MECHANISMS OF SSX GENE EXPRESSION REGULATION AT THE PROMOTER LEVEL

A THESIS SUBMITTED TO
THE DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS
AND THE INSTITUTE OF ENGINEERING AND SCIENCE OF
BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

BY
DERYA DÖNERTAŞ
JULY 2009

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

_____
Dr. Sreeparna Banerjee

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

_____
Dr. Cengiz Yakıcıer

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

_____
Dr. Ali O. Güre

Approved for the Institute of Engineering and Science

_____
Director of Institute of Engineering and Science
Prof. Mehmet Baray

ABSTRACT

MECHANISMS OF SSX GENE EXPRESSION REGULATION AT THE PROMOTER LEVEL

Derya Dönertaş

M.S. in Molecular Biology and Genetics

Supervisor: Dr. Ali O. Güre

July 2009, 101 pages

Cancer Testis (CT) Antigen Genes are not transcribed in any of the adult tissues except spermatogonia, oogonia and trophoblasts. This tight regulation of expression is reversed resulting in the reactivation of CT transcription in a wide variety of cancers. CT genes are coordinately expressed and known to be regulated epigenetically. CT genes are reactivated in cancers by a mechanism that leads to the specific hypomethylation of their promoter-proximal sequences. The mechanisms leading to this phenomenon are unknown.

The main objective of this thesis was to further unravel the mechanisms regulating CT gene expression at the promoter level. For this purpose, SSX4 ,a typical CT-X gene known to be under the control of a bidirectional promoter, was chosen as a model. We characterized the minimal critical sequences controlling the sense and antisense promoter and discovered a bidirectional promoter with overlapping promoter activities within a 40 bp region. To study how the antisense promoter could mediate sense promoter repression and vice versa, we used two different reporter genes for each of the promoters in a single construct and found that measurable antisense promoter activity was dramatically reduced upon the introduction of a reporter for the sense promoter.

The SSX4 antisense promoter is capable of producing a noncoding transcript from the neighboring ornitine aminotransferase-like pseudogene in vivo. This, however, wasn't confirmed in this study. The possibility of transcriptional interference or the production of a small dsRNA that could affect the regulation of SSX4 gene expression is discussed in the context of the data. Results from experiments where the down-regulation of DICER was studied as a mechanism that could influence CT gene expression are also discussed

ÖZET

SSX GENİ İFADESİNİN DÜZENLENMESİNDE ROL ALAN PROMOTOR MERKEZLİ MEKANİZMALAR

Derya Dönertaş

Moleküler Biyoloji ve Genetik Yüksek Lisansı

Tez Yöneticisi: Dr. Ali O. Güre

Temmuz 2009, 101 sayfa

Kanser Testis (KT) antijen genleri spermatogonya, oogonya ve trofoblast haricindeki yetişkin dokularında ifade edilmeyen genlerdir. İfade edilimdeki bu sıkı düzenleme birçok kanser türünde tersine dönerek KT genlerinin yeniden ifade edilmesine yol açar. KT genleri eşgüdümlü olarak ifade edilir ve ifadeleri epigenetik etmenler tarafından düzenlenir. KT genlerinin kanserlerde yeniden ifade edilmesinin altında yatan neden promotor bölgelerinin özel olarak hipometilasyona uğramasıdır. Bu olaya yol açan mekanizmalar bilinmemektedir.

Bu tezin asıl amacı KT geni ifadesinin düzenlenmesinde rol alan promotor merkezli mekanizmaları açığa çıkartmaktı. Bu amaçla tipik bir KT geni olan ve çift taraflı bir promotora sahip olduğu bilinen SSX4 geni model olarak seçildi. Bu bağlamda, anlamlı (sense) ve anlamsız (antisense) promotor aktivitelerinden sorumlu en küçük DNA dizilerini karakterize ettik ve 40 baz çiftini kapsayan bir bölgede çift yönlü ve örtüşen bir promotor ortaya çıkardık. Anlamsız promotor aktivitesinin anlamlı promotor aktivitesini nasıl baskıladığını anlayabilmek için, bir vektör yapısında her iki promotor için farklı birer haberci gen kullandık. Ölçülebilen anlamsız promotor aktivitesinin, vektör yapısına anlamlı promotorun aktivitesini ölçen haberci genini eklediğimizde önemli ölçüde azaldığını gözlemledik.

Canlı dokularda SSX4 geninin anlamsız promotor aktivitesi, komşu ornitine aminotransferaza benzer yalancı-gen bölgesi üzerinden bir kodlamayan transkript ifade edebilme kapasitesine sahiptir fakat bu transkript ifadesi bu çalışmada onaylanamamıştır. Yazılım karışması ya da küçük çift dizili RNA (dsRNA) oluşumunun SSX4 geni ifadesinin düzenlenmesindeki muhtemel etkileri, bulunan veriler bağlamında ele alınmış ve tartışılmıştır. Ayrıca, DICER geninin bağlı olduğu mekanizmaların KT genleri ifadesinin düzenlenmesine etkisi, DICER geninin azaltarak düzenlendiği deneylerin sonuçları üzerinden tartışılmıştır.

To My Family…

ACKNOWLEDGEMENTS

First of all I would like to thank my supervisor Dr. Ali O. Güre for his supervision, patience, encouragement and support throughout this study. He always helped me through the difficulties that I have experienced during this study and always looked on the bright side, unlike me. I would like to thank AOG group members for their support and friendship. I also would like to thank all of my friends in the department for technical support and for their friendship especially to Sinan, Tamer, Fırat, Onur, Nilufer, Gurbet and Atıl. I would like to thank Dr. Mayda Gürsel for her experimental support.

Lastly my deepest thanks are to my family. They have supported me in every decision I made and I am more than grateful to have them in my life.

Throughout my M.S. study, I was personally supported by TÜBİTAK-BİDEB scholarship 2228. This work was also supported by grants from TÜBİTAK and European Commission to Dr. Ali O. Güre.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

Ago	Argonaute
bp	Base pair
CAGE	Cap Analysis Gene Expression
CT	Cancer Testis
DNMT	DNA Methyltransferase
dsRNA	Double stranded RNA
ENCODE	Encyclopedia of DNA Elements
L1	LINE1 Repeat
MAR	Matrix Associating Region
miRNA	MicroRNA
ncRNA	Noncoding RNA
nt	Nucleotide
OATL	Ornithine Amino Transferase Like
ORF	Open Reading Frame
piRNA	Piwi-interacting RNA
rasiRNA	Repeat-associated RNA
RdRP	RNA dependent RNA Polymerase
RISC	RNA-induced Silencing Complex
RNA Pol II	RNA Polymerase II
SEREX	Serological Screening of Expression Libraries
siRNA	Small Interfering RNA
SSX	Synovial Sarcoma X-Translocation
TF	Transcription Factor
TSA	Trichostatin A
TSS	Transcription Start Site
UTR	Untranslated Region

1. INTRODUCTION

1.1 Epigenetic Regulation of Gene Expression

1.1.1 DNA Methylation

Epigenetics refers to heritable phenotypic alterations in the absence of DNA sequence changes. DNA methylation is one of the extensively studied epigenetic alterations. In eukaryotes, ranging from plants to humans, DNA methylation is found exclusively at cytosine residues. In mammals, cytosines can be methylated only if they are followed by a guanine residue (CpG). CpG methylation is vital for organisms (mice that lack DNA Methyltransferases (DNMT) do not live) (Li E, 1992). In mammalian systems the genomic DNA methylation is found throughout the genome with the exception of short unmethylated regions called CpG islands, which contain larger than expected rate of CpG residues over a span of more than 500 nucleotides (Suzuki MM, 2008; Takai D, 2002). DNA methylation generally associated with suppression of transcription as well as with cellular processes like genomic imprinting, X chromosome inactivation, gene- and tissue-specific expression (Bernstein BE, 2007). DNA methylation is a dynamic, yet heritable trait enabling long-term memory of transcription rates. Three DNA Methyltransferases act on CpGs in mammals: Dnmt1 as the maintenance methyltransferase which acts on hemi-methylated DNA during DNA replication (Bestor T, 1988), Dnmt3a and Dnmt3b as *de novo* methyltransferases playing both collaborative and unique roles mostly in development (Okano M, 1999).

DNA methylation analysis can be performed by sequencing or restriction digestion preceded by sodium bisulfite treatment (COBRA assay: COmbined Bisulfite Restriction Analysis) (Frommer M, 1992) More recently developed techniques enable, genome-wide analysis of methylation status by microarray hybridization or high-throughput sequencing of bisulfite treated DNA samples.

1.1.2 Histone Modifications

Histone modifications that result in alterations of gene expression are also major epigenetic mechanisms. Nucleosome is the basic structural unit that consists of four core histones – H2A, H2B, H3 and H4 – around which 147 bp DNA is wrapped. Another histone protein, H1, is called the linker histone brings the nucleosomes closer into a more compact 30 nm fiber structure. The core histones have critical roles in epigenetic regulation of gene expression. N-terminal and C-terminal tails of the

histone proteins can be covalently modified by lysine acetylation, lysine and arginine mono-, di- and tri-methylation, serine and threonine phosphorylation, lysine ubiquitination, lysine sumoylation and proline isomerization (Kouzarides T, 2007). Specific combinations of these modifications are thought to constitute a "histone code". The histone code determines the euchromatic/heterochromatic regions of the DNA (Barski A, 2007), regulates biological processes like transcription, replication and repair through recruitment of chromatin remodeling complexes and confers a long-term memory of the transcriptional state, which have roles in development and differentiation (Jenuwein T, 2001).

Histone modifications are usually analyzed by chromatin immunoprecipitation (ChIP) using antibodies specifically recognizing modified residues. The generation of histone maps of human, mouse and yeast have been successfully performed by high throughput assays based on hybridizing ChIP samples on arrays (ChIP on chip) or by massively parallel signature sequencing of ChIP samples (ChIP-Seq) (Barski A, 2007)

1.1.3 Chromatin Remodeling

Another important concept in the epigenetic regulation of gene expression is chromatin remodeling. The most important complexes involved in chromatin remodeling are Polycomb (PcG) and Trithorax (trxG) Complexes which are initially found in *D.melanogaster* Hox gene repressors and activators respectively (Schuettengruber B, 2007). There are two classes of PcG proteins; Class II is directly involved in repression of transcription while class I is methyltransferase complexes that maintain the repressed state (Levine SS, 2004). PcG proteins are recruited to target genes' polycomb response elements (PREs) via sequence-specific DNA-binding factors and/or by specific histone modifications such as H3K27me3. PREs are typically several kilobases long and also are bound by the trxG proteins (Schuettengruber B, 2007). There are also the ATP-dependent chromatin remodelers SWI/SNF (a TrxG complex), NuRD/Mi-2/CHD and INO80 families which have been found to play different roles in chromatin remodeling (Saha A, 2006).

All these epigenetic events are interconnected: Methylated DNA binding proteins can be recruited to the cytosine methylated DNA along with histone deacetylases (Nan X, 1998) and/or nucleosome modifiers as a complex (Zhang Y, 1999). Covalent modifications of histones are found to occur coupled chromatin remodeling by ATP-dependent remodeling machinery, which supports the notion that DNA cytosine methylation, histone modification, and nucleosomal remodeling are intimately linked (Jones PA, Baylin SB 2007)

1.2 Cancer Epigenomics

For decades, scientists have been engaged in dissecting the origins of human cancer in genetics, and now with an explosion of data, it has been realized that genetics and epigenetics cooperate at all stages of cancer development. A very well known epigenetic event in human cancers is the transcriptional silencing of tumour suppressor genes by CpG-island-promoter hypermethylation which dates back to 1989 when it is discovered in Retinoblastoma gene (Greger V, 1989) and becomes widely accepted when p16INK4a was also shown to be inactivated by hypermethylation (Gonzalez-Zulueta M, 1995; Herman JG, 1995). Today, 100-400 CpG islands are estimated to be hypermethylated in a given tumor (Esteller M, 2007). Along with DNA methylation alteration, global histone modifications are also seen in many cancers. For instance, loss of acetylation at lysine 16 and trimethylation at lysine 20 of histone H4 are common hallmarks of human cancer (Fraga MF, 2005)

A second kind of alteration of DNA methylation common in human cancers is "global" or "genome-wide" hypomethylation, defined as an overall decrease in methylcytosine content of up to 70% compared with the level in normal somatic cells (Ehrlich M, 2002). Retrotransposon sequences of the LINE (long interspersed nuclear element) and SINE (short interspersed nuclear element) classes, as well as classical satellites are major targets of methylation in normal circumstances, but they become demethylated to various degrees in cancer cells (Hoffmann MJ, 2005). Global hypomethylation levels show variation both within and between the cancer types. Global hypomethylation appears to be an early event for colon and breast cancer as well as chronic lymphocytic leukemia. For other cancers like hepatocellular carcinoma, the degree of hypomethylation seems to increase with stage or histological grade but it is almost ubiquitous in metastases (Wilson AS, 2007). Along global hypomethylation, region specific hypomethylation of single copy genes also occurs, as exemplified by cancer testis antigen genes. The mechanisms underlying DNA hypomethylation are still under study. It has been suggested that the global hypomethylation can be a result of methyl group metabolism (a decrease in the methyl donor substrate S-adenosylmethionine (SAM)), a defect in DNA modifying enzymes like DNMTs or a defect in chromatin remodeling enzymes which can affect DNA methylation (Hoffmann MJ, 2005). However, we find it difficult to reconcile these explanations with gene specific hypomethylation. It has been suggested that functional selection during the random hypomethylation can result in gene-specific hypomethylation; however, given the fact that most of the cancer testis antigen genes do not have similar functions, we find it hard to support this hypothesis. Another possibility is that noncoding endogenous antisense RNAs might be playing a role in this type of hypomethylation, which was demonstrated for some cases (Imamura T, 2004). The mechanisms by which DNA hypomethylation occurs in cancer cells still needs further research.

1.3 Eukaryotic promoters

1.3.1 General Information

The RNA polymerase II core promoter comprises the sequences that direct the initiation of transcription. In the classical view, for the transcription of mRNAs, RNA Polymerase II recruitment to DNA is a major step that is facilitated by the general transcription factors (GTFs), including TFIIB, TFIID, TFIIE, TFIIF and TFIIH which is termed initiation; followed by promoter escape, abortive transcription, elongation and finally transcription termination. Transcriptional activators/ repressors binding to several proximal and distal control regions and the "histone code" can facilitate or repress transcription (Koch F, 2008).

In fact, mammalian promoters can be separated into two classes. The classic view of the promoters constitute one class that have a TATA box and other elements such as BRE (TFIIB recognition element), Inr (Initiator), MTE (motif ten element), DPE (downstream core promoter element). These elements are not universal and each is present in only a subset of core promoters. These elements are recognized by the factors that will eventually recruit transcription machinery. The other type of promoters can be characterized as being CpG rich, lacking the TATA Box and having multiple transcription start sites distributed over a broad region (Juven-Gershon T, 2008). Recently in the largest TSS identification study to date, using CAGE technology (as part of the ENCODE project(**Enc**yclopedia **O**f **D**NA **E**lements)), the TATA-box promoter architecture was found to represent only a minority of mammalian promoters in mouse and humans and this kind of promoter is commonly associated with tissue-specific genes (Carninci P, 2006). This finding was later confirmed by the results of different technologies in the context of The ENCODE Project Consortium 2007.

1.3.2 ENCODE Findings and Their Importance

The ENCODE project was initiated in September 2003, and aimed to identify all functional elements in the human genome by using various technologies in a number of laboratories. The project started with two components - a pilot phase and a technology development phase. In the pilot phase, 1% of the genome that was randomly selected was analyzed in detail. An initial finding emerging from the ENCODE consortium was that a vast amount of DNA, not annotated as known genes, was transcribed and, therefore, named transcriptionally active regions (TARs). Most of these RNAs were not found to encode for a protein and were therefore, considered as non-coding RNAs (ncRNAs). Except those that are evolutionary and/or structural conserved, we know very little about ncRNAs or their roles

inside the cell. We now know thanks to ENCODE that most of the genome is actually transcribed. Various unannotated TSSs were identified by ENCODE as well as the fact that the location of sequences that have a role in regulation of gene expression are very dispersed and in some cases can be located within the first exon, intron or the entire body of the gene or actually reside closer to another gene. Additionally, a significant fraction (up to 20%) of pseudogenes was found to be transcriptionally active contributing to TARs.

1.3.3 Bidirectional Promoters

Some of the non-TATA box containing, high GC containing, more evolvable promoters are bidirectional (Juven-Gershon T, 2008). This annotation refers to regions where two transcripts with opposite orientations occur simultaneously, or when a single promoter initiates two transcripts that might or might not overlap. In a genome wide study, it has been found that 10% of all genes are bidirectional genes controlled by bidirectional promoters. Many of the bidirectional transcript pairs are coexpressed but some are antiregulated. By transient transfection experiments using reporter genes, more than half of all human promoters were found not to exhibit strong directionality in transcript initiation and that the regulatory elements for the transcription in both directions are shared (Trinklein ND, 2004). It has been suggested that these nucleosome-free promoters are sites for cryptic noncoding RNA transcription which might have a regulatory function. (Neil H, 2009)

1.4.Cancer Testis Antigens

1.4.1 General Information

Cancer testis (CT) antigen genes are normally expressed mainly in germ cells but in no other adult tissues, with rare exception of trophoblast. This gene expression pattern is disrupted in malignancies leading to CT expression in various types of cancers. Since the protein products of these genes are generally immunogenic, they are considered as potential biomarkers and targets for immunotherapy (Scanlan MJ, 2002). The first CT antigens were found in the early 1900s by T-cell epitope cloning. By this method MAGE-1 (van der Bruggen P, 1991 and Traversari C, 1992), BAGE (Boël P, 1995) and GAGE (Van den Eynde B, 1995) antigens were identified. With the development of SEREX (Serological screening of expression libraries) technique many other CT antigens were identified such as SSX-2 (Türeci O, 1998) and NY-ESO-1 (Chen YT, 1997). Subsequently, in silico approaches led to the identification of most other CT antigen genes such as SSX4 (Gure AO, 2002). The CT antigen genes

were found to be members of multigene families and mostly localized to the X chromosome. Currently more than 70 CT gene families that are testis-restricted and immunogenic are known that have varied properties, localizations and expression profiles (Almeida LG, 2009). All of the data about the gene, gene expression, protein, protein expression, immunogenic response and pubmed links of the known CTs are combined and stored in a recent database, Ctpedia, which can be accessed from the website of Ludwig Institute for Cancer Research (http://www.cta.lncc.br/index.php).

There are certain characteristics of CT antigens. Their expression is restricted to gametogenic tissues and cancer, the coding genes of the antigens frequently map to chromosome X, they exist as multigene families and they are immunogenic in cancer patients. Their expression is heterogenous in cancers, they are activated by DNMT inhibitors and/or histone deacetylase inhibitors in vitro and their expression seems to be associated with tumor progression and with tumors of high metastastic potential (Scanlan MJ, 2002). Subsequent to the discovery of the immunogenic CT antigen genes, several other genes were coined with the same designation. However, it seems that those CT antigen genes that reside on the X chromosome are distinct from those on somatic chromosomes and are thus referred to as CT-X genes (Simpson AJ, 2005). Non-X chromosome CT genes do not necessarily code for immunogenic proteins, they are not composed of large families with highly homologous members and their expression pattern is not as restricted as that of CT-X genes. While CT-X gene expression is primarily found in spermatogonia and oogonia, non-X CT gene expression is frequently absent in these cells and present in gametes of later stages (Simpson AJ, 2005).

1.4.2 Function and Conservation

The biological functions of many of the CT antigens are not known although some exceptional CT antigens have been found to have roles during meiosis like OY-TES-1 (Baba T, 1994), CT-15/Fertilinβ (Vidaeus CM, 1997) and SCP-1 (Meuwissen RL, 1992). Unlike other CT antigens that are mostly expressed in spermatogonia, both OY-TES-1 and CT15/Fertilinβ are expressed in late stage sperm and they are not localized in the X chromosome. Among the major CT gene families, only some MAGE orthologs have defined functions. Mouse MAGE3-B4 plays role in germ cell development; mouse necdin binds to E2F1 and negatively regulates G1 to S progression. MAGE-A4 binds to the Gankyrin or p28 protein and inhibits the adhesion-independent growth of Gankyrin-overexpressed cells (Xiao J, 2004). Recently MAGE-A3/6 was identified as a novel target of fibroblast growth factor 2-IIIb (FGFR2-IIIb) signaling in thyroid cancer cells (Kondo T, 2007). Other functional CT genes are BORIS and CAGE. BORIS (brother of the regulator of imprinted sites) is the paralog of the abundant transcription factor

CTCF, and has been proposed to play role in CT-regulation according to the three studies showing BORIS and CTCF binding to NY-ESO-1 and MAGE-A1 promoters, resulting in derepression of both antigens (Vatolin S, 2005; Hong JA, 2005 and Kang Y, 2007) Some other CT antigens have some functional domains that may be an indication of their potential role but still these functions need to be explored. Along with the unknown functions, CT genes are almost exclusively specific to primates, with few exceptions. Human CT genes have orthologs in primates, especially in great apes and they are generally located on the same chromosomes (Stevenson BJ, 2007).

1.4.3 Regulation of Expression

All typical cancer testis genes located on the X chromosome are expressed only in the spermatogonia of the adult human but in no other tissue. Attempts to identify additional CT antigen genes based on the hypothesis that they should be expressed in the testis resulted in the discovery of genes that were classified as testis-restricted, tissue-restricted (expressed in ≤2 of 13 non-gametogenic tissues tested), differentially expressed (expressed in 3-6 non-gametogenic tissues, among 13 tested) and ubiquitously expressed (Scanlan MJ, 2004). These are clearly very different than CT-X genes as explained above. The initially-discovered, major CT gene families fall in the testis- and tissue-restricted groups and the rest of this work will focus on only these genes. We use the CT abbreviation in this context.

CT genes are reactivated heterogenously in a wide variety of cancers with the broadest distribution in germ cell tumors, melanomas and lung carcinomas, various adenocarcinomas and chondrosarcomas (Hofmann O, 2008) No genetic mutations have been found in these genes that can be the cause of this reactivation . The first clue about the regulation of CT antigens emerged as the upregulation of the expression of the MAGE-1 tumor antigen (later designated MAGE-A1) upon induction with the demethylating agent 5-aza-2'-deoxycytidine (Weber J, 1994) DNA demethylation as a reactivation mechanism for MAGE-A1 was confirmed soon after (De Smet C, 1996). Along with MAGE-A1, other CT antigens (De Smet C, 1999; Sigalotti, 2002; LimJH, 2005) were also found to be regulated by DNA methylation, supporting the idea that CT genes are coordinately expressed which was subsequently demonstrated (Gure AO, 2005). This, in turn, supported the idea that CT gene expression could be governed by common epigenetic mechanisms (Scanlan MJ, 2002). In another experiment HCT-116 cancer cell lines lacking DNMT genes were used to assess the expression of MAGE-A1, NY-ESO-1, and XAGE1. The results showed that the genetic knockout of both DNMT1 and DNMT3b could robustly induce CT antigen gene expression; whereas individual DNMT1 or DNMT3b knockout had a modest or negligible effect (for XAGE they were found to act independently). This is

another indicative of DNA methylation as a major mechanism in the regulation of CT gene expression (James SR, 2006). Along with the DNA methylation, it was found that histone acetylation plays a secondary role as trichostatin A was able to significantly upregulate 5-aza-2'-deoxycytidine induced MAGE or SSX gene expression (Wischnewski F, 2006; Gure AO, 2002). Interestingly, most CT genes are known to harbor CpG islands. We currently know that although most of the CpG islands from other genes are hypomethylated in the adult tissue, promoters of CT genes are hypermethylated in the normal adult tissue except testis and become hypomethylated in malignancies. The mechanism of hypomethylation of CT genes might be related to global DNA hypomethylation observed in cancer cells (De Smet, 1996; Kaneda A, 2004). But it has been shown that in tumor cells expressing MAGE-A1, the 5' region is significantly less methylated than the other parts of the gene showing a promoter specific hypomethylation rather than an artifact of global hypomethylation (De Smet C, 2004). It was suggested that site-specific hypomethylation of MAGE-A1 in tumor cells relies on a transient process of demethylation followed by a persistent local inhibition of remethylation due to the presence of transcription (De Smet C, 2004).

Another insight for the regulation of expression of CT antigens comes from experiments where the effect of BORIS protein (brother of the regulator of imprinted sites, a homologue of CTCF) on CT gene expression was studied. During spermatogenesis, expression of BORIS is restricted to germ cells, and coincides with a marked decrease in CTCF expression, erasure of methylation patterns, and up-regulation of CT genes (Loukinov DI, 2002). In one study, conditional expression of BORIS was shown to activate expression of CT antigen genes, and activation of BORIS expression following 5-aza-2'-deoxycytidine treatment occurred hours prior to transcriptional activation of other CT genes. Furthermore, downregulation of BORIS by RNA interference prior to 5-aza-2'-deoxycytidine treatment reduced the capacity of 5-aza-2'-deoxycytidine to activate MAGE-A1 expression suggesting a role for BORIS in the regulation of CT gene expression (Vatolin S, 2005). BORIS was also shown to bind directly to the MAGE-A1 and NY-ESO-1 promoters and to displace CTCF at these loci (Vatolin S, 2005; Hong JA, 2005).

1.4.4 Promoters of Cancer Testis Antigens

CT genes have in common a TATAless promoter, which is heavily methylated and thus silent in normal tissues. The first promoter analysis of a CT gene was performed in 1995 with MAGE-A1 gene. Two inverted Ets motifs were found to drive 90% of the activity of the MAGE-A1 promoter. When episomally expressed, the MAGE-A1 promoter was transcriptionally active even in tumor cells which did not express MAGE-A1 gene (De Smet C, 1995; Scanlan MJ, 2002). This indicated that in tumor cell

lines that do not express MAGE-A1, transcription factors capable of inducing MAGE-A1 promoter activity are present but that the gene is insensitive to their action.

1.4.5 SSX Gene Family

Synovial sarcoma X-translocation (SSX) genes were first identified as fusion counterparts to SYT in in t(X;18)(p11.2;q11.2) chromosomal translocation that is present in 70% of synovial sarcomas (Clark J, 1994). The first found member of the SSX as a CT antigen (HOM-MEL-40) by SEREX was SSX2 (Sahin U,1995; Türeci O, 1996). By genome homology searches all 9 members of the SSX family together with 10 pseudogenes were identified (Gure AO, 1997). Like most of the other cancer testis antigens, SSX also mapped to X chromosome, in particular to the OATL1 and 2 regions on Xp11.2 (Clark J, 1994) SSX family members have high homology ranging from 89 to 95% at the nucleotide level and 77 to 91% at the amino acid level (Gure AO, 1997). The 10 SSX genes listed by Ctpedia include; SSX1, SSX2, SSX2B, SSX3, SSX4, SSX4B, SSX5, SSX6, SSX7 and SSX9. There are 2 SSX2 and 2 SSX4 genes located tail to tail and head to head orientation, respectively (Gure AO, 2002). Normal testis tissue expresses SSX1, 2, 3, 4, 5 and 7 but not 6, 8 or 9. SSX6 expression, however, can be induced by TSA in some tumor cell lines (Gure AO, 2002). SSX proteins appear to be transcriptional regulators, whose actions are mediated primarily through association with or recruitment of Polycomb group repressors by the SSX-RD domain, mapped to the C-terminal of SSX proteins (Ladanyi M, 2001). LIM homeobox protein LHX4 was identified among the proteins that interact with the SSX C-terminal repression domain (de Bruijn DR, 2008).

SSX4 and SSX4B are nearly identical copies that are positioned in a head to head manner in the Xp11.23 (**Figure 1.1**). There are two splice variants of SSX4 (Gure AO, 2002).

FIGURE 1.1 SSX4 in the genomic context

SSX4 has a testis selective expression pattern. It has been shown to be expressed in a percent of hematologic malignacies, brain cancer, hepatocellular carcinoma, non small cell lung carcinoma, melanoma, ovarian carcinoma and synovial sarcoma (Ctpedia).

1.5 Small RNA Mediated Epigenetic Regulation

The first small RNA, lin-4, was discovered in 1993 by genetic screens in nematode worms (Lee RC, 1993; Wightman B, 1993). After this discovery, with the help of new technologies like deep-sequencing and strong computational predictions, the number and types of small RNAs increased extensively. The known functions of small RNAs range from heterochromatin formation to mRNA destabilization and translational control (Chu CY, 2007; Filipowicz W, 2008) and they are involved in almost every biological process, including developmental timing, cell differentiation, cell proliferation, cell death, metabolic control, transposon silencing and antiviral defense.

The distinguishing patterns of well known eukaryotic small RNAs are their limited size (~20–30 nucleotides (nt)) and their association with Argonaute (Ago) family proteins that lie in the heart of small RNA pathways. Ago proteins bind various <32nt small RNAs which guide the Argonaute complexes to their regulatory targets. The Ago family proteins can be grouped into two clades: the Ago subfamily and the Piwi subfamily. At least three classes of small RNAs are encoded in our genome, based on their biogenesis mechanism and the type of Ago protein that they are associated with: microRNAs (miRNAs), endogenous small interfering RNAs (endosiRNAs or esiRNAs) and Piwi-interacting RNAs (piRNAs). Although these are the three main small RNAs that we know for sure, numerous other small RNAs are being discovered in the light of the recent developments (Kim VN, 2009).

1.5.1 miRNAs

The best understood among the three small RNA classes are miRNAs which are single-stranded RNAs (ssRNAs) of ~22 nt in length and generated from endogenous hairpin shaped transcripts (Kim VN, 2005) Two RNase III-type proteins, Drosha and Dicer, are known to be critical for miRNA biogenesis. Mature miRNAs are then loaded into the RNA-induced silencing complex (RISC) whose core component is a protein belonging to the Ago family. The miRNAs guide the RISC complex to its target mRNA by basepairing with the 3′-untranslated region (3′UTR) of the target mRNA which typically leads to translational repression and exonucleolytic mRNA decay (Kim VN, 2009). Although the vast majority of miRNAs seem to act exclusively in the cytoplasm and mediate mRNA degradation or translational arrest (Filipowicz W, 2008), some plant miRNAs may act directly in promoting DNA methylation (Bao N, 2004). Furthermore, recent studies have described promoter-directed human miRNAs that can lead to repressive chromatin modifications and transcriptional gene silencing

(Gonzalez S, 2008; Kim DH, 2008). Over one third of human genes are predicted to be directly targeted by miRNAs according to computational analysis (Kim VN, 2009).

1.5.2 piRNAs

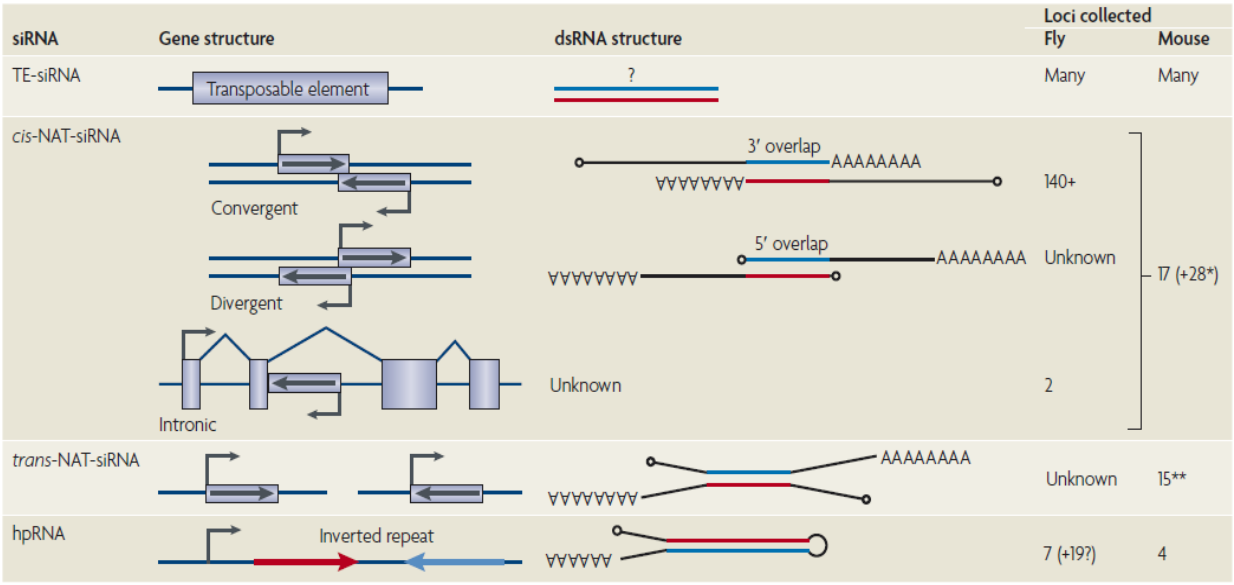
piRNAs were originally discovered during small RNA profiling studies of *D. melanogaster* development (Aravin AA 2001, 2003). piRNAs are endogenous, germ cell-specific small RNAs, generally longer than miRNAs (24–29 nt). Most of the piRNAs correspond to intergenic repetitive elements including transposons, thus initally named repeat-associated small interfering RNAs (rasiRNAs) (Aravin AA, 2003). The name piRNAs was coined after the interaction of Piwi subgroup proteins with these small RNAs was demonstrated. piRNAs are highly abundant in germ cells and at least some of them are involved in transposon silencing through heterochromatin formation or RNA destabilization (Kim VN, 2009). Intriguingly, the biogenesis of piRNAs does not depend on Dicer (Vagin VV, 2006) and occurs in a so called "ping-pong" mechanism in which formation of 5′ terminus within piRNA precursors is guided by piRNAs originating from transcripts of the other strand in concert with the Slicer activity of PIWI. Along with the fact that Aub and AGO3 (members of PIWI family) exhibited Slicer activity *in vitro*, this mechanism relies on the observation that Aub- and Piwi-associated piRNAs show a strong preference for uracil at their 5′ ends while AGO3-associated piRNAs show a strong preference for adenine at nucleotide 10 and AGO3- and Aub-associated piRNAs revealed pairs of RNAs showing complementarities in their first 10 nucleotides (Gunawardane LS, 2007). piRNAs have been identified in *D. melanogaster*, *C. elegans* and mammals, but seem to be absent in fungi and plants.

1.5.3 EndosiRNAs and Other Small RNAs

The first endosiRNAs were detected in plants and *C. elegans* and found to be products of RNA-dependent RNA polymerase (RdRP). A mammalian counterpart to RdRP has not been identified yet, and thus, these small RNAs were thought to be absent in mammalian cells. However, recently endosiRNAs have been shown to be ubiquitously present among higher eukaryote cells. The first mammalian endosiRNAs to be reported correspond to the long interspersed nuclear element (L1) retrotransposon and were detected in cultured human cells (Yang N, 2006). Full-length L1 contains both sense and antisense promoters in its 5′ UTR that could, in principle, drive bidirectional transcription of L1, producing overlapping complementary transcripts that can be processed into

siRNAs by Dicer. However, the precise mechanism by which transposons trigger siRNA production in mammals remains unknown (Yang N, 2006).

More recently, deep sequencing of small RNAs in *D. melanogaster* somatic tissue, cultured cells and ovaries has revealed a population of small RNAs (~21 nt long) that could readily be distinguished from miRNAs and piRNA as explained below. These small RNAs were found to be derived from transposon transcripts, sense–antisense transcript pairs and long stem-loop structures, **Figure 1.2** (Babiarz JE, 2008; Okamura K, 2008; Kawamura Y, 2008; Czech B, 2008 ; Ghildiyal M, 2008; Chung WJ, 2008; Okamura K, 2008)



Okamura K, Lai EC. Nat Rev Mol Cell Biol. 2008

FIGURE 1.2 The substrates for endosiRNA production in flies and mouse**.** Four classes of siRNA are shown: TE-siRNA, generated by overlapping transcripts corresponding to a transposable element, generated either in cis or trans; cis-NAT-siRNA, generated by overlapping transcripts that result due to the simultaneous activity a bidirectional RNA PolII promoter; trans-NAT-siRNAs, that result from mRNA homology forming in trans; hairpin RNAs (hpRNA) that generate siRNAs

The precise structure of the double-stranded RNA substrates of siRNAs derived from transposable elements is unknown, but hundreds or thousands of transposable elements are inferred to directly generate siRNAs. siRNAs derived from *cis*-natural antisense transcripts (*cis*-NATs) involve bidirectional transcription across the same genomic DNA, and can be convergent, divergent or involve annotated introns and/or internal exons. *Trans*-NAT dsRNAs form between transcripts that are produced from distinct genomic locations, and usually comprise an mRNA and an antisense-

transcribed pseudogene. siRNAs that are derived from hairpin RNA (hpRNAs) are long, inverted repeat transcripts whose double-stranded segment is typically much longer than that of miRNA precursors.

EndosiRNA production is Dicer dependent, Drosha independent and they are associated specifically with Ago2 in Drosophila (Kawamura Y, 2008; Czech B, 2008). Suppression of retrotransposons (and also some mRNAs) by endosiRNAs in *Drosophila* were observed mostly in the germ cells, and particularly in oocytes. This tissue specificity might relate to the fact that, other cell types may possibly invoke an interferon response to the dsRNAs produced intracellularly (Okamura K, Lai EC, 2008). EndosiRNAs have also been identified in mouse oocytes. As in flies, mouse endosiRNAs are 21 nucleotides, Dicer-dependent and derived from a variety of genomic sources (Tam OH, 2008 ; Watanabe T, 2008) A subset of mouse oocyte endosiRNAs map to regions of protein-coding genes that are capable of pairing to their cognate pseudogenes, and to regions of pseudogenes that are capable of forming inverted repeat structures which strengthens the theory that some pseudogene sequences may be under evolutionary selection to retain the ability to produce antisense transcripts that can pair with their cognate genes to produce endosiRNAs (Sasidharan R, 2008).

There are also promoter associated dsRNAs that have been shown to induce transcriptional gene silencing in DNA methylation dependent manner in plants (RNA dependent DNA Methylation) (Wassenegger M, 1994). Gene silencing by promoter associated dsRNAs has also been observed in human cells but in the absence of DNA methylation (Morris KV, 2004; Park CW, 2004; Ting HA, 2005). Along with endosiRNAs, synthethic antigene RNAs (agRNAs) complementary to the transcription sites of the genes, are also potent inhibitors (Janowski BA, 2005) or activators of gene expression (Li LC, 2006; Janowski BA, 2007). The agRNAs use AGO proteins and target antisense transcripts transcribed from the transcription start sites and bring different protein complexes to the transcription start sites (Schwartz JC, 2008).

1.5.4 Small RNA-mediated Control of Transposons

The PIWI-clade proteins and their associated piRNAs have important roles in the control of transposons in the germline — and possibly somatic cells — of D. melanogaster and mammals (Aravin AA, 2007). Three groups of data suggest that endosiRNAs are also a part of transposon regulation. First, in plants, viroids that possess host homologous sequences in their genomes cause fully methylation of the homologous sequences in the host (Wassenegger, 1994). Second, RNA viruses which contain homologous sequences to the host elicited the same effect (Jones L , 1998;

Wang MB, 2001) and lastly, the introduction of an inverted repeat containing transgene triggers methylation of both the transgene and the homologous sequences elsewhere in the genome of *Arabidopsis thaliana* (Mette MF, 2000). There are certain theories as to how dsRNAs that regulate transposon expression may arise. It is thought that such RNA should be similar to a read-through transcript from a single promoter that results in an inverted repeat thus producing an RNA hairpin or cryptic antisense promoters that confer bidirectional transcription.

There exists large amounts of data about small RNA mediated transcriptional silencing and chromatin remodeling; mostly in plants or fungi. In Drosophila, it is known that the role of the PIWI protein in repeat-induced gene silencing and heterochromatin assembly seems to involve a direct association between PIWI and HP1 (Brower-Toland B, 2007) There are adaptor proteins that bind to the Argonaute family of proteins and to chromatin or other DNA associated molecules. A mammalian counterpart to these adaptor proteins, however, has not been determined to date (Moazed D, 2009).

1.6 Long Noncoding RNAs (ncRNAs)

Although the current literature seems to be mostly focused on small RNAs, there are increasing numbers of reports describing long transcripts (longer than ~200 nucleotides) that are noncoding but yet functional in gene regulation. The transcription of these long ncRNAs is not random but strictly regulated. When compared with mRNAs, sequence conservation is low among most ncRNAs. One reason for this could be that process rather than the product of transcription has a functional consequence (Mercer TR, 2009). Functions of such ncRNAs are being identified and include roles in high order chromosomal dynamics, telomere biology and subcellular structural organization (Amaral PP, 2008). Although ncRNA mediated gene expression regulation would be expected to occur in trans, it does affect expression of neighboring genes. This cis-effect-like trans-regulation is referred to as transvection (Mercer TR, 2009)

Recently some examples have been found indicating the ability of long ncRNAs to regulate gene expression at the level of chromatin modification, transcription and post-transcriptional processing. Long ncRNAs can mediate epigenetic changes by recruiting chromatin remodelling complexes to specific genomic loci. For example, one of the ncRNAs expressed from human homeobox (Hox) loci, silences transcription across 40 kb of the HOXD locus in trans by inducing a repressive chromatin state by recruitment of the Polycomb chromatin remodelling complex PRC2 (Rin JL, 2007) For the transcriptional regulation, long ncRNAs can recruit RNA binding proteins and integrate them to the the transcriptional programme as in the case of CyclinD1. Long ncRNA recruits RNA binding protein

TLS which in turn inhibits the histone acetyltransferase activities of CReB binding protein and p300 and thus inhibits CyclinD1 expression (Wang X, 2008). With recruiting and affecting different proteins into the transcriptional programme, long ncRNAs can also activate transcription. Furthermore ncRNAs can basepair with mRNAs in a highly specific manner and regulate various steps in the post-transcriptional processing of mRNAs, including their splicing, editing, transport, translation and degradation (Mercer TR, 2009).

There is an increasing interest in the potential involvement of ncRNAs in disease etiology. For example, an antisense ncRNA transcribed from the p15 tumor suppressor locus induces local changes primarily of histone modifications, thereby regulating p15 expression. Thus, this ncRNA is potentially involved in oncogenesis as its expression is inversely correlated with that of p15 in leukemia. In this particular example, it is important to note that the effect of the noncoding antisense RNA was: 1. Dicer independent; 2.Its continued presence was not required for heterochromatinization, suggesting the sufficiency of a temporary presence for long lasting epigenetic alteration (Yu W, 2008).

2. MATERIALS & METHODS

2.1 Cell lines and Tissue Culture

2.1.1 Growth and Passage of the cell lines

The lung cancer cell line SK-LC-17, and two colon cancer cell lines, HCT 116 (ATCC #CCL-247) and HT-29 (ATCC #HTB-38) were grown in RPMI medium (GIBCO # 22400089) supplemented with 10% Fetal Bovine Serum (GIBCO #10106151), 1% Penicillin/Streptomycin (GIBCO #15070063) and 1% Non-essential Aminoacids (GIBCO #11140035 in 5% CO2 at 37°C. Cell culture medium was changed every two to four days. For subculturing, cells were washed with 1X PBS (see recipes) which was prepared and filtered before use. Trypsin-0.25%EDTA (GIBCO #25200056) was used to detach the cells. Cells were split at a dilution of 1:2 to 1:5 depending on the conditions of the cells. Media and aliquoted trypsin were kept at 4°C, Trypsin-0.25%EDTA was kept at -20°C and all solutions were warmed in 37°C water bath before use.

2.1.2 Long Term Storage of Cells (freezing)

For long term storage in liquid nitrogen, exponentially growing cells were harvested and centrifuged at 1300 rpm for 3 minutes at 4°C. Cells were counted and resuspended in freezing medium (10%DMSO, 20%FBS and 70%RPMI) at a concentration of approximately $3X10^6$ cells/ml. The suspension was aliquoted into 1 ml per cryotube and incubated at -20°C for a few hours, followed by -70°C for overnight, and finally placed in liquid nitrogen.

2.1.3 Thawing cells

The cryotube containing cells stored in liquid nitrogen was placed into an ice bucket immediately. The cryotube was then transferred into a 37°C water bath until only a small piece of ice remained visible. The cell suspension was then diluted into 5 ml of RPMI by decanting and not by pippetting and centrifuged for 3 minutes at 1300 rpm. Cells were resuspended in 5 ml of medium and seeded into a 25 cm^3 flask. Cells were incubated overnight in the incubator and were passaged next day.

2.2. Commonly Used Methods

2.2.1 RNA Extraction

RNA extraction was performed using TRI-Reagent (MRC #TR 118) according to the protocol given by manufacturer. Briefly, cells that were grown in 75cm^3 flasks were collected by trypsinization and centrifuged at 1300 rpm for 3 minutes at 4°C. Media was removed and cells were homogenized in 1ml TRI-Reagent. 200µl chloroform was added, vortexed for 15 seconds and the mixture was centrifuged 15 minutes at 13000rpm. The supernatant was collected and 500µl Isopropanol was added. The mixture was centrifuged 10 minutes at 14000 rpm and washed once with 75% and twice with 100% Ethanol. The pellet was dissolved in 250µl DNase-RNase free water and its concentration was read by NanoDrop ND-1000 Fullspectrum UV/Vis Spectrophotometer (Thermo Fisher Scientific). RNAse/DNAase free tubes and tips were used throughout the procedure.

2.2.2 DNase treatment of RNA

For DNase treatment of RNA, DNA-*free* ™Kit (Applied Biosystems #AM1906) was used. 1µl of rDNaseI was used per 10µg of RNA in 50µl including 10X DNaseI Buffer and incubated at 37°C for 30 minutes. For some of the samples that were known to be contaminated by genomic DNA an additional 1µl of rDNaseI was added to the mixture after this step and incubated for a further 30 minutes. 5µL of DNase Inactivation Reagent was added, vortexed and incubated at room temperature for 2 minutes by mixing occasionally. The sample was then centrifuged at 10000g for 1.5 minutes and the supernatant, including RNA, was transferred to another tube. The final concentration was 200ng/µl. RNAse/DNAase free tubes and tips were used throughout the procedure.

2.2.3 cDNA synthesis

cDNA synthesis was performed by DyNAmo™ cDNA Synthesis Kit (Finnzymes #F- 470S) according to manufacturer's instructions. During cDNA synthesis random hexamers were used to be able to detect noncoding RNAs.

2.2.4 5-aza-2'-deoxycytidine treatment

5-aza-2'-deoxycytidine (Sigma A3656) was used to treat cells at a concentration of 5mg/ml final per day for 4 days. The drug was prepared at a concentration of 5 mg/ml and stored at -70°C in aliquots of 10µl.

2.2.5 Agarose Gel Electrophoresis

Agarose gels were prepared by completely dissolving agarose in 1x TAE electrophoresis buffer at a percentage of 1.5% in microwave and ethidium bromide was added to final concentration of 30 µg/ml. 20µl of samples were mixed with 5X DNA loading buffer and loaded onto gels for the PCR products. The gel was run in 1x TAE at 90V for 45minutes.

2.2.6 Restriction Enzyme Digestions

The restriction enzyme digestions were performed in 20µl to 50µl reaction volumes with appropriate buffer at 37°C overnight.

2.2.7 Ligation Reactions

For ligation reactions, T4 DNA ligase (Promega # M1804) was used. For ligation reactions the necessary amounts of vector and insert were calculated by the following formula with 3:1insert to vector ratio and for 100 ng of vector.

$$\left(\frac{\text{ng of vector x kb size of insert}}{\text{kb size of vector}}\right) \text{X molar ratio of insert to vector} = \text{ng of insert}$$

Ligation reactions were typically incubated at 16°C, overnight and additionally for 3 hours at 4°C.

2.2.8 Transformation

For transformation, *E.coli* DH5α strain was used. Competent bacteria were prepared by the calcium chloride method and competent cells were stocked at -70°C in 150µl aliquots. For transformation, 50 to 100ng of ligation products were incubated with thawed competent cells on ice for 30 minutes. Cells were then incubated at 42°C for 30 seconds and immediately placed on ice again and incubated for 2-3 minutes. 850µl of LB was then added and the cells were grown in a 37°C shaker (220rpm) for one hour. Cells were centrifuged at 13000 rpm for 20 seconds, and resuspended in 100µl LB which was then plated by the spread plate technique on an LB-agar plate containing the selective antibiotic and was incubated overnight at 37°C. pGL3 plasmid is resistant to ampicillin, pTurboRFP-PRL plasmid is resistant to kanamycin and pSicoR PGK Puro is resistant to Ampicillin.

2.2.9 Purification of plasmid DNA

Purification of plasmid DNA was performed using the QIAprep Spin Miniprep Kit (Qiagen #27106) or home-made method of plasmid purification for miniprep. Briefly, 1.5 ml of overnight culture of a single colony was centrifuged at 13000rpm for 1 minute. Cells were resuspended in 100µl of solution P1 (recipes) and then 200µl P2 and 150µl P3 were added respectively. The mixture was centrifuged at 13000rpm for 10 minutes and the supernatant was transferred to a clean tube. The DNA was extracted by 100% EtOH and washed with 75% EtOH and dissolved in 50µl DNase-RNase free water. For the large scale plasmid DNA purification (maxiprep) the same home-made protocol was used with the appropriate amount of the solutions adjusted to maxiprep

2.3 Analysis of OATL transcript

The PCR primers used to detect OATL transcripts are listed in **Table1.** Primer Tms were calculated according to the following formula: Tm = 69.3°C + 0.41 (%GC) - 535/n, where n is primer length. For PCR reactions DyNAzyme II HS DNA polymerase (Finnzymes #F-504), dNTP mix (Finnzymes #F-560), forward and reverse primers at final concentrations of 0.03 U/µl, 250 µM and 500 nM, respectively were used in a total volume of 20µl. GAPDH primers were used as a positive control for testing RNA integrity and cDNA amplification. PCR was carried out under the conditions of 94°C for 10 min followed by 35 cycles of 94°C for 60s, 60 s at the appropriate melting temperature, 72°C for 60s, with a final extension at 72°C for 10 min, in an Applied Biosystems 9700 PCR machine.

2.4 Mapping the SSX4 minimal promoters in the sense and antisense orientations

2.4.1 Promoter Construct Generation 1: Insert Amplification by PCR

Individual promoter fragments were amplified by PCR to generate products that would then be cloned into the luciferase reporter vector pGL3. The primers used for forward (sense) promoter analysis are listed in **Table2** and the primers used for the reverse (antisense) promoter analysis are listed in **Table3**.

PCR with the appropriate primers were performed with 10 ng of a construct containing the SSX4 promoter region corresponding to -2000 to +1 in reference to the transcription start site as a template. Two PCR reactions for each fragment were performed and all of the PCR products were loaded to 1.5% agarose gel and run as explained before (**Figure S1**). The fragments were isolated from gel using QIAGEN Gel Extraction Kit (QIAGEN # 28706) according to the manufacturer's instructions.

2.4.2 Promoter Construct Generation 2: Vectors for luciferase experiments

pGL3-Basic (Promega #E1751) vector was used for the promoter activity analyses. The vector lacks eukaryotic promoter and enhancer sequences that control the firefly luciferase gene so that the expression of luciferase depends on the putative promoter that can be cloned within the MCS immediately upstream the luciferase ORF. The pGL3-Basic vector lacking an insert cotransfected with the pRL-TK (Promega E2241) vector was used as an internal control reporter. pRL-TK contains herpes simplex virus thymidine kinase (HSV-TK) promoter 5' to *Renilla* luciferase to provide low to moderate levels of *Renilla* luciferase expression in co-transfected mammalian cells.

2.4.3 Promoter Construct Generation 3: Restriction Enzyme Digestion and Ligation

XhoI (New England Biolabs #R0146L), HindIII (New England Biolabs # R0104L) restriction enzymes were used to double digest both the vector and the PCR fragments using NEB Buffer4 (**Figure S2**). Ligation reactions were performed as explained in section 2.6

2.4.4 Promoter Construct Generation 4: Vector construction

For the fine mapping of the minimal forward promoter pGL3-SSX4A3.3, pGL3-SSX4A3.31, pGL3-SSX4A3.32, pGL3-SSX4A3.33, pGL3-SSX4A3.34, pGL3-SSX4A3.35, pGL3-SSX4A3.36, pGL3-SSX4A3.37 constructs were prepared. For the fine mapping of the minimal reverse promoter, pGL3-SSX4A3.3R, pGL3-SSX4A3.31R, pGL3-SSX4A3.32R, pGL3-SSX4A3.33R, pGL3-SSX4A3.34R constructs were prepared. (**Figure 4.2** and **Figure 4.4**) To control the integrity of the constructs, XhoI and HindIII double digestions were performed to excise the fragments (**Figure S3**).

2.4.5 Transient Transfection of the Reporter Constructs

For transfection experiments Lipofectamine™ 2000 (Invitrogen #11668-019) was used. For luciferase reporter experiments, cells were plated in a 24-well culture plate according to the manifacturer's instructions. 0.8 µg pGL3 vector containing the promoter fragment to be tested was diluted in 50µl OPTI-MEM along with 20ng pRL-TK. In a separate tube 1.6µl of Lipofectamine 2000 was diluted in 50µl of OPTI-MEM to achieve 1:2 DNA:Lipofectamine ratio and incubated for 5 minutes in room temperature. Diluted DNA and Lipofectamine solutions were then mixed gently and incubated 30 minutes at room temperature. Meanwhile the growth medium of the cells to be transfected was replaced with 500µl OPTI-MEM serum free medium. The 100µl DNA-Lipofectamine mixture was then added to the cells and mixed by rocking the plate back and forth. After 5 hours of incubation at 37°C in a CO_2 incubator, OPTI-MEM was replaced by RPMI1640 and cells were left for incubation for another 19 hours and collected for luciferase analysis.

2.4.6 Luciferase Reporter Assay System

The Dual-Luciferase® Reporter Assay System from Promega (Cat# E1910) was used for these experiments. Upon delivery of the system The Luciferase Assay Substrate (LARII) was reconstituted according to manufacturer's instructions and divided into 1ml aliquots and stored at -70°C. Before each use, the necessary amount of LARII was thawed at room temperature in a water bath. Passive Lysis Buffer (PLB) and Stop & Glo® Reagent 1X working solutions were prepared just before use from 5X and 50X concentrates respectively, according to the manufacturer's instructions. The assays for firefly luciferase activity and *Renilla* luciferase activity were performed sequentially using one well of a 96 well opaque plate and by using The Reporter Microplate Luminometer (Turner Biosystems). After lysing the cells with 100µl PLB the day after the transfection, 10µl from the lysate was used for

the measurement. First 50µl of LARII was added and the luminescence was read immediately. Then 50µl Stop & Glo® Reagent was added, mixed by pippeting and Renilla luciferase activity was determined by reading the luminescence immediately again. All the samples were read one by one to avoid any decrease in the luminescence. Luciferase activities were acquired as RLU (relative luciferase unit), corrected by renilla luciferase activity, adjusted to pGL3-Control activity, and expressed in fold of pGL3-Basic activity +/- standard error of the mean.

2.5. Analysis of bidirectional SSX4 promoter activity by flow cytometry and fluorescent microscopy

2.5.1 Vectors used

pTurboRFP_PRL vector was purchased from Evrogen (#FP235). This is a promoterless vector encoding red fluorescent protein TurboRFP that can be used as an *in vivo* reporter of gene expression (**Figure 2.1**).

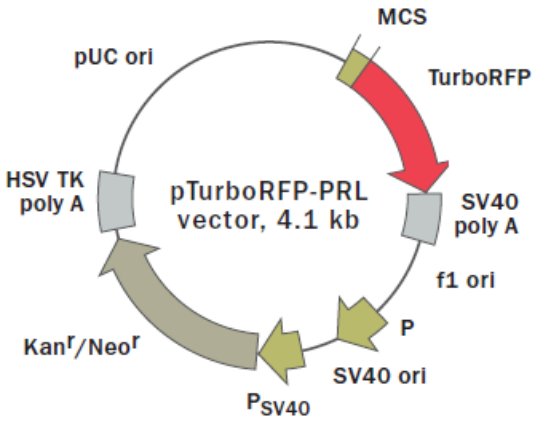


FIGURE 2.1 Map of pTurboRFP_PRL construct

pHygEGFP vector (Clontech #6014-1) was used as a control that expresses high levels of Green Flourescent Protein(GFP). pHygEGFP expresses a fusion of the hygromycin resistance gene with enhanced green fluorescent protein (EGFP) under control of the strong immediate early promoter of human cytomegalovirus (CMV) (**Figure 2.2**)

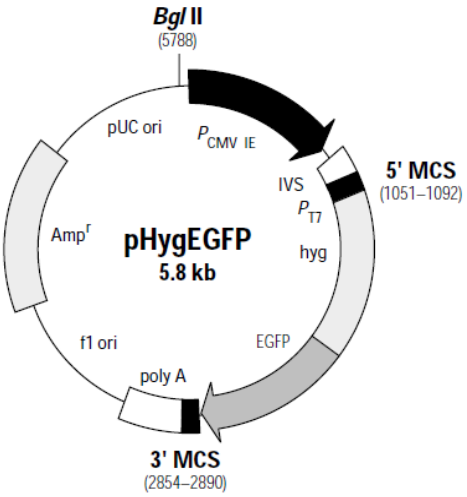


FIGURE 2.2 Map of pHygEGFP that expresses Hyg-EGFP under control of the strong immediate early promoter of human cytomegalovirus (CMV)

2.5.2 Reporter constructs

2.5.2.1 pCMV-TRFP (TurboRFP under the control of CMV Promoter)

The BglII/KpnI restricted CMV promoter from pHygEGFP was cloned into the same site of pTurboRFP-PRL to generate pCMV-TRFP that expresses TRFP under the control of the strong immediate-early promoter of human cytomegalovirus (CMV) (**Figure S6**) The map of the resulting construct can be seen in **Figure 2.3**. The vector was used as a control to measure TRFP.

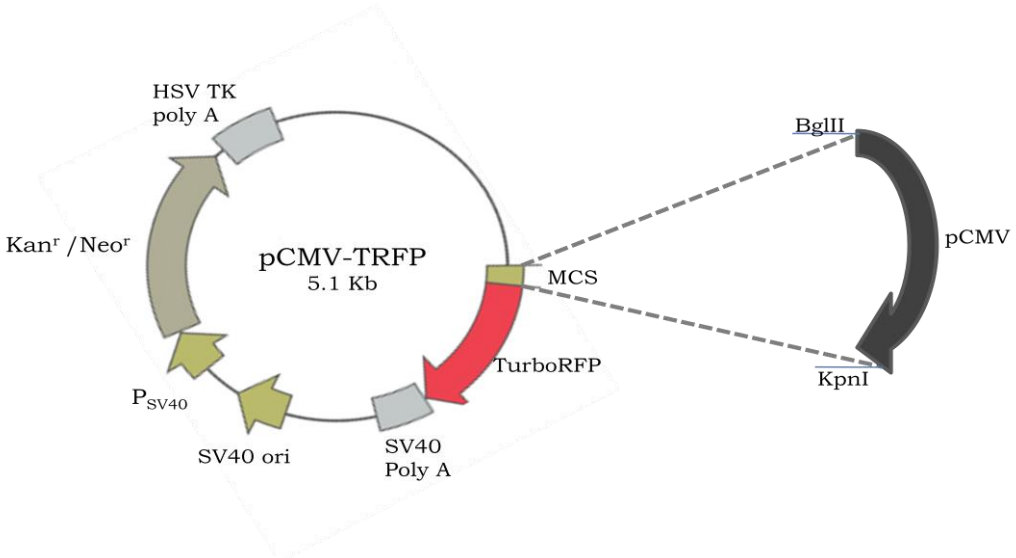


FIGURE 2.3 Map of pCMV-TRFP construct that expresses high levels of TRFP under the control of strong immediate-early promoter of human cytomegalovirus (CMV)

2.5.2.2 pSSX4P-TRFP (TurboRFP under the control of SSX4 Promoter)

XhoI and HindIII restricted SSX4 minimal promoter from pGL3-SSX4A3.3 was cloned into the same site of pTurboRFP-PRL to generate pSS4P-TRFP (**Figure S7**). The integrity of the construct was tested by XhoI and HindIII double digestion (**FigureS8**). The map of the resulting construct can be seen in **Figure 2.4**. This results in the vector where TRFP is under the control of the SSX4 promoter in the sense orientation.

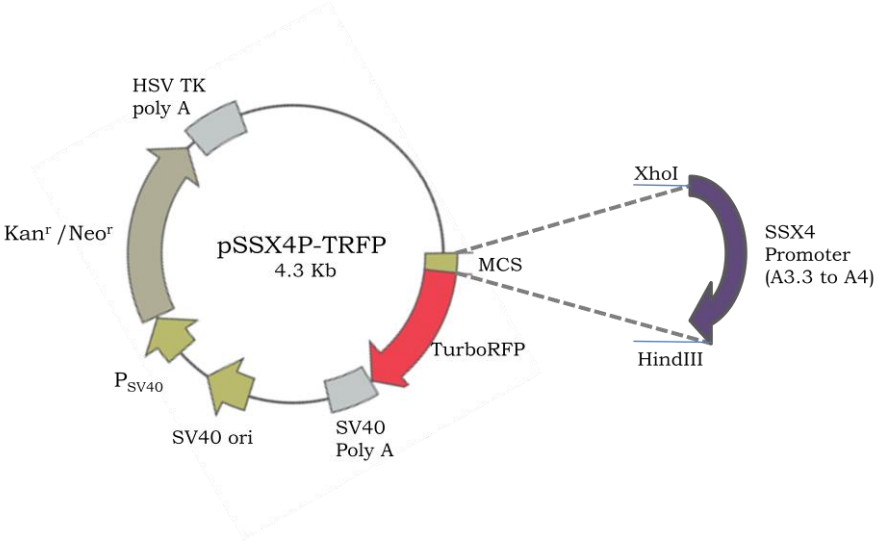


FIGURE 2.4. Map of pSSX4P-TRFP vector where TRFP is under the control of the SSX4 sense promoter

2.5.2.3 pSSX4PI-TRFP (TurboRFP under the control of inverted SSX4 Promoter)

XhoI and HindIII restricted inverted SSX4 promoter from the previous construct pGL3-SSX4A3.3R was cloned into the same site of pTurboRFP-PRL resulting in a vector with TRFP under the control of the SSX4 promoter in the antisense orientation(**Figure S7**). The integrity of the construct was tested by XhoI and HindIII double digestion (**FigureS8**). The map of the resulting construct can be seen in **Figure 2.5**

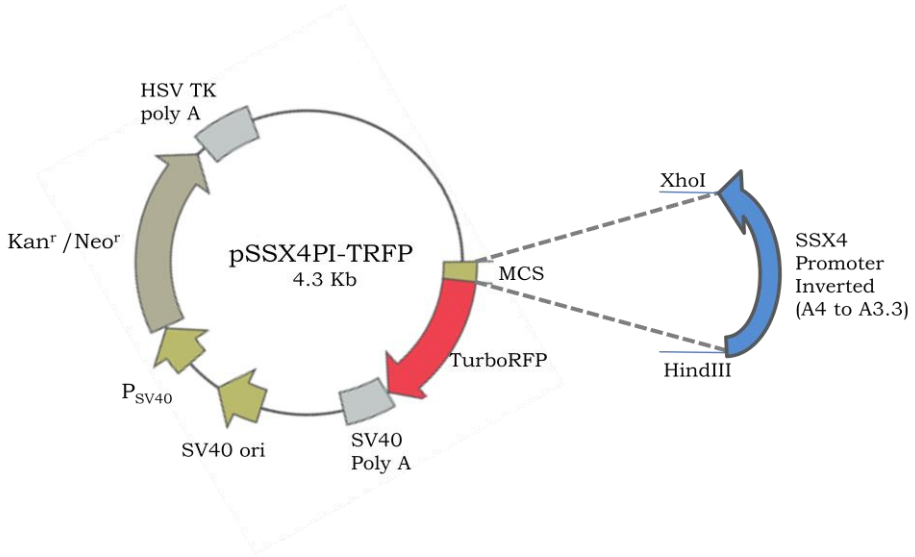


FIGURE 2.5 Map of pSSX4PI-TRFP vector where TRFP is under the control of the SSX4 antisense promoter

2.5.2.4 pGFP-SSX4PI-TRFP and pGFPI-SSX4PI-TRFP

pSSX4PI-TRFP was digested with BglII and treated with Arctic Phosphatase to prevent self ligation (**Figure S9**). pHygEGFP vector was digested with BamHI to excise HygEGFP (**Figure S10**) which was inserted into the BglII site of pSSX4PI-TRFP so that HygEGFP was either in the sense (pGFP-SSX4-TRFP) or antisense (pGFPI-SSX4-TRFP) orientation (**Figures 2.6.a** and **2.6.b**). The orientation of HygEGFP was confirmed by both NheI&XhoI and PvuI&XhoI double digestions (**Figure S11**). The constructs with the antisense GFP were used as a control.

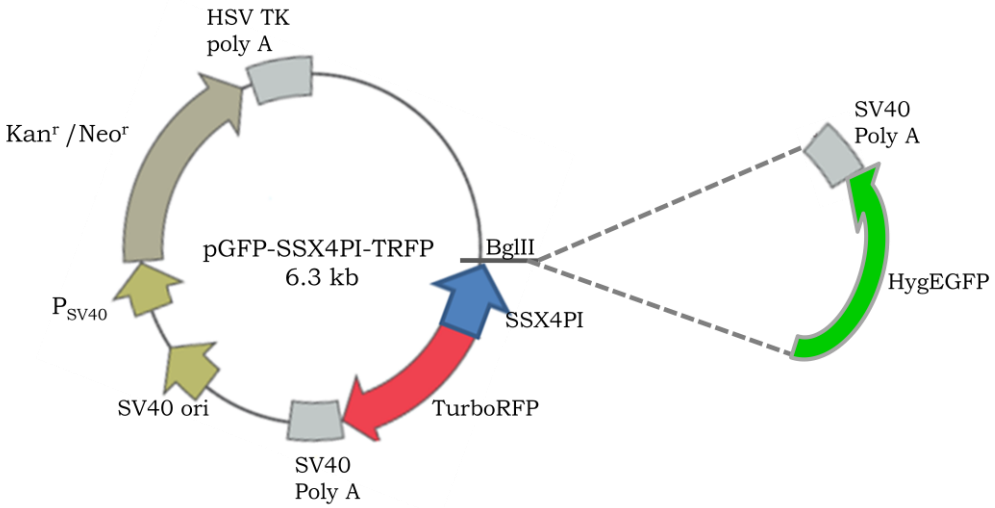


FIGURE 2.6.a. Map of pGFP-SSX4-TRFP vector in which HygEGFP expression is under control of SSX4 sense promoter and TRFP expression is under control of SSX4 antisense promoter

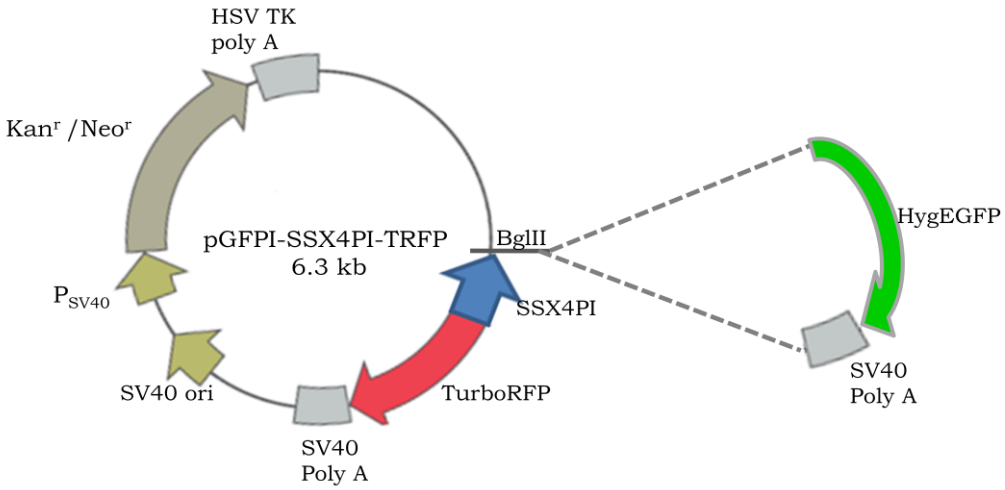


FIGURE 2.6.b. Map of pGFPI-SSX4-TRFP vector in which orientation of HEGFP is reversed. TRFP expression is under control of SSX4 antisense promoter.

2.5.3 Flow cytometry

Flow cytometry was performed in a BD FACSCalibur system. SK-LC-17 cells were transiently transfected with either pHygEGFP, pCMV-TRFP, pSSX4P-TRFP, pSSX4PI-TRFP, pGFPI-SSX4PI-TRFP or pGFP-SSX4PI-TRFP**.** Two days after the transfection, the cells were harvested by trypsinization and resuspended in 1X cold PBS. After resuspension, the cells were kept on ice and in the dark throughout the procedure. pHygEGFP and pCMV-TRFP constructs were used to adjust the instrument settings so that false positives for FL-1(GFP) and FL-2(RFP) were kept at a minimum. After the correct adjustments 200000 transiently transfected SK-LC-17 cells were used for the reading and the results were gated and evaluated accordingly.

2.5.4 Fluorescent Microscopy

For fluorescent microscopy analysis, SK-LC-17 cells that were seeded on slides and transiently transfected with pHygEGFP, pCMV-TRFP, pSSX4P-TRFP, pSSX4PI-TRFP, pGFPI-SSX4PI-TRFP and pGFP-SSX4PI-TRFP were collected two days after the transfection, washed two times with 1X PBS and fixed with cold methanol for 10 minutes, followed by two more washes PBS and observed under a Zeiss fluorescent microscope.

To observe nuclei, cells were stained with DAPI and a permeabilization step was added to the procedure. Following fixation, cells were incubated with permeabilization solution for 30 minutes and washed 2 times with 1X PBS before DAPI staining and fluorescence microscopy.

2.6. DICER Knock down experiment by shDICER plasmids

2.6.1 Vectors used

pSicoR human Dicer1 (Evrogen Plasmid#14763), pSicoR human Dicer2 (Evrogen Plasmid#14764), pSicoR human Dicer3 (Evrogen Plasmid#14765) were purchased from Addgene and were used previously as described (Kumar MS, 2007)

2.6.2 Generating stable clones

pSicoR vectors were transfected into SK-LC-17 cells as described in the section 4.5 but using a 6-well plate format. The day after the transfection, cells were trypsinized and reseeded in 15 cm^3 dishes in 1:2, 1:5 and 1:10 dilutions. Clone selection was initiated the following day by adding 0.25µg/µl puromycin to the cells. Cell death was obvious after 24 hours and resistant colonies were found to appear at day 7 to 10 (Optimization of selection dosage was performed independently). Individual colonies were picked under the microscope and transferred with a micropipette into a single well of a 96-well plate. Growing colonies were transferred 75 cm^3 dishes. RNA samples were collected as explained in section 2.1.

2.6.3 Real-Time PCR Analysis of DICER1

The primers that were used to detect DICER expression are shown in **Table 4.** For real-time PCR analysis, efficiency of DICER1 primers were determined (**Appendix B**) after plotting the efficiency curve (**Figure S5**). For the analysis, 1:5 diluted SK-LC-17 cDNA was used. 2 housekeeping genes, GAPDH and 18S rRNA were used as internal controls. Real-time PCR experiments were carried out in a Bio-Rad iCycler under the conditions of 94°C for 10 min followed by 35 cycles of 94°C for 60s, 60°C for 60 s, 72°C for 60s, with a final extension at 72°C for 10 min followed by melting curve analyses. SyBR Green dye (DyNAmo HS SYBR Green qPCR Kit, Finnzymes #F-410L) was used to quantitate the PCR product. Genes were normalized according to delta-delta Ct method (**Appendix B**)

2.7 Bioinformatic Analyses

2.7.1 Promoter Analysis

Promoter Analysis of SSX4 and NY-ESO-1 was performed by Genomatix Software Gene2Promoter Database. The predicted promoters for SSX4 and NY-ESO-1 were retrieved from this database and analyzed in detail.

2.7.2 Noncoding RNA Analysis

Noncoding RNA sequences were compared with sequences between the SSX4-A2 to A4 in various noncoding RNA databases using the BLAST algorithm. These databases include "Noncoding RNA database" ((http://biobases.ibch.poznan.pl/ncRNA/) and "Mammalian Noncoding RNA Database" ((http://research.imb.uq.edu.au/rnadb/default.aspx). The latter includes all the noncoding RNAs found to date by different methods and offers a very convenient BLAST option along with the UCSC track.

2.7.3 Microarray Analysis

The microarray data has been retrieved from Geo Accession Number: GSE6427 and analyzed by GENESPRING according to the "Analyzing Agilent Two Color Expression Data" section of GENESPRING manual. The data were then specifically filtered for CT genes.

2.8 RECIPES

50X TAE	242g Tris, 57.1 glacial CH_3COOH, 37.2G $Na_2EDTA.2H_2O$ in 1lt Working solution 1X diluted pH: 8,5
Luria Bertani Broth (LB)	10g tryptone, 10g NaCl, 10g Bacto Yeast Extract in 1lt 15g Bactoagar is added for LB agar plates
10X PBS	80g NaCl, 2g KCl, 17.8g $Na_2HPO4.2H_2O$, 2.4g KH_2PO_4 in 1lt Working solution 1X diluted pH: 7,4
Solution P1	50mM Glucose, 25mM Tris-HCl(pH: 8,0) 10mM EDTA (pH: 8,0)
Solution P2	0.2N NaOH, 1%SDS (from stock of 10%SDS) Prepare fresh
Solution P3	60ml 5M Potassium Acetate, 11.5ml glacial acetic acid, 28.5ml deionized water
Permeabilization Solution	10ml 2%BSA, 10ml 1X PBS, 50µl Tween-20

2.9 TABLES

TABLE 1. OATL and GAPDH PCR primers

Primer	Sequence	Tm	Amplicon length
SSX4/OATL-A2	5’-GTCCCTGGATCTTGGGTTAAAC	59.53	750 bp
SSX4/OATL-B2	5’-GTAATTGTGCGCACCACACTTAC	60.30	
SSX7/OATL-A	5’-AACAGATTTTTAACCCAAGATCCAGG	62.40	184 bp
SSX7/OATL-B	5’-GGAGGGCCTTGGACTGTTTTTTCAGTA	67.70	
OATL-B4	5’-CGCAAATGTGCTAGTTTGGAA	59.71	175 bp
OATL-A4	5’-CTGAGGACCACCCACTTTGT	58.43	
GAPDH-F	5’-GGCTGAGAACGGGAAGCTTGTCAT	67.48	150 bp
GAPDH-R	5’-CAGCCTTCTCCATGGTGGTGAAGA	67.39	

TABLE 2. PCR primers for SSX4 promoter mapping (sense orientation)

Name of the primer	Sequence of the primer	Tm
SSX4 A3.3 XhoI*	5’-ATATTTCTCGAGCTGATGGCGCCTGAGG	72.16
SSX4 A3.31 XhoI*	5’-ATATTTCTCGAGACCACCCACTTTGTCGCACA	71.80
SSX4 A3.32 XhoI*	5’-ATATTTCTCGAGCTGGAGGAGGCGACAACATT	71.80
SSX4 A3.33 XhoI*	5’-ATATTTCTCGAGCTGCAATGTCACTGCCCAAG	71.80
SSX4 A3.34 XhoI*	5’-ATATTTCTCGAGGATGATGGACCAATCAGGGC	69.24
SSX4 A3.35 XhoI*	5’-ATATTTCTCGAGAGTGAACTCCATCTGGCCAA	70.52
SSX4 A3.36 XhoI*	5’-ATATTTCTCGAGGTCAGAACAGTAGGCGGAAC	71.80
SSX4 A3.37 XhoI*	5’-ATATTTCTCGAGGAACAAGGGAAGCTGATGTG	69.24
SSX4 B4 HindIII**	5’-CACACAAAGCTTCCTAGAGCCTGGACTGACAGACG	73.93

*XhoI site CTCGAG (with the additional ATATTT sequence for restriction enzyme digestion)

**HindIII site AAGCTT (with the additional CACACA sequence for restriction enzyme digestion)

TABLE 3. PCR Primers for SSX4 Promoter Mapping (Antisense Orientation)

Primer	Sequence	Tm
SSX4 A3.3 HindIII**	5'-CACACAAAGCTTCTGATGGCGCCTGAGG	72.16
SSX4 A3.31 HindIII**	5'-CACACAAAGCTTACCACCCACTTTGTCGCACA	71.80
SSX4 A3.32 HindIII**	5'-CACACAAAGCTTCTGGAGGAGGCGACAACATT	71.80
SSX4 A3.33 HindIII**	5'-CACACAAAGCTTCTGCAATGTCACTGCCCAAG	71.80
SSX4 A3.34 HindIII**	5'-CACACAAAGCTTGATGATGGACCAATCAGGGC	69.24
SSX4 B4 XhoI *	5'- ATATTTCTCGAGCCTAGAGCCTGGACTGACAGACG	72.76

*XhoI site CTCGAG (with the additional ATATTT sequence for restriction enzyme digestion)

**HindIII site AAGCTT (with the additional CACACA sequence for restriction enzyme digestion)

TABLE 4. Quantitative RT-PCR Primers for DICER1

Primer Name	Sequence	Tm	Amplicon size
DICER1_FW	5'-GAAGCTGGCAAACAAGATCC	58.31	236 bp
DICER1_R	5'-GTGGGCAAATCAAAACGAACC	62.25	

3. PRELIMINARY DATA AND RATIONALE

SSX4 gene was identified in 1997 (Gure AO, 1997). Since SSX4 is a typical CT gene with frequent expression in tumors (Tureci O, 1998), and co-expressed with other CT genes (Gure AO, 2005), it was chosen as a model to decipher the mechanisms relevant to CT gene expression in cancer. Subsequent to the characterization of the gene, its transcription start site(s), as well as promoter and enhancer regions were identified, and DNA methylation levels of the promoter-proximal regions were studied in context of its expression (unpublished data) The rationale of this thesis is based on these data. These earlier results that eventually lead to the hypothesis of this thesis are summarized below.

3.1 General architecture of the SSX4 promoter

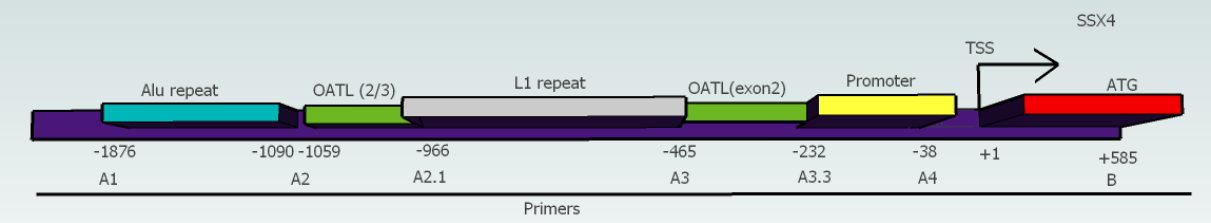


FIGURE 3.1. The general architecture of SSX4 promoter. Previously identified functional regions are shown in boxes. Sequences with promoter activity are depicted in yellow. Exon1 and 2 are shown as a single block in red. Ornithine Aminotransferase Like (OATL) sequences are in green. An L1 repeat (depicted in grey) disrupts the OATL sequence. Alu repeats at the 5' of the promoter are shown in blue. Primers used for generating reporter construct (A1-4 [forward], B [reverse]) and the major transcription start site (TSS) is indicated.

All SSX genes (SSX1-9) share common promoter architecture as depicted in **Figure 3.1.** The previously characterized SSX4 promoter was defined by generating luciferase reporter constructs with the primers shown in **Figures 3.1 and 3.2**. 5' to the promoter an OATL pseudogene, corresponding to the exon2 and part of the exon3 of the ornithine aminotransferase gene, followed by Alu repeats and finally by CT repeats exists (**Figure 3.1**).

3.2 Characterization of the SSX4 minimal promoter

Experiments aiming the characterization of the SSX4 promoter utilized those primers shown in **Figures 3.1 and 3.2**. As can be seen in Figure 3.2, the choise of "forward" primers was based on the inclusion of individual structural elements that possibly could affect SSX4 promoter activity. An initial set of luciferase reporter experiments utilizing constructs generated by the A1, -2, 3 and -4 forward primers with the B reverse primers as tested in the melanoma cell line SK-MEL-37 indicated that the minimal promoter had to be within A4 and B, and that the elements 5' to the promoter had repressive activity (**Figure 3.3**). Generating shorter constructs with A4.1, -4.2, -4.3, or -4.4 in combinations also resulted in loss of promoter activity (data not shown). However, using A3.3 primers in combination with A4 (in reverse orientation) resulted in maximum promoter activity in the same cell line (**Figure 3.3**). Interestingly reversing this construct such that the same promoter sequence was in antisense orientation in reference to the reporter, resulted in low, albeit significant promoter activity (**Figure 3.3**).

FIGURE 3.2. Sequence of the SSX4 promoter-proximal region. Exons 1 and 2 (containing the translation initiation codon) are in yellow. OATL pseudogene sequences in blue. Primers used for reporter vector construction are shown in boxes. Primers used for the initial experiments are in red characters, those used to define the minimal promoter are in blue and pink. The right and left arms of the 8 Alu repeats are in bold and italics, respectively.

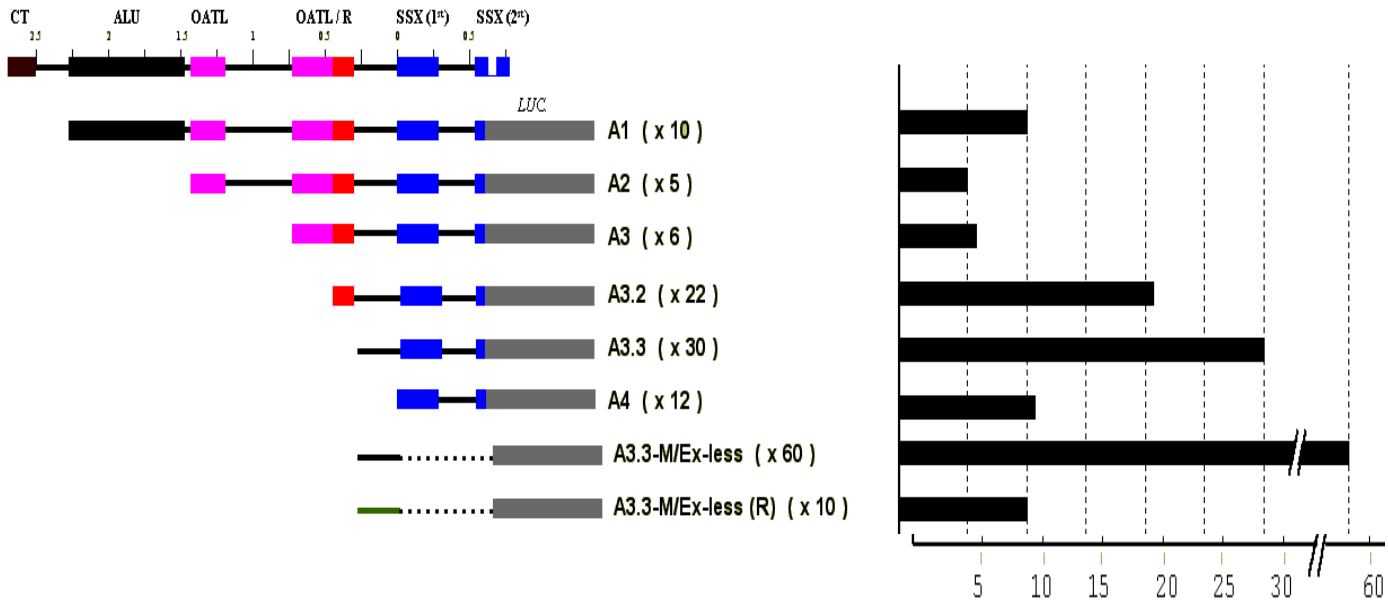


FIGURE 3.3. Luciferase activity of SSX4 promoter-reporter constructs. SSX4 gene architecture is shown on top. The corresponding reporter constructs contain the indicated portion of the promoter and the luciferase reporter (grey box). The A3.3-m/Ex-less reporter vector contains SSX4 promoter sequences between primers A3.3 and A4. The A3.3-m/Ex-less(R) vector has the same sequences but in antisense orientation. Luciferase activity of individual vectors is shown on the right as relative luciferase units.

3.3 Promoter activity of various SSX genes in different cell lines

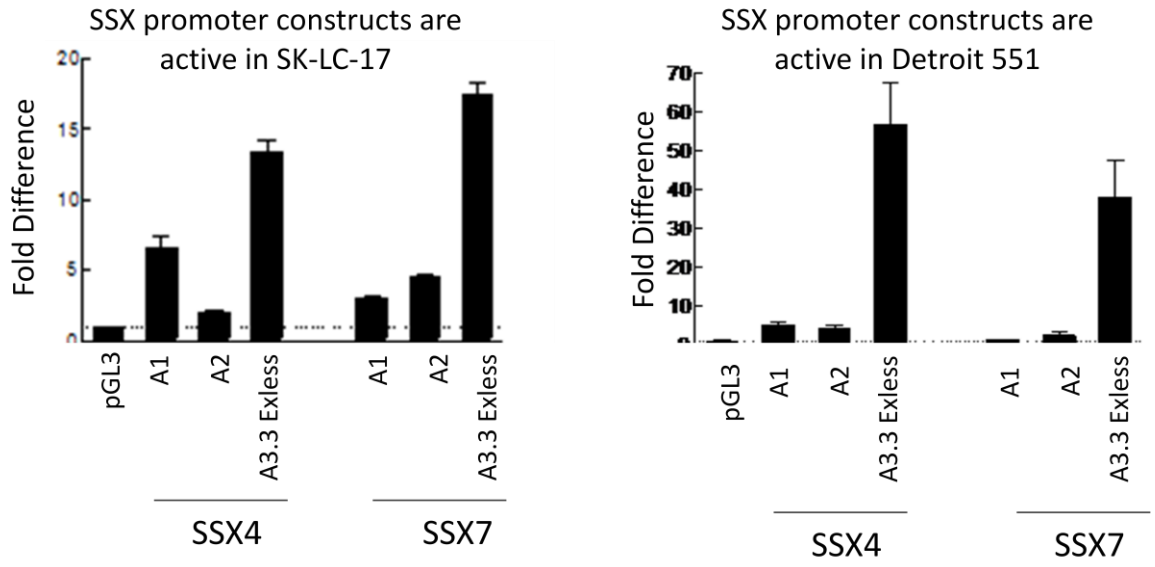


FIGURE 3.4. Luciferase activity of SSX4 and SSX7 promoter constucts in SK-LC-17 (left) and Detroit 551 (right). Vectors are identical to those described in **Figures 3.2** and **3.3**. Promoter activity is shown as relative luciferase units on the Y axis.

It was of interest to compare SSX promoter activity among different SSX genes, since their sequences are almost identical. To test for such potential differences, reporter constructs were generated from promoter sequences of the SSX7 gene corresponding to A1-B, A2-B and A3.3-A3 sequences of SSX4, and tested for reporter activity. SSX7 has been shown to be expressed in testis, but its expression is absent from all other normal and cancer tissues and cell lines including SK-LC-17 (Güre AO, 2002 & unpublished data not shown). As shown in **Figure 3.4**, luciferase activity of the reporter constructs with promoter sequences of SSX4 and SSX7 were indistinguishable. This experiment was performed in two cell lines: SK-LC-17 and the untransformed fibroblastoid cell line Detroit 551 (**Figure 3.4**). SK-LC-17 does express most CT genes including SSX1, -2, -3, -4 and 5 readily, whereas Detroit 551 has no detectable expression of any CT gene (data not shown).

In summary, the SSX4 and -7 promoters show reproducible and ubiquitous activity when part of episomal vectors, in contrast to their strict regulation *in vivo*. The two promoters also share a minimal promoter region that functions bidirectionally, and also various repressive elements that, nevertheless, do not block SSX promoter activity fully when part of the same reporter construct.

3.4. Characterization of repressive elements of the SSX4 promoter

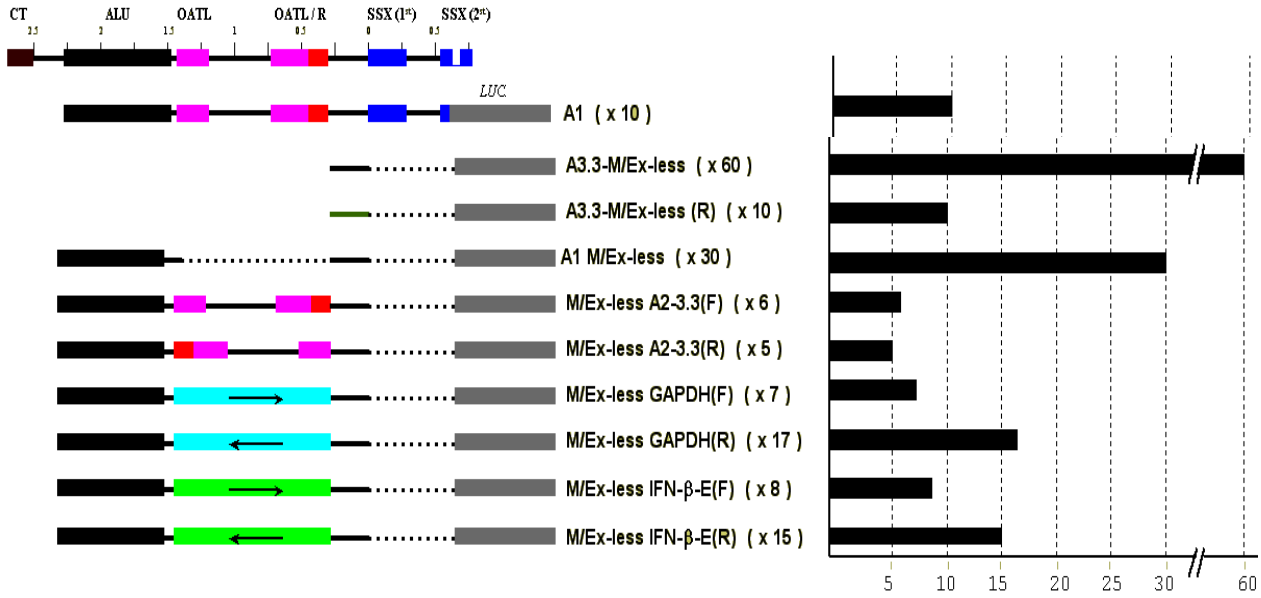


FIGURE 3.5 Analysis of the repressive element 5' to the minimal promoter. The repressive element (A2-A3.3) containing the OATL sequence and the L1 repeat was eliminated or replaced with various other sequences, including GAPDH (red) and the IFN -MAR element (purple), in the sense or anti-sense orientation.

Since the above described experiments identified the A3.3-A4 region as the minimal promoter, and the presence of repressive elements, to characterize the nature of this repressive element a second set of constructs, all sharing the A4 site as their most 3' end were constructed. The most repressive element in the SSX4 promoter (A2-A3.3) was then replaced with various sequences to test for a change in repression. Summary of these experiments are shown in **Figure 3.5**. These experiments demonstrated that reversing the orientation of the repressive sequence, or replacing it with one that contained a matrix-associating region (MAR) did not change repression significantly. In fact, cloning a house-keeping gene's (GAPDH) coding regions in either orientation could reconstitute the repressive effect as well. This suggested that the presence of "any" sequence, or the presence of a transcript spanning these regions itself could be related to the transcriptional control of the promoter.

3.5 Characterization of the transcription start sites of the bidirectional SSX4 promoter *in vivo*

To test for the presence of a transcript initiated in either orientation from the SSX4 promoter, 5'RACE analyses that could detect trancripts generated in either orientation were performed. Transcripts corresponding to 4 minor and 1 major transcription start sites were identified in the sense orientation. In addition, an antisense transcript originating within the minimal promoter was also identified (**Figure 3.6**).

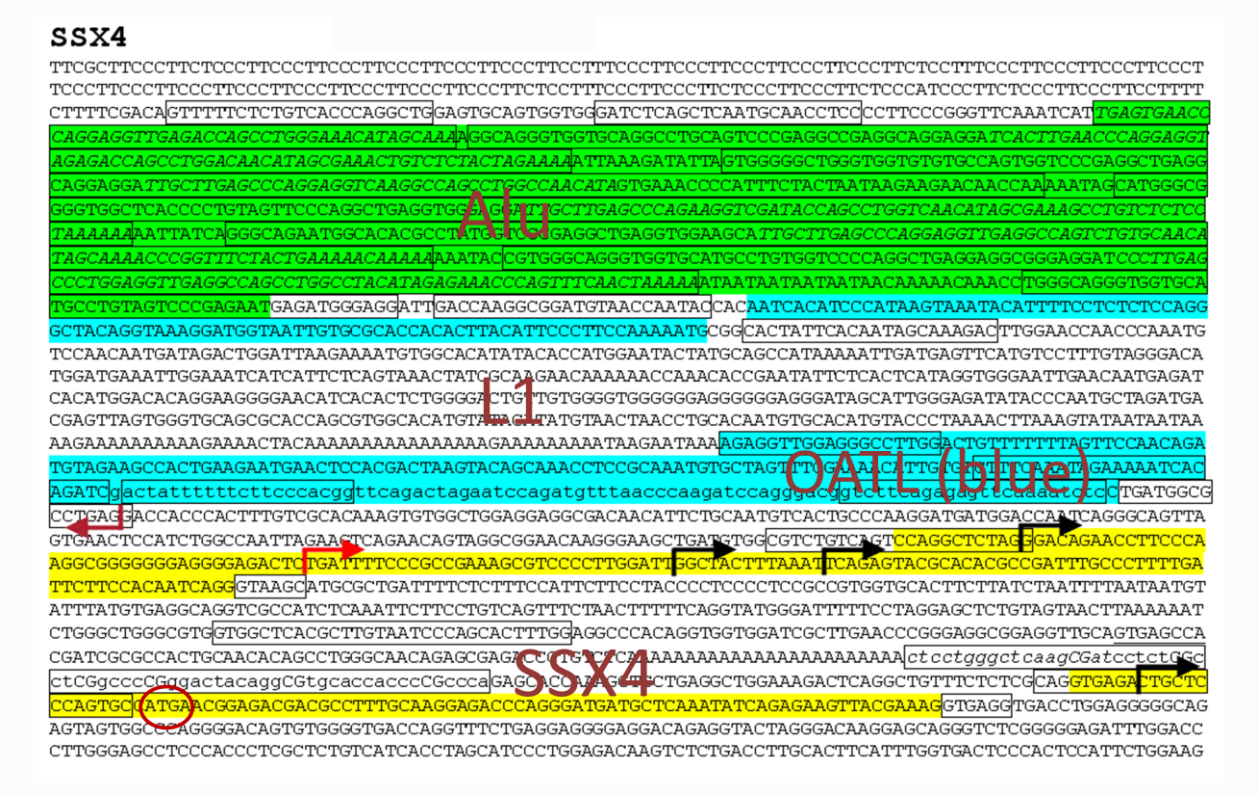


FIGURE 3.6. Transcription Start Sites identified by RACE analysis

The minor TSSs are depicted in black and the major TSS is depicted in bright red. The TSS of the OATL transcript can be seen in brown. Green sequences are Alu repeats and blue sequences are OATL sequences. The first ATG of the SSX4 gene is shown in a red circle.

3.6 Objectives and Rationale

The main objective in our research is to understand the mechanisms underlying the regulation of cancer-testis antigen gene expression. Previous data showed that CT antigen gene regulation has an epigenetic component. Specifically, CT gene promoter proximal sequences are consistently hypermethylated in normal tissues other than testis and become hypomethylated in cancer (Scanlan MJ, 2002). Similarly, TSA or other histone deacetylase inhibitors have consistently shown to induce CT gene expression (Scanlan MJ, 2002). Although hypomethylation phenomena, either global or gene specific, is known for years, the mechanisms underlying this process are not known. Although repeat regions are typical targets of tumorigenesis-associated hypomethylation, they are extremely difficult to study due to the fact that their genes and transcripts are so numerous. Although cancer testis antigens are generally known for their immunogenicity in cancer, they are also ideal tools to study the mechanisms of hypomethylation since they are individual genes.

The first promoter analysis of a CT gene was performed in 1995 with MAGE-A1 gene and MAGE-A1 promoter was transcriptionally active in tumor cells that did not express the MAGE-A1 gene (De Smet C, 1995). Considering that CT antigens are coordinately expressed (Gure AO, 2005), this observation should be valid for other CT genes as supported by our results with SSX4 and SSX7 promoter constructs (**Figure 3.4**). This suggests the ubiquitous presence of transcription factors acting on these promoters and that the genes are epigenetically regulated. We hypothesized that the mechanism involved in the epigenetic regulation of CT antigen genes partially be contained with the promoter itself. Previous data demonstrated a bidirectional promoter (see **Figure 3.3**) and a repressive regulatory mechanism that could be under the control of antisense promoter activity. To further understand the structure of the bidirectional promoter, we wanted to characterize the critical sequences responsible for the promoter activity in each direction within the minimal promoter. We then wanted to test if the SSX4 promoter could generate sense and antisense transcripts simultaneously and whether these could result in the generation of an intra-promoter dsRNA molecule. If SSX4 gene expression regulation is related to the generation of a ncRNA by the bidirectional promoter, then we wanted to determine whether this was a DICER-dependent event. We also aimed a bioinformatic search in hope of identifying structural similarities among CT-gene promoters that could possibly explain a pattern of similar behaviour within this group of genes.

4. RESULTS

4.1 Mapping the Bidirectional Elements of the SSX4 Basal Promoter

Earlier experiments had mapped the minimal SSX4 promoter to sequences -232 to -38, with reference to the major TSS. To characterize the minimal elements necessary for transcription in both the sense and antisense orientations, I designed reporter constructs containing various subregions of the SSX4 minimal promoter as shown in **Figure 4.1** and tested their reporter activity when transiently transfected into the cell line SK-LC-17.

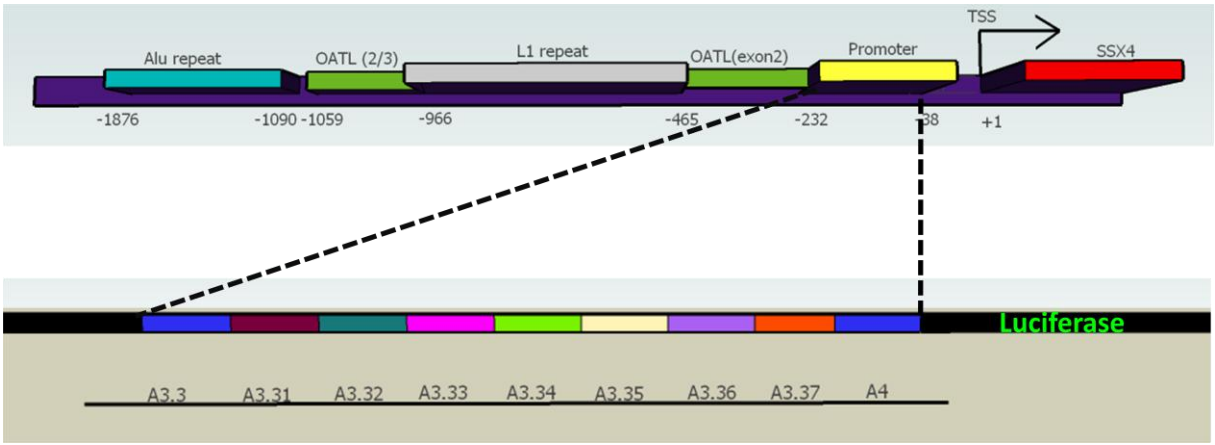


FIGURE 4.1 Sub-regions within the SSX4 promoter that were individually analyzed for promoter activity. Oligonucleotide primers corresponding to various sites within the minimal promoter are shown as colored boxes.

4.1.1 Mapping SSX4 promoter elements responsible for transcription in the sense orientation

I generated the deletion constructs shown in **Figure 4.2** by PCR to identify the important elements in the sense promoter using the primers shown in **Table 2.** The resulting constructs were pGL3-SSX4A3.3, pGL3-SSX4A3.31, pGL3-SSX4A3.32, pGL3-SSX4A3.33, pGL3-SSX4A3.34, pGL3-SSX4A3.35, pGL3-SSX4A3.36, pGL3-SSX4A3.37, where the designation refers to forward primer used for PCR. All constructs shared the same reverse primer (A4). Resulting PCR fragments were cloned into the pGL3-Basic plasmid and then assayed for their ability to transcribe luciferase gene. The relative luciferase units obtained from the constructs are shown as fold difference in **Figure 4.3**.

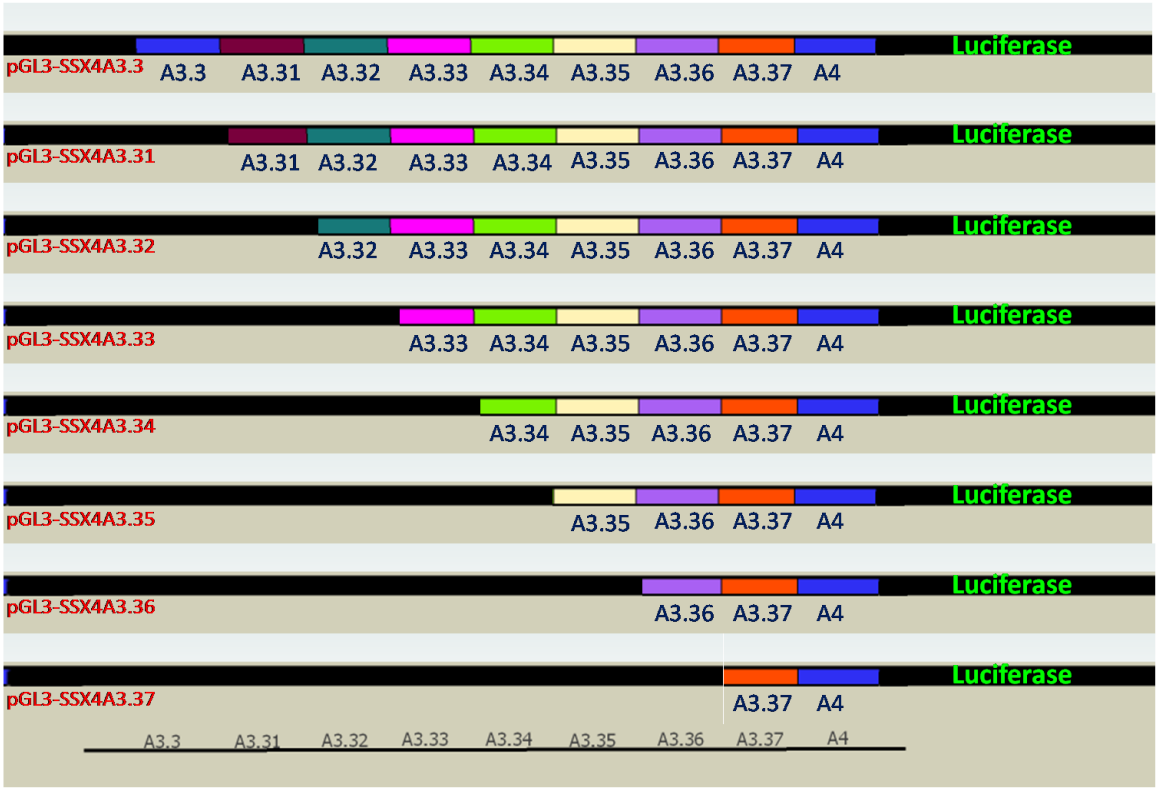


FIGURE 4.2. Sense promoter mapping constructs. The SSX4 minimal promoter (A3.3 to A4, 194bp) is shown on top. Mapping constructs share in common the most 3' region of the minimal promoter (A4), and are 177bp (A3.31 to A4), 150 bp (A3.32 to A4), 130 (A3.33 to A4), 110 bp (A3.34 to A4), 86 bp (A3.35 to A4), 60 bp (A3.36 to A4) and 44 bp (A3.37 to A4).

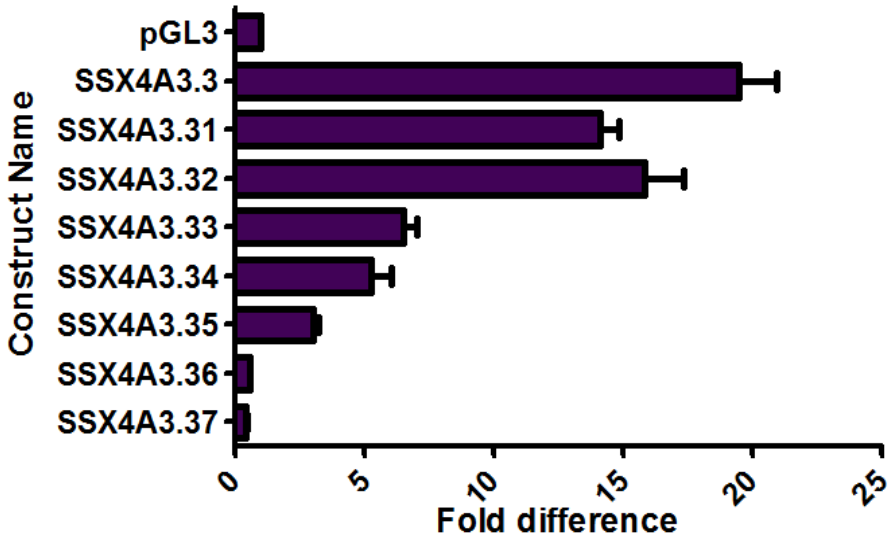


FIGURE 4.3 Promoter activity in the sense orientation of A3.3-A4 sub-fragments. Bars represent relative luciferase units. The average values obtained from three experiments are shown.

Data presented in **Figure 4.3** shows that the luciferase reporter activity of SSX4 sense promoter loses about 20% of its activity upon the eliminating A3.3 and A3.31 sequences but decreases to about 30% of its total activity if the A3.32 sequence is missing from the construct as well. Additional sequences within which scarce promoter activity reside in were A3.34 and A3.35. In summary, the major promoter activity in the sense orientation resides within the A3.32 sequences, with lesser activity in A3, A3.34 and A3.35.

4.1.2 Mapping SSX4 promoter elements responsible for transcription in the antisense orientation

I generated a second group of constructs shown in **Figure 4.4** by PCR to identify the important elements in the antisense promoter shown in **Table 3**. The resulting constructs were pGL3-SSX4A3.3R, pGL3-SSX4A3.31R, pGL3-SSX4A3.32R, pGL3-SSX4A3.33R and pGL3-SSX4A3, where the designation refers to the forward primer used for PCR. All constructs shared the same reverse primer (A4). The letter R designates that these sequences are in antisense orientation (reverse). The PCR fragments were cloned into the pGL3-Basic plasmid to test their promoter activity.

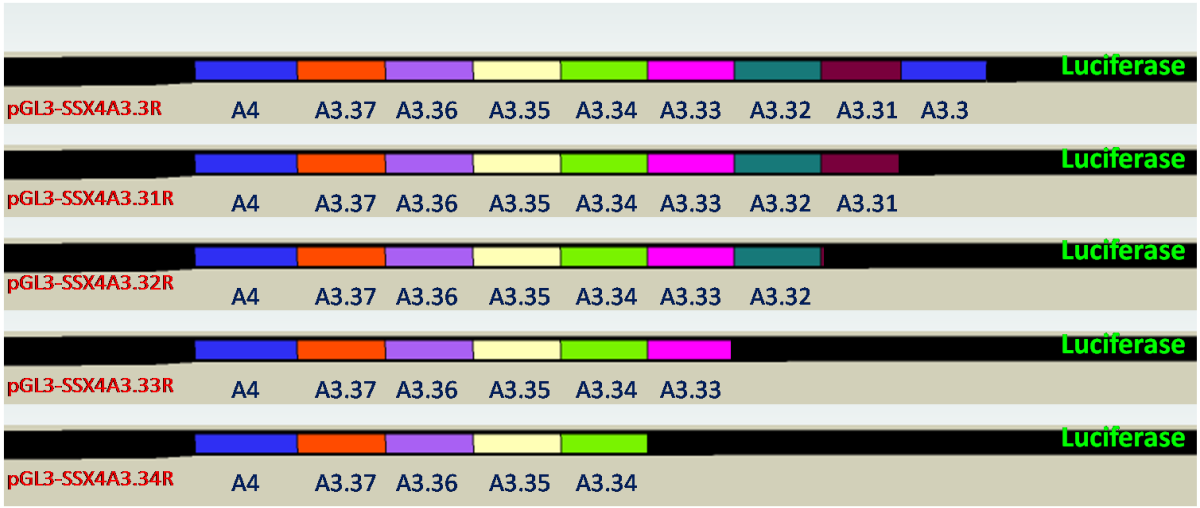


FIGURE 4.4 Antisense promoter mapping constructs. Mapping constructs share in common the most 5' region of the minimal promoter (A4) and are 177bp (A4 to A3.31), 150 bp (A4 to A3.32), 130 (A4 to A3.33), and 110 bp (A4 to A3.34) in length.

Since the SSX4 promoter activity in antisense orientation is lower compared to that in sense (**Figure 3.4**), I hypothesized that the promoter activity would more likely reside close to the 3' end of the inverted promoter. Therefore, constructs prepared for these experiments were designed to maintain 5' sequences, while gradually eliminating 3' sequences (**Figure 4.4**). The constructs were then assayed for their ability to transcribe the luciferase reporter gene. The relative luciferase units obtained from the constructs are shown as fold difference in **Figure 4.5**.

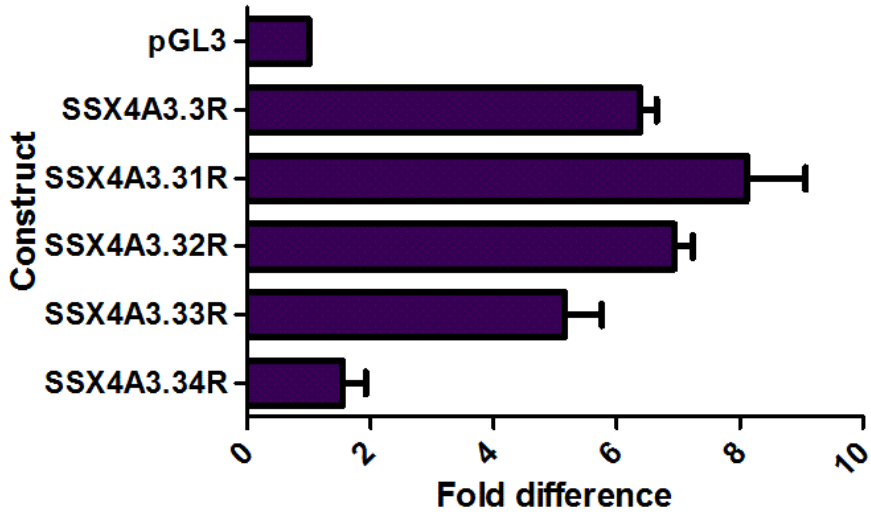


FIGURE 4.5. Promoter activity in the antisense orientation of A3.3-A4 sub-fragments. The average of three experiments is shown.

As hypothesized, the promoter activity of SSX4 antisense promoter was found to reside mostly in 3' regions (**Figure 4.5**). Antisense promoter activity was not significantly affected by the lack of A3.3 or A3.31, in fact there was a slight gain of activity when sequences within A3.3 region were lost. The loss of A3.32 decreased the activity of the antisense promoter about 20% but less than 25% of the promoter activity remained if A3.33 sequences were eliminated (**Figure 4.5**). In summary, the major promoter element for anti-sense transcription from the SSX4 minimal promoter resides within A3.33 sequences with minor activity in A3.32.

4.1.3 Overlapping promoter elements direct transcription in opposite orientations

As a result of these experiments, the sequences that show promoter activity can be summarized as shown in **Figure 4.6**

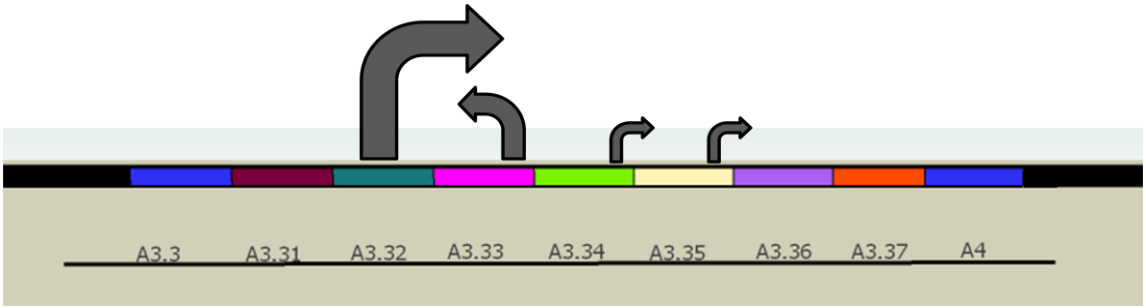


FIGURE 4.6. Summary of all results from sense and antisense promoter mapping experiments. The major promoter activity in the sense orientation resides within sequences within the A3.32 primer. Additional minor promoter activity is observed for A3.34 and -35 regions. Antisense transcription is primarily dependent on the A3.33

Our data shows that the SSX4 minimal promoter is clearly a bidirectional promoter with overlapping promoter activities within the same region. The main sense promoter activity and main antisense promoter activity actually resides in a 40 bp sequence. Thus it is likely that the SSX4 sense and antisense promoters interfere with eachother's promoter activity. Thus these experiments identified specific sequences that are primarily responsible for SSX4 promoter activity in both directions.

4.2 OATL Transcript Analysis

SSX4 promoter is found to be a bidirectional promoter that shows a promoter activity both for the SSX4 transcript and the putative OATL transcript that would contain sequences from the OATL pseudogene (**Figure 4.7**).

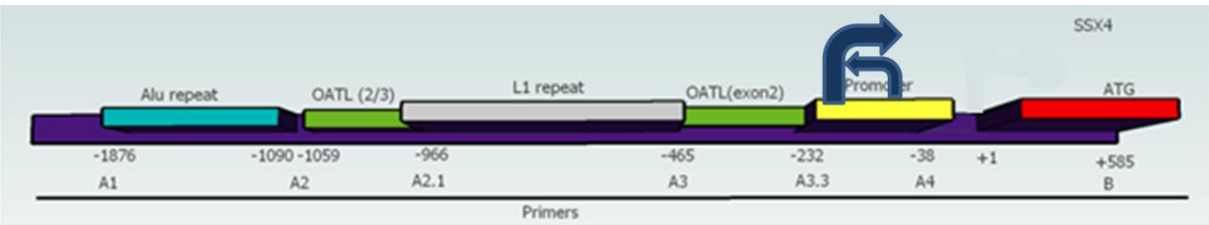


FIGURE 4.7 The two most active promoters in sense and antisense orientation upstream the SSX4 coding regions (red) are shown in context of the genomic DNA within which the promoter resides (in yellow). OATL pseudogene sequnces are in green, the L1 repeat in grey and the Alu repeat region in blue.

Considering that the antisense promoter is active *in vivo*, we expected to find an antisense transcript initiating from the SSX4 promoter that would correspond to OATL pseudogene sequences. A very short transcript of this nature had previously been identified by RACE analysis (**Figure 3.6 and 4.8**).

FIGURE 4.8. OATL transcript identified by 5'RACE analysis. The transcript is highlighted in blue. Primer sequences used to construct the promoter reporter constructs are shown in blue font color and the name of the primers are shown in red.

I wanted to generate PCR primers by which I could determine and quantify the OATL transcript, which I hypothesized would be longer than that identified by RACE analysis. Such primers could then be used to test if the OATL transcript co-existed with the SSX4 mRNA and whether their transcription would occur in a mutually exclusive fashion. OATL sequences are highly conserved among other SSX genes (**Figure 4.9**). Therefore, the OATL sequences can only be distinguished by a few base pair differences that need to be considered if an OATL transcript initiating from a specific SSX gene is to

be amplified. The conservation between the OATL sequences of SSX4, SSX7, SSX9, SSX2, SSX3 and OAT is summarized in **FIGURE 4.9**

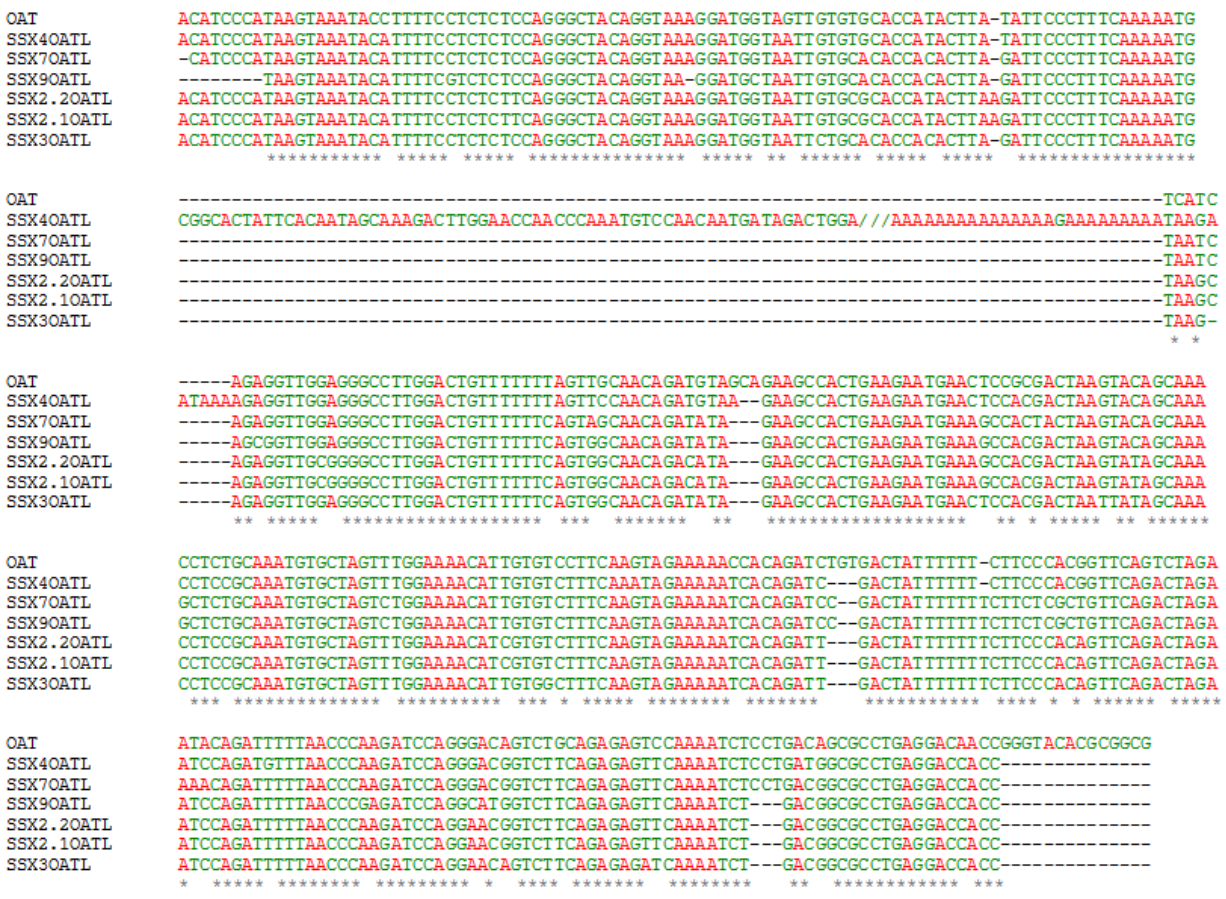


FIGURE 4.9 Homology of OATL sequences within the 5' regions of SSX2, -3, -4, -7 and -9 and to that of the Ornitine aminotransferase (OAT) mRNA are shown. There are two copies of SSX2, therefore, two OATL sequences are shown. Residues shared by all genes are highlighted by an asterisk. The sequence stretch unique to SSX4 corresponds to the LINE1 element.

The SSX4 OATL is disrupted by a LINE1 insertion. This repeat element is not present within OATL sequences of SSX genes other than SSX4 and therefore, seems to be a recent evolutionary event.

4.2.1 Detection of the SSX4/OATL transcript

Since SSX4 was the most intensely studied among this family of genes and since I already identified the minimal promoter element that could generate an OATL transcript from the SSX4 promoter (SSX4/OATL transcript), I decided to focus on the SSX4-SSX4/OATL transcript pair to study their combined regulation. I thus, designed primers (OATLA2-OATLB2) for SSX4/OATL such that the 3' end of the primer would be uniquely complementary to the OATL sequence of SSX4 (**Figure 4.10**).

FIGURE 4.10. OATLA2 (forward) and OATLB2 (reverse) primers in the context of homologous OATL sequences. OATLA2 was designed to contain a cytidine at its 3' end (top of figure). OATLB2 contained cytidine at its 3' end (bottom of figure); to ensure amplification of the specific sequence.

The OATLA2 and -B2 primers were designed to amplify sequences immediately upstream of the A3.3 primer site since these sequences corresponded to the 5' region of the SSX4/OATL transcript previously identified RACE (**Figure 4.8**). And, since the antisense SSX4 promoter was found to be contained primarily within sequences 3' to A3.3 (**Figure 4.6**).

FIGURE 4.11 OATLA2-OATLB2 primers in context of the SSX4 promoter region. OATLA2 and OATLB2 primers are highlighted in yellow. Primers used to construct reporter vectors are shown in blue letters and their designations are in red. The SSX4 minimal promoter is between the primers A3.3 and A4.

I hypothesized that this transcript would be uninterrupted; therefore, all the samples to be tested were treated with DNase prior to cDNA generation and genomic DNA contamination was assessed by reverse transcriptase-lacking controls. The quality and specificity of the primers were tested for their ability to specifically amplify the SSX4 reporter constructs (**Figure 3.3**) that possess SSX4-specific OATL sequences, and not those constructs containing SSX7 promoter regions that contain SSX7-specific OATL sequences (**Figure 4.10**). As expected, only when SSX4-A1 construct was used as the template, PCR amplification was observed using OATLA2-OATLB2 primer pair (**Figure S12**). To detect the OATL transcript specific for SSX4, different cell lines and conditions were used. SK-LC-17 cell line, which is known to express a high level of SSX4, was used in case the bidirectional promoter is active and working in both directions. HT29 cell line which does not express SSX4 unless it is treated with demethylating agent 5-aza-2'-deoxycytidine was also tested as this reagent would be expected to lead to the dereppression of the SSX4 promoter, possibly affecting the antisense promoter as well. However, we were unable to detect the presence of a SSX4/OATL transcript by these primers in any of the cell lines and/or conditions tested.

4.2.2 Detection of the SSX7/OATL transcript

SSX7 gene expression is seen in testis but SSX7 gene remains silent even in SSX4 gene expressing cancer cell lines. One possible explanation for this observation is that in tumors lacking SSX7 expression might be transcribing SSX7/OATL instead. In this manner, I wanted to design primers (SSX7/OATLA-SSX7/OATLB) to the OATL transcript specific for SSX7 (SSX7/OATL). In both SSX7/OATLA and SSX7/OATLB primers, the 3' end of the primer is uniquely complementary to OATL sequence of SSX7 (**Figure 4.12**).

FIGURE 4.12 SSX7/OATLA (forward) and SSX7/OATLB (reverse) primers in the context of homologous OATL sequences. SSX7/OATLA was designed to contain an adenine at its 3' end (top of figure). SSX7/OATLB contained a thymidine at its 3' end (bottom of figure); to ensure amplification of the specific sequence.

The quality and specificity of the primers were tested using the same methodology as described for OATLA2 & -B2 primers. As expected, the primers only gave a product when the SSX7 construct was

used as the template but not when SSX4 was used. To detect the OATL transcript specific for SSX7, the same cell lines and conditions (SK-LC-17, HT29 and 5-aza-2'-deoxycytidine treated HT29) were used. However, no transcript could be detected in any of the cell lines or conditions tested.

4.2.3 Detection of a universal OATL transcript

Since we could not detect the SSX4/OATL transcript using the OATLA2-OATLB2 primer pair, a new primer pair that could amplify any SSX/OATL transcript was designed. I thought that a reason the OATLA2, -B2 primer was not able to amplify a product was possibly that the target RNA was shorter. The new primer pair was, therefore, designed to yield a shorter product (**Figure 4.13 and 4.14**).

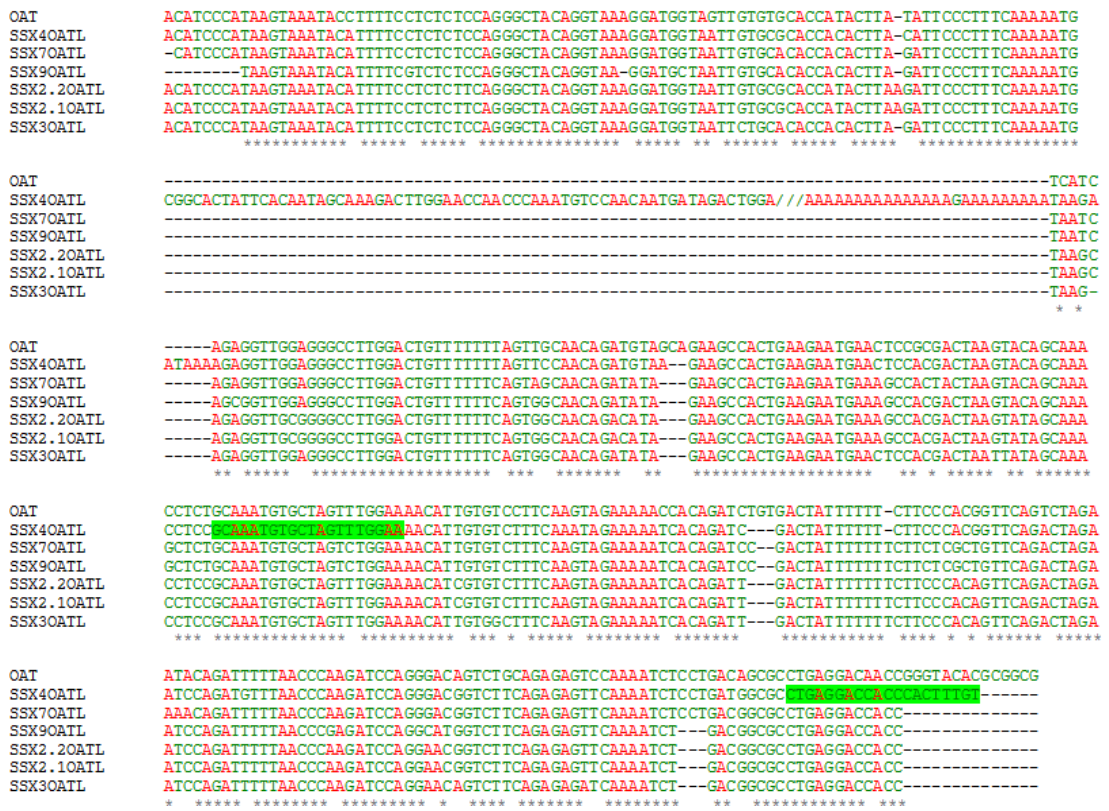


FIGURE 4.13 OATLA4 (forward) and OATLB4 (reverse) primers in the context of homologous OATL sequences. OATLA4 and OATLB4 are highlighted in green

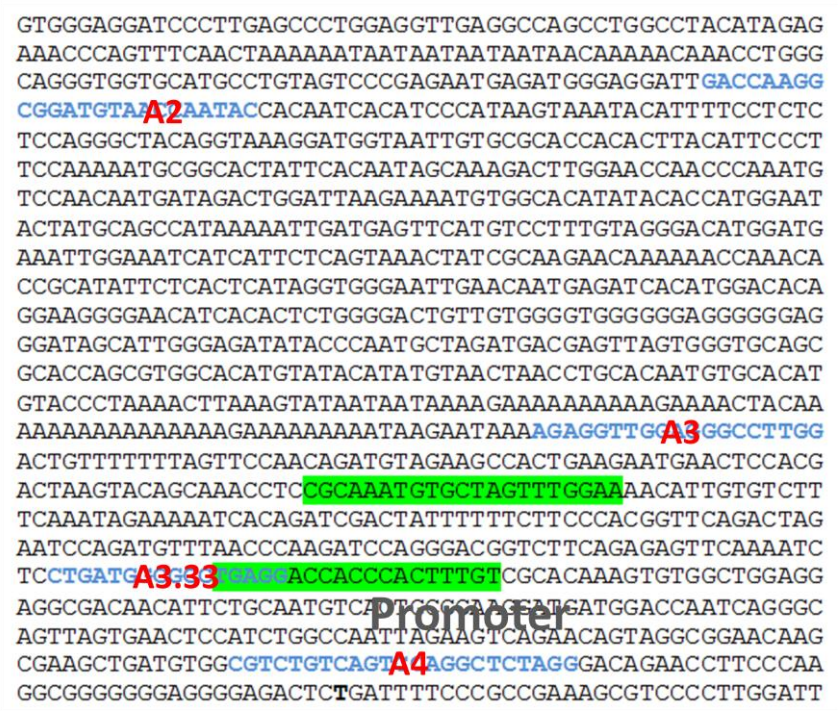


FIGURE 4.14 OATLA4-OATLB4 primers in the context of SSX4 promoter region. OATLA2 and OATLB2 are highlighted in yellow.

The quality of the primers were tested by using the luciferase reporter construct SSX4-A1 (**Figure 3.3**) which possesses OATL sequences of SSX4 and SSX7-A1 which possesses OATL sequences of SSX7 as templates. As expected, PCR amplification is detected in both templates. The same cell lines and conditions (SK-LC-17, HT29 and 5-aza-2'-deoxycytidine treated HT29) were used but the transcript could not be detected in any of the cell lines or conditions tested.

4.2.4 Detection of OATL transcripts using bioinformatics as a tool

Transcripts deposited into noncoding RNA (ncRNA) databases were searched for a transcript corresponding to the OATL sequences. These databases offer BLAST algorithms to compare putative ncRNAs with verified ncRNAs. One of these databases is the noncoding RNA database (http://biobases.ibch.poznan.pl/ncRNA/). The database includes over 30,000 individual sequences from 99 species of Bacteria, Archaea and Eukaryota. Sequences between the SSX4-A2 to -A4 were used to search for homologous ncRNAs using the BLAST algorithm, but no human ncRNA matched to these sequences.

Another database I used was mammalian ncRNA database (RNAdb) (http://research.imb.uq.edu.au/rnadb/default.aspx). This database includes sequences and annotations for tens of thousands of ncRNAs including microRNAs, small nucleolar RNAs, larger mRNA-like ncRNAs, PIWI-interacting RNAs, ncRNAs identified from the latest rounds of large-scale cDNA sequencing projects, putative antisense transcripts, as well as ncRNAs predicted on the basis of structural features and alignments. Sequences homologous to sequences between the SSX4-A2 to -A4 were searched by using the BLAST algorithm but no human ncRNA matched these sequences. I also searched for the presence of ncRNAs in the vicinity of the SSX4 and SSX3 gene by the UCSC BED files. I was not able to identify any ncRNA transcripts in the vicinity of the SSX genes using this approach either.

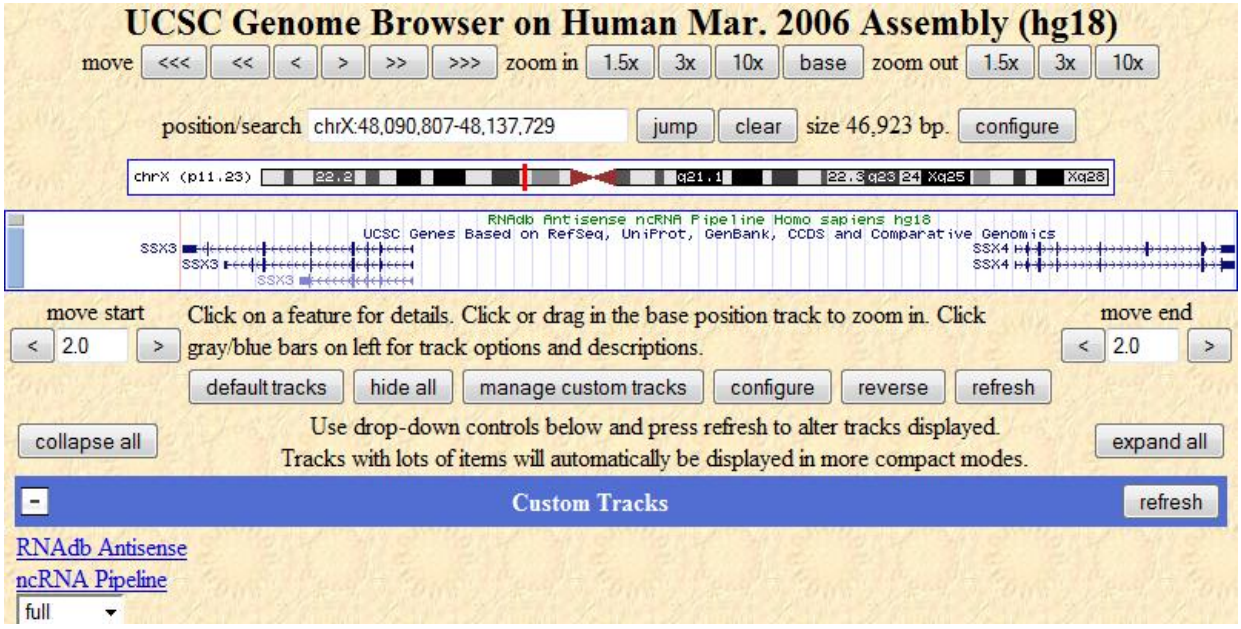


FIGURE 4.15. UCSC Browser showing UCSC genes and RNAdb Antisense ncRNA in the location chrX:48,090,0807-48,137,729. USCS genes are shown in blue in the second row. In this region of X chromosome SSX3 (another SSX gene family member) and SSX4 genes reside. The ncRNAs found in the database normally shown in green in the first row but the track is empty indicating that no ncRNA in the database corresponds to the 5' sequences of SSX4 or SSX3.

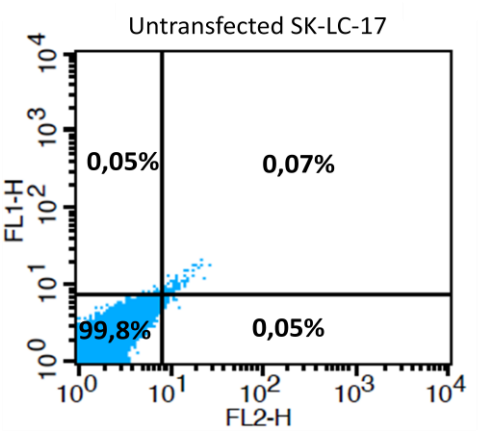
4.3. Analysis of bidirectional SSX4 promoter activity by flow cytometry and fluorescent microscopy

Although our reporter construct analysis of promoter activity revealed a bidirectional SSX4 promoter, we were not able to identify an OATL transcript and were, therefore, unable to determine whether the promoters are antiregulated so that when the sense promoter is active, antisense promoter is not active or vice versa, or whether they were co-regulated so that when the promoter is active as a whole, both sense and antisense promoter activity would be observed and when the promoter is not active no promoter activity would be seen. We, therefore, hypothesized that the antisense transcription could contribute to the regulation of sense promoter expression. This contribution may stem from "transcriptional interference" where the regulation of one promoter depends on the transcription from other promoter or a possible dsRNA from the activity of both of the promoters simultaneously. To test these possibilities, we generated a bi-reporter construct in which the activity of SSX4 minimal promoter is assessed both in the sense and antisense direction with the help of two different reporter genes. To assess the sense promoter activity, we used Enhanced Green Flourescent Protein (EGFP) and to assess the antisense promoter activity we have used Turbo Red Flourescent Protein (TRFP). Since TRFP is a more powerful reporter than the EGFP, it was chosen as the reporter of the weaker antisense promoter.

The generation of the bi-reporter system utilized two controls. The first contruct is pHygEGFP in which the EGFP is under the control of the strong immediate early promoter of human cytomegalovirus (CMV). The second control construct is generated in our laboratory, pCMV-TRFP, in which TRFP is under the control of strong immediate early promoter of human cytomegalovirus (CMV) (**Figure 2.3**). These two controls are used not only to assess the integrity of EGFP and TRFP but for calibration purposes, mosly of flow cytometry. The basic plasmid that has been used in the generation of SSX4 promoter constructs was pTurboRFP-PRL. This plasmid does not contain a promoter sequence and enables evaluation of the promoter activity by ligating the promoter of interest into the multiple cloning site of the plasmid. pSSX4P-TRFP and pSSX4PI-TRFP constructs were generated for testing SSX4 promoter activity in the sense and antisense orientation respectively (Figures 2.2 and 2.3), and served as quality control experiments as well. An additional reporter vector was generated by inserting the HygEGFP reporter downstream the SSX4 promoter of pSSX4PI-TRFP, generating pHEGFP-SSX4PI-TRFP (**Figure 2.6a**) that was suitable to test both promoter activities simultaneously. A control of this vector that contains the EGFP in antisense is shown in **Figure 2.6b**.

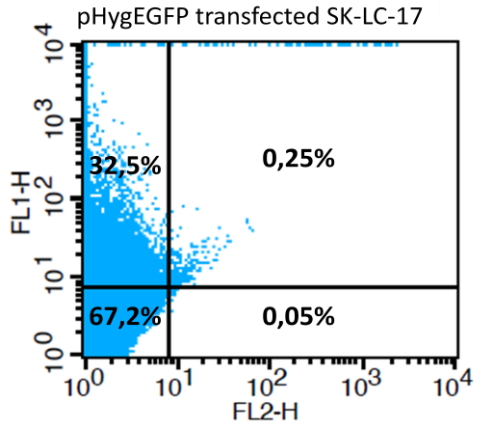
4.3.1 Flow Cytometry

For flow cytometry analysis SK-LC-17 cell line was used. Maximum excitation and emission values of TurboRFP protein are 553 nm and 574 nm respectively. Maximum excitation and emission values of EGFP protein are 488 nm and 509 nm respectively. Accordingly, two different lasers were used to detect GFP and RFP. The first laser (FL1) is used to detect EGFP and the second laser (FL2) has been used to detect TRFP. The calibration for compensation between these lasers was performed by using pHygEGFP and pCMV-TRFP to minimize autofluorescense. After the calibration according to the cell line and dual-fluorescence, 200.000 cells were read and gated for further analysis.



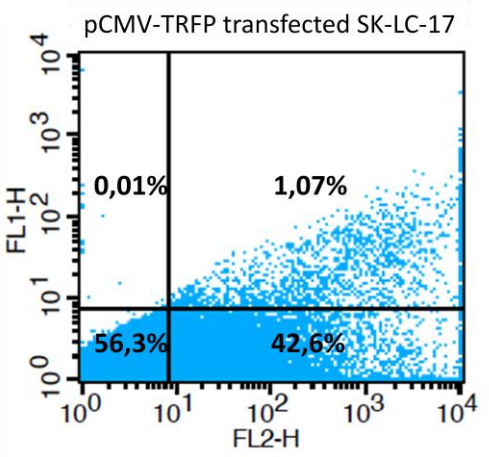
Quad	Events	% Gated	% Total
UL	48	0.05	0.03
UR	73	0.07	0.05
LL	106357	99.83	77.35
LR	57	0.05	0.04

FIGURE 4.16 Flow cytometry results of untransfected SK-LC-17 cells. The upper left quadrant, lower right quadrant and upper right quadrant refers to GFP positive cells, RFP positive cells and double positive cells respectively.The percentages and the details of the reading can be observed in the table.



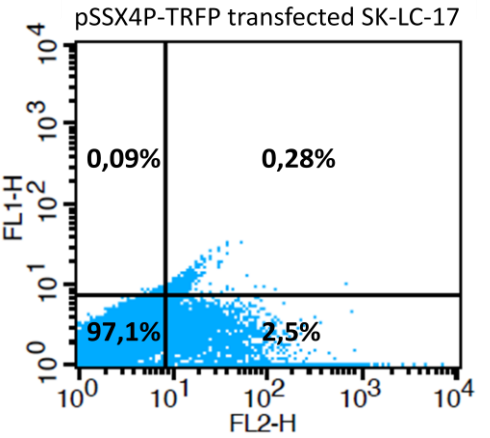
Quad	Events	% Gated	% Total
UL	50563	32.54	25.28
UR	387	0.25	0.19
LL	104387	67.17	52.19
LR	73	0.05	0.04

FIGURE 4.17 Flow cytometry results of pHygEGFP transfected SK-LC-17 cells. The upper left quadrant, lower right quadrant and upper right quadrant refers to GFP positive cells, RFP positive cells and double positive cells respectively.The percentages and the details of the reading can be observed in the table.



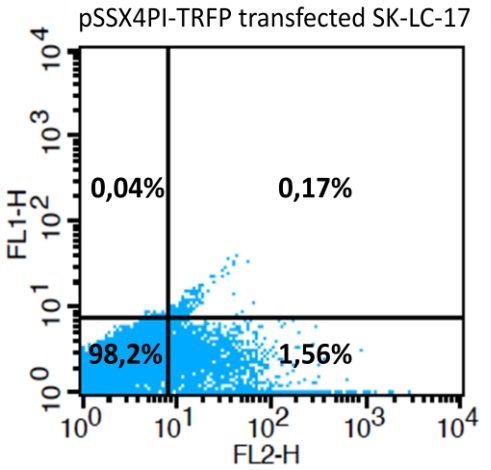
Quad	Events	% Gated	% Total
UL	20	0.01	0.01
UR	1721	1.07	0.86
LL	90641	56.32	45.32
LR	68570	42.60	34.28

FIGURE 4.18 Flow cytometry results of pCMV-TRFP transfected SK-LC-17 cells. The upper left quadrant, lower right quadrant and upper right quadrant refers to GFP positive cells, RFP positive cells and double positive cells respectively.The percentages and the details of the reading can be observed in the table.



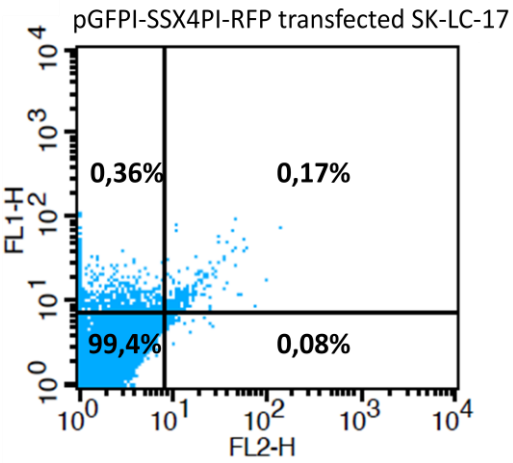
Quad	Events	% Gated	% Total
UL	136	0.09	0.07
UR	418	0.28	0.21
LL	145582	97.15	72.79
LR	3724	2.48	1.86

FIGURE 4.19 Flow cytometry results of pSSX4P-TRFP transfected SK-LC-17 cells. The upper left quadrant, lower right quadrant and upper right quadrant refers to GFP positive cells, RFP positive cells and double positive cells respectively.The percentages and the details of the reading can be observed in the table



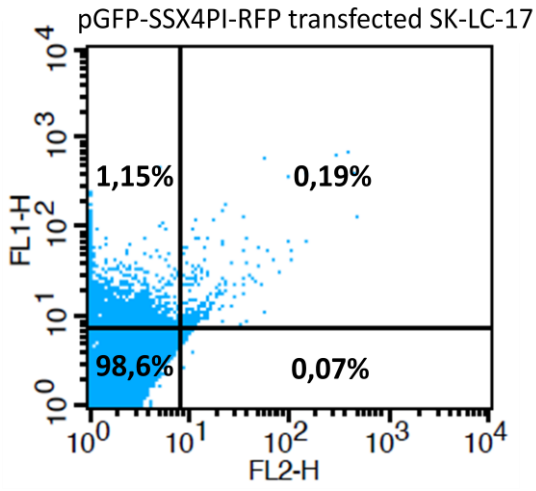
Quad	Events	% Gated	% Total
UL	71	0.04	0.04
UR	276	0.17	0.14
LL	155169	98.22	77.58
LR	2462	1.56	1.23

FIGURE 4.20 Flow cytometry results of pSSX4PI-TRFP transfected SK-LC-17 cells. The upper left quadrant, lower right quadrant and upper right quadrant refers to GFP positive cells, RFP positive cells and double positive cells respectively.The percentages and the details of the reading can be observed in the table



Quad	Events	% Gated	% Total
UL	590	0.36	0.29
UR	282	0.17	0.14
LL	162580	99.39	81.29
LR	123	0.08	0.06

FIGURE 4.21 Flow cytometry results of pGFPI-SSX4PI-TRFP transfected SK-LC-17 cells. The upper left quadrant, lower right quadrant and upper right quadrant refers to GFP positive cells, RFP positive cells and double positive cells respectively.The percentages and the details of the reading can be observed in the table



Quad	Events	% Gated	% Total
UL	1877	1.15	0.94
UR	307	0.19	0.15
LL	160688	98.59	80.34
LR	114	0.07	0.06

FIGURE 4.22 Flow cytometry results of pGFP-SSX4PI-TRFP transfected SK-LC-17 cells. The upper left quadrant, lower right quadrant and upper right quadrant refers to GFP positive cells, RFP positive cells and double positive cells respectively.The percentages and the details of the reading can be observed in the table

These results show that both EGFP and TRFP are expressed under the control of CMV promoter and that the transfection efficiencies for these constructs are 30 to 45% (**Figure 4.17 and 4.18**). TurboRFP is expressed under the control of both the SSX4 sense (**Figure 4.19**) as well as the antisense promoter (**Figure 4.20**). As expected, the antisense promoter activity is lower than the sense promoter activity. When we compare the fluorescence that our double reporter constructs show, we can clearly see that EGFP is expressed in the pGFP-SSX4PI-TRFP construct concomitant to the loss of TRFP transcription (**Figure 4.22**). The EGFP expression is significantly less in the control pGFPI-SSX4PI-TRFP construct where the direction of EGFP is reversed (**Figure 4.21**). Interestingly when we compare the results for pSSX4PI-TRFP and pGFP-SSX4PI-TRFP constructs, we can see that although the antisense promoter is capable of expressing TRFP when there is no other reporter gene in the construct, upon the addition of EGFP sequence (under the control of sense SSX4 promoter), TRFP expression is abolished.

4.3.2 Fluorescent Microscopy

Fluorescent microscopy is a very convenient method to observe the changes in fluorescence activities. In our case, the aim for using fluorescent microscopy experiments was to investigate the activities of sense and antisense promoter in a single cell basis enabling us to observe if both of the promoters are active in one cell simultaneously; and if they are active to observe the localizations of the activities. Unfortunately, the low sensitivity of the microscopy prevented us from observing low promoter activities conferred by SSX4 promoter. Only EGFP and TRFP expression under CMV promoter were detectable under the microscopy considering that CMV promoter is a far more powerful promoter than SSX4 promoter.

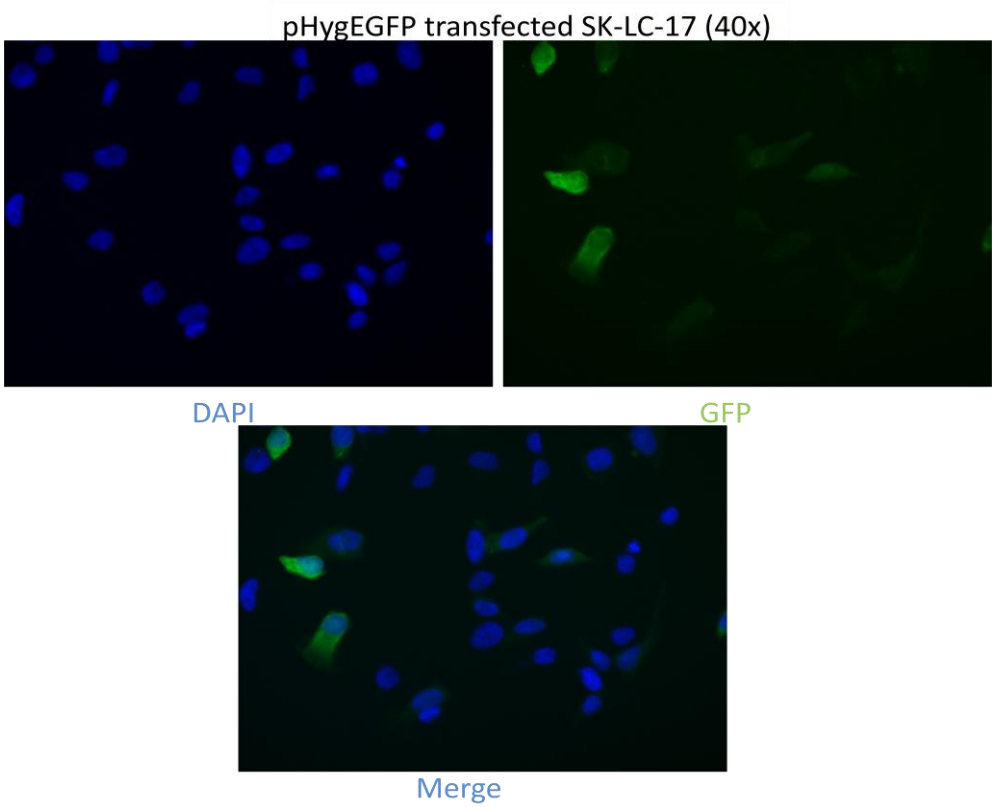


FIGURE 4.23 Fluorescent microscopy images of pHygEGFP transfected SK-LC-17 cells. To visualize the cells, nuclei are stained with DAPI (upper left) (4OX)

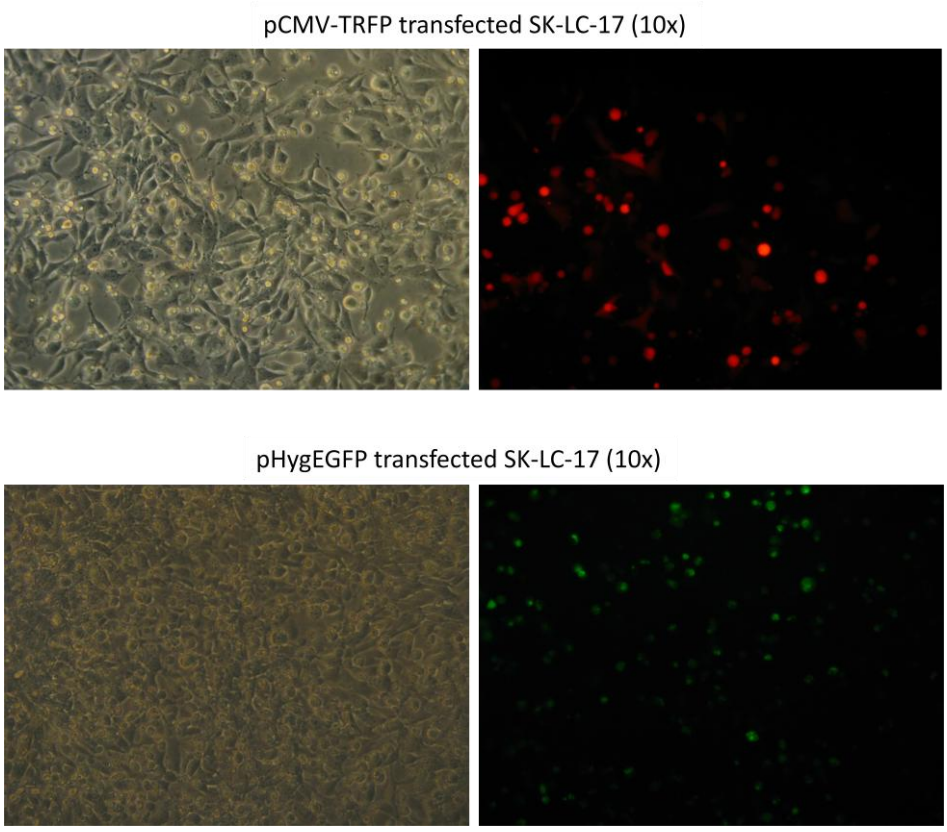


FIGURE 4.24 Images from light and fluorescence microscopy of pHygEGFP or pCMV-TRFP transfected SK-LC-17 cells. Light and fluorescence microscopy images were taken from the same samples.

Although the signals from other constructs were barely detectable, we could observe a low level of EGFP expression in the pGFP-SSX4PI-TRFP transfected SK-LC-17 cells.

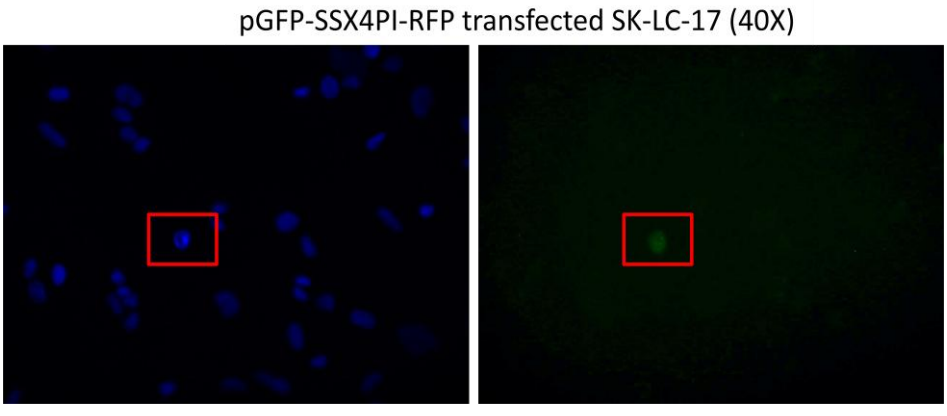


FIGURE 4.25 Fluorescent microscopy images of pGFP-SSX4PI-TRFP transfected SK-LC-17 cells. The red rectangular indicates a GFP positive cell.

4.4 Search for bidirectional promoter by using bioinformatics as a tool

Bioinformatics emerge to be a very powerful tool for promoter based analysis. Along with many other tools, bioinformatics enable us to predict the promoters of genes by offering promoter prediction tools. For the promoter prediction, I chose the Genomatix Software. For the prediction of SSX4 promoter, -2107 to +875 bp of transcription start site of the SSX4 gene was used. This sequence was given to the Genomatix Software and three potential promoters were identified by this analysis. Two of them are in the sense strand of SSX4 and one of them is in the antisense strand. The predicted promoter regions regarding the A3, A3.3 and A4 primers are shown in **Figure 4.26**. Note that the SSX4 minimal promoter sequence is between A3.3 and A4 primers.

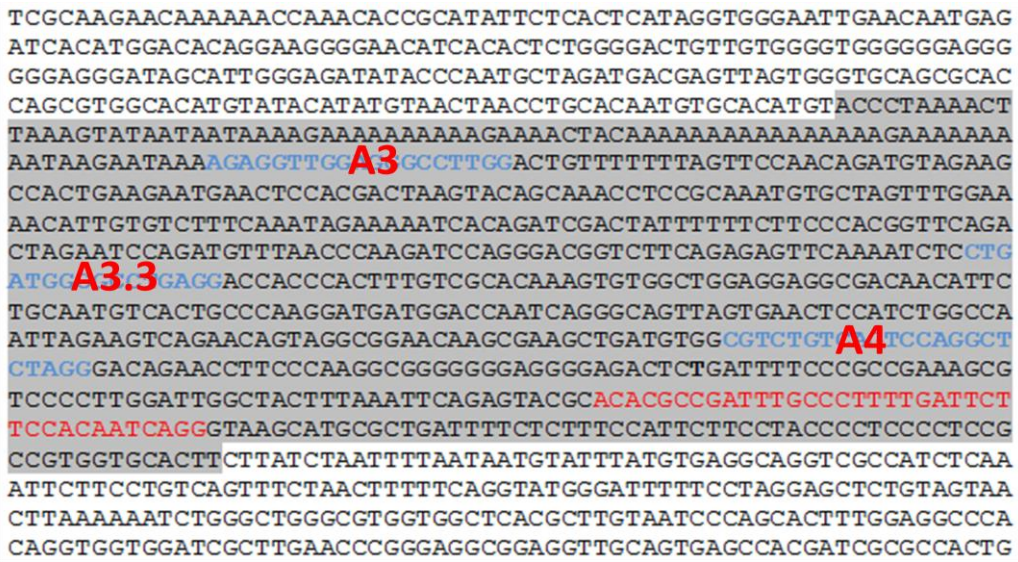


FIGURE 4.26. Predicted SSX4 promoters by Genomatix Software. Predicted sense strand promoters are highlighted in yellow and predicted antisense promoter is highlighted in gray. The first exon of the SSX4 gene is depicted in red font. (NCBI NM_005636)

Besides being located inside the transcripted sequences and exons of SSX4 gene, the first predicted sense promoter by Genomatix do not include SSX4 minimal promoter (A3.3 to A4) which was verified by luciferase reporter experiments and thus this predicted promoter will be ignored. The second predicted sense promoter actually includes SSX4 minimal promoter (A3.3 to A4). Also the genomatix software accurately predicts the antisense SSX4 promoter activity. This shows us that the promoter prediction tool of the Genomatix software can be considered as a strong promoter prediction tool. This tool can be used to predict the promoters of other CT genes and investigate the possibility that promoters of other CT genes are also bidirectional. The first CT to be investigated is the NY-ESO-1 gene since it is a very well known example of classic CT-X genes. Similar to the case with SSX4, the promoter prediction software of Genomatix predicted two promoters for NY-ESO-1. There is one sense and one antisense orientation (**Figure 4.27**). Thus, it seems like the promoter of NY-ESO-1 can be bidirectional like SSX4 promoter. This possibility should be confirmed by gene reporter systems.

CGACTAGGACAGGGACAGAACCCGTTGAACCCAGGAGTGAGATCCGGCCCCGGGTCCCGCTG
GGCCCTCCCGTCCACCTTGGCTGGACCTGGCGCCTGGGAGACCTTGGCTGGCGCGAGGCCAC
GCCCACCAGACATGCAGTTCCAGCTACCCCACCAGCTGGGCGACCAGGACAGGGACGGAGGC
TGCTGAGCCCAGTTAGAGGCCTGCCCCCCGGGGTCTGTCCTGGGCGCTCCCCCAAGGACGGA
CAGGGCAGGCAGGGTCCGGGACGATGGCCGCACAGTCCCGGCCCCGTGTTCCCAGGCCCGTC
TTGCTCCTCGATGTGAGGGAGACCCGGGGGATGGGACAGGCTGGGCCCCGCAGTGCCTGACT
CCCTGCAGGGCTCCCGGGACAGGGGTCCGGCGGACAGCCGGCTGCTCAGGGGTGAGGGGTCC
AAGCTGGCATTGCGGCCACCTTCCGGCCCGGGCTCTCTTGGGGAGGGGCGGGGTTGGTGAGA
ACCGGTCACGTGCTCCGGGGCTCACTCGGGGTCTCCCAGGGCCGGAAGTAGGGCCCCTGTGC
GCAGGCGCCCTGAGGATCCCGGGCTGCCCATCTCACGCCAGGGGGCGGAACTTCCTGCAGCC
TCTCTGCCTCCGCATCCTCGTGGGCCCTGACCTTCTCTCTGAGAGCCGGGCAGAGGCTCCGG
AGCCATGCAGGCCGAAGGCCGGGGCACAGGGGGTTCGACGGGCGATGCTGATGGCCCAGGAG
GCCCTGGCATTCCTGATGGCCCAGGGGGCAATGCTGGCGGCCCAGGAGAGGCGGGTGCCACG
GGCGGCAGAGGTCCCCGGGGCGCAGGGGCAGCAAGGGCCTCGGGGCCGGGAGGAGGCGCCCC
GCGGGGTCCGCATGGCGGCGCGGCTTCAGGGCTGAATGGATGCTGCAGATGCGGGGCCAGGG
GGCCGGAGAGCCGCCTGCTTGAGTTGTATTCTGTTCTGTTCTGTTCTGTTCTGACAGTTCTG
GTGGCGAGGTGGGGGCCGGGAGATGGGGAGGGCAGGGCCAGGTGGGGGAGGAGGCGGGGGAG

GCTGGTGGGAGGCCACCCGCAACCCACCCCACACGCAGCTCCAGCTCCCCCACCAGGCGGGG
CGACTAGGACAGGGACAGAACCCGTTGAACCCAGGAGTGAGATCCGGCCCCGGGTCCCGCTG
GGCCCTCCCGTCCACCTTGGCTGGACCTGGCGCCTGGGAGACCTTGGCTGGCGCGAGGCCAC
GCCCACCAGACATGCAGTTCCAGCTACCCCACCAGCTGGGCGACCAGGACAGGGACGGAGGC
TGCTGAGCCCAGTTAGAGGCCTGCCCCCCGGGGTCTGTCCTGGGCGCTCCCCCAAGGACGGA
CAGGGCAGGCAGGGTCCGGGACGATGGCCGCACAGTCCCGGCCCCGTGTTCCCAGGCCCGTC
TTGCTCCTCGATGTGAGGGAGACCCGGGGGATGGGACAGGCTGGGCCCCGCAGTGCCTGACT
CCCTGCAGGGCTCCCGGGACAGGGGTCCGGCGGACAGCCGGCTGCTCAGGGGTGAGGGGTCC
AAGCTGGCATTGCGGCCACCTTCCGGCCCGGGCTCTCTTGGGGAGGGGCGGGGTTGGTGAGA
ACCGGTCACGTGCTCCGGGGCTCACTCGGGGTCTCCCAGGGCCGGAAGTAGGGCCCCTGTGC
GCAGGCGCCCTGAGGATCCCGGGCTGCCCATCTCACGCCAGGGGGCGGAACTTCCTGCAGCC
TCTCTGCCTCCGCATCCTCGTGGGCCCTGACCTTCTCTCTGAGAGCCGGGCAGAGGCTCCGG
AGCCATGCAGGCCGAAGGCCGGGGCACAGGGGGTTCGACGGGCGATGCTGATGGCCCAGGAG
GCCCTGGCATTCCTGATGGCCCAGGGGGCAATGCTGGCGGCCCAGGAGAGGCGGGTGCCACG
GGCGGCAGAGGTCCCCGGGGCGCAGGGGCAGCAAGGGCCTCGGGGCCGGGAGGAGGCGCCCC
GCGGGGTCCGCATGGCGGCGCGGCTTCAGGGCTGAATGGATGCTGCAGATGCGGGGCCAGGG
GGCCGGAGAGCCGCCTGCTTGAGTTGTATTCTGTTCTGTTCTGTTCTGTTCTGACAGTTCTG
GTGGCGAGGTGGGGGCCGGGAGATGGGGAGGGCAGGGCCAGGTGGGGGAGGAGGCGGGGGAG
ATGCGAGTAAGTGGTTGGGTGGGGGTGGGGGTGGGGGTTGGGAGGTGGGGATATGAGAGGCC

FIGURE 4.27. Predicted NY-ESO-1 promoters by Genomatix Software. Predicted sense strand promoter is highlighted in yellow and predicted antisense promoter is highlighted in gray. The 5'UTR and the first exon of the gene is depicted in red font (NCBI NM_001327)

4.5 Investigating the Role of DICER1 in Regulation CT gene expression

DICER1 plays a vital role in most of the small RNA pathways we know. If small RNAs can have a role in the regulation of CT antigen expression, DICER1 may have a role. To test this possibility, we have planned to stably knock down DICER1 with the shRNA plasmids and investigate its effect on regulation of CT antigen expression.

pSicoR human Dicer1, pSicoR human Dicer2, pSicoR human Dicer3 plasmids were used to transfect the SK-LC-17 cells and stable clones were generated in the search for a DICER1 knockdown. These plasmids were used before with human cancer call lines and DICER1 knockdown clones were generated (Kumar MS, 2008), however, lentiviral infection, rather than transfection was usedin those experiments. Since DICER1 is required for the generation of its target small interfering RNAs, an incomplete knock down is expected. Kumar MS, 2008 used U20S osteosarcoma cells to infect with these plasmids and, thus, reported and DICER1 knock down in this cell line.

In our experiments, we transfected SK-LC-17 cells with these plasmids and generated stable clones but we were unable to find a clone in which DICER1 expression was downregulated as tested by Quantitative Real Time PCR to assess the mRNA levels of DICER1 (**Table S2**).

4.5.1 Investigating the effect of DICER on regulation of CT gene expression using bioinformatics as a tool

From NCBI GEO Database microarray data of the gene expression changes in human colorectal cell line HCT116 DICER Exon5 knockout cells have been acquired (Geo Accession Number: GSE6427). In this experiment Agilent-012391 Whole Human Genome Oligo Microarray G4112A platform has been used (Geo Accession Number: GPL1708). This data is from a two color microarray which is typically hybridized with cDNA prepared from two samples to be compared (control wild type HCT116 cells and HCT116 DICER Exon5 knockout cells in this case). The samples are labeled with two different fluorophores: Cy3, which has a fluorescence emission wavelength of 570 nm (corresponding to the green part of the light spectrum), and Cy5 with a fluorescence emission wavelength of 670 nm (corresponding to the red part of the light spectrum). These two Cy-labelled cDNA samples are mixed and hybridized to a single microarray that is then scanned in a microarray scanner to visualize fluorescence of the two fluorophores after excitation with a laser beam of a defined wavelength. Relative intensities of each fluorophore are then used in ratio-based analysis to identify up-regulated

and down-regulated genes. The data includes two samples, one of which is a dye –swap to prevent the dye bias in experiment.

The gene expression analysis has been performed according to the GENESPRING Manual "Analyzing Agilent Two Color Expression Data". The data files were imported and specifically filtered for CT genes. The dye swap data have also been analyzed and filtered for CT genes. Genes that were up or down regulated in this analysis are shown in **Figure 4.28** and **Figure 4.29** .

Gene Symbol	Common Name	Fold change(Cy3 vs Cy5)	Regulation(Cy3 vs Cy5)
MAGEE2	NM_138703	1,0723215	down
MAGEA12	NM_005367	1,1644258	down
SSX2	NM_175698	1,1867114	down
GAGE3	U19144	1,0351098	up
MAGEA6	NM_175868	1,187928	down
MAGEB2	NM_002364	1,059612	up
MAGEC1	NM_005462	1,0034466	down
MAGEC2	NM_016249	1,1625254	down
SSX3	NM_175711	1,1043183	down
MAGEA4	NM_002362	1,1130652	down
XAGE2	NM_130777	1,1039978	up
MAGED1	AF217963	1,1688178	down
SPANXA1	NM_013453	1,3766248	up
MAGEA9	NM_005365	1,4973171	down
MAGEC3	NM_138702	1,1367828	down
SSX4	NM_005636	1,1232446	down
MAGED1	NM_00100533	1,0754699	down
MAGED2	NM_201222	1,2406985	up
GAGE1	BC036094	1,3426495	down
MAGEA5	NM_021049	1,0328555	up
MAGEA8	NM_005364	1,2241	down
MAGEH1	NM_014061	1,1698048	down
XAGE3	NM_130776	1,4699597	up
SSX5	NM_021015	1,0430244	down
MAGEA10	NM_001011543	1,2618699	down
MAGEE1	NM_020932	1,0413634	up
MAGEA11	NM_005366	1,0641445	down
MAGEA10	ENS00000370323	1,0306301	down
MAGEL2	AJ243531	1,0356313	down
MAGEB1	NM_002363	1,180184	down
MAGED2	NM_201222	1,3382381	up
MAGEF1	NM_022149	1,2466332	down
MAGEA1	NM_004988	1,1318392	down

FIGURE 4.28 . GENESPRING analysis of the GSE6427 data filtered for CT genes

Gene Symbol	Common Name	Fold change(Cy3 vs Cy5)	Regulation(Cy3 vs Cy5)
MAGEE2	NM_138703	1,2440244	Down
MAGEA12	NM_005367	1,1792434	Up
SSX2	NM_175698	1,0615646	Down
GAGE3	U19144	1,1805035	Up
MAGEA6	NM_175868	1,17695	Up
MAGEB2	NM_002364	1,0937167	Down
MAGEC1	NM_005462	1,0580491	Down
MAGEC2	NM_016249	1,3717254	Down
SSX3	NM_175711	1,0832577	Up
MAGEA4	NM_002362	1,2823062	Up
XAGE2	NM_130777	1,242939	Up
MAGED1	AF217963	1,1033945	Down
MAGEA9	NM_005365	2,1945105	Down
MAGEC3	NM_138702	1,2561129	Down
SSX4	NM_005636	1,2788457	Down
MAGED1	NM_00100533	1,0690029	Up
MAGED2	NM_201222	1,1642103	Down
GAGE1	BC036094	1,2351307	Up
MAGEA5	NM_021049	1,2821825	Up
MAGEA8	NM_005364	1,0614607	Down
MAGEH1	NM_014061	1,2153622	Down
SSX5	NM_021015	1,1192068	Up
MAGEA10	NM_001011543	1,107301	Down
MAGEE1	NM_020932	1,1391786	Up
MAGEA11	NM_005366	1,4328613	Down
MAGEA10	ENS00000370323	1,276002	Down
MAGEL2	AJ243531	1,2093933	Down
MAGEB1	NM_002363	1,086816	Up
MAGED2	NM_201222	1,1140487	Down
MAGEF1	NM_022149	1,3169265	Up
MAGEA1	NM_004988	1,1839479	Up
NTN4	NM_021229	1,099939	Down

FIGURE 4.29 . GENESPRING analysis of dye-swap GSE6427 data filtered for CT genes

We hypothesized that if the deregulation of a small-RNA based mechanism led to the upregulation of CT genes in cancer, and if DICER had a role in maintaining this suppressive effect, that the loss of DICER protein could lead to the coordinate upregulation of CT gene expression. However, as can be seen from **Figures 4.28** and **4.29**, the knock-out of DICER does not result in a difference of CT gene expression larger than 1.5 fold (with the exception of MAGE-A9) and the effect is in both directions, suggesting non-specificity. Based on this analysis, it can be suggested that DICER knockout does not have an apparent effect on CT gene expression in HCT116 human colorectal cell line.

5. DISCUSSION AND FUTURE PERSPECTIVES

The main objective in this study was to characterize the SSX4 promoter in detail to understand the mechanisms underlying the regulation of CT antigen gene expression. Although it is clear that the mechanisms regulating CT gene expression (in a coordinate fashion) are primarily epigenetic in nature, the "initial signal" that initiates those events leading to the gene-specific epigenetic silencing of these genes is unknown. The primary hypothesis that led to all of the experiments contained within this thesis is that, the SSX4 promoter itself, is the initiator of the epigenetic silencing of the gene under normal conditions. The hypothesis is based on the fact that transcriptional activity of the antisense promoter interferes with the sense promoter, possibly involving the generation of a dsRNA molecule during the process. Very recently, noncoding RNAs are being studied for their role in regulating mammalian gene expression. By performing those experiments detailed in this thesis, several mechanisms that could lead to a better understanding of how the convergent bidirectional SSX4 promoter were tested or uncovered.

5.1 Mapping of the SSX4 Promoter

To characterize the bidirectional SSX4 promoter in detail, we identified the minimal sequences responsible for the sense and antisense transcription by using the luciferase reporter system. Two very short streches of DNA (~40bp) are responsible for most of the sense or the antisense transcriptional activity in the SSX4 promoter. There are other examples of such short sequences which are capable of transcribing both in the sense and antisense directions (Trinklein ND, 2004). The functions and working principles of these bidirectional promoters are not known exactly but mostly they seem to differ depending on the nature of the promoter. In fact, functionality of the bidirectional promoter can be affected by an antisense RNA which can mediate regulation of sense promoter or steric interference of RNA Polymerase may lead to inhibition of transcription in the sense or antisense promoter. Since we characterized the SSX4 promoter thoroughly (at 20 bps intervals), we believe that dissecting the promoters into shorter sequences is unlikely to reveal extra information. Instead, inhibiting the antisense promoter activity by site directed mutagenesis directed to A3.33 primer sequence might reveal a direct effect of the antisense promoter on the activity of the sense promoter. The mutant promoter could be tested as part of the luciferase constructs described. We would expect that this would lead to the derepression of the sense transcription by A2-A3.3 sequences.

We used two approaches to study the activity of the bidirectional promoter. By luciferase reporter experiments we were able to show that the antisense activity of the promoter is about 2/5th of that of the sense promoter. Previous experiments where the whole promoter was tested had revealed this ratio to be 1/10. However, the fluorescence ratios as tested by our EGFP/TRFP construct show that in this system the antisense promoter has an activity that is at most 1-2% of the sense promoter. In other words, we can observe TRFP expression under the control of antisense promoter alone but when the EGFP reporter is cloned under the control of sense promoter in the same construct, TRFP expression is dramatically reduced (**Figure 4.22** vs. **Figure 4.24**). Previous data had shown that an additional but nonspecific stretch of DNA added 3' to the antisense promoter would result in the inhibition of the sense promoter. Our data with the double fluorescent constructs show that this is true for the sense promoter as well: the activity of the antisense promoter is inhibited more than 90% by the addition of a reporter sequence 3' to the sense promoter. Therefore, it is possible that the mechanism of the corepression observed between the promoters in opposite direction is either dependent on the ability of the promoter to recruit RNA polymerase, or its ability to generate a transcript that has a function only if it is of a certain length or both. A third possibility is that the promoters, working simultaneously or in turns, lead to the generation of a dsRNA molecule

The loss of TRFP in the pGFP-SSX4PI-TRFP construct could be a result of transcriptional interference. The transcriptional interference is known to be asymmetric - a strong promoter (aggressive) reduces the expression level of a weak promoter (sensitive) as in our case, and the act of transcription itself rather than the sequence of the transcribed RNA is important. Transcriptional interference has been demonstrated convincingly before, by using viral promoters using the viral lytic promoter pR and the lysogenic promoter pL. When these two promoters were arranged face to face with their start sites 62 bp apart, pR activity was found to decrease pL activity 5.6 fold; however, pL activity did not affect pR driven transcription in this system (Callen BP, 2004). In this context, transcription from pR interferes with the transcription from pL. The situation is very similar to our observation but with mammalian promoters. There are a couple of theories for the reason under this transcriptional interference. One of the theories argues that this interference is a result of basic RNA Polymerase II activity and it can be caused by steric hindrance or sequestering of transcriptional machinery in these promoters that prevents the activity of other. Another argument supports that the interference is caused by the effect of movement of RNA Polymerase from the strong promoter toward the weak promoter. In this process the RNA Polymerases can collide head to head causing the weakly bound one to fall off from DNA. Another argument states that antisense transcribed DNA can be important. In the case of SSX4, we know that if an antisense transcript is functional in the regulation of the promoter, that this is independent of the sequence content of the transcript (**Figure 3.5**). However,

since dsRNA that forms as a result of convergent bidirectional promoters may tether one elongating RNA Polymerase to the other and lead to topological stress that result in transcriptional interference (Callen BP, 2004), this also can be a possible explanation of what we observe for SSX4. Another effect of the movement of RNA Polymerase II can be the elongation of RNA Polymerase II from the strong promoter over the weak promoter disrupting the initiation intermediates residing on the weak promoter. This theory presumes that the weak promoter is slower in both promoter escape and elongation. This would then cause open complexes of RNA Polymerase II to wait in the weak promoter for the escape while the RNA Polymerase II from the strong promoter has already begun elongation. In an experimental system with pR and pL these possibilities were tested and it was shown that the major cause of the interference in the experimental system is a sitting duck mechanism, in which pR convergent transcription over pL negatively affects promoter initiation intermediates that are present at pL (Callen BP,2004). This type of transcriptional interference is also found to contribute significantly to the HIV latency as viral integration into the actively transcribed host genes led to transcriptional interference caused by the elongating RNA polymerase II transcribing through the viral promoter (Lenasi T, 2008). Another example to transcriptional interference is demonstrated by the tissue specific epigenetic repression of one of two promoters of the enzyme folylpoly-γ-glutamate synthetase that then results in the derepression of a weaker downstream promoter which avoids methylation (Racanelli AC, 2008).

However, mechanisms other than transcriptional interference may lead to the inhibition of antisense SSX4 promoter since transcript identified by the 5'RACE experiments previously conducted do not overlap (**Figure 3.6**). In this case there can be a mechanism that forces the promoter to choose between the sense and antisense activities. If this is the case, the sense promoter being the stronger promoter would probably be chosen.

Another possibility is that the sense transcription may inhibit the antisense transcription by a feedback mechanism. Although this hypothesis is not supported by literature, the possibility could also be assessed by interfering with the sense transcript (by using a very early stop signal) and observing the effect on the antisense transcription. The opposite experiment, early termination of antisense transcript, would also be expected to have a result, given that both promoters are active for SSX4.

The exact mechanism of the transcriptional interference in our system where the sense promoter activity inhibits the activity of antisense promoter, which is normally active, could be any of these theories. Further experiments with different constructs to test the exact mechanism can be

constructed like the ones with the viral promoters but certain points need to be kept in mind. One is the validity of these *in vitro* experiments in *in vivo* conditions. Although this transcriptional interference seems to reside in the SSX4 promoter, we cannot know the actual situation in the cells when many other different parameters are added like higher order chromatin conformation or histone code around the promoter. Another important point to consider is that these promoters are rather weak promoters and the minor changes in the promoter activities are really hard to observe. In fact, this reporter system may not be appropriate for detecting the changes in the weak promoters like SSX4 antisense promoter. It may be useful to try other reporter systems to understand the principles of this promoter. Since luciferase reporter systems seem to be the most effective reporter systems up to date, new double reporter systems can be constructed using the luciferase system. In the future, I believe a double reporter construct with *Renilla* and firefly luciferase could be useful since the commercial luciferase assay systems are designed to allow high-throughput analysis of mammalian cells containing genes for firefly and luciferases. Understanding the working principle of this bidirectional promoter can give us further explanations about the regulation of gene expression of CTs.

Another important point about this promoter based regulation is whether other CT antigens also have bidirectional promoters or not. From bioinformatics analysis, NY-ESO-1 seems to have a bidirectional promoter like SSX4 but preliminary data about MAGEA1 indicates no antisense promoter in MAGEA1 promoter. It would be interesting to construct reporter vectors for promoters derived from other CT genes to test for the presence of bidirectional promoters in those genes.

5.2 RNA Based Silencing Mechanisms and the Role of DICER in the Regulation of CT Gene Expression

The RNA world is expanding day by day by emergence of new type of functional RNAs. There are also many more so called nonfunctional RNAs that we do not know the function for now. In this huge world, different RNAs are responsible for regulating the expression of different genes or gene groups. We are just beginning to understand the transcriptional nature of the genome or the so called transcriptome, described recently. The traditional view of many transcriptional processes is changing rapidly as genome-wide studies increase. We know much about the RNA and its role in the transcriptional regulation but I believe that this is just the tip of an iceberg.

The possible role of noncoding RNAs in the regulation of CT antigens emerged from the analysis of bidirectional promoter of SSX4 gene during this study. The fact that sense and antisense promoter regions are convergent raised the idea that transcription from this promoter may lead to dsRNA biogenesis.

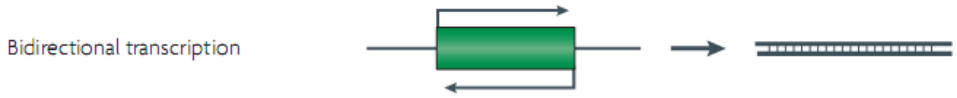


Figure 5.1 Endogenous small interfering RNAs can be produced by bidirectional transcription. From: Okamura K, Nat Rev Mol Cell Biol. 2008

These dsRNAs may have a regulatory role on the SSX4 gene expression. These dsRNAs are supposedly belong to endosiRNA small RNAs and DICER1 has been thought to have a role in the biogenesis of endosiRNAs.

Another clue that CT genes can be regulated by small RNAs come from similar expression and methylation patterns of LINE1 repeats and CT genes. The expression of active repeat elements is normally restricted to a specific period during germ cell development, when active retrotransposition can take place. In adult tissues, expression of repeats is inhibited by DNA methylation but they are reactivated in a wide variety of cancers, like CTs, especially in metastasis (Boo YJ, 2007). The reason for this reactivation is thought to be global hypomethylation of the genomic DNA, observed during the course of cancer progression. Reexpression of repeats is thought to contribute to cancer progression by causing genomic instability (Hoffmann MJ, 2005). Indeed, CT genes might be regarded as repeats since they are in gene families with 90% or more homology.

Data supporting the fact that repeat sequences and CT genes were similarly regulated came from the methylation analysis of L1 repeats and SSX4.

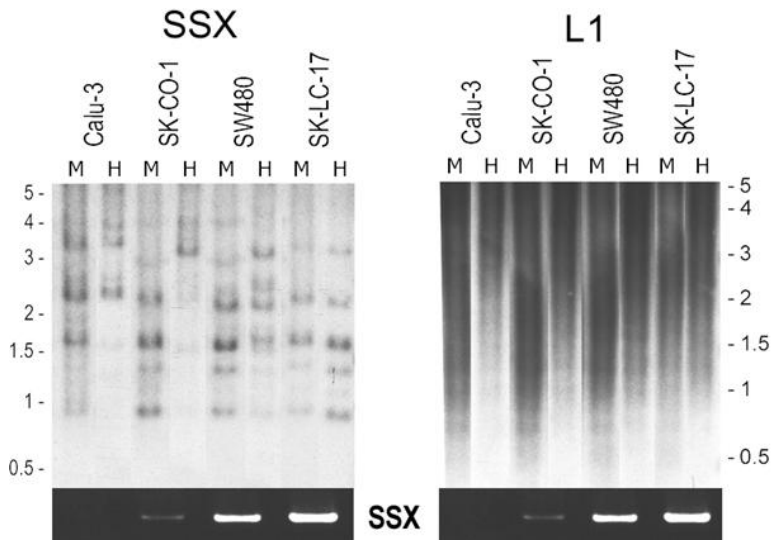


Figure 5.2. Methylation analysis of genomic DNA in cancer cell lines. The top part of the figure shows Southern analyses of genomic DNA digested by MspI(M) and HpaII(H), probed by a universal SSX probe (left) or a labeled L1 repeat (right). The SSX gene expression rates as determined by RT-PCR for cell lines are shown on the bottom of the figure (Gure AO, unpublished data)

The probes used in **Figure 5.2** are for the SSX genes that can detect SSX 1,2,3,4,5 and L1 probe that can possibly hybridize to all L1 repeat sequences in the cell. Genomic DNA from different cancer cell lines was cleaved by MspI and HpaII. MspI and HpaII recognize the same sites (CCGG) but HpaII digests the sequence only when it is unmethylated whereas MspI is insensitive to methylation. As a result, these data suggest that when the SSX4 gene is hypomethylated (the gene expression is higher), L1 repeats are also hypomethylated coordinately. This may suggest that L1 repeats and CT genes can be regulated by the same mechanism. Of course another explanation may be that they both are affected from the global hypomethylation but a promoter specific hypomethylation is seen in the MAGEA1 promoter (De Smet C, 2004) and possibly in all other CTs. It is difficult to precisely study only one LINE-1 sequence in the human genome since the sequence similarity among these sequences makes the design of specific primers difficult to test for methylation. Yet they are not identical and therefore, it is surprising that they should behave in an identical manner with respect to DNA methylation. Another similarity between L1 repeats and CT antigens is that the LINE-1 promoter of intact elements is equally active in different cell types as well as in tumor and normal cells from the same tissue (Steinhoff C, 2003). This results show similarity with the activity of the MAGEA1 and SSX4 promoter in tumor cells where the MAGE A1 or SSX4 gene is not expressed. (De Smet C, 1995 and **Figure 3.3**) Furthermore the promoters of active L1 repeats are also bidirectional.

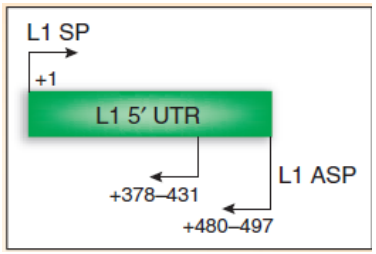


Figure 5.3 Illustration of sense and antisense promoters of an active L1 repeat. Taken from *Harris S Soifer & John J Rossi Small interfering RNAs to the rescue: blocking L1 retrotransposition*

Endogenous L1-derived siRNAs, are triggers for RNAi, and are naturally present in human cells (Yang N, 2006). It has been suggested that simultaneous bidirectional transcription from the L1 antisense promoter (ASP) and L1 sense promoter (SP) can give rise to L1 doublestranded RNAs (dsRNAs) which then could serve as a substrate for processing into siRNAs by DICER (**Figure 5.3**). To understand the picture completely, we also have to be aware of the other ways for cell to produce dsRNA and their association with CT genes.

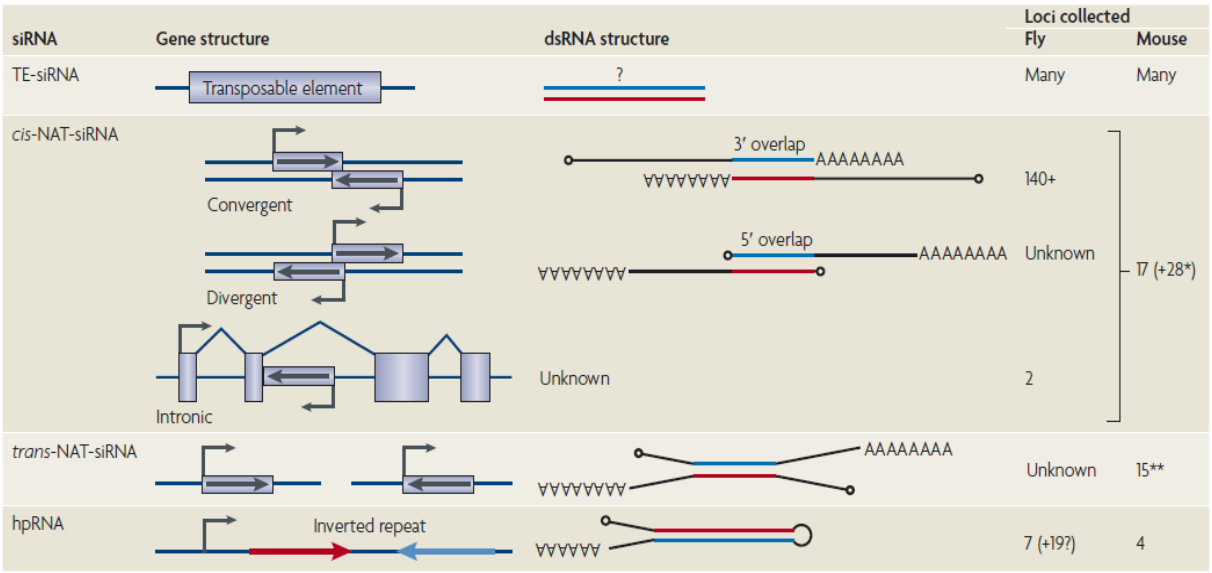


Figure 5.4 Endogenous small interfering RNAs. Taken from *Okamura K, Lai EC. Nat Rev Mol Cell Biol. 2008*

Figure 5.4 summarizes the known mechanisms by which the cells can generate dsRNA. The number of cases identified to date in flies and mice is indicated on the most right side of the figure. Transposable element siRNAs (TE-siRNAs) have been found extensively in both fly and mice. Considering the similarities in the expression patterns of LINE-1 repeats and CT antigens, CT-endosiRNAs might also exist in this context. Another event that generates dsRNA does so by *Trans*-NAT dsRNAs formed between transcripts that are produced from distinct genomic locations that are

generally comprised of an mRNA and an antisense-transcribed pseudogene. *Trans*-NAT dsRNAs can also be possible for CT antigens given the high homology between different members of CT antigen gene families. For example any SSX gene family member transcript can base pair with the transcript of another SSX gene family member if one of them is an antisense transcript generated by a cryptic ptomoter. Furthermore, another dsRNA producing mechanism based on the generation of hairpin RNA (hpRNA) could be the case for CT antigen genes. Most of the CT-X antigen genes are found to be organized into large inverted repeats and transcription through these genes can produce large stem-loop structures which then could serve as a substrate for endosiRNA production. As can be seen, regulation of CT antigen gene expression by endosiRNAs is very likely to occur in the cellular context.

EndosiRNA pathway is known to depend on Dicer activity; thus, we decided to knockdown human DICER1 and see its effect on the expression patterns of CT genes.

However, when we tried to investigate the effect of DICER1 on the regulation of expression of CT antigen genes using previously generated microarray data of HCT116 DICER Exon5 knockout cells (Ting AH, 2008), we did not observe any differences in CT gene expression among parental and knockout cells (**Figure4.28** and **Figure 4.29**). This suggests that CT gene expression may be DICER-independent. To eliminate the possible role of DICER in the regulation of CT gene expression, this result has to be confirmed. Furthermore, there can be a DICER dependent mechanism working on the regulation of CT antigens in a different cell line and in different conditions.

We were unable to knockdown DICER1 in SK-LC-17 cells (section 4.4). Although we had a couple of clones, we were unable to detect any decrease in DICER1 expression in Quantitative Real Time PCR. Different interpretations may explain this phenomenon. First of all, we have compared the DICER levels of clones with the parental SK-LC-17 but we needed to compare it with scrambled shRNA transfected SK-LC-17 cells as a control. The DICER1 shRNA constructs have been purchased but shLuc (control shRNA that codes for shRNA against firefly luciferase) construct could not be prepared because of the time constraints. Another reason could be the methodology that we have used. Although we have purchased the shDICER1 constructs and they have been reported to silence DICER1 successfully in mammalian cell lines previously (Kumar MS, 2007), the used method for introducing these plasmids into cells was lentiviral transduction which is a more effective way than transfection. Although shRNA constructs can be transfected, the obligation of a high efficiency transduction can be a DICER1 specific requirement since in this case DICER1 is also necessary for its own knockdown. Interestingly, knockdown of DICER1 by siRNA did not result in a biologically important increase in expression of endogenous active L1s. However when an active L1 is introduced

HeLa cells where DICER1 expression was reduced, a cell-based assay showed a strong increase in retrotransposition activity (Yang N, 2006) which suggests that cells rely on the RNAi machinery to take action against only active L1s, rather than using RNAi to maintain L1s in a transcriptionally silent state. This scene could also be true for the CT genes and RNAi machinery can become active only after CT genes are activated by hypomethylation. Another consideration about the DICER1 knockdown is that the cell may be sensitive to DICER knockdown. We have to take into account that loss of DICER1 or miRNA-associated Argonaute proteins is nearly always lethal in animals, and such mutants show severe developmental defects in both plants and animals. We do not know the effect of DICER1 knockdown on the cells that we are working on and it is possible that some cells can be more sensitive to DICER1 levels than others. There is also an interesting finding about the HCT116 DICER1 Exon5 knockout cells (Ting AH, 2008): reexpression of DICER seemed to be lethal. This may be an indication that DICER function in these cells is rendered redundant by the over activity of another mechanism and reexpression of DICER may lead to an overdose which leads to death of these cells. The compensation of DICER loss by other mechanisms may also explain the lack of change in the gene expression profiles for CT antigens (**Figure4.28** and **Figure 4.29**).

I believe that small RNA pathways have roles in regulation of CT gene expression and knocking down DICER1 is a wise way to control the effect of DICER1 in CT gene expression. For this purpose, DICER1 knockdown experiments should be performed in different cell lines (known expressCT genes at different levels [strong, weak, none]) by using pSicoRshLuc construct as a control and by using lentiviral transduction system instead of regular transfection. This experiment would show us if the regulation of CT gene expression is DICER1 dependent or not and even a negative result would be very informative since only then we can eliminate the possible role of DICER1 dependent small RNA pathways on the regulation of CT gene expression.

Also I believe more experiments should be performed to investigate the similarities of LINE-1 repeats and CT antigen gene expression patterns. If they are found to be coordinately expressed, known data about the LINE-1 expression and regulation of expression could be used to uncover the unknowns of CT antigens and vice versa.

5.3 Analysis of OATL transcript

The proof that SSX4 promoter actually can work bidirectionally *in vivo* is the OATL transcript. This transcript has been found once by 5'RACE but despite several attempts with different primers and conditions, we were unable to detect this transcript.

There can be several different reasons for these results. First, if there is an OATL transcript, it has to be a very scarce RNA considering that the SSX4 RNA is also not very abundant even in SK-LC-17 and that the antisense promoter activity is several fold weaker than the sense promoter activity. The detection of this transcript may not be possible with RT-PCR since most of the RNAs can be lost during the harsh RNA isolation procedures and also some RNA may be lost during the cDNA synthesis. For these reasons, new primers can be designed specifically for the transcript found by 5'RACE to extend sequence information in both orientations, since that portion has been detected before. Secondly, the cDNA synthesis can be performed by using primers specific to the OATL sequence to increase sensitivity. PCR with nested primers can be done to detect even very low amounts of transcript.

However, noncoding RNAs are very hard to detect by regular PCR techniques since these RNAs are not as stable as mRNAs, since they might not have polyA tails, they might not be transcibed fully and they are not transported to the cytoplasm. These RNAs are mostly degraded by cell's exonucleolytic RNA exosome. Recently, by silencing a core component of human 3′ to 5′ exoribonucleolytic exosome new noncoding transcipts called promoter upstream transcripts (PROMPTs) have been identified. These transcripts are produced ~0.5 to 2.5 kilobases upstream of active transcription start sites (Preker P, 2008). Inhibiting the RNA exosome can reveal low-abundant transcripts by eliminating RNA degradation which may help us to detect our OATL transcript.

There are specific techniques for detecting noncoding RNAs. One of them is of course northern blot analysis. For the detection of the OATL transcript, northern blot analysis can be a very informative technique to use. A probe corresponding to the RACE found OATL transcript can be designed and used in the analysis. By this analysis the size of the OATL transcript can also be inferred. Northern blot analysis is still the most common technique used to investigate small RNAs and different probes corresponding to different sequences of promoter may provide us a map of transcripts around the SSX4 promoter. Another recent technique that is used for the detection of ncRNAs is cap analysis gene expression (CAGE), which is based on preparation and sequencing of concatamers of DNA tags derived from the initial 20 nucleotides of the 5′ ends of mRNAs. This technique allows us to detect all RNA Pol II transcripts including ncRNAs (they are also 5′ capped) and can be used for high-throughput analysis. This technique is quite sensitive and can be used to detect the OATL transcripts. By using CAGE analysis overlapping bidirectional promoters have been identified where there are at least two TSSs on opposite strands within 40bps and bidirectional promoters have been proven to function as an origin of natural dsRNAs (Kawaji H, 2008).

All in all, the possible role of SSX4 promoter in the regulation of expression of SSX4 gene was investigated in this study. Bidirectional and convergent nature of SSX4 promoter has been described in detail by addressing the individual sequences responsible for sense and antisense transcriptional activity. Preliminary data about the bidirectionality of other CT genes have been collected. A model system for testing both the sense and antisense promoter activities has been constructed and a possible transcriptional interference between sense and antisense promoters has been identified which might have a role in the regulation of SSX4 gene expression. Further experiments, detailed above, should be performed to fully understand the nature of the promoter and its relevance to CT gene expression.

6. REFERENCES

Almeida LG, Sakabe NJ, deOliveira AR, Silva MC, Mundstein AS, Cohen T, Chen YT, Chua R, Gurung S, Gnjatic S, Jungbluth AA, Caballero OL, Bairoch A, Kiesler E, White SL, Simpson AJ, Old LJ, Camargo AA, Vasconcelos AT. *CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens* **Nucleic Acids Res.** 2009 Jan;37(Database issue):D816-9

Amaral PP, Mattick JS. *Noncoding RNA in development*. **Mamm Genome**. 2008 Aug;19(7-8):454-92

Aravin AA, Naumova NM, Tulin AV, Vagin VV, Rozovsky YM, Gvozdev VA. *Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline***. Curr. Biol.** 2001 11, 1017–1027 137**.**

Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, Snyder B, Gaasterland T, Meyer J, Tuschl T. *The small RNA profile during Drosophila melanogaster development*. **Dev. Cell** 2003 5, 337–350

Aravin AA, Hannon GJ, Brennecke J. *The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race***. Science.** 2007 Nov 2;318(5851)

Baba T, Niida Y, Michikawa Y, Kashiwabara S, Kodaira K, Takenaka M, Kohno N, Gerton GL, Arai Y. *An acrosomal protein, sp32, in mammalian sperm is a binding protein specific for two proacrosins and an acrosin intermediate*. **J Biol Chem**. 1994 Apr 1;269(13):10133-40.

Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R. *Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs*. **Genes Dev.** 2008 22, 2773–2785

Bao N, Lye KW, Barton MK. *MicroRNA binding sites in Arabidopsis class III HD-ZIP mRNAs are required for methylation of the template chromosome*. **Dev Cell**. 2004 Nov;7(5):653-62.

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. *High-resolution profiling of histone methylations in the human genome*. **Cell.** 2007 May 18;129(4):823-37

Bestor T, Laudano A, Mattaliano R, Ingram V. *Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases*. **J Mol Biol.** 1988 Oct 20;203(4):971-83

Boël P, Wildmann C, Sensi ML, Brasseur R, Renauld JC, Coulie P, Boon T, van der Bruggen P. *BAGE: a new gene encoding an antigen recognized on human melanomas by cytolytic T lymphocytes* **Immunity.** 1995 Feb;2(2):167-75

Boo YJ, Park JM, Kim J, Chae YS, Min BW, Um JW, Moon HY. *L1 expression as a marker for poor prognosis, tumor progression, and short survival in patients with colorectal cancer*. **Ann Surg Oncol.** 2007 May;14(5):1703-11

Brower-Toland B, Findley SD, Jiang L, Liu L, Yin H, Dus M, Zhou P, Elgin SC, Lin H. *Drosophila PIWI associates with chromatin and interacts directly with HP1a.* **Genes Dev**. 2007 Sep 15;21(18):2300-11.

Callen BP, Shearwin KE, Egan JB. *Transcriptional interference between convergent promoters caused by elongation over the promoter***. Mol Cell**. 2004 Jun 4;14(5):647-56.

Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engström PG, Frith MC, Forrest AR, Alkema WB, Tan SL, Plessy C, Kodzius R, Ravasi T, Kasukawa T, Fukuda S, Kanamori-Katayama M, Kitazume Y, Kawaji H, Kai C, Nakamura M, Konno H, Nakano K, Mottagui-Tabar S, Arner P, Chesi A, Gustincich S, Persichetti F, Suzuki H, Grimmond SM, Wells CA, Orlando V, Wahlestedt C, Liu ET, Harbers M, Kawai J, Bajic VB, Hume DA, Hayashizaki Y. *Genome-wide analysis of mammalian promoter architecture and evolution.* **Nat Genet.** 2006 Jun;38(6):626-35

Chen YT, Boyer AD, Viars CS, Tsang S, Old LJ, Arden KC. *Genomic cloning and localization of CTAG, a gene encoding an autoimmunogenic cancer-testis antigen NY-ESO-1, to human chromosome Xq28*. **Cytogenet Cell Genet**. 1997;79(3-4):237-40

Chu CY, RanaTM *Small RNAs: regulators and guardians of the genome*. **J. Cell Physiol**. 2007 213,412–419

Chung WJ, Okamura K, Martin R, Lai EC. *Endogenous RNA interference provides a somatic defense against Drosophila transposons*. **Curr. Biol.** 2008 18, 795–802.

Clark J, Rocques PJ, Crew AJ, Gill S, Shipley J, Chan AM, Gusterson BA, Cooper CS. *Identification of novel genes, SYT and SSX, involved in the t(X;18)(p11.2;q11.2) translocation found in human synovial sarcoma.* **Nat Genet.** 1994 Aug;7(4):502-8.

Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, Hannon GJ, Brennecke J. *An endogenous small interfering RNA pathway in Drosophila***. Nature** 2008 453, 798–802

de Bruijn DR, van Dijk AH, Willemse MP, van Kessel AG. *The C terminus of the synovial sarcoma-associated SSX proteins interacts with the LIM homeobox protein LHX4*. **Oncogene**. 2008 Jan 24;27(5):653-62

De Smet C, Courtois SJ, Faraoni I, Lurquin C, Szikora JP, De Backer O, Boon T. *Involvement of two Ets binding sites in the transcriptional activation of the MAGE1 gene*. **Immunogenetics**. 1995;42(4):282-90

De Smet C, De Backer O, Faraoni I, Lurquin C, Brasseur F, Boon T. *The activation of human gene MAGE-1 in tumor cells is correlated with genome-wide demethylation*. **Proc Natl Acad Sci U S A**. 1996 Jul 9;93(14):7149-53

De Smet C, Lurquin C, Lethe B, Martelange V, Boon T. *DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter*. **Mol Cell Biol** 1999; 19: 7327–35.

De Smet C, Loriot A, Boon T. *Promoter-dependent mechanism leading to selective hypomethylation within the 5' region of gene MAGE-A1 in tumor cells*. **Mol Cell Biol.** 2004 Jun;24(11):4781-90.

Ehrlich M. *DNA methylation in cancer: too much, but also too little*. **Oncogene**. 2002 Aug 12;21(35):5400-13

Esteller M. *Cancer epigenomics: DNA methylomes and histone-modification maps*. **Nat Rev Genet**. 2007 Apr;8(4):286-98

Filipowicz W., Bhattacharyya S. N., Sonenberg N. *Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?* **Nature Rev. Genet**. 2008 9, 102–114

Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, Bonaldi T, Haydon C, Ropero S, Petrie K. *Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer*. **Nat Genet**. 2005 Apr;37(4):391-400.

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL. *A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands.* **Proc Natl Acad Sci U S A**. 1992 Mar 1;89(5):1827-31.

Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler EL, Zapp ML, Weng Z, Zamore PD.. *Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells*. **Science** 2008 320, 1077–1081

Gonzalez S, Pisano DG, Serrano M. *Mechanistic principles of chromatin remodeling guided by siRNAs and miRNAs.* **Cell Cycle.** 2008 Aug 15;7(16):2601-8

Gonzalez-Zulueta M, Bender CM, Yang AS, Nguyen T, Beart RW, Van Tornout JM, Jones PA. *Methylation of the 5' CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing*. **Cancer Res**. 1995 Oct 15;55(20):4531-5.

Greger V, Passarge E, Höpping W, Messmer E, Horsthemke B. *Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma*. **Hum Genet.** 1989 Sep;83(2):155-8.

Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC. *A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila.* **Science** 2007 315, 1587–1590

Gure AO, Türeci O, Sahin U, Tsang S, Scanlan MJ, Jäger E, Knuth A, Pfreundschuh M, Old LJ, Chen YT. *SSX: a multigene family with several members transcribed in normal testis and human cancer*. **Int J Cancer.** 1997 Sep 17;72(6):965-71.

Gure AO, Chua R, Williamson B, Gonen M, Ferrera CA, Gnjatic S, Ritter G, Simpson AJ, Chen YT, Old LJ, Altorki NK. *Cancer-testis genes are coordinately expressed and are markers of poor outcome in non-small cell lung cancer*. **Clin Cancer Res.** 2005 Nov 15;11(22):8055-62

Herman JG, Merlo A, Mao L, Lapidus RG, Issa JP, Davidson NE, Sidransky D, Baylin SB. *Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers*. **Cancer Res.** 1995 Oct 15;55(20):4525-30.

Hoffmann MJ, Schulz WA. *Causes and consequences of DNA hypomethylation in human cancer*. **Biochem Cell Biol**. 2005 Jun;83(3):296-321

Hofmann O, Caballero OL, Stevenson BJ, Chen YT, Cohen T, Chua R, Maher CA, Panji S, Schaefer U, Kruger A, Lehvaslaiho M, Carninci P, Hayashizaki Y, Jongeneel CV, Simpson AJ, Old LJ, Hide W. *Genome-wide analysis of cancer/testis gene expression*. **Proc Natl Acad Sci U S A**. 2008

Hong JA, Kang Y, Abdullaev Z, Flanagan PT, Pack SD, Fischette MR, Adnani MT, Loukinov DI, Vatolin S, Risinger JI, Custer M, Chen GA, Zhao M, Nguyen DM, Barrett JC, Lobanenkov VV, Schrump DS., *Reciprocal binding of CTCF and BORIS to the NY-ESO-1 promoter coincides with derepression of this cancer-testis gene in lung cancer cells.* **Cancer Res,** 2005. 65(17): p. 7763-74

Imamura T, Yamamoto S, Ohgane J, Hattori N, Tanaka S, Shiota K. *Non-coding RNA directed DNA demethylation of Sphk1 CpG island.* **Biochem Biophys Res Commun**. 2004 Sep 17;322(2):593-600

James SR, Link PA, Karpf AR. Epigenetic regulation of X-linked cancer/germline antigen genes by DNMT1 and DNMT3b. **Oncogene**. 2006 Nov 2;25(52):6975-85

Janowski BA, Huffman KE, Schwartz JC, Ram R, Hardy D, Shames DS, Minna JD, Corey DR. *Inhibiting gene expression at transcription start sites in chromosomal DNA with antigene RNAs.* **Nat Chem Biol**. 2005 Sep;1(4):216-22

Janowski BA, Younger ST, Hardy DB, Ram R, Huffman KE, Corey DR. *Activating gene expression in mammalian cells with promoter targeted duplex RNAs*. **Nat. Chem. Biol**. 2007 3, 166–173

Jenuwein T, Allis CD. *Translating the histone code*. **Science.** 2001 Aug 10;293(5532):1074-80.

Jones AL, Thomas CL, Maule AJ. *De novo methylation and co-suppression induced by a cytoplasmically replicating plant RNA virus*. **EMBO J.** 1998 Nov 2;17(21):6385-93

Jones PA, Baylin SB. *The epigenomics of cancer*. **Cell.** 2007 Feb 23;128(4):683-92.

Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT. *The RNA polymerase II core promoter* - the gateway to transcription. **Curr Opin Cell Biol.** 2008 Jun;20(3):253-9

Kaneda A, Tsukamoto T, Takamura-Enya T, Watanabe N, Kaminishi M, Sugimura T, Tatematsu M, Ushijima T. *Frequent hypomethylation in multiple promoter CpG islands is associated with global hypomethylation, but not with frequent promoter hypermethylation*. **Cancer Sci**. 2004 Jan;95(1):58-64.

Kang Y, Hong JA, Chen GA, Nguyen DM, Schrump DS., *Dynamic transcriptional regulatory complexes including BORIS, CTCF and Sp1 modulate NY-ESO-1 expression in lung cancer cells.* **Oncogene**, 2007. 26(30): p. 4394-403.

Kawaji H, Nakamura M, Takahashi Y, Sandelin A, Katayama S, Fukuda S, Daub CO, Kai C, Kawai J, Yasuda J, Carninci P, Hayashizaki Y. *Hidden layers of human small RNAs*. **BMC Genomics**. 2008 Apr 10;9:157.

Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H. *Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells*. **Nature**. 2008 Jun 5;453(7196):793-7

Kim DH, Saetrom P, Snøve O Jr, Rossi JJ. *MicroRNA-directed transcriptional gene silencing in mammalian cells*. **Proc Natl Acad Sci U S A.** 2008 Oct 21;105(42)

Kim VN. *MicroRNA biogenesis: coordinated cropping and dicing*. **Nature Rev. Mol. Cell Biol**. 2005 6, 376–385

Kim VN, Han J, Siomi MC. *Biogenesis of small RNAs in animals* **Nat Rev Mol Cell Biol.** 2009 Feb;10(2):126-39.

Koch F, Jourquin F, Ferrier P, Andrau JC. *Genome-wide RNA polymerase II: not genes only!* **Trends Biochem Sci.** 2008 Jun;33(6):265-73

Kondo T, Zhu X, Asa SL, Ezzat S., *The cancer/testis antigen melanoma-associated antigen-A3/A6 is a novel target of fibroblast growth factor receptor 2-IIIb through histone H3 modifications in thyroid cancer.* **Clin Cancer Res**, 2007. 13(16): p. 4713-20.

Kouzarides T. *Chromatin modifications and their function*. **Cell**. 2007 Feb 23;128(4):693-705

Kumar MS, Lu J, Mercer KL, Golub TR, Jacks T. *Impaired microRNA processing enhances cellular transformation and tumorigenesis*. **Nat Genet**. 2007 May;39(5):673-7

Ladanyi M. *Fusions of the SYT and SSX genes in synovial sarcoma*. **Oncogene.** 2001 Sep 10;20(40):5755-62.

Lee RC, Feinbaum RL, Ambros V. *The C. Elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14***. Cell** 1993 75, 843–854

Lenasi T, Contreras X, Peterlin BM. *Transcriptional interference antagonizes proviral gene expression to promote HIV latency.* **Cell Host Microbe**. 2008 Aug 14;4(2):123-33

Levine SS, King IF, Kingston RE. *Division of labor in polycomb group repression*. **Trends Biochem Sci.** 2004 Sep;29(9):478-85.

Li E, Bestor TH, Jaenisch R. Targeted mutation of the *DNA methyltransferase gene results in embryonic lethality***. Cell.** 1992 Jun 12;69(6):915-26.

Li LC, Okino ST, Zhao H, Pookot D, Place RF, Urakami S, Enokida H, Dahiya R. *Small dsRNAs induce transcriptional activation in human cells.* **Proc. Natl. Acad. Sci. USA** 2006 103, 17337–17342

Lim JH, Kim SP, Gabrielson E, Park YB, Park JW, Kwon TK. *Activation of human cancer/testis antigen gene, XAGE-1, in tumor cells is correlated with CpG island hypomethylation*. **Int J Cancer** 2005; 116:200-6.

Loukinov DI, Pugacheva E, Vatolin S. *BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma*. **Proc Natl Acad Sci U S A** 2002;99:6806–11

Mercer TR, Dinger ME, Mattick JS. *Long non-coding RNAs: insights into functions*. **Nat Rev Genet.** 2009 Mar;10(3):155-9NA

Mette MF, Aufsatz W, van der Winden J, Matzke MA, Matzke AJ. *Transcriptional silencing and promoter methylation triggered by double-stranded RNA.* **EMBO J.** 2000 Oct 2;19(19):5194-20

Meuwissen RL, Offenberg HH, Dietrich AJ, Riesewijk A, van Iersel M, Heyting C. *A coiled-coil related protein specific for synapsed regions of meiotic prophase chromosomes.* **EMBO J.** 1992 Dec;11(13):5091-100

Moazed D. *Small RNAs in transcriptional gene silencing and genome defence*. **Nature.** 2009 Jan 22;457(7228):413-20

Morris KV, Chan SW, Jacobsen SE, Looney DJ. *Small interfering RNA-induced transcriptional gene silencing in human cells.* **Science** 2004 305, 1289–1292

Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, and Bird A. *Transcriptional repression by themethyl-CpG-binding protein MeCP2 involves a histone deacetylase complex*. **Nature** 1998 393, 386–389

Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM, Jacquier A. *Widespread bidirectional promoters are the major source of cryptic transcripts in yeast.* **Nature.** 2009 Feb 19;457(7232):1038-42

Okamura K, Balla S, Martin R, Liu N, Lai EC. *Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in Drosophila melanogaster*. **Nature Struct. Mol. Biol**. 2008 15, 581–590

Okamura K, Chung WJ, Ruby JG, Guo H, Bartel DP, Lai EC. *The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs.* **Nature** 2008 453, 803–806

Okamura K, Lai EC. *Endogenous small interfering RNAs in animals.* **Nat Rev Mol Cell Biol.** 2008 Sep;9(9):673-8

Okano M, Bell DW, Haber DA, Li E. *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development***. Cell.** 1999 Oct 29;99(3):247-57.

Park CW, Chen Z, Kren BT, Steer CJ. *Double-stranded siRNA targeted to the huntingtin gene does not induce DNA methylation*. **Biochem. Biophys. Res. Commun**.2004 323, 275–280 (2004).

Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, Schierup MH, Jensen TH. *RNA exosome depletion reveals transcription upstream of active human promoters.* **Science**. 2008 Dec 19;322(5909):1851-4.

Racanelli AC, Turner FB, Xie LY, Taylor SM, Moran RG. *A mouse gene that coordinates epigenetic controls and transcriptional interference to achieve tissue-specific expression*. **Mol Cell Biol.** 2008 Jan;28(2):836-4

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY. *Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs***. Cell.** 2007 Jun 29;129(7):1311-23

Saha A, Wittmeyer J, Cairns BR. *Chromatin remodelling: the industrial revolution of DNA around histones.* **Nat Rev Mol Cell Biol**. 2006 Jun;7(6):437-47

Sahin U, Türeci O, Schmitt H, Cochlovius B, Johannes T, Schmits R, Stenner F, Luo G, Schobert I, Pfreundschuh M. *Human neoplasms elicit multiple specific immune responses in the autologous host*. **Proc Natl Acad Sci U S A.** 1995 Dec 5;92(25):11810-3

Sasidharan R, Gerstein M. *Genomics: protein fossils live on as RNA.* **Nature.** 2008 Jun 5;453(7196):729-31

Scanlan MJ, Gure AO, Jungbluth AA, Old LJ, Chen YT. *Cancer/Testis Antigens: an expanding family of targets for cancer immunotherapy.* **Immunological Reviews** 2002 Vol 188: 22–32

Scanlan MJ, AJ Simpson, and LJ Old, *The cancer/testis genes: review, standardization, and commentary.* **Cancer Immun**, 2004. 4: p. 1

Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. *Genome regulation by polycomb and trithorax proteins.* **Cell.** 2007 Feb 23;128(4):735-45.

Schwartz JC, Younger ST, Nguyen NB, Hardy DB, Monia BP, Corey DR, Janowski BA. *Antisense transcripts are targets for activating small RNAs.* **Nat Struct Mol Biol**. 2008 Aug;15(8):842-8

Sigalotti L, Coral S, Nardi G, Spessotto A, Cortini E, Cattarossi I, Colizzi F, Altomonte M, Maio M. *Promoter methylation controls the expression of MAGE2, 3 and 4 genes in human cutaneous melanoma*. **J Immunother** 2002; 25:16-26

Simpson AJ, Caballero OL, Jungbluth A, Chen YT, Old LJ. *Cancer/testis antigens, gametogenesis and cancer*. **Nat Rev Cancer**. 2005 Aug;5(8):615-25

Steinhoff C, Schulz WA. Transcriptional regulation of the human LINE-1 retrotransposon L1.2B. Mol Genet Genomics. 2003 Dec;270(5):394-402

Stevenson BJ, Iseli C, Panji S, Zahn-Zabal M, Hide W, Old LJ, Simpson AJ, Jongeneel CV. *Rapid evolution of cancer/testis genes on the X chromosome*. **BMC Genomics**. 2007 May 23;8:129.

Suzuki MM, Bird A. *DNA methylation landscapes: provocative insights from epigenomics*. **Nat Rev Genet**. 2008 Jun;9(6):465-76.

Takai D, Jones PA. *Comprehensive analysis of CpG islands in human chromosomes 21 and 22*. **Proc Natl Acad Sci U S A.** 2002 Mar 19;99(6):3740-5.

Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, Hannon GJ. *Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes.* **Nature.** 2008 May 22;453(7194):534-8

Ting AH, Schuebel KE, Herman JG, Baylin SB. *Short double-stranded RNA induces transcriptional gene silencing in human cancer cells in the absence of DNA methylation***. Nat Genet.** 2005 Aug;37(8):906-10

Ting AH, Suzuki H, Cope L, Schuebel KE, Lee BH, Toyota M, Imai K, Shinomura Y, Tokino T, Baylin SB. *A requirement for DICER to maintain full promoter CpG island hypermethylation in human cancer cells*. **Cancer Res.** 2008 Apr 15;68(8):2570-5

Traversari C, van der Bruggen P, Van den Eynde B, Hainaut P, Lemoine C, Ohta N, Old L, Boon T. *Transfection and expression of a gene coding for a human melanoma antigen recognized by autologous cytolytic T lymphocytes.* **Immunogenetics**. 1992;35(3):145-52

Trinklein ND, Aldred SF, Hartman SJ, Schroeder DI, Otillar RP, Myers RM. *An abundance of bidirectional promoters in the human genome*. **Genome Res.** 2004 Jan;14(1):62-6.

Türeci O, Sahin U, Schobert I, Koslowski M, Scmitt H, Schild HJ, Stenner F, Seitz G, Rammensee HG, Pfreundschuh M. *The SSX-2 gene, which is involved in the t(X;18) translocation of synovial sarcomas, codes for the human tumor antigen HOM-MEL-40***. Cancer Res**. 1996 Oct 15;56(20):4766-72

Türeci O, Chen YT, Sahin U, Güre AO, Zwick C, Villena C, Tsang S, Seitz G, Old LJ, Pfreundschuh M. *Expression of SSX genes in human tumors*. **Int J Cancer.** 1998 Jul 3;77(1):19-23

Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. *A distinct small RNA pathway silences selfish genetic elements in the germline.* **Science** 2006 313, 320–324

Van den Eynde B, Peeters O, De Backer O, Gaugler B, Lucas S, Boon T. *A new family of genes coding for an antigen recognized by autologous cytolytic T lymphocytes on a human melanoma*. **J Exp Med**. 1995 Sep 1;182(3):689-98

Van der Bruggen P, Traversari C, Chomez P, Lurquin C, De Plaen E, Van den Eynde B, Knuth A, Boon T *A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma* **Science.** 1991 Dec 13;254(5038):1643-7

Vatolin S, Abdullaev Z, Pack SD, Flanagan PT, Custer M, Loukinov DI, Pugacheva E, Hong JA, Morse H 3rd, Schrump DS, Risinger JI, Barrett JC, Lobanenkov VV., *Conditional expression of the CTCF-paralogous transcriptional factor BORIS in normal cells results in demethylation and derepression of MAGE-A1 and reactivation of other cancer-testis genes.* **Cancer Res**, 2005. 65(17): p. 7751-62.

Vidaeus CM, von Kapp-Herr C, Golden WL, Eddy RL, Shows TB, Herr JC. *Human fertilin beta: identification, characterization, and chromosomal mapping of an ADAM gene family member*. **Mol Reprod Dev.** 1997 Mar;46(3):363-9

Wang MB, Wesley SV, Finnegan EJ, Smith NA, Waterhouse PM. *Replicating satellite RNA induces sequence-specific DNA methylation and truncated transcripts in plants*. **RNA.** 2001 Jan;7(1):16-28.

Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, Tempst P, Rosenfeld MG, Glass CK, Kurokawa R. *Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription*. **Nature**. 2008 Jul 3;454(7200):126-30.

Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, Surani MA, Sakaki Y, Sasaki H. *Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes*. **Nature.** 2008 May 22;453(7194):539-43.

Wassenegger M, Heimes S, Riedel L , Sanger HL. *RNA-directed de novo methylation of genomic sequences in plants.* **Cell** 1994 76, 567–576

Weber J, Salgaller M, Samid D, Johnson B, Herlyn M, Lassam N, Treisman J, Rosenberg SA. *Expression of the MAGE-1 tumor antigen is up-regulated by the demethylating agent 5-aza-2'-deoxycytidine*. **Cancer Res**. 1994 Apr 1;54(7):1766-71

Wightman B, Ha I , Ruvkun G. *Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans*. **Cell** 1993 75, 855–862.

Wilson AS, Power BE, Molloy PL. *DNA hypomethylation and human diseases*. **Biochim Biophys Acta.** 2007 Jan;1775(1):138-62

Wischnewski F, Pantel K, Schwarzenbach H. *Promoter demethylation and histone acetylation mediate gene expression of MAGE-A1, -A2, -A3, and -A12 in human cancer cells*. **Mol Cancer Res**. 2006 May;4(5):339-49

Xiao, J. and H.S. Chen, *Biological functions of melanoma-associated antigens.* **World J Gastroenterol**, 2004. 10(13): p. 1849-53.

Yang N, Kazazian HH Jr. *L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells.* **Nat Struct Mol Biol.** 2006 Sep;13(9):763-71.

Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, Cui H. *Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA.* **Nature.** 2008 Jan 10;451(7175):202-6

Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A,Reinberg, D. *Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation*. **Genes Dev.** 1999 13, 1924–1935

7. APPENDICES

APPENDIX A. GENERATION OF SSX4 PROMOTER FINE TUNING CONSTRUCTS

APPENDIX B. DICER QUANTITATIVE REAL TIME PCR EXPERIMENTS

APPENDIX C. GENERATION OF DOUBLE REPORTER CONSTRUCTS

APPENDIX A. GENERATION OF SSX4 SENSE AND ANTISENSE PROMOTER MAPPING CONSTRUCTS

Agarose gel photos of PCR amplification using the primers indicated.

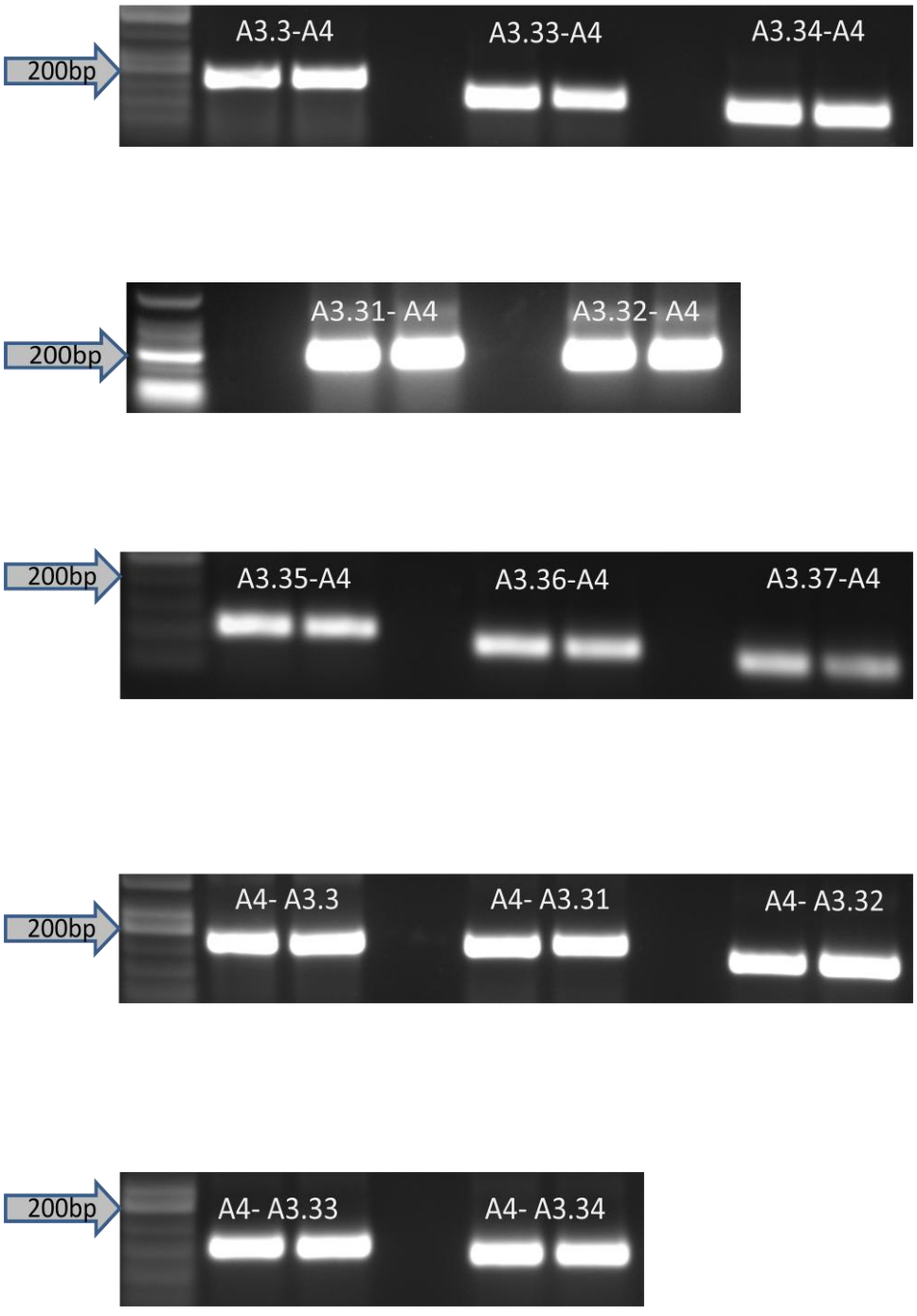


FIGURE S1. PCR products obtained by the denoted primer pairs

PCR products of the given primers were excised and eluted from the gel. All of the PCR primers are intact and working as can be seen in the figures. The PCR products were restricted with XhoI and HindIII and run on 1.5% agarose gel again.

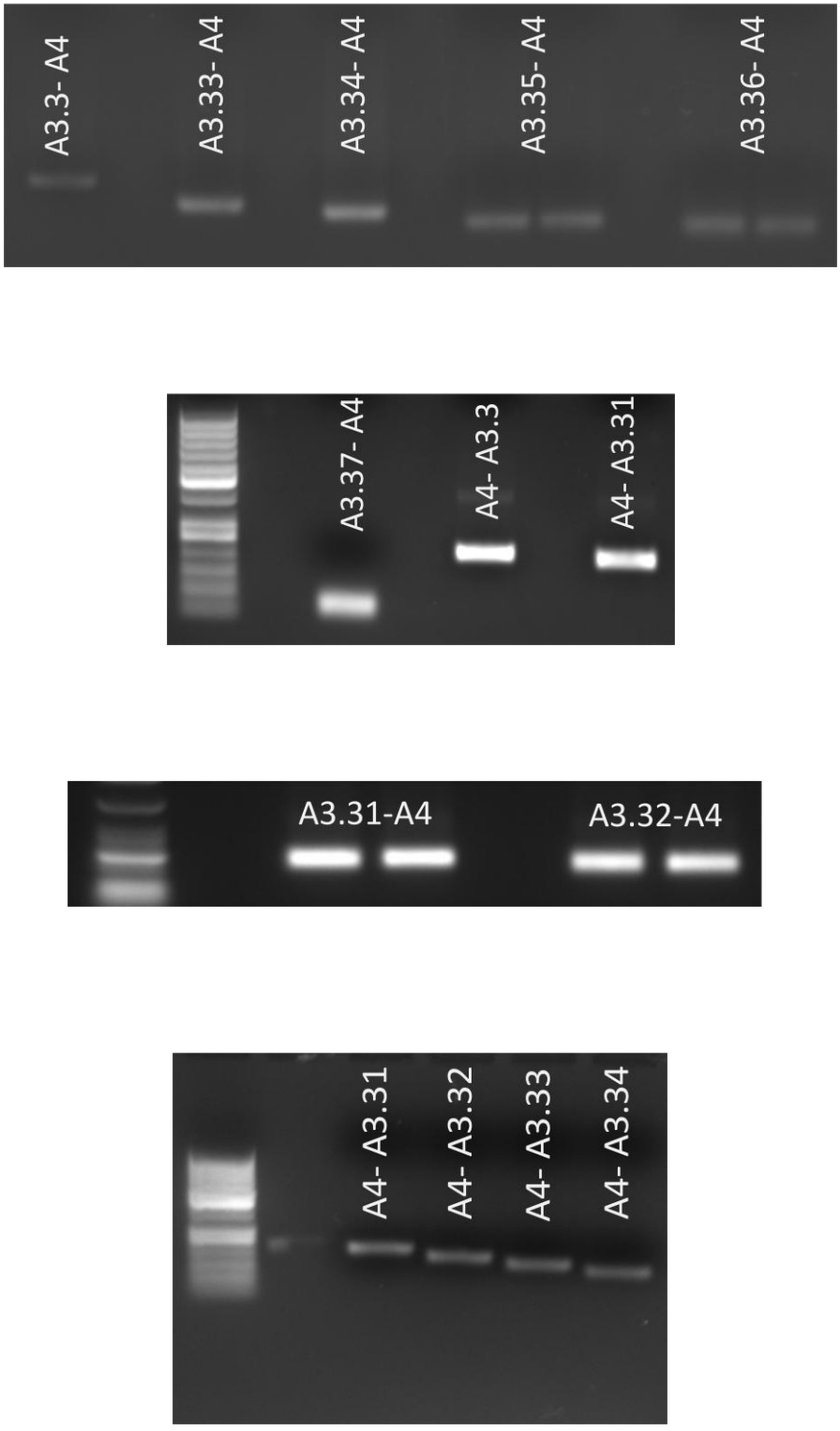


FIGURE S2. XhoI and HindIII restricted PCR products from FIGURE S1

These XhoI and HindIII restricted sequences have been ligated to the same site of pGL3 plasmid. After transformation the resulted constructs have been cut with XhoI and HindIII to excise the promoter fragment from the pGL3 construct for control.

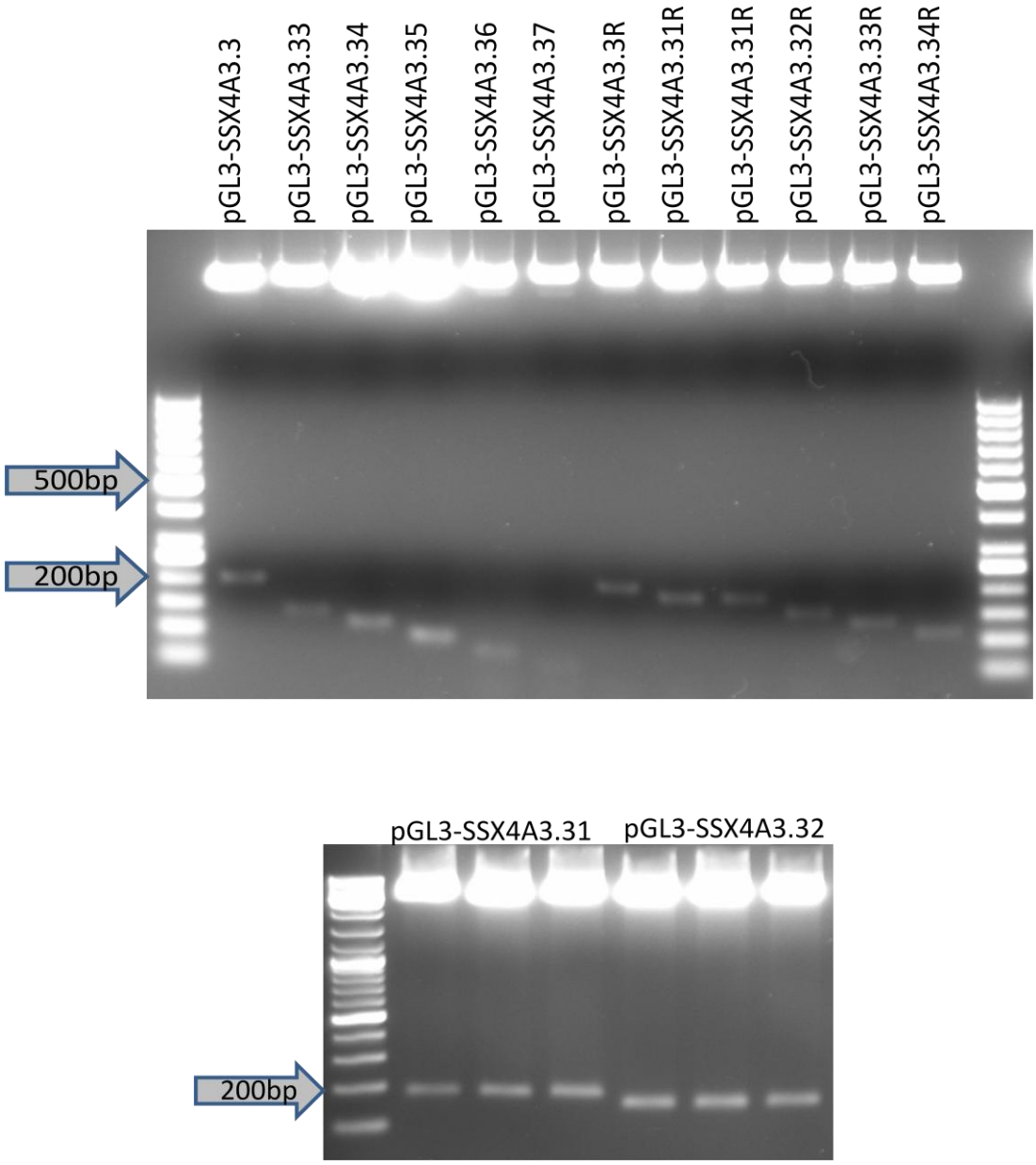


FIGURE S3. Control XhoI and HindIII restriction enzyme digestions for mapping of SSX4 promoter constructs

APPENDIX B. DICER1 QUANTITATIVE REAL TIME PCR EXPERIMENTS

B1. Efficiency of the Primers

To calculate the efficiency of primers 10-fold dilutions of SK-LC-17 cDNA has been used. Samples were analyzed using SyBr Green Dye, run on 1.5% agarose gel (**Figure S4**) and the efficiency graphs were plotted (**Figure S5**).

TABLE S1 Detailed Quantitative Real Time PCR results for the efficiency of DICER1 primers

	1st READ	2nd READ	Average
LC17 DICER1 1:1	19,40	20,1	19,75
LC17 DICER1 1:10	22,30	22,4	22,35
LC17 DICER1 1:100	27,30	27,1	27,20
LC17 DICER1 1:1000	29,40	30,5	29,95
LC17 DICER1 1:10000	33,30	35,2	34,25

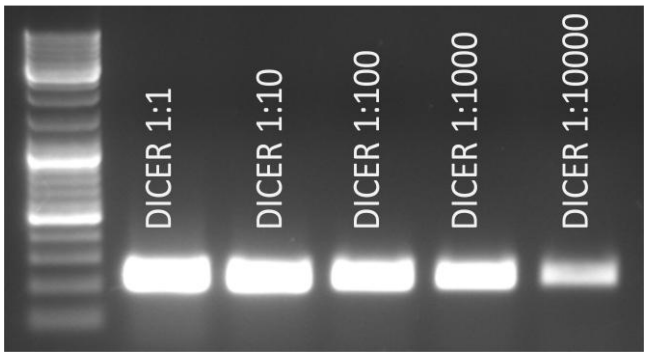


FIGURE S4 Agarose gel photo of the SK-LC-17 cell cDNA dilutions using DICER1primers. The cDNA dilutions are 1:1, 1:10, 1:100, 1:1000 and 1:10000 depicted in the gel photo.

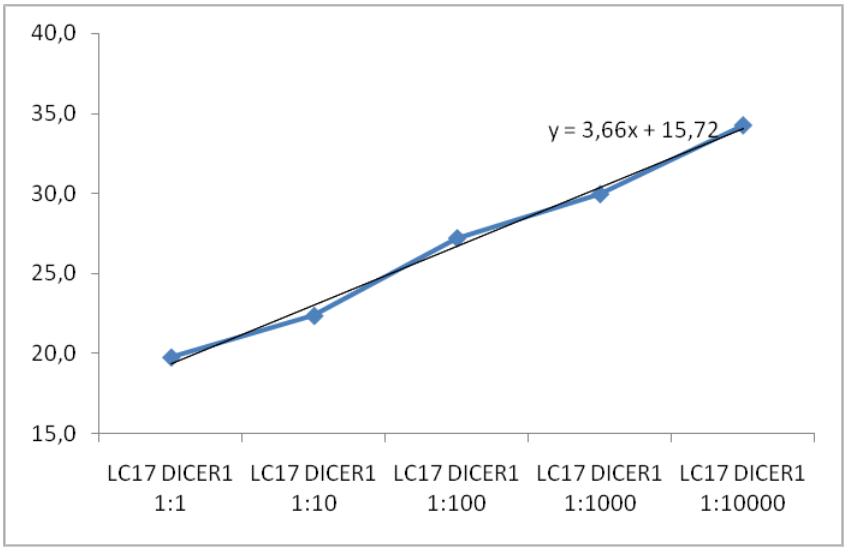


FIGURE S5 Efficiency graph of DICER1 primers.

The efficiency can be calculated from the formula:

Efficiency (E) = $D^{(1/m)}$

Where D is fold-dilution of cDNA

m is the slope of the trend line of data

Efficiency = $10^{-3.66}$ = 1,876

DICER1 mRNA levels in the shDICER clones

The calculations have been performed by delta CT method.

Q (sample quantity) = E (efficiency of the primer of interest)^{deltaCt}

The fold change differences relative to the parental cell line has been calculated as follows

$E_{DICER}^{(Ct\ parental\ -Ct\ clone)}/E_{referencegene}^{(Ct\ parental\ -Ct\ clone)}$

B2. Results

Two different reference genes have been used; GAPDH (E=1.9) and 18SrRNA (E=2.0) (the efficiencies of the reference genes were calculated before, MSc Thesis, Bulut A) and the DICER1 levels in the pSicoR-PGK-shDICER clones were calculated relative to the DICER1 levels in the parental SK-LC-17 cells.

TABLE S2. Quantitative Real Time PCR Results for pSicoR-PGK-shDICER constructs

DICER						
	1st read	**2nd read**	**3rd read**	**Average**	**Q**	**Fold change**
Parental SK-LC-17	23,8	24,6	23,5	23,97		
Clone 1.11	25,3	25,5	25,9	25,57	0,36	1,60
Clone 2.1	26,1	25,8	25,9	25,93	0,28	1,78
Clone 2.5	24,5	24,3	24,4	24,40	0,76	1,90
Clone 2.9	24,5	25,5	25	25,00	0,52	1,92
Clone 2.21	26	25,8	26,2	26,00	0,27	2,40
Clone 3.3	24,5	24,9	24,6	24,67	0,64	2,47
Clone 3.8	24,8	24,4	24,1	24,43	0,74	2,48
Clone 3.9	25,2	24,9		25,05	0,50	2,67
Clone 3.23	24,1	24,2	23,9	24,07	0,94	2,68
18SRNA					**Q**	
Parental SK-LC-17	8,9	8,9	8,3	8,70		
Clone 1.11	15,7	16,2	15,9	15,93	0,01	
Clone 2.1	14,3	14,9	14,6	14,60	0,02	
Clone 2.5	12,2	13,3	13	12,83	0,06	
Clone 2.9	12,9	12,8	13,3	13,00	0,05	
Clone 2.21	10,6	10,7	11,2	10,83	0,23	
Clone 3.3	9,8	10,3	9,9	10,00	0,41	
Clone 3.8	9,4	9,9	10,2	9,83	0,46	
Clone 3.9	9,3	9,2	9,6	9,37	0,63	
Clone 3.23	8,8	8,8	9,3	8,97	0,83	

DICER						
	1st read	**2nd read**	**3rd read**	**Average**	**Q**	**Fold Change**
Parental SK-LC-17	26,10	26,70	26,30	26,37		
Clone 1.11	29,80	27,80	27,60	28,40	0,27	1,36
Clone 2.1	27,00	28,60	28,00	27,87	0,38	1,35
Clone 2.5	26,70	26,40	27,60	26,90	0,71	1,34
Clone 2.9	25,90	25,90	26,90	26,23	1,09	1,28
Clone 2.21	26,70	26,90	26,90	26,83	0,74	1,47
Clone 3.3	25,50	24,90	25,20	25,20	2,11	1,42
Clone 3.8	25,30	24,60	25,10	25,00	2,40	1,34
Clone 3.9	25,60	25,20	25,10	25,30	1,98	1,39
Clone 3.23	24,90	24,40	24,20	24,50	3,31	1,45
GAPDH					**Q**	
Parental SK-LC-17	18,40	18,50	18,50	18,47		
Clone 1.11	21,20	20,60	21,00	20,93	0,21	
Clone 2.1	20,80	20,30	20,70	20,60	0,25	
Clone 2.5	19,90	20,00	20,20	20,03	0,37	
Clone 2.9	20,80	20,30	20,40	20,50	0,24	
Clone 2.21	18,50	18,30	17,90	18,23	1,16	
Clone 3.3	17,70	17,80	17,70	17,73	1,60	
Clone 3.8	18,80	18,70	18,50	18,67	0,88	
Clone 3.9	18,40	18,30	18,10	18,27	1,14	
Clone 3.23	16,80	17,20	16,60	16,87	2,79	

DICER						
	1st Read	**2nd Read**	**3rd Read**		**Q**	**Fold change**
Parental SK-LC-17	25,80	25,70	24,90	25,47		
Clone 1.11	29,80	27,80	27,60	28,40	0,13	1,38
Clone 2.1	27,00	28,60	28,00	27,87	0,19	1,59
Clone 2.5	26,70	26,40	27,60	26,90	0,37	2,09
Clone 2.9	25,90	25,90	26,90	26,23	0,59	4,59
Clone 2.21	26,70	26,90	26,90	26,83	0,39	0,63
Clone 3.3	25,50	24,90	25,20	25,20	1,20	1,38
Clone 3.8	25,30	24,60	25,10	25,00	1,38	3,03
Clone 3.9	25,60	25,20	25,10	25,30	1,12	1,87
Clone 3.23	24,90	24,40	24,20	24,50	1,95	1,23
GAPDH					**Q**	
Parental SK-LC-17	17,50	17,70	17,40	17,53		
Clone 1.11	21,20	20,60	21,00	20,93	0,09	
Clone 2.1	20,80	20,30	20,70	20,60	0,12	
Clone 2.5	19,90	20,00	20,20	20,03	0,18	
Clone 2.9	20,80	20,30	20,40	20,50	0,13	
Clone 2.21	18,50	18,30	17,90	18,23	0,62	
Clone 3.3	17,70	17,80	17,70	17,73	0,87	
Clone 3.8	18,80	18,70	18,50	18,67	0,46	
Clone 3.9	18,40	18,30	18,10	18,27	0,60	
Clone 3.23	16,80	17,20	16,60	16,87	1,59	

APPENDIX C. GENERATION OF DOUBLE REPORTER CONSTRUCTS

Vectors constructed in the analysis of bidirectional SSX4 promoter activity by flow cytometry and fluorescent miscoscopy.

C1. pCMV-TRFP

BglII/KpnI restricted CMV promoter cloned into the same site of pTurboRFP-PRL.

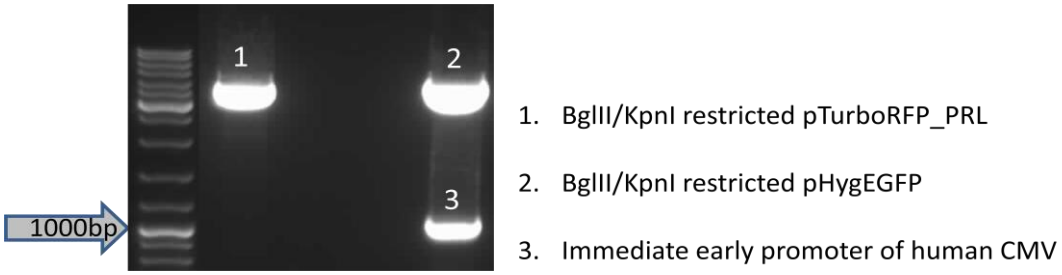
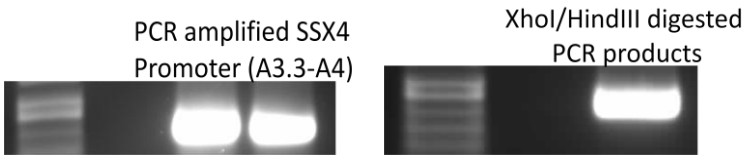


FIGURE S6. BglII and KpnI double digetsed pTurboRFP_PRL, pHygEGFP anf immediate early promoter of human cytomegalovirus (CMV).

C2. pSSX4P-TRFP and pSSX4PI-TRFP construct (TurboRFP under the control of SSX4 Promoter)

SSX4 promoter has been amplified by PCR by using primers A3.3 and A4. SSX4 Inverse promoter has been amplified by PCR by using primers A4 and A3.33. The PCR amplifications ar followed by XhoI and HindIII restriction enzymes double digestion.



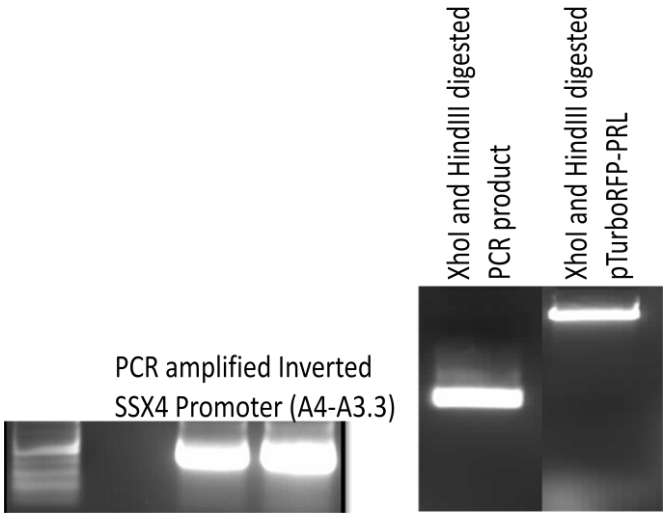


FIGURE S7. PCR amplified SSX4 sense and antisense promoter using the primers indicated and XhoI and HindIII double digested same PCR products.

XhoI and HindIII digested PCR products have been ligated to the same site of pTurboRFP-PRL. To control the resulting constructs XhoI and HindIII double digestions have been performed to excise the promoter fragment from the pTurboRFP-PRL construct. Also a control double digestion was performed to excise the same sequence from pGL3-SSX4-A3.3 construct.

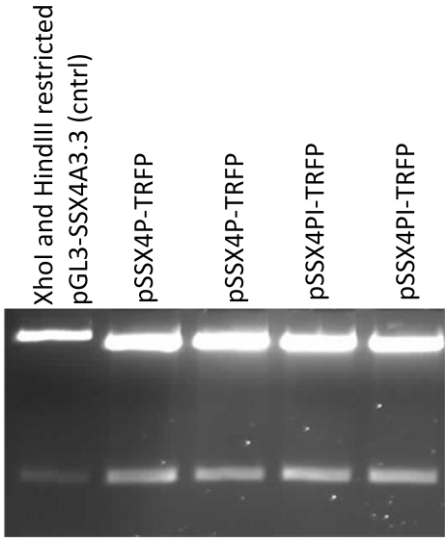


FIGURE S8. XhoI and HindIII double digested pSSX4P-TRFP and pSSX4PI-TRFP and excised fragments of sense and antisense SSX4 promoters. Control XhoI and HindIII double digested pGL3-SSX4A3.3 can also be seen in this gel photo.

C3. pGFP-SSX4PI-TRFP constructs

pHygEGFP vector is restricted with BamHI to excise HygEGFP from the construct.

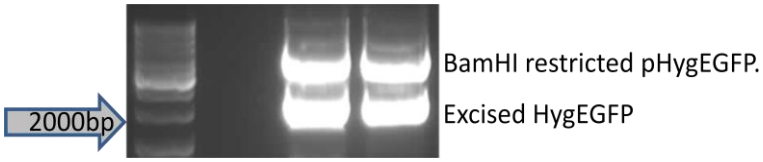


FIGURE S9. BamHI restricted pHygEGFP and excised HygEGFP are seen in this gel photo.

pTurboRFP-PRL construct was digested with BglII and treated with Arctic Phosphatase to prevent self ligation.

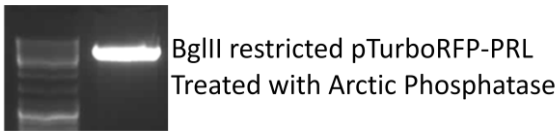


FIGURE S10. BglII restricted pTurboRFP-PRL is seen in this gel photo.

BamHI digested HygEGFP is ligated into BglII sites of pSSX4PI-TRFP construct. The direction of the HygEGFP was confirmed by NheI&XhoI double digestion. After NheI and XhoI double digestion if the orientation of the HygEGFP is correct, expected band size is 475 bp.

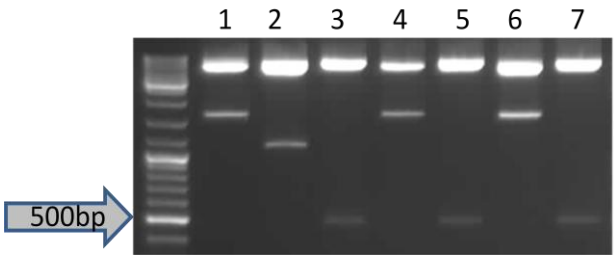


FIGURE S11. After NheI and XhoI double digested consructs showing the correctly oriented HygEGFP showing the expected digestion product of 475 bp .

Accordingly, the direction of HygEGFP in constructs 3,5,7 are correct and they are named as pGFP-SSX4PI-TRFP. On the other hand, the 1st, 4th and 6th constructs are named as pGFPI-SSX4PI-TRFP and used as control constructs.

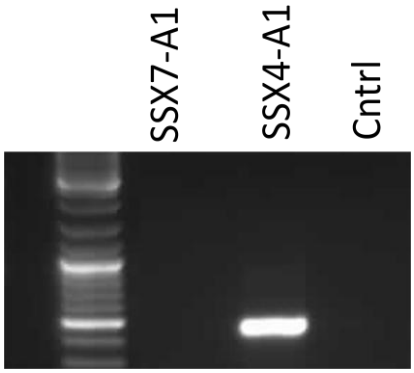


FIGURE S12. Control PCR of OATLA2/B2 primer pair to show that SSX4/OATL is specifically amplified.