, 2013).
in turn interacts with KRAS, activating it. Through GDP-GTP exchange, KRAS changes conformation and binds effector molecules, inducing the downstream phosphorylation sequence ending with ERK phosphorylation and translocation to the nucleus; here, p-ERK phosphorylates and activates transcription factors such as c-FOS and c-JUN, inducing the expression of genes regulating proliferation, differentiation, apoptosis (Figure 2.27). After GTP hydrolysis, KRAS is inactivated. All the different KRAS point mutations to date identified cause the loss of hydrolytic function, resulting in a constitutively active protein which induces an uncontrolled cell proliferation. KRAS mutations also alter Wnt pathway: KRAS activation induces an increased expression of beta-catenin, which accumulates in the cytoplasm; KRAS also phosphorylates and activates PI3K, causing GSK3β phosphorylation and inducing beta-catenin nuclear translocation. Hence, in physiological conditions, GSK3β acts as negative regulator of two fundamental pathways involved in CRC pathogenesis (196).

Figure 2.27: schematic representation of KRAS and downstream MAPK signaling pathway (from Kim and Bar-Sagi, Nat Rev Mol Cell Biol. 2004).
PI3K/AKT pathway also plays a crucial role in CRC carcinogenesis, inducing malignant transformation, cell proliferation, migration and survival, and promoting tumor progression. PI3K (phosphoinositide 3-kinase), belonging to a lipid kinase family, is made up of a regulatory subunit (p85) and a catalytic subunit (p110); there are three variants of p85 subunit (p85α, p85β and p85γ, coded respectively by PIK3R1, PIK3R2 and PIK3R3 genes) and three variants of p110 subunit (p110α, p110β and p110γ coded respectively by PIK3CA, PIK3CB and PIK3CD genes). Mutations inducing aberrant pathway activation have been observed in 30-35% of CRC patients, most frequently in PIK3CA and PIK3CB genes (196). Normally, PI3K is activated by external stimuli causing RTK activation, among which also KRAS-induced stimuli; p85 subunit recognizes phosphorylated tyrosine residues in the receptor intracellular domain, binding them and releasing p110 subunit, thus activated; p110 subunit phosphorylates phosphatidylinositol 4,5 biphosphate (PIP2) to phosphatidylinositol 3,4,5 triphosphate (PIP3), inducing PIP3 accumulation at cell membrane. PIP3 recruits AKT (AKT serine/threonine kinase 1), which is phosphorylated by PDK1 (pyruvate dehydrogenase kinase 1) and mTORC2 complex, including among others MTOR (mechanistic target of rapamycin). AKT promotes growth and survival in several ways: i) inhibition of pro-apoptotic proteins belonging to Bcl-2 family; ii) increased cytoplasmic availability of MDM2 (MDM2 proto-oncogene) and subsequent increase of p53 degradation; iii) induced expression of anti-apoptotic genes through NF-kB (nuclear factor kappa B) activation. PI3K pathway is inhibited by PTEN (phosphatase and tensin homologue), a tumor suppressor phosphatase which dephosphorylates PIP3 to PIP2, inhibiting AKT activation; therefore, PTEN is associated with inhibition of cell cycle progression and promotion of cell death and angiogenesis. CRC cases with PTEN inactivation, due to truncating mutations or promoter methylation-induced silencing, have been reported (198).
In late CRC carcinogenesis, p53 mutation has been reported in 50% of patients; p53 is a tumor suppressor protein which acts as transcription factor, thus stopping cell cycle and activating apoptosis in presence of DNA damage; for this reason, p53 has been defined the “genome guardian”. In physiological conditions, low p53 protein levels are maintained through ubiquitin-proteasome degradation: MDM2, an ubiquitin-ligase enzyme, binds p53 and causes its degradation. Interestingly, MDM2 itself is a target of p53, which activates its expression; this mechanism represents a negative feedback regulation of p53 expression and function. After DNA damage, p53 phosphorylation within the MDM2-binding domain impedes p53-MDM2 interaction; thus, p53 is more stable, translocates to the nucleus and binds DNA, activating the transcription of several genes, including p21 (cyclin-dependent kinase inhibitor 1A, CDKN1A) and the pro-apoptotic factor BAX (BCL2-associated X protein); in this way, p53 arrests cell cycle in G1 phase, allowing DNA repair, or induces apoptosis if DNA damage cannot be repaired. About 80% of p53 mutations are missense and cause a stable but non-functional protein, which accumulates in the cytoplasm; this cytoplasmic accumulation of p53 is strongly correlated to metastatization and poor prognosis (196, 197).

Finally, epigenetic alterations also participate to tumor progression and have been associated with CRC; they include DNA methylation, histone modifications and chromatin remodeling, reversible alterations modulating gene expression without modifications in DNA sequence. DNA methylation is mediated by DNA methyltransferases; in eukaryotic genomes, it affects prevalently cytosines of CpG islands (CpG dinucleotides enriched regions), located within the promoter of the 60% of human genes; typically, methylation of promoter CpG islands provokes silencing, while methylation of CpG islands external to promoter fosters transcriptional activation. Carcinogenesis induces remarkable modification in the genome methylation pattern, with demethylation and hypermethylation of many promoters: among hypermethylated genes in CRC there are genes involved in Wnt pathway (APC and AXIN2), DNA repair (MLH1 and MLH2), cell cycle regulation (CDKN2A), and KRAS pathway (RASSF1A and RASSF1B). Histone modifications consist of
addition or removal of functional groups, such as acetyl groups (acetylation or deacetylation) or methyl groups (methylation or demethylation), on specific amino acid residues of N-terminal tails of histones, proteins around which DNA is wrapped within the nucleus. Specific histone modifications are associated with silencing or transcription activation: deacetylation, catalyzed by histone deacetylases (HDACs), induces transcription repression, while acetylation, catalyzed by histone acetyltransferases (HATs), promotes gene expression. The upregulation of many HDAC, including HDAC1-3, HDAC5 and HDAC7, has been reported in CRC, together with the downregulation of genes involved in Wnt pathway. Moreover, HDAC2 upregulation observed in early carcinogenesis is associated with H4K12 and H3K18 histone ipoacetylation during the adenoma-carcinoma sequence: this evidence suggests that increased HDAC2 expression and subsequent ipoacetylation participate in CRC progression (199).

2.5.2 CRC diagnosis and therapy

It has been established that favorable prognosis of cancer patients depends on many factor, including early diagnosis; for this reason, several screening projects are implemented every year, aiming to diagnose early stage tumors, before symptom onset. Screening tests are usually minimally invasive and easy to perform; only patients positive to these tests undergo more deepened and often more invasive analyses.

Many epidemiological studies showed that environmental factors, such as diet and lifestyle, are associated with CRC pathogenesis: red meat high consumption, reduced physical exercise, obesity, alcohol and smoke increase the risk of tumor. It is believed that in most cases colorectal cancer begins as adenoma, a benign lesion that can evolve into a tumor; this process can require a long time, even ten years, making CRC a good candidate for screening and prevention. In addition, adenoma resection can prevent tumor development, reducing incidence and mortality. To date, two
principal screening strategies are used: faecal occult blood testing (FOBT) and endoscopy (flexible sigmoidoscopy and colonoscopy).

Tumors bleed, pouring into the intestinal lumen small amounts of blood that can be detected in stool. Therefore, FOBT represents a non-invasive and cheap test, easy to perform at home by the patient himself. This test has to be repeated every one or two years to increase its sensitivity for cancer. However, this test is characterized by low sensitivity and specificity, giving false positive results when bleeding is caused by other pathologies, such as ulcers or inflammatory diseases of the intestinal tract.

Flexible sigmoidoscopy (FS) allows to inspect the mucosa, perform tissue biopsies and remove polyps in the distal part of the colon, but not in the region proximal to sigmoid colon because of the endoscope limited length. FS is a high sensitivity and specificity test both for adenomas and carcinomas within the inspected area.

On the contrary, colonoscopy allows to inspect the entire colon mucosa, perform tissue biopsies and remove polyps in the whole colon-rectum in one single session; therefore, it is the ideal test for CRC prevention and diagnosis. Colonoscopy sensitivity and specificity to identify advanced adenomas and tumors are about 100%, and it is indeed the last test performed when all other screening tests gave positive results. For these reasons, colonoscopy is considered the gold standard for colonic disease diagnosis, but a great variability depending on endoscopists has been observed. Plus, it is an invasive test, requiring a long time, expensive and painful for the patient. Serious complications, such as bleeding or perforation, even if rare, occur with higher frequency compared to flexible sigmoidoscopy (200, 201).

Thus, diagnosis tests used today are invasive or have low sensitivity. The identification of early genetic and epigenetic markers (in stool and blood) specific for adenomas or invasive tumors is still under investigation. It is known that mucosal cells physiologically shed from gastrointestinal epithelium, allowing to detect tumor DNA and epigenetic modifications in stool samples, which can
be used for diagnostic and prognostic analyses. Similarly, markers can be detected in blood: to date, CRC-associated markers include KRAS, APC, p53 and methylation markers, such as vimentin and septin-9; however, these markers are currently used only in clinical trials, not in CRC screening (201).

The first and most important therapeutic approach for colorectal cancer is represented by surgical resection of the affected colon tract; relapse probability is correlated with several factors, including depth of tumor penetration through the bowel wall and lymph node involvement. Surgical resection efficacy is high for stage I and II tumors (90% and 75% respectively); stage III tumors with lymph node involvement have 60% of recurrence (202).

Many therapeutic strategies are today used in CRC treatment; some of them aim to impair DNA replication and thus cell proliferation. A drug commonly administered in CRC therapy is 5-fluorouracil, a fluorinated pyrimidine which inhibits thymidylate synthase, the rate-limiting enzyme of pyrimidine nucleotide synthesis. Fluorouracil is usually administered with leucovorin (or folinic acid), a reduced folate stabilising the drug-enzyme interaction and increasing the inhibition of DNA replication: simultaneous administration of both molecules induces 50% tumor size reduction in 20% of patients and increases survival from 6 to 11 months. Side effects of fluorouracil depends on administration method (intravenous or bolus); they include neutropenia (reduced number of neutrophils), stomatitis and diarrhea. Irinotecan is a semisynthetic derivative of natural alkaloid camptothecin, extracted from the bark of *Camptotheca acuminata*; this drug exerts cytotoxic effect by interacting with topoisomerase I, the enzyme that uncoils DNA through single-stranded breaks during replication. These DNA breaks are normally repaired, but camptothecin stabilises them causing DNA fragmentation and cell death. Irinotecan is administered as a prodrug, which is successively hydrolysed to its active metabolite SN-38 by hepatic carboxylesterases. Side effects include diarrhea, myelotoxicity (bone marrow suppression), nausea, vomiting and alopecia. Oxaliplatin is a third-generation derivative of platinum which forms big DNA adducts, thus
inducing apoptosis. Oxaliplatin and fluorouracil act synergistically: since the former causes thymidylate synthase downregulation, it increases fluorouracil efficacy. Oxaliplatin toxicity provokes renal dysfunctions, alopecia, ototoxic effects and neuropathy (203).

Another common therapeutic approach is based on molecules specifically targeting the epidermal growth factor receptor (EGFR), which acts upstream of several signal transduction pathways associated with CRC, such as Wnt, MAPK, and PI3K. EGFR is a 170-kDa transmembrane glycoprotein characterized by an extracellular domain, containing the ligand-binding site and the sequences involved in dimerization, and an intracellular domain with tyrosine kinase activity. EGFR upregulation has been observed in many cancer types, including CRC; therefore, therapies based on inhibitors of the intracellular domain tyrosine kinase activity, or on monoclonal antibodies targeting the EGFR extracellular domain have been created. The higher the expression of EGFR, the better the response to the anti-EGFR therapy. Tyrosine kinase inhibitors include gefitinib and erlotinib, which reversibly inhibits EGFR autophosphorylation; among the best known monoclonal antibodies there are cetuximab, an IgG1 recombinant human/mouse chimeric antibody, and panitumumab, an IgG2 fully human antibody; cetuximab and panitumumab are the only monoclonal antibodies used in clinical practise. Both in normal and tumor cells, cetuximab binds the extracellular domain of EGFR, causing the internalization of the receptor and thus impeding receptor-ligand interaction; panitumumab also interacts with the EGFR extracellular domain in both cell types, thus inhibiting ligand binding, receptor dimerization, autophosphorylation and signal transduction (204). Monoclonal antibody side effects are mild, including acne-like rash and skin drying and fissuring among the most common (203). The anti-EGFR therapy is useless in patients bearing KRAS mutations: in fact, mutated KRAS is constitutively active, triggering MAPK pathway downstream of EGFR even in absence of EGFR ligands; therefore, KRAS mutations confer resistance to anti-EGFR therapies upon tumor cells
(205). For this reason, it is necessary to perform KRAS genotyping on surgically removed tumors before administering these drugs to the patient.

To promote its progression, tumor induces angiogenesis, that is the formation of new blood vessels in order to guarantee itself a constant supply of nutrients; given the importance of this phenomenon, several angiogenesis inhibitors are under investigation for therapeutic applications. One of the most important angiogenic factors is VEGF (vascular endothelial growth factor), associated with cellular proliferation and blood vessel maturation; it has been observed that VEGF expression correlates not only with angiogenesis, but also with tumor progression. Several studies demonstrated that anti-VEGF monoclonal antibodies associated with chemotherapy increase patient survival (206). Bevacizumab is an anti-VEGF monoclonal antibody; its action combined with chemotherapeutic agents has been evaluated in several clinical trials in patients affected by advanced CRC: results showed a considerable improvement in patient response to therapy and an increase in overall survival of 4-7 months; drug administration has been quite well tolerated and caused mild toxic effects, including reversible hypertension and proteinuria (203).

The possible use of vitamin A and its retinoid derivatives in therapeutic approaches is attracting growing interest; currently, retinoid are used in the treatment of epithelial tumors because they regulate several pathways involved in carcinogenesis, including those typically associated with CRC. Retinoic acid receptors (RARs), located within the nucleus, bind all-trans retinoid acid (ATRA), forming a complex which can bind DNA and promote transcription of genes regulating cell differentiation. Retinoid and vitamin A administration, even through nutrition, increases ATRA levels, reducing proliferation and increasing apoptosis; moreover, retinoids induce decreased levels of cyclin D1 and matrix metalloproteases (MMPs), enzymes involved in invasion, migration and metastasis (196).
2.6 NcRNAs as mediators of tumor progression

Since ncRNA discovery, a lot of scientific papers have been published about their functions; it was immediately clear that ncRNAs regulate a wide range of fundamental biological processes, such as cell proliferation, differentiation, migration, angiogenesis and apoptosis; therefore, ncRNAs play a crucial role in cell physiology. Since their fundamental role, the association between ncRNA deregulation and several diseases, including tumors, is not surprising: it has been shown that ncRNAs act as oncogenes or tumor suppressor genes, contributing to tumor onset and progression.

2.6.1 LncRNA role in CRC pathogenesis

Despite their recent discovery, in the last decade many studies proved lncRNA involvement in carcinogenesis, and in particular in CRC, through several molecular mechanisms: lncRNAs modulate numerous signal transduction pathways classically associated with CRC; moreover, they interact with other molecules, such as DNA, RNA and proteins. It has been demonstrated that expression levels of several lncRNAs are altered in CRC; however, to date the molecular mechanism of only a few lncRNAs has been understood.

One of most studied cancer-associated lncRNAs is MALAT1, upregulated in metastases of many tumors, including non-small cell lung cancer, hepatocellular carcinoma, and endometrial stromal sarcoma (207, 208, 209). Several MALAT1 functions have been hypothesized, including the regulation of pre-mRNA processing because of the association between the lncRNA and SC35 splicing domains within the nucleus (210); MALAT1 accumulates in the nucleolus, where it acts as “riboregulator” modulating the expression of target genes, including numerous tumor suppressors (72, 211). In addition, MALAT1 controls the activity of E2F1 (E2F transcription factor 1), a key transcription factor regulating cell cycle progression: MALAT1 interacts with non-methylated Polycomb 2 protein (Pc2), thus activating interchromatin granule and promoting E2F1 sumoylation and activation (212). In vitro silencing experiments in bladder cancer cells demonstrated the
involvement of MALAT1 in cell migration through EMT regulation (213); in addition, MALAT1 depletion causes aberrant mitosis, during which numerous cells accumulate at G2/M boundary, and cell death (214). Finally, the tRNA-like 61-nucleotide fragment produced by MALAT1 3’-end, called mascRNA, induces proliferation and invasion in CRC cells (215).

The lncRNA BANCR (BRAF-activated non-protein coding RNA) seems to be associated with the BRAF V600E mutation, one of the most frequent mutations observed in several tumors, including CRC (216). BANCR upregulation, frequently observed in CRC tissues, correlates with lymph node metastasis and tumor stage; moreover, BANCR modulates migration and EMT in CRC cell lines (217). LncRNA-ATB (long non-coding RNA-activated by TGF-β) is also upregulated in CRC and promotes tumor progression by inducing EMT: in particular, lncRNA-ATB high levels cause reduce the expression of E-cadherin, while increasing ZEB1 (zinc finger E-box binding homeobox 1), a mesenchymal cell marker (45).

Through genome wide association studies (GWASs) a region of approximately 2 Mb on human chromosome 8q24 has been associated with increased susceptibility to several diseases, including CRC (218); this region is located within a large protein-coding gene desert, but it includes many ncRNA genes, such as CCAT1, CCAT2, PCAT2, PRNCR1 (219). In particular, the rs6983267 SNP (single nucleotide polymorphism) has been associated with an increased risk of CRC (220); this SNP maps to a region containing DNA enhancer elements and affects the binding affinity of TCF7L2 (transcription factor 7 like 2), a transcription factor regulating the expression of Wnt pathway target genes (221). The rs6983267 SNP lies within the gene coding for CCAT2 (colon cancer associated transcript 2), affecting its expression: the allele G induces a higher CCAT2 expression, thus increasing CRC risk; CCAT2 itself interacts with TCF7L2, inducing the upregulation of MYC, a proto-oncogene involved in CRC, and overactivating Wnt pathway; also, CCAT2 is itself a Wnt downstream target, in a positive feedback loop regulation mechanism (222). The upregulation of CCAT1-L, the long isoform of CCAT1 (cancer associated transcript 1), is
associated with tumor stage and progression in CRC (223): CCAT1-L accumulates within the nucleus, where it interacts with a MYC enhancer (MYC-335), which in turn interacts with MYC promoter through chromatin looping, thus positively regulating MYC expression. Moreover, CCAT1-L knockdown reduces MYC mRNA levels, suggesting that this lncRNA regulates MYC expression in cis (224). Also PVT1 (plasmacytoma variant translocation 1) is associated with CRC and seems to be necessary to maintain high levels of MYC protein; PVT1 is upregulated in CRC because of a copy number amplification on chromosome 8q24; knockdown of PVT1 decreases proliferation and migration in vitro by reducing MYC tumorigenic activity (225). In addition, PVT1 is the precursor of several spliced ncRNAs, including a cluster of six miRNAs (miR-1204, miR-1205, miR-1206, miR-1207-5p, miR-1207-3p, miR-1208) involved in carcinogenesis (226). PVT1 is also a target of p53, which binds and activates a DNA canonical response element near to miR-1204 coding region, thus increasing miRNA expression; ectopic expression of miR-1204 increases p53 levels inducing cell death in CRC cell lines (227).

Another lncRNA widely studied in several tumors is HOTAIR, which regulates gene expression through PRC2-induced chromatin remodeling, thus promoting metastasis in breast cancer (228); HOTAIR and PRC2 levels are correlated also in CRC, suggesting an involvement of the lncRNA in maintaining the mesenchymal and undifferentiated status in CRC cells (229). CRNDE (colorectal neoplasia differentially expressed) also interacts with PRC2, thus modulating chromatin remodeling (60). CRNDE promoter is bound by pluripotency-related transcription factors, including MYC, and CRNDE knockdown induces decreased expression of many pluripotency markers, such as SOX2, KLF4, NANOG, and OCT4; indeed, CRNDE expression is highest during the early stages of mammals development, decreasing progressively thereafter (230).

The lncRNA TUSC7 is regulated by p53 and acts as tumor suppressor, impairing tumor growth both in vivo and in vitro. In particular, this function is due to TUSC7 fourth exon, which contains two binding sites for miR-211, involved in cell growth regulation: TUSC7 acts as miRNA
sponge, sequestering miR-211 and causing its downregulation. Moreover, TUSC7 is a miR-211 target, creating a negative feedback regulation mechanism (71). Similarly, FER1L4 (fer-1 like family member 4), a lncRNA downregulated in CRC, acts as tumor suppressor by sequestering miR-106a, a well known oncogene in colorectal cancer. Ectopic expression of FER1L4 decreases miR-106a levels and inhibits proliferation, invasion and migration in CRC cell lines (231).

H19 upregulation has been observed in several tumors, including CRC (232). H19 function is associated with miRNAs: it is the precursor of miR-675, which regulates cell cycle and EMT by targeting RB1 (69); on the other hand, H19 contains binding sites for let-7 miRNA family, key tumor suppressors in CRC (70).

A CRC-associated lncRNA recently discovered is UPAT (UHRF1 protein associated transcript), essential for surviving and tumorigenicity of CRC cell lines. UPAT interacts with UHRF1 (ubiquitin like with PHD and ring finger domains 1), an ubiquitin ligase regulating chromatin remodeling and DNA methylation through the degradation of DNA-methyltransferase 1 (DNMT1) and histons; the interaction with UPAT stabilizes UHRF1 and inhibits its degradation via ubiquitination. Moreover, UPAT interacts with the promoter of Sprouty 4 (SPRY), a key regulator of MAPK pathway, which expression is induced by high levels of the lncRNA (233). Among other lncRNA interacting with DNMT1, DACOR1 (DNMT1-associated colon cancer repressed lncRNA 1) is downregulated in CRC cell lines. Ectopic expression of DACOR1 induces a gain of DNA methylation, although it doesn’t alter DNMT1 protein levels: this evidence suggests that DACOR1 recruits DNMT1 on specific loci of the genome. In addition, increased expression of DACOR1 inhibits cell growth through TGF-β (transforming growth factor beta 1) pathway, a tumor suppressive pathway associated with CRC, and induces the downregulation of CBS (cystathionine β-synthase) and the subsequent increment of SAM (S-adenosyl methionine), essential donor of methyl groups for the activity of DNA methyltransferases (234).
2.6.2 CircRNAs and colorectal cancer

CircRNAs represent a class of quite unknown gene expression regulators; they compete with other RNAs for the interaction with miRNAs and RNA-binding proteins (RBPs), probably modulating their concentration (91). Reduced circRNA expression has been often observed in several tumors compared to normal tissues: this could be due to the absence of linear ends in circRNAs, which are resistant to exonucleases and tend to accumulate within cells with low proliferative rate (79).

Few data have been reported on circRNA involvement in colorectal cancer carcinogenesis. A recent study showed that circRNA expression levels are globally reduced in CRC tissues compared to normal mucosa; the negative correlation observed between circRNA expression and cell proliferation rate could be explained by several factors: i) an altered backsplicing machinery in tumor cells; ii) an increased circRNA degradation mediated by miRNAs deregulated in tumors; iii) circRNA “dilution” provoked by high proliferative rate (235). According to this evidence, the downregulation of hsa_circ_001988, circRNA transcribed from the FBXW7 (F-box and WD repeat domain containing 7) gene, has been recently reported in CRC tissues compared to normal adjacent tissues (236). In addition, it has been reported that circRNA expression is globally reduced in CRC cell lines bearing KRAS mutations compared to KRAS wild type CRC cells, while the expression of circRNA host genes is unvaried; authors hypothesized that reduced circRNA levels could be explained by their increased export toward exosomes (237).

In contrast, another study showed that hsa_circ_001569, circRNA transcribed from the ABCC1 (ATP binding cassette subfamily C member 1) gene, is upregulated in CRC tissues and positively correlates with aggressive characteristics, such as tumor stage, lymph node involvement, distant metastasis and poor differentiation. *In vitro* experiments demonstrated that hsa_circ_001569 promotes cell proliferation and migration and inhibits apoptosis, thus fostering tumor progression.
In particular, hsa_circ_001569 plays its oncogenic action by sequestering miR-145, well known tumor suppressor miRNA in CRC, thus inducing upregulation of miR-145 targets (238).

2.6.3 MiRNAs involved in CRC carcinogenesis

It is now clear that miRNAs play a crucial role in the regulation of cell physiology by interacting with their mRNA targets; it has been widely demonstrated that these ncRNAs regulate key processes in carcinogenesis, such as proliferation, invasion, migration, EMT and apoptosis. A lot of miRNA-coding genes are located at genomic regions frequently rearranged in tumors, such as fragile sites, minimal regions of loss of heterozygosity (LOH), minimal regions of amplification (minimal amplicons), or common breakpoint regions, strengthening miRNA role in carcinogenesis (239, 240).

The altered expression of several miRNAs has been reported in many different tumors, including colorectal cancer (15, 240). Depending on biological function, determined by the functions of mRNA targets, miRNAs can be classified as oncogenes (oncomiRs) or tumor suppressors (241, 242): i) oncomiRs are transcribed from genes localized in genomic regions which are amplified or upregulated in the tumor, and induce increased cell proliferation, angiogenesis, invasiveness, and reduced apoptosis; ii) tumor suppressor miRNAs are transcribed from genes located in regions which are deleted or silenced in the tumor, having an opposite biological effect. Several miRNAs acting as oncogenes or tumor suppressor have been associated with CRC: among the most known oncogenic miRNAs there are the miR-17-92 cluster, miR-21, miR-31, miR-106a, while among the tumor suppressor miRNAs there are the let-7 and miR-34 miRNA families, miR-143 and miR-145 (15).
2.7 Exosomes in tumor progression

The role of mediators of intercellular communication makes exosomes fundamental tools used by the tumor to spread pro-tumor stimuli. The tumor promotes its own progression by secreting several signals into the extracellular environment, thus building around itself a niche suited to its development and survival; recently it has been demonstrated that many of these signals are delivered by exosomes. Exosome involvement in carcinogenesis is supported by numerous studies reporting a greater amount of circulating vesicles in patients affected by several tumors compared with healthy individuals (243, 244, 245).

Exosomes secreted by tumor cells promote tumor progression in many ways; one of the most important mechanisms is certainly represented by the suppression of the immune response. Tumor-secreted exosomes can interact with immune system cells causing their death through several molecules, such as FasL (Fas ligand) and TRAIL (TNF-related apoptoisi-inducing ligand) expressed on vesicle membranes, which bind their receptor on CD8^+ T lymphocytes (246, 247); in addition, the direct fusion of exosomes secreted by skin melanoma and CRC cells with monocytes inhibits monocyte differentiation into dendritic cells, instead inducing the secretion of cytokines causing T lymphocyte death (248).

Exosomes can also foster tumor progression by spreading pro-proliferative, pro-metastatic and pro-angiogenic signals. Exosomes express on their surface growth factor receptors, such as HER (human epidermal growth factor receptor) receptor family, and carry their ligands, which can be transferred to recipient cells and promote proliferation (249). Exosomes secreted by breast cancer and CRC cells express on their surface amphiregulin, an EGFR ligand, which interacts with its receptor on the plasma membrane of tumor cells, thus increasing their invasiveness (250). Finally, EGFR molecules contained within tumor-derived exosomes can be transferred to endothelial cells, inducing VEGF expression and activating angiogenesis (251).
In addition, exosomes can confer upon tumor cells resistance to immunotherapy treatments. Standard treatment for HER2-overexpressing breast carcinoma is represented by a monoclonal antibody called trastuzumab: the drug binds the extracellular domain of HER2/ERBB2 (erb-b2 receptor tyrosine kinase 2) receptor, resulting in antitumor effects, such as the inhibition of mitogenic signal transduction, the induction of antibody-dependent cell-mediated cytotoxicity (ADCC) by natural killer (NK) cells, and probably the downregulation of the receptor itself through its internalization. Tumor-derived exosomes express on their surface HER2, which binds and sequesters trastuzumab molecules, thus reducing their availability and efficiency on tumor cells; it has been observed that exosome-mediated protective effect is higher in advanced tumors because they produce more vesicles, which strongly limit drug availability thus promoting tumor progression and metastatization (252).

2.8 NcRNAs in biological fluids

Although they play a crucial regulatory role within cells, ncRNAs are also secreted into the extracellular environment; frequently, ncRNAs are contained within exosomes, which make them resistant to degradation, more stable and able to reach distant body areas. NcRNAs have been found in all biological fluids analysed to date, first of all blood. Searching for non-invasive methods to diagnose tumor, blood represents the most interesting sample, because: i) it can be obtained in a simple and non-invasive way; ii) since it circulates throughout the entire body, blood collects exosomes produced by all cell hystotypes, including brain cells, because exosomes are small enough to cross the blood-brain barrier. Recently, several studies reported that altered serum levels of ncRNAs are associated with many tumors, suggesting that ncRNAs could be used as diagnostic markers; circulating miRNAs have been widely analysed, both in neoplastic and non-neoplastic diseases, while research on circulating IncRNAs and circRNAs is just beginning.
Few data have been produced on CRC and lncRNAs. The lncRNA CRNDE is upregulated in CRC tissues; serum levels of the splicing variant CRNDE-h are significantly higher in CRC patients compared to healthy individuals, and showed a good diagnostic accuracy (87% sensitivity and 93% specificity) (253). Another study identified through microarray experiments a molecular signature composed by three new lncRNAs (XLOC_006844, LOC152578 and XLOC_000303) in plasma of CRC patients (254). Successively, it has been reported that CCAT1 and HOTAIR levels are increased, while lincRNA-p21/TP53COR1 (tumor protein p53 pathway corepressor 1) levels are decreased in plasma of CRC patients compared to healthy individuals; the predictive power of these lncRNAs raises by combining HOTAIR and CCAT1. Moreover, HOTAIR and CCAT1 plasma levels decrease and lincRNA-p21 levels increase after surgical excision of the tumor (255).

No data are to date available about circulating circRNAs and CRC.

The study presented in this doctoral thesis aims to investigate lncRNA and circRNA involvement in colorectal cancer through gene expression analysis in tissues and serum exosomes of CRC patients compared to normal colon mucosa and healthy individuals respectively. Through ROC curve computation, we assessed the diagnostic accuracy of ncRNAs in serum exosomes to evaluate their possible application as CRC biomarkers. By treating CRC cell lines with MAPK inhibitors, we also investigated ncRNAs involvement in MAPK pathway, which regulates cell cycle and has been associated with CRC. Finally, we performed ASO-mediated silencing of UCA1 in HCT-116 cells and analysed the expression of its mRNA and miRNA targets, aiming to investigate the effect of lncRNA-mRNA/miRNA and circRNA-miRNA interactions on the stability and the eventual degradation of the same RNA molecules.
3. Materials and methods

3.1 Cell lines

HCT-116 and Caco-2 cell lines were obtained from the Interlab Cell Line Collection (ICLC), an International Repository Authority within the IRCCS Azienda Ospedaliera Universitaria San Martino-IST Istituto Nazionale per la Ricerca sul Cancro (Genova, Italy); both cell lines derive from primary tumor. HCT-116 cells, bearing KRAS mutation, were cultured in Roswell Park Memorial Institute-1640 (RPMI-1640) medium (Gibco), supplemented with 10% fetal bovine serum (FBS) (Gibco) and 2mM L-glutamine (Lonza); Caco-2 cells, bearing wild type KRAS, were cultured in Dulbecco’s Modified Eagle Medium (DMEM) (Sigma), supplemented with 20% FBS (Gibco), 2mM L-glutamine (Lonza) and 1% non-essential amino acids (NEAA) (Lonza); both medium were supplemented with penicillin/streptomycin (10.000 U/ml) (Gibco). Cells were cultured at 37°C with 5% CO₂.

3.2 Isolation of exosomes secreted by CRC cell lines

Exosomes secreted by both cell lines were isolated from culture medium 3 days after seeding in 75 cm² flasks: the medium was harvested and centrifuged at 300 g (4°C, 10 minutes) on a Beckman J-6M/E centrifuge (rotor TY.JS 5.2) to pellet debris. Supernatant was then centrifuged at 16,500 g (4°C, 30 minutes) on Beckman L8 70M ultracentrifuge (rotor SW28), and then filtered through a 0.2 μm filter. The final supernatant was ultracentrifuged at 120,000 g (4°C, 70 minutes) on Beckman L8-70M ultracentrifuge. Exosome pellet was directly lysed for RNA isolation. FBS used to culture the cells was deprived of exosomes through the same protocol, removing exosome pellet and harvesting FBS supernatant for subsequent use.
3.3 RNA isolation from cell lines and exosomes

Total RNA was extracted from cell lines and exosomes with TriZol (Invitrogen), according to manufacturer instructions, and quantified by GenQuant pro spectrophotometer (Biochrom) and Qubit fluorescence quantification system (Invitrogen).

3.4 RNA isolation from CRC formalin-fixed, paraffin-embedded samples

For our study, 20 CRC patients were recruited at Azienda Ospedaliero-Universitaria Policlinico Vittorio Emanuele (Catania, Italy); formalin-fixed, paraffin-embedded (FFPE) tumor tissues and distal normal mucosa were retrieved at Unit of Anatomical Pathology, Department G.F. Ingrassia, University of Catania (Catania, Italy).

RNA was extracted from FFPE samples through PureLink FFPE RNA Isolation Kit (Ambion), according to manufacturer instructions, and quantified by GenQuant pro spectrophotometer and Qubit.

3.5 Exosome and RNA isolation from CRC patient serum

For serum exosome analysis, 20 CRC patients and 20 healthy individuals were recruited at Istituto Oncologico del Mediterraneo (IOM) (Viagrande, Catania, Italy) and Azienda Ospedaliero-Universitaria Policlinico Vittorio Emanuele (Catania, Italy). Following formal written consent, patients underwent fasting venous blood sampling. To separate serum from cellular component, blood samples were centrifuged at 15,000 g (4°C, 15 minutes) on Beckman J-6M/E centrifuge; supernatant was harvested and centrifuged again under the same conditions to remove eventual circulating cells or debris which could alter subsequent analysis. Supernatant serum was harvested and stored at −80°C until analysis.

Patient serum circulating exosomes were isolated with ExoQuick (System Byosciences), a proprietary polymer that gently precipitates exosomes, according to manufacturer instructions.
Exosome pellet was directly lysed with TriZol for total RNA isolation; extracted RNA was quantified by GenQuant pro spectrophotometer and Qubit.

### 3.6 PCR primer design

By literature data mining, we selected for our study 17 lncRNAs involved in CRC or in other tumors, and regulating carcinogenesis processes, such as cell cycle, apoptosis, gene expression, splicing (Table 3.1).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCAT1</td>
<td>GGAAAGGTGCCGAGACATGA</td>
<td>GCCATACAGAGCCAACCTGG</td>
</tr>
<tr>
<td>CCAT2</td>
<td>CCTGCAGAGGGGCACTAGACT</td>
<td>CCAGGGTCAGGCAATTGGTC</td>
</tr>
<tr>
<td>CRNDE</td>
<td>TCGATCGGCGCTATTGTCATGG</td>
<td>GCCTCGCTTAGACATGGC</td>
</tr>
<tr>
<td>H19</td>
<td>GAGTCGGCACACACATGGCT</td>
<td>GTCCGGATTCAAAGGCCAG</td>
</tr>
<tr>
<td>HOTAIR</td>
<td>GTGGGGGTAAGAGAGCACCA</td>
<td>GAGGGTCTAAGTCCGCGT</td>
</tr>
<tr>
<td>LINK-ROR</td>
<td>CTTCAGGCACTCAGGAATGGG</td>
<td>AGGGTGCTGAGAGGATCC</td>
</tr>
<tr>
<td>LIT1</td>
<td>CAGGACATGCCCCCCACTAGC</td>
<td>CTCTTCTACTGTGAGGCC</td>
</tr>
<tr>
<td>MALAT1</td>
<td>TAGTCTGATCTCCTGTGGG</td>
<td>AACCCCAAAAGACCTCGAC</td>
</tr>
<tr>
<td>MEG3</td>
<td>GCCAGCTGCCTCTCTACCT</td>
<td>GCATAGCAAGGTCAGGCT</td>
</tr>
<tr>
<td>MIR17HG</td>
<td>TGCCACGTGGATGTGAAGAT</td>
<td>GCTTCTCAGGAGTGAGGC</td>
</tr>
</tbody>
</table>

**Table 3.1:** cancer-related lncRNAs selected by literature data mining.

To analyse the expression of selected lncRNAs, we designed PCR primers; in addition, we designed specific primers for two housekeeping genes used for normalization: PPIA (peptidylprolyl isomerase A) and GAPDH (glyceraldehyde 3-phosphate dehydrogenase). Primer design was performed by using the online tool PrimerBlast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). Primers are shown in table 3.2.
### Materials and methods

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCAT1</td>
<td>AGGCCAAGCTACAGGGAGAAA</td>
<td>TGGGATTATCTTTGGGAGGTCC</td>
</tr>
<tr>
<td>PCAT6</td>
<td>CTCTTGGACACAAACTCCGCC</td>
<td>ATCTTGGGCTGCACAAATGAGGA</td>
</tr>
<tr>
<td>PTENP1</td>
<td>CATTCTTGCATGTATTTGGGATTGGA</td>
<td>GGTATATGTTCCAGAGTCCAC</td>
</tr>
<tr>
<td>TUG1</td>
<td>CTTTCACCATGGGGTTGGCTG</td>
<td>AGCTTGTCTTACTGGGTTGCC</td>
</tr>
<tr>
<td>UCA1</td>
<td>TGCAACCTTAGACGTTCCAAA</td>
<td>CAAATGTGACCCAGGGACTG</td>
</tr>
<tr>
<td>WRAP53</td>
<td>ACCGGACTGGCTGAGGTTTTTTTT</td>
<td>CTGGGCTGAAAGGGCTATGCA</td>
</tr>
<tr>
<td>GAPDH</td>
<td>GGCTGACTGCGGCCCCAGTTTT</td>
<td>AGCGGCCCAATACGACAAAA</td>
</tr>
<tr>
<td>PPIA</td>
<td>ACTGAGTGGTTGGGATGGAAGGCA</td>
<td>AGCGCTCCATGGGCTCCAC</td>
</tr>
</tbody>
</table>

**Table 3.2:** PCR primers designed for selected lncRNAs and housekeeping genes.

Primers for 31 circRNAs were retrieved by a paper published by Memczak and colleagues on Nature in 2013 (80) (Table 3.3).
Materials and methods

3.7 LncRNA and circRNA expression analysis by Real-Time PCR

We investigated the expression of selected lncRNAs and circRNAs in CRC cell lines and exosomes, and in CRC patient tissues and serum exosomes though Real-Time PCR by using Power SYBR Green RNA-to-CT™ 1-Step kit (Applied Biosystems), according to manufacturer instructions. All Real-Time PCR reactions were performed on a 7900HT Fast Real-Time PCR System (Applied Biosystems).

Differentially expressed (DE) lncRNAs and circRNAs were identified through SDS RQ Manager 1.2 software (Applied Biosystem); normalization was performed using PPIA and GAPDH for tumor tissues and serum exosomes respectively. Expression fold changes of DE ncRNAs were calculated by applying the $2^{-\Delta\Delta Ct}$ method. Statistical analysis performed by paired T-test was used to compare tissue sample $\Delta Ct$s, while unpaired T test was used for serum exosome $\Delta Ct$s; statistical significance was established at a p-value $\leq 0.05$.

3.8 Computational analysis

To confirm our results, we analysed DE lncRNA expression in several datasets derived from microarray experiments deposited in GEO DataSets (https://www.ncbi.nlm.nih.gov/gds/) (Table 3.4).

<table>
<thead>
<tr>
<th>GEO DataSets ID</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE10715</td>
<td>peripheral blood (19 tumors vs 11 controls), blood smears (10 tumors vs 10 controls)</td>
</tr>
<tr>
<td>GSE13294</td>
<td>fresh-frozen primary tumor tissues (74)</td>
</tr>
<tr>
<td>GSE14333</td>
<td>fresh-frozen primary tumor tissues (290)</td>
</tr>
<tr>
<td>GSE17538</td>
<td>fresh-frozen primary tumor tissues (232)</td>
</tr>
</tbody>
</table>

Table 3.3: circRNA primers retrieved by Memczak et al., Nature 2013.
Materials and methods

<table>
<thead>
<tr>
<th>GEO Series</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE18088</td>
<td>frozen or formalin-fixed and paraffin-embedded tissues (53)</td>
</tr>
<tr>
<td>GSE18105</td>
<td>cancer tissues and normal adjacent tissues (17), laser capture microdissected primary tumors (77)</td>
</tr>
<tr>
<td>GSE20916</td>
<td>distant full-thickness normal colon, normal colon mucosa dissected from tumor, adenoma, and carcinoma (40)</td>
</tr>
<tr>
<td>GSE2109</td>
<td>tumor tissues (315)</td>
</tr>
<tr>
<td>GSE21510</td>
<td>laser capture microdissected and homogenized tumors (148)</td>
</tr>
<tr>
<td>GSE23878</td>
<td>tumor tissues (35 tumors vs 24 matched controls)</td>
</tr>
<tr>
<td>GSE33114</td>
<td>fresh frozen or formalin-fixed and paraffin-embedded tissues (90 tumors vs 6 controls)</td>
</tr>
<tr>
<td>GSE35896</td>
<td>primary tumor tissues (62)</td>
</tr>
<tr>
<td>GSE3629</td>
<td>53 ulcerative colitis tissues (10 UC-associated cancer, 43 non-cancer), 60 sporadic cancer tissues</td>
</tr>
<tr>
<td>GSE39582</td>
<td>fresh-frozen primary tumor tissues (566)</td>
</tr>
<tr>
<td>GSE41258</td>
<td>primary colon adenocarcinomas, adenomas, metastasis and corresponding normal mucosae (390)</td>
</tr>
</tbody>
</table>

Table 3.4: CRC GEO DataSets used to confirm DE lncRNA expression.

In addition, we performed a computational analysis to identify potential mRNA and miRNA targets of DE ncRNAs.

For DE lncRNAs, data about correlation of expression were retrieved in three databases containing microarray data: MEM (http://biit.cs.ut.ee/mem/), which shows data about positive and negative correlation, lncRNAtor (http://lncrnator.ewha.ac.kr/index.htm) and BioGPS (http://biogps.org/), which only provide data about positive correlation; through these databases we obtained three list of mRNAs (one from each database) with correlation of expression for each DE lncRNA. These lists were submitted to the database cBioPortal (http://www.cbioportal.org/): according to CRC data published on Nature in 2012 within the TCGA (The Cancer Genome Atlas) project (256), cBioPortal database allowed us to identify those genes within the lists which have been shown to be altered in CRC (because of point mutations, deletions, gene amplification,
upregulation or downregulation); we selected genes altered in a percentage of cases corresponding to the first and the second quartiles. Gene lists filtered through cBioPortal were then unified for each DE lncRNA; for each gene, we retrieved from Gene NCBI (https://www.ncbi.nlm.nih.gov/gene/) the RefSeq accessions, which were used to perform a sequence alignment between the DE lncRNA and the mRNA targets by using Nucleotide BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (settings: “Somewhat similar sequence” and “Species-specific repeats for Homo sapiens”). We selected only those genes showing a complementary (not homologous) region longer than 25 nucleotides (Figure 3.1). Therefore, for each DE lncRNA, we obtained a list of genes involved in CRC and characterized by correlation of expression and sequence complementarity with the DE lncRNA.

![Sequence Alignment Example](image)

**Figure 3.1:** selection criteria for target genes of DE lncRNAs after sequence alignment: “strand plus/plus” indicates sequence homology, while “strand plus/minus” indicates sequence complementarity.
In addition, we aligned each DE lncRNA sequence with the whole genome sequence by using two different databases: NR and Human genomic and transcripts; this additional analysis allowed us to identify genomic regions characterized by sequence complementarity with each DE lncRNA, which could bind DNA modulating the expression of near target genes. Using again Nucleotide BLAST, we obtained genes, transcripts and chromosome tracts complementary to each DE lncRNA; again, we selected only those complementary tracts longer than 25-30 nucleotides. The list including complementary genes (or the genes nearest to the complementary region) was analysed to identify genes which expression was correlated to lncRNA expression and genes involved in CRC; for this purpose, we used again MEM, selecting only genes with p-values lower that $10^{-6}$, and IncRNAtor, comparing the two gene lists; selected genes were then submitted to cBioPortal to investigate their involvement in CRC, again selecting only genes altered in a percentage of cases corresponding to the first and the second quartiles. This second analysis provided us a second list (for each DE lncRNA) of genes involved in CRC and characterized by correlation of expression and sequence complementarity with the DE lncRNA. The two final lists were unified for each DE lncRNA; finally, a literature search was performed to further filter the list, selecting only genes involved in carcinogenesis and cancer-related processes, such as regulation of splicing, chromatin status, cell cycle, apoptosis and metastasis.

Interactions with miRNAs have been investigated for both DE lncRNAs and circRNAs. First of all, miRNAs involved in CRC were identified by mining two databases, miRCancer (http://mircancer.ecu.edu/) and miR2Disease (http://www.mir2disease.org/); mature sequences of these miRNAs were retrieved from miRBase (http://www.mirbase.org/) and then used to perform a sequence alignment on each DE ncRNA again through Nucleotide BLAST (settings: “Word size 7”, “Somewhat similar sequence” and “Species-specific repeats for Homo sapiens”). We selected only those miRNAs showing at least four complementary tracts located throughout the whole ncRNA sequence but always within the miRNA seed region. Successively, selected miRNAs were further
filtered selecting only those characterized by negative correlation of expression (anticorrelation) with each DE ncRNA.

### 3.9 Biological network construction and Gene ontology analysis

To understand the effects of DE ncRNA deregulation, we built two biological networks based on the interactions between all DE ncRNAs and their mRNA targets and all DE ncRNAs and their miRNA targets identified through the computational analysis; in addition, the software retrieved the nearest neighbours of target genes. Networks were built by using the software Cytoscape 2.8 (http://www.cytoscape.org/) and the MIMI Plugin. To build the network based on ncRNA-miRNA interactions, miRNA validated target retrieved from the database TarBase v7.0 (http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=tarbase/index) were used.

Functional classification of Gene ontologies and pathway enrichment analysis were performed using all genes included in the two networks by using: i) the FatiGO online tool available on Babelomics 4.3 (http://v4.babelomics.org/functional.html), applying the Fisher’s exact test with false discovery rate (FDR) Correction (p-value < 0.005); ii) the GeneMania online tool (http://www.genemania.org/), calculating q-values with FDR corrected hypergeometric test.

### 3.10 ROC curve analysis

To evaluate the diagnostic accuracy of ncRNAs deregulated in patient serum exosomes, ROC (receiver operating characteristic) curves were computed through the software MedCalc v 15.11.4. The area under the curve (AUC) and 95% confidence intervals (95% CIs) were calculated to assess the accuracy of sensitivity and specificity parameters and to find an appropriate cut-off point. Statistical significance was established at a p-value ≤ 0.05.
3.11 CRC cell lines treatment with MAPK inhibitors

We investigated the effect of MAPK pathway inhibition on DE ncRNAs in HCT-116 cells, characterized by overactivation of the pathway because of KRAS mutation. MAPK inhibition was achieved through treatment with FR180204 and U0126. FR180204 is an ATP-competitive inhibitor of ERK1 and ERK2 and a weaker inhibitor of the related kinase p38α. 2x10^5 cells per well were seeded in 12-well plates and cultured in serum starvation conditions (1% FBS) for 6 hours; cells were then treated with 5 µM, 10 µM, 25 µM, 50 µM FR180204 (ERK Inhibitor II, Calbiochem). U0126 is a highly selective ATP-non-competitive inhibitor of both MEK1 and MEK2 kinases, specifically preventing MEK1 and MEK2 phosphorylation and activation in vitro by binding the inactive enzymes and blocking the recruitment of Raf-1 kinase. 3x10^5 cells per well were seeded in 6-well plates and cultured in serum starvation conditions (1% FBS) for 14 hours; successively, cells were treated with 5 µM, 10 µM, 20 µM U0126 (MEK1/2 inhibitor, Merck). Cells were exposed to inhibitors for 24 hours. Control samples were treated with an equivalent volume of DMSO (solvent of the drugs used for treatments). All experiments were performed in biological triplicates.

Cell viability was assessed through MTT assay at 24 hours post-treatment.

3.12 Protein extraction and Western Blot analysis

MAPK pathway inhibition was verified by analysing phospho-ERK (pERK) levels through Western Blot. Proteins were extracted with RIPA buffer [2% protease-inhibitor cocktail (Sigma), 1% of both phosphatase-inhibitor cocktail 2, phosphatase-inhibitor cocktail 3 (Sigma)] and quantified by Qubit (Invitrogen). Blotting was performed by iBlot Dry Blotting System (Invitrogen). Proteins were detected using the ECL Plus Western Blotting detection (Amersham). Membranes were bound with monoclonal antibodies to p44/42 MAPK (Erk1/2) and phospho-p44/p42 MAPK (Erk1/2) (Cell Signaling), using beta-actin (Abcam) as loading control. Bound antibodies were detected with peroxidase-labeled goat antirabbit IgG-HRP secondary antibody (Santa Cruz).
3.13 In vitro silencing of UCA1

We performed UCA1 silencing in HCT-116 cell line through antisense oligonucleotide (ASO) technology. 4x10⁴ cells per well were seeded in 24-well plates and simultaneously reverse-transfected with 50 pmoles of UCA1 specific Antisense LNA™ GapmeR (Exiqon), using Lipofectamine RNAiMAX Reagent (Invitrogen), according to manufacturer instruction. All experiments were performed in biological triplicates. 24 and 48 hours after transfection cells were lysed with TriZol for RNA isolation. UCA1 mRNA and miRNA targets were selected through our computational analysis, the analysis of GEO DataSets (GSE15960, GSE33114, GSE39582, GSE8671), and literature search. Specific PCR primers were designed and used to evaluate the expression of mRNA targets by Real-Time PCR (Table 3.5), while miRNA expression was evaluated through TaqMan miRNA assays (Applied Biosystems). PPIA and miR-29a were used as endogenous controls.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>AREG</td>
<td>CCTCACAGCTGTTGCTGTTATT</td>
<td>TTCGTTCCCTAGCTTCTCCCTT</td>
</tr>
<tr>
<td>ASCL2</td>
<td>CTTGGAGCTGGCCCCCATAA</td>
<td>CACTGCACTGAGAAGCTGT</td>
</tr>
<tr>
<td>LINC01764</td>
<td>CTCCGGACTGCTTCAAGTGT</td>
<td>GTTATTCACCGTCAACAGTTCTCA</td>
</tr>
<tr>
<td>MACC1</td>
<td>GCCACACAGAGAAATACAGG</td>
<td>CGTGCGACTAAACTCTGGCA</td>
</tr>
<tr>
<td>NGL</td>
<td>CCCAGGGGATCACCTAATGC</td>
<td>CAGTGGCATCTCCAGACAGGC</td>
</tr>
<tr>
<td>PDP1</td>
<td>GTAGCCACAGGCACATCAT</td>
<td>AAGAAGTACGATTCGCGG</td>
</tr>
<tr>
<td>SYNCRI</td>
<td>TACAAAAGAAGCAGCTCAGGAGG</td>
<td>ATTGGTGCAACTGAGATGCAG</td>
</tr>
<tr>
<td>UGDH</td>
<td>GCAACAGGGTGGGAATGGA</td>
<td>TTTTGAACAGCTCNCACCA</td>
</tr>
</tbody>
</table>

Table 3.5: PCR primers for UCA1 mRNA targets.
4. Results

4.1 NcRNA expression in CRC cell lines and exosomes

The expression of selected IncRNAs and circRNAs was evaluated through Real-Time PCR in two CRC cell lines, HCT-116 and Caco-2. The results of this preliminary screening showed that almost all analysed molecules are expressed in both cell lines (Figure 4.1).

![IncRNA expression in CRC cell lines](image1)

**Figure 4.1:** expression of selected IncRNAs and circRNAs in CRC cell lines HCT-116 and Caco-2; expression is shown as 40-Ct.
NcRNAs showing the highest expression in CRC cell lines were also tested in exosomes secreted by the same cells, aiming to evaluate their presence within vesicles. Except for circ2, all tested molecules were detected into exosomes secreted by both CRC cell lines (Figure 4.2).

Figure 4.2: expression of lncRNAs and circRNAs in exosomes compared to donor cells; expression is shown as 40-Ct.

4.2 NcRNA expression in FFPE CRC biopsies

Through a trial experiment on three FFPE biopsy pairs, we selected 15 out of 17 lncRNAs (CCAT1, CCAT2, CRNDE, H19, HOTAIR, MALAT1, MEG3, MIR17HG, PCAT1, PCAT6,
PTENP1, TUG1, UCA1, WRAP53, ZEB2AS1) and 6 out of 31 circRNAs (circ2, circ4, circ13, circ17, circ22, CDR1AS) to be analysed in all FFPE samples. We performed Real-Time PCR analysis in 20 FFPE CRC biopsies and their normal adjacent tissues (NATs); results showed significant differential expression of 7 lncRNAs and 1 circRNA (Table 4.1).

<table>
<thead>
<tr>
<th>lncRNA</th>
<th>Fold Change</th>
<th>T test p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCAT1</td>
<td>8.37</td>
<td>0.0007</td>
</tr>
<tr>
<td>CCAT2</td>
<td>4.67</td>
<td>0.001</td>
</tr>
<tr>
<td>CDR1AS</td>
<td>-3.3</td>
<td>0.002</td>
</tr>
<tr>
<td>circ13</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>circ17</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>circ2</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>circ22</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>circ4</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>CRNDE</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>H19</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>HOTAIR</td>
<td>9.17</td>
<td>0.0008</td>
</tr>
<tr>
<td>MALAT1</td>
<td>-1.8</td>
<td>0.004</td>
</tr>
<tr>
<td>MEG3</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>MR17HG</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td>PCAT1</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>PCAT6</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>PTENP1</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>TUG1</td>
<td>-2.04</td>
<td>0.036</td>
</tr>
<tr>
<td>UCA1</td>
<td>7.14</td>
<td>0.001</td>
</tr>
<tr>
<td>WRAP53</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>ZEB2AS1</td>
<td>12.13</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Table 4.1: statistical results of lncRNA and circRNA expression analysis in 20 tissue biopsies of CRC patients compared to normal adjacent tissues. For each ncRNA, paired T test p-value and fold change are shown.

Our results showed that 5 lncRNAs (CCAT1, CCAT2, HOTAIR, UCA1 and ZEB2AS1) were upregulated, while 2 lncRNAs (MALAT1 and TUG1) and 1 circRNA (CDR1AS) were downregulated in CRC tumor tissues compared to NATs (Figure 4.3).
Figure 4.3: box plots showing differential expression of 7 lncRNAs and 1 cirRNA in 20 CRC tumor tissues compared to normal adjacent tissues.
4.3 NcRNA expression in serum exosomes of CRC patients

We performed a trial experiment with a small group of samples to identify lncRNAs and circRNAs to be tested in the entire cohort of 20 CRC patients and 20 healthy individuals. The trial experiment allowed us to select 10 lncRNAs (CCAT2, CRNDE, HOTAIR, LIT1, MALAT1, PCAT1, PCAT6, TUG1, UCA1, ZEB2AS1) and 3 circRNAs (circ4, circ7, circ16), which expression was evaluated through Real-Time PCR. Our results showed that 3 ncRNAs were differentially expressed in serum exosomes of CRC patients compared to healthy individuals: in particular, UCA1 was downregulated, while circ16 and TUG1 were upregulated (Table 4.2) (Figure 4.4).

<table>
<thead>
<tr>
<th></th>
<th>Fold Change</th>
<th>T test p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCAT2</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>circ16</td>
<td>3.82</td>
<td>0.036</td>
</tr>
<tr>
<td>circ4</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>circ7</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>CRNDE</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>HOTAIR</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>LIT1</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>MALAT1</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>PCAT1</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>PCAT6</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>TUG1</td>
<td>1.69</td>
<td>0.029</td>
</tr>
<tr>
<td>UCA1</td>
<td>-4.22</td>
<td>0.03</td>
</tr>
<tr>
<td>ZEB2AS1</td>
<td>0.95</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2: statistical results of lncRNA and circRNA expression analysis in serum exosomes of 20 CRC patients compared to 20 healthy individuals. For each ncRNA, unpaired T test p-value and fold change are shown.
Figure 4.4: box plots showing differential expression of 2 lncRNAs and 1 circRNA in exosomes isolated from serum of 20 CRC patients and 20 healthy individuals.

We computed Pearson’s correlation coefficient (r) to evaluate the correlation between TUG1 and UCA1 expression in serum exosomes of all CRC patients and healthy individuals, and in serum exosomes of CRC patients only. The r values showed in both cases a negative correlation, stronger in CRC patient group; T test revealed that this anticorrelation is strongly significant in both sample groups (Figure 4.5).

Figure 4.5: Pearson’s correlation coefficient (r) showed a negative correlation between TUG1 and UCA1 expression in serum exosomes of CRC patients and healthy individuals (A), and in serum exosomes of CRC patients only (B).
No statistical correlation was observed between UCA1 and circ16 and between TUG1 and circ16.

4.4 Analysis of GEO DataSets and identification of mRNA and miRNA targets of DE ncRNAs

Overall, our study revealed the differential expression of 8 ncRNAs in CRC FFPE biopsies compared to NATs and serum exosomes of CRC patients compared to healthy individuals. To validate our results, we analysed the expression of these ncRNAs in several CRC GEO DataSets; however, we obtained no results for CCAT1, CCAT2, CDR1AS and circ16, because no probe specific for these ncRNAs was retrieved in the analysed datasets. Datasets based on CRC tissues and normal colon mucosa confirmed the same deregulation trends here reported and also revealed a positive correlation between the up- or downregulation of each DE ncRNA and CRC progression and features (Figure 4.6).
Figure 4.6: GEO DataSets analysis confirmed the upregulation of HOTAIR and UCA1 and the downregulation of MALAT1 and TUG1 in CRC tissues compared to NATs; moreover, the same analysis revealed a positive correlation between ncRNA up- or downregulation and CRC progression; statistical significance is shown as \(-\log_{10}(p\text{-value})\). Each dataset is represented by a different colour. MSI: microsatellite instability; MSS: microsatellite stability; CIN: chromosomal instability.

In addition, we analysed a dataset based on a study upon peripheral blood of CRC patients compared to healthy individuals. Again, the analysis confirmed the upregulation of TUG1 and the downregulation of UCA1 that we observed in exosomes isolated from serum of CRC patients and
healthy individuals (Figure 4.7a). In particular, the same deregulation trend of both lncRNAs was observed in two different groups of CRC patients based on Dukes’ classification: i) Dukes’ stages A and B, ii) Dukes’ stages C and D. Unfortunately, no clinical data was available to further analyse the association between ncRNA deregulation and CRC. Plus, TUG1 and UCA1 showed a significant negative correlation of expression (r-value=-0.439 p-value=0.02) (Figure 4.7b).

**Figure 4.7:** analysis of a GEO dataset reporting expression data of peripheral blood of CRC patients compared to healthy individuals (GSE10715). A) the dataset confirmed the upregulation of TUG1 and the downregulation of UCA1; the same trend is observed in two different groups of CRC patients, composed by i) Dukes’ stages A and B, and ii) Dukes’ stages C and D. B) TUG1 and UCA1 showed a significant negative correlation of expression.
We identified hypothetical mRNA targets of lncRNAs deregulated in tumor tissues using three databases (lncRNAtor, MEM, BioGPS); however, we obtained data about HOTAIR, MALAT1, TUG1 and UCA1 from all the three databases, while CCAT1 and ZEB2AS1 data were retrieved only from lncRNAtor; no data about CCAT1 were found in any analysed database. Genes with correlation of expression were submitted to cBioPortal database to verify their involvement in CRC; the sequences of selected genes were aligned with each DE lncRNA sequence through Nucleotide BLAST. Also, each lncRNA sequence was aligned with whole genome sequence; the existence of a correlation of expression of selected complementary genes with each DE lncRNA was investigated: i) genes complementary to HOTAIR, MALAT1, TUG1 and UCA1 were submitted to MEM; only genes with p-value inferior or equal to $10^{-6}$ were selected; ii) genes complementary to CCAT1 and ZEB2AS1 were compared to genes retrieved from lncRNAtor; only common genes were selected. Only the alignment with the whole genome provided data about CCAT2.

We also identified hypothetical miRNA targets of lncRNAs and circRNAs deregulated in tumor tissues through the alignment of mature sequences of miRNAs associated with CRC and the sequence of each DE ncRNA; only miRNAs showing sequence complementarity and negative correlation of expression with DE ncRNAs were selected.

Tables 4.3 and 4.4 show mRNA and miRNA targets selected for each DE ncRNA.

<table>
<thead>
<tr>
<th>DE ncRNA</th>
<th>mRNA gene name</th>
<th>gene name</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCAT1</td>
<td>ACTL6A</td>
<td>actin-like 6°</td>
</tr>
<tr>
<td></td>
<td>PDZD2</td>
<td>PDZ domain containing 2</td>
</tr>
<tr>
<td>CCAT2</td>
<td>MKI67</td>
<td>marker of proliferation Ki-67</td>
</tr>
<tr>
<td></td>
<td>PTPRU</td>
<td>protein tyrosine phosphatase, receptor type, U</td>
</tr>
<tr>
<td></td>
<td>RASEF</td>
<td>RAS and EF-hand domain containing</td>
</tr>
<tr>
<td>HOTAIR</td>
<td>EDNRB</td>
<td>endothelin receptor type B</td>
</tr>
<tr>
<td>SRC</td>
<td>SRC proto-oncogene, non-receptor tyrosine kinase</td>
<td></td>
</tr>
<tr>
<td>MALAT1</td>
<td>NCOA2</td>
<td>nuclear receptor coactivator 2</td>
</tr>
<tr>
<td></td>
<td>PAX3</td>
<td>paired box 3</td>
</tr>
<tr>
<td></td>
<td>RAC1</td>
<td>ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)</td>
</tr>
</tbody>
</table>
### 4.5 Analysis of gene ontologies

Gene ontology and pathway enrichment analyses were performed aiming to understand the effects of lncRNA and circRNA deregulation within tumor cells. We used mRNA and miRNA
targets of DE ncRNAs and their nearest neighbours to build two biological networks; all genes included in these network were then used to perform the analyses. The network based on miRNA targets comprised 3792 nodes and 26870 edges, while the network based on mRNA targets was made of 939 nodes and 12710 edges. Results of gene ontology and pathway enrichment analyses are shown in figure 4.8.

Figure 4.8: pathways regulated by miRNA and mRNA targets of DE ncRNAs in tumor tissues; chromatic scale represents $-\log_{10}(p\text{-value})$ of Fisher's exact test.

Gene ontology and pathway enrichment analyses were also performed for ncRNAs deregulated in serum exosomes of CRC patients; we performed different analysis for each DE ncRNA: FatiGO tool was used for circ16 and UCA1, while GeneMania was used for TUG1. Pathways regulated by ncRNAs deregulated in serum exosomes are shown in figure 4.9.
4.6 Evaluation of the diagnostic accuracy of DE ncRNAs in serum exosomes

Differentially expressed ncRNAs in serum exosomes of CRC patients compared to healthy individuals represent potential tumor markers; to assess their diagnostic accuracy we computed ROC curves using \( \Delta Ct \) values respect to endogenous control GAPDH. Circ16 ROC curve showed an AUC of 0.771 (95% CIs, 0.508-0.936; \( p = 0.02 \)) with 71% of sensitivity and 80% of specificity.
Results

(ΔCt cut-off value ≤ 8.88) (Figure 4.10a); TUG1 curve showed an AUC of 0.640 (95% CIs, 0.455-0.799) with 56% of sensitivity and 88% of specificity (ΔCt cut-off value > -2.22), but it was not statistically significant (p = 0.2) (Figure 4.10b); UCA1 curve showed an AUC of 0.719 (95% CIs, 0.533-0.863; p = 0.01) with 100% of sensitivity and 43% of specificity (ΔCt cut-off value ≤ 3.14) (Figure 4.10c).

Figure 4.10: ROC curves of DE ncRNAs in serum exosomes of CRC patients compared to healthy individuals.

Although these results derived from a limited sample size and further analyses are needed, they suggest that ncRNAs in serum exosomes could be used as diagnostic markers for CRC with a good diagnostic accuracy.
4.7 *In vitro* inhibition of MAPKs

4.7.1 Cell viability assay

Cell viability was assessed through MTT assay 24 hours after treatment with MAPK inhibitors FR180204 and U0126. Results showed that both inhibitors significantly reduced cell proliferation of treated cells compared to negative controls treated with an equal volume of DMSO (Figure 4.11).

![MTT assay 24h post treatment](image)

**Figure 4.11:** treatment with both MAPK inhibitors (FR180204 and U0126) significantly reduced cell viability, measured by MTT assay; cells treated with DMSO were used as controls. *: p-value < 0.05; **: p-value < 0.005; ***: p-value < 0.0005.

4.7.2 Confirmation of MAPK inhibition through Western Blot analysis

Western Blot analysis was performed 24 hours after treatment with MAPK inhibitors FR180204 and U0126 to confirm the inhibition of MAPK pathway. Results showed a significant reduction in phospho-ERK (pERK) levels in treated cells compared to control cells, treated with equal volume of DMSO (Figure 4.12).
Figure 4.12: Western Blot showing ERK and pERK levels in HCT-116 cells treated with MAPK inhibitors FR180204 and U0126 compared to control cells treated with DMSO: MAPK inhibition is showed by reduced levels of pERK compared to ERK and beta-actin, used as loading control.

4.7.3 MAPK inhibition affects ncRNA expression

The expression of ncRNAs deregulated in CRC tissues and serum exosomes was investigated in HCT-116 cells 24 hours after treatment with U0126. Expression analysis was performed in Real-Time PCR using PPIA as endogenous control. Results showed that the expression of 4 out of 9 ncRNAs was affected by the inhibition of MAPK pathway (Table 4.5). The analysis will be also performed in HCT-116 cells treated with FR180204.

<table>
<thead>
<tr>
<th>ncRNA</th>
<th>Fold Change</th>
<th>T test p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCAT1</td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td>CCAT2</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>CDR1AS</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>circ16</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>HOTAIR</td>
<td>-15</td>
<td>0.037</td>
</tr>
<tr>
<td>MALAT1</td>
<td>3.66</td>
<td>0.002</td>
</tr>
<tr>
<td>TUG1</td>
<td>5.85</td>
<td>0.00009</td>
</tr>
<tr>
<td>UCA1</td>
<td>3.14</td>
<td>0.033</td>
</tr>
<tr>
<td>ZEB2AS1</td>
<td>0.59</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.5: expression of ncRNAs in HCT-116 cells treated with U0126. For each ncRNA, median fold change and T test p-value are shown.

In particular, U0126 treatment induced the upregulation of MALAT1, TUG1 and UCA1 and the downregulation of HOTAIR (Figure 4.13).
4.7.4 Evaluation of correlation between ncRNA expression and ERK activation

We analysed GEO DataSets to investigate the correlation between the expression of ncRNAs deregulated by MAPK inhibition and the activation of the pathway; the analysis was performed by retrieving from CRC datasets correlation data between each DE ncRNA and HSPA5 [heat shock protein family A (Hsp70) member 5], a transcriptional target of pERK (257). Results showed a positive correlation between HOTAIR and HSPA5, and a negative correlation between MALAT1, TUG1 or UCA1 and HSPA5, confirming our observation (Table 4.6).

<table>
<thead>
<tr>
<th></th>
<th>Geo DataSets ID</th>
<th>p-value</th>
<th>r-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOTAIR</td>
<td>GSE13294</td>
<td>3.00E-02</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>GSE33114</td>
<td>3.00E-02</td>
<td>0.212</td>
</tr>
<tr>
<td>MALAT1</td>
<td>GSE37892</td>
<td>1.10E-08</td>
<td>-0.475</td>
</tr>
<tr>
<td></td>
<td>GSE55896</td>
<td>1.30E-03</td>
<td>-0.4</td>
</tr>
<tr>
<td></td>
<td>GSE8671</td>
<td>7.70E-03</td>
<td>-0.33</td>
</tr>
<tr>
<td></td>
<td>GSE13294</td>
<td>2.70E-03</td>
<td>-0.24</td>
</tr>
<tr>
<td></td>
<td>GSE4554</td>
<td>3.00E-02</td>
<td>-0.233</td>
</tr>
<tr>
<td></td>
<td>GSE33114</td>
<td>2.00E-02</td>
<td>-0.217</td>
</tr>
<tr>
<td>TUG1</td>
<td>GSE18088</td>
<td>1.20E-05</td>
<td>-0.561</td>
</tr>
<tr>
<td></td>
<td>GSE13294</td>
<td>4.00E-14</td>
<td>-0.559</td>
</tr>
</tbody>
</table>
Results

<table>
<thead>
<tr>
<th></th>
<th>GSE33114</th>
<th>1.90E-09</th>
<th>-0.538</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GSE37892</td>
<td>1.30E-08</td>
<td>-0.473</td>
</tr>
<tr>
<td></td>
<td>GSE35896</td>
<td>2.50E-04</td>
<td>-0.449</td>
</tr>
<tr>
<td></td>
<td>GSE8671</td>
<td>2.20E-03</td>
<td>-0.376</td>
</tr>
<tr>
<td></td>
<td>GSE4554</td>
<td>5.10E-03</td>
<td>-0.303</td>
</tr>
<tr>
<td></td>
<td>GSE28702</td>
<td>5.80E-03</td>
<td>-0.301</td>
</tr>
<tr>
<td></td>
<td>GSE2109</td>
<td>1.70E-07</td>
<td>-0.29</td>
</tr>
<tr>
<td></td>
<td>GSE17538</td>
<td>1.50E-05</td>
<td>-0.28</td>
</tr>
<tr>
<td></td>
<td>GSE18105</td>
<td>3.00E-03</td>
<td>-0.279</td>
</tr>
<tr>
<td></td>
<td>GSE21510</td>
<td>2.90E-03</td>
<td>-0.243</td>
</tr>
<tr>
<td></td>
<td>GSE41258</td>
<td>2.10E-05</td>
<td>-0.213</td>
</tr>
<tr>
<td>UCA1</td>
<td>GSE33114</td>
<td>4.40E-03</td>
<td>-0.272</td>
</tr>
<tr>
<td></td>
<td>GSE14333</td>
<td>2.10E-03</td>
<td>-0.18</td>
</tr>
</tbody>
</table>

Table 4.6: analysis of CRC GEO DataSets showing correlation between ncRNAs deregulated by MAPK inhibition and HSPA5, a transcriptional target of pERK. Results confirmed the positive correlation observed between HOTAIR expression and MAPK activation, and the negative correlation between MALAT1, TUG1 or UCA1 and pathway activation.

4.8 UCA1 target expression after in vitro silencing

UCA1 transient silencing was achieved through transfection with ASOs; RNA was isolated from treated cells and analysed through Real-Time PCR to evaluate the expression of miRNA and mRNA targets of the lncRNA. Transfection efficiency is shown in table 4.7; only replicates with transfection efficiency higher than 70% were considered for target expression analysis.

<table>
<thead>
<tr>
<th></th>
<th>24h</th>
<th>48h</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTRL1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TREAT1</td>
<td>88.91443</td>
<td>95.28384</td>
</tr>
<tr>
<td>TREAT2</td>
<td>51.28108</td>
<td>79.27341</td>
</tr>
<tr>
<td>TREAT3</td>
<td>28.71566</td>
<td>69.67335</td>
</tr>
<tr>
<td>CTRL2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TREAT1</td>
<td>96.3376</td>
<td>6.129288</td>
</tr>
<tr>
<td>TREAT2</td>
<td>83.90183</td>
<td>3.434258</td>
</tr>
<tr>
<td>TREAT3</td>
<td>76.44554</td>
<td>21.41815</td>
</tr>
<tr>
<td>CTRL3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TREAT1</td>
<td>23.58149</td>
<td>23.58149</td>
</tr>
<tr>
<td>TREAT2</td>
<td>3.434258</td>
<td>2.200903</td>
</tr>
<tr>
<td>TREAT3</td>
<td>27.61162</td>
<td>27.61162</td>
</tr>
</tbody>
</table>

Table 4.7: efficiency of UCA1 ASO transfection in HCT-116 cells at 24 and 48 hours.

Results showed that UCA1 transient silencing induced the upregulation of 3 mRNAs (MACC1, NCL, UGDH) and the downregulation of 1 miRNA (miR-328-3p) (Tables 4.8 and 4.9).
### Table 4.8: expression of UCA1 mRNA targets after transient silencing of the lncRNA. Median fold change and T test p-value are shown for each target.

<table>
<thead>
<tr>
<th></th>
<th>Median fold change</th>
<th>T test p-value</th>
<th>Median fold change</th>
<th>T test p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AREG</td>
<td>0.72</td>
<td>0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASCL2</td>
<td>0.06</td>
<td>0.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LINC01764</td>
<td>0.06</td>
<td>0.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MACC1</td>
<td><strong>1.74</strong></td>
<td><strong>0.04</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCL</td>
<td><strong>2.11</strong></td>
<td><strong>0.02</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PDP1</td>
<td>0.92</td>
<td>0.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SYNCRIP</td>
<td>0.22</td>
<td>0.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UGDH</td>
<td><strong>2.89</strong></td>
<td><strong>0.031</strong></td>
<td></td>
<td>0.87</td>
</tr>
</tbody>
</table>

### Table 4.9: expression of UCA1 miRNA targets after transient silencing of the lncRNA. Median fold change and T test p-value are shown for each target.

<table>
<thead>
<tr>
<th></th>
<th>Median fold change</th>
<th>T test p-value</th>
<th>Median fold change</th>
<th>T test p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-let-7b</td>
<td>0.83</td>
<td>0.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hsa-miR-106a</td>
<td>0.81</td>
<td>0.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hsa-miR-145</td>
<td>0.73</td>
<td>0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hsa-miR-16</td>
<td>0.36</td>
<td>0.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hsa-miR-18a</td>
<td>0.56</td>
<td>0.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hsa-miR-328-3p</td>
<td><strong>-2.91</strong></td>
<td><strong>0.05</strong></td>
<td></td>
<td>0.72</td>
</tr>
<tr>
<td>hsa-miR-339-5p</td>
<td>0.34</td>
<td>0.89</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5. Discussion

Despite their recent discovery, lncRNAs have been intensively studied to understand their functions in cellular physiology and their involvement in pathological processes, including carcinogenesis. The few characterized mechanisms of action suggest that lncRNAs play a crucial role in regulatory cellular networks, modulating fundamental biological processes such as proliferation, gene expression, apoptosis, migration. Less explored but equally fascinating and promising is the field of circRNAs, about which few data are to date available. The characterization of lncRNA and circRNA functions in physiological and pathological conditions may promote the development of new diagnostic and therapeutic approaches.

5.1 lncRNAs contribute to CRC progression

Through Real-Time PCR we identified 7 lncRNAs (CCAT1, CCAT2, HOTAIR, MALAT1, TUG1, UCA1, ZEB2AS1) and 1 circRNA (CDR1AS) differentially expressed in tumor tissues compared to normal mucosa; in particular, we observed the upregulation of CCAT1, CCAT2, HOTAIR, UCA1 and ZEB2AS1, and the downregulation of CDR1AS, MALAT1 and TUG1.

Several studies have previously reported the upregulation of CCAT1 in CRC tissues compared to normal mucosa (223, 258, 259): CCAT1 levels are increased through all disease stages, and also in liver, peritoneal, and lymph node metastasis (223, 258); high levels of CCAT1 promote proliferation, invasion and migration in CRC cell lines (259). CCAT1 is transcribed from a locus near to MYC gene, and it also acts as super-enhancer of this oncogene: the CCAT1-L isoform fosters MYC transcription inducing the formation of a DNA loop that brings MYC promoter near to an enhancer located 335 kb (MYC-335) upstream and to a super-enhancer located 515 kb (MYC-515) upstream of MYC gene (224).
The upregulation of CCAT2 has also been reported by a previous study: higher levels of the IncRNA have been observed in CRC tissues characterized by microsatellite stability compared to CRC tissues with microsatellite instability and normal tissues. CCAT2 is also upregulated in primary tumors of metastatic patients compared to non-metastatic, suggesting a role of CCAT2 in metastatization: CCAT2 upregulation promotes tumor growth, invasion and metastasis both \textit{in vitro} and \textit{in vivo}. The CCAT2 gene is located within the region 8q24, the same of CCAT1 and MYC genes: CCAT2 is retained within the nucleus where it activates MYC transcription, thus regulating its expression and tumorigenic effects. CCAT2-induced expression of MYC could be mediated by TCF7L2, a transcription factor regulating the expression of Wnt target genes, including MYC itself; moreover, CCAT2 itself is a target of Wnt pathway. Therefore, this complex network represents a feedback regulation mechanism (222). CCAT2 upregulation and involvement in tumor progression have also been observed in other tumors, such as breast cancer (260), non-small lung cell carcinoma (261), and esophageal squamous cell carcinoma (262).

Few data are available about CDR1AS in CRC: its function of “molecular sponge” for miR-7 has been demonstrated in brain tissue (80, 91), and miR-7 downregulation has also been associated with CRC (263). However, our results showed the downregulation of CDR1AS in CRC tissues, suggesting a different function of this circRNA in CRC model; CDR1AS downregulation has also been observed in glioblastoma multiforme (264). A recent paper has reported a global downregulation of circRNAs in CRC, except for CDR1AS which showed increased levels (235). Further analyses on larger cohorts will be necessary to fully understand CDR1AS expression and function in CRC.

High HOTAIR levels have been observed in several tumors, including CRC; HOTAIR upregulation has also shown a positive correlation with poor prognosis, EMT and metastasis in CRC patients (229, 265, 266). \textit{In vitro} knockout of HOTAIR affects migration and invasion of CRC cell lines, reduces vimentin and MMP9 metalloprotease expression and increases E-cadherin levels:
these evidences suggest a strong involvement of HOTAIR in metastatization (265). Recently, it has been suggested that HOTAIR contributes to apoptosis block in several tumors: HOTAIR knockdown in vitro reduces miR-125a-5p levels, inducing the consequent increase of one of its target, CASP2 (caspase 2), which in turn triggers apoptosis through the mitochondrial pathway; HOTAIR-mediated regulation of miR-125a-5p levels is based on epigenetic mechanisms modulated by EZH2 (enhancer of zeste 2 polycomb repressive complex 2 subunit) and LSD1, cofactors of the lncRNA in chromatin remodeling (267). A key role of HOTAIR has been reported in cancer stem cells (CSCs): TGF-β treatment of CD133+ CSCs increases HOTAIR expression and triggers EMT (268); plus, HOTAIR knockdown impairs proliferation, migration, invasion and colony forming capacity in CD133+ CSCs (268, 269).

Recent studies have reported the upregulation of MALAT1 in CRC primary tumors compared to normal mucosa and its association with metastasis, suggesting a possible application of MALAT1 as prognostic marker: MALAT1 levels are higher in tumor tissues of metastatic patients compared to non-metastatic and to normal mucosa (270, 271). Our analysis showed a reduced expression of MALAT1 in CRC tissues compared to normal adjacent tissues; similar evidences have been reported in other tumors. MALAT1 downregulation has been observed in glioma compared to normal brain tissues; in vitro silencing of MALAT1 promotes cell proliferation and invasion, increasing MKI67 (marker of proliferation Ki-67) and MMP2 (matrix metallopeptidase 2) expression respectively, while the overexpression of the lncRNA provokes cell cycle block at G0/G1 phase; similar results were obtained through in vivo experiments. Interestingly, MALAT1 upregulation in vitro reduces pERK levels, suppressing MAPK pathway, while low MALAT1 levels activate the same pathway (272), which is crucial in CRC carcinogenesis. MALAT1 downregulation has also been observed in breast cancer: low expression of the lncRNA is associated with axillary lymph node metastasis and shorter relapse-free survival. In vitro knockdown of MALAT1 promotes migration and invasion and triggers EMT through PI3K/AKT
Discussion

pathway (273). Reduced levels of MALAT1 here reported differ from data available in literature about CRC, where MALAT1 upregulation is associated with poor prognosis and metastasis; to date, we know that only 3 out of 20 patients analysed developed lymph node metastases, and this limited number could explain this discordance. Moreover, it is noteworthy that literature reports a very complex scenario about MALAT1 expression in several tumors and that lncRNA mechanisms of action have not been fully understood yet; therefore, the downregulation observed in glioma and breast cancer, where MALAT1 participates in signaling pathways classically associated with colorectal cancer, could occur in CRC too. In addition, our analysis of several CRC GEO DataSets confirmed the downregulation of MALAT1 in CRC tissues compared to normal mucosa, and its association with CRC progression and features. Further studies will be necessary to understand MALAT1 involvement in carcinogenesis of CRC and other tumors.

Similarly, we reported the downregulation of TUG1 in CRC patients, while recent studies have reported the its upregulation and association with poor prognosis, EMT and metastasis in CRC (274, 275). The tumor suppressor activity of TUG1 has been proved in glioma and non-small lung cell cancer, where TUG1 has shown reduced levels. In glioma the downregulation of TUG1 is associated with tumor stage and poor prognosis; TUG1 overexpression in glioma cell lines significantly increases the percentage of apoptotic cells through caspase activation and Bcl-2 downregulation, while TUG1 knockdown promotes cell proliferation (276). Also in non-small lung cell cancer reduced levels of TUG1 are associated with poor prognosis, tumor stage and size; in vitro knockdown of TUG1 increases cell proliferation and colony forming capacity. In addition, TUG1 is a p53 target which preferentially accumulates within the nucleus, where it interacts with PRC2, thus epigenetically modulating the expression of HOXB7 (homeobox B7), a key transcription factor in cell proliferation belonging to HOXB gene cluster (277); HOXB7 promotes cell proliferation through AKT and MAPK pathways, also in CRC (278). Again, our data on TUG1 differs from CRC literature; this discordance could be explained by the limited cohort with a small
Discussion

number of metastatic patients; again, GEO DataSets analysis confirmed the downregulation of TUG1 in CRC tissues compared to normal adjacent mucosa. Therefore, further analyses will be necessary.

The upregulation of UCA1 (urothelial cancer associated 1) here reported agrees with literature data. UCA1 shows higher levels in tumor tissues compared to normal mucosa, and its levels increase with tumor stage; UCA1 expression is also increased in tumor tissues of patients with lymph node metastasis compared to non-metastatic (279). UCA1 upregulation correlates with tumor stage, poor prognosis and metastasis; in vitro knockdown of UCA1 inhibits cell proliferation, invasion and migration and promotes apoptosis in CRC cell lines (280). High levels of UCA1 are also associated with resistance to 5-fluorouracil treatment through the inhibition of apoptosis (281).

No literature data are available about ZEB2AS1 in CRC or in other tumors compared to normal tissues. Recently it has been proposed a mechanism according to which the antisense transcript of ZEB2 gene regulates the translation of ZEB2 mRNA in several cell lines, including CRC cells. In epithelial cells, the main promoter of ZEB2 gene produces a transcript characterized by a 3-kb 5’-UTR region; after the removal of a 2.5-kb intron, only 481 nucleotides of the 5’-UTR region are kept. Within this residual 5’-UTR region lies a sequence that impairs ribosome scanning of the transcript and provokes its release, thus inhibiting ZEB2 mRNA translation. In mesenchymal cells, a second downstream promoter is activated, generating a greater amount of antisense transcript compared to ZEB2 mRNA; the antisense transcript binds the mRNA and masks its splicing sites, thus inhibiting intron removal; within the retained intron lies an internal ribosome entry site (IRES), close to the transcription start site: the antisense transcript thus modulates ZEB2 mRNA splicing, promoting its translation (20) (see section 2.1.2, figure 2.8). Therefore, the upregulation of ZEB2AS1 induces increased levels of Zeb2 protein. ZEB2 upregulation has been observed in primary tumor and metastases of CRC patients, and at invasion front compared to tumor centre; high levels of ZEB2 protein have also been associated with tumor stage and reduced
survival. *In vitro* experiments have proved that ZEB2 silencing reduces migration and invasion in CRC cell lines (282). Several studies have also demonstrated that some ZEB2-targeting miRNAs are downregulated in CRC (283, 284, 285).

It is clear that ncRNAs deregulated in CRC tissues contribute to carcinogenesis in several tumors, including CRC. Gene ontology and pathway enrichment analyses performed on the 8 ncRNAs deregulated in CRC tissues showed the involvement of these molecules and of their miRNA and mRNA targets in numerous processes associated with carcinogenesis, such as cell cycle and apoptosis regulation, adhesion and migration, angiogenesis; ncRNA deregulation also affects pathways classically associated with CRC, such as MAPK, Wnt, EGFR, PI3K, and mismatch repair pathways, and processes regulated by IncRNAs, such as splicing, chromatin remodeling and histone modifications, all associated with tumor progression.

Hence, both literature data and gene ontology analysis suggest a strong involvement of the DE ncRNAs in CRC progression.

### 5.2 NcRNAs in serum exosomes of CRC patients: new diagnostic biomarkers?

Expression analysis of ncRNAs in serum exosomes of CRC patients compared to healthy individuals showed altered expression of 2 IncRNAs (TUG1 and UCA1) and 1 circRNA (circ16); in particular, we observed the upregulation of circ16 and TUG1, and the downregulation of UCA1.

The presence of circ16 in serum exosomes has been previously reported (286); circ16, identified in circBase as hsa_circ_0000284, is a 1099 bp circular transcript produced from the second exon HIPK3 (homeodomain interacting protein kinase 3) gene, which is flanked by two long introns in the linear transcript. HIPK3 gene codes for a 170-kDa protein kinase which phosphorylates FADD (Fas associated via death domain) thus increasing resistance of prostate carcinoma cells to Fas-mediated apoptosis (287). Circ16 expression has been investigated in several cancer cell lines, including CRC cells, showing higher levels compared to the linear HIPK3
transcripts and a typically cytoplasmic localization; it has been demonstrated that circ16 knockdown significantly reduces cell growth in cell lines of several cancer histotypes, including CRC; in addition, this circRNA has Ago2-binding sites and acts as miRNA sponge for some miRNAs acting as tumor suppressor in several tumors (288); among them, miR-124, miR-193a, miR-29b, and miR-338 are downregulated in CRC (http://mircancer.ecu.edu/).

We reported for the first time the presence of TUG1 within exosomes isolated from human serum; previously, high TUG1 levels have been shown in plasma of patients affected by chronic lymphocytic leukemia and multiple myeloma (289), and in serum of multiple sclerosis patients (290); also, it has been reported that TUG1 is upregulated in serum of osteosarcoma patients, and that its levels significantly decrease after surgical resection of the tumor (291). Interestingly, our data showed an opposite deregulation trend in CRC tissues and in serum exosomes for TUG1, which is downregulated within tumor cells and upregulated in serum exosomes: this observation could suggest that tumor cells protect themselves from TUG1 tumor suppressive activity by secreting it through exosomes. Similar results were previously reported by our research group for miRNAs: we observed that exosomes secreted by Caco-2 and HCT-116 cells were enriched in tumor suppressive miRNAs, involved in blocking proliferation and immune escape (170).

Several studies have demonstrated the presence of UCA1 in different biological fluids, such as urine, gastric juice, plasma and serum of oncologic patients, but no one ever reported UCA1 presence within serum exosomes. High levels of UCA1 have been observed in plasma of gastric cancer patients (292) and in serum of hepatocellular carcinoma patients (293). Moreover, a signature of three lncRNAs (UCA1/CUDR, LSINCT-5 and PTENP1) was identified as diagnostic marker in serum of patients affected by gastric cancer compared to healthy individuals (294). UCA1 upregulation has been reported in plasma of CRC patients; the same study has shown that its levels decreased 14 days after tumor resection (279). It is noteworthy that UCA1 showed asymmetric distribution between tumor tissues, where the lncRNA is upregulated, and serum exosomes, where
it is downregulated; a similar observation was reported in this study for TUG1, which showed an opposite asymmetry. In this case, this observation could suggest that the oncogenic function of UCA1 is crucial for tumor progression, inducing tumor cells to retain the lncRNA by limiting its secretion through exosomes.

Hence, literature is enriched of evidences supporting the involvement of these ncRNAs in tumor progression and their possible application as disease biomarkers. We evaluated the diagnostic accuracy of DE ncRNAs through ROC curves: our data showed good accuracy for 2 out of 3 DE ncRNAs (circ16: 71% sensitivity, 80% specificity; UCA1: 100% sensitivity, 43% specificity), with significant p-values. On the contrary, TUG1 ROC curve showed no statistical significance, likely because of the small number of recruited patients. Surely further analyses will be necessary to evaluate the diagnostic accuracy of these molecules as CRC biomarkers.

Gene ontology and pathway enrichment analyses confirmed that ncRNAs deregulated in serum exosomes are involved in cancer-related processes and in pathways associated with CRC onset and progression. It has been shown that exosomes secreted by tumor cells can target several cell histotypes, including immune cells; exosomal pro-cancer effect may be accomplished by ncRNAs, as suggested by their involvement in pathways related to immune escape and suppression of immune response.

5.3 Identification of miRNA and mRNA targets of DE ncRNAs: future perspectives for the understanding of miRNA-ncRNA and mRNA-ncRNA interaction effects

Several evidences have proved that lncRNAs and circRNAs act through a plethora of molecular mechanisms to regulate various and fundamental cellular processes; on the contrary, miRNAs perform their function of post-transcriptional repressors through a unique mechanism, widely studied, which provokes mRNA target degradation or translation block. We wondered what effects could derive from lncRNA/circRNA-miRNA or lncRNA-mRNA interactions. According to
miRNA degradative function on mRNAs, we could hypothesize that the interaction between ncRNAs and miRNAs induces ribonuclease activation, with the consequent degradation of the ncRNA or the miRNA: it is likely that the miRNA, together with RISC complex and Ago proteins, binds the ncRNA as an mRNA target, causing its degradation; alternatively, the ncRNA could decrease miRNA levels sequestering it or inducing its degradation. Effects of lncRNA-mRNA interactions also need to be investigated: it has been demonstrated that this interaction may be involved in splicing regulation, but other effects may occur.

We performed a computational analysis to identify miRNA and mRNA targets of ncRNA deregulated in tumor tissues aiming to evaluate the effects of their interactions: we foresee to perform functional assays to silence lncRNAs or their mRNA targets (by using siRNA) in CRC cell lines aiming to evaluate the effects on mRNA and lncRNA expression respectively; also, the same approach will be used for miRNAs, modulating their levels with miRNA mimics/inhibitor and analysing the effects on lncRNA and circRNA expression.

Interestingly, no miRNA satisfying our selection criteria was identified for ZEB2AS1; this lncRNA acts as splicing regulator for ZEB2 mRNA, modulating ZEB2 protein levels. Therefore, we hypothesized that ZEB2AS1 doesn’t act as miRNA sponge, as suggested by the lack of miRNA-binding sites within its sequence.

Several studies investigated the effect of miRNA-lncRNA/circRNA and mRNA-lncRNA interactions reporting many possible effects, including the degradation of one of the two ncRNAs: miR-9 binds MALAT1 causing its Ago2-mediated degradation (295), as observed for miR-671 and CDR1AS (296); the lncRNA HULC (hepatocellular carcinoma up-regulated long non-coding RNA) sequesters miR-372, reducing its activity and its expression post-transcriptionally (297); HULC is also destabilized by the interaction with IGF2BP1 (insulin like growth factor 2 mRNA binding protein 1), which recruits the cytoplasmic RNA decay machinery (298); GAS5 (growth arrest specific 5) binds c-Myc mRNA impairing its association with polysomes (299), while GHET1
(gastric carcinoma proliferation enhancing transcript 1) stabilizes the interaction between c-Myc mRNA and IGF2BP1, increasing mRNA stability and translation (300); gadd7 (growth-arrested DNA damage-inducible gene 7) interacts with TDP-43 (TAR DNA binding protein), thus causing the degradation of CDK6 (cyclin-dependent kinase 6) mRNA (301).

5.4 Altered expression of ncRNAs after treatment with MAPK inhibitors

MAPK pathway is crucial in CRC onset and progression; it has been observed that CRC patients frequently bare point mutations in KRAS, the G protein acting upstream the pathway, causing its constitutive activation and the consequent iperactivation of signal transduction.

We investigated the effects of MAPK pathway inhibition on DE ncRNA expression; we treated with two inhibitors HCT-116 cells, characterized by mutated KRAS and MAPK iperactivation. We used FR180204, ERK1/2 inhibitor with weak activity on p38α, and U0126, MEK1/2 specific inhibitor.

As showed by MTT assay and Western Blot, both inhibitors reduced cell viability and pERK levels in treated cells compared to control cells. Therefore, both inhibitors showed their efficiency in pathway inhibition.

Expression analysis was performed through Real-Time PCR; results showed that U0126 treatment altered the expression of 4 out of 9 ncRNAs (HOTAIR, MALAT1, TUG1 and UCA1); in particular, MALAT1, TUG1 and UCA1 were upregulated, while HOTAIR was strongly downregulated.

The association between HOTAIR and MAPK pathway has been previously reported in breast cancer: cells treated with SB203580, a specific inhibitor of p38/MAPK14 (mitogen-activated protein kinase 14), showed reduced expression of HOTAIR, suggesting an involvement of p38 in the regulation of its expression (302); our data also showed a strong downregulation of HOTAIR after MAPK inhibition, supporting the same hypothesis. A few evidences suggested that MALAT1
may be involved in MAPK pathway activation, although two different studies reported opposite roles of this lncRNA in glioma and gallbladder cancer cell lines: MALAT1 acts as tumor suppressor in glioma, and its upregulation causes reduced pERK levels, suppressing MAPK pathway (272); on the contrary, in gallbladder cancer, where MALAT1 acts as oncogene, its knockdown inactivates the pathway (303); therefore, despite the connection between MALAT1 and MAPKs seems to be a fact, further analysis will be necessary to understand the molecular mechanism. Plus, our data showed that MAPK inhibition induced increased expression of MALAT1, suggesting the involvement of the pathway in the regulation of MALAT1 expression. TUG1 has been suggested as MAPK activator in non-small cell lung cancer cells: the lncRNA epigenetically regulates HOXB7 expression, which in turn activates AKT and MAPK pathways; conversely, TUG1 knockdown induced increased levels of pERK (277). Our data confirmed the negative correlation between TUG1 and pERK levels, suggesting that MAPKs may regulate TUG1 expression in a feedback mechanism. A recent study reported that UCA1 knockdown significantly reduced pERK levels, but did not affect p38 phosphorylation, in hepatocarcinoma cell lines; UCA1 sequestered miR-216b, increasing FGFR1 (fibroblast growth factor receptor 1) protein levels and thus inducing ERK activation (304). However, our data showed a negative correlation between UCA1 levels and MAPK activation, suggesting that UCA1 expression is not regulated by MAPK pathway; we could speculate that UCA1 significant upregulation after MAPK inhibition is due to increased stability, and thus reduced degradation, of the lncRNA.

Interestingly, U0126 treatment induced an opposite alteration of HOTAIR, MALAT1 and TUG1 expression compared to that observed in CRC tumor tissues: HOTAIR, upregulated in CRC, was downregulated by U0126, while MALAT1 and TUG1, downregulated in tumor tissues, were upregulated after MAPK inhibition; on the contrary, UCA1 upregulation occurred both in tumor tissues compared to normal mucosa and in cells treated with U0126 compared to controls. These data suggest an involvement of HOTAIR, MALAT1 and TUG1 in cell cycle regulation mediated by
MAPKs; instead, UCA1 may perform its oncogenic function through other pathways. Moreover, our data confirmed the anticorrelation reported in literature between MALAT1 or TUG1 knockdown and MAPK pathway activation: the inhibition of the pathway caused increased expression of both lncRNAs, suggesting a feedback regulation mechanism between these lncRNAs and a still unknown member of the MAPK pathway.

5.5 **In vitro silencing of UCA1 affects the expression of its targets**

Our results showed that UCA1 was upregulated in tumor cells and downregulated in serum exosomes of CRC patients; because of this asymmetric distribution, we chose to focus our attention on this lncRNA to perform *in vitro* functional assays. Through ASO technology, we performed UCA1 transient silencing in HCT-116 cells aiming to investigate the effects of lncRNA downregulation on the expression of its targets; among them, we analysed miRNAs and mRNAs characterized by positive or negative correlation of expression with UCA1, retrieved by computational analysis and literature data. Also, we analysed the expression of LINC01764, a ncRNA transcribed from the antisense strand of UCA1 gene, hypothesizing that the two ncRNAs could be involved in a reciprocal mechanism of expression regulation. Results showed the upregulation of 3 mRNA targets (MACC1, NCL, UGDH) and the downregulation of 1 miRNA target (miR-328-3p).

Several studies reported the upregulation of MACC1 (metastasis associated in colon cancer 1) in CRC tissues and its association with metastatization, suggesting its application as a prognostic marker (305, 306, 307). MACC1 acts as transcriptional factor inducing the expression of MET (MET proto-oncogene, receptor tyrosine kinase), the hepatocyte growth factor (HGF) receptor; thus, MACC1 participates in various signal transduction pathways, including MAPK pathway, and regulates proliferation, migration and metastasis in CRC cells (305, 308). NCL (nucleolin) is an abundant non-ribosomal nucleolar protein, crucial in many steps of ribosome biogenesis; it is also
located in nucleus and cytoplasm, and on cell membrane. The localization of NCL within cells is associated with different functions, performed through the interaction with nucleic acids (both RNA and DNA): i) the nuclear NCL pool controls the expression of several oncogenes, genome stability and splicing; ii) within the cytoplasm, NCL binds various mRNAs coding for proteins involved in cell proliferation and apoptosis, thus regulating their translation; among NCL target there is also p53, which translation is inhibited; iii) on cell surface, NCL interacts with several ligands regulating proliferation, apoptosis and angiogenesis, such as Fas, Ras, HGF, VEGF (309, 310). Several studies recently reported NCL deregulation in many tumors, including its upregulation in CRC (311).

UGDH (UDP-glucose 6-dehydrogenase) catalyzes the conversion of uridine diphosphate glucose (UDP-glucose) to uridine diphosphate glucuronic acid (UDP-glucuronic acid), which is a key precursor for the synthesis of glycosaminoglycans, common components of extracellular matrix. Among glycosaminoglycans, hyaluronic acid (HA) is involved in physiological cell growth and migration, but its increased levels have been associated with many tumors; including CRC, where it is an unfavorable prognostic factor (312). The incremented synthesis of HA is induced by UGDH upregulation: indeed, UGDH knockdown affects HA production, resulting in reduced motility of colon cancer cells (313). Among miRNAs, miR-328-3p showed reduced levels in CRC tissues and correlation with patient survival (314, 315). Experiment on CRC side population cells showed that miR-328-3p is critical in the maintenance of cancer stem-like phenotype of SP cells: the upregulation of the miRNA impairs drug resistance and invasion. Among its target, MMP16 (matrix metallopeptidase 16) plays a crucial role in metastatization (315). MACC1, NCL and UGDH all showed increased expression after UCA1 transient silencing; miR-328-3p levels were decreased.

However, it has been recently shown that IncRNA cellular localization affects the silencing efficiency of different methods: indeed, ASOs showed to be more suitable for nuclear IncRNA silencing, while cytoplasmic IncRNAs were more effectively silenced by RNAi (316). For this
reason, we will confirm these preliminary results with further experiments, performed by using siRNAs to achieve transient silencing of UCA1 in HCT-116 cells.
6. Conclusions and future perspectives

The onset and progression of a tumor require mutations causing inactivation or aberrant expression of various molecules; in recent years it has become clear that these molecules are represented not only by proteins, but also by non-coding RNAs. Research on ncRNAs, already advanced in the case of miRNAs but still at the beginning for lncRNAs and circRNAs, proved that altered expression of specific molecules can be associated with one or more tumors, promoting their progression through several molecular mechanisms. To understand these mechanisms and evaluate the possible application of ncRNAs as diagnostic markers or new therapeutic targets it will be necessary to investigate the involvement of these molecules in carcinogenesis.

Our study showed the aberrant expression of 8 ncRNAs (CCAT1, CCAT2, CDR1AS, HOTAIR, MALAT1, TUG1, UCA1 and ZEB2AS1) in CRC tumor tissues compared to normal mucosa and their involvement in cancer-related pathways; our results also suggested a connection between MAPK pathway, altered in CRC, and some of these ncRNAs (HOTAIR, MALAT1, TUG1 and UCA1) after the inhibition of MAPKs in HCT-116 cells by U0126 treatment. These data will be confirmed in the same cells treated with FR180204. We also demonstrated the deregulation of 3 ncRNAs (circ16, TUG1 and UCA1) in exosomes isolated from serum of CRC patients compared to healthy individuals, and evaluated their possible application as diagnostic biomarkers through ROC curve computation. Finally, through ASO-mediated silencing of UCA1 in HCT116, we investigated the effects of ncRNA-miRNA/mRNA interactions, aiming to understand if these interactions trigger degradation mechanisms of any of the interacting molecules. Preliminary results showed that MACC1, NCL, UGDH were upregulated by UCA1 silencing, while miR-328-3p was downregulated.
7. Bibliography


44. He C, Jiang B, Ma J, Li Q. *Aberrant NEAT1 expression is associated with clinical outcome in high grade glioma patients.* *APMIS.* 2016; 124:169-74.


102. Cai X, Hagedorn CH, Cullen BR. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. RNA. 2004; 10:1957-66.


   **EMBO J.** 2006; 25:522-32.
108. MacRae IJ, Ma E, Zhou M, Robinson CV, Doudna JA. In vitro reconstitution of the human RISC-loading complex. 
109. Filipowicz W, Bhattacharyya SN, Sonenberg N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? 
110. Kim VN. MicroRNA biogenesis: coordinated cropping and dicing. 
   **RNA.** 2011; 17:312-26.
114. Ruby JG, Jan CH, Bartel DP. Intronic microRNA precursors that bypass Drosha processing. 
   **Cell.** 2009; 136:642-55.
   **Cell Cycle.** 2015; 14:289-90.
   **J Neurooncol.** 2013 May; 113:1-11.
   **Genes (Basel).** 2013; 4:152-70.


278. Liao WT, Jiang D, Yuan J, Cui YM, Shi XW, Chen CM, Bian XW, Deng YJ, Ding YQ. HOXB7 as a prognostic factor and mediator of colorectal cancer progression. **Clin Cancer Res.** 2011; 17:3569-78.


