UNIVERSITY OF TROMSØ UIT
FACULTY OF HEALTH SCIENCES
DEPARTMENT OF MEDICAL BIOLOGY
RNA AND TRANSCRIPTOMICS GROUP

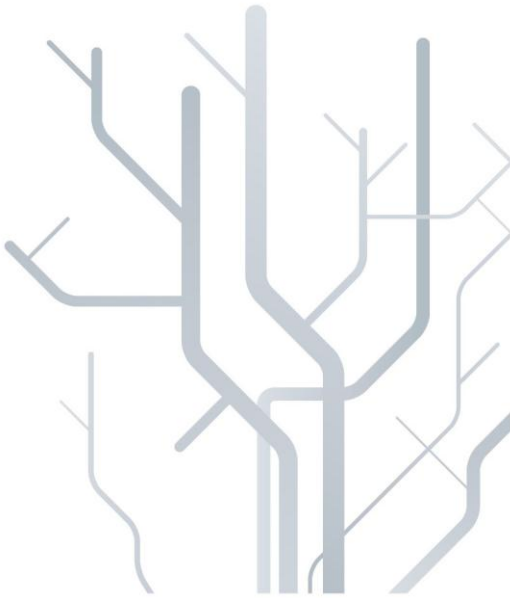
Establishing a method for RNA immunoprecipitation to study interactions between long non-coding RNAs and proteins

Elsebeth Sophie Brun

Thesis for the degree Master in Molecular Biotechnology

MBI-3941

December 2012



Acknowledgements

This work was performed at the RNA and Transcriptomics Research Group, Department of Medical Biology (IMB), Faculty of Health Science at the University of Tromsø from August 2012 to December 2012.

I would like to start by thanking my supervisor Maria Perander for her excellent guidance during the practical work, and for all her time spent guiding and helping me through the writing process of this thesis.

I also would like to thank Erik Knutsen for his guidance and support both in the laboratory, and in the writing process of this thesis. For always answering my questions, and sharing his knowledge with me.

Thanks to Professor Steinar Johansen for his excellent knowledge about RNA, and for giving me the opportunity to write this thesis.

I would like to thank all the members of the RNA and Transcriptomics Research Group for a good working environment, and special thanks to Anita Ursvik and Kari Haugli for helping me in the lab and giving me practical advices.

Finally, I would like to thank my friends and family for supporting me during this period, with coffee breaks and phone calls.

Elsebeth Sophie Brun

Tromsø, December 2012

Table of content

Summary

Long non-coding RNAs (lncRNAs) are involved in several important biological processes, and several lncRNAs have been associated with human diseases, particularly cancer. LncRNAs regulate gene expression at different levels. LncRNAs exert their functions through interactions with DNA, RNA or proteins. In this study, we established a method for investigating interactions between lncRNAs and protein complexes in human breast cancer cell lines. We used the well-known interaction between *HOTAIR* and the PRC2 member Suz12 to work out a RNA immunoprecipitation protocol in the lab, using an antibody targeting Suz12 to co-immunoprecipitate *HOTAIR* from nuclear lysates from the Hs578 breast cancer cells. We achieved a significant enrichment of *HOTAIR* in the immunoprecipitate from nuclear extracts from Hs578T. Differential expression of *HOTAIR* in the three breast cancer cell lines Hs578T, MDA-MB-231 and MCF-7 were also observed.

Introduction

Long non-coding RNA

The central dogma of gene expression is that genetic information is stored in protein coding genes; DNA is transcribed into messenger RNAs (mRNA) which functions as template for protein synthesis (1; 2). In this dogma, the proteins are the main protagonists of cellular functions while the RNA functions only as an intermediate between the DNA sequence and the protein that it encodes. Only 2% of the human genome encodes for proteins, but as much as 90% of the genome is being actively transcribed (3). The resent year's discovery of the extensive productions of transcripts that do not have any protein-coding potential provides new perspective to this central dogma (1). These transcripts are called noncoding RNAs (ncRNAs), and are divided into two groups based on their sizes; the small ncRNAs are shorter than 200 nucleotides, whereas the long non-coding RNAs (lncRNAs) have transcripts that range from 200 nucleotides up to 100 kilobases (2). Most lncRNAs are mRNA-like transcripts but lack significant open reading frames. Many lncRNAs are transcribed by RNA polymerase II, are 5'capped and 3' polyadenylated. However, many lncRNAs are also non-polyadenlyated transcripts, and different lncRNAs localize either in the cytoplasm or in the nucleus (4). According to their genomic localization, lncRNA can either be intergenic, intronic, antisense, overlapping or divergent relative to other lncRNAs or protein-coding genes (Figure 1).

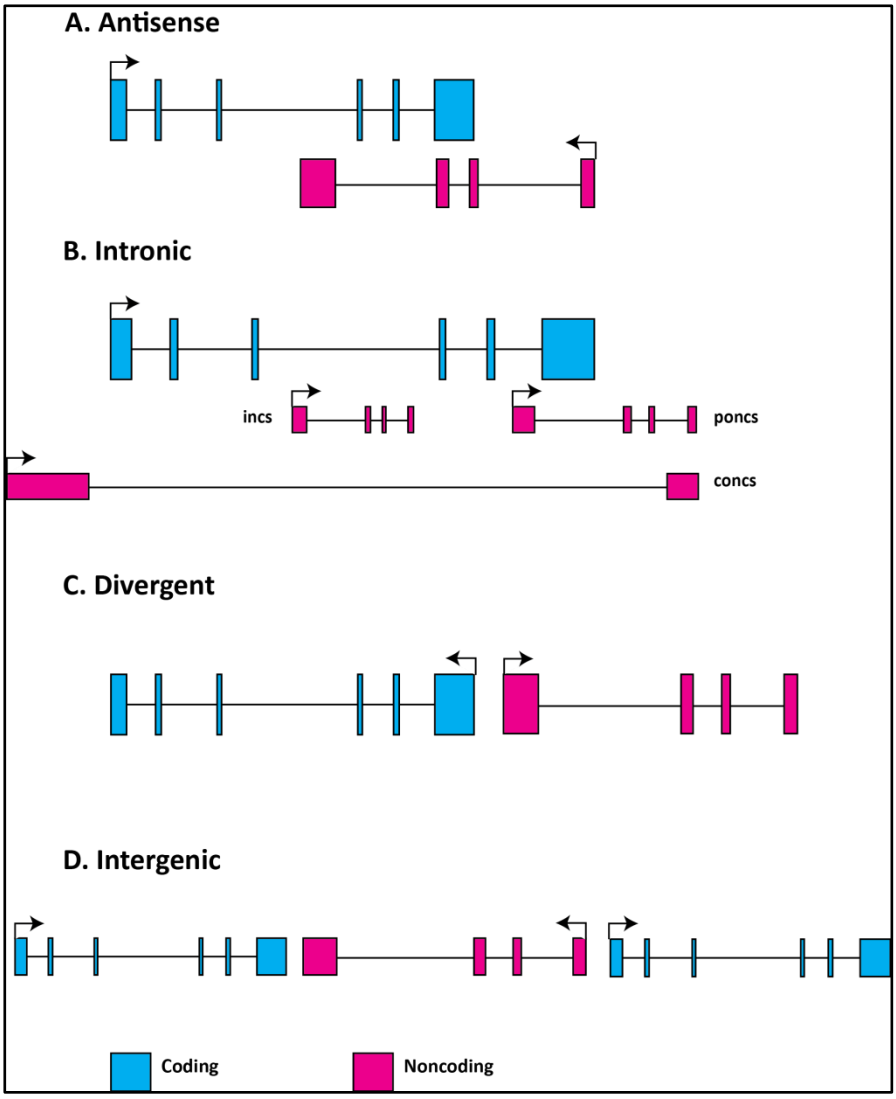


Figure 1. The different genomic locations of lncRNA loci. A. Antisense lncRNA is transcribed in the opposite direction of a protein coding gene and overlap with at least one of the exons of the protein coding gene. B. Intronic lncRNAs overlap with a protein coding gene. The overlap can be internal, partial, or completely resulting in, respectively, *incs*, *poncs* and *concs*. Intronic lncRNAs are located within the intron of a protein coding gene and terminates without any overlap with exons. C. Divergent lncRNAs is transcribed in the opposite direction of a nearby protein coding. D. Intergenic lncRNA (lincRNAs) is transcribed from loci localized between protein coding genes. Modified from (1; 5).

Biological functions of lncRNAs

LncRNAs participate in several biological processes such as chromosome dosage compensation, imprinting, cell cycle control, nucleo-cytoplasmic transport, protein synthesis and RNA maturation (2; 6). In general, lncRNAs exert their function by regulating the expression of other genes at different levels. The role of lncRNAs in the regulation of

transcription is most thoroughly studied. Even though some lncRNAs have been suggested to function as enhancers of transcription, many lncRNAs mediate gene silencing by inducing epigenetic modifications. LncRNAs can interact directly with DNA or RNA through base pairing, but through secondary and tertiary structures the lncRNAs can also interact with proteins. Determining the structural and functional domains of lncRNAs will lay the detailed foundation for a better mechanical understanding of lncRNAs (7).

An example of a lncRNA that is involved in epigenetic silencing is ANRIL (antisense lncRNA of the *INK4* locus). This locus encodes for three tumor suppressor genes that have been linked to various types of cancer. ANRIL mediates transcriptional silencing of the *INK4* locus in *cis* by interacting with the Pc/Chromobox 7 (CBX7) protein which is a member of the polycomb repressive complex 1 (PRC-1). The *INK4* locus is regulated by methylation of histone H3 at lysine 27 which is directed by the polycomb group proteins (8). Altered activity of ANRIL may result in dysregulated silencing of the *INK4* locus, and disruption of the interaction impacts the ability of CBX7 to repress the *INK4* locus and control senescence (8). Another and maybe the most prominent examples of lncRNA-mediated epigenetic control of chromatin is the silencing of the inactive X chromosome by the lncRNA *XIST*. Transcription of *XIST* from one of the female X chromosome is involved in recruiting polycomb group proteins that methylates histone H3 on lysine 27, which leads to silencing of the other X chromosome (9).

Another regulating lncRNA is the metastasis-associated in lung adenocarcinoma transcript, MALAT1. Studies have shown that MALAT-1 regulates alternative splicing by interacting with the SR (serin/arginine) rich family that regulates alternative splicing (10). LncRNAs can also be involved in translational control, one example is BACE1-AS (β-site amyloid precursor protein cleaving enzyme). *BASE1-AS* becomes elevated when exposed to cellular stress, and this increases *BASE1* mRNA stability and generation of amyloid-β 1-42 through a post-transcriptional feed-forward mechanism (11). LncRNAs can also be involved in regulation of apoptosis and cell cycle control. The two lncRNAs PANDA and lincRNA-p21 both respond to the binding of p53 to the *CDKN1A* locus after DNA damage. PANDA block apoptosis through the transcription factor NF-YA, and linc-p21 mediates gene silencing through recruitment of the protein hnRPK (12).

Some of the functions of the lncRNAs given above share the same molecular mechanism. Both PANDA and linc-p21 are induced after binding of p53, and functions as *signals* in response to the DNA damage that causes p53 to bind the *CDKN1A* locus. PANDA also functions in addition to being a signal, as a *decoy* by binding to the transcription factor NF-YA. This protein activates apoptosis, and the binding to PANDA results in promotion of cell survival (12). XIST functions as a *guide* by recruiting the PcG proteins and guiding the protein complex to the X chromosome that is to be inactivated. ANRIL functions as a *scaffold* by binding to components from both the PRC-1 complex and the PRC-2 complex (8; 13). The different mechanisms listed above are illustrated in Figure 2.

The examples above indicate the important roles of lncRNAs in several biological processes. Some of these lncRNAs have been linked to several human diseases. ANRIL has been shown to be up-regulated in prostate cancer (8), MALAT1 is over-expressed in several cancer types; lung, breast, pancreas, colon prostate and liver (14; 15), and concentrations of *BACE1*-AS are elevated in subjects with Alzheimer's disease (11). Large-scale transcriptomic analysis has been used to identify lncRNAs that are correlated with cancer. These results may lead to the discovery of lncRNAs that can function as bio-markers for cancer (14; 16; 17). The potential of lncRNAs as molecular markers can be important as diagnostic and prognostic tools. Already today, there are several lncRNAs that have been identified as prognostic parameters for patient survival (14; 17).

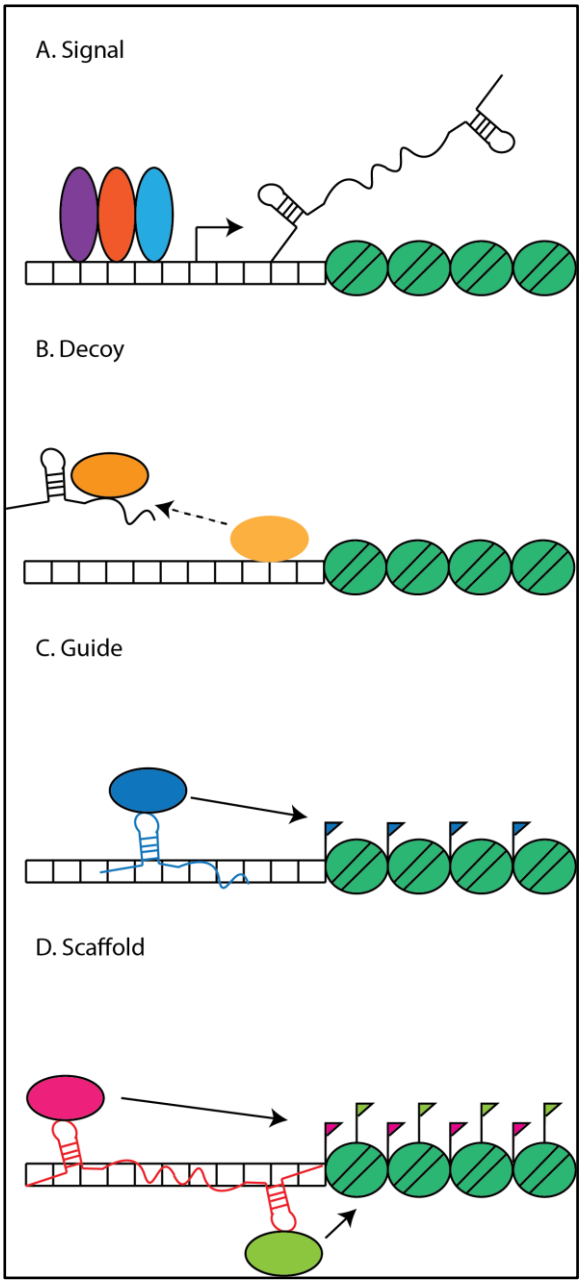


Figure 2. LncRNAs can function by several different molecular mechanisms. A. LncRNAs can function as *signals* where their expression can give information about gene regulation in space and time by the combination of transcription factors or signaling pathways. B. As *decoys*, the lncRNAs can bind proteins such as transcriptor factors and titrate them away from the chromatin. C. LncRNAs that function as *guides* can recruit chromosome modifying enzymes to their target genes. D. As *scaffolds*, the lncRNAs can bring together proteins that form ribonucleprotein complexes, in this figure shown as histone modifying proteins. One specific lncRNA may function by more than one of these molecular mechanisms. (Modified from (6)).

HOTAIR

As mentioned above, lncRNAs have diverse roles in both gene regulation and epigenetic control of chromatin. They have been found to be key regulators of different chromatin

states for several biological processes, such as imprinting, developmental gene expression and dosage compensation. Thousands of lincRNAs (long intergenic noncoding RNAs) have in recent studies been shown to be in association with specific chromatin modifications. One of these lincRNAs is the *HOX* Antisense Intergenic RNA, *HOTAIR* (18).

In mammals there are 39 *HOX* transcriptions factors that are clustered on four chromosomal loci, from *HOXA* to *HOXD*. These transcription factors are essential for specifying the positional identities of cells, and there is a complex epigenetic regulation of the maintenance of *HOX* expression patterns (18).

HOTAIR is transcribed in an antisense manner with the *HOXC* genes on chromosome 12, and it is a 2158 nucleotide long, polyadenlyated, and spliced transcript (18). The 5' end of human *HOTAIR* binds to and targets the PRC2 complex (Polycomb repressive complex 2) to the *HOXD* locus on chromosome 2. The PRC2 complex silences specific sites of genes at specific time points and has an important regulatory role in embryonic development. *HOTAIR* also binds to a second complex, the LSD1/CoREST/REST complex, through its 3'end. Studies with deletion mutants have demonstrated that nucleotides within the first 300 nucleotides of *HOTAIR* is necessary and sufficient to bind the PRC2 complex, while the nucleotides within the last 646 nucleotides (position 1500-2146) is necessary and sufficient to bind the LSD1 complex (19).

The findings of *HOTAIR* binding to both the PRC2 complex and the LSD1/CoREST/REST complex show that *HOTAIR* targets multiple chromatin modifying complexes and suggest that *HOTAIR* functions as a scaffold between PRC2 and LSD1/CoREST/REST, and that *HOTAIR*/PRC2/LSD1 can suppress gene expression via multiple mechanisms at the same time (6). In addition to acting as a scaffold for the two complexes, *HOTAIR* also acts as a guide by targeting PRC2 to its genomic location, showing the complexity of lncRNAs.

The PRC2 complex consists of the proteins EZH2, SUZ12, EED, and RbAP46/RbAP48. The EZH2 subunit is a histone H3 lysine 27 (H3K27) methyl transferase and trimethylation of H3K27 is associated with transcriptional silencing. The PRC2 complex has been found to be involved in both cancer progression and gene silencing (20; 21).

HOTAIR has been found to be highly upregulated in both primary and metastatic breast tumors, with up to a 2000-fold increased transcription level over normal breast tissue (17). The high levels of *HOTAIR* have also been found to be correlated with both metastasis and poor survival rate, which links *HOTAIR* to patient prognosis and cancer invasiveness.

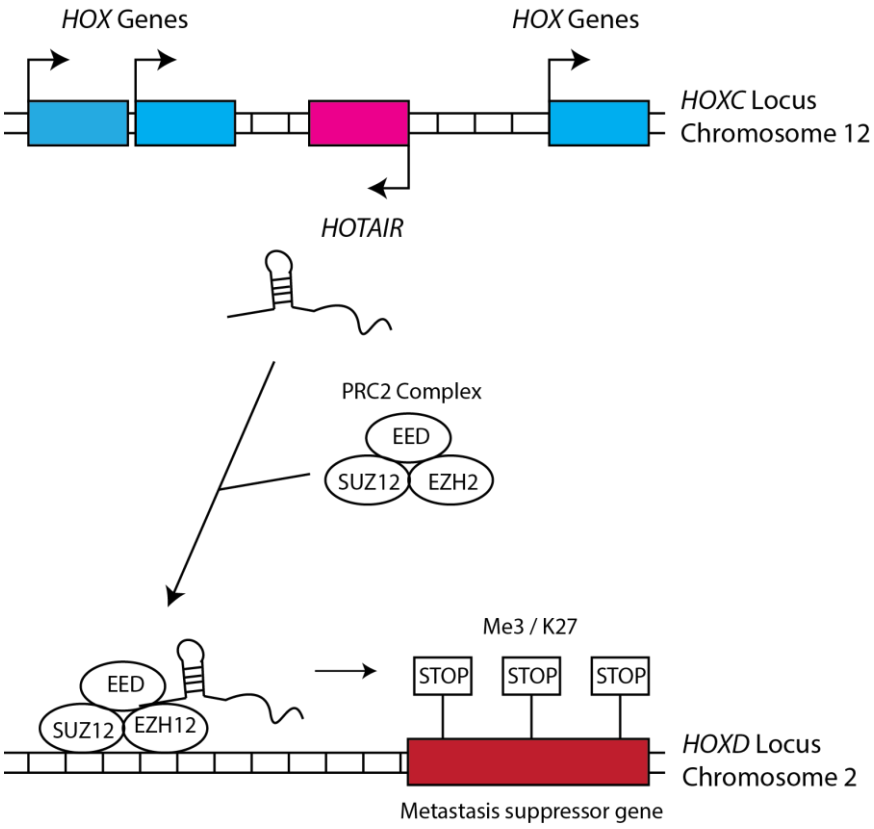


Figure 3. The proposed mechanism of *HOTAIR* mediated gene silencing, in *trans*, of the *HOXD* locus. *HOTAIR* is transcribed in an antisense, intergenic manner from the *HOXC* locus on chromosome 12, and results in a H3K27 methylation and gene repression on the *HOXD* locus on chromosome 2. (Modifed from (3; 22).

RNA immunoprecipitation

RNA can interact with proteins through their secondary or tertiary structure to create ribonucleoprotein complexes (RNPs). RNPs have instrumental functions in the cells regulating all aspects of gene expression. The regulatory and catalytic roles of RNPs in RNA processing and protein synthesis are well known. Regulation of gene expression by both small regulatory RNAs, like microRNAs, and lncRNAs involve RNA-protein interactions (23). RNA immunoprecipitation (RIP) is an antibody-based technique used to map RNA-protein interactions in vivo. Specific protein complexes can be immunoprecipitated from a cellular homogenate with an antibody raised against the protein of interest. Any RNA that is

associated with this protein complex will also be isolated and can be further analyzed by polymerase chain reaction-based methods, hybridization or sequencing (24).

Ribonucleoproteins can re-associate after cell-lysis, and therefore not give an accurate reflection of the association that occur in vivo (25). To prevent the re-assortment of proteins and RNA during cell lysis and to “freeze” the interactions in the cell, cross-linking agents can be used to fix all the interactions. UV-cross-linking is widely used to identify direct interactions between RNA and proteins. However, UV-cross-linking is not reversible. Cross-linking agents that are reversible may be more beneficial for subsequent characterization of the associated molecules (23). One of the reversible cross-linking agents is formaldehyde, which is the smallest aldehyde possible. Formaldehyde has the ability to rapidly preserve cellular complexes in their native state, and to penetrate the cell membrane in a rapid manner. These are the qualities that have led to its application in methods such as RNA immunoprecipitation (26).

In a RIP assay, as shown in Figure 4, involve either cross-linked or non-cross-linked RNA-protein complexes from living cells. These complexes are isolated using a specific antibody towards the protein of interest. After reversing the cross-links, the interacting RNA can be analyzed by reverse transcription, followed by a polymerase chain-reaction. The RIP method also has some limitations; the result is highly dependent on the antibody used and the abundance of the target ribonucleoprotein. Furthermore, RNA molecules are known to be “sticky” and unspecific binding represents a challenge (23).

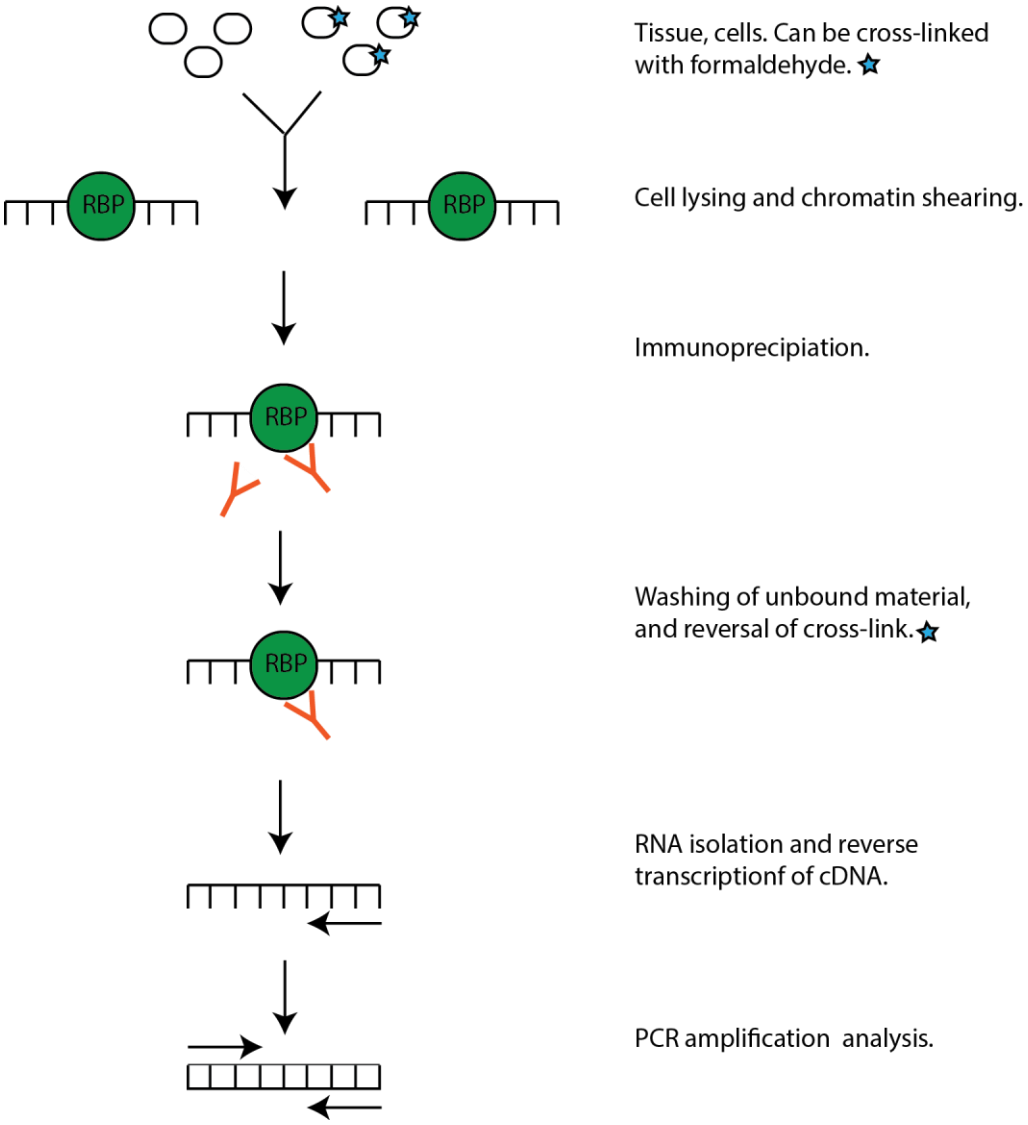


Figure 4. Schematic representation and summary of a RNA immunoprecipitation assay. The RNA binding proteins (RBPs) are immunoprecipitated with antibodies, followed by isolation of co-immunoprecipitated RNA and PCR amplification analysis. Modified from (23; 26).

Aim

The aim of this study was to establish a method in the lab for investigating interactions between lncRNAs and proteins in human breast cancer cell lines. To accomplish this, the well-known interaction between the lncRNA *HOTAIR* and the PRC2 member Suz12 was studied.

Materials and methods

Cell lines

In this study the cell lines MDA-MB-231, Hs578T and MCF-7 were used. Both MDA-MB-231 and Hs578T are triple-negative Basal-B breast cancer cell lines that lack progesterone receptor (PR), estrogen receptor (ER), and Her2 (human epidermal growth factor 2). The triple-negative cell lines are generally more aggressive than other breast cancer cell lines (27). MCF-7 is a luminal epithelial-like breast cancer cell line that is less aggressive than MDA-MB-231 and Hs578T (28).

All three cell lines were cultured and sub-cultured according to the recommendations given by the American Type Culture Collection (ATCC). Both MDA-MB-231 (ATCC® Number: HTB-26) and MCF-7 (ATCC® Number: HTB-22) cells were maintained in high glucose Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 1% penicillin streptomycin, and 1% L-Gluthamin. The Hs578T (ATCC® Number: HTB-126) cells were maintained in high glucose DMEM supplemented with 10% FBS, 1% penicillin streptomycin, 1% L-Gluthamin and 0.01 mg/ml bovine insulin. The medium was renewed two times per week with a sub-cultivation ratio of 1:4 for all the three cell lines. The three cell lines were incubated at 37°C under 5% CO_2.

Total RNA isolation

Total RNA was isolated using the TRIzol® protocol[1] from Invitrogen™ with modifications by the RNA and Transcriptomics group at the University of Tromsø (see below). TRIzol® is a monophasic solution of phenol and guanidine isothiocyanate that was used to disrupt cells and dissolving cell components while maintaining the integrity of the RNA during sample homogenization. RNA was separated into an aqueous layer by addition of chloroform, and the RNA was precipitated from the aqueous layer with isopropanol. To remove impurities from the RNA, the RNA was washed with ethanol.

Total RNA was isolated from two 15 cm^2 culture dishes where the cells had grown to a ~90% confluence. The cells were washed two times with ice-cold PBS buffer before they were lysed by adding 1.0 ml TRIzol followed by scraping of the cells. The lysate was transferred to a clean 1.5 ml microcentrifuge tube and incubated for 5 min at room temperature for a

[1] *Cat. No: 15596-026*

complete dissolution of nucleoprotein complexes. 0.2 x volumes of chloroform was added to the lysate and mixed well. The sample was incubated for 20 min on ice and centrifuged at 9 000 rpm for 20 min at 4°C. The aqueous phase was transferred to a clean microcentrifuge tube before 1.0 x volume of isopropanol was added. The tube was incubated overnight at -20°C and centrifuged at 15 000 x g for 30 min at 4°C. The supernatant was removed, and the RNA pellet was washed with 1.0 ml ice-cold 80% EtOH, followed by centrifugation at full speed for 5 min. The EtOH was removed and the RNA pellet was air-dried. The isolated RNA was resuspended in RNase free water, and stored at -70°C. The total RNA was DNase treated with the "Heat&Run gDNA Removal Kit" from Arcitzymes, according to the manufacturer's instruction.

Qubit

Nucleic acids can be quantified using spectrophotometry by measuring the absorbance at 260 nm. This is done by passing ultraviolet light at 260 nm through the sample in a spectrophotometer. Nucleic acids absorb light at 260 nm, and the higher the concentration of nucleic acid is in a sample; the higher the absorbance of light will occur. This method does not allow for discrimination between DNA and RNA present in the same sample, so in this study the Quant-iT™ Technology[2] from Invitrogen™ was used. The Qubit® Fluorometer uses Quant-iT™ dye that binds specifically to either RNA or DNA, and the Flurorometer detects the absorbance of the dye, which increases upon binding to nucleic acids.The concentration of the samples were measured according to the manufacturer's instructions.

Agilent Bioanalyzer

The Agilent 2100 Bioanalyzer[3] is a platform that can be used for quantification, sizing and quality control of DNA, RNA, proteins, and cells. This is done by loading samples onto a chip that contains micro channels that are filled with fluorescent dye and a gel like matrix. By an electrophoretic separation, the sample move through the gel. The Agilent Bioanalyzer valuates the ratio between the 18S and the 28S ribosomal subunits, as well as the presence of degraded short fragments. This is used to calculate a RIN value (RNA integrity value), which is an evaluation of the intactness of the RNA sample. An intact total RNA sample

[2] http://www.invitrogen.com/site/us/en/home/brands/Product-Brand/Quant-iT.html

[3] http://www.genomics.agilent.com/CollectionOverview.aspx?PageType=Application&SubPageType=ApplicationOverview&PageID=275

would have a RIN value of 10. In this study the integrity of total RNA samples was analyzed by the use of the Total RNA 6000 Nano Chip, and the concentration of the isolated RNA from the RNA immunoprecipitation was quantified by the use of RNA 6000 Pico Chip. All the assays were performed according to the manufacturer's instruction.

Reverse transcription polymerase chain reaction (RT-PCR)

The total RNA from the RNA isolation was reverse transcribed into its complementary DNA (cDNA) by using the Transcriptor First Strand cDNA Synthesis Kit from Roche. The cDNA was transcribed using both random hexamer primers and anchored-oligo(dT)$_{18}$ primers.

Table 1. The components and concentrations of the denaturation of template-primer step.

Component	Concentration
Total RNA	500 ng
Anchored-oligo(dt)$_{18}$ primer	2.5 µM
Random hexamer primer	60 µM
Water	-
Total volume	13

The tubes were incubated at 65°C for 10 min to denature the template-primer mixture and immediately placed on ice after the incubation.

The following components were added to the sample:

Table 2. The components of the RT-PCR reaction and their concentration.

Component	Concentration
Transcriptor Reverse Transcriptase Buffer, 5X	1x
Protector RNase Inhibitor	20 U
dNTP mix	1 mM each
Transcriptor Reverse Transcriptase	10 U
Total volume	20

The sample was incubated for 10 min at 25°C followed by 30 min at 55°C. The Transcriptor Reverse Transcricptase was inactivated by heating the sample at 85°C for 5 min before the sample was placed on ice. cDNA was stored at -20°C.

Polymerace chain reaction (PCR)

The cDNA transcribed from the total RNA was further amplified in a polymerase chain reaction. This is thermal cycling technique that is used for direct amplification of specific regions of DNA. One amplification cycle includes three steps; a denaturation step where double stranded DNA is separated, a hybridization step where primers hybridize to the target sequence, and an elongation step where a DNA polymerase replicates the single stranded DNA from the 3'end of the annealed primers to the complementary sequence. The number of cycles used depends on the concentration of the product of interests within the sample. The different polymerases used for PCR amplification in this study were the LA Taq™ DNA polymerase from TaKaRa and AmpliTaq® DNA polymerase from Applied Biosystems. Both DNA polymerases have 5'→3' exonuclease activity but only LaTaq™ DNA polymerase has 3'→5' exonuclease activity which is used for proofreading.

The conditions for the PCR reactions were determined by the product, primers, and DNA polymerase used in the reaction.

Table 3. The PCR conditions used in this study where the different condition of each step is dependent on PCR-product, primers, and DNA polymerase used.

Thermal-cycling conditions			
Step	**Temperature**	**Duration**	**Cycles**
Initialization step	94°C	2 min	Hold
Denaturation	94°C	30 sec	Hold
Annealing	X1	X2	-
Elongation	Y1	Y2	
Final elongation	Z1	Z2	
Final hold	4°C	∞	Hold

- X1+X2: The temperature and duration of the annealing step is dependent on the Tm of the primers, and the PCR-product. Annealing conditions used in this study:
    - Primers for HOTAIR: 30 sec at 55°C.
    - Primers for housekeeping genes: 30 sec at 60°C.
- Y1+Y2: The temperature and the duration of the elongation step is dependent on the polymerase used in the reaction, elongation conditions used in this study were;

- AmpliTaq® DNA polymerase: 1 min at 72°C.
- LaTaq™ DNA polymerase: 30 sec at 68°C.

- Z1+Z2: The conditions for the final elongation step is, in addition to the first elongation step, dependent on the DNA polymerases used:
  - AmpliTaq® DNA polymerase: 10 min at 72°C.
  - LaTaq™ DNA polymerase: 7 min at 72°C.

Primers for *HOTAIR* were designed by using primer3 (frodo.wi.mit.edu). Primer3 calculates the melting point of the primer (Tm) and the GC content of the primer sequence. The Tm should not be much higher than 60°C, and the GC content should not be higher than 50%, and with most of the GC's in the 3' end. The length of the product should be between 200-800 bp, and the length of the primer should be between 22-33 nt.

To evaluate if the primers would result in a match in the human genome, UCSC In-Silico PCR was used to blast the sequences. (genome.ucsc.edu).

Table 4. The primers used in this study, their oligonucleotide sequences and their sizes.

Name	Oligonucleotide primer Sequence	Size in nucleotides
OP1822	CGAAAGCTTCCACAGTGAGGACTG	24
OP1823	GCTTCCTTGCTCTTCTTATCATC	23
OP1824	CAACCACGAAGCTAGAGAGAGAG	23
OP1825	GCATCTTGAGACACATGGGTAAC	23
OP1838	GGTGAAAAAGCAACCACGAAGC	23
OP1839	ACATAAACCTCTGTCTGTGAGTGCC	25
GAPDH[4]	Cat. No: QT01192646	-
18S rRNA4	Cat no: QT00199367	-
β-actin4	Cat no: QT01680476	-

Quantitative real time PCR

In contrast to traditional PCR, the quantitative real time PCR (qPCR) is both qualitative and quantitative. The principle of qPCR is the same as for the traditional PCR, but the amplified DNA is detected in real time (between each round of amplification) instead of a detection of the amplified DNA only at the end of the reaction. The first papers on qPCR were published

[4] These primers for the house keeping genes were for QuantiTect® Primer Assay by Qiagen for use with SYBR®Green-based qPCR.

in 1996, and this method avoided the problem of end point analysis by traditional PCR (29; 30).

In this study the SYBR®Green Select Master Mix from Applied Biosystems® was used according to the manufacturer's instructions.

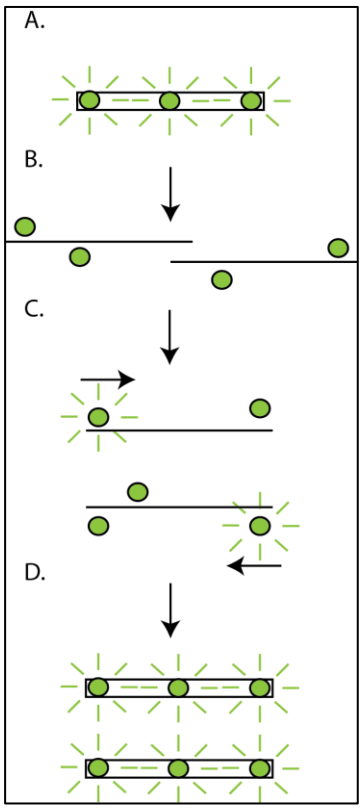


Figure 5. The mechanism for the SYBR®Green Dye I from Applied Biosystems®. A. In the beginning of the reaction, the SYBR®Green Dye I is fluorescent when it is bound to all double stranded DNA. B. During denaturation, the dye is released, and the fluorescence is reduced. C. Forward and reverse primers anneals during the annealing, and the PCR product is generated. D. When the amplification is complete, the dye will again bind to the double stranded DNA and the increase in fluorescence can be detected. (Modified from the Invitrogen™ web page[5])

For the qPCR reaction a 96-well plate was used, and 20 µl aliquots of reactions mix were pipetted into each well. Three technical replicates of cDNA were made to avoid technical variation.

http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/PCR/real-time-pcr/qpcr-education/taqman-assays-vs-sybr-green-dye-for-qpcr.html

[5] http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/PCR/real-time-pcr/qpcr-education/taqman-assays-vs-sybr-green-dye-for-qpcr.html

Table 5. Set up for q-PCR reaction.

Component	20 µl reaction
SYBR® Select Master Mix 2X	10 µl
Forward and reverse primers	Final concentration: 0,25 µM each.
cDNA	25 ng
Water	-
Total volume	20 µl

Table 6. The thermal-cycling conditions used in the qPCR reaction where the annealing temperature is based on the Tm of the primers (≥60°C).

Thermal-cycling conditions			
Step	**Temperature**	**Duration**	**Cycles**
UDG Activation	50°C	2 min	Hold
AmpliTaq® DNA Polymerase, UP Activation	90°C	2 min	Hold
Denature	95°C	15 sec	40
Anneal/Extend	60°C	2 min	

After the 40 cycles the instrument was also programmed to perform a default dissociation step.

The threshold was manually set with the SDS Plate Utility Software, and every reaction was checked to have an amplification curve that had reached log-phase amplification. Raw data was adjusted for technical variations and background noise. cDNA with Cq value greater than 37.5 was set to 37.5 and reported as not detected. The data was normalized by the mean expression value (ΔCq), and the expression fold change was calculated (-ΔΔCq).

Sanger sequencing

Sanger sequencing is a technology that uses fluorescent dideoxynucleotides (ddNTPs) in the amplification. The ddNTPs are modified deoxynucleotides that lack the hydroxyl group at the 3'end. This results in truncated DNA fragments because the DNA polymerase ceases the extension of DNA when a ddNTP is incorporated since the hydroxyl group is required to form the phosphodiester bond between two nucleotides. The four ddNTPS have each a different fluorescent color tag and the base sequence can be read by a fluorescent scanner when the fragments of different sizes wanders through a gel. PCR products were purified using the

QIAQuick® PCR Purification Kit from Qiagen. The purified PCR products were sequenced from both ends by using the BigDye® Terminator v3.1 Cycle Sequencing Kit from Applied Biosystems. Cycle parameters used:

96°C	1 min	
96°C	10 sec	35 cycles.
50°C	5 sec	
60°C	4 min	

Preparation of whole cell extract

Cells were grown to a 90% confluence in cell culture dishes. The cells were placed on ice, and washed two times with ice-cold PBS. MKK lysis buffer was added to the cell culture dishes, followed by scraping of the cells and transfer to microcentrifuge tubes. The samples were incubated on ice for 20 min before centrifugation at 13 000 rpm for 15 min at 4°C. After the centrifugation, the supernatant containing the whole cell extract was transferred to a clean tube and stored at -70°C.

MKK Lysis Buffer:
20 mM Tris HCl pH 7.0
1% Triton X-100
5 mM NaPPi
50 mM NaF
1 mM EDTA
1 mM EGTA
1 mM VO_4^{2-}
0.27 M Sucrose
10 mM β-Glycerophosphat

Sodium dodecyl polyacrylamide gel electrophoresis

Sodium dodecyl polyacrylamide gel electrophoresis (SDS-PAGE) is a method used to separate proteins according to their ability to move within an electrical current according to their molecular weight. SDS is an anionic detergent that denatures the proteins, and gives the proteins a negative charge that is evenly distributed according to the proteins' mass. This distribution of charge results in the size fractionation during the electrophoresis. In this study, the PowerEase® 500 Power Supply and the XCell *SureLock*™ Mini-Cell systems from Invitrogen™ were used. The gels used in the electrophoresis were the NuPAGE® Bis-Tris gels with a 4-12% gradient from Life Science. In short, cell lysates for SDS-PAGE were added 4x LDS Sample buffer and 10x Sample Reducing to a final concentration of 1x for both the

components. Samples were heated at 70°C for 10 min before loaded to the gel. Gels were run for 40 min with a constant voltage of 200V.

Western blot

The proteins that have been separated during an SDS-PAGE can be analyzed with western blot, which is a method used to detect the proteins of interest. Proteins are transferred to a nitrocellulose membrane where they are immobilized. The proteins can be detected by using antibodies that target the proteins of interest, followed by secondary antibodies containing an infra read dye. In this study the XCell II™ Blot Module from Invitrogen™ was used to perform the western blot. Western blots were performed to identify the presence of Suz12 in the MDA-MB-231, Hs578T and MCF-7 cell lines and in the samples from RNA immunoprecipitation. A control blot was also performed with antibody for actin.

Antibodies used:

Table 7. The antibodies used to detect the proteins of interest.

Antibody	Concentration	Dilution	Company
Anti-Actin, clone C4	-	1:1000	Millipore
Anti-SUZ12	0.90 mg/ml	1:1000	Abcam
Anti-p62	1 mg/ml	1:1000	-
IRDye® Goat (polyclonal) Anti-Rabbit IgG (H+L)	1.0 mg/ml	1:10 000	LI-COR
IRDye® 800CW Conjugated Goat (polyclonal) Anti-Mouse IgG (H+L)	1.0 mg/ml	1:7000	LI-COR

Transfer buffer:

20% MetOH

0.05 M TrisBase

0.4 M Glycine

TBS-T:

20% TBS

0.1% Tween®- 20

TBS:

0.03 M KCl

1.4 M NaCl

0.04 M Tris-HCl pH = 7.5

Cross-linking of an antibody to magnetic protein A/G beads

100 µl of Magna ChIP™ Protein A/G Magnetic Beads from Millipore™ were washed 3 times with 500 µl MKK Lysis Buffer. 30 µg of antibody was added to the magnetic beads after resuspending the beads in 200 µl MKK Lysis Buffer. The solution with antibody and magnetic beads were incubated at room temperature for 1 hour with rotation, followed by 3 washings steps with 500 µl MKK Lysis Buffer and one washing step with 500 µl PBS. 83 µl 5mM[6] BS^3 was added to the solution and incubated for 30 min at room temperature with rotation. The cross-linking was quenched by addition of Tris-HCl pH 7.5 to a final concentration of 50 mM and incubation for 15 min at room temperature with rotation. The beads were washed 3 times with 500 µl MKK Lysis Buffer before addition of 500 µl 0.1 M Glycin pH 2.5 followed by incubation at room temperature for 10 min with rotation. Finally, the beads were washed 3 times with 500 µl MKK Lysis Buffer and stored in 67 µl PBS at 4°C.

RNA Immunoprecipitation

As mentioned in the Introduction, RIP is a method for studying RNA-protein interactions in vivo. In this study the interaction between the Suz12 protein of the PRC2 complex and the lncRNA *HOTAIR* was of interest.

The method used in this study was modified from Abcam® web page[7] and V.A Moran et.al (31). The RIP was performed both with and without cross-linking of the cells.

Cell harvesting

Cells were grown to 90% confluence in three 175 cm^2 culture flasks and trypsiniated by first removing the media followed by one wash with PBS. 2.0 ml of trypsin was added, and the reaction was quenched by adding 8 ml of complete medium. The cells from the three flasks were collected in one 50 ml tube before centrifugation at 500 x g for 10 min. The cell pellet was washed by resuspending the pellet in 1.0 ml ice-cold PBS and transferred to a 15 ml tube, followed by one additional wash with 10.0 ml ice-cold PBS. The cells were collected by pelleting by centrifugation at 500 x g for 5 min.

Nuclei isolation and nuclear pellets lysis

The cell pellet was resupended in 2.0 ml PBS, 2.0 ml freshly prepared nuclear isolation buffer and 6.0 ml water and kept on ice for 20 min with frequent mixing. The nuclei were pelleted

[6] The 2 mg BS^3 in the No-Weigh Format from Thermo Scientific were solved in 700 µl conjugation buffer containing 20 mM NaH_2PO_4 and 0.15 M NaCl.
[7] http://www.abcam.com/index.html?pageconfig=resource&rid=14913

by centrifugation at 2 500 x g for 15 min. The supernatant was stored at -70°C, and the nuclear pellet was resuspended in 1.0 ml of freshly prepared RIP buffer. A dounce homogenizer was used for mechanically shearing (20 strokes). Nuclear membranes and debris were pelleted by centrifugation at 13 000 rpm for 10 min, and the nuclear lysates were split in two fractions; one for Mock with IgG and for IP with Suz12. At this stage a small fraction of the input was taken out for SDS-PAEGE and western blot analysis.

Preparation of beads

50 µl of protein A/G magnetic beads was used per reaction. The beads were placed in a magnetic rack before the supernatant was removed. 500 µl of RIP buffer was used to wash the beads, and then the tubes were briefly vortexed and quick spun. The tube was placed in a magnetic rack where the RIP buffer was removed, followed by two more washing steps performed in the same manner. RIP buffer was added to the beads to make a 50% slurry, where 20 µl was used to pre-clear the nuclear lysate to avoid unspecific binding: The slurry was incubated with the nuclear extract for 1 h at 4°C with rotation. The supernatant was transferred to a new microcentrifuge tube by using the magnetic rack.

RNA immunoprecipitation

5 µg of antibodies (Suz12 or IgG) were incubated with the nuclear extracts overnight at 4°C with rotation before 25 µl of the bead-slurry was added followed by incubation for another 2 hours at 4°C with rotation.

The supernatant was removed by placing the tube in the magnetic rack before the beads were resuspended in 500 µl RIP buffer, this step was repeated three times, followed by two washing steps with 500 µl PBS. In the last washing step it was taken out sample for SDS-PAGE and Western Blot analysis.

Purification of RNA that was bound to immunoprecipitated RBP

For isolation of co-precipitated RNAs, the magnetic protein A/G p beads were resuspended in 1.0 ml TRIzol®. The RNA was isolated by the same approach as described above. The isolated RNA was resuspended in 7 µl RNase free H_2O.

Nuclear isolation buffer:

1.28 M Sucrose

40 mM Tris-HCl pH 7.5

20 mM $MgCl_2$

4% Triton X-100

RIP Buffer:

150 mM KCl

25 mM Tris pH 7.5

5 mM EDTA

0.5 mM DTT

0.5% NP40RNase inhibitor [8]

Protease inhibitor[9]

Cross-linking with formaldehyde

The RNA-protein interactions between HOTAIR and the PRC2 complex in the cells were "frozen" by using formaldehyde as the reversible cross-linker.

Cells were grown to 90% confluence in three 175 cm^2 culture flasks and trypsiniated as described earlier. The cells were pooled in one 50 ml tube and centrifuged at 500 x g for 10 min. The cell pellet was resuspended in 1.0 ml ice-cold PBS and transferred to a 15 ml tube before the cells were washed with 10 ml ice-cold PBS. The cells were collected by centrifugation at 500 x g for 5 min. 37% formaldehyde was added to a final concentration of 0.3% and incubated for 10 min with gentle rotation at room temperature. The cross-linking reaction was quenched by adding 1.25 M glycine to a final concentration of 0.125 M and incubated for 5 min at room temperature, followed by to washing steps with 10 ml ice-cold PBS as described above.

Reversing the formaldehyde cross-links.

The magnetic protein A/G beads were resuspended in 100 µl Buffer C, followed by adding 10 µg proteinase K and incubated at 42°C for 30 min for proteinase K digestion. The formaldehyde cross-links were reversed by incubation at 65°C with rotation for 4 h.

[8] Added fresh before use, (100 U/ml)
[9] Added fresh before use, 1 tablet solved in 10 ml RIP buffer.

Buffer C (for reversing the crosslink):

150 mM NaCl

50 mM Tris-HCl (pH=7.5)

5 mM EDTA

10 mM DTT

1% SDS

Transfection

Transfection is a method used to introduce nucleic acids into cells. Lipofectamine was used in this study to transfect siRNA into the cells. Lipofectamine is cationic liposome formulation that binds to negatively charged nucleic acids, and then fuses through the cell membrane transporting the nucleic acids into the cell. siRNA (silencing RNA) targets and interferes with the expression of genes that have a complementary sequence to the siRNA.

Lipofectamine® 2000 Reagent from Invitrogen™ was used to transfect siRNA targeting Suz12 into the cells. In short, cells were grown to a 70% confluence before the media was replaced with Opti-MEM®. SiRNA and Lipofectamin® were diluted in Opti-MEM® before added to the cells. Media was changed after 4 hours, and the cells were further analyzed after 24 hours. The experiment was performed according to the manufacturer's instructions. Transfections with scramble RNA and siRNA targeting p62 were used as controls.

Table 8. The siRNA targeting SUZ12 were provided from Ambion® by Life Technologies ™.

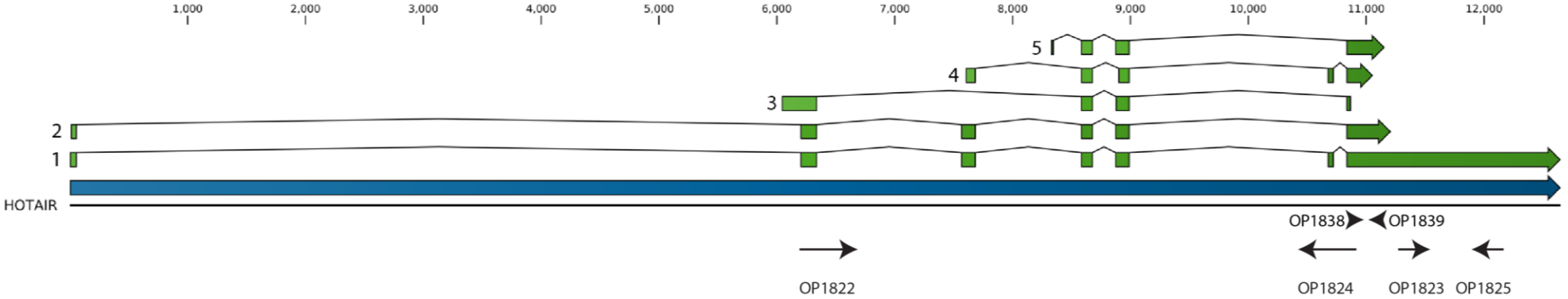
	Sequnce (5'→3')	
siRNA	Sense	Antisense
S23967	GGAUGUAAGUUGUCCAAUAtt	UAUUGGACAACUUACAUCCtt
S23968	GGACCUACGUUGCAGUUCAtt	UGAACUGCAACGUAGGUCCct
S23969	GGAUAGAUGUUUCUAUCAAtt	UUGAUAGAAACAUCUAUCCta

Results

The main purpose of this study was to establish a protocol for analyzing the interaction between lncRNAs and proteins in breast cancer cell lines by using RNA immunoprecipitation (RIP). To establish this method in the lab, we chose to study the well-known interaction between *HOTAIR* and the Suz12 member of the PRC2 complex. For all the figures in the results section, the raw data including complete images of PCR-products resolved on agarose gels and western blot analyses can be found in the Appendix.

Expression of *HOTAIR* in the breast cancer cell lines MDA-MB-231, MCF-7 and Hs578T

In order to analyze the expression of *HOTAIR* in the breast cancer cell lines, cDNA was synthesized from total RNA isolated from MDA-MB-231, MCF-7 and Hs578T cells. Before cDNA synthesis, the quality and concentration of the RNA were determined by the Aglient 2100 Bioanalyzer and Quant-IT™ technology (Figure A 1 and Table A 1). In general, the quality of the RNA isolations using the TRIzol protocol with modifications from the RNA and transcriptomics group is very high, with RIN values (RNA integrity number) close to 10. The cDNA was subjected to PCR analyses using three different primer sets complimentary to different regions of *HOTAIR*. OP1822/1824 was expected to result in a PCR product at 504 bp, OP1823/OP1825 to result in a 765 bp PCR product, and the OP1838/OP1839 was expected to result in a 122 bp PCR product. For verification, the PCR-products obtained by primer sets OP1822/OP1824 and OP1823/OP1825 were sequenced by Sanger Sequencing, and the sequences were mapped to *HOTAIR* (Figure 6). *HOTAIR* consist of seven different exons, and the OP1822/OP1824 primer set expanded six of these, raising the opportunity to detect alternative splice products. This primer set resulted in two bands, which were found by Sanger sequencing to represent two different splicing variants of *HOTAIR* where one of the PCR products contained one more exon (501 bp) than the other (451 bp).



Size of PCR products:

OP1822 + OP1824 = 504/451 bp

OP1823 + OP1825 = 765 bp

OP1838 + OP1839 = 122 bp

Mapping of Sanger sequenced primers:

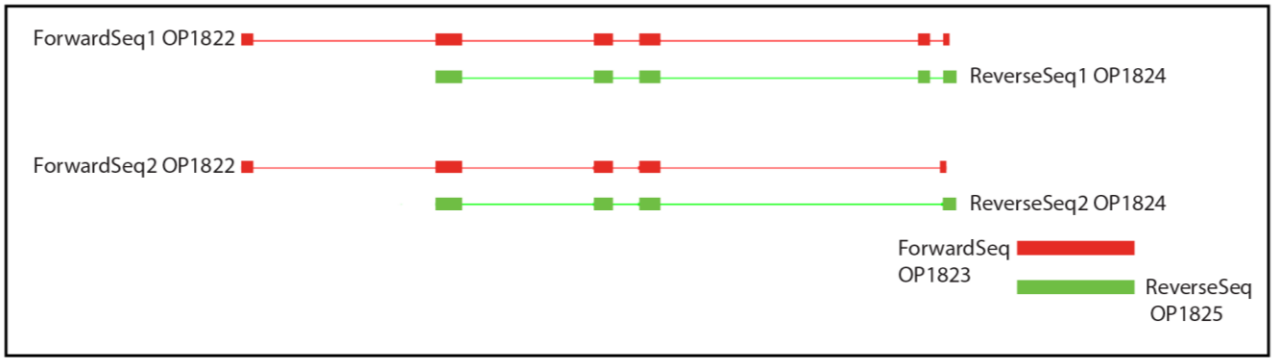


Figure 6. For the annotation of *HOTAIR* in the Ensemble database, there are a total of five different splice variants of *HOTAIR*. All the different variants are shown in the figure, marked 1-5. The blue arrow indicates the gene, while the green arrows indicate the exons included in the transcript. The locations of the primers are illustrated by arrows. The sequences obtained from Sanger sequencing were mapped to *HOTAIR*. Red lines illustrate sequences with forward primers, while the green lines illustrations sequences with reverse primers. The dense line indicates the PCR product, while the weaker line indicates the gaps between the exons. The two sequencing reactions of OP1822/1824 resulted in two different PCR products. The first sequencing resulted in a product consisting of one more exon than the other (504 bp and 451 bp). The complete sequence of the primer sets and their consensus sequence can be found in the Appendix (p. 56-57).

HOTAIR expression was detected in the MCF-7 and Hs578T cell lines. In contrast, no expression was seen in MDA-MB-231 cells. The expression of three housekeeping genes, coding for Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), β-actin, and 18S rRNA, in the cell lines were determined as control (Figure 7). Both the splicing variants mentioned above are shown in the figure. The intensity of the bands indicates the expression level of the 451 bp slicing variant is higher than that of the 504 bp splicing variant, seen as a weaker band over the 451 bp band.

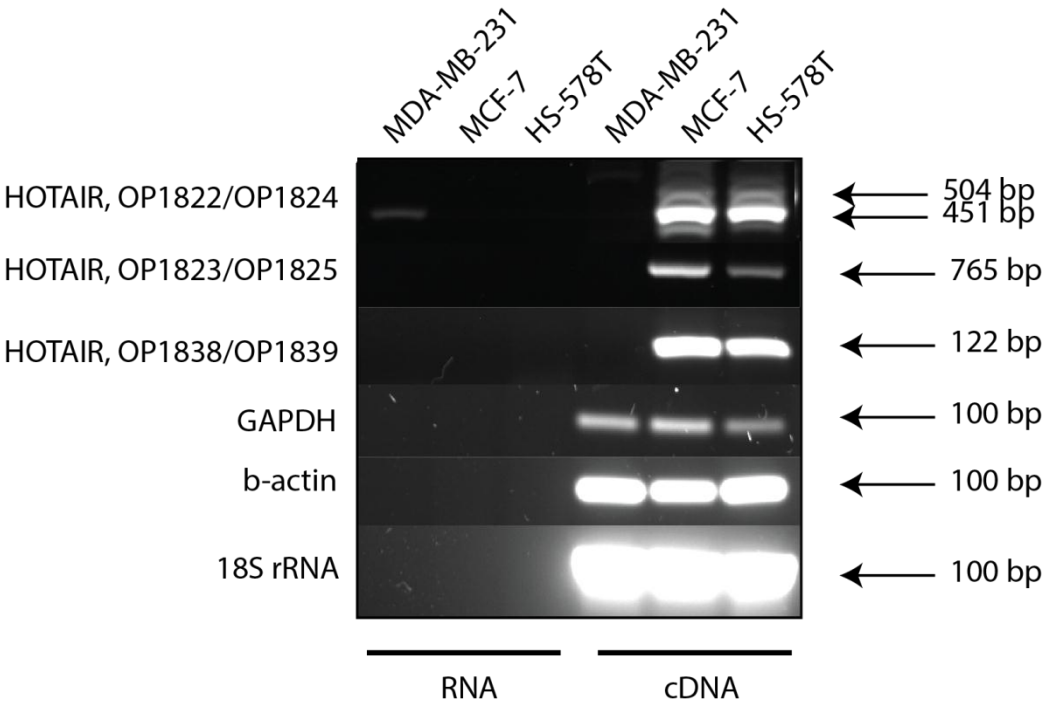


Figure 7. *HOTAIR* was detected in MCF-7 and Hs578T, but not in MDA-MB-231. The different use of primers resulted in different band sizes according to the size of the PCR product. The house-keeping genes GAPDH, β-actin, and 18S rRNA were all present in the three cell lines. These primer sets resulted in bands at around 100 bp. Negative control was included using total RNA.

The traditional PCR results described above are only qualitative, so quantitative real-time PCR (qPCR) was performed in addition. Both cDNA synthesis and qPCR reactions were performed in triplicates. Analysis of the qPCR results indicated that the expression level of *HOTAIR* is higher in the MCF-7 cell line than in the Hs578T cell line. Again, no expression of *HOTAIR* was observed in MDA-MB-231. This correlates with the results from the traditional PCR results shown above.

Additional information can be found in the raw data from the qPCR in the Appendix (Figure A 6 and Table A 2).

Expression of Suz12 protein in the breast cancer cell lines

In order to analyze the expression of the Suz12 protein in the three breast cancer cell lines, western blot analyses of whole cell extracts (WCE) using an antibody raised against Suz12 was performed. Suz12 was detected in all the three cell lines. Equal protein loading was verified by probing membranes with anti-actin antibody.

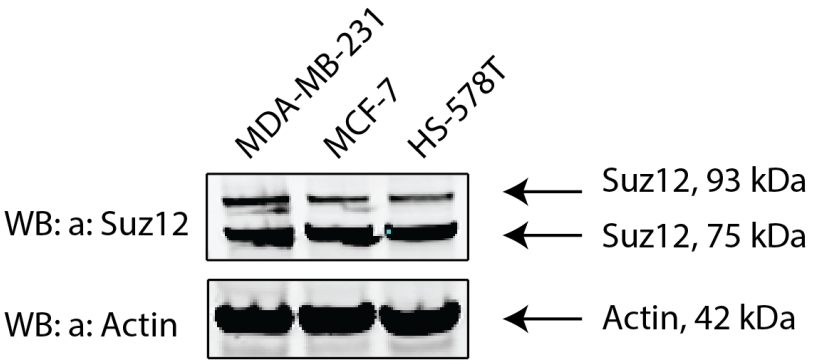


Figure 8. Suz12 was identified in the three cell lines MDA-MB-231, MCF-7, and Hs578T by western blot analysis. The antibody used resulted in to distinct bands, one at 93 kDa and one at 75 kDa.

The lncRNA *HOTAIR* and the protein Suz12 were detected in the breast cancer cell lines MCF-7 and Hs578T. In line with ongoing experiments in the lab, Hs578T was chosen for further experiments with immunoprecipitation and RNA immunoprecipitation.

RNA immunoprecipitation of the lncRNA *HOTAIR* and the protein Suz12 from the breast cancer cell line Hs578T.

As demonstrated above, expression of Suz12 was detected in Hs578T cells. To determine if the antibody for Suz12 was suitable for immunoprecipitation (IP), whole cell extracts were incubated with either anti-Suz12 antibody or purified IgG as control. Immunoprecipitation was also performed using anti-Suz12 antibody or IgG that had been cross-linked to magnetic protein A/G beads. This was done because the same antibody was used in both IP and western blot analysis. A secondary antibody will detect antibodies produced in the same organism, resulting in detection of the heavy chain of IgG. This can potentially interfere with the specific signals of

interest obtained by a secondary antibody. Suz12 was detected by IP in Hs578T. Cross-linking of the antibody to the magnetic protein A/G beads led to a substantially decrease in the intensity of the unspecific heavy chain signal. However, this method did not result in a better immunoprecipitation of Suz2. Therefore, in future RNA immunoprecipitation experiments, no cross-linking step of antibodies to protein A/G beads was included.

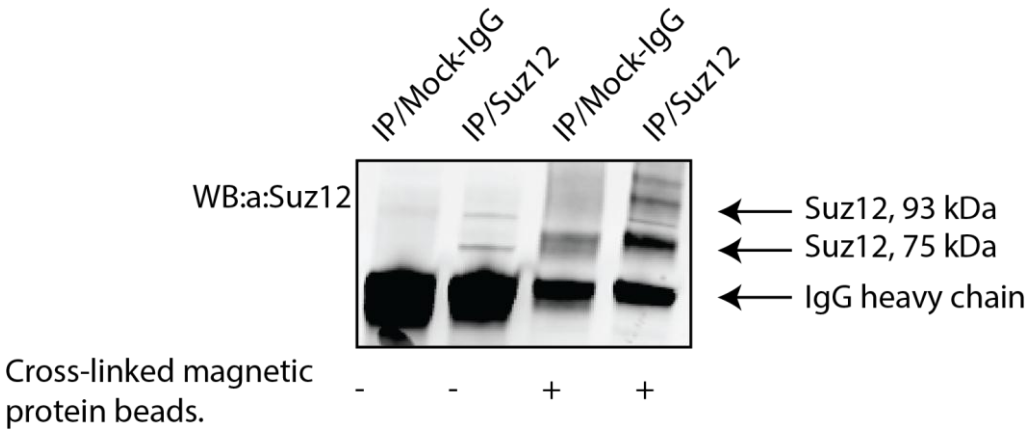


Figure 9. A western blot analysis confirmed the presence of Suz12 after immunoprecipitation from the breast cancer cell line Hs578T. The experiment was performed with or without magnetic proteins beads cross-linked with the antibody against Suz12.

The protein Suz12 was detected by immunoprecipitation from whole cell extract from the breast cancer cell line Hs578T. Since Suz12 is located in the nucleus, the RNA immunoprecipitation was performed from nuclear extracts from Hs578T to improve the yield of the protein, and to decrease the amount of unspecific binding of "sticky" RNA that might be present in the cytoplasm. The method was performed with or without cross-linking the cells with formaldehyde. The cross-linking increases the specificity of the method, and allows for higher stringency during washing steps to decrease unspecific binding. Suz12 was detected as a 93 kDa band in the input from cells that had not been cross-linked, and as a band at 75 kDa in the input from cells that had been cross-linked. However, Suz12 was not detected after the RNA immunoprecipitation in either cross-linked or not cross-linked nuclear extracts (Figure 10).

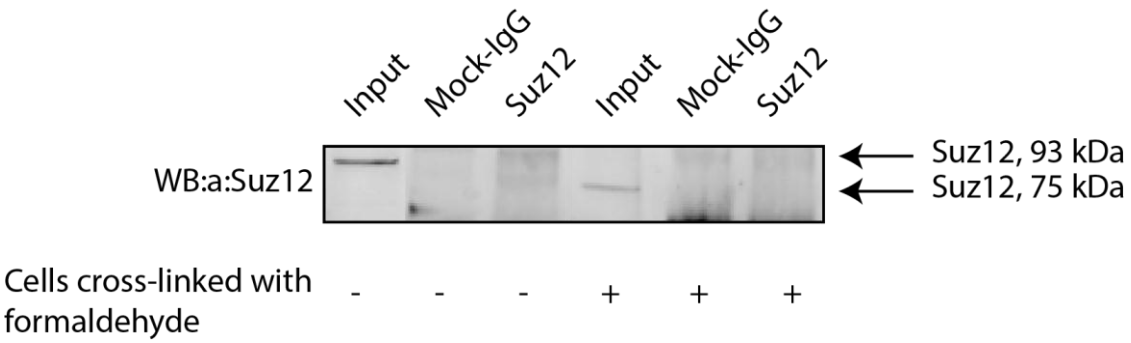
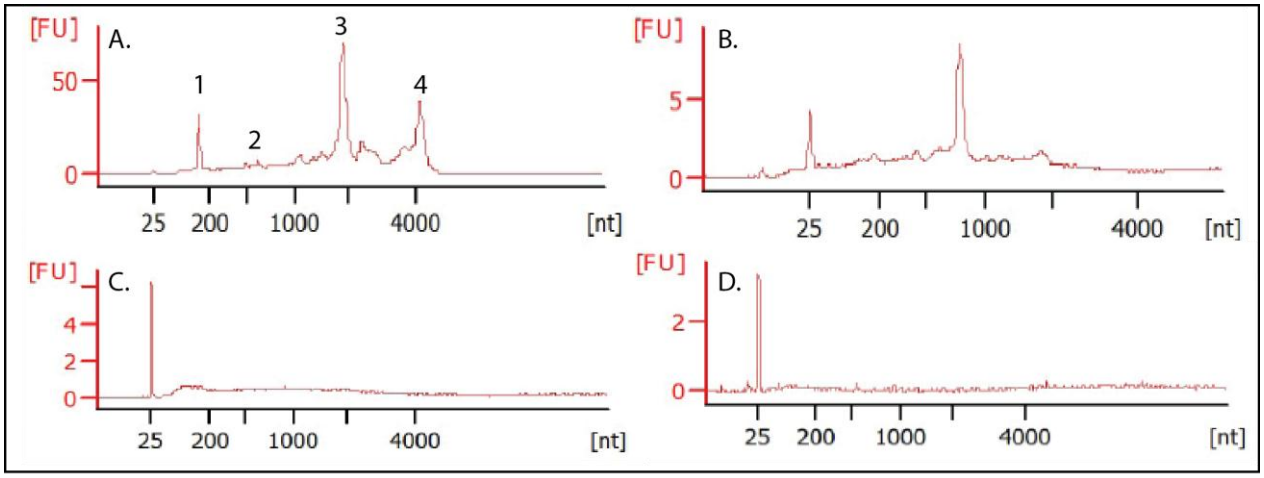


Figure 10. Suz12 is present in the nuclear lysates before RNA immunoprecipitation, but could not be identified after the RIP. Suz12 was detected at 93 kDa in nuclear extract from cells that had not been cross-linked; while the protein was detected at 75 kDa in nuclear extract from cells that been cross-linked with formaldehyde.

To obtain qualitative and quantitative information about the RNA co-immunoprecipitated with anti-Suz12 from the cell line Hs578T, the RNA was analyzed and evaluated using the Agilent 2100 Bioanalyzer. The fragment size distribution of RNA present in the samples is shown in Figure 11, and the RNA concentrations are shown in Table 9. By cross-linking the cells, the amount of RNA isolated from the nuclear extracts by RIP is lower than the amount of RNA isolated from nuclear extracts from cells that had not been cross-linked. Comparing RNA isolation from RNA immunoprecipitation with the purified anti-IgG to the RIP with anti-Suz12 shows that the amount of RNA decreases with use of anti-Suz12 for both cross-linked and not cross-linked cells. Due to the high amounts of ribosomal RNA and other RNA species present in the samples, the Agilent was able to calculate RIN values by the ration between the 18S and 28S ribosomal units present in the samples that had not been cross-linked. The presences of these two rRNA species are easily detected in Figure 11 as two large peaks. This indicates binding of "sticky" RNA to the IgG, and illustrates the problem mentioned above with unspecific binding of "sticky" RNA.

Table 9. The RNA concentration of isolated RNA by RNA immunoprecipitation from nuclear extracts from Hs578T.

RNA sample	Concentration (pg/µl)
IgG, not cross-linked	1498
Suz12, not cross-linked	174
IgG, cross-linked	59
Suz12, cross-linked	17



1. Marker
2. tRNA
3. 18S
4. 28S

Figure 11. Agilent results from the RNA immunoprecipitation with Suz12. A. The Agilent results show a substantial amount of RNA present in the sample from RIP with anti-IgG from nuclear extracts that had not been cross-linked. B. From the RIP with anti-Suz12 from nuclear extracts that had not been cross-linked, there is less RNA than with the anti-IgG. C+D. Very low amounts of RNA were detected after cross-linking the cells with formaldehyde.

In order to analyze the expression of *HOTAIR* in the nuclear extracts from Hs578T after the RIP, cDNA was synthesized from the isolated RNA from the RIP. The RNA was tested for the presence of genomic DNA, and did not give any up-amplification of any PCR product (Figure A 5). *HOTAIR* was detected in the nuclear lysate of Hs578T cells that had not been cross-linked. The strength of the bands indicates a higher amount of *HOTAIR* in the RIP from the nuclear lysate by using

anti-Suz12, illustrating that there was achieved a significant enrichment of co-immunoprecipitated *HOTAIR.* However, there was no detection of *HOTAIR* in the nuclear extracts from Hs578T cells that had been cross-linked. The use of primer sets for the housekeeping genes Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and 18S rRNA confirmed the suspicion of unspecific binding of RNA. Both GAPDH and 18S rRNA were detected in all the samples of nuclear extract from cells that had not been cross-linked. Neither *HOTAIR* nor the housekeeping mRNAs were detected in the immunoprecipitates from formaldehyde-treated cells, except from a very faint 18S rRNA band in the Suz12 IP (Figure 12).

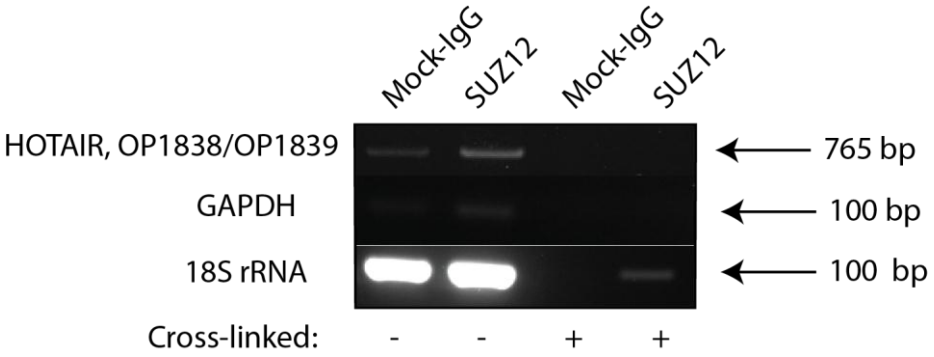


Figure 12. *HOTAIR* was only detected in the samples that had not been cross-linked, but is present in both the control with purified anti-IgG, as well as the sample with anti-SUZ12. The PCR reaction with house-keeping primers revealed that both GAPDH and 18S rRNA were present in the samples that had not been cross-linked with formaldehyde.

As mentioned above, traditional PCR analysis does not give any quantitative information about amplification, but only end-time information. A qPCR analysis was performed to see if an enrichment of *HOTAIR* in the nuclear cell lysate from Hs578T using anti-Suz12 in the RIP was achieved. This was done by calculating the fold change between the cross-linked and the not cross-linked samples from normalized Cq-values, setting the fold change value from the RIP with anti-IgG to 1. The RIP with anti-Suz12 from nuclear lysate from cells that had not been cross-linked showed a significant enrichment of *HOTAIR*, and all though 18S rRNA and GAPDH were detected in Suz12 immunoprecpitates, the enrichment of *HOTAIR* is clearly more substantial than that of the housekeeping genes (Figure 13). This illustrates the interaction between

HOTAIR and Suz12, and how *HOTAIR* can be achieved by co-immunoprecipitating it with Suz12. There was no significant enrichment of *HOTAIR* in the nuclear extracts from cross-linked cells.

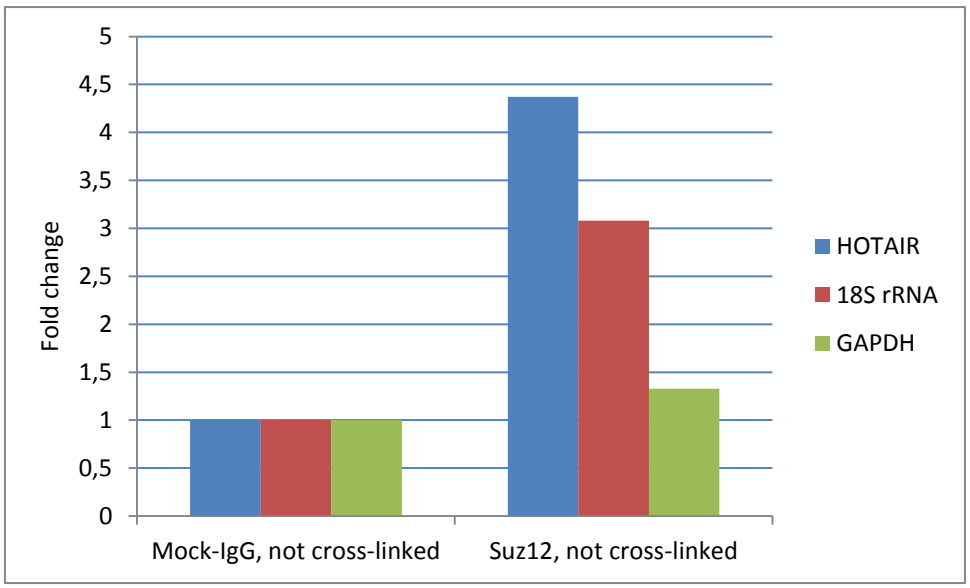


Figure 13. The fold changes between samples containing nuclear extracts from Hs578T cells that had not been cross-linked. *HOTAIR* is enriched by using anti-Suz12 in the RIP.

Transfection of siRNA targeting Suz12 in the breast cancer cell line Hs578T.

To evaluate the specificity of the interaction between *HOTAIR* and Suz12, the Hs578T cell line was transfected with three different siRNAs against the *suz12* mRNA, or scramble RNA as a control. The use of one specific siRNA did not result in silencing of Suz12, independent of the concentration (varied from 5 nM to 50 nM). A titration with the three different siRNAs, S23967, S23968 and S23969, showed that a combination of the three siRNAs with a concentration at 20 nM was most efficient for silencing Suz12. The silencing was detected by a decrease in the 93 kDa band of Suz12, while the 75 kDa band was unaffected by the transfection with siRNA, (Figure 14). Equal protein loading was verified by probing membranes with anti-actin antibody.

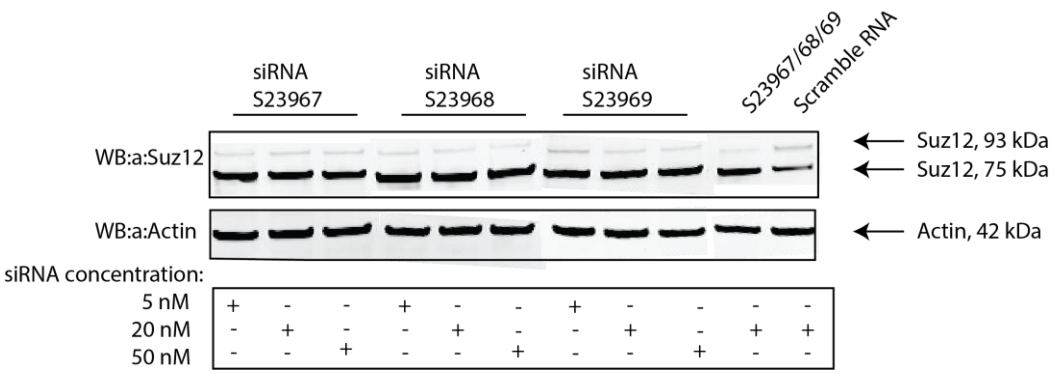


Figure 14. Titration of siRNA targetic Suz12 showed that the combination of all the three siRNAs (S23967, S23968 and S23969) in a concentration of 20 nM were most effective to silence Suz12. This is shown by a decrease in strength of the 93 kDa Suz12 band.

Since Suz12 was not completely removed by the specific siRNA-transfections in Hs578T cells, the experiment was repeated also including MCF-7 cells that have previously been used in successful transfections. A siRNA robustly knocking down the expression of the p62 protein (Eva Sjøttem, personal communication) was included to evaluate the transfection efficiency. The expression of both Suz12 and p62 was reduced in MCF-7 and Hs578T cells following siRNA transfections, but not completely abolished, indicating a low transfection efficiency of the transfection method used in this study (Figure 15A and B).

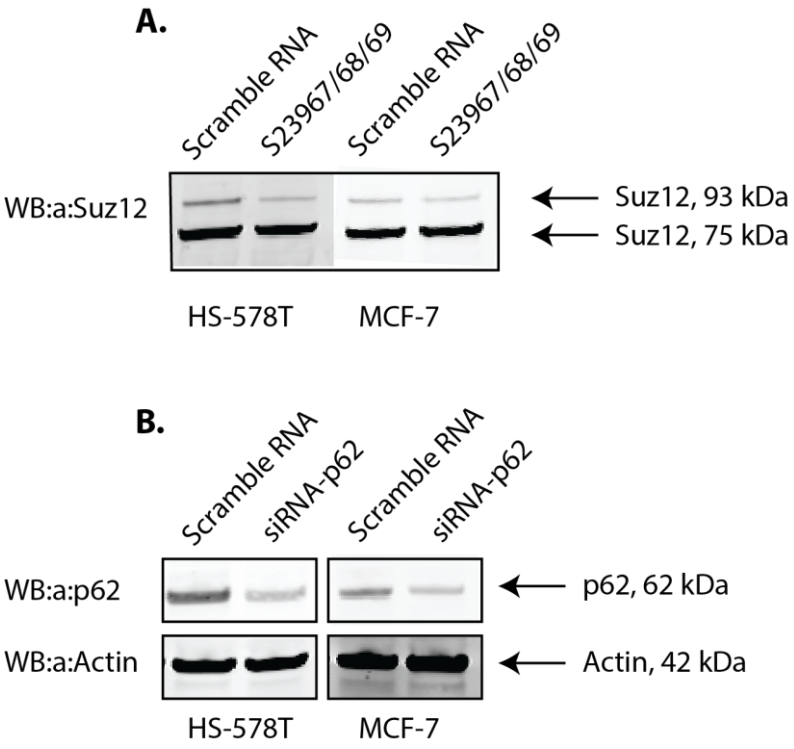


Figure 15. Transfection efficiency between the cell lines Hs578T and MCF-7 were compared by comparing the transfection with siRNA for both Suz12, A, and a siRNA targeting p62, B. The concentration of siRNA targeting p62 was 20 nM, and the concentration of siRNA targeting Suz12 was 3 x 20 nM.

Discussion

The functional roles of long non-coding RNAs (lncRNAs) in human diseases have become clearer over the past few years. Several studies have revealed lncRNAs that are up-regulated in cancer (8; 14; 15; 17). One of these lncRNAs is the *HOX* Antisense Intergenic RNA *HOTAIR*. *HOTAIR* binds to the polycomb repressive complex 2, leading to redirection of the complex to specific genes and epigenetic silencing through trimethylation of H3K27 (19). Previous studies have shown that the expression of *HOTAIR* is highly up-regulated in metastatic breast tumors, and correlated with metastasis and poor survival rate (17). In this study, the interaction between *HOTAIR* and the Suz12 member of the PRC2 complex was of interest. The aim of the study was to establish a protocol for RNA immunoprecipitation in the lab. The expression of *HOTAIR* and Suz12 was investigated in three breast cancer cell lines, MDA-MB-231, MCF-7 and Hs578T.

Three different primer sets to detect *HOTAIR* were used in the three breast cancer cell lines. One of the primer sets resulted in two bands, which were cut out from the gel and analyzed by Sanger sequencing. The bands turned out to be two different splicing variants in the cell line Hs578T (Figure 6). *HOTAIR* was detected in Hs578T and MCF-7 cell lines, but not in MDA-MB-231 (Figure 7). Hs578T and MDA-MB-231 are triple negative Basal-B-like breast cancer cell line, while MCF-7 is a luminal epithelial-like breast cancer cell line (27). Triple negative breast cancer cell lines are generally more aggressive than others, and especially MDA-MB-231 is associated with cancer metastasis (17). The finding of *HOTAIR* not being expressed in MDA-MB-231 illustrates that *HOTAIR* is not just associated with Basal-B breast cancer cell lines, but can be expressed in luminal breast cancer cell lines as well. Previous studies have shown that the expression level of *HOTAIR* is significantly lower in cell lines compared to what is observed in either primary or metastatic breast tumors (17).

The PRC2 member, Suz12, was detected in all the three breast cancer cell lines (Figure 8). Western blot analysis of Suz12 resulted in two bands, one at 75 kDa and one at 93 kDa. The manufacturer informed that the 75 kDa was the Suz12 protein, while there was some uncertainty around the 93 kDa band[10]. Several studies have indicated that the 93 kDa band in

[10] *Anti-SUZ12 antibody - ChIP Grade. Cat. No: ab12073*

fact is the actual Suz12 protein, while the lower band is a Suz12 degradation product (32) (33). These studies also indicated that the protein degradation is a product from preparing whole cell lysates, since the lower band is not present in nuclear lysates. Our results confirm these indications; both the 93 kDa band and the 75 kDa band were detected by western blot analysis in whole cell lysate, while only the 93 kDa band was detected in a nuclear lysate (Figure 8 and Figure 10). However, our siRNA transfection using three siRNAs targeting Suz12 only reduced the expression of the 93 kDa band (Figure 14). Therefore, we strongly believe that the 75 kDa is not a Suz12 degradation product, but an entirely different protein, probably cytoplasmic, that cross-reacts with the antibody for Suz12.

In our study, the RIP was performed using nuclear extracts from cells that had either been left un-treated or treated with the cross-linking agent formaldehyde. There was a substantial amount of unspecific binding of "sticky" ribosomal RNA in Suz12 immunoprecipitates from un-treated cells that could be seen in the Agilent analysis (Figure 11). However, we achieved a significant enrichment of *HOTAIR* in the Suz12-IPs compared to GAPDH and 18S rRNA. Cross-linking the cells resulted in very small amounts of RNA co-immunoprecipitated with Suz12, and *HOTAIR* was not detected. This could indicate that the RNA was degraded at certain steps in the protocol. The reversal process of the cross-linking was performed at a high temperature (65°C) for rather a long period (4 hours), which could have led to RNA degradation. Previous studies have indicated that an incubation times as short as 0.5 hour to 1 hour are sufficient for reversal of RNA-protein interactions (34). Therefore, in future studies, a more gentle reversal step should be employed.

Another explanation can be that there is an interaction between the DNA and the *HOTAIR*-Suz12 complex. The fragmentation method used in our study might not be good enough for a proper fragmentation of the DNA, and the *HOTAIR*-Suz12 complex could be pelleted together with the DNA in a centrifugation step. For a better fragmentation of the DNA, sonication might be more proper than the dounce homogenizer used in this study. This possibility was strengthened by the fact that the 93 kDa band of Suz12 was not present in the input of nuclear lysates from cross-linked cells, only the "mysterious" 75 kDa protein. The 93 kDa Suz12 might simply be associated with DNA which is lost in the preparation.

In this project we detected *HOTAIR* in MCF-7 and Hs578T, and Suz12 in all the three cell lines. We indeed got a significant enrichment of *HOTAIR* by using anti-Suz12 in RIP in Hs578T cells that were un-treated with formaldehyde even though we experienced the problem with unspecific binding of "sticky" RNA. Our cross-linking protocol surely needs further optimization. Following immunoprecipitation of Suz12, we were not able to detect the protein by immunoblotting. This could be solved by using a more sensitive Suz12-antibody, preferably one that is raised in another organism than the one used in the IP.

We were not able to achieve a complete knock-down of Suz12 by transfection with siRNA targeting Suz12, and optimizing the transfection efficiency would be important to further investigate the specificity of the interaction between *HOTAIR* and Suz12. A successful knockdown of Suz12 could be used in a RIP to investigate if *HOTAIR* still would be co-immunoprecipitated when Suz12 is silenced, or if the silencing of Suz12 would result in no co-immunoprecipitated *HOTAIR.* The specificity of the interaction between *HOTAIR* and Suz12 would be illustrated if there was no enrichment of *HOTAIR* by co-immunoprecipitation with silenced Suz12 in siRNA-transfected cells.

To further investigate the interaction between *HOTAIR* and Suz12, the RNA co-immunoprecipitated with Suz12 could be analyzed by deep-sequencing technology to examine the expression of *HOTAIR* in more detail. The protocol could also be applied on projects aimed at studying other lncRNAs and protein complexes.

References

1. **Rinn, L John and Chang, Howard Y.** Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry.* 81, 2012, pp. 145-166.

2. **Wapinski, Orly and Chang, Howard Y.** Long noncoding RNAs and human disease. *Trends in Cell Biology.* 21 (6) , 2011, pp. 354-361.

3. **Gibb, Ewan A, Brown, Carolyn J and Lam, Wan L.** The functional role of long non-coding RNA in human carcinomas. *Molecular Cancer.* 10 (38) , 2011, pp. 1-17.

4. **Reis, Eduardo M and Verjovski-Almeida, Sergio.** Perspectives of long non-coding RNAs in cancer diagnostics. *Frontiers in Genetics.* 3 (32), 2012, pp. 1-8.

5. **Pauli, Andrea, et al.** Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Research.* 22, 2012, pp. 577-591.

6. **Wang, Kevin C and Howard, Chang Y.** Molecular Mechanisms of Long Noncoding RNAs. *Molecular Cell.* 43 (6), 2011, pp. 904-914.

7. **Novikova, Irina V, Hennelly, Scott P and Sanbonmatsu, Karissa Y.** Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Research.* 40 (11), 2012, pp. 5034-5051.

8. **Yap, Kyoko 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 (5), 2010, pp. 662-674.

9. **Plath, Kathrin, et al.** Role of Histone H3 Lysine 27 Methylation in X Inactivation. *Science.* 300 (5616), 2003, pp. 131-135.

10. **Bernard, Delphine, et al.** A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *The EMBO Journal.* 29, 2010, pp. 3082-3093.

11. **Faghihi, Mohammad Ali, et al.** Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β-secretase. *Nature Medicine.* 14 (7), 2008, pp. 723-730.

12. **Hung, Tiffany, et al.** Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nature Genetics.* 43 (7), 2011, pp. 621-629.

13. **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 (16), 2011, pp. 1956-1962.

14. **Ji, Ping, et al.** MALAT-1, a novel noncoding RNA, and thymosin β4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene.* 22 (39), 2003, pp. 8031-8041.

15. **Lin, R, et al.** A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. *Oncogene.* 26 (6), 2007, pp. 851-858.

16. **Bussemakers, Marion JG, et al.** DD3: A New Prostate-specific Gene, Highly Overexpressed in Prostate Cancer. *Cancer Research.* 59 (23), 1999, pp. 5975-5979.

17. **Gupta, Rajnish A, et al.** Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature.* 464 (7291), 2010, pp. 1071-1076.

18. **Rinn, John L, et al.** Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell.* 129 (7), 2007, pp. 1311-1323.

19. **Tsai, Miao-Chih, et al.** Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science.* 329 (5992), 2010, pp. 689-693.

20. **Sparmann, Anke and van Lohuizen, Maarten.** Polycomb silencers control cell fate, development and cancer. *Nature Reviews. Cancer.* 6 (11), 2006, pp. 846-856.

21. **Bracken, Adrian P, et al.** Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes & Development.* 20 (9), 2006, pp. 1123-1136.

22. **Prensner, John R and Chinnaiyan, Arul M.** The Emergence of lncRNAs in Cancer Biology. *Cancer Discovery.* 1 (5), 2011, pp. 391-407.

23. **Niranjanakumari, Somashe, et al.** Reversible cross-linking combined with immunoprecipitation to study RNA–protein interactions in vivo. *Elsevier. Methods. .* 26 (2), 2002, pp. 182-190.

24. **Gong, Chenguang, Wei-Lin Popp, Maximilian and Maquat, Lynne E.** Biochemical analysis of long non-coding RNA-containing ribonucleoprotein complexes. *Elsevier. Methods.* 58 (2), 2012, pp. 88-93.

25. **Mili, Stavroula and Steitz, Joan A.** Evidence for reassociation of RNA-binding proteins after cell lysis: Implications for the interpretation of immunoprecipitation analyses. *RNA.* 10 (11), 2004, pp. 1692-1694.

26. **Klockenbusch, Cordula, O'Hara, Jane E and Kast, Juergen.** Advancing formaldehyde cross-linking toward quantitative cross-linking toward quantitative. *Analytical and Bioanalytical Chemistry.* 404 (4), 2012, pp. 1057-1067.

27. **Lehmann, Brian D, et al.** Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *The Journal of Clinical Investigation.* 121 (7), 2011, pp. 2750-2767.

28. **Lacroix, Marc and Leclercq, Guy.** Relevance of breast cancer cell lines as models for breast tumours: an update. *Breast Cancer Research & Treatment.* 83 (3), 2004, pp. 249-289.

29. **VanGuilder, Heather D, Vrana, Kent E and Freeman, Willard M.** Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques.* 44 (5), 2008, pp. 619-626.

30. **Heid, Christian A, et al.** Real Time Quantitative PCR. *Genome Research.* 6 (10), 1996, pp. 986-994.

31. **Moran, Victoria A, Niland, Courtney N and Khalil, Ahmad M.** Co-Immunoprecipitation of Long Noncoding RNAs. *Methods in Molecular Biology.* 925 (15), 2012, pp. 219-228.

32. **Kirmizis, Antonis, et al.** Silencing of human polycomb target Silencing of human polycomb target histone H3 Lys 27. *Genes & Development.* 18 (13), 2004, pp. 1592-1605.

33. **Tie, Feng, et al.** A 1-Megadalton ESC/E(Z) Complex from Drosophila That Contains Polycomblike and RPD3. *Molecular and Cellular Biology.* 23 (9), 2003, pp. 3352-3362.

34. **Gilbert, Chris and Svejstrup, Jesper Q.** RNA Immunoprecipitation for Determining RNA Immunoprecipitation for Determining. *Current Protocols in Molecular Biology.* 27 (4), 2006, pp. 4.1-4.11.

Appendix

HOTAIR sequence

The complete sequence of *HOTAIR.* Obtained from www.ncbi.nlm.nih.gov

ACATTCTGCCCTGATTTCCGGAACCTGGAAGCCTAGGCAGGCAGTGGGGAACTCTGACTCGCCTGTGCTC
TGGAGCTTGATCCGAAAGCTTCCACAGTGAGGACTGCTCCGTGGGGGTAAGAGAGCACCAGGCACTGAGG
CCTGGGAGTTCCACAGACCAACACCCCTGCTCCTGGCGGCTCCCACCCGGGACTTAGACCCTCAGGTCCC
TAATATCCCGGAGGTGCTCTCAATCAGAAAGGTCCTGCTCCGCTTCGCAGTGGAATGGAACGGATTTAGA
AGCCTGCAGTAGGGGAGTGGGGAGTGGAGAGAGGGAGCCCAGAGTTACAGACGGCGGCGAGAGGAAGGAG
GGGCGTCTTTATTTTTTTAAGGCCCCAAAGAGTCTGATGTTTACAAGACCAGAAATGCCACGGCCGCGTC
CTGGCAGAGAAAAGGCTGAAATGGAGGACCGGCGCCTTCCTTATAAGTATGCACATTGGCGAGAGAAGTG
CTGCAACCTAAACCAGCAATTACACCCAAGCTCGTTGGGGCCTAAGCCAGTACCGACCTGGTAGAAAAAG
CAACCACGAAGCTAGAGAGAGAGCCAGAGGAGGGAAGAGAGCGCCAGACGAAGGTGAAAGCGAACCACGC
AGAGAAATGCAGGCAAGGGAGCAAGGCGGCAGTTCCCGGAACAAACGTGGCAGAGGGCAAGACGGGCACT
CACAGACAGAGGTTTATGTATTTTTATTTTTTAAAATCTGATTTGGTGTTCCATGAGGAAAAGGGAAAAT
CTAGGGAACGGGAGTACAGAGAGAATAATCCGGGTCCTAGCTCGCCACATGAACGCCCAGAGAACGCTGG
AAAAACCTGAGCGGGTGCCGGGGCAGCACCCGGCTCGGGTCAGCCACTGCCCCACACCGGGCCCACCAAG
CCCCGCCCCTCGCGGCCACCGGGGCTTCCTTGCTCTTCTTATCATCTCCATCTTTATGATGAGGCTTGTT
AACAAGACCAGAGAGCTGGCCAAGCACCTCTATCTCAGCCGCGCCCGCTCAGCCGAGCAGCGGTCGGTGG
GGGGACTGGGAGGCGCTAATTAATTGATTCCTTTGGACTGTAAAATATGGCGGCGTCTACACGGAACCCA
TGGACTCATAAACAATATATCTGTTGGGCGTGAGTGCACTGTCTCTCAAATAATTTTTCCATAGGCAAAT
GTCAGAGGGTTCTGGATTTTTAGTTGCTAAGGAAAGATCCAAATGGGACCAATTTTAGGAGGCCCAAACA
GAGTCCGTTCAGTGTCAGAAAATGCTTCCCCAAAGGGGTTGGGAGTGTGTTTTGTTGGAAAAAAGCTTGG
GTTATAGGAAAGCCTTTCCCTGCTACTTGTGTAGACCCAGCCCAATTTAAGAATTACAAGGAAGCGAAGG
GGTTGTGTAGGCCGGAAGCCTCTCTGTCCCGGCTGGATGCAGGGGACTTGAGCTGCTCCGGAATTTGAGA
GGAACATAGAAGCAAAGGTCCAGCCTTTGCTTCGTGCTGATTCCTAGACTTAAGATTCAAAAACAAATTT
TTAAAAGTGAAACCAGCCCTAGCCTTTGGAAGCTCTTGAAGGTTCAGCACCCACCCAGGAATCCACCTGC
CTGTTACACGCCTCTCCAAGACACAGTGGCACCGCTTTTCTAACTGGCAGCACAGAGCAACTCTATAATA
TGCTTATATTAGGTCTAGAAGAATGCATCTTGAGACACATGGGTAACCTAATTATATAATGCTTGTTCCA
TACAGGAGTGATTATGCAGTGGGACCCTGCTGCAAACGGGACTTTGCACTCTAAATATAGACCCCAGCTT
GGGACAAAAGTTGCAGTAGAAAAATAGACATAGGAGAACACTTAAATAAGTGATGCATGTAGACACAGAA
GGGGTATTTAAAAGACAGAAATAATAGAAGTACAGAAGAACAGAAAAAAAATCAGCAGATGGAGATTACC
ATTCCCAATGCCTGAACTTCCTCCTGCTATTAAGATTGCTAGAGAATTGTGTCTTAAACAGTTCATGAAC
CCAGAAGAATGCAATTTCAATGTATTTAGTACACACACAGTATGTATATAAACACAACTCACAGAATATA
TTTTCCATACATTGGGTAGGTATGCACTTTGTGTATATATAATAATGTATTTTCCATGCAGTTTTAAAAT
GTAGATATATTAATATCTGGATGCATTTTCTGTGCACTGGTTTTATATGCCTTATGGAGTATATACTCAC
ATGTAGCTAAATAGACTCAGGACTGCACATTCCTTGTGTAGGTTGTGTGTGTGTGGTGGTTTTATGCATA
AATAAAGTTTTACATGTGGTGAAAAAA

Sanger sequences

The sequences obtained from Sanger sequences with HOTAIR primers.

OP1822_1 (forward)

TYMTYKYKCAWGAGAGCACAGGCACTGAGGCCTGGGAGTTCCACAGACCAACACCCCTGCTCCTGGCGGCTCCCACCCGGGA
CTTAGACCCTCAGGTCCCTAATATCCCGGAGGTGCTCTCAATCAGAAAGGTCCTGCTCCGCTTCGCAGTGGAATGGAACGGATT
TAGAAGCCTGCAGTAGGGGAGTGGGGAGTGGAGAGAGGGAGCCCAGAGTTACAGACGGCGGCGAGAGGAAGGAGGGGCG
TCTTTATTTTTTTAAGGCCCCAAAGAGTCTGATGTTTACAAGACCARAAATGCCACGGCCGCGTCCTGGCAGAGAAAAGGCTGA
AATGGAGGACCGGCGCCTTCCTTATAAGTATGCACATTGSCRAGAGAAGTGCTGCAACCTAAACCARCAATTACMCCCAAGCTC
GTTGGGGMSTAAGCCAGTACCGACCTGGTAGAAAAAGCAACCACGAAGCTAGAGAAGAGAGAGA

OP1822_2 (forward)

TYGKSYYGCWGAGAGCMCAGGCACTGAGGCCTGGGAGTTCCACAGACCAACACCCCTGCTCCTGGCGGCTCCCACCCGGGAC
TTAGACCCTCAGGTCCCTAATATCCCGGAGGTGCTCTCAATCAGAAAGGTCCTGCTCCGCTTCGCAGTGGAATGGAACGGATTT
AGAAGCCTGCAGTAGGGGAGTGGGGAGTGGAGAGAGGGAGCCCAGAGTTACAGACGGCGGCGAGAGGAAGGAGGGGCGT
CTTTATTTTTTTAAGGCCCCAAAGAGTCTGATGTTTACAAGACCAGAAATGCCACGGCCGCGTCCTGGCARAGAAAAGGCTGAA
ATGGAGGACCGGCGCCTTCCTTATAAGCTCGTTGGGGCCTAAGCCAGTACCGACCTGGTARAAAAAGCAACCACGAAGCTAGA
RRARAGAGAG

OP1824_1 (reverse)

TYAGTYGGTMTGGCTTAGGCCCAACGAGCTTGGGTGTAATTGCTGGTTTAGGTTGCAGCACTTCTCTCGCCAATGTGCATACTT
ATRAGGAAKGCGCCSGTCCTCCATTTCARCCTTTTCTCTGCCARGACRCGGCCGTGGCATTTCTGGTCTTGTAAACATCAGACTCT
TTGGGGCCTTAAAAAAATAAAGACSCCCCTCCTTCCTCTCGCCGCCGTCTGTAACTCTGGGCTCCCTCTCTCCACTCCCCACTCCC
CTACTGCAGGCTTCTAAATCCGTTCCWTTCCACTGCGAAGCGSARCAKGACCTTTCTGATTGASAGCACCTCCGGGATATTAGG
GACCTGASGGYCTAAGTCCCGGGTGGGAGCCKCCMKGASCASGGGTGTTGGTCTGTGGAACTCCCAGGCCTCAATGCCTGGTG
CTCTCTTACCCCCACGGAGCAGTCCTCACTGTGRAAACCTTTCGA

OP1824_2 (reverse)

TYRGGGGYGGGTMTGGGCTTAGGCCCAACGAGCTTATAAGGAAGGCGCCGGTCCTCCATTTCAGCCTTTTCYCTGCCAGGACG
CGGCCGTGGCATTTCTGGTCTTGTAAACATCAGACTCTTTGGGGCCTTAAAAAAATAAAGACGCCCCTCCTTCCTCTCGCCGCCG
TCTGTAACTCTGGGCTCCCTCTCTCCACTCCCCACTCCCCTACTGCAGGCTTCTAAATCCGTTCCATTCCACTGCGAAGCGGAGCA
GGACCTTTCTGATTGAGAGCACCTCCGGGATATTAGGGACCTGAGGGTCTAAGTCCCGGGTGGGAGCCGCCAGGAGCAGGGG
TGTTGGTCTGTGGAACTCCCAGGCCTCAGTGCCTGGTGCTCTCTTACCCCCACGGAGCAGTCCTCACTGTGRAAAGCTTTCRA

OP1823 (forward)

TAGAGCTGTTACAGACCAGAGAGCTGGCCAAGCACCTCTATCTCAGCCGCGCCCGCTCAGCCGAGCAGCGGTCGGTGGGGGG
ACTGGGAGGCGCTAATTAATTGATTCCTTTGGACTGTAAAATATGGCGGCGTCTACACGGAACCCATGGACTCATAAACAATAT
ATCTGTTGGGCGTGAGTGCACTGTCTCTCAAATAATTTTTCCATAGGCAAATGTCAAAGGGTTCTGGATTTTTAGTTGCTAAGGA
AAGATCCAAATGGGACCAATTTTAGGAGGCCCAAACAGAGTCCGTTCAGTGTCAGAAAATGCTTCCCCAAAGGGGTTGGGAGT
GTGTTTTGTTGGAAAAAAGCTTGGGTTATAGGAAAGCCTTTCCCTGCTACTTGTGTAGACCCAGCCCAATTTAAGAATTACAAG
GAAGCGAAGGGGTTGTGTAGGCCGGAAGCCTCTCTGTCCCGGCTGGATGCAGGGGACTTGAGCTGCTCCGGAATTTGAGAGG
AACATAGAAGCAAAGGTCCAGCCTTTGCTTCGTGCTGATTCCTAGACTTAAGATTCAAAAACAAATTTTTAAAAGTGAAACCAG
CCCTAGCCTTTGGAAGCTCTTGAAGGTTCAGCACCCACCCAGGAATCCACCTGCCTGTTACACGCCTCTCCAAGACACAGTGGC
ACCGCTTTTCTAACTGGCAGCACAGAGCAACTCTATAATATGCTTATATTAGGTCTAGAAGAATGCATCTTACA

OP1825 (reverse)

GTAGTATTAGCATATTATAGAGTTGCTCTGTGCTGCCAGTTAGAAAAGCGGTGCCACTGTGTCTTGGAGAGGCGTGTAACAGG
CAGGTGGATTCCTGGGTGGGTGCTGAACCTTCAAGAGCTTCCAAAGGCTAGGGCTGGTTTCACTTTTAAAAATTTGTTTTTGAA
TCTTAAGTCTAGGAATCAGCACGAAGCAAAGGCTGGACCTTTGCTTCTATGTTCCTCTCAAATTCCGGAGCAGCTCAAGTCCCCT
GCATCCAGCCGGGACAGAGAGGCTTCCGGCCTACACAACCCCTTCGCTTCCTTGTAATTCTTAAATTGGGCTGGGTCTACACAA
GTAGCAGGGAAAGGCTTTCCTATAACCCAAGCTTTTTTCCAACAAAACACACTCCCAACCCCTTTGGGGAAGCATTTTCTGACAC
TGAACGGACTCTGTTTGGGCCTCCTAAAATTGGTCCCATTTGGATCTTTCCTTAGCAACTAAAAATCCAGAACCCTCTGACATTT
GCCTATGGAAAAATTATTTGAGAGACAGTGCACTCACGCCCAACAGATATATTGTTTATGAGTCCATGGGTTCCGTGTAGACGC
CGCCATATTTTACAGTCCAAAGGAATCAATTAATTAGCGCCTCCCAGTCCCCCCACCGACCGCTGCTCGGCTGAGCGGGCGCGG
CTGAGATAGAGGTGCTTGGCCAGCTCTCTGGTCTTGTTAACAAGCCTCATCATAAAGATGGAGTGATAGGAAAA

Consensus sequences from Sanger sequences primer sets

Consensus sequences of primer set OP1822/OP1824 (the italic letters illustrates the extra exon in of the splicing variants):

CGAAAGGCTTTNCACAGTGAGGACTGCTCCGTGGGGGTAAGAGAGCACCAGGCACTGAGGCCTGGGAGTTCCACAGACCAAC
ACCCCTGCTCCTGGCGGCTCCCACCCGGGACTTAGACCCTCAGGTCCCTAATATCCCGGAGGTGCTCTCAATCAGAAAGGTCCTGC
TCCGCTTCGCAGTGGAATGGAACGGATTTAGAAGCCTGCAGTAGGGGAGTGGGGAGTGGAGAGAGGGAGCCCAGAGTTACAG
ACGGCGGCGAGAGGAAGGAGGGGCGTCTTTATTTTTTTAAGGCCCCAAAGAGTCTGATGTTTACAAGACCAGAAATGCCACGGC
CGCGTCCTGGCAGAGAAAAGGCTGAAATGGAGGACCGGCGCCTTCCTTATAAG***TATGCACATTGGCGAGAGAAGTGCTGCAAC***
CTAAACCAGCAATTACACCCAAGCTCGTTGGGGCCTAAGCCAGTACCGACCTGGTAGAAAAAGCAACCACGAAGCTAGAGAAGA
GAG

Consensus sequence of primer set OP1823/OP1825
TTCCTATCACTCCATCTTTATGATGAGGCTTGTTAACAAGACCAGAGAGCTGGCCAAGCACCTCTATCTCAGCCGCGCCCGCTCAGC
CGAGCAGCGGTCGGTGGGGGGACTGGGAGGCGCTAATTAATTGATTCCTTTGGACTGTAAAATATGGCGGCGTCTACACGGAAC
CCATGGACTCATAAACAATATATCTGTTGGGCGTGAGTGCACTGTCTCTCAAATAATTTTTCCATAGGCAAATGTCAAAGGGTTCTG
GATTTTTAGTTGCTAAGGAAAGATCCAAATGGGACCAATTTTAGGAGGCCCAAACAGAGTCCGTTCAGTGTCAGAAAATGCTTCC
CCAAAGGGGTTGGGAGTGTGTTTTGTTGGAAAAAAGCTTGGGTTATAGGAAAGCCTTTCCCTGCTACTTGTGTAGACCCAGCCCA
ATTTAAGAATTACAAGGAAGCGAAGGGGTTGTGTAGGCCGGAAGCCTCTCTGTCCCGGCTGGATGCAGGGGACTTGAGCTGCTC
CGGAATTTGAGAGGAACATAGAAGCAAAGGTCCAGCCTTTGCTTCGTGCTGATTCCTAGACTTAAGATTCAAAAACAAATTTTTAA
AAGTGAAACCAGCCCTAGCCTTTGGAAGCTCTTGAAGGTTCAGCACCCACCCAGGAATCCACCTGCCTGTTACACGCCTCTCCAA
GACACAGTGGCACCGCTTTTCTAACTGGCAGCACAGAGCAACTCTATAATATGCTAATATTAGGTCTAGAAGAATGCATCTT

Quality control of total RNA isolated from breast cancer cell lines

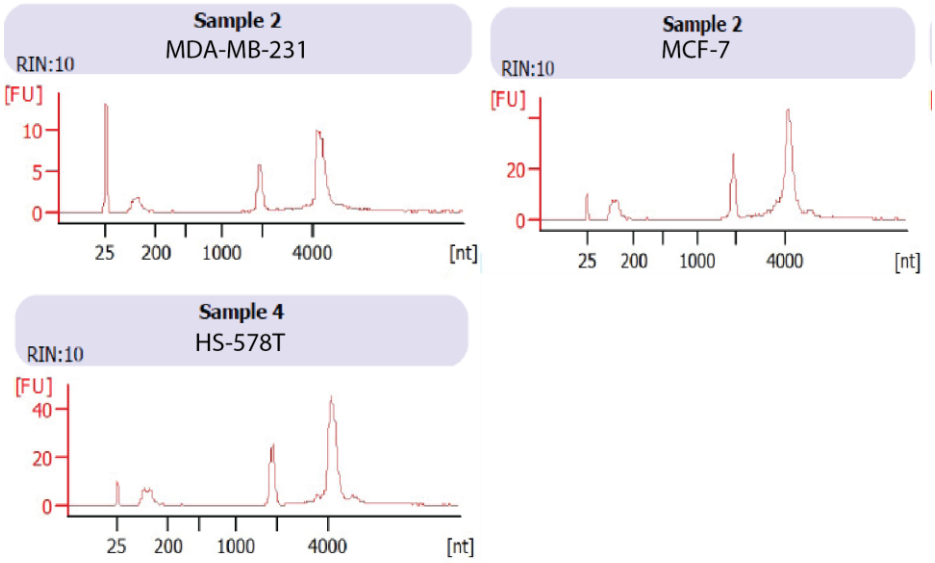


Figure A 1. The Agilent results from the total RNA isolated from the breast cancer cell lines MDA-MB-231, MCF-7 and Hs578T.

Table A 1. The RNA concentrations from Qubit analysis of the total RNA isolated from the three breast cancer cell lines.

RNA sample	Concentration (µg/µl)
MDA-MB-231	3.79
MCF-7	2.86
Hs578T	4.08

Expression of *HOTAIR* in the breast cancer cell lines

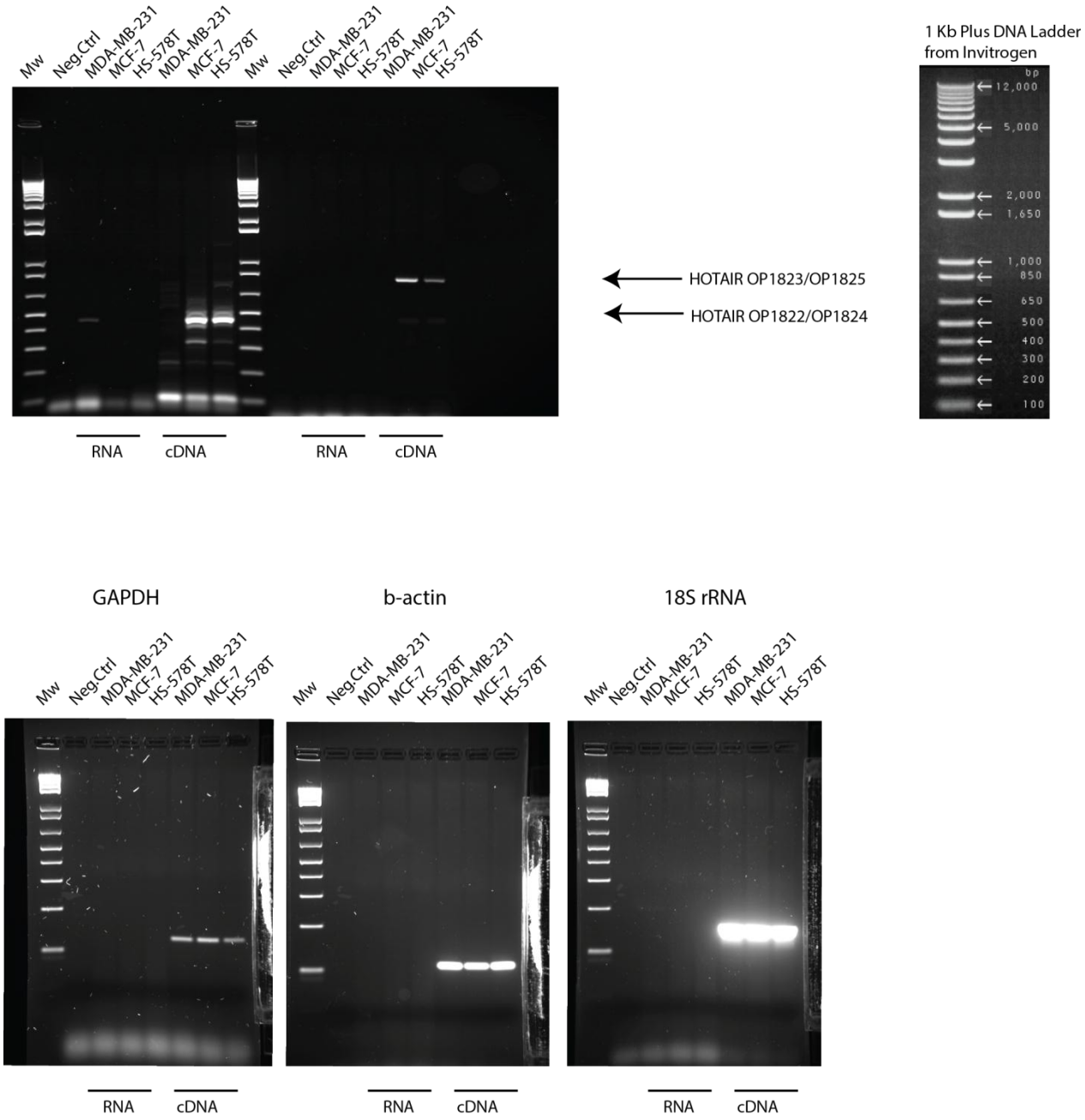


Figure A 2. The complete pictures of PCR products resolved on agarose gels showing the expression of *HOTAIR* and the housekeeping genes GAPDH, β-actin and 18S rRNA in the three cell lines.

Expression of Suz12 in the three breast cancer cell lines

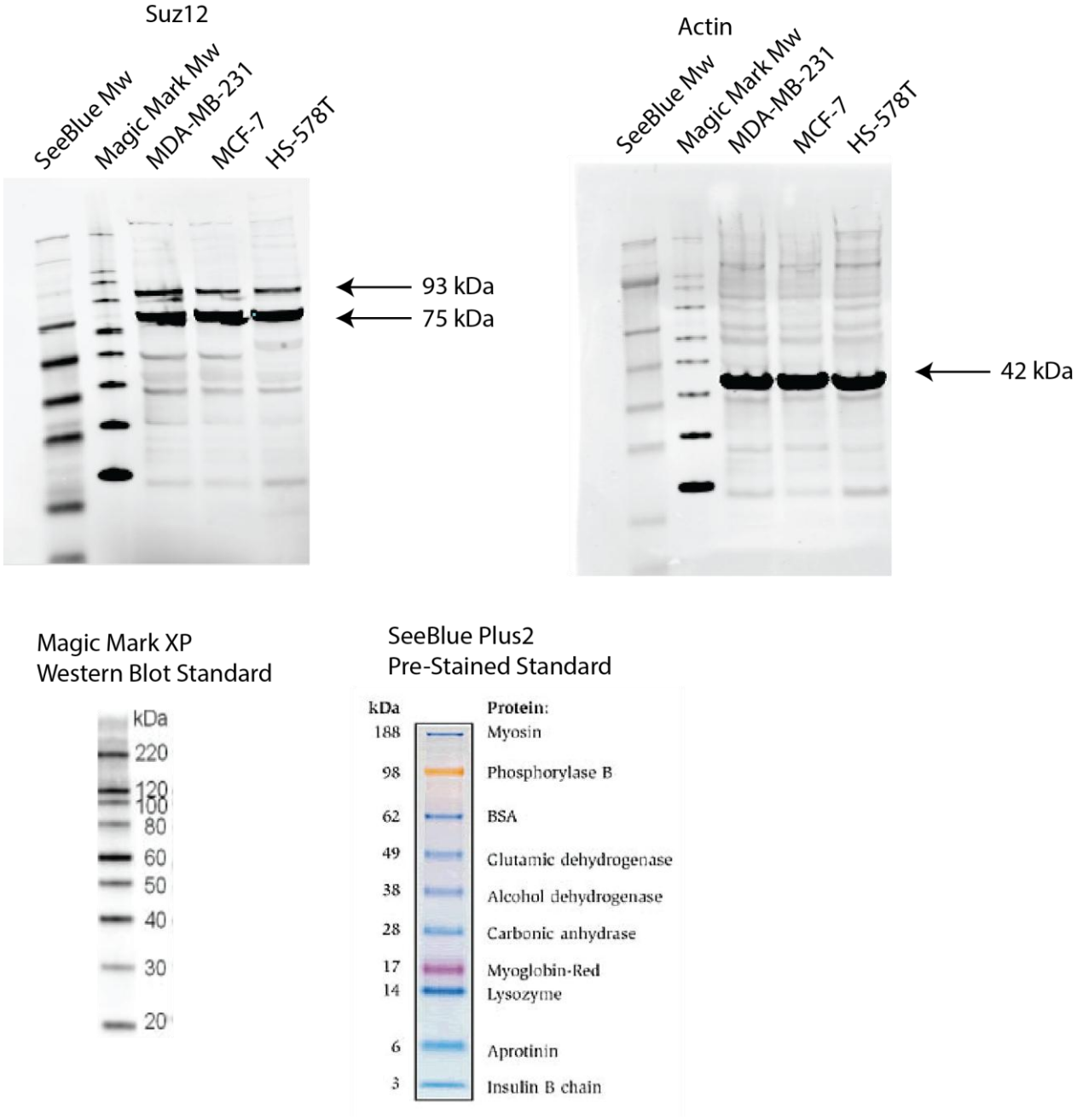


Figure A 3. The complete pictures from western blot analysis of Suz12 expression in the three breast cancer cell and the control blot with actin.

IP and RIP of Suz12 in Hs578T

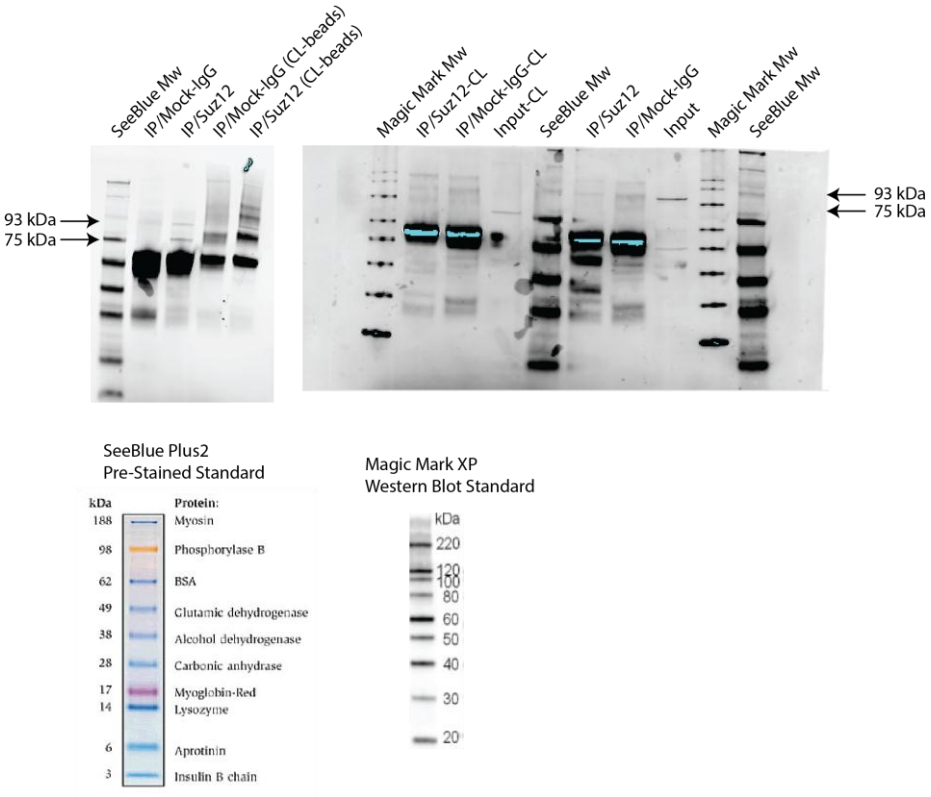


Figure A 4. The complete pictures from western blot analysis of detection of Suz12 in IP from WCE and in RIP from nuclear extract both treated and un-treated with formaldehyde.

Expression of *HOTAIR* by RIP in Hs578T

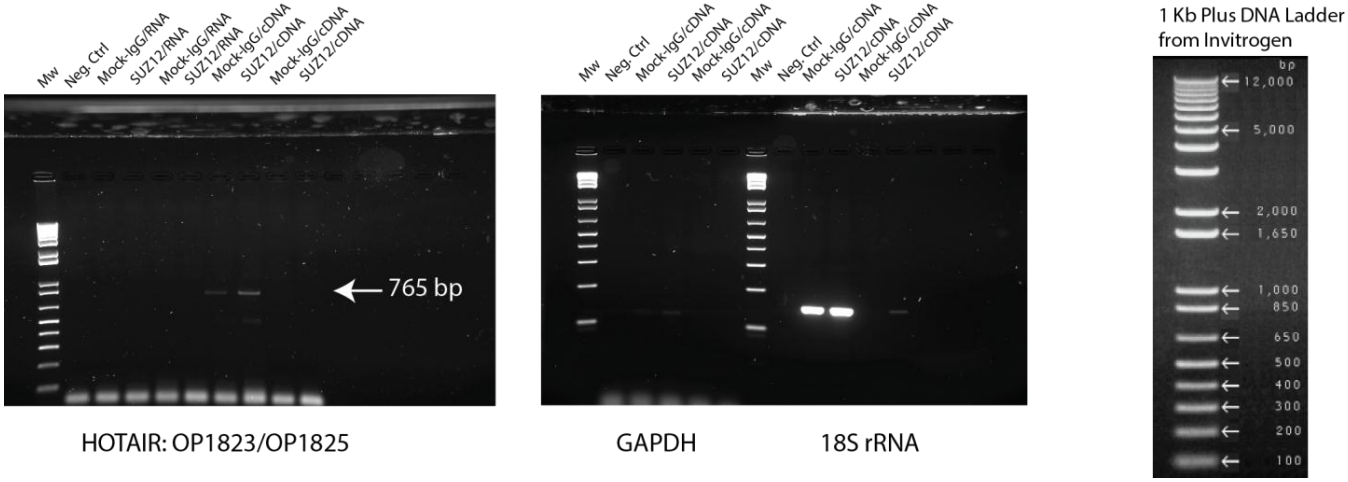


Figure A 5. The complete gel pictures from PCR products resolved on agarose gels showing the expression of *HOTAIR* and the housekeeping mRNA of GAPDH and 18S rRNA in RIP with anti-Suz12 and anti-IgG from nuclear extracts from cells both treated and un-treated with formaldehyde.

Raw data from qPCR

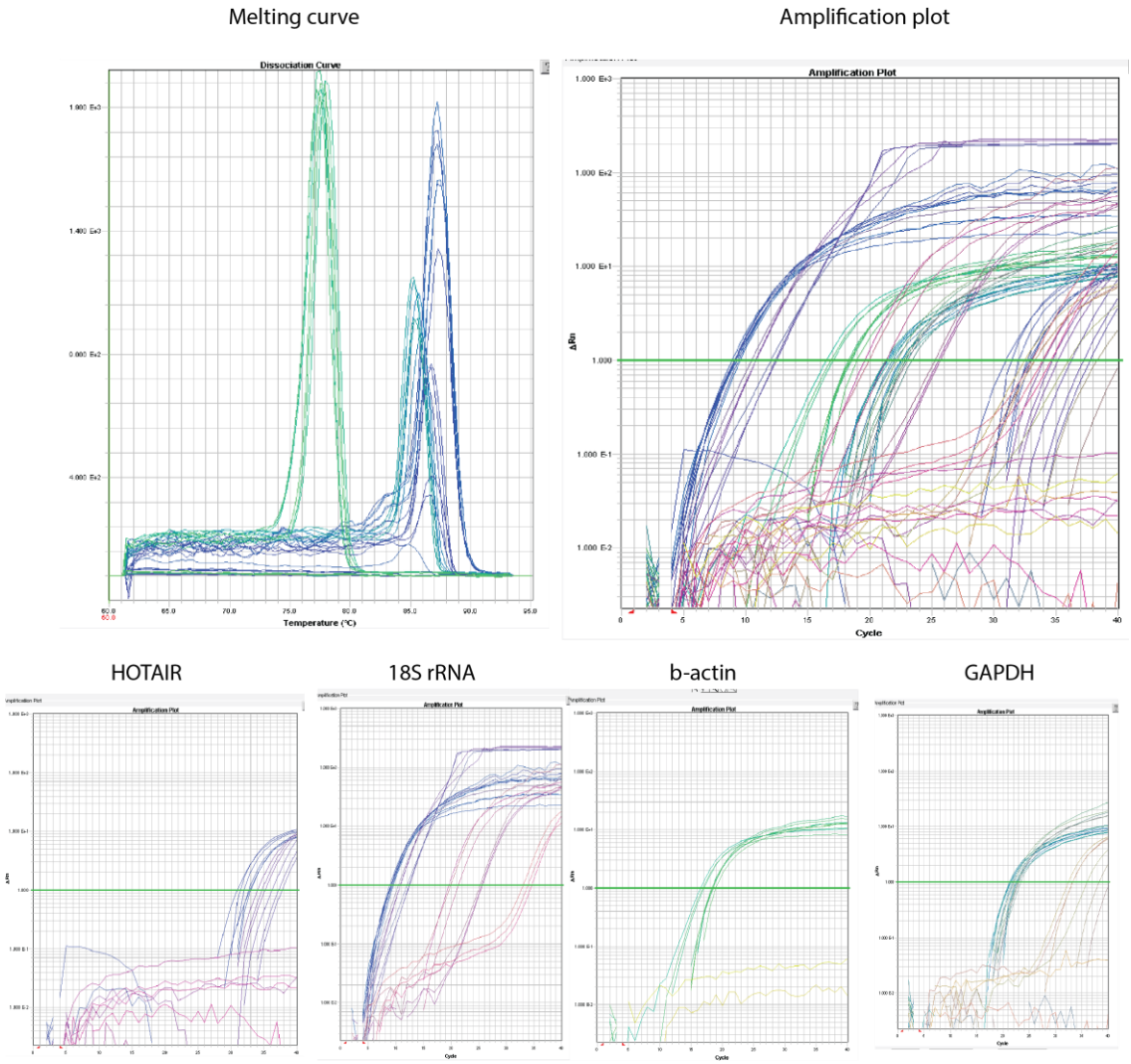


Figure A 6. The raw data from the qPCR analysis of the expression of *HOTAIR* and housekeeping genes in cDNA synthesized from total RNA from the three cell lines, and the expression of *HOTAIR*, GAPDH and 18S rRNA from RIP.

Table A 2. The normalized Cq-values from the qPCR analysis.

	Hs 578T totRNA	MCF7 totRNA	MDA-MB-231 totRNA	IgG, non-crosslinked	SUZ12, non-corsslinked	IgG, crosslinked	SUZ12, corsslinked	Blank
HOTAIR	32,84	31,43	N/A	36,75	34,62	N/A	N/A	N/A
18s rRNA	9,08	9,26	9,23	12,35	10,73	25,51	20,31	33,90
GAPDH	21,44	22,63	21,65	23,05	22,64	N/A	32,99	N/A
b-actin	16,76	18,50	18,39					N/A

Transfection of siRNA in Hs578T

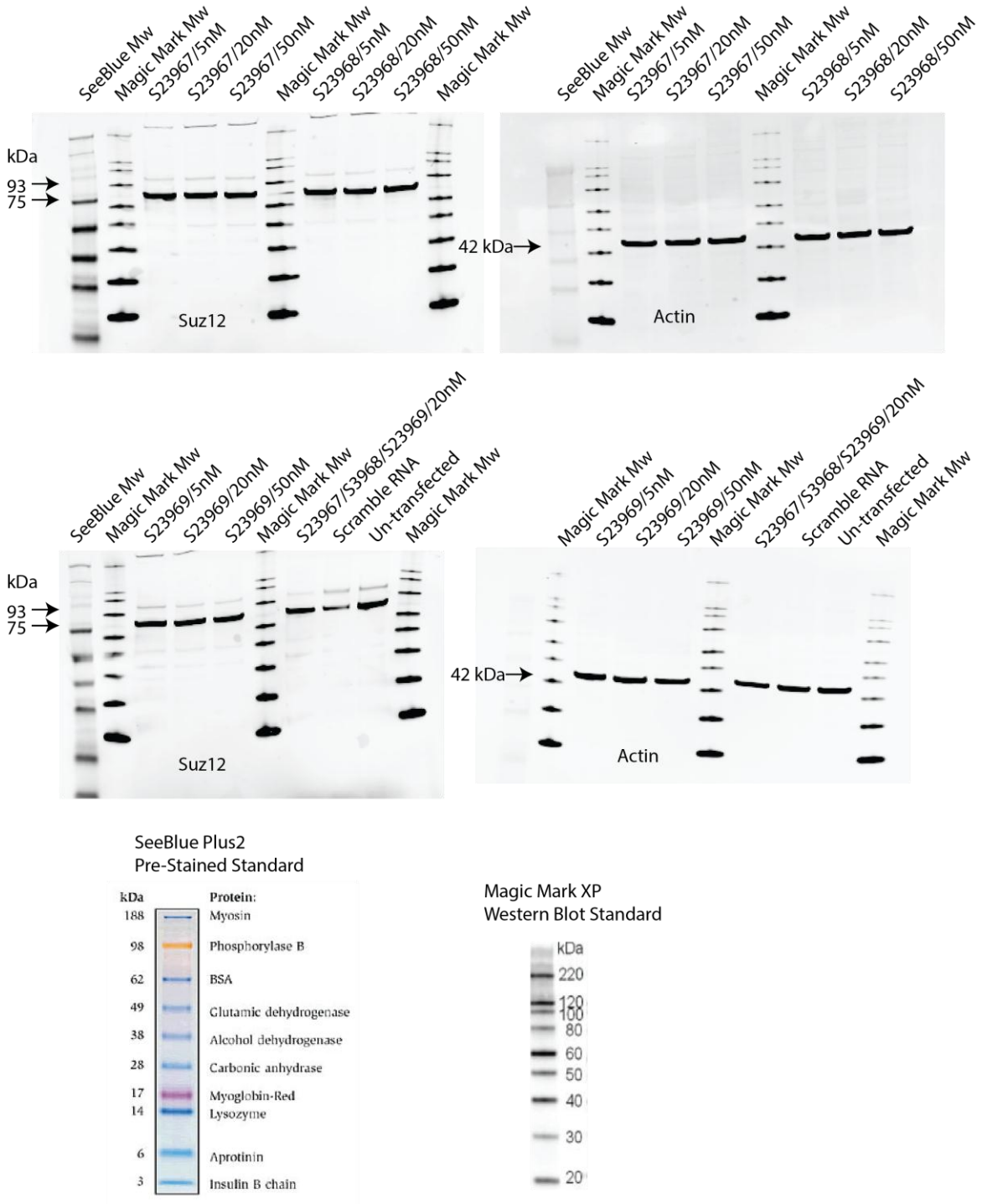


Figure A 7. The complete pictures from western blot analysis of transfection of siRNA targeting Suz12 in the Hs578T cell line.

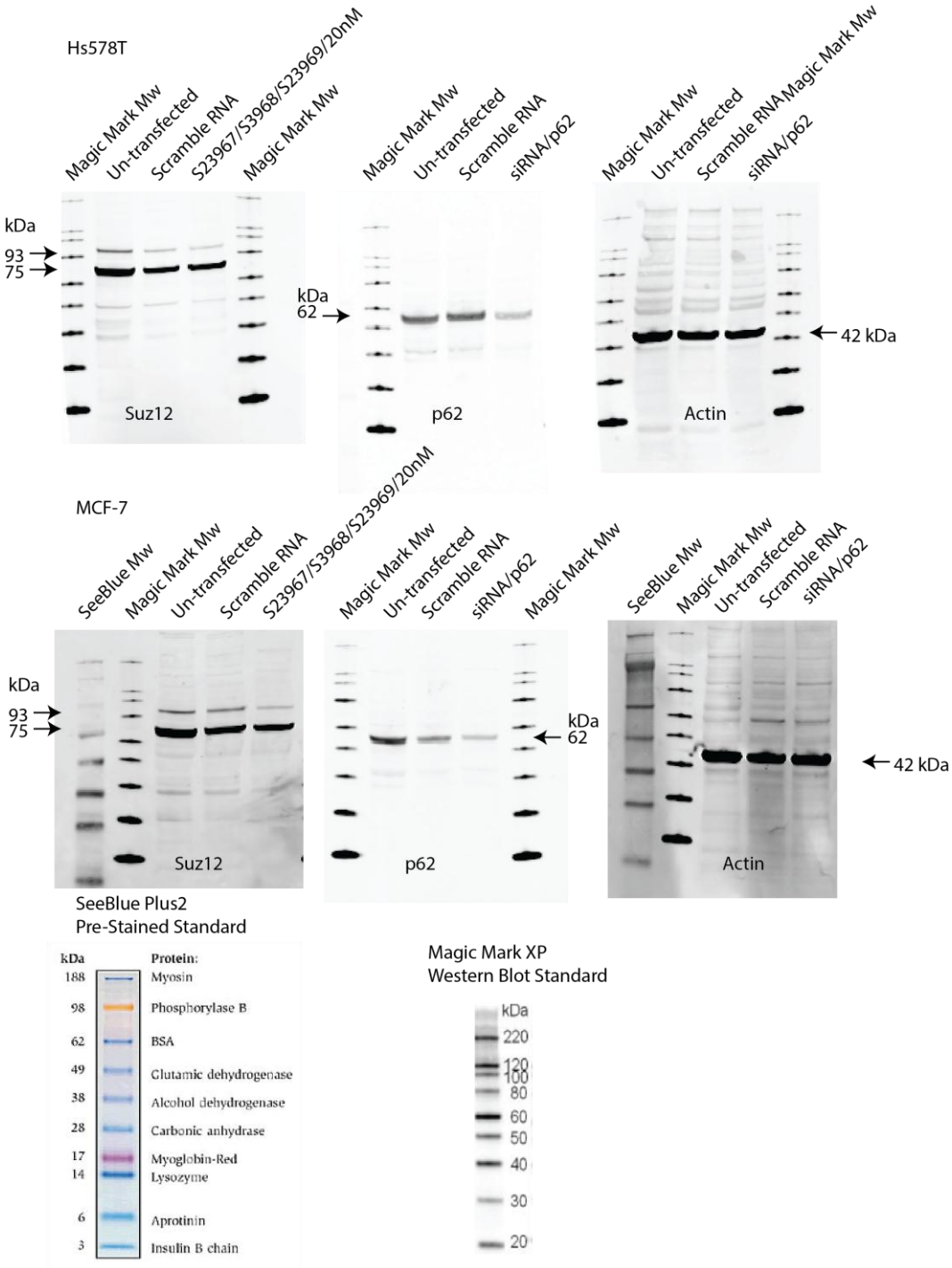


Figure A 8. The complete pictures from western blot analysis of transfections of siRNA targeting Suz12 and p62 in Hs578 and MCF-7 cell lines and their control blots with actin.