



Skolkovo Institute of Science and Technology
Skolkovo Institute of Science and Technology

FUNCTIONAL STUDY OF HUMAN AND MURINE MORRBID LNCRNA *IN VITRO*

Doctoral Thesis

By

ANNA FEFILOVA

DOCTORAL PROGRAM IN LIFE SCIENCES

Supervisor
Associate Professor Timofei Zatsepin

Moscow – 2020

© Anna Fefilova, 2020

I hereby declare that the work presented in this thesis was carried out by myself at Skolkovo Institute of Science and Technology, Moscow, except where due acknowledgment is made and has not been submitted for any other degree.

Candidate (Anna Fefilova)

Supervisor (Associate Professor Timofei Zatsepin)



Skolkovo Institute of Science and Technology
Skolkovo Institute of Science and Technology

THESIS DEFENSE COMMITTEE

Professor Yuri Kotelevtsev, PhD, Skoltech, Russia

Dmitri Pervouchine, PhD, Skoltech, Russia

Professor Konstantin Lukyanov, PhD, Skoltech, Russia

Pavel Ivanov, PhD, Harvard Medical School, U.S.A.

Rory Johnson, PhD, University of Bern, Switzerland

Moscow – 2020

© Anna Fefilova, 2020

Abstract

Long non-coding RNA (lncRNA) are RNA transcripts longer than 200 nt that participate in various cellular processes, such as chromatin remodeling, transcription, splicing, protein translation, etc. However, approximately 99% of annotated lncRNAs have not been functionally characterized. Some lncRNA have several transcripts, which may possess tissue specificity or have different functions that also may depend on external and internal conditions. Many lncRNAs participate in the development of various diseases, including cardiovascular, neurological, autoimmune, and cancers. Recently over three dozen papers reported a strong association of MIR4435-2HG (human Morrbrid or hMorrbrid) and CYTOR (LINC00152) long non-coding RNAs upregulation with the progression of multiple cancers. Also, the murine ortholog of these lncRNAs was proposed – mMorrbrid that contributes to adaptive immunity and antiviral response.

In this study, we evaluated the roles of hMorrbrid/CYTOR and mMorrbrid lncRNAs in hepatocyte cell lines *in vitro*. hMorrbrid and CYTOR are human paralogous genes located at the opposite arms of the second chromosome. We generated a cell line with the deletion of parts of both hMorrbrid and CYTOR paralogues, characterized its phenotype, and demonstrated that disruption of these lncRNA genes leads to small alterations in cell cycle progression and cell migration. Also, we overexpressed the hMorrbrid/CYTOR transcript (M-217) with the conserved region (exonCh) in the knockout cell line and observed a delay in proliferation and increase in apoptosis. Analysis of pro-survival and pro-apoptotic proteins in the M-217 expressing cells revealed elevated levels of pro-apoptotic proteins. Treatment of the M-217 expressing cells with MCL1 inhibitor S-63845 additionally shifted the ratio between pro-survival and pro-apoptotic proteins and confirmed an increased apoptosis sensitivity of the M-217 expressing cells. Thus, we found that the deletion of hMorrbrid and CYTOR lncRNAs did not lead to the activation of the mitochondrial apoptosis pathway, while overexpression recovery of the M-217 transcript primed liver cells to apoptosis.

Transient depletion of the murine ortholog of Morrbrid lncRNA in hepatocytes leads to slight upregulation of Bim cellular levels without apoptosis increase. To uncover alternative roles of mMorrbrid lncRNA in normal hepatocytes, we performed RNA-seq analysis of Morrbrid

depleted cells. We found that mMorrbid participates in the regulation of proto-oncogene NRAS mRNA splicing, including the formation of the isoform with a premature termination codon (PTC). The depletion of murine Morrbrid lncRNA led to a significant increase of the NRAS isoform with PTC in hepatocytes. We found that the NRAS isoform with PTC is degraded via the NMD pathway. By a modified capture hybridization (CHART) analysis of the protein targets, we uncovered interactions of Morrbrid lncRNA with the SFPQ-NONO splicing complex. Finally, we propose the regulation mechanism of NRAS splicing in murine hepatocytes by alternative splicing coupled with the NMD pathway with the input of Morrbrid lncRNA.

These findings clarified moonlight functions of human and murine Morrbrid lncRNAs in hepatocytes *in vitro*.

Publications

1. **Anna Fefilova**, Pavel Melnikov, Tatiana Prikazchikova, Tatiana Abakumova, Ilya Kurochkin, Pavel V. Mazin, Rustam Ziganshin, Olga Sergeeva, Timofei S. Zatsepin. Murine Long Noncoding RNA Morrbrid Contributes in the Regulation of NRAS Splicing in Hepatocytes in Vitro. *International Journal of Molecular Sciences*. 2020; 21(16):E5605, DOI: 10.3390/ijms21165605.
2. Elena M. Smekalova, Maxim V. Gerashchenko, Patrick O'Connor, Charles A. Whittaker, Kevin J. Kauffman, **Anna S. Fefilova**, Timofei S. Zatsepin, Roman L. Bogorad, Pavel V. Baranov, Robert Langer, Vadim N. Gladyshev, Daniel G. Anderson, Victor Koteliansky. *In Vivo* RNAi-Mediated eIF3m Knockdown Affects Ribosome Biogenesis and Transcription but Has Limited Impact on mRNA-Specific Translation. *Molecular Therapy: Nucleic Acids*. 2020; 19:252-266, DOI: 10.1016/j.omtn.2019.11.009

Conferences

1. Functional study of human and murine Morrbrid lncRNA *in vitro*. **Fefilova A.**, Sergeeva O., Mazin P., Kireev I., Zatsepin T.; EMBL Symposia "The Non-Coding Genome", Heidelberg, Germany, 16-19 October 2019 (Poster).
2. Morrbrid lncRNA is related to pre-mRNA processing in hepatocytes. **A. Fefilova**, O. Sergeeva, M. Nazarova, P. Melnikov, I. Kireev, T. Zatsepin; 44th FEBS (Federation of European Biochemical Societies), Krakow, Poland, 6-11 July 2019 (Short talk).

Acknowledgments

First and foremost, I would like to express my deepest appreciation to my supervisor Timofei S. Zatsepin. I truly value the opportunity to conduct this research in his lab, his guidance and advice, our many valuable scientific discussions, and for the time he devoted to helping me grow and progress through my PhD journey.

I would like to extend my deepest gratitude to Olga Sergeeva for teaching and mentoring me at every single research step during the last three years. I am thankful for all of the help troubleshooting experiments, for our scientific brainstorming, for advice and ideas, and for being a great example from whom to learn.

I also want to thank all the thesis defense committee members Dr. Pavel Ivanov, Dr. Rory Johnson, Dr. Dmitri Pervouchine, Prof. Konstantin Lukyanov, and Prof. Yuri Kotelevtsev for taking the time to read and give a critical review of this manuscript.

I want to acknowledge my lab members Olga Burenina, Tatiana Prikazchikova, Mikhail Nesterchuk, Tatiana Abakumova, Svetlana Dukova, Katerina Nagornova, Natalia Logvina, and Denis Melnik. It has been a pleasure to work with this group of talented researchers. I would like to extend special thanks as well to Evgenia Klimova, for being the greatest lab mate and friend both in Moscow and overseas, to Dominique Leboeuf, for your kind friendship and for sharing with me your vast research experience, and to Maria Nazarova, who joined this research work for a year and was incredible help with the hMorrbid/CYTOR project as well as a wonderful roommate.

I want to thank my first Skoltech supervisor, Prof. Victor Kotelianski, for opening doors for me during my PhD and for providing me with so many research opportunities. I am greatly indebted to Daniel Anderson and Arturo Vegas and members of Anderson and Vegas laboratories for giving me a chance to perform my research work with them. It was my great honor to work in these laboratories and to learn from the incredible people I have met there. I want to address special thanks to Elena Smekalova, who was my mentor during the very first research steps of my PhD and who taught me cell culture and animal work.

I am particularly grateful to Prof. Yuri Kotelevtsev for always being open to discuss my research projects and always instilling me with encouragement, inspiration, and motivation.

I want to acknowledge the Skoltech Life Sciences team for the immense work they are doing and for always being there to help and answer my questions.

I thank my parents for tolerating my choice of doing a PhD and trying to be as supportive as they can despite never quite understanding why and exactly what it is that I am doing.

And last but not least, with all of my heart, I am thankful to this group of people: Adam Damiano, Natalia Glazkova, Olga Chervinskaia, Kirill Abrosimov, Aysylu Askarova, Dina Bek, Maria Goncharova, Anna Moroz, Maxim Zakharkin, Kristina Zakharkina, and Alexander Vaniev. Thank you all for always being there for me, encouraging me, sharing with me laughter, joy, and sadness, and keeping me sane during the hard times. Your love and support were precious, and I could never do it without you.

Table of Contents

Abstract	4
Publications	6
Acknowledgments.....	7
Table of Contents.....	9
List of symbols, abbreviations.....	14
List of Figures	17
List of Tables	20
Chapter I. Introduction	21
Chapter II. Literature Review.....	25
2.1 Discovery of lncRNAs.....	25
2.2 LncRNA biological diversity	26
2.2.1. LncRNA classification based on relative position towards PCGs.....	27
2.2.2. LncRNA stability and localization.....	31
2.2.3. Coding potential of lncRNA.....	32
2.3 LncRNA functional diversity.....	32
2.3.1 In cis regulation by lncRNAs	33
2.3.2. In trans regulation by lncRNAs.....	38
2.4. LncRNAs in cancers and other pathologies.....	46

2.4.1. LncRNAs in hepatocellular carcinoma.....	49
2.5. LncRNA tissue specificity.....	51
2.5.1 LncRNA isoforms may exhibit functionally distinct and tissue-specific properties.....	52
2.6 Evolution and inter-species homology of lncRNAs	55
2.6.1. Conservation of nucleotide sequence.....	56
2.6.2. Conservation of gene structure.....	58
2.6.3 Positional conservation of lncRNAs	59
2.6.4 Functional homology of lncRNAs	59
2.7 Morrbid lncRNA possess positional and partial sequence conservation between human and mouse.....	61
2.7.1 Morrbid lncRNA positional conservation between human and mouse	62
2.7.2 Morrbid sequence conservation between human and mouse	64
2.8 Human Morrbid has a paralogue lncRNA named CYTOR	66
2.9 Human Morrbid and CYTOR lncRNAs are strongly associated with the progression of various cancers.....	69
2.10 Human Morrbid and CYTOR lncRNAs are important for liver cancer development and proposed as HCC biomarkers.....	72
2.11 Human Morrbid and CYTOR lncRNAs were demonstrated to participate in a plethora of cellular networks.....	74

2.11.1 Regulation of cell cycle by hMorrbid/CYTOR	75
2.11.2. Regulation of PI3k/AKT1 Pathway	80
2.11.3 Regulation of Wnt Pathway.....	81
2.11.4 Regulation of NF-kb pathway	82
2.11.5 hMorrbid/CYTOR in hypoxia	82
2.11.6 TGFb Pathway.....	82
2.11.7 Involvement in apoptosis	83
2.12 Evidence that Human Morrbrid may exhibit transcript-specific functions	84
2.13 In mouse Morrbrid lncRNA regulates inflammation response in immune cells.....	85
Chapter III. Materials and Methods	88
Chapter IV. Results. Characterization of hMorrbrid/CYTOR knockout and investigation of exonCh functional significance	109
4.1 Generation of hMorrbrid/CYTOR CRISPR-Cas9 knockout cell line	110
4.2 hMorrbrid/CYTOR is dispensable for hepatocellular carcinoma cells proliferation, migration, and apoptosis resistance	113
4.3 Proteomic changes in response to hMorrbrid/CYTOR knockout	115
4.4 Previously reported downstream target genes of hMorrbrid/CYTOR mostly remained unaffected in the knockout.....	117
4.5 Overexpression of hMorrbrid transcript in knockout cell line resulted in the delayed proliferation and increased sensitivity to apoptosis.	119

4.5.1 MIR4435-2HG (M-217) is a transcript of the hMorrbid gene containing evolutionary conserved region exonCh.....	119
4.5.2. Generation of the M-217 and mutated M-217 cell lines.....	122
4.5.3. M-217oe cell line but not R-217oe cell line had impaired cell growth and increased apoptosis	123
4.6 hMorrbid/CYTOR KO and M-217oe cells demonstrate differential shifts in expression of pro-apoptotic and pro-survival BCL2 family proteins	125
4.7 Expression of exonCh primes cells for apoptosis and increases cells sensitivity to MCL1 inhibition.....	128
Chapter V. Results. Functional roles of Morrbid lncRNA in murine hepatocytes.....	133
5.1 Inhibition of Morrbid lncRNA decreases viability and migration of normal but not cancerous murine hepatocytes.....	134
5.2 Inhibition of mMorrbid enhances expression of pro-apoptotic protein Bim	137
5.3 mMorrbid KD leads to differential changes in the genes expression related to apoptosis and cell cycle progression at transcriptome and proteome levels.	140
5.4 mMorrbid contributes to the regulation of NRAS oncogene alternative splicing in murine hepatocytes.....	141
5.4.1 NRAS PTC transcript undergoes NMD decay in the cytoplasm	144
5.4.2 KRAS and HRAS members of Ras family are increased in response to mMorrbid KD	146
5.4.3 Morrbid lncRNA interacts with the SFPQ-NONO heterodimer	147

VI. Discussion.	151
6.1 hMorrbid/CYTOR genetic knockout is compatible with normal viability and growth of hepatocellular carcinoma cells.....	151
6.2 Evolutionary conserved fragment exonCh has a role in hepatocellular carcinoma cells priming for apoptosis.....	159
6.3 In murine hepatocytes Morrbrid is involved in the regulation of NRAS alternative splicing coupled with the NMD pathway	165
6.4 A question of whether Morrbrid gene orthologs in mouse and human are functional homologs remains open.....	172
Conclusions	175
Bibliography	177
Supplementary Figures	211
Supplementary Tables	215

List of symbols, abbreviations

AS – alternative splicing

ASO – antisense oligonucleotide

CHART – capture hybridization analysis of RNA targets

ceRNA – competing endogenous RNA

ciRNA – circular intronic RNA derived from spliced intron lariats

circRNA – circular lncRNA produced by back-splicing

CRISPR – clustered regularly interspaced short palindromic repeats

EMT – epithelial-mesenchymal transition

eRNAs – enhancer RNA

FISH – fluorescence in situ hybridization

FITC - fluorescein isothiocyanate

gRNA – guide RNA

HCC – hepatocellular carcinoma

hMorrbid – human lncRNA Morrbid

IRG – immune response genes

KD – knockdown

KO – knockout

LC-MS – liquid chromatography – mass spectrometry

LOF – loss-of-function

lncRNA – long non-coding RNA

lincRNA – long intergenic RNA

LUC – luciferase

M-217oe – cell line overexpressing transcript MIR4435-2HG-217

miRNA – microRNA

mMorrbid – murine lncRNA Morrbid

mRNA – messenger RNA

NAT – natural antisense transcript

NGS – next generation sequencing

NMD – nonsense-mediated mRNA decay

PAGE – polyacrylamide gel electrophoresis

PCG – protein coding gene

PCR – polymerase chain reaction

PI – propidium iodine

PROMPT – promoter upstream transcript

PTC – premature termination codon

R-217oe – cell line overexpressing mutated transcript MIR4435-2HG-217

RAP – RNA antisense purification

RNA - ribonucleic acid

rRNA – ribosomal RNA

RNAi – RNA interference

RIP - RNA immunoprecipitation

RT– room temperature

RT-qPCR – reverse-transcription quantitative polymerase chain reaction

RUST – regulated unproductive splicing and translation

s.d. – standard deviation

snRNA – small nuclear RNA

snoRNA – small nucleolar RNA

sno-lncRNA – snoRNA-ended lncRNA

SPA – 5' snoRNA-ended and 3'-polyadenylated lncRNA

scaRNAs – small Cajal body-specific RNA

siRNA – small interfering RNA

tRNA – transport RNA

TSS – transcription start site

WT – wild-type

WB – western blot

List of Figures

Figure 1. The abundance of protein-coding and non-coding genes in the human genome.-----	27
Figure 2. Classification of lncRNA transcripts based on their position relative to the PCGs.----	28
Figure 3. Production of sno-lncRNAs, ciRNA and circRNAs. -----	30
Figure 4. In cis and in trans mechanisms of lncRNA action.-----	33
Figure 5. Mechanisms of lncRNA regulation in cis. -----	35
Figure 6. lncRNA HOTAIR serves as a scaffold for repressive chromatin remodeling complexes to inhibit the transcription of its target genes.-----	39
Figure 7. Organization of nuclear architecture by lncRNAs.-----	42
Figure 8. Circular RNA CDR1as. -----	43
Figure 9. Examples of alternative splicing regulation by lncRNAs. -----	45
Figure 10. lncRNA regulators of various hallmarks of cancer. -----	48
Figure 11. Alternative splicing of MALAT1 generates a longer nuclear speckles-retained transcript and a shorter tRNA-like cytoplasmic isoform. -----	53
Figure 12. Human and mouse RNA-seq tracks at Morrbid locus-----	62
Figure 13. Schematic representation of genomic environments of murine and human Morrbid	63
Figure 14. Comparison of Morrbid transcripts between human and in the mouse.-----	65
Figure 15. Genomic location of CYTOR and hMorrbid on chromosome 2. -----	66
Figure 16. Alignment of CYTOR (red) and hMorrbid (blue) genomic sequences. -----	67
Figure 17. Xenograft tumor in nude mice injected subcutaneously with HCCLM3 cells inhibited in hMorrbid/CYTOR shRNA grow slower than control-----	74
Figure 18. In HCC hMorrbid/CYTOR sponges miR-193b to promote expression of CyclinD1.	77

Figure 19. hMorrbid/CYTOR interacts with EpCam promoter to drive its expression and subsequently expression of its target genes.-----	78
Figure 20. hMorrbid/CYTOR regulates p21 and p15.-----	79
Figure 21. hMorrbid/CYTOR promotes expression of MCL1 by sponging miR-193a-3p in gastric cancer and miR-125b in ovarian cancer.-----	83
Figure 22. Mechanism of Morrbrid regulation of Bim expression in mice. -----	87
Figure 23. RT-qPCR analysis of hMorrbrid and CYTOR transcripts expression in Huh7 cells.	110
Figure 24. Design and validation of hMorrbrid/CYTOR knockout. -----	112
Figure 25. Characterization of hMorrbrid/CYTOR knockout in Huh7 cells.-----	115
Figure 26. Summary of LC-MS analysis of genes deferentially expressed in cells with hMorrbrid/CYTOR depletion. -----	116
Figure 27. Analysis of hMorrbrid/CYTOR literature targets in knockout cells. -----	118
Figure 28. Establishment of cell lines overexpressing transcripts M-217 and R-217. -----	122
Figure 29. Characterization of M-217 overexpression.-----	124
Figure 30. Mechanism of mitochondria-mediated apoptosis -----	125
Figure 31. Analysis of expression of pro-apoptotic and pro-survival BCL2 family members in WT, KO, M-217oe, and R-217oe cells. -----	127
Figure 32. Analysis of expression of pro-apoptotic protein Bim. -----	128
Figure 33. Schematic representation of the strategy used in the study to characterize M-217oe cells priming for apoptosis.-----	129
Figure 34. Flow cytometry analysis of AnnexinV-FITC/PI stained cells after MCL1 or BCL2/BCL-xL/BCL-W inhibition. -----	132
Figure 35. mMorrbrid cellular localization and LOF phenotype.-----	136

Figure 36. Analysis of mitochondrial apoptosis in mMorrbid depleted cells.-----	139
Figure 37. Summary of RNA-seq and LC-MS analyzes of genes deferentially expressed in cells with mMorrbid depletion. -----	140
Figure 38. mMorrbid downregulation leads to changes in alternative splicing of NRAS proto-oncogene. -----	143
Figure 39. NRAS PTC transcript is degraded in the cytoplasm via non-sense mediated decay. -----	145
Figure 40. mMorrbid directly interacts with SFPQ-NONO heterodimer to modulate AS of NRAS PTC isoform.-----	149
Figure 41. Effect of M-217/exonCh expression on the intrinsic apoptosis pathway. -----	164
Figure 42. Proposed mechanism of the Morrbid lncRNA contribution in the regulation of NRAS mRNA alternative splicing. -----	172

List of Tables

Table 1. Examples of lncRNA in cis and in trans regulators. -----	37
Table 2. Examples of lncRNAs dysregulated in HCC. -----	50
Table 3. Alternative names of hMorrbid/CYTOR lncRNAs used in literature.-----	67
Table 4. hMorrbid/CYTOR roles in cancers.-----	70
Table 5. Regulation of cellular processes by hMorrbid/CYTOR. -----	76
Table 6. gRNAs used for Cas9-mediated knockout of hMorrbid/CYTOR genes.-----	105
Table 7. List of primers used for PCR and cloning.-----	105

Chapter I. Introduction

During the last decades, advances in high-throughput next-generation sequencing (NGS) have allowed for the identification of thousands of previously unknown long non-coding RNA (lncRNA) genes [1,2]. These discoveries reshaped our understanding of the central dogma of molecular biology DNA \leftrightarrow RNA \rightarrow protein, adding a complex network of non-coding regulators on the scheme. Moreover, a thorough analysis of a total of 1,627 prokaryotic and 153 eukaryotic genomes demonstrated that the proportion of the non-coding DNA increases in more complex multicellular organisms, while protein-coding genes (PCG) show little variation across animal lineages [3]. This expansion of non-coding RNAs in higher organisms suggests their impact on biological complexity.

Long non-coding RNAs are distinguished from the other non-coding RNAs by length longer than 200 nt. To date, lncRNAs have been identified in all species studied including eukaryotes, plants, bacteria, and viruses. LncRNAs were found to participate in the regulation of various cellular processes. They can interact with DNA, RNA or proteins modulating transcription, chromatin remodeling, formation, targeting, and stabilization of functional ribonucleoproteins [4]. In addition to that, lncRNAs were linked to diseases such as cardiovascular diseases, neurological diseases, diabetes, autoimmune diseases, and cancer [5,6]. In human cells, the number of lncRNA genes is significant and potentially exceeds the number of PCGs. Recent estimates suggest more than 50 thousand lncRNA genes [7] with >15 thousand annotated genes [8]. At the same time, only ~1% of lncRNAs have been studied and functionally characterized. Thus, lncRNAs represent an abundant part of the expressed transcriptome produced in parallel with protein-coding sequences, that is still insufficiently studied. Extensive research of

the last decade suggests their role in controlling higher-order regulatory events within cells, enabling for a more rapid and fine-tuned interactions between regulators of gene expression and signaling pathways. Deeper investigation and functional characterization of lncRNA is essential for understanding the full complexity of cellular processes and diseases.

In this work, we performed a functional study of orthologue long non-coding genes human *Morrbid* (h*Morrbid*) and murine *Morrbid* (m*Morrbid*) in hepatocytes. In human cells, *Morrbid* has a paralogous gene named *CYTOR*. Paralogs h*Morrbid* and *CYTOR* have been intensively studied and were reported to promote cell proliferation, migration, and invasiveness of gastric cancer [9–13], hepatocellular carcinoma [14–19], lung cancer [20–25], colon cancer [26], gallbladder cancer [27], glioma [28], ovarian sarcoma [29,30], pancreatic cancer [31], clear renal cells sarcoma [32,33], breast cancer [34–37], etc. Literature suggests that these lncRNAs are involved in regulation of genes via a variety of mechanisms, like epigenetic gene regulation (e.g. *EpCAM* [16], *p15*, *p21* [38], *p16* [33], *PI3KCA* [19]), post-transcriptional modifications (β -catenin [39]), regulation of mRNA translation (*SNAIL1* [19]), etc. An attempt to characterize genetic knockout of h*Morrbid*/*CYTOR* genes has not been done previously. Inconsistencies between loss-of-function (LOF) phenotypes discovered using RNAi and CRISPR approaches were reported for several lncRNAs (e.g. *MALAT1* [40], *lincRNA-p21* [41], etc.). These inconsistencies highlight the importance of performing complementary functional studies using different research tools. In this research work, we developed a knockout system to study LOF phenotype of these cancer-related lncRNAs in human hepatocellular carcinoma cells.

lncRNA genes gain and lose functional properties throughout evolution at a much higher rate than protein-coding genes (PCG) and demonstrate rapid turnover of their primary nucleotide sequence. It was suggested that in vertebrates more than 70% of lncRNA evolved less than 50

million years ago [42-44]. Even widely conserved between species and essential for cell viability lncRNAs like Xist demonstrated poor inter-species sequence homology, usually bound to short sequence regions within exons [42]. It has also been suggested that such conserved regions may represent functional units essential for lncRNA activity [45]. One example is a highly conserved region of lncRNA LINC-PINT responsible for interaction with PCR2 [46]. This view is consistent with the hypothesis of the modular structure of lncRNAs, which suggests that lncRNAs contain discrete sequence domains responsible for functional properties [47, 48]. Such domains are relatively short compared to the entire length of the transcript and the rest of the sequence might be of little importance. For example, only one-tenth of the RoX1 RNA nucleotide sequence is responsible for sex dosage compensation in *Drosophila* [49]. Similar short patches of conserved nucleotide sequence were found in several exons of human and murine Morrbid lncRNAs. hMorrbid isoform MIR44352HG-217 (M-217) contains a region of ~300nt (exonCh) alignable to the corresponding exon in mice (exonCm). To investigate if sequence conservation of exonCh indicates functional activity, we used hMorrbid/CYTOR knockout system to study M-217 functions in the absence of other transcripts. To clarify the role of the conserved exonCh we overexpressed M-217 and mutated M-217 with reversed exonCh in hMorrbid/CYTOR knockout cells.

In mice, Morrbid was reported to be essential for the viability of immune cells [50,51]. Morrbid knockout mice have reduced levels of eosinophils, neutrophils, and Ly6Chi classical monocytes in the peripheral blood and tissues [51]. In the mouse genome, Morrbid is located close to the proapoptotic gene *Bim* and functions *in cis* to regulate *Bim* promoter through the recruitment of the polycomb repressive complex 2

(PRC2) [51]. In preleukemic Hematopoietic Stem and Progenitor Cells (HSPC) depleted of the Tet2 gene, murine Morrbid is a key component of the signaling pathway promoting the enhanced survival and reduced apoptosis of Tet2-KO cells [52]. Despite the apparent inhibitory role of Morrbid in the regulation of Bim in myeloid cells and HSPC, in CD8+ T cells, Morrbid was shown to be essential for the upregulation of Bim expression in response to Lymphocytic Choriomeningitis Virus (LCMV) infection signaling [50]. Despite a thorough characterization of murine Morrbid lncRNA in myeloid cells, nothing is known about its functions in other cell types and tissues. Many lncRNA demonstrated tissue-specific expression and functions due to varied cell-specific regulatory networks [53]. Thus, functions of lncRNA murine Morrbid may significantly differ between immune cells and hepatocytes. We identified mMorrbid as a mostly nuclear lncRNA and used antisense oligonucleotides to downregulate its expression in normal and cancerous hepatocytes and study its functional roles in liver cells.

The main objectives of this research project are:

- Phenotype characterization of hepatocellular carcinoma cells depleted in hMorrbid/CYTOR lncRNAs expression using CRISPR/Cas9 knockout system
- Functional study of the evolutionarily conserved fragment exonCh of Morrbid lncRNA in hepatocellular carcinoma *in vitro*
- Investigation of murine Morrbid lncRNA functions in hepatocytes *in vitro*

Chapter II. Literature Review

2.1 Discovery of lncRNAs

Over 100 years stand between the beginning of nucleic acid research and identification of non-coding transcripts as functional units essential for cell viability. First regulatory non-coding RNAs were identified and described in bacteria and belonged to a class of molecules, which will be later called small (s)RNAs, regulatory ncRNA in bacteria. The first lncRNA found in eukaryotes was H19 in mice [54]. This gene was abundantly expressed, transcribed by polymerase II, polyadenylated, spliced, important for embryonic development, however, was found to lack any translation [55]. Another lncRNA that had been described before complete sequencing of the human genome took place, and is probably the most well-known lncRNA, was Xist (X-inactive specific transcript), which is responsible for the inactivation for the entire X chromosome (Xi) in the female cells.

The revolution in the perception of non-coding transcripts happened in 2001 when the complete human genome sequence was published as a result of the Human Genome Project (HGP) [56,57]. Published data showed that the human genome contains only 1.2% of protein-coding regions, while the rest was called the “non-coding” part. The HGP, together with FANTOM, RIKEN and ENCODE consortia, also demonstrated that, even though non-protein-coding, this DNA is actively transcribed [58,59]. The HGP triggered debates on the biological relevance of non-coding transcripts. Even though there was a lot of skepticism about ncRNAs functional relevance. For example, one study reported that the deletion of 1.5 Mb and 0.8 Mb long genomic regions, containing over a

thousand evolutionarily conserved non-coding sequences resulted in viable mice [60]. Many scientists saw this transcriptional diversity of non-coding transcripts as a potential higher-level regulatory network of the cell and also a key to an explanation for phenotypic diversity between organisms. Even though most of the lncRNAs are poorly conserved, 3% of them are conservative and can be found from *Xenopus* to man [61]. Currently, ~15 thousand lncRNA transcripts were annotated in the human genome [8], the majority of them are capped, polyadenylated, and spliced in the same manner as mRNAs (Figure 1). Highly conserved lncRNAs are believed to be involved in biological processes shared by multiple species, such as embryonic development. On the other hand, poorly conserved transcripts were suggested to drive phenotypic and functional variations at individual and interspecies levels [62]. Intensive work in the field of lncRNA research resulted in the discovery of the multitude of regulatory functions performed by these genes. And despite the initial hypothesis of ‘transcriptomic noise’, the non-coding RNAs were later demonstrated to perform a variety of master regulatory functions within the cell, including epigenetics, biological inheritance, gene regulation, cell protection, etc.

2.2 LncRNA biological diversity

Non-coding RNA is a large group of non-protein-coding transcripts of various lengths encoded in the genome and pervasively transcribed. Based on the length, non-coding transcripts are classified as small (<200 nt) and long (>200 nt) [63]. Small non-coding transcripts include tRNA involved in translation; snoRNAs (small nucleolar RNAs) responsible for chemical modifications of other RNAs in the cell (like rRNA, tRNA, snRNA); snRNA (small nuclear) mainly involved in splicing; miRNAs, siRNAs, and PIWI-interacting RNAs responsible for gene silencing; scaRNAs (small Cajal body-specific RNAs) involved in

snoRNAs posttranslational modification; transcription initiation RNAs, etc [64]. Together small non-coding and long non-coding transcripts constitute the most abundant part of the transcribed genome (Figure 1).

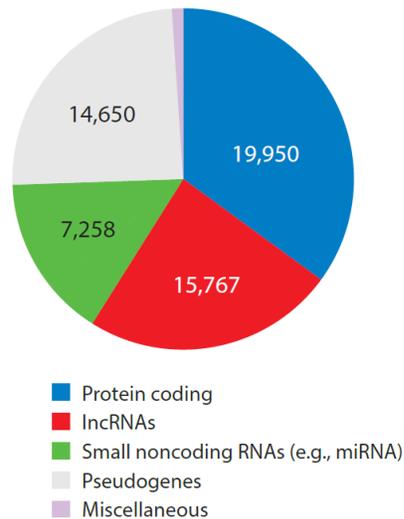


Figure 1. The abundance of protein-coding and non-coding genes in the human genome (in brackets number of genes) Schematic representation of gene types annotated by GENCODE (version 25) [65] © 2017 by Annual Reviews.

Long non-coding RNA differ from other non-coding transcripts based on their length, which can range from 200nt to up to 100kbs. On average lncRNAs are expressed at lower levels than protein-coding genes, nevertheless, most of them share common features with mRNAs: transcription by RNA polymerase II, polyadenylation, 5' prime cap, and splicing [61].

2.2.1. lncRNA classification based on relative position towards PCGs

lncRNA genes can overlap or not overlap protein-coding sequences in the

genome. Not overlapping lncRNAs are located from several to >3Mb away from the nearest PCG, with an average of 40kbs and ~28% located less than 10kb away [66]. If there is no intersection with protein-coding genes, the lncRNA is called long intergenic RNA (lincRNA), otherwise, lncRNA is called intragenic. Intragenic lncRNA genes are further classified as overlapping sense, intronic, antisense, and bidirectional (Figure 2).

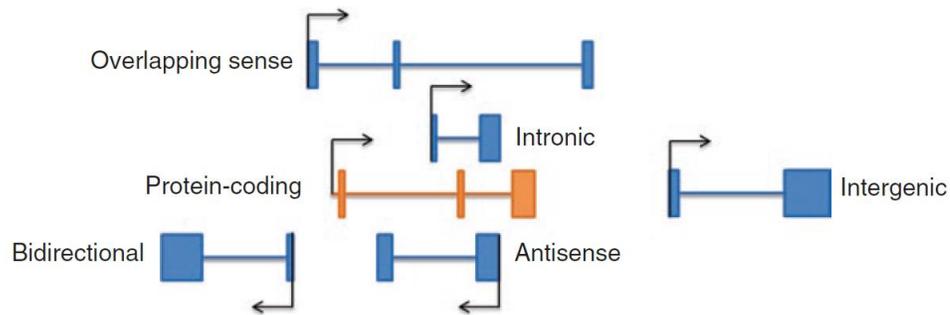


Figure 2. Classification of lncRNA transcripts based on their position relative to the protein-coding gene [62] © 2017, Springer Nature Singapore Pte Ltd.

Antisense lncRNAs are transcribed from the antisense strand and overlap paired PCG. A subgroup of antisense transcripts are **NATs** – natural antisense transcripts, which are transcribed from the same gene locus, but in the opposite direction from paired PCGs. NATs may regulate expression of genes they overlap (cis-NAT), or genes transcribed from other genomic locations (trans-NAT). Due to the complementarity rule, these RNAs may form RNA-RNA interactions with a pair mRNA thereby regulating its expression and stability [62]. NATs containing inverted short interspersed nuclear element B2 (SINEB2) and are called **SINEUPs** [67]. This lncRNA subgroup can promote the translation of paired mRNAs. First discovered SINEUP was an antisense **lncRNA AS Uchl1**, which was found to increase expression of its complementary gene UCHL1, a protein involved in proper brain function, in

a post-transcriptional manner [68].

Most eukaryotic promoters are bidirectional and initiate Polymerase II transcription in two opposite directions (divergent transcription), which leads to the production of many **bidirectional lncRNA**. Bidirectional lncRNAs are transcribed from the antisense strand immediately upstream of PCG promoter and in some cases partially overlap it. **eRNAs** (enhancer RNA) are bidirectionally transcribed from enhancers and believed to control the activity of the corresponding enhancer and in some cases promoter, creating a higher-order chromatin organization at the site of their transcription. Generated via bidirectional transcription eRNA transcripts seem to be irrelevant for function and are quickly degraded by exosomes [69]. Another example of short-lived lncRNAs, related to promoters of protein-coding genes, are **PROMPTs** (promoter upstream transcripts). They are transcribed in the sense or antisense orientation, approximately 0.5–2.5 kb upstream of the active transcription start sites (TSSs) of most protein-coding genes in mammals. Their functionality remains controversial. Both eRNAs and PROMPTs are retained in the nucleus and quickly degraded by nuclear RNA exosomes [69].

Intronic lncRNAs are located within the introns of PCGs. As introns are highly unstable, lncRNA produced is protected from degradation via several peculiar mechanisms. A subgroup of intronic lncRNAs is excised intron-derived snoRNA-ended lncRNAs (**sno-lncRNAs**), which are produced when a single intron contains two snoRNAs (e.g. **SLERT** [70]). The region between snoRNAs is not degraded forming a lncRNA body, protected from both ends by snoRNA secondary structures and/or snRNPs (Figure 3A), which makes their half-lives comparable to mRNAs [69].

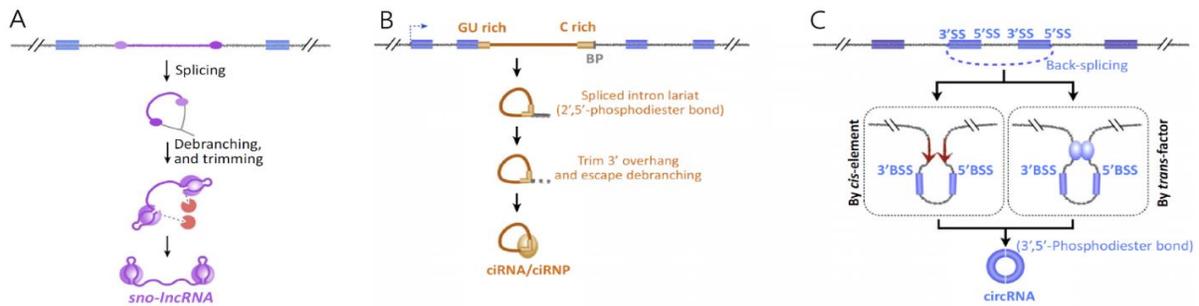


Figure 3. Production of *sno-lncRNAs*, *ciRNA* and *circRNAs*. **(A)** Production of *sno-lncRNAs*. *sno-lncRNAs* are formed when one intron contains two *snoRNA* genes. During splicing the sequence between the *snoRNAs* are not degraded, leading to the production of *lncRNA* flanked by *snoRNAs/snoRNPs*. **(B)** Production of circular intronic RNA (*ciRNA*). *ciRNA* is derived from excised introns and depends on consensus RNA sequences (orange bars) to avoid debranching of the lariat intron. **(C)** *circRNA* produced by back-splicing circularization is catalyzed by the spliceosome machinery. Back-splicing is enhanced by *cis* orientation-opposite complementary sequences (red arrows) in intron flanking circularized exons (left panel) or *trans* protein factors that can facilitate the positioning of distal back-splicing sites in close proximity (right panel). BP – branch point, 3'SS-3' splice site, 5'SS-5' splice site. Adopted from [69] © 2017 Elsevier Ltd.

Yet another variation of *snoRNA*-containing *lncRNAs* is when only one *snoRNA* is formed on the 5' end of *lncRNA*, while the 3' end is protected by polyadenylation signal. These types of *lncRNAs* are called 5' *snoRNA*-ended and 3'-polyadenylated *lncRNAs* (**SPAs**) [69]. Besides linear intron-derived *lncRNA* species, there are also circular intronic *lncRNAs* derived from spliced intron lariats – **ciRNAs**. These species are produced as a result of intron lariats debranching failure (Figure 3B), which is dependent on two consensus motifs: 7 nt GU-rich elements near the 5' splice site and an 11 nt C-rich element close to the branchpoint site [71]. It was demonstrated that *ciRNA* accumulate in the nucleus and may positively regulate

the transcription of their host PCG [71]. There is yet another type of circular lncRNAs in cells, called **circRNAs**, which differ from ciRNAs in generation mechanism, which involves back-splicing (Figure 3C). circRNAs can be both intronic and overlapping sense lncRNAs, therefore can be processed from internal exons of pre-mRNA and contain several exons. The back-splicing mechanism competes in cells with conventional splicing and normally circRNAs are presented in cells in lower numbers than their mRNA counterparts [69]. Complex alternative splicing regulation may result in the production of multiple circRNAs species from the same gene, in addition to linear isoforms. A well-known example of this is lncRNA ANRIL, which processing results in a family of linear isoforms, localized in the nucleus and a family of circular isoforms, functioning in the cytoplasm [72]. Several circRNAs were found to regulate miRNAs via sponging mechanism and gained name competing endogenous RNAs (**ceRNAs**), for example, **CDR1as** lncRNA [73].

The above mentioned are only a fraction of discovered lncRNAs types and their generation mechanisms. The most abundant and most studied group of lncRNAs is long intergenic RNAs (**lincRNAs**), and most of the cases discussed further are examples of lincRNAs.

2.2.2. LncRNA stability and localization

LncRNA half-lives are extremely heterogeneous and on average shorter than mRNAs half-lives (4.8 h versus 7.7 h). The minority (about 22%) of lncRNAs are unstable (half-life less than 2 hours) and about 6% of them are highly stable (half-life over 18 hours). lincRNAs and NATs were found to be more stable than intronic

lncRNA, spliced lncRNA are more stable than unsliced and cytoplasmic more stable than nuclear [74,75]. Some lncRNAs are restricted to the nucleus (e.g. **NEAT1**), some exported and function in the cytoplasm (e.g. **DANCR**) [1], and many are known to function both in the nucleus and in the cytoplasm, potentially continuously shuttling between them (e.g. **HOTAIR**, **TUG1**, **MALAT1**). Cases are known of lncRNAs delocalization as a part of the stress response, for example, oxidative stress. [76].

2.2.3. Coding potential of lncRNA

Some lncRNAs contain short open reading frames (sORF) and translate short peptides with important biological functions. For example, mitochondria-localized 56aa peptide MtlN involved in the regulation of mitochondria respiration is encoded by a gene annotated as lncRNA **LINC00116** [77]. Another example is 34aa long peptide encoded by lncRNA **DWORF** and involved in the regulation of muscle performance [78]. Thus, while studying functions of lncRNA, careful verification of its protein-coding potential has to take place to ensure that the lncRNA gene is not a misannotated mRNA. However, the studies show that while most of the lncRNA indeed interact with ribosomal machinery, the majority of them are not translated into functional peptides [79].

2.3 LncRNA functional diversity

LncRNAs orchestrate cellular metabolism at various levels through regulation of gene expression, transcription, pre-mRNA processing, translation, and interaction with proteins to modulate signaling. Mechanisms of lncRNA action can be divided into two major types: *in cis* and *in trans*. *In cis* mechanism of action means that lncRNA acts at its genomic loci and does

not leave the site of transcription, solely modulating the expression of its neighboring genes (Figure 4A).

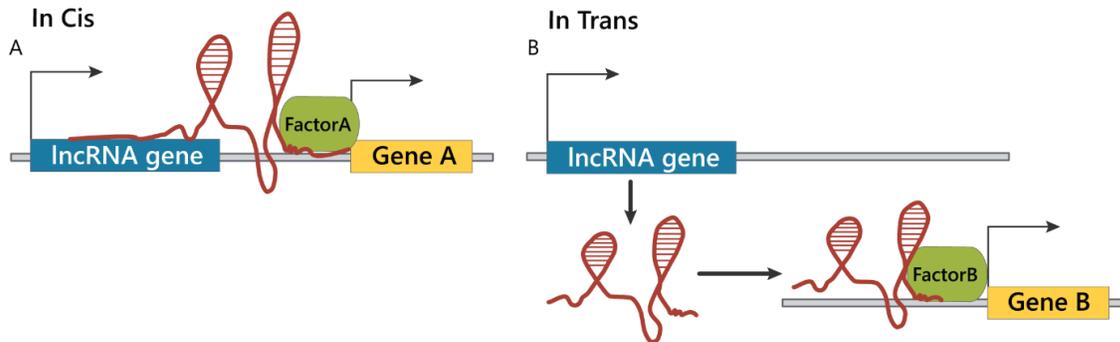


Figure 4. *In cis* and *in trans* mechanisms of lncRNA action. (A) lncRNA acting *in cis* does not leave its genomic locus and regulates transcription of neighbor genes. (B) lncRNA acting *in trans* leaves the site of its transcription and performs regulatory roles at distant cellular locations.

Therefore, lncRNA, which acts *in cis* never leaves the cell nucleus. *In trans* mechanism involves the relocation of the mature lncRNA transcript away from the site of its transcription to perform functions (Figure 4B). *In trans* lncRNA moves within the nucleus to regulate the expression of distant genes or becomes a structural backbone of various nuclear bodies, such as speckles and paraspeckles. Alternatively, *in trans* lncRNAs are being exported into the cytoplasm to perform their roles in protein scaffolding, miRNAs sponging, regulation of mRNA translation, and post-translational modification of proteins. We will now look closely at different *in cis* and *in trans* mechanisms of regulation exploited by lncRNA.

2.3.1 *In cis* regulation by lncRNAs

When it became clear that the majority of DNA is being transcribed into RNA at least to some extent, including many thousands of lncRNA genes, reasonable questions about the functional relevance of these transcripts were raised. It soon became clear, that many lncRNA genes are restricted to the nucleus and are expressed at very low levels [1]. Identification of bidirectional expression from transcription regulatory elements, such as promoters and enhancers [80], led to the proposal that the majority of lncRNA transcripts are not functionally relevant and transcription itself represents a regulatory unit, which was later proved to be true for some lncRNAs. Therefore, while describing *in cis* regulation, we should distinguish two possibilities: 1) lncRNA transcript itself is not essential and regulatory function depends on either DNA sequence within lncRNA locus or transcription/splicing activity of the locus (Figure 5A, B); 2) regulation is performed by the newly transcribed lncRNA transcript, which specifically binds to protein regulators, attracting them to the target gene (Figure 5C). Let us review major examples of each of these mechanisms (Table 1).

Active transcription from lncRNA **Upperhand** (Uph) locus, but not Uph transcripts themselves, plays an important role in embryonic cardiac tissue development by activating transcription of Hand2 transcription factor. Uph is transcribed from a bi-directional promoter in the opposite direction from Hand2 and spans across several enhancer elements responsible for Hand2 activation and conserved in multiple species. Premature termination of Uph transcription via the introduction of early polyadenylation sites into the non-conserved region of murine Uph did not alter Hand2 expression. The same absence of phenotype was observed when Uph transcripts were knocked down by more than 90%. At the same time, the termination of Uph transcription resulted in embryonically lethal mice failing to develop a right ventricular chamber [81]. The same functionality of transcription, but not the transcript itself was shown for **Airn**

lncRNA. Airn is an antisense lncRNA spanning across an insulin-like growth factor receptor Igf2r promoter. Igf2r is paternally imprinted by the mechanism of transcriptional interference caused by Airn expression from the opposite strand from Igfr2 [82].

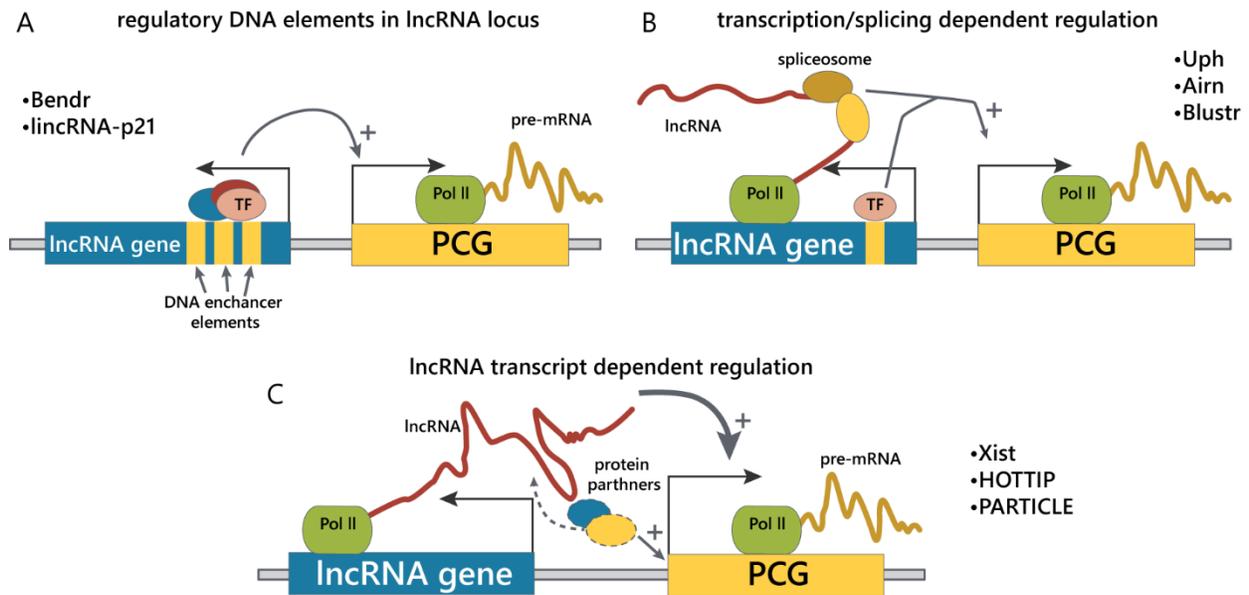


Figure 5. Mechanisms of lncRNA regulation in cis. (A) Regulation via DNA elements imbedded into lncRNA locus sequence. (B) Regulation via active transcription/splicing of from lncRNA locus. (C) Synthesized lncRNA transcript attracts regulatory factors in a sequence-dependent manner.

In some cases, even transcription activity of the lncRNA locus seems to be not essential and regulation depends on DNA sequence within the lncRNA locus only. In a thorough study done by Engreitz and colleagues, cell lines carrying heterozygous deletions of 12 lncRNAs promoter regions were generated and it was found that in 5 out of 12 deletions affected (increased or decreased) the expression of neighboring genes in an allele-specific manner [83]. Premature polyadenylation signals (pAS) were inserted downstream of promoters, into the genomic sequence of 4 out these 5 lncRNAs to abrupt production of lncRNA transcripts, while

leaving promoter sequences intact. Careful examination showed that out of 4 examined cases, in 3 a DNA sequence, but not an RNA transcript was essential for regulation. One of them was lncRNA **Bendr**. The deletion of its promoter reduced expression of the adjacent Bend4 gene, while the premature polyadenylation site did not affect Bend4. In addition to that, no transcriptional activity was detected upstream of pAS, suggesting that solely DNA elements contained within Bendr promoter are required for activation of Bend4 expression [83]. The same study proved that *in cis* activation of expression of DNA-binding protein Sfmbt2 by lncRNA **Blustr** relies on both intact promoter sequence and active transcription from the Blustr locus. Moreover, the deletion of the 5'-splice site from the first intron of Blustr resulted in a drastic reduction in overall Blustr expression and expression of Sfmbt2 as a result. At the same time, a specific Blustr RNA sequence was important for Sfmbt2 regulation. Therefore, Blustr is an example of how active transcription from lncRNA locus and splicing of lncRNA transcripts can be essential units for *in cis* regulation [83].

Another example of regulatory DNA elements within lncRNA locus is **lincRNA-p21**, which initially was mistakenly identified to be acting *in trans* [84]. More precise studies using knockout mouse models demonstrated that lincRNA-p21 transcripts are dispensable for function and at least partially lincRNA-p21 regulation is performed via active DNA enhancer elements within its sequence [85] independently of lincRNA-p21 transcript production. At the same time, it was clearly shown that the ASO-mediated knockdown of lincRNA-p21 leads to the downregulation of expression of neighboring cyclin-dependent kinase inhibitor CDKN1A (p21) in mouse embryonic fibroblasts. This fact confirms that under some conditions, either lincRNA-p21 transcripts or active transcription are also involved in regulation [41] (Table 1).

2) LncRNA of this type bind to transcription factors or chromatin remodeling complexes in a sequence-dependent way and attract them to their target DNA sequences, which may be gene promoters or in the case of **Xist**, the entire X chromosome. Xist transcription is triggered during cell differentiation from the X-inactivation center (Xic), located on the X chromosome which is destined to become imprinted. Intensively transcribed lncRNA propagates onto the entire chromosome, attracting protein complexes, responsible for transcription inhibition, including PRC1 and PRC2, resulting in chromosome reorganization and complete chromosome silencing [86,87].

Table 1. Examples of lncRNA in cis and in trans regulators.

In cis acting lncRNAs		
Type of regulation	LncRNA [ref]	Mechanism
Regulation through enhancer DNA elements in lncRNA locus	Bendr [83]	Promoter DNA elements are required for activation of adjacent Bend4 expression
	lincRNA-p21 [85]	DNA enhancer elements within the locus sequence are essential for adjacent genes regulation
Regulation via active transcription/splicing	Upperhand [81]	Transcription from Uph locus activates transcription of Hand2 TF.
	Airn [82]	AS lncRNA transcription results in silencing of pair Igf2r gene transcription
	Blustr [83]	Active transcription of lncRNA is required for Sfmbt2 gene transcription
Sequence-dependent regulation by lncRNA transcript	Xist [88]	Binds protein partners in a sequence-dependent way to inactivate X chromosome
In trans acting lncRNAs		
Chromatin remodeling by lncRNAs	HOTTAIR [89]	Acts as a scaffold bringing inhibitory proteins to HoxD gene locus
	LincRNA-EPS [90]	Maintains repressed chromatin states of IRGs promoters by promoting nucleosome occupancy
	MEG3	
LncRNA interact with proteins to regulate their function (scaffolds, decoys, etc)	SLERT [70]	Intronic sno-lncRNA promotes PolII transcription of pre-rRNA genes by binding to PolII inhibitor DDX21
	NORAD [91,92]	Serves as a decoy for PUMILIO proteins, an inhibitor of DNA repair and replication

The organization of nuclear compartments by lncRNAs	NEAT1 [93,94]	Scaffold core for paraspeckles
	Firre [95]	Participates in 3D chromatin organization, bringing distant chromosome loci together
	MALAT1 [96,97]	Involved in the organization of nuclear speckles
LncRNAs acting as competing endogenous RNAs (ceRNAs)	linc-MD1 [98]	Regulates muscle differentiation by sponging miR-133
	CDR1as [99,100]	Regulates brain development by sponging miR-7
	Sry [73]	Sponge for miR-138
LncRNA as regulators of alternative splicing	SAF [101]	Regulates AS of Fas receptor via direct RNA-RNA interaction
	MALAT1 [102]	Promotes tumor growth by binding SFPQ releasing proto-oncogene PTBP2 from SFPQ/PTBP2 inhibitor complex. 70 Modulating distribution of SR factors at nuclear speckles 68
	NEAT1 [103]	Interacts with Clk kinase to promote phosphorylation of SRp40 protein and regulation of PPAR γ AS
	LINC01133 [104]	Binds SRSF6 and inhibits its oncogenic properties
	GOMAFU [105]	Binds QK1 and SRSF1 splicing factors to regulate AS of Schizophrenia associated genes
	LINC-HELLP [106]	Binds several proteins involved in the regulation of RNA splicing: YBX1, PCBP1, and PCBP2

2.3.2. *In trans* regulation by lncRNAs

LncRNAs, which are adopted *in trans* mechanism of regulation leave the site of transcription to perform their functions in distant parts of the nucleus or in the cytoplasm. *In trans* lncRNAs were found to perform a wide range of functions, which include: 1) regulation of distant genes expression via chromatin remodeling and modulation of transcription; 2) structural basis for nuclear compartments; 3) protein scaffolds; 4) regulators of protein post-translational modifications; 5) acting as competing endogenous RNAs via sponging miRNAs; 6) regulators of alternative splicing. Each type of regulation is represented by dozens of examples, let us name a couple of them (Table 1).

2.3.2.1. Chromatin remodeling by lncRNAs

LncRNA may act as scaffolds, navigating chromatin-remodeling protein complexes to their target sites. For example, lncRNA **HOTAIR** is 2.2 kilobases long transcript spliced and polyadenylated, expressed from HoxC locus in the antisense orientation, which was found to regulate HoxD gene cluster located 40 kb away from the site of its transcription. HOTAIR interacts with Polycomb Repressive Complex 2 (PRC2) and is required for repressive H3K27 trimethylation of HoxD locus genes [89] (Figure 6).

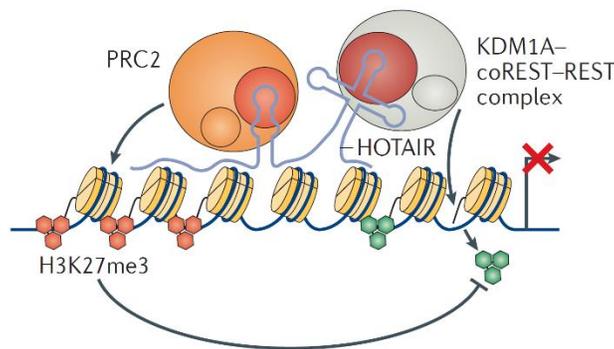


Figure 6. lncRNA HOTAIR serves as a scaffold for repressive chromatin remodeling complexes to inhibit the transcription of its target genes [66] © 2017, Springer Nature.

LincRNA-EPS is a spliced and polyadenylated 2.5-kb long transcript that specifically represses the activity of immune response genes (IRGs) in macrophages. Mice with the generated knockout of the entire 4 kb lincRNA-EPS genomic locus showed increased expression of IRGs, which was then rescued by overexpressing lincRNA-EPS from the vector. RNA antisense purification (RAP) assay identified that lincRNA-EPS interacts directly with promoters of repressed genes and ATAC-Seq showed that lincRNA-EPS depleted cells experienced

alterations of nucleosome fingerprints around TSSs of IGRs. These results suggest that lincRNA-EPS maintains repressed chromatin states of IGRs promoters by promoting nucleosome occupancy [90].

2.3.2.2. *LncRNA interact with proteins to regulate their function (scaffolds, decoys, etc)*

LncRNA **SLERT** is a rare example of lncRNA controlling expression of genes transcribed by polymerase I. SLERT is 694-nt RNA that is transcribed from the intron of the human TBRG4 gene and stabilized by two box H/ACA snoRNAs at its ends. SLERT promotes PolI transcription of pre-rRNA genes by binding to PolI inhibitor DDX21. Direct interaction of SLERT and DDX21 changes the conformation of DDX21 individual subunits, which results in dissociation of DDX21 from PolI and increases PolI occupancy at rRNA clusters [70].

NORAD is unspliced, conserved lncRNA induced by DNA damage in both human and mouse. NORAD depleted cells exhibited loss and gain of chromosomal numbers [91]. NORAD contains ~400 nt repetitive element recurring 5 times within its RNA sequence, which serves as a binding platform for PUMILIO proteins. PUMILIO proteins were found to inhibit mitotic regulators, DNA repair factors, and DNA replication factors negatively affecting the process of chromosome segregation during mitosis. NORAD acts as a molecular decoy sequestering PUMILIO proteins maintaining chromosomal stability [91,92].

2.3.2.3 *The organization of nuclear compartments by lncRNAs*

Nuclear speckles are nuclear domains enriched in pre-mRNA splicing factors including snRNP and SR proteins, located in the interchromatin regions of the nucleoplasm of mammalian

cells. **MALAT1** is a single-exon transcript over 7 kb in length, which is localized to nuclear speckles through its specific interaction with speckle retained proteins. Both MALAT1 knockdown and delocalization from the speckles resulted in the downregulation of a subset of target genes suggesting that MALAT1 regulates transcription or RNA processing through its localization to speckles [96] (Figure 7A). Additionally, MALAT1 was found to localize in speckles in an RNA transcription-dependent manner and MALAT1 knockdown studies showed that MALAT1 modulates localization of SR proteins at the transcriptionally active sites [97]. However, MALAT1 is not essential for nuclear speckles formation, moreover, MALAT1-deficient mice did not have abnormalities in alternative splicing patterns [107,108].

NEAT1 is a single-exon transcript that is alternatively spliced in human cells to produce 3.7-kb and 22.7-kb isoforms. Together with protein partners - SFPQ and NONO, NEAT1 forms a scaffolding core of the subnuclear structures called paraspeckles. Paraspeckles are dynamic nuclear compartments, which regulate cellular processes like circadian cycling and stress response through sequestering proteins involved in transcription regulation and pre-mRNA processing [93]. The knockdown of NEAT1 with ASOs resulted in the complete disintegration of paraspeckles in both human and murine cell lines [94]. NEAT1 function as a structural scaffold in paraspeckles depends on interaction with SFPQ and NONO proteins (Figure 7B). Initial binding of SFPQ and NONO to NEAT1 is essential for NEAT1 stability. Additionally, SFPQ and NONO play the role of basic paraspeckle scaffold themselves as RNA-protein interaction results in oligomerization of SFPQ-NONO dimers into longer chains of polymers along NEAT1, essential for paraspeckles integrity. The knockdown of either SFPQ or NONO completely oblates paraspeckle formation [94].

LncRNA **Firre** was found to participate in a 3D chromatin organization. Firre contains a 156 nt repeating sequence, which directly interacts with the nuclear-matrix factor hnRNPU. Firre is expressed from the X chromosome avoiding X chromosome inactivation and localizes across an ~5-Mb domain on the X chromosome. Simultaneously, Firre was found by RAP to localize at multiple trans locations on various autosomes and bridging the X chromosome with these distant genome loci (Figure 7C). This colocalization was completely released after the depletion of Firre [95]. Additionally, **Xist** lncRNA also plays a role in the nuclear organization by targeting an inactive X chromosome to the nuclear periphery via direct interaction with the lamin B receptor [109].

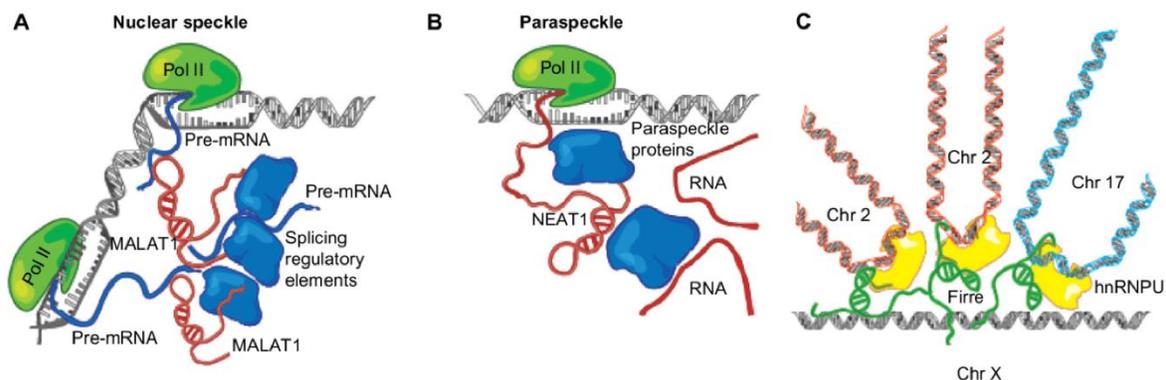


Figure 7. Organization of nuclear architecture by lncRNAs. (A) MALAT1 is a part of nuclear speckles. (B) NEAT1 is an essential component of nuclear paraspeckles. (C) Firre participates in 3D chromatin organization bridging Xi chromosome and distant autosomes. Adopted from [110] © Copyright 2020 Dove Press Ltd.

2.3.2.4 LncRNAs acting as competing endogenous RNAs (ceRNAs)

LncRNA were reported to modulate not only protein activity, but also the activity of RNA molecules. Competing endogenous RNAs (ceRNAs) bind miRNAs and titrate their

availability for target mRNAs. For instance, this mechanism is crucial for muscle differentiation where lncRNA **linc-MD1** governs the whole process by sponging miR-133 and preventing it from the regulation of the expression of MAML1 and MEF2C, transcription factors of muscle-specific genes [98]. **CDR1as** is a circular cytoplasmic long-noncoding RNA abundantly expressed in the human and mouse brain. A very unique feature of CDR1 is that it has ~70 seed matches for miR-7 and it was demonstrated to bind AGO2 in a miR-7 dependent manner [73] (Figure 8). Expression of human CDR1 in zebrafish embryos resulted in an abnormal brain development phenotype similar to miR-7 knockdown [99], suggesting that CDR1 sponges miR-7 to increase the expression of its target genes. However, knockout of CDR1 from the mouse genome results in a drastic decrease of miR-7 expression, indicating that interplay between these RNAs is more sophisticated [100].

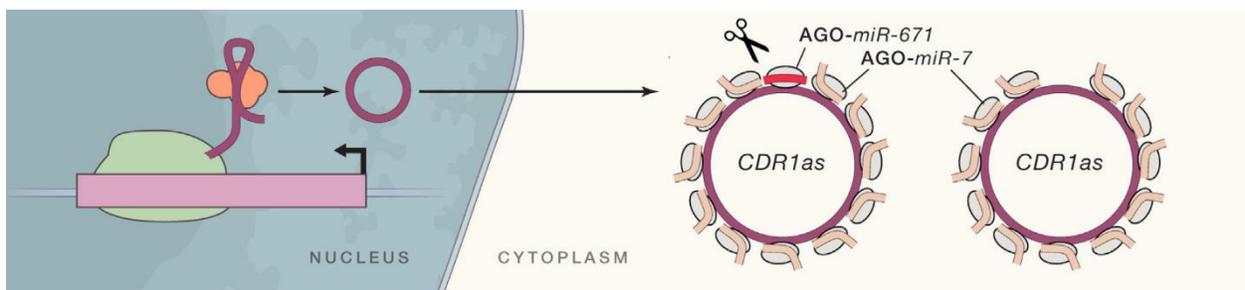


Figure 8. Circular RNA *CDR1as* contains ~70 seed matches for miR-7 playing the role of *ceRNA* for miR-7.

Adopted from [84] © 2018 Elsevier Inc.

Another example is testis-specific circular RNA **Sry**, which acts as a sponge for miR-138 [73]. Circular lncRNAs are more stable towards miRNA-induced degradation since they lack poly(A) tail and are not subjected to deadenylation.

2.3.2.5. *LncRNA as regulators of alternative splicing*

Many lncRNAs were identified to regulate mRNA splicing through several mechanisms, among them is the modulation of the phosphorylation state of various splicing factors (SFs) (MALAT1 [111]), NEAT1 [103], competitive binding (MALAT1 acting on SFPQ [102], RNA-RNA duplex formation with pre-mRNA molecules [101]. The RNA-RNA mechanism of regulation is typically attributed to NATs. One example of this is the NAT called **SAF**, which is transcribed from the opposite strand of intron 1 of the Fas receptor – activator of apoptosis. SAF directly interacts with Fas pre-mRNA, promoting skipping of exon coding for the transmembrane domain of the Fas receptor, which results in a production of soluble Fas protein (sFas), which is a prominent suppressor of apoptosis. This mechanism of exon skipping and the production of sFas was found to be used by the cancer cells to acquire resistance to apoptosis [101].

In most cases, lncRNAs team up with proteins to modulate AS patterns of various mRNAs. This regulation is often involved in a progression of various diseases, such as cancer [102,104] or neurological disorders [105]. For example, **NEAT1** has been shown to interact with Clk kinase to promote phosphorylation of SRp40 protein. Phosphorylated SRp40 binds to one of the exons on PPAR γ pre-mRNA driving its inclusion and therefore production of a longer version of PPAR γ transcription factor, PPAR γ 2, over the shorter version PPAR γ 1, which manifests a switch in a differentiation program of adipocytes [103] (Figure 9A). Another lncRNA, which modulates phosphorylation of SR proteins is **MALAT1**. SR proteins cycle between phosphorylated and dephosphorylated states, which is essential for pre-mRNA splicing. MALAT1 depletion results in both dephosphorylation of SR proteins and differential changes in

AS events in several mRNAs, mostly exon inclusions [111]. Additionally, MALAT1 interacts with splicing factors to promote tumor growth. Proto-oncogene PTBP2 resides in the cytoplasm inhibited by direct binding of the SFPQ splicing factor. MALAT1 was shown to compete with PTBP2 for SFPQ binding, disrupting the SFPQ/PTBP2 complex and releasing PTBP2, which then promotes tumorigenesis [102] (Figure 9B).

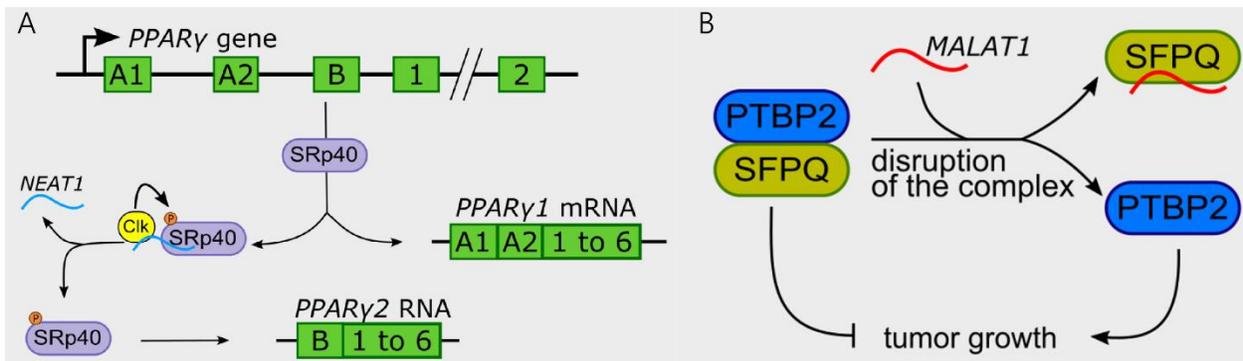


Figure 9. Examples of alternative splicing regulation by lncRNAs. (A) NEAT1 promotes phosphorylation of SRp40 protein and production of a longer version of PPARγ transcription factor, PPARγ2, over the shorter version PPARγ1. (B) MALAT1 competes with PTPB2 proto-oncogene for binding to the SFPQ splicing factor. Adopted from [112] © 2018, Oxford University Press.

Another lncRNA **LINC01133** binds to the regulator of alternative splicing SRSF6 to inhibit its oncogenic properties like the promotion of EMT and metastatic properties of colorectal cancer cells [104]. LncRNA **GOMAFU** was found to be involved in Schizophrenia (SZ) associated with defective alternative splicing. GOMAFU is expressed from SZ related genomic locus and it was found to bind splicing factors QKI and SRSF1. GOMAFU knockdown results in differential AS patterns of SZ-associated genes DISC1 and ERBB4, which correspond to AS patterns found in SZ patients [105].

LncRNA **LINC-HELLP** is associated with pregnancy-specific HELLP syndrome in Dutch families and is mutated in individuals carrying this syndrome. LINC-HELLP binds several proteins involved in the regulation of RNA splicing: YBX1, PCBP1, and PCBP2. Involvement of LINC-HELLP in AS regulation remains elusive, however, upon the occurrence of mutations associated with HELLP syndrome, LINC-HELLP transcripts lose the ability to interact with its RNA splicing related protein partners [106].

These are only several examples of lncRNAs teaming up with splicing factors to fine-tune AS of mRNAs. This type of lncRNA regulatory role emerges to be quite common across different organisms, although mechanisms of action are extremely diverse, and we can expect that future research will uncover many new variations. As many reported cases of lncRNAs regulation of AS are linked to disease progression, lncRNAs are essential for AS fine-tuning under specific conditions and opening a prospective research road toward novel tools for disease treatment [112].

2.4. LncRNAs in cancers and other pathologies

LncRNA are involved in virtually every aspect of the regulation of gene expression and evidence suggests that malfunction of lncRNA signaling leads to various pathologies. It was identified in genome-wide association studies that 88% of disease-associated SNPs (single-nucleotide polymorphism) lay outside protein-coding regions, in non-coding regions including lncRNA genes [58]. By 2018 around 10000 experimentally validated lncRNA-disease associations were published on NCBI Pubmed [113]. Differential expression of lncRNA was linked to various complex diseases, such as cardiovascular, neurological diseases, diabetes, autoimmune diseases, cancer, etc. [65,114] and in some cases, functional roles were described.

For example, lncRNA **Flicr** promotes autoimmune diabetes by *in cis* inhibition of Foxp3 transcription factor in regulatory T cells [115]. And mutations in **lnc-NR2F1** locus were associated with autism [116].

LncRNA expression was associated with carcinogenesis and lncRNA potential to become cancer biomarkers and therapeutic targets is widely appreciated [117]. LncRNAs are found to regulate hallmarks of cancer: proliferative signaling, evasion of growth suppression, metastasis, replicative immortality, induction of angiogenesis, apoptosis resistance. (Figure 10). LncRNAs **PVT1** and **CCAT1** (also known as CARLo-5) were found to promote the proliferation of tumor cells by enhancing protein production of Myc oncogene [118,119] (Figure 10A). To continue uncontrollable growth, the pre-cancer cell switches off built-in mechanisms of growth suppression. Several lncRNAs are involved in the p53-tumor suppressing pathway and were found downregulated in different cancers. For example, **MEG3** lncRNA that binds p53 to stimulate transcription of p53 downstream targets [120]. Another p53 induced lncRNA **LED** activates p21 enhancer transcription, promoting cell cycle arrest in tumor cells [121]. **lincRNA-p21** also positively regulates p21 via *in cis* mechanisms [41]. On the other hand, lncRNA **FAL1** inhibits p21 expression by attracting BMI1 chromatin repressor to p21 promoter [122] (Figure 10B). Promoters of cancer cell motility were also found among lncRNA. BRAF oncogene-induced lncRNA **BANCR** is a driver of melanoma cell migration [123].

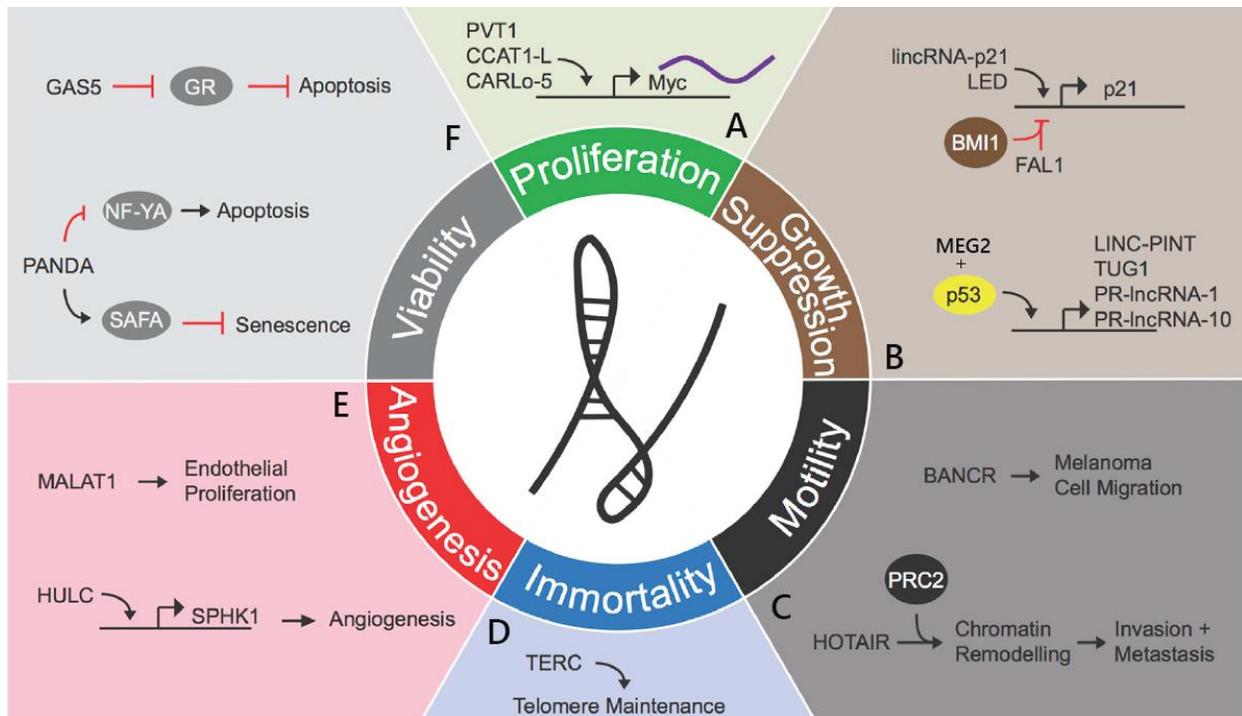


Figure 10. LncRNA regulators of various hallmarks of cancer. Adopted from [117] © 2016 Elsevier Inc.

Overexpression of **HOTAIR** reprograms breast cancer cells via PCR2 dependent chromatin remodeling that results in metastasis [124] (Figure 10C). For fast-dividing cancer cells, telomere maintenance is essential to establish replicative immortality. SNPs in Telomere RNA component (**TERC**) were associated with longer telomere lengths in glioma patients [125] (Figure 10D). Angiogenesis is the formation of new blood vessels through migration, growth, and differentiation of endothelial cells. **MALAT1** overexpression was previously associated with cancer progression. **MALAT1** was also confirmed to play a role in angiogenesis. Mice lacking **MALAT1** had a delayed vessel sprouting compared to wild type mice [126]. Another lncRNA **HULC** promoted tumor angiogenesis in liver cancer via **SPHK1** kinase [127] (Figure 10E). To survive and develop into tumors, pre-cancer cells develop mechanisms to avoid apoptosis. **PANDA** is a lncRNA induced in response to DNA damage in a p53-dependent manner that can

bind the NF-YA transcription factor to limit the expression of pro-apoptotic genes. PANDA depletion sensitizes cells to DNA-damage induced apoptosis [128]. Cell senescence is an irreversible cell cycle arrest that restricts the proliferation of pre-cancerous cells. lncRNA PANDA was demonstrated to bind scaffold-attachment-factor A (SAFA) to regulate senescence-promoting genes in proliferating cells via recruiting PRC1 and PRC2 to their promoters [129] (Figure 10F). lncRNA **GAS5** acts as a decoy for a glucocorticoid receptor (GR) preventing its binding to target genes and activation of their expression. GAS5 is downregulated in prostate cancer and breast cancer. Overexpression of the wild type GAS5 in both cancers induces apoptotic cell death [130] (Figure 10F).

2.4.1. lncRNAs in hepatocellular carcinoma

Liver cancer is the fifth most common and third deadliest cancer in the world, with cases per year currently increasing. About three-fourths of liver cancer cases are hepatocellular carcinoma (HCC). Genome-wide lncRNA profiling studies performed using tissue samples from HCC patients demonstrated that dozens of lncRNAs were dysregulated in this type of liver cancer. In the study involving 20 HCC patients, 917 lncRNAs were found recurrently dysregulated, correlated with clinical data from previous studies, and associated with clinicopathologic features [131]. Several lncRNAs were found to drive specific mechanisms of HCC progression, among them are HULC, HOTTIP, HOTAIR, MALAT1, NEAT1, H19, CYTOR. These lncRNAs were found upregulated in HCC tissues in comparison to adjacent healthy tissues and were associated with poor patient prognosis [132] (Table 2).

Table 2. Examples of lncRNAs dysregulated in HCC.

lncRNA	class	Expression in HCC
HULC	intergenic	upregulated
HOTTIP	intergenic	upregulated
HOTAIR	intergenic	upregulated
MALAT1	antisense	upregulated
NEAT1	intergenic	upregulated
H19	intergenic	Downregulated/upregulated
MEG3	intergenic	downregulated
CYTOR	intergenic	upregulated

LncRNA **HULC** is one of the most upregulated genes [133] found in the blood samples of HCC patients, proposed as a novel HCC biomarker, and promoted angiogenesis in liver cancer [127]. LncRNA **HOTTIP**, which serves as a scaffold for chromatin remodeling complexes, is essential for cancer cell proliferation in vitro [134]. It was demonstrated that HOTTIP promotes tumorigenesis at least in part via regulation of the glutaminase GLS1 gene [135]. **HOTAIR** is a scaffold lncRNA acting *in trans* promotes the expression of oncogene BMI1 via competitive binding to its inhibitor miR-218 [136]. MALAT1 and NEAT1 are frequently mutated in HCC. **MALAT1** enhances glycolysis in cancer cells, one of the cancer hallmarks, via TCF7L2 transcription factor, a negative regulator of gluconeogenesis [137]. **NEAT1** is associated with chemoresistance and overexpressed in both sorafenib and doxorubicin-resistant cells [138]. **H19** is transcribed from imprinted locus IGF2/H19, which dysregulated expression is associated with many cancers, including HCC. In mice, PHB1 and CTCF cooperatively inhibit the IGF2/H19 locus control region and PHB1 inhibition leads to an increase in cancer cell proliferation via H19 upregulation [139]. LncRNA **CYTOR** and its paralogue MIR4435-2HG were also found

upregulated in HCC and proposed as biomarkers of HCC progression [14]. Details of their involvement in HCC are discussed further.

MEG3 expression is significantly reduced in HCC. DLK1-MEG3 is an imprinted locus consisting of multiple maternally expressed noncoding RNA genes and paternally expressed protein-coding genes. DLK1-MEG3 locus plays a role of a tumor suppressor in multiple cancers, including HCC. miR-493-5p silenced in DLK1-MEG3 locus plays a tumor-promoting role partially by inhibiting the IGF2-derived intronic miR-483-3p interconnecting two genetically imprinted gene loci in the regulation of cancer progression [140].

2.5. LncRNA tissue specificity.

Global transcriptome profiling of various organisms and tissues revealed highly specific cell-type and tissue-type expression patterns of lncRNAs in comparison to mRNAs [141,142]. Analysis of 24 different human tissues and cell types showed that 78% of lincRNAs are tissue-specific (versus ~19% for PCGs), including 35% of the most highly expressed lincRNAs. Only the minority of lncRNAs is expressed in all human tissues. Illumina Human Body Map Project demonstrated that this number is approximately 11% of lncRNA (e.g. **MALAT1**, **TUG1**) [8,143]. Moreover, lncRNA demonstrated higher than mRNA expression variability between individual humans [144]. Differential tissue and cell type expression of lncRNA might reflect the variability of their functions in different cell types and under different conditions. It was observed that the roles of lncRNA in cell culture and in the whole organism seem to contradict each other in numerous cases. Moreover, there are precedents of different loss-of-function (LOF) phenotypes in the cases of lncRNA knockout and knockdown. For example, knockout of

MALAT1 in human tumor cell lines (lung and liver cancer) and in mice does not demonstrate any changes that were previously reported for MALAT1 knockdowns, such as the cell cycle arrest and a disrupted nuclear architecture.

2.5.1 LncRNA isoforms may exhibit functionally distinct and tissue-specific properties

LncRNAs undergo splicing events and most of them are polyadenylated, like mRNA. Studies report that lncRNA may switch performed regulatory functions via expression of several distinct functionally active transcripts from the same genomic locus. One example of this is an evolutionarily conserved lncRNA called **lnc-NR2F1** that was found to be mutated in neurodevelopmental disorders and involved in neuronal cell maturation and regulation of transcription of genes linked to autism. ChIPR-Seq method demonstrated that isoforms of lnc-NR2F1 have different binding sites in the genome and different effects on gene transcription. Specifically, only the longest isoform of lnc-NR2F1 is functionally active in neurogenesis and patients with neurodevelopmental disorders like autism [116].

Another example is **MALAT1**, ~7000nt long, conserved across 33 mammalian species abundantly expressed at the levels comparable to protein-coding house-keeping genes. MALAT1 premature transcript is not polyadenylated, instead, it has poly(A) tail encoded in its genome, which is then being processed by t-RNA biogenesis factors RNase Z and RNase P into two transcripts: long transcript (around 6.7 kb) and a short transcript (61 nt). The long transcript is retained in the nucleus, localized to the nuclear speckles, and is thought to play a scaffold role bridging together transcription and pre-mRNA processing machinery (Figure 11). The short transcript is exported into the cytoplasm and is called MALAT1-associated small cytoplasmic

RNA (mascRNA) [145]. mascRNA adopts a tRNA-like secondary structure, which protects it from degradation.

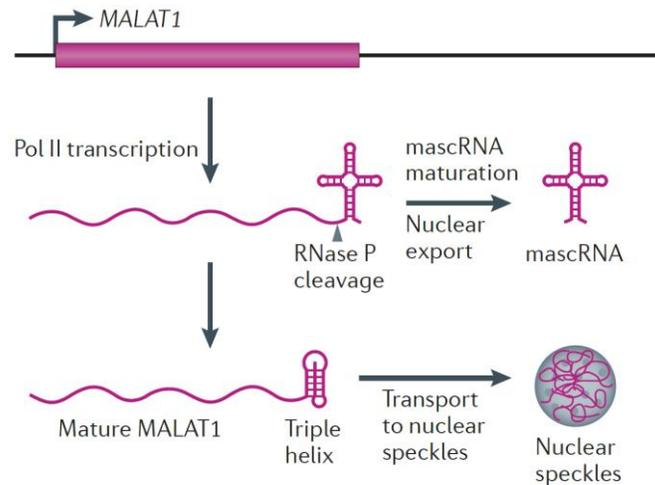


Figure 11. Alternative splicing of *MALAT1* generates a longer nuclear speckles-retained transcript and a shorter tRNA-like cytoplasmic isoform. Adopted from [146] © 2015, Springer Nature.

ANRIL lncRNA genomic locus gives rise to many linear and circular transcripts, which are expressed in a tissue-dependent manner. ANRIL is an antisense lncRNA, spanning across the gene *CDKN2B*. ANRIL expression is deregulated in a multitude of illnesses such as cardiovascular disease, glaucoma, diabetes, cancers (gastric, breast, lung, bladder) and represents a frequent mutational spot for disease-associated polymorphisms [72]. ANRIL locus codes for 21 exons and therefore has the potential to give rise to a great number of splice variants. Analysis of ANRIL expression in patient-derived human melanoma demonstrated differential expression of ANRIL exons suggesting a tissue-specific distribution of ANRIL transcripts, as well as separate cellular localization and degradation rates of ANRIL linear and circular isoforms, with circular

transcripts being more stable and localized to the cytoplasm and linear transcripts less stable and retained in the nucleus [72]. Disease-associated polymorphism found in ANRIL locus may lead to alterations in transcript expression by affecting the splicing or stability of the transcripts. Indeed, several studies demonstrate ANRIL isoform-specific effects [147]. For example, a polymorphism associated with elevated risk for Coronary artery disease (CAD) was linked to the upregulation of expression of some ANRIL transcripts and downregulation of others [148].

Circular ANRIL transcript (circANRIL) was reported to be a component of the pre-ribosomal assembly complex and function in maturation of the ribosome. circANRIL competes with pre-rRNA for binding to PES1 protein, which is a key component of ribosomal protein machinery PeBoW, preventing it from binding pre-rRNA and inhibiting further pre-rRNA maturation [149]. Linear nuclear retained ANRIL transcript broadly regulates gene expression, acting *in cis* and *in trans*. *In cis* ANRIL regulates expression of its neighboring tumor suppressor genes: cyclin-dependent kinase inhibitor 2A (CDKN2A) and cyclin-dependent kinase 4 inhibitor B (CDKN2B) [150,151]. *In trans* ANRIL regulates genes in PCR1/2 dependent manner via Alu motifs found in ANRIL transcripts and in promoters of target genes [152].

lncRNA GNG12-AS1 represents an unusual event of allele-specific expression of isoforms. lncRNA GNG12-AS1 is transcribed in the antisense orientation from the tumor suppressor gene DIRAS GTPase, which is located within GNG12-AS1 intron. It was shown that transcription from the GNG12-AS1 locus and not the mature transcript itself is required for suppression of DIRAS transcription [153]. Incorporation of a second exon (exon which is the closest to DIRAS locus) into the mature GNG12-AS1 isoform is strictly maternal while expression of isoforms that do not contain this exon is biallelic in normal cells. In breast cancer, the expression from the paternal allele is completely silenced and transcription of all transcripts

is monoallelic. The inclusion of the second exon is probably regulated by cohesin and related to DIRAS3 expression, as knockdown of cohesin resulted in biallelic incorporation of this exon and inhibition of DIRAS expression [154]. Therefore, lncRNAs regulatory roles should be considered in the transcript specific and tissue-specific context.

2.6 Evolution and inter-species homology of lncRNAs

The degree of evolutionary conservation of lncRNAs and evidence for functional homologs in different species is an essential question from both research and application points of view. Human lncRNA involved in regulatory networks and associated with pathologies needs to be studied in model organisms before the mechanism of its action can be used for therapeutic purposes. In the case of lncRNA found in model organisms, search for its human homologs is important to clarify potential implications for humans.

A common feature of lncRNA genes is fast evolutionary turnover, which means that they are acquired and lost throughout evolution more rapidly than protein-coding genes and some other types of non-coding RNAs, such as tRNAs, rRNAs, snoRNAs. However, comparison of lncRNA between human and mouse proved evolutionary conservation of thousands of lncRNAs (about 30% of all human lncRNAs [155]) between these species, demonstrating evidence for purifying selection of lncRNA genomic loci, promoter sequences, exonic sequences, and positions of consensus splicing motifs [43,44]. This evolutionary conservation has also been extended to other mammals as well as nonmammalian vertebrates [156]. However, unlike mRNA genes and other non-coding RNA genes, mammalian lncRNAs do not have orthologs outside vertebrates [45]. In

vertebrates, transcriptome analysis of 17 species discovered that more than 70% lncRNA evolved less than 50 million years ago [42] and no common lncRNAs were found between species diverged more than 500 million years ago [45].

The high rate of lncRNA turnover is better observed on closely related species. About 20% of the total human lncRNA pool seem to not have any detectable orthologs in any other mammal except chimpanzees [143]. Analysis of lncRNA expression in the livers of closely related rodents demonstrated that nearly 40% of lncRNAs have been acquired or lost between species as evolutionary close as *Mus musculus domesticus* and *Mus musculus castaneus*, which shared a common ancestor only 1 million years ago [155]. Interestingly, tissue-specificity is a well-preserved feature of lncRNAs and multiple lncRNAs conserved between species demonstrate the same expression profile in various tissues in different organisms. [143].

Three types of lncRNA homology between species can be identified: sequence conservation, gene structure conservation, positional conservation.

2.6.1. Conservation of nucleotide sequence

The strongest evidence for lncRNA inter-species homology is the conservation of their primary nucleotide sequences. However, the primary sequence of most lncRNAs does not seem to be under significant selection and is subject to frequent rearrangements and single nucleotide substitutions, resulting in overall poor sequence conservation between organisms even for conserved and well-studied lncRNAs such as MALAT1 and XIST. lncRNAs conserved between human and mouse on average demonstrate only around 20% sequence homology, which decreases down to 5% homology when the same lncRNAs are compared between human and zebrafish [42].

One probable explanation of this phenomenon is that lncRNA function is dependent on their secondary structure, not primary or tertiary structure. Therefore, while the primary structure evolves the secondary structure remains unchanged. As all previously studied types of noncoding RNAs function through adopting a specific and highly conserved secondary structure (tRNAs, snRNAs, rRNAs, etc), this could be a reasonable assumption. However, while for some lncRNAs it was demonstrated that their secondary structure indeed plays an important role and remains conserved throughout evolution (e.g. 3'-ends of MALAT1 and NEAT1, roxX lncRNA in *Drosophila*, GAS5, Xist, HOTAIR) [157], it seems that these are rare exceptions rather than a general rule. Most lncRNAs do not exhibit conservation of their secondary structure in the absence of primary sequence conservation [45]. Additionally, no correlation was found between the amount of lncRNA secondary structure and its evolutionary conservation and generally, the complexity of lncRNA folding is the same as for mRNAs [158].

However, lncRNA exon sequences do not evolve randomly and do undergo a certain degree of selection. Comparison of human and mouse genomes to zebrafishes showed that lncRNA exons have higher turnover than coding regions of mRNAs but significantly better preserved than introns and intergenic regions [156]. It was observed that homologous lncRNAs usually contain only short patches of conserved sequence, much smaller than the complete length of the transcript and usually spanning across one or two exons [42]. Analysis of 29 mammalian genomes showed that only 22% of lncRNAs bases were located within regions of the conserved sequence [159]. This observation suggests that only a limited amount of lncRNA sequence may be responsible for its function, and therefore remains preserved throughout evolution, while the rest of

the transcript is dispensable. Such a sequence-dependent function can be attributed to binding to a certain functional partner, such as miRNA or a protein. It was even demonstrated that short segments of lncRNAs can represent functional modules capable of performing functions in the absence of the rest of the transcript. **RoX1** RNA responsible for sex dosage compensation in *Drosophila* has three protein binding domains and one of them only one-tenth the entire transcript in size is sufficient for complete dosage compensation in roX-null flies [49].

Xist lncRNA has overall gene structure conservation, such as positions of tandem-repeats and exon/intron architecture between human and mouse, but conservation of primary sequence is constricted to specific functional regions. Analysis of closely related rodent species revealed that the sequence of Xist lncRNA is conserved over Xist-related tandem repeats and five short sequence patches, while the rest of the sequence experience higher than expected rate of turnover [160].

2.6.2. Conservation of gene structure

lncRNA genes, which either lack or demonstrate limited conservation of their sequence, may have conservation of their overall genome architecture: size, number, and positions of exons, length of the transcript, splicing patterns [156]. One example is **MALAT1**, which sequence conservation is limited to the 3' end of the transcript, however, length of ~7kb and single-exon structure is conserved across all vertebrates. Another example is **NEAT1** lncRNA, which in both human and mouse is characterized by the expression of two long and short non-polyadenylated isoforms, containing RNase P cleavage at 3' site. While preserving their genomic characteristics in both human and mouse, NEAT1 isoforms lack any significant nucleotide sequence conservation between species [161].

2.6.3 Positional conservation of lncRNAs

Finally, lncRNA genes also have positional conservation, which has been shown to have prevalence over sequence conservation [156]. A thorough comparison of zebrafish lncRNAs to murine and human lncRNAs found a large number of lncRNA genes in these species having similar genomic environments and same orientation while lacking any detectable sequence conservation [156]. Such lncRNAs may function through transcript-independent *in cis* mechanisms when the transcriptional activity is by itself is a regulatory factor and the mature RNA transcript is not relevant for function. For these lncRNA, it is typical to have a non-conserved length of the transcript and exon-intron architecture. The same lncRNAs may simultaneously have all three types of homology or just one or two of them.

2.6.4 Functional homology of lncRNAs

Sequence, structural or positional homology of lncRNA may or may not be an indicator of functional homology. Generally, sequence homology is a greater indicator of a similar functionality than positional or structural homology. The functional repertoire and mechanism of actions of lncRNA homologs between species must be determined experimentally for each specific lncRNA through studying loss of function phenotypes and potential protein partners.

Two great examples of functional inter-species homology between lncRNAs lacking profound primary sequence conservation are *Cyrano* and *Megamind* lncRNAs. **Cyrano** lncRNA possesses structural conservation and sequence similarity of one short region but does not have positional conservation. The 67nt sequence region of *Cyrano* is

conserved between human, mouse, and zebrafish. *Cyrano* also shares a similar gene structure between orthologs (two to three short exons followed by a long terminal exon of 4–8 kb) and a similar genetic environment. Zebrafish embryos depleted in *Cyrano* demonstrated abnormal morphological phenotypes, which were partially rescued by expression of human or murine *Cyrano* lncRNA orthologs, demonstrating that such partial sequence conservation is enough to compensate for function [156].

Megamind is a 2.4 kb transcript consistent of 3 exons, contains ~340nt region conserved among zebrafish and mammals. *Megamind* overlaps the intronic sequence of the protein-coding gene *birc6* in an antisense orientation. Deleterious effects in the brain and eyes of zebrafish embryos after *Megamind* knockdown were completely alleviated in embryos co-injected with human or murine *Megamind*, demonstrating functional homology between three distant species [156].

While it is reasonable to suggest that many lncRNA orthologs have changed or lost their initial function throughout evolution, many lncRNA have been reported to play essentially the same roles in various species, suggesting that there are yet more functionally significant and evolutionarily conserved lncRNAs to be uncovered. **NEAT1** lncRNA was demonstrated to be essential for paraspeckle formation in various organisms, including human, mouse, and opossum [161,162]. lncRNA regulator of cardiac cells differentiation **CARMEN** is conserved across vertebrates and its human and murine homologs are expressed from orthologous genomic loci, exhibit high conservation of its promoter region, and perform a conserved function in the pathological remodeling of human and mouse hearts [163]. **Xist** lncRNA is essential for X chromosome inactivation in both human and mouse [160].

Cases of inter-species conserved lncRNAs deviating from their original function have also been demonstrated. In human lncRNA **HOTAIR** negatively regulates transcription of a HoxD locus [89], while in mice HoxD cluster was completely unaffected after by HOTAIR deletion, suggesting that lncRNA rapidly evolved in mammals with the acquisition of novel functions [164]. Another example is maternally expressed imprinted genes murine **Glt2** lncRNA and human **MEG3** lncRNA, which are homologs identified using gene trapping [165]. Despite significant differences in the size of the lncRNAs (~7kB for Meg3 and ~2kB for Glt2) MEG3 and Glt2 have the same number and sizes of their exons, therefore sharing both positional and structural homology. Human MEG3 has been found to participate in tumor suppression via activation of p53 in human colon cancer cells [166], while murine Glt2 plays role in the embryonic development of skeletal muscles via its implication in Dlk2-Glt2 locus imprinting in mice [167].

To conclude, even though lncRNA inter-species functional homology of should be experimentally verified for each lncRNA, the larger the number of analogous features genes share (like a position in the genome, genomic environment, sequence homology, gene structure, etc) the higher the chances orthologous lncRNAs also perform evolutionarily conserved functions.

2.7 Morrbid lncRNA possess positional and partial sequence conservation between human and mouse

Murine lncRNA Gm14005 and human lncRNA MIR4435-2HG were first mentioned as inter-species homologs by Kotzin and colleagues [51]. Kotzin et. al. also

found that this lncRNA is actively expressed in mature eosinophils, neutrophils, and classical monocytes where it inhibits expression of pro-apoptotic Bim protein and suggested the name for it: Morrbid (myeloid RNA regulator of Bim-induced death). Hereafter we are going to refer to Gm14005 lncRNA as mMorrbid (murine Morrbid) and to human MIR4435-2HG lncRNA as hMorrbid (human Morrbid). RNA-seq data from human neutrophil, mouse granulocyte and cow peripheral blood showed that Morrbid is conserved between species (Figure 12). Same study showed that Morrbid is mainly localized to the nucleus and is bound to chromatin according to CHIP-seq data. Analysis of mMorrbid and hMorrbid genes loci demonstrates that these lncRNAs share positional conservation and partial sequence conservation.

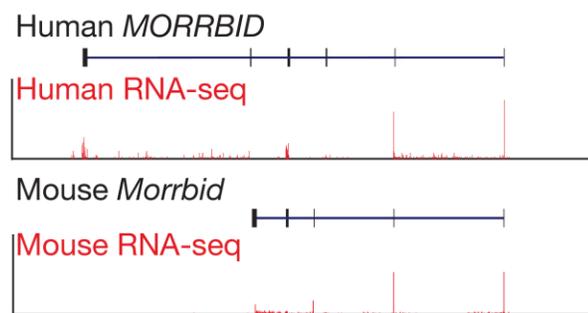


Figure 12. Human neutrophil and mouse granulocyte normalized RNA-seq tracks at Morrbid locus from [51] © 2020 Springer Nature Limited.

2.7.1 Morrbid lncRNA positional conservation between human and mouse

Both human and murine Morrbid are located on the second chromosome in reverse orientation and adjacent to the Bim gene. Morrbid has a highly similar genomic environment between human and mouse. In both species, Morrbid locus is located downstream from BUB1,

ACOXL, and BCL2L11 (Bim) and upstream from ANAPC1 and MERTK genes. ANAPC1 and BUB both are important for progression through the cell cycle, BCL2L11 or Bim acts as an inducer of apoptosis, ACOXL plays a role in peroxisomal lipid metabolism and MERTK is a transmembrane tyrosine kinase important for phagocytic pathway in retinal epithelium tissue. (Figure13).

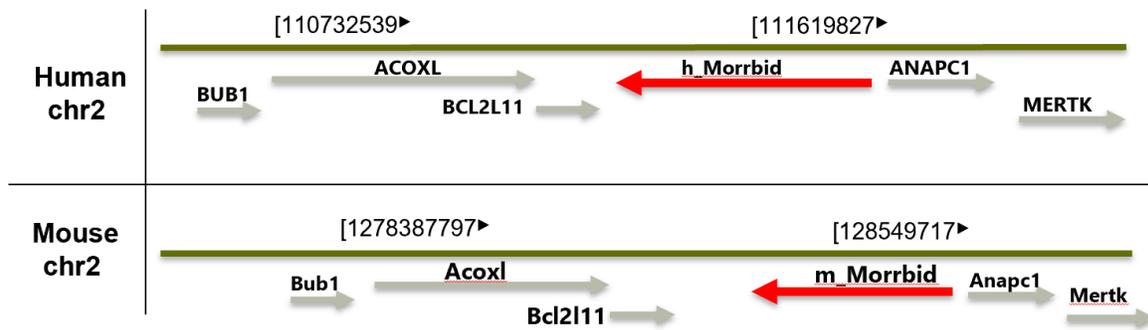


Figure 13. Schematic representation of genomic environments of murine *Morrbid* and human *Morrbid* (*MIR4435-2HG*).

mMorrbid genomic locus length is 3.24 Mb (position chr2:128,178,319-128,502,765) and hMorrbid is 4.86 Mb (positions chr2:111,036,776-111,523,376) (Ensembl human genome assembly GRCh38.p12 and mouse genome assembly GRCm38.p6).

The number, length, and exon composition of *Morrbid* transcripts vary greatly between different resources (genome browsers and papers). According to Ensembl genome browser hMorrbid potentially has 42 exons, which may assemble into 108 transcript models, while mMorrbid may have 33 exons with 50 potential transcript models.

Ensembl applies a method for classification of multi-exon transcript models, aimed to distinguish between well-supported and poorly supported transcript models, called Transcript Support Level (TSL). The TSL method utilizes alignment and comparison of transcript sequences between [GENCODE](#), [International Nucleotide Sequence Database Collaboration](#) (GenBank, ENA, and DDBJ), RNA alignments from Ensembl, BLAT RNA, and EST alignments from the [UCSC](#) Genome Browser Database. TSL is only applied to multi-exon transcripts. The method for evaluation of single-exon transcripts is still being developed by the community. We filtered Morrbrid transcripts in human and mouse based on their TSL classification, leaving only transcripts with TSL1,2 or 3 (TSL1 – all splice junctions of the transcript are supported by at least one non-suspect* EST; TSL2 – the best supporting EST is flagged as suspect* or the support is from multiple ESTs; TSL3 – the only support is from a single EST) (Figure 14) and single-exon transcripts since their quality cannot be assessed by the TSL method. Filtering leaves 12 hMorrbrid multi-exon transcripts and 3 single-exon and 7 mMorrbrid multi-exon transcripts, which are represented in Figure 14.

2.7.2 Morrbrid sequence conservation between human and mouse

Genomic DNA sequences of hMorrbrid and mMorrbrid genes have moderate conservation of the primary nucleotide sequence. ~13% of the hMorrbrid genomic sequence is aligned to mMorrbrid with ~69% identity. While comparison of exon sequences shows that 34% of mMorrbrid exon sequence aligned to the exon sequence of hMorrbrid with ~68% identity, which is higher than an average coding sequence conservation of known conserved lncRNAs [42].

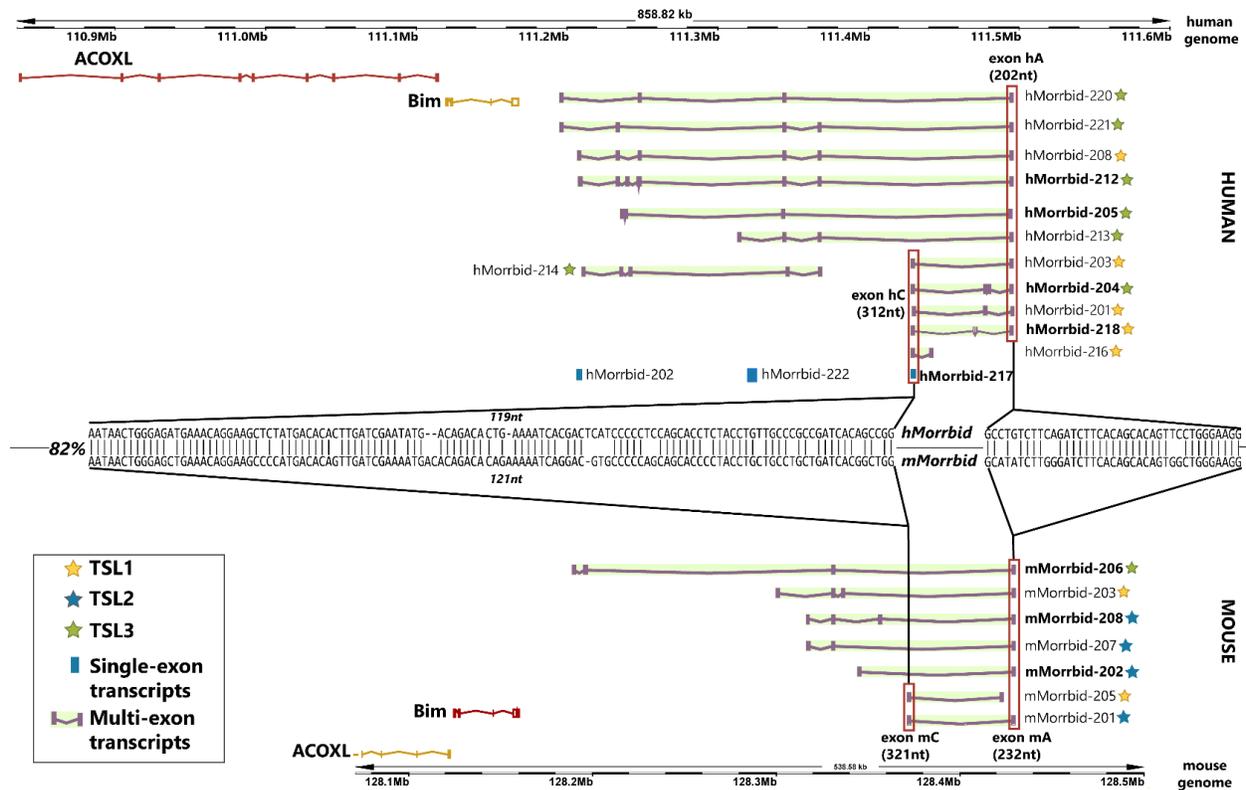


Figure 14. Comparison of Morrbid transcripts between human and in the mouse. Transcripts were filtered based on Ensembl TSL classification and only TSL1, TSL2, and TSL3 transcripts were mapped. Additionally, single-exon transcripts (blue boxes) are also represented on the figure, as they are not subjected to TSL classification. Morrbid has partial sequence homology of its exons A and C (framed in the red box) between human and mouse. 119 nt region of exonCh is aligned to the 121nt region of exonCm with 82% sequence homology.

The two most notable regions of sequence similarity are shared by two exons, which we named exonA and exonC. Human exonC is (312 nt long) and murine exonC (321 nt long) share 119/121 nt region with 82% of sequence identity. Human exonA (exonAh) and murine exonA (exonAm) overall are ~50% identical, sharing homological sequence patches the longest of which is 17nt long (Figure 14). Evolutionarily conservation of primary nucleotide sequences may serve as an indicator of the functional significance of exonA and exonC. Overall, positional

conservation and sequence conservation between hMorrbid and mMorrbid suggest that these lncRNAs are indeed inter-species orthologs.

2.8 Human Morrbid has a paralogue lncRNA named CYTOR

In the human genome lncRNA Morrbid has a paralogue lncRNA gene called CYTOR. hMorrbid and CYTOR are located on the different arms of the chromosome 2 about 24 Mb apart. hMorrbid gene is about 300 kb longer than CYTOR and is located on the antisense strand, while CYTOR is located on the sense strand (Figure 15).

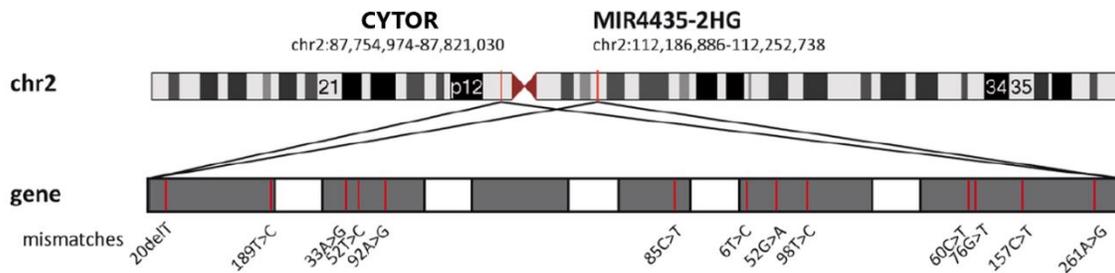


Figure 15. Genomic location of CYTOR and hMorrbid on chromosome 2. Adopted from [168] © 2020 Springer Nature Limited.

A region 99.3% identical to the full genome sequence of CYTOR is contained closer to the 5' end of the hMorrbid gene oriented in antisense direction (Figure 16). Both hMorrbid and CYTOR are actively expressed in human cells of various types.

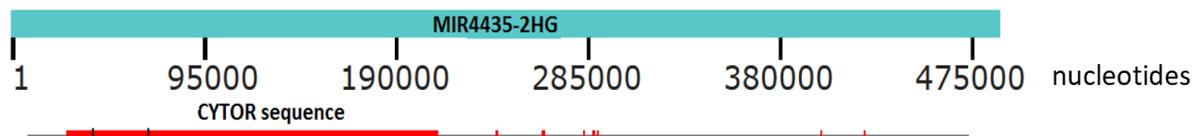


Figure 16. Alignment of *CYTOR* (red) and *hMorrbid* (blue) genomic sequences.

This duplication is only present in humans and *Morrbid* does not have any detectable paralogs in other species. It is known that human chromosome 2 evolved from the fusion of the two ancestral ape chromosomes – 2p and 2q [169]. Evidence suggests that this chromosomal fusion made adjacent chromosome 2 regions evolutionarily unstable as this area is enriched in pseudogenes and gene duplications [170]. Interestingly, *CYTOR* gene is located near the site of the fusion. It suggests that a peculiar genomic localization of paralogous genes *hMorrbid* and *CYTOR* on chromosome 2 might be a consequence of a partial duplication of *hMorrbid* genomic sequence in reverse orientation in humans.

hMorrbid and *CYTOR* first attracted attention of the scientific community due to their cancer-promoting properties and soon were discovered to have plenty of regulatory roles, like regulation of cell proliferation, migration, invasiveness, progression through the cell cycle, etc. Both lncRNAs are known under several different names, which must be considered when doing literature research (Table 3).

Table 3. Alternative names of hMorrbid/CYTOR lncRNAs used in literature.

CYTOR	LINC00152, NCRNA00152, C2orf59, MGC4677
hMORRBID	MIR4435-2HG, LINC00978, MIR4435-1HG, lncRNA-AWPPH, AGD2, AK001796

The question remains open whether paralogous lncRNAs perform the same or distinct functions. One hypothesis suggests that evolutionary selection against gene duplication takes place and is an explanation of the single-copy status of most of the genes. The acquisition of novel molecular functions would be a way to escape this evolutionary pressure [170] and the existence of paralogs may be a result of the acquisition of one or several gene-specific functions.

Unfortunately, most of the hMorrbid or CYTOR papers published to date do not take into consideration the paralogous gene in the design of the research tools (primers, shRNAs, siRNAs, etc.), creating confusion in the literature. For example, in paper [171] siRNAs and primers were designed to target both hMorrbid and CYTOR transcripts. However, results were reported for CYTOR only. In the paper, which discusses the regulation of Cyclin D1 in hepatocellular carcinoma [15] primers were designed to target specifically hMorrbid exons but not CYTOR exons, however, the paper only mentions CYTOR. Many articles lack sequences of primers that makes it impossible to verify which particular transcript was studied and if it is expressed from both genes or specific to one lncRNA gene. As a result, despite an increasing number of papers reporting functional roles, it is impossible to make certain conclusions about CYTOR and hMorrbid gene-specific functions. We will next describe previously discovered roles of hMorrbid and CYTOR genes, however, it has to be noted that none of the mentioned works made an effort to separately target paralogue genes and, therefore, mentioned functions should be assumed to be performed by both genes, which we will refer as hMorrbid/CYTOR.

2.9 Human Morrbrid and CYTOR lncRNAs are strongly associated with the progression of various cancers

LncRNAs hMorrbrid and CYTOR drew scientific interest after the burst of publications reporting a strong association between hMorrbrid/CYTOR expression and progression of cancer. The first publication dating back to 2014 [172] proposed CYTOR to be a biomarker for gastric cancer, was followed by many reports demonstrating upregulation of hMorrbrid/CYTOR levels both in patient-derived tissue samples and in cultured cell lines in a variety of cancers: gastric cancer [9–13], hepatocellular carcinoma [14–19], lung cancer [20–25], colon cancer [26], gallbladder cancer [27], glioma [28], ovarian sarcoma [29,30], pancreatic cancer [31], clear renal cells sarcoma [32,33], breast cancer [34–37] (Table 4). In a recent study [10], samples of gastric cancer tissues and adjacent healthy tissues collected from 150 patients (mixed age and gender), who underwent surgical removal of the tumor, were analyzed and compared. hMorrbrid/CYTOR expression was found to be significantly elevated depending on the stage of the disease and in correlation with a lower survival rate. Another group of 72 patients demonstrated upregulation of hMorrbrid/CYTOR expression and correlation with the size of the tumor, however not with other parameters, like tumor number, differentiation grade, or TNM stage [13].

Overall, a typical hMorrbrid/CYTOR loss of function phenotype in cancerous cells lines is characterized by: decreased cell proliferation [9,10,12,13], lower cell invasion [9,12], decreased colony formation [10], slower cell migration [9,12], inhibition of EMT (epithelial to mesenchymal transition) [9,10], increase in apoptosis [9,10,12,22,23,25], cell cycle arrest in G1 phase [9,12]. On the other hand, cells transfected with vectors stably overexpressing hMorrbrid/CYTOR demonstrated an increased rate of proliferation

[10]. Multiple experiments with xenograft models in nude mice demonstrated that subcutaneous injection of tumor cells with stable hMorrbid/CYTOR KD results in tumors smaller in size and volume in comparison to wild-type xenografts [12,13]. On the opposite, xenografts created with gastric cancer cells stably overexpressing hMorrbid/CYTOR led to increased growth of tumors *in vivo* [10,22] (Table 4). Just several studies had deviations from these common phenotypical trends. For instance, knockdown or overexpression of hMorrbid/CYTOR in glioblastoma cell line U87 did not result in any changes in cell proliferation, however, it did affect cell invasion [173].

Overall, there is overwhelming evidence from the literature that hMorrbid/CYTOR promotes tumorigenesis of multiple cancers.

Table 4. *hMorrbid/CYTOR* roles in cancers.

	Correlation between cancer progression and hMorrbid/CYTOR expression in cancer patients	hMorrbid/CYTOR expression in cell lines	hMorrbid/CYTOR knockdown phenotype	hMorrbid/CYTOR overexpression phenotype
Liver cancer [14–19]	-Increased in HCC and significantly related to the tumor size (90 patients) [14]. -Associated with HCC progression, tumor size, survival prognosis (102 samples [16], 72 pairs of samples [18] and 88 pairs of samples [19] derived from HCC patients) -Not associated with tumor metastasis [17]	Upregulated in HepG2, MHCC-97H, Huh7, SMMC-7721, Hep3B, and SNU-423 cell lines [16].	-Decreased cell viability, proliferation, and migration [15,16,18] -Upregulation of E-cadherin protein levels and downregulation in N-cadherin and Vimentin protein levels (inhibition of EMT) [18]	Enhanced cell viability, proliferation, and migration [19].

Gastric cancer [9–13]	<ul style="list-style-type: none"> -Correlated with survival rate (150 patients) [10] -Correlated with the size of the tumor, but not the tumor number, differentiation grade, or TNM stage (72 patients) [13] -Elevated in blood serum of cancer 72 cancer patients [12] 	Upregulated in cell lines HGC-27[9,12,13], SGC-7901 [9,10,12], BGC-823 [10,12], AGS[10], MGC-803[10,12,13]	Decreased cell proliferation [9,10,12,13], lower cell invasion [9,12], decreased colony formation [10], slower cell migration [9,12], inhibition of EMT [9,10], increase in apoptosis [9,10,12], cell cycle arrest in G1 phase [9,12].	-Increased rate of proliferation [10].
Breast cancer [34–37]	<ul style="list-style-type: none"> -Upregulated in TBNC (triple-negative breast cancer) and correlated with late TNM stage and node metastases in 52 patients [34] and 68 patients [37] - Negatively associated with hormone receptor status of breast cancer patients [34,35] 	-Elevated in MCF7, T74D, MDA-MB-468 and MDA-MB-231, but not BT474 cell lines[34,35]	<ul style="list-style-type: none"> -Decreased colony formation and cell invasion of cells [36] - Increase in apoptosis [36] -Reduced growth of xenograft tumors [36] 	-Promoted breast cancer cells proliferation [37]
Clear renal cells sarcoma [32,33]	Upregulated in cancer tissues. Associated with the TNM stage and number of lymph node metastasis. Predictor of overall survival (in 77 patients) [32,33]	Elevated in cell lines 786-O, ACHN, A498, Caki-1, and Caki-2 [32,33]	-Decreased cell proliferation and colony formation. Arrest in G0/G1 phase [33]	-Significantly promoted cell proliferation, colony formation, higher migration. A decrease in the number of cells in G0/G1 phase [32,33]
Ovarian sarcoma [29,30]	<ul style="list-style-type: none"> -Elevated expression in plasma, correlated with tumor metastasis but not tumor size [29] -Increased expression in cancer tissues and serum of OC patients. Associated with tumor size and distant metastasis [30] 			Promoted cell proliferation, invasion, and migration [29,30]

Lung cancer [20-25]	Associated with the presence of distant metastasis in patients [20,21] (based on 138 biopsies and blood serum sample pairs)		-Decreased proliferation, G1 phase arrest of cell cycle [22,23,25] and enhanced apoptosis [23]	-Increased cell proliferation and EMT [22,23,25]
Colon cancer [26]	Upregulated in colon cancer and significant negative correlation with patient's survival (138 patients) [26]	Higher expression in RKO, SW480 and SW620 cells and lower in HCT116 and HCT8 cells	Decreased colony formation, cell migration, and invasion -Inhibition of EMT -Decreased growth of xenograft tumors	Promoted colony formation, migration, and invasion -Promotion of EMT via E-cadherin inhibition and increase of Vimentin expression
Gallbladder cancer [27]	Upregulated in cancer tissues positively correlated with tumor status progression and lymph node invasion [27]	Upregulated in GBC-SD and NOZ cells	-Decreased cell migration and invasion, increased E-cadherin expression, decrease Vimentin expression [27]	Promotion of cell migration and invasion, decreased E-cadherin expression, increased Vimentin expression [27]

2.10 Human Morrbrid and CYTOR lncRNAs are important for liver cancer development and proposed as HCC biomarkers

For the first time, hMorrbrid/CYTOR were described as potential biomarkers for human hepatocellular carcinoma (HCC), together with HULC lncRNA (Highly Upregulated in Liver Cancer), in 2015 in a study which screened for promising HCC biomarkers among lncRNA based on gene expression analysis of overall 90 pairs of patient-derived HCC samples [14]. In this study, hMorrbrid/CYTOR levels were significantly related to the tumor size and ROC (receiver operating characteristic) curve analysis confirmed that implementation of hMorrbrid/CYTOR could be valuable in liver cancer diagnosis, with the highest predictive value of simultaneous analysis of hMorrbrid/CYTOR, HULC, and AFP. The following studies confirmed a significant association of hMorrbrid/CYTOR expression with HCC progression

based on groups of 102 samples [16], 72 pairs of samples [18], and 88 pairs of samples [19] derived from HCC patients. Expression of hMorrbid/CYTOR was higher in HBV positive samples compared to HBV negative samples [18] and was also higher in patients with portal vein tumor thrombosis (PVTT) [19]. The lncRNAs expression levels correlated with the tumor size and survival prognosis [16,17] but not with tumor metastasis [17]. Cox proportional hazards regression analysis was done based on 88 pairs of HCC tissues revealed that high hMorrbid/CYTOR expression in HCC tissues is an independent prognostic factor for poor recurrence-free survival and overall survival [19]. Elevated expression of hMorrbid/CYTOR was also confirmed in HCC cell lines relative to L02 normal human liver cell line: HepG2, MHCC-97H, Huh7, SMMC-7721, Hep3B, and SNU-423 [16].

Xenograft model experiments demonstrated that various liver cancer cell lines depleted in hMorrbid/CYTOR form tumors smaller in size and volume *in vivo* than their wild type counterparts, with reduced proliferation index Ki-67[15,16] (Figure 17). On the other hand, SMMC-7721 cells stably overexpressing hMorrbid/CYTOR had an enhanced growth rate and proliferation index *in vivo* [169]. Interestingly, hMorrbid/CYTOR expression was affected by levels of the key regulator of hepatitis B virus-driven hepatocarcinogenesis protein HBx, it was upregulated in cells overexpressing HBx and downregulated in cells with HBx inhibition [18].



Figure 17. Xenograft tumor in nude mice injected subcutaneously with HCCLM3 cells inhibited in *hMorrbid/CYTOR* shRNA grow slower than control [15] © 2018 Informa UK Limited.

One of the studies included 80 pairs of HCC samples derived from patients [15] specifically used primers detecting exons of *hMorrbid* and not exons of *CYTOR*. The results of the study can be used to conclude that at least some transcripts of *hMorrbid* alone are required for proliferation of cells *in vitro*, tumorigenesis of cells *in vivo*, and increased progression G0/G1 of the cell cycle.

Therefore, paralogs lncRNAs *hMorrbid/CYTOR* are important promoters of human liver cancer, and deciphering their specific role is important for potential implementation in combating this disease.

2.11 Human *Morrbid* and *CYTOR* lncRNAs were demonstrated to participate in a plethora of cellular networks.

Association of *hMorrbid/CYTOR* with cancer progression stimulated intensive research of their functional involvement in regulatory circuits. *hMorrbid/CYTOR* were found to promote cancer cell proliferation, migration, and invasiveness via positive regulation of growth signaling related pathways, like PI3K/AKT1, mTOR, and TGFbeta. EMT stimulating effect can be explained by the activation of EMT promoting pathways like NF-kb and Wnt, as well as direct

regulation of EMT proteins, like E-cadherin and SNAIL. Deregulation of the cell cycle was found to be related to modulation of expression of key cell cycle players: CyclinD1, p15, p21, and p16. Mechanisms of hMorrbid/CYTOR gene regulation are remarkably diverse. LncRNAs were found to localize both in the nucleus and in the cytoplasm, which enables them to perform a wide range of functions, like epigenetic regulation of transcription (e.g. EpCam, E-cadherin, p15, I124, etc.), scaffold for protein complexes (NCL, Sam68), promotion of target gene translation (SNAIL), upregulation of target genes expression by sponging inhibitory miRNAs, and even regulation of proteins post-translational modifications, like prevention of beta-catenin phosphorylation by CK-1 or promotion of PTEN ubiquitylation by NEDD4 ligase. Table 5 contains a summary of hMorrbid/CYTOR targets and proposed mechanisms of their regulation.

2.11.1 Regulation of cell cycle by hMorrbid/CYTOR

hMorrbid/CYTOR were demonstrated to drive progression through the cell cycle. They negatively regulate transcription of cell cycle inhibitors: p21 which induces cell cycle arrest in response to various stresses and p15 tumor suppressor gene, which inhibits progression through the G1 phase [174]. In addition to repressive action against cell cycle inhibitors, hMorrbid/CYTOR also positively modulate the expression of Cyclin D1.

Table 5. Regulation of cellular processes by hMorrbid/CYTOR.

hMorrbid/CYTOR regulatory role	Proposed mechanism	Type of tissue
Regulation of cell cycle progression	Promotion of CyclinD1 expression via sponging of miR-193b	Gastric cancer [10], HCC [15], Lung cancer [22]
	EZH2 dependent silencing of p21 promoter	Gastric cancer [174], non-small lung cancer [25]
	EZH2 dependent silencing of p15 promoter	Gastric cancer [174]
	EZH2 dependent silencing of p16 promoter	Renal cell carcinoma [33]
EMT	EZH2 dependent silencing of E-cadherin	Liver cancer [18]
	Promotion of SNAIL mRNA translation by YBX1	Liver cancer [19]
Inhibition of apoptosis	Positive regulation of MCL1 by sponging miR-193a-3p	Gastric cancer [171]
	Positive regulation of MCL1 by sponging miR-125b	Ovarian cancer [175]
Activation of PI3K/AKT1 signaling	Attraction of YBX1 protein to the promoter of PI3K kinase catalytic subunit PI3KCA	Liver cancer [19]
	Sponging miR-103a-3p to activate FEZF1/CDC25A axis	Glioma cells [28]
	Direct binding to EGFR	Gastric cancer [13]
	Promoting PTEN ubiquitinylation by NEDD4-1 for proteasomal degradation	Triple-negative breast cancer [34]
Activation of Wnt	Protection of beta-catenin from inhibitory	Colon cancer [39],

pathway	phosphorylation by CK1	ovarian cancer [30], lung cancer [176]
Activation of NF-kb pathway	Formation of a regulatory complex with NCL and Sam68 proteins	Colon cancer [29]
	Sponging miR-612 to elevate AKT2 levels	Glioblastoma [177]
	Sponging miR-155 to elevate IKKi levels	Cardiomyocytes [178]
Activation of mTOR pathway	Binding to EpCam promoter to upregulate its expression	HCC [16]
Cell proliferation	Enhancing the expression of HIF-1a by sponging miR-138	Gallbladder [27]
	Repression of IL24 via EZH2	Lung cancer [23]
	Upregulation of TGFbeta protein expression	Gastric cancer [12], Non-small lung cancer [20,21]

hMorrbid was found to increase Cyclin D1 levels in cells by sponging Cyclin D1 inhibitor miR-193b in hepatocellular carcinoma [15] and is induced by cigarette smoke lung cancer [22] (Figure 18). Overexpression of Cyclin D1 reversed the effect of reduced proliferation caused by hMorrbid/CYTOR depletion, suggesting that in HCC at least partially hMorrbid/CYTOR oncogenicity is explained via its promotion of Cyclin D1 expression [15].

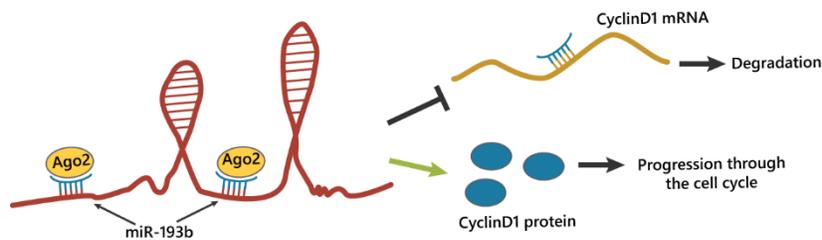


Figure 18. In HCC hMorrbid/CYTOR sponges miR-193b to promote expression of CyclinD1.

EpCam is a gene located in the genomic environment of CYTOR, known to promote tumorigenesis by driving expression of multiple oncogenes: C-MYC, Cyclins A, and E. In hepatocellular carcinoma, the luciferase reporter system demonstrated that hMorrbid/CYTOR bind to the promoter of EpCam driving its expression and expression of its target genes and leading to the activation of mTOR signaling pathway (Figure 19).

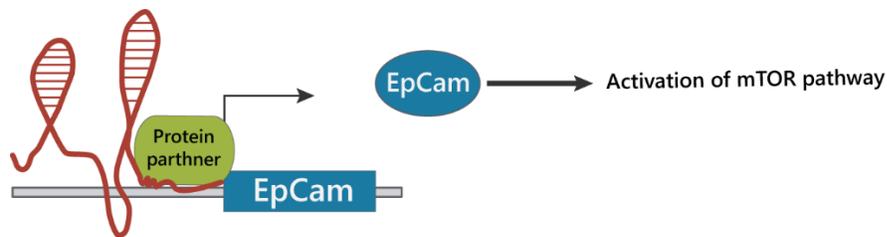


Figure 19. *hMorrbid/CYTOR interacts with EpCam promoter to drive its expression and subsequently expression of its target genes.*

Evidence suggests that hMorrbid/CYTOR regulates cell proliferation and progression through the cell cycle in part via its collaboration with EZH2. Reciprocal RNA Pulldown and RIP (RNA immunoprecipitation) assays revealed that hMorrbid/CYTOR directly interacts with EZH2 in HCC [19] and lung cancer [23].

EZH2 (Enhancer of zeste homolog 2) histone-lysine N-methyltransferase is an enzymatic component of the PRC2 (Polycomb repressive complex 2). PRC2 inhibits transcription of many genes by catalyzing trimethylation of a histone H3 at lysine 27 at the promoter of its target genes, which leads to epigenetic silencing of gene transcription. EZH2 binds many lncRNAs to establish a regulatory unit where lncRNA acts as a locus-specific “guide” acting both *in cis* and *in trans* and EZH2 as an “executor” performing locus silencing. EZH2 has a high RNA binding

affinity and was reported to interact with more than 9000 lncRNA in embryonic stem cells [179]. Another study, which performed global screening for EZH2 bound lncRNAs in different organs, demonstrated that the majority of identified lncRNAs interact with EZH2 in a tissue-specific manner [180]. hMorrbid/CYTOR was found to directly interact with EZH2 and facilitate its binding to E-cadherin promoter in hepatocellular carcinoma cell lines Huh-7 and SMCC-7721 promoting epithelial to mesenchymal transition [18]. In gastric cancer, proliferative effects of hMorrbid/CYTOR overexpression were reverted by knockdown of EZH2 confirming that at least part of hMorrbid/CYTOR functions depends on EZH2 [173]. In the same study CHIP analysis and functional assays confirmed hMorrbid/CYTOR and EZH2 coregulation of cell cycle proteins p15 and p21, which are both targets of EZH2 (Figure 20B).

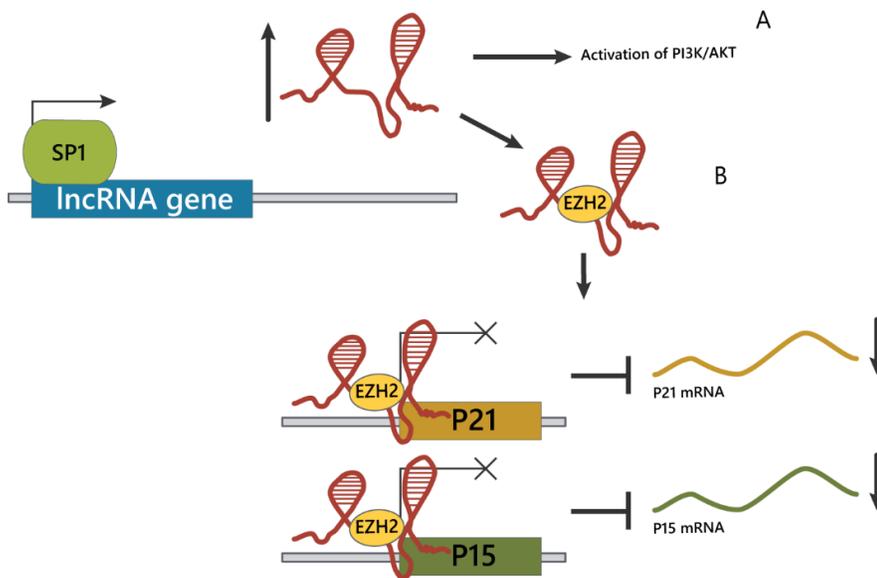


Figure 20. hMorrbid/CYTOR regulates p21 and p15. (A) SP1 binding to the CYTOR promoter leads to the upregulation of lncRNA expression in gallbladder cancer. (B) In gastric cancer tandem between hMorrbid/CYTOR and EZH2 binds to promoters of p21 and p15 to inhibit their expression.

Expression of the tumor suppressor p16, which detains progression through the cell cycle from G1 to S phase, was reported to be negatively regulated by EZH2-hMorrbid functional collaboration in the same manner in renal cell carcinoma [33]. In lung adenocarcinoma, hMorrbid/CYTOR expression is inversely correlated to IL24 expression. By binding to EZH2 lncRNA inhibits IL24 and promotes the transition to the G2/M stage of mitosis, cell proliferation, and reduces sensitivity to apoptosis [181].

2.11.2. Regulation of PI3k/AKT1 Pathway

PI3K-AKT1 pathway promotes metabolism, proliferation, and survival in response to extracellular signals. hMorrbid/CYTOR involvement in the activation of EGFR/PI3K/AKT1 pathway was demonstrated for gastric cancer [13], lung cancer [25], gallbladder cancer [182], and for liver cancer via attracting YBX1 to PI3KCA promoter [19].

It was demonstrated that transcription factor SP1 binds to CYTOR promoter leading to the upregulation of CYTOR expression in NOZ (gallbladder cancer) cell line, which potentially leads to the activation of the PI3K/AKT pathway as demonstrated by western blots [182] (Figure 20A).

M. Yu and colleagues demonstrated that hMorrbid/CYTOR sponges an oncosuppressor miR-103a-3p, leading to enhanced expression of its target gene FEZF1, which in turn was demonstrated to bind the promoter of CDC25A facilitating its transcription. CDC25A phosphatase is an essential catalyst for cell cycle progression and a well-known oncogene. hMorrbid/CYTOR /miR-103a-3p/FEZF1 axis dependent overexpression of CDC25A also resulted in activation of PI3K/AKT1 pathway [28]. Regulation of transcription is not the only functional outcome of YBX1- hMorrbid/CYTOR partnership. The tandem was demonstrated to

activate the cap-independent translation of SNAIL1, a transcriptional repressor of E-cadherin and promoter of epithelial and mesenchymal transition [19,183].

Another protein-dependent epigenetic gene regulation reported for hMorrbid is the regulation of expression of PI3K kinase catalytic subunit PI3KCA by attracting YBX1 protein to its promoter. It was demonstrated that YBX1 binds to the PI3KCA promoter and induces its transcriptional activation [184]. The importance of PI3KCA for cell oncogenic transformation and growth factors signaling was demonstrated and its enhanced expression correlates with increased activation of PI3K/Akt1 pathway, which enhances cell survival, proliferation, and tumorigenesis of cancer cells. hMorrbid/CYTOR facilitates YBX1 binding to the promoter of PI3KCA in hepatocellular carcinoma, promoting tumor growth and metastasis *in vitro* and *in vivo* [19].

PTEN phosphatase is a tumor suppressor gene, which negatively regulates PI3K/AKT1 signaling by dephosphorylating PIP₃ to inactive PIP₂, therefore preventing cells from the uncontrollable proliferation. hMorrbid/CYTOR decreases PTEN protein levels in triple-negative breast cancer by promoting its NEDD4-1 E3 ligase dependent proteasomal degradation of PTEN [34].

2.11.3 Regulation of Wnt Pathway

Wnt pathway activation is an important driver of EMT and cell proliferation. hMorrbid/CYTOR overexpression promoted b-catenin protein levels in ovarian cancer resulting in enhanced oncogenic properties of cells [30]. In a precise study on colon cancer cells, it was demonstrated that paralogous lncRNAs activate the Wnt pathway by direct binding to serine 45 of beta-catenin, preventing CK1 (casein kinase-1) access to its

phosphorylation site, therefore protecting beta-catenin subsequent degradation [39]. Similar conclusions were made in lung cancer, where it was found that hMorrbid/CYTOR directly binds to b-catenin hindering its proteasomal degradation [176], which led to increased tumor growth *in vivo*.

2.11.4 Regulation of NF- κ b pathway

NF- κ b pathway was demonstrated to be the major promoter of EMT in various cancers [185–187]. In colon cancer cells it was demonstrated that hMorrbid/CYTOR acts as a scaffold to form a regulatory complex with proteins NCL and Sam68, which are involved in NF- κ b activation, resulting in increased phosphorylation of p65 [26]. In glioblastoma, hMorrbid/CYTOR acted as a molecular sponge of miR-612 to impair its inhibition of AKT2, which led to activated NF- κ b pathway and enhanced proneural-mesenchymal transition [177]. Several studies indicate that hMorrbid/CYTOR also participate in the progression of non-malignant diseases. In pathological cardiac hypertrophy, hMorrbid/CYTOR was shown to have negative dual regulation with miR-155, which is a repressor of noncanonical I κ B-related kinase, IKBKE. hMorrbid/CYTOR sponges miR-155, which leads to the elevation of IKBKE levels and enhanced NF- κ b signaling [178].

2.11.5 hMorrbid/CYTOR in hypoxia

Hypoxia-inducible factor 1-alpha (HIF-1a) is a protein involved in metastasis formation. hMorrbid/CYTOR positively regulate HIF-1a via sponging miR-138, which targets HIF-1a for degradation, promoting EMT and metastasis in gallbladder cancer [27].

2.11.6 TGF β Pathway

TGF-beta pathway induces EMT in cancer cells facilitating cell migration and invasion. Moreover, mitogenic growth factors production stimulated by TGF-β promotes tumor proliferation and survival [188]. Overexpression of hMorrbid/CYTOR in lung cancer led to the upregulation of protein levels of TGF-b [20,21], while in gastric cancer inhibition of TGF-beta/SMAD2 axis was demonstrated in response to the lncRNAs knockdown [12], however, the mechanism of this process has not been studied.

2.11.7 Involvement in apoptosis

Another axis of hMorrbid/CYTOR malignant function is the inhibition of sensitivity of cancer cells to apoptosis. In several studies, protein levels of pro-apoptosis genes Bax and Bim were reported to be upregulated, while anti-apoptosis genes like Bcl-2, Bcl-xl, Mcl1 were downregulated after the lncRNAs knockdown in gastric and lung cancers [10,12].

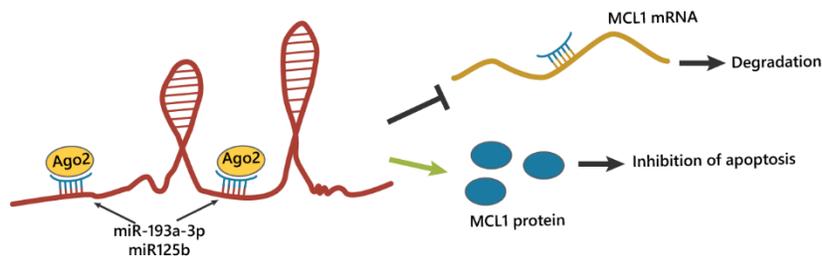


Figure 21. hMorrbid/CYTOR promotes expression of MCL1 by sponging miR-193a-3p in gastric cancer and miR-125b in ovarian cancer.

A study done in gastric cancer identified that hMorrbid/CYTOR enhances expression of MCL1 by acting as a ceRNA (competing endogenous RNA) on MCL1

repressor miR-193a-3p [171] (Figure 21). Another study done in ovarian cancer cells also found that MCL1 is positively regulated by hMorrbid/CYTOR in miRNA dependent manner, however, a different miRNA, miR-125b, was found to be a regulatory mediator [175].

2.12 Evidence that Human Morrbrid may exhibit transcript-specific functions

While studying literature on hMorrbrid/CYTOR lncRNAs it can be noted that they have an unusually broad range of functional roles both in the nucleus (epigenetic regulators and activators of genes transcription) and in the cytoplasm (post-translational regulation of proteins, protein scaffolding, sponging of miRNA). Such an abundant functional diversity hardly can be performed by a single transcript shuttling between the nucleus and cytoplasm. Most probably, hMorrbrid/CYTOR genes give rise to several functionally distinct transcripts, performing transcript-specific roles, some of which are nuclear bound and some cytoplasm bound.

In a favor of this assumption speaks an inconsistency between reports on the cellular localization of hMorrbrid/CYTOR. They were reported as both nuclear and cytoplasmic, mostly nuclear [16] or mostly cytoplasmic [13]. There is also a discrepancy in reports on the functional roles of hMorrbrid/CYTOR. A study done by Ji J. et al showed that invasiveness and cell cycle progression of hepatocellular carcinoma cell lines HepG2 and MHCC-97H are not affected by hMorrbrid/CYTOR inhibition [16]. On the other hand, different groups found an impaired invasion of Huh-7, SMCC-7721 and HepG2 cells [18] and arrest in the G1 phase of the cell cycle of Huh7 and HCCLM3 cells [15] associated with hMorrbrid/CYTOR knockdown.

Additionally, exons of hMorrbrid were demonstrated to be expressed at a different level in a study that compared gastric cancer patients to healthy controls, suggesting that hMorrbrid isoforms might have different oncogenic properties [11].

Other contradictions between reports suggest tissue specificity for the hMorrbid/CYTOR transcripts. In lung cancer, it was demonstrated that hMorrbid/CYTOR bind miR-193b but not miR-193a-3p to regulate Cyclin D1, while in HCC both miR-193a/b-3p were found to be sponged and deactivated by paralogous lncRNAs.

The duplicated genomic region covers the entire CYTOR locus and only part of the hMorrbid locus (Figure 16). As a result, every CYTOR exon has an identical twin exon in hMorrbid, while hMorrbid is a larger gene and has several gene-specific exons. According to the current Ensembl annotation, CYTOR has 38 transcripts, constituted of 24 exons and hMorrbid has 108 splice variants constituted of 43 exons. According to RefSeq CYTOR has 5 and hMorrbid 8 splice variants. Therefore, some transcripts of hMorrbid may possess isoform-specific functions, not interplaying with CYTOR transcripts.

2.13 In mouse Morrbrid lncRNA regulates inflammation response in immune cells

Both human and murine Morrbrid lncRNAs are located on the second chromosome in reverse orientation, have a similar genomic environment, and share conserved sequence fragments in exons. In the mouse genome, Morrbrid does not have a paralogue gene. The role of mMorrbrid lncRNA for the regulation of apoptosis in short-lived myeloid cells was intensively studied by J.J. Kotzin and colleagues [50,51]. They demonstrated that CRISPR/Cas Morrbrid knockout leads to a significant reduction in the population of eosinophils, neutrophils, and classical monocytes in murine blood due to

increased susceptibility to apoptosis. In the mouse genome, *Morrbid* is located close to the proapoptotic gene *Bim* and functions *in cis* to regulate *Bim* promoter through the recruitment of the polycomb repressive complex 2 (PRC2), which results in repressive H3K27me3 histone modification of *Bim* promoter (Figure 22). *Morrbid* dependent *Bim* inhibition was demonstrated to be triggered by LPS stimulation and happening in an allele-specific manner. Thus, the expression of a long noncoding RNA serves as a locus-specific controller of the life span of short-lived myeloid cells.

Another study demonstrated that hematopoietic stem cells (HPSCs) with the stable knockout of *Tet2* protein, which is frequently mutated in patients with myeloid malignancies, have improved resistance to inflammation-induced damage via NF- κ B/IL-6/Stat3/*Morrbid* signaling [52]. Preleukemic HSPCs contain mutations in AML (acute myelogenous leukemia)-related genes, such as *Tet2*. A single mutation by itself is not enough to cause leukemia and needs additional environmental factors, such as inflammation [52].

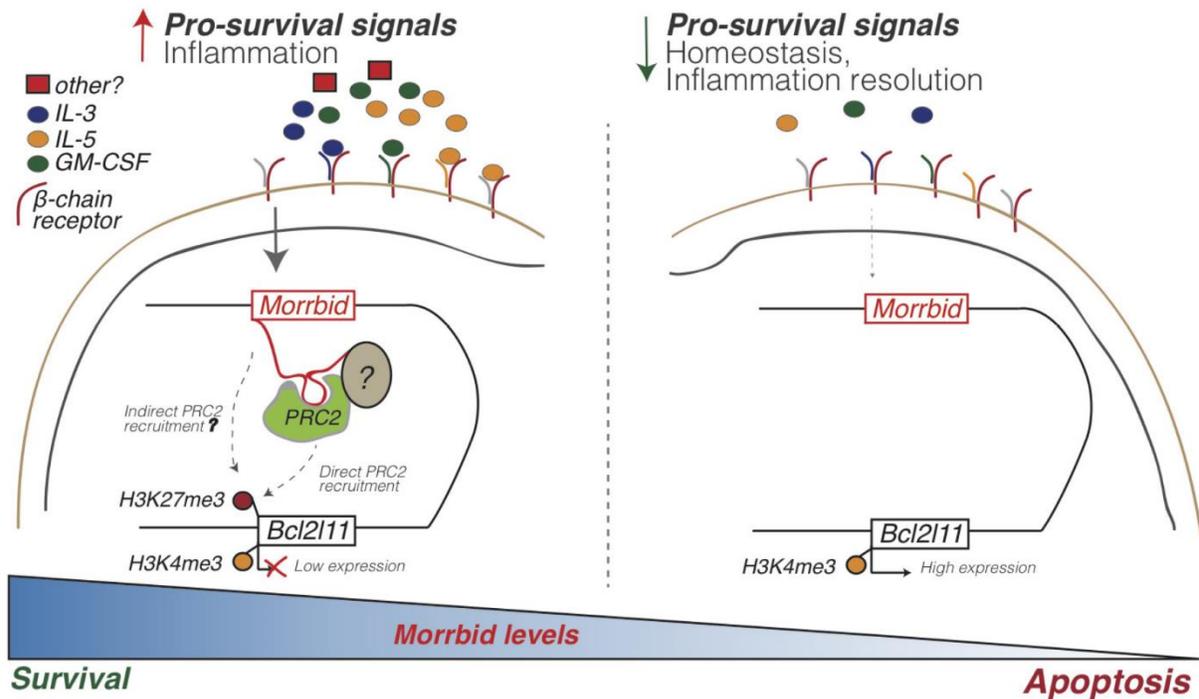


Figure 22. Mechanism of Morrbid regulation of Bim expression in mice. Morrbid integrates extracellular signals to control the lifespan of eosinophils, neutrophils, and classical monocytes through the allele-specific regulation of Bim (Bcl2l11). Pro-survival cytokines induce mMorbid expression, which promotes enrichment of the PRC2 complex within the bivalent Bcl2l11 promoter through direct and potentially indirect mechanisms to maintain this gene in a poised state. Adopted from [51] © 2020 Springer Nature Limited.

HSPCs from Tet2-KO mice had survival advantage and apoptosis resistance over HSPCs from wild-type mice after the LPS challenge, due to the activation of the Stat3 transcription factor and upregulation of Morrbid expression. The region in Morrbid promoter, which was demonstrated by Kotzin et al [51] to be essential for Morrbid activity, was found to bind Stat3 transcription factor [52]. While in short-lived myeloid cells and HSPC cells mMorbid enhances viability, in CD8 T cells TCR signaling-dependent activation of mMorbid expression leads to survival disadvantage via negative regulation of PI3K-Akt signaling [50]. The same group also evaluated the mMorbid role in leukemia tumorigenesis in vivo [190,191]. In mouse models of myelomonocytic leukemia (CMML), myeloproliferative neoplasm (MPN), juvenile myelomonocytic leukemia (JMML), and AML cellular levels of mMorbid were upregulated while Bim expression was inhibited [190,191]. Depletion of Morrbid in all mentioned mouse models resulted in increased Bim and improved overall survival [190,191].

Despite a thorough characterization of Morrbid lncRNA in myeloid cells, nothing is known about its functions in other cells and tissues. Many lncRNA demonstrated tissue-specific expression and functions. Thus, functions of lncRNA Morrbid may significantly differ between immune cells and hepatocytes. The present study is focused on uncovering mMorbid functions in murine hepatocytes.

Chapter III. Materials and Methods

3.1 Mammalian cell culture

Human hepatocellular carcinoma cell line Huh7, murine liver cell lines—normal AML12 (ATCC number: CRL-2254) and hepatoma Hepa1-6 (provided by Prof. O. Dontsova, Moscow State University, Moscow, Russia)—were cultured in Dulbecco's Modified Eagle Medium (or Nutrient Mixture F-12 (DMEM/F12) medium for AML12) (Thermo Fisher Scientific, Waltham, MA, U.S.) and supplemented with 10% fetal bovine serum (Thermo Scientific Gibco) and 1% penicillin-streptomycin (Thermo Fisher Scientific, Waltham, MA, U.S.). Transfections were done in the Opti-MEM Reduced Serum Medium (Gibco). Cells were maintained in a Water-Jacketed Autoflow Automatic CO₂ Incubator at 37°C with 5% CO₂ in DMEM medium. Reseeding with TrypLE Express Enzyme, phenol red (Thermo Fisher Scientific, Waltham, MA, USA).

3.2 Mammalian cells transfection

First, siRNA or ASO (Supplementary Table S1) were premixed with Lipofectamine RNAi or Lipofectamine 2000 (Thermo Fisher Scientific, Waltham, MA, U.S.), respectively, and opti-MEM reduced serum medium (Thermo Fisher Scientific, Waltham, MA, U.S.) according to the manufacturer's protocol. To perform cell transfection, the premix was added to adherent cells to achieve a 10-nM final concentration of ASO/siRNA. Cells were incubated for 1 day to measure mRNA knockdown efficacy, for 2 days to perform the knockdown of mMorrbid or for 6 days to perform SFPQ and UPF1 protein knockdowns, with repeated transfections at days 2 and 4. The efficacy of RNA downregulation was quantified by RT-qPCR and efficacy of protein

knockdown by Western blot analysis.

3.3 DNA isolation and PCR

Genomic DNA was isolated with QuickExtract DNA Extraction Solution, Lucigen. 20 μ L of the DNA Extract Solution is added to the 100k cells, then incubated consequently 65°C for 15 minutes, 68°C – 15 minutes, 98°C – 10 minutes. Then the mixture was centrifuged at 20,000 \times g for 5 minutes, the supernatant was transferred to a new tube and diluted with MQ water. DNA concentration was measured on NanoDrop One Microvolume UV-Vis.

For the PCR reaction, we used PCR Master Mix (2X) (Thermo Scientific, #K0171), Phusion High-Fidelity DNA Polymerase (Thermo Scientific, #F530S), DreamTaq DNA Polymerase (Thermo Scientific, #EP0701). The reaction was mixed and placed in Bio-Rad T100 PCR Thermal Cycler according to the manufacturer protocol.

3.4 DNA electrophoresis

Electrophoresis of PCR products and restricted plasmids was carried in 1,5% agarose gel with the addition of SYBR Green (Lumiprobe) in a TAE buffer under 100V. Analysis of product length with GeneRuler 1Kb or 100bp DNA ladder (Thermo Fisher Scientific, Waltham, MA, U.S.), images were obtained with ChemiDoc™ Imaging Systems, Life Science Research, Bio-Rad (Bio-Rad Laboratories, Inc., Hercules, California, USA).

3.5 RNA Isolation and RT-qPCR

Total RNA was isolated using TRIzol (Thermo Fisher Scientific, Waltham, MA, U.S.), according to the manufacturer's instructions. To remove any residual DNA from samples,

isolated total RNA (0.5 µg) was further treated with 1U of DNase I (Thermo Fisher Scientific, Waltham, MA, U.S.), supplied with a RiboLock RNase Inhibitor (0.4 U/µL). cDNA was synthesized using a Maxima First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, Waltham, MA, U.S.), followed by qPCR using a qPCR mix-HS LowROX Kit (Evrogen, Moscow, Russia), according to the manufacturer's protocols. qPCR was performed using primers listed in Supplementary Table S2. Positions of NRAS primers are shown in Figure 38C. Gapdh was used as a reference gene for the RT-qPCR analysis.

3.6 Primers for PCR and RT-qPCR

Primers for PCR and RT-qPCR were designed using the OligoAnalyzer and Primer-BLAST web tools and ordered from Evrogen, Moscow, Russia. PCR products were run on a 1,5%-agarose gel to confirm the length of the product. RT-qPCR primers efficiency was validated with a Standard Curve method, only primers with 80-120% efficiency were used for further functional assays.

3.7 Proteins isolation and concentration measurement

Cell extracts were prepared in triplicates and lysed using a radioimmunoprecipitation assay buffer (RIPA) (Sigma Aldrich, St. Louis, MO, U.S.) supplied with a 100xHalt Protease Inhibitor Cocktail (Thermo Fisher Scientific, Waltham, MA, U.S.), according to the manufacturer's protocol. The cell lysate was centrifuged at 14,000×g for 5 min at +4°C. The supernatant was transferred to clean tubes and placed on ice for 30 minutes. The total protein concentration in the lysate was measured using the Bradford Assay. 8 solutions of 10 times diluted RIPA with 0 to 2 mg/mL concentrations of BSA were prepared for the standard curve.

BSA standards and samples were mixed with Bradford Solution in ration 1:30 and then the optical density of each mix was measured on NanoDrop One Microvolume UV-Vis Spectrophotometer.

3.8 Western Blot

Cell extracts (~30 µg) were denatured in Laemmli buffer (Bio-Rad Laboratories Inc., Hercules, CA, U.S.) by heating at 95 °C for 10 min and separated by electrophoresis in 10% SDS—polyacrylamide gel using PageRuler Prestained Protein Ladder (Thermo Fisher Scientific, Waltham, MA, U.S.) as the standard. Proteins were transferred to PVDF membranes (Bio-Rad Laboratories Inc., Hercules, CA, U.S.) using a Mini Trans-Blot Cell and Criterion Blotter (Bio-Rad Laboratories Inc., Hercules, CA, U.S.) at 80 V for 40 min at RT in the transfer buffer (25-mM Tris, 250-mM glycine (pH 8.3) and 10% ethanol). Then, PVDF membranes were blocked by incubation in a 0.05% Tween 20-TBS solution (10-mM Tris–HCl, pH 7.5 and 150-mM NaCl) with 5% bovine serum albumin (Sigma Aldrich, St. Louis, MO, U.S.) at 4 °C overnight. The blocked membrane was incubated with primary antibody for 1 h at room temperature. After 3 washes with the 0.05% Tween 20-TBS solution, the appropriate secondary antibody was added for 1h. Membrane washed 3× with the 0.05% Tween 20-TBS solution and visualized in the ChemiDoc imager (Bio-Rad Laboratories Inc., Hercules, CA, U.S.). In the case of HRP-linked secondary antibodies using Clarity Western ECL blotting substrates (Bio-Rad Laboratories Inc., Hercules, CA, U.S.)). Images were analyzed using ImageJ software.

Antibodies:

Primary antibodies:

Ezh2 (D2C9) XP® Rabbit mAb from Cell Signaling, #5246

NF- κ B1 p105/p50 (D4P4D) Rabbit mAb from Cell Signaling, #13586

Bim Antibody from Cell Signaling, #2819

Bcl-xL (54H6) Rabbit mAb from Cell Signaling, #2764

Bax Antibody from Cell Signaling, #2774

Cyclin D1 (92G2) Rabbit mAb from Cell Signaling, #2978

GAPDH (14C10) Rabbit mAb Cell Signaling, #2118

anti-SFPQ SAB4200501 Sigma Aldrich, St. Louis, MO, U.S. (1:1000)

beta-Actin Monoclonal Antibody (15G5A11/E2) from Thermo Fisher Scientific, catalog # MA1-140, RRID AB_2536844 (1:5000)

alpha-Tubulin Monoclonal Antibody (B-5-1-2), Alexa Fluor 488 from Thermo Fisher Scientific for IHC.

Secondary antibodies:

Anti-rabbit IgG, HRP-linked Antibody from Cell Signaling, #7074

Goat anti-Mouse IgG (H+L) Secondary Antibody, DyLight 650 from Thermo Fisher Scientific, catalog # 84545, RRID AB_10942301.

Donkey anti-rabbit IgG H&L Alexa Fluor 680 ab175772, Abcam, Cambridge, U.K)

3.9. Separation of Nuclear and Cytoplasmic Cell Fractions

For separation of the cytoplasmic and nuclear fractions, cells were resuspended in a low-salt buffer (20-mM HEPES (pH 7.9), 10% glycerol, 1.5-mM MgCl₂, 0.05% NP-40 and 0.4-U/ μ L RiboLock RNase inhibitor). After intensive mixing, cells were kept on ice for 5 min and centrifuged at 1500 \times g for 5 min at 4 °C. Cytoplasmic fraction was collected as a supernatant. The remaining pellet was then resuspended in high-salt buffer (20-mM HEPES (pH 7.9), 10%

glycerol, 1.5-mM MgCl₂, 0.05% NP-40, 0.4-U/μL RiboLock RNase inhibitor (40 U/μL) and 0.5-M NaCl), kept on ice for 10–15 min and centrifuged at 12,000×g for 5 min at 4 °C to isolate supernatant with the nuclear fraction. Then, RNA was isolated both from cytoplasmic and nuclear fractions using TRIzol, as described above. GAPDH was used as a reference for the cytoplasmic fraction, and snRNA U6 was used as a reference for nuclear fraction.

3.10 PI staining of cells for the cell cycle analysis

Approximately 1.5mln cells were harvested for one sample, washed in PBS twice and fixed in the 75% ethanol and PBS, then placed at +4°C overnight. The next day fixed cells were centrifuged at 4000×g for 10 minutes, the precipitate was resuspended in 2 mL of PBS, then centrifuged again. Then cells were resuspended in 0,5 mL of PBS with 2.5 μg/mL RNase A and 1.5 μg/mL PI and incubated at RT for 30 minutes. Samples were protected from light and stored at +4°C. Analysis of the samples was done by Dr. Olga Sergeeva using Cell Analyzer Ze5, Bio-Rad, and further data analysis with FlowJo v.10, LLC.

3.11. Cell Viability Assay

Cells were plated into 48-well plates in triplicates, ~25*10³ cells per well. Viability was measured using 1% (w/v) resazurin sodium salt stock solution in PBS, which was further diluted 250-fold with PBS and added to adherent cells (preliminary washed twice with PBS), followed by a 1h incubation at 37 °C. Then, the fluorescent signal was measured using a Varioscan Microplate reader with a 540/590-nm (ex/em) filter set (Thermo Fisher Scientific, Waltham, MA, U.S.). To calculate the results, the ratio between tested samples and corresponding control samples was calculated for each time point and then normalized to the zero time point value. In

the case of AML12 and Hepa1-6 cells were transfected with mMorrbid or control ASO (final concentration 10nM) and the viability of cells was measured at 24h, 36h, 48h, and 72h time points after initial transfection.

3.12. Wound-Healing Assay

Cells were cultured in 6-well plates until confluence reached 70–80%, then several wounds with a width of approximately 1.5 mm were introduced in cell monolayers using a pipette tip, and a first image of the wounds was made at the zero time point. To trace the wound closure, pictures were taken at 24 h, 48h, and 72h after wound introduction in 6-8 replicas per each sample. The total wounding area was measured using ImageJ software, as was described previously [189]. The results were normalized to the corresponding wound areas at the zero time point. In the case of AML12 and Hepa1-6 cells, transfection was done with mMorrbid or control ASO (final concentration 10nM). The wound was introduced and 0h time point pictures were taken right after transfection,

3.13. Antisense Oligonucleotides (ASO) and Small Interfering RNA (siRNA)

To perform the downregulation of mMorrbid lncRNA expression, we designed 13 chemically modified gapmer ASO (Supplementary Table S1). First, the secondary structure of mMorrbid transcript variant 3 (NCBI Reference Sequence NR_028591.1) was simulated using ViennaRNA RNAfold WebServer [192]. Then, ASO binding sites in mMorrbid lncRNA were chosen based on the accessibility within the secondary structure. We aligned RNase H cleavage sites at loops, multiloops, or internal loops. ASO were synthesized as gapmers to improve exonucleolytic stability and binding to the RNA target. Ten central nucleotides were

phosphorothioate 2'-deoxynucleotides to maintain RNase H catalytic activity, while 5'- and 3'- flank four to six nucleotides—2'-OMe phosphorothioate nucleotides. The control ASO was designed to target the Firefly Luciferase gene. Additionally, we designed six chemically modified siRNAs (Supplementary Table S1) targeting the murine SFPQ sequence (NCBI Reference Sequence NP_076092.1). The siRNAs were selected to avoid off-target activity based on several known criteria [193–195]. To estimate the off-target binding capacity, siRNA 19-mer sequences were screened against the RefSeq mRNA database. Specifically, siRNAs were filtered based on the number/positions of the mismatches in the seed region, mismatches in the non-seed region, and mismatches in the cleavage site position. Then, chosen candidate sequences were checked for the presence of known miRNA motifs and immune stimulatory sequence motifs [194] that should be avoided. Resulting siRNAs were further chemically modified with 2'-OMe pyrimidine nucleotides and single 3'-internucleotide phosphorothioates to reduce the immune response and off-target effects and increase stability against nucleases [194,196]. The control siRNA targets the Firefly Luciferase gene. siRNA against UPF1 were previously designed using the same protocol and tested in our lab. siRNA UPF1 target both murine UPF1 transcript variant 1 and transcript variant 2 (NCBI Reference Sequences NM_001122829.2 and NM_030680.3) (Supplementary Table S1). siRNA design was performed by Tatiana Prikazchikova.

3.14. Fluorescent in Situ Hybridization (FISH)

AML12 cells were cultivated on poly-L-lysine-coated microscopy glasses. First, cells were washed twice with phosphate-buffered saline (PBS) and fixed with 4% paraformaldehyde (PFA) (Sigma Aldrich, St. Louis, MO, U.S.) at room temperature (RT) for 20 min. Then, the cells were washed with PBS twice for 10 min and permeabilized for 10 min with 0.5% Triton X-

100 in PBS. Then, cells were washed with PBS for 10 min, followed by two washes in “FA wash” buffer (40% formamide, 2.4×SSC (36-mM sodium chloride and 36-mM sodium citrate in water)). Negative-control glasses were additionally incubated with RNase A at 37 °C for 1 h (200 ug/mL of RNase A in 2X SSC buffer (30-mM sodium chloride and 30-mM sodium citrate in water)). Then, all glasses with cells were placed in the hybridization chamber and stained with 40 ng of a Cy5 oligonucleotide probe TTGCCTGGAAAGTCACTTTG-Cy5 in the hybridization buffer (0.4% BSA (Sigma Aldrich, St. Louis, MO, U.S.), 36-mM sodium chloride and 36-mM sodium citrate in water and 44% formamide (v/v) supplied with a tRNA+ssRNA mix (5.5 µg/µL) overnight at 37 °C. After that, cells were washed twice with FA wash buffer, preheated up to 37 °C for 15 min, and then washed once with PBS at RT for 10 min. Thereafter, the cells were stained with 4',6-diamidino-2-phenylindole (DAPI) (Thermo Fisher Scientific, Waltham, MA, U.S.) (300 nM in DMF) for 2 min, washed with PBS, and studied by confocal microscopy. Confocal microscopy was performed using a Nikon A1+MP confocal imaging system using a Plan Apo 20×/0.75 Dic N objective (numerical aperture = 0.75, Nikon, Japan), Apo LWD 40×/1.15 S water immersion objective (numerical aperture = 0.15, Nikon, Japan), and Apo tirtf 60×/1.49 DIC oil immersion objective (numerical aperture = 0.49, Nikon, Japan). Images were scanned sequentially using 561-nm diode lasers in combination with a DM561-nm dichroic beam splitter. FISH slides were analyzed by Pavel Melnikov (Serbsky National Medical Research Center for Psychiatry and Narcology).

3.15. RNA-seq Data Processing and Analysis

For transcriptome analysis, we used $\sim 6 \times 10^6$ AML12 cells per sample after 48h of ASO-mediated Morrbid KD or control LUC KD; 4 replicates per experiment were used. Total RNA

was extracted using a TRIzol reagent (Thermo Fisher Scientific, Waltham, MA, U.S.) according to the manufacturer's instructions. Six micrograms of total RNA (quantified using a NanoDrop OneC Spectrophotometer ((Thermo Fisher Scientific, Waltham, MA, U.S.))) was fragmented using conditions optimized to result in average 200-nt RNA fragments: incubation for 7 min at 95 °C in RNA fragmentation buffer (100-mM Tris (pH 8.0) and 2-mM MgCl₂). Fragmented RNA was purified by precipitation in 100% ethanol with a 1/10 volume of 3-M sodium acetate, and 1 µg of RNA (measured using a NanoDrop OneC Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, U.S.) was used for an rRNA depletion reaction using a NEBNext rRNA Depletion Kit (NEB E6310L, New England Biolabs, Ipswich, MA, U.S.) according to the manufacturer's protocol. Then, the RNA solution was diluted with a 1/10 volume of 3-M sodium acetate, RNA precipitated with ethanol, and 300 ng of RNA (measured using a NanoDrop™ OneC Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, U.S.) were used for the sequencing library preparation with a NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB 7760, New England Biolabs, Ipswich, MA, U.S.), according to the manufacturer's protocol, and the resulting double-stranded cDNA was purified using AMPure XP magnetic beads (A63881, Beckman Coulter, Brea, CA, U.S.). Efficient concentrations of libraries were determined using RT-qPCR. Library quality (length distribution and the absence of primer-dimers) was assessed on a Bioanalyzer2100 (Agilent Technologies, Santa Clara, CA, U.S.).

Libraries were pooled in equal amounts and sequenced using a HiSeq4000 (Illumina, San Diego, U.S.) instrument in 50-nt single-read mode, according to the manufacturer's protocol. Conversion to fastq format and demultiplexing was performed using bcl2fastq2 software (Illumina, San Diego, U.S.). Morrbid lncRNA KD and control KD samples were sequenced,

returning a variable number of paired reads.

For mapping, those samples, genome annotations were obtained from Ensembl. Paired-end reads were mapped using STAR v2.5.3a [197] with default settings, except for the following one: `--quantMode GeneCounts`. The resulting gene counts were further processed with R package DESeq2 [198], where it was further normalized using the RLE method. Additionally, to take into account unwanted data variations, we introduced additional variables, obtained by sva package [199] that capture the unwanted variations into a design matrix. The DESeq2 package was used for performing a differential expression analysis based on the Wald test. We defined genes as differentially expressed if they passed the thresholds: $FDR < 0.05$ and $|\log_2\text{foldChange}| > 0.5$.

For the alternative splicing analysis, we mapped all RNA-seq reads using the HISAT2 (v2.1.0) program [200], with the `--no-softclip` parameter on the mouse genome (assembly GRCm38) using the splice site coordinates from the Ensembl annotation. Then, the data was processed using the SAJR pipeline [201]. Briefly, each gene was split into the regions between two adjacent splice sites (segments)—based on the exon/intron coordinates from the Ensembl annotation. Alternative segments (that are included in some transcripts and excluded from others) were classified into three types according to the combinations of types of splice sites that define their borders: (i) cassette exons are segments that start from acceptor sites and end with donor sites, (ii) alternative acceptor (donor) segments are segments that both start and end with acceptor (donor) sites and (iii) retained introns are segments that start at a donor site and end with an acceptor site. For each segment and each sample, we calculated the number of inclusion reads (i.e., reads that overlap exons by at least one nucleotide) and the number of exclusion reads (i.e., reads that are mapped to exon-exon junctions that span a given segment). Reads mapped to multiple genomic locations were excluded from the analysis (i.e., only uniquely mapping reads

were used). The percent spliced in (PSI, a fraction of transcripts of a given gene that includes the exon) was calculated based on the inclusion and exclusion reads counts. Inclusion and exclusion read counts were modeled using the generalized linear model (GLM function in R) with a binomial distribution. Segments with FDR-corrected $p < 0.05$ (test for quasi-likelihood) were considered significant. Transcriptome analysis was performed by Ilya Kurochkin. Bioinformatic analysis of the alternative splicing was executed by Pavel Mazin.

3.16. Transcripts Degradation Rate Measurement

To measure the degradation rates of NRAS PTC and NRAS total transcripts, $\sim 1.5 \times 10^5$ AML12 cells were seeded in 12-well plates and treated with actinomycin D (final concentration 5 $\mu\text{g}/\text{mL}$) for 1, 2, 3, 5, 6, or 9 h. After incubation, cells were harvested, and RNA was extracted with a TRIzol reagent (Thermo Fisher Scientific, Waltham, MA, U.S.) using the manufacturer's protocol, followed by RT-qPCR analysis. To estimate the half-lives of the NRAS PTC and NRAS total transcripts, we used a nonlinear regression curve to fit experimental data points to the function $y(x) = \exp[-K \times (x - x_0)]$, where $-K$ is an exponential decay, and x_0 is the time offset. Nonlinear regression curves were plotted together with raw datapoints; the standard deviation was calculated based on 3 replicates. For calculations, GraphPad Prism 6.0 software was used.

3.17. Formaldehyde-Crosslinked RNA-Immunoprecipitation (RIP)

Two 10cm^2 plates with $\sim 5 \times 10^6$ AML12 cells were used to prepare each sample (performed in two replicates). Cells were harvested, resuspended in 2 mL of phosphate-buffered saline (PBS), and crosslinked by adding 37% methanol-free formaldehyde (143 μL) and

incubated for 10 min at room temperature. Crosslinking was terminated by the addition of 2M glycine in water (685 μ L). Cells were washed twice with ice-cold PBS, followed by centrifugation at 1000 \times g for 5 min at 4 $^{\circ}$ C. Cell pellets were resuspended in 1 mL of IP lysis buffer (50-mM HEPES (pH 7.5), 0.4M NaCl, 1mM EDTA, 1mM DTT, 0.5% Triton X-100, and 10% glycerol), with an added 20 μ L of 0.1M phenylmethylsulfonyl fluoride (PMSF), 10 μ L of a 100 \times Halt protease inhibitor cocktail (Thermo Fisher Scientific, Waltham, MA, U.S.) and 5 μ L of a RiboLock RNase inhibitor (40 U/ μ L) (Thermo Fisher Scientific, Waltham, MA, U.S.). The lysates were sonicated (10 s ON, 10 s OFF, amplitude 20 μ m, 10 cycles; QSonica sonicator, amplitude 20%) and centrifugated at 14,000 \times g for 3 min. Supernatant with crosslinked protein-RNA complexes was subjected to IP overnight at 4 $^{\circ}$ C with an anti-SFPQ (ab195352, Abcam, Cambridge, U.K.), anti-NONO (N8664, Sigma Aldrich, St. Louis, MO, U.S.), anti-DDX3 (A300-476A, Bethyl laboratories, Montgomery, TX, U.S.) or human IgG (control) antibody bound to pre-blocked Sepharose G-Beads (ab193259, Abcam, Cambridge, U.K.). Then, beads were 5 \times washed with the IP buffer, followed by a wash with the RIP buffer (50-mM HEPES (pH 7.5), 0.1-M NaCl, 5-mM EDTA, 10-mM DTT, 0.5% Triton X-100, 10% glycerol, and 1% SDS). Samples were incubated at 70 $^{\circ}$ C for 1 h and centrifugated at 1000 \times g for 5 min. RNA samples were extracted using TRIzol following the manufacturer's protocol and analyzed by reverse-transcription qPCR amplification using the primers listed in Supplementary Table S2 (List of PCR primers used in the study).

3.18. Capture Hybridization Analysis of Morrbid lncRNA Targets (CHART) [202]

For cell lysate preparation, $\sim 4 \times 10^7$ AM12 cells per each sample were washed with ice-cold PBS and then resuspended in a 1% paraformaldehyde solution in PBS. PFA crosslinking

was done under agitation for 20 min at RT and quenched by adding 1/10 of the volume of 1.25M glycine for 5 min at RT under agitation. Crosslinked cells were collected by centrifuging at 1000×g for 5 min and then twice-washed with PBS. Cells were lysed by adding 1 mL of WB100 buffer (100-mM NaCl, 10-mM HEPES (pH 7.5), 2-mM EDTA, 1-mM EGTA, 0.2% SDS and 0.1% N-lauroylsarcosine, supplemented with 1x protease inhibitors cocktail, 0.8μM PMSF and 4 μl of RNase inhibitor (40 U/μl)) per 100 mg of tissue. The lysates were sonicated (30 s ON, 30 s OFF, at 20% amplitude, 5 cycles; sonicator QSonica, Newtown, CN, U.S.), followed by 10 μl of RNase inhibitor (40 U/μl), 5 μl of 1M DDT in water and 5 μl of 100X protease inhibitors immediately after sonication.

For the CHART experiment, 250 μL of denaturant buffer (8M urea; 200mM NaCl; 100mM HEPES (pH 7.5); 2% (w/v) SDS) and 750 μL of 2X hybridization buffer (1.5M NaCl, 1.12M urea, 10X Denhardt solution and 10mM EDTA, pH 8) were added per 500 μL of lysate [202]. At this point, 100 μL from each sample was collected for the no-oligos control. For both the Morrbid CHART and for the control CHART, four biotinylated probes were used (Supplementary Table S3). Fifty-four picomoles of each probe per 100μL of extract were added to each sample and incubated at RT with gentle agitation for 7 h. After incubation, samples were centrifuged for 10 min at 16,000×g at RT, and the supernatant was transferred to a fresh tube supplemented with 50 μL of denaturant buffer and 100 μL of pre-rinsed MyOne Dynabeads Streptavidin C1 (65001, Thermo Fisher Scientific, Waltham, MA, U.S.). To allow a biotin-streptavidin interaction, samples were incubated overnight at room temperature with gentle agitation.

To minimize nonspecific binding, beads with Morrbid RNA-protein complexes were captured with the magnet and 5× washed with 1 mL of WB250 buffer. After the wash step, beads

were captured with the magnet capture and resuspended in 30 μ L of water. Samples were fractionated by SDS-PAGE; each Morrbid-specific band and corresponding gel position within the control lane were cut out of the gel and analyzed by mass spectrometry after the in-gel tryptic digestion of the proteins. Excised protein bands were cut into 1 \times 1 \times 1-mm cubes, transferred into sample tubes, and destained with 50% acetonitrile (ACN) in 100mM ammonium bicarbonate and dehydrated by the addition of 100% ACN. After ACN removal, samples were subjected to in-gel digestion by trypsin overnight at 37 °C. The digestion buffer solution contained 13-ng/ μ L Promega sequencing-grade modified trypsin in 10mM ammonium bicarbonate containing 10% ACN. The resulting tryptic peptides were extracted from the gel by adding two volumes (in comparison to the digestion buffer solution) of 0.5% aqueous trifluoroacetic acid and incubating for 1 h. Then, an equal volume of ACN was added, and the samples were incubated for another hour. The samples were vacuum-dried and dissolved in a solution containing 3% ACN and 0.1% aqueous TFA before LC-MS/MS analyses.

For the LC-MS/MS analysis, peptides were separated on a 50-cm 75- μ m inner diameter column packed in-house with Aeris Peptide XB-C18 2.6- μ m resin (Phenomenex, Torrance, CA, U.S.). Reverse-phase chromatography was performed with an Ultimate 3000 Nano LC System (Thermo Fisher Scientific, Waltham, MA, U.S.), which was coupled to a QExactive HF benchtop Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, U.S.) via a nanoelectrospray source (Thermo Fisher Scientific, Waltham, MA, U.S.). The mobile phases were: (A) 0.1% (v/v) formic acid in water and (B) 0.1% (v/v) formic acid, 80% (v/v) acetonitrile and 19.9% (v/v) water. Samples were loaded onto a trapping column (100- μ m internal diameter, 20-mm length and packed in-house with Aeris Peptide XB-C18 2.6- μ m resin (Phenomenex, Torrance, CA, U.S.)) in mobile phase A at the flow rate 6 μ l/min for 5 min and eluted with a

linear gradient of mobile phase B (5–45% B in 60 min) at a flow rate of 350 nl/min. The column temperature was kept at 40 °C. Peptides were analyzed on the QExactive HF benchtop Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, U.S.), with one full scan (300–1400 m/z, R = 60,000 at 200 m/z) at a target of 3e6 ions, followed by up to 15 data-dependent MS/MS scans with higher-energy collisional dissociation (HCD) (target 1e5 ions, max ion fill time 60 ms, isolation window 1.2 m/z, normalized collision energy (NCE) 28%, underfill ratio 2%) detected in the Orbitrap (R = 15,000 at fixed first mass 100 m/z). Other settings: charge exclusions—unassigned, 1 and > 6; peptide match—preferred; exclude isotopes—on and dynamic exclusion—30 s was enabled. MS raw files were analyzed by PEAKS Studio 8.5 (Bioinformatics Solutions Inc., Canada) [203], and peak lists were searched against the Uniprot-Tremble FASTA (canonical and isoform) database version of March 2018 (84,951 entries) with methionine oxidation and asparagine and glutamine deamidation as the variable modifications. The false discovery rate was set to 0.01 for the peptide-spectrum matches and was determined by searching a reverse database. The enzyme specificity was set to trypsin in the database search. Peptide identification was performed with an allowed initial precursor mass deviation up to 10 ppm and an allowed fragment mass deviation of 0.05 Da. LC-MS analysis was done by Rustam Ziganshin (Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry).

3.19. Affinity Pulldown of Biotinylated RNA for the Detection of NRAS mRNA-Protein Complexes

To prepare the NRAS total and NRAS NMD minigenes, we performed cDNA synthesis and amplification of the total RNA from wild-type AML12 cells using Pfu DNA polymerase (EP0501, Thermo Fisher Scientific, Waltham, MA, U.S.) and primers from Supplementary Table

S2. To synthesize NRAS minigenes, we used a forward primer containing the T7-promoter sequence at its 5' end and a gene-specific reverse primer; for control reverse NRAS minigenes, we used a gene-specific forward primer and reverse primer containing the T7-promoter sequence at its 3' end (Figure 40B). Two products of each PCR reaction—one corresponding to the NRAS total transcript and the second corresponding to the NRAS NMD transcript—were purified from the gel and confirmed by Sanger sequencing. To create the biotinylated NRAS minigenes, an in vitro T7 transcription was done using a Pierce RNA 3' Desthiobiotinylation Kit, (20163, Thermo Fisher Scientific, Waltham, MA, U.S.) according to the manufacturer's instructions, with the incubation time 4 h at 37 °C. The resulting biotinylated RNA products of the T7 transcription reaction were treated with DNase I to digest any traces of DNA and purified by 7.5% denaturing polyacrylamide gel (PAGE). Bands were excised, crushed, and soaked in the buffer (500-mM sodium acetate, 89-mM Tris, 89-mM boric acid, and 2-mM EDTA, pH 8.3) at 4 °C overnight; the supernatant was collected, and RNA was precipitated by adding 3 volumes of ethanol and isolated by centrifugation for 1 h at 10,000×g. Purified RNA was used to perform an RNA pull-down assay using a Pierce Magnetic RNA-Protein Pull-Down Kit (20164, Thermo Fisher Scientific, Waltham, MA, U.S.) according to the manufacturer's instructions and then analyzed by RT-qPCR.

3.20 Generation of knockout cell line

For targeted deletion of CYTOR and MIR4435-2HG promoter regions pairs of gRNAs (Table 6) were designed using the CRISPR design tool (<http://crispr.mit.edu/>) and cloned into pX458 vector plasmid containing GFP gene as a selection marker (Addgene plasmid #48138 <http://n2t.net/addgene:48138>; RRID: Addgene_48138). Each guide sequence was cloned into

the sgRNA scaffold according to the Zhang Lab general cloning protocol [204]. The plasmid was restricted with BbsI enzyme. Ligation was carried out at RT for 10 minutes with Rapid DNA Ligation Kit (Thermo Fisher Scientific) according to the manufacturer's instructions.

Obtained plasmids were validated by Sanger sequencing and then used for transfection of hepatocellular carcinoma cell line Huh7 with Lipofectamine 2000 reagent (Thermo Fisher Scientific). Two days after transfection single cells were isolated for generation of single cell-derived clones by dilution plating as described in [205]. Expanded colonies were subjected to further analysis. Generation of and maintenance of the hMorrbid/CYTOR knockout cell line was done with the help of Maria Nazarova and Olga Sergeeva.

Table 6. gRNAs used for Cas9-mediated knockout of hMorrbid/CYTOR genes.

Name	Sequence, 5'-3'
gRNA2	TGTCCTTTAGTGTGACTGTC
gRNA1	ATCTTTGAATGCGACACTGG

Table 7. List of primers used for PCR and cloning.

Target	Forward Primer Sequence, 5'-3'	Reverse Primer Sequence, 5'-3'
gRNA_C1	CTGTCCTTTAGTGTGACTGTCAG	CTGACAGTCACACTAAAGGACAG
gRNA_C3	GATTATCTTTGAATGCGACACTGG	CCAGTGTCGCATTCAAAGATAATC
MIR17_KpnI	GCTGCTAGCTCCAAAATCACATGCCTTCA	
MIR17_XbaI	TCAGCGGCCGCGAGACTGGCCAGACAAATGG	

3.21 *Next-generation sequencing analysis of knockout cells*

582M paired input reads were filtered and trimmed by Trimmomatic [206] for adapter sequences and low-quality scored bases. 471M paired reads were mapped by bowtie2 [207] with max 5 alignments per reading and reported an overall 98% alignment rate. Mapping data analysis and visualization performed with Integrative Genomics Viewer [208], SAMtools [209].

3.22 *Flow cytometry analysis of AnnexinV-FITC/PI stained cells.*

Cells were stained with Dead Cell Apoptosis Kit with Annexin V FITC and PI, for flow cytometry (Thermo Fisher Scientific, Waltham, MA, U.S.) according to the manufacturer's instructions. Apoptosis was analyzed by flow cytometry with BD FACS Canto II Analyzer™ (BD Biosciences) by Viacheslav V. Senichkin (Faculty of Basic Medicine, MV Lomonosov Moscow State University, Moscow, Russia). Inhibitors S-63845 and ABT-737 were kindly provided by Boris Zhivotovsky laboratory (Faculty of Basic Medicine, MV Lomonosov Moscow State University, Moscow, Russia). Flow cytometry data were processed with FlowJo software by Olga Sergeeva.

3.23 *Bacterial culture*

Mix & Go competent *E. coli* strain from Zymo Research were cultured in Luria-Bertani Broth (LB): 1% Bacto tryptone; 0.5% Bacto Yeast Extract; 1% NaCl; Luria-Bertani Broth Agar (LB-Agar): 1% Bacto tryptone; 0.5% Bacto Yeast Extract; 1% NaCl; 1.5% Agar.

3.24 *Bacterial transformation*

Frozen competent *E. coli* Mix&Go cells were thawed on ice and mixed with 5ul of

ligation mixture, then kept on ice for 5 minutes and then 50µL of cells were seed on 100 mm Petry Dish with LB Agar with 100µg/mL of Ampicillin and maintained at 37°C in the shaker-incubator. The next day individual colonies were transferred in 2 ml of LB media and growth overnight. Next day cells were tested for plasmid presence by PCR of isolated DNA. Plasmids were isolated with the Evrogen Miniprep kit and then sequenced by the Sanger method by Genome Centre, Moscow, Russia.

3.25 Plasmid construction for the MIR4435-2HG-217 RNA overexpression

For cloning of MIR4435-2HG-217 transcript (Ensemble ID ENST00000603827.1) cDNA, we used pcDNA3.1 Mammalian Expression Vector. PCR of total Huh7 cDNA with primers containing KpnI and XbaI restriction sites was done (Table 7), and the product was purified from the mixture with the Evrogen Cleanup Mini kit. Restriction digestion of cDNA and plasmid was carried out with KpnI and XbaI enzymes from the Fast Digest Thermo Scientific kit according to the manufacturer's instructions. 10 µL of the reaction mixture was incubated at the appropriate temperature for 60 minutes. The restricted DNA product was analyzed by electrophoresis and purified from the gel with the Evrogene Cleanup Mini kit. Ligation (20ng of vector and 100ng of cDNA) was carried out at 22°C for 5 minutes with Rapid DNA Ligation Kit (Thermo Fisher Scientific) according to manufacturer's instructions. The validated plasmid was used to obtain inverted control. The 300nt-long segment of MIR4435-2HG-217 (M-217) transcript was PCR out and then ligated to its location as a reverse complement sequence to create an R-217 vector. Selection was done based on zeocin resistance. The optimal concentration of the antibiotic (900µg/mL) was defined by plating Huh7 cells and treating them with zeocin containing media in concertation range 20µM – 200µM. Vectors M-217 and R-217

were transfected into hMorrbid/CYTOR knockout cells with Lipofectamine 2000 reagent. Cells were selected for 4 weeks in 50 μ M Zeocin to establish cell lines with stable M-217 and R-217 overexpression. Generation M-217oe and R-217oe cell lines were done by Maria Nazarova and Olga Sergeeva.

3.26. Statistical Analysis of the Experimental Data

Most of the diagrams are based on at least three independent experiments. Statistical data processing was performed using the GraphPad Prism software (version 6) with a two-sample t-test. The data were considered statistically significant at $p < 0.05$.

Chapter IV. Results. Characterization of hMorrbid/CYTOR knockout and investigation of exonCh functional significance

The work described in this chapter was performed by the author at Timofei Zatsepin's laboratory, at the Center of Life Sciences, Skolkovo Institute of Science and Technology. The author performed all molecular biology experiments described in this chapter (except for knockout PCR verification); cell culture work; generation of knockout cell line; sample preparation for LC-MS, flow cytometry, and microscopy analysis. Olga Sergeeva (Center of Life Sciences, Skolkovo Institute of Science and Technology) designed and cloned guide RNAs for CRISPR-Cas9 knockout. Maria Nazarova (Center of Life Sciences, Skolkovo Institute of Science and Technology) maintained monoclonal knockout cell lines, verified knockout cell lines by PCR, cloned M-217 and R-217 overexpressing vectors. Library preparation and high throughput sequencing of the complete genome of knockout cells were performed by Anna Fedotova and Maria Logacheva at Skoltech Genomics Core Facility on Skoltech sequencing grant resources. Rustam Ziganshin (Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry) conducted LC-MS analysis. Flow cytometry of PI-stained cells and AnnexinV-FITC/PI stained cells was done by Viacheslav Senichkin (Faculty of Basic Medicine, MV Lomonosov Moscow State University) and analyzed with the help of Olga Sergeeva. Pavel Melnikov (Serbsky National Medical Research Center for Psychiatry and Narcology) conducted a microscopy analysis of TUNEL samples.

4.1 Generation of hMorrbid/CYTOR CRISPR-Cas9 knockout cell line

hMorrbid and CYTOR genes are located on the different arms of human chromosome 2. CYTOR is a sense lncRNA and hMorrbid is an antisense lncRNA. According to the Ensembl genome browser, hMorrbid has 42 exons, which may assemble into 108 transcripts, while CYTOR has 24 exons, which may assemble into 38 transcripts. hMorrbid and CYTOR exons are abundantly transcribed, examples of just some of them measured in Huh7 cells are represented in Figure 23.

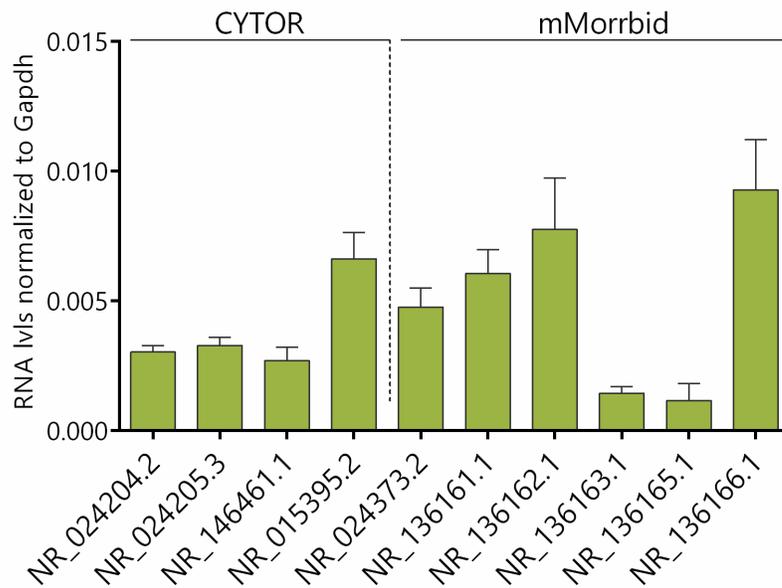


Figure 23. RT-qPCR analysis of hMorrbid and CYTOR transcripts expression in Huh7 cells.

Results show mean \pm SD.

To investigate hMorrbid and CYTOR roles a generation of cells with the loss-of-function (LOF) phenotype is useful. Previously, hMorrbid/CYTOR functions were studied via RNAi-mediated downregulation of their expression [9,10,12,13,22,23,25]. However, lncRNA

knockdown with siRNA or shRNA does not guarantee immediate and complete transcript decay. Also, lncRNA may preserve a complete functional role even when part of its sequence is degraded like it was for RoX1 RNA responsible for sex dosage compensation in *Drosophila* [49]. Additionally, many isoforms of hMorrbid/CYTOR do not have overlapping exons and a pool of siRNA/shRNA is required to target all actively expressed transcripts. The CRISPR-Cas9 knockout approach allows us to address these issues by introducing genomic deletion, that completely abrogates expression from the gene locus.

Here, for the first time, we performed genome editing of hepatocellular carcinoma Huh7 cells and excised predicted TSSs of both genes to completely terminate their expression. Knockout of hMorrbid/CYTOR genes was done with CRISPR-Cas9 mediated deletion of ~2200bp area around hMorrbid/CYTOR promoter and first exon regions (Figure 24). CYTOR genomic sequence is 99.3% alignable to hMorrbid gene (Figure 16). This allowed us to target both genes loci with one pair of gRNAs.

Exact transcription start sites of the lncRNA paralogs are unknown. We wanted to confirm that we either significantly damage or entirely deplete hMorrbid /CYTOR promoters. So, we used a pair of gRNAs targeting loci ~1200bp upstream and ~1000bp downstream of the first exon of CYTOR, which corresponds to the area ~ 2000bp upstream and 200 bp downstream of the first hMorrbid exon (Figure 24A). Guide sequences were cloned into the sgRNA scaffold of pX458 plasmid according to the Zhang Lab general cloning protocol [204]. A total of 7 different gRNAs were tested: 3 upstream and 4 downstream. Huh7 cell line was transfected with the different

combinations of gRNA-caring plasmids: one upstream + one downstream. After transfection single cell-derived knockout clones were expanded by dilution plating [205].

After several weeks of the growth survived monoclonal cell lines were tested by PCR using genome DNA followed by Sanger sequencing (Figure 24 B, C, D) and NGS. Primer pair 1 and primer pair 2 spanned across the gRNA-guided cuts 1 and 2 correspondingly. No PCR products from primers 1 and 2 were detected in the case of KO (Figure 24 B and C).

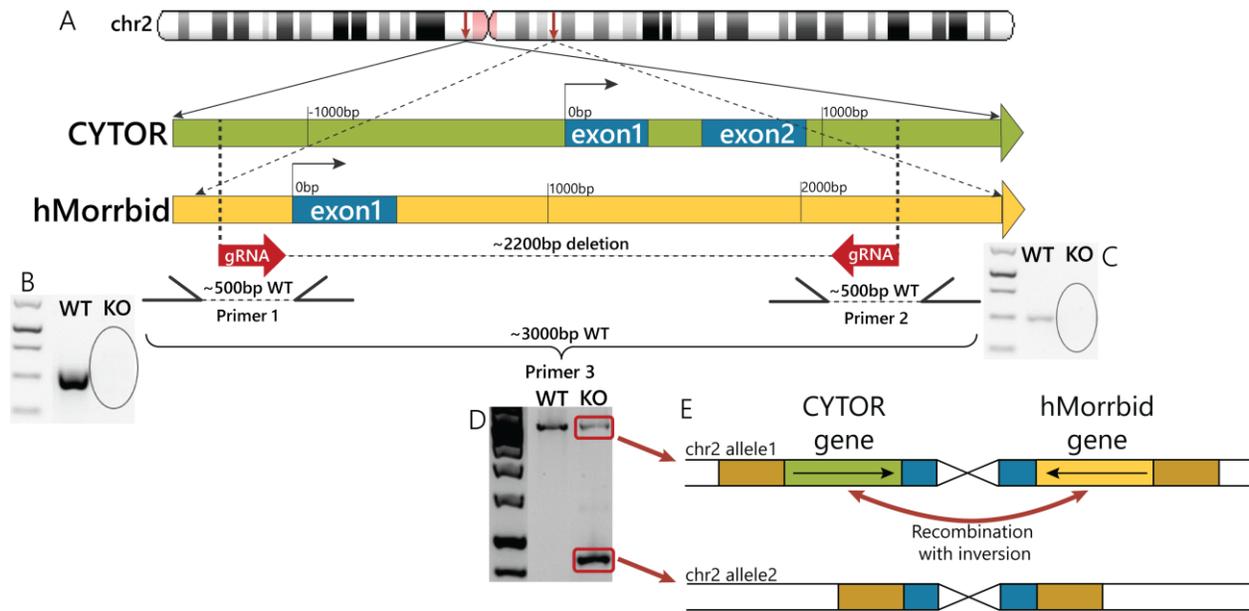


Figure 24. Design and validation of hMorrbid/CYTOR knockout. (A) Schematic representation of gRNAs and primers positions on the hMorrbid (MIR4435-2HG) and CYTOR genome DNA. (B) Image of amplicons separation by agarose gel electrophoresis. Amplicons were obtained with primers pair 1, spanning across the gRNA 1 target site. No amplicon was detected in KO samples. (C) Image of amplicon separation by agarose gel electrophoresis. Amplicons were obtained with primers pair 2, spanning across the gRNA 2 target site. No amplicon was detected in KO samples. (D) Image of amplicons separation by agarose gel electrophoresis. Amplicons were obtained with primers pair 3, spanning across the entire deletion region. (E) Schematic representation of the whole genome NGS analysis of Huh7 hMorrbid/CYTOR KO cells.

Primer pair 3 spanned across ~3000bp gDNA region comprising a complete 2200bp deletion region. Separation by agarose gel showed two amplicons in the KO cells: the same size

as wild type PCR product and a band about 800bp shorter than the WT PCR product (Figure 24D). To confirm successful knockout and eliminate off-target effects we performed whole-genome next-generation sequencing (NGS) of the KO Huh7 cell line. NGS analysis showed concurrence of excision with end-joining of the genomic regions with coordinates chr2:111494701 chr2:111496860 and chr2:87453667 chr2:87455829 and genomic recombination with inversion of the target regions (Figure 24E, Supplementary Figures S1 and S2). The first situation corresponds to the lighter band on the gel, which was confirmed by Sanger sequencing and the second situation corresponds to the heavier band on the gel, also confirmed by Sanger (Figure 24D). Thus, a successful double excision of two 2kbs regions at two loci on chromosome 2 using one pair of gRNAs was unambiguously confirmed.

4.2 hMorrbid/CYTOR is dispensable for hepatocellular carcinoma cells proliferation, migration, and apoptosis resistance

Complete loss of transcription from the hMorrbid and CYTOR loci of KO cells was confirmed with RT-qPCR analysis (Figure 25A). Next, we looked at the phenotype changes in knockout cells. Previously, hMorrbid/CYTOR-depletion led to decreased proliferation in various cancer cells [9,10,12,13], including HCC [17]. Surprisingly, Huh7 hMorrbid/CYTOR KO cell line demonstrated the same proliferation rate as wild type Huh7 cells over the time course of 4 days as measured with MTS viability test (Figure 25B).

Cell cycle arrest in the G1 phase was typically reported for the hMorrbid/CYTOR loss-of-function phenotype [9,12]. In the case of KO Huh7 cells a slight but significant

decrease ($\downarrow 3.6\%$) of cells in G1 phase and accumulation ($2.8\% \uparrow$) of cells in G2 phase were detected (progression through the cell cycle was measured using flow cytometry analysis with PI (propidium iodide) staining) (Figure 25C).

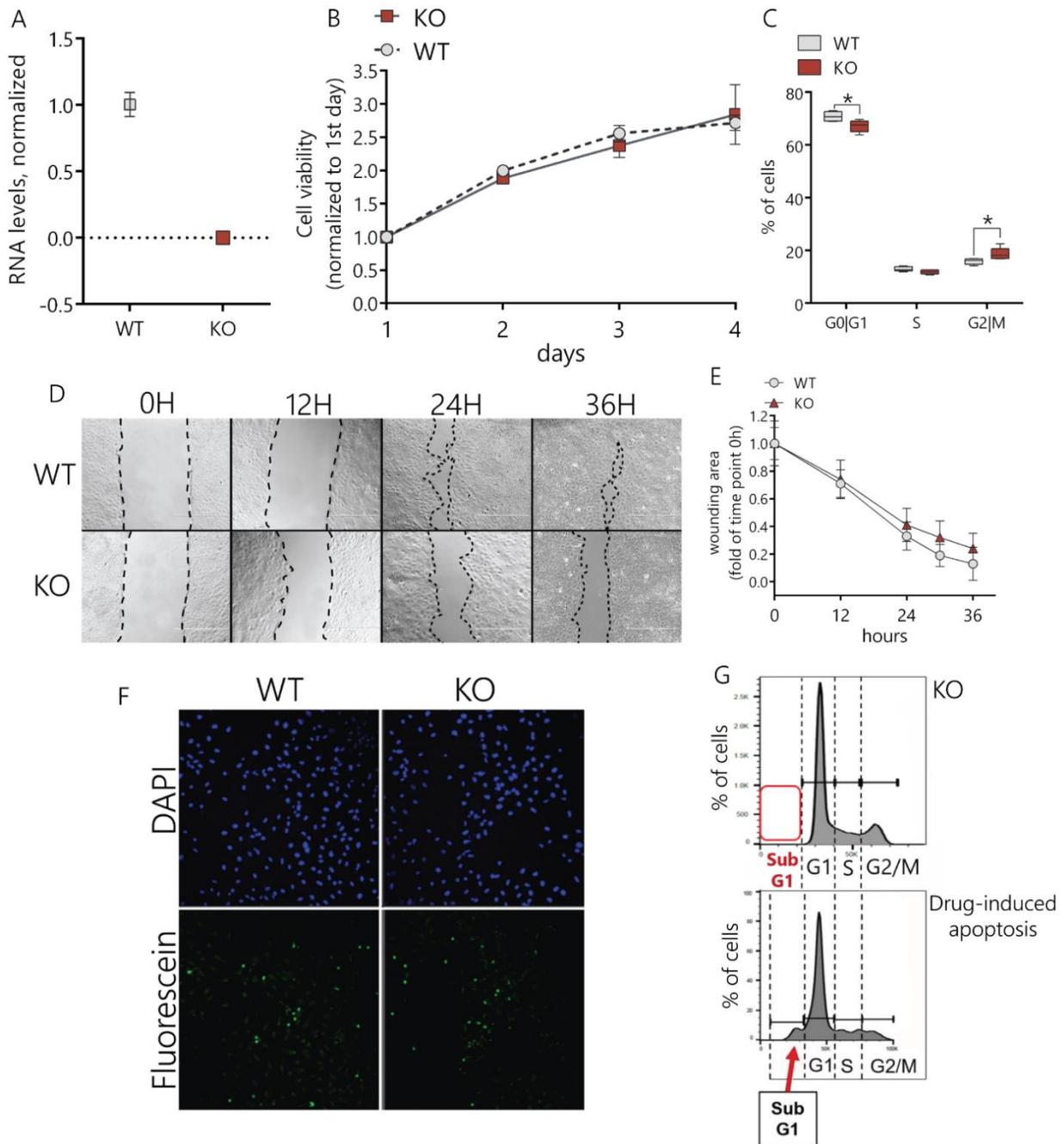


Figure 25. *Characterization of hMorrbid/CYTOR knockout in Huh7 cells. (A) RT-qPCR analysis of hMorrbid/CYTOR transcripts expression in WT and KO cells. (B) Viability assay of hMorrbid/CYTOR KO and WT Huh7 cells on the 1, 2, 3, and 4 days after plating, normalized to viability at day 1. (C) Flow cytometry analysis of cell cycle of PI-stained KO and WT Huh7 cells. (D) Microscopy images of wound healing assay in hMorrbid/CYTOR KO and WT Huh7 cells. (E) Wound-healing assay of hMorrbid/CYTOR KO and WT Huh7 cells, normalized to the wound area at the 0h time point. (F) Microscopy images of TUNEL assay of hMorrbid/CYTOR KO and WT Huh7 control cells. (G) Flow cytometry analysis of the SubG1 phase of PI-stained hMorrbid/CYTOR KO Huh7 cells. Cells treated with apoptosis-inducing drugs were used as a positive control. Results show mean \pm SD. n.s.—not significant. * $p < 0.05$.*

It was previously reported that knockdown of hMorrbid/CYTOR resulted in reduced HCC cell migration if measured by wound healing assay [19]. In our case knockout resulted in reduced, however non-significantly, wound closure rate after 36H (Figure 25D, E). Another feature of RNAi-mediated hMorrbid/CYTOR knockdown is increased cellular apoptosis [9,10,12,22,23,25]. In the case of hMorrbid/CYTOR KO, no apoptosis was detected when measured by TUNEL assay (Terminal deoxynucleotidyl transferase dUTP nick end labeling) (Figure 25F) or by flow cytometry analysis of PI-stained KO cells SubG1 phase (Figure 25G).

Thus, we demonstrated that complete knockout of hMorrbid/CYTOR lncRNA genes is characterized by a different phenotype, than RNAi mediated knockdown of hMorrbid/CYTOR transcripts and only slightly varies from wild-type cells in cell cycle and migration.

4.3 Proteomic changes in response to hMorrbid/CYTOR knockout

We analyzed proteome changes in response to hMorrbid/CYTOR knockout with quantitative LC-MS proteome analysis. In total, 3281 proteins were identified with 329 of the differentially expressed between KO and WT. 53 proteins were upregulated and 43

were downregulated with at least 2-fold difference between KO and WT.

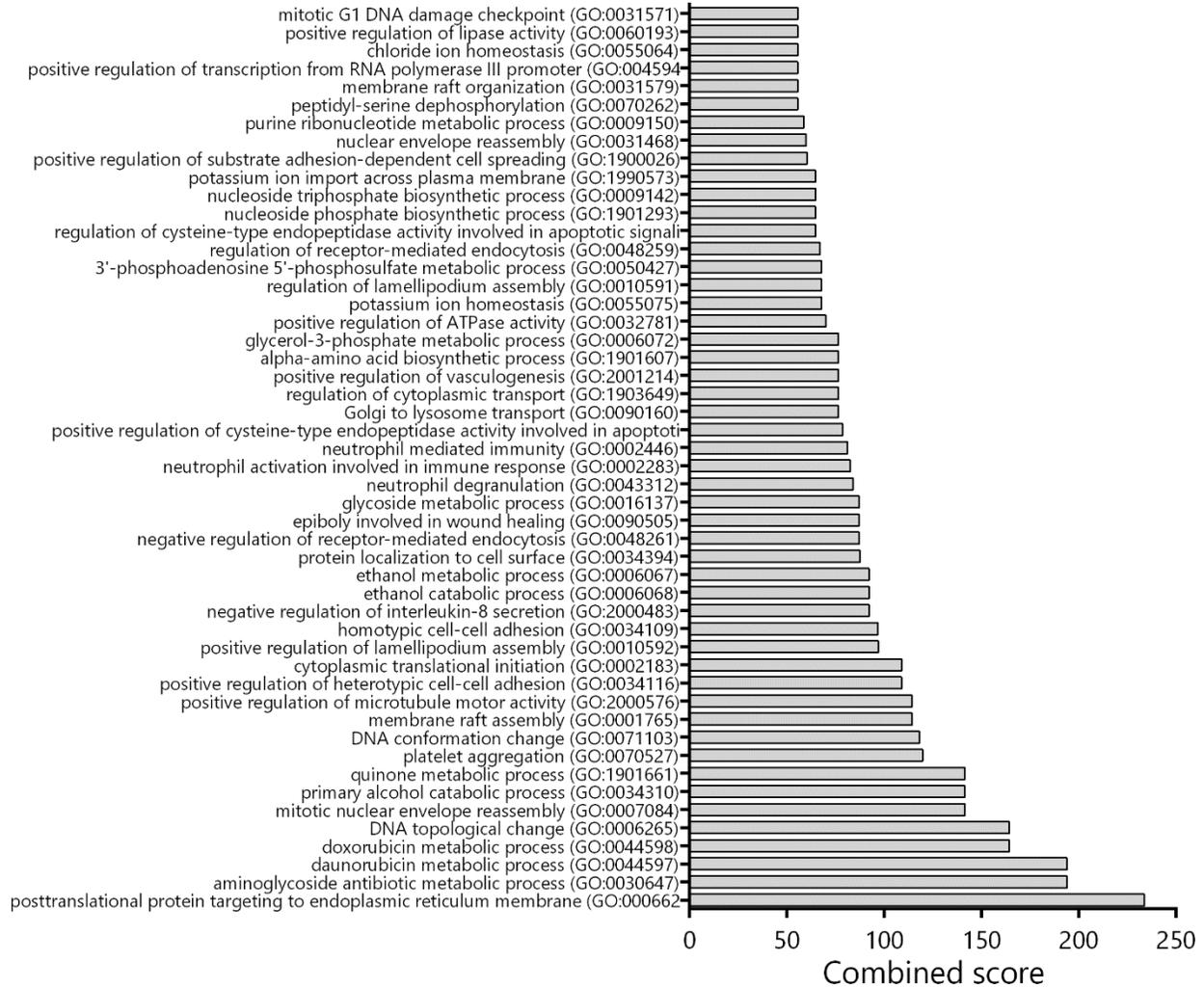


Figure 26. Summary of LC-MS analysis of genes differentially expressed in cells with hMorrbid/CYTOR depletion.

Identified differential changes affected proteins involved in various cellular processes: regulation of signal transduction, cytoskeleton, mitochondrial and lipid metabolism, etc (Figure 26). This includes cell cycle checkpoint proteins RAD9A, NEK9, SHPRH; regulators of p53

HEXIM1, and DDX24; regulators of transcription and elongation ELP2, GTF2F2, ELP3. These shifts in the proteome of the knockout cells may be a part of the compensation mechanism for hMorrbid/CYTOR loss.

4.4 Previously reported downstream target genes of hMorrbid/CYTOR mostly remained unaffected in the knockout

We analyzed the expression of genes, which were previously reported as direct or indirect downstream targets of hMorrbid/CYTOR in knockout Huh7 cells. hMorrbid/CYTOR was repeatedly reported to promote EMT (epithelial to mesenchymal transition) [9,10], through activation of various pathways, like Wnt [30]. The knockdown of hMorrbid/CYTOR was characterized by the upregulation of E-cadherin and downregulation of Vimentin and N-cadherin [30]. However, knockout of hMorrbid/CYTOR in hepatocytes has led to an opposite (but slight) dynamic of E-cadherin and Vimentin expression (Figure 27A). hMorrbid/CYTOR was reported to promote progression through the cell cycle via downregulation of cell cycle inhibitors p21 and p15 [174]. Knockdown of hMorrbid/CYTOR with siRNA led to significant upregulation of p21 and p15 levels in gastric cancer cell lines at mRNA and protein levels [174]. However, p15 and p21 were slightly downregulated in the Huh7 knockout cell line (Figure 27A). hMorrbid/CYTOR was reported to positively regulate CyclinD1 expression in HCC via sponging miR-193b [15]. In hMorrbid/CYTOR KO Huh7 cells CyclinD1 is upregulated 1.5-fold at both mRNA and protein levels (Figure 27A-C). Downregulation of p21 and p15 and simultaneous upregulation of CyclinD1 correspond to a slight accumulation of hMorrbid/CYTOR KO cells in the G2 phase of the cell cycle (Figure 25C). A positive correlation between the expression of hMorrbid and EGFR was detected in gastric cancer and lung cancer patients: EGFR protein level

was downregulated in cells with hMorrbid/CYTOR knockdown [13,25]. In Huh7 hMorrbid/CYTOR KO cells EGFR mRNA expression is about 1.2-fold elevated in comparison to wild-type cells (Figure 27A). Another study performed in HCC cell lines demonstrated that hMorrbid/CYTOR binds EpCam promoter and drives EpCam expression [16]. ShRNA knockdown of hMorrbid/CYTOR resulted in the downregulation of EpCam [16].

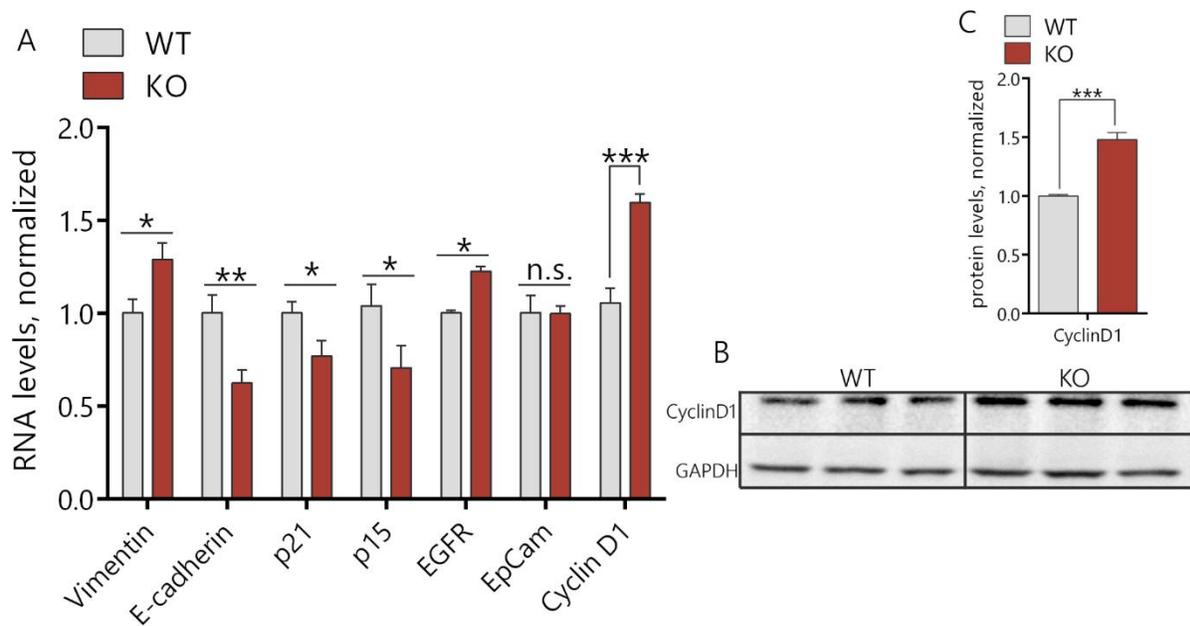


Figure 27. Analysis of hMorrbid/CYTOR literature targets in knockout cells. (A) RT-qPCR analysis of expression of genes reported as hMorrbid/CYTOR downstream targets. (B) Estimation of CyclinD1 protein level after hMorrbid/CYTOR depletion with western blot. (C) Quantitative analysis of CyclinD1 protein level using ImageJ software. Results show mean \pm SD. n.s.—not significant. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

In hMorrbid/CYTOR KO cells EpCam remains unchanged in comparison to WT (Figure 27A). The upregulation of Bim expression in hMorrbid/CYTOR RNAi knockdown was reported in [10]. We observed an increase of Bim expression in hMorrbid/CYTOR knockout, which will

be discussed further (Figure 30). The same research together with several others reported downregulation of MCL1 in hMorrbid/CYTOR knockdown [10,175], however, we did not notice any change in MCL1 levels in KO cells (Figure 32).

Overall, previously reported hMorrbid/CYTOR targets were not confirmed in Huh7 KO cells. One possible explanation for that is tissue specificity of lncRNAs, as some of the effects we attempted to verify were obtained not in the liver cancer. Another reason is that the previous studies used RNAi-mediated knockdown to downregulate hMorrbid/CYTOR and discrepancies between phenotypes obtained with RNAi and CRISPR methods were previously described [40,41,210–212]. For example, MALAT1 genetic knockout in mice did not confirm any of the previously reported effects on cell proliferation and viability acquired with RNAi methods [40].

4.5 Overexpression of hMorrbid transcript in knockout cell line resulted in the delayed proliferation and increased sensitivity to apoptosis.

4.5.1 MIR4435-2HG (M-217) is a transcript of the hMorrbid gene containing evolutionary conserved region exonCh

Unlike PCGs, lncRNA genes demonstrate poor conservation of their primary nucleotide sequence. lncRNA genes conserved between human and mouse generally share around 20% alignable nucleotide sequence attributed to short sequence regions usually spanning across one or two exons [42]. It is possible that these short patches of sequence avoided evolutionary turnover due to their functional features, while the rest of the transcript sequence can undergo genomic rearrangements and nucleotide

substitutions without an impact on the overall transcript function [42]. Short lncRNA regions were previously demonstrated to be responsible for the lncRNA functional role [49, 156].

Around 300 nt long exon of hMorrbid, which we called exonCh, contains 119 nt long region that is 82% alignable to murine exonCm (Figure 14). Thus, exonCh may be specifically involved in regulation. ExonCh is shared by 17 hMorrbid transcripts (genome assembly Human GRCh38.p13, 5.8.20) (Figure 28A and Figure14). Transcript MIR4435-2HG-217 (M-217), Ensemble ID ENST00000603827.1 (genome assembly Human GRCh38.p13, 5.8.20) is 1587 nt long single-exon transcript. The last 299 3' end nucleotides of M-217 correspond to exonCh. Additionally, 1288 5' nucleotides of M-217 transcript represent an M-217-specific sequence, that cannot be aligned to any other hMorrbid or CYTOR transcripts (Figure 28A), suggesting potential transcript-specific roles for M-217.

M-217 is expressed in various human cell lines (Figure 28B), including hepatocellular carcinoma cells HepG2 and Huh7 where its expression is higher than in other tested tissues (A459-human adenocarcinoma, HEK-human embryonic kidney, VA-13-human fibroblasts). Therefore, M-217 is a broadly expressed transcript, upregulated in human hepatocellular carcinoma cell lines. Further M-217 investigation we performed in Huh7 cells.

M-217 expression was validated by PCR with M-217 full-length primers (Figure 28C) and following Sanger sequencing. Thus, we decided to clarify the impact of the conserved region exonCh into the functional role of M-217 transcript in human liver cancer cells.

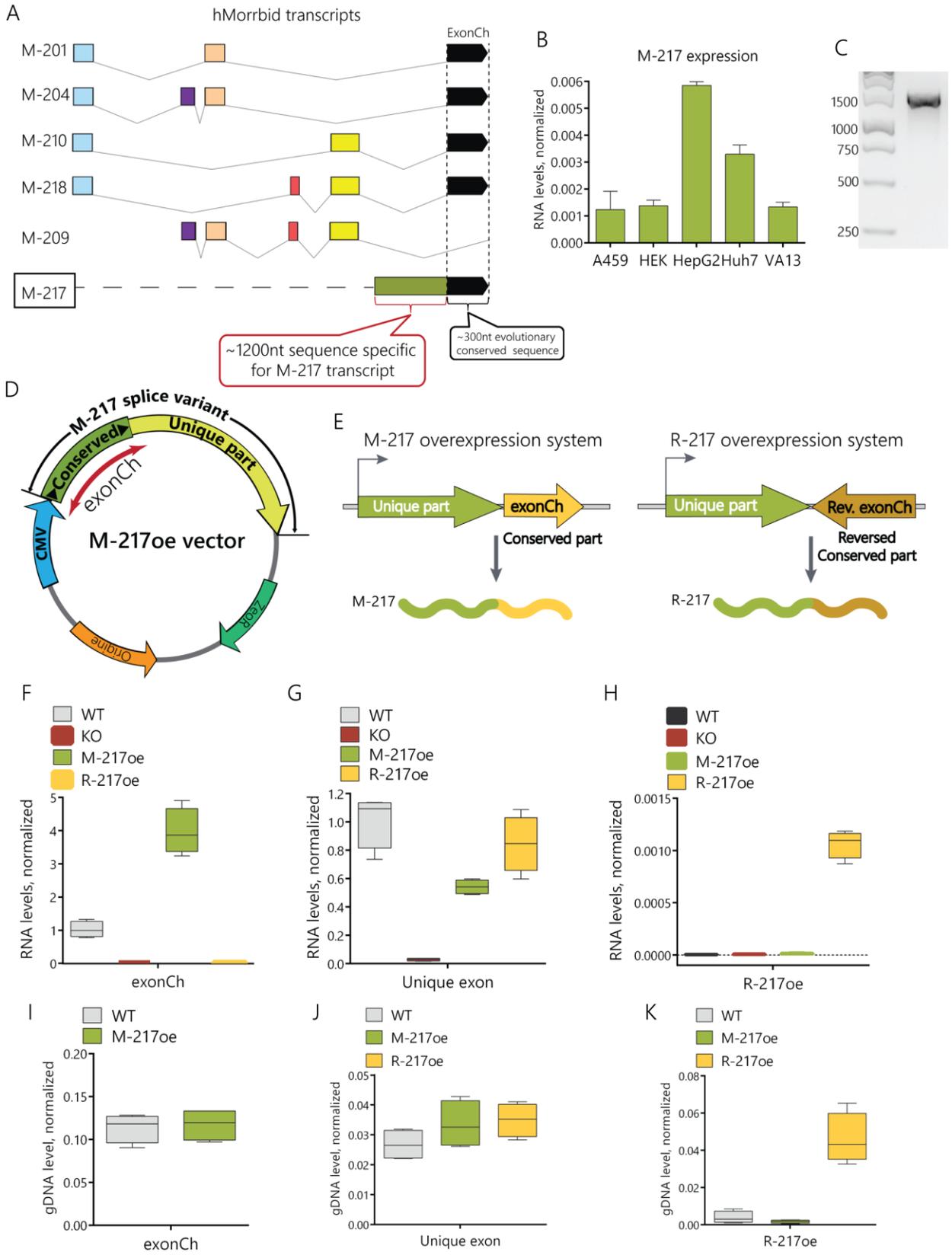


Figure 28. Establishment of cell lines overexpressing transcripts M-217 and R-217. (A) Schematic representation of several hMorrbid transcripts. Transcript M-217 is a single-exon transcript that consists of exonCh (~300nt) and a unique part (~1200nt). (B) RT-qPCR analysis of M-217 expression in human cell lines: A459 (lung adenocarcinoma), HEK (embryonic kidney), HepG2 (hepatocellular carcinoma), Huh7 (hepatocellular carcinoma), Va13 (lung). (C) Image of agarose gel electrophoresis of amplicon obtained with primers spanning across the entire length of the M-217 transcript. (D) Schematic representation of M-217oe vector expressing M-217 under CMV promoter and containing zeocin resistance gene. (E) Schematic representation of the difference between M-217oe vector expressing intact M-217 transcript (right) and R-217oe vector expressing R-217 transcript with reversed exonCh(left). (F) RT-qPCR analysis of exonCh expression in WT cells, KO cells, and knockout cell lines expressing M-217 and R-217. (G) RT-qPCR analysis of M-217 unique region expression in WT cells, KO cells, and knockout cell lines expressing M-217 and R-217. (H) RT-qPCR analysis of reversed exonCh (R-217oe) expression in WT cells, KO cells, and knockout cell lines expressing M-217 and R-217. (I) RT-qPCR analysis of exonCh incorporation into the genomic DNA of M-217 expressing cell line. (J) RT-qPCR analysis of M-217 unique region incorporation into genomic DNA of M-217 and R-217 expressing cell lines. (K) RT-qPCR analysis of reversed exonCh (R-217oe) incorporation into the genome of R-217 expressing cell lines. Results show mean \pm SD.

4.5.2. Generation of the M-217 and mutated M-217 cell lines

To investigate M-217 transcript functions we overexpressed it in hMorrbid/CYTOR knockout Huh7 cell line. M-217 transcript was cloned into pcDNA3.1 plasmid under CMV promoter (Figure 28D and E left). Successful cloning was confirmed with Sanger sequencing. To identify the input of conserved exonCh into M-217 functional properties we created a vector expressing mutated M-217 transcript with inverted (reverse complement sequence) exonCh. The mutated M-217 transcript was called Reversed-217 or R-217 (Figure 28E right). Validated vectors were transfected into KO cells. Cells were selected based on zeocin resistance for 4 weeks. As a result, we obtained stable cell lines M-217oe and R-217oe overexpressing transcripts M-217 and R-217 respectively as was confirmed with RT-qPCR (Figure 28F-H) and incorporation of both vectors into the genome was confirmed with RT-qPCR with genome DNA as a template (Figure 28I-K).

4.5.3. M-217oe cell line but not R-217oe cell line had impaired cell growth and increased apoptosis

We noticed that M-217oe cell growth is significantly delayed in comparison to WT, KO, and R-217 cells. Resazurin analysis demonstrated that M-217oe cells proliferate ~30% slower than other cell lines, while R-217oe cells proliferation is similar to the WT cells (Figure 29A).

Flow cytometry analysis of Annexin V-FITC/PI stained cells is a tool to distinguish between intact live cells, cells undergoing early apoptosis, or late apoptosis by monitoring translocation of phosphatidylserine (PS) and cellular membrane permeabilization. Early apoptosis signature: translocation of PS from cytosolic to extracellular part of the cellular membrane is detected by fluorescein isothiocyanate (FITC) conjugated AnnexinV (AnnexinV-FITC). In the late apoptosis stage cellular membrane becomes permeable for Propidium iodide (PI) [213]. We measured cell death with AnnexinV-FITC/PI flow cytometry and found that M-217oe cells experience an increase in the rate of total apoptosis (11.7% versus 6.7% in WT Huh7 cells), while R-217oe cells have it at the similar level as the wild-type and knockout cells: around 5-6% (Figure 29B, C). Thus, exonCh expression may negatively impact cell proliferation and cause apoptosis susceptibility of human malignant hepatocytes.

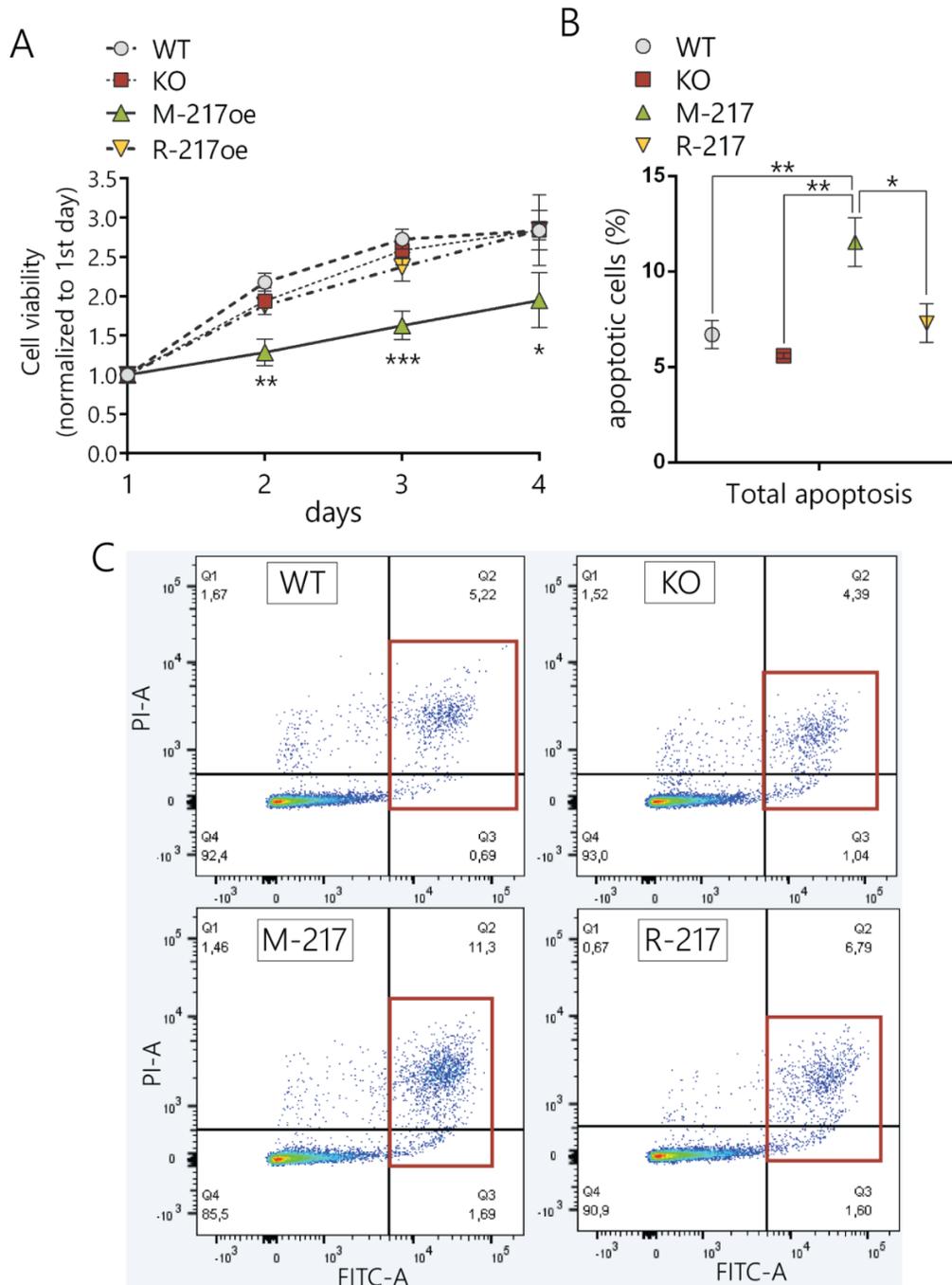


Figure 29. Characterization of M-217 overexpression. **(A)** Resazurin viability and growth assay of WT, KO M-217oe, and R-217oe cells on the 1, 2, 3, and 4 days after plating, normalized to viability at day 1. **(B)** Flow cytometry analysis of AnnexinV-FITC/PI-stained WT, KO, M-217oe, and R-217oe cells. **(C)** Flow cytometry images of AnnexinV-FITC/PI-stained WT, KO, M-217oe, and R-217oe cells. Framed in red are apoptotic cells, Q3 – early apoptosis and Q2-late apoptosis. Results show mean \pm SD. n.s.—not significant. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

4.6 hMorrbid/CYTOR KO and M-217oe cells demonstrate differential shifts in expression of pro-apoptotic and pro-survival BCL2 family proteins

Regulation of apoptosis is a complicated mechanism, which relies on the equilibrium between pro-apoptotic and pro-survival (or anti-apoptotic) proteins [214]. Pro-apoptotic and pro-survival proteins belong to the BCL2 family of proteins. A group of pro-survival proteins includes BCL-2, BCL-W, BCL-XL, BFL1, MCL1. Pro-apoptotic proteins are divided into groups: BH3-only ‘activators’ (Bim, PUMA, tBID); BH3-only ‘sensitizers’ (BAD, Noxa, BMF, Hrk) and pore-forming proteins (BAX and BAK) [214]. BH3-only ‘activators’ directly bind and activate pore-forming proteins located at the surface of the mitochondria. This binding leads to conformational changes in pore-forming proteins, which allows them to oligomerize and form macropores in the mitochondrial membrane, resulting in mitochondrial outer membrane permeabilization (MOMP) (Figure 30). MOMP leads to the release of apoptogenic proteins from the intermembrane space, such as Cytochrome C, which trigger caspase cascade and inevitable cell death (Figure 30).

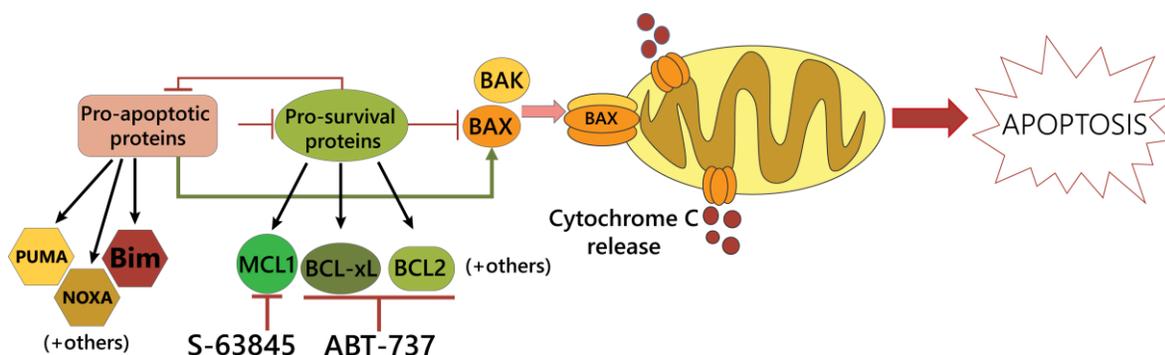
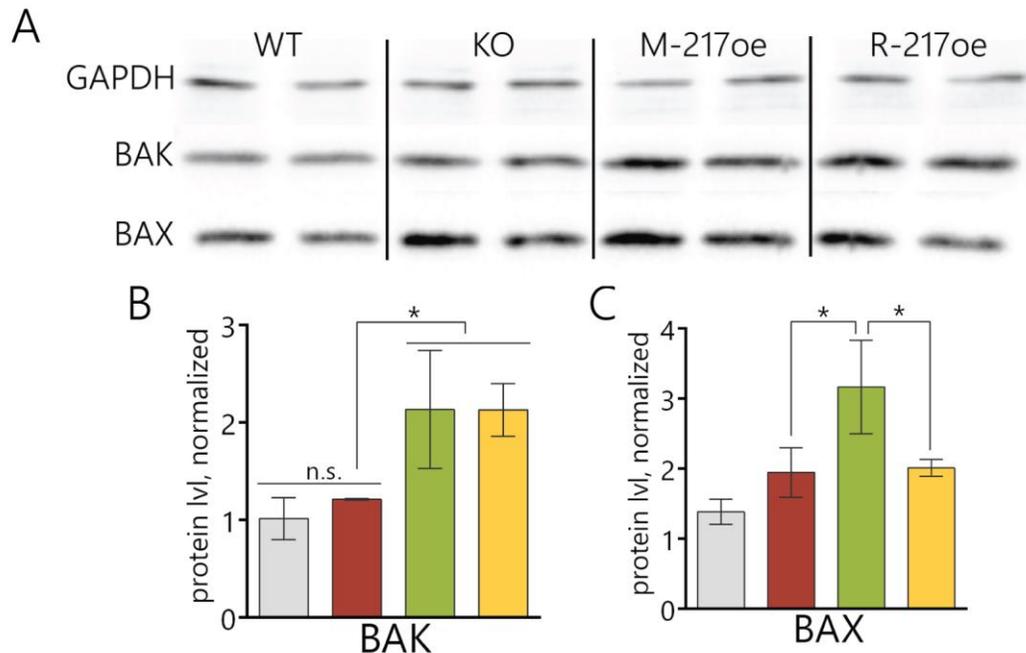


Figure 30. Mechanism of mitochondria-mediated apoptosis

Pro-survival BCL2 proteins prevent apoptosis by binding to pore-forming and BH3-only apoptotic activators and blocking their activity. BH3-only pro-apoptotic sensitizers are less efficient in activating pore-forming proteins than activators, so their main function is considered to be inhibition of pro-survival proteins [214]. For a cell to commit to apoptosis cellular concentration of active pro-apoptotic proteins must surpass the cellular concentration of active pro-survival proteins. If pro-survival proteins become overwhelmed and BAX and BAK get activated, apoptosis occurs (Figure 30). We measured protein expression of BCL2 family proteins in wild-type, knockout, and overexpression cell lines (Figure 31). Western blot demonstrated that expression of BAK is elevated in both overexpressing cell line M-217 and control R-217 (Figure 31A, B), while BAX is significantly increased only in M-217oe cells (Figure 31A, C). Analysis of pro-survival proteins showed that protein levels of BCL-xL and MCL1 remain relatively high in all four cell lines with BCL-xL only slightly decreasing in R-217oe cells. Previously, hMorrbid was characterized as a negative regulator of Bim in murine short-lived myeloid cells, and *in cis* mechanism of regulation was proposed (Figure 22) [51]. Pro-apoptotic protein Bim has three main isoforms produced by alternative splicing: BimEL, BimL, and BimS. We analyzed Bim protein expression with western blot and found that it was elevated in hMorrbid/CYTOR KO cell line: BimEL 1.5-fold, BimL 2.6-fold, BimS 2.8-fold (Figure 32A, B). Interestingly, in M-217oe cell lines Bim was downregulated: BimEL by ~60%, BimL, and BimS by ~80%. On the other hand, in control R-217 cells BimEL was partially rescued by 10% and was about 50% lower than the WT level; BimL increases to ~90% of WT and BimS to ~70% of WT level (Figure 34A, B). Therefore, hMorrbid is a Bim repressor in human hepatocytes, and exonCh plays a role in this regulation.

Pro-apoptotic proteins



Pro-survival proteins

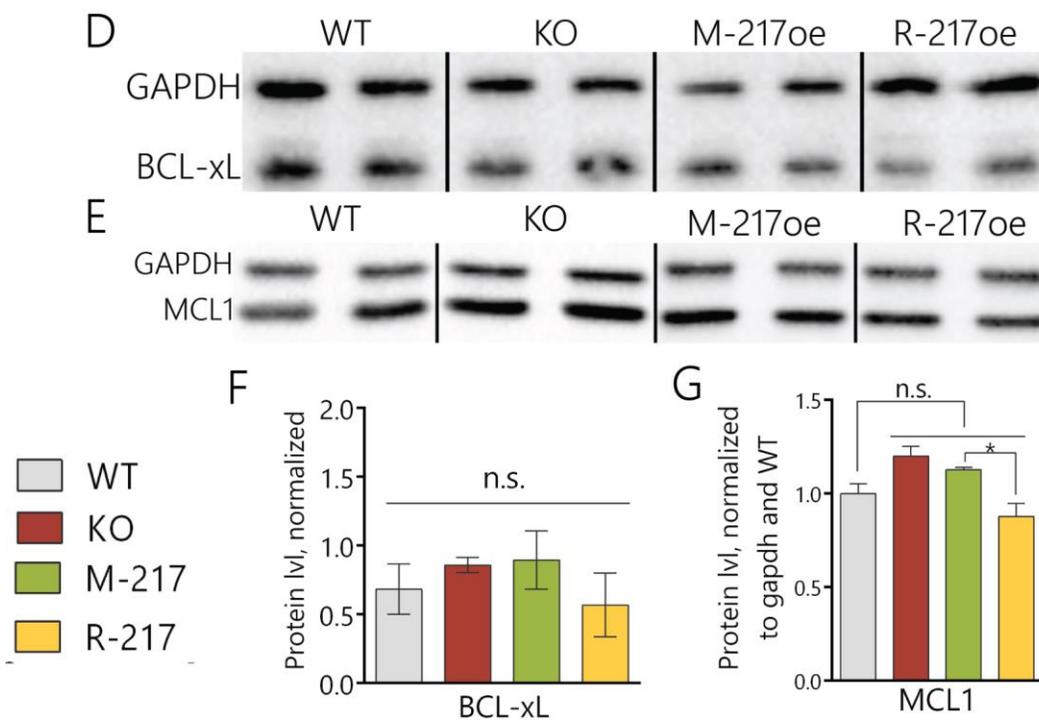


Figure 31. Expression of some pro-apoptotic and pro-survival members of BCL2 family in WT, KO, M-217oe, and R-217oe cells. (A) Estimation of BAX and BAK protein level in WT, KO, M-217oe, and R-217oe cells

with western blot. **(B)** Quantitative analysis of BAK protein level using ImageJ software normalized to GAPDH. **(C)** Quantitative analysis of BAX protein level using ImageJ software normalized to GAPDH. **(D, E)** Estimation of BCL-xL and MCL1 protein levels in WT, KO, M-217oe, and R-217oe cells with western blot. **(F)** Quantitative analysis of BCL-xL protein level using ImageJ software normalized to GAPDH. **(G)** Quantitative analysis of MCL1 protein level using ImageJ software normalized to GAPDH. Results show mean \pm SD. n.s.—not significant. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

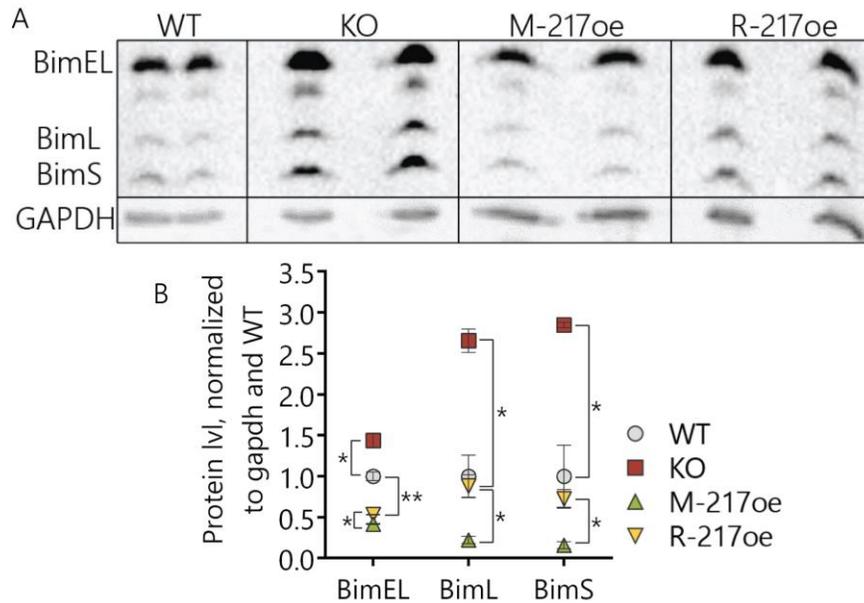


Figure 32. Analysis of expression of pro-apoptotic protein Bim. **(A)** Estimation of Bim protein level in WT, KO, M-217oe, and R-217oe cells with western blot. **(B)** Quantitative analysis of Bim protein level measured with western blot using ImageJ software normalized to GAPDH and protein level in WT. Results show mean \pm SD. n.s.—not significant. * $p < 0.05$, ** $p < 0.01$.

4.7 Expression of exonCh primes cells for apoptosis and increases cells sensitivity to MCL1 inhibition

Cells can exhibit different susceptibility to apoptosis. Cells primed for apoptosis express just enough pro-survival proteins to buffer existing pro-apoptotic proteins. Such cells are sensitized to apoptosis and would undergo MOMP in response to even weak cellular stress

or damage [214]. On the other hand, cells unprimed for apoptosis express pro-survival proteins in excess and easily withstand mild stress or damage (Figure 33).

To clarify what primes for apoptosis cells expressing exonCh, we used selective small-molecule suppressors of MCL1 (S-63845) and BCL2, BCL-xL, and BCL-W proteins (ABT-737) (Figure 33). S-63845 and ABT-737 were designed to competitively bind long hydrophobic groove of pro-survival BCL2 family proteins, preventing their inhibitory interaction with pro-apoptotic members. The deactivation of pro-survival proteins is cellular stress that shifts the balance towards apoptosis (Figure 33).

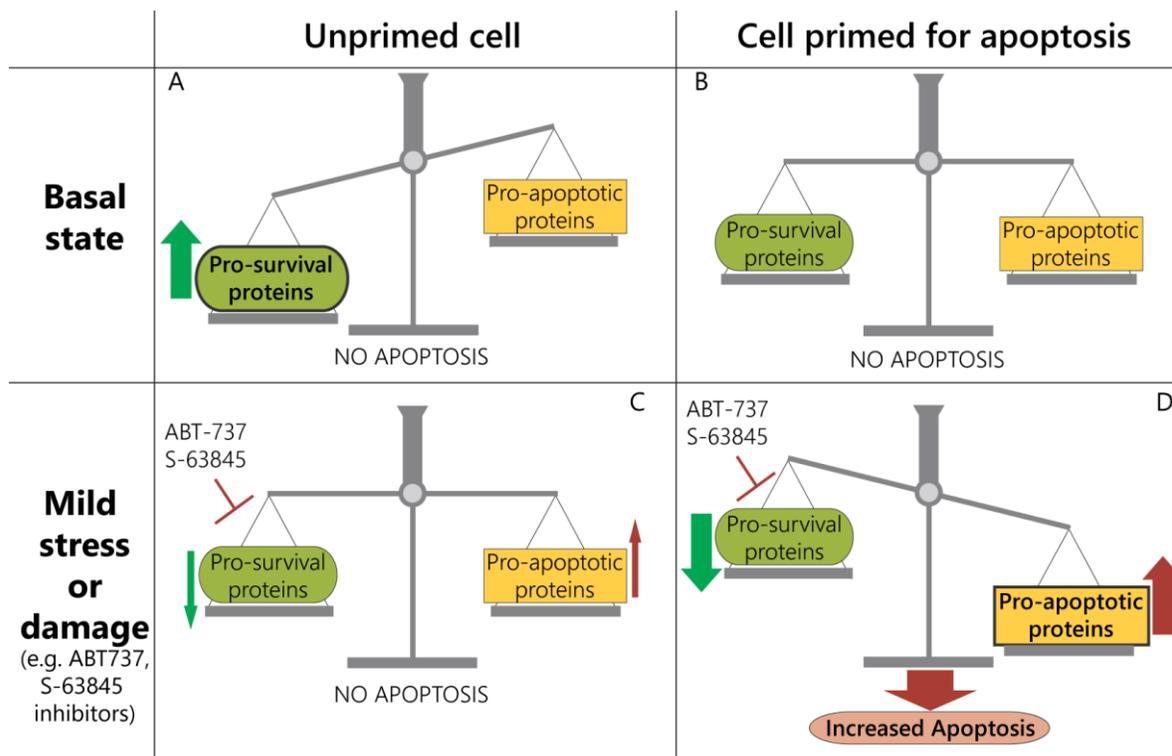


Figure 33. Schematic representation of the strategy used in the study to characterize M-217oe cells priming for apoptosis. (A) Cells unprimed for apoptosis in their basal state express pro-survival proteins in surplus. (B) Cells primed for apoptosis in their basal state express just enough pro-survival proteins to compensate pro-

apoptotic proteins. (C) In a situation of mild stress or damage, excessive pro-survival proteins allow unprimed cells to demonstrate apoptotic resistance. (D) Even mild stress or damage in primed for apoptosis cells results in a balance shift toward apoptosis. We used small molecule inhibitors S-63845 and ABT-737 to deactivate various pro-survival proteins and characterize apoptosis priming of hMorrbid/CYTOR depleted cells and M-217 overexpressing cells.

First, we treated WT, KO, M-217oe, R-217oe cell lines with MCL1 inhibitor S-63845 for 24h, stained them with AnnexinV-FITC/PI, and analyzed with flow cytometry. We observed a significant increase in the number of cells undergoing both early and late apoptosis in the M-217oe cell line (Figure 34A, C). Total apoptosis was calculated as a sum of early apoptosis (quadrant Q3) and late apoptosis (quadrant Q2) (Figure 34C). Total apoptosis increased by around 13% in M-217oe cells, 2% in KO, and R-217oe cells, while WT cells were resistant to MCL1 inhibition (Figure 34A, C). Normalization to non-treated condition demonstrated that M-217oe cells are 2.5-fold more sensitive to S-63845-induced apoptosis than WT cells, 1.8-fold than KO cells, and 2-fold than R-217oe cells (Figure 34B). Thus, overexpression of M-217 transcript results in cell susceptibility towards MCL1 inhibition, and exonCh is important for this effect. Then, we inhibited proteins BCL2, BCL-xL, and BCL-W with ABT-737 and found enhanced cell death in both M-217oe (15% upregulation to non-treated condition) and R-217oe (11% upregulation to non-treated condition) cell lines, while KO and WT cells showed little sensitivity to ABT-737 (4% in WT and 5% in KO) (Figure 34D-F). Normalization to non-treated control showed that both M-217oe and R-217oe cell lines are 1.5-fold more sensitive to ABT-737-induced apoptosis than wild-type cells (Figure 34F). Thus, active MCL1 is essential for M-217oe survival, while is unnecessary for R-217oe, and BCL2, BCL-xL, and BCL-W are equally important for both cell lines.

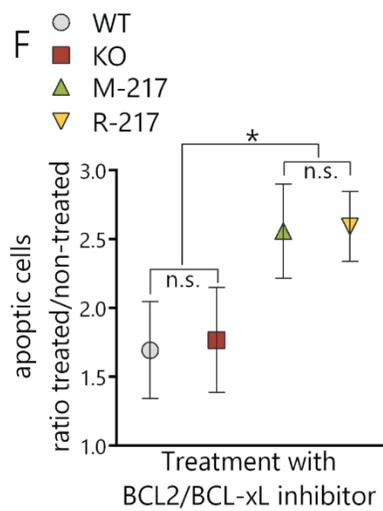
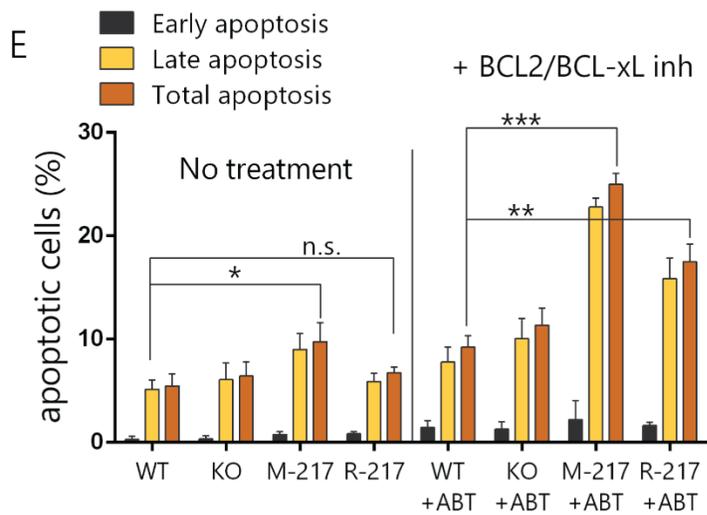
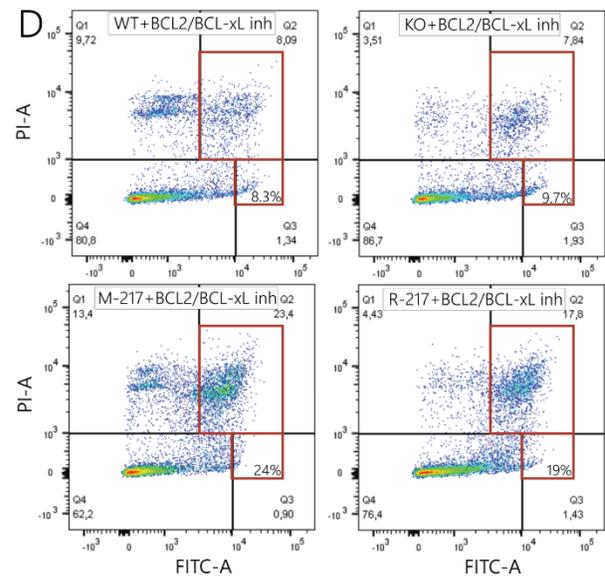
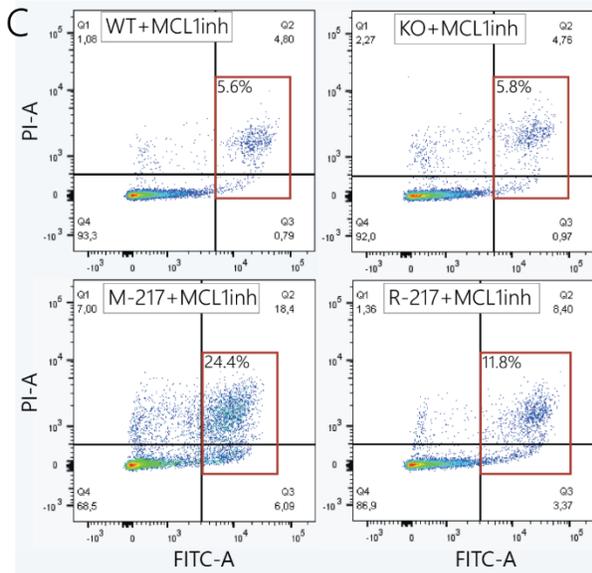
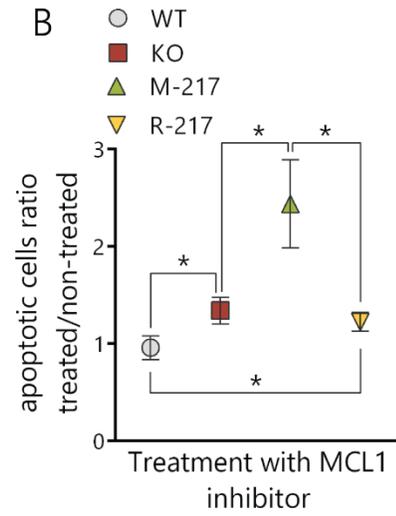
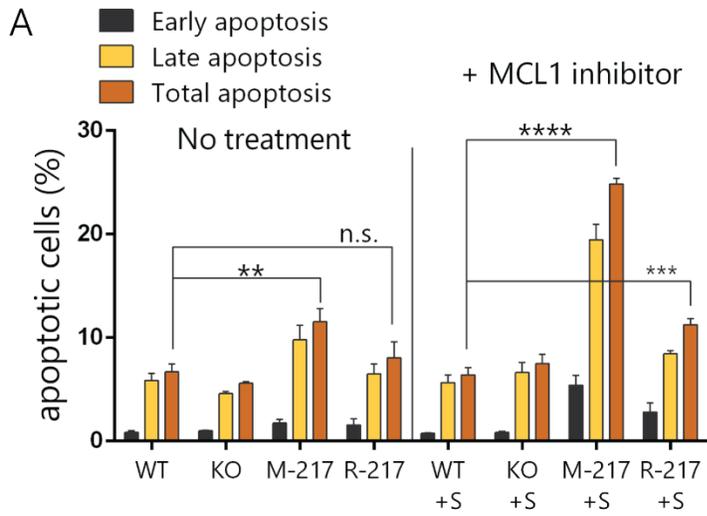


Figure 34. Flow cytometry analysis of AnnexinV-FITC/PI stained cells after MCL1 or BCL2/BCL-xL/BCL-W inhibition. (A-C) Flow cytometry analysis of WT, KO, M-217oe, and R-217oe cells after treatment with specific MCL1 inhibitor S-63845. (A) Percentage of early, late, and total apoptotic cells in WT, KO, M-217oe, and R-217oe cell lines treated and not treated with MCL1 inhibitor. (B) Total apoptosis in WT, KO, M-217oe, and R-217oe cells treated with S-63845, normalized to the non-treated control. (C) Flow cytometry images of cells treated with MCL1 inhibitor. Framed in red are apoptotic cells, Q3 – early apoptosis, and Q2 – late apoptosis. (D-F) Flow cytometry analysis of WT, KO, M-217oe, and R-217oe cells after treatment with BCL2/BCL-xL/BCL-W inhibitor ABT-737. (D) Flow cytometry images of cells treated with ABT-737 inhibitor. Framed in red are apoptotic cells, Q3 – early apoptosis, and Q2. (E) Percentage of early, late, and total apoptotic cells in WT, KO, M-217oe, and R-217oe cell lines treated and not treated with ABT-737 inhibitor. (F) Total apoptosis in WT, KO, M-217oe, and R-217oe cells treated with ABT-737, normalized to the non-treated control. Results show mean \pm SD. n.s.—not significant. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ and **** $p < 0.0001$

Chapter V. Results. Functional roles of Morrbid lncRNA in murine hepatocytes

The work described in this chapter was performed by the author at Timofei Zatsepin's laboratory, at the Center of Life Sciences, Skolkovo Institute of Science and Technology. The author performed all molecular biology experiments described in this chapter; cell culture manipulations; design of antisense oligonucleotides; sample preparation for RNA sequencing, LC-MS, flow cytometry, and microscopy analyzes. Library quality control and high throughput RNA sequencing were performed by Anna Fedotova and Maria Logacheva at Skoltech Genomics Core Facility. Ilya Kurochkin (Center for Neurobiology and Brain Restoration, Skolkovo Institute of Science and Technology) conducted the bioinformatic analysis of the transcriptome. Pavel Mazin (Center of Life Sciences, Skolkovo Institute of Science and Technology) performed the bioinformatic analysis of the alternative splicing. LC-MS analysis was done by Rustam Ziganshin at Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry. Flow cytometry of PI-stained cells and AnnexinV-FITC/PI stained cells was done by Viacheslav Senichkin (Faculty of Basic Medicine, MV Lomonosov Moscow State University) and analyzed with the help of Olga Sergeeva (Center of Life Sciences, Skolkovo Institute of Science and Technology). Pavel Melnikov (Serbsky National Medical Research Center for Psychiatry and Narcology) performed a microscopy analysis of FISH samples. Tatiana Prikazchikova (Center of Life Sciences, Skolkovo Institute of Science and Technology) performed siRNA design.

5.1 Inhibition of Morrbrid lncRNA decreases viability and migration of normal but not cancerous murine hepatocytes

mMorrbrid lncRNA expression was investigated in murine cancer (Hepa1-6) and normal (AML12) hepatocytes. First, we studied the subcellular localization of mMorrbrid lncRNA in hepatocytes by fluorescence in situ hybridization analysis (FISH) and found that the lncRNA predominantly localizes in the cell nucleus (Figure 35A). To confirm this data, we separated the nuclear and cytoplasmic fractions of AML12 and Hepa1-6 cells and measured the mMorrbrid lncRNA levels by RT-qPCR. In both cell lines, ~70% of mMorrbrid lncRNA was detected in the nuclear RNA fraction (Figure 35B). This corresponds to studies done in eosinophils and B cells, where mMorrbrid was also found predominantly localized in the cell nucleus [51]. Human Morrbrid was also found to be mostly nuclear in HCC [16]. Then, we compared mMorrbrid lncRNA levels in murine cancer (Hepa1-6) and normal (AML12) cell lines and found that mMorrbrid expression was ~2.2-fold decreased in Hepa1-6 (Figure 35C). Additionally, we checked the mMorrbrid expression in diethylnitrosamine (DENa) - induced hepatocellular carcinoma mouse model [215]. mMorrbrid was expressed approximately 1.4-fold higher in a murine wild-type liver than in the DENa model (Figure 35D). To elucidate mMorrbrid functions we designed antisense oligonucleotides (ASO) (Supplementary Table S1), validated each of them *in vitro*, and used a combination of the four most active ASOs in further studies (Figure 35E, F). Specifically, ASOs were designed to target mMorrbrid exons possessing sequence homology to human Morrbrid (Figure 14).

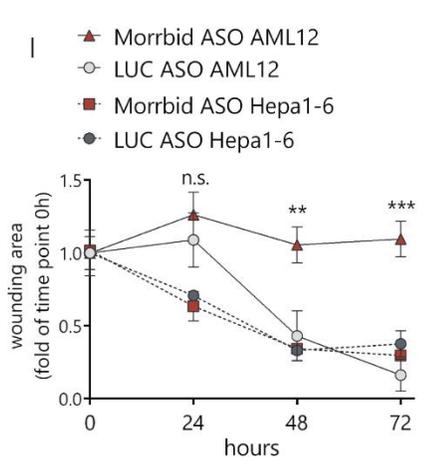
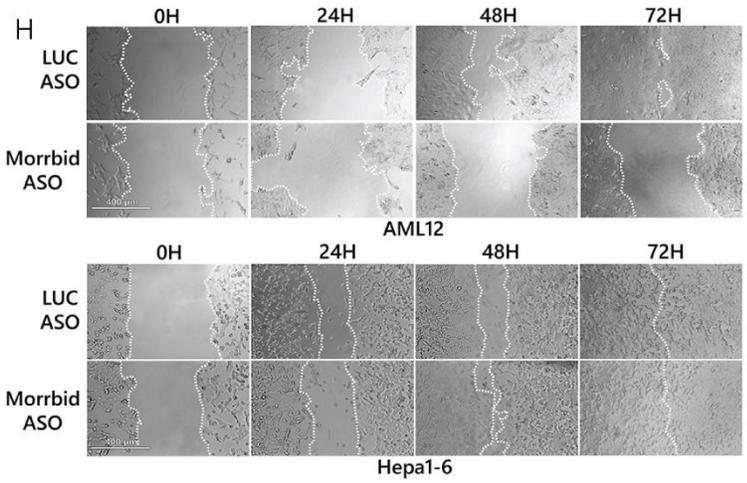
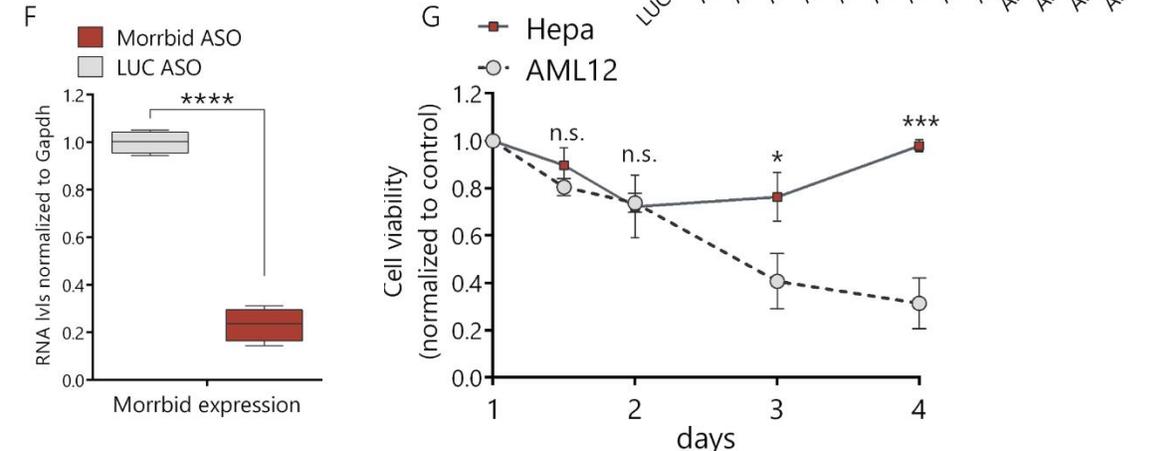
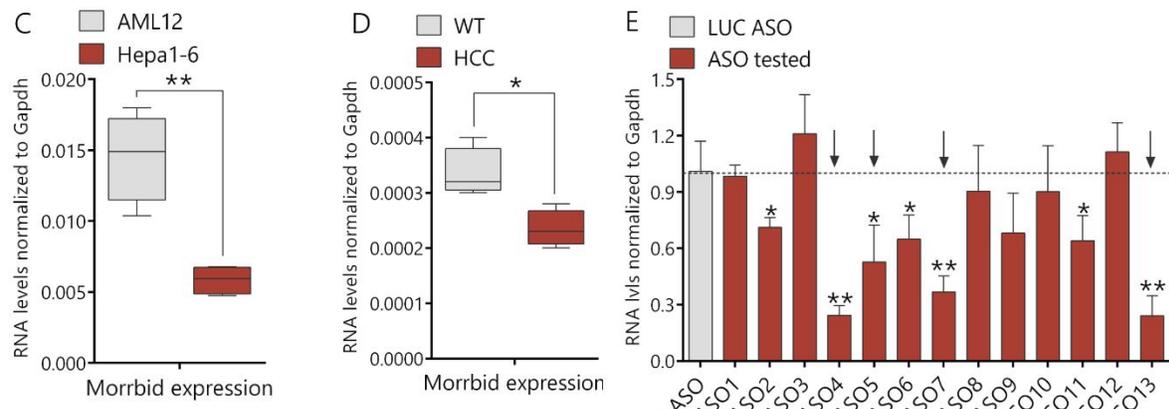
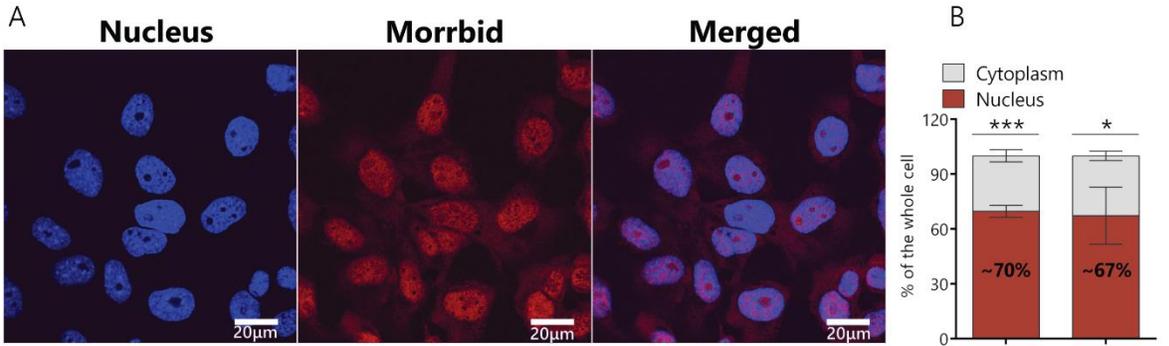


Figure 35. *mMorrbid* cellular localization and LOF phenotype. (A) Fluorescence in situ hybridization analysis (FISH) of *mMorrbid* localization in AML12 cells (DNA was stained with Dapi and *mMorrbid* was stained with a Cy5-labeled probe). (B) RT-qPCR analysis of *mMorrbid* expression in the nuclear and cytoplasmic fractions extracted from AML12 cells. (C) Comparison of *mMorrbid* expression levels in AML12 normal murine hepatocytes and Hepa1-6 hepatoma cell line with RT-qPCR. (D) Comparison of *mMorrbid* expression levels in wild-type mouse liver and DENA model of liver cancer with RT-qPCR (E) Estimation of the efficacy of ASOs targeting *mMorrbid* in AML12 cells after 24 h of transfection with RT-qPCR. (F) Efficacy of *mMorrbid* lncRNA inhibition using the mix of 4 most efficient ASOs (ASO4, ASO5, ASO7, ASO13) was analyzed with RT-qPCR. (G) Viability assay of Hepa1-6 and AML12 cells depleted in *mMorrbid*, on the 1, 1.5, 2, 3, and 4 days of knockdown, normalized to control the luciferase antisense oligonucleotides (LUC ASO) treatment and viability at day 1 after the initial transfection. (H) Microscopy images of wound healing assay in AML12 (top) and Hepa1-6 (bottom) *mMorrbid* KD cells. (I) Wound-healing assay of *mMorrbid* knockdown (KD) and LUC control KD in AML12 and Hepa1-6 cells. The wound was introduced right after the initial transfection, and data was normalized to the wound area at the 0h time point. Results show mean \pm SD. n.s.—not significant. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

We demonstrated that the transfection of the ASO mix results in an ~80% decrease of the *Morrbid* lncRNA expression in AML12 cells in comparison to control ASO (targets the firefly luciferase gene, marked as LUC control) (Figure 35F). The efficacy of the lncRNA downregulation after 24 h of ASO treatment was confirmed with the FISH analysis (Supplementary Figure S3). We estimated the viability of *mMorrbid* depleted Hepa1-6 and AML12 cells by the resazurin assay in optimized conditions. In the case of normal AML12 cells, viability gradually decreased following ASO transfection, reaching a 70% reduction on day four (Figure 35G). On the other hand, cancer cells after an initial slight decrease of viability (about 20%), completely recovered on day four after ASO transfection (Figure 35G). We estimated the impact of *mMorrbid* lncRNA on the migration ability of AML12 and Hepa1-6 cells using the wound-healing assay. In the case of AML12 cells, the migration rate of hepatocytes with *mMorrbid* knockdown (KD) was significantly reduced in comparison to the control cells

(Figures 35H-I). On the other hand, Hepa1-6 cells depleted in mMorrbid showed no changes in migration (Figures 35H-I). Thus, mMorrbid lncRNA is highly represented in a non-cancerous hepatocyte cell line and promotes the cell viability and migration of normal liver cells. Unlike its human homolog, murine Morrbrid lncRNA was not upregulated in the cancer cell line, and its inhibition did not negatively affect the proliferation or migration of murine cancerous hepatocytes.

5.2 Inhibition of mMorrbid enhances expression of pro-apoptotic protein Bim

mMorrbid lncRNA was reported to promote the H3K27me3 histone modification of the Bim gene promoter through interaction with the polycomb repressive complex 2 (PRC2) in short-lived myeloid cells [51]. We studied mMorrbid involvement in the regulation of Bim in murine hepatocytes. Like in immune cells downregulation of mMorrbid resulted in the upregulation of Bim at mRNA (up to 1.7-fold for all Bim isoforms) and protein levels (3.5-fold) (Figure 36A-B). In addition to Bim, several other pro-apoptotic proteins were also elevated, like Bax, Bid, Bok, and most significantly PUMA and NOXA (Figure 36D). Despite the upregulation of pro-apoptotic proteins, the TUNEL assay demonstrated no apoptosis in cells depleted in mMorrbid (cell treated with Doxorubicin were used as a positive control) (Figure 36E). The activity of pro-apoptotic proteins can be blocked by pro-survival proteins. We measured the expression of MCL1 and BCL-xL and found them upregulated in mMorrbid KD cells (Figure 36F-H).

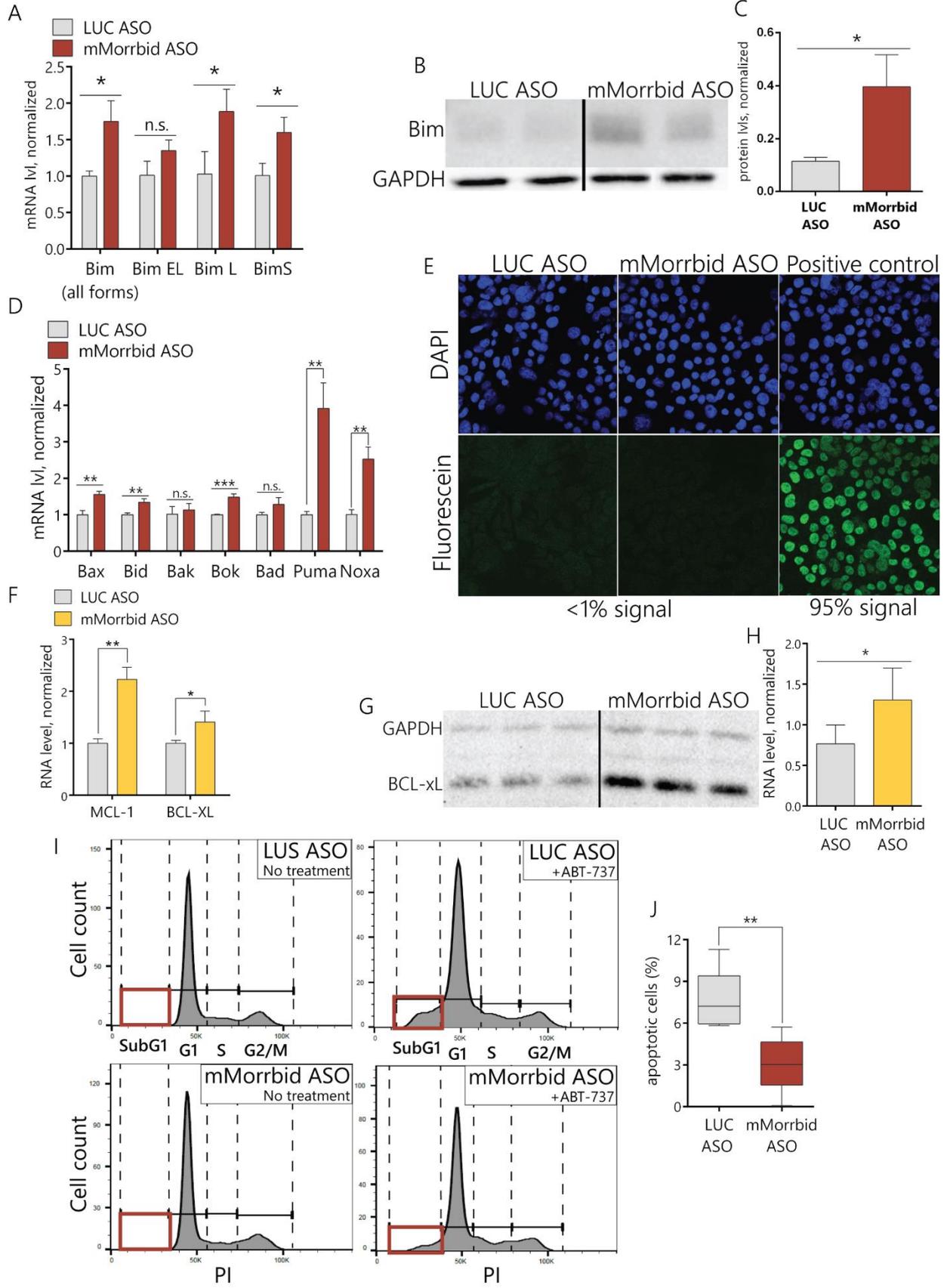


Figure 36. Analysis of mitochondrial apoptosis in mMorrbid depleted cells. (A) RT-qPCR analysis of Bim isoforms in mMorrbid KD cells. (B) Estimation of Bim protein level in mMorrbid KD cells with western blot. (C) Quantitative analysis of Bim protein level measured with western blot using ImageJ software normalized to GAPDH. (D) RT-qPCR analysis of pro-apoptotic genes expression in mMorrbid KD cells. (E) Microscopy images of the TUNEL assay of mMorrbid KD, LUC KD, and positive control treated with Doxorubicin cells. (F) RT-qPCR analysis of Bcl-xL and MCL1 genes expression in mMorrbid KD cells. (G) Estimation of BCL-xL protein level in mMorrbid KD cells with western blot. (H) Quantitative analysis of BCL-xL protein level measured with western blot using ImageJ software normalized to GAPDH. (I) Flow cytometry analysis of the SubG1 phase of PI-stained mMorrbid KD and LUC control KD cells with and without ABT-737 inhibitor treatment. Framed in red are apoptotic cells in the SubG1 phase. (J) Quantitative analysis of the SubG1 phase of PI-stained mMorrbid KD and LUC control KD cells treated with ABT-737 inhibitor. Results show mean \pm SD. n.s.—not significant. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

Western blot analysis demonstrated that BCL-xL is elevated about 1.3-fold after mMorrbid KD (Figure 35G, H). To investigate if apoptosis is blocked in mMorrbid KD cells by upregulation of pro-survival proteins, we treated knockdown and control cells with small molecule inhibitor of BCL2, BCL-xL, and BCL-W ABT-737 for 24 h, and then fixed cells and stained with PI. The apoptosis rate was measured using flow cytometry analysis as a SubG1 phase of the cells cycle. Surprisingly, we could not detect any increase in the apoptosis rate following BCL-xL inhibition (Figure 36I, J). Moreover, a slight improvement in cell viability (~4.4%) was detected in mMorrbid KD in comparison to LUC control treated with ABT-737 (Figure 35J). Thus, depletion of mMorrbid leads to slight upregulation of Bim cellular levels, however, it does not lead to an increase in apoptosis.

5.3 mMorrbid KD leads to differential changes in the genes expression related to apoptosis and cell cycle progression at transcriptome and proteome levels.

To uncover alternative roles of mMorrbid lncRNA in normal hepatocytes, we performed RNA-seq analysis of Morrbrid depleted AML12 cells and defined 1988 genes with significantly altered expressions (p -value < 0.05). Among them, 1244 were upregulated and 744 downregulated (the two-fold change was used as a threshold). Most of these genes are involved in signal transduction in cancer (MAPK pathway, p53, and NF- κ B pathways); apoptosis, peroxisome, and mitochondrial metabolisms (Figure 37A). Thus, Morrbrid lncRNA is significant for the functioning of normal hepatocytes and participates in multiple processes that influence the decreased viability and motility of KD.

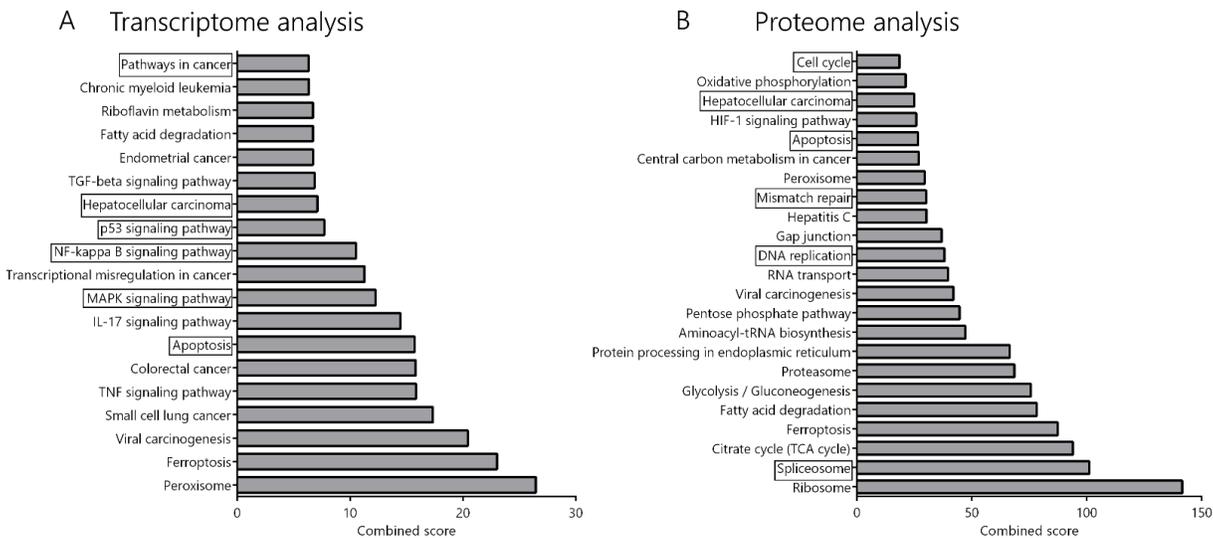


Figure 37. Summary of RNA-seq (A) and LC-MS (B) analyzes of genes differentially expressed in cells with mMorrbrid depletion.

LC-MS quantitative proteome analysis of mMorrbrid KD cells identified 1818 proteins, of them 363 were significantly upregulated and 157 downregulated. Differential expression was observed for proteins involved in cancer progression (TGFb2, SMAD4, MYC, PIK3R3, etc),

mitochondrial(NDUFA8, ATP5B, etc), lipid metabolism (HADH, ALDH3A2, etc.), and for proteins involved in the regulation of apoptosis, cell cycle, DNA replication, mismatch repair, RNA transport, components of the ribosome (Figure 37B). In addition to that, proteome analysis identified differential changes in components of the spliceosome machinery.

5.4 mMorrbid contributes to the regulation of NRAS oncogene alternative splicing in murine hepatocytes

We mapped the RNA-seq data against the splicing graph to detect novel alternative splicing (AS) events. We found 84 AS events for 79 genes with significant changes (generalized linear model (GLM), quasi-likelihood ratio test and BH correction; $p < 0.05$): 38 of them represent the differential expression of alternative 5'-donors, 30—cassette exons, 15—3'-acceptors and one—retained intron (Figure 38A and Supplementary Table S4). Among them, we focused our attention on the proto-oncogene NRAS, which demonstrated a significantly increased inclusion of cassette exon in hepatocytes (Figure 38B). NRAS is a member of the Ras gene family (NRAS, KRAS, and HRAS), which is involved in the regulation of cell proliferation and migration by inducing downstream signaling cascades, such as MAPK/Erk and PI3K/Akt [216]. The downregulation of mMorrbid leads to the increased incorporation of a 96-nt-long alternative exon into NRAS mRNA (Ensembl ID ENSMUSE00000742446) between the first and second exons. This cassette exon contains a premature stop codon. It has been previously shown that alternative splicing can produce exons that introduce premature translation termination codons (PTC) in transcripts. Such PTC-containing splice variants are degraded through the nonsense-mediated mRNA decay (NMD) pathway [217]. We, therefore, named this

96-nt exon the NRAS PTC exon. Its inclusion into mature mRNA can result in the nonsense-mediated mRNA decay degradation of the NRAS transcript, which we called NRAS PTC transcript. The main NRAS transcript that leads to the production of the mature NRAS protein is named as the NRAS-no-PTC transcript or simply the NRAS transcript.

NMD pathway was found to contribute to the fine-tuning of gene expression via so-called regulated unproductive splicing and translation (RUST) [218]. RUST modulation of transcript levels is achieved by the enhanced or decreased production of PTC-containing splice variants further degraded by NMD without protein synthesis. Lewis et al. [219] estimated that ~30% of alternatively spliced exons introduce PTC. This fact demonstrates the widespread coupling of alternative splicing and NMD for the regulation of gene expression.

To differentiate and compare the expression levels between various NRAS isoforms, we designed a set of primers (Figure 38C) targeting either the region shared by all NRAS transcripts (NRAS total), the isoform that lacks the PTC exon (no PTC), or the NRAS PTC transcript (Figure 38C). To detect NRAS PTC transcripts, we used primers that amplify the entire PTC exon (primer pair PTC 96nt), primers laying across the junctions with neighboring exons 1 and 2 (primer pair PTC junction), and primers that span across the PTC-exon 2 junction (PTC downstream or PTC down) (Figure 38C). Using RT-qPCR, we confirmed that the relative expression of the NRAS PTC transcript to the NRAS total is approximately 3.5-fold higher in cells with depleted mMorrbid lncRNA than in control cells (Figure 38D), while the expression level of the isoform lacking the PTC exon is unaffected by mMorrbid KD (Figure 38D).

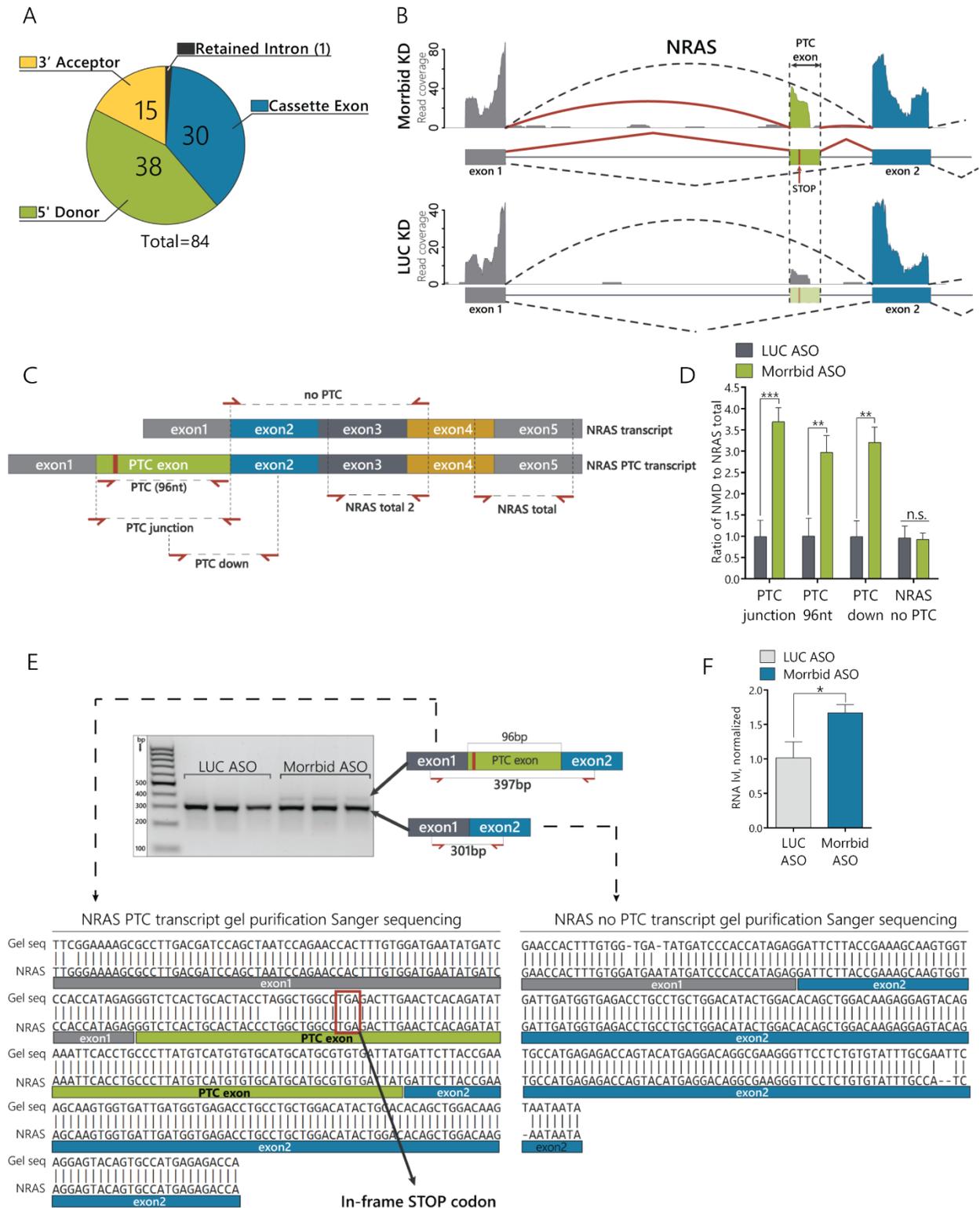


Figure 38. mMorbid downregulation leads to changes in alternative splicing of NRAS proto-oncogene. (A) Summary of differential changes in alternative splicing patterns in mMorbid KD cells. (B) RNA coverage of the

zoomed region of the NRAS transcript in mMorbid and LUC control knockdowns. The solid and dashed arced lines represent the RNA coverage of the splice junctions. (C) Schematic representation of qPCR primers positions used in the study to detect various isoforms of NRAS. (D) RT-qPCR analysis of the NRAS isoform expression after mMorbid KD and control LUC KD. (E) Top. Image of amplicon separation by agarose gel electrophoresis. Amplicons were obtained with primers spanning across the alternative NRAS exon to amplify the PTC and no PTC NRAS transcripts. Bottom. Alignment of the Sanger sequence data of PCR products spanning across NRAS cassette exon on the NRAS genome data (NCBI Gene ID: 18176). (F) RT-qPCR analysis of the NRAS pre-mRNA level in the control LUC and mMorbid KD cells using primers that amplify the fragment with exon1 and intron. NMD: nonsense-mediated mRNA decay. Results show mean \pm SD. n.s.—not significant. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. © 1996-2020 MDPI (Basel, Switzerland).

Then, we amplified complementary DNA (cDNA) using primers laying within exon 1 and exon 2 and spanning across the PTC exon (Figure 38E) and analyzed products on the agarose gel. There are two evident bands (Figure 38E) that correspond to NRAS transcripts with (397nt) and without alternative PTC exons (301nt), as was confirmed by Sanger sequencing (Figure 38E). Additionally, we estimated the amount of NRAS pre-mRNA by RT-qPCR and found an increase in the mMorbid KD cells (Figure 38F). Therefore, mMorbid lncRNA downregulation leads to the increased expression of NRAS pre-mRNA and an increased portion of the alternative NRAS PTC transcript with a cassette PTC exon.

5.4.1 NRAS PTC transcript undergoes NMD decay in the cytoplasm

To explore whether the NRAS PTC transcript is indeed degraded in the cytoplasm via the NMD pathway, we separated the nuclear and cytoplasmic fractions of the RNA from the mMorbid KD and control cells.

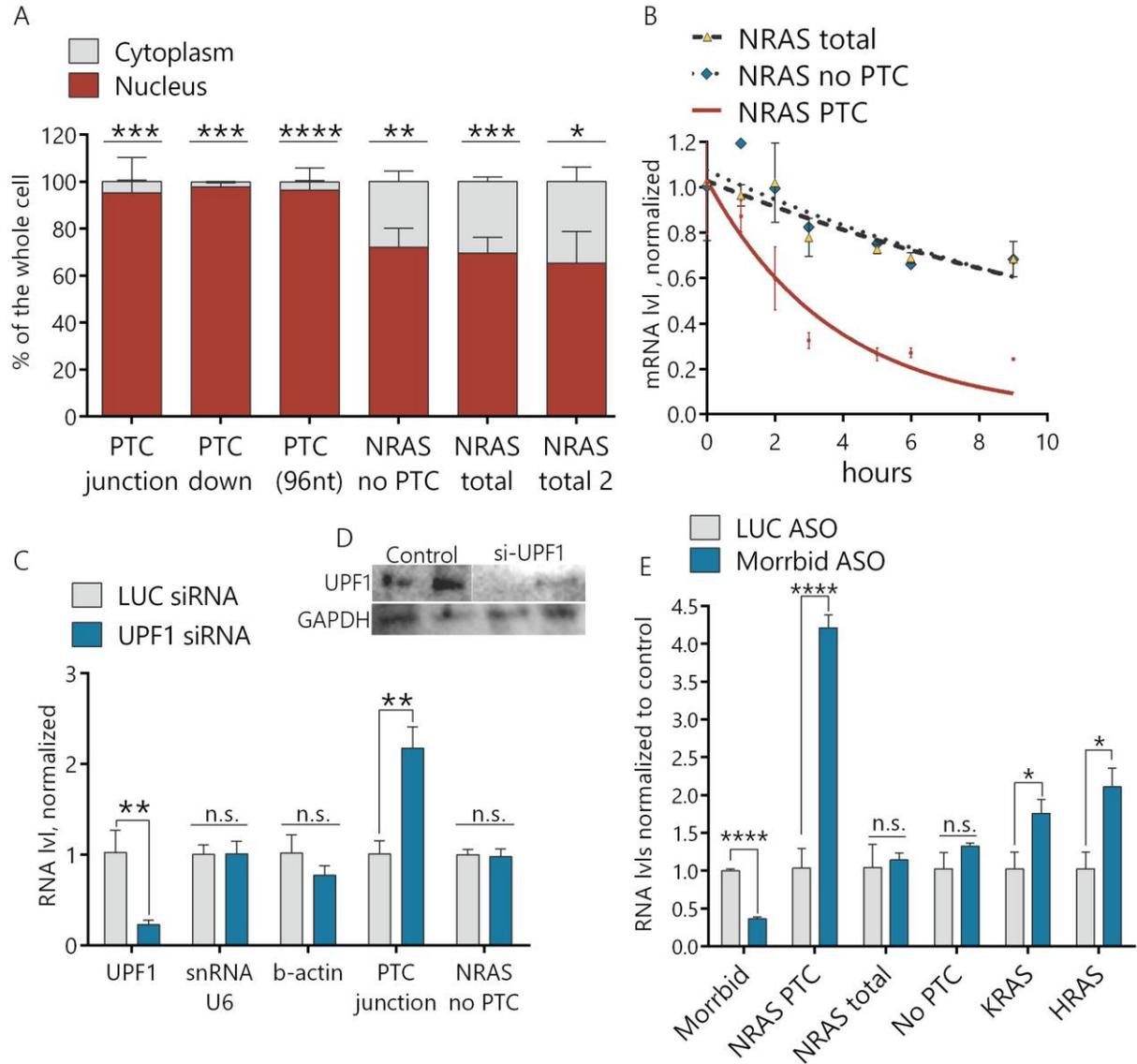


Figure 39. NRAS PTC transcript is degraded in the cytoplasm via non-sense mediated decay. (A) RT-qPCR analysis of NRAS PTC, NRAS no PTC, and NRAS total in the nuclear and cytoplasmic fractions extracted from AML12 cells. (B) Estimation of the NRAS transcripts degradation rate by an actinomycin D assay. (C) RT-qPCR analysis of gene expressions after 6 days of knockdown of the UPF1 protein. (D) Western blot for UPF1 protein after 6 days of RNAi-mediated inhibition. (E) RT-qPCR analysis of expression of Ras family proteins, NRAS (total and NMD forms), KRAS, and HRAS after Morrbid KD Results show mean \pm SD. n.s.—not significant. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ and **** $p < 0.0001$.

We demonstrated by RT-qPCR that the distribution of both the NRAS total and NRAS-no-PTC transcripts in the cytoplasm versus nucleus was ~3:7. At the same time, the NRAS PTC transcript was almost undetectable in the cytoplasmic fraction (> 95% was localized in the nucleus) (Figure 39A). To confirm the NMD nature of NRAS PTC transcript degradation, we measured the half-life of transcripts by the actinomycin D assay. Blocked transcription in AML12 cells by actinomycin D led to a time-dependent gradual decrease of all NRAS transcripts, but the degradation rate of the NRAS PTC transcript was ~four times higher than for the total NRAS and NRAS-no-PTC transcripts. The half-life of the NRAS PTC transcript was 2.6 h, while the half-life of the NRAS transcript with the omitted PTC exon was 10.9 h and NRAS total—11.8 h (Figure 39B). These results support the hypothesis that the NRAS PTC transcript is quickly degraded in the cytosol by the NMD mechanism. For additional proof of the NMD pathway involvement, we used siRNAs to downregulate the key NMD factor UPF1 (siRNA-targeting firefly luciferase gene (Luc siRNA) was used as a control). Downregulation of the UPF1 protein increased the NRAS PTC transcript, while the NRAS transcript lacking the PTC exon remained unchanged in UPF1 KD cells (Figure 39C). This data confirms that the level of the NRAS PTC transcript depends on the NMD factor UPF1 and finally proves the NMD of the NRAS PTC transcript in the cytosol.

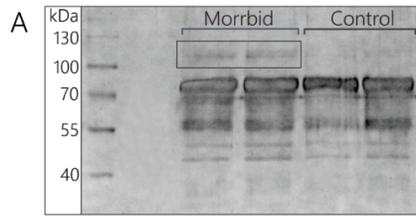
5.4.2 KRAS and HRAS members of Ras family are increased in response to mMorrbid KD

Other members of Ras protein family undergo AS regulation of their expression, HRAS via stress-induced AS-dependent NMD degradation [220], and KRAS via AS switch between isoforms KRAS4A and KRAS4B [221]. Ras proteins are known to functionally compensate each other. For example, in the case of NRAS deficiency, KRAS signaling coordinated cellular

processes through Raf and Akt [222]. We analyzed the expression of KRAS and HRAS following mMorbidity KD. We found that simultaneously with NRAS PTC transcript the mRNA levels of KRAS and HRAS were also increased two-fold (Figure 39D). Thus, the deregulation of NRAS AS promotes the upregulation of KRAS and HRAS expression.

5.4.3 Morbidity lncRNA interacts with the SFPQ-NONO heterodimer

LncRNA can regulate splicing by multiple mechanisms [112]. To find functional protein partners of mMorbidity lncRNA, we performed a modified capture hybridization analysis of RNA targets (CHART) in AML12 cells followed by a liquid chromatography-mass spectrometry (LC-MS) analysis of proteins crosslinked with mMorbidity [202]. We used several biotinylated probes complementary to mMorbidity lncRNA exons that cover most of the Morbidity-annotated transcripts. Two independent Morbidity-CHART experiments followed by protein separation using Laemmli polyacrylamide gel electrophoresis (PAGE) resulted in the enrichment of specific protein bands (molecular mass ~120 kDa) in comparison to the controls (Figure 40A). In-gel trypsinolysis followed by an LC-MS analysis showed that these proteins are SFPQ and NONO (Figure 40B). SFPQ and NONO are members of the *Drosophila* behavior/human splicing (DBHS) protein family, each containing two RNA-binding motifs. Together, they form the SFPQ-NONO heterodimer that regulates multiple steps of the RNA metabolism, including alternative splicing [223]. To confirm that SFPQ and NONO directly interact with Morbidity lncRNA in cells, we performed a RIP (RNA immunoprecipitation) analysis using antibodies against SFPQ and NONO proteins.



B

Type	Scheme	SFPQ	NONO
Morrbid IncRNA		YES	YES
Control IncRNA		NO	NO
NRAS PTC forward		NO	NO
NRAS PTC reverse		NO	NO
NRAS no PTC forward		YES	NO
NRAS no PTC reverse		NO	NO

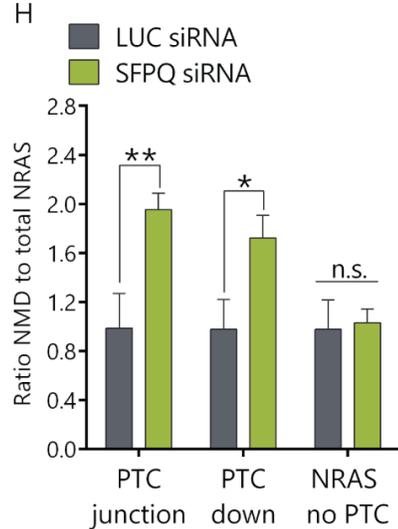
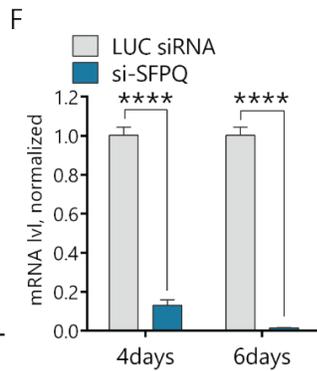
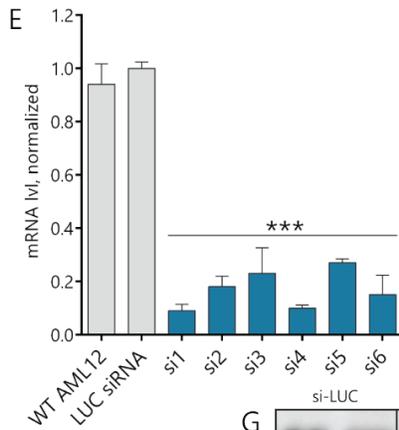
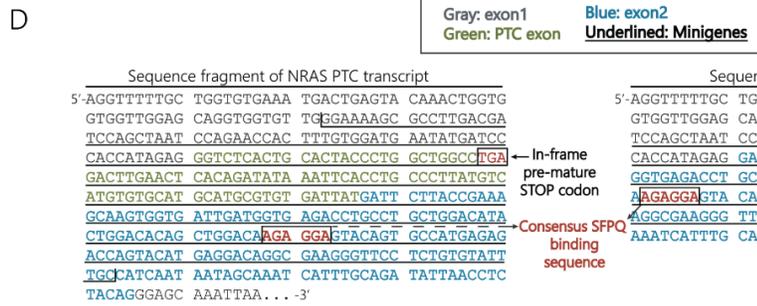
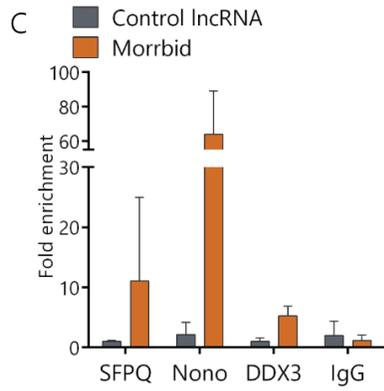


Figure 40. *mMorrbid* directly interacts with SFPQ-NONO heterodimer to modulate AS of NRAS PTC isoform. (A) SDS-PAGE of samples obtained in the CHART experiment, framed are *Morrbid* specific bands, excised and analyzed by LC-MS. (B) Summary table of the Capture Hybridization Analysis of RNA Targets (CHART) and RNA pulldown assay results. (C) Fold enrichment of *Morrbid* lncRNA in the RNA immunoprecipitation assay (RIP) performed with SFPQ and NONO antibodies, as well as DDX3 and IgG antibodies as controls, quantified with RT-qPCR. (D) Fragments of NRAS PTC and NRAS no PTC transcripts with designated exon1, PTC exon, exon2, positions of pre-mature STOP-codon, consensus SFPQ binding sites, generated NRAS minigenes. (E) Efficiency test of siRNAs targeting SFPQ mRNA in AML12 cells after 24h of KD analyzed by RT-qPCR (F-G) Estimation SFPQ mRNA (F) and protein (G) level after RNAi-mediated inhibition by RT-qPCR and Western-blot respectively. (H) Relative expression of NRAS isoforms after 6 days of inhibition of the SFPQ protein (RT-qPCR analysis). siRNA: small interfering RNA. Results show mean \pm SD. n.s.—not significant. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ and **** $p < 0.0001$. © 1996-2020 MDPI (Basel, Switzerland).

mMorrbid lncRNA was enriched in SFPQ (12-fold) and NONO (60-fold) fractions in comparison to nonspecific IgG control (Figure 40C). RNA-binding helicase DDX3 was used as additional control and demonstrated only five-fold enrichment for *mMorrbid* lncRNA.

A control RNA did not show any enrichment in SFPQ and NONO protein fractions in comparison to the IgG control. Thus, the results of the combined CHART LC-MS and RNA pulldown assays clearly show that *Morrbid* lncRNA directly and specifically interacts with SFPQ and NONO proteins. Previously, it was shown that SFPQ and NONO proteins can interact with pre-mRNA regulatory elements and affect alternative splicing [224,225]. To verify if the SFPQ-NONO heterodimer can directly interact with NRAS pre-mRNA, we performed an RNA pulldown analysis using biotinylated transcripts of four minigenes. These constructs were: (1) minigene coding a full NRAS PTC exon flanked with 68 nt of exon1 upstream and 137 nt of exon2 downstream, (2) minigene coding an exon1-exon2 fusion with a missing PTC exon, and (3) and (4) controlling minigenes that were created by inverting sequences of the first two constructs (Figure 40B, D). A mass-spectrometry analysis of RNA pulldown samples

unambiguously confirmed the interaction of SFPQ proteins with the NRAS total minigene (Figure 40B). Therefore, the SFPQ protein interacts directly with both lncRNA mMorrbid and NRAS mRNA, while the NONO protein interacts with mMorrbid lncRNA. To study if SFPQ is involved in the regulation of NRAS splicing, we depleted SFPQ using the RNA interference (RNAi) technique. We designed and validated six small interfering RNAs (siRNAs) targeting SFPQ mRNA and selected the most efficient ones (Figure 40E). The SFPQ protein was not detected by Western blot analysis after six days of siRNA treatment, while some protein remained in the cells after four days of knockdown (Figure 40F, G). In SFPQ-depleted cells, the ratio of the NRAS PTC transcript to the NRAS total was increased approximately two-fold in comparison to cells transfected with control siRNA, while the expression of NRAS-no-PTC remained unchanged (Figure 40H). Thus, the SFPQ protein is involved in the regulation of the splicing of the NRAS PTC transcript.

VI. Discussion.

6.1 hMorrbid/CYTOR genetic knockout is compatible with normal viability and growth of hepatocellular carcinoma cells

Most long non-coding RNA genes have not been studied yet and a complete picture of their regulatory features and the way they interact with each other and other molecules is yet to be revealed. One approach for the investigation of lncRNAs functions are loss-of-function methods, which specifically reduce the expression of a gene of interest and are powerful tools for the discovery of its biological functions. The LOF phenotype is obtained through the execution of gene silencing techniques either involving genomic manipulations (such as CRISPR interference (CRISPRi), CRISPR/Cas9-mediated deletions, knock-ins, etc.) or targeted cleavage of an RNA transcript of interest (RNA interference or antisense oligonucleotides) [226]. A comparison of effectiveness and off-target effects of different LOF strategies demonstrated that each method has its limitations and weaknesses, and simultaneous application of several approaches is advised [211,227]. Unique biological features of lncRNAs must be considered when choosing a LOF method and interpretation the results. For example, lncRNA localization in some cases may influence the efficiency of the knockdown. Also, lncRNA may action in cis, in trans or be non-functional. Nuclear localization and association with chromatin are indicators of cis-acting lncRNA regulator, cytoplasmic or nucleoplasmic lncRNAs most probably function in trans, while a possibility that a nuclear lncRNA is a non-functional by-product of transcription should also be considered. Additionally, regulatory DNA elements within the lncRNA locus or process of active transcription may both

affect the expression of adjacent genes and lead to the phenotype that is unrelated to the lncRNA transcript itself.

An important advantage of knockdown techniques, which specifically degrade RNA molecule, while leaving lncRNA genomic locus and its transcription intact, is that they allow for the identification of lncRNA transcript-specific functions. RNA interference and antisense nucleotides are universally used for knockdown methods. RNAi utilizes transfected siRNAs molecules or endogenously expressed shRNA molecules to guide the RNA-induced silencing complex (RISC) to the target RNA molecule via complementary base pairing. Once a target lncRNA binds RISC-loaded siRNA/shRNA, it is cleaved by Ago2 and then degraded. A lot of evidence for lncRNA functional roles came from RNAi LOF studies. For example, the first affirmation that NEAT1 is a structural basis of paraspeckles came from siRNA-based phenotype assay in HeLa cells [228]. An efficient alternative to RNAi is provided by antisense oligonucleotides (ASOs), in which inhibitory activity is either promoted by catalytic RNaseH cleavage of the DNA:RNA substrates or blockage of pre-mRNA processing or translation by direct binding of modified oligonucleotides (steric-blocker oligonucleotides). It was demonstrated that RNaseH-dependent ASOs are more efficient in the nucleus than RNAi [229] and thus, for nuclear-bound lncRNAs ASOs represent a preferred knockdown method. Moreover, it was shown that ASOs lead to degradation of a nascent RNA transcript in the nucleus, thus mature lncRNA is never produced [230]. Locked nucleic acids (LNAs) are an important type of antisense oligonucleotides acting as steric blockers. LNAs are nucleic acid analogs, which are modified to have a ribose ring “locked” by a methylene bridge between the 2' oxygen and the 4' carbon. This modification results in a drastic increase of affinity to RNA and DNA. Xist targeting with LNAs resulted in its displacement from the inactive X chromosome

and allowed for the identification of Repeat C as the essential for RNA localization at Xi [231]. Another steric blocking oligonucleotide approach is offered by antisense morpholino oligonucleotides (MOs), which are widely used to inhibit the expression of protein genes by preventing mRNA translation. MOs can also be designed to bind lncRNA splice-sites and therefore deactivate RNA maturation or to target lncRNA functional regions, as it was done to investigate Cyrano and Megamind roles in zebrafish [156]. However, gene downregulation with synthetic nucleic acids possesses serious limitations. One of them is the risk of oversaturation of endogenous small RNA pathways, which may lead to abnormalities in cellular homeostasis, which can be mistaken for lncRNA dependent LOF phenotype [232]. Moreover, supraphysiological amounts of oligos may result in accumulation of aberrant RNA species inside the cell that can lead to non-specific changes in gene expression [233]. Minimization of these undesirable effects can be achieved through careful optimization of oligonucleotides sequences and doses before the transfection. Additionally, RNAi and ASOs techniques exhibit significant sequence-dependent off-target effects, that are extremely challenging to completely avoid [234,235]. One strategy, that was suggested to reduce off-target effects is targeting multiple regions within lncRNA transcript with a set of siRNAs/ASOs and pinpointing the concordance in phenotypes from individual probes [236]. Development of the novel methods targeting lncRNA transcripts for degradation continues. For example, the insertion of self-cleaving ribozymes into lncRNA sequences recently demonstrated promising results and proved to have efficiency similar to RNAi [237].

Various CRISPR-based methods have also been widely utilized in lncRNA studies. CRISPRi is a method that allows inhibition of lncRNA transcription without introducing changes into the underlying DNA sequence. CRISPRi uses a nuclease-deficient version of Cas9 (dCas9) which still possesses its RNA-dependent DNA-binding activity fused to the KRAB (Krüppel-associated box), which is recruited to the TSS of the target lncRNA by sgRNA and catalyzes repressive chromatin modifications around the TSS. CRISPRi has been successfully applied in high-throughput functional lncRNA characterization, such as a screen for growth regulators among more than 16 thousand lncRNAs in 7 different cell lines [238]

CRISPR/Cas9 is used to generate knockout of the target lncRNA via genomic excision of the entire lncRNA locus [239] or regulatory elements in the DNA sequence, like promoters [239]. Deletion of the entire gene does not allow to distinguish between lncRNA role versus DNA sequence. Also, depletion of such significant genome region may cause perturbations in the chromosome architecture, affect neighbor genes, resulting in phenotype changes. On the other hand, removal of just the promoter region terminates transcription, while leaving most of the DNA sequence unchanged. Another method Homology-Directed Repair (HDR) of the Cas9-induced DSBs (CRISPRn HR) can be used to knock-in DNA elements into lncRNA locus. Transcriptional terminator sequences inserted after TSS abolish the transcription while leaving the DNA sequence almost intact, providing a method, which allows distinguishing between the role of active DNA elements and the transcript/transcription [86].

CRISPR-based methods cannot be used to manipulate lncRNA loci which overlap PCGs, located antisense, expressed from a bidirectional promoter, or located close to a protein-coding gene. A genome-wide analysis found that only 38% of all lncRNA can be safely genetically modified without the risk of disrupting the expression of neighbor genes [240]. For example, the

CRISPRi approach was used to knockdown lncRNA NOP14-AS1 expressed from a bidirectional promoter of MFSD10 protein. All tested sgRNAs targeting NOP14-AS1 also lead to downregulation of MFSD10. However, NOP14-AS1 knockdown using antisense LNA GapmeRs did not affect NOP14-AS1 expression. A similar effect was obtained for 4 other lncRNA, including HOTAIR, as well as for protein-coding mRNA TP53 [240]. Like RNAi, CRISPR/Cas9 and CRISPRi technologies also exhibit off-target effects. Although they reported having higher fidelity [226], studies suggest off-target cannot be neglected when using CRISPR and dCas9 may bind up to 1000 off-target sites depending on sgRNA sequence [241]. Comparison between transcriptome profiles of the same lncRNA inhibited by RNAi, ASOs, or CRISPRi revealed little overlap in differentially expressed genes between these methods, suggesting significant method-specific off-target effects [226].

In this study we chose to terminate transcription of hMorrbid and CYTOR by performing CRISPR/Cas9 excision of the promoter regions leaving the rest of the loci complete to avoid serious perturbations in the genome architecture. Also, we confirmed that deleted regions do not overlap any protein coding genes and located more than 2000 base pairs from the closest promoter as is recommended [240].

Paralogous human lncRNAs Morrbid and CYTOR are highly overexpressed in many types of cancers. Multiple studies of the hMorrbid/CYTOR RNAi-based LOF phenotype suggested that hMorrbid/CYTOR are crucial for cancer cell proliferation [9,10,12,13], invasion [9,12], migration [9,12], EMT (epithelial to mesenchymal transition) progression [9,13]. Human Morrbid/CYTOR give rise to multiple abundantly expressed transcripts, some of them may execute transcript-specific roles. Analysis of

literature showed that most of the research groups used only one or two siRNAs/shRNAs targeting just a fraction of all expressed hMorrbid/CYTOR transcripts. To completely repress expression of both hMorrbid and CYTOR genes and clarify their functions in hepatocellular carcinoma cells we generated, for the first time, hMorrbid/CYTOR genetic knockout in Huh7 cells. CRISPR-Cas9 mediated double excision of two 2kbs regions around promoters of hMorrbid and CYTOR genomic loci was performed and confirmed with whole-genome next-generation sequencing, PCR, and Sanger sequencing. The acquired knockout cell line was put through standard tests: viability assay, cell cycle flow cytometry analysis of PI-stained cells, migration wound healing assay, TUNEL assay for apoptosis (Figure 25). We demonstrated that abundantly expressed lncRNAs hMorrbid and CYTOR are dispensable for HCC cells viability, but slightly influence the cell cycle and migration. Despite previous reports suggested delayed progression through the cell cycle of hMorrbid/CYTOR knockdown cells, we observed a slight accumulation of cells in the G2 phase of hMorrbid/CYTOR KO cells (Figure 25C). It corresponds to the upregulation of CyclinD1 in KO cells, which potentially could be a compensatory cell mechanism to overcome hMorrbid/CYTOR loss (Figure 27B, C). Other potential compensatory mechanisms were determined by proteome analysis of the KO cell line, as it demonstrated a significant increase of the check-point proteins and p53 regulators, e.g. SHPRH, RAD9A, Nek9, HEXIM, and downregulation of transcription regulators Elp2, Elp3, GTF2F (Figure 26). Deregulation of these players may explain close to normal cell cycle progression of KO cells.

Analysis of expression of previously reported hMorrbid/CYTOR targets showed that these genes are either not affected or their expression slightly changes in the direction opposite of what was suggested by the literature (Figure 27). One explanation of these discrepancies is the

tissue-dependent function of hMorrbid/CYTOR. However, several cases were previously reported specifically for hepatocellular carcinoma cells. For example, hMorrbid/CYTOR was proposed to act as a ceRNA for miR-193b in HCC cells thereby enhancing the expression of CyclinD1 [15]. Our study shows upregulation of CyclinD1 after depletion of hMorrbid/CYTOR (Figure 27B, C). Similarly, knockdown of hMorrbid/CYTOR in HCC cells was shown to cause EpCam inhibition [16], however, in Huh7 KO cells EpCam was unaffected (Figure 27A).

All previously published hMorrbid/CYTOR studies used either siRNA or shRNA to promote lncRNA downregulation [9,10,12,13,22,23,25]. It was demonstrated that RNAi methods are less effective than ASOs when targeting nuclear lncRNAs [242]. As hMorrbid/CYTOR was found to localize both in nucleus and cytoplasm it is possible that some of the nuclear-localized transcripts were mistargeted by RNAi methods. Additionally, RNAi agents must target all the multiple expressed isoforms, which may or may not have overlapping regions. L. Nötzold et.al used a pool of 30 siRNAs in their LOF study to target hMorrbid/CYTOR [168]. They also reported that several earlier identified hMorrbid/CYTOR targets (EGFR, Vimentin, p15, p21, etc) were not confirmed in their system [168].

Most probably, the phenotype differences with previously published results stem from RNAi/CRISPR method-specific off-target effects. RNA interference post-transcriptionally deactivates a target molecule (mRNA or lncRNA) by introducing an Ago2-mediated cleavage of a target transcript at a region complementary to siRNA or shRNA. The biological specificity of this mechanism produces sequence-dependent off-target effects [211]. Additionally, if a transcript is missing an RNAi targeted sequence it

escapes the cleavage and therefore, expression from the gene locus is only partially repressed. Another issue with RNAi is a knockdown efficiency, which often is less than a hundred percent, while for abundantly expressed lncRNAs even 20% of a wild-type expression level may be enough to compensate for the functional role. CRISPR genome editing helps to avoid the last two issues: it interrupts gene expression at the genome level completely disrupting expression of all transcripts expressed from the locus. At the same time, the CRISPR-Cas9 system was reported to have off-target DNA cleavage activity with as little as 3-5 mismatches in PAM or sgRNA sequences [242]. In our work, we eliminated the possibility of CRISPR-Cas9 off-target cutting event by complete genome sequencing of knockout cells. NGS analysis verified the absence of off-target cleavages and confirmed an excision of intended 2kbs from hMorbid and CYTOR promoter regions (Supplementary Figures S1 and S2).

Poor correlation between LOF studies performed using different methods in the same biological background has been extensively discussed [211,227,243]. Direct comparison between phenotypes obtained using CRISPR/Cas9 and shRNA-based screening technologies identified different biological groups of genes, which showed little correlation [243]. Discrepancies between RNAi-based and CRISPR-based lncRNA studies have been reported for MALAT1, NEAT1, lincRNA-p21, Megamind, and HOTAIR lncRNAs [40,41,210–212]. lincRNA-p21 depletion with knockout methods [41] and with RNAi methods [212] in the same cell type triggered the differential expression of not overlapping sets of genes and led to different conclusions. The RNAi-based study concluded that lincRNA-p21 acts *in trans* and regulates a wide range of genes [212] and the knockout study deduced that lincRNA-p21 *in cis* regulates expression of p21 [41]. MALAT1 genetic knockout in mice did not confirm any of the previously reported effects on cell proliferation and viability acquired with RNAi methods [40].

Knockout of MALAT1 in liver and lung cancer cells demonstrated that MALAT1 is dispensable for cell proliferation and viability [40], despite many MALAT1 knockdown studies reporting the opposite [244,245]. Multiple lncRNA NEAT1 knockdown studies reported its positive regulation of cell viability and proliferation [246,247], however, knockout of NEAT1 in mice resulted in viable and fertile animals with the only reported phenotype to be the absence of paraspeckles [210]. Overall, this suggests that RNA interference-based knockdown methods need to be carefully analyzed for efficiency and possible off-target effects, as well as reproduced with many siRNAs or shRNAs to validate a gene-specific effect. Additionally, knockout models should also be applied for the final validation of gene functions and its importance for cell viability.

Regulatory roles of some lncRNA genes depend on the expression of the mature RNA transcript, while others function through DNA sequence, transcription/splicing activity, and independently of the RNA transcript itself. A combination of methods targeting genomic locus (CRISPR) or lncRNA transcript (CRISPRi, RNAi, ASO, LNA) should be used to fully distinguish transcript-dependent and transcript-independent lncRNA gene roles.

6.2 Evolutionary conserved fragment exonCh has a role in hepatocellular carcinoma cells priming for apoptosis

LncRNA genes lack strong interspecies sequence homology [45]. However, many well-studied lncRNAs, such as Xist, Air, MALAT1, NEAT1, etc perform the same functions in different organisms despite poor conservation of primary nucleotide sequence. This phenomenon may be explained by the hypothesis that lncRNA sequence is discrete and contains “functional

blocks”, important for regulation while the rest of the sequence is arbitrary and is not under selection pressure [45]. The hypothesis of the modular structure of lncRNA suggests that these molecules contain distinct sequence motifs responsible for interaction with functional partners, such as proteins, miRNAs, or carry other regulatory roles [47,48]. For example, it was proposed that lncRNA sequences that originated from neofunctionalised transposable elements (TE) are examples of such functional domains [48]. They were demonstrated to undergo evolutionary selection as parts of lncRNA exons and are likely to promote nuclear localization of the host lncRNA [248]. A sequence motif within lncRNA BORG was found to define its nuclear localization [249] and G-rich element within lncRNA Braveheart was responsible for its interaction with a zinc-finger protein CNBP [250]. Sequence conservation has been proposed to be one of the indicators of such lncRNA functional modules [7]. Experimental evidence that conserved lncRNA fragments carry a specific function has been obtained for several cases. For example tumor suppressor lncRNA LINC-PINT that transcriptionally represses a set of tumor-promoting genes. LINC-PINT function is dependent on a highly conserved sequence motif, that interacts with PRC2 to silence LINC-PINT target genes [46]. Another example is the 67 nt sequence region of Cyrano lncRNA conserved between human, mouse, and zebrafish. Overexpression of both human or mouse Cyrano transcripts was enough to rescue morphological changes in Zebrafish embryos caused by Cyrano depletion [156].

One of the Morrbid exons, which we named exonCh, contains a region conserved between human and mouse. We looked at exonCh specific functional roles and found that it is implicated in apoptosis priming of HCC cells increasing their susceptibility to MCL1 inhibition. Generated cell line overexpressing exonCh-containing transcript M-217 had a delayed proliferation, a ~1.7-fold increase in apoptosis (Figure 29), and elevated expression of pro-

apoptotic genes (Figure 31). On the other hand, cells overexpressing isoform with mutated exonCh had growth and apoptosis rates at the same level as WT (Figure 29).

Apoptosis, a mechanism of programmed cell death, is tightly regulated to prevent both unnecessary cell death and survival of cells potentially harmful to the organism. The two major apoptosis pathways are the extrinsic pathway (activated by extracellular stimuli) and the intrinsic or mitochondrial apoptosis pathway, activated by internal cellular stresses (DNA damage, oxidative stress, etc.). The deficiency of the mitochondrial apoptosis pathway has been implicated in the development of various diseases, particularly cancer and autoimmunity. Apoptosis resistance is one of the major hallmarks of cancer and components of the apoptotic pathway are targets for anticancer therapy. In the mitochondrial apoptosis pathway, MOMP (mitochondria outer membrane permeabilization) is a critical point at which a cell irreversibly commits to cell death, precisely regulated by BCL2 family proteins. Damaging factors trigger the expression of pro-apoptotic BCL2 proteins via different mechanisms, for example, DNA-damage activates p53, which in turn upregulates the expression of PUMA and NOXA. MOMP is a switch-like event that occurs when the concentration of active pro-survival BCL2 proteins becomes insufficient to inactivate pro-apoptotic BCL2 proteins (Figure 30). This makes an equilibrium between pro-survival and pro-apoptotic proteins one of the key determinants of cell viability. Cells expressing surplus amounts of pro-survival proteins, which buffer apoptotic signals, or insufficient amounts of pro-apoptotic proteins become unprimed for apoptosis and can withstand mild or even strong cellular stresses (Figure 33). This mechanism is widely exploited by cancers. A genetic mutation in ~90% of follicular center B cell lymphomas results in drastic BCL2 overexpression [251]. MCL1 and BCL-xL are upregulated in multiple cancers [252]. Loss

of the pro-apoptotic BCL-2 family members (BIM, PUMA, BAD, BMF, BAX) also has been linked to tumorigenesis [253].

Cells, which express pro-survival proteins in amounts that are just enough to block pro-apoptotic proteins to survive, have poor apoptosis resistance, and commit apoptosis upon mild stress or damage (Figure 33). In other words, the more cell is primed, the closer it is to cross the apoptotic MOMP threshold. Priming depends on many physiological conditions such as growth factor deprivation, changes in metabolism, or genetic aberrations. Enhanced cell priming for apoptosis has been tested and proved beneficial as anticancer drug treatment [254]. In this study, we demonstrated that the expression of conserved exonCh primes cancerous hepatocytes for intrinsic apoptosis and increases cell sensitivity to MCL1 deactivation.

BCL2 pro-survival proteins directly bind BH3-only pro-apoptotic BCL2 proteins, this interaction is mutually deactivating. BCL2, MCL1, BCL-xL, BCL-W are key pro-survival proteins working in concord to inactivate all known pro-apoptotic BCL2 family members [208]. S-63845 and ABT-737 are BH3-only mimetics designed as anticancer drugs that competitively bind pro-survival proteins, releasing pro-apoptotic proteins from inhibition. S-63845 specifically targets MCL1, while ABT-737 targets BCL2, BCL-xL, and BCL-W. ABT-737 treatment revealed that overexpression of M-217oe sensitizes cells for BCL2/BCL-xL/BCL-W deactivation, however, exonCh is not involved (Figure 34 D-F). Both M-217 and control R-217 cell lines expressing mutated exonCh reacted with a similar increase in apoptosis (1.5-fold) in response to inhibition of a BCL2/BCL-xL/BCL-W complex (Figure 34F). However, S-63845 treatment demonstrated that cells expressing exonCh specifically rely on MCL1 activity for survival, while control R-217 cells are resistant to MCL1 inhibition (Figure 34 A-C). Expression

of M-217 transcript resulted in around a 2-fold increase in apoptosis of S-63845 treated cells, against a 1.2-fold increase in the case of R-217 transcript (Figure 34B).

Deactivation of one of the pro-survival proteins is an activity test for the remaining pro-survival proteins to withstand the increased burden of the released pro-apoptotic members. R-217^{oe} control cells showed the same apoptosis rate as M-217^{oe} cells following ABT-737 treatment, suggesting that R-217^{oe} cells rely on BCL2/BCL-xL/BCL-W to resist apoptosis. Similar reactions of M-217 and R-217 cells to ABT-737 as well as different reactions to S-63845 suggest that cells expressing exonCh become susceptible to apoptosis due to the deregulation of BCL2/BCL-xL/BCL-W, which is partially compensated by MCL1 upregulation (Figure 41). ABT-737 inhibitor has an equally high affinity to BCL2, BCL-xL, and BCL-W, however, it was demonstrated to preferentially bind BCL2 rather than BCL-xL and BCL-W [255]. Other members of the BCL2 protein family may also contribute to the resulting phenotype (Figure 41).

LncRNAs have been previously reported as cancer repressors due to their apoptosis promoting properties. Such lncRNAs are typically downregulated in malignancies [256-257]. In HCC (Huh7, HepG2) overexpression of lncRNA CASC2 caused decreased cell viability and a ~15% increase in apoptosis [256]. A similar phenotype was observed in osteosarcoma cells after FER1L4 lncRNA upregulation [257] and in the human brain microvascular endothelial cells (HBMECs) as a result of FENDRR lncRNA overexpression [258]. Regulator of NF- κ b pathway NKILA lncRNA sensitizes T lymphocytes to activation-induced cell death (AICD) and ectopic expression of NKILA in T cells enhanced their susceptibility to apoptosis [259].

Surprisingly, despite the evidence that M-217 transcript overexpression primes HCC cells for apoptosis, hMorrbid and its paralogue CYTOR are generally overexpressed in cancers including hepatocellular carcinoma. Cancer LncRNA Census project included CYTOR into the list of oncogene lncRNAs confidently implicated into cancer progression [260]. Therefore, hMorrbid/CYTOR most likely are not tumor suppressors and in wild-type cancer cell setting cumulative expression of all hMorrbid/CYTOR transcripts positively affects tumorigenesis.

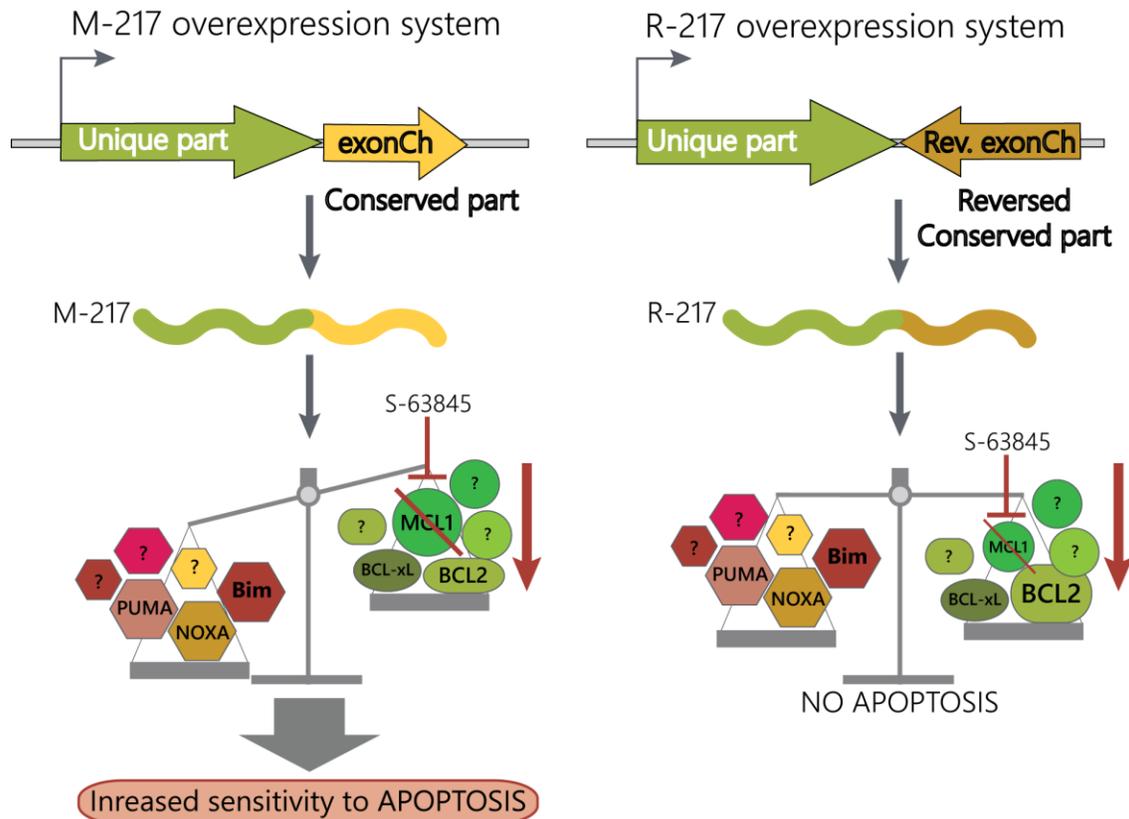


Figure 41. Effect of M-217/exonCh expression on the intrinsic apoptosis pathway. Expression of evolutionarily conserved region exonCh primed knockout cells to cell death and sensitized cells to MCL1 inhibition. Cell expressing mutated exonCh (R-217) had lower apoptosis rates and were resistant to MCL1 inhibition.

Besides its effect on apoptosis susceptibility, we also found that exonCh is involved in

the regulation of Bim expression (Figure 32). Bim levels were significantly elevated in hMorrbid/CYTOR knockout cells, however, overexpression of exonCh reversed upregulation of Bim protein. In M-217oe cells, Bim was downregulated ~60% below the wild-type level, while mutation of exonCh partially alleviated inhibition of Bim (Figure 32). In murine myeloid cells, Morrbrid regulates Bim via allele-dependent *in cis* mechanism [51]. In Morrbrid-heterozygous mice, deletion of Bim on the Morrbrid-deficient chromosome resulted in normalization of Bim expression and rescued short-lived myeloid cell numbers, while Bim deletion on a different chromosome did not [51]. In this work, the transfection of the overexpressing vector into the hMorrbid/CYTOR knockout system demonstrated that hMorrbid mature transcript can relocate to the nucleus and act *in trans* to modulate Bim expression (Figure 32).

Overall, we can conclude that involvement in the activation of the mitochondrial apoptosis is not the main function of Morrbrid lncRNA in human and mouse cells, but some transcripts may be involved in its modulation.

6.3 In murine hepatocytes Morrbrid is involved in the regulation of NRAS alternative splicing coupled with the NMD pathway

The actual amount of specific mRNA relies on the rates of its synthesis and degradation in the cell. Among many regulatory mechanisms, the degradation of some mRNA transcripts is performed by alternative splicing coupled with the nonsense-mediated mRNA decay pathway (AS-NMD). Particularly, AS results in transcript isoforms with PTC followed by the degradation of these mRNA by NMD [261]. Genome-wide studies have shown that 5% to 10% of the

Saccharomyces cerevisiae [262], *Caenorhabditis elegans* [263], and *Drosophila melanogaster* [264] transcriptomes are changed when NMD is inactivated.

Nonsense-mediated mRNA decay (NMD) is a protein synthesis quality control mechanism in cells that eliminates mRNAs containing premature termination codons (PTC) or other NMD-triggering factors. mRNA containing PTCs carry a potential threat to cell homeostasis since they can result in bulk production of non-functional proteins [265]. Detection of fault mRNAs occurs during the first round of mRNA translation. To discriminate between NMD-target and non-target mRNAs, cellular machinery utilizes either of two mechanisms: exon junction complex (EJC)-independent NMD and 3' UTR EJC- dependent NMD, where the last is a more efficient process. All types of NMD pathways rely on RNA- dependent helicase and ATPase called UPF1 in addition to several other enzymes which can vary depending on a specific mechanism [266]. UPF1 binds to a single-stranded RNA disregarding the sequence and uses ATP hydrolysis to move along the mRNA molecule in the 5' → 3' direction. The core of the nuclear EJC is composed of eukaryotic initiation factor 4A3 (eIF4A3), which is a helicase that anchors the EJC to the RNA, RNA- binding protein 8A (RBM8A), and MAGOH52. This core is joined by other proteins, including UPF3X, which directly interacts with UPF2. EJC is positioned onto mRNA after it has been spliced in the nucleus ~20-24nt upstream of the exon-exon junction. During the first round of translation, the ribosome dislocates any protein complexes bound to mRNA, including EJC and UPF1. When translating ribosome encounters termination codon, the translation termination complex composed of eukaryotic release factor 1 (eRF1) and eRF3 is formed. UPF1 is recruited to eRF1–eRF3 translation termination complex together with the serine/ threonine kinase SMG1, forming an SMG1–UPF1–eRFs (SURF) complex. UPF3X in complex with UPF2 recruits UPF1 from the termination complex, bridging

SURF to the EJC to form the decay-inducing (DECID) complex. This stimulates UPF1 helicase activity and promotes UPF1 phosphorylation by SMG1. UPF1 phosphorylation by SMG1 at multiple residues within its amino-terminal and carboxy-terminal regions is a commitment step in NMD. It prevents further rounds of translation initiation and is crucial for mRNA decay serving as a platform for recruitment of RNA degrading factors such as SMG5–SMG7 and SMG6. In EJC-dependent NMD, the EJC complex positioned downstream from the termination codon is required for stimulation of the NMD pathway. This feature justifies a ‘50–55 nucleotide rule’: NMD occurs if a PTC located ≥ 50 –55 nucleotides upstream of an exon-exon junction so that the leading edge of the terminating ribosome can’t physically remove the EJC off the junction. On the other hand, the EJC-independent NMD pathway is activated in case mRNA contains longer than ~1 kb, unstructured 3’ untranslated region (3’ UTR), and mechanistically explained by the remote location of NMD inhibitor PABPC1 from the termination codon and therefore abolished initiation of a proper translation termination mechanism [265]. NMD is an essential modulator of gene expression, implicated in various physiological processes. It facilitates cellular response to environmental changes. Deregulation of NMD signaling in humans is associated with intellectual disability and cancer.

Multiple non-coding RNAs are involved in the regulation of alternative splicing and nonsense-mediated mRNA decay pathways. LncRNA may interact with splicing factors; form duplexes with pre-mRNA or perform chromatin remodeling, modulating transcription, and splicing. Recently, the lncRNAs NEAT1 and MALAT1 were shown to be colocalized with nuclear speckles containing splicing factor SC35 [267]. Additionally, lncRNA may interact with the target mRNA (for example, half-STAU1-binding site RNAs (1/2-sbsRNAs)) and create a

double-stranded transactivation motif that binds to the STAU1 double-stranded (ds) RNA-binding protein and induce mRNA degradation [268].

In this work, we found that murine *Morrbid* lncRNA is involved in the regulation of the AS-NMD pathway for the proto-oncogene *NRAS*. Previously, lncRNA *mMorrbid* was identified as a PRC2-dependent inhibitor of the proapoptotic gene *Bim* in myeloid cells [51]. Here, we demonstrated that the expression of *mMorrbid* lncRNA in normal hepatocytes is higher in comparison with cancer cells. This data contradicts with previous reports on the upregulation of human *Morrbid* lncRNA in HCC and can be the result of differed lncRNA functions between species. As *mMorrbid* is preferably localized in the nucleus, we used modified antisense oligonucleotides (ASO) to downregulate *mMorrbid* lncRNA in AML12 cells [242]. The knockdown of *mMorrbid* lncRNA by ASO led to a decrease of hepatocyte proliferation and migration rates (Figure 35G, I). This data correlates with the already published function of murine *Morrbid* lncRNA in neutrophils, eosinophils, and classical monocytes, in which *mMorrbid* is crucial for the physiologic control of the lifespan [51]. However, lncRNA functions can vary in different cell types, so we performed RNA-seq analysis of *mMorrbid*-depleted hepatocytes and found the dysregulation of many transcripts involved in several signaling pathways. Further analysis of the splicing variants in the RNA-seq data of hepatocytes with *mMorrbid* depletion revealed >80 alternative splicing events. We were interested in the accumulated *NRAS* transcript with a PTC cassette exon (“poison exon”), which could be a target of nonsense-mediated mRNA decay based on the Ensembl database annotation (Figure 37B). To confirm this feature, we measured the relative expression of the *NRAS* PTC transcript in *mMorrbid* KD cells and found its significant upregulation in comparison to the total *NRAS* (Figure 38D). During the next step, we demonstrated that this *NRAS* PTC transcript with a

cassette exon is a target of NMD, as this transcript is undetectable in the cytosol (Figure 39A). The NRAS PTC transcript shows almost three times decreased half-life in comparison to the total NRAS in normal hepatocytes, but it is rather stable—2.6 h in comparison with the published data of the NMD targeted transcripts [269] (Figure 39B).

To further verify NMD nature of NRAS-PTC cellular degradation, we silenced the central NMD factor UPF1 to simultaneously disrupt all circuitries leading to NMD activation in cells. Proteins UPF1, UPF2, UPF3 are well-studied key factors of the NMD pathway. While all three proteins were found essential for NMD signaling in yeast cells, there have been several reports indicating that UPF2 and UPF3 can be dispensable in vertebrates under specific conditions and/or cell types. Several branches are leading to NMD activation and there is evidence that degradation of some PTC-containing mRNAs does not require UPF2 and UPF3 while all known NMD pathway converge at UPF1, making it a master regulator of NMD in vertebrates [266]. For example, in HeLa cells mRNP composed of Y14, MAGOH, and eIF4A3 was demonstrated to activate UPF1 phosphorylation independently of UPF2 and only a small subset of known NMD targets were upregulated in UPF3-depleted HeLa cells [270] UPF1 is essential for all NMD steps from PTC recognition to mRNA degradation.

Interestingly, in hepatocytes with depleted UPF1, the NRAS PTC transcript is upregulated (Figure 39C), which additionally proves that this NRAS transcript is a target for NMD. Previously, Barbie et al. showed that alternative splicing is an important mechanism for RAS regulation [220]. The RAS protein family member HRAS has a cassette exon containing PTC, which, upon inclusion, leads to a quick NMD degradation in the cytosol. The incorporation of this HRAS NMD exon was favored in response to

genotoxic stress in a p53-dependent manner, suggesting a stress response mechanism for the regulation of HRAS cellular levels. In the case of NRAS, it was shown previously that different isoforms possess different oncogenic activity. Human melanoma cells produce five NRAS isoforms, expressed at different levels, resulting in a varied activation of downstream pathways, levels of phosphorylated p-Erk and p-Akt, and resistance to anticancer drugs [271,272].

To identify how mMorrbid lncRNA may be involved in the regulation of NRAS splicing, we performed protein pull-down assays and found that mMorrbid interacts with the SFPQ-NONO heterodimer (Figure 40B, C), which is a splicing-related protein complex [273]. SFPQ and NONO bind as a complex to the conserved stem-loop within snRNA U5 and under splicing conditions and assemble U5 snRNA together with other U5-specific factors like the U5/U5/U6 tri-SNP [274]. Despite direct binding, DBHS proteins do not represent the essential components of the splicing machinery; rather, they function as alternative splicing regulators [223]. SFPQ and NONO regulate both the inclusion (for example, the N30 exon of NMHC mRNA and exon7 of the SMN2 gene) and exclusion (for example, exon4 of PPT, Tau, and CD45) of multiple cassette exons [224,225,275–277]. Thus, we propose that SFPQ and NONO interact with lncRNA mMorrbid to mediate the exclusion of the cassette exon, with a premature stop codon in NRAS pre-mRNA. This result correlates with published data showing that SFPQ binds to the stem-loop structure of the 5' splice site of microtubule-associated Tau pre-mRNA, thus promoting the exclusion of one of the microtubule-binding repeat regions [225]. Another example includes SFPQ binding to the ESS1 (exonic splicing silencer) regulatory element in transmembrane tyrosine phosphatase CD45 pre-mRNA. This interaction results in the skipping of three exons and the production of a catalytically inactive CD45 protein [224]. Several known lncRNAs bind to SFPQ and NONO to perform regulatory functions. The most well-known

lncRNA partner of DHBS proteins is NEAT1. The SFPQ-NONO-NEAT1 complex forms a scaffold core of the subnuclear structures called paraspeckles [93]. Another example is lncRNA VL30, which regulates the transcription of several genes (GAGE6 and Rab23) via an RNA protein decoy-like mechanism [278,279].

To prove the direct interaction of the SFPQ-NONO complex with NRAS pre-mRNA, we performed a protein pull-down assay using the coding parts of NRAS transcripts and found that the SFPQ-NONO complex interacts only with the main NRAS mRNA transcript but not with the PTC isoform (Figure 40B). Interestingly, SFPQ binds to the consensus sequence UGGAGAGGAAC on pre-mRNA to promote splicing, with the middle nucleotides (AGAGGA) representing patterns that interact with SFPQ more frequently [274]. We found an AGAGGA sequence within exon 2 of NRAS mRNA (Figure 40D). We propose that an SFPQ binding site may be located at the junction of exon1 and exon2 of the total NRAS mRNA, while the NRAS NMD transcript may form secondary structures that prevent SFPQ binding.

We propose the mechanism of the regulation of NRAS splicing variants by alternative splicing coupled with the nonsense-mediated mRNA decay pathway under control of the murine Morrbid lncRNA in hepatocytes (Figure 41). Under normal conditions, mMorrbid lncRNA can interact with the SFPQ-NONO heterodimer and participate in interactions of the SFPQ-NONO complex with NRAS pre-mRNA to improve the maturation of NRAS mRNA. In the case of mMorrbid depletion, the efficacy of the SFPQ-NONO heterodimer binding to NRAS pre-mRNA decreases, resulting in alternative splicing of NRAS mRNA. These events promote the inclusion of the “poison exon” with PTC, leading to the degradation of such a transcript by the NMD pathway.

Thus, Morrbrid lncRNA contributes to the correct splicing of the main NRAS mRNA isoform.

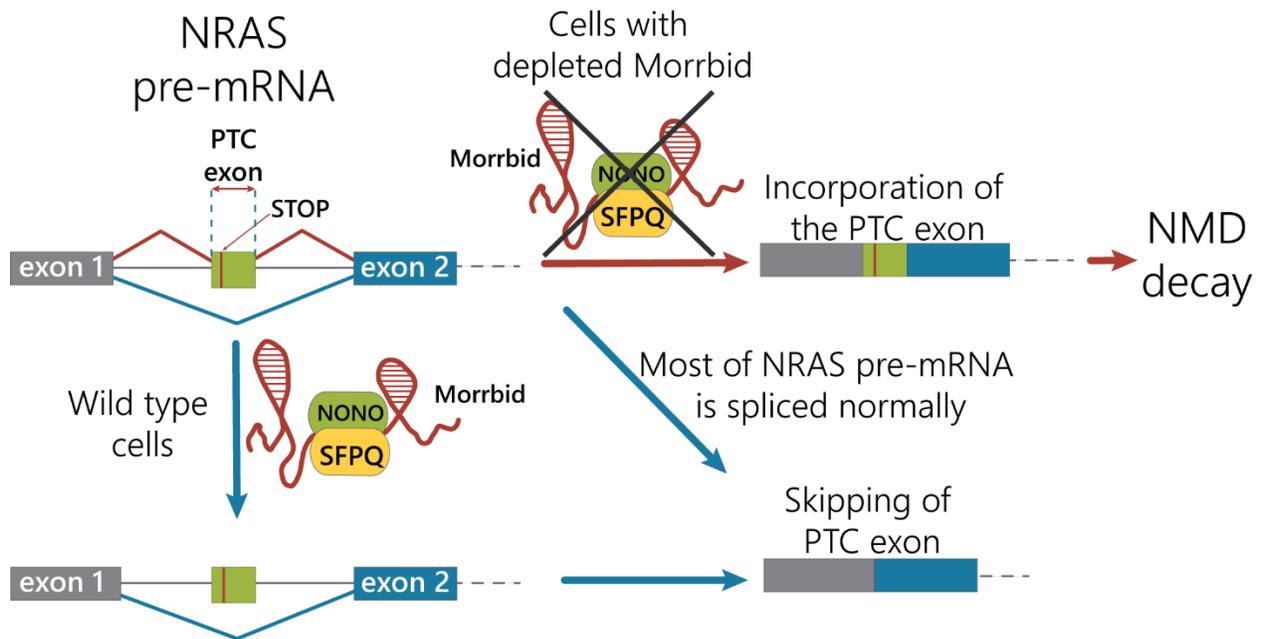


Figure 42. Proposed mechanism of the Morrbrid lncRNA contribution in the regulation of NRAS mRNA alternative splicing.

6.4 A question of whether Morrbrid gene orthologs in mouse and human are functional homologs remains open

Murine and human Morrbrid lncRNAs were first named homologs in 2016 in a study devoted to Bim regulation in immune cells [51]. There is good evidence that these genes are orthologs: similar genomic environment (Figure 13), position in the genome, partial sequence conservation (Figure 14). A comparison of exon sequences shows that 34% of mMorrbrid sequence aligns to hMorrbrid with ~68% identity, which is close to coding sequence conservation of known conserved lncRNAs [42]. However, essential loss-of-function and gain-of-function experiments validating that these two genes are functional homologs have not been done.

In case of protein-coding genes it used to be generally be assumed that orthologous genes have the same functions [280]. However, research demonstrated that phenotypes associated with orthologous genes is oftentimes different between species [281]. One example is SPTLC2 gene, which disfunction causes non-identical physiological problems in human and in mouse [282]. Thus, even protein-coding orthologs can functionally diverge and an even more significant deviation can be expected from much faster evolving non-coding genes. A recent analysis of sequences and LOF transcriptomes of ortholog genes exhibiting phenotypic differences demonstrated correlation with changes in noncoding regulatory elements and tissue-specific expression profiles rather than changes in protein-coding sequences [282]. Cases, when functional repertoires of lncRNA orthologs with partial sequence conservation deviated throughout evolution, have been reported. For example, in human cells, HOTAIR lncRNA acquired novel functions in the regulation of HoxD locus, absent in mice [89]. Orthologs MEG3/Glt2 have been found to participate in a variety of non-overlapping cellular processes [166, 167]. Therefore, careful investigation of each individual case is required for a full understanding of lncRNA functional interspecies homology.

Multiple studies reported that hMorrbid expression is highly upregulated and is important for the proliferation of human cancers [9,10,12,13]. In our lab, we also confirmed the upregulation of human Morrbid in HCC tissues (not published), a more detailed analysis demonstrated that hMorrbid is related to HCC cells migration and cell cycle (Figure 25). At the same time analysis of murine Morrbid in HCC cells and HCC mouse model showed downregulation of mMorrbid in cancer cells in comparison to normal cells (Figure 34 C, D). mMorrbid downregulation affected growth and migration rates of normal liver cells but not cancerous liver cells (Figure 35 F-I). Interestingly, in preleukemic and leukemic murine cells Morrbid expression is elevated and Morrbid

inhibition results in improved viability of mice with AML (acute myelogenous leukemia) [190] as well as JMML (juvenile myelomonocytic leukemia) [191]. Therefore, murine Morrbid role in cancers depends on the cellular background.

In both human and murine hepatocytes, we observed upregulation of Bim protein level in response to Morrbid downregulation (Figures 32, 36). Thus, Morrbid lncRNA might have a universal function of a negative modulator of Bim expression in hepatocytes. Interestingly, human and murine Morrbid depleted hepatocytes did not demonstrate any mitochondrial apoptosis activity in response to the Morrbid-mediated increase of Bim, unlike myeloid cell lineages [51]. Kotzin and colleagues reported that Morrbid promotes cell survival via Bim inhibition in murine myeloid cells [51] and leukemia cells [190,191]. On the other hand, inhibition of mMorrbid in CD8 T cells did not alter Bim levels and cell survival [50]. It suggests that the degree of Morrbid involvement in the regulation of Bim as well as cell survival depends on the tissue and conditions.

In mice, Morrbid appears to be involved in the regulation of alternative splicing of some genes. In particular, it is involved in the production of NRAS oncogene PTC-containing splice form. In human, no evidence for AS differential changes was found and the human NRAS gene does not produce PTC forms.

Overall, there has been little comparison between human and murine Morrbid functions in the same cellular background. Here we attempted to compare human and murine Morrbid LOF phenotypes in hepatocellular carcinoma cells. Our comparison suggests that Morrbid has a conserved function as negative regulation of Bim protein, while also performing organism-specific roles. However, further exploration of this research question is required.

Conclusions

In this study, we focused on the role of human and murine Morrbid lncRNA in cancer hepatocyte cell lines *in vitro*. Generation of CRISPR/Cas9 knockout (KO) of human Morrbid and its paralogous gene CYTOR allowed us to investigate the functions of these lncRNAs in human hepatocellular carcinoma cells. In the case of the murine cells, we downregulated mMorrbid with antisense oligonucleotides. Phenotype characterization of human and murine Morrbid-depleted cell lines demonstrated that Morrbid is not essential for the proliferation of liver cancer cells - we observed only a slight influence on migration and cell cycle in the case of human cells. We measured levels of several pro-apoptotic and pro-survival proteins and confirmed the upregulation of pro-apoptotic Bim protein previously described for immune cells. However, this upregulation was not sufficient for the activation of apoptosis. We overexpressed the unique transcript of human Morrbid (M-217) with the evolutionarily conserved sequence (exonCh) between human and mouse in the human knockout cell line and observed impaired cell proliferation and activation of the mitochondrial apoptosis pathway. Interestingly, the upregulation of Bim was reversed in cells overexpressing M-217, partially due to recovered expression of conserved fragment exonCh. Inhibition of pro-survival proteins in cells overexpressing exonCh with small molecule inhibitors ABT-737 and S-63845 revealed increased sensitivity to MCL1 inhibition. Thus, we uncovered that the contribution of Morrbid lncRNA in apoptosis regulation in human and murine hepatocellular carcinoma cells is related mainly to fine-tuning rather than to driving this process.

We investigated the function of Morrbid lncRNA in murine hepatocytes and found its importance for proliferation and migration of normal but not cancerous liver cells. Transcriptome

analysis of the cells with mMorrbid depletion by ASO demonstrated changes in the alternative splicing. We found the upregulation of NRAS pre-mRNA and NRAS isoform containing premature translation stop codon and proved its degradation via the nonsense-mediated mRNA decay pathway. We confirmed the direct interaction of Morrbrid lncRNA with splicing complex SFPQ-NONO and proposed the mechanism for the regulation of NRAS splicing in murine hepatocytes by alternative splicing coupled with the NMD pathway with the input of Morrbrid lncRNA.

Our findings clarified moonlight functions of Morrbrid lncRNAs in human and murine hepatocytes in vitro.

Bibliography

1. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. *Nature*. 2012 489(7414), 101–8.
2. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008 5(7), 621–8.
3. Ganqiang L, Mattick J S, Taft R J. A meta-analysis of the genomic and transcriptomic composition of complex life. *Cell Cycle*. 2013 12(13), 2061-72.
4. He RZ, Luo DX, Mo YY. Emerging roles of lncRNAs in the post-transcriptional regulation in cancer. *Genes and Diseases*. 2019 6(1), 6–15.
5. Smekalova EM, Kotelevtsev YV, Leboeuf D, Shcherbinina EY, Fefilova AS, Zatsepin TS, et al. lncRNA in the liver: Prospects for fundamental research and therapy by RNA interference. *Biochimie*. 2016 131, 159–72.
6. Xie H, Ma B, Gao Q, Zhan H, Liu Y, Chen Z, et al. Long non-coding RNA CRNDE in cancer prognosis: Review and meta-analysis. *Clinica Chimica Acta*. 2018 485, 262–71.
7. Uszczyńska-Ratajczak B, Lagarde J, Frankish A, Guigó R, Johnson R. Towards a complete map of the human long non-coding RNA transcriptome. *Nat Rev Genet*. 2018 19(9), 535-548.
8. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res*. 2012 22(9), 1775–89.
9. Zhao J, Liu Y, Zhang W, Zhou Z, Wu J, Cui P, et al. Long non-coding RNA Linc00152 is involved in cell cycle arrest, apoptosis, epithelial to mesenchymal transition, cell migration and invasion in gastric cancer. *Cell Cycle*. 2015 14(19), 3112–23.

10. Bu J-Y, Lv W-Z, Liao Y-F, Xiao X-Y, Lv B-J. Long non-coding RNA LINC00978 promotes cell proliferation and tumorigenesis via regulating microRNA-497/NTRK3 axis in gastric cancer. *Int J Biol Macromol.* 2019 123, 1106–14.
11. Ke D, Li H, Zhang Y, An Y, Fu H, Fang X, et al. The combination of circulating long noncoding RNAs AK001058, INHBA-AS1, MIR4435-2HG, and CEBPA-AS1 fragments in plasma serve as diagnostic markers for gastric cancer. *Oncotarget.* 2017 8(13), 21516–25.
12. Fu M, Huang Z, Zang X, Pan L, Liang W, Chen J, et al. Long noncoding RNA LINC00978 promotes cancer growth and acts as a diagnostic biomarker in gastric cancer. *Cell Prolif.* 2018 51(1), e12425.
- (13. Zhou J, Zhi X, Wang L, Wang W, Li Z, Tang J, et al. Linc00152 promotes proliferation in gastric cancer through the EGFR-dependent pathway. *J Exp Clin Cancer Res.* 2015 34, 135.
14. Li J, Wang X, Tang J, Jiang R, Zhang W, Ji J, et al. HULC and Linc00152 act as novel biomarkers in predicting diagnosis of hepatocellular carcinoma. *Cell Physiol Biochem.* 2015 37(2), 687–96.
15. Ma P, Wang H, Sun J, Liu H, Zheng C, Zhou X, et al. LINC00152 promotes cell cycle progression in hepatocellular carcinoma via miR-193a/b-3p/CCND1 axis. *Cell Cycle.* 2018 17(8), 974–84.
16. Ji J, Tang J, Deng L, Xie Y, Jiang R, Li G, et al. LINC00152 promotes proliferation in hepatocellular carcinoma by targeting EpCAM via the mTOR signaling pathway. *Oncotarget.* 2015 6(40), 42813–24.
17. Kong Q, Liang C, Jin Y, Pan Y, Tong D, Kong Q, et al. The lncRNA MIR4435-2HG is

- upregulated in hepatocellular carcinoma and promotes cancer cell proliferation by upregulating miRNA-487a. *Cell Mol Biol Lett*. 2019 24(1), 26.
18. Deng X, Zhao X fang, Liang X qiu, Chen R, Pan Y feng, Liang J. Linc00152 promotes cancer progression in hepatitis B virus-associated hepatocellular carcinoma. *Biomed Pharmacother*. 2017 90, 100–8.
 19. Zhao X, Liu Y, Yu S. Long noncoding RNA AWPPH promotes hepatocellular carcinoma progression through YBX1 and serves as a prognostic biomarker. *Biochim Biophys Acta - Mol Basis Dis*. 2017 1863(7), 1805–16.
 20. Tang L, Wang T, Zhang Y, Zhang J, Zhao H, Wang H, et al. Long Non-Coding RNA AWPPH Promotes Postoperative Distant Recurrence in Resected Non-Small Cell Lung Cancer by Upregulating Transforming Growth Factor beta 1 (TGF- β 1). *Med Sci Monit*. 2019 25(1), 2535–41.
 21. Huo Y, Li A, Wang Z. LncRNA AWPPH participates in the metastasis of non-small cell lung cancer by upregulating TGF- β 1 expression. *Oncol Lett*. 2019 18(4), 4246–52.
 22. Liu Z, Liu A, Nan A, Cheng Y, Yang T, Dai X, et al. The linc00152 Controls Cell Cycle Progression by Regulating CCND1 in 16HBE Cells Malignantly Transformed by Cigarette Smoke Extract. *Toxicol Sci*. 2019 167(2), 496–508.
 23. Chen Q-N, Chen X, Chen Z-Y, Nie F-Q, Wei C-C, Ma H-W, et al. Long intergenic non-coding RNA 00152 promotes lung adenocarcinoma proliferation via interacting with EZH2 and repressing IL24 expression. *Mol Cancer*. 2017 16(1), 17.
 24. Zhang J, Li W. Long noncoding RNA CYTOR sponges miR-195 to modulate proliferation, migration, invasion and radiosensitivity in nonsmall cell lung cancer cells. *Biosci Rep*. 2018, 38(6).

25. Zhang Y, Xiang C, Wang Y, Duan Y, Liu C, Jin Y, et al. lncRNA LINC00152 knockdown had effects to suppress biological activity of lung cancer via EGFR/PI3K/AKT pathway. *Biomed Pharmacother.* 2017 94, 644–51.
26. Wang X, Yu H, Sun W, Kong J, Zhang L, Tang J, et al. The long non-coding RNA CYTOR drives colorectal cancer progression by interacting with NCL and Sam68. *Mol Cancer.* 2018 17(1).
27. Cai Q, Wang Z, Wang S, Weng M, Zhou D, Li C, et al. Long non-coding RNA LINC00152 promotes gallbladder cancer metastasis and epithelial-mesenchymal transition by regulating HIF-1 α via miR-138. *Open Biol.* 2017 7(1).
28. Yu M, Xue Y, Zheng J, Liu X, Yu H, Liu L, et al. Linc00152 promotes malignant progression of glioma stem cells by regulating miR-103a-3p/FEZF1/CDC25A pathway. *Mol Cancer.* 2017 16(1).
29. Gong J, Xu X, Zhang X, Zhou Y. LncRNA MIR4435-2HG is a potential early diagnostic marker for ovarian carcinoma. *Acta Biochim Biophys Sin.* 2019 51(9), 953-959.
30. Yu G, Wang W, Deng J, Dong S. LncRNA AWPPH promotes the proliferation, migration and invasion of ovarian carcinoma cells via activation of the Wnt/ β -catenin signaling pathway. *Mol Med Rep.* 2019 19(5), 3615–21.
31. Zhang H, Zhu M, Du Y, Zhang H, Zhang Q, Liu Q, et al. A panel of 12-lncRNA signature predicts survival of pancreatic adenocarcinoma. *J Cancer.* 2019 10(6), 1550–9.
32. Wu Y, Tan C, Weng W-W, Deng Y, Zhang Q-Y, Yang X-Q, et al. Long non-coding RNA Linc00152 is a positive prognostic factor for and demonstrates malignant biological behavior in clear cell renal cell carcinoma. *Am J Cancer Res.* 2016 6(2), 285–99.
33. Wang Y, Liu J, Bai H, Dang Y, Lv P, Wu S. Long intergenic non-coding RNA 00152

- promotes renal cell carcinoma progression by epigenetically suppressing P16 and negatively regulates miR-205. *Am J Cancer Res.* 2017 7(2), 312–22.
34. Shen X, Zhong J, Yu P, Zhao Q, Huang T. YY1-regulated LINC00152 promotes triple negative breast cancer progression by affecting on stability of PTEN protein. *Biochem Biophys Res Commun.* 2019 509(2), 448–54.
 35. Deng L, Chi Y, Liu L, Huang N, Wang L, Wu J. LINC00978 predicts poor prognosis in breast cancer patients. *Sci Rep.* 2016 6(1), 37936.
 36. Wu J, Shuang Z, Zhao J, Tang H, Liu P, Zhang L, et al. Linc00152 promotes tumorigenesis by regulating DNMTs in triple-negative breast cancer. *Biomed Pharmacother.* 2018 97, 1275–81.
 37. Wang K, Li X, Song C, Li M. LncRNA AWPPH promotes the growth of triple-negative breast cancer by up-regulating frizzled homolog 7 (FZD7). *Biosci Rep.* 2018 38 (6), BSR20181223.
 38. Chen W, Huang M, Sun D, Kong R, Xu T, Xia R, et al. Long intergenic non-coding RNA 00152 promotes tumor cell cycle progression by binding to EZH2 and repressing p15 and p21 in gastric cancer. *Oncotarget.* 2016 7(9), 9773–87.
 39. Yue B, Liu C, Sun H, Liu M, Song C, Cui R, et al. A positive feed-forward loop between lncRNA-CYTOR and Wnt/ β -catenin signaling promotes metastasis of colon cancer. *Mol Ther.* 2018 26(5), 1287–98.
 40. Eißmann M, Gutschner T, Hämmerle M, Günther S, Caudron-Herger M, Groß M, et al. Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. *RNA Biol.* 2012 9(8), 1076–87.
 41. Dimitrova N, Zamudio JR, Jong RM, Soukup D, Resnick R, Sarma K, et al. LincRNA-

- p21 activates p21 in cis to promote polycomb target gene expression and to enforce the G1/S checkpoint. *Mol Cell*. 2014 54(5), 777–90.
42. Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I. Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. *Cell Rep*. 2015 11(7), 1110–22.
 43. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. 2009 458(7235), 223–7.
 44. Ponjavic J, Ponting CP, Lunter G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res*. 2007 17(5), 556–65.
 45. Ulitsky I. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat Rev Genet*. 2016 17(10), 601–14.
 46. Marín-Béjar O, Mas AM, González J, Martínez D, Athie A, Morales X, Galduroz, M, Raimondi I, Grossi E, Guo S et al. The human lncRNA LINC-PINT inhibits tumor cell invasion through a highly conserved sequence element. *Genome Biol*. 2017 18, 202.
 47. Guttman M, Rinn J. Modular regulatory principles of large non-coding RNAs. *Nature*. 2012 482(7385), 339-46.
 48. Johnson R, Guigó R. The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. *RNA*. 2014 20(7), 959-76.
 49. Quinn JJ, Ilik IA, Qu K, Georgiev P, Chu C, Akhtar A, et al. Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. *Nat Biotechnol*. 2014 32(9), 933–40.
 50. Kotzin JJ, Iseka F, Wright J, Basavappa MG, Clark ML, Ali M-A, et al. The long

- noncoding RNA Morrbid regulates CD8 T cells in response to viral infection. *Proc Natl Acad Sci U S A*. 2019 116(24), 11916–25.
51. Kotzin JJ, Spencer SP, McCright SJ, Kumar DBU, Collet MA, Mowel WK, et al. The long non-coding RNA Morrbid regulates Bim and short-lived myeloid cell lifespan. *Nature*. 2016 537(7619), 239–43.
 52. Cai Z, Kotzin JJ, Ramdas B, Chen S, Nelanuthala S, Palam LR, et al. Inhibition of Inflammatory Signaling in Tet2 Mutant Preleukemic Cells Mitigates Stress-Induced Abnormalities and Clonal Hematopoiesis. *Cell Stem Cell*. 2018 23(6), 833-849.
 53. Brunner AL, Beck AH, Edris B, Sweeney RT, Zhu SX, Li R, et al. Transcriptional profiling of long non-coding RNAs and novel transcribed regions across a diverse panel of archived human cancers. *Genome Biol*. 2012 13(8), R75.
 54. Bartolomei MS, Zemel S, Tilghman SM. Parental imprinting of the mouse H19 gene. *Nature*. 1991 351(6322), 153–5.
 55. Brannan CI, Dees EC, Ingram RS, Tilghman SM. The product of the H19 gene may function as an RNA. *Mol Cell Biol*. 1990 10(1), 28–36.
 56. Lander, E., Linton, L., Birren, B. et al. Initial sequencing and analysis of the human genome. *Nature*. 2001 409(6822), 860–921.
 57. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The Sequence of the Human Genome. *Science*. 2001 291(5507), 1304–51.
 58. Dunham, I., Kundaje, A., Aldred, S. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012 489(7414), 57–74.
 59. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The Transcriptional Landscape of the Mammalian Genome. *Science*. 2005 309(5740), 1559–

- 63.
60. Nóbrega MA, Zhu Y, Plajzer-Frick I, Afzal V, Rubin EM. Megabase deletions of gene deserts result in viable mice. *Nature*. 2004 431(7011), 988–93.
61. Fang Y, Fullwood MJ. Roles, Functions, and Mechanisms of Long Non-coding RNAs in Cancer. *Genomics Proteomics Bioinformatics*. 2016 14(1), 42–54.
62. Jarroux J, Morillon A, Pinskaya M. History, Discovery, and Classification of lncRNAs. *Adv Exp Med Biol*. 2017 1008, 1–46.
63. Dinger ME, Pang KC, Mercer TR, Mattick JS. Differentiating protein-coding and noncoding RNA: Challenges and ambiguities. *PLoS Comput Biol*. 2008 4(11), e1000176.
64. Blokhin I, Khorkova O, Hsiao J, Wahlestedt C. Developments in lncRNA drug discovery: where are we heading? *Expert Opin Drug Discov*. 2018 13(9), 837–49.
65. Atianand MK, Caffrey DR, Fitzgerald KA. Immunobiology of Long Noncoding RNAs. *Annu Rev Immunol*. 2017 35, 177–98.
66. Ransohoff JD, Wei Y, Khavari PA. The functions and unique features of long intergenic non-coding RNA. *Nat Rev Mol Cell Biol*. 2018 19 (3), 143–57.
67. Zucchelli S, Fasolo F, Russo R, Cimatti L, Patrucco L, Takahashi H, et al. SINEUPs are modular antisense long non-coding RNAs that increase synthesis of target proteins in cells. *Front Cell Neurosci*. 2015 9, 174.
68. Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, et al. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature*. 2012 491(7424), 454–7.
69. Wu H, Yang L, Chen L-L. The diversity of long noncoding RNAs and their generation. *Trends Genet*. 2017 33(8), 540–52.

70. Xing YH, Yao RW, Zhang Y, Guo CJ, Jiang S, Xu G, et al. SLERT regulates DDX21 rings associated with Pol I transcription. *Cell*. 2017 169(4), 664-678.e16.
71. Zhang Y, Zhang XO, Chen T, Xiang JF, Yin QF, Xing YH, et al. Circular intronic long noncoding RNAs. *Mol Cell*. 2013 51(6), 792–806.
72. Sarkar D, Oghabian A, Bodiyaabadu PK, Joseph WR, Leung EY, Finlay GJ, et al. Multiple isoforms of ANRIL in melanoma cells: structural complexity suggests variations in processing. *Int J Mol Sci*. 2017 18(7), 1378.
73. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, et al. Natural RNA circles function as efficient microRNA sponges. *Nature*. 2013 495(7441), 384–8.
74. Clark MB, Johnston RL, Inostroza-Ponta M, Fox AH, Fortini E, Moscato P, et al. Genome-wide analysis of long noncoding RNA stability. *Genome Res*. 2012 22(5), 885–98.
75. Ayupe AC, Tahira AC, Camargo L, Beckedorff FC, Verjovski-Almeida S, Reis EM. Global analysis of biogenesis, stability and sub-cellular localization of lncRNAs mapping to intragenic regions of the human genome. *RNA Biol*. 2015 12(8), 877–92.
76. Giannakakis A, Zhang J, Jenjaroenpun P, Nama S, Zainolabidin N, Aau MY, et al. Contrasting expression patterns of coding and noncoding parts of the human genome upon oxidative stress. *Sci Rep*. 2015 5(1), 1–16.
77. Chugunova A, Loseva E, Mazin P, Mitina A, Navalayeu T, Bilan D, et al. LINC00116 codes for a mitochondrial peptide linking respiration and lipid metabolism. *Proc Natl Acad Sci U S A*. 2019 116(11), 4940–5.
78. Nelson BR, Makarewich CA, Anderson DM, Winders BR, Troupes CD, Wu F, et al. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA

- activity in muscle. *Science*. 2016 351(6270), 271–5.
79. Housman G, Ulitsky I. Methods for distinguishing between protein-coding and long noncoding RNAs and the elusive biological purpose of translation of long noncoding RNAs. *Biochim Biophys Acta*. 2016, 1859(1), 31–40.
80. Li W, Notani D, Rosenfeld MG. Enhancers as non-coding RNA transcription units: Recent insights and future perspectives. *Nat Rev Genet*. 2016 17(4), 207–23.
81. Anderson KM, Anderson DM, McAnally JR, Shelton JM, Bassel-Duby R, Olson EN. Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. *Nature*. 2016 539(7629), 433–6.
82. Latos PA, Pauler FM, Koerner M V, Şenergin HB, Hudson QJ, Stocsits RR, et al. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science*. 2012 338(6113), 1469–72.
83. Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*. 2016 539(7629), 452–5.
84. Kopp F, Mendell JT. Functional classification and experimental dissection of long noncoding RNAs. *Cell*. 2018 172(3), 393–407.
85. Groff AF, Sanchez-Gomez DB, Soruco MML, Gerhardinger C, Barutcu AR, Li E, et al. In vivo characterization of linc-p21 reveals functional cis-regulatory DNA elements. *Cell Rep*. 2016 16(8), 2178–86.
86. Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, et al. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*. 1991 349(6304), 38–44.

87. Borsani G, Tonlorenzi R, Simmler MC, Dandolo L, Arnaud D, Capra V, et al. Characterization of a murine gene expressed from the inactive X chromosome. *Nature*. 1991 351(6324), 325–9.
88. Cerase A, Pintacuda G, Tattermusch A, Avner P. Xist localization and function: new insights from multiple levels. *Genome Biol*. 2015. 16(1), 166.
89. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*. 2007 129(7), 1311–23.
90. Atianand MK, Hu W, Satpathy AT, Shen Y, Ricci EP, Alvarez-Dominguez JR, et al. A long noncoding RNA lincRNA-EP5 acts as a transcriptional brake to restrain inflammation. *Cell*. 2016 165(7), 1672–85.
91. Lee S, Kopp F, Chang TC, Sataluri A, Chen B, Sivakumar S, et al. Noncoding RNA NORAD regulates genomic stability by sequestering PUMILIO proteins. *Cell*. 2016 164(1–2), 69–80.
92. Tichon A, Gil N, Lubelsky Y, Solomon TH, Lemze D, Itzkovitz S, et al. A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nat Commun*. 2016 7(1), 1–10.
93. Fox AH, Nakagawa S, Hirose T, Bond CS. Paraspeckles: where long noncoding RNA meets phase separation. *Trends in Biochem Sci*. 2018. 43(2), 124–35.
94. Sasaki YTF, Ideue T, Sano M, Mituyama T, Hirose T. MENε/β noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc Natl Acad Sci U S A*. 2009 106(8), 2525–30.
95. Hacısuleyman E, Goff LA, Trapnell C, Williams A, Henao-Mejia J, Sun L, et al.

- Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat Struct Mol Biol.* 2014 21(2), 198–206.
96. Miyagawa R, Tano K, Mizuno RIE, Nakamura YO, Ijiri K, Rakwal R, et al. Identification of cis- and trans-acting factors involved in the localization of MALAT-1 noncoding RNA to nuclear speckles. *RNA.* 2012 18(4), 738–51.
 97. Bernard D, Prasanth K V, Tripathi V, Colasse S, Nakamura T, Xuan Z, et al. A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J.* 2010 29(18), 3082–93.
 98. Legnini I, Morlando M, Mangiavacchi A, Fatica A, Bozzoni I. A feedforward regulatory loop between HuR and the long noncoding RNA linc-MD1 controls early phases of myogenesis. *Mol Cell.* 2014 53(3), 506–14.
 99. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature.* 2013 495(7441), 333–8.
 100. Piwecka M, Glažar P, Hernandez-Miranda LR, Memczak S, Wolf SA, Rybak-Wolf A, et al. Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. *Science.* 2017 357(6357), eaam8526.
 101. Villamizar O, Chambers CB, Riberdy JM, Persons DA, Wilber A. Long noncoding RNA Saf and splicing factor 45 increase soluble Fas and resistance to apoptosis. *Oncotarget.* 2016 7(12), 13810–26.
 102. Ji Q, Zhang L, Liu X, Zhou L, Wang W, Han Z, et al. Long non-coding RNA MALAT1 promotes tumour growth and metastasis in colorectal cancer through binding to SFPQ and releasing oncogene PTBP2 from SFPQ/PTBP2 complex. *Br J Cancer.* 2014 111(4), 736–48.

103. Cooper D, Carter G, Li P, Patel R, Watson J, Patel N. Long non-coding RNA NEAT1 associates with SRp40 to temporally regulate PPAR γ 2 splicing during adipogenesis in 3T3-L1 cells. *Genes (Basel)*. 2014 5(4), 1050–63.
104. Kong J, Sun W, Li C, Wan L, Wang S, Wu Y, et al. Long non-coding RNA LINC01133 inhibits epithelial–mesenchymal transition and metastasis in colorectal cancer by interacting with SRSF6. *Cancer Lett*. 2016 380(2), 476–84.
105. Barry G, Briggs JA, Vanichkina DP, Poth EM, Beveridge NJ, Ratnu VS, et al. The long non-coding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing. *Mol Psychiatry*. 2014 19(4), 486–94.
106. Van Dijk M, Visser A, Buabeng KML, Poutsma A, Van Der Schors RC, Oudejans CBM. Mutations within the LINC-HELLP non-coding RNA differentially bind ribosomal and RNA splicing complexes and negatively affect trophoblast differentiation. *Hum Mol Genet*. 2015 24(19), 5475-85.
107. Zhang X, Hamblin MH, Yin K-J. The long noncoding RNA Malat1: its physiological and pathophysiological functions. *RNA Biol*. 2017 14(12), 1705–14.
108. Zhang B, Arun G, Mao YS, Lazar Z, Hung G, Bhattacharjee G, et al. The lncRNA malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep*. 2012 2(1), 111–23.
109. Chen CK, Blanco M, Jackson C, Aznauryan E, Ollikainen N, Surka C, et al. Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science*. 2016 354(6311), 468–72.
110. Huarte M, Marín-Béjar O. Long noncoding RNAs: from identification to functions and

- mechanisms. *Adv Genom Genet.* 2015 5, 257-74.
111. Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell.* 2010 39(6), 925–38.
 112. Romero-Barrios N, Legascue MF, Benhamed M, Ariel F, Crespi M. Splicing regulation by long noncoding RNAs. *Nucleic Acids Res.* 2018 46(5), 2169–84.
 113. Bao Z, Yang Z, Huang Z, Zhou Y, Cui Q, Dong D. LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.* 2019 47(D1), D1034-D1037.
 114. Fernandes J, Acuña S, Aoki J, Floeter-Winter L, Muxel S. Long non-coding RNAs in the regulation of gene expression: physiology and disease. *Noncoding RNA.* 2019 5(1), 17.
 115. Zemmour D, Pratama A, Loughhead SM, Mathis Di, Benoist C. Flicr, a long noncoding RNA, modulates Foxp3 expression and autoimmunity. *Proc Natl Acad Sci U S A.* 2017 114(17), E3472–80.
 116. Ang CE, Ma Q, Wapinski OL, Fan S, Flynn RA, Lee QY, et al. The novel lncRNA lnc-NR2F1 is pro-neurogenic and mutated in human neurodevelopmental disorders. *Elife.* 2019 8.
 117. Schmitt AM, Chang HY. Long noncoding RNAs in cancer pathways. *Cancer Cell.* 2016 29(4), 452–63.
 118. Xiang JF, Yin QF, Chen T, Zhang Y, Zhang XO, Wu Z, et al. Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res.* 2014 24(5), 513–31.
 119. Tseng YY, Moriarity BS, Gong W, Akiyama R, Tiwari A, Kawakami H, et al. PVT1

- dependence in cancer with MYC copy-number increase. *Nature*. 2014 512(1), 82–6.
120. Zhu J, Liu S, Ye F, Shen Y, Tie Y, Zhu J, et al. Long noncoding RNA MEG3 interacts with p53 protein and regulates partial p53 target genes in hepatoma cells. *PLoS One*. 2015 10(10), e0139790.
 121. Léveillé N, Melo CA, Rooijers K, Díaz-Lagares A, Melo SA, Korkmaz G, et al. Genome-wide profiling of p53-regulated enhancer RNAs uncovers a subset of enhancers controlled by a lncRNA. *Nat Commun*. 2015 6(1), 1–12.
 122. Hu X, Feng Y, Zhang D, Zhao SD, Hu Z, Greshock J, et al. A functional genomic approach identifies FAL1 as an oncogenic long noncoding RNA that associates with BMI1 and represses p21 expression in cancer. *Cancer Cell*. 2014 26(3), 344–57.
 123. Flockhart RJ, Webster DE, Qu K, Mascarenhas N, Kovalski J, Kretz M, et al. BRAF^{V600E} remodels the melanocyte transcriptome and induces BANCR to regulate melanoma cell migration. *Genome Res*. 2012 22(6), 1006–14.
 124. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*. 2010 464(7291), 1071–6.
 125. Walsh KM, Codd V, Smirnov I V., Rice T, Decker PA, Hansen HM, et al. Variants near TERT and TERC influencing telomere length are associated with high-grade glioma risk. *Nat Genet*. 2014 46(7), 731–5.
 126. Michalik KM, You X, Manavski Y, Doddaballapur A, Zörnig M, Braun T, et al. Long noncoding RNA MALAT1 regulates endothelial cell function and vessel growth. *Circ Res*. 2014 114(9), 1389–97.
 127. Lu Z, Xiao Z, Liu F, Cui M, Li W, Yang Z, et al. Long non-coding RNA HULC promotes

- tumor angiogenesis in liver cancer by up-regulating sphingosine kinase 1 (SPHK1). *Oncotarget*. 2016 7(1), 241–54.
128. Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet*. 2011 43(7), 621–9.
 129. Puvvula PK, Desetty RD, Pineau P, Marchio A, Moon A, Dejean A, et al. Long noncoding RNA PANDA and scaffold-attachment-factor SAFA control senescence entry and exit. *Nat Commun*. 2014 5(1), 1–16.
 130. Hudson WH, Pickard MR, De Vera IMS, Kuiper EG, Mourtada-Maarabouni M, Conn GL, et al. Conserved sequence-specific lincRNA-steroid receptor interactions drive transcriptional repression and direct cell fate. *Nat Commun*. 2014 5(1), 1–13.
 131. Yang Y, Chen L, Gu J, Zhang H, Yuan J, Lian Q, et al. Recurrently deregulated lincRNAs in hepatocellular carcinoma. *Nat Commun*. 2017 8(1), 1–13.
 132. Lanzafame M, Bianco G, Terracciano LM, Ng CKY, Piscuoglio S. The role of long non-coding RNAs in hepatocarcinogenesis. *Int J Mol Sci*. 2018 19(3), 682.
 133. Panzitt K, Tschernatsch MMO, Guelly C, Moustafa T, Stradner M, Strohmaier HM, et al. Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as Noncoding RNA. *Gastroenterology*. 2007 132(1), 330-42.
 134. Quagliata L, Matter MS, Piscuoglio S, Arabi L, Ruiz C, Procino A, et al. Long noncoding RNA HOTTIP/HOXA13 expression is associated with disease progression and predicts outcome in hepatocellular carcinoma patients. *Hepatology*. 2014 59(3), 911–23.
 135. Ge Y, Yan X, Jin Y, Yang X, Yu X, Zhou L, et al. fMiRNA-192 and miRNA-204 directly suppress lincRNA HOTTIP and interrupt GLS1-mediated glutaminolysis in hepatocellular

- carcinoma. *PLoS Genet.* 2015 11(12), e1005726.
136. Fu WM, Zhu X, Wang WM, Lu YF, Hu BG, Wang H, et al. Hotair mediates hepatocarcinogenesis through suppressing miRNA-218 expression and activating P14 and P16 signaling. *J Hepatol.* 2015 63(4), 886–95.
137. Malakar P, Stein I, Saragovi A, Winkler R, Stern-Ginossar N, Berger M, et al. Long noncoding RNA MALAT1 regulates cancer glucose metabolism by enhancing mTOR-mediated translation of TCF7L2. *Cancer Res.* 2019 79(10), 2480–93.
138. Kessler SM, Hosseini K, Hussein UK, Kim KM, List M, Schultheiß CS, et al. Hepatocellular carcinoma and nuclear paraspeckles: induction in chemoresistance and prediction for poor survival. *Cell Physiol Biochem.* 2019 52(4), 787–801.
139. Ramani K, Mavila N, Ko KS, Mato JM, Lu SC. Prohibitin 1 regulates the H19-Igf2 axis and proliferation in hepatocytes. *J Biol Chem.* 2016 291(46), 24148–59.
140. Gailhouste L, Liew LC, Yasukawa K, Hatada I, Tanaka Y, Kato T, et al. MEG3-derived miR-493-5p overcomes the oncogenic feature of IGF2-miR-483 loss of imprinting in hepatic cancer cells. *Cell Death Dis.* 2019 10(8), 553.
141. Cabili M, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 2011 25(18), 1915–27.
142. Jiang C, Li Y, Zhao Z, Lu J, Chen H, Ding N, et al. Identifying and functionally characterizing tissue-specific and ubiquitously expressed human lncRNAs. *Oncotarget.* 2016 7(6), 7120–33.
143. Washietl S, Kellis M, Garber M. Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res.* 2014 24(4), 616–28.

144. Kornienko AE, Dotter CP, Guenzl PM, Gisslinger H, Gisslinger B, Cleary C, et al. Long non-coding RNAs display higher natural expression variation than protein-coding genes in healthy humans. *Genome Biol.* 2016 17, 14.
145. Wilusz JE, Freier SM, Spector DL. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell.* 2008 135(5), 919–32.
146. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet.* 2016 17(1), 47–62.
147. Kong Y, Hsieh C-H, Alonso LC. ANRIL: A lncRNA at the CDKN2A/B locus with roles in cancer and metabolic disease. *Front Endocrinol.* 2018 9, 405.
148. Jarinova O, Stewart AFR, Roberts R, Wells G, Lau P, Naing T, et al. Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Arterioscler Thromb Vasc Biol.* 2009 29(10), 1671–7.
149. Holdt LM, Stahringer A, Sass K, Pichler G, Kulak NA, Wilfert W, et al. Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans. *Nat Commun.* 2016 7(1), 12429.
150. Yap KL, Li S, Muñoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, et al. Molecular Interplay of the Noncoding RNA ANRIL and Methylated Histone H3 Lysine 27 by Polycomb CBX7 in Transcriptional Silencing of INK4a. *Mol Cell.* 2010 38(5), 662–74.
151. Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, et al. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15INK4B tumor suppressor gene. *Oncogene.* 2011 30(16), 1956–62.
152. Holdt LM, Hoffmann S, Sass K, Langenberger D, Scholz M, Krohn K, et al. Alu Elements in ANRIL Non-Coding RNA at Chromosome 9p21 Modulate Atherogenic Cell Functions

- through Trans-Regulation of Gene Networks. *PLoS Genet.* 2013 9(7), e1003588.
153. Stojic L, Niemczyk M, Orjalo A, Ito Y, Ruijter AEM, Uribe-Lewis S, et al. Transcriptional silencing of long noncoding RNA GNG12-AS1 uncouples its transcriptional and product-related functions. *Nat Commun.* 2016 7, 10406.
 154. Niemczyk M, Ito Y, Huddleston J, Git A, Abu-Amero S, Caldas C, et al. Imprinted chromatin around DIRAS3 regulates alternative splicing of GNG12-AS1, a long noncoding RNA. *Am J Hum Genet.* 2013 93(2), 224–35.
 155. Kutter C, Watt S, Stefflova K, Wilson MD, Goncalves A, Ponting CP, et al. Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet.* 2012 8(7), e1002841.
 156. Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell.* 2011 147(7), 1537–50.
 157. Johnsson P, Lipovich L, Grandér D, Morris K V. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim Biophys Acta.* 2014 1840(3), 1063–71.
 158. Managadze D, Rogozin IB, Chernikova D, Shabalina SA, Koonin E V. Negative correlation between expression level and evolutionary rate of long intergenic noncoding RNAs. *Genome Biol Evol.* 2011 3, 1390-1404.
 159. Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol.* 2010 28(5), 503–10.
 160. Nesterova TB, Slobodyanyuk SY, Elisaphenko EA, Shevchenko AI, Johnston C, Pavlova

- ME, et al. Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Res.* 2001 11(5), 833–49.
161. Lin Y, Schmidt BF, Bruchez MP, Mcmanus CJ. Structural analyses of NEAT1 lncRNAs suggest long-range RNA interactions that may contribute to paraspeckle architecture. *Nucleic Acids Res.* 2018 46(7), 31.
162. Cornelis G, Souquere S, Vernochet C, Heidmann T, Pierron G. Functional conservation of the lncRNA NEAT1 in the ancestrally diverged marsupial lineage: evidence for NEAT1 expression and associated paraspeckle assembly during late gestation in the opossum *Monodelphis domestica*. *RNA Biol.* 2016 13(9), 826–36.
163. Ounzain S, Micheletti R, Arnan C, Plaisance I, Cecchi D, Schroen B, et al. CARMEN, a human super enhancer-associated long noncoding RNA controlling cardiac specification, differentiation and homeostasis. *J Mol Cell Cardiol.* 2015 89, 98–112.
164. Schorderet P, Duboule D. Structural and functional differences in the long non-coding RNA hotair in mouse and human. *PLoS Genet.* 2011 7(5), e1002071.
165. Miyoshi N, Wagatsuma H, Wakana S, Shiroishi T, Nomura M, Aisaka K, et al. Identification of an imprinted gene, Meg3/Gtl2 and its human homologue MEG3, first mapped on mouse distal chromosome 12 and human chromosome 14q. *Genes Cells.* 2000 5(3), 211–20.
166. Zhou Y, Zhong Y, Wang Y, Zhang X, Batista DL, Gejman R, et al. Activation of p53 by MEG3 non-coding RNA. *J Biol Chem.* 2007 282(34), 24731–42.
167. Zhou Y, Cheunsuchon P, Nakayama Y, Lawlor MW, Zhong Y, Rice KA, et al. Activation of paternally expressed genes and perinatal death caused by deletion of the Gtl2 gene.

- Development*. 2010 137(16), 2643–52.
168. Nötzold L, Frank L, Gandhi M, Polycarpou-Schwarz M, Groß M, Gunkel M, et al. The long non-coding RNA LINC00152 is essential for cell cycle progression through mitosis in HeLa cells. *Sci Rep*. 2017 7(1), 2265.
 169. Yunis JJ, Prakash O. The origin of man: a chromosomal pictorial legacy. *Science*. 1982 215(4539), 1525–30.
 170. Ciccarelli FD, von Mering C, Suyama M, Harrington ED, Izaurralde E, Bork P. Complex genomic rearrangements lead to novel primate gene function. *Genome Res*. 2005 15(3), 343–51.
 171. Huang Y, Luo H, Li F, Yang Y, Ou G, Ye X, et al. LINC00152 down-regulated miR-193a-3p to enhance MCL1 expression and promote gastric cancer cells proliferation. *Biosci Rep*. 2018 38(3), BSR20171607.
 172. Pang Q, Ge J, Shao Y, Sun W, Song H, Xia T, et al. Increased expression of long intergenic non-coding RNA LINC00152 in gastric cancer and its clinical significance. *Tumour Biol*. 2014 35(6), 5441–7.
 173. Reon BJ, Takao Real Karia B, Kiran M, Dutta A. LINC00152 promotes invasion through a 3'-hairpin structure and associates with prognosis in glioblastoma. *Mol Cancer Res*. 2018 16(10), 1470–82.
 174. Chen W, Huang M, Sun D, Kong R, Xu T, Xia R, et al. Long intergenic non-coding RNA 00152 promotes tumor cell cycle progression by binding to EZH2 and repressing p15 and p21 in gastric cancer. *Oncotarget*. 2016 7(9), 9773–87.
 175. Chen P, Fang X, Xia B, Zhao Y, Li Q, Wu X. Long noncoding RNA LINC00152 promotes cell proliferation through competitively binding endogenous miR-125b with

- MCL-1 by regulating mitochondrial apoptosis pathways in ovarian cancer. *Cancer Med.* 2018 7(9), 4530–41.
176. Qian H, Chen L, Huang J, Wang X, Ma S, Cui F, et al. The lncRNA MIR4435-2HG promotes lung cancer progression by activating β -catenin signalling. *J Mol Med.* 2018 96(8), 753–64.
177. Cai J, Zhang J, Wu P, Yang W, Ye Q, Chen Q, et al. Blocking LINC00152 suppresses glioblastoma malignancy by impairing mesenchymal phenotype through the miR-612/AKT2/NF- κ B pathway. *J Neurooncol.* 2018 140(2), 225–36.
178. Yuan Y, Wang J, Chen Q, Wu Q, Deng W, Zhou H, et al. Long non-coding RNA cytoskeleton regulator RNA (CYTOR) modulates pathological cardiac hypertrophy through miR-155-mediated IKKi signaling. *Biochim Biophys Acta Mol Basis Dis.* 2019 1865(6), 1421–7.
179. Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell.* 2010 40(6), 939–53.
180. Wang Y, Xie Y, Li L, He Y, Zheng D, Yu P, et al. EZH2 RIP-seq identifies tissue-specific long non-coding RNAs. *Curr Gene Ther.* 2018 18(5), 275–85.
181. Chen Q nan, Chen X, Chen Z yao, Nie F qi, Wei C chen, Ma H wei, et al. Long intergenic non-coding RNA 00152 promotes lung adenocarcinoma proliferation via interacting with EZH2 and repressing IL24 expression. *Mol Cancer.* 2017 16(1), 17.
182. Cai Q, Wang Z-Q, Wang S-H, Li C, Zhu Z-G, Quan Z-W, et al. Upregulation of long non-coding RNA LINC00152 by SP1 contributes to gallbladder cancer cell growth and tumor metastasis via PI3K/AKT pathway. *Am J Transl Res.* 2016 8(10), 4068.
183. Evdokimova V, Tognon C, Ng T, Ruzanov P, Melnyk N, Fink D, et al. Translational

- activation of Snail1 and other developmentally regulated transcription factors by YB-1 promotes an epithelial-mesenchymal transition. *Cancer Cell*. 2009 15(5), 402–15.
184. Astanehe A, Finkbeiner MR, Hojabrpour P, To K, Fotovati A, Shadeo A, et al. The transcriptional induction of PIK3CA in tumor cells is dependent on the oncoprotein Y-box binding protein-1. *Oncogene*. 2009 28(25), 2406–18.
185. Huber MA, Azoitei N, Baumann B, Grünert S, Sommer A, Pehamberger H, et al. NF-kappaB is essential for epithelial-mesenchymal transition and metastasis in a model of breast cancer progression. *J Clin Invest*. 2004 114(4), 569–81.
186. Pires BRB, Mencialha AL, Ferreira GM, De Souza WF, Morgado-Díaz JA, Maia AM, et al. NF-kappaB is involved in the regulation of EMT genes in breast cancer cells. *PLoS One*. 2017 12(1), e0169622.
187. Min C, Eddy SF, Sherr DH, Sonenshein GE. NF-kappaB and epithelial to mesenchymal transition of cancer. *J Cell Biochem*. 2008 104(3), 733–44.
188. Fabregat I, Fernando J, Mainez J, Sancho P. TGF-beta signaling in cancer treatment. *Curr Pharm Des*. 2014 20(17), 2934–47.
189. Venter C, Niesler C. Rapid quantification of cellular proliferation and migration using ImageJ. *Biotechniques*. 2019 66(2), 99–102.
190. Cai Z, Aguilera F, Ramdas B, Daulatabad SV, Srivastava R, Kotzin JJ, Carroll M, Wertheim G, Williams A, Janga SC et al. Targeting Bim via a lncRNA Morrbid Regulates the Survival of Preleukemic and Leukemic Cells. *Cell Rep*. 2020 31(12), 107816.
191. Cai Z, Zhang C, Kotzin JJ, Williams A, Henao-Mejia J, Kapur R. Role of lncRNA Morrbid in PTPN11(Shp2)E76K-driven juvenile myelomonocytic leukemia. *Blood Adv*. 2020 4(14), 3246-3251.

192. RNAfold web server. Available online: <http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi> (accessed on 6 July 2020).
193. Anderson EM, Birmingham A, Baskerville S, Reynolds A, Maksimova E, Leake D, et al. Experimental validation of the importance of seed complement frequency to siRNA specificity. *RNA*. 2008 14(5), 853–61.
194. Pei Y, Tuschl T. On the art of identifying effective and specific siRNAs. *Nat Methods*. 2006 3(9), 670–6.
195. Reynolds A, Leake D, Boese Q, Scaringe S, Marshall WS, Khvorova A. Rational siRNA design for RNA interference. *Nat Biotechnol*. 2004 22(3), 326–30.
196. Jackson AL, Burchard J, Leake D, Reynolds A, Schelter J, Guo J, et al. Position-specific chemical modification of siRNAs reduces “off-target” transcript silencing. *RNA*. 2006 12(7), 1197–205.
197. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013 29(1), 15–21.
198. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014 15(12), 550.
199. Leek JT. svaseq: removing batch effects and other unwanted noise from sequencing data. *Nucleic Acids Res*. 2014 42(21), e161.
200. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019 37(8), 907–15.
201. Mazin P, Xiong J, Liu X, Yan Z, Zhang X, Li M, et al. Widespread splicing changes in human brain development and aging. *Mol Syst Biol*. 2013 9(1), 633.
202. Simon MD. Capture Hybridization Analysis of RNA Targets (CHART). *Curr Protoc Mol*

- Biol.* 2013 21, 21.25.
203. Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, et al. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom.* 2003 17(20), 2337–42.
 204. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc.* 2013 8(11), 2281–308.
 205. Giuliano CJ, Lin A, Girish V, Sheltzer JM. Generating single cell–derived knockout clones in mammalian cells with CRISPR/Cas9. *Curr Protoc Mol Biol.* 2019 128(1), e100.
 206. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014 30(15), 2114–20.
 207. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012 9(4), 357–9.
 208. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol.* 2011 29(1), 24–6.
 209. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009 25(16), 2078–9.
 210. Nakagawa S, Naganuma T, Shioi G, Hirose T. Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. *J Cell Biol.* 2011 193(1), 31–9.
 211. Stojic L, Lun ATL, Mangei J, Mascalchi P, Quarantotti V, Barr AR, et al. Specificity of RNAi, LNA and CRISPRi as loss-of-function methods in transcriptional analysis. *Nucleic Acids Res.* 2018 46(12), 5950–66.
 212. Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the

- p53 response. *Cell*. 2010 142(3), 409–19.
213. Wlodkowic D, Skommer J, Darzynkiewicz Z. Flow cytometry-based apoptosis detection. *Methods Mol Biol*. 2009 559, 19–32.
214. Singh R, Letai A, Sarosiek K. Regulation of apoptosis in health and disease: the balancing act of BCL-2 family proteins. *Nat Rev Mol Cell Biol*. 2019 20(3), 175–93.
215. Tolba R, Kraus T, Liedtke C, Schwarz M, Weiskirchen R, Weiskirchen R. Diethylnitrosamine (DEN)-induced carcinogenic liver injury in mice. *Lab Anim*. 2015 49(S1), 59–69.
216. Simanshu DK, Nissley D V., McCormick F. RAS proteins and their regulators in human disease. *Cell*. 2017 170(1), 17–33.
217. Maquat LE. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat Rev Mol Cell Biol*. 2004. 5(2), 89–99.
218. Holbrook JA, Neu-Yilik G, Hentze MW, Kulozik AE. Nonsense-mediated decay approaches the clinic. *Nat Genet*. 2004. 36(8), 801–8.
219. Lewis BP, Green RE, Brenner SE. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci U S A*. 2003 100(1), 189–92.
220. Barbier J, Dutertre M, Bittencourt D, Sanchez G, Gratadou L, de la Grange P, et al. Regulation of H-ras splice variant expression by cross talk between the p53 and nonsense-mediated mRNA decay pathways. *Mol Cell Biol*. 2007 27(20), 7315–33.
221. Chen W-C, To MD, Westcott PM, Delrosario R, Kim I-J, Philips M, et al. Regulation of KRAS4A/B splicing in cancer stem cells by the RBM39 splicing complex. *bioRxiv*. 2019, 646125.

222. Fotiadou PP, Takahashi C, Rajabi HN, Ewen ME. Wild-type Nras and Kras perform distinct functions during transformation. *Mol Cell Biol.* 2007 27(19), 6742–55.
223. Knott GJ, Bond CS, Fox AH. The DBHS proteins SFPQ, NONO and PSPC1: a multipurpose molecular scaffold. *Nucleic Acids Res.* 2016 44(9), 3989–4004.
224. Heyd F, Lynch KW. Phosphorylation-dependent regulation of PSF by GSK3 controls CD45 alternative splicing. *Mol Cell.* 2010 40(1), 126–37.
225. Ray P, Kar A, Fushimi K, Havlioglu N, Chen X, Wu JY. PSF suppresses tau exon 10 inclusion by interacting with a stem-loop structure downstream of exon 10. *J Mol Neurosci.* 2011 45(3), 453–66.
226. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell.* 2013 152(5), 1173–83.
227. Rosenbluh J, Xu H, Harrington W, Gill S, Wang X, Vazquez F, et al. Complementary information derived from CRISPR Cas9 mediated gene deletion and suppression. *Nat Commun.* 2017 8, 15403.
228. Kim A, Lennox Kim A. BMA. Cellular localization of long non-coding RNAs affects silencing by RNAi more than by antisense oligonucleotides. *Nucleic Acids Res.* 2016 44(2), 863–77.
229. Zhang XH, Tee LY, Wang XG, Huang QS, Yang SH. Off-target effects in CRISPR/Cas9-mediated genome engineering. *Mol Ther Nucleic Acids.* 2015. 4(11), e264.
230. Morgens DW, Deans RM, Li A, Bassik MC. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat Biotechnol.* 2016 34(6), 634–6.
231. Guo F, Li Y, Liu Y, Wang J, Li Y, Li G. Inhibition of metastasis-associated lung

- adenocarcinoma transcript 1 in CaSki human cervical cancer cells suppresses cell proliferation and invasion. *Acta Biochim Biophys Sin (Shanghai)*. 2010 42(3), 224–9.
232. Feng C, Zhao Y, Li Y, Zhang T, Ma Y, Liu Y. LncRNA MALAT1 promotes lung cancer proliferation and gefitinib resistance by acting as a miR-200a sponge. *Arch Bronconeumol*. 2019 55(12), 627–33.
233. Liu Q, Shi H, Yang J, Jiang N. Long non-coding RNA NEAT1 promoted hepatocellular carcinoma cell proliferation and reduced apoptosis through the regulation of let-7b-IGF-1R axis. *Onco Targets Ther*. 2019 12(1), 10401–13.
234. Li D, Sun Y, Wan Y, Wu X, Yang W. LncRNA NEAT1 promotes proliferation of chondrocytes via down-regulation of miR-16-5p in osteoarthritis. *J Gene Med*. 2020
235. Carlevaro-Fita J, Polidori T, Das M, Navarro C, Zoller TI, Johnson R. Ancient exapted transposable elements promote nuclear enrichment of human long noncoding RNAs. *Genome Res*. 2019 29(2), 208-222.
236. Zhang B, Gunawardane L, Niazi F, Jahanbani F, Chen X, Valadkhan S. A Novel RNA Motif Mediates the Strict Nuclear Localization of a Long Noncoding RNA. *Mol Cell Biol*. 2014 34(12), 2318–2329.
237. Xue Z, Hennelly S, Doyle B, Gulati AA, Novikova IV, Sanbonmatsu KY, Boyer LA. A G-Rich Motif in the lncRNA Braveheart Interacts with a Zinc-Finger Transcription Factor to Specify the Cardiovascular Lineage. *Mol Cell*. 2016 64(1), 37-50
238. Tsujimoto Y, Finger LR, Yunis J, Nowell PC, Croce CM. Cloning of the chromosome breakpoint of neoplastic B cells with the t(14;18) chromosome translocation. *Science*. 1984 226(4678), 1097-9.
239. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J,

- Boehm JS, Dobson J, Urashima M et al. The landscape of somatic copy-number alteration across human cancers. *Nature* 2010 463(7283), 899-905.
240. Czabotar PE, Lessene G, Strasser A, Adams JM. Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. *Nat Rev Mol Cell Biol.* 2014 15(1), 49-63
241. Kuscu C, Arslan S, Singh R, Thorpe J, Adli M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol.* 2014, 32(7):677-83
242. Zhang XH, Tee LY, Wang XG, Huang QS, Yang SH. Off-target effects in CRISPR/Cas9-mediated genome engineering. *Mol Ther Nucleic Acids.* 2015. 4(11), e264.
243. Morgens DW, Deans RM, Li A, Bassik MC. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat Biotechnol.* 2016 34(6), 634–6.
244. Guo F, Li Y, Liu Y, Wang J, Li Y, Li G. Inhibition of metastasis-associated lung adenocarcinoma transcript 1 in CaSki human cervical cancer cells suppresses cell proliferation and invasion. *Acta Biochim Biophys Sin (Shanghai).* 2010 42(3), 224–9.
245. Feng C, Zhao Y, Li Y, Zhang T, Ma Y, Liu Y. LncRNA MALAT1 promotes lung cancer proliferation and gefitinib resistance by acting as a miR-200a sponge. *Arch Bronconeumol.* 2019 55(12), 627–33.
246. Liu Q, Shi H, Yang J, Jiang N. Long non-coding RNA NEAT1 promoted hepatocellular carcinoma cell proliferation and reduced apoptosis through the regulation of let-7b-IGF-1R axis. *Onco Targets Ther.* 2019 12(1), 10401–13.
247. Li D, Sun Y, Wan Y, Wu X, Yang W. LncRNA NEAT1 promotes proliferation of chondrocytes via down-regulation of miR-16-5p in osteoarthritis. *J Gene Med.* 2020

248. Carlevaro-Fita J, Polidori T, Das M, Navarro C, Zoller TI, Johnson R. Ancient exapted transposable elements promote nuclear enrichment of human long noncoding RNAs. *Genome Res.* 2019 29(2), 208-222.
249. Zhang B, Gunawardane L, Niazi F, Jahanbani F, Chen X, Valadkhan S. A Novel RNA Motif Mediates the Strict Nuclear Localization of a Long Noncoding RNA. *Mol Cell Biol.* 2014 34(12), 2318–2329.
250. Xue Z, Hennelly S, Doyle B, Gulati AA, Novikova IV, Sanbonmatsu KY, Boyer LA. A G-Rich Motif in the lncRNA Braveheart Interacts with a Zinc-Finger Transcription Factor to Specify the Cardiovascular Lineage. *Mol Cell.* 2016 64(1), 37-50
251. Tsujimoto Y, Finger LR, Yunis J, Nowell PC, Croce CM. Cloning of the chromosome breakpoint of neoplastic B cells with the t(14;18) chromosome translocation. *Science.* 1984 226(4678), 1097-9.
252. Beroukhi R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M et al. The landscape of somatic copy-number alteration across human cancers. *Nature* 2010 463(7283), 899-905.
253. Czabotar PE, Lessene G, Strasser A, Adams JM. Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. *Nat Rev Mol Cell Biol.* 2014 15(1), 49-63
254. Potter DS, Letai A. To Prime, or Not to Prime: That Is the Question. *Cold Spring Harb Symp Quant Biol.* 2016 81, 131-140.
255. Rooswinkel RW, Van De Kooij B, Verheij M, Borst J. Bcl-2 is a better ABT-737 target than Bcl-xL or Bcl-w and only Noxa overcomes resistance mediated by Mcl-1, Bfl-1, or Bcl-B. *Cell Death Dis.* 2012 3(8), e366.

256. Fan JC, Zeng F, Le YG, Xin L. LncRNA CASC2 inhibited the viability and induced the apoptosis of hepatocellular carcinoma cells through regulating miR-24-3p. *J Cell Biochem.* 2018 119(8), 6391-6397.
257. Ye F, Tian L, Zhou Q, Feng D. LncRNA FER1L4 induces apoptosis and suppresses EMT and the activation of PI3K/AKT pathway in osteosarcoma cells via inhibiting miR-18a-5p to promote SOCS5. *Gene.* 2019 721, 144093.
258. Dong B, Zhou B, Sun Z, Huang S, Han L, Nie H, Chen G, Liu S, Zhang Y, Bao N. LncRNA-FENDRR mediates VEGFA to promote the apoptosis of brain microvascular endothelial cells via regulating miR-126 in mice with hypertensive intracerebral hemorrhage. *Microcirculation.* 2018 25(8), e12499.
259. Huang D, Chen J, Yang L, Ouyang Q, Li J, Lao L, Zhao J, Liu J, Lu Y, Xing Y et al. NKILA lncRNA promotes tumor immune evasion by sensitizing T cells to activation-induced cell death. *Nat Immunol.* 2018 19(10), 1112-1125.
260. Carlevaro-Fita J, Lanzós A, Feuerbach L, Hong C, Mas-Ponte D, Pedersen JS, PCAWG Drivers and Functional Interpretation Group, Johnson R, PCAWG Consortium. Cancer LncRNA Census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis. *Commun Biol.* 2020 3(1), 56.
261. Schoenberg DR, Maquat LE. Regulation of cytoplasmic mRNA decay *Nat Rev Genet.* 2012 13(4), 246–59.
262. Lelivelt MJ, Culbertson MR. Yeast Upf proteins required for RNA surveillance affect global expression of the yeast transcriptome. *Mol Cell Biol.* 1999 19(10), 6710–9.
263. Ramani AK, Nelson AC, Kapranov P, Bell I, Gingeras TR, Fraser AG. High resolution transcriptome maps for wild-type and nonsense-mediated decay-defective *Caenorhabditis*

- elegans*. *Genome Biol.* 2009 10(9), R101.
264. Rehwinkel J, Letunic I, Raes J, Bork P, Izaurralde E. Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. *RNA*. 2005 11(10), 1530–44.
265. Kurosaki T, Popp MW, Maquat LE. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat Rev Mol Cell Biol.* 2019. 20(7):406-420.
266. Yi Z, Sanjeev M, Singh G. The Branched Nature of the Nonsense-Mediated mRNA Decay Pathway. *Trends Genet.* 2020. S0168-9525(20)30210-9.
267. Hutchinson JN, Ensminger AW, Clemson CM, Lynch CR, Lawrence JB, Chess A. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics.* 2007 8(1), 39.
268. Gong C, Maquat LE. LncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu element. *Nature.* 2011 470(7333), 284–90.
269. Guan Q, Zheng W, Tang S, Liu X, Zinkel RA, Tsui K-W, et al. Impact of nonsense-mediated mRNA decay on the global expression profile of budding yeast. *PLoS Genet.* 2006 2(11), e203.
270. Gehring NH, Kunz JB, Neu-Yilik G, Breit S, Viegas MH. Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. *Mol Cell.* 2005. 20(1):65-75
271. Einfeld AK, Schwind S, Hoag KW, Walker CJ, Liyanarachchi S, Patel R, et al. NRAS isoforms differentially affect downstream pathways, cell growth, and cell transformation. *Proc Natl Acad Sci U S A.* 2014 111(11), 4179–84.
272. Duggan MC, Regan-Fendt K, Olaverria Salavaggione GN, Howard JH, Stiff AR, Sabella J, et al. Neuroblastoma RAS viral oncogene homolog mRNA is differentially spliced to

- give five distinct isoforms: Implications for melanoma therapy. *Melanoma Res.* 2019 29(5), 491–500.
273. Snijders AP, Hautbergue GM, Bloom A, Williamsom JC, Minshull TC, Phillips HL, et al. Arginine methylation and citrullination of splicing factor proline- and glutamine-rich (SFPQ/PSF) regulates its association with mRNA. *RNA.* 2015 21(3), 347–59.
274. Peng R, Dye BT, Pérez I, Barnard DC, Thompson AB, Patton JG. PSF and p54nrb bind a conserved stem in U5 snRNA. *RNA.* 2002 8(10), 1334–47.
275. Marko M, Leichter M, Patrinoú-Georgoula M, Guialis A. hnRNP M interacts with PSF and p54nrb and co-localizes within defined nuclear structures. *Exp Cell Res.* 2010 316(3), 390–400.
276. Kim KK, Kim YC, Adelstein RS, Kawamoto S. Fox-3 and PSF interact to activate neural cell-specific alternative splicing. *Nucleic Acids Res.* 2011 39(8), 3064–78.
277. Cho S, Moon H, Loh TJ, Oh HK, Williams DR, Liao DJ, et al. PSF contacts exon 7 of SMN2 pre-mRNA to promote exon 7 inclusion. *Biochim Biophys Acta.* 2014 1839(6), 517–25.
278. Wang G, Cui Y, Zhang G, Garen A, Song X. Regulation of proto-oncogene transcription, cell proliferation, and tumorigenesis in mice by PSF protein and a VL30 noncoding RNA. *Proc Natl Acad Sci U S A.* 2009 106(39), 16794–8.
279. Song X, Sun Y, Garen A. Roles of PSF protein and VL30 RNA in reversible gene regulation. *Proc Natl Acad Sci U S A.* 2005 102(34), 12189–93.
280. Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. *Science.* 1997. 278(5338):631–7.
281. Kim J, Kim I, Han SK, Bowie JU, Kim S. Network rewiring is an important mechanism of

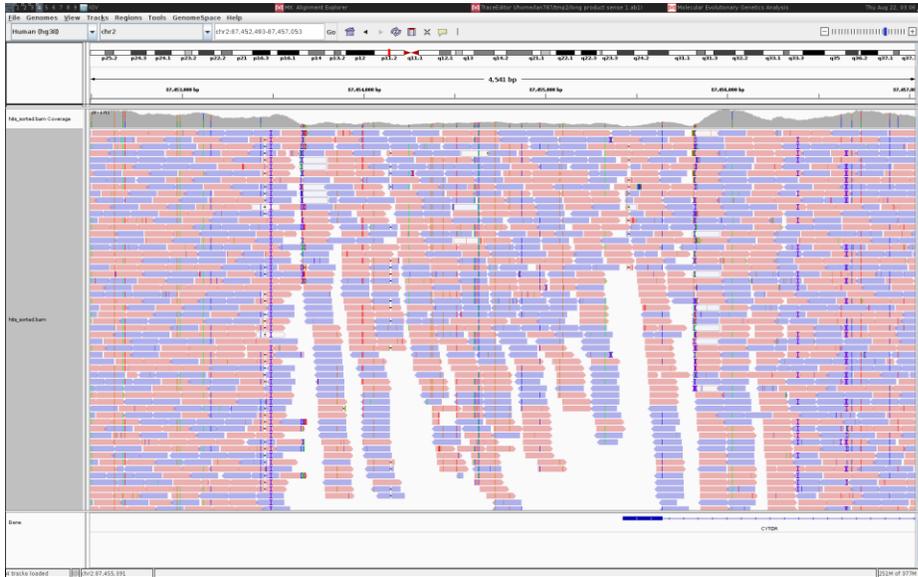
gene essentiality change. *Sci Rep.* 2012. 2:900.

282. Han SK, Kim D, Lee H, Kim I, Kim S. Divergence of Noncoding Regulatory Elements Explains Gene-Phenotype Differences between Human and Mouse Orthologous Genes. *Mol Biol Evol.* 2018. 35(7):1653-1667.

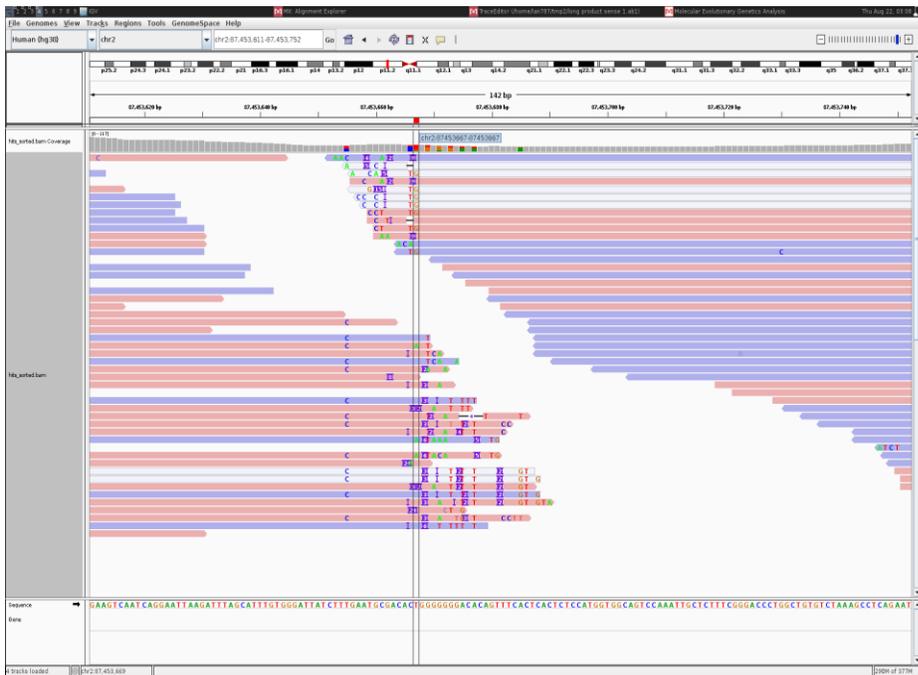
Supplementary Figures

Figure S1. CYTOR gene locus. NGS validation of the knockout.

A



B



C

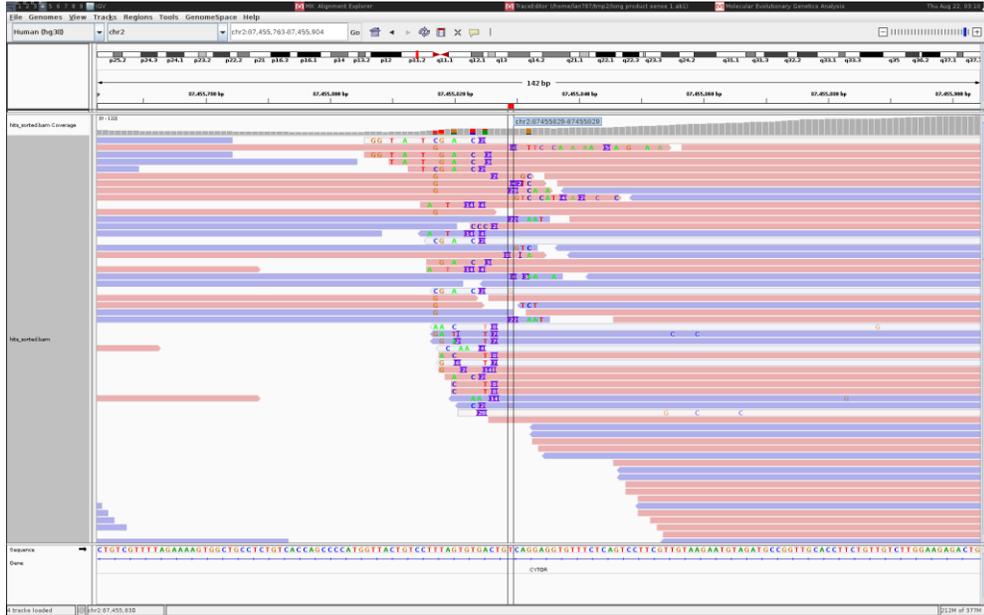
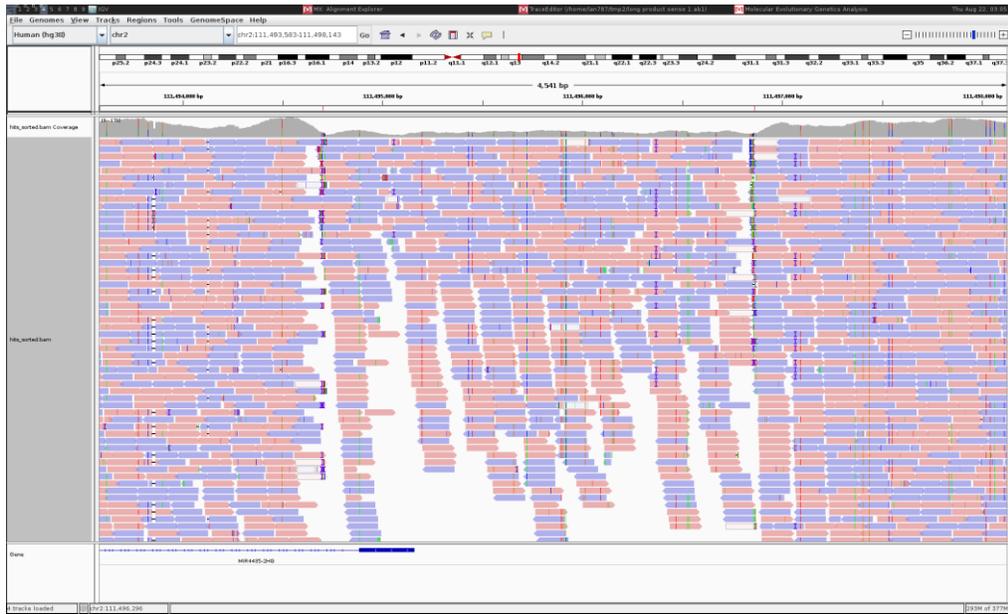
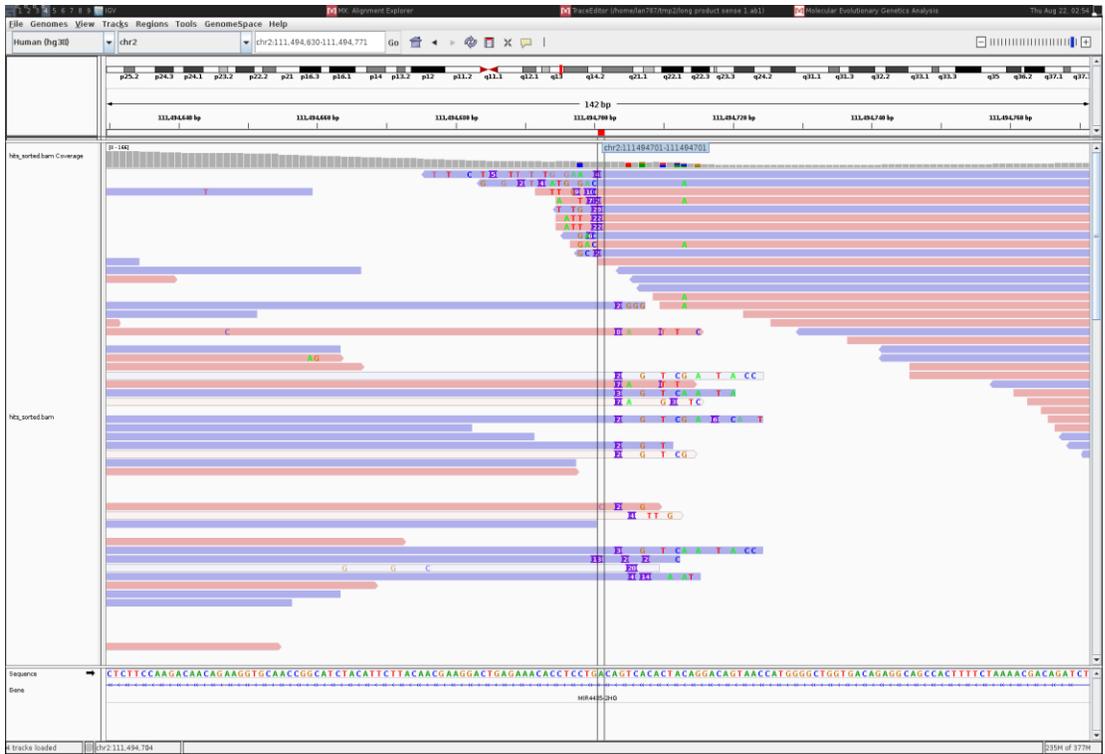


Figure S2. hMorrbid gene locus. NGS validation of the knockout

A



B



C

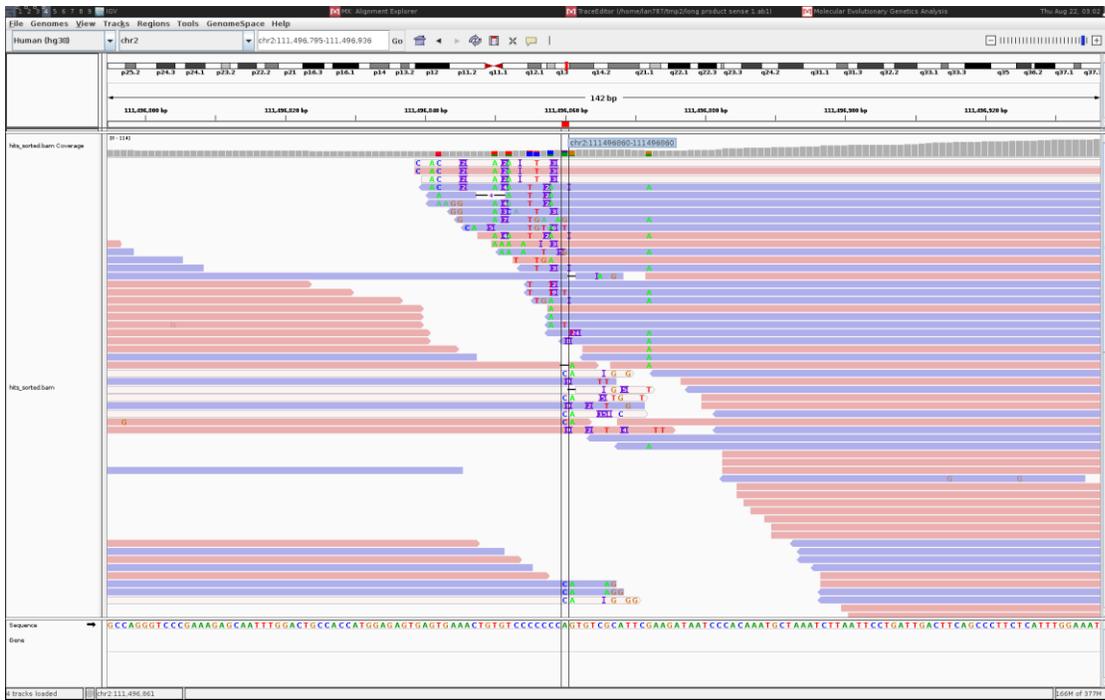
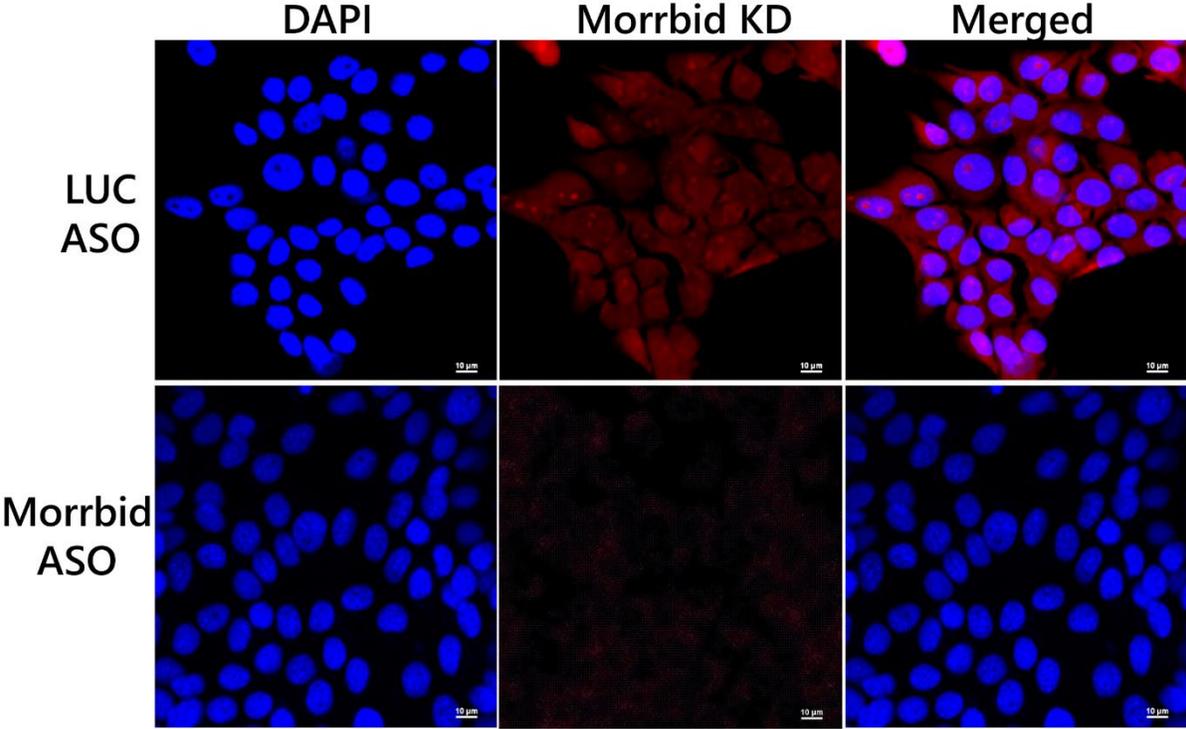


Figure S3. Fluorescent in situ hybridization analysis of Morrbrid depletion in AML12 cells after 24h of inhibition by ASOs. DNA was stained with Dapi, Morrbrid was stained with Cy5-labeled probes.



Supplementary Tables

Supplementary Table S1. List of ASOs and siRNAs used in the study.

Name	Sequence
Morrbid ASOs	
Morrbid ASO-1	usgsuscscsAsCsAsTsGsAsTsTsAsGsgsasasasas
Morrbid ASO-2	gsuscscscsCsTsCsAsTsTsCsTsCsAsgsasgsasus
Morrbid ASO-3	ususgscsusTsTsTsTsAsAsTsGsAsAsgsasasasgs
Morrbid ASO-4	csusgsusgsasAsGsAsTsCsCsCsAsAsGsasusasusgscs
Morrbid ASO-5	gsasusasgsAsCsGsGsGsTsCsCsGsCscsusgscscs
Morrbid ASO-6	asgsasgscsAsTsCsCsGsAsAsAsAsGscsususcusgs
Morrbid ASO-7	cscscsasgscsAsCsCsCsGsTsGsAsGsCsascsusgsasgs
Morrbid ASO-8	cscsgsususCsCsCsGsGsGsTsGsGsGsasascscscs
Morrbid ASO-9	gscsasgscsAsGsGsTsAsGsGsGsTsTsgscsuscs
Morrbid ASO-10	ascsgsuscsCsTsGsAsTsTsTsTsTsCsusgsusgsus
Morrbid ASO-11	gsasusasCsAsAsAsCsCsAsGsCsTsgsusasasgs
Morrbid ASO-12	gscscsasasusAsAsAsTsAsAsAsGsTsAsasusasgsasus
Morrbid ASO-13	csuscscscsAsGsCsCsGsTsGsAsTsCsasgscsasgs
Upper case – 2'-deoxynucleotide, lower case – 2'-O-methylribonucleotide, s –phosphorothioate group.	
SFPQ siRNAs	
msSfpQ-1	GuAuGAAGGGccAAAuAAATsT UUuAUUUGGCCCUUCAUACTsT
msSfpQ-2	cAuuAAGcuuGAAucuAGATsT UCuAGAUUcAAGCUuAAUGTsT
msSfpQ-3	ccAGAAGAAuccAAuGuAuTsT AuAcAUUGGAUUCUUCUGGTsT
msSfpQ-4	AAAcAuGAAGGAuGcuAAATsT UUuAGcAUCCUUCaUGUUUTsT
msSfpQ-5	uAuuGAAAGGGcuuGuuGuATsT uAcAAcAGCCCUUUCAAuATsT
msSfpQ-6	cuGuCuGuucGAAaucucuTsT AGAGAUUUCGAACAGAcAGTsT
control	cuuAcGcuGAGuAcuucGATsT UCGAAGuACUCAGCGuAAGTsT

UPF siRNA	
UPF-1 siRNA	cuGcGuGGuuuAcuGuAAuTsT uAUGUUCUGGuACUGGuAGTsT

Upper case – ribonucleotide, lower case – 2'-O-methylribonucleotide, s – phosphorothioate group.

Supplementary Table S2. List of PCR primers used in the study.

Name	Sequence, 5'→3'
GAPDH_FWD	TGCACCACCAACTGCTTAGC
GAPDH_REV	GGATGCAGGGATGATG
CYTOR/hMorrbid all transcripts FWD	ACAGCACAGTTCCTGGGAAG
CYTOR/hMorrbid all transcripts REV	ACAGGTAGAGGTGCTGGAGG
hMorrbid-217 FWD	TCCAAAATCACATGCCTTCA
hMorrbid-217 REV	CCAAGATGCAACCCTCAGTT
hE-Cadherin FWD	TTTGACGCCGAGAGCTACAC
hE-Cadherin REV	GCTGTCCTTTGTCGACCG
hVimentin FWD	TCTGGATTCACTCCCTCTGG
hVimentin REV	TCAAGGTCATCGTGATGCTG
hEpCAM FWD	TTGCTGGAATTGTTGTGCTGG
hEpCAM REV	GTTCCCTATGCATCTCACCCA
hEGFR FWD	AACTGTGAGGTGGTCCTTGG
hEGFR REV	TGAGGACATAACCAGCCACC
hp15 RWD	CTAGTGGAGAAGGTGCGACAG
hP15 REV	TCATCATGACCTGGATCGCG
hp21 FWD	TGGAGACTCTCAGGGTCGAA
hP21 REV	GGATTAGGGCTTCCTCTTGG
hBim FWD	TCTGCCATCCCTGCTGATTTAG
hBim REV	ACCCCAACTGATTTAGCGTCAT
hCyclin D1FWD	GCTGCGAAGTGGAACCATC
hCyclin D1 REV	CCTCCTTCTGCACACATTTGAA
mMorrib_FWD	AAATGACACAGACACAGAAAAATCA
mMorrbid_REV	ACTGAGTAGCTAAGAGTCCGTTCC
PTC junction_FWD	CCCACCATAGAGGATTCTTACC
PTC junction_REV	GATCCCACCATAGAGGATTCTTACCG
PTC down_FWD	CAGATATAAATTCACCTGCCCTTATGT
PTC down_REV	GTAGAGGTTAATATCTGCAAATGATTT G
NRAS no PTC_FWD	GGTCTCACTGCACTACCCTG
NRAS no PTC_REV	ATAATCACACGCATGCATGCAC
NRAS total_FWD	ACGAACTGGCCAAGAGTTACG

NRAS total_REV	CATTCGGTACTGGCGTATCTCC
NRAS total 2_FWD	ACATGAGGACAGGCGAAGG
NRAS total 2_REV	AGCACCATGGGGACATCATC
m_SFPQ_FWD	GGAAGCGACATGCGTACTGA
m_SFPQ_REV	TTCCAGGCCCCATTCCTCTA
mUPF1_FWD	AGATCACGGCACAGCAGAT
mUPF1_REV	TGGCAGAAGGGTTTTTCCTT
msnU6_FWD	CGCTTCGGCAGCACATATAC
msnU6_REV	AAATATGGAACGCTTCACGA
Nras_T7_PCR_fwd1	TAATACGACTCACTATAGGGAAAAGCG CCTTGACGATCC
Nras_noT7_PCR_rev1	GCAAATACACAGAGGAACCCTTCGCC
Nras_noT7_PCR_fwd1	GGAAAAGCGCCTTGACGATCC
Nras_T7_PCR_rev1	TAATACGACTCACTATAGGCAAATACA CAGAGGAACCCTTCGCC
NRAS_ex1-intr1_fwd	CCACTTTGTGGATGAATATGATCCCAC
NRAS_ex1-intr1_rev	CCTCCTTGCTTTCTCTCTCTTTACT

Supplementary Table S3. Biotinylated Probes used in CHART and RIP protocols.

Name	Sequence
Morrbid_biot_1	Biotin-s-csusgsusgsasAsGsAsTsCsCsCsAsAsGsasusasusgsc-s-NH2
Morrbid_biot_2	Biotin-s-gsasusasgsAsCsGsGsGsTsCsCsGsCscsusgscsc-s-NH2
Morrbid_biot_3	Biotin-s- cscscsasgscsAsCsCsCsGsTsGsAsGsCsascsusgsasg-s-NH2
Morrbid_biot_4	Biotin-s- csuscscscsAsGsCsCsGsTsGsAsTsCsasgscsasg-s-NH2
Control_1	Biotin-s-csasgsasgsasgscsTsCsAsCsAsCsTsTsCsAsasasasusgsuscsc-s-NH2
Control_2	Biotin-s-csusgsasgsasgsusAsGsGsTsTsTsGsTsTsTscscsasgsgsasa-s-NH2
Control_3	Biotin-s- gsgsususasusgsusTsCsCsTsAsGsTsGsAsCsAsgsasasasgsasgsu-s-NH2
Control_4	Biotin-s- gscsasususcscscsAsTsCsAsCsAsTsCsTsCsTscsusasgsusgsusg-s-NH2

Upper case – 2'-deoxynucleotide, lower case – 2'-O-methylribonucleotide, s –phosphorothioate group, NH2 – 3'-hexylamine.

Supplementary Table S4. Differential alternative splicing genes in response to Morrbid depletion

AS Type	Gene name	Gene ID
---------	-----------	---------

Alternative Acceptor	Chd4	ENSMUSG00000063870
Alternative Acceptor	Atp2c1	ENSMUSG00000032570
Alternative Acceptor	Myrf	ENSMUSG00000036098
Alternative Acceptor	Brd2	ENSMUSG00000024335
Alternative Acceptor	Map4k4	ENSMUSG00000026074
Alternative Acceptor	Sbds	ENSMUSG00000025337
Alternative Acceptor	Hax1	ENSMUSG00000027944
Alternative Acceptor	Atxn2l	ENSMUSG00000032637
Alternative Acceptor	Hnrnpa2b1	ENSMUSG00000004980
Alternative Acceptor	Oaz1	ENSMUSG00000035242
Alternative Acceptor	Rpl3	ENSMUSG00000060036
Alternative Acceptor	Psm2	ENSMUSG00000006998
Alternative Acceptor	Msln	ENSMUSG00000063011
Alternative Acceptor	Use1	ENSMUSG00000002395
Alternative Acceptor	1600012H06Rik	ENSMUSG00000050088
Alternative Donor	Rbm25	ENSMUSG00000010608
Alternative Donor	Smardc1	ENSMUSG00000029920
Alternative Donor	Herc2	ENSMUSG00000030451
Alternative Donor	Taf1d	ENSMUSG00000031939
Alternative Donor	Gnl3	ENSMUSG00000042354
Alternative Donor	Gapdh	ENSMUSG00000057666
Alternative Donor	Akt1s1	ENSMUSG00000011096
Alternative Donor	Rnasek	ENSMUSG00000093989
Alternative Donor	Gapdh	ENSMUSG00000057666
Alternative Donor	Zfp574	ENSMUSG00000045252
Alternative Donor	Calu	ENSMUSG00000029767
Alternative Donor	2610507B11Rik	ENSMUSG00000010277
Alternative Donor	Spg20	ENSMUSG00000036580
Alternative Donor	Chmp2a	ENSMUSG00000033916
Alternative Donor	Gapdh	ENSMUSG00000057666
Alternative Donor	Ei24	ENSMUSG00000062762
Alternative Donor	Dnaja1	ENSMUSG00000028410
Alternative Donor	Tomm34	ENSMUSG00000018322
Alternative Donor	Spg20	ENSMUSG00000036580
Alternative Donor	Dcaf8	ENSMUSG00000026554
Alternative Donor	Emc6	ENSMUSG00000047260
Alternative Donor	Mboat7	ENSMUSG00000035596
Alternative Donor	Ptprn	ENSMUSG00000026204

Alternative Donor	Zkscan17	ENSMUSG00000020472
Alternative Donor	Sf1	ENSMUSG00000024949
Alternative Donor	Cog8	ENSMUSG00000031916
Alternative Donor	Cited2	ENSMUSG00000039910
Alternative Donor	Cyc1	ENSMUSG00000022551
Alternative Donor	Snhg1	ENSMUSG000000108414
Alternative Donor	Trim3	ENSMUSG00000036989
Alternative Donor	Hhex	ENSMUSG00000024986
Alternative Donor	Srsf5	ENSMUSG00000021134
Alternative Donor	Nin	ENSMUSG00000021068
Alternative Donor	Dnaja1	ENSMUSG00000028410
Alternative Donor	Pafah1b2	ENSMUSG00000003131
Alternative Donor	Rack1	ENSMUSG00000020372
Alternative Donor	Pttg1	ENSMUSG00000020415
Alternative Donor	Stk16	ENSMUSG00000026201
Cassette Exon	Commd1	ENSMUSG00000051355
Cassette Exon	Sh3pxd2a	ENSMUSG00000053617
Cassette Exon	Kmt5a	ENSMUSG00000049327
Cassette Exon	Tra2b	ENSMUSG00000022858
Cassette Exon	Mfge8	ENSMUSG00000030605
Cassette Exon	Tpm1	ENSMUSG00000032366
Cassette Exon	Gnb1	ENSMUSG00000029064
Cassette Exon	Ptpn12	ENSMUSG00000028771
Cassette Exon	Bub3	ENSMUSG00000066979
Cassette Exon	Dhx9	ENSMUSG00000042699
Cassette Exon	Ube2i	ENSMUSG00000015120
Cassette Exon	Vkorc111	ENSMUSG00000066735
Cassette Exon	Chmp2a	ENSMUSG00000033916
Cassette Exon	Fam20b	ENSMUSG00000033557
Cassette Exon	Nras	ENSMUSG00000027852
Cassette Exon	Cystm1	ENSMUSG00000046727
Cassette Exon	Srsf3	ENSMUSG00000071172
Cassette Exon	Xiap	ENSMUSG00000025860
Cassette Exon	Arl6ip1	ENSMUSG00000030654
Cassette Exon	Actn1	ENSMUSG00000015143
Cassette Exon	Slc25a39	ENSMUSG00000018677
Cassette Exon	Hdgf	ENSMUSG00000004897
Cassette Exon	Impdh2	ENSMUSG00000062867

Cassette Exon	Actr3	ENSMUSG00000026341
Cassette Exon	Snrbp	ENSMUSG00000027404
Cassette Exon	Epn1	ENSMUSG00000035203
Cassette Exon	Srrm2	ENSMUSG00000039218
Cassette Exon	Furin	ENSMUSG00000030530
Cassette Exon	Ndufa6	ENSMUSG00000022450
Cassette Exon	Pkm	ENSMUSG00000032294
Retained Intron	Aurkaip1	ENSMUSG00000065990