Università degli Studi di Milano


PhD program
**Experimental Medicine and Medical Biotechnologies**

Department
**Oncology and Hemato-Oncology**
**MED/15**

Curriculum
**Immunology and Hemato-Oncology**

**Cycle XXXI**

# The relevance of long non-coding RNA in the biological and clinical heterogeneity of Multiple Myeloma

PhD thesis:

Cristina VINCI

R11425

Tutor: Professor Antonino NERI

---

**Academic Year: 2017/2018**

# Table of contents

# Abstract

Multiple myeloma (MM) is a malignant proliferation of antibody-secreting bone marrow plasma cells that accounts for 10% of all hematological malignancies. MM is characterized by a wide clinical spectrum ranging from the presumed pre-malignant condition called monoclonal gammopathy of undetermined significance, to extra-medullary myeloma/plasma cell leukemia (PCL). Despite the remarkable improvements in treatment and patient care, MM remains an incurable disease.

In the last few years it has been demonstrated that the vast majority (>90 %) of the human genome sequence is actively transcribed, but only less than 2% of transcripts serve as mRNA to encode protein. The non-coding RNAs (ncRNAs) are involved in various biological processes and are commonly divided into short (<200nt), including microRNAs (miRNAs), and long (>200 nt, lncRNAs) transcripts. The dysregulation of ncRNAs has been reported to occur virtually in all types of cancer including MM. The role of miRNA in malignant transformation has been widely explored and recently different lncRNA roles has been the identified in normal and malignant cells.

With the aim to identify lncRNAs modulated in MM, we set up the condition to the use of GeneChip® Human Gene 2.0 ST microarray, able to distinguish more than 10000 lncRNAs, and RNAseq methods. Based on these analyses, on our cohort of MM patients (50 MM primary tumors, 15 PCL and 4 normal donors), we evidenced *ST3GAL6-AS1* as the unique lncRNA up-modulated in pathological samples compared to normal controls. GEP data evidenced the up-modulation of this lncRNA and a significant correlation with *ST3GAL6*, neighboring gene of *ST3GAL6-AS1*, expression levels. *ST3GAL6* codes for a protein able to produce proteins and lipids glycosylation, and a role for ST3GAL6 in homing and engraftment of MM has been described.

In our study, *ST3GAL6-AS1* and *ST3GAL6* were overexpressed in human MM cell lines (HMCLs), and showed equally localization in nuclear and cytoplasmic fractions and a long half-life.

Molecular analysis of *ST3GAL6-AS1* in HMCLs showed the presence of the splice single nuclear polymorphic (SNP) variation rs13065271 with the production of unpredicted splice events (retention of intron 3 and skipping of exon 3). In particular, homozygous minor allele HMCLs showed a nuclear localization of the lncRNA, a low expression and half-life of

transcripts shorter than the HMCLs with at least one major allele. Furthermore, the SNP rs13065271 analysis made on MM patients confirmed the significant down-regulation of *ST3GAL6-AS1* expression levels in homozygous minor allele patients.

Additionally, silencing of *ST3GAL6-AS1* in HMCLs by GapmeR approach evidenced a significant down-regulation of the lncRNA in all the HMCLs investigated. SNP homozygous minor allele HMCLs did not evidenced biological effects; by contrast, HMCLs with at least one SNP major allele evidenced biological effects in terms of decrease of cell growth and viability, changes in percentage of cell cycle phases, increase of percentage of apoptotic cells, and loss of capability to form colonies. GeneChip® Human Gene 2.0 ST microarray analysis on silenced HMCLs evidenced the up-regulation of *FUCA1*, coding for a lysosomal enzyme involved in the degradation of fucose-containing glycoproteins and glycolipids. Additionally, the expression levels of *FUCA1* in our cohort of patients resulted significantly inversely correlated with *ST3GAL6-AS1* expression levels.

Overall, these data suggest the hypothesis that *ST3GAL6-AS1* may act on the down-regulation of *FUCA1* in MM.

# Sommario

Il mieloma multiplo (MM) è una proliferazione maligna di plasmacellule del midollo osseo secernenti anticorpi, che rappresenta il 10% di tutte le neoplasie ematologiche. Il MM è caratterizzato da un ampio spettro clinico che va da una presunta condizione pre-maligna chiamata gammopatia monoclonale di significato indeterminato a mieloma extra-midollare/ leucemia plasmacellulare (PCL). Nonostante i notevoli miglioramenti nel trattamento e nella cura dei pazienti, il MM è ancora una patologia incurabile.

Negli ultimi anni è stato dimostrato che la stragrande maggioranza (> 90%) della sequenza del genoma umano viene attivamente trascritta, ma solo meno del 2% dei trascritti è rappresentato da RNA messaggero e codifica per proteine. Gli RNA non codificanti (ncRNA) sono coinvolti in vari processi biologici e sono comunemente suddivisi in RNA non codificanti corti (<200nt), come i microRNA (miRNA), e lunghi (> 200 nt, lncRNA).

La de-regolazione dei ncRNA è stata evidenziata in tutti i tipi di tumore, compreso il MM. In particolar modo il ruolo dei miRNA nella trasformazione maligna è stato ampiamente esplorato e solo di recente sono stati evidenziati diversi ruoli dei lncRNA sia nelle cellule normali che patologiche.

Con l'obiettivo di identificare i lncRNA modulati nel MM, abbiamo messo a punto le condizioni per l'utilizzo del GeneChip® Human Gene 2.0 ST microarray, in grado di distinguere più di 10000 lncRNA, e la metodica dell'RNAseq. Sulla base di queste analisi, effettuate sulla nostra casistica di pazienti (50 tumori primari di MM, 15 PCL e 4 donatori sani), abbiamo evidenziato *ST3GAL6-AS1* come unico lncRNA che presenta un significativo incremento di espressione nei campioni patologici rispetto ai controlli sani. I dati GEP hanno evidenziato l'incremento di espressione di questo lncRNA e una correlazione significativa con i livelli di espressione di ST3GAL6, gene adiacente a *ST3GAL6-AS1*. *ST3GAL6* codifica per una proteina che agisce nel processo di produzione di proteine e lipidi glicosilati, ed è stato descritto un suo ruolo nel MM come attore nello spostamento delle cellule (fenomeno dell'"homing") e nel loro attecchimento nel midollo osseo.

Nel nostro studio, *ST3GAL6-AS1* e ST3GAL6 hanno mostrato una sovra espressione nelle linee cellulari umane di MM (HMCLs), una localizzazione omogenea nelle frazioni nucleari e citoplasmatiche e una lunga emivita.

L'analisi molecolare di *ST3GAL6-AS1* nelle HMCLs ha evidenziato la presenza di un polimorfismo a singolo nucleotide (SNP) con codice rs13065271 che interessa un nucleotide fondamentale per gli eventi di splicing. Infatti, la presenza di questo SNP produce 2 eventi di splicing non predetti: la ritenzione dell'introne 3 e il salto dell'esone 3. In particolare, le analisi molecolari su HMCLs omozigoti per l'allele minore hanno mostrato una localizzazione nucleare del lncRNA, un'espressione bassa e un'emivita dei trascritti più breve rispetto alle HMCLs che presentano almeno un allele maggiore. Inoltre, l'analisi dello SNP rs13065271, effettuata su pazienti affetti da MM, ha confermato una significativa riduzione dell'espressione di *ST3GAL6-AS1* in pazienti omozigoti per l'allele mutato.

Il silenziamento di *ST3GAL6-AS1* in HMCLs mediante l'utilizzo di GapmeR ha messo in luce una significativa diminuzione del lncRNA in tutte le HMCLs esaminate. Le HMCLs in omozigosi per l'allele minore dello SNP non hanno presentato effetti biologici; al contrario, le HMCLs con almeno un allele maggiore dello SNP hanno evidenziato effetti biologici in termini di diminuzione della crescita e vitalità cellulare, variazioni della percentuale delle cellule nelle fasi del ciclo cellulare, aumento della percentuale delle cellule apoptotiche e perdita della capacità di formare colonie. L'analisi di *microarray* con GeneChip® Human Gene 2.0 ST sulle HMCLs silenziate ha rivelato un aumento dell'espressione di *FUCA1*, gene codificante per un enzima lisosomiale coinvolto nella degradazione delle glicoproteine e dei glicolipidi contenenti fucosio. Inoltre, i livelli di espressione di *FUCA1* nella nostra casistica di pazienti di MM sono risultati inversamente correlati, in maniera significativa, con i livelli di espressione di *ST3GAL6-AS1*.

Nel complesso, questi dati suggeriscono l'ipotesi che *ST3GAL6-AS1* possa agire sulla regolazione negativa di *FUCA1* nel MM.

# Introduction

# 1. *Multiple myeloma*

## 1.1 Etiology

Multiple Myeloma (MM) is a malignant proliferation of antibody-secreting bone marrow plasma cells (PCs). MM accounted for 1.7% of all malignancies and 10% of hematological malignancies in the United States in 2017[1]. Italian population represents 4.71% of world population affected by MM. (Figure 1, GLOBOCAN 2012. Graph production: Global Cancer Observatory (http://gco.iarc.fr/)© International Agency for Research on Cancer 2018).

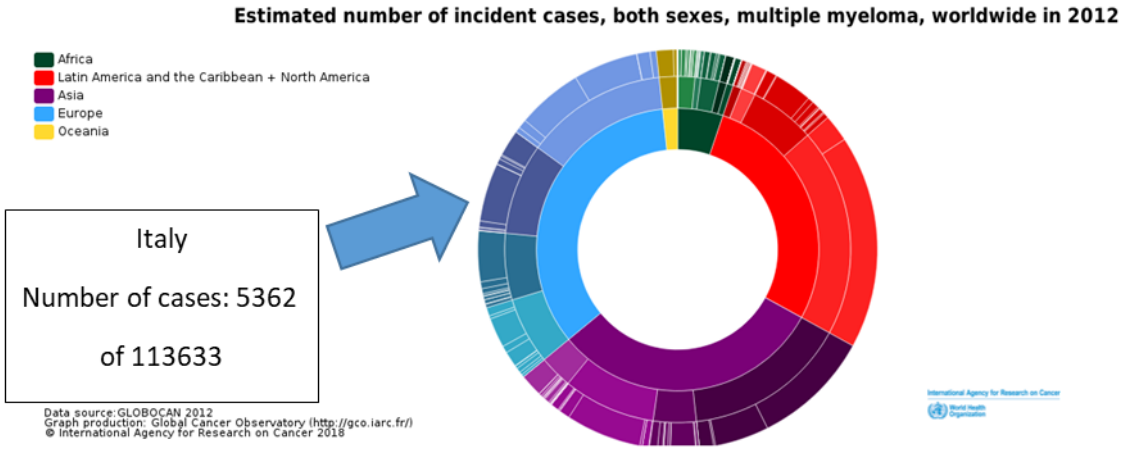



**Figure 1**: **sunburst schematic visualization of MM estimated number of incident cases worldwide 2012**. Data source: GLOBOCAN 2012 http://gco.iarc.fr/today/home accessed August 2018

Based on these data, MM can be considered a rare cancer disease. The first MM identification and description was in 1844 by Solly S.[2]. During these 174 years, the knowledge and therapy of MM were rapidly improved but it remains an incurable disease.

## 1.2 Progression of Multiple Myeloma

For the understanding of MM origin, it is important to recall the fundamental steps of B-cells development. The development of B-cells starts in bone marrow, where they generate a functional B-cell receptor precursor by the rearrangement of immunoglobulin (Ig) genes. The immature/virgin B-cells leave the bone marrow and migrate via the blood and the lymph to the secondary lymphoid organs where they are activated by specific antigens. In the periphery B-cells complete their development to the mature or naïve stage, which is signaled by the appearance of IgD in addition to IgM on the cell surface. This development sequence occurs in

the absence of any contact with exogenous antigen, a stage known as antigen-independent B cell development.

The second phase of B-cell development occurs after encounter with antigen and activation, and is called the antigen-dependent phase. After the meeting with the specific antigen, virgin B-cells move towards the germinal center of peripheral lymphoid tissues where they undergo affinity maturation in response to the antigen present on antigen-presenting cells. This mechanism happens thanks to somatic hypermutation of the Ig heavy chain locus, a process mainly implying single nucleotide substitutions into the IgV (Variable) gene, to produce antigen-specific antibodies. Subsequently, B-cells go through class switch recombination, a mechanism able to switch the isotype of the Ig from IgM and IgD to IgG, IgA and IgE (Figure 2). Specifically, the variable portion of the heavy chain (VH exon) is brought adjacent to with different Ig constant regions (CH exons) to generate antibodies with different effector functions. The B-cells may then leave the GC rapidly differentiating into long-lived PCs or memory B cells. Both somatic hypermutation and class switch recombination are mediated by the generation of double strand DNA breaks in the Ig loci[3].

MM may arise during these processes of PC differentiation. Indeed B-cells may accumulate mutations and translocations able to change genomic sequences using double strand DNA breaks[4-7]. MM cells could be defined as a post-switch B-lineage cell; indeed IgG and IgA isotype are the most common Igs secreted by PCs in MM patients.

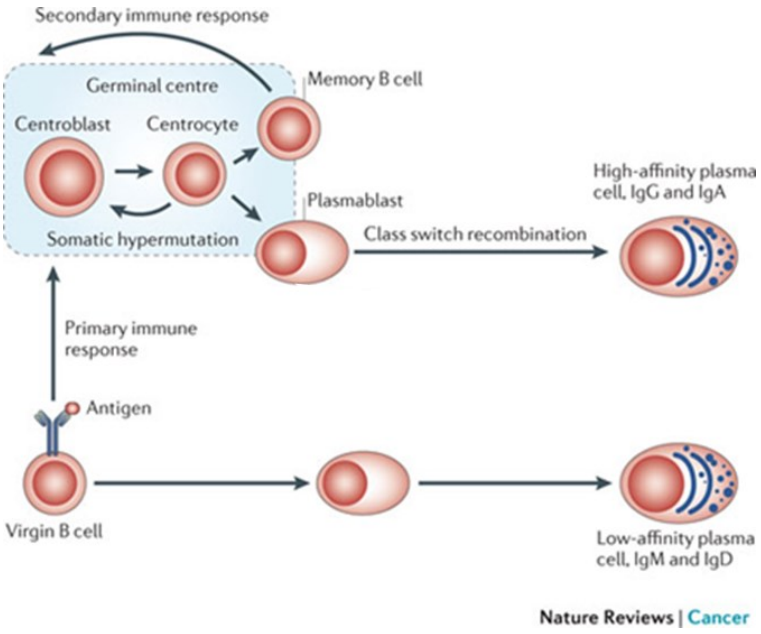



**Figure 2**: **maturation of a Virgin B cell after antigen recognition.** Image modified by Morgan, G. J. et al, Nature reviews. Cancer (2012).

MM is characterized by a wide clinical spectrum ranging from the presumed pre-malignant condition called monoclonal gammopathy of undetermined significance (MGUS), to extra-medullary myeloma/plasma cell leukemia (PCL). Chromosomal translocations, involving important and known oncogenes, and hyperdiploidy events interesting chromosomes 3, 5, 7, 9, 11, 15, 19 and 21, considered as primary genetic events, are the first causes of MGUS progression. Asymptomatic MGUS status is characterized by PC bone marrow infiltration and secretion of monoclonal protein. Rasmussen T. et al. in 2010 suggested that all MM patients pass through an MGUS phase[8]. After MGUS phase we can evidenced a smoldering MM phase (SMM). SMM is defined by the presence of higher secretion of monoclonal protein than in MGUS, with no evidence of specific MM end-organ damage and higher risk of progression to MM[9-11]. The accumulation of secondary genetic events as copy numbers abnormalities, DNA hypomethylation and acquired mutation could cause the progression of MM

MM patients, after an extramedullary accumulation of PCs, could become PCL[12].

Furthermore, MM disease is characterized by the presences of CRAB (acronym for the most common symptoms of MM) features, which are hypercalcemia, renal insufficiency, anemia, and/or bone disease with lytic lesions[9]. CRAB symptoms are caused by PCs proliferative processes.

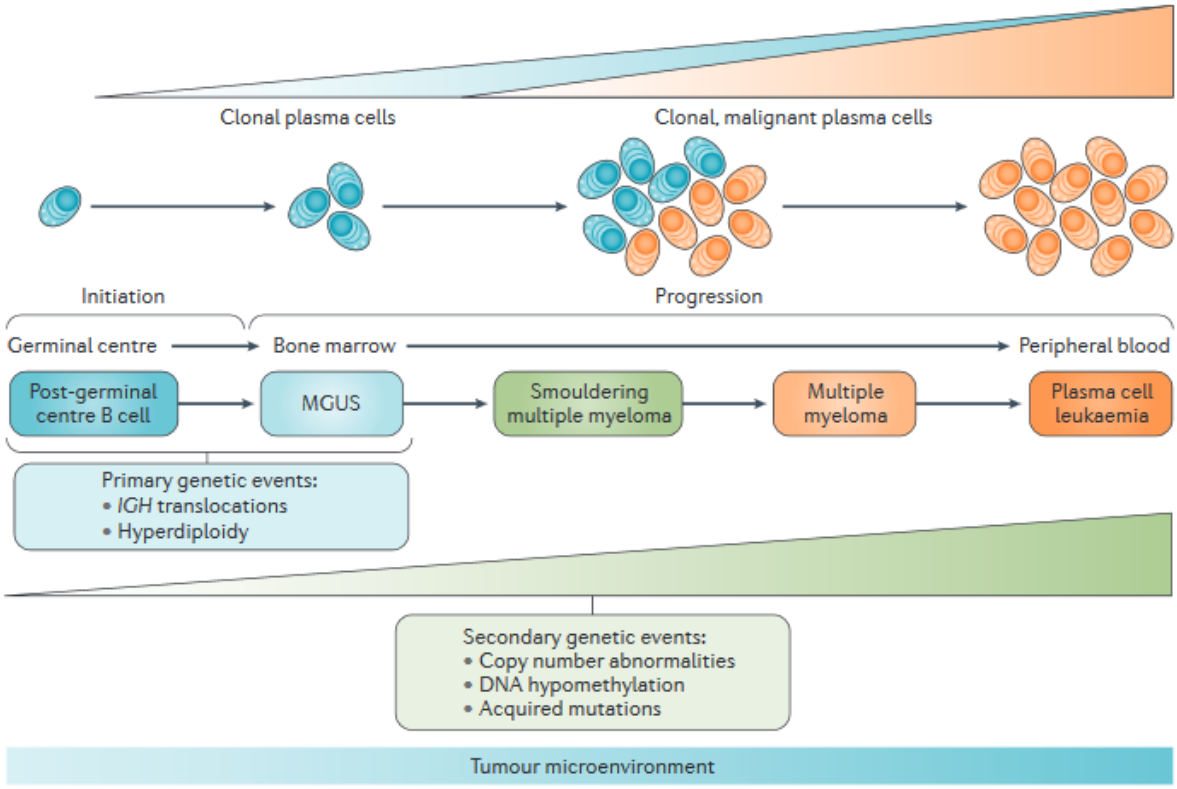



**Figure 3**: **multiple myeloma development[13].** Image from Kumar, S. K. et al. Nature reviews. Disease primers 3, (2017).

## 1.3 Genetic aberrations of MM

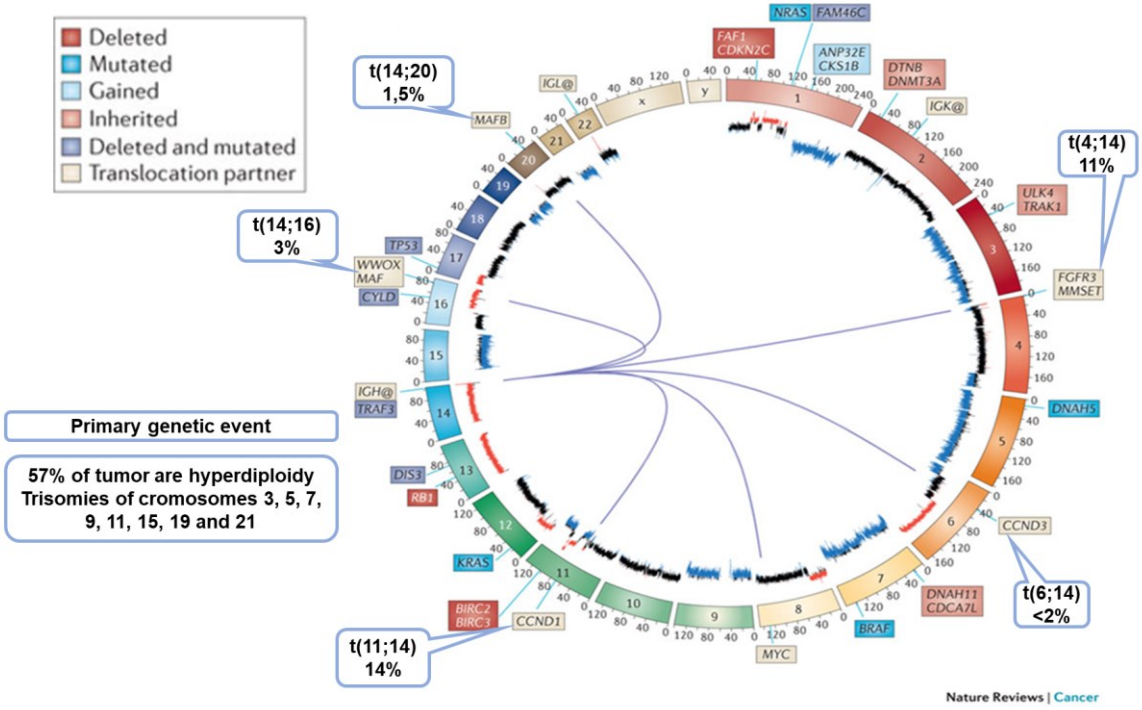



**Figure 3**: **circol plot of key translocations in MM.** Image modify by by Morgan, G. J. et al, Nature reviews. Cancer (2012).

MM is characterized by a large number of different chromosomal abnormalities. In particular, it is important to focus our interest on the following primary genetic events, often related to Ig heavy chain (IGH) enhancer, mapping on 14q32:

- 14% of MM patients present t(11;14)(q13;q32) translocation. This translocation causes the control of Cyclin D1 (CCND1) expression under immunoglobulin heavy chain enhancer[14]. Because of this, PCs overexpress CCND1 that is an important regulator of G1 to S phase progression[15].
- 11% of MM patients present t(4;14)(p16;q32) translocation. This translocation cause Nuclear Receptor Binding SET Domain Protein 2 (MMSET/WHSC1) and Fibroblast Growth Factor Receptor 3 (FGFR3) deregulation. This translocation is important in malignant transformation[16] and is a mark of cytogenetically high-risk disease.[17]
- 4.5% of MM patients present t(14;16)(q32;q23) or t(14;20)(q32;q11) with the deregulation of *c-MAF* (Avian Musculoaponeurotic Fibrosarcoma) and *MAFB* gene, respectively. The deregulation of *MAF* genes can change the homing and adhesive proprieties of the cells and in this way promotes the survival of PCs in the bone marrow[18] (MAF-trx).
- less than 1% of MM patients present t(6;14)(p21;q32) with modulation of Cyclin D3[19], an important protein of cell cycle.
- 57% of MM patients present the hyperdiploidy status (HD), characterized by the trisomy of chromosomes 3, 5, 7, 9, 11, 15, 19 and 21, that seems to confer a better prognosis[20].

The secondary genetic events are important in driving MGUS progression versus MM pathology.

- Deletion of chromosome 13 (del(13)(q14q21)) seems to act in the haploinsufficency of RB1 (Retinoblastoma 1) that is a negative regulator of cell cycle[21].
- Deletion of chromosome 17 (del(17p)) is considered an independent negative predictor for survival of patients because of the TP53 (Tumor Protein P53) involvement. Indeed 37% of del(17) MM present TP53 mutations[17,22,23].
- Gain of chromosome 1q is associated with poor prognosis[21,24]; the role of this gain of function is unknown but in this region is localized MCL1 (Myeloid Cell Leukemia Sequence 1), an important anti-apoptotic gene.
- RAS (Rat Sarcoma Viral Oncogene Homolog) and BRAF (B-Raf Proto-Oncogene, Serine/Threonine Kinase) are genes involved in RAS signaling cascade that regulate cell survival and proliferation, as well as cytoskeletal organization[25]. N-RAS (Neuroblastoma

RAS) and K-RAS (Kirsten RAS), overexpress RAD (Ras Related Glycolysis Inhibitor And Calcium Channel Regulator) gene and are associated with disease progression, aggressive phenotype, and shorter survival.[26,27] Literature on BRAF activating mutation (V600E) are very unclear of the relationship between this mutation and survival of MM patients[28]. BRAF mutation is very stable between diagnosis and relapse[29].

- *DIS3* (DIS3 Homolog, Exosome Endoribonuclease And 3'-5' Exoribonuclease) is a gene coding for a protein that is involved in RNA degradation and quality control pathways. DIS3 mutation was associated with IGH translocations and deletions of 13q14. The relationship between this mutation and MM prognosis are still unclear.[30,31]
- FAM46C (Family With Sequence Similarity 46, Member C), presented in 1p12, is often loss or mutated in MM patients but its role is unknown[32].

## 1.4 Diagnosis of MM

MM patients can be diagnosed on the base of symptomatic or asymptomatic features. In particular, patients with asymptomatic MM are characterized by the presence of hypercalcemia, anemia, or proteinuria in laboratory test[33]. Instead, symptomatic patients are characterized by specific symptoms like bone disease, renal pain, anemia or hyperviscosity like deep venous thrombosis, or by unspecific symptoms like nausea, vomiting, recurrent infections, weight loss and weakness[34].

Specifically, to the diagnosis of MM patients it is important to identify two criteria:

- clonal PCs have to be more than 60% in bone marrow
- one or more of end-organ damage have to be present and due to PCs proliferation and Igs secretion. These symptoms are hyprcalcemia with calcium serum higher than 1mg/dL; renal insufficiency with variation of serum creatinine and creatinine clearance, anemia with hemoglobin deregulation and bone lesion identifiable with skeletal radiography, CT, or positron emission tomography/CT.[35-38]

## 1.5 Treatment of MM

Up to the second part of XX century, MM had no therapeutic possibility. Indeed, in these years was described the use of autologous stem cell transplantation (ASCT) for MM patients[39-41]. This ASCT is possible for patients less than 70 years old and without cardiac, pulmonary and hepatic problems.

Chemotherapy agents used in clinical for MM treatment can be divided in several classes[42]:

- proteasome inhibitors: Bortezomib, Carfilzomib, and Ixazomib
- immunomodulatory drugs (IMiDs): Lenalidomide
- monoclonal antibodies (MoAbs): Daratumumab, targeting CD38 and Elotuzumab, directed against human MM cell surface glycoprotein receptor CS1/SLAMF7
- histone-deacetylase (HDAC) inhibitors: Vorinostat, Panobinostat, Ricolinostat.

The use of these chemotherapy agents in concomitant or in single use is able to improve the outcome of MM patients but MM remains an incurable disease.

# 2. *Long non coding RNA*

## 2.1 Identification, classification and structures

From 1977, the Sanger sequencing increased the knowledge of DNA sequence[43] and worldwide researchers investigated to understand what all DNA is coding for. Thanks to the new discovery, researchers describe genome as composed of protein-coding region and non-protein coding region. Figure 4 shows the timeline of discoveries of RNAs in biological regulation, and the role of some genome sequences.

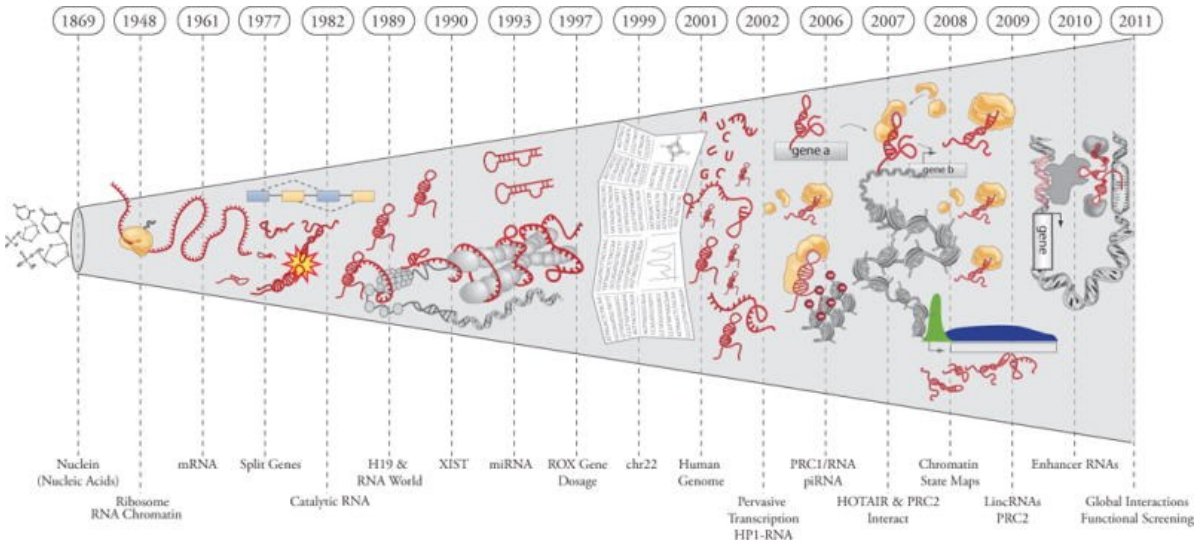



**Figure 4**: **timeline of discoveries of RNAs in biological regulation** by John L. Rinn and Howard Y. Chang, Annu Rev Biochem (2013)[44].

In this landscape, our interest is focused on non-coding RNAs (ncRNAs). The non-protein coding regions of genome transcribes for two different classes of ncRNA: short non-coding RNAs, molecules longer less than 200 nt, and long non-coding RNA (lncRNA), molecules longer that 200nt. In the class of short-ncRNA we can evidence molecules like microRNA (miRNA). From 1993, several group focused on these small molecules and evidenced their role in the regulation of gene expressions. LncRNAs are defined as RNAs longer than 200nt unable to code for proteins[44]. The first lncRNA identified, in 1991, was *XIST* ( X Inactive Specific Transcript) that was described as a RNA expressed only from inactive X chromosome, in nucleus, and unable to produce any kind of protein[45,46]. In these 27 years, and in particular in the last 10 years, the knowledge of lncRNAs rapidly increased. Indeed, in 2017, Ensembl described a total number of 58381 genes, of which 15779 lncRNAs and 19901 protein coding genes (Version 28 (November 2017 freeze, GRCh38) - Ensembl 92, 93, Figure 5). In 2012,

Kelly D. and Rinn j. defined lncRNAs as less conserved in mammalian species than protein coding genes because of the presence of transposable elements in their sequences[47]. In addition, the last class of ncRNA are pseudogenes, defined as a copy of a DNA coding sequences that have lost some characteristics, as the coding protein capability, and could accumulate mutations.

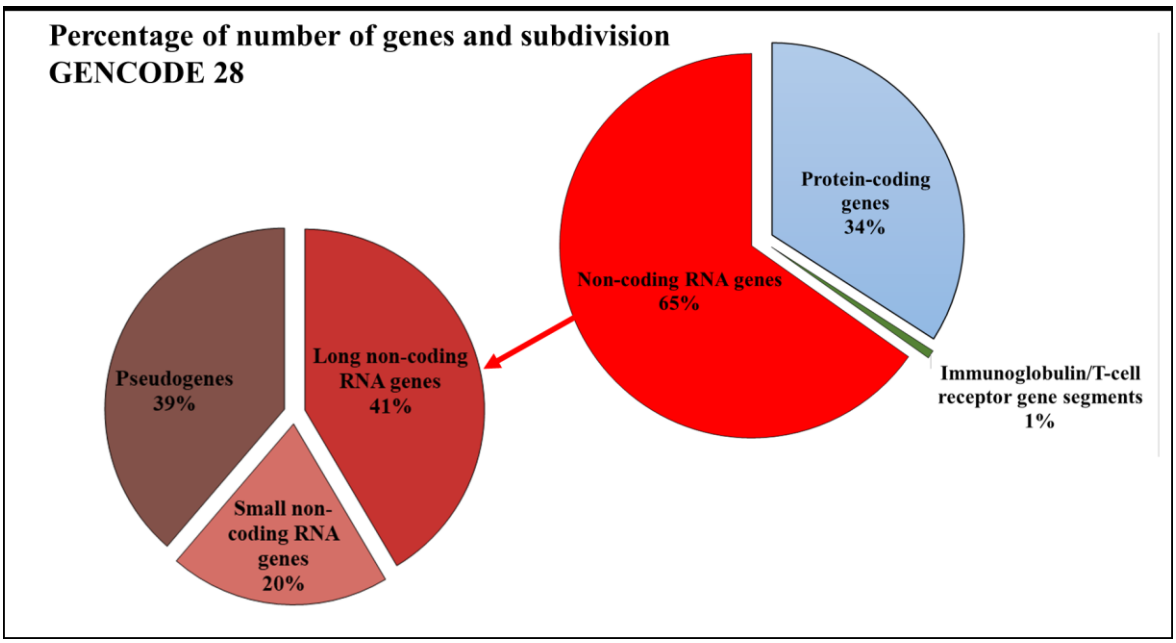



**Figure 5**: **statistics about all Human GENCODE releases.** Pie chart of the percentage of number of genes defined by GENCODE annotation version 27. Data from https://www.gencodegenes.org/stats/archive.html#a28

The focus of this thesis is on lncRNAs. Based on their genomic position, lncRNAs can be classified in broad categories. In particular, we can divided lncRNAs in[48-51](Figure 6):

- intronic lncRNA, transcribed by the intron region of a protein-coding gene, in sense or antisense direction;
- intergenic lncRNA, transcribed by a region between two protein-coding genes;
- bidirectional lncRNA, when the start transcription point of lncRNA is near to that of a protein-coding gene;
- sense lncRNA when the lncRNA transcript region is in the same strand of a protein coding gene exon;
- antisense lncRNA when the lncRNA transcript region is in the opposite strand of a protein coding gene exon.

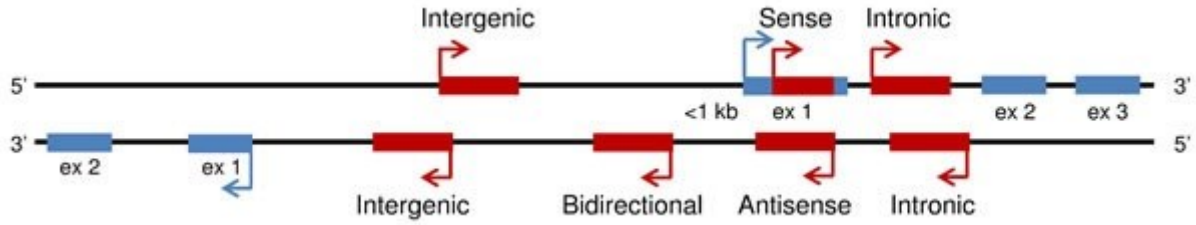



**Figure 6**: **classification of lncRNAs.** lncRNAs categories (sense, antisense, bidirectional, intronic, intergenic; depicted in red) based on their location relative to nearby protein-coding genes (depicted in blue)[51].

After the transcription, by RNA polymerase II and in few cases by RNA polymerase III[52], lncRNAs can undergo polyadenylation and splicing. The mature lncRNAs can interact with RNA, protein or DNA, thanks to their sequences and/or their secondary and tertiary structures[53,54] to produce biological effects.

## 2.2 Functional roles

Very important for the identification of functional roles of lncRNAs are the subcellular localization. Indeed, lncRNAs could be localized in nuclear or cytoplasmic fraction and, based on this localization, could have different roles[55]. In this list, we evidenced different roles of lncRNAs and some examples in the nucleus:

- *Scaffold for chromatin modifying complexes*. Prensner and Chinnaiyan in 2011 evidenced the capability of *ANRIL* (Antisense Noncoding RNA In The INK4 Locus) to interact with PRC1 and PRC2 (Protein Regulator Of Cytokinesis 1/2), two repressive histone modifier complexes for the repression of INK4A and INK4B (Cyclin Dependent Kinase Inhibitor 2A/B, p16 and p15) expression[56,57].
- *Controlled recruitment of transcription factor*. *PANDA* (Promoter Of CDKN1A Antisense DNA Damage Activated RNA) is a lncRNA able to sequester transcription factor NF-YA (Nuclear Transcription Factor Y Subunit Alpha) and inhibit the expression of proliferation genes under the control of NF-YA[58].
- *Enhancer RNAs* (eRNAs), transcribed by active enhancers and with high levels of histone H3 lysine 4 mono-methylate in genomic locus, facilitate enhancer–promoter interaction or activate promoter-driven transcription[59].
- *Splice modulator. MALAT1* (Metastasis Associated Lung Adenocarcinoma Transcript 1) is able to interact with the serine/arginine-rich splicing regulatory factors and modulate alternative splicing of pre-mRNA in nuclear speckle domains[60].

- *Protein sub-nuclear sequestration*. *NEAT1* (Nuclear Paraspeckle Assembly Transcript 1) is able to sequestrate the protein NONO (Non-POU Domain Containing Octamer Binding) and SFPQ (Splicing Factor Proline And Glutamine Rich) for the elongation of paraspeckles[61].
- *Spatial conformation of chromosomes*. To allow the interaction between chromosomal *loci* like Hox *loci*, *HOTTIP* (HOXA Distal Transcript Antisense RNA) is able to change the 3D structure of the chromosomes[62].

LncRNAs are able to act in cytoplasm fraction, particularly in post-transcription events.

- *Control of mRNA decay*. ½-sbsRNAs are a class of lncRNAs with a lot of ALU sequences inside, that activate the decay of specific target mRNAs[63], while Faghihi et al. described *BACE1-AS* (Beta-Secretase 1 Antisense RNA), a lncRNA able to decrease the RNA degradation of BACE1 (Beta-Secretase 1) in neuronal cells[64].
- *microRNA (miRNA) sponge*. *CCAT1* (Colon Cancer Associated Transcript 1) is described, in hepatocellular carcinoma, as a let-7 sponge to inhibit its interaction with HMGA2 (High Mobility Group AT-Hook 2) and c-Myc[65].
- *Control of translation*. *lincRNA-p21*, when the concentration of *HuR* (ELAV Like RNA Binding Protein 1) is low, is able to interact with CTNNB1 (Catenin Beta 1) and JUNB (JunB Proto-Oncogene, AP-1 Transcription Factor Subunit) mRNAs and represses their translation into proteins[66].

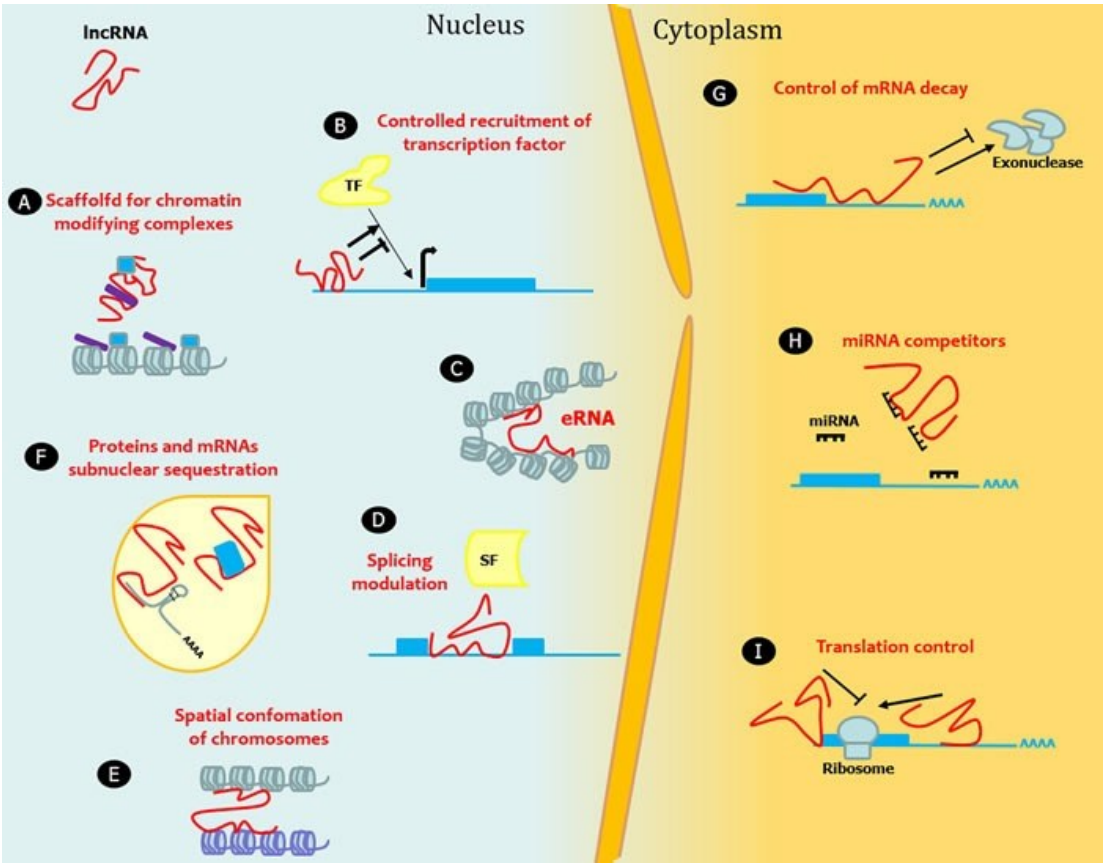



**Figure 7**: **roles of lncRNAs.** In this figure are reported 9 role of lncRNAs. In particular, in the left part of the figure is described the nucleus compound in which lncRNAs could act as modifier of chromatine scaffold (A), activator or repressor of transcription factor (B), enhancer RNA(C), modulator of splicing events (D), modulation of spatial conformation of chromosomes (E), and sequester of protein and mRNA (F). In the right part of the figure is evidenced cytoplasmic roles of lncRNAs. In particolular lncRNA in cytoplasm could act in the control of RNA decay(G), as a miRNA sponge (H) and in the control of mRNA translation (I)[67].

## 2.3 LncRNAs and cancer

Hanahan and Weinberg described the hallmarks of cancer for the first time in 2000. They indicated that cell could undergo several modifying essential processes and progressively go towards a malignant transformation. In particular, they indicated six processes that are: sustaining proliferative signalling, evading growth suppressors, enabling replicative immortality, activating invasion and metastasis, inducing angiogenesis and resisting cell death[68]. Additionally, in 2011 they added to these six processes other four that are: deregulating cellular energetics, avoiding immune destruction, genome instability and mutation, and tumor-promoting inflammatory[69].

Based on previously reported description of lncRNAs roles (paragraph 2.2), it is clear that lncRNAs could play important roles in malignant transformation. Indeed, in 2012, Gutschner

and Diederichs illuminated scientific community about roles of lncRNA in cancer transformation[70].

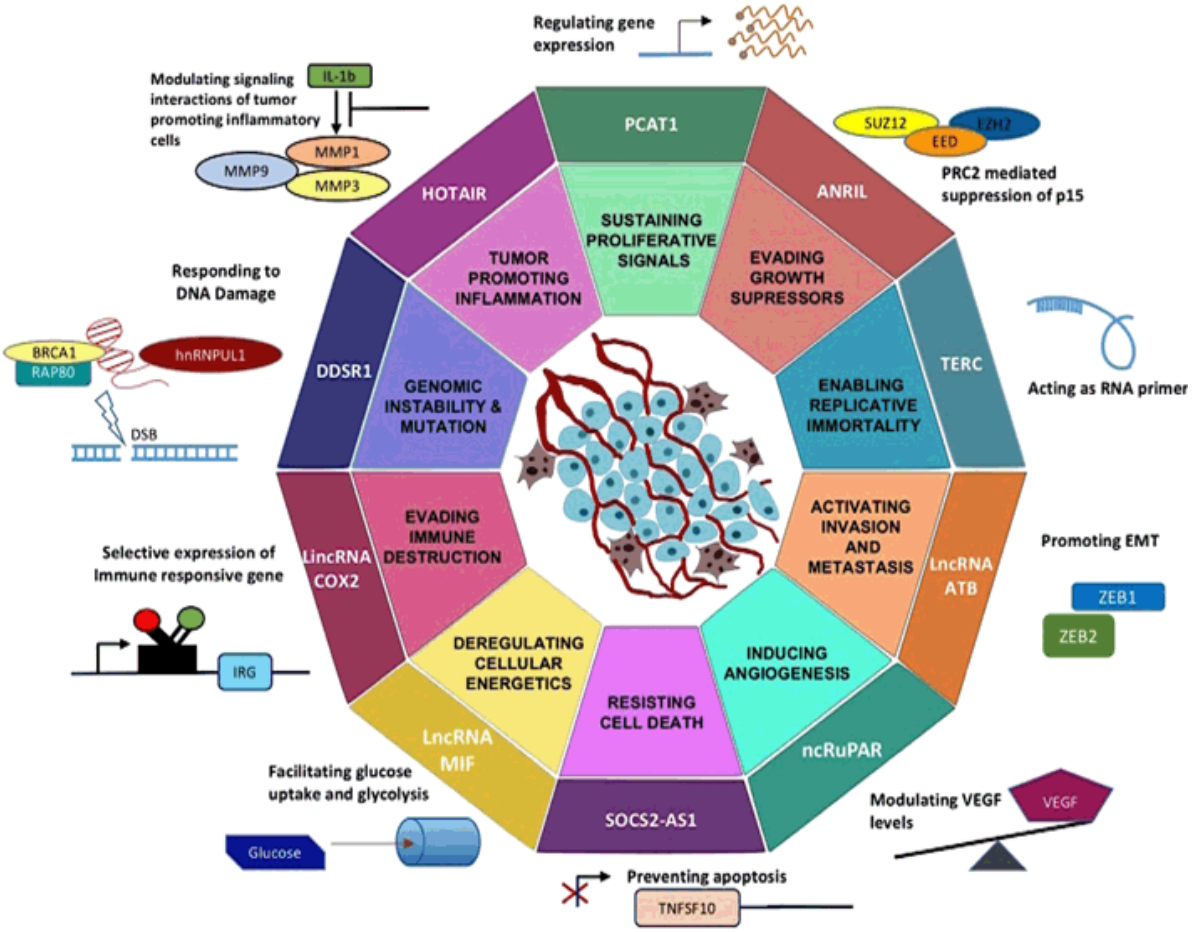



**Figure 8**: **lncRNAs in hallmarks of cancer.** In this figure are reported examples of lncRNAs that act in malignant transformation[71].

For instance, *PCAT1* (Prostate Cancer Associated Transcript 1) is a lncRNA up-modulated in prostate cancer and thanks to the expression regulation of the tumor suppressor BRACA1 (BRCA1-Interacting Protein 1) and oncoprotein cMYC, is capable to *modulate sustaining proliferative signals*[72].

*ANRIL*, a lncRNA described previously, thanks to the suppression of p15 and p16 makes cell able to *evading growth suppressors*[56]. *ANRIL* has a role in several kind of tumor like gastric and breast cancer [73,74].

An example of lncRNAs with a role in *enabling replicative immortality* is *TERC* (Telomerase RNA Component). *TERC* is the telomerase RNA component important for the length of telomeres because of it is use as template[75].

A lncRNA important in *invasion and metastasis is lncRNA-ATB*. In hepatocellular carcinoma it induced invasion with the up-modulation of ZEB1 and ZEB2 (Zinc Finger E-Box Binding Homeobox 1/2)[76].

The expression levels of PAR1 (Serine/threonine-protein kinase par-1) and VEGF (Vascular Endothelial Growth Factor) in gastric cancer are very important for *angiogenetic induction*. In 2014, Liu et al. evidenced *ncRuPAR* (Non-Protein Coding RNA, Upstream Of F2R/PAR1), a lncRNA negatively correlated with the expression level of PAR1 and VEGF and showed its important role in angiogenetic events[77].

In prostate cancer cells, Misawa A. et al. evidenced *SOCS-AS1* ( Suppressor Of Cytokine Signaling-antisense1) as an actor of *apoptosis repression*[78].

*lncRNA-MIF* (lncRNA- Macrophage Migration Inhibitory Factor), a lncRNA transcribed by c-MYC, is important in a loop in which it is able to suppress c-Myc and *miR-586* and inhibit the aerobic glycolysis. This lncRNA act in the *deregulation of energetic cellular processes*[79].

Fitsgerald KA.'s laboratory, in 2013 describer *lincRNA-Cox2* (lincRNA- Mitochondrially Encoded Cytochrome C Oxidase II). *lincRNA-Cox2* works in immune genes expression to activate or repress them[80]. This lncRNA could have a role in *tumor evading immune system*.

For the *genetic instability and mutation* an important lncRNA is *DDRS1* (Discoidin Domain Receptors 1) that is expressed after DNA damage. The loss of this lncRNA is correlated to reduction of DNA repair[81].

*Tumor promoting inflammation* hallmark is represented by *HOTAIR* (HOX Transcript Antisense RNA) because of its capability to inhibit IL-1b, an important factor expressed in inflammatory cells[82].

## 2.4 SNPs and lncRNAs

One of the hallmarks of cancer is genomic instability and mutation. In particular, we want to focus on Single Nucleotide polymorphisms (SNPs), alterations in a DNA nucleotide that have a role in the gene sequences. Indeed, SNPs could produce different events like stop of transcription, variation of splicing events, change of amino acid in protein sequences (nonsynonymous SNPs) or un-produce any kind of effect (silent SNPs).

It is important to underline that non-coding genes, as coding genes, are affected by these kind of SNPs. Indeed, in 2015, Verma A. and colleagues evidenced the presence of several mutation in exonic regions of lncRNAs in B cell lymphoma. Furthermore, they defined these mutations as stabilizer for lncRNA structures[83]. In Table1 are reported some examples of SNPs variation with a role in association with different cancer types.

| SNP | lncRNA | Function | Associated cancers |
|---|---|---|---|
| rs11655237 | LINC00673 | Creates binding site for miR-1231 causing reduced activity of LINC00673 | Pancreatic cancer |
| rs7763881 | HULC | The variant decreases the risk of HBV related HCC | Hepatocellular carcinoma |
| rs619586 | MALAT1 | Protective for HCC risk in non-alcoholics | Hepatocellular carcinoma |
| rs7463708 | PCAT1 | Increases binding affinity of ONECUT2 and androgen receptor (AR) leading an enhanced PCAT1 | Prostate cancer |
| rs6434568, rs16834898 | PCGEM1 | Variant increases susceptiblity but molecular mechanism needs to be investigated | Prostate cancer |
| rs944289 | PTCSC3 | Diminishes binding of p42 C/EBPα and C/EBPβ reducing PTCSC3 transcriptional levels | Papillary thyroid carcinoma |
| rs2157719 | ANRIL | An intronic SNP found to correlate with increased expression of ANRIL | Esophageal cancer |
| rs564398 | ANRIL | Variant is linked to ANRIL overexpression | Acute lymphoblastic leukemia |
| rs1063192, rs1011970 | ANRIL | The risk variant is linked to allelic expression of ANRIL | Melanoma |
| rs2839698 | H19 | Identified with increased cancer risk involving H19 expression | Colorectal and gastric cancer |
| rs145204276 | GAS5 | Determines methylation status of GAS5 promoter | Hepatocellular carcinoma, colorectal cancer |

**Table 1: SNP variation in lncRNAs.** Examples of important SNP variation affected lncRNAs. Table modify by Rao A. et al. Molecular biology reports, 2017 [71].

## 2.5 LncRNAs in multiple myeloma

The first identification of lncRNAs roles in MM was published by Benetatos L. et al. in 2008. They evidenced a modulation of the methylation of MEG3 promoter and the methylation status was described as correlated with disease stage and subtype[84]. To date, the number of lncRNAs deregulated in MM is increased. Ronchetti et al. in 2015, made a gene expression profile (GEP) analysis on a cohort of 268 patients with PC dyscrasias. Their analysis, made using GeneChip® Human Gene 1.0 ST array (Affymetrix INC. Santa Clara, CA) was able to investigate 1852 lncRNAs. Their results showed a list of 21 lncRNAs whose expression levels were significant correlated with the progression of the disease. Furthermore, 31 lncRNAs resulted specifically deregulated in pathological samples compared to normal controls[85].

One of the most interesting lncRNA deregulated in MM, in Ronchetti et al. 2015, is *MALAT1*. Amodio N. et al. in collaboration with Neri A.'s laboratory focused his attention on *MALAT1*. This lncRNA has a highly conserved DNA sequence and is one of the most studied lncRNAs in scientific community. The expression levels of *MALAT1* resulted deregulated in a wide variety of haematological malignancies and solid tumors[86]. Amodio N. et colleagues silenced *MALAT1* expression in MM cell lines and identified an important roles of this lncRNA in proteasome machinery regulation and focused the attention of scientific community on the possibility of treating MM patients with specific molecules directed against *MALAT1*[87].

Other interesting lncRNAs modulated in MM are *CRNDE* and *UCA1*. *CRNDE* (Colorectal Neoplasia Differentially Expressed) expression resulted associated with poor overall survival of MM patients and, *in vitro*, its silencing resulted correlate with a modulation of cellular proliferation and colony formation[88]. The expression level of *UCA1* (Urothelial Cancer Associated 1) in MM patients compared to healthy donors resulted down-regulated[89].

In 2018, Shen et al., investigated five patients with MM and five Iron deficiency anemia patients, as control, with Arraystar Human lncRNA Microarray V4.0 (KangChen Biotech) for GEP analysis. This Microarray is able to investigate 40173 lncRNA transcripts and 20730 coding transcripts. Despite the few number of cases, they evidenced *ST3GAL6-AS1* (*ST3GAL6* Antisense RNA 1), ZNF663P (Zinc Finger Protein 663, Pseudogene), LAMA5-AS1(Laminin Subunit Alpha 5 Antisense RNA 1) and RP11-175D17.3 significantly up-modulated in their MM samples compared to controls. In particular, they focused on *ST3GAL6-AS1* that is the higher expressed lncRNA of the four modulated lncRNAs and they hypothesized a relationship between *ST3GAL6-AS1* and its neighbor gene *ST3GAL6* (Sialyltransferase 3 Beta-Galactoside Alpha-2,3-Sialyltransferase 6 )[90].

# 3. *ST3GAL6 and DCBLD2*

*ST3GAL6* and *DCBLD2* (Discoidin, CUB and LCCL Domain Containing 2), both protein coding genes, and *ST3GAL6-AS1*, a non–coding gene, map on 3q12.1. In particular, *ST3GAL6* maps on sense strand while *DCBLD2* and *ST3GAL6-AS1* in antisense strand.

*ST3GAL6* transcribes for 12 protein coding RNA full length from 489 to 3571nt (from 121 to 331 aa), 4 nonsense mediated decay full length from 685 to 3148nt (all of 90aa), 9 non-coding processed transcripts full length from 314 to 1102nt and two transcript with retained intron of 582 and 804bp (GENCODE annotation).

ST3GAL6 is a sialyltransferase belonging to the glycosyltransferase superfamily. Four distinct families, based on the combination of the acceptor saccharide residue and the precise hydroxyl group on this residue, identify Glycosyltransferase superfamily[91-93]. These enzymes, with other enzymes like Fucosyltransferases (FUT3-7), are involved in the formation of sialyl-Lewis$^x$ (sLe$^x$) and sialyl-Lewis$^a$ (sLe$^a$) that represent the minimal recognition motif for selectin ligands. Selectin ligands are important molecules for the interactions of cells with all components of microenvironment. Modulation of the cellular expression of these residues in cancer could alter processes such as adhesion, mobilization and migration[94]. sLe$^x$ and sLe$^a$ presence, in concomitance with high expression levels of sialyltransferases, are associated with an increase in the risk of metastatic disease in solid tumors[95,96]. In particular, ST3GAL6 enzymes catalyse the transfer of sialic acid to 3'hydroxyl group of a Gal residue of proteins and lipids. ST3GAL6 results highly expressed in placenta, liver, heart and skeletal muscle[97]. In gastric carcinoma *ST3GAL6* resulted hyper-methylated in concurrence with *B4GALNT2* (Beta-1,4-N-Acetyl-Galactosaminyltransferase 2), another important genes for the process of glycosylation[98]. Milde-Langosh K. et al. in 2014 investigated the prognostic relevance of glycosylation-associated genes in breast cancer. They claimed that the expression levels of *ST3GAL6* significantly correlated with the formation of distant metastasis while FUCA1 (Alpha-L-Fucosidase 1), a lysosomal enzymes able to remove terminal l-Fucose residues in glycoproteins[99], was significantly inversely correlated with distant metastasis[100].

Glavey SV. et al. in 2015 described the higher expression levels of *STGAL6* in MM patients and its significant correlation with lower overall survival. The knockdown of *ST3GAL6* expression levels in MM cell lines evidenced no biological effect in terms of cell viability and growth and a decrease of sialyl residues on MM cells surface with inferior adhesion and

migration capability. Their *in vivo* experiments on mice, with the down-modulation of *ST3GAL6*, demonstrated an important role of this gene in homing and engraftment of MM.

*DCBLD2* transcribes for 3 protein coding RNA full length from 554 to 6122nt (from 108 to 775 aa), 2 non-coding processed transcript full length of 582 and 584nt, and 5 transcript with retained intron of 418 and 4435bp (GENCODE annotation).

Identified for the first time as up-regulated in vascular injury, *DCBLD2* is defined as a neurophilin-like membrane protein[101]. Guo X. et al. showed that *DCBLD2* plays a part in PDGFR- β stimulation inhibition by ubiquitination of PDGFR-β through c-CBL E3 ligase[102]. It seems to be associated with acquisition of a metastatic phenotype in lung cancer[103]. Moreover, DCBLD2 was identified by proteomic studies of EGFR/EGFRvIII stimulation of various types of cancer cells as a phosphorylated protein at several tyrosine residues[104], suggesting a potential involvement of DCBLD2 in EGFR stimulation of cancer cell behaviour. In 2017, Feng H. et al. described DCBLD2 as an actor in enhancing the oncogenic activity of the EGFR/AKT pathway in human cancers[105].

# Aim of the study

To date, MM is still defined as an incurable cancer disease that represents 10% of hematological malignancies in United States. Recent study on MM defined the gene expression modulation of important coding and non-coding genes. In particular, in the last 10 years, scientific community focused its attention on lncRNAs. These molecules are defined as able to act in different ways and are deregulated in tumors. Several studies evidenced the deregulation of lncRNAs with important role in the progression of MM.

Our hypothesis is that lncRNAs could have an essential role in the progression and maintaining of MM status.

The aim of this thesis is to identify a lncRNA up-modulated in a large cohort of MM patients, characterize its proprieties and sequences. Furthermore, we want to describe its role in MM cell lines and identify some molecules able to down-modulate this lncRNA in *in vitro* experiments to make out possible new functional therapy for MM.

# Material and methods

# 1. *MM and PCL samples*

PCs were purified, as described in Mattioli et al. 2005[106], from 115 samples, of which 94 MM and 21 PCL, obtaining more than 90% enrichment in all the cases. All samples were characterized for the presence of the most frequent chromosomal translocations and the ploidy status based on fluorescence in situ hybridization (FISH) evaluation criteria as already reported[107], and by next-generation sequencing analysis[108-111] for the presence of several mutations important in MM and PCL pathology. The characteristics of the patients are reported in Table 2.

| sample features | Positive | | Negative | | not available | |
|---|---|---|---|---|---|---|
| | n | % | n | % | n | % |
| HD | 35 | 30.4 | 68 | 59.1 | 12 | 10.4 |
| t(11;14) | 28 | 24.3 | 86 | 74.8 | 1 | 0.9 |
| t(6;14) | 4 | 3.5 | 108 | 93.9 | 3 | 2.6 |
| t(4;14) | 19 | 16.5 | 96 | 83.5 | 0 | 0.0 |
| MAF-trx | 11 | 9.6 | 103 | 89.6 | 1 | 0.9 |
| del(17) | 13 | 11.3 | 101 | 87.8 | 1 | 0.9 |
| del(13) | 61 | 53.0 | 54 | 47.0 | 0 | 0.0 |
| 1q-gain | 53 | 46.1 | 55 | 47.8 | 7 | 6.1 |
| N-RAS | 24 | 20.9 | 77 | 67.0 | 14 | 12.2 |
| K-RAS | 29 | 25.2 | 72 | 62.6 | 14 | 12.2 |
| BRAF | 10 | 8.7 | 91 | 79.1 | 14 | 12.2 |
| DIS3 | 20 | 17.4 | 81 | 70.4 | 14 | 12.2 |
| P53 | 6 | 5.2 | 94 | 81.7 | 15 | 13.0 |
| FAM46C | 10 | 8.7 | 89 | 77.4 | 16 | 13.9 |

**Table 2: molecular characteristics of 115 samples (94 MM and 21 PCL).** The table shows the molecular characteristics of our cohort of patients. In particular, 28 samples showed t(11;14) translocation with consequent CCND1 overexpression, 4 had t(6;14) and CCND3 up-modulations, 19 were with t(4;14) and CCND2 overexpression, and 11 with MAF genes translocation (t(14;16) and t(14;20) with the highest levels of CCND2 (MAF-trx). 35 samples were hyperdiploid (HD).

In accordance with the declaration of Helsinki, all the patients signed a written informed consent. The study was approved by the Ethical Committee of the University of Milan (N°24/15, May 06 2015).

# 2. *Gene expression analysis (GEP)*

Total RNA samples were obtained from purified PCs of 50 MM, 15 PCL, and 4 normal controls (Voden, Medical Instrument IT) and analyzed by GeneChip® Human Gene 2.0 ST array

(Affymetrix INC. Santa Clara, CA). Robust Multi Array Average (RMA) procedure allowed us to calculate normalized expression values.

| sample features | Positive | | Negative | | not available | |
|---|---|---|---|---|---|---|
| | n | % | n | % | n | % |
| HD | 17 | 26.2 | 41 | 63.1 | 7 | 10.8 |
| t(11;14) | 15 | 23.1 | 50 | 76.9 | 0 | 0.0 |
| t(6;14) | 0 | 0.0 | 65 | 100.0 | 0 | 0.0 |
| t(4;14) | 15 | 23.1 | 50 | 76.9 | 0 | 0.0 |
| MAF-trx | 11 | 16.9 | 54 | 83.1 | 0 | 0.0 |
| del(17) | 8 | 12.3 | 56 | 86.2 | 1 | 1.5 |
| del(13) | 36 | 55.4 | 29 | 44.6 | 0 | 0.0 |
| 1q-gain | 33 | 50.8 | 27 | 41.5 | 5 | 7.7 |
| N-RAS | 10 | 15.4 | 44 | 67.7 | 11 | 16.9 |
| K-RAS | 13 | 20.0 | 41 | 63.1 | 11 | 16.9 |
| BRAF | 8 | 12.3 | 46 | 70.8 | 11 | 16.9 |
| DIS3 | 12 | 18.5 | 41 | 63.1 | 12 | 18.5 |
| P53 | 5 | 7.7 | 48 | 73.8 | 12 | 18.5 |
| FAM46C | 7 | 10.8 | 45 | 69.2 | 13 | 20.0 |

**Table 3: molecular characteristics of 65 samples (50 MM and 15 PCL).** The table shows the molecular characteristics of our cohort of patients. In particular 28 samples showed t(11;14) translocation with consequent CCND1 overexpression, 4 with t(6;14) and CCND3 up-modulations, 19 with t(4;14) and CCND2 overexpression, and 11 with MAF genes translocation (t(14;16) and t(14;20) with the highest levels of CCND2 (MAF-trx). 35 samples were hyperdiploid (HD).

With the aim to take away probes that map to regions where ambiguous detection due to transcript overlap might occur, we set up a custom annotation pipeline able to combine GENCODE v25 (Ensembl v87) annotations with the CDF (Chip Definition File) version 21 for gene annotations freely available at

http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/21.0.0/GENCODEg.asp.

We achieved the expression levels of Ensembl genes specific for 10138 unique lncRNAs. All the data have been deposited in the NCBI Gene Expression Omnibus database (GEO; http://www.ncbi.nlm.nih.gov/geo) and are accessible under accession #GSE109116.[112]

Supervised analyses were carried out using the Significant Analysis of Microarrays software version 5.00 [113] using the web application provided in the shiny package of the R software (https://github.com/MikeJSeo/SAM). The cutoff point for statistical significance (at a q-value 0) was determined by tuning the Δ parameter on the false discovery rate and controlling the q-

value of the selected probes. Hierarchical agglomerative clustering of patients based on the most significant probesets found was performed adopting Pearson's correlation and average as distance and linkage methods, respectively.

We analyzed RNA samples from experiments with the same method.

## 3. *RNAseq analysis*

As explained in Ronchetti et al. 2018, 400 ng of total RNA from 30 MM samples, analyzed by GeneChip® Human Gene 2.0 ST array too, were used to prepare paired-end (PE) cDNA libraries using TruSeq® sample preparation kit (Illumina)[112].

| sample features | Positive | | Negative | | not available | |
|---|---|---|---|---|---|---|
| | n | % | n | % | n | % |
| HD | 8 | 26.7 | 20 | 66.7 | 2 | 6.7 |
| t(11;14) | 8 | 26.7 | 22 | 73.3 | 0 | 0.0 |
| t(6;14) | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 |
| t(4;14) | 7 | 23.3 | 23 | 76.7 | 0 | 0.0 |
| MAF-trx | 4 | 13.3 | 26 | 86.7 | 0 | 0.0 |
| del(17) | 3 | 10.0 | 27 | 90.0 | 0 | 0.0 |
| del(13) | 18 | 60.0 | 12 | 40.0 | 0 | 0.0 |
| 1q-gain | 15 | 50.0 | 13 | 43.3 | 2 | 6.7 |
| N-RAS | 3 | 10.0 | 20 | 66.7 | 7 | 23.3 |
| K-RAS | 7 | 23.3 | 16 | 53.3 | 7 | 23.3 |
| BRAF | 4 | 13.3 | 19 | 63.3 | 7 | 23.3 |
| DIS3 | 6 | 20.0 | 17 | 56.7 | 7 | 23.3 |
| P53 | 2 | 6.7 | 20 | 66.7 | 8 | 26.7 |
| FAM46C | 1 | 3.3 | 21 | 70.0 | 8 | 26.7 |

**Table 4: molecular characteristics of 30 MM samples.** The table shows the molecular characteristics of our cohort of patients. In particular 8 samples showed t(11;14) translocation with consequent CCND1 overexpression, 7 were with t(4;14) and CCND2 overexpression, and 4 with MAF genes translocation (t(14;16) and t(14;20) with the highest levels of CCND2 (MAF-trx). 8 samples were hyperdiploid (HD).

We obtained 100 bp PE reads strand-specific by sequencing of the libraries on HiSeq.2000 (Illumina). STAR aligner, based on splice junction from Ensembl database version 87, and GENCODE v25 GTF file were used to aligned reads at human genome. FeatureCounts (default parameters) for the estimation of transcript abundance and cufflinks default procedure for Fragments Per Kilobase Milion (FPKM) quantification on sorted BAM files were used. Differentially expressed genes were identified using DeSeq at FDR < 0.01, provided that expression across the whole dataset was not null. *Multiqc* tool and the QC metrics were used

for Quality control (QC) analysis on all samples. With this method, we are able to detect 14,202 lncRNAs but the expression filter retained 9,540 lncRNAs in our dataset.

## 4. *Cell line cultures*

KMS12, AMO1, KMS26, RPMI-8226, DELTA47, U266, JJN3, INA6, KMS34, KMS11, KMS18, OPM2, CMA-03, UTMC2, MM1.S, MM1.44, LP1, H929, CMA-03/06, KMS20, MEC1, U937, MOLT4, JURKAT, MAVER, MINO cell lines were cultured in RPMI-1640 (Gibco, Thermo Fisher) with 10% FBS (Gibco, Thermo Fisher) and 2 mM L-glutamine (Invitrogen Corporation Carlsbad, CA, USA). Cells were cultured in 5% CO2 at 37°C, maintaining the optimum concentration at $5x10^5$ cells/ml with a complete change of medium every 72 hours after counting cells in Burker's chamber using Trypan Blue Solution 1:1(Thermo Fisher). CMA-03 cells were cultured in presence of IL-6 10 ng/ml.

HeLa, Hs5, and HEK-293T cell lines were cultured in Dulbecco's modified Eagle's medium (DMEM) with 10% FBS (Gibco, USA), 2 mM L-glutamine (Invitrogen Corporation Carlsbad, CA, USA). Having a doubling time of 48-72h, the cells were splitted twice a week.

## 5. *Total RNA extraction*

$5–10 \times 10^6$ cells were collected and centrifuged at 250 rcf 10 min at 4°C; medium was removed and the pellet was suspended in 1 ml of TRIzol™ Reagent (Thermo Fisher). RNA was extracted following the protocol supplied by the manufacturer.

5 breast cancer cell lines (SKBR3, MCF7, SUM-149, MDA-MB-468, MDA-MB-231) and 5 thyroid carcinoma cell lines (BCPAP, NHY-OKI, TPC1, WRO, HTCJC3) RNA were obtained from laboratory of Doctor Tagliabue and Doctor Greco from Istituto Nazionale dei Tumori di Milano, respectively.

## 6. *DNA extraction*

$5x10^6$ cells were collected and centrifuged at 900 x g 5 min at RT; medium was removed and the pellet was washed in 0.5 ML of PBS. Samples were centrifuged at 250 rcf 10 min at 4°C; medium was removed and the pellet could be freezed at -80°C. DNA extraction was performed

using Wizard® Genomic DNA Purification Kit following the protocol supplied by the manufacturer.

## 7. *DNA and RNA quantification*

Nanodrop (spectrophotometric measure, Thermo Fisher) was used to quantify RNA and DNA. We used 1.5 µl of each sample. Absorbance was read at two different wavelengths, at 260 nm (A1) and 280 nm (A2). High quality RNA was used (A1/A2 ratio closed to 1.9-2).

## 8. *Reverse transcription*

cDNA was obtained by reverse transcription of 200 ng of RNA with RevertAid M-MuLV Reverse Transcriptase (ThermoScientific). The reaction was prepared in 20 µl using Random Primers following manufacturer's instructions.

## 9. *Quantitative RT-PCR*

Quantitative Real-Time PCR (RT-PCR) reactions were carried out on a Step-One Plus PCR system (Applied Biosystems, Life Technologies Italia, Italy) using the Maxima SYBR Green/ROX qPCR Master Mix (2X) (ThermoFisher). The reaction, in a final volume of 20 µl in a 48-well plate, was made as described in the protocol supplied by the manufacturer.

Each sample was analyzed in triplicate with no template controls. Results were calculated with $2^{-\Delta\Delta Ct}$ method and the copy numbers of the analyzed mRNA were normalized using GAPDH or 18S mRNA levels. Primer sequences used for cDNA amplification are shown in Table 5.

| **RT-qPCR primers** | **Forward Primer 5'-3'** | **Reverse Primer 5'-3'** |
|---|---|---|
| GAPDH | TTGGCCGTATTGGGCGCCTG | CACCCTTCAAGTGGGCCCCG |
| 18S | GTAACCCGTTGAACCCCATT | CCATCCAATCGGTAGTAGCG |
| *ST3GAL6-AS1* | GCAGCACAGAATCCTGACAA | GCCAGCATTTTGGTAAGAGC |
| *ST3GAL6* | TTGCCTCTCTGCTGAGGTTT | CCTCCATTACCAACCACCAC |
| *DCBLD2* | CTCAGCCACTGGTAGGAGGA | GGCACCTGGTACACCAATTC |

**Table 5: primer sequences for Real-time PCR**

RT PCR for known genes were carried out using TaqMan assay using TaqMan™ Fast Universal PCR Master Mix (2X), no AmpErase™ UNG (Thermo Fisher). The reaction, in a final volume

of 20µl in a 48-well plate, was made as described in the protocol supplied by the manufacturer. Each sample was analyzed in triplicate with no template controls. Results were calculated with $2^{-\Delta\Delta Ct}$ method and the copy numbers of the analyzed mRNA were normalized using GAPDH or 18S mRNA levels. TaqMan assay for each gene are reported in Table 6.

| **Gene name** | **TaqMan Assay** |
|---|---|
| GAPDH | Hs02758991_g1 |
| 18S | Hs99999901_s1 |
| MYC | Hs00153408_m1 |
| OAS1 | Hs00973637_m1 |
| ISG15 | Hs01921425_s1 |
| SOCS1 | Hs00705164_s1 |

**Table 6: TaqMan Assay for Real-time PCR**

# 10. *Qualitative RT-PCR*

Qualitative RT-PCR (qPCR) was used for the identification of *ST3GAL6-AS1* splicing isoforms, sub-cellular localization and DNA sequencing. qPCR was performed using TaKaRa Taq™ Hot Start Version (Takara bio Inc.). We used 25 ng cDNA or DNA in a final volume of 25 µl for each reaction and 0.4 µM of each primer, with no template controls. The sequences of the primers used are shown in Table 7. After the qPCR reaction, sample and control were loaded on an agarose gel (1-4%) and electrophoresis was carried out. The DNA band pattern was visualized using Atlas Clear sight DNA staining with the detection of ChemiDoc™ MP Imaging System (BIO-RAD).

| *ST3GAL6-AS1* primers | 5'-3' sequences |
|---|---|
| 1_F | GAAGGGAGGGACGAGGAAAG |
| 1_5'end_F | CCAAATTCGAGGAGGAAGGTGC |
| 2_F | GCCCATCCAGATCATCCAGG |
| 3_R | CTGGCTTCATCCTCAGGTCC |
| 4_3'end_R | TTCCCCATGTTTTACTCAAGAGA |
| 5_R | CTAACTGGAAGAGGGCCTGG |

**Table 7: primer sequences for qPCR.** The reported primers were used for the identification of *ST3GAL6-AS1* splicing isoforms. (primers code: F= forward primer; R=reverse primer; 1-5= exon number)

# 11. *Nuclear and cytoplasmic RNA extraction and analysis*

$6x10^6$ cells were collected and centrifuged at 1500 rpm 5 min at RT, medium was removed and the pellet was washed in 10 ml of cold PBS 1X. The supernatant was discarded and the cell pellet was resuspend in 0.2 ml of Hypotonic Buffer A (10mM TRIS HCl pH7.9, 10 mM KCl, 15 mM $MgCl_2, H_2O$) with 200 U/ml of RNAsi OUT (Invitrogen) by passing the solution up and down several times through a pipette tip for 90 sec. After 10 min incubation in ice, the sample was vortexed for 10 sec and 0.5% NP-40 (Abcam) was added. We centrifuged the sample at 4000 rpm, 30 min at 4°C to pellet nuclei. We collected supernatant in a new tube, bring the volume to 0.25 ml, add 0.75ml of TRIzol™ LS Reagent (Thermo Fisher) and freeze at -80°C. In the same time, we washed nuclei twice with 0.2ml of Hypotonic Buffer A plus RNAsi OUT, centrifuged sample at 4000 rpm, 10min at 4°C, discarded supernatant, resuspended nuclei pellet in 1 ml of TRIzol™ Reagent and freeze at -80°C. Nuclear and cytoplasmic RNA were extracted following the protocol supplied by the manufacturer. RNA was quantified and RT-PCR was made. Nuclear localization was analyzed by qPCR using primers in Table 8.

| qPCR primers | Forward Primer 5'-3' | Reverse Primer 5'-3' |
|---|---|---|
| circPVT1 | GGTTCCACCAGCGTTC | CAACTTCCTTTGGGTCTCC |
| 18S | GTAACCCGTTGAACCCCATT | CCATCCAATCGGTAGTAGCG |
| NEAT1 | AGGGTGGTGGCAGTGCTCCT | AATACCGACTCCAACAGCCACTC |
| *ST3GAL6-AS1*_PCR1 | GAAGCCAGCATCTGAGGAAG | GCCAGCATTTTGGTAAGAGC |
| *ST3GAL6-AS1*_PCR2 | GCAGCACAGAATCCTGACAA | GCCAGCATTTTGGTAAGAGC |
| *ST3GAL6-AS1*_PCR3 | GCAGCACAGAATCCTGACAA | CTAACTGGAAGAGGGCCTGG |

**Table 8: primer sequences for qPCR**

# 12. *Sanger sequencing*

qPCR was controlled on Agarose gel 1.5%. qPCR product was purified with Pure PCR purification Kit (Invitrogen) following manufacturer's instructions. If the qPCR product evidenced more than one band, all the sample were loaded on an agarose gel 0.8% and electrophoresis was carried out. When the band were separated, we cut off and purified them with DNA gel Extraction kit (MERK) following manufacturer's instructions.

Purified product sample was loaded on an agarose gel 1.5% and electrophoresis was carried out to check the band. Purified product was amplified with BigDye™ Terminator v1.1 Cycle Sequencing Kit (Thermo Fisher) and amplified product was purified by Performa® DTR Gel Filtration Cartridges (Edge Bio) following manufacturer's instructions. This sample was denatured at 95°C for 1min and loaded on Abi Prism 310 Genetic Analyzer (Thermo Fisher) to be sequenced. The sequences were analyzed using Chromas software (Technelysium).

## 13. *SNP Genotyping Analysis*

SNP rs13065271 variation status of *ST3GAL6-AS1* was investigated, at the beginning, on human myeloma cell lines (HMCLs) DNA with Sanger sequencing. For the analysis of all the HMCLs and MM patients for which DNA was available, we set up conditions for SNP Genotyping Analysis Using TaqMan Assays (Thermo Fisher) (C___2165147_10). In particular, we used MM1.S, KMS20 and KMS34 cell lines as control for C/C, C/T and T/T presence, respectively. PCR reactions were carried out on a Step-One Plus PCR system (Applied Biosystems, Life Technologies Italia, Italy) using the TaqMan™ Genotyping Master Mix (2X) (ThermoFisher). The reaction, in a final volume of 5 µl for a 48-well plate, was made as described in the protocol supplied by the manufacturer on 20 ng of DNA. Each sample was analyzed in triplicate with no template controls.

## 14. *RNA half-lives identification*

To detect RNA half-lives, HMCLs were seeded at $1x10^6$ cell/ml in RPMI-1640 (Gibco, Thermo Fisher) with 10% FBS (Gibco, Thermo Fisher) and 2 mM L-glutamine (Invitrogen Corporation Carlsbad, CA, USA) in 12-multiwell culture plates. One well for each time point was treated with Actinomycin D (which could inhibit transcription, Thermo Fisher) at a final concentration of 1 µg/ml or DMSO as control vehicle (SCR treatment) and cells were cultured in 5% $CO_2$ at 37°C. Cells were harvested at different time points after treatment (20, 40, 60, 90, 120, 240, 360, and 540 minutes) and total RNAs were extracted [114]. qPCR was made on these sample and the results were analyzed by comparing expression level of treated cells with expression level in Scramble (SCR) treated cells at the same times points using as housekeeping gene 18S, which was stable for more than 24 hours. In addition, we verified the transcription inhibition using cMYC half-life reported as minor of 1hour.[115,116]

# 15. *Transfection methods*

## 15.1 GapmeRs and siRNA

LNA$^{TM}$ longRNA GapmeR (GapmeR) were designed by EXIQON on-line tool and their efficiency was tested by lipofection (RNAiMAX) and electroporation (NEON transfection system) tecniques. In Table 9 we report GapmeR Design ID.

| GapmeR Design ID | Nome of GapmeR |
|---|---|
| Negative Control B | Gap SCR |
| 613937-2 | Gap 1 |
| 676401-1 | Gap 4 |
| 676401-4 | Gap 5 |
| 676401-7 | Gap 6 |

**Table 9: GapmeRs Design ID and Name**

siRNAs were purchased from Dharmacon.

- *ST3GAL6* siRNA: *ST3GAL6* L-018009-01-0005, ON-TARGETplus Human *ST3GAL6* (10402) siRNA SMARTpool
- Negative control siRNA: ON-TARGETplus Non-targeting Pool

## 15.2 RNAiMAX lipofection method

HEK-293T e Hs5 were seeded at $5x10^5$ cells in 0.45 ml of DMEM with 10% FBS (Gibco, USA), 2 mM L-glutamine (Invitrogen Corporation Carlsbad, CA, USA) in 12-multiwell culture plates. After 24 hours, cell were treated with 0.05 ml of transfected solution made with Lipofectamine™ RNAiMAX Transfection Reagent (Thermo Fisher), Opti-MEM™ I Reduced Serum Medium (Thermo Fisher) and 100 nM of GapmeR following manufacturer's instructions. After 72 hours from treatment, cells were detached using Trypsin-EDTA (0.25%), phenol red (Thermo Fisher) and counted in Burker's chamber. RNA was extracted and silencing was analyzed by RT-PCR. All the transfection conditions were made in triplicate.

## 15.3 NEON transfection system

HMCLs were counted and centrifuged at 900 x g 5 min at RT. Medium was removed and the pellet was washed in 1 ml of PBS 1X. The supernatant was discarded and the cell pellet was

resuspended in Resuspension Buffer R at $1x10^6$ cells/0.1 ml and 100nM of siRNA or 100 nM of GapmeRs were added. In the same time, we connected Neon® Pipette station to the Neon®device and inserted Neon®Tube with 3 ml of Electrolytic Buffer in Neon® Pipette station. We connected Neon® Pipette and Tip, aspirated the cell suspension without air bubbles and inserted Neon® Pipette in Neon® Pipette station. We started the electroporation with Voltage 1100V, Width 30 ms, Pulses 2. We transferred samples into prepared cultured plate containing prewarmed medium (RPMI-1640 with 10% FBS and 2 mM L-glutamine) to have $0.5x10^6$ cells/ml. After 6, 48 and 72 hours of incubation from electroporation we collected the cells, counted them, extracted RNA and analyzed it by RT-PCR (Thermo Fisher). All the transfection conditions were made in triplicate.

### 15.4 Gymnotic delivery

HMCLs were counted and centrifuged at 900 x g 5 min at RT. The supernatant was discarded and the cell pellet was resuspended in RPMI-1640 (with 10% FBS and 2 mM L-glutamine) to a concentration of $5x10^4$ cells/ml and seeded in multi-well. We added GapmeR in the right well at a final concentration of 5 µM. Every day, for nine days from Gimnotic delivery, cells were re-suspended and counted. We confirmed biological effect using CellTiter-Glo® Luminescent Cell Viability Assay (Promega) following manufacturer's instructions. After 2, 4, 5, 7, and 9 days from electroporation, RNA was extracted and gene expression levels analyzed. All the transfection conditions were made in triplicate.

## 16. *Flow cytometry analysis*

All the Flow cytometry analyses were performed with BD FACSVerse™ System (BD Biosciences, USA) and data were analyzed using flow cytometry analysis program "Cytomics FC500" (Beckman Coulter) in collaboration with Professor Chiaramonte's laboratory.

### 16.1 Cell cycle assay

$3x10^5$ cells of each sample were washed in PBS 1X and fixed in 70% cold ethanol. We centrifuged fixed cell suspension, discarded supernatant and added 0.5 ml of FxCycle™ PI/RNAse Staining Solution (Thermo Fisher). Samples were vortexed and incubated for 30 min at RT in the dark before processing.

### 16.2 Apoptosis assay

$3x10^5$ cells of each sample were washed twice in PBS 1X and then resuspended in 1X Binding Buffer. We added PE Annexin V and 7-AAD (BD Pharmingen™), vortexed, and incubated samples for 15 min at RT in the dark before adding 400 µl of 1X Binding Buffer and analyzing by flow cytometry.

## 17. *Colony-forming assay*

After 24 hours from GapmeR gymnotic delivery, LP1 and H929 cells were counted and 200cells/well (24-multiwell) were seeded in 1ml of MethoCult™ (STEMCELL ™ TECHNOLOGIES) plus RPMI-1640 with 10% FBS and 5µM of GapmeR, following manufacturer's instructions. Cells were cultured in 5% CO2 at 37°C for 12/21 days and pictures were taken by IPhone 6.

## 18. *Statistical analysis*

All the statistical analyses were performed on Graph Pad Prism 5 (Graph Pad Software) and R software. We used Kruskal-Wallis test to analyze differential expression levels of genes and lncRNAs, in GEP and PCR data; we performed nonparametric pairwise multiple comparisons between the independent groups by Dunn's test. We used Spearman nonparametric correlation to analyze correlation between expression levels of genes and lncRNAs. Experimental data are represented as means ± SD of all the replicates and for the statistical analysis, we used two-way ANOVA.

# Results

# 1. *Identification of lncRNAs deregulated in MM patients compared with normal donors*

In order to investigate deregulated lncRNAs in MM, highly purified bone marrow PCs from 50 MM primary tumors, 15 PCL and 4 normal donors were profiled on GeneChip® Human Gene 2.0 ST microarray (Affymetrix). A custom annotation pipeline using GENCODE annotation was developed to investigate 10138 lncRNAs present in the microarray.

The gene expression data were analyzed by SAM supervised analysis with very high stringency (q value=0) by comparing MM and PCL samples with healthy donor samples. As shown in the heatmap in Figure 9, this analysis generated a list of 9 specific differentially expressed (DE) lncRNAs: *ST3GAL6-AS1* (ST3 beta-galactoside alpha-2,3 sialyltransferase 6 antisense RNA 1) was the unique lncRNA upregulated in MM and PCL patients compared to healthy donor samples; lncRNAs AP000345.2, AC079610.1, WASIR2, RP11-843A23.1, RP1-35C21.1 and AP002954.4 were downregulated in all pathological samples; LINC00582 and RP11-87F15.2 were upregulated in MM and downregulated in PCL patients.

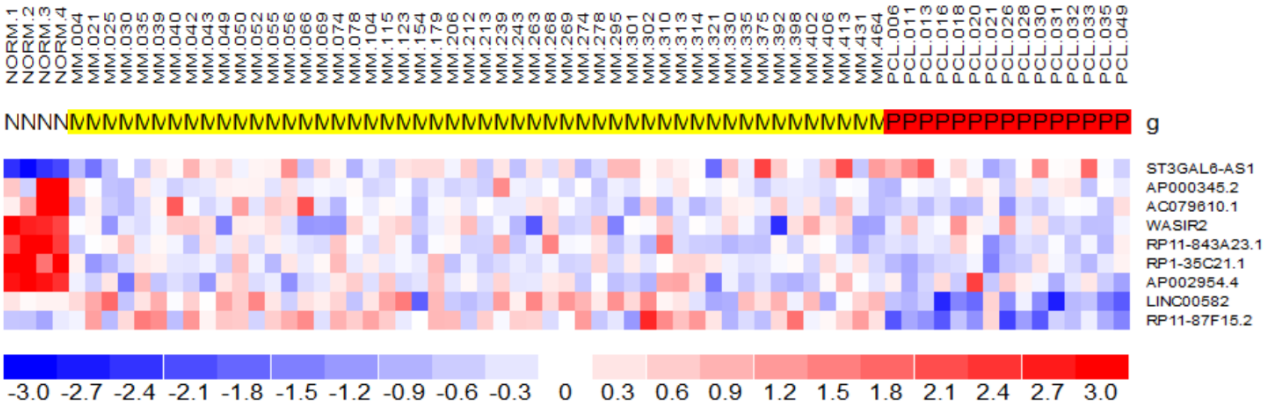



**Figure 9**: **heatmap of differentially expressed lncRNA in MM compared with normal controls**. The Heatmap shows results of SAM supervised analysis by comparing 50 MM, 15 PCL and 4 normal control samples profiled on GeneChip Gene 2.0 ST arrays. We show nine differentially expressed lncRNAs. In particular, *ST3GAL6-AS1* is the unique lncRNA up modulated in MM and PCL patients compared to healthy donor samples.

Table 10 provides detailed information concerning the DE lncRNAs list. *ST3GAL6-AS1* appears as the only lncRNA previously described in literature as modulated in MM[90]. Furthermore, Glavey SV et al, in 2014, described the neighboring gene of *ST3GAL6-AS1*,

*ST3GAL6*, as involved in homing and in *in vivo* engraftment in MM[117]. The *ST3GAL6-AS1* is predicted to have 5 exon generating three alternative spliced transcripts. Concerning the remaining lncRNAs modulated in our analysis, all of them showed one isoforms each and their neighboring genes have not been reported to be involved in MM pathology.

In addition, we investigated the RNAseq expression pattern of these lncRNAs available in a proprietary data set of 30 MM and 2 HMCLs by means of the IGV software. To note, these 30 MM samples were included in the series investigated with GeneChip® Human Gene 2.0 ST array previously. As reported in Table 10, and in accordance with GEP data, the *ST3GAL6-AS1* showed very high expression levels compared with the other differentially expressed lncRNAs (see RNAseq data scale). As shown in Figure 10, the *ST3GAL6-AS1* evidenced the presence of reads mapped on the five predicted exons, all of them showing similar expression levels. Furthermore, we identify the presence of unpredicted longer sequences in 3'end of exon five.

| Name | Alias name | Number of isoform | Lenght in bp | nEX-nINT | Mapping region | Chr location | Neighboring genes in 11000bp | RNAseq data scale |
|---|---|---|---|---|---|---|---|---|
| ST3GAL6-AS1 | | ENST00000461931.1 | 669 | 4-3 | 3q12.1- | chr 3: 98,714,330-98,732,651 | ST3GAL6 + DCBLD2- | 0.0-359.711 |
| | | ENST00000475816.5 | 524 | 4-3 | | | | |
| | | ENST00000488132.5 | 769 | 5-4 | | | | |
| AP000345.2 | | ENST00000608615.1 | 1081 | 3-2 | 22q11.2- | chr22: 23,580,880-23,583,859 | IGLL1 - | 0.0-11.6979 |
| AC079610.1 | | ENST00000360083.3 | 1727 | 3-2 | 2q34- | chr 2: 213,276,552-213,284,205 | SPAG16+ IKZF2- | 0.0-17.8442 |
| WASIR2 | | ENST00000527434.1 | 1054 | 2-1 | 16p13.3+ | chr 16: 22,910-25,123 | | 0.0-4.19858 |
| RP11-843A23.1 | AP003110.1 | ENST00000526976.1 | 717 | 4-3 | 11q14.1+ | chr11: 78,533,176-78,558,565 | NARS2- | 0.0-1.61208 |
| RP1-35C21.1 | LINC01645 | ENST00000451341.1 | 491 | 4-3 | 1q25.2+ | chr1: 177,351,586-177,366,272 | BRINP2+ ASTN1- | 0.0-1.80704 |
| AP002954.4 | AP002954.1 | ENST00000526274.1 | 425 | 3-2 | 11q23.3+ | chr 11: 118,700,427-118,750,263 | DDX6-TREH-PHLDB1+ | 0.0-5.19219 |
| LINC00582 | | ENST00000448058.1 | 597 | 2-1 | 1q42.2- | chr1: 231,591,292-231,612,090 | TSNAX-DISC1+ TSNAX+ | 0.0-668.159 |
| RP11-87F15.2 | AC019163.1 | ENST00000512634.1 | 610 | 3-2 | 4q34.2- | chr 4: 176,308,268-176,320,458 | SPCS3+ ASB5- | 0.0-117.71 |

**Table 10: lncRNAs modulated in MM and PCL patients compared with normal controls.** For each lncRNA we reported name and alias name of ENST spliced transcripts, length of all the predicted spliced transcripts, number of predicted exons (EX) and introns (INT), chromosome mapping region and strand (forward +, reverse -), chromosome location, and neighboring genes in 11000bp and RNAseq DATA-SCALE. This information was extracted by Ensembl dataset and the UCSC browser.

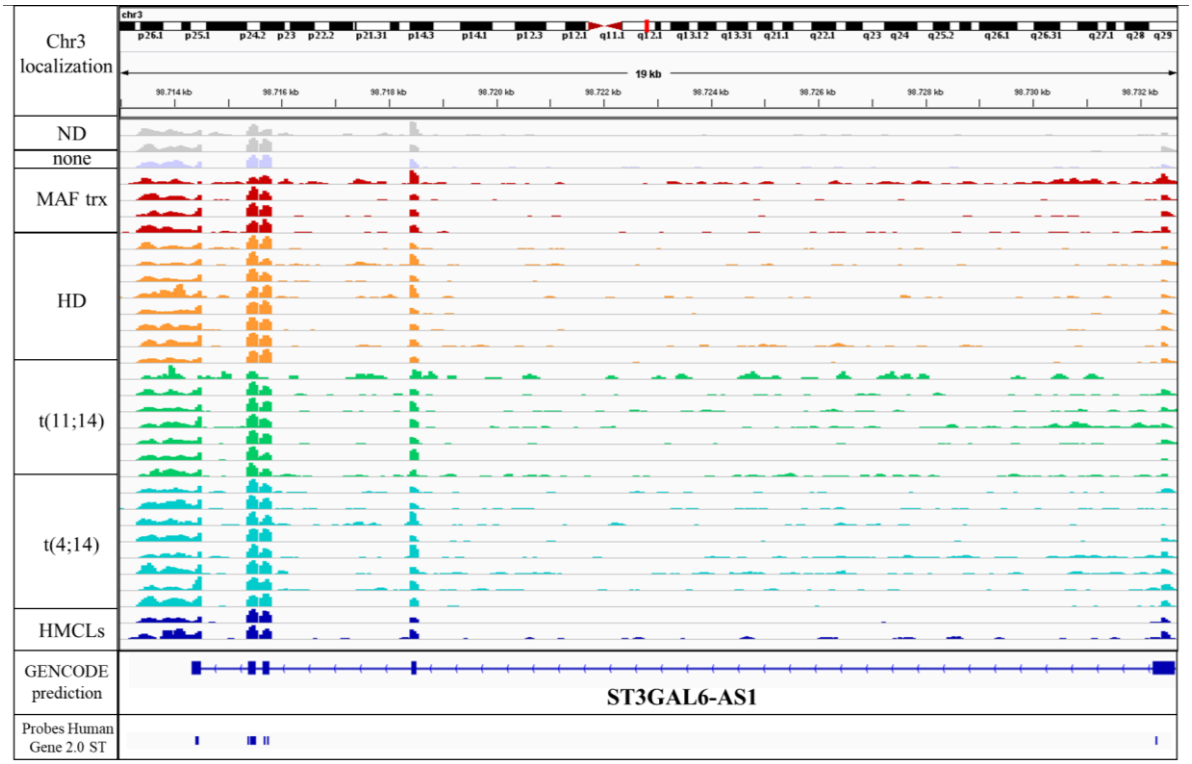



**Figure 10: IGV visualization of *ST3GAL6-AS1* reads mapping.** The upper part of the figure shows chromosomal localization of *ST3GAL6-AS1*. In the middle part of the figure are evidenced the mapping reads on *ST3GAL6-AS1* made on RNAseq data of 30 MM patients and 2 MM cell lines (KMS-11 and U266). In the figure the 5 exons of *ST3GAL6-AS1* are evidenced. The lower part of the figure shows GENCODE predicted structure of *ST3GAL6-AS1* and *ST3GAL6* gene and the mapping of Gene 2.0 array probes.

Therefore, we focused our attention on *ST3GAL6-AS1* based on the evidence that it has been reported as up-modulated in MM[90], it shows a significant RNAseq expression pattern and maps near *ST3GAL6,* a protein-coding gene involved in MM pathology[117] .

## 2. *GENCODE annotation: predicted ST3GAL6-AS1 transcripts*

As shown in Figure 11A, the lncRNA *ST3GAL6-AS1* is located on the reverse strand of chromosome 3q12.1 and has two neighboring genes: *DCBLD2* (Discoidin, CUB and LCCL Domain Containing 2) located on the same strand with its far 3' end at 63kb from the *ST3GAL6-AS1*, and coding for a receptor tyrosine kinase with aberrant expression in malignant tumors[118]; and *ST3GAL6* (ST3 Beta-Galactoside Alpha-2,3-Sialyltransferase 6) located on the sense strand in head-to-head orientation, sharing a 415 bp complementary sequence with *ST3GAL6*.AS1. *ST3GAL6* codes for an enzyme able to transfer sialic acid for the production of glycolipids and glycoproteins. Furthermore, this enzyme is involved in the formation of selectin ligands and sialyl Lewis X, a blood group antigen[117].

GENCODE annotation reports three variants of *ST3GAL6-AS1* derived from different splicing events involving the five predicted exons (Figure 11B).

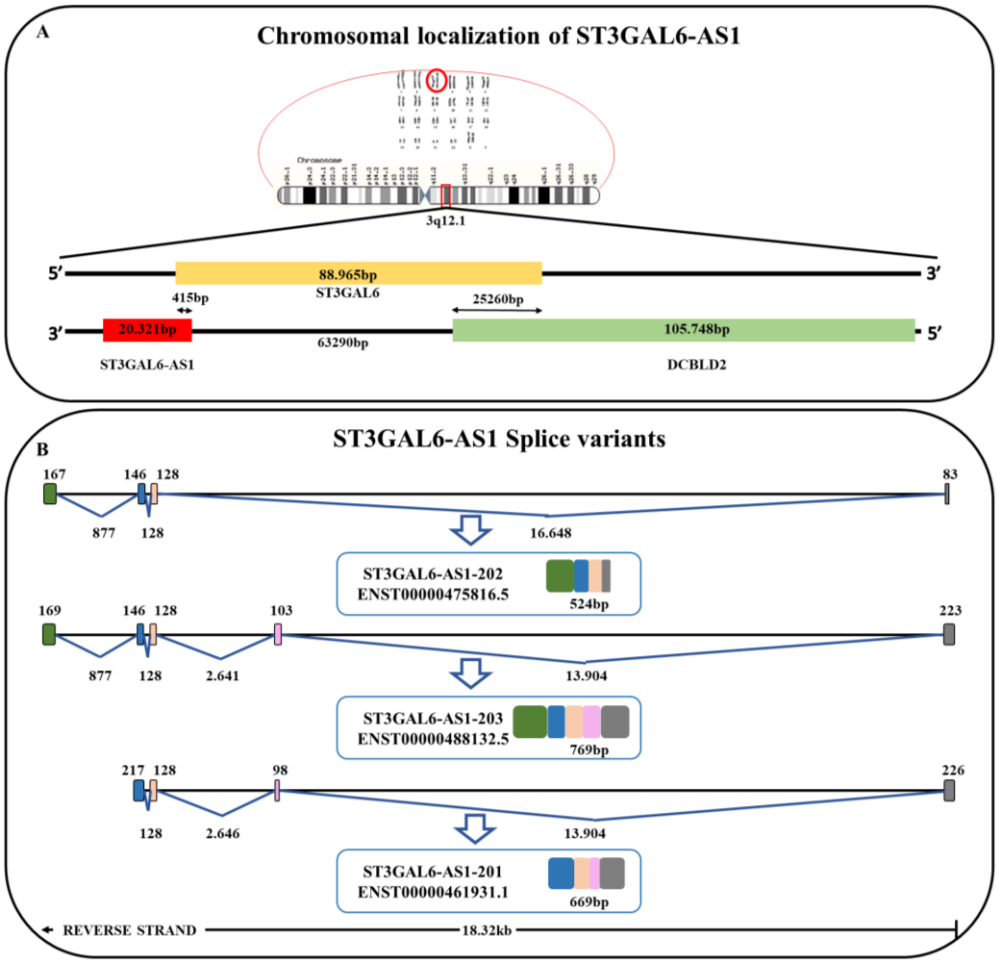


**Figure 11: A. Representation of chromosomal region 3q12.1:** in the upper part of the figure is reported a schematic view of gene chromosomal localization of *ST3GAL6-AS1* and its neighboring genes. Mapping on the reverse strand of chromosome 3q12.1, *ST3GAL6-AS1* has as neighboring genes *ST3GAL6* and *DCBLD2*. In particular, *ST3GAL6-AS1* is in head-to-head orientation and a region of 415bp in common with *ST3GAL6* and *DCBLD2* locates at 63kbp from the lncRNA. *DCBLD2* shares with *ST3GAL6* 2560bp. **B**. **Representation of splice variants of *ST3GAL6-AS1*:** we show a schematic view of GENCODE annotation *ST3GAL6-AS1* splice variant products by the splicing of five exons.

Based on the Ensembl predictions, we could evidence that exons may have different length (see Table 11). Specifically:

- exon 1 is present in all the predicted isoforms with a variable length between 83 to 226 bp at its 5' region;
- exon 2 is present in two isoforms with 5bp of difference at its 3'end (98 or 103 bp);
- exon 3 is present in all the isoforms with 128 bp in lenght;
- exon 4 is present in all the isoforms; in *ST3GAL6-AS1*-201 it has predicted to be 217 bp in length extending to its 3' end;
- exon 5 is present in two isoforms with a length of 167 or 169 bp.

| Exon | Length of the different predicted isoforms | | |
|---|---|---|---|
| | *ST3GAL6-AS1*-202 ENST00000475816.5 | *ST3GAL6-AS1*-203 ENST00000488132.5 | *ST3GAL6-AS1*-201 ENST00000461931.1 |
| 1 | 83 | 223 | 226 |
| 2 | / | 103 | 98 |
| 3 | 128 | 128 | 128 |
| 4 | 146 | 146 | 217 |
| 5 | 167 | 169 | / |
| Total length | 524 | 769 | 669 |

**Table 11: in the table are indicated the specific exons of *ST3GAL6-AS1* isoforms and their specific length.**

## 3. *Expression of ST3GAL6-AS1 in MM patients and correlation with its neighboring genes.*

Based on GEP data, *ST3GAL6-AS1* was found overexpressed in MM compared to normal donors. We then analyzed the *ST3GAL6-AS1*, *ST3GAL6* and *DCBLD2* expression levels according to the major molecular subtypes of patients, including 17 with hyperdiploidy status, 15 with t(11;14), 15 with t(4;14), and 11 harboring MAF translocations (MAF trx). As shown by box plots in Figure 12, the expression levels of *ST3GAL6-AS1* appeared similar in all four molecular groups of MM patients, as well as the expression levels of *ST3GAL6*. Furthermore, we evidenced a significant correlation between the expression level of *ST3GAL6* and its neighboring antisense lncRNA (spearman r=0.68). On the contrary, the expression pattern of *DCBLD2* was quite different from those of *ST3GAL6-AS1* and *ST3GAL6*; of note, no correlation with expression levels of *ST3GAL6-AS1* was found (spearman r=0.39).

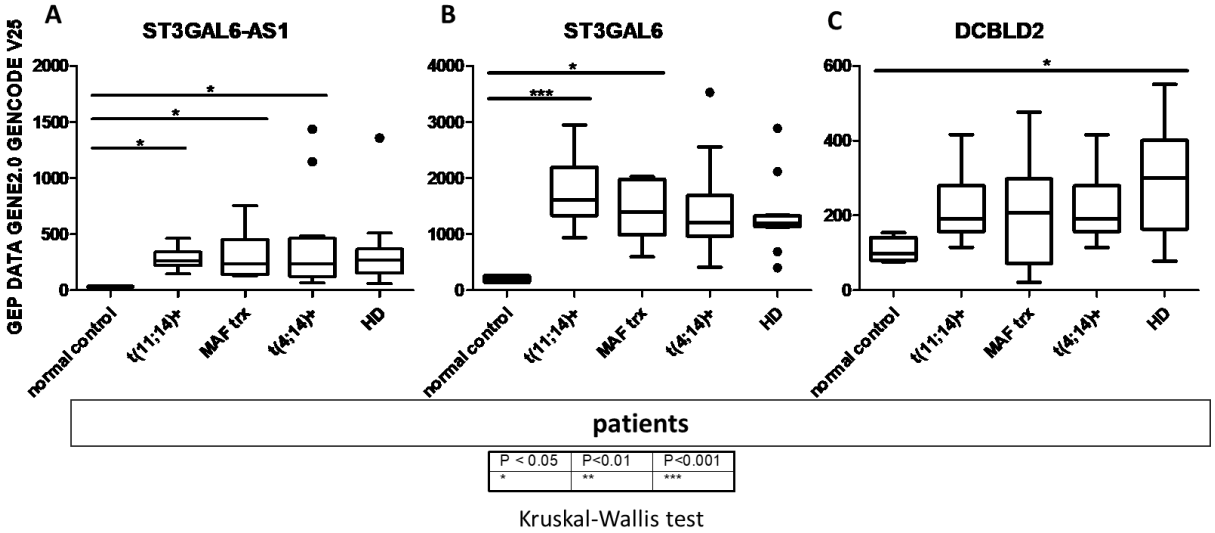



**Figure 12: box plot distribution.** Box plot distribution of *ST3GAL6-AS1* (A), *ST3GAL6* (B) and *DCBLD2* (C) expression levels in our cohort of patients; patients were divided for the presence of hyperdiploidy status (17 patients), t(11;14) translocation (15 patients), translocation of MAF genes (11 patients) or t(4;14) translocation (15 patients). Kruskal-Wallis Test was used for the calculation of p value. *ST3GAL6-AS1* and *ST3GAL6* were overexpressed in MM and PCL samples.

# 4. *ST3GAL6-AS1, ST3GAL6 and DCBLD2 expression in human cell lines*

In order to evaluate the expression levels of *ST3GAL6-AS1*, *ST3GAL6* and *DCBLD2* in human cell lines we performed RT-PCR analyses. We investigated 20 HMCLs, 5 breast cancer cell lines, 5 thyroid carcinoma cell lines, 6 hematological neoplastic non-MM cell lines, Hs5, derived by bone marrow stroma, and HeLa and HEK-293T cell lines.

The RT- PCR data, normalized on *GAPDH* expression, showed a variable expression level of *ST3GAL6-AS1* and *ST3GAL6* in MM cell lines, that were higher compared to the other analyzed cell lines. Moreover, we evidenced a significant correlation between *ST3GAL6-AS1* and *ST3GAL6* in all the cell lines (spearman r=0.93). *DCBLD2* showed a different expression pattern in investigated cell lines compared with *ST3GAL6* and *ST3GAL6-AS1*. In particular, *DCBLD2* displayed a significant up-modulation in thyroid carcinoma compared with MM cell lines (Pvalue<0.001). We did not evidence any type of correlation between *ST3GAL6-AS1* and *DCBLD2* expression levels (spearman r=-0.25).

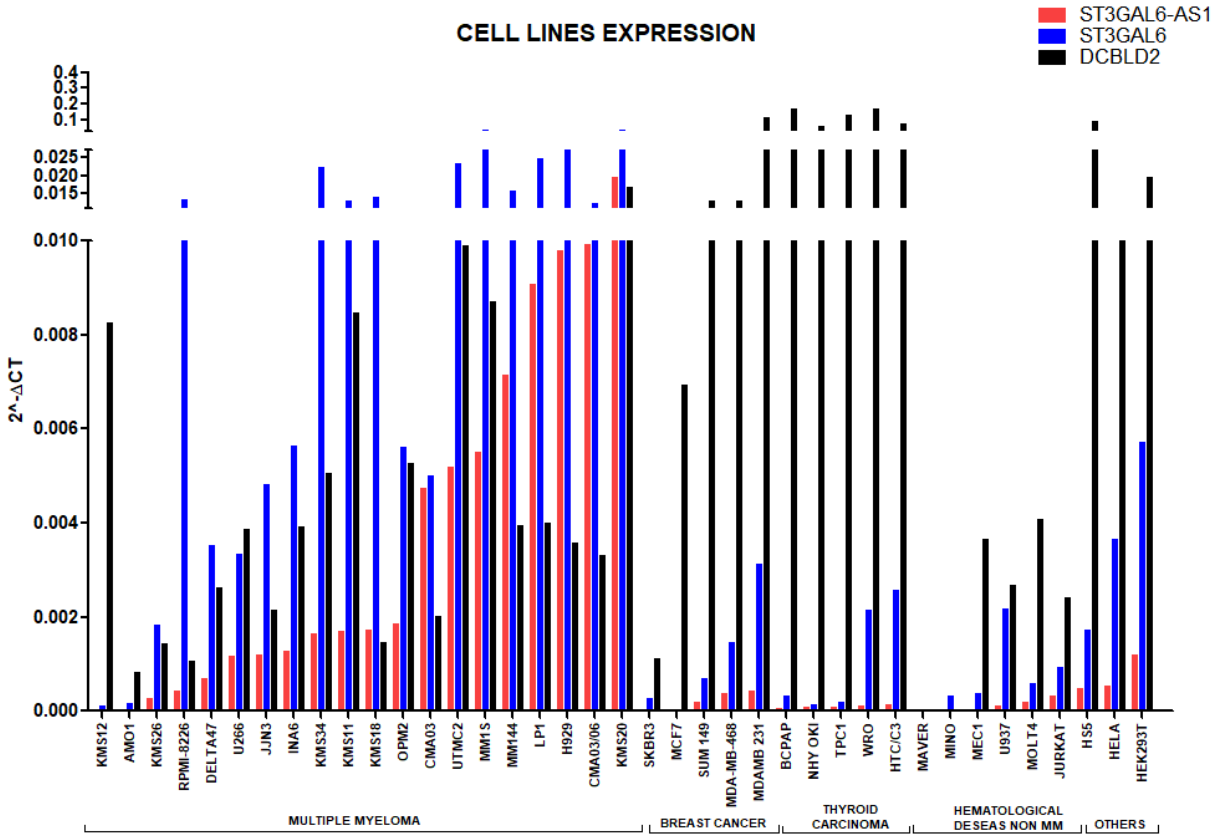


**Figure 13: cell lines expression levels.** The histogram shows expression levels of *ST3GAL6-AS1*, *ST3GAL6* and *DCBLD2* in 44 cell lines. In particular, we analyzed 20 HMCLs (KMS12, AMO1, KMS26, RPMI-8226, DELTA47, U266, JJN3, INA6, KMS34, KMS11, KMS18, OPM2, CMA-03, UTMC2, MM1.S, MM1.144, LP1, H929, CMA-03/06, KMS20), 10 breast cancer cell lines (SKBR3, MCF7, SUM-149, MDA-MB-468, MDA-MB-231), 5 thyroid carcinoma cell lines (BCPAP, NHY-OKI, TPC1, WRO, HTCJC3), 6 hematological neoplastic non-MM cell lines (MAVER, MINO, MEC1, U937, MOLT4, JURKAT), Hs5 derived from bone marrow stroma, HeLa and HEK-293T. We straightened up cell lines first in their tumor source and in a second time for *ST3GAL6-AS1* expression levels.

# 5. *Characterization of ST3GAL6-AS1 isoforms in Myeloma*

## 5.1 Study of splicing isoforms

Based on the Ensembl prediction (see paragraph 2), *ST3GAL6-AS1* is characterized by three alternative splicing isoforms.

With the aim to confirm *ST3GAL6-AS1* alternative splicing isoforms, we designed primers in its first and fifth exons (primers 1_F and 5_R; table 7) and performed qualitative RT-PCR. Based on this approach, 2 bands were expected, 536 bp derived from the splicing of all the introns, and 433 bp from the skipping of exon 2.

The 4% agarose gel visualization of the qPCR made on cDNA of 12 HMCLs (MM1.S, H929, U266, CMA-03, KMS11, KMS20, RPMI-8226, LP1, DELTA47, KMS34, KMS26 and INA-6) surprisingly evidenced different bands compared to what it would be expected. We could distinguish three different expression patterns of this lncRNA and in particular, the presence of unknown isoforms of *ST3GAL6-AS1* in all patterns. Specifically, the cell lines MM1.S, H929, U266 and CMA-03 (blue box) evidenced a predominant band of approximately 536 bp like the predicted one; KMS34, KMS26, INA-6 and DELTA47 (green box) evidenced a predominant band of approximately 390 bp lower than the predicted band of 433 bp; KMS18, KMS11, KMS20 and RPMI-8226 (orange box) an intermediate pattern between the other two classes (see Figure 14)

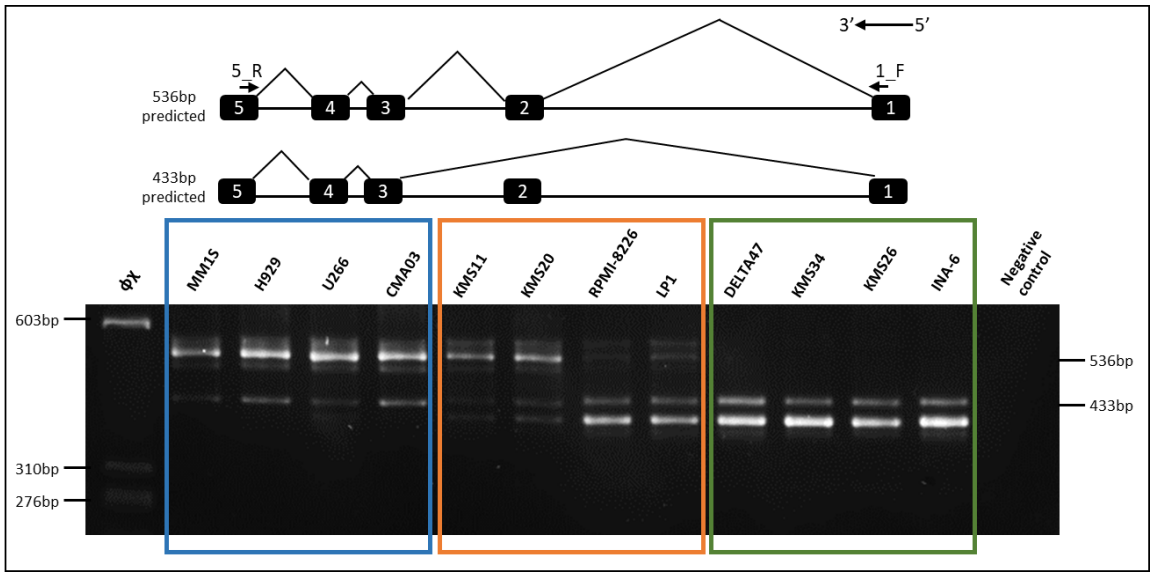


**Figure 14: *ST3GAL6-AS1* splicing isoform identification.** In the upper part of the figure these is a schematic view of *ST3GAL6-AS1* with the indications of primers in exon 1 and 5 and predicted splicing. In the bottom, agarose 4% gel visualization of the qPCR, made on cDNA of 12 HMCLs with primers on exons 1 and 5. Colored boxes showed different pattern of splicing.

Based on these findings, we next investigated the structure of these fragments by using different couples of primers. We analyzed two HMCLs for each different pattern: MM1.S and H929 (green box), KMS20 and KMS11 (orange box), and KMS34 and DELTA47 (red box).

At first, qPCR was performed by using the forward primer already described on exon 1 (1_F) and a reverse primer on exon 3 (3_R), in order to analyze the 5' region of the transcripts. It is predicted that this PCR should evidence 2 bands: the first of 239 bp in which we have the splice of intron 1 and intron 2 and the second of 118 bp in which there is the skipping of exon 2 (Figure 15A).

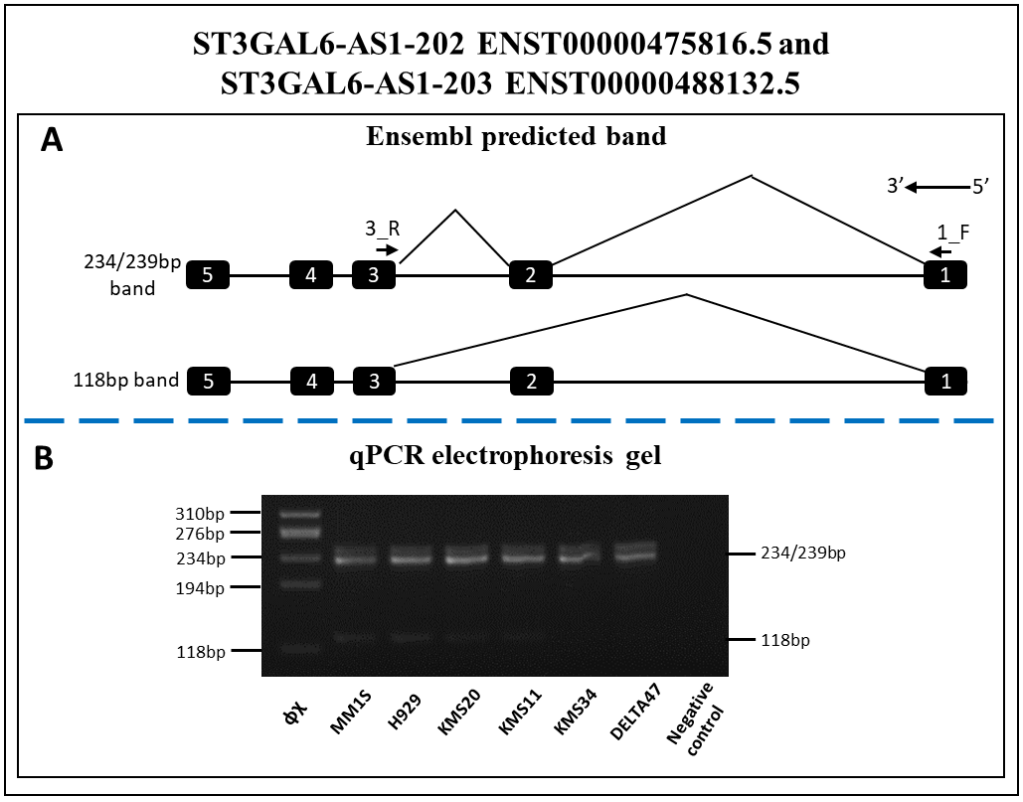



**Figure 15: *ST3GAL6-AS1*-202 and *ST3GAL6-AS1*-203 5'end region analysis. A:** schematic view of *ST3GAL6-AS1* with the indications of primers in exon 1 (1_F) and 3 (3_R) and predicted splicing. **B:** agarose 1.5% gel visualization of the qPCR, made on cDNA of 6 HMCLs; in the right part are indicated marker bands size. In the right part are indicated qPCR products band size.

Notably, this qPCR analysis showed the absence of the predicted splicing isoform of 118 bp in KMS34 and DELTA47 cell lines. The other HMCLs examined presented all the splicing isoforms predicted (Figure 15B). Additionally, this qPCR evidenced in all the HMCLs the

double band of nearly 230-240bp that is caused by the alternative length of exon 2 (see above Table 11).

The second qPCR (Figure 16A) analyzed the 3'region of *ST3GAL6-AS1* by means of a forward primer on exon 2 (2_F) and the reverse primer already described on exon 5 (5_R). We expected a predicted band of 474/479bp (the difference is caused 5bp length variation of exon2).

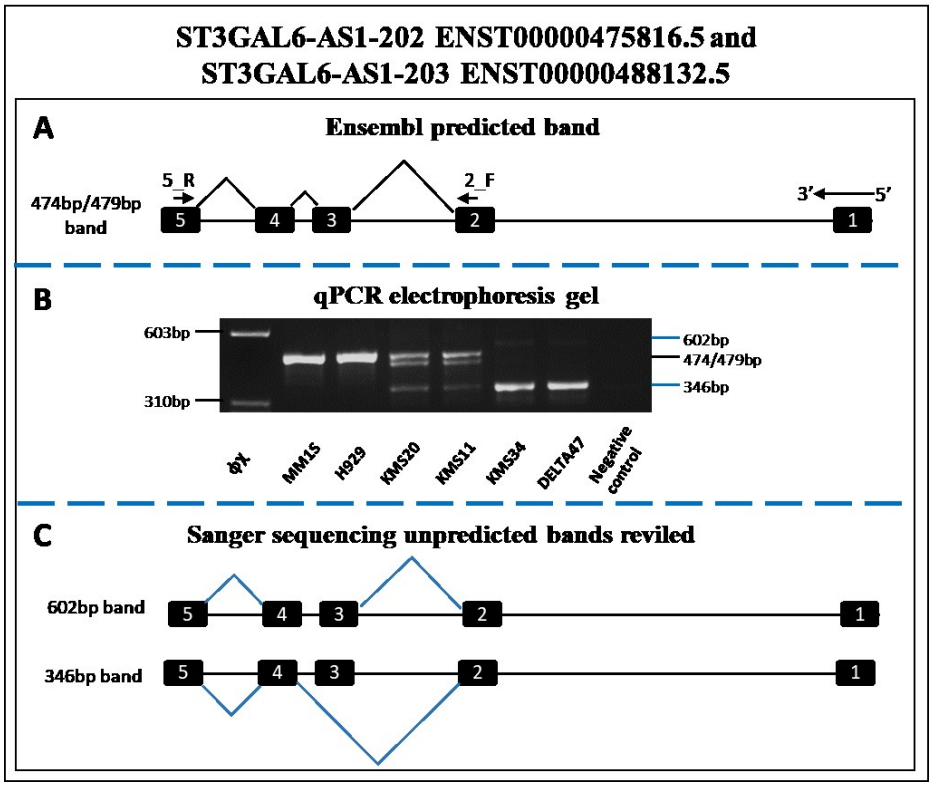



**Figure 16: *ST3GAL6-AS1*-202 and *ST3GAL6-AS1*-203 3'end region analysis. A:** schematic view of *ST3GAL6-AS1* with the indications of primers in exon 5 (5_R) and 2 (2_F) and predicted splicing product attended. **B:** agarose 1.5% gel visualization of the qPCR, made on cDNA of 6 HMCLs; in the right part are indicated marker bands size. In the right part are indicated qPCR products band size. **C:** Schematic view of the alternative splicing evidenced after Sanger sequencing of the bands.

As shown in Fig 16B, MM1.S and H929 presented the only predicted band of 474/479 bp; KMS34 and DELTA47 showed the presence of two unpredicted bands of approximately 602 bp and 346 bp; KMS20 and KMS11 showed the presence of the predicted band of 474/479 bp and the lower band of 346 bp as in KMS34 and DELTA47.

Ensembl prediction of *ST3GAL6-AS1* isoforms described an additional isoform composed by the splicing of exons 1 to 4 and the absence of exon 5. This isoform, named *ST3GAL6-AS1*-201, resulted composed by a longer exon 1 of 226 bp and a longer exon 4 of 217 bp (for details see table 11).

For the analysis of 5' end region of this isoform it is important to underline that the primer already used on exon 1 is not able to characterize all the length of this exon so we designed a new primer on exon 1 located at its far 5' end called *1_5'_F*.

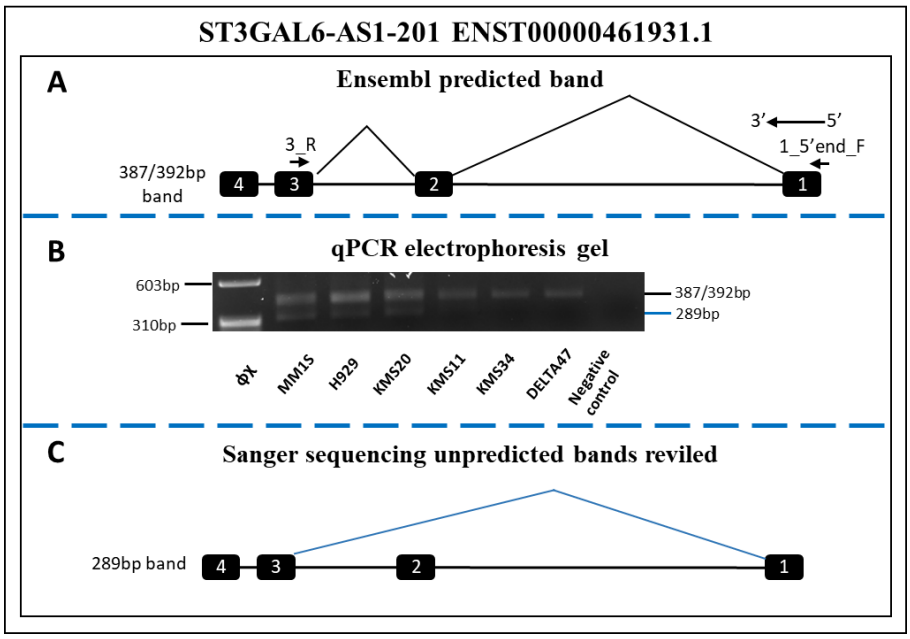



**Figure 17: *ST3GAL6-AS1*-201 5'end region analysis. A:** schematic view of *ST3GAL6-AS1*-201 with the indications of primers in 5'end of exon 1 (*1_5'end_F*) and in exon 3 (3_R) and predicted splicing product attended. . **B:** agarose 1.5% gel visualization of the qPCR, made on cDNA of 6 HMCLs; in the left part are indicated marker bands size. In the right part are indicated qPCR products band size. **C:** schematic view of the alternative splicing evidenced after Sanger sequencing of the bands.

The use of *1_5'end_F* forward primer and the already used reverse primer on exon 3 (3_R) should predict a band of 387 bp caused by splicing events of intron 1 and intron 2 (Figure 17A). Actually, this analysis showed the presence of predicted band of 387 bp in all HMCLs analyzed; however, MM1.S, H929 and KMS20 showed the presence of an unpredicted band of approximately 289 bp (Figure 17B).

Furthermore, in *ST3GAL6-AS1*-201, exon 4 is 217 bp in lenght, 71 bp longer than exon 4 in the other isoforms. For the analysis of 3'end of *ST3GAL6-AS1*-201 we designed a new reward primer at the 3' end of exon 4 called *4_3'end_R* (Figure 18A)

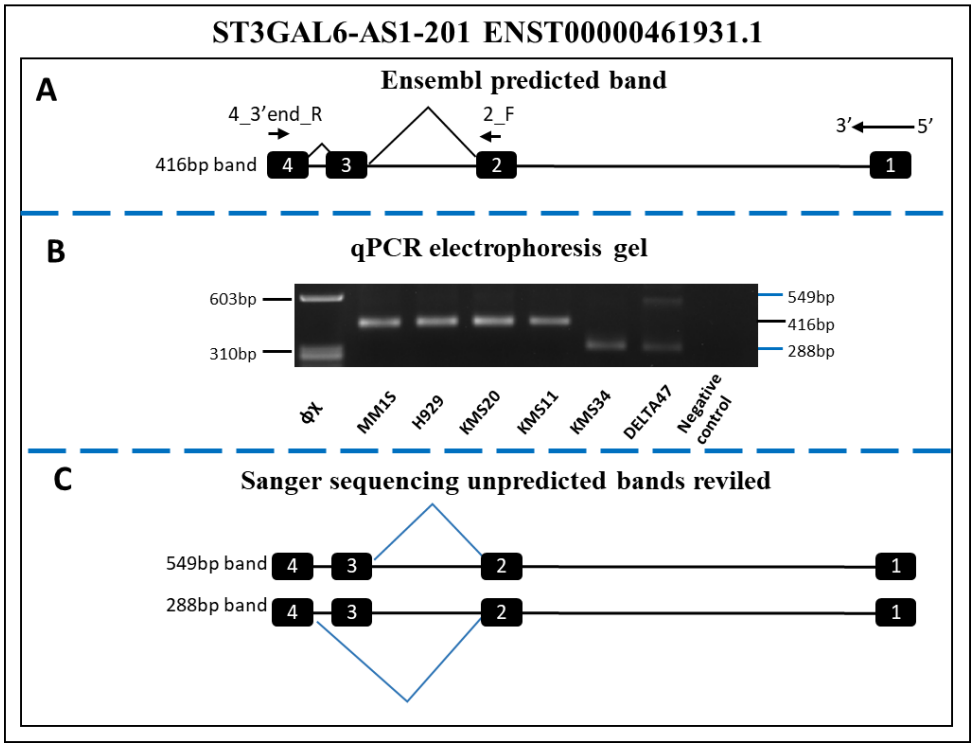



**Figure 18: *ST3GAL6-AS1*-201 3'end region analysis. A:** a schematic view of *ST3GAL6-AS1*-201 with the indications of primers in exon 2 (2_F) and in 3'end of exon 4 (*4_3'end_R*) and predicted splicing product attended. **B:** agarose 1.5% gel visualization of the qPCR, made on cDNA of 6 HMCLs; in the left part are indicated marker bands size. In the right part are indicated qPCR products band size. **C:** schematic view of the alternative splicing evidenced after Sanger sequencing of the bands.

Next, we performed a qPCR with the described forward primer on exon 2 (2_F) and the reverse primer *4_3'end_R* predicted to give a band of 416 bp caused by splicing events of intron 2 and intron 3. MM1.S, H929, KMS20 and KMS11 cell lines analyzed with this couple of primers evidenced the only predicted band of 416 bp; instead, KMS34 and DELTA47 evidenced the absence of the predicted band and the presence of a higher band of approximately 549 bp and a lower band of approximately 288bp.(Figure 18B).

Based on these findings concerning the presence of unpredicted *ST3GAL6-AS1* fragments we investigated their nucleotide sequences. For this purpose, we isolated the single fragments by agarose gel and performed Sanger Sequencing using the same primers used in qPCR. In Figure 19 are reported representative electropherograms of the nucleotide sequencing at the border on involved exons and introns of novel *ST3GAL6-AS1* alternative splicing.

We observed:

- Retention of intron 3 in the fragments of 602 bp in Figure 16B, and 549 bp in Figure 18B with correct splicing of intron 4; in particular these two fragments were evidenced in

DELTA47 and KMS34 cell lines(Figure 19A). The Figures 16C and 18C represent a schematic view of these new alternative splicing events.

- Skipping of exon 3 in the fragments of 346 bp in Figure 16B and 288 bp in Figure 18B; in particular these two bands were evidenced in DELTA47 and KMS34 cell lines (Figure 19B). The Figures 16C and 18C represent a schematic view of these new alternative splicing events.
- Skipping of exon 2 in 289 bp band in Figure 17B. In particular, this band was evidenced in H929, MM1.S and KMS20 cell lines (Figure 19C). The Figure 17C represents a schematic view of these new alternative splicing events.

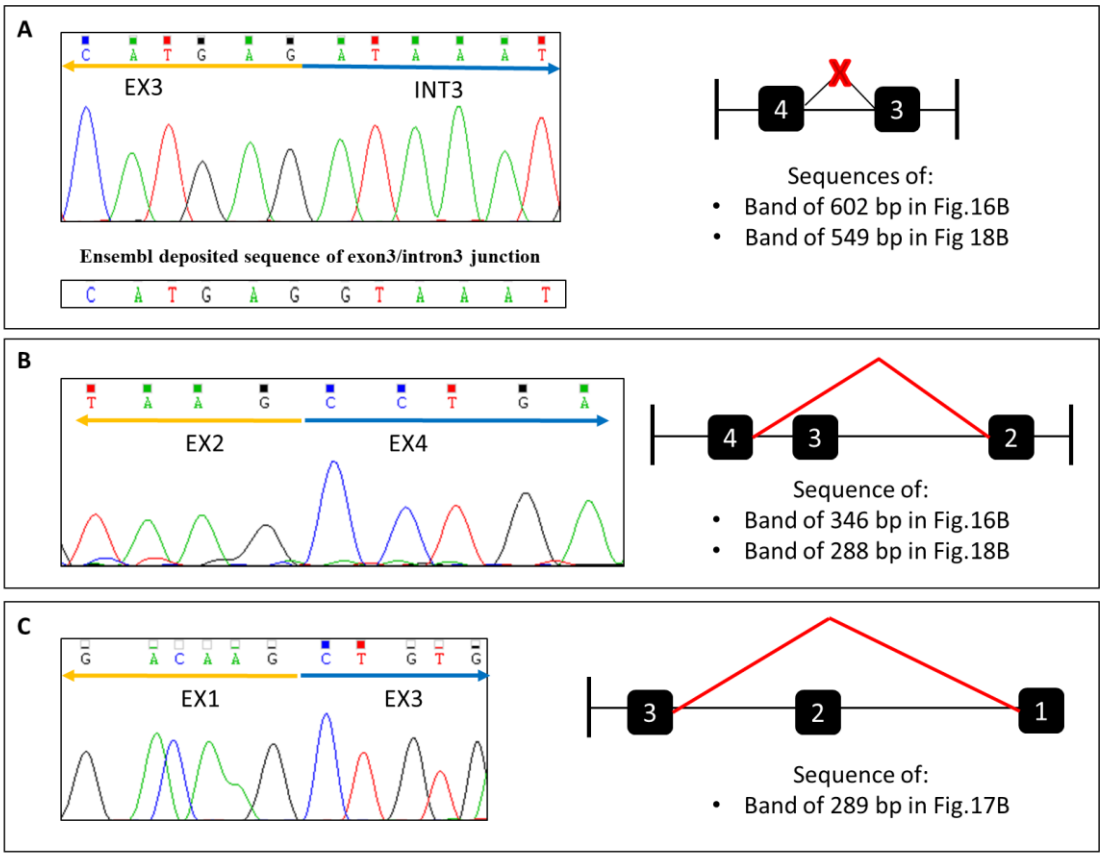



**Figure 19: *ST3GAL6-AS1* splicing variation sequences.** Examples of sequences evidenced by Sanger sequence of the unpredicted bands in the Figure 16, 17 and 18. In the left part of this figure are presented the sequence patterns while in the right there are schematic views of the unpredicted splicing observed. **A:** the sequence evidences the retention of intron 3 in the band of 602bp of Figure 16B and in the band of 549bp of Figure 18B. Below the electropherogram is indicated the Ensembl deposited sequences of exon3/intron3 junction. **B:** the sequence evidences the skipping of exon 3 in the band of 346bp in Figure 16B and in the band of 288bp in Figure 18B. **C:** the sequence evidences the skipping of exon2 in the band of 289bp of Figure 17B.

Interestingly, all the sequences analyzed evidenced the same perfect sequences of exons and introns of genomic *ST3GAL6-AS1* Ensembl sequences. The only exception was Retained intron

3 Sanger sequencing analysis showed in the first base of the intron a variation between G to A in DELTA47 and KMS34 bands (Figure 19A).

Based on this intriguing finding, we decided to sequence the genomic DNA region encompassing exon3/intron3 junction in the 6 HMCLs previously investigated by using a primer forward in exon 3 and a primer reverse in exon 4 (*ST3GAL6-AS1*_PCR1; Table 8).

The Sanger sequencing data (Figure 20) evidenced:

- in MM1.S and H929 the confirmation of Ensembl deposited sequence (G) of the first base of intron 3 (blue box);
- in KMS20 and KMS11 the presence of a double peak of G/A (orange box);
- in KMS34 and DELTA47 the presence of a G to A substitution (green box).

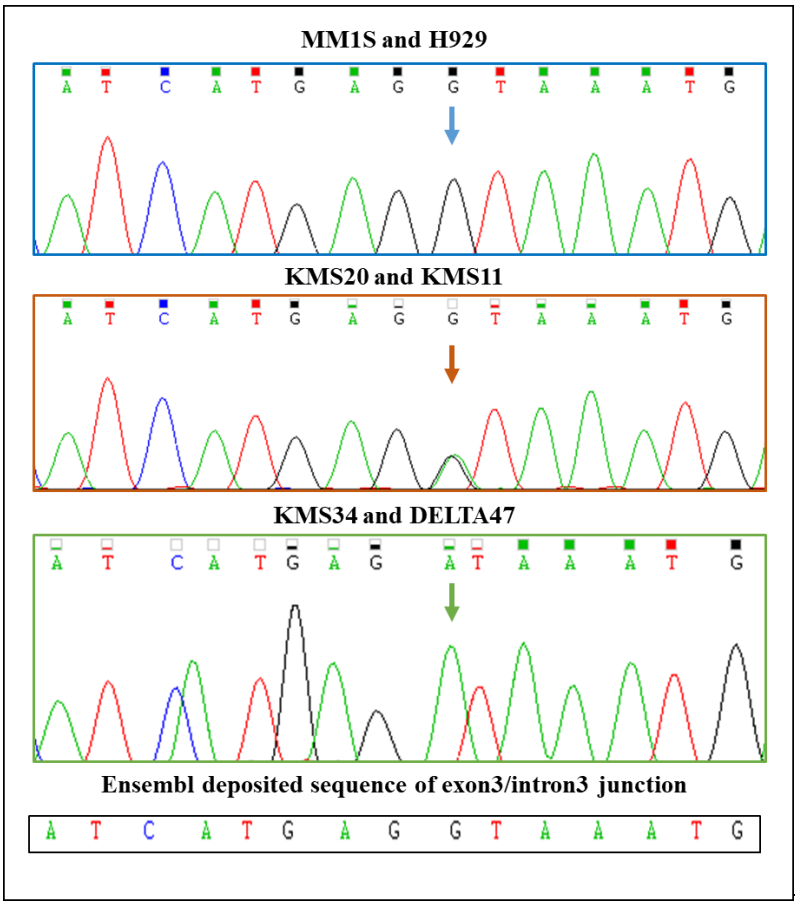



**Figure 20: DNA sequencing of exon3/intron3 junction:** Sanger sequencing analysis of exon3/intron 3 junction in HMCLs (MM1.S, H929, KMS20, KMS11, KMS34, DELTA47). We evidenced the perfect sequences in MM1.S and H929 (blue box) as the Ensembl deposited sequences (black box). KMS20 and KMS11 present a double peak in the ninth base of these represented sequences (orange box), and KMS34 and DELTA47 present the variation of the sequences in the ninth base in which we evidenced A despite to G (green box).

We matched this sequence against the Ensembl database showing that it represent a common single nucleotide polymorphism (SNP) variation, rs13065271, which may represent a splice donor variant. The analysis of the *ST3GAL6-AS1* locus identified other 5 SNPs as potential splice variants but all of them showed a very low minor allelic frequency (MAF) (see Table 12).

| Name | type of SNP | ST3GAL6-AS1 localization | ANCESTRAL/MUT | MAF |
|---|---|---|---|---|
| rs1046985161 | splice acceptor variant | intron1 | T/A | <0.01 A |
| rs187189959 | splice region variant | intron2 | G/A | <0.01 A |
| rs13065271 | splice donor variant | intron3 | C/T | 0.38 T |
| rs1026218454 | splice acceptor variant | intron3 | C/G | / |
| rs972016331 | splice region variant | exon 4 | G/C | / |
| rs1036478067 | splice acceptor variant | intron4 | T/C | <0.01 C |

**Table 12: SNP splicing variation in *ST3GAL6-AS1*.** The table shows SNP variation indicated on Ensembl database. The columns indicate SNP rs code, type of SNP, localization in *ST3GAL6-AS1* transcript, kind of variation, minor allele frequencies (MAF).

Therefore, based on our data, the evaluation by Sanger Sequencing of the presences of rs13065271 in our six HMCLs investigated was the following

- MM1.S and H929 are homozygous for the major allele (C/C);
- KMS20 and KMS11 are heterozygous (C/T);
- KMS34 and DELTA47 are homozygous for the minor allele (T/T).

## 5.2 SNP rs13065271 variation status in HMCLs

To better understand the presences of SNP rs13065271 in HMCLs we analyzed DNA with TaqMan SNP Genotyping assay. The use of this method allowed us to identify clearly the presence or absence of the variation. We analyzed 19 HMCLs including the six cell line investigated by Sanger Sequencing and we identify 7 un-variated cell lines (C/C), 7 heterozygous variated (C/T) and 5 homozygous variated cell lines (T/T) (see data in Table 13).

Interestingly, this method was able to confirm the Sanger sequencing evidences on the SNP variation rs13065271.

| C/C | C/T | T/T |
|---|---|---|
| KMS-12 BM | KMS18 | AMO1 |

| | | |
|---|---|---|
| MM.1.144 | RPMI-8226 | DELTA47 |
| U266 | KMS11 | KMS34 |
| CMA-03/06 | OPM2 | KMS-26 |
| CMA-03 | UTMC2 | INA-6 |
| MM1.S | KMS20 | |
| H929 | LP-1 | |

**Table 13: SNP rs13065271 splicing variation in *ST3GAL6-AS1*.** The table shows the results of TaqMan SNP Genotyping assay SNP on rs13065271 variation indicated on Ensembl database in MM cell lines. The columns C/C indicates un-variated cell lines, C/T a single allele variation and T/T a double allele variated cell lines.

Based on the SNP variation status we analyzed the quantitative expression of *ST3GAL6-AS1* and *ST3GAL6*. We evidenced a significant down-modulation of *ST3GAL6-AS1* and *ST3GAL6* in T/T cell lines compared to major allele homozygous cell lines. These results showed us that the presence of homozygous minor allele status significantly correlated with the down modulation of the lncRNA (Figure 21).

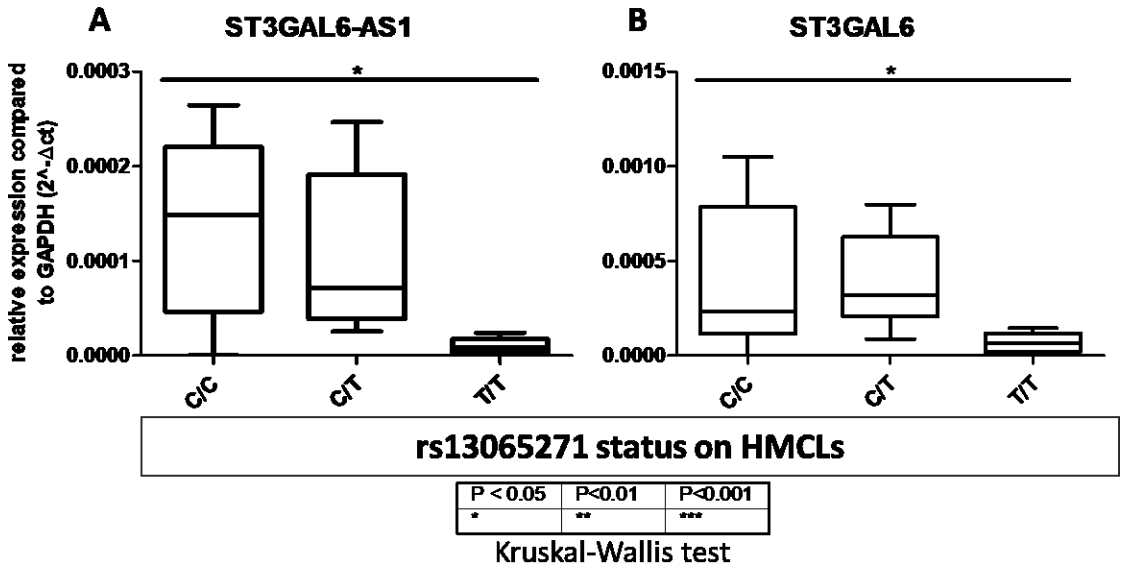



**Figure 21: box plot distribution.** Box plot distribution of *ST3GAL6-AS1* (A) and *ST3GAL6* (B) expression levels in HMCLs. HMCLs were divided based on SNP rs13065271 variation (7 un-variated cell lines (C/C), 7 heterozygous variated (C/T) and 5 homozygous variated cell lines (T/T) ). Kruskal-Wallis Test was used for the calculation of p value.

# 6. *Cellular properties of ST3GAL6-AS1*

With the aim to identify the molecular proprieties of *ST3GAL6-AS1* in HMCLs, we investigated the cellular localization in nuclear and cytoplasmic RNA fractions. MM1.S, H929, KMS20, KMS11, KMS34 and DELTA47 (representative of C/C, C/C, C/T, C/T, T/T and T/T of SNP rs13065271 splicing variation, respectively) were chosen as representative of all the HMCLs expressing *ST3GAL6-AS1* and their RNA were fractionated in nuclear and cytoplasmic portions.

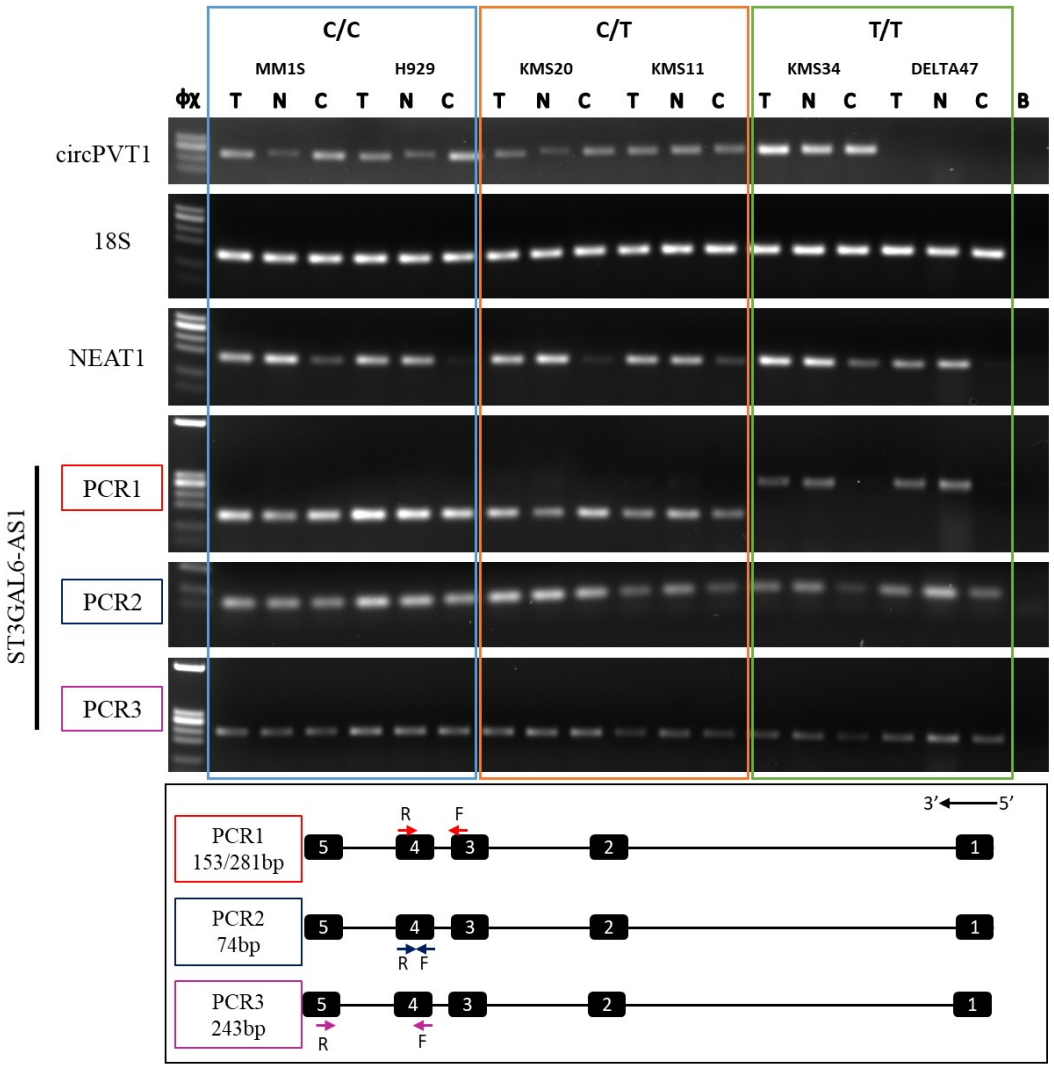
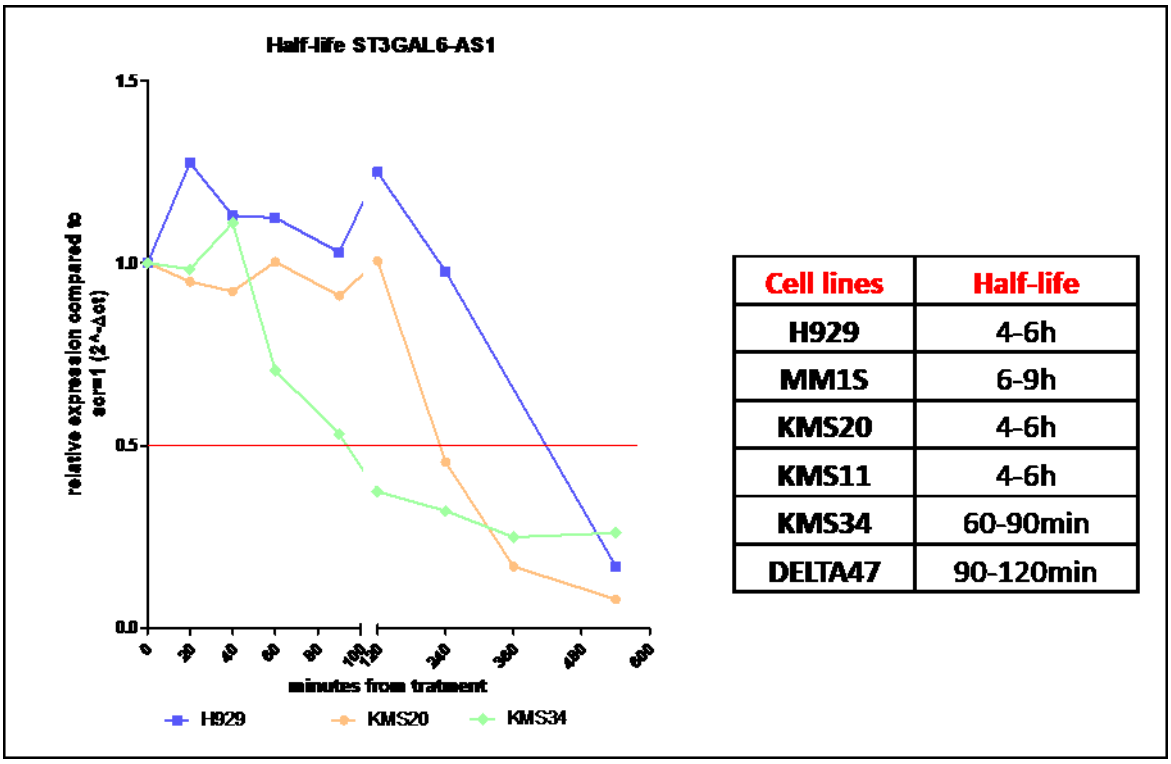



**Figure 22: RNA sub-cellular localization:** RNA of six MM cell lines were fractionated in nuclear and cytoplasmic RNA by a hypotonic buffer. qPCR with different primers were made to evaluate the localization of RNA. Blue box shows C/C cell lines, orange box shows C/T cell lines and green box shows T/T cell lines. 18S is analyzed as a co-localized control; NEAT1 as a nuclear control, circPVT1 as a cytoplasmic control. We analyzed *ST3GAL6-AS1* with three different couple of primers.

The 1.5% agarose gel on qPCR products evidenced in C/C and C/T cell lines a co-localization of the lncRNA in all the fractions. In contrast, T/T cell lines (KMS34 and DELTA47) showed a prevalent nuclear localization. In particular, the qPCR made with primers close to the intron 3 (PCR1) showed the unique nuclear localization of the isoforms (Figure 22).

In order to identify lncRNA half-life, we treated the same MM cell lines with Actinomycin D and we analyzed the RNA expression at different time-points.




| Cell lines | Half-life |
|---|---|
| H929 | 4-6h |
| MM1S | 6-9h |
| KMS20 | 4-6h |
| KMS11 | 4-6h |
| KMS34 | 60-90min |
| DELTA47 | 90-120min |

**Figure 23: *ST3GAL6-AS1* half-life:** HMCLs were treated with actinomycin D and RNAs were collected and extracted at different time points. Real-time PCR with 18s as housekeeping genes were made to evaluate half-life of *ST3GAL6-AS1*. The graph in the left of the figure shows the relative expression compared to untreated cell collected at the same time-points of three cell lines representative of homozygous major allele (C/C, H929), heterozygous (C/T, KMS20) and homozygous minor allele cell lines (T/T, KMS34). The table on the right reports the half-life range in 6 HMCLs analyzed.

The RT-PCR made on these RNAs showed that C/C and C/T cell lines (H929, MM1.S, and KMS20, KMS11) presented a stability of *ST3GAL6-AS1* transcripts with a half-life of 4-9 hours and T/T cell lines (KMS34 and DELTA47) evidenced an instability of transcripts with 60-90 minutes half-life (Figure 23).

# 7. *rs13065271 SNP variation in MM patients*

## 7.1 Correlation with Gene Expression

To better understand the role of *ST3GAL6-AS1* SNP variation in MM we analyzed 53 MM patients of which the DNA sample was available. The SNP variation appeared in homozygous status in 8 patients (T/T) while 29 present only one minor allele (C/T). Indeed, 16 patients present the Major allele in homozygous status (C/C).

The expression level of *ST3GAL6-AS1* in these patients divided following the variation status showed a significant down-modulation of the transcript in minor allele homozygous status of the SNP samples.

The expression level of *ST3GAL6* in the same patients did not display any significant difference in the three sub-groups (Figure 24).

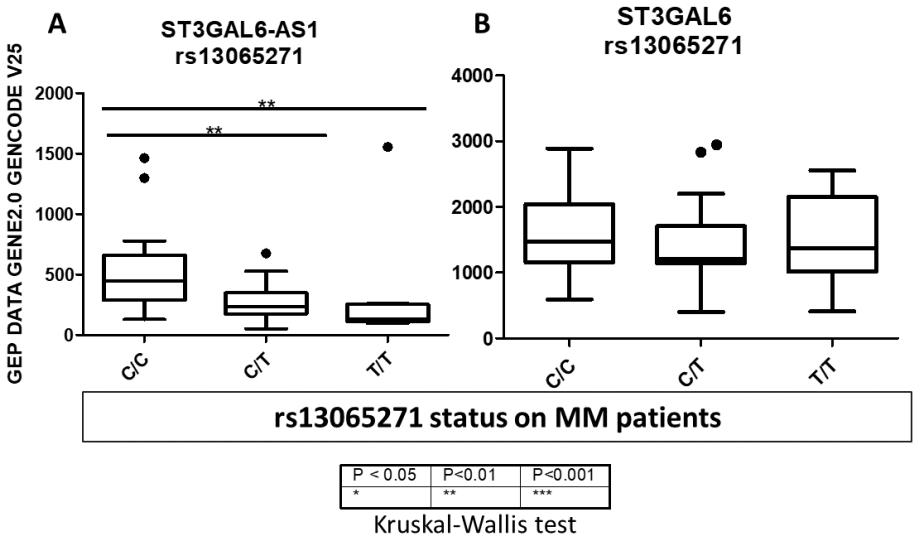



**Figure 24: box plot distribution.** Box plot distribution of *ST3GAL6-AS1* (A) and *ST3GAL6* (B) expression levels in MM patients. MM patients were divided based on SNP rs13065271 variation (29 C/C, 16 C/T and 8 T/T) . Kruskal-Wallis Test was used for the calculation of p value.

## 7.2 Correlation with MM prognosis

In our laboratory, we had the possibility to have the biological samples and the prognosis data of several MM patients. In particular, we have the Overall Survival (OS) data of 67 patients and the Time to First Relapse (TFR) of 47 patients of which DNA samples were available.

The Genotyping assay analysis of SNP rs13065271 variation in the 67 patients of which OS data were available, evidenced 17 C/C patients, 41 C/T patients and 9 T/T patients. No differences in terms of OS were evidenced after 192 months (Figure 25).

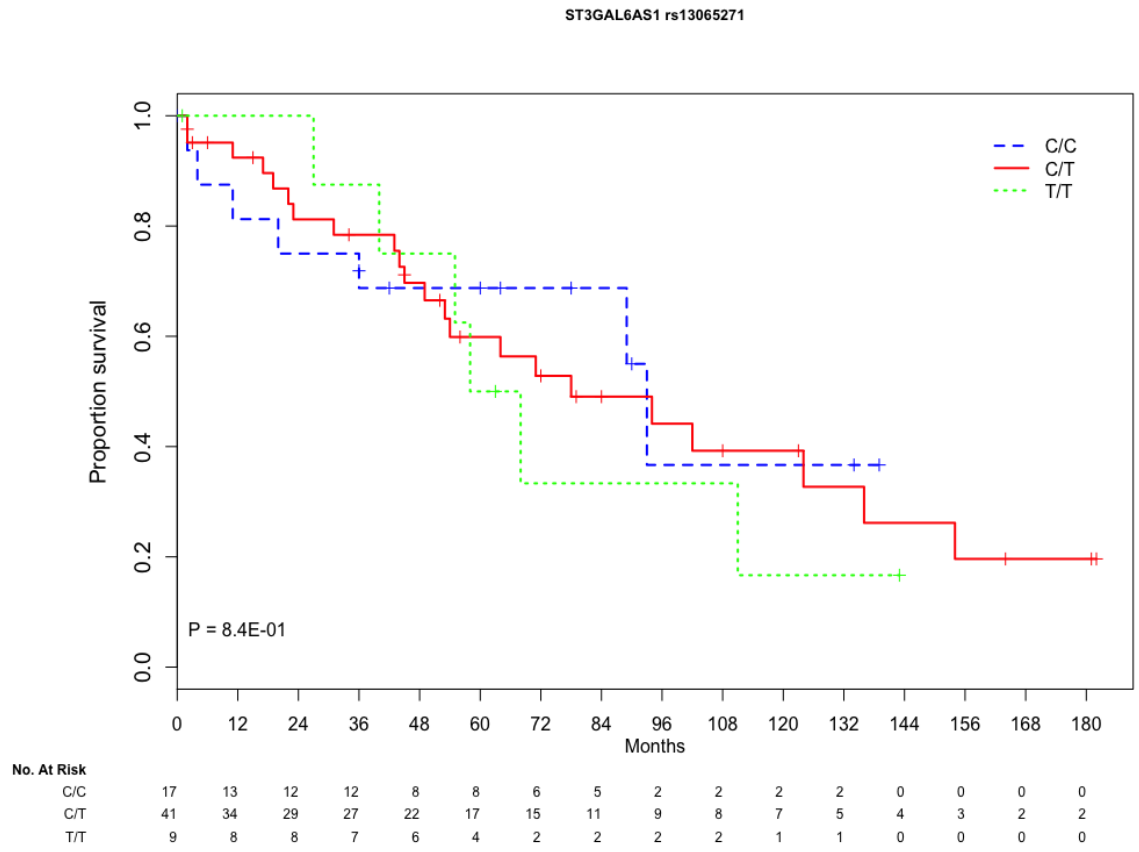



**Figure 25: Kaplan–Meier distribution.** Kaplan-meier distribution of patients divided for SNP rs13065271 variation of *ST3GAL6-AS1* in three groups: without SNP (C/C, Blue line), one-allele SNP samples (C/T, red line) and bi-allelic SNP samples (T/T, Green line). Kaplan-Meier distribution of Overall Survival (OS) data of 67 patients by which we have the DNA samples too.

The Genotyping assay analysis of SNP rs13065271 variation in the 47 patients of which TFS data were available, evidenced 12 C/C patients, 26 C/T patients and 7 T/T patients. No differences in terms of TFS were evidenced after 132 months (Figure 26).

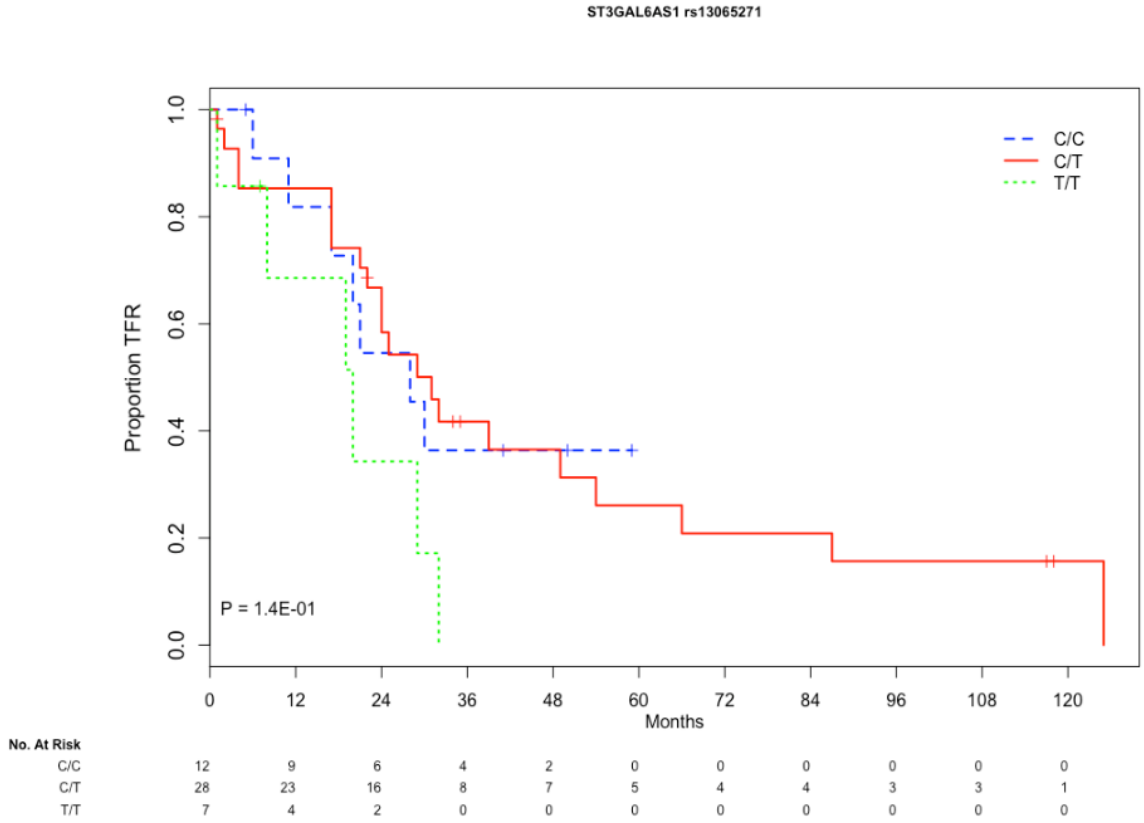
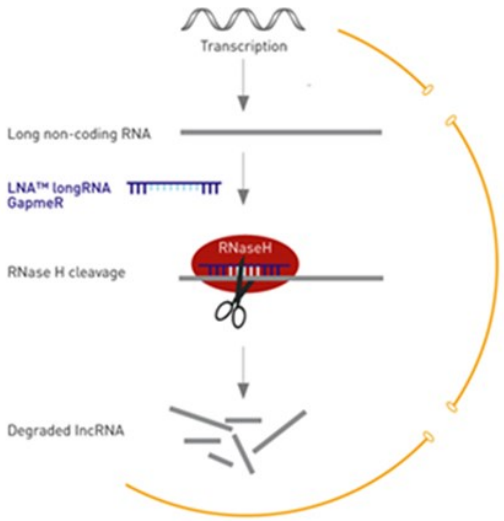



**Figure 26: Kaplan–Meier distribution.** Kaplan-meier distribution of patients divided for SNP rs13065271 variation of *ST3GAL6-AS1* in three groups: without SNP (C/C, Blue line), one-allele SNP samples (C/T, red line) and bi-allelic SNP samples (T/T, Green line). Kaplan-Meier distribution of the Time to First Relapse (TFR) of 47 patients by which we have the DNA samples too.

# 8. *ST3GAL6-AS1 silencing*

## 8.1 GapmeRs design

With the aim of investigating the functional consequences of *ST3GAL6-AS1* silencing in myeloma cells, we designed and set up the experimental conditions for the use of *LNA™ longRNA GapmeR* (GapmeR) in HMCLs. GapmeRs are specific DNA/LNA (Locked Nucleic Acid Nucleotides) hybrid molecules of 16-mer able to activate RNase H cleavage of DNA-RNA heteroduplex. Thanks to the specific sequences of DNA, GapmeRs are capable to identify the RNA of interest; the presences of LNA nucleotides, at the end of these molecules are essential for the stability and resistance to RNase H cleavage. These molecules can be used for *in vitro* and *in vivo* assay. GapmeR are able to enter cell membrane without any specific vehicles (Gymnotic delivery).




At first, we designed, by means of EXIQON tool, 4 different GapmeRs; specifically, one in exon 4, two in intron 1 (the longer intron of 13904bp) and one in intron 2 (2641bp) (Figure 27).

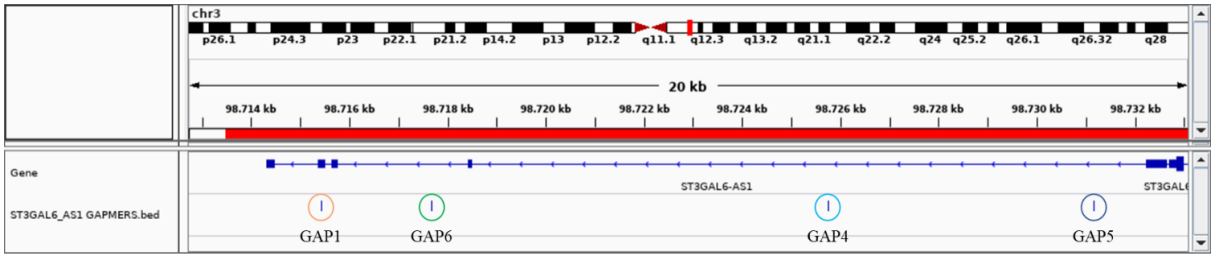



**Figure 27: schematic view of GapmeR.** In the upper part of the figure schematic view of GapmeR RNA degradation. (Image by Exiqon web site). In the lower part of the figure there is a schematic IGV visualization of *ST3GAL6-AS1* and GapmeRs we decided to use.

## 8.2 GapmeRs efficacy evaluation

To evaluate the efficacy of the 4 selected GapmeRs, to down-modulate *ST3GAL6-AS1*, we at first tested them by transfection in Hs5 and HEK-293T cell lines, by means of RNAiMAX method (see Material and Methods), which express the lncRNA at considerable levels (see

Figure 13). We made experiments in triplicate using 100 nM of each GapmeR and after 72 hours from the transfection, we extracted RNA to investigate the efficiency of GapmeR to silencing *ST3GAL6-AS1*. We evidenced a significant down-modulation of the lncRNA between 20 to 70% (left part of Figure 28), and GapmeR 4 giving the best results in terms of silencing.

Based on these preliminary data, we tested all four GapmeRs in 3 HMCLs (H929, LP1 and U266). Since HMCLs are difficult to transfect, we decided to use the Neon Transfection system, a method able to create membrane temporary pores. After 6 hours from the transfection[119], we evaluated by RT-PCR the expression level of *ST3GAL6-AS1*. We found a down-modulation between 30-75%, confirmed after 72 hours (right part of Figure 28).

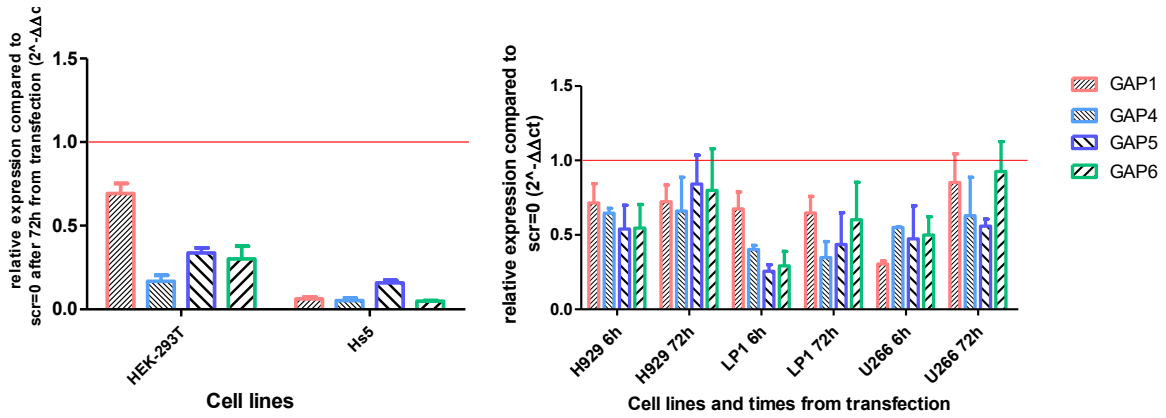



**Figure 28: histograms of expression levels of *ST3GAL6-AS1* in HMCLs treated with *ST3GAL6-AS1* GapmeRs.** Silencing of *ST3GAL6-AS1* by using specific GapmeRs showed in Figure 27. In the left part of the figure we showed relative expression levels of *ST3GAL6-AS1* after 72 hours post RNAiMAX transfection in HEK-293T and Hs5. In the right part of the figure we showed relative expression levels of *ST3GAL6-AS1* after 6 and 72 hours from NEON transfection of H929, LP1 and U266. Silencing was showed in comparison with GapmeR SCR transfection (SCR=1 showed by red lines; $2^{-\Delta\Delta Ct}$ method).

All the four GapmeRs showed a good efficiency in down-modulation of *ST3GAL6-AS1* in HMCLs. However, the *ST3GAL6-AS1* down-modulation made by GapmeR4 exhibited the ability to keep the silencing of the lncRNA for more than 72 hours. For this reason, we choose GapmeR 4 to perform functional experiment in HMCLs by gymnotic delivery.

## 8.3 Functional effect of *ST3GAL6-AS1* silencing

Based on our experience on lncRNA silencing, we ordered *LNA™ longRNA GapmeR in vivo Ready* with the sequence of GapmeR 4. This kind of GapmeR can be used for gymnotic Delivery [120] and *in vivo* experiments. Gymnotic delivery is used in low-density seeded cells and in this way, we are able to analyze cells after day 7-9 from the GapmeR treatment.

We treated with gymnotic delivery of GapmeR4 H929 and U266, with homozygous major allele of SNP rs13065271 variation (C/C), KMS20 and LP1, with heterozygous status (C/T) and DELTA47 and KMS34, with homozygous minor allele (T/T). After treatments of these 6 HMCLs with 5µM of GapmeR for 7-9 days, we evidenced a significant down-modulation of transcripts from day 2 to day 9 (≈95% of silencing). As a representative example, in Figure 29 are shown the expression levels of *ST3GAL6-AS1* at days 5 after silencing; for completeness, we evaluated the expression levels of *ST3GAL6* and *DCBLD2* genes, which did not show any modulation.

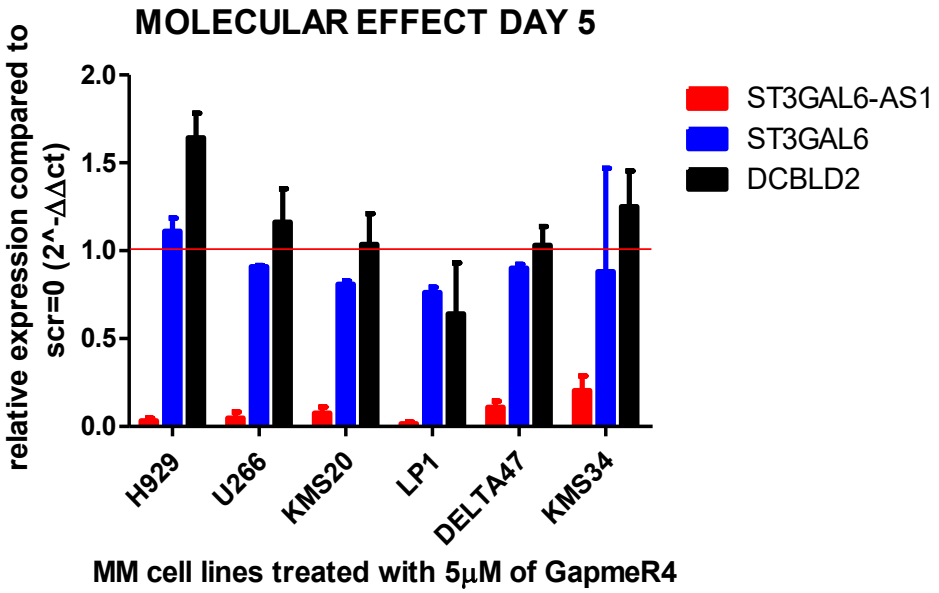



**Figure 29: *ST3GAL6-AS1*, *ST3GAL6* and *DCBLD2* expression levels after *ST3GAL6-AS1* silencing.** Histogram shows relative expression level of *ST3GAL6-AS1*, *ST3GAL6* and *DCBLD2* after 5 days of GapmeR 4 treatment, by gymnotic delivery. Silencing was showed in comparison with GapmeR SCR transfection (SCR=1 showed by red lines; $2^{-\Delta\Delta Ct}$ method).

In addition, samples cell count, in all the days of treatment, showed in H929, U266, KMS20 and LP1 (representative of C/C, C/C, C/T, and C/T of SNP rs13065271 splicing variation, respectively) significant biological effect in terms of cell growth after day 4, and in term of viability after day 5. In DELTA47 and KMS34, representative of SNP rs13065271 splicing variation T/T cells, we did not evidence any kind of biological effect (Figure 30). These biological effects, at day 5 and 6 after GapmeR treatment, were confirmed by luminescent assay for cell growth, showing the decreasing of cell growth in H929, U266, KMS20 and LP1 after *ST3GAL6-AS1* silencing and no effect for DELTA47 and KMS34 (Data not show).

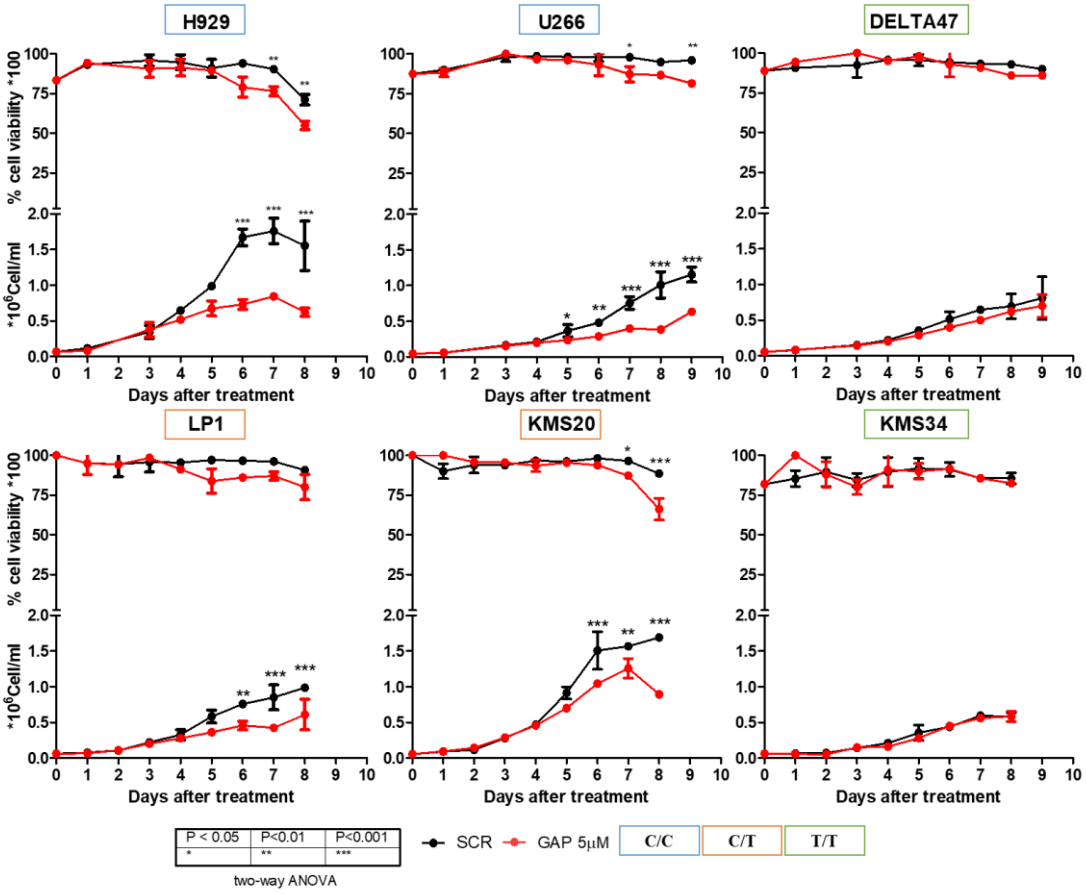



**Figure 30: percentage of viability and cell growth graph after *ST3GAL6-AS1* silencing:** this figure shows biological effect of GapmeR 4 treatment, by gymnotic delivery. In the upper part of these graphs is reported the percentage of cell viability and in the lower part cell growth curves. In black are reported SCR treated and in red GapmeR treated cells. Name of the cell lines are in box. The colored boxes indicate the SNP rs13065271 variation status in H929 and U266 (C/C), KMS20 and LP1 (C/T), and DELTA47 and KMS34 (T/T). (SCR: GapmeR negative control treatment and GAP: GapmeR 4 treatment)

To better understand the biological effect of *ST3GAL6-AS1* silencing, we analyzed, by flow cytometry, cell cycle phases after 5 and 6 days of GapmeR treatment. In particular, H929, U266, KMS20 and LP1 showed an increase of cells in sub G0/G1 phases and a decrease of percentage of cells in S-G2/M phases (Figure 31). DELTA47 and KMS34 cell lines did not evidence any cell cycle variation after *ST3GAL6-AS1* silencing (Data not shown).

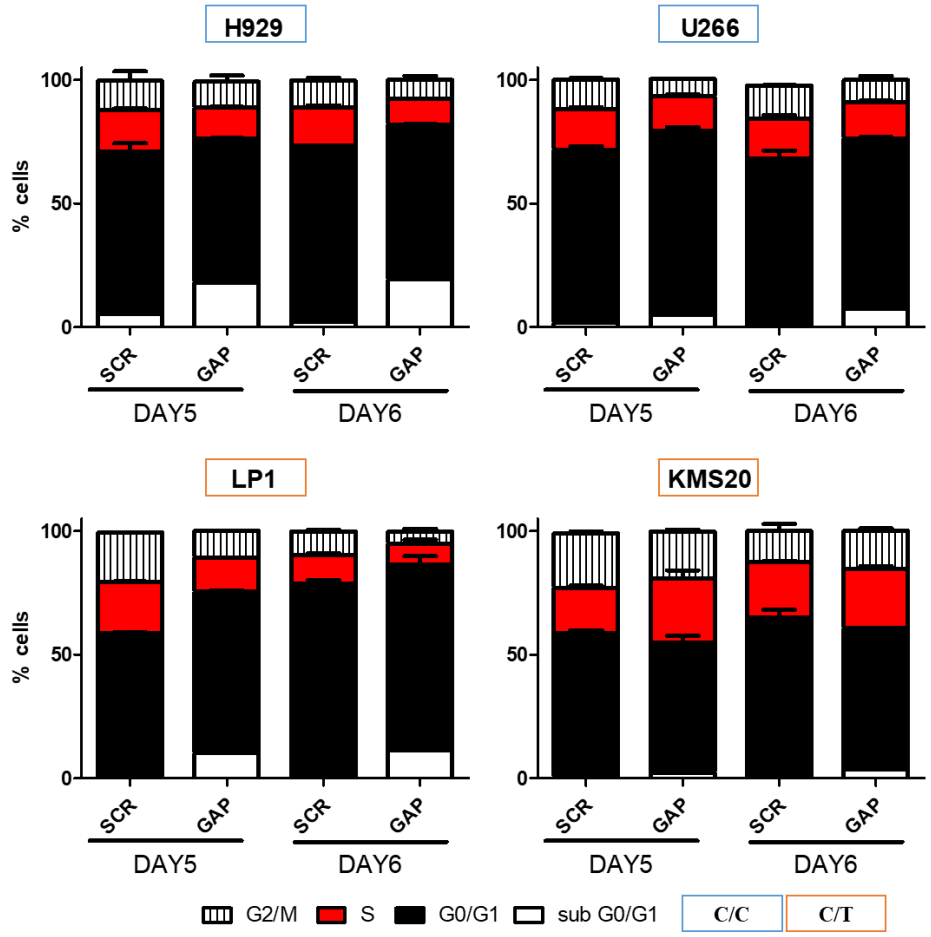



**Figure 31: cell cycle after *ST3GAL6-AS1* silencing:** This figure shows biological effect in terms of cell cycle of GapmeR 4 treatment, by gymnotic delivery at day 5 and 6. Percentage of cell in the 4 phases of cell cycle are reported in y-axis. Name of the cell lines are in boxes. The color of the boxes are indicative of the SNP rs13065271 variation status (H929 and U266 are C/C, KMS20 and LP1 are C/T). (SCR: GapmeR negative control treatment and GAP: GapmeR 4 treatment)

Furthermore, H929, C/C cell line for SNP rs13065271 splicing variation and LP1, C/T cell line, were investigated for apoptotic cell percentage, with flow cytometry test, at day 5 and 6 after *ST3GAL6-AS1* silencing. The two HMCLs showed an increase of percentage of apoptotic cells in particular after day 6 albeit not reaching statistical significance (Figure 32).

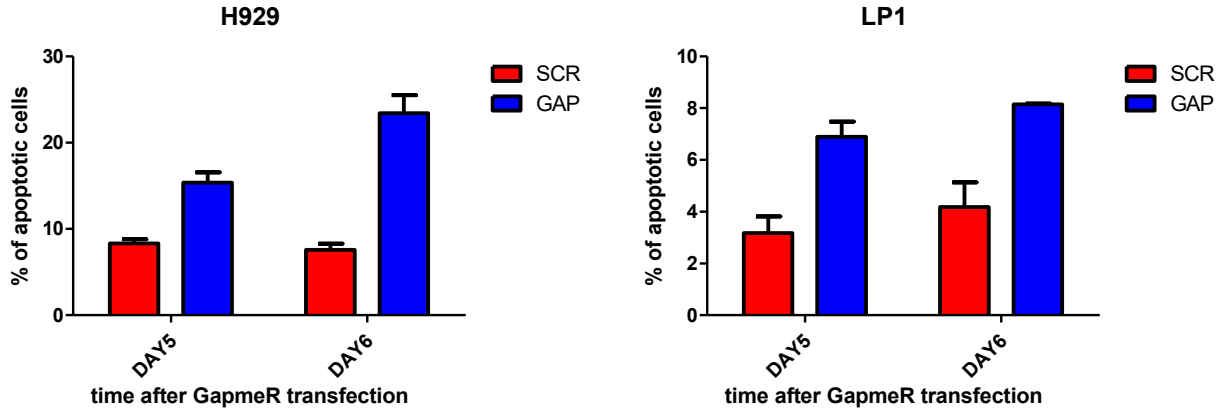



**Figure 32: percentage of apoptotic cells after *ST3GAL6-AS1* silencing:** these histograms show biological effect, in terms of apoptotic cell, of GapmeR 4 treatment, by gymnotic delivery at day 5 and 6. Percentage of apoptotic cells are reported in y-axis. (SCR: GapmeR negative control treatment and GAP: GapmeR 4 treatment)

Finally, we analyzed the colony formation ability of LP1 and H929 cells after 12 days and 20 days respectively, from *ST3GAL6-AS1* silencing. Notably, silenced LP1 and H929 cells were not able to make colonies (Figure 33).

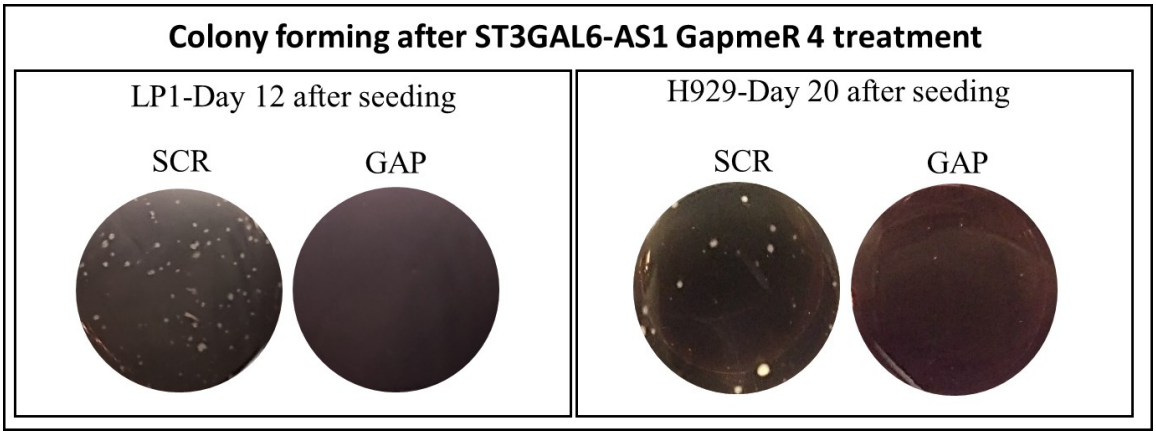



**Figure 33: LP1 and H929 colony formation assay after *ST3GAL6-AS1* Silencing:** These image shows biological effect, in terms of colony formation, of GapmeR 4 treatment, in the left part shows LP1 colony formation after 12 days from gymnotic delivery; in the right part shows H929 colony formation after 20 days from gymnotic delivery. Experiments were performed in triplicate. (SCR: GapmeR negative control treatment and GAP: GapmeR 4 treatment)

# 9. *Gene expression analysis of LP1 and H929 after ST3GAL6-AS1 silencing*

## 9.1 GEP assays on H929 and LP1 *ST3GAL6-AS1* silenced

H929, as representative of HMCLs cell lines major allele homozygous for SNP rs13065271, and LP1, as representative of heterozygous cell lines, were treated with GapmeR4 by gymnotic delivery and after 4 days from the treatment RNA were extracted and analyzed by GeneChip® Human Gene 2.0 ST array.

The *ST3GAL6-AS1* silencing was confirmed by GEP data in both the cell lines (Figure 34). For completeness, we evaluated the expression levels of *ST3GAL6* and *DCBLD2* genes, which did not show any modulation as in RT-PCR data (Figure 34).

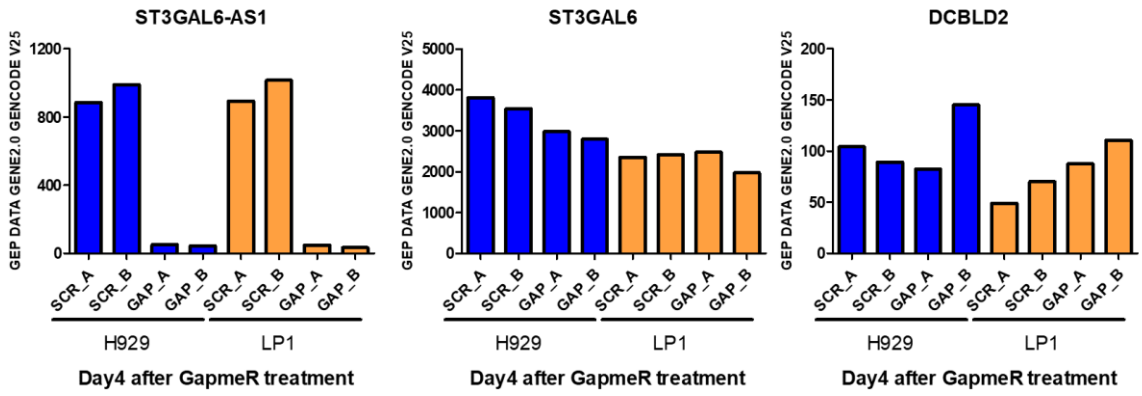



**Figure 34: GEP data of *ST3GAL6-AS1* and its neighboring genes:** The histograms show expression levels of *ST3GAL6-AS1*, *ST3GAL6* and *DCBLD2* in H929 (colored in blue) and LP1 (colored in orange) after treatment by gymnotic delivery of GapmeR. (SCR: GapmeR negative control treatment and GAP: GapmeR 4 treatment; A and B are the two replicate) we evidence ≈97% of *ST3GAL6-AS1* silencing.

## 9.2 Supervised analysis of silenced-*ST3GAL6-AS1* versus control in LP1 and H929

With the aim to identify the role of *ST3GAL6-AS1* in HMCLs, we analyzed our silenced samples and controls with Rank Product analysis (10% percentage of false positives (pfp), RANK PROD). This is a supervised analysis by which we analyzed H929 and LP1 samples separately.

In particular, in LP1 we evidenced 26 modulated genes, 5 up-regulated and 21 down-regulated, after *ST3GAL6-AS1* silencing. In H929 we evidenced 73 modulated genes, 49 up-regulated and 24 down- regulated, after *ST3GAL6-AS1* silencing (Figure 35).

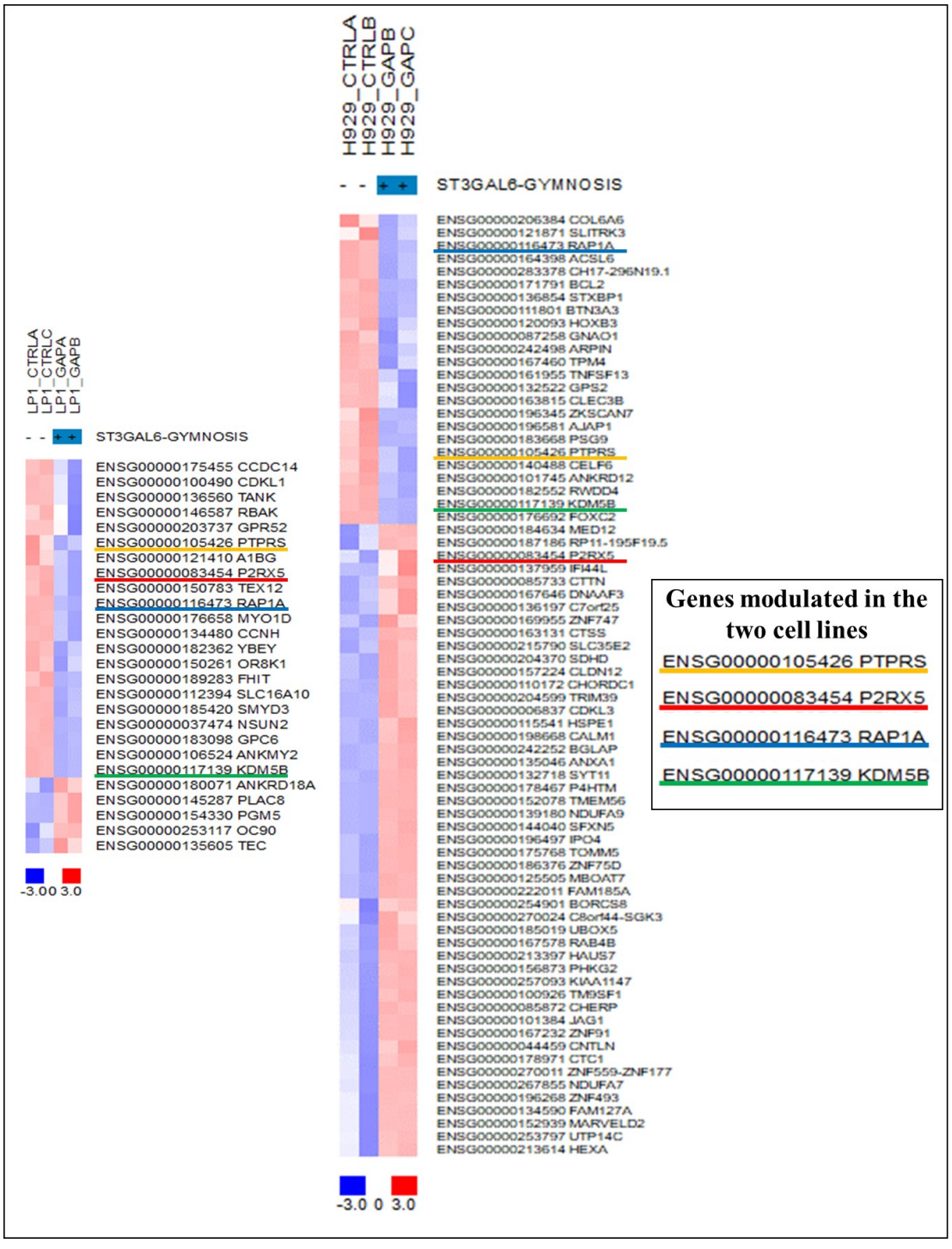



**Figure 35: RANK PROD analysi.:** The figure shows RANK PROD test made with FDR 0.1 on LP1 and H929 GEP data after *ST3GAL6-AS1* silencing. Black boxes evidence with different colors the genes resulted modulate in both the HMCLs analyzed.

Among these modulated genes, we looked at common modulated ones. We evidenced PTPRS, RAP1 and KDM5B down-regulated genes after *ST3GAL6-AS1* silencing in both cell lines analyzed, while P2RX5 is down-regulated in LP1 and up-regulated in H929 (Figure 35 and Figure 36).

The expression levels of these 4 genes in our cohort of patients, described previously in paragraph 3 of Results, did not correlate in any way with the expression levels of *ST3GAL6-AS1*.

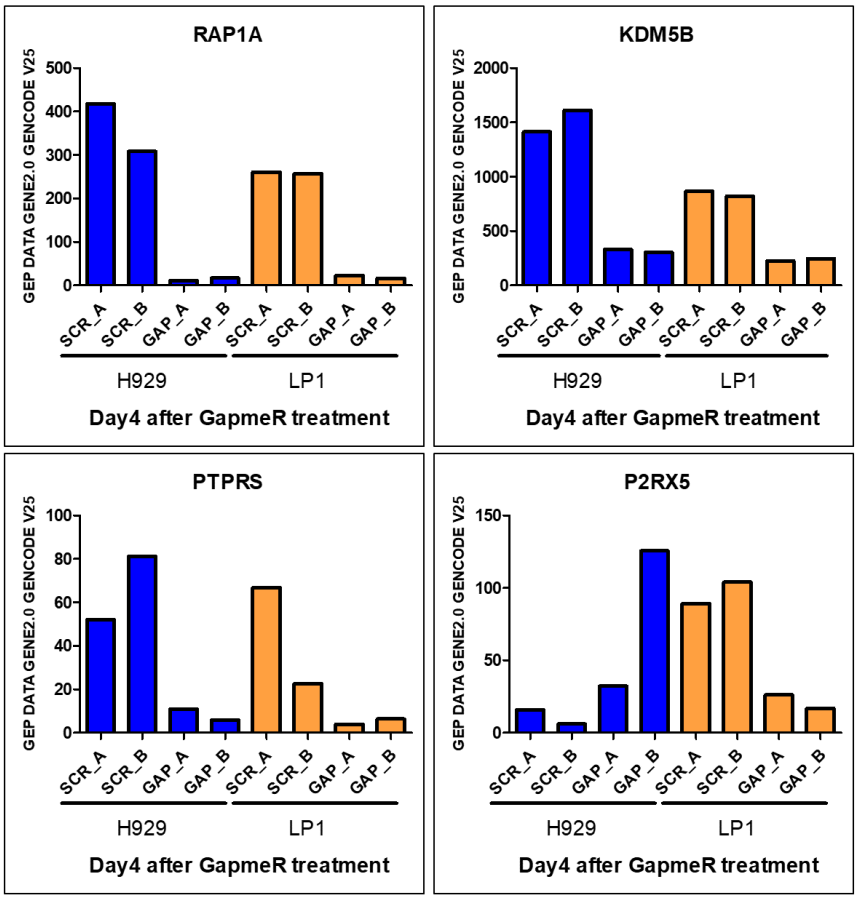



**Figure 36: GEP data of genes modulate in Figure 35.** The histograms show expression levels of RAP1A, KDM5B, PTPRS and P2RX5 in H929 (colored in blue) and LP1 (colored in orange) after treatment by gymnotic delivery of GapmeR. (SCR: GapmeR negative control treatment and GAP: GapmeR 4 treatment; A and B are the two replicate).

## 9.3 GSEA analysis on global gene expression data respectively in LP1 and H929 cell lines

GSEA analyses were performed on both LP1 and H929 cell lines, by using different collections of gene sets that were representative of molecular pathways (KEGG, REACTOME), curated from literature (myeloma or leukemia associated), or obtained by applying computational

methodologies, finding coordinately expressed genes in association to specific biological states (hallmark).

Five common gene sets were found significantly modulated in both treated cell lines: the upregulated *ARMS mediated activation* gene set and the downregulated *p53*, *lysosome actions*, *other glycan degradation*, and *NOTCH signaling*.

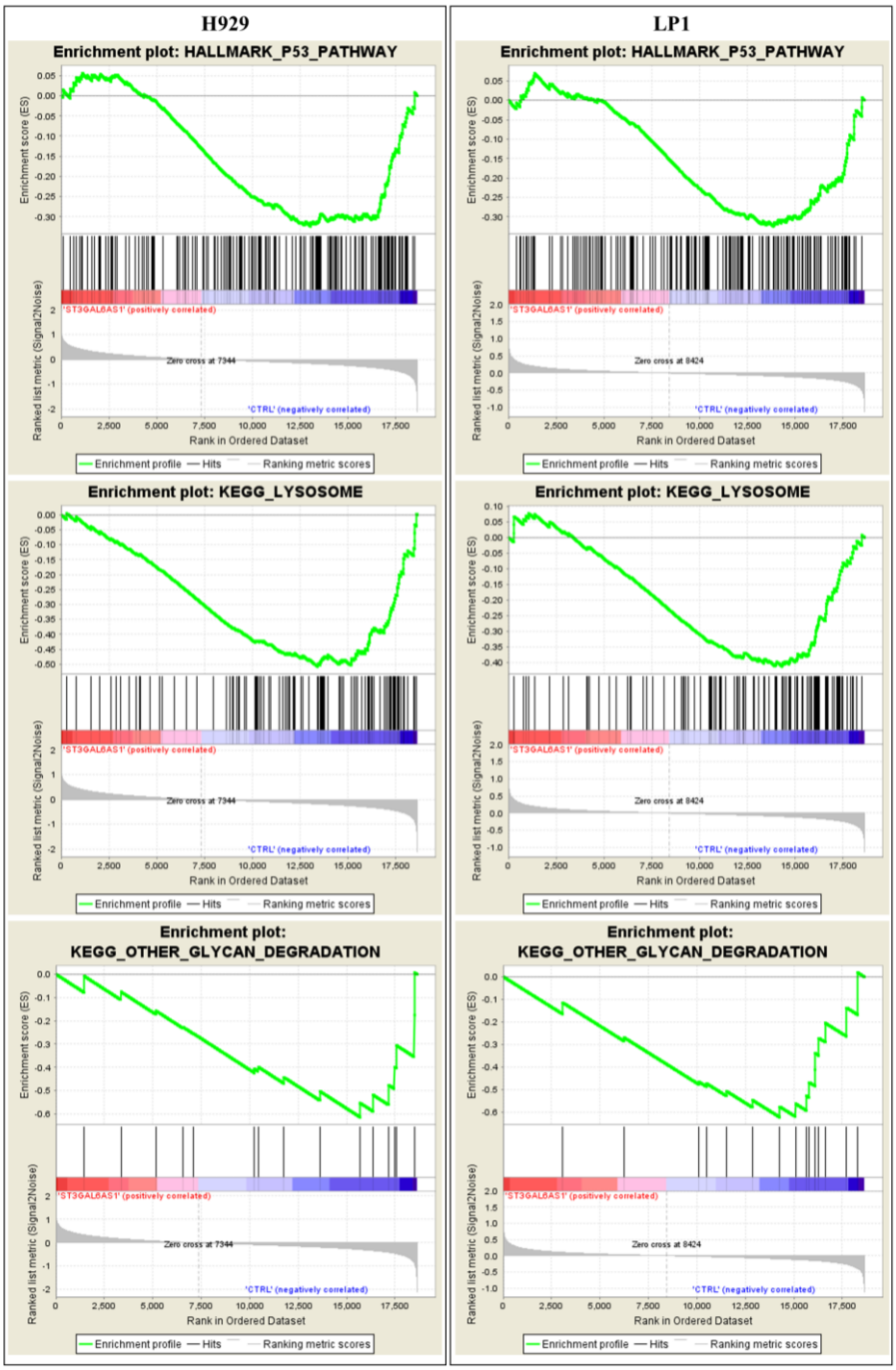



**Figure 37: GSEA enrichment analysis of the modulated genes in *ST3GAL6-AS1* silenced LP1 and H929.** The figure shows GSEA enrichment analysis of the modulated genes in *ST3GAL6-AS1* silenced LP1 and H929 after 4 days from treatment. The pathways analyzed are

in the upper part P53 pathway, in the middle part lysosome and in the lower part other glycan degradation pathway. All these pathway result down-modulated.

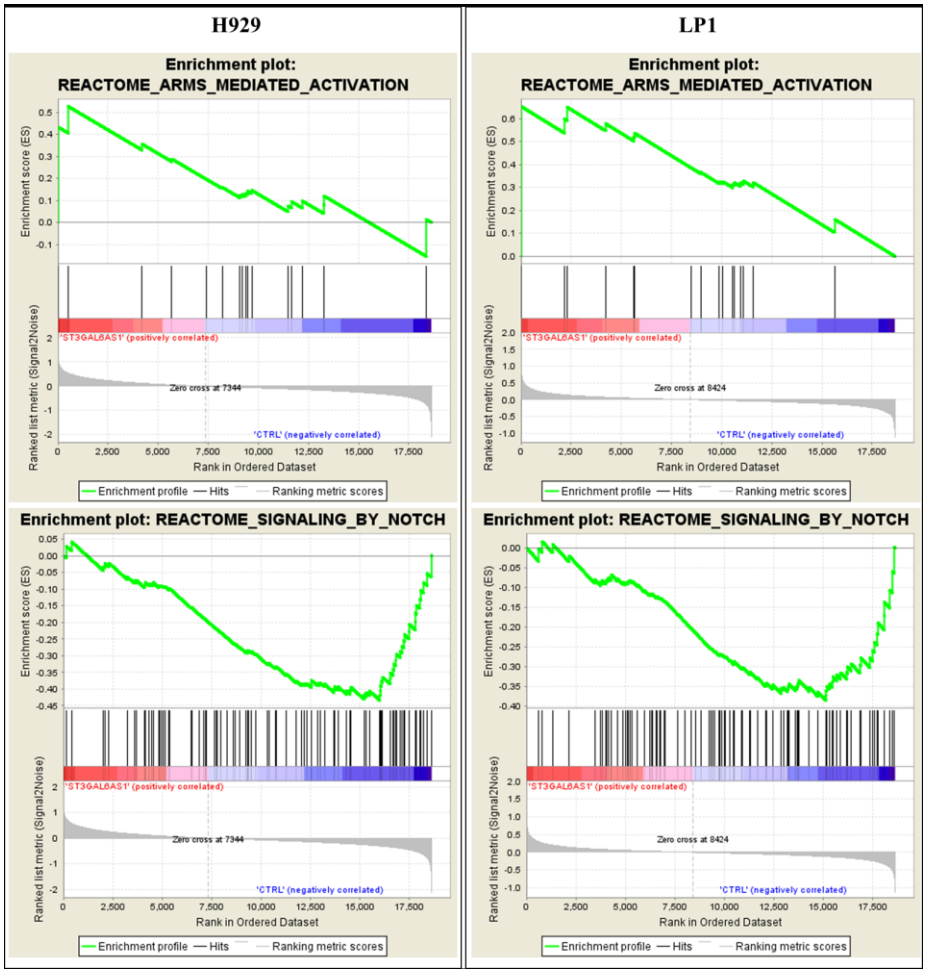



**Figure 38: GSEA enrichment analysis of the modulated genes in *ST3GAL6-AS1* silenced LP1 and H929.** The figure shows GSEA enrichment analysis of the modulated genes in *ST3GAL6-AS1* silenced LP1 and H929 after 4 days from treatment. The pathways analyzed are in the upper part ARMS mediated activation and in the lower part NOTCH signaling. Only ARMS mediated activation resulted up-modulated while the other is down-modulated after *ST3GAL6-AS1* silencing.

Interestingly, a more careful analysis of these five gene sets showed the common deregulation of *FUCA1* in three of them (*P53 pathway*, *lysosome* and *other glycan degradation)*. The analysis of *FUCA1* expression levels in our cohort of MM patients evidenced a significant down-modulation of the transcript of this gene in pathological samples compared to normal controls (Figure 39B). Notably, *FUCA1* showed an opposite pattern of expression in MM samples compared to *ST3GAL6-AS1* (see above in Figure 12A). Additionally, *FUCA1* expression levels in MM patients resulted significant anti-correlated with *ST3GAL6-AS1* expression (r=-0.28; Figure 39C).

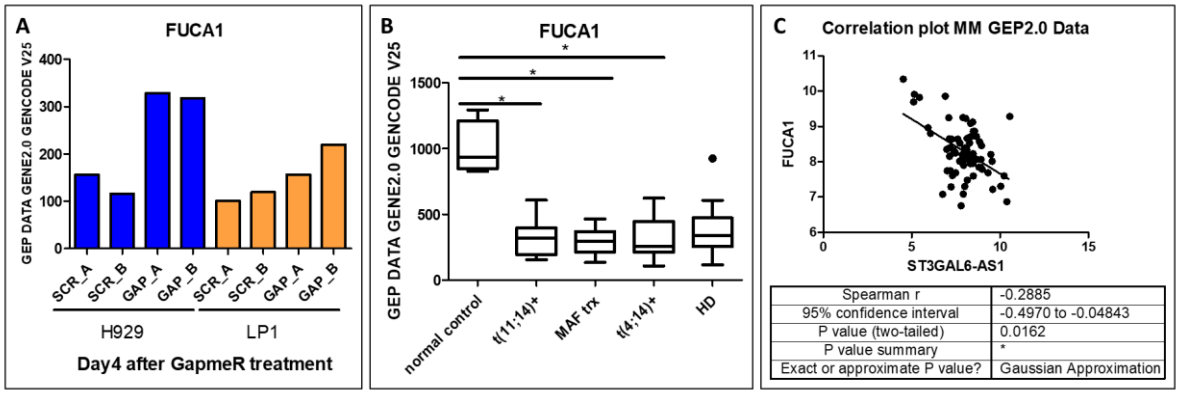



**Figure 39: *FUCA1* expression levels and correlation. A:** the histograms show expression levels of *FUCA1* in H929 (colored in blue) and LP1 (colored in orange); after treatment by gymnotic delivery of GapmeR. (SCR: GapmeR negative control treatment and GAP: GapmeR 4 treatment; A and B are the two replicate). **B:** box plot distribution of *FUCA1* expression levels in our cohort of patients; patients were divided for the presence of hyperdiploidy status (17 patients), t(11;14) translocation (15 patients), translocation of MAF genes (11 patients) or t(4;14) translocation (15 patients). Kruskal-Wallis Test was used for the calculation of p value (*→pvalue<0.05). *FUCA1* is downregulated in MM and PCL samples. C: correlation plot between *ST3GAL6-AS1* and *FUCA1* expression levels in our cohort of MM patients.

# Discussion

MM is the second common hematologic malignancy in the US[121]. Even if in the last years new drugs and therapeutic regimens improved the treatment with response rates approaching 100%[122,123], MM remains an incurable disease.

Since 2008 different Authors evidenced the role of lncRNAs in MM progression and pathology. LncRNAs are transcripts longer that 200 nt, unable to encode protein. LncRNAs have crucial role in a variety of processes accompanying the biology of normal and cancer cells. Several examples of lncRNAs evidenced their possible potential usage as biomarkers in the diagnosis and prognosis of many cancer types because of their tissue-specific expression[124].

Importantly, Samur K.M. et al. have recently investigated the correlation between lncRNAs deregulation by RNAseq method and disease outcome on MM patients. They identified a signature composed by protein coding genes and lncRNAs along with traditional risk features that can identify ultra-high-risk patients, as well as those who could potentially have excellent chances of survival[125]. Even if many lncRNAs were defined as deregulated in MM, only a few of them have been, at least in part, characterized.

In our laboratory, we firstly analyzed lncRNAs expression levels in MM using the GeneChip® Human Gene 1.0 ST microarray (Affymetrix) able to distinguish more than 1800 unique lncRNAs. Thanks to this method we identified 21 lncRNAs whose expression was progressively deregulated through the more aggressive stages of PC dyscrasia (Ronchetti et al 2016)[85]. In the last years, the number of lncRNAs is progressively increased to more than 15000. Based on this evidence, we set up the conditions to investigate lncRNA expression in MM by using the GeneChip® Human Gene 2.0 ST microarray (Affymetrix) able to distinguished more than 10000 lncRNAs and the RNAseq approach.

The analysis performed here on patients affected by PC dysrasias (50 MM and 15 PCL patients samples along with 4 normal controls) focused our attention on *ST3GAL6-AS1* as the unique lncRNA up modulated in pathological samples compared to normal controls (Figure 9). Of note, recently Shen et al. described the up-modulation of *ST3GAL6-AS1* by microarray analysis in a very limited series of 5 MM patients.

Furthermore, RNAseq analysis, made on 30 MM patients included in the same GEP analysis, evidenced the expression of all the predicted 5 exons of *ST3GAL6-AS1* with no difference between the molecular subgroups of MM (Figure 10).

In 2010, Ørom UA et al., described that after the depletion of lncRNA loci from genome, a down-modulation of the expression levels of genes mapping in the same genomic regions were obtained[126]. Additionally, Guil S. and Esteller M summarized that antisense lncRNAs generally act in the regulation of their sense gene expression[127]. Based on this notion, we analyzed the expression levels of *ST3GAL6-AS1,* which maps in 3q12.1 in the reverse strand, and two genes, *ST3GAL6* and *DCBLD2* located within 2 Mbp encompassing the *ST3GAL6-AS1* locus. The *ST3GAL6* gene is located in the sense strand and in a head to head orientation with the lncRNA; the *DCBLD2* gene is located in the reverse strand at 63 Kbp from *ST3GAL6-AS1* (see schematic view of Figure 11A). Notably, Glavey et al., recently described *ST3GAL6* as an oncogene for MM. ST3GAL6 is able to act in protein and lipid glycosylation. Its action resulted essentially for the homing and engraftment of MM[117].

The expression levels of these protein-coding genes by GEP, showed that *ST3GAL6* was significantly correlated with *ST3GAL6-AS1* in our MM cohort of patients while *DCBLD2* did not correlate. Additionally, the analysis of the *ST3GAL6-AS1* and *ST3GAL6* expression levels in MM patients did not display any significant difference between the molecular groups of MM patients. The only significant difference was between pathological samples and normal controls (see box plot in Figure 12).

The investigation of the expression levels of *ST3GAL6-AS1* and its neighboring genes in HMCLs confirmed the significant correlation with *ST3GAL6* but not with *DCBLD2*. Furthermore, we investigated the same expression levels in a limited number of non-hematological cancer cell lines. By this analysis, we obtained a significant overexpression of *ST3GAL6-AS1* in HMCLs compared to other non-hematological cell lines (Figure 13).

Focused on the possible role of *ST3GAL6-AS1* in MM pathology, we decided to start our investigation trying to understand its splicing events occurring in HMCLs. It is an accepted notion that alternative splicing affected 90% of multi-exon human genes[128]. In GENCODE annotation, *ST3GAL6-AS1* is predicted to generate three isoforms that are produced by alternative splicing of the 5 exons (see schematic representation in Figure 11B). Based on this evidence we decided to validate the presence of these isoforms in HMCLs by qPCR. Interestingly, in all the HMCLs tested unpredicted alternative splicing isoforms (Figure 14) were revealed, thus stratifying HMCLs in three groups based on the observed pattern (MM1.S, H929, U266 and CMA-03 in first group in blue box in the figure; KMS18, KMS11, KMS20

and RPMI-8226 second group in orange box of the figure and KMS34, KMS26, INA-6 and DELTA47 third group in green box of figure, Figure 14).

This investigation, along with Sanger sequencing, confirmed all the Ensembl predicted splice events (see Figure 14-18) with the addition of 2 novel splice events: the skipping of exon 3 (Figure 19B) and the retention of intron 3 (Figure 19A) in *ST3GAL6-AS1* fragments *in* DELTA47 and KMS34 cell lines.

The Sanger analysis of the fragment with the retention of intron 3 revealed a variation between G to A in the first base of the intron 3 (Figure 19A). Thus, we investigated the genomic sequences of this intron. DNA Sanger sequencing (Figure 20) evidenced a homozygous variation G to A in DELTA47 and KMS34 (cell lines of green box in Figure 14), a heterozygous variation in KMS20 and KMS11 (cell lines of orange box in Figure 14), and no variation in MM1.S and H929 (cell lines of blue box in Figure 14). In-depth analysis of Ensembl database revealed that this change matches with the rs13065271 SNP variation, reported to have a MAF of 0.38. Based on these data, we can suggest that the presence of homozygous or heterozygous rs13065271 minor allele affects significantly the splicing events in *ST3GAL6-AS1*.

Importantly, several studies have found that SNPs in lncRNAs are linked to their possible abnormal expression and dysregulation[129,130]. The examination of SNP rs13065271 variation in 19 HMCls evidenced the presence of the homozygous minor allele in 5 (26%) and a heterozygous condition in 7 (37%) of the cell lines. Notably, our analysis showed a significant down regulation of *ST3GAL6-AS1* in HMCLs harboring a homozygous minor allele condition (Figure 21).

The study of lncRNAs in the last years has been focused on the investigation of their role. An important association has been found between function and subcellular localization[131]. In our study we reported that *ST3GAL6-AS1* resulted equally localized in nucleus and cytoplasmic compartments. Additionally, the analysis of the subcellular localization in HMCLs carrying the homozygous minor allele evidenced that in these cell lines the presence of rs13065271 modified *ST3GAL6-AS1* subcellular localization with a prevalent nuclear localization (Figure 22).

Another important property of lncRNAs is the half-life. In 2012, Tani H et al. showed the importance of lncRNAs half-life in the characterization of their role[132]. Our analysis showed that *ST3GAL6-AS1* half-life could be modulated by the presence of rs13065271 SNP variation,

which appears to be highly correlated with the loss of stability of the lncRNA transcripts (Figure 23).

Furthermore, the SNP rs13065271 analysis made on MM patients confirmed the significant down-regulation of *ST3GAL6-AS1* expression levels in homozygous minor allele patients who represent 15% of our analyzed cohort (Figure 24).

Along with the characterization of the structural and molecular features of *ST3GAL6-AS1,* we decided to investigate its functional role by down-modulating its expression levels in HMCLs by GapmeR approach. We obtained a remarkable down-modulation in all the 6 HMCLs investigated (2 major allele homozygous, 2 heterozygous and 2 minor allele homozygous for SNP rs13065271, see Figure 29). However, we observed a biological effect in terms of cell viability and cell growth only in cell lines that expressed at least one major allele of SNP rs13065271 (Figure 30). In particular, the analysis of cell cycle and apoptosis pointed out an important increase of the percentage of cells in sub-G0/G1 phase and of apoptotic cells in the same *ST3GAL6-AS1* silenced HMCLs (H929, U266, LP1 and KMS20; Figure 31 and 32). HMCLs, such as H929 and LP1, homozygous for the major allele and heterozygous for SNP rs13065271 respectively, were examined for the colony formation after *ST3GAL6-AS1* silencing. These two cell lines lost the capability to form colonies after *ST3GAL6-AS1* silencing (Figure 33). Furthermore, in all the analyzed cell lines silenced for *ST3GAL6-AS1* we evidenced no variation of neighboring genes expression levels (*ST3GAL6* and *DCBLD2*; Figure 29).

Based on our results, we could suggest that *ST3GAL6-AS1* could be involved in MM pathology. For this purpose, we performed GEP analysis on H929 and LP1 silenced for *ST3GAL6-AS1* to identify possible targets and associated molecular pathways. Supervised analysis evidenced several genes modulated after *ST3GAL6-AS1* silencing in each of the two cell lines. However, only three genes resulted significantly down-modulated in both HMCLs: *RAP1A*, *KDM5B* and *PTPRS* (Figure 35). No one of these genes presented, in our cohort of MM patients, a correlation with *ST3GAL6-AS1* expression levels.

GSEA analysis made on GEP data evidenced in both the HMCLs the down-modulation of four gene sets. Interestingly, three of these gene sets, related to *P53 pathway*, *lysosome* and *other glycan degradation*, showed the up-regulation of *FUCA1* (Figure 37 and 39), a gene mapping on 1p36.11 coding for a lysosomal enzyme involved in the degradation of fucose-containing

glycoproteins and glycolipids, and suggested to act as a tumor suppressor gene[133]. Interestingly, *FUCA 1* was found inversely correlated with *ST3GAL6* in breast cancer[100].

Recently, Glavey et al, indicated in a Patent application (n. 20170327899) *FUCA1* and *ST3GAL6* genes as useful in the diagnosis and assessment of prognosis of MM. In that application, *ST3GAL6* and *FUCA1* were indicated as biomarkers of MM with inferior survival rate; the Authors described a cohort of patients in which the high expression of *FUCA1* correlated with a better prognosis, while high levels of *ST3GAL6* and low levels of *FUCA1* with a worse prognosis.

Based on these findings, we investigated the expression levels *FUCA1* in our cohort of patients and identified an important down-regulation of the gene in pathological samples compared to normal controls and a significant inverse correlation with the expression levels of ST3GAL6-AS1 and *ST3GAL6.* Supported by these results, we can speculate that *ST3GAL6-AS1* as well as the *ST3GAL6* expression levels could be related to *FUCA1* expression in MM patients.

Overall, these data led to the hypothesis that *ST3GAL6-AS1* may act on the down-regulation of *FUCA1* allowing an increase of protein glycosylation and thus contributing to some oncogenic pathways in MM. They represent the bases for further studies able to dissect the role and functions of *ST3GAL6-AS1* in MM.

# References


1 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA: a cancer journal for clinicians* **66**, 7-30, doi:10.3322/caac.21332 (2016).

2 Solly, S. Remarks on the pathology of mollities ossium; with cases. *Medico-chirurgical transactions* **27**, 435-498 438 (1844).

3 Reiman, T. *et al.* Persistent preswitch clonotypic myeloma cells correlate with decreased survival: evidence for isotype switching within the myeloma clone. *Blood* **98**, 2791-2799 (2001).

4 Morgan, G. J., Walker, B. A. & Davies, F. E. The genetic architecture of multiple myeloma. *Nature reviews. Cancer* **12**, 335-348, doi:10.1038/nrc3257 (2012).

5 Gonzalez, D. *et al.* Immunoglobulin gene rearrangements and the pathogenesis of multiple myeloma. *Blood* **110**, 3112-3121, doi:10.1182/blood-2007-02-069625 (2007).

6 Bergsagel, P. L. & Kuehl, W. M. Critical roles for immunoglobulin translocations and cyclin D dysregulation in multiple myeloma. *Immunological reviews* **194**, 96-104 (2003).

7 Hakim, O. *et al.* DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. *Nature* **484**, 69-74, doi:10.1038/nature10909 (2012).

8 Rasmussen, T. *et al.* Identification of translocation products but not K-RAS mutations in memory B cells from patients with multiple myeloma. *Haematologica* **95**, 1730-1737, doi:10.3324/haematol.2010.024778 (2010).

9 Rajkumar, S. V. *et al.* International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *The Lancet. Oncology* **15**, e538-548, doi:10.1016/S1470-2045(14)70442-5 (2014).

10 Rajkumar, S. V., Larson, D. & Kyle, R. A. Diagnosis of smoldering multiple myeloma. *The New England journal of medicine* **365**, 474-475, doi:10.1056/NEJMc1106428 (2011).

11 Rajkumar, S. V., Landgren, O. & Mateos, M. V. Smoldering multiple myeloma. *Blood* **125**, 3069-3075, doi:10.1182/blood-2014-09-568899 (2015).

12 Sant, M. *et al.* Incidence of hematologic malignancies in Europe by morphologic subtype: results of the HAEMACARE project. *Blood* **116**, 3724-3734, doi:10.1182/blood-2010-05-282632 (2010).

13 Kumar, S. K. *et al.* Multiple myeloma. *Nature reviews. Disease primers* **3**, 17046, doi:10.1038/nrdp.2017.46 (2017).

14 Ronchetti, D. *et al.* Molecular analysis of 11q13 breakpoints in multiple myeloma. *Blood* **93**, 1330-1337 (1999).

15 Tasidou, A. *et al.* Increased expression of cyclin-D1 on trephine bone marrow biopsies independently predicts for shorter overall survival in patients with multiple myeloma treated with novel agents. *American journal of hematology* **87**, 734-736, doi:10.1002/ajh.23223 (2012).

16 Ayton, P. M. & Cleary, M. L. Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins. *Oncogene* **20**, 5695-5707, doi:10.1038/sj.onc.1204639 (2001).

17 Chng, W. J. *et al.* IMWG consensus on risk stratification in multiple myeloma. *Leukemia* **28**, 269-277, doi:10.1038/leu.2013.247 (2014).

18 Annunziata, C. M. *et al.* A mechanistic rationale for MEK inhibitor therapy in myeloma based on blockade of MAF oncogene expression. *Blood* **117**, 2396-2404, doi:10.1182/blood-2010-04-278788 (2011).

19 Shaughnessy, J., Jr. *et al.* Cyclin D3 at 6p21 is dysregulated by recurrent chromosomal translocations to immunoglobulin loci in multiple myeloma. *Blood* **98**, 217-223 (2001).

20 Kumar, S. *et al.* Trisomies in multiple myeloma: impact on survival in patients with high-risk cytogenetics. *Blood* **119**, 2100-2105, doi:10.1182/blood-2011-11-390658 (2012).

21 Fonseca, R. *et al.* International Myeloma Working Group molecular classification of multiple myeloma: spotlight review. *Leukemia* **23**, 2210-2221, doi:10.1038/leu.2009.174 (2009).
22 Tiedemann, R. E. *et al.* Genetic aberrations and survival in plasma cell leukemia. *Leukemia* **22**, 1044-1052, doi:10.1038/leu.2008.4 (2008).
23 Lode, L. *et al.* Mutations in TP53 are exclusively associated with del(17p) in multiple myeloma. *Haematologica* **95**, 1973-1976, doi:10.3324/haematol.2010.023697 (2010).
24 Walker, B. A. *et al.* A compendium of myeloma-associated chromosomal copy number abnormalities and their prognostic value. *Blood* **116**, e56-65, doi:10.1182/blood-2010-04-279596 (2010).
25 Cully, M. & Downward, J. SnapShot: Ras Signaling. *Cell* **133**, 1292-1292 e1291, doi:10.1016/j.cell.2008.06.020 (2008).
26 Rasmussen, T., Kuehl, M., Lodahl, M., Johnsen, H. E. & Dahl, I. M. Possible roles for activating RAS mutations in the MGUS to MM transition and in the intramedullary to extramedullary transition in some plasma cell tumors. *Blood* **105**, 317-323, doi:10.1182/blood-2004-03-0833 (2005).
27 Neri, A. *et al.* Ras oncogene mutation in multiple myeloma. *The Journal of experimental medicine* **170**, 1715-1725 (1989).
28 Cheung, C. H. Y. *et al.* Prevalence and Clinicopathologic Significance of BRAF V600E Mutation in Chinese Multiple Myeloma Patients. *Clinical lymphoma, myeloma & leukemia* **18**, e315-e325, doi:10.1016/j.clml.2018.05.008 (2018).
29 Corre, J. *et al.* Multiple myeloma clonal evolution in homogeneously treated patients. *Leukemia*, doi:10.1038/s41375-018-0153-6 (2018).
30 Bolli, N. *et al.* Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nature communications* **5**, 2997, doi:10.1038/ncomms3997 (2014).
31 Chapman, M. A. *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature* **471**, 467-472, doi:10.1038/nature09837 (2011).
32 Boyd, K. D. *et al.* Mapping of chromosome 1p deletions in myeloma identifies FAM46C at 1p12 and CDKN2C at 1p32.3 as being genes in regions associated with adverse survival. *Clinical cancer research : an official journal of the American Association for Cancer Research* **17**, 7776-7784, doi:10.1158/1078-0432.CCR-11-1791 (2011).
33 Riccardi, A. *et al.* Changing clinical presentation of multiple myeloma. *European journal of cancer* **27**, 1401-1405 (1991).
34 Kyle, R. A. *et al.* Review of 1027 patients with newly diagnosed multiple myeloma. *Mayo Clinic proceedings* **78**, 21-33, doi:10.4065/78.1.21 (2003).
35 Bianchi, G. & Anderson, K. C. Understanding biology to tackle the disease: Multiple myeloma from bench to bedside, and back. *CA: a cancer journal for clinicians* **64**, 422-444, doi:10.3322/caac.21252 (2014).
36 Rollig, C., Knop, S. & Bornhauser, M. Multiple myeloma. *Lancet* **385**, 2197-2208, doi:10.1016/S0140-6736(14)60493-1 (2015).
37 Rajkumar, S. V. Myeloma today: Disease definitions and treatment advances. *American journal of hematology* **91**, 90-100, doi:10.1002/ajh.24236 (2016).
38 Michels, T. C. & Petersen, K. E. Multiple Myeloma: Diagnosis and Treatment. *American family physician* **95**, 373-383 (2017).
39 Thomas, E. D., Lochte, H. L., Jr., Lu, W. C. & Ferrebee, J. W. Intravenous infusion of bone marrow in patients receiving radiation and chemotherapy. *The New England journal of medicine* **257**, 491-496, doi:10.1056/NEJM195709122571102 (1957).
40 Osserman, E. F. *et al.* Identical twin marrow transplantation in multiple myeloma. *Acta haematologica* **68**, 215-223, doi:10.1159/000206984 (1982).

41 Gahrton, G. *et al.* Bone marrow transplantation in multiple myeloma: report from the European Cooperative Group for Bone Marrow Transplantation. *Blood* **69**, 1262-1264 (1987).
42 Willenbacher, W. *et al.* Towards Molecular Profiling in Multiple Myeloma: A Literature Review and Early Indications of Its Efficacy for Informing Treatment Strategies. *International journal of molecular sciences* **19**, doi:10.3390/ijms19072087 (2018).
43 Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463-5467 (1977).
44 Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annual review of biochemistry* **81**, 145-166, doi:10.1146/annurev-biochem-051410-092902 (2012).
45 Brown, C. J. *et al.* A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**, 38-44, doi:10.1038/349038a0 (1991).
46 Brown, C. J. *et al.* Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature* **349**, 82-84, doi:10.1038/349082a0 (1991).
47 Kelley, D. & Rinn, J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome biology* **13**, R107, doi:10.1186/gb-2012-13-11-r107 (2012).
48 Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **136**, 629-641, doi:10.1016/j.cell.2009.02.006 (2009).
49 Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & development* **25**, 1915-1927, doi:10.1101/gad.17446611 (2011).
50 Mercer, T. R. & Mattick, J. S. Structure and function of long noncoding RNAs in epigenetic regulation. *Nature structural & molecular biology* **20**, 300-307, doi:10.1038/nsmb.2480 (2013).
51 Nobili, L., Lionetti, M. & Neri, A. Long non-coding RNAs in normal and malignant hematopoiesis. *Oncotarget* **7**, 50666-50681, doi:10.18632/oncotarget.9308 (2016).
52 Wang, K. C. & Chang, H. Y. Molecular mechanisms of long noncoding RNAs. *Molecular cell* **43**, 904-914, doi:10.1016/j.molcel.2011.08.018 (2011).
53 Johnsson, P., Lipovich, L., Grander, D. & Morris, K. V. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochimica et biophysica acta* **1840**, 1063-1071, doi:10.1016/j.bbagen.2013.10.035 (2014).
54 Diederichs, S. The four dimensions of noncoding RNA conservation. *Trends in genetics : TIG* **30**, 121-123, doi:10.1016/j.tig.2014.01.004 (2014).
55 Lipovich, L., Johnson, R. & Lin, C. Y. MacroRNA underdogs in a microRNA world: evolutionary, regulatory, and biomedical significance of mammalian long non-protein-coding RNA. *Biochimica et biophysica acta* **1799**, 597-615, doi:10.1016/j.bbagrm.2010.10.001 (2010).
56 Yap, K. L. *et al.* Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Molecular cell* **38**, 662-674, doi:10.1016/j.molcel.2010.03.021 (2010).
57 Kotake, Y. *et al.* Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* **30**, 1956-1962, doi:10.1038/onc.2010.568 (2011).
58 Puvvula, P. K. *et al.* Long noncoding RNA PANDA and scaffold-attachment-factor SAFA control senescence entry and exit. *Nature communications* **5**, 5323, doi:10.1038/ncomms6323 (2014).
59 Kim, T. K., Hemberg, M. & Gray, J. M. Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers. *Cold Spring Harbor perspectives in biology* **7**, a018622, doi:10.1101/cshperspect.a018622 (2015).

60 Tripathi, V. *et al.* The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Molecular cell* **39**, 925-938, doi:10.1016/j.molcel.2010.08.011 (2010).
61 Hirose, T. *et al.* NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Molecular biology of the cell* **25**, 169-183, doi:10.1091/mbc.E13-09-0558 (2014).
62 Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120-124, doi:10.1038/nature09819 (2011).
63 Gong, C. & Maquat, L. E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* **470**, 284-288, doi:10.1038/nature09701 (2011).
64 Faghihi, M. A. *et al.* Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nature medicine* **14**, 723-730, doi:10.1038/nm1784 (2008).
65 Deng, L., Yang, S. B., Xu, F. F. & Zhang, J. H. Long noncoding RNA CCAT1 promotes hepatocellular carcinoma progression by functioning as let-7 sponge. *Journal of experimental & clinical cancer research : CR* **34**, 18, doi:10.1186/s13046-015-0136-7 (2015).
66 Yoon, J. H. *et al.* LincRNA-p21 suppresses target mRNA translation. *Molecular cell* **47**, 648-655, doi:10.1016/j.molcel.2012.06.027 (2012).
67 Nobili, L., Ronchetti, D., Taiana, E. & Neri, A. Long non-coding RNAs in B-cell malignancies: a comprehensive overview. *Oncotarget* **8**, 60605-60623, doi:10.18632/oncotarget.17303 (2017).
68 Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
69 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).
70 Gutschner, T. & Diederichs, S. The hallmarks of cancer: a long non-coding RNA point of view. *RNA biology* **9**, 703-719, doi:10.4161/rna.20481 (2012).
71 Rao, A., Rajkumar, T. & Mani, S. Perspectives of long non-coding RNAs in cancer. *Molecular biology reports* **44**, 203-218, doi:10.1007/s11033-017-4103-6 (2017).
72 Prensner, J. R. *et al.* Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nature biotechnology* **29**, 742-749, doi:10.1038/nbt.1914 (2011).
73 Zhang, E. B. *et al.* Long noncoding RNA ANRIL indicates a poor prognosis of gastric cancer and promotes tumor growth by epigenetically silencing of miR-99a/miR-449a. *Oncotarget* **5**, 2276-2292, doi:10.18632/oncotarget.1902 (2014).
74 Iranpour, M. *et al.* Expression analysis of four long noncoding RNAs in breast cancer. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* **37**, 2933-2940, doi:10.1007/s13277-015-4135-2 (2016).
75 Theimer, C. A. & Feigon, J. Structure and function of telomerase RNA. *Current opinion in structural biology* **16**, 307-318, doi:10.1016/j.sbi.2006.05.005 (2006).
76 Yuan, J. H. *et al.* A long noncoding RNA activated by TGF-beta promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer cell* **25**, 666-681, doi:10.1016/j.ccr.2014.03.010 (2014).
77 Liu, L. *et al.* ncRuPAR inhibits gastric cancer progression by down-regulating protease-activated receptor-1. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* **35**, 7821-7829, doi:10.1007/s13277-014-2042-6 (2014).
78 Misawa, A., Takayama, K., Urano, T. & Inoue, S. Androgen-induced Long Noncoding RNA (lncRNA) SOCS2-AS1 Promotes Cell Growth and Inhibits Apoptosis in Prostate Cancer

Cells. *The Journal of biological chemistry* **291**, 17861-17880, doi:10.1074/jbc.M116.718536 (2016).
79 Zhang, P., Cao, L., Fan, P., Mei, Y. & Wu, M. LncRNA-MIF, a c-Myc-activated long non-coding RNA, suppresses glycolysis by promoting Fbxw7-mediated c-Myc degradation. *EMBO reports* **17**, 1204-1220, doi:10.15252/embr.201642067 (2016).
80 Carpenter, S. *et al.* A long noncoding RNA mediates both activation and repression of immune response genes. *Science* **341**, 789-792, doi:10.1126/science.1240925 (2013).
81 Sharma, V. *et al.* A BRCA1-interacting lncRNA regulates homologous recombination. *EMBO reports* **16**, 1520-1534, doi:10.15252/embr.201540437 (2015).
82 Zhang, C. *et al.* Upregulation of lncRNA HOTAIR contributes to IL-1beta-induced MMP overexpression and chondrocytes apoptosis in temporomandibular joint osteoarthritis. *Gene* **586**, 248-253, doi:10.1016/j.gene.2016.04.016 (2016).
83 Verma, A. *et al.* Transcriptome sequencing reveals thousands of novel long non-coding RNAs in B cell lymphoma. *Genome medicine* **7**, 110, doi:10.1186/s13073-015-0230-7 (2015).
84 Benetatos, L. *et al.* Promoter hypermethylation of the MEG3 (DLK1/MEG3) imprinted gene in multiple myeloma. *Clinical lymphoma & myeloma* **8**, 171-175, doi:10.3816/CLM.2008.n.021 (2008).
85 Ronchetti, D. *et al.* Distinct lncRNA transcriptional fingerprints characterize progressive stages of multiple myeloma. *Oncotarget* **7**, 14814-14830, doi:10.18632/oncotarget.7442 (2016).
86 Dhamija, S. & Diederichs, S. From junk to master regulators of invasion: lncRNA functions in migration, EMT and metastasis. *International journal of cancer* **139**, 269-280, doi:10.1002/ijc.30039 (2016).
87 Amodio, N. *et al.* Drugging the lncRNA MALAT1 via LNA gapmeR ASO inhibits gene expression of proteasome subunits and triggers anti-multiple myeloma activity. *Leukemia*, doi:10.1038/s41375-018-0067-3 (2018).
88 Meng, Y., Li, Q., Li, L. & Ma, R. The long non-coding RNA CRNDE promotes cervical cancer cell growth and metastasis. *Biological chemistry* **399**, 93-100, doi:10.1515/hsz-2017-0199 (2017).
89 Sedlarikova, L. *et al.* Deregulated expression of long non-coding RNA UCA1 in multiple myeloma. *European journal of haematology* **99**, 223-233, doi:10.1111/ejh.12908 (2017).
90 Shen, Y. *et al.* Focusing on long non-coding RNA dysregulation in newly diagnosed multiple myeloma. *Life sciences* **196**, 133-142, doi:10.1016/j.lfs.2018.01.025 (2018).
91 Harduin-Lepers, A. *et al.* The human sialyltransferase family. *Biochimie* **83**, 727-737 (2001).
92 Datta, A. K. Comparative sequence analysis in the sialyltransferase protein family: analysis of motifs. *Current drug targets* **10**, 483-498 (2009).
93 Natoni, A., Macauley, M. S. & O'Dwyer, M. E. Targeting Selectins and Their Ligands in Cancer. *Frontiers in oncology* **6**, 93, doi:10.3389/fonc.2016.00093 (2016).
94 Reymond, N., d'Agua, B. B. & Ridley, A. J. Crossing the endothelial barrier during metastasis. *Nature reviews. Cancer* **13**, 858-870, doi:10.1038/nrc3628 (2013).
95 Julien, S. *et al.* Selectin ligand sialyl-Lewis x antigen drives metastasis of hormone-dependent breast cancers. *Cancer research* **71**, 7683-7693, doi:10.1158/0008-5472.CAN-11-1139 (2011).
96 Perez-Garay, M. *et al.* alpha2,3-Sialyltransferase ST3Gal IV promotes migration and metastasis in pancreatic adenocarcinoma cells and tends to be highly expressed in pancreatic adenocarcinoma tissues. *The international journal of biochemistry & cell biology* **45**, 1748-1757, doi:10.1016/j.biocel.2013.05.015 (2013).

97 Okajima, T. *et al.* Molecular cloning of a novel alpha2,3-sialyltransferase (ST3Gal VI) that sialylates type II lactosamine structures on glycoproteins and glycolipids. *The Journal of biological chemistry* **274**, 11479-11486 (1999).
98 Kawamura, Y. I. *et al.* DNA hypermethylation contributes to incomplete synthesis of carbohydrate determinants in gastrointestinal cancer. *Gastroenterology* **135**, 142-151 e143, doi:10.1053/j.gastro.2008.03.031 (2008).
99 Darby, J. K. *et al.* Pvu II RFLP at the human chromosome 1 alpha-L-fucosidase gene locus (FUCA1). *Nucleic acids research* **14**, 9543 (1986).
100 Milde-Langosch, K. *et al.* Prognostic relevance of glycosylation-associated genes in breast cancer. *Breast cancer research and treatment* **145**, 295-305, doi:10.1007/s10549-014-2949-z (2014).
101 Kobuke, K. *et al.* ESDN, a novel neuropilin-like membrane protein cloned from vascular cells with the longest secretory signal sequence among eukaryotes, is up-regulated after vascular injury. *The Journal of biological chemistry* **276**, 34105-34114, doi:10.1074/jbc.M105293200 (2001).
102 Guo, X. *et al.* Endothelial and smooth muscle-derived neuropilin-like protein regulates platelet-derived growth factor signaling in human vascular smooth muscle cells by modulating receptor ubiquitination. *The Journal of biological chemistry* **284**, 29376-29382, doi:10.1074/jbc.M109.049684 (2009).
103 Koshikawa, K. *et al.* Significant up-regulation of a novel gene, CLCP1, in a highly metastatic lung cancer subline as well as in lung cancers in vivo. *Oncogene* **21**, 2822-2828, doi:10.1038/sj.onc.1205405 (2002).
104 Guo, A. *et al.* Signaling networks assembled by oncogenic EGFR and c-Met. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 692-697, doi:10.1073/pnas.0707270105 (2008).
105 Feng, H. *et al.* EGFR phosphorylation of DCBLD2 recruits TRAF6 and stimulates AKT-promoted tumorigenesis. *The Journal of clinical investigation* **124**, 3741-3756, doi:10.1172/JCI73093 (2014).
106 Mattioli, M. *et al.* Gene expression profiling of plasma cell dyscrasias reveals molecular patterns associated with distinct IGH translocations in multiple myeloma. *Oncogene* **24**, 2461-2473, doi:10.1038/sj.onc.1208447 (2005).
107 Agnelli, L. *et al.* Upregulation of translational machinery and distinct genetic subgroups characterise hyperdiploidy in multiple myeloma. *British journal of haematology* **136**, 565-573, doi:10.1111/j.1365-2141.2006.06467.x (2007).
108 Lionetti, M. *et al.* Molecular spectrum of BRAF, NRAS and KRAS gene mutations in plasma cell dyscrasias: implication for MEK-ERK pathway activation. *Oncotarget* **6**, 24205-24217, doi:10.18632/oncotarget.4434 (2015).
109 Lionetti, M. *et al.* A compendium of DIS3 mutations and associated transcriptional signatures in plasma cell dyscrasias. *Oncotarget* **6**, 26129-26141, doi:10.18632/oncotarget.4674 (2015).
110 Lionetti, M. *et al.* Molecular spectrum of TP53 mutations in plasma cell dyscrasias by next generation sequencing: an Italian cohort study and overview of the literature. *Oncotarget* **7**, 21353-21361, doi:10.18632/oncotarget.7241 (2016).
111 Barbieri, M. *et al.* Compendium of FAM46C gene mutations in plasma cell dyscrasias. *British journal of haematology* **174**, 642-645, doi:10.1111/bjh.13793 (2016).
112 Ronchetti, D. *et al.* A compendium of long non-coding RNAs transcriptional fingerprint in multiple myeloma. *Scientific reports* **8**, 6557, doi:10.1038/s41598-018-24701-8 (2018).
113 Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5116-5121, doi:10.1073/pnas.091062498 (2001).

114 Ayupe, A. C. & Reis, E. M. Evaluating the Stability of mRNAs and Noncoding RNAs. *Methods in molecular biology* **1468**, 139-153, doi:10.1007/978-1-4939-4035-6_11 (2017).
115 Sharova, L. V. *et al.* Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA research : an international journal for rapid publication of reports on genes and genomes* **16**, 45-58, doi:10.1093/dnares/dsn030 (2009).
116 Ayupe, A. C. *et al.* Global analysis of biogenesis, stability and sub-cellular localization of lncRNAs mapping to intragenic regions of the human genome. *RNA biology* **12**, 877-892, doi:10.1080/15476286.2015.1062960 (2015).
117 Glavey, S. V. *et al.* The sialyltransferase ST3GAL6 influences homing and survival in multiple myeloma. *Blood* **124**, 1765-1776, doi:10.1182/blood-2014-03-560862 (2014).
118 Kikuta, K. *et al.* Discoidin, CUB and LCCL domain-containing protein 2 (DCBLD2) is a novel biomarker of myxofibrosarcoma invasion identified by global protein expression profiling. *Biochimica et biophysica acta* **1865**, 1160-1166, doi:10.1016/j.bbapap.2017.06.023 (2017).
119 Fluiter, K. *et al.* On the in vitro and in vivo properties of four locked nucleic acid nucleotides incorporated into an anti-H-Ras antisense oligonucleotide. *Chembiochem : a European journal of chemical biology* **6**, 1104-1109, doi:10.1002/cbic.200400419 (2005).
120 Soifer, H. S. *et al.* Silencing of gene expression by gymnotic delivery of antisense oligonucleotides. *Methods in molecular biology* **815**, 333-346, doi:10.1007/978-1-61779-424-7_25 (2012).
121 Kumar, S. K. *et al.* Improved survival in multiple myeloma and the impact of novel therapies. *Blood* **111**, 2516-2520, doi:10.1182/blood-2007-10-116129 (2008).
122 Mailankody, S. *et al.* Minimal residual disease in multiple myeloma: bringing the bench to the bedside. *Nature reviews. Clinical oncology* **12**, 286-295, doi:10.1038/nrclinonc.2014.239 (2015).
123 Kazandjian, D. & Landgren, O. A look backward and forward in the regulatory and treatment history of multiple myeloma: Approval of novel-novel agents, new drug development, and longer patient survival. *Seminars in oncology* **43**, 682-689, doi:10.1053/j.seminoncol.2016.10.008 (2016).
124 Nobili, L. *et al.* Long Non-Coding RNAs in Multiple Myeloma. *Genes* **9**, doi:10.3390/genes9020069 (2018).
125 Samur, M. K. *et al.* Long intergenic non-coding RNAs have an independent impact on survival in multiple myeloma. *Leukemia*, doi:10.1038/s41375-018-0116-y (2018).
126 Orom, U. A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46-58, doi:10.1016/j.cell.2010.09.001 (2010).
127 Guil, S. & Esteller, M. Cis-acting noncoding RNAs: friends and foes. *Nature structural & molecular biology* **19**, 1068-1075, doi:10.1038/nsmb.2428 (2012).
128 Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics* **40**, 1413-1415, doi:10.1038/ng.259 (2008).
129 Tang, J. Y. *et al.* Long noncoding RNAs-related diseases, cancers, and drugs. *TheScientificWorldJournal* **2013**, 943539, doi:10.1155/2013/943539 (2013).
130 Zhang, X. *et al.* The identification of an ESCC susceptibility SNP rs920778 that regulates the expression of lncRNA HOTAIR via a novel intronic enhancer. *Carcinogenesis* **35**, 2062-2067, doi:10.1093/carcin/bgu103 (2014).
131 Chen, L. L. Linking Long Noncoding RNA Localization and Function. *Trends in biochemical sciences* **41**, 761-772, doi:10.1016/j.tibs.2016.07.003 (2016).

132 Tani, H. *et al.* Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome research* **22**, 947-956, doi:10.1101/gr.130559.111 (2012).
133 Ezawa, I. *et al.* Novel p53 target gene FUCA1 encodes a fucosidase and regulates growth and survival of cancer cells. *Cancer science* **107**, 734-745, doi:10.1111/cas.12933 (2016).