Department of Experimental and Health Sciences
Faculty of Health and Life Sciences
Universitat Pompeu Fabra
Tesi Doctoral 2014

Natural Antisense Transcripts control *LEF1* gene expression

Dissertation presented by
Estel Aparicio i Prat
For the degree of Doctor in Philosophy

Work carried out under the supervision of Dr. Antonio García de Herreros Madueño in the Epithelial to Mesenchymal Transition and Tumor Progression Group in the Cancer Research Program in the Institut Hospital del Mar d'Investigacions Mèdiques (IMIM).

Dr. Antonio García de Herreros Madueño
(Thesis Director)

Estel Aparicio i Prat
(PhD candidate)

Als meus pares

A la Marina

Al Víctor

ACKNOWLEDGEMENTS

En aquests anys de tesi he après un munt! I si fer el doctorat ha estat una experiència tan positiva, és gràcies a tots vosaltres.

En primer lloc, gràcies a tu, Antonio. Perquè després d'ajudar-me a buscar grups a la UAB, quan vaig saber que oferies plaça, em vas dir que, si jo volia, era per mi. Gràcies, Mireia, per no crear un conflicte matrimonial en aquell moment! Antonio, ets un pou de saviesa i s'aprèn molt de tu. Recordes els experiments que he fet millor que jo mateixa i discutir de ciència amb tu és molt enriquidor. Gràcies per estar sempre tan disponible! I no només per la part científica, sinó també les xerrades sobre viatges, els castells a la piscina o la deliciosa barbacoa que ens vas preparar! Un plaer.

Manolo, el meu jefecillo. Mil gràcies per ser el meu mestre! Gràcies per la paciència als meus inicis, per les explicacions amb dibuixos al paper de poiata, pels riures i complicitats, per compartir experiments surrealistes i per fer el dia a dia al laboratori molt més divertit! I pels vespres al Bitàcora, passant pel Tres Flores, acabant amb dits de Baileys i nits del mariner! Vaig tenir molta sort de caure en les teves mans quan vaig entrar al laboratori! Mil gràcies!

Un dels agraïments més grans i sentits en aquesta tesi és per tu, Alba. Quan et vaig conèixer fa més de 9 anys, poc m'imaginava que les nostres vides agafarien camins tan propers i que series un dels meus punts de suport més fort durant aquesta tesi. Grandíssima amiga i excel·lent científica: persona ideal per a rebre els millors consells, personals i professionals. Sóc molt afortunada de tenir-te! Els moments plegades són incomptables! Però sens dubte aquest darrer any de convivència, mimant-nos sempre, també en aquest

darrer temps de tesi, t'han convertit pràcticament en una germana per mi. T'estimo molt!

Lore, apareciste un poco tarde, pero suerte que lo hiciste! Nuestras "primeras citas" para comprar ropa, confesiones muy personales en la terraza, noches de viernes en petit comité y un día a día muy cálido, tejieron una de mis mejores amistades. De ti aprendo cada día: tu transparencia, tu gran entrega por los tuyos; siempre estás, cuando ni una misma sabe que lo que necesita oír, es exactamente lo que tú has dicho. Tus consejos son oro para mí y es un privilegio tenerte como amiga. No me equivoqué decidiendo de vivir con vosotras; complicidad 100%, no importa si con un bikini de desayuno o un vino en la cena. Gracias por este inmejorable año de convivencia y lo que aún seguiremos compartiendo! Miki, et quedes una perla, però sembla que tu no ets menys. Encara ens queda molt per compartir, però la teva proximitat, disponibilitat, carinyo en les paelles i tiramisús, apunten molt bones formes! I un agraïment molt sincer per l'ajuda tècnica per fer la tapa d'aquesta tesi!! Mil gràcies, parella!

Núria, moltíssimes gràcies per les converses de tot, pel suport diari, per la teva gran vitalitat i per ser-hi quan més ho he necessitat. Me n'alegro un munt que una metgessa com tu hagi decidit aventurar-se amb un doctorat, perquè en trec una amiga! Faig extensiu l'agraïment al Guillem i a l'Esteleta, i segur que seguirem compartint molts dinars, excursions i concerts!

Jou, el teu bon rollo i abraçades en el moment més necessari han fet de tu la millor companya d'escriptori! Discreta, però pendent dels qui t'envolten,... un plaer tenir-te a prop! Rosa, gràcies per ser un punt de suport personal i laboral en molts moments. Compartint poiata, excursions i viatges, m'ho he passat molt bé amb tu! Àlex, vam entrar quasi a l'hora i el dia a dia a l'escriptori amb tu, les beersessions que vam tancar al Sidecar i el teu tarannà de fer fàcils les coses, només em porten bons records. Jelena,

gracias por tu soporte en días bajos y por las risas en el Bitàcora! Ane, un placer conocer un poco más de Donosti de mano de una auténtica vasca y gracias por las charlas sobre ropa femenina! Joan Pau, ets la bondat personificada! Un plaer que hagis entrat a Snail! Laura, el mateix dic! Molta sort amb la cromatina! Cristina, compatriota, un plaer haver coincidit aquests mesos amb tu (curts però intensos!). Raúl, mil gràcies per haver pogut comptar amb tu sempre. Allò que està al teu abast ho has fet molt fàcil i és de molt agrair. Gràcies també per les converses de tot a les hores de dinar! Jessica, mil gràcies per deixar-me posar PCRs quan no hi havia cap forat! Víctor, m'ho he passat fantàsticament amb tu. Gràcies pels teus consells científics, per sempre tenir una broma a la boca i per alegrar-me el dia a dia! Rocco, encajamos desde el principio. Contigo todo es fácil y pasa mejor. Seguiremos compartiendo corales, comidas en la playa y conversaciones en la terraza! Sílvia, gràcies per les converses a l'hora de dinar. Sandra, moltes gràcies per estar sempre disposada a donar idees fresques al projecte després de seminaris "encallats". Jepi, un plaer haver pogut compartir amb tu el món bastoner i els balls folk a Manresa! Tots vosaltres feu el dia a dia al laboratori més agradable, divertit i enriquidor!

Però també n'hi ha que van ser Snail un dia i van marxar... Niko, gràcies pel suport científic i personal que he tingut en tu sempre que he necessitat. Natàlia, recordo amb especial carinyo certes converses i confessions al Bitàcora. Gràcies per la teva vitalitat! Patri, vam coincidir poc temps però la connexió va ser molt fàcil. Gràcies per les reflexions sobre al ciència i pels bons consells en mètode científic! Montse, també vam coincidir molt poc, però les converses amb tu sempre han estat molt positives. David, un autèntic plaer tenir-te al lab, quan eres la nostra Goldypedia! Ali, la alegria de la "huerta". Mil gràcies pels riures en el dia a dia i fer-ho tot tan fàcil i feliç! Clara, gràcies per ser tan propera i per ser un suport tan gran per les persones que estan al teu càrrec. Raquel,

gràcies per les converses a cultius i a l'escriptori. Josu, ets un "catxondo" encantador. Un plaer poder fer broma de tot amb tu! Fani, sencillamente encantadora. Fue genial compartir contigo los NATs, las conversaciones de todo, la descubierta de nuevos sitios en Barcelona… Se te echa en falta! Alba Azagra, un plaer també haver coincidit amb tu aquest temps! I no m'oblido de tu, Pere, el meu mindu! Mil gràcies per tota l'ajuda en el projecte: vas treballar molt i amb ganes! Però també pel suport personal en els moments necessaris. He estat afortunada amb tu!

I els Snails s'acaben, però no el suport dels companys del dia a dia. I tu, Silvi, ets el meu puntal fora del laboratori. Ha estat genial descobrir-nos i ets una de les millors persones amb qui m'he creuat en aquest doctorat. Per totes les converses i les que vindran, mil gràcies, crack!! Neus, a tu també et vull fer un agraïment especial. És un plaer compartir el dia a dia amb tu, sempre amb el somriure i bona predisposició amb tot, científic i personal. Mil gràcies pels viatges a Matadepera en cotxe, que a més de ser el taxi perfecte, ens duien a converses de tot. Moltes gràcies també a tu, Joan, per l'aire fresc, ganes de fer de tot i estar sempre disposat a escoltar! Conchi i Laura, sou estupendes! Les veïnes perfectes per alegrar el dia! Elena, David i Fabien, gràcies per les cerveses, converses i estones de cultius plegats! Mireia i Tània, gràcies pels somriure i estar sempre disposades a ajudar en tot! I també per fer-me cagar el tió al lab!! Erika, Pol, Leonor, Cristina, Jessica, Júlia, Jordi, Teresa, Vero i Vane, moltíssimes gràcies per fer-ho tot tan fàcil, des de buscar-me plàsmids control o enzims de restricció fins a les converses a cultius!

Un agraïment especial a l'antic lab 4. Mercè, Lara, Marta Garrido, Laura, Marta Moreno, Alex i els nouvinguts Marc i Kathi. Uns inicis en ciència així, només fan que venir ganes de seguir! Gràcies per tants bons moments! També gràcies a tu, Carme, per dirigir-me als

meus inicis! Gràcies a la bona experiència amb vosaltres em vaig decidir a fer el doctorat a Snail!

I fora de l'antiga UBCM com a tal, també hi ha molta gent que ha fet més fàcil el dia a dia. Vull donar-te les gràcies a tu, Eulàlia, perquè trobar-te al nanodrop sempre és una alegria, i compartir passions viatgeres, problemes d'RNA i preocupacions científiques sempre ha estat un autèntic plaer! També a l'Alba Dalmases, per les converses durant la celebració de la tesi del Niko i per les aturades al passadís per comentar la jugada.

I hi ha somriures que alegren el dia a dia: com el del Carlos quan ens porta els paquets, el de l'Òscar de manteniment, el de la Montse de la neteja a les tardes, el de la Paqui i la gent de la cuina, el de l'Heleia, el del Sergi, el de l'Aina, el del David, el del Maurizzio, el del David i el de l'Irene. També el de l'Ignasi i la Marta, el del Thomas, la Cuti, l'Arnau i l'Amado i el de tots aquells amb qui hem compartit dijous de beersession.

També un agraïment especial a la gent del volei, per fer unes tardes de primavera tan divertides com saludables! I a la Coral del PRBB, per amenitzar els dilluns al lab!

Abans d'anar a parar a l'IMIM, vaig tenir la sort de poder col·laborar amb l'Elena Bosch i l'Arcadi Navarro a Biologia Evolutiva de la UPF. Estic molt agraïda d'haver pogut fer els primers passos científics amb vosaltres i que confiéssiu tant amb mi. Moltes gràcies!

Gràcies també al Rory Johnson i a la Carme Arnan pels intents de down-regular el NAT. Ha estat un plaer haver pogut col·laborar amb vosaltres.

També vull agrair al Comitè de Tesi que acceptessin donar el vist-i-plau al projecte en els seus inicis.

I also want to thank to the thesis board members to accept to revise this piece of work. Thank you for your time and dedication, for the presence on the defense day and for the discussion of the work.

I also want to thank to AICR (Association of International Cancer Research) for the funding for the PhD thesis during three years.

Voldria donar les gràcies a la Fundació IMIM per l'ajuda de publicació de tesi doctoral.

Però res d'això hagués estat possible si fora del PRBB no hagués tingut l'estima de tots vosaltres.

Colla pessigolla, mil gràcies per tot! Sopars, escapades a Camprodon o la Molina, aniversaris, caps d'any i viatge al Sud-est asiàtic! Amistats com les vostres fa temps que m'he adonat que són molt especials, així que per molts anys! També vull agrair a la resta de la promoció 2004-2009 pels moments en què ens hem anat retrobant al llarg d'aquests anys.

I a Terrassa hi tinc amigues fantàstiques. Gràcies Gemma per ser-hi tots aquests anys. Pels cafès al Progrés, sopars a casa teva i converses de tot que m'han desestressat en molts moments! També gràcies a les Íngrids, pels sopars, casaments i retrobaments. Glòria, gràcies per les excursions, sopars a casa, converses i moments d'esbarjo durant aquests anys. Lara, els nostres soparets amb converses eternes són la millor teràpia en dies baixos de doctorat. Molts ànims en la recta final de la teva tesi! Anna, Núria i Adrià; Festes Majors, concerts, cafès,... gràcies per ser-hi!

Quan et trobes a gust a un lloc, et fan sentir bé facis més o menys, et valoren la feina i sempre hi ha somriures quan t'hi deixes caure, és la millor teràpia! Gràcies amics i amigues de Bastoners, per les enèrgiques Estaperes, bona entesa de la Junta, l'espectacle del Principal, colònies, sopars i cerveses d'aniversari!

Gràcies també als nous amics dels Joves del CEC, per les excursions, sopars, guitarres i esquiades durant aquest darrer any.

També vull agrair al professorat de la Tecnos per formar-me durant 14 anys i treure el millor de mi. Sense la bona base de l'escola no hagués arribat fins aquí! I gràcies també a la gent del CECA per fer-me millor persona!

Moltíssimes gràcies pel recolzament fidel de la família. Avis, cosins i tiets heu estat sempre un coixí on recolzar-me, poder compartir els meus neguits i alegries i fer els caps de setmana molt més càlids.

I ja per acabar, però no menys important... a qui va dedicada aquesta tesi.

Gràcies, Víctor, per ser el meu punt de suport incondicional en el darrer any. Abraçades en el moment necessari, llargues converses al telèfon i els millors consells en moments baixos, han estat vitals per aconseguir escriure aquesta tesi. Amb tu al costat he sentit l'energia per fer-ho possible. Mil gràcies per creure en mi!

Marineta, gràcies per ser-hi sempre, per poder comptar amb tu i poder-ho compartir tot. Tu i jo ens entenem sense paraules, amb una simple mirada o amb telepatia! La família no es tria, però la relació tan única que tinc amb tu no la canviaria! També gràcies a tu, Pol, per l'ajuda inestimable amb els trasllats i per tots els bons consells.

I finalment vosaltres, mama i papa. La vostra estima infinita i el vostre recolzament incondicional fan de casa vostra el millor niu. L'agraïment és infinit, però vull destacar que sempre heu vetllat perquè tinguéssim la millor educació, formal i no formal. I no ho haguéssiu pogut fer millor. Aquesta tesi no podria estar a les vostres mans sense el vostre suport diari. Sempre em recolzeu, m'aconselleu i em doneu suport amb allò que decideixo. No podria

tenir millors pares. I mil gràcies també pel dia a dia a Matadepera durant aquests dos mesos d'escriptura, m'ho heu posat tot molt fàcil i heu estat un gran suport. Us estimo molt i us ho dec tot!

A tots i totes, MOLTES GRÀCIES!

“Sigues lliure i camina cap a l’estel que vulguis guanyar”

(cançó popular catalana)

TABLE OF CONTENTS

FIGURE INDEX

INTRODUCTION

RESULTS

DISCUSSION

TABLE INDEX

MATERIALS AND METHODS

NOTES TO THE READER

All exogenous snail1 used throughout this study is the murine homolog (mmsnail1).

Genes appear in italic case.

ACRONYMS AND ABBREVIATIONS

AGO	Argonaute protein
BOPA	Biotinylated Oligo Pull-down Assay
BSA	Bovine Serum Albumin
CDH1	E-cadherin
CDS	coding Sequence
ChIP	Chromatin Immunoprecipitation
CT	Control
DNA	Deoxyribonucleic acid
dsDNA	Double strand DNA
dsRNA	Double strand RNA
EM	Epigenetic modifier
EMT	Epithelial-to-Mesenchymal Transition
Ezh2	Enhancer of zeste 2
FN1	Fibronectin 1
H3K27me3	Histone H3 lysine-27 trimethylation

H3K4me2	Histone H3 lysine-4 dimethylation
H3K4me3	Histone H3 lysine-4 trimethylation
IP	Immunoprecipitation
KO	Knock-out
LEF1	Lymphoid enhancer factor 1
lincRNA	Long intergenic non-coding RNA
lncRNA	Long non-coding RNA
MET	Mesenchymal-to-Epithelial Transition
miRNA	MicroRNA
mRNA	Messenger RNA
NAT	Natural Antisense Transcript
ncRNA	Non-coding RNA
o/n	Over night
ORF	Open reading frame
PBS	Phosphate Buffered Saline
piRNA	Piwi-interacting RNA
PRC2	Polycomb Repressive Complex 2
qPCR	Quantitative PCR
RBP	RNA-binding protein
RHP	Random Hexamer Primer
RIP	RNA Immunoprecipitation

RNA	Ribonucleic acid
SDS	Sodium Dodecyl Sulfate
siRNA	Small interfering RNA
Sna	Snail1
ssDNA	Single strand DNA
ssRNA	Single strand RNA
TCF	T-cell factor
TGF-β	Transforming growth factor β
TSS	Transcription start site
UTR	Untranslated region
WRE	Wnt response element
w/v	Weight/volume
XIST	X-inactive specific transcript

ABSTRACT

Non-coding RNA functions are emerging in the recent years. In this thesis we describe a Natural Antisense Transcript (NAT) that controls the expression of LEF1 transcriptional factor. This *LEF1* NAT is transcribed from a promoter present in the first *LEF1* intron and undergoes splicing in mesenchymal cells. In epithelial cells, there is no expression of *LEF1* NAT. However, in metastable epithelial cells, *LEF1* NAT is transcribed and a significant part of it remains unspliced and, contrarily to the spliced NAT, down-regulates the main *LEF1* promoter and *LEF1* mRNA and protein expression. Moreover, unspliced NAT also down-regulates cell migration and up-regulates E-cadherin expression. Unspliced *LEF1* NAT interacts with *LEF1* promoter and physically associates with Polycomb Repressive Complex 2 (PRC2) inducing its binding to the *LEF1* promoter and trimethylating Lysine 27 in Histone 3. Spliced *LEF1* NAT prevents the binding between unspliced *LEF1* NAT and *LEF1* promoter, inhibiting *LEF1* promoter repression. Thus, these results indicate that *LEF1* gene expression is finely controlled by splicing of the *LEF1* NAT that, when is not processed, recruits PRC2 to *LEF1* promoter to inhibit it.

RESUM

En els darrers anys, les funcions exercides pels ARN no codificants estan creixent. En aquesta tesi es descriu un Natural Antisense Transcript (NAT) que controla l'expressió del factor de transcripció LEF1. Aquest NAT de *LEF1* és transcrit des del promotor que es troba al primer intró de *LEF1* i es processa mitjançant splicing en les cèl·lules mesenquimals. En les cèl·lules epitelials no hi ha expressió del NAT de *LEF1*. No obstant, en les cèl·lules epitelials que inicien la Transició Epiteli-Mesènquima (EMT), una part significativa de NAT no es processa i, contràriament al NAT que ha estat processat, fa baixar l'activitat del principal promotor de *LEF1* i disminueix l'expressió de *LEF1*, a nivell d'ARN i proteïna. A més, el NAT que no ha estat processat també disminueix la migració cel·lular i incrementa l'expressió de l'E-caderina. El NAT de *LEF1* interactua amb el promotor de *LEF1* i s'uneix físicament amb Polycomb Repressive Complex 2 (PRC2) induint-ne la seva unió al promotor de *LEF1* i trimetilant la Lisina 27 de l'Histona 3. El NAT de *LEF1* que ha estat processat prevé la unió entre el NAT que no ho ha estat i el promotor de *LEF1*, prevenint la repressió del promotor de *LEF1*. Per tant, aquests resultats indiquen que l'expressió de *LEF1* està finament controlada pel processament del NAT de *LEF1* que, quan no ha patit splicing, recluta PRC2 al promotor de *LEF1* per inhibir-lo.

INTRODUCTION

1. Non-coding RNAs

The sequencing of the human genome provided quite a surprise to many when it was determined that there are only ~20.000 protein-coding genes, representing <2% of the total genomic sequence (International Human Genome Sequencing Consortium 2004). Since other less complex eukaryotes have very similar number of protein-coding genes, it quickly became clear that the developmental and physiological complexity of humans probably cannot be explained only by the number of protein-coding genes. Alternative pre-mRNA splicing of protein-coding transcripts as well as post-translational modifications of proteins increase the diversity and functionality of the proteome, likely explaining part of this increased complexity.

The conventional view of gene regulation in biology has centered on protein-coding genes via the central dogma of DNA → RNA → protein. However, over the past decade, evidence from numerous high-throughput genomic platforms suggests that the evolution of developmental processes regulating the complexity of the organism is mainly due to the expansion of regulatory potential of the non-coding portions of the genome[1]. The recent explosion in knowledge demonstrating the importance of non-coding RNAs (ncRNAs) in the regulation of multiple major biological processes impacting development, differentiation and metabolism have brought ncRNAs to one of the main topics of interest of the present biomedical research[2–4] (**Figure I1**).

In contrast to the small ncRNAs such as siRNAs, miRNAs and piRNAs, which are highly conserved and involved in transcriptional and posttranscriptional gene silencing through specific base pairing with their targets, long ncRNAs (lncRNAs) are poorly conserved and

regulate gene expression by diverse mechanisms that are not yet fully understood[5–8].

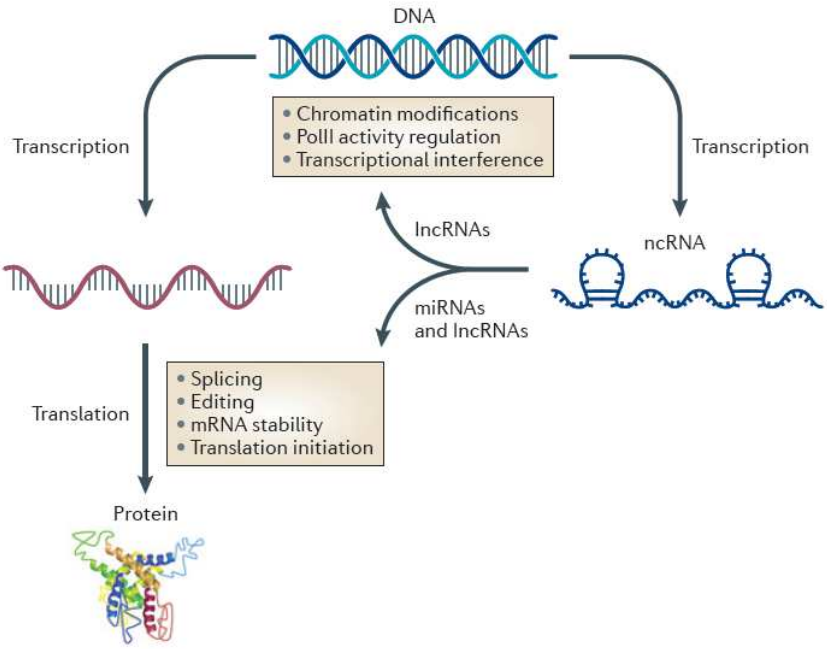


Figure I1. "Central dogma" in the context of regulatory non-coding RNAs[9]. The discovery of a large number of ncRNAs, many of which have the capacity to regulate gene expression at transcriptional and translational level, have changed the "central dogma" of biology. Here it is complemented with aspects of ncRNA functions. As it was first formulated by Francis Crick in 1958: "Once information has passed into protein, it cannot get out again. In more detail, the transfer of information from nucleic acid to nucleic acid or from nucleic acid to protein may be possible, but transfer from protein to protein or from protein to nucleic acid is impossible".

1.1. Long non-coding RNAs

The concept of functional long non-coding RNAs (lncRNAs) was first introduced over 20 years ago, following the description of the X-inactive specific transcript (XIST), the gene that is responsible for X-chromosome inactivation and lacks an open reading frame[10]. Long

non-coding RNAs (lncRNAs) are described as RNA transcripts which can range from 200 nucleotides to multiple kilobases in length, that lack an open reading frame and therefore do not encode protein[9]. The majority of lncRNAs are transcribed by RNA polymerase II, as evidenced by Pol II occupancy, 5' caps, histone modifications associated with Pol II transcriptional elongation and polyadenylation[11]. Their expression levels are frequently at least one order of magnitude lower than mRNAs[12–14]. Furthermore, lncRNAs exhibit poor sequence conservation across species[15]. However, RNA structure is more highly conserved than its sequences[16], which indicates that conventional sequence alignment across species may not reveal functionally conserved RNA motifs. Therefore, poorly conserved lncRNAs may still be functionally active.

It has been argued by Mattick[15] and others that the numbers of lncRNAs increase with evolutionary complexity of the organisms. LncRNAs are the biggest class of ncRNAs with approximately 10.000 lncRNA genes so far annotated in humans[12]. Although the function of most lncRNAs is unknown, the number of characterized lncRNAs is growing and many publications suggest they play roles in negatively or positively regulating gene expression in development, differentiation and human disease[4,17–23].

LncRNAs regulate protein-coding gene expression at both the posttranscriptional and transcriptional level. Posttranscriptional regulation by lncRNAs could occur by competing with endogenous RNAs, like miRNAs do, as well as by modulating mRNA stability and translation by homologous base pairing, or promoting nuclear retention of mRNAs. Transcriptional regulation by lncRNAs could work either in *cis* or in *trans*, and could negatively or positively control protein-coding gene expression[24,25]. LncRNAs work in *cis* when their effects are restricted to the chromosome from which

they are transcribed, and work in *trans* when they affect genes on the other chromosomes[26].

1.1.1. Regulation in *trans*

Some significant examples of lncRNAs that act in *trans* are those that can influence the general transcriptional output of a cell by directly affecting RNAPII activity. One example is the 331 nucleotide 7SK lncRNA, which represses transcription elongation by preventing the PTEFβ transcription factor from phosphorylating the RNAPII carboxy-terminal domain[27] (**Figure I2a**). Another example is the 178 nucleotide B2 lncRNA, a general repressor of RNAPII activity upon heat shock[28]. The B2 lncRNA acts by binding RNAPII and inhibiting phosphorylation of its carboxy-terminal domain by TFIIH, thus disturbing the ability of RNAPII to bind DNA[29,30] (**Figure I2b**).

Regulation in *trans* can also be locus-specific. An example of that is the HOTAIR lncRNA, that it is expressed from the *HOXC* cluster and represses transcription in *trans* across 40 kb of the *HOXD* cluster[31] (**Figure I2d**). HOTAIR interacts with Polycomb repressive complex 2 (PRC2) (epigenetic modifiers (EMs) in the figure) and is required for repressive histone H3 lysine-27 trimethylation (H3K27me3) of the *HOXD* cluster (pc gene in the figure). Targeting EMs by lncRNAs provided a good explanation to explain how EMs gain locus specificity (**Figure I2d**), and it has since been suggested as a general mechanism for *trans*-acting lncRNAs[24,32].

1.1.2. Regulation in *cis*

In contrast to *trans*-acting lncRNAs, which act via their RNA product, *cis*-acting lncRNAs have the possibility to act in two fundamentally different modes.

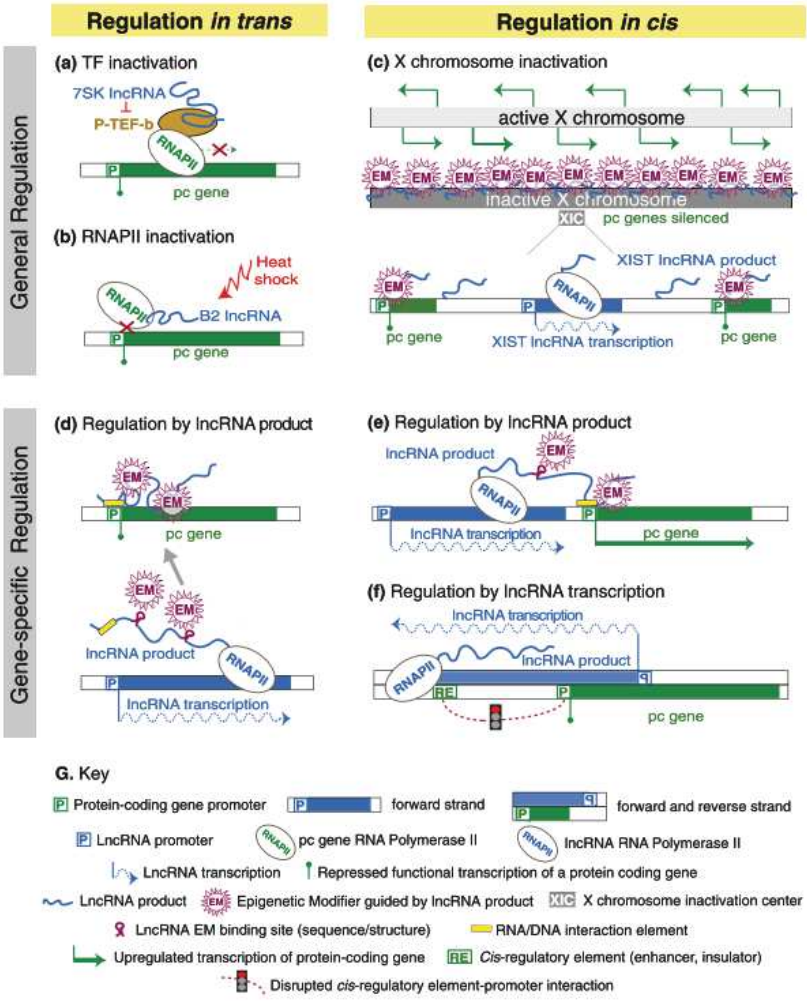


Figure I2. Long non-coding RNAs (lncRNAs) act at different levels to regulate protein coding gene expression[26]. LncRNAs can inhibit general protein-coding (pc) gene expression in *trans* (a) by preventing transcription factor (TF) activity (7SK lncRNA) or (b) by inhibiting RNAPII binding to DNA (B2 lncRNA). (c) XIST lncRNA is transcribed from the X inactivation center (XIC) and inactivates a whole chromosome in *cis* by recruiting epigenetic modifiers (EM). (d) LncRNAs can regulate specific genes, acting in *trans* like HOTAIR or (e) in *cis* like HOTTIP by directly recruiting epigenetic modifiers to certain genomic loci. (f) Transcription of lncRNA can affect protein-coding gene expression in *cis* independent of the lncRNA product, by the transcription process itself.

The first mode depends on the lncRNA product. The main example is the induction of chromosome X inactivation by the XIST lncRNA in female mammals. XIST is expressed from one of the two X chromosomes and induces silencing of the whole chromosome by recruiting epigenetic modifiers[33,34] (**Figure I2c**).

As an example of locus-specific regulation it has been proposed that enhancer RNAs activate corresponding genes in *cis* via their products[35]. A well-studied *cis*-acting lncRNA acting through its product is the human HOTTIP lncRNA that is expressed in the HOXA cluster and activates transcription of flanking genes (**Figure I2e**). HOTTIP was shown to act by binding WDR5 in the MLL histone modifier complex, thereby bringing histone H3 lysine-4 trimethylation (H3K4me3) to promoters of the flanking genes[36]. Such a mechanism in which a nascent lncRNA transcript binds and delivers epigenetic modifiers to its target genes while still attached to the elongating RNAPII is generally termed "tethering" and is often used to explain *cis*-regulation by lncRNAs[24,37] (**Figure I2e**).

In contrast, the second mode of *cis* regulation by lncRNAs involves the process of the transcription itself (**Figure I2f**). Several lines of evidence suggest that the mere process of lncRNA transcription can affect gene expression if RNAPII traverses a regulatory element or changes general chromatin organization of the locus[38].

1.1.3. Long non-coding RNAs mechanisms of action

A different way of classifying lncRNAs is depending on how they do their action. Their molecular ways of action could be classified in signals, decoys, guides and scaffolds[39] (**Figure I3**).

LncRNAs show cell type-specific expression and respond to diverse stimuli, suggesting that their expression is under considerable transcriptional control. As such, lncRNAs can serve as molecular

signals, because transcription of individual lncRNAs occurs at a very specific time and place to integrate developmental signals, interpret cellular context or respond to diverse stimuli. The chromatin state can change merely by the expression of the associated lncRNAs. The advantage of using RNA as a medium suggests that potential regulatory functions can be performed quickly without protein translation. Some examples for this way of action are KCNQ1ot1[40], Air[41], XIST[42], HOTAIR[31,43], LincRNA-p21[44] and COLDAIR[45].

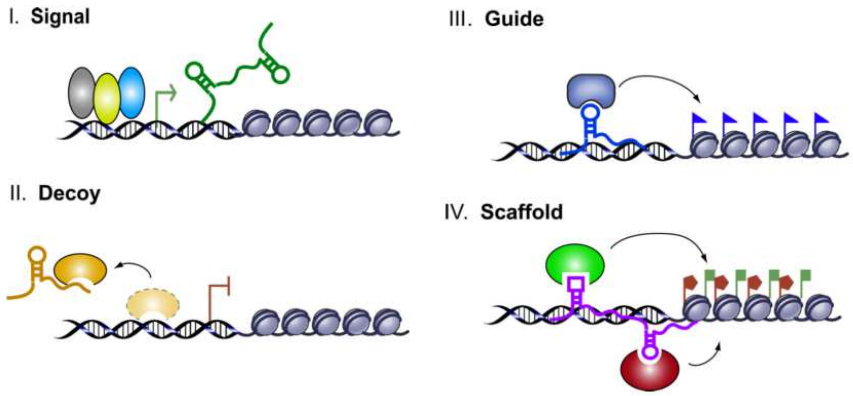


Figure I3. Schematic diagram of the four types of lncRNA mechanism[39]. Type I: as signals, lncRNA expression can faithfully reflect the combinatorial actions of transcription factors (colored ovals) or signaling pathways to indicate gene regulation in space and time. Type II: as decoys, lncRNAs can remove away transcription factors and other proteins from chromatin, or remove the protein factors into nuclear subdomains. Type III: as guides, lncRNAs can recruit chromatin modifying enzymes to target genes, either in *cis* or in *trans*. Type IV: as scaffolds, lncRNAs can bring together multiple proteins to form ribonucleoprotein complexes. The lncRNA-RNP may act on chromatin as illustrated to affect histone modifications. In other instances, the lncRNA scaffold is structural and stabilizes nuclear structures or signaling complexes.

The pervasive transcription of enhancers and promoters[46] hints at a central role for lncRNAs in regulating transcription, both positively and negatively. A major mechanism to do so is acting as molecular **decoys**. These lncRNAs are transcribed and then bind

and remove away a protein target. The RNAs act as a "molecular sink" for RNA-binding proteins (RBPs), which are themselves transcription factors, chromatin modifiers or other regulatory factors. Decoy lncRNAs usually act by negatively regulating an effector. Some examples of this way of action are TERRA[47] and MALAT1[48].

The third mechanism is lncRNA acting as a **guide**. LncRNA binds a protein that directs the localization of ribonucleoprotein complex to specific targets. LncRNAs can guide changes in gene expression either in *cis* or in *trans* in a manner that is not easily predicted based on lncRNA sequence. In principle, lncRNAs can guide chromatin change in *cis* in a cotranscriptional manner (tethered by RNA polymerase) or as a complementary target for small regulatory RNAs; guidance in *trans* can occur by lncRNA binding to target DNA as a RNA:DNA heteroduplex, as RNA:DNA:DNA triplex, or RNA recognition of complex surface of specific chromatin features[49,50]. Some examples of this way of action are XIST[42], Air[41] and COLDAIR[45] for guides in *cis* and HOTAIR[31,43] and LincRNA-p21[44] for guides in *trans*.

The fourth mechanism is lncRNA acting as a **scaffold**. LncRNAs can serve as central platforms upon which relevant molecular components are assembled. In many diverse biological signaling processes, the precise control is vital for specificity and dynamics of intermolecular interactions and signaling events[51]. Traditionally, proteins were thought to be the major players in various scaffolding complexes[52]. Recent evidence, however, raises the possibility that lncRNAs may also play a similar role. For that, lncRNAs acting as scaffold possess different domains that bind distinct effector molecules. Some examples of this way of action are HOTAIR[31,43] and ANRIL[23,53].

1.1.4. LncRNAs and PRC2

Many ncRNAs have been linked to Polycomb Repressive Complex 2 (PRC2) protein complex[54–56]. The recruitment of PRC2 by lncRNAs was first demonstrated for XIST RNA[42,57,58], expressed from the X chromosome. It mediates the recruitment of PRC2 that in turn catalyzes the heterochromatinization of the entire X chromosome. It was then extrapolated to Kcnqtlot1, a lncRNA required for silencing a cluster of imprinted genes on mouse chromosome 7[59]. It was also described for the lncRNA HOTAIR transcribed from the *HOXD* locus on human chromosome 12, required to direct PRC2 in *trans* to the *HOXC* locus on chromosome 2[31], as well as for a class of short ncRNAs produced at CpG island loci in mammalian cells and implicated more widely in PRC2 recruitment to target genes throughout the genome[60]. Finally, a novel ncRNA COLDAIR transcribed from an intron in the *Arabidopsis thaliana* flowering control locus *FLC* has recently been implicated in the control of vernalization (regulation of flowering time by periods of cold) through direct recruitment of the *A. thaliana* PRC2 complex[45]. A key concept to emerge from these studies is that of a direct biochemical interaction between specific ncRNAs and proteins of the PRC2 complex[42], an idea that has generated considerable excitement in the field.

1.1.5. Classification of long non-coding RNAs

Several subclasses of lncRNAs have been catalogued, some of which may overlap.

A *Natural Antisense Transcript (NAT)* to a particular gene is transcribed in the opposite direction, often overlapping with the sense RNA and commonly not coding for protein. Several large-scale analyses have suggested that antisense transcription is very

common in the mammalian genome. Notably, the international FANTOM Consortium demonstrated that at least 25-40% of mammalian protein-coding genes display overlapping transcription[61,62].

In contrast to intragenic ncRNAs such as NATs, the term *"long intergenic non-coding RNAs" (lincRNAs)*[63] is used to refer to lncRNAs that are located between protein-coding genes. However, as it was found that such intergenic RNAs can also be found in introns[64], so the term "lincRNA" is controversial. Recent informatics and experimental approaches have demonstrated that these RNA species can be involved in various processes, including cell cycle regulation and nuclear factor-κB (NF-κB) signalling[11], pluripotency and differentiation[20], as well as somatic tissue differentiation[65]; furthermore, some can regulate gene expression through their association with chromatin-modifying complexes such as PRC2[63].

Another subtype of lncRNA are the *sense overlapping RNAs*, that are RNA transcripts that are encoded on the same DNA strand as another RNA transcripts but have a different sequence.

There also exist the *sense intronic RNAs* that are RNA transcripts that are encoded within an intron of a coding gene but does not overlap with any exons on the same strand.

Finally, we must consider that all these subclasses of lncRNAs can be processed transcripts (those that are spliced) and/or polyadenylated RNAs.

1.2. Natural Antisense Transcripts

Natural Antisense Transcripts (NATs) are described as RNA transcripts originating from the opposite strand of sense RNA

transcripts with which they share sequence complementarity[61,66–68]. The most prominent form of antisense transcription in the mammalian genome is a non-protein-coding antisense RNA partner of a protein-coding transcript[61]. It has been demonstrated that a large number of transcription units contain antisense transcripts. For instance, the FANTOM3 mouse transcriptome sequencing consortium identified Natural Antisense Transcripts for more than 70% of transcription units, most of which represent non-protein-coding RNAs[61].

1.2.1. Characteristics of antisense RNAs

Antisense transcripts are not evenly distributed across the genome. Both ends of protein-coding genes have a propensity for natural antisense transcription[69,70]; specifically, antisense transcription is enriched 250 nucleotides upstream of the transcription start site (TSS)[71,72] and 1.5 kilobases downstream of sense genes[71,73]. The basal expression levels of sense and antisense transcripts in different tissues and cell lines can be either positively or negatively correlated[61,74].

Antisense RNAs have a tendency to undergo fewer splicing events and typically show lower abundance than sense transcripts[69].

There are a number of proposed mechanisms for antisense-mediated regulation of sense mRNA. A way of classifying is: mechanisms related to transcription (including epigenetic interactions), RNA-DNA interactions, RNA-RNA interactions in the nucleus and RNA-RNA interactions in the cytoplasm[7]. Among these four mechanisms, RNA-mediated epigenetic modification has received and increasing amount of experimental support. Antisense transcripts can provide a scaffold for effector proteins to interact with DNA and chromatin in a locus specific way.

A large portion of NATs could exert their regulatory role by binding to chromatin enzymes and recruiting them in *cis* to their targets (**Figure I4**). In favor of this hypothesis, RNA immunoprecipitation (RIP) experiments targeting Ezh2, a subunit of PRC2, coupled with directional RNA sequencing (RIP-seq), revealed that the PRC2 complex associates with almost 10000 RNAs in mouse embryonic stem cells[55]. Almost 3000 of these RNAs are NATs, and around 1000 are bidirectional transcripts. Interestingly, some NATs linked to disease loci were found to immunoprecipitate with Ezh2 such as *Hspa1α*-AS, *Bgn*-AS, *Foxn2*-AS and *Malat1*-AS. Thus, the presence of NATs associated with PRC2 suggests the importance of these RNA transcripts in mediating the recruitment of chromatin-modifying complexes.

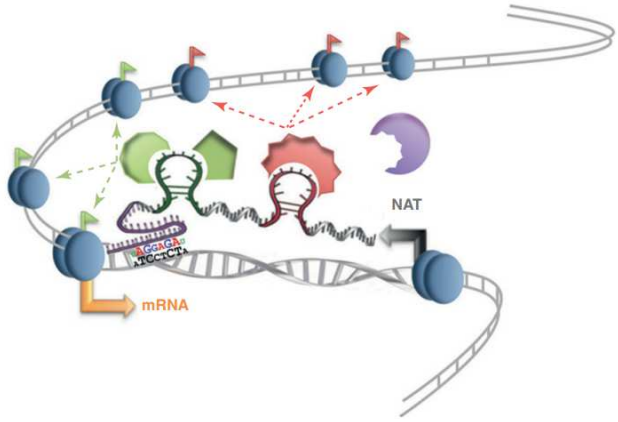


Figure I4. Epigenetic regulation induced by NATs[37]. NATs regulate the epigenetic landscape of genomic loci from which they are transcribed (*cis* regulation). A specific secondary structure permits the NAT to interact with different chromatin-modifying enzymes (green, red and purple shapes), thereby coordinating their action and directing specific epigenetic modifications of the nearby chromatin (green and red flags). Locus specificity may be achieved through sequence-specific interactions between the NAT and the DNA.

Evidence of a functional interaction between NATs and PRC2 comes from a study on the cyclin-dependent kinase inhibitor p21, another important tumor-suppressor gene. Bidirectional

transcription at the *p21* locus generates an antisense transcript and *p21* mRNA. The *p21* NAT represses *p21* mRNA in a process involving the deposition of the repressive histone mark H3K27me3[75]. This mechanism is AGO1-dependent. Thus, depending on the cellular context, an imbalanced expression of NATs can result in the silencing or activation of partner protein-coding genes, providing an interesting potential mechanism to explain the aberrant up-regulation of silencing of cancer-related genes.

2. Epithelial-to-Mesenchymal Transition (EMT)

Epithelial-to-Mesenchymal Transition (EMT) is the process that includes all the cellular and molecular changes undergone by a well-differentiated epithelial cell in order to be converted into a mesenchymal-like cell. The process of EMT was originally detected at specific stages of embryo development in which epithelial cells migrate and invade other territories in order to form new structures. However, it has also been described to be crucial in pathological situations such as fibrosis and cancer[76,77]. This process is characterized by three major changes in the cellular phenotype[78,79]:

- Morphological changes from a cobblestone-like monolayer of epithelial cells with an apical-basal polarity to a disperse, spindle-shaped mesenchymal cells with migratory protrusions.
- Changes of differentiation markers from cell-cell junction proteins and cytokeratin intermediate filaments to vimentin filaments and fibronectin.
- Functional changes associated with the conversion of stationary cells to motile cells with the capacity of invasion through the extracellular matrix.

Epithelial cells are characterized by the presence of the molecular marker E-cadherin, a transmembrane protein that plays a crucial role in the establishment of adherens junctions. This molecule forms homotypic calcium-dependent interactions between adjacent cells. Its intracellular domain associates with actin cytoskeleton through the catenin family proteins. Adherens junctions, together with the tight junctions and desmosomes in the apical part of the basolateral membrane, are essential to seal the intercellular space between cells and to form the permeability barrier[80,81].

On the other hand, mesenchymal cells are characterized for their fibroblastic phenotype and low cell-cell contacts. Mesenchymal cells have an elongated morphology, with a front-back end asymmetry that facilitates motility and locomotion. They also have filopodia at the leading edge and are enriched with integrin receptor and metalloproteinases that digest the basement membrane providing invasive motility[82].

2.1. Importance of EMT

2.1.1. Physiological EMT

Physiological EMT takes places at different moments during embryonic development in mammals. The formation of the mesoderm and neural crest delamination are the two main examples where EMT is crucial[83].

In adults, EMT is essential for wound healing. Repair of the dermis require the recruitment of active fibroblasts to heal the wound and it has been described that at least a significant part of these active fibroblasts proceed from EMT[84].

Moreover, it is generally accepted that EMT can be reversible in a process called Mesenchymal-to-Epithelial transition (MET). This process is naturally occurring during development, for example during nephron formation in developing kidney[85].

Both EMT and MET processes reflect cell plasticity and suggest that interconversion between epithelial and mesenchymal cells is a common feature during development.

2.1.2. Pathological EMT

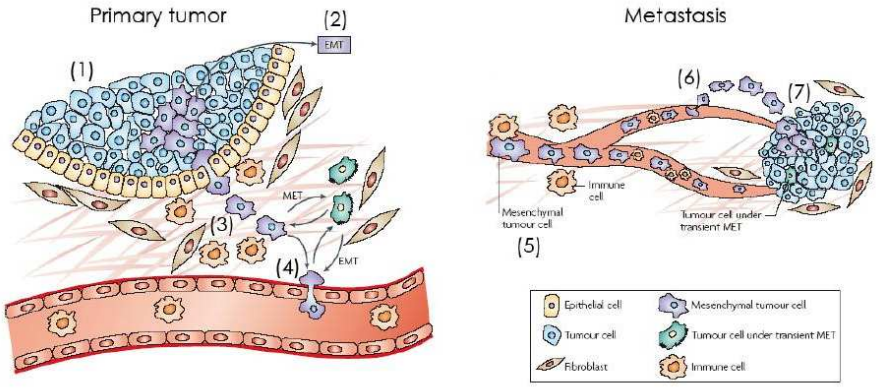


Figure I5. EMT and MET in tumor progression and metastasis (adapted from Peinado et al. 2007[86]). Representation of the process of metastatic formation. (1) Advanced stage carcinoma, (2) some cells in the tumor undergo EMT and start migrating to other tissues (3) and intravasating. (4) Mesenchymal tumor circulating cell (5) arrives to selected organ and extravasates (6) where cells undergo MET and start to grow as metastasis (7).

EMT plays a key role during cancer progression (**Figure I5**). An invasive malignant tumor derived from an epithelial tissue which tends to invade other areas is called carcinoma. During carcinoma progression, advanced tumor cells frequently show down-regulation of epithelial markers, loss of cell polarity and reduced intercellular adhesions. This process is often accompanied by an

increase in cell motility and an up-regulation of mesenchymal gene expression. This change of phenotype is EMT and it is associated to carcinoma features, such as loss of contact inhibition, altered growth control and enhanced invasiveness. EMT correlates with poor prognosis of the disease and poor histologic differentiation, as well as to destruction of tissue integrity and metastasis. Thus, it has been considered a crucial event in tumor progression[87–90].

This data show that EMT underlies critical steps of tumor progression and it shows the high relevance of EMT study in cancer research.

2.2. Key molecules in EMT

2.2.1. E-cadherin

E-cadherin is the prototypic type I cadherin that mediates homophylic intercellular interactions by formation of bonds between one or more immunoglobulin domains in the extracellular domain. It binds indirectly to actin microfilaments through α-catenin and β-catenin in the cytoplasm[91–93]. E-cadherin is present in the adherens junctions and is one of the most important epithelial markers: its functional loss is considered a hallmark of EMT.

When it is down-regulated, E-cadherin-mediated sequestering of β-catenin is abolished, stimulating transcriptional activity through LEF/TCF[94]. Moreover, disruption of E-cadherin contacts allows the activation of several signaling pathways involved in EMT such as MAPK[95], RhoA[96], ILK[97] and NF-κB[98,99].

Nevertheless, this disruption is not enough to trigger EMT since the loss of E-cadherin is not sufficient to achieve the full activation of mesenchymal genes[100]. During tumor progression, *CDH1* (E-

cadherin) can be functionally inactivated by different mechanisms including somatic mutation and promoter methylation, but the most important and prevalent mechanism consists in the transcriptional repression[101]. Studies performed on the *CDH1* gene have identified short bases elements named E-boxes (5'-CACCTG-3' or 5'-CAGGTG-3') that are responsible for *CDH1* transcriptional repression in mesenchymal cells[102,103]. Those E-box elements are used as direct targets by several transcription factors acting downstream of the different pathways promoting EMT and are located in *CDH1* promoter.

2.2.2. Snail1

Snail transcriptional repressors, and particularly Snail1, are the most widely studied effectors of EMT and *CDH1* repression. Snail family members belong to the Snail superfamily of transcription factors, composed by the Snail and the Scratch families[104]. The three vertebrate members belonging to the Snail family are known as Snail1 (properly Snail), Snail2 (formerly Slug) and the less characterized Snail3 (Smuc). All the family members encode transcription factors of the zinc-finger type and all share a similar organization with a highly conserved C-terminal domain, which contains from four to six C_2H_2 type zinc fingers responsible for DNA binding through the E-boxes referred before. Several studies have demonstrated that Snail1 blocks expression of E-cadherin by directly binding to the E-boxes present on its promoter[105,106]. This correlates with an increase of the interaction with histone deacetilaces 1 / 2 (HDAC 1 / 2) through the Sin3A correpressor[101] and with the recruitment of PRC2 to the *CDH1* promoter, and posterior trimethylation of the Lysin 27 in the Histone 3 (H3K27me3), the most typical transcriptional repression mark[107].

Snail1 pro-metastatic function was strongly highlighted when our lab and others reported that transfection of *snail1* in epithelial cells promotes down-regulation of E-cadherin and is sufficient to trigger EMT[105,106]. Accordingly, Snail1 was also pointed as key factor in developmental EMT, as *snail1* KO mouse shows defects in mesoderm formation due to impaired E-cadherin down-regulation[108].

Interestingly, Snail1 is also able to stimulate mesenchymal genes transcription, although little is known about its mechanism. It has been demonstrated that Snail1 is able to increase the levels of extracellular matrix proteins such as fibronectin (FN1)[106,109,110], matrix metalloproteinases[111] and cytoskeleton proteins such as vimentin[106], regulatory proteins like RhoB[112] or transcription factors such as LEF1 or Zeb1 and Zeb2[109,113]. Snail1 is also involved in survival by down-regulating PTEN, caspases and p53[114–117].

2.2.3. LEF1

Lymphoid enhancer factor 1 (*LEF1*) is one of the genes up-regulated during EMT. It can be up-regulated via different pathways, such as TGF-β or Wnt signaling.

LEF1 mediates Wnt signaling via recruitment of β-catenin to target genes[118]. It has been reported that about 80% of colon tumors exhibit aberrant activation of *LEF1* gene expression[119,120]. Genetic mutations in the Wnt pathway lead to stabilization of β-catenin, a cytoplasmic-nuclear shuttling protein with a potent transcription activation domain. Stabilization and subsequent nuclear localization of β-catenin produces aberrant, Wnt-independent signals to target genes, an activity tightly linked to the genesis of colon cancers[121]. In the nucleus, the transcription factor family of LEF1/TCF proteins transmits Wnt signals by binding to β-catenin and recruiting it to target genes for activation.

LEF1 is also a key molecule in the EMT triggered by TGF-β (**Figure I6**). TGF-β up-regulates Snail and Slug expression via MAPK signaling. Snail and Slug bind to *CDH1* promoter and inhibit its gene expression, limiting β-catenin substrate binding at the cell membrane. β-catenin is then free in the cytosol and can translocate to the nucleus, where it binds to TCF4 to form β-catenin-TCF4 complexes that will promote transcription of mesenchymal genes such as *LEF1*, *Fibronectin*, *Vimentin* and *α-SMA*. LEF1 is also a substrate for β-catenin and they also form a complex that promotes transcription of mesenchymal genes. As shown in **Figure I6**, β-catenin is further stabilized in the cytoplasm by the action of TGF-β through PI3K. PI3K can signal molecules such as ILK and AKT, which can phosphorylate and inactivate GSK-3β, a protein that targets both Snail1 and β-catenin for degradation through the ubiquitin proteasome pathway. This high amount of cytoplasmic β-catenin achieved through GSK-3β phosphorylation and *CDH1* repression will also promote association with LEF1, upon which these β-catenin-LEF1 complexes will translocate to the nucleus to promote transcription of mesenchymal genes.

LEF1 mRNA is generated from chromosome 4 (specifically from 4q23-q25) from two different Transcriptional Start Sites (TSS). Four different transcripts are produced: transcripts 1, 2 and 3 start at the position -1189 from the Translational Start Site (from now on: position +1. NP_057353.1) (**Figure I7**), they share the great majority of the sequence and they finish differently; transcript 4 starts at position +969 (after the Translational Start Site) and then shares a big part with the other transcripts.

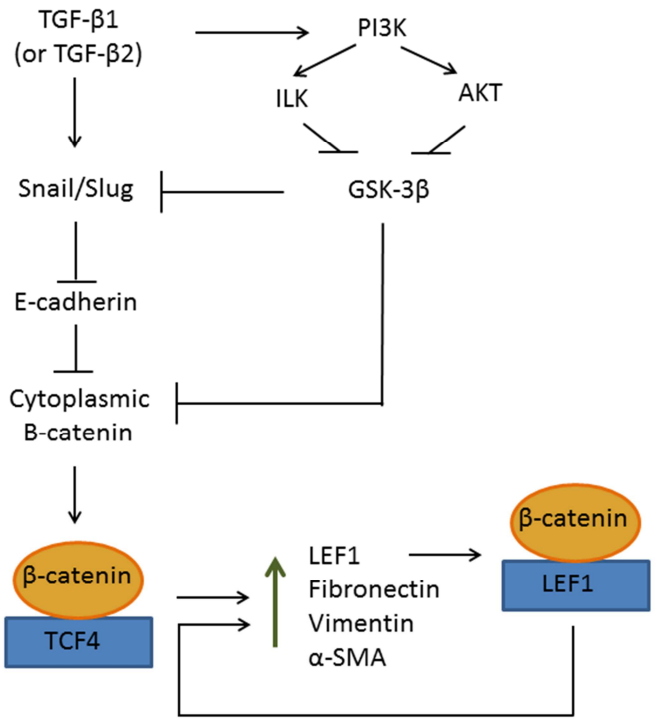


Figure I6. Schematic diagram of the signalling mechanism that stimulates LEF1[122]. TGF-β1 (or TGF-β2) promotes the up-regulation of *CDH1* repressors Snail and Slug via MAPK signaling. Snail and Slug inhibit *CDH1* gene expression, limiting β-catenin substrate binding at the cell membrane. Through this mechanism, β-catenin-TCF4 complexes up-regulate synthesis of genes such as *LEF1*, *Fibronectin*, *Vimentin* and *α-SMA*. LEF1 is also a substrate for β-catenin, and together they promote transcription of mesenchymal genes. β-catenin is further stabilized in the cytoplasm by the actions of TGF-β1, TGF-β2 and TGF-β3 through PI3K. Activation of ILK and AKT, which can phosphorylate and inactivate GSK-3β, increases the levels of β-catenin. This high amount of cytoplasmic β-catenin will also promote association with LEF1, upon which these β-catenin-LEF1 complexes will translocate to the nucleus to promote transcription of mesenchymal genes.

The full length *LEF1* mRNAs (transcripts 1, 2 and 3) contain a β-catenin binding domain at the N-terminus and a DNA-binding domain near the C-terminus. This form acts in Wnt signaling as its DNA-binding domain recognizes Wnt response elements (WREs) in

target genes and its N-terminal domain recruits β-catenin to those target genes for activation.

Transcript 4 encodes a smaller form of LEF1 missing the β-catenin domain. This isoform is a dominant negative of LEF1 as it retains the ability to bind to WRE but it cannot recruit β-catenin[123]. Therefore, the dominant negative of LEF1 suppresses Wnt target gene activation and opposes the actions of full length LEF1 and other LEF/TCF factors.

Thus, understanding *LEF1* gene regulation during EMT is a very interesting point in cancer research.

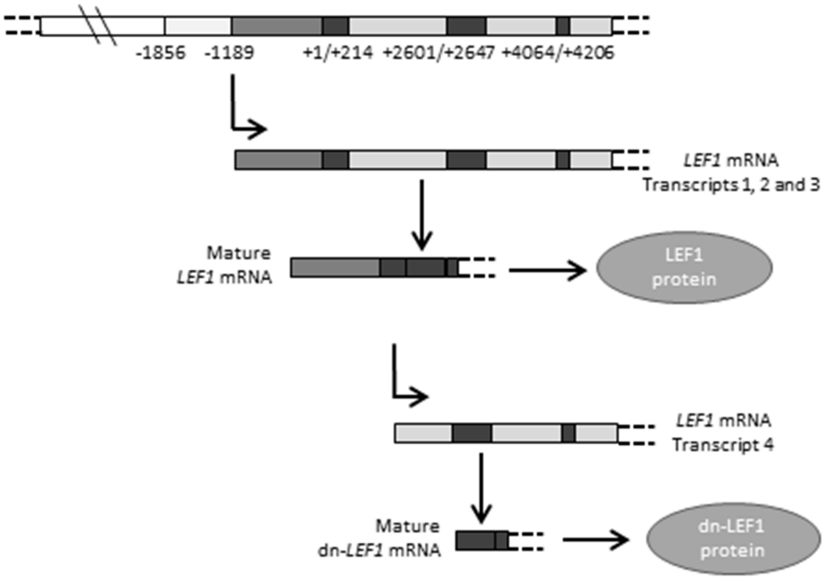


Figure I7. Diagram of the *LEF1* DNA locus and the *LEF1* mRNA transcripts that encodes. In chromosome 4q23-q25, different RNAs are transcribed. The 5'UTR (from -1189 to +1) is indicated in grey. Exons are indicated in dark grey and introns in light grey.

2.2.4. LncRNAs and EMT

Very little is known about how lncRNAs regulate EMT. Previous work in our lab has demonstrated that a Natural Antisense Transcript (NAT) is required for the expression of Zeb2, a transcriptional repressor of *CDH1*[113].

During EMT, Snail1 does not affect the synthesis of *Zeb2* mRNA, but prevents the processing of a large intron located in its 5'-untranslated region (5'-UTR) (**Figure I8**). This intron contains an internal ribosome entry site (IRES) necessary for the expression of Zeb2. Maintenance of 5'-UTR *Zeb2* intron is dependent on the expression of a NAT that overlaps the 5' splice site in the intron. Ectopic overexpression of this NAT in epithelial cells prevents splicing of the *Zeb2* 5'-UTR, increases the levels of Zeb2 protein, and consequently down-regulates E-cadherin mRNA and protein. It has been demonstrated a strong association between NAT presence and conservation of the 5'-UTR intron in cells that have undergone EMT and also in human tumors with low E-cadherin expression.

Thus, this is the unique study published to date relating NATs and EMT, where the existence of a NAT capable of activating Zeb2 expression explains the mechanism involved in this activation and demonstrates that this NAT regulates E-cadherin expression. For that, we wanted to further study the regulation of other genes implicated in EMT by Natural Antisense Transcripts.

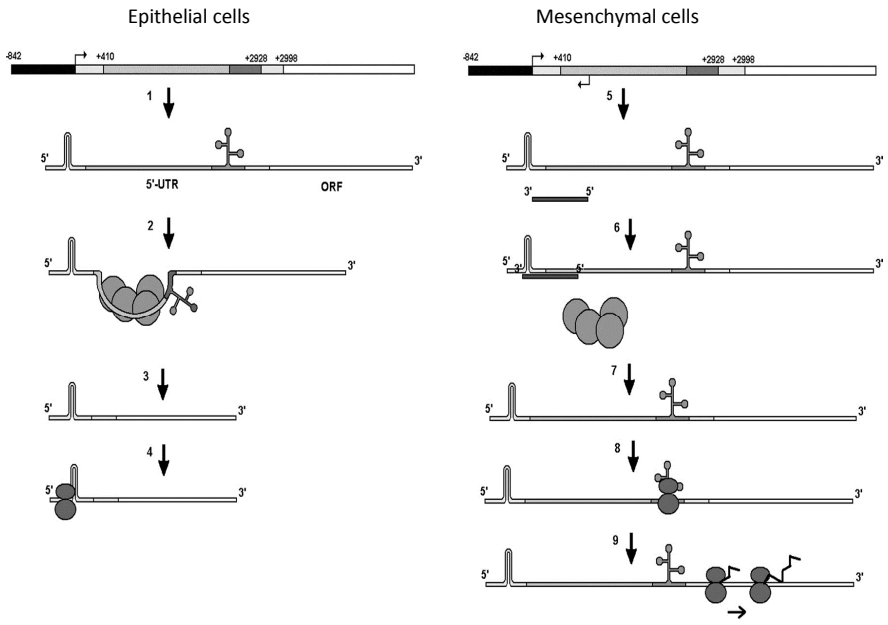


Figure I8. A model for the regulation of Zeb2 expression in epithelial and mesenchymal cells[113] (Left side) (1) In epithelial cells, transcription of the main promoter of *Zeb2* gene (in black) generates an RNA composed of a 3-kb 5'-UTR (in grey) and the ORF (in white). Upon binding of the spliceosome (2), an intron corresponding to 2.5 kb is eliminated, generating a processed transcript with a 5'-UTR of 481 nucleotides (3). (4) This 5'-UTR contains a sequence that inhibits scanning by the ribosomes and therefore prevents translation of Zeb2. (Right side) (5) After completion of the EMT, transcription of Zeb2 is accompanied by expression of a NAT depending of the activation of a different promoter placed 5' downstream. Expression of this NAT is greater than the long transcript and prevents binding of the spliceosome to the 5' splice site (6) and, consequently, the intron present in the 5'-UTR is conserved (7). This intron contains an IRES situated close to the start of translation. Binding of the ribosomes to this IRES (8) makes the translation of Zeb2 protein possible (9).

OBJECTIVES

The general objective of this PhD thesis was to investigate the regulation by Natural Antisense Transcripts of genes involved in Epithelial-to-Mesenchymal Transition.

In particular, we focused on understanding how different isoforms of a Natural Antisense Transcript modulate *LEF1* gene expression.

RESULTS

1. Natural Antisense Transcripts in Epithelial-to-Mesenchymal Transition

As mentioned in the introduction, little is known about how Natural Antisense Transcripts regulate Epithelial-to-Mesenchymal Transition (EMT).

Using the database NATsDB (http://natsdb.cbi.pku.edu.cn/), where NATs are listed and associated with their mRNA partners, we looked for NATs corresponding to key molecules in EMT. We analyzed these NATs performing semi-quantitative RT-PCRs using strand specific oligo for the RetroTranscription. For the PCR amplification, the second oligo was added in the reaction. These RT-PCRs were named **oligo-specific** or **strand-specific oligo RT-PCRs**. We used RNA extracts of RWP1 cell line (liver metastasis of ductal pancreatic adenocarcinoma), with or without stable expression of snail1, that induces an EMT. *CDH1* and *PTEN* are epithelial genes, whereas *Fibronectin 1* (*FN1*), *LEF1* and *Twist* are mesenchymal. As observed in **Figure R1**, NATs corresponding to these genes are regulated during EMT. *CDH1* NAT is more expressed in mesenchymal than in epithelial cells, so *CDH1* NAT and mRNA have an inverse correlation. This may indicate that *CDH1* NAT could have a negative effect of the E-cadherin expression. *PTEN* NAT is more expressed in epithelial cell, as *PTEN* mRNA; thus it can have a positive effect of it. *FN1*, *LEF1* and *Twist* NATs are more expressed in epithelial cells, although in the case of *Fibronectin 1* (*FN1*) the down-regulation during EMT is weak, whereas their mRNAs partners are expressed in mesenchymal cells. Thus, we could say that these NATs might have a negative effect on the expression of their mRNA partners.

Several NATs such as *PTEN*, *Twist* and *LEF1* were candidates to play a relevant role in the regulation of their partner mRNA. In these

cases, the overlapping sequence with the partner mRNA was completely different from one to each other. In the case of *PTEN* and *LEF1*, the overlapping was between the 5' ends of both transcripts, whereas for *Twist*, the overlapping was between the 3'ends of the transcripts. This suggested that the possible NAT mechanism of regulation would be different in each case.

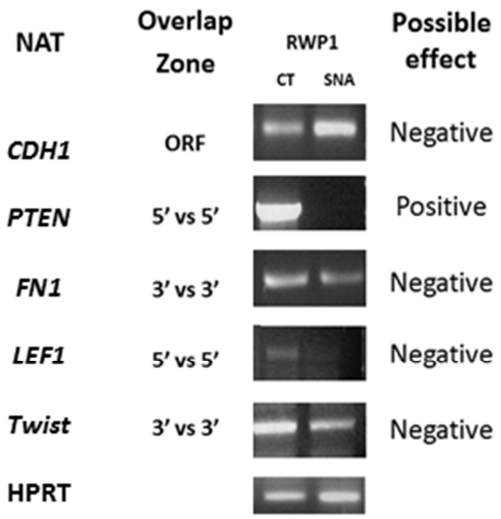


Figure R1. NATs of various genes are regulated during EMT. Total RNA extracts of RWP1 control (CT) and RWP1 snail1 (SNA) cell lines were extracted and analyzed by semi-quantitative RT-PCR using specific oligos in order to detect the NAT of the diverse genes. Picture shows the results of one of three experiments performed.

2. *LEF1* locus drives sense and antisense transcription

LEF1 NAT was chosen for further analysis. As explained in the introduction, *LEF1* mRNA is generated from chromosome 4 (4q23-q25) from two different Transcriptional Start Sites (TSS). The full-length *LEF1* mRNA starts at the position -1189 from the

Translational Start Site (position +1)[124] (**Figure R2**), whereas a dominant negative form starts at position +969 (not shown in Figure R2). This protein retains the ability to bind to Wnt response elements (WRE) but it cannot recruit β-catenin. Moreover, the full-length *LEF1* mRNA has different exons and it is processed by splicing before being translated into LEF1 protein.

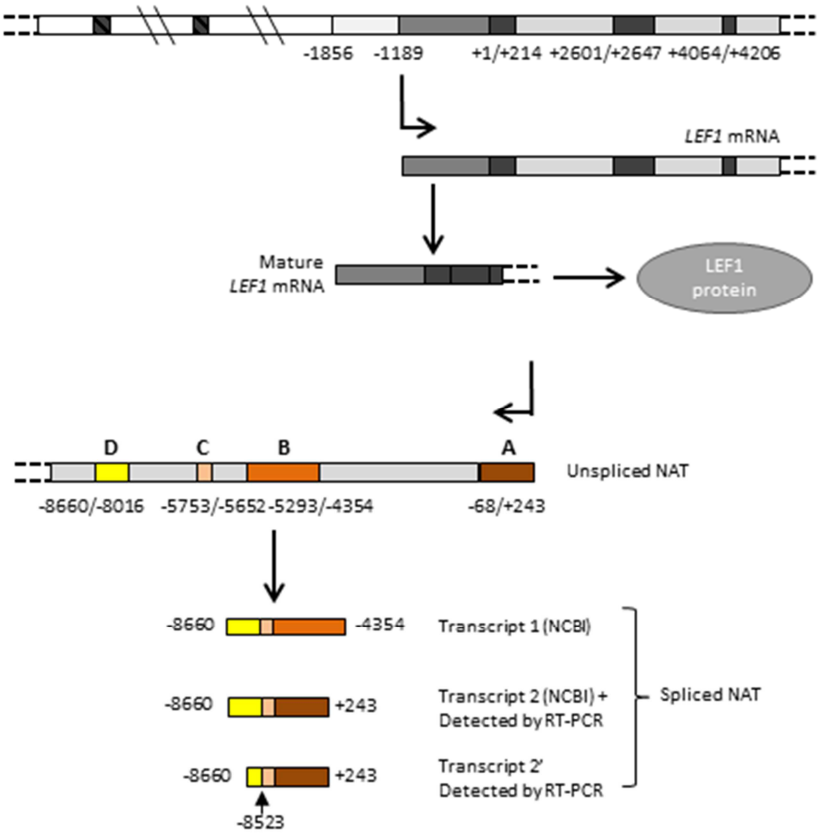
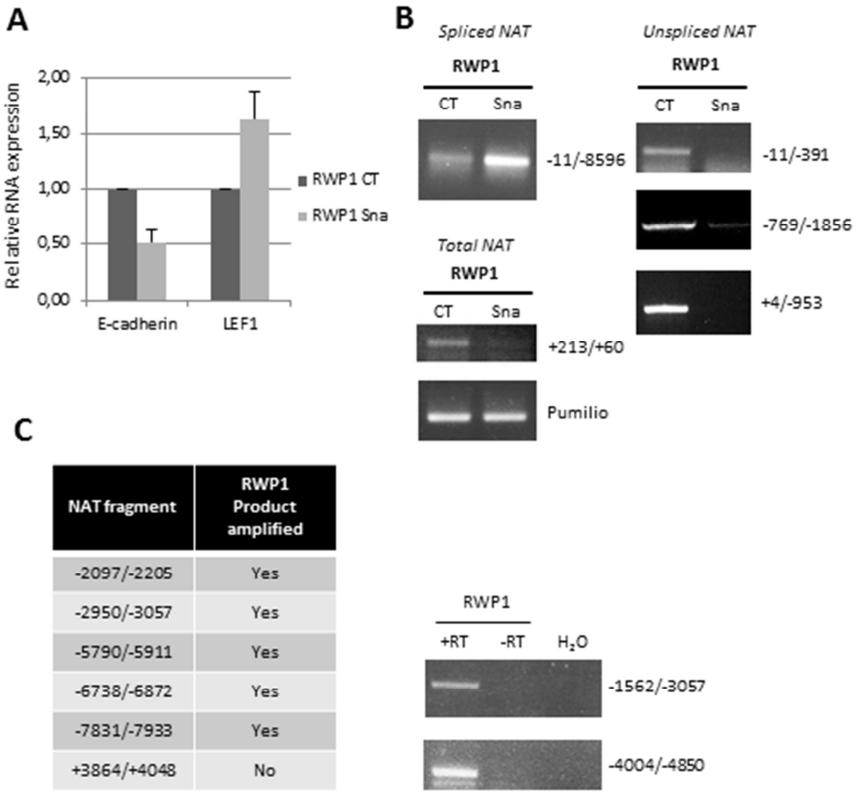


Figure R2. *LEF1* locus contains sense and antisense transcription. In chromosome 4q23-q25, different RNAs are transcribed. *LEF1* mRNA (sense) starts at position -1189 and presents a big 5'UTR (dark grey). The exons are indicated in black and introns in light grey. *LEF1* mRNA is spliced and translated into LEF1 protein. *LEF1* NAT (antisense) starts at position +243 and finishes at position -8660. The different transcripts are indicated.

From the same DNA locus, antisense transcription also occurs. In NCBI, different antisense transcripts for *LEF1* gene are described. Besides the unspliced form of the Natural Antisense Transcript, detected in our analysis presented in **Figure R1**, *LEF1* NAT can undergo alternative splicing. Transcript 1 contains exons B, C and D; and transcript 2 contains exons A, C and D (**Figure R2**). Transcript 1 does not overlap with *LEF1* mRNA, so that is why we focused our view on transcript 2 (spliced *LEF1* NAT). *LEF1* NAT starts at the position +243 (from the Translational Start Site) in the opposite strand of the *LEF1* mRNA and extends to up to 9 kb. It overlaps with the first exon, the 5'UTR and the promoter of *LEF1* mRNA. We detected by RT-PCR both the unspliced and spliced NAT. For the spliced NAT, we detected the transcript 2 (+243/-68, -5652/-5753, -8016/-8660) and also a shorter form 2' (+243/-68, -5652/-5753, -8523/-8660) (**Figure R2**).

We analyzed the expression pattern of the transcripts generated in this locus. We used total RNA extracts from the RWP1 cell line with or without stable *snail1* transfection. As observed in **Figure R3A**, in cells expressing snail1, E-cadherin levels were down-regulated and *LEF1* mRNA levels up-regulated as expected. In **Figure R3B**, we determined the relative amount of the different forms of the NAT between RWP1 control and RWP1 snail1 cells. For that, we performed strand-specific oligo RT-PCRs. As shown in **Figure R3B**, spliced and unspliced NAT show opposite patterns. Spliced NAT was clearly detected in RWP1 snail1 cells whereas unspliced NAT was detected in RWP1 control cells. When looking to the global levels of NAT (PCR product +213/+60 that is common for the spliced and unspliced NAT), we observed that there is more NAT in RWP1 control cells, indicating that there is more unspliced NAT in RWP1 control cells than spliced NAT in RWP1 snail1 cells.



NAT fragment	RWP1 Product amplified
-2097/-2205	Yes
-2950/-3057	Yes
-5790/-5911	Yes
-6738/-6872	Yes
-7831/-7933	Yes
+3864/+4048	No

Figure R3. Unspliced *LEF1* NAT is expressed in RWP1 cells whereas spliced NAT is in RWP1 snail1. **A**. RNA extracts form the indicated cell lines were analyzed by RT with Random Hexamer Primer (RHP) and qPCR with oligos for E-cadherin and *LEF1* mRNA. The RNA levels are corrected for the levels of Pumilio. Results show the average ± standard desviation of two experiments performed in duplicate. **B**. Oligo-specific RT-PCRs were performed in order to detect spliced, unspliced and total NAT. The products were sequenced in order to confirm them. The spliced NAT product -11/-8596 is 252bp long and consists of -11/-68, -5652/-5753 and 8523/-8596 (transcript 2'). **C**. *LEF1* NAT fragments were analyzed in RNA from RWP1 cell line. In the left, RT with RHP followed of qPCR was performed. +3864/+4048 fragment was amplified as negative control of expression. In the right, oligo-specific RT-PCR was performed in order to detect bigger products of the NAT.

Using Rapid Amplification cDNA 5'end technique we verified that the Transcription Start Site (TSS) is located at position +243 (data not shown). Performing RT with Random Hexamer Primer (RHP)

and qPCRs in RNAs form RWP1 cell line, we were able to detect that *LEF1* NAT extended to at least position -7933 (**Figure R3C** left). By oligo-specific semi-quantitative RT-PCR, bigger fragments of the NAT were also detected (**Figure R3C** right). This suggested the continuity of *LEF1* NAT transcript as we detected more than 1kb long fragments overlapping between them (+4/-953, -769/-1856 and -1562/-3057).

In order to determine if the unspliced *LEF1* NAT was polyadenylated or not, we performed RetroTranscription with Random Hexamer Primer (RHP) or with oligo dT in RWP1 RNAs. We then performed qPCR for the NAT (region -1562/-1688) and also HPRT to normalize. Calculating the coefficient oligo dT/RHP for each transcript, we could estimate the percentage of polyadenylation of the transcripts. As shown in **Figure R4**, HPRT was equally amplified after RetroTranscription with oligo dT and with RHP, as it is 100% polyadenylated. With this method, we observed that polyadenylation of unspliced *LEF1* NAT was 34%, as we amplified it better when RetroTranscription was performed with RHP than with oligo dT. We could conclude that unspliced *LEF1* NAT is not extensively polyadenylated.

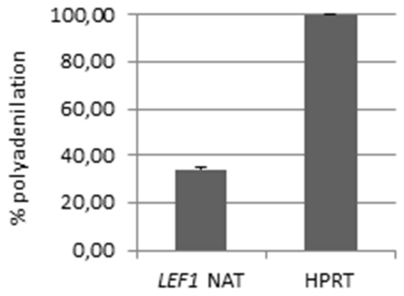


Figure R4. Unspliced *LEF1* NAT is not extensively polyadenylated. RNA extracts from RWP1 cells were analyzed. RT was performed with RHP or Oligo dT, and qPCR with *LEF1* NAT oligos -1562/-1688 and HPRT. The coefficient oligo dT/RHP was performed and 100% was given to HPRT. Results show the average ± standard desviation of two experiments performed in duplicate.

We then separated nuclei from cytoplasms from RWP1 cells. We extracted RNA from both fractions and analyzed the levels of unspliced *LEF1* NAT and HPRT (**Figure R5**). As expected, HPRT was clearly enriched in the cytoplasmic fraction, as it is exported there for translation. Unspliced *LEF1* NAT was clearly enriched in the nuclear fraction, where we could detect 93.5% of the unspliced *LEF1* NAT in the cell. Thus, unspliced *LEF1* NAT is mainly localized in the nucleus.

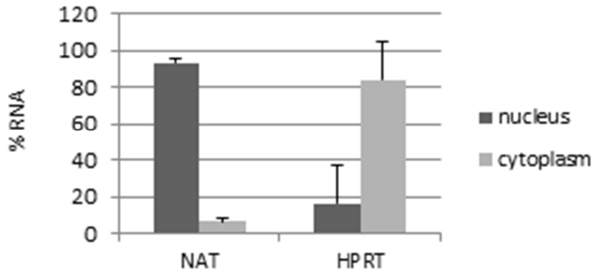


Figure R5. Unspliced *LEF1* NAT is enriched in the nucleus. Nuclei and cytoplasms from RWP1 cells were separated and RNA was extracted from both fractions. RT was performed with RHP and qPCR with *LEF1* NAT oligos -1562/-1688 and HPRT. The percentage of each RNA specie in each fraction was calculated. Results show the average ± standard desviation of three experiments performed in triplicate.

We analyzed the *LEF1* NAT levels in two different cell lines in order to determine if it was also differently expressed in epithelial versus mesenchymal cells. We used HT29 M6 (colon adenocarcinoma) with or without stably transfected *snail1* and SW480 (colon adenocarcinoma) with or without stably transfected *CDH1*. SW480 is an epithelial cell line but with an intermediate morphology. At low confluence, it presents low E-cadherin levels and as it has a truncated APC, it has the β-catenin/TCF pathway highly active. Thus, SW480 control cells express high levels of *LEF1* mRNA. Therefore, SW480 with stably-transfected *CDH1* cells have an epithelial phenotype.

In SW480 cells, when overexpressing E-cadherin, E-cadherin levels were highly increased and LEF1 levels decreased, as expected (**Figure R6**). We analyzed the *LEF1* NAT levels and found that both total NAT and unspliced NAT were only present in SW480 control cells.

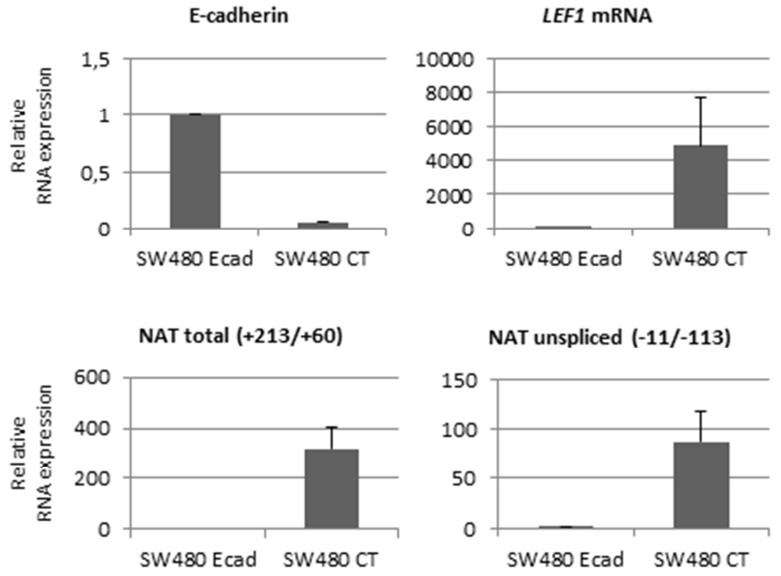


Figure R6. *LEF1* NAT is present in SW480 control cells. RNA extracts from SW480 E-cadherin and SW480 control cells were analyzed. RT with RHP was performed for qPCR of E-cadherin, *LEF1* mRNA and Pumilio. Oligo-specific RT was performed for total *LEF1* NAT and unspliced *LEF1* NAT analyses. The RNA levels were normalized to Pumilio. Results show the average ± standard desviation of the values of the experiment.

We then analyzed RNAs from HT29 M6 cell line. As shown in **Figure R7A**, E-cadherin levels were down-regulated with snail1 overexpression and *LEF1* mRNA levels up-regulated, as expected. When analyzing the *LEF1* NAT levels, we also found that both total NAT and unspliced NAT were up-regulated in snail1 expressing cells.

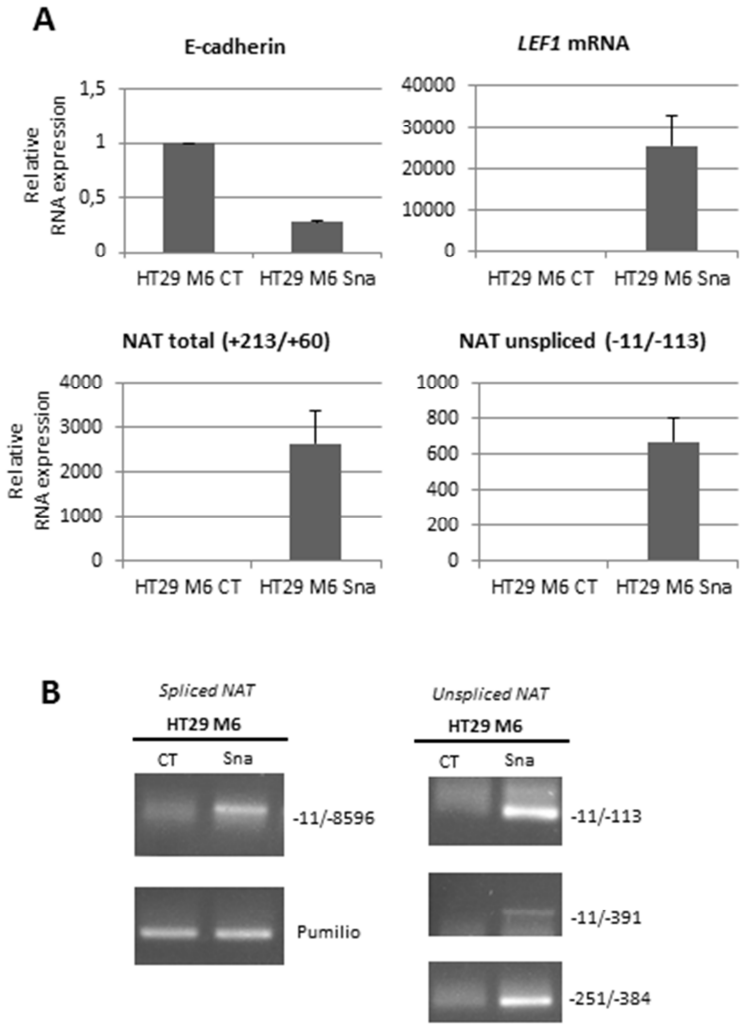


Figure R7. *LEF1* NAT is present in snail1 expressing cells for HT29 M6. RNA extracts from HT29 M6 control and HT29 M6 snail1 cells were analyzed. **A.** RT with RHP was performed for qPCR of E-cadherin, *LEF1* mRNA and Pumilio. Oligo-specific RT was performed for total NAT and unspliced NAT analyses. The RNA levels were normalized to Pumilio. Results show the average ± standard desviation of the values of the experiment. **B**. Oligo-specific RT-PCRs were performed in order to detect spliced and unspliced NAT. Pumilio is shown as loading control.

We analyzed more carefully the expression of spliced NAT and unspliced NAT by oligo-specific semi-quantitative RT-PCR (**Figure R7B**). We detected all *LEF1* NAT species in HT29 M6 snail1 cells and not in HT29 M6 control cells.

The results of *LEF1* NAT expression in SW480 and HT29 M6 cells were not the expected, as we estimated that unspliced *LEF1* NAT and *LEF1* mRNA would have an inverse correlation like in RWP1 cells.

SW480 control cells also expressed more spliced NAT than SW480 E-cadherin. We wondered which of the two species of the *LEF1* NAT was more abundant for these cell lines where the spliced and unspliced NAT coexist. For that, we designed a semi-quantitative RT-PCR with 3 oligonucleotides (**Figure R8**). We used two reverse oligos [one retrotranscribing the spliced NAT (-8596) and the other retrotranscribing the unspliced NAT (-113)] and a forward oligo, common for both NATs (-11). For the retrotranscription process, the two reverse primers were added to the reaction. When the PCR cycles started, the common forward primer was added. In that way, both spliced and unspliced NAT competed for being amplified and we could observe the relative abundance between them.

In our positive control, equal amounts of DNA (pcDNA3 spliced NAT and pcDNA3 +243/-1856 NAT) were amplified. We observed that the smaller band, corresponding to the unspliced NAT, was more easily amplified. We used the oligos -11/-113/-8596, where the unspliced NAT product weights ~100bp and the spliced NAT ~250bp. (**Figure R8**).

We could observe that in HT29 M6 snail1 cells and SW480 control cells, the upper band, corresponding to the spliced NAT, was clearly more abundant (even that the efficiency of the oligos would favor the lower band). Thus, when the unspliced and spliced NAT

coexist in snail1 expressing cells, the spliced form of the NAT is clearly more abundant than the unspliced one.

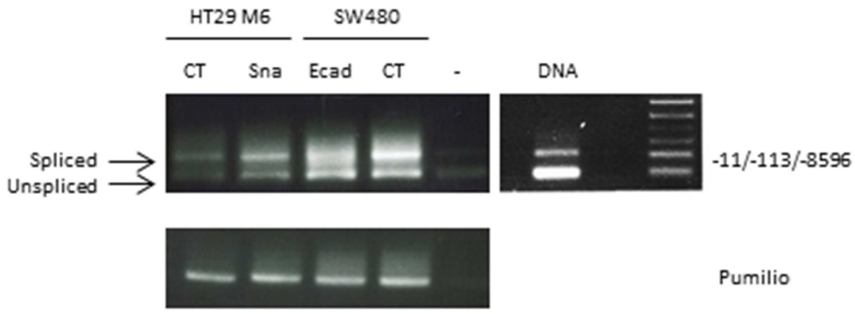


Figure R8. Spliced NAT is more abundant than unspliced NAT in HT29 M6 snail1 and SW480 control cells. RNA extracts from the indicated cell lines were analyzed. Semi-quantitative RT-PCRs with 3 oligos were performed as explained in the text (to detect spliced and unspliced NAT at the same time). In the RT-PCR -11/-113/-8596, the upper band corresponds to the spliced NAT and the lower to the unspliced NAT. Pumilio is shown as loading control.

3. *LEF1* NAT promoter is down-regulated during EMT

As shown in **Figure R3**, in RWP1 cells, total *LEF1* NAT levels were down-regulated during EMT. We analyzed the region preceding the NAT and that could be the element controlling its expression. We isolated the DNA fragment +66/+857 and inserted it into pGL3 vector in antisense direction in order to perform luciferase reporter assays.

As observed in **Figure R9**, this fragment presented promoter activity in RWP1 control cells, and it was down-regulated in cells with stable expression of snail1, correlatively with *LEF1* NAT expression. This promoter sequence is located at the first intron of the mRNA sequence, something usual in NATs promoters[125].

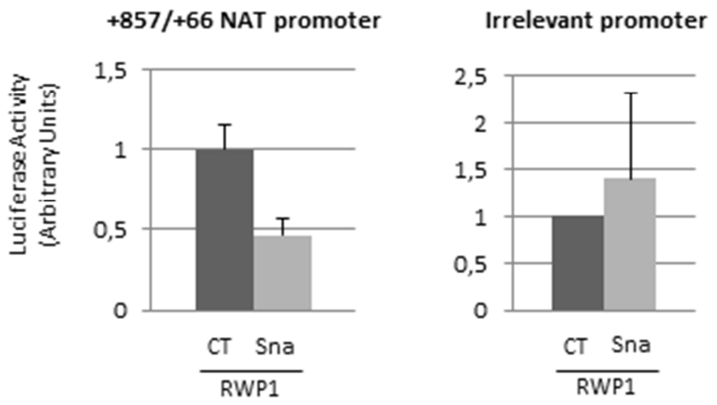


Figure R9. *LEF1* NAT promoter is down-regulated during EMT. pGL3 NAT promoter +857/+66 was transfected in RWP1 control and snail1 cells and luciferase activity was measured after 48 hours of transfection. Firefly Luciferase was standardized to the value of Renilla Luciferase. The *HES1* promoter was used as a negative control of a promoter not being altered by EMT. Results show the average ± standard desviation of three experiments performed in triplicate.

4. *LEF1* NAT represses *LEF1* promoter activity

We wanted to determine if *LEF1* NAT had any effect on *LEF1* mRNA. For that, we cloned the *LEF1* NAT fragment (+58/-1856) in pBabe vector and overexpressed unspliced *LEF1* NAT in RWP1 snail1 cells.

We first tested if unspliced *LEF1* NAT affected *LEF1* mRNA stability. For that, we treated RWP1 snail1 cells, with or without NAT, with Actinomycin D, an antibiotic that inhibits transcription. As shown in **Figure R10**, both cell lines degraded *LEF1* mRNA equally. We could conclude from this experiment, that *LEF1* NAT does not affect *LEF1* mRNA stability.

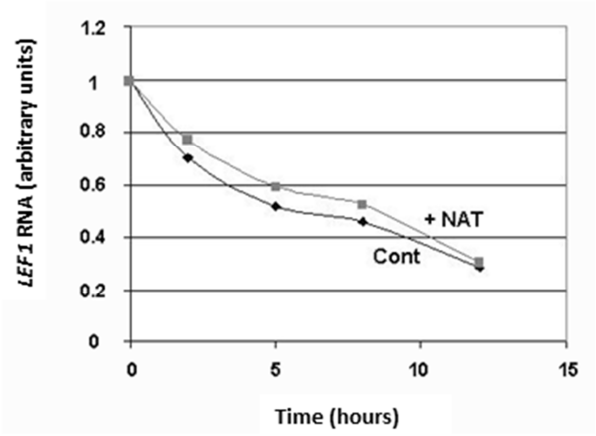


Figure R10. *LEF1* NAT does not affect *LEF1* mRNA stability. RWP1 snail1 cells were transfected with pBabe *LEF1* NAT (+58/-1856) or empty pBabe and selected with puromycin for 48 hours. Cells were supplemented with actinomycin D (2 µg/mL) to inhibit transcription. RNA was collected at the indicated times after actinomycin D addition and *LEF1* mRNA levels were determined by RT-qPCR. The relative amount of this RNA with respect to the initial time is shown. The results are representative of two experiments.

We thought that as the first part of the *LEF1* NAT overlapped with the 5'UTR and the promoter region of the *LEF1* gene, the NAT could be acting on the *LEF1* promoter (diagram in **Figure R11A**). We cloned *LEF1* promoter -1856/+58 into pGL3 plasmid to perform luciferase assays. *LEF1* promoter activity was increased in snail1 cells compared to the control in both cell lines tested (**Figure R11B**). When *LEF1* NAT (+58/-1856) was overexpressed, *LEF1* promoter activity in snail1 cells was inhibited.

On the other hand, spliced *LEF1* NAT was unable to do it (**Figure R12B**). When we overexpressed it in RWP1 snail1 cells, the luciferase activity of *LEF1* promoter did not change. On the contrary, spliced *LEF1* NAT stimulated *LEF1* promoter activity in RWP1 cells as shown in the left part of **Figure R12B**, suggesting that spliced *LEF1* NAT is "pro-mesenchymal".

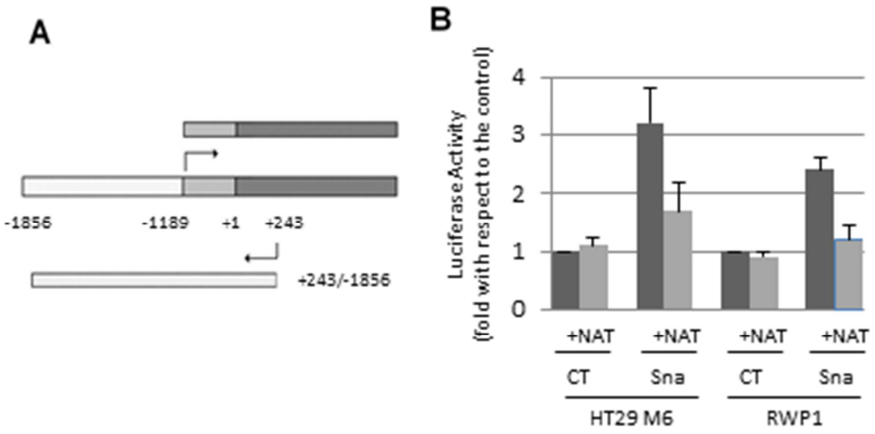


Figure R11. *LEF1* promoter is repressed by *LEF1* NAT. **A**. Diagram of *LEF1* locus where *LEF1* promoter and *LEF1* NAT overlapping can be observed. **B**. pGL3 *LEF1* promoter -1856/+58 was transfected in control and snail1 expressing cells transfected with pBabe NAT (+58/-1856) or empty pBabe. Luciferase activity was measured after 48 hours of transfection. Firefly Luciferase was standardized to the value of Renilla Luciferase. Here we show the average ± standard deviation of three experiments performed in triplicate.

But which part of the unspliced NAT is the most important for the repression of *LEF1* promoter activity? For answering that, we performed luciferase assays on RWP1 snail1 cells transfecting different fragments of the NAT (diagram **Figure R12A**) cloned in pcDNA3 vector. As shown in **Figure R12B**, both the +58/-1856 and +1/-1463 inhibited *LEF1* promoter activity to a similar extent than the +243/-1856 *LEF1* NAT. A +1/-879 NAT also significantly repressed the promoter whereas shorter fragments, with elimination in 5' or 3' sequences, were not active. Therefore, the +1/-1463 NAT contains all the elements required for *LEF1* promoter inhibition.

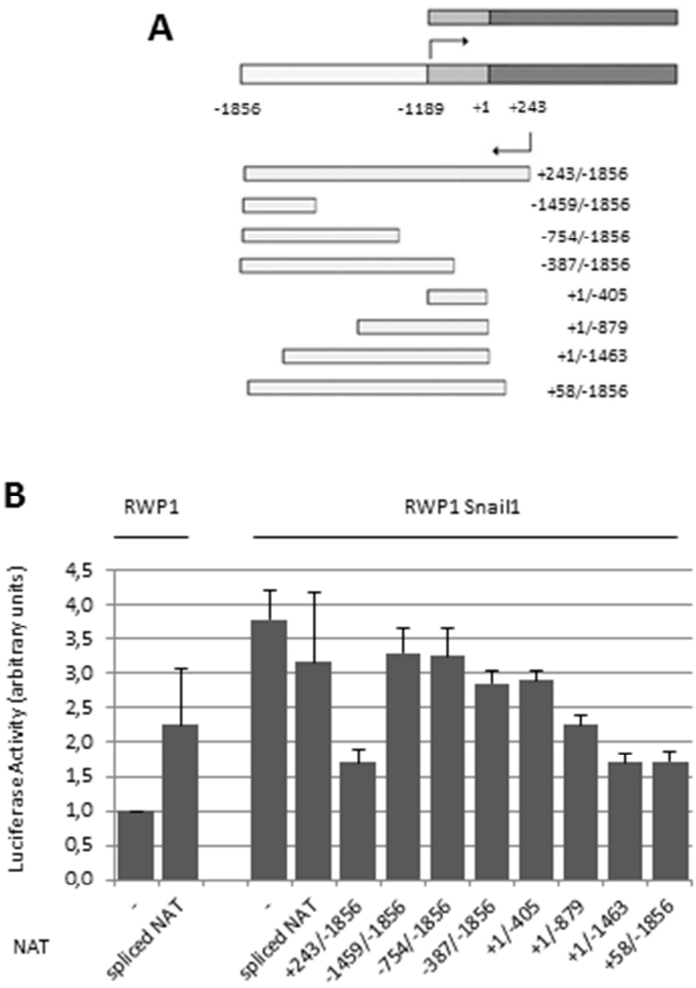


Figure R12. Spliced NAT does not inhibit *LEF1* promoter activity and +1/-1463 NAT contains all the elements required for *LEF1* promoter inhibition. **A**. Diagram of *LEF1* NAT fragments cloned in pcDNA3 used for inhibiting *LEF1* promoter activity. **B**. pGL3 *LEF1* promoter -1856/+58 was transfected in RWP1 and RWP1 snail1 expressing cells together with different constructs of pcDNA3 *LEF1* NAT (fragments of it). Luciferase activity was measured after 48 hours of transfection. Firefly Luciferase was standardized to the value of Renilla Luciferase. Here we show the average ± standard deviation of three experiments performed in triplicate.

5. *LEF1* NAT regulates LEF1 levels

We wondered if the repression of *LEF1* promoter by unspliced *LEF1* NAT was accompanied also with a regulation of LEF1 expression. For that, we overexpressed +58/-1856 *LEF1* NAT in RWP1 and HT29 M6 stably transfected with *snail1*. As observed in **Figure R13**, exogenous unspliced NAT expression repressed *LEF1* mRNA (**Figure R13A**).

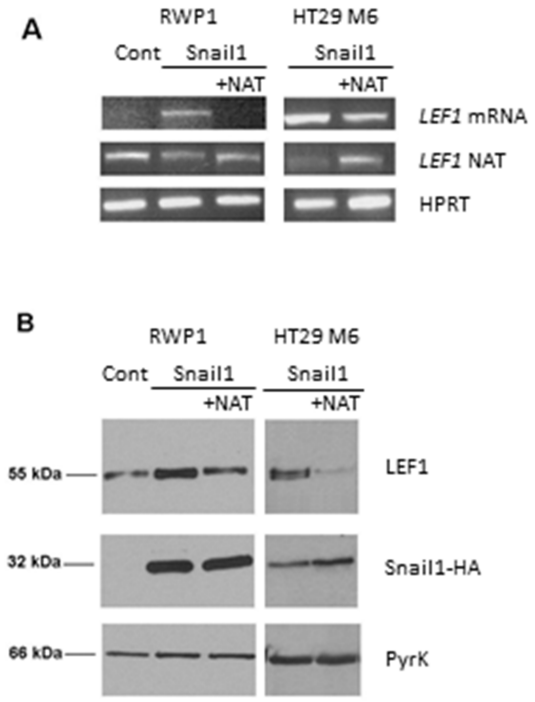
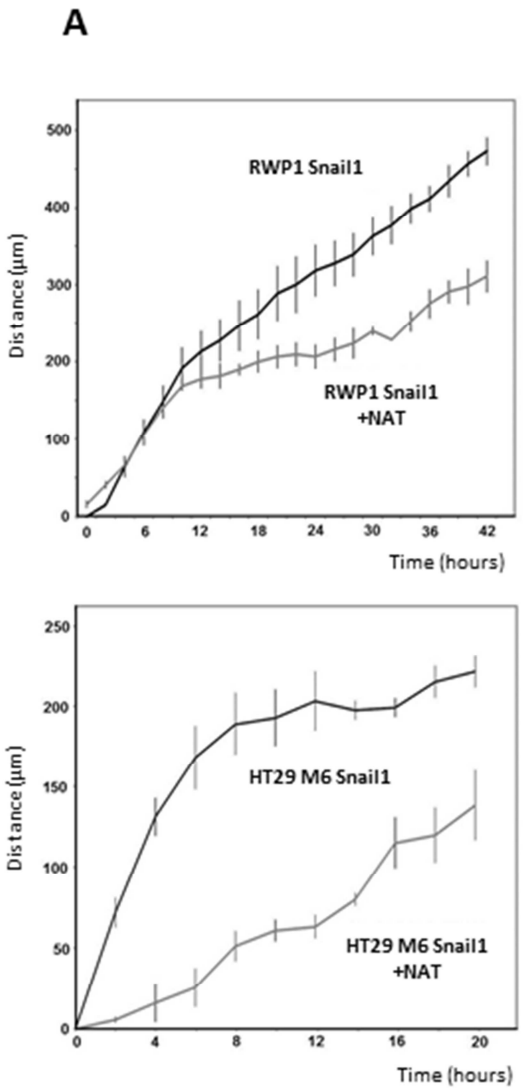


Figure R13. *LEF1* NAT down-regulates LEF1 expression. **A**. RNA was obtained from RWP1 control and snail1 stably transfected cells, from HT29 M6 snail1 cells and from the snail1 cells transfected with pBabe *LEF1* NAT (+58/-1856). *LEF1* mRNA and total *LEF1* NAT levels were analyzed by RT-PCR and HPRT was analyzed to normalize RNA levels. **B**. Total lysates from the same cells as in A were obtained and Western Blot was performed with LEF1 antibody. Anti-HA antibody was used to detect snail1 presence. Pyruvate kinase was analyzed as loading control.

Similar results were observed at the protein level in both cell lines tested (**Figure R13B**). The levels of this ectopic expression were comparable to the NAT levels detected in epithelial cells (without snail1) and they significantly decreased *LEF1* mRNA and protein levels.

We analyzed the relevance of this inhibition. For that, we performed migration assays with the same cells used in **Figure R13**. We seeded cells at high confluence and once the cells were attached, we performed a scratch. Images were taken each two hours and the changes of position of cell frontline were analyzed. As shown in **Figure R14A**, snail1 cells overexpressing exogenous NAT migrate less than snail1 cells. This difference was seen faster in HT29 M6 snail1 cells than in RWP1 snail1 cells. So we can say that unspliced *LEF1* NAT overexpression reduces cell migration.

We performed microarrays with the RNAs from RWP1 snail1 + pBabe empty and RWP1 snail1 + pBabe NAT in order to detect genes up-regulated and down-regulated by the unspliced *LEF1* NAT. The analysis were performed substracting the genes expressed in RWP1 snail1+empty from the ones expressed in RWP1 snail1+NAT (RWP1 snail1+NAT – RWP1 snail1+empty). In **Figure R14B** we show a list of genes up-regulated by *LEF1* NAT and that are implicated in reducing cell migration. We also show that *LEF1* mRNA levels were down-regulated when *LEF1* NAT was overexpressed, as expected.



B

genes downregulating cell migration

ID	Log Ratio
LEF1	**-0,646**
CD82	1,623
LAMC2	1,072
SERPINA3	1,056
TNC	1,003
TNFSF10	0,837
SPRY2	0,764
GZMB	0,708
TIMP3	0,601
TNFSF15	0,578
ACVRL1	0,554
PPARA	0,515
NDRG1	0,510
BMP6	0,504
CDH1	**0,440**

Figure R14. *LEF1* NAT down-regulates cell migration. **A**. Migration assays were performed with HT29 M6 snail1 and RWP1 snail1 cells transfected with pBabe *LEF1* NAT (+58/-1856) or an empty vector and selected with Puromycin. Cells were seeded at high confluence in 24-well plate. Once cells were attached, a scratch was performed. Three high definition phase contrast image of the same area were taken each two hours. Changes of position of cell frontline were analyzed in each image and represented as average of three different replicates respect to each time point. Error bars depict standard error of the mean. **B**. RNAs purified from RWP1 snail1+pBabe *LEF1* NAT (+58/-1856) and RWP1 snail1+pBabe empty cells were used to perform microarray analysis on Human

Gene 1.0 ST (Affymetrix®). Here we show a set of genes up-regulated by unspliced *LEF1* NAT and implicated in down-regulating cell migration. We also show that *LEF1* mRNA is down-regulated when unspliced *LEF1* NAT is overexpressed.

We then performed proliferation assays with the same cells than in **Figure R14**. Cells were seeded at low confluence and images were taken each two hours to calculate confluences. As shown in **Figure R15**, cells overexpressing snail1 do not change their proliferation when overexpressing the unspliced *LEF1* NAT.

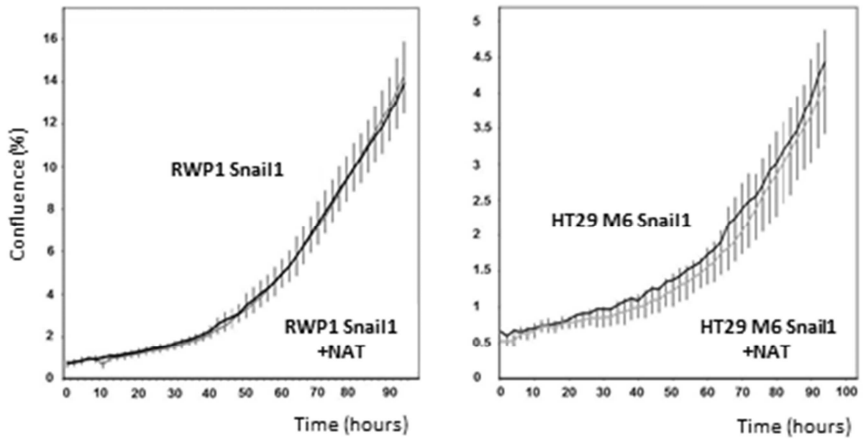


Figure R15. *LEF1* NAT does not change cell proliferation. Proliferation assays were performed with HT29 M6 snail1 and RWP1 snail1 cells transfected with pBabe *LEF1* NAT (+58/-1856) or an empty vector and selected with Puromycin. Cells were seeded at low confluence in 96-well plate. Three high definition phase contrast image of the same area were taken each two hours. Images were analyzed for calculating average confluence in each image and represented as average of three different replicates respect to each point. Error bars depict standard error of the mean.

As shown in **Figure 14B**, E-cadherin mRNA was up-regulated by *LEF1* NAT expression. For validating this up-regulation of the E-cadherin expression with unspliced *LEF1* NAT overexpression, we performed luciferase assays on *CDH1* promoter cloned in pGL3. We transfected it in RWP1 control or overexpressing snail1, and with

or without *LEF1* NAT. In the results shown in **Figure R16**, we can observe that snail1 repressed *CDH1* promoter as expected and that *LEF1* NAT slightly increased *CDH1* promoter activity. Thus, *LEF1* NAT increases *CDH1* promoter activity in RWP1 snail1 cells.

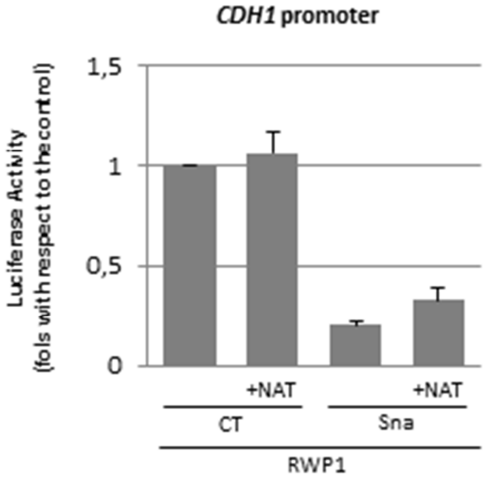


Figure R16. *LEF1* NAT increases *CDH1* promoter activity in RWP1 snail1 cells. pGL3 *CDH1* promoter was transfected in RWP1 control and snail1 cells, overexpressing pBabe *LEF1* NAT (+58/-1856) or empty pBabe. Luciferase activity was measured after 48 hours of transfection. Firefly Luciferase was standardized to the value of Renilla Luciferase. Results show the average ± standard desviation of three experiments performed in triplicate.

We had seen up-regulation of E-cadherin with *LEF1* NAT overexpression with promoter activity experiments and at mRNA level. We wondered if the E-cadherin up-regulation was also occurring at the protein level, so we performed Western Blot with E-cadherin antibody. As shown in **Figure R17**, E-cadherin protein is up-regulated when the *LEF1* NAT is overexpressed. Thus, unspliced NAT up-regulates E-cadherin expression both at mRNA and protein level. In **Figure R17** we could also observe that *LEF1* NAT also down-regulated LEF1 protein, as previously demonstrated.

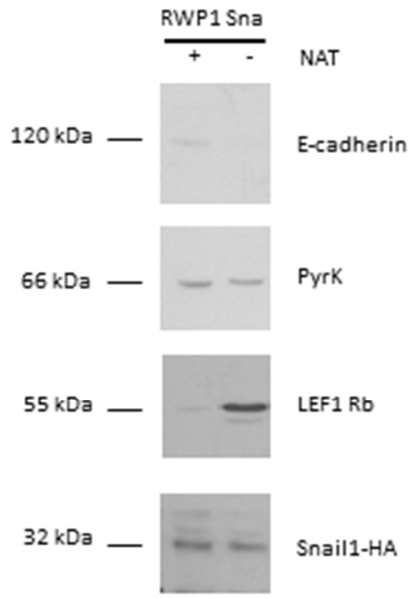


Figure R17. *LEF1* NAT up-regulates E-cadherin protein expression. Total protein extracts of RWP1 snail1+pBabe empty and RWP1 snail1+pBabe +58/-1856 *LEF1* NAT cells were used to perform Western Blot. Pyruvate kinase was used as loading control. Anti-HA antibody was used to detect the stable transfected snail1.

6. *LEF1* NAT associates to *LEF1* promoter

We performed Chromatin Isolation by RNA Purification (ChIRP) in order to test if *LEF1* NAT associated to *LEF1* promoter. Biotinylated *in vitro* synthesized *LEF1* NAT (+58/-1856) and pGL3 *LEF1* promoter were transfected to RWP1 cells. Cells were crosslinked with formaldehyde and NAT was precipitated from total extracts using an anti-biotin antibody. As shown in **Figure R18**, sequences corresponding to the proximal *LEF1* promoter (-1806/-1626, -1306/-1188 and -904/-703) were enriched in these complexes indicating that NAT binds to these elements. Very little interaction was detected with another amplicon corresponding to luciferase of pGL3 (+570/+744). The specificity of the NAT binding to *LEF1*

promoter was further demonstrated by the absence of binding detected with two irrelevant RNAs to *LEF1* promoter and the absence of binding of *LEF1* NAT to another different co-transfected promoter, *CDH1* promoter.

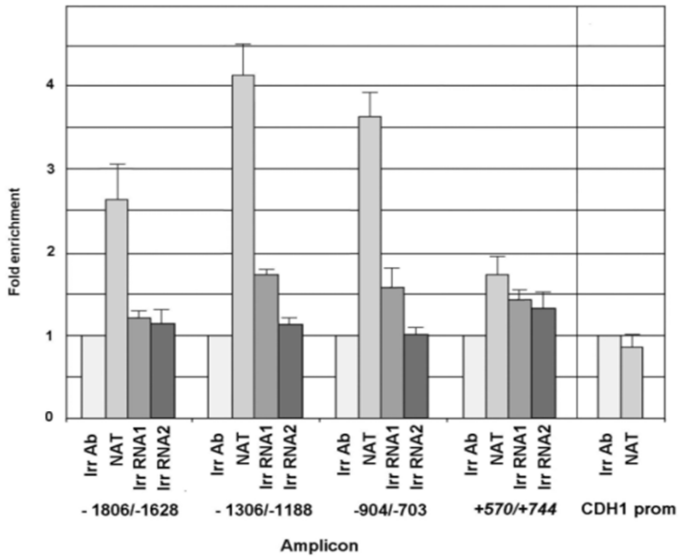


Figure R18. *LEF1* NAT binds to *LEF1* promoter. RWP1 cells were seeded in 150mm plates. After 24 hours, cells were transfected with 17.5µg pGL3 *LEF1* promoter (-1856/+58) plasmid and either 14µg *in vitro* synthesized biotinylated NAT (+58/-1856) or an irrelevant RNA, corresponding to Cre (IrrRNA1) and a fragment of pcDNA3 empty (IrrRNA2). *CDH1* promoter was alternatively transfected when indicated. After 24 hours, cells were fixed with formaldehyde as described for the ChIP assays and chromatins were prepared as in ChIP with RNAse out to do RNAse free chromatins. Extracts were incubated with an anti-biotin antibody and immunoprecipitated with protein A-agarose. Presence of the indicated amplicons was carried out by qPCR as described in Methods. Results show the average ± standard desviation of three experiments performed in triplicate.

We then performed the same experiment but using two fragments of the biotinylated NAT. pGL3 *LEF1* promoter was cotransfected with +1/-1463 or -387/-1856 *in vitro* synthesized biotinylated NAT to RWP1 cells. The rest of the experiment was performed as in

Figure R18. As shown in **Figure R19**, +1/-1463 NAT was able to bind to *LEF1* promoter whereas -387/-1856 NAT did not. This result is in accordance with **Figure R12**, where +/-1463 NAT was the one having the capability to inhibit *LEF1* promoter activity. Thus, the fragment +1/-1463 of *LEF1* NAT binds to *LEF1* promoter to inhibit it.

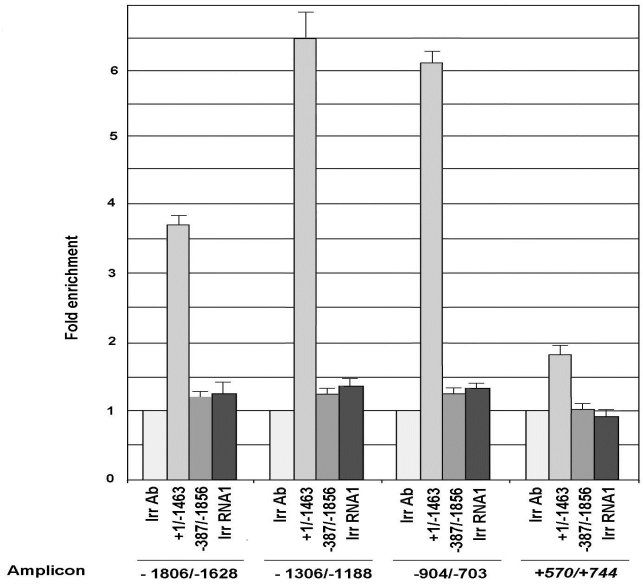


Figure R19. +1/-1463 *LEF1* NAT binds to *LEF1* promoter. RWP1 cells were seeded in 100mm plates. After 24 hours, cells were transfected with 5µg pGL3 *LEF1* promoter (-1856/+58) plasmid and either 6µg *in vitro* synthesized biotinylated NAT fragment (-1/-1463 or -387/-1856) or an irrelevant RNA, corresponding to a fragment of pcDNA3 empty. After 24 hours, cells were fixed with formaldehyde as described for the ChIP assays and chromatins were prepared as in ChIP with RNAse out to do RNAse free chromatins. Extracts were incubated with an anti-biotin antibody and immunoprecipitated with protein A-agarose. Presence of the indicated amplicons was carried out by qPCR as described in Methods. Results show the average ± standard desviation of three experiments performed in triplicate.

We also analyzed binding to the endogenous *LEF1* promoter; for that we just transfected the *in vitro* synthesized biotinylated NAT (+58/-1856) in RWP1 cells. Cells were crosslinked with

formaldehyde and NAT was precipitated from total extracts using an anti-biotin antibody. As shown in **Figure R20**, sequences corresponding to the proximal *LEF1* promoter were enriched indicating that *LEF1* NAT binds to them. Very little interaction was detected with another amplicon corresponding to a downstream region of *LEF1* sense (+3864/+4048). An irrelevant antibody was used in order to detect unspecific binding of biotin.

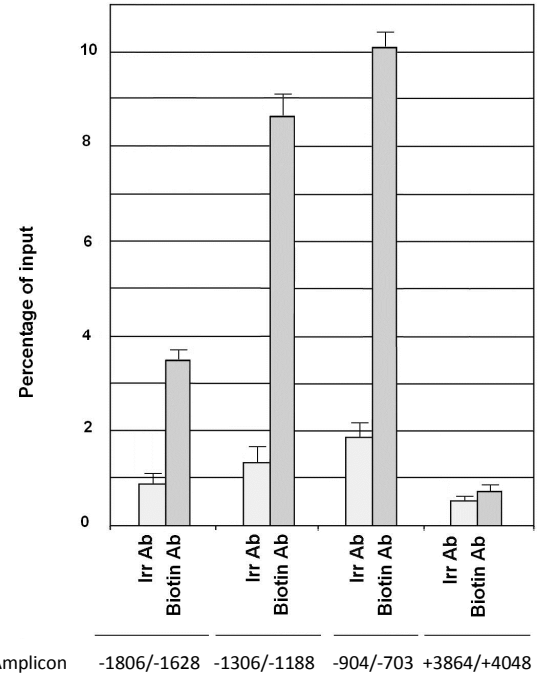


Figure R20. Endogenous *LEF1* NAT binds to *LEF1* promoter. RWP1 cells were seeded in 100mm plates. After 24 hours, cells were transfected with 9.5µg *in vitro* synthesized biotinylated NAT. After 24 hours, cells were fixed with formaldehyde as described for the ChIP assays and chromatins were prepared as in ChIP with RNAse out to do RNAse free chromatins. Extracts were incubated with an anti-biotin antibody or an irrelevant antibody and immunoprecipitated with protein A-agarose. Presence of the indicated amplicons was carried out by qPCR as described in Methods. Results show the average ± standard desviation of three experiments performed in triplicate.

As we had seen in **Figure R12** that spliced NAT does not inhibit *LEF1* promoter activity, we wondered if the spliced NAT affected unspliced NAT binding to *LEF1* promoter. For that, we synthesized *in vitro* spliced NAT (not biotinylated) and *in vitro* unspliced NAT (biotinylated). We also synthesized an *in vitro* irrelevant RNA (not biotinylated). We performed the experiment with endogenous and exogenous *LEF1* promoter, as indicated. Cells were crosslinked with formaldehyde and unspliced NAT was precipitated from total extracts using an anti-biotin antibody. As shown in **Figure R21**, *LEF1* promoter sequences binding to biotinylated unspliced NAT was markedly inhibited by spliced NAT but not by an irrelevant RNA. The same result was observed for endogenous *LEF1* promoter (**Figure R21A**) and exogenous *LEF1* promoter (**Figure R21B**). Thus, we can say that spliced NAT prevents unspliced *LEF1* NAT binding to *LEF1* promoter.

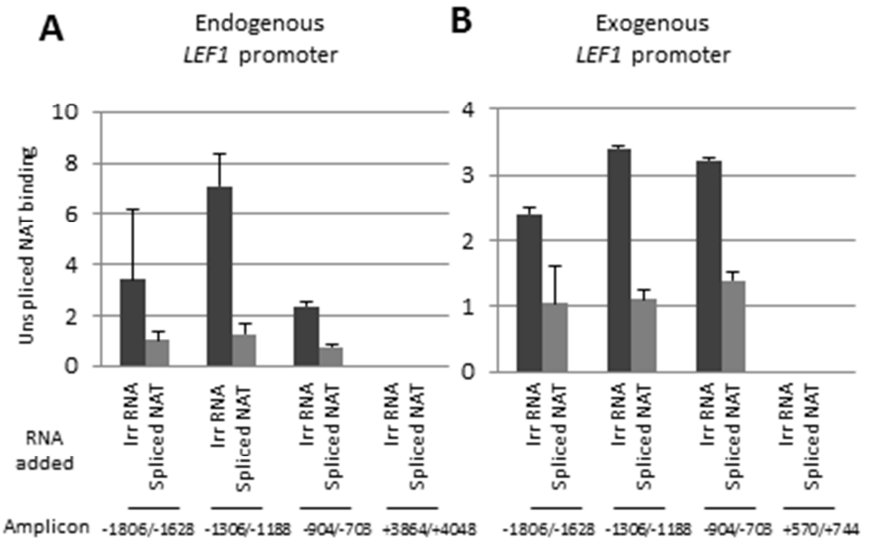


Figure R21. Spliced NAT prevents unspliced *LEF1* NAT binding to *LEF1* promoter. RWP1 cells were seeded in 100mm plates. After 24 hours, cells were transfected with 1µg pGL3 *LEF1* promoter (-1856/+58) plasmid in panel B, with 1.5µg *in vitro* synthesized biotinylated unspliced *LEF1* NAT and either 3µg *in vitro* synthesized not biotinylated spliced NAT or an irrelevant RNA, corresponding to a fragment of pcDNA3 empty. After 24 hours, cells were fixed with formaldehyde as described for the ChIP assays and chromatins were prepared as in ChIP with RNAse out to do RNAse free chromatins. Chromatins were sonicated with Biorruptor (Diagenode). Extracts were incubated with an anti-biotin antibody and immunoprecipitated with protein A-magnetic beads. Presence of the indicated amplicons bound to unspliced *LEF1* NAT was carried out by qPCR as described in Methods. Results show the average ± standard desviation of two experiments performed in duplicate.

7. *LEF1* NAT targets PRC2 to *LEF1* promoter

We then performed Chromatin Immunoprecipitation (ChIP) against the methylation marks present in histones since they reflect the activation or repression status of the promoter. We determined the presence of two marks in *LEF1* promoter; dimethylation of Lys4 of Histone 3 (H3K4me2) and trimethylation at Lys27 in Histone 3 (H3K27me3), which are associated with promoter activation or repression, respectively[126]. In these assays we amplified sequences corresponding to the *LEF1* promoter (-931/-750 amplicon), the NAT promoter (+266/+435 amplicon), or a control DNA (+3864/+4048). As seen in **Figure R22**, the H3K4me2 mark was present at the *LEF1* promoter to a much greater extent in RWP1 snail1 cells than in RWP1 cells, as *LEF1* promoter is active in mesenchymal cells. Conversely, H3K4me2 was detected at the NAT promoter to a greater degree in RWP1 than in RWP1 snail1 cells, correlating with a higher expression of the NAT in these cells. Expressing NAT in RWP1 snail1 cells reversed this pattern to a similar to that observed in RWP1 cells. Therefore, the presence of the NAT down-regulated H3K4me2 mark at the *LEF1* promoter and up-regulated it at the NAT promoter, indicating that the NAT activates its own promoter. As expected, no changes in methylation marks were found in amplicon +3864/+4048 that corresponds to control DNA.

The opposite results were obtained when the promoters were analyzed for H3K27me3. *LEF1* NAT overexpression increased the presence of this repressive mark at the *LEF1* promoter and decreased it at the NAT promoter (**Figure R23**). Since methylation of Lys27 is a consequence of the catalytic activity of the Polycomb Repressive Complex 2 (PRC2)[127], we investigated the binding of subunits of this complex to these two promoters. Both of the core subunits of this complex, named Suz12 and Ezh2, displayed a binding pattern similar to the pattern obtained with the H3K27me3 mark (**Figure R23**). Therefore, expression of NAT up-regulates the

association of Suz12 and Ezh2 to the *LEF1* promoter. In agreement with what we found for activation marks, H3K27me3 levels in *LEF1* NAT promoter (amplicon +266/+433) and also PRC2 binding were down-regulated in RWP1 snail1 cells when the *LEF1* NAT was overexpressed. No changes in methylation marks were observed in amplicon +3864/+4048.

Therefore, our results clearly show that *LEF1* NAT represses the *LEF1* promoter increasing the presence of the PRC2 on it and subsequently up-regulating the repressive mark H3K27me3.

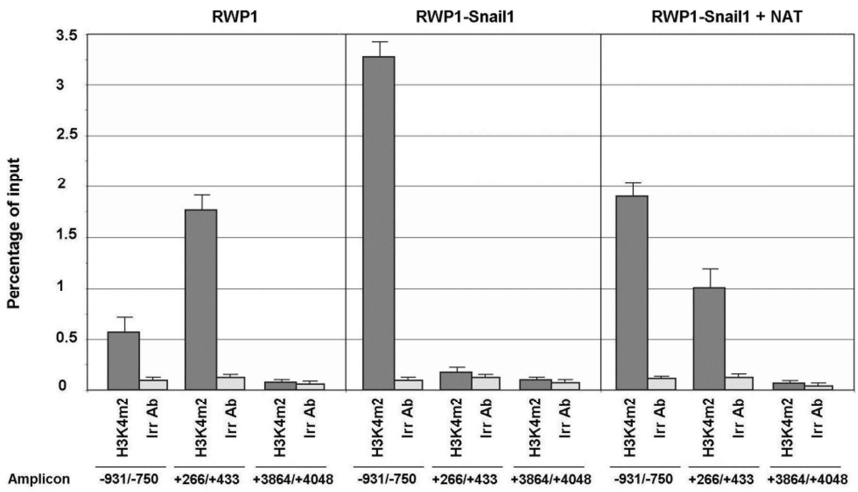


Figure R22. *LEF1* NAT decreases the activation mark H3K4me2 in *LEF1* promoter and increased it in NAT promoter. ChIP assays were carried out with chromatin obtained from RWP1, RWP1 snail1 and RWP1 snail1 cells transfected with pBabe *LEF1* NAT (+58/-1856). The immunoprecipitations were performed with antibody against H3K4me2 or with an irrelevant IgG. The presence of amplicons corresponding to the *LEF1* promoter (-931/-750), the NAT promoter (+266/+433) or an irrelevant DNA sequence (+3864/+4048) was determined. Results were quantified relative to the input amount. The results show the average ± standard deviation of three experiments performed in triplicate.

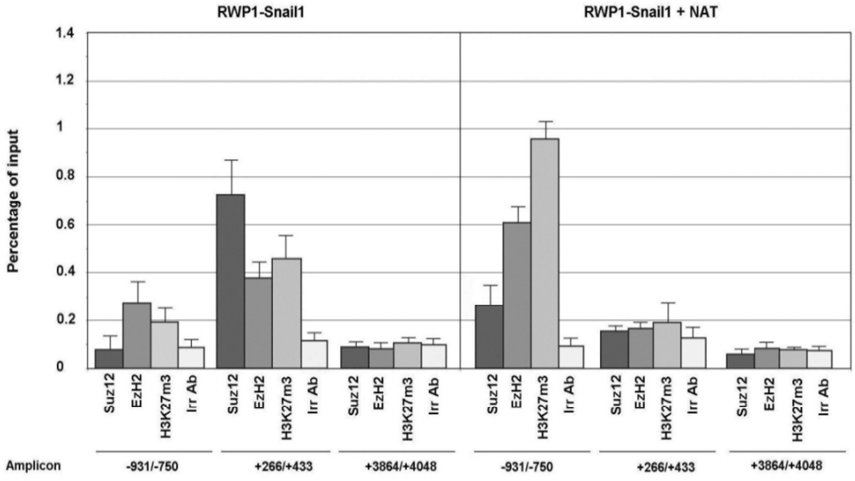


Figure R23. *LEF1* NAT increases the repressive mark H3K27me3 and PRC2 presence in *LEF1* promoter. ChIP assays were carried out with chromatin obtained from RWP1 snail1 and RWP1 snail1 cells transfected with pBabe *LEF1* NAT (+58/-1856). The immunoprecipitations were performed with the indicated antibodies and with an irrelevant IgG. The presence of amplicons corresponding to the *LEF1* promoter (-931/-750), the NAT promoter (+266/+433) or an irrelevant DNA sequence (+3864/+4048) was determined. Results were quantified relative to the input amount. The results show the average ± standard deviation of three experiments performed in triplicate.

We also checked whether the presence of the NAT can promote PRC2 binding to *LEF1* promoter using two *in vitro* assays. For this, we synthesized a biotinylated DNA corresponding to *LEF1* promoter (-1856/+58). A cell extract, prepared in RNAse free conditions from RWP1 snail1 cells transfected with empty pBabe or pBabe *LEF1* NAT plasmids, was incubated with biotinylated *LEF1* promoter or a fragment of the same size of the *Fibronectin 1* promoter as control. As shown in **Figure R24A**, binding of PRC2 component Suz12 was only detected to *LEF1* promoter and in NAT-expressing cells. Finally, a cell extract from RWP1 snail1 cells was incubated with biotinylated *LEF1* promoter in the presence of *in vitro* synthesized *LEF1* NAT or an irrelevant RNA. As shown in **Figure R24B**, PRC2 component Ezh2 interacted with the *LEF1*

promoter only when the NAT was present, indicating that the NAT facilitates the recruitment of PRC2 to this promoter.

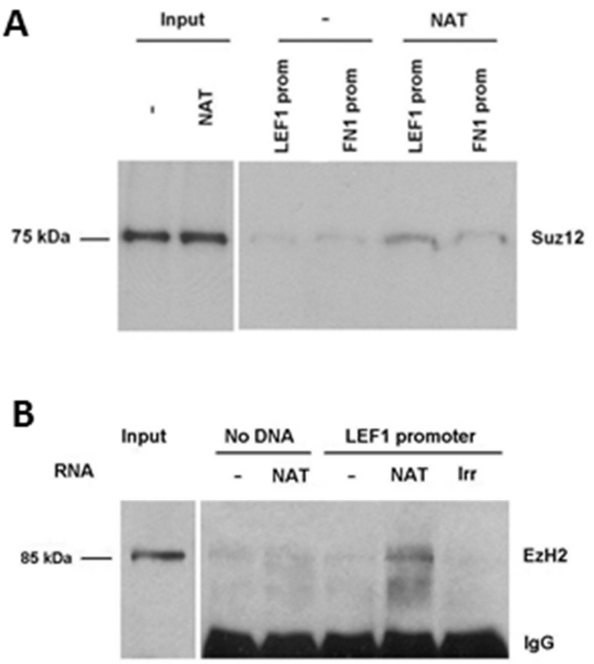


Figure R24. *LEF1* NAT is necessary for the recruitment of PRC2 in *LEF1* promoter. Biotinylated Oligo Pull-down Assays (BOPA) using biotinylated *LEF1* promoter (-1856/+58) were performed. **A**. RWP1 snail1 cells were transfected with empty pBabe or pBabe *LEF1* NAT (+58/-1856). Cells were lysated in RNAse free conditions with Polysomal Lysis Buffer. Cell extracts were incubated with biotinylated *LEF1* promoter or a fragment of the same size of *FN1* promoter. Immunoprecipitation was performed with anti-biotin antibody, the pull-down with protein G-agarose, and the presence of Suz12 in the precipitate was analyzed by Western Blot. **B**. Cell extracts from RWP1 snail1 cells were incubated with biotinylated *LEF1* promoter. When indicated, *in vitro* synthesized NAT (+58/-1856) or an irrelevant RNA (YBX1) was added. Samples were incubated with an anti-biotin antibody, biotinylated probes were pulled-down with protein G-agarose, and the presence of Ezh2 in the precipitate was analyzed by Western Blot.

8. *LEF1* NAT binds to the PRC2 complex

It has been described that PRC2 can directly bind to RNAs[42,63]. Therefore, we checked its putative interaction with *LEF1* NAT using RNA Immunoprecipitation assays (RIP) coupled to RT-PCR. Total extracts of RWP1 snail1 transfected with *LEF1* NAT (+58/-1856) were incubated with anti-Suz12 antibody. After immunoprecipitation, total RNA was extracted and semi-quantitative RT-PCR was performed to detect transcripts coimmunoprecipitated with Suz12. As shown in **Figure R25A**, *LEF1* NAT was effectively coimmunoprecipitated with anti-Suz12 but not by an irrelevant control antibody.

The assay was repeated using quantitative PCR and pretreating the immunocomplexes with different nucleases in order to elucidate which nuclease structure is needed for *LEF1* NAT binding to PRC2 (**Figure R25B**).

When samples were not treated with nucleases, both Ezh2 and Suz12 antibodies immunoprecipitated *LEF1* NAT. DNAse Turbo hydrolyses double strand DNA (dsDNA); when added, PRC2 was still able to immunoprecipitate *LEF1* NAT, indicating that *LEF1* NAT binds to PRC2 independently of the presence of *LEF1* promoter. In order to determine if PRC2 binding required DNA-RNA heteroduplexes, as that generated in the Transcriptional Start Sites, RNAse H was added as it hydrolyses the RNA in these hybrids. RNAse H did not affect PRC2-*LEF1* NAT immunoprecipitation, indicating that the NAT-PRC2 binding does not require DNA-RNA heteroduplex. We also checked if dsRNA structures were needed for the *LEF1* NAT-PRC2 interaction. For that we used RNAse VI, which hydrolyses dsRNA. This treatment affected the binding, indicating that NAT forms secondary structures with itself or it needs another RNA in order to interact

with PRC2. Finally, RNAse A1, that hydrolyses ssRNA, was used as a negative control, as *LEF1* NAT is hydrolyzed in this condition.

Therefore, these RIP assays reveled that *LEF1* NAT binding to PRC2 is independent of any DNA and that a secondary structure of dsRNA is needed for this binding.

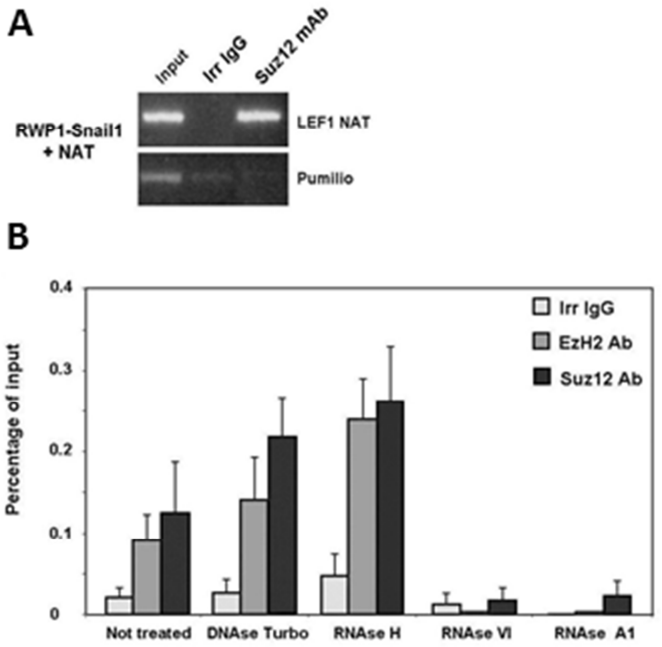


Figure R25. *LEF1* NAT binds to PRC2 and needs a dsRNA structure. A. RNA Immunoprecipitation (RIP) with Suz12 antibody was performed in RWP1 snail1 cells transfected with pBabe *LEF1* NAT (+58/-1856). From the RNA extracted, *LEF1* NAT was amplified by oligo-specific RT-PCR with oligos corresponding to +213/+60 sequence. Pumilio primers were used to detect unspecific binding. **B**. RIP with Suz12 and Ezh2 antibodies was performed in RWP1 snail1 cells transfected with pBabe *LEF1* NAT (+58/-1856). Treatment with DNAse and RNAses was performed when indicated. *LEF1* NAT binding was analyzed by RT-qPCR. RT was performed with *LEF1* NAT antisense oligo corresponding to +60/+80 sequence. qPCR was performed with oligos corresponding to *LEF1* NAT sequence +213/+60. Results were quantified relative to the input amount. Results show the average ± standard deviation of three experiments performed in triplicate.

We also looked to *LEF1* NAT-PRC2 interaction *in vitro*. For that, biotinylated-oligo pull-down (BOPA) assays were also performed using *in vitro* synthesized biotinylated *LEF1* NAT. RNAse free cell extracts were incubated with biotinylated *LEF1* NAT (+58/-1856). No RNA or a biotinylated irrelevant RNA were used as negative controls. We immunoprecipitated with anti-biotin antibody, pulled down with protein G-agarose and performed Western Blot against PRC2 subunits. The PRC2 component Suz12 copurified with *LEF1* NAT (+58/-1856) when it was incubated with cell extracts (**Figure R26A**). The presence of the *LEF1* promoter did not up-regulate Suz12-NAT binding, supporting our finding that DNA is not required for this association.

We next mapped the element in *LEF1* NAT required for PRC2 binding. The NAT +243/+1 sequence was not required for the inhibition of *LEF1* promoter, suggesting that this segment did not contain the binding element. RNA-BOPA assays confirmed this conclusion since Ezh2 coprecipitated with the +1/-1463 segment of *LEF1* NAT (**Figure R26B**). Progressive deletion of elements downstream in the NAT affected the interaction, since Ezh2 did not bind as efficiently to the +1/-879 and +1/-405 fragments as it did to +1/-1463. No association was observed between PRC2 and -754/-1856 or -387/-1856 sequences. Therefore, we conclude that binding requires an element located between +1 and -405, although other sequences situated downstream are also necessary to maximize the interaction.

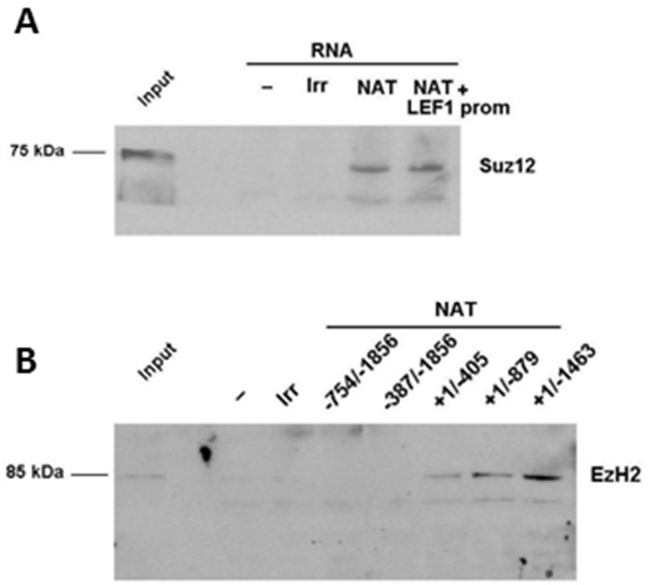


Figure R26. *LEF1* NAT binds to PRC2 by +1/-405 NAT sequence. RNA BOPA assays were performed. HT29 M6 cells were lysated in RNAse free conditions with Polysomal Lysis Buffer. **A**. Cell extracts were incubated with 6 µg of *in vitro* synthesized biotinylated *LEF1* NAT, biotinylated irrelevant RNA or with no RNA. pGL3 *LEF1* promoter was added in the reaction when indicated. Immunoprecipitation was performed with anti-biotin antibody, the pull-down with protein G-agarose, and the presence of Suz12 in the precipitate was analyzed by Western Blot. **B**. Cell extracts were incubated with *in vitro* synthesized biotinylated fragments of the *LEF1* NAT, as indicated. Immunoprecipitation was performed with anti-biotin antibody, the pull-down with protein G-agarose, and the presence of Ezh2 in the precipitate was analyzed by Western Blot.

DISCUSSION

Binding of *LEF1* NAT to Polycomb Repressive Complex 2

In recent years, the function of long non-coding RNAs has started to be unveiled. The broad functional repertoire of these RNAs includes roles in high-order chromosomal dynamics, telomere biology and subcellular structural organization[2,128]. One major theme emerging is the involvement of these ncRNAs in regulating the transcription of neighbor protein-coding genes[7]. In this thesis we have characterized a Natural Antisense Transcript that depending on the splicing process, it inhibits its partner expression.

The almost 9kb-long *LEF1* NAT is expressed in epithelial cells such as RWP1 using a promoter located in the first intron of the *LEF1* gene, a feature common to most NATs. It overlaps the 5'UTR region of the *LEF1* mRNA and it is not extensively polyadenylated. We did not expect polyadenylation in *LEF1* NAT as it would not give a significant advantage to it because it is not translated and is not exported to the cytoplasm since it makes its function in the nucleus.

Although *LEF1* NAT extends to almost 9kb, we have determined that the functional action of *LEF1* NAT is confined to the first 2kb; ectopic expression of this transcript down-regulates *LEF1* mRNA and protein.

As we have described, *LEF1* NAT decreases the activity of the *LEF1* promoter by binding to this element and promoting the recruitment of the Polycomb Repressive Complex 2 complex (PRC2). Recent reports have demonstrated that non-coding RNAs can direct PRC2 binding to promoters. This is the case for HOTAIR[31,43], XIST[42], Kcnq1ot1[40] and Gtl2[55]. Perhaps the most prominent example is silencing of the inactive X-chromosome by the ncRNA XIST. To normalize the copy number of X-chromosome

between male and female cells, XIST RNA from one of the two female X-chromosome recruits PRC2 to trimethylate H3 at lysine 27 (H3K27me3), rendering the chromosome transcriptionally silent[57]. More specifically, a 1.6kb ncRNA (RepA) within XIST is responsible for the PRC2 interaction, with Ezh2 serving as the RNA-binding subunit[42]. Another example is HOTAIR, a long intergenic RNA transcribed from the *HOXC* cluster that represses genes in the *HOXD* cluster by binding to the PRC2 complex[31]. ChIRP-seq results indicate that HOTAIR also recruits PRC2 to many other genes, suggesting that it participates in the mechanism of repression by this complex[43]. Moreover, HOTAIR also interacts with LSD1 (lysine-specific demethylase 1)[129], thus assembling PRC2 with the LSD1/CoREST/REST complex, and maybe with other proteins associated with LSD1, such as the Snail1 transcriptional repressor[130].

As explained, *LEF1* NAT is an example of another long non-coding RNA capable of interacting with PRC2. The fact that more than the 90% of unspliced *LEF1* NAT is localized in the nucleus is in accordance to that because it is in the nucleus where it performs its action. Furthermore, the binding element to PRC2 in *LEF1* NAT has been investigated in this work using luciferase assays with NAT fragments, ChIRP assays to identify the NAT region binding to *LEF1* promoter, and RNA-BOPA assays for localizing the NAT sequence that binds to PRC2. According to our results, optimal interaction requires a long sequence corresponding to the +1/-1463, although a significant binding was detected with +1/-405 element. Therefore, it is possible that several elements scattered through this sequence are required for high affinity binding.

Our results also suggest that PRC2 binds to a double strand RNA structure in the NAT. It is possible that this corresponds to a highly structured region in the NAT. However, the double strand RNA might be generated through the interaction of the NAT and short

fragments of the sense RNA. Recent studies carried out in embryonic stem cells have revealed that PRC2 is also physically associated with a short non-coding *LEF1* RNA[60]. This non-coding RNA, shorter than 200bp, was detected with a probe corresponding to -341/-402 and therefore belongs to the 5'-UTR of *LEF1* mRNA. The complementary sequence of this short ncRNA is included within the region of the *LEF1* NAT involved in PRC2 binding (-1/-405). We have not been able to detect this short sense RNA in an experiment similar to ChIRP. It could be that the experiment conditions should be optimized, that the PCR product amplified is too short (-403/-353) or that in RWP1 cell line this short sense RNA is not present as in embryonic stem cells is.

Another option for the double strand RNA structure needed for the interaction between NAT and PRC2 might be the requirement of *LEF1* mRNA. In this case, we propose that *LEF1* mRNA would be expressed in low levels, less than *LEF1* NAT, but sufficient to be part of this complex. In cells where LEF1 protein is not expressed (this is the case for RWP1 cells), low levels of *LEF1* mRNA in the nucleus would bind to *LEF1* NAT and form the double strand RNA structure required for PRC2 binding to *LEF1* NAT. As the short *LEF1* RNA -403/-353 is contained in *LEF1* mRNA, this hypothesis could be only possible if the experiment conditions of the variation of the ChIRP were not the appropriate, making impossible the detection of the sense *LEF1* RNA.

The interaction between ncRNAs and the PRC2 subunits might be alternatively due through a double RNA hairpin[42]. We performed bioinformatics analysis in *LEF1* NAT with *mFold*[131] looking for structures similar to those present in RepA, which are necessary to interact with PRC2. We identified a possible binding sequence doing a double hairpin in -195/-225. However, more experiments have to be performed in order to experimentally demonstrate this

result. One possibility would be to mutate this region in order to try to avoid PRC2 binding to the mutated RNA.

As our results show, no DNA is needed for NAT-PRC2 interaction. RNA-BOPA assays show that the presence of *LEF1* promoter does not increase NAT-PRC2 interaction and RIP assays with nucleases show that the removal of DNA does not affect to NAT-PRC2 binding. These results suggest that first occurs the binding of PRC2 to *LEF1* NAT and then this complex binds to *LEF1* promoter. As in other cases occurs (HOTAIR[31,43], XIST[42], Kcnq1ot1[40] or Gtl2[55]), the lncRNA serve as a guide for PRC2 to localize in the proper place where to perform its action.

What is still unclear is whether the binding between *LEF1* NAT and PRC2 is direct or indirect. Our *in vitro* experiments are all performed with cell extracts and not with purified protein; so we cannot discard that other proteins would be needed for the binding of PRC2 to *LEF1* NAT. The involvement of Argonaute (AGO) proteins in this complex could be a possibility. Corey's group have observed that Argonaute 2 (AGO2) binds to duplexes of RNA and promotes trimethylation of Lysine 27 in Histone 3 in the corresponding DNA sequence[132]. They studied the *Progesterone Receptor* locus, more than 100kb long, and they detected the same pattern of AGO2 binding in the 3'UTR of the gene than to the promoter, suggesting that the gene would be folded approximating the two regions (**Figure D1**). Gene looping has also been observed at the X-inactivation center[133]. We could test if, in RWP1 cells, RIP with anti-AGO2 immunoprecipitates *LEF1* NAT, in order to see if this critical component in the RNAi pathway for the miRNA biogenesis is involved in our model.

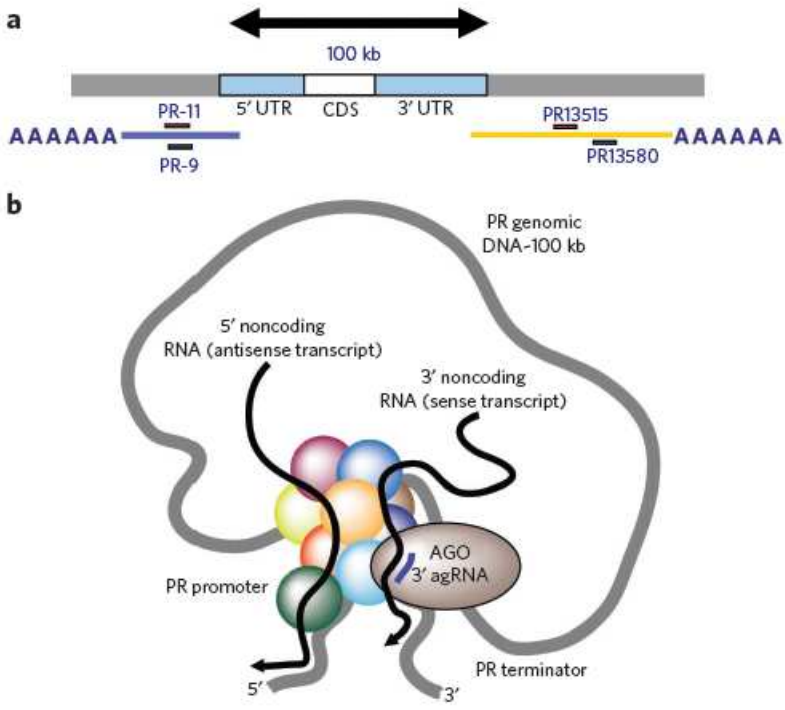


Figure D1. Model for modulation of transcription by AGO in *Progesterone* gene[132] **A**. 100.000 bases separate the genomic locations of the promoter and 3' terminal regions of the *PR* gene. **B**. Gene looping juxtaposes the 5' promoter and 3' terminator, bringing DNA sequences into close proximity. A small RNA (3' agRNA) recruits AGO2 to the 3' non-coding transcript. The arrival of AGO2 may affect other proteins (here shown as unlabeled spheres) at the gene promoter and alter regulation of transcription. The proximity of 3' and 5' non-coding transcripts allows them to co-immunoprecipitate during RIP with anti-AGO antibodies. CDS, coding sequence.

It is also still unclear if the binding between *LEF1* NAT and *LEF1* promoter is mediated by intermediaries or is direct. In the case of XIST RNA, it has been described that YY1 protein is needed for the RNA interaction with the chromatin[134]. YY1 is a "bivalent" protein, capable of binding both RNA and DNA through different sequence motifs. Specific YY1-to-RNA and YY1-to-DNA contacts are required to load XIST particles onto X-chromosome. The authors propose

that YY1 acts as an adaptor between regulatory RNA and chromatin targets. In the case of *LEF1*, it is tempting to think that complementarity of sequences between *LEF1* NAT and *LEF1* promoter might be enough for the binding between them. Matching of the sequences could give the specificity of this binding.

Which subunit of PRC2 is interacting with lncRNAs is still unclear. In 2008 and also in 2010, Lee's group proposed that Ezh2 served as the RNA-binding subunit of PRC2 to the double hairpin of RepA[42] and also to many RNA-protein interactions[55]. In 2010, Jenner's group[60] proposed that Suz12 interacts most strongly with both XIST-RepA and short RNA stem-loops. Both possibilities could be occurring in our model, so further experiments should be performed to elucidate which subunit of PRC2 is interacting with *LEF1* NAT.

It has to be mentioned that a controversial paper has been recently published by Cech and colleagues where they affirmed that the binding of RNA to PRC2 is promiscuous[135]. They say that the binding of human PRC2 to various RNAs is the same than to irrelevant transcripts from ciliates and bacteria. They also demonstrate that PRC2 binding is size dependent, with lower affinity for shorter RNAs. *In vivo*, they observe PRC2 predominantly on repressed genes, but also PRC2 is associated with active genes that are not regulated by PRC2. Based on our results, *LEF1* NAT-to-PRC2 binding would not be unspecific for several reasons. *LEF1* NAT binding to PRC2 is not size dependent as demonstrated in the RNA-BOPA of **Figure R26**, where a longer RNA (-387/-1856) does not bind to PRC2 whereas shorter ones (+1/-405 and +1/-879) do. Moreover, in the ChIPs for PRC2 of **Figure R23**, PRC2 subunits only bind to repressed promoters and not to the active ones. Thus, PRC2 binding to *LEF1* locus is specific for repressing promoter regions. In RIP and BOPA assays we also have demonstrated that

PRC2 does not bind to irrelevant RNAs such Pumilio and others. These evidences suggest that binding between *LEF1* NAT and PRC2 is more specific than other RNA-to-PRC2 interactions.

Regulation of *LEF1* NAT expression

LEF1 NAT overlaps a zone corresponding to the first exon, the 5'UTR and the main promoter of *LEF1* gene. We have determined that the region +857/+66 has promoter activity and is active in RWP1 cells. We wondered how this promoter region is regulated. Snail1, as a repressive transcriptional factor, could be inhibiting *LEF1* NAT promoter activity. Although canonical E-boxes were not located in this promoter, ChIP experiments were performed in order to detect snail1 bound to the NAT promoter in RWP1 snail1 cells. Results were negative. However, NAT promoter contains binding sequences for several transcription factors such as Twist1 or Wilms Tumor 1 protein. We expect that Twist1 would bind to *LEF1* NAT promoter in snail1 expressing cells and repress it, and that Wilms Tumor 1 would bind *LEF1* NAT promoter in RWP1 cells and activate it. Further experiments should be performed to determine if some of these proteins control *LEF1* NAT expression.

A ChIP-Sequencing for Snail1 in SW620 cells (a human colon adenocarcinoma cell line with high levels of Snail1) was performed in our lab by Alba Millanes-Romero and Sandra Peiró. Two Snail1 binding sites on the *LEF1* locus were found in this ChIP-Seq: -519/-481 and -1159/-981. Both regions correspond to the *LEF1* NAT transcript, not to the promoter region, confirming that absence of Snail1 in *LEF1* NAT promoter obtained in our ChIP. It is possible that these two regions where Snail1 binds could be acting negatively, repressing *LEF1* NAT promoter.

An additional possibility would be that Snail1 is regulating splicing of *LEF1* NAT. It has been described that transcription factors can regulate splicing and that methylation marks change in splicing regions[136,137]. In particular, Snail1 has also been seen to favor splicing during EMT by repressing Epithelial Splicing Regulatory Protein 1 (ESRP1)[138]. Both binding regions in ChIP-Seq for Snail1 in *LEF1* NAT correspond to the first intron, so it is tempting to speculate that the presence of Snail1 in the first NAT intron would modify the presence of splicing factors on it increasing splicing of this intron. This would explain that for all three cell lines tested, in the presence of Snail1, spliced NAT is the most predominant form. Thus, further experiments should be performed to better understand what Snail1 is doing in the DNA region corresponding to the first intron of *LEF1* NAT.

Another point to be discussed referred to *LEF1* NAT regulation is the fact that when we overexpress the 2kb *LEF1* NAT (+58/-1856), the epigenetic methylation marks in *LEF1* NAT promoter change. As we have showed in the ChIPs of methylation marks, *LEF1* NAT up-regulates H3K4me2 on its own promoter and removes PRC2 and H3K27me3 from it. It could be that a mutual repression between *LEF1* mRNA and *LEF1* NAT exists, where in RWP1 snail1 cells, *LEF1* mRNA is transcribed and inhibits *LEF1* NAT promoter. When *LEF1* NAT is overexpressed, it inhibits *LEF1* mRNA transcription so *LEF1* mRNA cannot repress *LEF1* NAT promoter, leading to a positive feedback.

Relevance of *LEF1* NAT

In this thesis we have widely demonstrated that *LEF1* NAT can regulate *LEF1* gene expression. It is also interesting that *LEF1* NAT can regulate also other genes as shown in the microarray, such as *CDH1*. E-cadherin is a key molecule in Epithelial-to-Mesenchymal

Transition (EMT) and its expression affects the migration and invasion capacity of tumor cells. We propose the following mechanism for E-cadherin regulation: *LEF1* NAT down-regulates LEF1 protein levels decreasing β-catenin-LEF1 complexes. Zeb1 is a direct target of β-catenin complexes[139]; thus, Zeb1 protein levels are reduced and *CDH1* is derepressed. The ability of the *LEF1* NAT to up-regulate E-cadherin expression at RNA and protein level gives to this lncRNA a very interesting potential for cancer treatment. From the cited microarray, more information could be extracted, suggesting other important genes implicated in cancer that are affected by *LEF1* NAT.

As explained in the beginning of this thesis, other possible NATs from genes implicated in EMT were analyzed by RT-PCR. It would be very interesting to go into detail with them and see if other NATs can regulate EMT genes like *LEF1* and *Zeb2*[113] NATs do. LncRNAs are being involved in many processes and crucial functions are assigned to them. Maybe regulating EMT is also one of these big challenges to them.

It would be very interesting to examine what happen in other cellular models; if the expression pattern of *LEF1* NATs is like in RWP1 cells or like in HT29 M6 and SW480. *In vivo* experiments could also be performed to give more relevance to the work. We would like to determine if *LEF1* mRNA and both *LEF1* NATs are present in human tumor samples where different levels of E-cadherin and Snail1 are present. On the one hand, we would expect that unspliced *LEF1* NAT would be found in the low-Snail1-expressing tumors, with high levels of E-cadherin. On the other hand, spliced NAT and *LEF1* mRNA would be expected to have a similar pattern and be found in tumors with high expression of Snail1 and low expression of E-cadherin.

As discussed in this thesis, one of the best-described functions of NATs and other lncRNAs is the regulation of chromatin states[37].

Specifically, many can regulate negatively their sense (protein-coding) partners. Strategies for therapeutic manipulation are arising and targeting the NATs is a promising strategy for those that have oncogenic functions. In our case, it would be interesting to target spliced *LEF1* NAT or *LEF1* mRNA. It would be also possible to overexpress unspliced *LEF1* NAT as a therapeutic strategy. To specifically inhibit the functions of NATs, single-stranded oligonucleotides can be designed to strand-specifically block the interaction of the antisense transcript with the sense gene mRNA or with the corresponding DNA.

This approach was originally introduced in 2005 by Wahlestedt and collegues[61,140] and has since been demonstrated by various investigators using different systems. Oligonucleotides that are designed to inhibit NAT functions in this manner have been named "antagoNATs"[141]. As with other single-stranded therapeutic oligonucleotides, a number of beneficial characteristics should ideally be built into an antagoNAT to enable its potential *in vivo* applications. These include introducing chemical modifications to promote metabolic stability and minimizing the length of the oligonucleotide to aid cellular uptake, while maintaining selectivity to minimize off-target activities.

As previously mentioned, in our case it would be interesting to down-regulate spliced *LEF1* NAT. If we targeted spliced NAT with an antagoNAT, it would not block unspliced NAT binding to *LEF1* promoter, permitting the inhibition of *LEF1* promoter by unspliced *LEF1* NAT (**Figure D2**).

There are many advantages of targeting long non-coding RNAs. By utilizing strategies that are not applicable to protein targets, we can hope to modulate disease pathways that have previously been considered to be intractable. Also, by targeting a regulatory ncRNA it may be possible to change the expression of an endogenous gene in a natural manner. The fact that NATs act in *cis* and gene-

locus specific is a clear advantage to gain specificity. Finally, a possible advantage of targeting lncRNAs is their relatively low abundance compared to mRNA. It would be easier to down-regulate their expression.

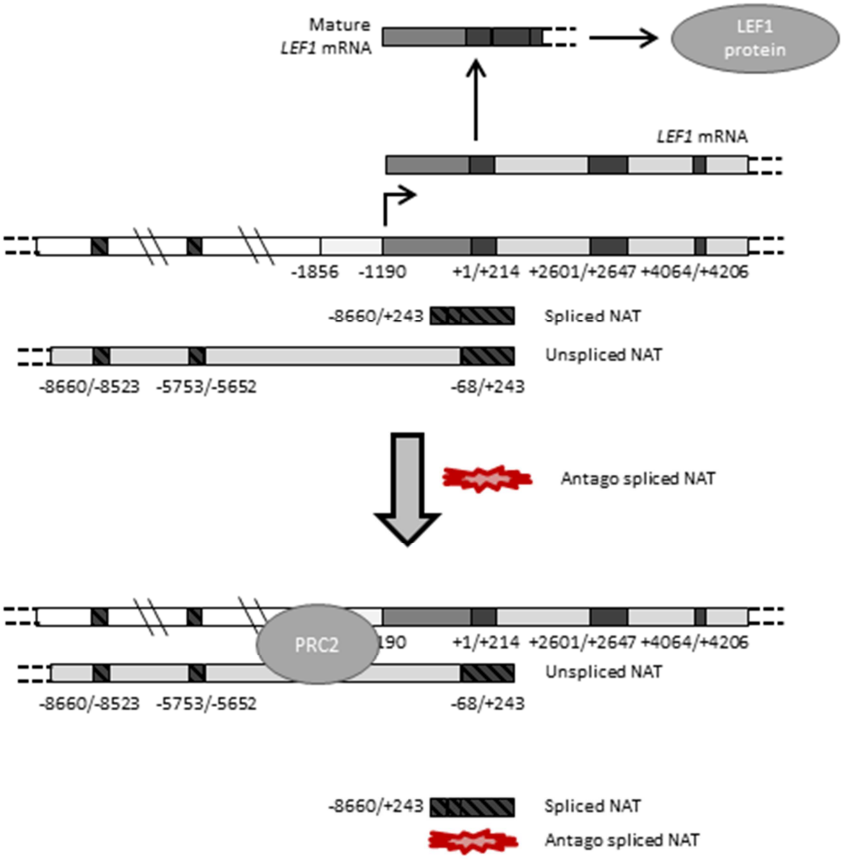


Figure D2. Hypothetical model of targeting spliced NAT with antagoNAT. As observed in the upper part of the figure, spliced NAT prevents the binding between unspliced NAT and *LEF1* promoter. Thus, *LEF1* promoter is active and *LEF1* mRNA and protein are produced. As shown in the lower part of the figure, if we introduced an antagoNAT for the spliced NAT (it could be targeting the sequence between exon A and C), spliced NAT would be blocked or degraded. Therefore, it would not block the binding between unspliced NAT and *LEF1* promoter and PRC2 would be recruited there for repressing it.

Promising results are being obtained when targeting ncRNAs[141,142], so we encourage future studies on cancer treatment to try to down-regulate spliced *LEF1* NAT and *LEF1* mRNA, or to introduce systemically unspliced *LEF1* NAT in order to block LEF1 protein expression and also E-cadherin expression.

Current model of regulation of *LEF1* mRNA by *LEF1* NAT

In this work, we have analyzed the complex locus of *LEF1* gene and the transcripts that it encodes. As explained in this thesis, different species of a Natural Antisense Transcript regulate *LEF1* mRNA expression. The splicing of this Natural Antisense Transcript is crucial for the regulation of *LEF1* mRNA expression, as the different species of NATs have different functions. The situation is also variable regarding the phenotype of the cell lines.

When *LEF1* NAT is transcribed and does not suffer splicing, this long non-coding RNA binds to Polycomb Repressive Complex 2 (PRC2), responsible for the epigenetic repressive mark H3K27me3. *LEF1* NAT associates to PRC2 and recruits it to *LEF1* promoter, inhibiting it. In consequence, *LEF1* mRNA is not transcribed; so LEF1 protein is not produced.

When *LEF1* NAT undergoes splicing, the low amount of unspliced NAT is not sufficient to inhibit *LEF1* mRNA promoter. Spliced NAT act as a dominant negative of the unspliced NAT as it can bind to *LEF1* promoter but it cannot inhibit it as it lacks the necessary sequence for that. As consequence, *LEF1* mRNA promoter is active; *LEF1* mRNA is transcribed and translated into LEF1 protein.

We have observed three different situations. In epithelial cells, such as HT29 M6 Control and SW480 E-cadherin cells, both *LEF1*

mRNA and *LEF1* NAT promoters are silenced and no transcripts are produced in this locus in these cell lines (**Figure D3**).

Figure D3. Proposed model for epithelial cells. In epithelial cells, no transcripts are synthesized from *LEF1* locus as both promoters are silenced.

In metastable epithelial cells (epithelial cells that have initiated EMT and have an intermediate phenotype), such as RWP1 control cells, *LEF1* NAT is transcribed and does not suffer splicing. This long non-coding RNA binds to Polycomb Repressive Complex 2 (PRC2) and recruits it to *LEF1* promoter, inhibiting it (**Figure D4**). In consequence, *LEF1* mRNA is not transcribed.

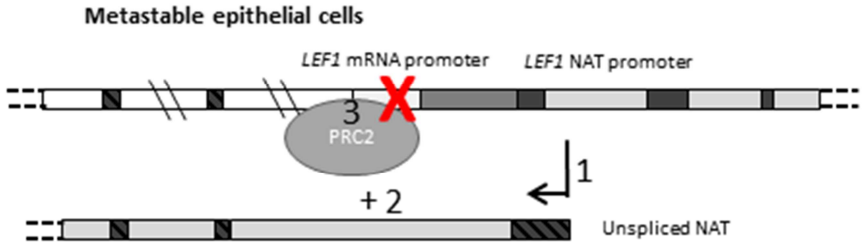


Figure D4. Proposed model for metastable epithelial cells. In epithelial cells with an intermediate phenotype, unspliced NAT is produced; it binds to PRC2 and recruits it to *LEF1* promoter, inhibiting it.

Finally, in mesenchymal and snail1 expressing cells; such as in HT29 M6 snail1, SW480 Control cells and RWP1 snail1; both *LEF1* mRNA and *LEF1* NAT promoters are active. *LEF1* NAT is transcribed, but most of it undergoes splicing. The low amount of unspliced NAT is not sufficient to inhibit *LEF1* mRNA promoter, and the spliced NAT act as a dominant negative of the unspliced NAT, allowing the activation of *LEF1* mRNA promoter in these cell lines. *LEF1* mRNA is

transcribed, and translated into *LEF1* full-length protein (**Figure D5**).

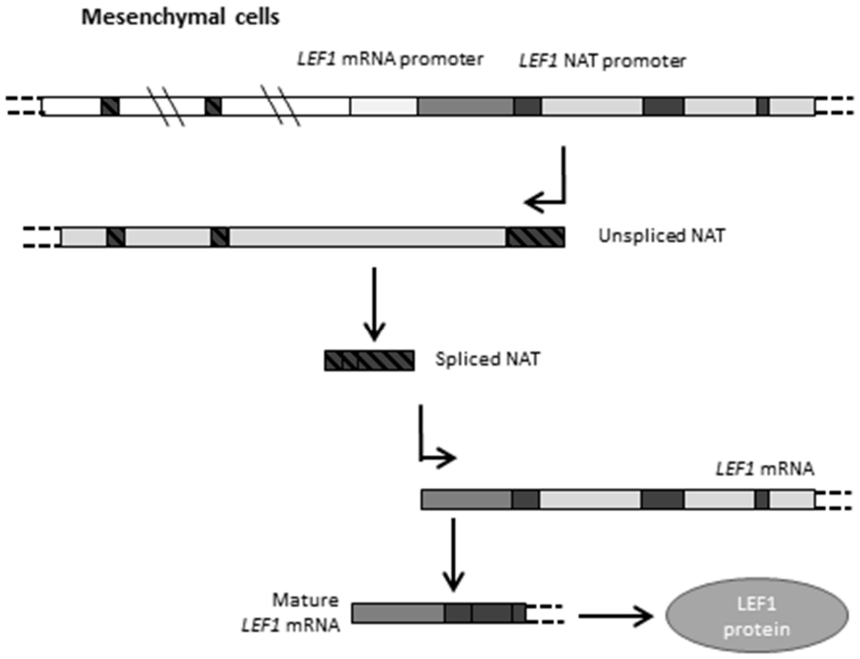


Figure D5. Proposed model for mesenchymal cells. In mesenchymal cells, more spliced NAT than unspliced NAT is produced, enabling *LEF1* mRNA to be synthesized and LEF1 protein is translated.

As we have observed, spliced NAT is "pro-mesenchymal". In **Figure R12**, when spliced NAT is transfected in epithelial RWP1 cells, *LEF1* promoter activity is increased. We have suggested that spliced NAT prevents unspliced NAT binding to *LEF1* promoter acting as a dominant negative transcript. Spliced and unspliced NAT share the first part of the sequence (+243/-68), so probably both can bind to this region in *LEF1* DNA locus because of the complementarity of sequences. However, only the unspliced NAT contains the sequence necessary for the binding to PCR2 as seen in **Figure R26** and necessary for the LEF1 promoter repression as seen in **Figure R12** (fragment +1/-1463). In snail1 expressing cells, the abundance

of more spliced NAT in comparison to unspliced NAT, would make that the spliced NAT preferentially binds to *LEF1* DNA locus and does not permit to PRC2 to bind to this region.

CONCLUSIONS

Conclusions

1. A *LEF1* Natural Antisense Transcript is expressed in *LEF1* locus, starting from the first intron of *LEF1* mRNA, and it undergoes alternative splicing.

2. Transcription of *LEF1* NAT changes during EMT. In epithelial cells, there is no expression of *LEF1* NAT. Unspliced *LEF1* NAT is expressed in metastable epithelial cells whereas spliced *LEF1* NAT is present in mesenchymal cells.

3. Unspliced *LEF1* NAT down-regulates *LEF1* promoter activity whereas spliced NAT does not. It also down-regulates LEF1 expression both at mRNA and protein level.

4. Unprocessed *LEF1* NAT decreases cell migration. It up-regulates E-cadherin expression both at mRNA and protein level.

5. Unspliced *LEF1* NAT binds to *LEF1* promoter. Sequence +1/-1463 is needed for this binding and also contains all the elements required for *LEF1* promoter inhibition. Spliced *LEF1* NAT prevents the binding between unspliced *LEF1* NAT and *LEF1* promoter, blocking *LEF1* promoter inhibition.

6. Unprocessed *LEF1* NAT recruits Polycomb Repressive Complex 2 (PRC2) to *LEF1* promoter to inhibit it. It changes methylation marks of *LEF1* promoter and of its own promoter. In *LEF1* promoter, it decreases the activation mark H3K4me2 and increases the repressive mark H3K27me3. In *LEF1* NAT promoter, it increases H3K4me2 and decreases H3K27me3.

7. Binding between PRC2 and unspliced *LEF1* NAT requires a double strand RNA structure. *LEF1* NAT binds to PRC2 by sequence +1/-405 but sequence +1/-1463 is necessary to maximize the interaction.

MATERIALS AND METHODS

1. Cell Culture

All cells were grown in Dulbecco's modified Eagle's medium (DMEM, Invitrogen), which were grown in Roswell Park Memorial Institute (RPMI) 1640 medium (Invitrogen). Media were supplemented with glucose 4.5 g/L (Life Technologies), glutamine 2 mM, penicillin 100 U/mL, streptomycin 100 µg/mL and 10% fetal bovine serum (FBS, GIBCO). Cells were maintained at 37ºC in a humid atmosphere containing 5% CO_2 and 95% air.

1.1. RWP1

This cell line comes from a human pancreatic carcinoma and present epithelial morphology and well-formed intercellular junctions.

Stable RWP1 mmsnail1-hemagglutinin (HA) clones were generated in our laboratory[143]. Stable expression of snail1 was maintained with the antibiotic G418 (500 µg/mL) (Gibco). Expression of HA tag was confirmed by Western Blot.

RWP1 snail1 cells were transfected with pBabe *LEF1* NAT using Lipofectamine Plus kit (Gibco) according to the manufacturer's instructions. Transfected cells were selected in medium containing 500 µg/mL G418 (Gibco) and 2.5 ng/mL Puromycin (Gibco).

Transitory transfections were performed on RWP1 and RWP1 snail1 cells using Lipofectamine Plus kit (Gibco) according to the manufacturer's instructions.

1.2. HT29 M6

This cell line comes from human colon adenocarcinoma. It presents epithelial morphology, high E-cadherin levels and it grows forming compact colonies. HT29 M6 cells are a subpopulation of HT29 cells selected with methotrexate 10^{-6}M that present mucosecretor phenotype.

Stable HT29 M6 clones for mmsnail1-HA were generated[144] and maintained in our laboratory. Stable expression of snail1 was conserved with the addition of 500 µg/mL G418 (Gibco) to the culture medium.

Generation of HT29 M6 snail1 *LEF1* NAT stable expression was performed transfecting pBabe *LEF1* NAT using Lipofectamine Plus reagents (Gibco) according to manufacturer instructions. Transfected cells were grown in medium supplemented with 500 µg/mL G418 (Gibco) and 1.5 ng/mL Puromycin (Gibco).

1.3. SW480

This cell line comes from human colon adenocarcinoma and it has intermediate morphology. At low confluence presents low E-cadherin levels, phenotype reverted at high confluence. This cell line has a truncated APC, therefore it has the β-catenin/TCF pathway highly active.

SW480 cells were kindly provided by Dr. Alberto Muñoz (Instituto de Investigaciones Biomédicas "Alberto Sols", Consejo Superior de Investigaciones Científicas – Universidad Autónoma de Madrid, Madrid, Spain). For the generation of E-cadherin transfectant clones, SW480-ADH (adherent) cells were transfected with *CDH1* cDNA in the eukaryotic vector pBATEM2 (kindly provided by M. Takeichi, Kyoto University, Kyoto, Japan)[145]. Stable transfectants

were obtained after selection with 2 mg/mL G418 and screened by Western Blot and immunofluorescence. The clones with higher E-cadherin expression were selected.

2. Plasmid construction

Unless otherwise specified, a general cloning protocol was used to perform the constructions detailed here:

i. **Production of the linear insert** (a linear piece of the desired DNA sequence was obtained either from PCR using *Pfx* Platinium polymerase (Invitrogen) from genomic DNA of HT29 M6 or RWP1 cells lines.
ii. **Cutting the insert** (if needed) **and target vector with appropriate restriction enzymes**
iii. **Ligating the linearized vector and insert together**
iv. **Transformation of the completed vector and screening** (in the DH5α strain of *E.coli*)

All linearized vectors were dephospholylated using calf intestine phosphatase (CIP) from New England Biolabs (NEB) for one hour at 37ºC and ligation was performed with T4 ligase (NEB) o/n on melting ice. Positive DNAs were confirmed by sequencing. For PCR fragments cloned blunt, phosphorylation was performed with T4 kinase and *Forward* buffer (Invitrogen) for one hour at 37ªC. If fill-in was needed, *Klenow fragment* (NEB) was used, fifteen minutes at 37ºC. When needed, DNA was purified either from solution or agarose gel (0.5, 1 or 2% depending on the size of the DNA) using the GFX PCR DNA and Gel Band Purification kit (GE Healthcare). All constructs were verified by enzyme digestion and sequencing (Big Dye from Life Technologies).

LEF1 constructs

Sequence of *LEF1* locus is attached in **Annex 1** (*LEF1* locus -2000/+1000). All numeration is referred to the Translational Start Site of the full-length protein (position +1. NP_057353.1). Oligonucleotides used for cloning are shown in **Table M1**.

Construct	*LEF1* NAT fragment +enzymes	Oligonucleotide (5'-3')	
pGL3 *LEF1* NAT promoter	+857/+66 HindIII	S	CCCAAGCTTGCTCTCCGGAGGAGGTGGCGG
		AS	CCCAAGCTTGGCCTGATCCGGTCCCAAACCC
pGL3 *LEF1* promoter	-1856/+58 KpnI/SmaI	S	ATGCGGTACCCTTGTCTCCAAAGAGCG
		AS	TGGCGCAGAGTTCCGG
pBabe *LEF1* NAT	+58/-1856 EcoRI	S	TGGCGCAGAGTTCCGG
		AS	ATGCGGTACCCTTGTCTCCAAAGAGCG
pcDNA3 *LEF1* NAT	+58/-1856 EcoRV	S	TGGCGCAGAGTTCCGG
		AS	ATGCGGTACCCTTGTCTCCAAAGAGCG
	+243/-1856 HindIII/KpnI	S	ACCAAAGCTTCTCTCGGAACTGGGGCAG
		AS	ATGCGGTACCCTTGTCTCCAAAGAGCG
pcDNA3 spliced NAT	+213/-8660 KpnI/XhoI	S	CATTGGTACCCTCGTGTCCGTTGCTGGC
		AS	GTTTCTCGAGGATGCTCTAGGTTGGTTTATTTAC
pcDNA3 NAT fragments	-1459/-1856 EcoRV	S	CCACCGAAAGGCACGC
		AS	ATGCGGTACCCTTGTCTCCAAAGAGCG
	-754/-1856 EcoRV	S	GATCTCGAGTTGCCAAGAATAAAGTTTTTGCC
		AS	ATGCGGTACCCTTGTCTCCAAAGAGCG
	-387/-1856 EcoRV	S	GGTTCTGGAGGGAGCACG
		AS	ATGCGGTACCCTTGTCTCCAAAGAGCG
	+1/-405 HindIII/EcoRV	S	TTTAAGCTTCCCGGCGGCTCTG
		AS	CGTGCTCCCTCCAGAACC
	+1/-879 HindIII/EcoRV	S	TTTAAGCTTCCCGGCGGCTCTG
		AS	CTTACGCGTCCGGGCAGAGGCATTT
	+1/-1463 EcoRV	S	TTTAAGCTTCCCGGCGGCTCTG
		AS	GTGGGTTATAAGCAGCCCCG

Table M1. Oligonucleotides used for cloning *LEF1* constructs. S: Sense, AS: Antisense

- ***LEF1* NAT promoter +857/+66** was cloned using oligonucleotides corresponding to the sequences +841/+857 and +66/+82 both provided with a HindIII sites. The fragment, after digestion with HindIII, was inserted in pGL3[109], previously digested with HindIII. Correct direction of the insert was checked by sequencing.

- ***LEF1* promoter -1856/+58** was cloned using oligonucleotides corresponding to the sequences -1837/-1856 and +37/+58 that contained restriction sites for KpnI and SmaI enzymes, respectively. The fragment was inserted in pGL3, in KpnI/SmaI sites.

- The expression plasmid **pBabe *LEF1* NAT** was obtained by inserting the **+58/-1856** amplification product in the EcoRI site of pBabe-Puro (Addgene) (previously re-filled to have blunt ends). Correct direction of the insert was checked by sequencing.

- The **pcDNA3 *LEF1* NAT +58/-1856** vector was obtained with +58/-1856 amplification product in the EcoRV site in pcDNA3 vector. Correct direction of the insert was checked by sequencing.

- The **pcDNA3 *LEF1* NAT +243/-1856** vector was cloned using oligonucleotides corresponding to sequences +243/+225 and -1837/-1856 that contained restriction sites for HindIII and KpnI, respectively. The fragment was inserted in pcDNA3, in HindIII/KpnI sites.

- The expression plasmid **pcDNA3 spliced NAT +213/-8660** was cloned from RNA of SW480 control cells. Oligonucleotides for the sequences +213/+195 with KpnI site and -8660/-8636 with XhoI site were used of semi-quantitative RT-PCR. Different length products were obtained and the shorter one was used as

it was the more abundant when detected by RT-PCRs. The product was 517 bp long, corresponding to +213/-68, -5652/-5753 and -8523/-8660 (checked by sequencing). The insert and pcDNA3 vector were digested with KpnI and XhoI for cloning.

- The different NAT deletions were constructed using oligonucleotides corresponding to the sequences -18/+1, -369/-387, -754/-769, -1439/-1459 as sense oligos; and -1837/-1856, -1463/-1445, -879/-859, -405/-382 as reverse oligos. Amplification products were obtained using high-fidelity *Pfx* Platinium polymerase (Invitrogen), isolated and inserted in a pcDNA3 vector digested with EcoRV. In the case of constructs +1/-405 and +1/-879, inserts where digested with HindIII and pCDNA3 vector, opened with EcoRV and HindIII.

Other vectors

pGL3 *CDH1* promoter and pcDNA3 snail1 vectors were cloned as described in Batlle et al. (2000)[105]. pGL2 *HES1* promoter vector was obtained from Alain Israel lab[146].

TK Renilla (pRL-TK, Promega, E2241) and SV40 Renilla (pRL-SV40, Promega, E2231) were used as internal controls of transfection in luciferase assays.

Solutions

Luria-Bertani (LB): 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl

LB-agar: LB, 1.5% agar (w/v)

Antibiotics: 0.05 mg/mL ampicillin

TAE: 40 mM Tris pH 7.6, 9.4 mM acetic acid, 1mM EDTA

Sample buffer for DNA (10x): 16.7 mM Tris-HCl pH 7.5, 83.3 mM EDTA, 16.7 Ficoll 400, 0.6 % orange green

3. Transcript analysis

RNA was extracted with Trizol™ method and treated with Turbo DNAse-*free*™ kit (Ambion) (DNAse treatment was for at least 30 minutes). RNA was quantified using Thermo Scientific NanoDrop™ 1000 Spectophotometer.

For semi-quantitative analysis 250 ng to 1 µg of RNA were used in reactions performed with One Step RT-PCR kit (Qiagen) and oligonucleotides corresponding to sequences described in **Table M2**. Expression levels of transcripts were also calculated by RT-qPCR. Retrotranscription was performed with Transcriptor First Strand cDNA Synthesis Kit (Roche) and quantitative PCR with SYBR Green (Roche Diagnostics). RNA levels were determined quantitatively in triplicate using the indicated primers (**Table M2**) on a LightCycler®480 Real-Time PCR System (Roche). All quantifications were normalized to an endogenous control (Hipoxanthine-guanine phosphoribosyl transferase, HPRT; or Pumilio). The relative quantification value for each target gene compared with the calibrator for that target is expressed as 2^(Ct-Cc) (Ct and Cc are the mean threshold cycle differences after normalizing to HPRT or Pumilio). Products from both semi-quantitative RT-PCRs and RT-qPCRs were loaded in 2% agarose gels to check that fragment size was that expected and the identity of the amplified fragments was verified by sequencing.

Figure	Gene	Position in *LEF1* locus	Oligonucleotide (5'-3')	
1	*CDH1* NAT		S	CACACGCTGACCTCTAAGG
			AS	GGTCGGGAGTTTGGGAC
	PTEN NAT		S	GGGGGCTTTAACTGTAGTATTTGG
			AS	GGGATGAGGCATTATCCTGTACAC
	FN1 NAT		S	CCACTATTTACTATTCTGAATGTCAC
			AS	GAATGTTAATTGCCCAATTGAG
	Twist NAT		S	GCAGGAGAGAGTAAGAGATGAAGAG
			AS	GCATGCATTCTCAAGAGGTC

Figure	Gene	Position in *LEF1* locus	Oligonucleotide (5'-3')	
	HPRT		S	GGCCAGACTTTGTTGGATTTG
			AS	TGCGCTCATCTTAGGCTTTGT
	Pumilio		S	CGGTCGTCCTGAGGATAAAA
			AS	CGTACGTGAGGCGTGAGTAA
3,6,7	*CDH1* mRNA		S	GAACGCATTGCCACATACAC
			AS	ATTCGGGCTTGTTGTCATTC
3,6,7, 10,13	*LEF1* mRNA	+132/+2614 *	S	CGAAGAGGAAGGCGATTTAG
			AS	GTCTGGCCACCTCGTGTC
3	*LEF1* sense	+3864/+4048	S	CATCTGGTTTGCTGCTAAGCTA
			AS	CAATGATGCACTGACTTCCCTTT
3,7,8	*LEF1* NAT	-11/-8596	S	TCTGTAATCTCCGCTCCGCT
			AS	CACTGTGCCTGTGTAGGATGTG
3,7		-11/-391	S	TCTGTAATCTCCGCTCCGCT
			AS	GCGTGCTCCCTCCAGAA
3		-769/-1856	S	TTGCCAAGAATAAAGTTTTTGCC
			AS	GTACCCTTGTCTCCAAAGAGCG
		+4/-953	S	TCCCGGCGGCTCTG
			AS	GCCGGCGAGCCAGG
1,3,6,7, 13,25		+213/+60	S	CCTCGTGTCCGTTGCTG
			AS	GACGAGATGATCCCCTTCAAG
3		-1562/-3057	S	CACTCCTTTTCCTCTGCCAGTC
			AS	CTGCATTTTCCACTTGTTTTCC
		-4004/-4850	S	CAATTGCTAGGGGCTGGCT
			AS	CCGAGCTACAGACGCCAA
		-2097/-2205	S	AAAGGACTCAGGAGAGCAAAGC
			AS	GGAGACTGGGGAATTTTTGAGG
		-2950/-3057	S	CAGAGAGGTGAACGTGTAGCTG
			AS	CTGCATTTTCCACTTGTTTTCC
		-5790/-5911	S	AGCTGGATCTGAAATGTGAATTGA
			AS	CACAGAAAAGAGTCAAAGGGAGC
		-6738/-6872	S	GGGACAGTGTTACGCTCTCG
			AS	ACGAGTTAAGGCACATTCACCC
		-7831/-7933	S	GGTCTCAAATTTCAAACCTCAGG
			AS	GGACATTTGGGAAAAGGTACATTG
4,5		-1562/-1688	S	CACTCCTTTTCCTCTGCCAGTC
			AS	CTCCCCACTGTCAGAGCATCT
7		-251/-384	S	GAGGCCGAAGGGCACT
			AS	TTCTCCGCGCGCAAAT
6,7,8		-11/-113	S	TCTGTAATCTCCGCTCCGCT
			AS	GCACGAACCCTTCCAACTCT

Table M2. Oligonucleotides used for transcript analysis. *Note that *LEF1* mRNA oligonucleotides correspond to the following sequences: +132/+151 for the sense oligo and +206/+216_+2603/+2614 for the antisense oligo (two exons are overlapped).

For **oligo-specific semi-quantitative RT-PCRs**, the reverse oligonucleotide was added at the beginning of the reaction, for the retrotranscription. Just before the PCR cycles started, the forward oligonucleotide was added to the reaction.

Nuclear/cytoplasmic RNA extraction

Cells were washed twice with PBS and scrapped in 200 µl of Soft Lysis Buffer (see ChIP buffers). Extracts were centrifuged at 3000 rpm for 10 minutes at 4ºC. Supernatant (cytoplasm) was separated in a new eppendorf and Trizol was added to extract cytoplasmic RNA. Pellet was resuspended again with 200 µl of Soft Lysis Buffer, centrifuged again and supernatant was discarded. Trizol was added to the pellet to extract nuclear RNA.

RNA stabilization (Actinomycin D treatment)

RWP1 snail1 cells were transfected with pBabe *LEF1* NAT or an empty vector. 24 hours after transfection actinomycin D was added at 2 µg/mL concentration and cells lysed for RNA extraction at 2, 5, 8 and 12 hours after the addition of the drug. A sample was taken without addition of actinomycin D as control. RNA was extracted with Trizol method.

4. Luciferase reporter assays

Reporter assays were performed using 100-400 ng of the indicated constructs containing the different reporter vectors pGL3*[109] plasmid (named pGL3 in all thesis). Transfections were performed using Lipofectamine Plus kit (Gibco) according to the manufacturer's instructions. 10 ng pRenilla-TK luciferase or 1 ng

pRenilla-SV40 was also cotransfected to control efficiency of transfection. Expression of Firefly and Renilla luciferases was analyzed 48 hours after transfection according to the manufacturer's instructions (Promega).

5. Western Blot

Total cell extracts were prepared by homogenizing cells in 1% SDS buffer after two washes with phosphate buffered saline (PBS). After passing cell extracts through a 20-gauge syringe, extracts were centrifuged at 20000 g for 5 minutes. Protein concentration from supernatants was determined by Lowry (Bio-Rad *DC™* Protein Assay).

Protein was loaded in sodium dodecyl sulfate 10% polyacrylamide gel electrophoresis (SDS-PAGE) and gels were run in TGS buffer (see Western Blot solutions). Proteins were transferred to a nitrocellulose membrane (Protran) for 75 minutes using Transfer Buffer (TB). Membranes were blocked with 5% non-fat milk in TBS-T buffer (25mM Tris-HCl pH 7.5, 137 mM NaCl, 0.1% Tween20) for one hour. Primary antibody was added to fresh blocking solution and incubated o/n at 4ºC. Antibodies used are listed in **Table M3**.

Protein detected	Commercial Provider	Code
LEF1	Cell Signalling	2230S
Pyruvate Kinase	Chemicon	AB1235
HA	Roche/Sigma	H6908
E-cadherin	BD-Biosciences	612131
Ezh2	Hybridome Supernatant [107]	
Suz12	Hybridome Supernatant [107]	

Table M3. Antibodies used for Western Blot.

After three ten minutes washes with TBS-T, secondary antibody peroxidase-combined (HRP) was incubated (in fresh blocking solution) for one hour at Room Temperature. Three more ten minutes washes were performed with TBS-T. Membranes were developed using substrate for HRP *Enhanced ChemiLuminiscene*, ECL, either incubating PIERCE® ECL Western Blotting substrate for one minute or Supersignal® West Dura Extended Duration substrate for five minutes. Membranes were exposed on Agfa-Curix autoradiographic films.

Solutions

1% SDS lysis buffer: 1% SDS, 10 mM EDTA, 50 mM Tris pH 8.0

TGS: 25 mM Tris pH 8.3, 192 mM Glycine, 5% SDS

Transfer Buffer: 50 mM Tris, 386 mM Glycine, 0.1% SDS

Sample buffer for proteins (Laemmli, 1x): 60 mM Tris-HCl pH 6.8, 2% SDS, 5% β-mercaptoethanol, 0.005% Bromophenol blue, 5% Glycerol

TBS: 25 mM Tris-HCl pH 7.5, 137 mM NaCl

TBS-T: TBS, 0.1% Tween20

Ponceau: 0.5% Ponceau (w/v), 1% Glacial acetic acid

SDS-PAGE recipie	Resolving	Stacking
% polyacrylamide	10	4
H_2O	4.9 mL	6 mL
Tris-HCl 1.5M pH 8.8	2.5 mL	-
Tris-HCl 0.5M pH 6.8	-	2.5 mL
10% SDS	100 µl	
Acrylamide/bisacrylamide (37.5:1)	2.5 mL	1.4 mL
10% APS	40 µl	
TEMED	20 µl	
Final volume	10 mL	

Table M4. Reagents used to prepare polyacrylamide gels.

6. Chromatin Immunoprecipitation (ChIP)

Cells seeded in 150-mm-diameter plates were washed twice with phosphate buffered saline (PBS) pre-warmed at 37ºC and crosslinked with 1% formaldehyde in DMEM for 10 minutes at 37ºC. Reaction was stopped by adding 250 µL of glycine 2.5 M (0.125 M final concentration) and incubating for 2 more minutes. Cells were washed twice with cold PBS and 1 mL of Soft Lysis Buffer (with Protease inhibitor cocktail, Complete Mini, Roche) was added to the plates on ice. After scrapping, lysates were centrifuged for 15 minutes at 3000 rpm at 4ºC. Supernatant was discarded, pellet resuspended in SDS lysis buffer and sonicated to generate fragments of DNA from 200 to 500 bp (40% amplitude in Bransol DIGITAL Sonifier® UNIT Model S-450D sonicator, 10 pulses of 10 seconds).

Lysates were incubated for 20 minutes on ice and centrifuged at maximum speed for 10 minutes. Protein concentration was determined by Lowry (Bio-rad) and the desired amount of protein per immunoprecipitation (IP) was diluted in Dilution Buffer (1:10). Pre-clearing was performed to reduce background with IgGs (Dako), salmon sperm (Ambion) and BSA (bovine serum albumin) blocked protein A or G (Upstate) for 3 hours at 4ºC and agitation.

Samples were then centrifuged at 2000 rpm, input was stored at 4ºC and samples for IP divided and incubated either with specific antibody (**Table M5**) or irrelevant antibody of the same specie overnight at 4ºC and agitation.

Antibody against	Commercial Provider	Code
H3K4me2	Millipore	07-030
H3K27me3	Diagenode	Ls-069-100
Suz12	Hybridome supernatant[107]	
Ezh2	Hybridome supernatant[107]	

Table M5. Antibodies used for ChIP analyses.

Blocked beads were added to each sample and incubated for one more hour at 4ºC.

Five washes were performed with each Low Salt Buffer, High Salt Buffer, LiCl Buffer and TE Buffer (see below). Samples were eluted after centrifuging them to eliminate all traces of buffer and incubating the remaining beads with Elution Buffer at 37ºC for 30 minutes. DNA was recovered by centrifugation (5 minutes at 2000 rpm). Decrosslinking was performed incubating samples at 65ºC overnight. After one hour digestion with proteinase K (Roche), DNA was purified by the GFX PCR DNA and Gel Band Purification kit (GE Healthcare). DNA fragments were analyzed using quantitative PCR. Promoter regions were detected by qPCR with LightCycler®480 Real-Time PCR System (Roche), using the oligonucleotides in **Table M6**. Data collection was performed with the LightCycler®480 System. The ChIP results were quantified relative to the input amount.

DNA amplified	Amplicon	Oligonucleotide (5'-3')
***LEF1* promoter**	**-951/-750**	CTCGAGCCGGGAACAAAGA
		GGGAAGAGAAAGAGAAGTTTGCC
NAT promoter	**+266/+433**	CAGAGAGGGAGGAAGGGAAC
		CCCCTCTACCTCCCATCCTA
Control sequence	**+3864/+4048**	CATCTGGTTTGCTGCTAAGCTA
		CAATGATGCACTGACTTCCCTTT

Table M6. Oligonucleotides used to analyze the DNA fragments in ChIP experiments.

Solutions

Soft Lysis Buffer: 50 mM Tris pH 8.0, 10 mM EDTA, 0.1% Nonidet P-40, 10% glycerol, Protease inhibitor cocktail (Complete Mini, Roche), Phosphatase Inhibitors (20 mM sodium fluoride phosphatase (NaF), 0.5 mM sodium orthovanadate (NaOV) and 2.5 mM sodic butyrate)

SDS Lysis Buffer: 1% SDS, 10 mM EDTA, 50 mM Tris pH 8.0

Dilution Buffer: 0.01% SDS, 1.1% Triton X-100, 16.7 mM Tris pH 8.0, 1.2 mM EDTA, 167 mM NaCl

Low Salt Buffer: 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris pH 8.0, 150 mM NaCl

High Salt Buffer: 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris pH 8.0, 500 mM NaCl

LiCl Buffer: 250 mM LiCl, 1% Nonidet P-40, 1% Sodium deoxycholate (NaDOC), 1 mM EDTA, 10 mM Tris pH 8.0

Elution Buffer: 100 mM Na_2CO_3, 1% SDS

7. Chromatin Isolation by RNA Purification (ChIRP)

RWP1 cells were transfected with pGL3 *LEF1* promoter when indicated and *in vitro* synthesized biotinylated NATs or irrelevant RNAs. pGL3 *CDH1* promoter[105] was alternatively transfected as control when indicated. Cells were crosslinked with formaldehyde as for the ChIP assays and chromatins were prepared with ChIP buffers with RNAse out (Fermentas) for preserving RNAse free conditions. Fragmentation of the chromatin was performed with Bransol DIGITAL Sonifier® UNIT Model S-450D sonicator as in ChIP in Figures R18, R19 and R20. In Figure R21, fragmentation of the chromatin was performed with Biorruptor (Diagenode).

Anti-biotin antibody (Sigma) and protein A-agarose were used to immunoprecipitate biotinylated RNAs. Samples were washed with ChIP buffers with RNAse out (Fermentas), treated with elution buffer and purified as for ChIP assays. The presence of the amplicons -1806/-1626, -1306/-1188 and -904/-703 (corresponding to *LEF1* promoter), +3864/+4048 (corresponding to a downstream region in *LEF1* mRNA, when endogenous *LEF1* promoter was detected) and +570/+744 (corresponding to Luciferase, when exogenous *LEF1* promoter was detected) was measured by qPCR.

The sequence of the Luciferase amplicon corresponds to +483/+657 with respect to Luciferase translation start site.

DNA amplified	Amplicon	Oligonucleotide (5'-3')
***LEF1* promoter**	**-1806/-1626**	AACTCTCTTTTCCTTGTCCTTCTG
		GCAGAGGGAGGAAGATGAAA
	-1306/-1188	AGACTCGTCCTACAGGATCTGG
		CGCTGAAAAGCTACCCACTT
	-904/-703	ACTGAGTGTGTGTGTCGGCT
		ATCTGCTAGAGAAGGAGGAGGAG
Control sequence	**+3864/+4048**	CATCTGGTTTGCTGCTAAGCTA
		CAATGATGCACTGACTTCCCTTT
Luciferase	**+570/+744**	GCACATATCGAGGTGGACATC
		CGCAACTGCAACTCCGAT
***CDH1* promoter**		AACCCTCAGCCAATCAGCGG
		GTTCCGACGCCACTGAGAGG

Table M7. Oligonucleotides used to analyze the DNA fragments in ChIRP experiments.

8. RNA Immunoprecipitation (RIP)

RWP1 snail1 *LEF1* NAT cells were washed with cold PBS and lysed with Polysomal Lysis Buffer (see below). When indicated, cell extract (500 µg) was treated with 400 units of the nucleases DNAse Turbo, RNAse H, RNAse V1 or RNAse A1 (Ambion), in a final volum of 1 mL. Cell extracts were pre-cleared with irrelevant IgGs, and protein G-magnetic beads previously blocked with salmon sperm (1 mg/mL), poly dI-dC (1 µg/mL) and BSA (100 µg/mL). Immunoprecipitation was carried out with hybridome supernatants from Suz12 and Ezh2 and with irrelevant immunoglobulin G (Sigma/DAKO) o/n at 4ºC with agitation. Binding was performed with protein G-magnetic beads for one hour at 4ºC with agitation. After 5 washes with NT2 Buffer (see below), RNA was extracted using Trizol (Invitrogen) method. Transcripts were

analyzed by RT-PCR of RT-qPCR using procedures described in the Transcript analysis section. RIP results were quantified relative to the input amount.

Solutions

Polysomal Lysis Buffer: 100 mM KCl, 5 mM $MgCl_2$, 10 mM Hepes pH 7.0, 0.5% Nonidet P-40, 1 mM DTT, 100 units/mL RNAse out (Fermentas), Protease inhibitor cocktail (Complete Mini, Roche)

NT2 Buffer: 50 mM Tris HCl pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$, 0.1% Nonidet P-40, 100 units/mL RNAse out (Fermentas), Protease inhibitor cocktail (Complete Mini, Roche)

9. *In vitro* transcription

pcDNA3 *LEF1* NAT plasmids were linearized using XbaI, located at the end of the polylinker. pcDNA3 Cre plasmid was also linearized using XbaI. pcDNA3 YBX1 plasmid was linearized using XhoI, also at the end of the polylinker. pcDNA3 empty vector was linearized with SmaI to generate a RNA fragment of 900 bp.

Transcription reactions were performed using T7MEGAscript kit (Ambion). When indicated, RNAs were biotinylated adding 30% biotin 16-UTP (Ambion) to the reaction.

10. Biotinylated Oligo Pull-down Assay (BOPA)

Biotinylated-DNA pull-down assays (**DNA-BOPA**) were carried out using a biotinylated DNA fragment corresponding to the *LEF1* promoter (-1856/+58). It was generated by high fidelity Pfx polymerase (Invitrogen) PCR using the corresponding DNA as template with the same specific oligonucleotides used for cloning, with a 5'-biotin label on the sense primer. PCR products were

loaded in 1% agarose gels and purified using GFX PCR DNA and gel band purification kit (Amersham).

In **Figure R24A**, RWP1 snail1 cells were transfected with pBabe empty or pBabe *LEF1* NAT. Cells were lysed with Polysomal Lysis Buffer (see RIP assays) and diluted in binding buffer (20 mM Hepes pH 7.6, 150 mM KCl, 3 mM MgCl2, 10% glycerol, 3 mg/ml BSA, 0.2 Triton X-100, 20 µg/ml poly dI-dC, 1 mM DDT, plus protease and RNase inhibitors). Pre-clearing was performed by incubating with protein G-agarose blocked with salmon sperm (1 mg/mL) and mouse IgG (10 µg/mL). After pre-clearing, samples (500 µg) were incubated 4 hours in binding buffer with the DNA-biotinylated probes (2 µg) and antibody against biotin (20 µg/mL) in a final volume of 1 mL. Samples were pulled-down with protein G-agarose, washed with binding buffer supplemented with 0.1% Tween-20 and analyzed by Western Blot for Suz12.

In **Figure R24B**, RWP1 snail1 cells were lysed with Polysomal Lysis Buffer and diluted in binding buffer. Pre-clearing was performed by incubating with protein G-agarose blocked with salmon sperm (1 mg/mL) and mouse IgG (10 µg/mL). After pre-clearing, samples (500 µg) were incubated 4 hours in binding buffer with the DNA-biotinylated probes (2 µg), *in vitro* synthesized RNA (4 µg), and antibody against biotin (20 µg/ml) in a final volume of 1 mL. Samples were pulled-down with protein G-agarose, washed with binding buffer supplemented with 0.1% Tween-20 and analyzed by Western Blot for Ezh2.

RNA-BOPA was performed using biotinylated *LEF1* NAT, generated by *in vitro* transcription adding biotin16-UTP (Ambion) to the reaction. HT29 M6 cells were lysed with Polysomal Lysis Buffer. Cell extracts were incubated with 6 µg of biotinylated NAT, biotinylated irrelevant RNA or with no RNA. pGL3 *LEF1* promoter was added in the reaction when indicated. The procedure was as described for DNA-BOPA. Immunoprecipitation was performed with anti-biotin

antibody, the pull-down with protein G-agarose and the presence of PRC2 subunits was analyzed by Western Blot.

11. Migration assays

HT29 M6 snail1 and RWP1 snail1 cells were transfected using 6 µg of pBabe *LEF1* NAT or empty vector according to Fugene HD manufacturer's instructions and selected for stable expression using Puromycin (Life Technologies).

Two dimension (2D) cell migration assays were performed seeding cells at high confluence in 24-well plate. Once cells were attached, a scratch was performed using Essen 24-well wound maker (Essen Bioscience). Cells were washed three times with PBS and plate was incubated with complete media in Incucyte Essen (Essen Bioscience) where three high definition phase contrast image of the same area were taken each two hours. Using Incucyte Essen software images were analyzed measuring changes of position of cell frontline in each image and represented as average of three different replicates respect to each time point. Error bars depict standard error of the mean.

12. Proliferation assays

HT29 M6 snail1 and RWP1 snail1 cells were transfected using 6 µg of pBabe *LEF1* NAT or empty vector according to Fugene HD manufacturer's instructions and selected for stable expression using Puromycin (Life Technologies).

Proliferation assays were performed seeding equal amount of cells at low confluence in a 96 well plate. Once cells were attached, plate was incubated in Incucyte Essen (Essen Bioscience) where

three high definition phase contrast image of the same area were taken each two hours. Using Incucyte Essen software, images were analyzed for calculating average confluence in each image and represented as average of three different replicates respect to each time point. Error bars depict standard error of the mean.

13. Microarrays

RWP1 snail1 cells were transfected using 6 µg of pBabe *LEF1* NAT or empty vector according to Fugene HD manufacturer's instructions and selected for stable expression using Puromycin (Life Technologies). RNA extraction was performed using Trizol (Invitrogen) and samples were treated with DNAse Turbo (Ambion) to avoid DNA contamination. RNA integrity was analyzed with Bioanalyzer 2100. Microarrays were performed on Human Gene 1.0 ST (Affymetrix®). Samples were processed according to Manual Ambion® WT Expression (P/N 4425209 Rev.B 05/2009) (Ambion/Applied Biosystems, Foster City, CA, USA) and to Manuals of Affymetrix WT Terminal Labeling and Hybridization User Manual (P/N 702808 Rev.1) and Expression Wash, Stain and Scan User Manual (P/N 702731 Rev. 3) (Affymetrix Inc., Santa Clara, CA, USA).

REFERENCES

1. Mattick, J. S. RNA regulation: a new genetics? *Nat Rev Genet* **5,** 316–23 (2004).

2. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nature reviews. Genetics* **10,** 155–9 (2009).

3. Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **136,** 629–41 (2009).

4. Wilusz, J. E., Sunwoo, H. & Spector, D. L. Long noncoding RNAs: functional surprises from the RNA world. *Genes & development* **23,** 1494–504 (2009).

5. Bernstein, E. & Allis, C. D. RNA meets chromatin. *Genes & development* **19,** 1635–55 (2005).

6. Bracken, A. P. & Helin, K. Polycomb group proteins: navigators of lineage pathways led astray in cancer. *Nature reviews. Cancer* **9,** 773–84 (2009).

7. Faghihi, M. A. & Wahlestedt, C. Regulatory roles of natural antisense transcripts. *Nature reviews. Molecular cell biology* **10,** 637–43 (2009).

8. Whitehead, J., Pandey, G. K. & Kanduri, C. Regulation of the mammalian epigenome by long noncoding RNAs. *Biochimica et biophysica acta* **1790,** 936–47 (2009).

9. Wahlestedt, C. Targeting long non-coding RNA to therapeutically upregulate gene expression. *Nature reviews. Drug discovery* **12,** 433–46 (2013).

10. 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 (1991).

11. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458,** 223–7 (2009).

12. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research* **22,** 1775–89 (2012).

13. Bernstein, B. E. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).

14. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489,** 101–8 (2012).

15. Taft, R. J., Pheasant, M. & Mattick, J. S. The relationship between non-protein-coding DNA and eukaryotic complexity. *BioEssays : news and reviews in molecular, cellular and developmental biology* **29,** 288–99 (2007).

16. Mathews, D. H., Moss, W. N. & Turner, D. H. Folding and finding RNA secondary structure. *Cold Spring Harbor perspectives in biology* **2,** a003665 (2010).

17. Wapinski, O. & Chang, H. Y. Long noncoding RNAs and human disease. *Trends in cell biology* **21,** 354–61 (2011).

18. Taft, R. J., Pang, K. C., Mercer, T. R., Dinger, M. & Mattick, J. S. Non-coding RNAs: regulators of disease. *The Journal of pathology* **220,** 126–39 (2010).

19. Huarte, M. & Rinn, J. L. Large non-coding RNAs: missing links in cancer? *Human molecular genetics* **19,** R152–61 (2010).

20. Guttman, M. *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477,** 295–300 (2011).

21. Gupta, R. A. *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464,** 1071–6 (2010).

22. 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,** 742–9 (2011).

23. 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,** 662–74 (2010).

24. Guttman, M. & Rinn, J. L. Modular regulatory principles of large non-coding RNAs. *Nature* **482,** 339–46 (2012).

25. Derrien, T., Guigó, R. & Johnson, R. The Long Non-Coding RNAs: A New (P)layer in the "Dark Matter". *Frontiers in genetics* **2,** 107 (2011).

26. Kornienko, A. E., Guenzl, P. M., Barlow, D. P. & Pauler, F. M. Gene regulation by the act of long non-coding RNA transcription. *BMC biology* **11,** 59 (2013).

27. Peterlin, B. M., Brogie, J. E. & Price, D. H. 7SK snRNA: a noncoding RNA that plays a major role in regulating eukaryotic transcription. *Wiley interdisciplinary reviews. RNA* **3,** 92–103

28. Espinoza, C. A., Allen, T. A., Hieb, A. R., Kugel, J. F. & Goodrich, J. A. B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nature structural & molecular biology* **11,** 822–9 (2004).

29. Espinoza, C. A., Goodrich, J. A. & Kugel, J. F. Characterization of the structure, function, and mechanism of B2 RNA, an ncRNA repressor of RNA polymerase II transcription. *RNA (New York, N.Y.)* **13,** 583–96 (2007).

30. Yakovchuk, P., Goodrich, J. A. & Kugel, J. F. B2 RNA represses TFIIH phosphorylation of RNA polymerase II. *Transcription* **2,** 45–9

31. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129,** 1311–23 (2007).

32. Ng, S.-Y., Johnson, R. & Stanton, L. W. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *The EMBO journal* **31,** 522–33 (2012).

33. Brown, C. J. *et al.* The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71,** 527–42 (1992).

34. Wutz, A. Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nature reviews. Genetics* **12,** 542–53 (2011).

35. Ørom, U. A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143,** 46–58 (2010).

36. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472,** 120–4 (2011).

37. Magistri, M., Faghihi, M. A., St Laurent, G. & Wahlestedt, C. Regulation of chromatin structure by long noncoding RNAs: focus on natural antisense transcripts. *Trends in genetics : TIG* **28,** 389–96 (2012).

38. Latos, P. a *et al.* Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science (New York, N.Y.)* **338,** 1469–72 (2012).

39. Wang, K. C. & Chang, H. Y. Molecular mechanisms of long noncoding RNAs. *Molecular cell* **43,** 904–14 (2011).

40. Pandey, R. R. *et al.* Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Molecular cell* **32,** 232–46 (2008).

41. Nagano, T. *et al.* The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science (New York, N.Y.)* **322,** 1717–20 (2008).

42. Zhao, J., Sun, B. K., Erwin, J. A., Song, J.-J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science (New York, N.Y.)* **322,** 750–6 (2008).

43. Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Molecular cell* **44,** 667–78 (2011).

44. Huarte, M. *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142,** 409–19 (2010).

45. Heo, J. B. & Sung, S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science (New York, N.Y.)* **331,** 76–9 (2011).

46. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130,** 77–88 (2007).

47. Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E. & Lingner, J. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science (New York, N.Y.)* **318,** 798–801 (2007).

48. Tripathi, V. *et al.* The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Molecular cell* **39,** 925–38 (2010).

49. Hung, T. & Chang, H. Y. Long noncoding RNA in genome regulation: prospects and mechanisms. *RNA biology* **7,** 582–5

50. Bonasio, R., Tu, S. & Reinberg, D. Molecular signals of epigenetic states. *Science (New York, N.Y.)* **330,** 612–6 (2010).

51. Spitale, R. C., Tsai, M.-C. & Chang, H. Y. RNA templating the epigenome: long noncoding RNAs as molecular scaffolds. *Epigenetics : official journal of the DNA Methylation Society* **6,** 539–43 (2011).

52. Good, M. C., Zalatan, J. G. & Lim, W. A. Scaffold proteins: hubs for controlling the flow of cellular information. *Science (New York, N.Y.)* **332,** 680–6 (2011).

53. 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,** 1956–62 (2011).

54. Brockdorff, N. Noncoding RNA and Polycomb recruitment. *RNA (New York, N.Y.)* **19,** 429–42 (2013).

55. Zhao, J. *et al.* Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Molecular cell* **40,** 939–53 (2010).

56. Lee, J. T. Epigenetic regulation by long noncoding RNAs. *Science (New York, N.Y.)* **338,** 1435–9 (2012).

57. Plath, K. *et al.* Role of histone H3 lysine 27 methylation in X inactivation. *Science (New York, N.Y.)* **300,** 131–5 (2003).

58. Silva, J. *et al.* Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Developmental cell* **4,** 481–95 (2003).

59. Fitzpatrick, G. V, Soloway, P. D. & Higgins, M. J. Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nature genetics* **32,** 426–31 (2002).

60. Kanhere, A. *et al.* Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Molecular cell* **38,** 675–88 (2010).

61. Katayama, S. *et al.* Antisense transcription in the mammalian transcriptome. *Science (New York, N.Y.)* **309,** 1564–6 (2005).

62. Engström, P. G. *et al.* Complex Loci in human and mouse genomes. *PLoS genetics* **2,** e47 (2006).

63. Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 11667–72 (2009).

64. St Laurent, G. *et al.* Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells. *BMC genomics* **13,** 504 (2012).

65. Kretz, M. *et al.* Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* **493,** 231–5 (2013).

66. Kiyosawa, H., Yamanaka, I., Osato, N., Kondo, S. & Hayashizaki, Y. Antisense transcripts with FANTOM2 clone set and their implications for gene regulation. *Genome research* **13,** 1324–34 (2003).

67. Chen, J., Sun, M., Hurst, L. D., Carmichael, G. G. & Rowley, J. D. Human antisense genes have unusually short introns: evidence for selection for rapid transcription. *Trends in genetics : TIG* **21,** 203–7 (2005).

68. Chen, J., Sun, M., Hurst, L. D., Carmichael, G. G. & Rowley, J. D. Genome-wide analysis of coordinate expression and evolution of human cis-encoded sense-antisense transcripts. *Trends in genetics : TIG* **21,** 326–9 (2005).

69. He, Y., Vogelstein, B., Velculescu, V. E., Papadopoulos, N. & Kinzler, K. W. The antisense transcriptomes of human cells. *Science (New York, N.Y.)* **322,** 1855–7 (2008).

70. Finocchiaro, G. *et al.* Localizing hotspots of antisense transcription. *Nucleic acids research* **35,** 1488–500 (2007).

71. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science (New York, N.Y.)* **322,** 1845–8 (2008).

72. Seila, A. C. *et al.* Divergent transcription from active promoters. *Science (New York, N.Y.)* **322,** 1849–51 (2008).

73. Sun, M., Hurst, L. D., Carmichael, G. G. & Chen, J. Evidence for a preferential targeting of 3'-UTRs by cis-encoded natural antisense transcripts. *Nucleic acids research* **33,** 5533–43 (2005).

74. Okada, Y. *et al.* Comparative expression analysis uncovers novel features of endogenous antisense transcription. *Human molecular genetics* **17,** 1631–40 (2008).

75. Morris, K. V, Santoso, S., Turner, A.-M., Pastori, C. & Hawkins, P. G. Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. *PLoS genetics* **4,** e1000258 (2008).

76. Bissell, M. J. & Radisky, D. Putting tumours in context. *Nature reviews. Cancer* **1,** 46–54 (2001).

77. Thiery, J. P. & Chopin, D. Epithelial cell plasticity in development and tumor progression. *Cancer metastasis reviews* **18,** 31–42 (1999).

78. Boyer, B. & Thiery, J. P. Epithelium-mesenchyme interconversion as example of epithelial plasticity. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica* **101,** 257–68 (1993).

79. Hay, E. D. An overview of epithelio-mesenchymal transformation. *Acta anatomica* **154,** 8–20 (1995).

80. Adams, C. L. & Nelson, W. J. Cytomechanics of cadherin-mediated cell-cell adhesion. *Current opinion in cell biology* **10,** 572–7 (1998).

81. Garrod, D., Chidgey, M. & North, A. Desmosomes: differentiation, development, dynamics and disease. *Current opinion in cell biology* **8,** 670–8 (1996).

82. LaGamba, D., Nawshad, A. & Hay, E. D. Microarray analysis of gene expression during epithelial-mesenchymal transformation. *Developmental dynamics : an official publication of the American Association of Anatomists* **234,** 132–42 (2005).

83. Yang, J. & Weinberg, R. A. Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Developmental cell* **14,** 818–29 (2008).

84. Schäfer, M. & Werner, S. Cancer as an overhealing wound: an old hypothesis revisited. *Nature reviews. Molecular cell biology* **9,** 628–38 (2008).

85. Davies, J. A. Mesenchyme to epithelium transition during development of the mammalian kidney tubule. *Acta anatomica* **156,** 187–201 (1996).

86. Peinado, H., Olmeda, D. & Cano, A. Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nature reviews. Cancer* **7,** 415–28 (2007).

87. Christiansen, J. J. & Rajasekaran, A. K. Reassessing epithelial to mesenchymal transition as a prerequisite for carcinoma invasion and metastasis. *Cancer research* **66,** 8319–26 (2006).

88. Gotzmann, J. *et al.* Molecular aspects of epithelial cell plasticity: implications for local tumor invasion and metastasis. *Mutation research* **566,** 9–20 (2004).

89. Thiery, J. P. Epithelial-mesenchymal transitions in tumour progression. *Nature reviews. Cancer* **2,** 442–54 (2002).

90. Tarin, D., Thompson, E. W. & Newgreen, D. F. The fallacy of epithelial mesenchymal transition in neoplasia. *Cancer research* **65,** 5996–6000; discussion 6000–1 (2005).

91. Kemler, R. From cadherins to catenins: cytoplasmic protein interactions and regulation of cell adhesion. *Trends in genetics : TIG* **9,** 317–21 (1993).

92. Takeichi, M. Morphogenetic roles of classic cadherins. *Current opinion in cell biology* **7,** 619–27 (1995).

93. Tepass, U., Truong, K., Godt, D., Ikura, M. & Peifer, M. Cadherins in embryonic and neural morphogenesis. *Nature reviews. Molecular cell biology* **1,** 91–100 (2000).

94. Thiery, J. P. & Sleeman, J. P. Complex networks orchestrate epithelial-mesenchymal transitions. *Nature reviews. Molecular cell biology* **7,** 131–42 (2006).

95. Ara, T. *et al.* Membrane type 1-matrix metalloproteinase expression is regulated by E-cadherin through the suppression of mitogen-activated protein kinase cascade. *Cancer letters* **157,** 115–21 (2000).

96. Noren, N. K., Arthur, W. T. & Burridge, K. Cadherin engagement inhibits RhoA via p190RhoGAP. *The Journal of biological chemistry* **278,** 13615–8 (2003).

97. Wu, H. *et al.* Positive expression of E-cadherin suppresses cell adhesion to fibronectin via reduction of alpha5beta1 integrin in human breast carcinoma cells. *Journal of cancer research and clinical oncology* **132,** 795–803 (2006).

98. Kuphal, S. & Bosserhoff, A. K. Influence of the cytoplasmic domain of E-cadherin on endogenous N-cadherin expression in malignant melanoma. *Oncogene* **25,** 248–59 (2006).

99. Kuphal, S., Poser, I., Jobin, C., Hellerbrand, C. & Bosserhoff, A. K. Loss of E-cadherin leads to upregulation of NFkappaB activity in malignant melanoma. *Oncogene* **23,** 8509–19 (2004).

100. Wells, A., Yates, C. & Shepard, C. R. E-cadherin as an indicator of mesenchymal to epithelial reverting transitions during the metastatic seeding of disseminated carcinomas. *Clinical & experimental metastasis* **25,** 621–8 (2008).

101. Peinado, H., Ballestar, E., Esteller, M. & Cano, A. Snail mediates E-cadherin repression by the recruitment of the Sin3A/histone deacetylase 1 (HDAC1)/HDAC2 complex. *Molecular and cellular biology* **24,** 306–19 (2004).

102. Giroldi, L. A. *et al.* Role of E boxes in the repression of E-cadherin expression. *Biochemical and biophysical research communications* **241,** 453–8 (1997).

103. Hennig, G. *et al.* Progression of carcinoma cells is associated with alterations in chromatin structure and factor binding at the E-cadherin promoter in vivo. *Oncogene* **11,** 475–84 (1995).

104. Nieto, M. A. The snail superfamily of zinc-finger transcription factors. *Nature reviews. Molecular cell biology* **3,** 155–66 (2002).

105. Batlle, E. *et al.* The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nature cell biology* **2,** 84–9 (2000).

106. Cano, A. *et al.* The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nature cell biology* **2,** 76–83 (2000).

107. Herranz, N. *et al.* Polycomb complex 2 is required for E-cadherin repression by the Snail1 transcription factor. *Molecular and cellular biology* **28,** 4772–81 (2008).

108. Carver, E. A., Jiang, R., Lan, Y., Oram, K. F. & Gridley, T. The mouse snail gene encodes a key regulator of the epithelial-mesenchymal transition. *Molecular and cellular biology* **21,** 8184–8 (2001).

109. Guaita, S. *et al.* Snail induction of epithelial to mesenchymal transition in tumor cells is accompanied by MUC1 repression and ZEB1 expression. *The Journal of biological chemistry* **277,** 39209–16 (2002).

110. Stanisavljevic, J., Porta-de-la-Riva, M., Batlle, R., de Herreros, A. G. & Baulida, J. The p65 subunit of NF-κB and PARP1 assist Snail1 in activating fibronectin transcription. *Journal of cell science* **124,** 4161–71 (2011).

111. Yokoyama, K. *et al.* Increased invasion and matrix metalloproteinase-2 expression by Snail-induced mesenchymal transition in squamous cell carcinomas. *International journal of oncology* **22,** 891–8 (2003).

112. Del Barrio, M. G. & Nieto, M. A. Overexpression of Snail family members highlights their ability to promote chick neural crest formation. *Development (Cambridge, England)* **129,** 1583–93 (2002).

113. Beltran, M. *et al.* A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-

mesenchymal transition. *Genes & development* **22,** 756–69 (2008).

114. Kajita, M., McClinic, K. N. & Wade, P. A. Aberrant expression of the transcription factors snail and slug alters the response to genotoxic stress. *Molecular and cellular biology* **24,** 7559–66 (2004).

115. Lee, S.-H. *et al.* Blocking of p53-Snail binding, promoted by oncogenic K-Ras, recovers p53 expression and function. *Neoplasia (New York, N.Y.)* **11,** 22–31, 6p following 31 (2009).

116. Vega, S. *et al.* Snail blocks the cell cycle and confers resistance to cell death. *Genes & development* **18,** 1131–43 (2004).

117. Escrivà, M. *et al.* Repression of PTEN phosphatase by Snail1 transcriptional factor during gamma radiation-induced apoptosis. *Molecular and cellular biology* **28,** 1528–40 (2008).

118. Colin Jamora, Ramanuj DasGupta, Pawel Kocieniewski, E. F. Links between signal transduction, transcription and adhesion in epithelial bud development. *Nature* **422,** 317–322 (2003).

119. Li, T. W. *et al.* Wnt Activation and Alternative Promoter Repression of LEF1 in Colon Cancer. **26,** 5284–5299 (2006).

120. Hovanes, K. *et al.* Beta-catenin-sensitive isoforms of lymphoid enhancer factor-1 are selectively expressed in colon cancer. *Nature genetics* **28,** 53–7 (2001).

121. Waterman, M. L. Lymphoid enhancer factor/T cell factor expression in colorectal cancer. *Cancer metastasis reviews* **23,** 41–52

122. Medici, D., Hay, E. D. & Olsen, B. R. Snail and Slug promote epithelial-mesenchymal transition through beta-catenin-T-cell factor-4-dependent expression of transforming growth factor-beta3. *Molecular biology of the cell* **19,** 4875–87 (2008).

123. Yokoyama, N. N., Pate, K. T., Sprowl, S. & Waterman, M. L. A role for YY1 in repression of dominant negative LEF-1 expression in colon cancer. *Nucleic acids research* **38,** 6375–88 (2010).

124. Hovanes, K., Li, T. W. & Waterman, M. L. The human LEF-1 gene contains a promoter preferentially active in lymphocytes and encodes multiple isoforms derived from alternative splicing. *Nucleic acids research* **28,** 1994–2003 (2000).

125. Lehner, B., Williams, G., Campbell, R. D. & Sanderson, C. M. Antisense transcripts in the human genome. *Trends in genetics : TIG* **18,** 63–5 (2002).

126. Kouzarides, T. Chromatin modifications and their function. *Cell* **128,** 693–705 (2007).

127. Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H. & Helin, K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes & development* **20,** 1123–36 (2006).

128. Amaral, P. P. & Mattick, J. S. Noncoding RNA in development. *Mammalian genome : official journal of the International Mammalian Genome Society* **19,** 454–92 (2008).

129. Tsai, M.-C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science (New York, N.Y.)* **329,** 689–93 (2010).

130. Lin, Y. *et al.* The SNAG domain of Snail1 functions as a molecular hook for recruiting lysine-specific demethylase 1. *The EMBO journal* **29,** 1803–16 (2010).

131. Zuker, M., Mathews, D. & Turner, D. *Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide in RNA Biochemistry and Biotechnology*. 11–43 (NATO ASI Series, Kluwer Academic Publishers, 1999).

132. Yue, X. *et al.* Transcriptional regulation by small RNAs at sequences downstream from 3' gene termini. *Nature chemical biology* **6,** 621–9 (2010).

133. Tsai, C.-L., Rowntree, R. K., Cohen, D. E. & Lee, J. T. Higher order chromatin structure at the X-inactivation center via looping DNA. *Developmental biology* **319,** 416–25 (2008).

134. Jeon, Y. & Lee, J. T. YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* **146,** 119–33 (2011).

135. Davidovich, C., Zheng, L., Goodrich, K. J. & Cech, T. R. Promiscuous RNA binding by Polycomb repressive complex 2. *Nature structural & molecular biology* (2013). doi:10.1038/nsmb.2679

136. Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R. & Misteli, T. Epigenetics in alternative pre-mRNA splicing. *Cell* **144,** 16–26 (2011).

137. Braunschweig, U., Gueroussov, S., Plocik, A. M., Graveley, B. R. & Blencowe, B. J. Dynamic integration of splicing within gene regulatory pathways. *Cell* **152,** 1252–69 (2013).

138. Reinke, L. M., Xu, Y. & Cheng, C. Snail represses the splicing regulator epithelial splicing regulatory protein 1 to promote epithelial-mesenchymal transition. *The Journal of biological chemistry* **287,** 36435–42 (2012).

139. Sánchez-Tilló, E. *et al.* β-catenin/TCF4 complex induces the epithelial-to-mesenchymal transition (EMT)-activator ZEB1 to regulate tumor invasiveness. *Proceedings of the National Academy of Sciences of the United States of America* **108,** 19204–9 (2011).

140. Wahlestedt, C. Natural antisense and noncoding RNA transcripts as potential drug targets. *Drug discovery today* **11,** 503–8 (2006).

141. Modarresi, F. *et al.* Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. *Nature biotechnology* **30,** 453–9 (2012).

142. Davis, M. E. *et al.* Evidence of RNAi in humans from systemically administered siRNA via targeted nanoparticles. *Nature* **464,** 1067–70 (2010).

143. Peiró, S. *et al.* Snail1 transcriptional repressor binds to its own promoter and controls its expression. *Nucleic acids research* **34,** 2077–84 (2006).

144. Lesuffleur, T., Barbat, A., Dussaulx, E. & Zweibaum, A. Growth adaptation to methotrexate of HT-29 human colon carcinoma cells is associated with their ability to differentiate into columnar absorptive and mucus-secreting cells. *Cancer research* **50,** 6334–43 (1990).

145. Nose, A., Nagafuchi, A. & Takeichi, M. Expressed recombinant cadherins mediate cell sorting in model systems. *Cell* **54,** 993–1001 (1988).

146. Jarriault, S. *et al.* Signalling downstream of activated mammalian Notch. *Nature* **377,** 355–8 (1995).

ANNEX

LEF1 locus (-2000/+1000)

```
-2000 CACAACTCTCCGTACATCCCGTGGTGAGAACAGAATGAAAGATATATTGT -1951
-1950 TTAAAAAGCAATAATTAAAATCTAGTCTTCAGTTCCTTCTTCCTCGTCAT -1901
-1900 ATTTTTTCTCCGTAAGCCTAGAGATTTTATTTTCACTAGTGTGTTCCCTT -1851
-1850 GTCTCCAAAGAGCGTGTGTGTGATATTATATTCGGGAAAGCTACAACTCT -1801
-1800 CTTTTCCTTGTCCTTCTGTTCTCTCTTAATAGTTGAGCAATGTCTGTTAT -1751
-1750 ATTTTCCCCTTTTCCTTTTTTTCTCAGTCCCAGATTCCCGCCTCTCCCCA -1701
-1700 CTGTCAGAGCATCTATCAATGTGGTGTCCATCACAGCGGCAGCGGCTTTC -1651
-1650 TCTTTCATCTTCCTCCCTCTGCCAGAGCCAGGGAGGGAGAGTGGGAGGCG -1601
-1600 TCAAGGAGGTAGGGGAGAGACTGGCAGAGGAAAAGGAGTGGGTGGGTGGG -1551
-1550 GGCCAAGTAAATAGATACTTAGATGATGAAGTCAAGCCACTGCGGCAATG -1501
-1500 TTTCTTGTCAGTTTCACGCGGGCAAAGCGTGCCTTTCGGTGGGTTATAAG -1451
-1450 CAGCGCCCGGTCCTTCCTTCTCTCGCCAAGTTGCCTGATCCTTCCCTCCA -1401
-1400 GGCGCGCGCGCACACACCACACTCACACACCCCAAAACCAAGACTCGTCC -1351
-1350 TACAGGATCTGGGAAAAGAAAAAGAAAAAAAAGCCCTCAATCACCACCTC -1301
-1300 CTTCTCGCCGACTCCCCCTCACCCCCCGCCTCCCCTCCAGCGGGCAGCCA -1251
-1250 AGGAGAGCTAGAGGCGGGGGAGGGGAGAGGGAGGAGAAGCGACGCAAGTG -1201
-1200 GGTAGCTTTTCAGCGCCGGCGAGGCGCGGGAGGAGGAGAAGCAGTGGGGA -1151
-1150 GGCGCAGCCGCTCACCTGCGGGGCAGGGCGCGGAGGAGGGACCCGGGCTG -1101
-1100 CGCGCTCTCGGGCCGAGGAACCAGGACGCGCCCGGAGCCTCGCACGCGGC -1051
-1050 CAAGCTCGGGGCGTCCCCTCCCCTCGGCCGGGCGAACTCAAGGGGCGCAG -1001
-1000 CTCTTTGCTTTGACAGAGCTGGCCGGCGGAGGCGTGCAGAGCGGCGAGCC -951
-950  GGCGAGCCAGGCTGAGAAACTCGAGCCGGGAACAAAGAGGGGTCGGACTG -901
-900  AGTGTGTGTGTCGGCTCGAGCTCCGGGCAGAGGCATTTGGGCCCGAGGCC -851
-850  CCCGCTGTGACTCCCCGAGACTCCGCAGTGCCCTCCACTGCGGAGTCCCC -801
-800  GCGCTTGCCGGCAAAAACTTTATTCTTGGCAAACTTCTCTTTCTCTTCCC -751
-750  CTCCTCCTCGGCCCCCATCTTCTGCTCCTCCTCCTTCTCTAGCAGATTAA -701
-700  ATGAGCCTCGAGAAGAAAAACCGAAGCGAAAGGGAAGAAAATAAGAAGAT -651
-650  CTAAAACGGACATCTCCAGCGTGGGTGGCTCCTTTTTCTTTTTCTTTTTT -601
-600  TCCCACCCTTCAGGAAGTGGACGTTTCGTTATCTTCTGATCCTTGCACCT -551
-550  TCTTTTGGGGCAAACGGGGCCCTTCTGCCCAGATCCCCTCTCTTTTCTCG -501
-500  GAAAACAAACTACTAAGTCGGCATCCGGGGTAACTACAGTGGAGAGGGTT -451
-450  TCCGCGGAGACGCGCCGCCGGACCCTCCTCTGCACTTTGGGGAGGCGTGC -401
-400  TCCCTCCAGAACCGGCGTTCTCCGCGCGCAAATCCCGGCGACGCGGGGTC -351
```

```
-350  GCGGGGTGGCCGCCGGGGCAGCCTCGTCTAGCGCGCGCCGCGCAGACGCC -301
-300  CCCGGAGTCGCCAGCTACCGCAGCCCTCGCCGCCCAGTGCCCTTCGGCCT -251
-250  CGGGGGCGGGCGCCTGCGTCGGTCTCCGCGAAGCGGGAAAGCGCGGCGGC -201
-200  CGCCGGGATTCGGGCGCCGCGGCAGCTGCTCCGGCTGCCGGCCGGCGGCC -151
-150  CCGCGCTCGCCCGCCCCGCTTCCGCCCGCTGTCCTGCTGCACGAACCCTT -101
-100  CCAACTCTCCTTTCCTCCCCCACCCTTGAGTTACCCCTCTGTCTTTCCTG -51
-50   CTGTTGCGCGGGTGCTCCCACAGCGGAGCGGAGATTACAGAGCCGCCGGG -1
+1    ATGCCCCAACTCTCCGGAGGAGGTGGCGGCGGCGGGGGGGACCCGGAACT +50
+51   CTGCGCCACGGACGAGATGATCCCCTTCAAGGACGAGGGCGATCCTCAGA +100
+101  AGGAAAAGATCTTCGCCGAGATCAGTCATCCCGAAGAGGAAGGCGATTTA +150
+151  GCTGACATCAAGTCTTCCTTGGTGAACGAGTCTGAAATCATCCCGGCCAG +200
+201  CAACGGACACGAGGTGAGCGGGCCGCTGCCCCAGTTCCGAGAGGCGTGGC +250
+251  GGGTCCCTGGGAAACTCAGAGAGGGAGGAAGGGAACGAAGGCGCCCGGCC +300
+301  GCCCGGCTGAGGGTCCCGAGTAGATCCCTCGTTTGTTTTGGGGTGTGTGC +350
+351  GTAGCTGGGCCCCCGGTCCGCAGTGTGTGTGTGTGTGTGTGTGTGTGTGT +400
+401  GTGTGTGTGTGTGTGTAGGATGGGAGGTAGAGGGGCAGGGTAACTTATTC +450
+451  CAAATGGATCAGTTTAGGGGGTTCGGTGGCTGAACGGTTGGTTCGCGGGC +500
+501  AGGAACCCACAATGTTTCGCCTTCGGTCTTTTTGTTGGGCTGAGGGGTGC +550
+551  CGTGCTTTAAAGCGAGTAGGGCTGAATAACGCTACTGATAAATTCTTCCG +600
+601  TGGGACGGAAAACAAAGTGCACACGCTATCTCTTGGTAGATGTCTCTGCA +650
+651  AAAGTGCGTTAAACCACCAAAGGCTTAGGTGGAGACGAGCAGAGCCACAG +700
+701  GCGGGAGGAAGGGGAGGCGGGCGTGGGCTGGGGCAGGCGGGGCCGTCCCT +750
+751  ACCACCGGTGAGCCGGGCCGCGGGGAGCAGGCCGGCCGATGGGGCGCGCA +800
+801  CCCGGGCTGCGGGGGCGGTGCGGGGTTCAGTCCCTGCGGCCCCAGCGCCT +850
+851  GGGCTAGCAGCTGGGAGCTGTGGGTGCGCCGAGCCGACCCACCGCAAGGG +900
+901  TACACCCTGGCTGTACCGCCCCCAACTCTAGTTTTCTCCAACCGCCATCA +950
+951  CTTGGAAACAAACCCACAACTTTTATTTGGGGTTTGGGACCGGATCAGGC +1000
```

A	*LEF1* mRNA Transcription Start Site	**-1189**
A	LEF1 Translational Start Site	**+1**
G	*LEF1* NAT Transcription Start Site	**+243**