UPPSALA
UNIVERSITET

Epigenetic regulation of the E-cadherin gene by the chromatin factor Hmga2

Gene silencing and its importance in cancer progression

Oskar Idås

Degree project in biology, Master of science (2 years), 2014
Examensarbete i biologi 45 hp till masterexamen, 2014
Biology Education Centre, Uppsala University, and Ludwiginstitutet för cancerforskning
Supervisors: E-Jean Tan and Aristidis Moustakas
External opponent: Panagiotis Papoutsoglou

Table of Contents

Abstract

Epithelial mesenchymal transition is important in embryonic development, but has also been associated with cancer progression and metastasis. The secreted protein TGFβ has an ambiguous role in cancer since it can act as both a tumor suppressor and a tumor promoter, depending on the cellular context. In this master thesis, TGFβ:s ability to induce EMT and promote tumor metastasis in mouse nemru murine mammary gland (NMuMG) cells will be addressed by measuring protein and mRNA levels of known epithelial and mesenchymal markers after TGFβ stimulation.

A hallmark of EMT is the loss of the cell-cell adhesion protein E-cadherin. The TGFβ/Smad pathway can induce repression of E-cadherin expression through the action of High mobility group A2 (Hmga2) and transcription factors Snail1 and Twist. This repression has been shown to remain in Hmga2 overexpressing NMuMG clones, even after silencing of Snail1 and Twist, which indicates that E-cadherin could be epigenetically silenced in these clones. Based on this result, the methylation status of the E-cadherin promoter region *Cdh1* has been analyzed by bisulfite sequencing.

We conclude that TGFβ induces an EMT response in NMuMG cells that can be reversed (MET) upon removal of TGFβ. We also show that the *Cdh1* promoter region is extensively methylated in Hmga2 overexpressing NMuMG clones. This result indicates that Hmga2, that previously been shown to regulate important EMT transcription factors, has an important role in silencing the E-cadherin promoter. In addition, we show that Hmga2 interacts with the genomic insulator CTCF that is known to protect tumor suppressor genes from repression.

Abbrevations

cDNA	Complementary DNA
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
EMT	Epithelial mesenchymal transition
GST	Glutathione S-transferase
HAT	Histone acetylase
HDAC	Histone deacetylase
Hmga2	High mobility group A2
MET	Mesenchymal epithelial transition
NMuMG	Nemru murine mammary gland
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
RT	Reverse transcriptase
shRNA	Short hairpin RNA
TF	Transcription factor
TGFβ	Transforming growth factor β
TGFβRI	TGFβ type I receptor
TGFβRII	TGFβ type II receptor

1. Introduction

Over 200 types of cancers have been found and in total more than 8 million people worldwide die every year beacause of it (de Martel *et al.* 2012). Although cancer is not one single disease but rather a group of diseases, they have at least one thing in common; unregulated cell growth. These are cells that no longer have the ability to respond to apoptotic signals and instead keep on proliferating and expand in a tissue or and organ. The expansion of a tumor can disrupt the function of an organ and eventually be fatal for the patient if not treated. Depending on the size and location of the tumor it can be removed surgically or treated with chemotherapy. The worst tumors are the ones that start metastasizing and forming secondary tumors in neighboring or distant tissues. When a primary tumor started to metastasize it is very hard to surgically remove the tumors and a complete removal of the tumors cannot be guaranteed. For a cell to become able to migrate and invade neighboring tissues, it needs to repress the expression of proteins like E-cadherin and coxsackie and adenovirus receptor (CAR), which form cell-cell junctions and increase expression of migratory proteins like fibronectin and N-cadherin, all of which will be discussed in more detail in this report. The process that generates migratory mesenchymal cells from epithelial cells is called epithelial-mesenchymal transition (EMT). Since this transition is tightly associated with cancer progression and metastasis, and thus of high importance in a pathology perspective, it has been a heavily studied mechanism in the field of molecular biology (Iwatsuki *et al.* 2010 and Guarino *et al.* 2007).

1.1 EMT

Epithelial-mesenchymal transition (EMT) is the process where static, polarized epithelial cells become invasive and mesenchymal. The transition emerges in a stepwise manner. First, the cell loses expression of cell-cell junction proteins and gains the expression of mesenchymal proteins. Secondly, the cell undergoes polarization followed by cytoskeleton remodeling, which involves the formation of actin stress fibers. Finally, the activation of proteplytic enzymes such as metalloproteases alters the cell-matrix adhesion, a step essential for mobility and tissue invasion (Iwatsuki *et al.* 2010). These cell alterations contribute to a more motile and invasive cell phenotype. After the transition, the mesenchymal cells can enter the circulation and migrate to surrounding tissues, an ability found in cells that are involved in embryonic development. The EMT process is reversible, enabling the mesenchymal cells to undergo mesenchymal-epithelial transition (MET) when they migrated to a new tissue location. These opposite processes are important for tissue formation during embryonic development and wound healing in adherent tissue, where they enable cells to migrate from one site to another (Iwatsuki *et al.* 2010).

In cancer the EMT/MET processes can be used to promote tumor progression (Heldin *et al* 2012). A primary tumor uses EMT to form migratory and invasive mesenchymal cells. These

cells are able to enter the blood stream and travel to distant tissues where they exit and undergo MET, forming secondary metastases. The tumor microenvironment can promote EMT through different cytokines that can activate several different cellular pathways. One of the major inducers of EMT during cancer progression is transforming growth factor β (TGFβ) (Piek *et al.* 1999).

1.2 TGFβ and its dual role in cancer

The TGFβ family consists of 33 members, all of them cytokines involved in developmental processes where they regulate cell growth, differentiation, migration, adhesion and death. During normal conditions and in early stage tumors TGFβ acts as a tumor suppressor by inducing growth arrest and apoptosis. However, in late stage tumors, the TGFβ signaling pathway starts promoting tumor progression and development instead (Heldin *et al.* 2012). This switch in TGFβ signaling function is known as the TGFβ paradox.

In mammalian cells three TGFβ isoforms are expressed (TGFβ-1, TGFβ-2 and TGFβ-3) (Javelaud and Mauviel 2004), with the most abundant isoform being TGFβ-1 that has been used in this project. The cytokine is secreted in an inactive form in a complex with latent transforming growth factor β binding proteins (LTBPs). These latent complexes can be cleaved in the extracellular matrix by metalloproteases, which activate TGFβ and enable its binding to the TGFβ receptor I and II (TGFβRI and II).

Initially TGFβ binds to TGFβR II, and then TGFβRI is recruited to the complex. Upon formation of the heteromeric complex, TGFβRII phosphorylates a TGFβRIs glycine-serine (GS) domain. This activates TGFβRI, which in turn will be able to phosphorylate intracellular substrates, among these receptor activated Smads (R-Smads) 2 and 3, which are the mediators of the TGFβ/Smad signaling pathway. The R-Smads can form complexes together with the common mediator Smad 4 (co-Smad) in the cytoplasm. These R-Smad/co-Smad complexes can enter the nucleus, recruit transcriptional activators to the DNA and promote transcription of certain genes (Dijke *et al.* 2002). In late stage cancers this TGFβ/Smad pathway has been shown to promote the transcription of EMT transcription factors (EMT TFs) that will help drive the EMT response (Fig. 1A).

1.3 Hmga2 is induced by TGFβ and regulates transcription of EMT TFs

One of the genes that are upregulated by the TGFβ/Smad pathway is high mobility group A2 (Hmga2). Hmga2 is a transcription factor that can bind to DNA through its AT-hooks (Fig. 1B). Normally Hmga2 is only expressed during embryonic development but very poorly in adult cells. When Hmga2 is abnormally expressed in adult cell it has been implicated in cancer progression (Fusco and Fedele 2007). Hmga2 can be upregulated by TGFβ and in turn activate a range of transcription factors important for EMT such as Snail1, Snail2, ZEB1, ZEB2 and Twist (Fig. 2A). These so called EMT TFs are responsible for repressing the transcription of epithelial genes and inducing the transcription of mesenchymal and pro-migratory genes. The fact that Hmga2 controls these EMT TFs makes it an important factor

in cancer progression and metastasis (Fig. 2C). Accordingly, cells that express Hmga2 are mesenchymal and do not express E-cadherin.

Although Hmga2 has been shown to act as a master regulator of these transcription factors, knock down experiments have shown that the E-cadherin expression cannot be rescued after silencing of Snail1 and Twist in Hmga2 overexpressing clones (Tan *et al.* 2012). In this project the role of Hmga2 in epigenetically silencing the E-cadherin gene promoter will be further addressed.

1.4 Transcriptional regulation of E-cadherin

As previously mentioned, one of the hallmarks of EMT is the loss of the transmembrane glycoprotein E-cadherin (Yoshiura *et al.* 1995 and Thiery *et al.* 2009). E-cadherin is important for cell-cell adhesion and the maintenance of epithelial cell architecture. Upon down regulation of E-cadherin the cell loses contact with neighboring cells and becomes more motile (Onder *et al.* 2008).

Snail1 is a zinc-finger transcription factor that, along with the ZEBs, is responsible for the repression of E-cadherin. Snail1 is induced by Hmga2, which, in turn is induced by TGFβ. Hmga2 has been proven to bind directly to the Snail1 promoter, forming a complex with Smads induced by the TGFβ/Smad pathway that promotes Snail1 expression. Snail1 has a highly conserved C-terminal domain containing tandem zinc-fingers that recognize E2-box type elements C/A (CAGGTG) (Nieto 2002). These types of E-boxes can be found within the promoter region of E-cadherin, enabling Snail1 to bind and repress its activity (Fig. 2B).

1.5 Epigenetic silencing in cancer

Gene expression can be regulated not only by transcription factors, but also by histone modifications and DNA methylation. This type of regulation is essential for development and tissue formation. Since all cells in an organism have the same genetic information to start with, they need to reduce or repress expression of certain genes in order to be able to differentiate into specific cell types. In terms of cancer progression, epigenetic rewiring is needed in order to turn of tumor suppressor genes, allowing cancer cells to proliferate and invade surrounding tissue (Suvá *et al.* 2013 and Mazzio *et al.* 2012).

Histone modifications offer dynamic ways of altering the local chromatin structure, enabling or disabling transcriptional complex to access the DNA. The most common histone modifications in terms of transcriptional potential are acetylation/deacetylation and methylation/demethylation. Histone acetylation, performed by histone acetyltransferases (HATs), introduces an ”open chromatin structure” that enhances binding of the transcriptional machinery (Mazzio *et al.* 2012). The reverse process is performed by histone deacetylaces (HDACs), which closes the chromatin structure and decreases the binding of the transcriptional machinery (Görisch *et al.* 2005). Histone methylation and demethylation, on the other hand, has been associated to both transcriptional induction and repression, depending on which histone that is modified (Stewart *et al.* 2005).

The more stable type of DNA modification in a long-term perspective is DNA methylation, which is carried out by members of the DNA methyltransferase (DNMT) family. DNMTs have the ability to add a methyl group to the 5th carbon ring of cytosine (Kumar *et al.* 1994). This modification is almost exclusively targeted towards cytosines that are bound to a guanine (CpG dinucleotides). Clusters of CpG sites, so called CpG islands, can be found at a low frequency throughout the genome, although they tend to aggregate within proximity to gene promoters. The CpG islands that are located close to the promoter region can be methylated by DNMTs, which inhibits the transcriptional machinery to bind and consequently silences the gene. In normal cells these promoters are generally unmethylated. A hallmark in cancer cells, however, is the hyper methylation of such promoter regions (Feinberg 1983).

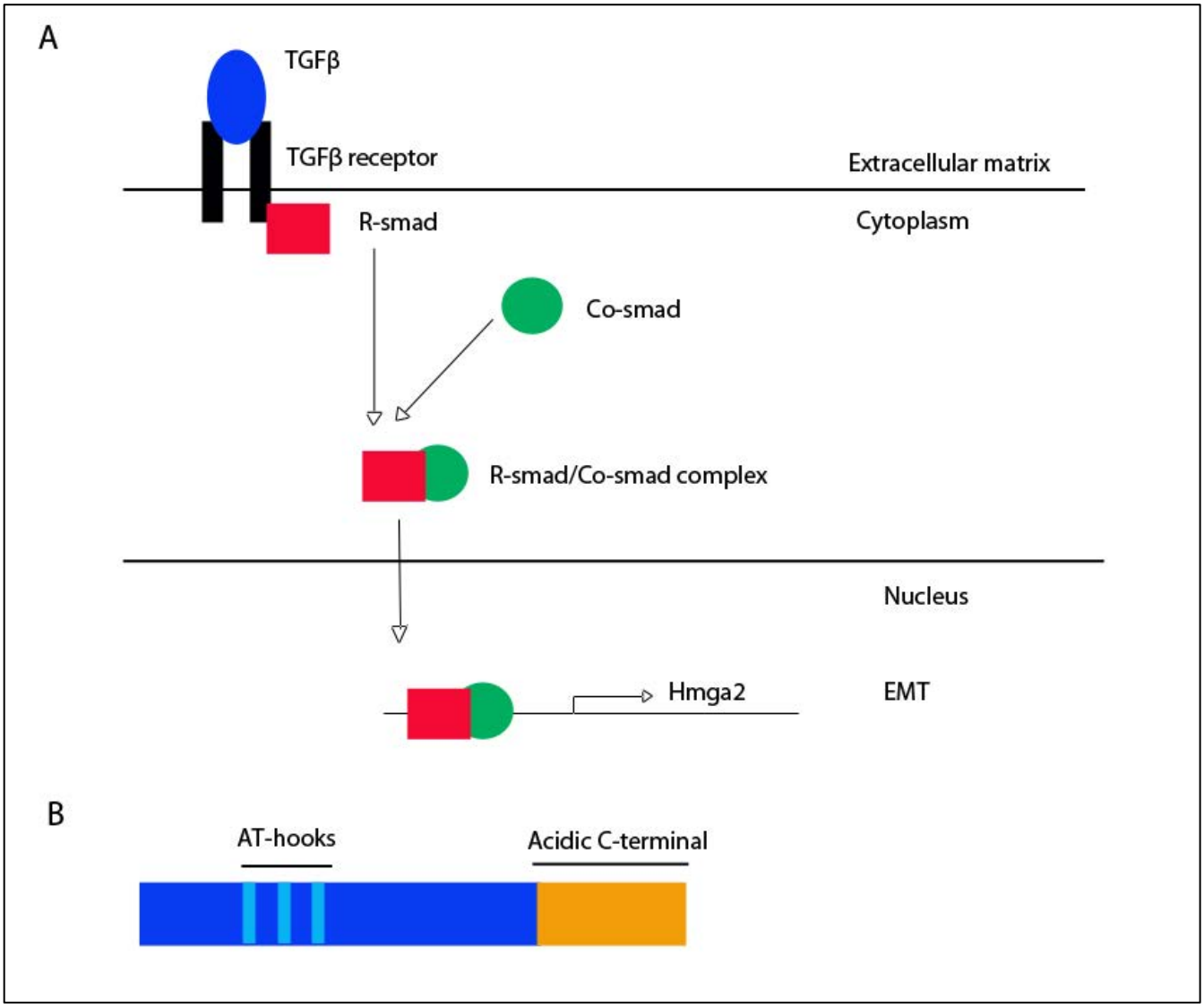

Figure 1. ***A****, The TGFβ signaling pathway upregulates, among other genes, Hmga2 that is a master regulator of EMT transcription factors, that promotes EMT.* ***B****, Schematic figure of Hmga2. The three central AT-hooks are used to bind DNA, the acidic C-terminal is important for protein interactions.*

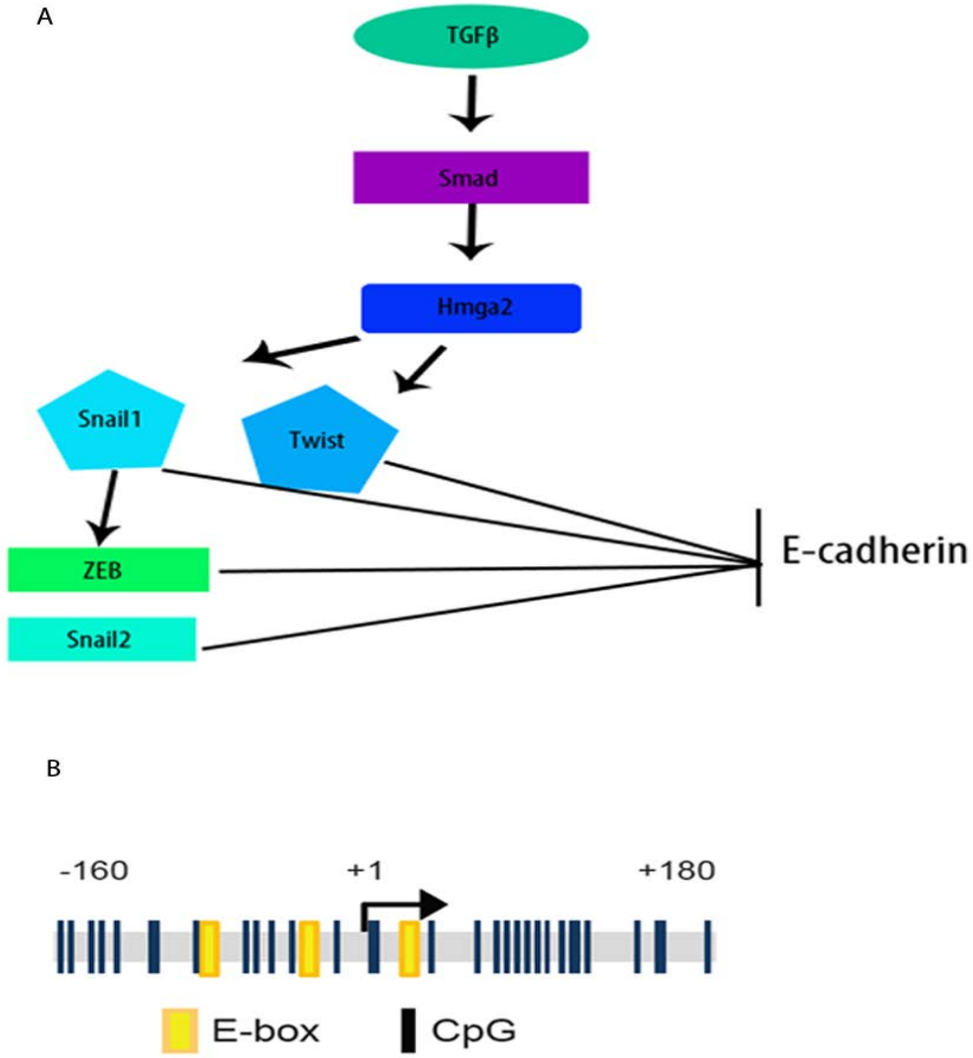

Figure 2. ***A****, The factors involved in E-cadherin repression, a hallmark of EMT.* ***B****, The promoter region of the E-cadherin gene (Cdh1).*

1.6 Aims of the project

The primary aims of this project were to analyze the EMT response in mouse NMuMG (nemru murine mammary gland) cells by looking at the mRNA and protein level response of well known EMT markers upon stimulation with TGFβ. In order to determine if the cells can undergo MET after the EMT response, TGFβ withdrawal experiments were performed.

Apart from EMT studies, the potential role of Hmga2 in the epigenetic regulation of the E-cadherin gene (*Cdh1*) was analyzed. To explore this, the methylation pattern of the Cdh1 promoter in Hmga2 overexpressing clones was analyzed.

Since DNMTs are important in DNA methylation, members from this family were studied and their involvement in epigenetic silencing of *Cdh1* and EMT was further explored.

2. Methods and materials

2.1 NMuMG cells

The cells used in this project have been exclusively mouse NMuMG cells. These are epithelial cells extracted from a mouse mammary gland tumor, which have a strong EMT response when stimulated with TGFβ (Miettinen *et al.* 1994).

2.2 Cell culture conditions and lysis

The NMuMG cells are grown in petri dishes with complete medium (Dulbecco´s modified eagle´s medium (DMEM), 10% fetal bovine serum and 10 ng/ml insulin) at 37°C. The TGFβ concentration the cells were treated with 5ng/ml. When the cells reach 80% confluency they were trypsinized and collected in complete medium. After that they were transferred (1:5 ratio) to a new plate.

The following procedure was used in order to extract the protein from treated cells:

- Media was removed.
- The cells wore washed with 1x PBS (phosphate buffered saline) buffer
  - 1x PBS buffer: 137 mM NaCl, 2,7 mM KCl, 10 mM Na_2HPO_4, 1,8 mM KH_2PO_4.
- Add 200µl 1x SDS (sodium dodecyl sulfate) buffer.
  - 1x SDS buffer: 100 mM Tris-Cl, 4% (w/v) sodium dodecyl sulfate, 0,2% (w/v) bromophenol blue, 20% (v/v) glycerol, 200 mM dithiothreitol.
- Use a cell scraper to spread SDS buffer on plate.
- Transfer cell lysate to 1.5 ml tube and sonicate samples for 5 min with a 30 sec interval.
- Boil samples for 5 minutes and store store them at -20°C.

2.3 Polyacrylamide gel casting

Depending on the size of the proteins of interest, the percentage of the running gel differed. Either an 8%, 10% or 12% polyacrylamide gel was used (table 1.). The ingredients for the stacking gel remained the same regardless the percentage of the running gel (table 2.).

Table 1. Polyacrylamide gel (12%, 10% or 8%) ingredients.

	12%	**10%**	**8%**
40% Acrylamide	3ml	2.5ml	2ml
2 M Tris-Cl pH 8.8	2ml	2ml	2ml
20% SDS	50µl	50µl	50µl
Water	5.6ml	5.6ml	6ml
TEMED	10µl	10µl	10µl
10% APS	40µl	40µl	40µl

Table 2. Stacking gel ingredients (4%).

	4%
40% Acrylamide	0.6ml
0,5M Tris-Cl pH 6,8	0.84ml
20% SDS	30µl
87% Glycerol	0.8ml
Water	3.8ml
TEMED	6µl
10% APS	30µl

2.4 Gel electrophoresis and semi-dry transfer (Western blot)

When the cell lysate samples are to be electrophoresed, they are boiled for 5 min before being loaded to the gel wells. In addition, a PageRuler prestained protein ladder was added to a neighboring well in order to monitor the size of the proteins during and after electrophoresis.

Running buffer composition (1x):

- 0,025 M Tris-base
- 0,192 M glycine
- 0.1% (w/v) SDS

To run the electrophoresis, the gel was scaffolded on a tray, and the tray was clamped into a small tank that was filled with running buffer. The sample buffer was added to the wells, for a 15-well gel, the usual sample volume was 30 µl. The volume of the protein ladder was around 5 µl protein ladder + 5 µl of 1x SDS loading buffer in one well. The gel tank was connected to a power pack and ran for 80 min at 150 V constant at room temperature.

When the protein bands had separated a semi-dry transfer was performed in order to transfer the proteins to a nitrocellulose membrane (Western blot). The Western blot sandwich was assembled by soaking six filter papers, a nitrocellulose membrane and the gel in 1x transfer buffer and putting them in a semi-dry immunoblotter in the following order: three filter papers – nitrocellulose membrane – gel – three filter papers. Air bubbles were removed by rolling a rod and the immunoblotter was connected to a power pack and ran it at constant 15V for 60 min.

After 60 minutes, the protein bands have been transferred to the nitrocellulose membrane, which could be visualized by soaking the membrane in 1x Ponceau red solution.

2.5 Immunodetection of proteins on a Western blot

First, the membrane was treated with a blocking solution (Bovine serum albulin (BSA) or dry, fat milk powder plus BSA, 5% w/v in TBS-T) for one hour. After blocking, the membrane was incubated along with the primary antibody over night. Then the membrane

was washed with TBS-T (50 mM Tris, 150 mM NaCl and 0.05% Tween 20) for five minutes (x3). After washing, the membrane was incubated with the secondary antibody for one hour and then washed again (3x) with TBS-T for 5 minutes. When the membrane had been washed, 1 ml mix of ECL reagent 1 (HRP substrate luminol reagent, Millipore) and 2 (HRP substrate proxide solution, Millipore) was added to the membrane, and incubated at room temperature for 1 minute. The excess liquid was drained off and the membrane put between two transparency sheets, proceeded to a dark room to expose and develop an X-ray film (Fujifilm) or to a charged coupled device (CCD) camera (Fujifulm Intelligent dark box II) to obtain a digital image.

Antibodies used: Fibronectin (anti-fibronectin rabbit, dilution 1:20 000, Sigma), E-cadherin (anti-E-cadherin mouse, dilution 1:20 000, BD Biosciences), N-cadherin (anti-N-cadherin mouse, dilution 1:20 000, BD Biosciences), Hmga2 (anti-Hmga2 rabbit, dilution 1:1000, Santa Cruz), Snail1 (anti-Snail1 rabbit, dilution 1:2000, Santa Cruz), α-tubulin (anti-α-tubulin mouse, dilution 1:5000, Santa Cruz).

2.6 Analysis of mRNA levels

2.6.1 Stimulation of NMuMG cells with TGFβ

NMuMG cells were plated in three wells of a 6-well plate ($8x10^3$ cells/well) and put in a 37°C, 5 % CO_2 incubator overnight. The next day the cells were washed twice with PBS. After washing the cells were starved with 1% FBS medium for 9 hours. After starvation the regular 10% media was added to all cells, and the samples were stimulated with 5 ng/ml TGFβ for 0, 24 and 48 hours respectilvey.

2.6.2 RNA extraction, purification and cDNA synthesis

After the TGFβ stimulation, the RNA was extracted and purified using QIAGEN RNeasy® Mini Kit.

The RNA was reverse transcribed into complementary DNA (cDNA) using iScript™ cDNA Synthesis Kit (table 3).

Table 3. cDNA synthesis reaction mix.

	Volume
5x iScript reaction mix	4µl
iScript reverse transcriptase	1µl
RNA template	1µg
Nuclease-free water	X µl
Total volume	20µl

Thermocycler program:
5 minutes, 24°C
30 minutes, 42°C
5 minutes, 85°C

Hold, 4°C

After cDNA synthesis, the samples were diluted 10x in dH_2O and stored at -20°C.

2.6.3 Real-time PCR

Primers (Forward/Reverse) used: Gapdh F/R (mouse, Sigma Genosys), Snail F/R (human, Sigma Genosys), E-cadherin F/R (mouse, Sigma Genosys) and Fibronectin F/R (mouse, Sigma Genosys).

Table 4. Real-time PCR reaction mix.

	Volume
10µM forward primer	3.6µl
10µM reverse primer	3.6µl
Sybergreen	90.6µl
Nuclease-free water	47.1µl
Total volume	145µl

Each master mix (table 4) was divided to three separate tubes, where 15µl of respective cDNA sample (0, 24, 48h) was added. From this mix, 15µl was added in triplicates to a 96-well qPCR plate and amplified in using qPCR (CFX96 Touch™ Real-Time PCR Detection System).

The collected triplicate-data was normalized against the housekeeping gene Gapdh:s quantitation values and plotted in a bar graph (Fig. 6A).

2.7 Phase contrast microscopy

NMuMG cells were grown on petri dishes (Ø10-cm) in complete medium and stimulated with TGFβ (5 ng/ml). The TGFβ stimulation was divided in two categories; short-term and long-term. In the short-term experiment the NMuMG cells were stimulated for 0, 24 and 48 hours. In the long-term experiment the time points were 0 and 3 weeks, in this experiment the cells were split to a new plate (1:5 ratio) upon reaching 80% confluency. After TGFβ stimulation the cells were photographed in a phase contrast microscope (Zeiss Axiovert 40 CFL) in order to see the morphology.

2.8 Luciferase promoter assay

2.8.1 Transfection and TGFβ stimulation

NMuMG cells were plated to 9 wells (10,000 cells/well) of 24-well plate, the plate was incubated in a 37°C, 5 % CO_2 incubator overnight.

After 24 hours the cells reached approximately 20% confluency. The following transfection mix setup was prepared:

A	Opti-MEM medium (GIBCO Lifesciences)	475µl
	Lipofectamine 2000	19µl
B	Opti-MEM medium	475µl
	p(-1359) Ecad-Inc plasmid DNA (0.1 µg/ µl)	9.5µl
	pCMV-βgal plasmid DNA (0.1 µg/µl)	9.5µl

The solutions were incubated for 5 minutes at room temperature, then pooled together and incubated for 20 minutes at room temperature.

The media was removed from the cells (20 % confluent) and 500µl complete medium + 104 µl A/B mix was added to each well. The 24-well plate was swirled and incubated in a 37°C, 5% CO_2 incubator overnight.

After 24 hours, the media was removed and new complete media was added. TGFβ1 was added (5 ng/ml) to three wells (48 hours of TGFβ stimulation triplicate sample). The plate was incubated in a 37°C, 5% CO_2 incubator for 24 hours.

The next day three wells (24 hours triplicate) were stimulated with 5 ng/ml TGFβ1. The plate was then again incubated in a 37°C, 5% CO_2 incubator for 24 hours.

2.8.2 Harvesting and luciferase assay

The cells were washed twice with PBS. 100µl 1x lysis buffer (0.4 mM DTT, 0.4 mM 1,2-diaminocyclohexanetetraacetic acid, 1 % Triton X-100, 5 mM Tris-phosphate pH 7.8) was added to each well and the cells were scraped and collected in separate tubes. The tubes were snap freezed in liquid nitrogen, thawed and centrifuged at 5,000 rpm for 4 minutes. The sample was divided to two 96-well plates.

First, 20µl from each cell lysate + 20µl 2x β- Gal (200 nM sodium phosphate pH 7.3, 2 mM $MgCl_2$, 100 mM β-mercaptoethanol, 1.33 mg/ml ortho-Nitrophenyl-β-galactoside) was added into wells of a transparent 96-well plate. In addition to the sample triplicates a blank triplicate (20µl 1x lysis buffer) was added (12 wells in total). The plate was incubated at 37°C until the reaction turned yellow upon which the absorbance was read at 450 nm by using an EnSpire® Multimode Plate Reader.

Second, 20µl from each cell lysate was added to the wells of a non-transparent 96-well plate, a triplicate of 1x lysis buffer was used as a blank. The cell lysate was mixed with the luciferase substrate (BD pharmingen) and the luciferase promoter activity was measured by using an EnSpire® Multimode Plate Reader.

The promoter activity data was normalized against the β- Gal values and plotted in a bar graph (Fig. 7B).

2.9 Immunofluorescence microscopy

2.9.1 Cell fixation

Glass cover slips (Ø16-mm) were placed into each well of a 12-well plate. The plate was then microwaved for 3 minutes at maximum power. Let the plate was the cooled down and the NMuMG cells were transferred onto the cover slips ($2x10^5$ cells/well). The plates were incubated in a 37 °C, 5% CO_2 incubator overnight.

The cells were washed with 1ml cold PBS (x2). The cells that would be stained for fibronectin were then fixed onto the cover slips with 4% (w/v) paraformaldehyde/PBS for 20 minutes at room temperature. The cells that would be stained for E-cadherin were fixed to the cover slips with MeOH/Acetone (1:1 v/v ratio) for 5 minutes at -20 °C. After the fixation the cells were washed again with 1ml cold PBS (x2). The samples were stored overnight in 1ml PBS at 4 °C.

2.9.2 Blocking and antibody incubation

The cover slips were transferred to the wells of a new 12-well plate, with each well containing 1ml TBS-T. The liquid was decanted between each transfer. Each cover slip was blocked with 1ml of 5% (w/v) dry milk in TBS-T, and were put on a shaker for one hour at room temperature. The milk was washed away by adding 1ml of TBS-T, and the cover slips were shaken for 5 minutes at room temperature (x3).

The primary antibodies were diluted in TBS-T;

- E-cadherin (mouse monoclonal antibody, BD Biosciences, 1:100 (v/v) dilution).
- Fibronectin (rabbit polyclonal antibody, Sigma, 1:200 (v/v) dilution).

A layer of parafilm was put in the bottom of a Ø15-cm petri dish, onto this layer the primary antibody was added (30µl/sample). Onto each drop the cover slip was added with the fixated cells facing down. In the corner of the petri dish a wet tissue was placed to keep a humid environment within the petri dish. The petri dish was then incubated overnight at 4 °C.

The cover slips were removed and added to new 12-well plates containing 1ml TBS-T. The cover slips were then washed (x2) with TBS-T on a shaker for 10 minutes at room temperature.

The secondary antibodies (fluorescently labeled and therefore light sensitive) were diluted in TBS-T;

- *anti-Mouse secondary antibody, 1:1000 dilution (Red fluorescence)-corresponded to the E-cadherin primary antibody.*
- *anti-Rabbit secondary antibody, 1:1000 dilution (Green fluorescence)-correspnded to the fibronectin primary antibody.*

A layer of parafilm was put in the bottom of a Ø15-cm petri dish, onto this layer the primary antibodywas added (30µl/sample). Onto each drop the cover slip was added with the

fixated cells facing down. In the corner of the petri dish a wet tissue was placed to keep a humid environment within the petri dish. The petri dish was then incubated for one hour at room temperature in a dark environment.

The cover slips were removed and added to new 12-well plates containing 1ml TBS-T. The cover slips are then washed (x2) with TBS-T on a shaker for 10 minutes at room temperature.

60µl of 4',6-diamidino-2-phenylindole (DAPI) solution (1µg/ml) was added to each cover slip and incubated for 10 minutes at room temperature. The cover slips were washed (x2) with 1ml of TBS-T for 5 minutes at room temperature. Approximately 15µl of mounting solution was added onto the glass slides, then the cover slips were added to the mounting solution, with the cells facing down. Slides were stored at 4 °C in dark environment.

The samples were analyzed and photographed digitally using a fluorescence microscope (Zeiss Imager M2)

2.10 DNMT1 and DNMT3A siRNA transfection

2.10.1 Transfection setup

10,000 cells/well were added to a 12-well plate, the plate was incubated in a 37°C, 5% CO_2 incubator overnight.

The following setup was prepared (12 tubes in total) in order to obtain a transfection mix that contained 100nM siRNA (short interfering RNA):

A1-6	Opti-MEM medium (GIBCO Lifesciences)	97.5µl
	Silentfect (BioRad Ltd)	2.5µl
B1-2	Opti-MEM medium	75µl
	siCtrl (2 µM)	25µl
B3-4	Opti-MEM medium	90µl
	siDNMT1 (5 µM)	10µl
B5-6	Opti-MEM medium	90µl
	siDNMT3A (5 µM)	10µl

A1 was pooled with B1, A2 was pooled with B2, etc. The tubes were then incubated at room temperature for 20 minutes. After the incubation, 300µl of complete medium was added to each transfection mix. The media were removed from the cells and the transfection mix was added to its respective well (6 wells in total). The plate was put in a 37°C, 5% CO_2 incubator overnight.

The transfection mix was removed and the cells were washed with 1x PBS. Starving media (complete media with 1% FBS) was added to each well and the plate was put in 37°C, 5% CO_2 incubator for 6 hours. The starvation media was replaced with complete media and at the

same time as TGFβ (5ng/ml) was added to one transfection set (3 wells in total). The plate was put in 37 °C, 5% CO_2 incubator for 48 hours.

The siRNAs used for these experiments were purchased from Dharmacon Inc. and were pools of 4 different siRNAs targeting the specific mouse gene. The three siRNAs described above were: siCtrl: control non-targeting siRNA (ON-TARGETplus siCONTROL, D-001810-10); siDNMT1: specific siRNA targeting mouse DNMT1 (Dharmacon ON-TARGETplus SMART pool L-056796-01); siDNMT3A: specific siRNA targeting mouse DNMT3A (Dharmacon ON-TARGETplus SMARTpool L-065433-01).

2.10.2 Protein analysis by Western blot

The transfected cells were washed with 1x PBS and 200µl of 1x SDS loading buffer was added to each well as described above. The cell lysate was scraped and collected in a 1.5ml eppendorf tube. The samples were sonicated (5 minutes with 30 second pulse) and boiled for 5 minutes before they were stored at -20°C.

The samples were run on 8% SDS-PAGE, immunoblotting for DNMT1 (anti-DNMT1 rabbit antibody, dilution1:1000, Santa Cruz) and DNMT3A (anti-DNMT3A mouse antibody, dilution 1:1000, Abcam) to validate the knock down. The samples were also immunoblotted for fibronectin (anti-fibronectin rabbit antibody, dilution 1:20 000, Sigma) and E-cadherin (anti-E-cadherin mouse antibody, dilution1:20 000, BD Biosciences).

2.11 Bisulfite sequencing

2.11.1 Bisulfite conversion of DNA

Genomic DNA from the five following stable clones was extracted:

- Mock (NMuMG)
- Hmga2
- Hmga2 + shSnail1
- Hmga2 + shSnail1 + shTwist #49
- Hmga2 + shSnail1 + shTwist #50

The bisulfite treatment was performed using the Epitect Bisulfite kit (QIAGEN). A bisulfite mix and DNA protection buffer was added to each genomic DNA sample (1ng/sample) according to the manufacturer´s protocol, the DNA was then incubated in a thermocycler at the following program:
5 minutes, 95°C
25 minutes, 60°C
5 minutes, 95°C
1 hour 25 minutes, 60°C
5 minutes, 95°C

2 hours 55 minutes, 60°C
Hold, 20°C

The bisulfite treated DNA was then purified using the Epitect Bisulfite kit (QIAGEN) according to the manufacturer´s protocol.

2.11.2 Amplification of bisulfite treated DNA by PCR

The region of interest was amplified by using the following single stranded DNAprimers: 5'-TGGGTTAGAGTATAGTTAGGTTAGG-3' (sense) and: 5'-AATCAAAACCCTCCACATACCTACA-3' (antisense). For each sample, the reaction mix stated in table 5 was used.

Table 5. PCR reaction mix.

	Volume
10x PCR buffer	5µl
10mM dNTP mix (all 4 dNTPs)	1µl
50nM $MgCl_2$	1.5µl
10µM sense primer	0.5µl
10µM antisense primer	0.5µl
Platinum Taq DNA polymerase	0.2µl
dH_2O	40.3µl
Template DNA	1µl
Total volume	50µl

The amplification condition:
2 minutes, 94°C
30 seconds, 94°C
30 seconds, 55°C
1 minute, 72°C
} 35 cycles
2 minutes, 72°C
Hold, 4°C

In order to purify the PCR product the total amount of amplified DNA was electrophoresed on a 0.8% agarose gel and the DNA band at 416 bp was cut out and extracted from the gel using a QIAquick Gel Extraction Kit according to the manufacturer´s protocol (Fig. 5A).

2.11.3 Cloning the PCR product into a vector

The 416bp sized PCR product was cloned into a TOPO® vector using the reaction mix stated in table 6.

Table 6. TOPO cloning reaction mix.

	Volume
PCR product	3µl
Salt solution (200nM NaCl, 10mM $MgCl_2$)	1µl
H_2O	1µl
PCR™ II TOPO® Vector	1µl
Total volume	6µl

The mixture was incubated at 25°C for 30 minutes and then stored at -20°C.

2.11.4 *E.coli* transformation and growth

The transformation was performed following the TOPO® TA Cloning® Kit protocol (Invitrogen) for transforming Shot® TOP10 Competent *E.coli*. At this point, a cloning control was added beyond the five samples, thus named sample "6". This control only contained an empty plasmid (pUC19). The transformed cells were spread on LB + Ampicillin (100µg/ml) plates and incubated at 37°C for 20 hours. The cloning control sample was spread on a LB + Kanamycin (50µg/ml) plate.

On each plate >20 colonies grew, bacteria from 10 colonies/plate were picked and inoculated in 3ml LB + Ampicillin (100µg/ml) media (Kanamycin (50µg/ml) for the cloning control), then incubated for 20 hours at 37°C. This resulted in 51 bacterial mini-cultures in total.

2.11.5 PCR analysis of transformants

First, DNA from one mini-culture per sample and the cloning control were separately amplified using the PCR reaction mix stated in table 7.

Table 7. PCR reaction mix.

	Volume
10x PCR buffer	2µl
10mM dNTP mix (all 4 dNTPs)	0.5µl
50nM $MgCl_2$	1µl
10µM sense primer	2µl
10µM antisense primer	2µl
DMSO	0.4µl
Platinum Taq DNA polymerase	0.5µl
dH_2O	10.6µl
Template DNA	1µl
Total volume	20µl

Amplification condition:

8 minutes, 94°C
30 seconds, 94°C } 30 cycles

30 seconds, 54°C
1 minute, 72°C
5 minutes, 72°C
Hold, 4°C

The PCR product was electrophoresed on a 1% agarose gel along with a 1kbp DNA ladder and DNA was stained with 10x Orange G loading dye and photographed digitally using a CCD camera (Fig. 5B). As shown in figure 5B, samples 1-5 are around 416bp in size, whereas the cloning vector control (sample 6) is 750bp. The other colonies were analyzed in the same way to confirm successful transformation. The plasmids were purified and isolated from the colonies using QIAprep Spin Miniprep Kit (QIAGEN) according to the manufacturers protocol. A small aliquot was sent for DNA sequencing to the company Eurofins.

2.11.6 Sequence analysis

The raw sequence data (Supplementary data fig. 11) were analyzed using the BiQ Analyzer, a program that helps one align and compare genetic sequences. By comparing the bisulfite treated sequence with the original sequence the methylation pattern was obtained. By using this feature we were able to see which cytosine had been converted into uracil, thus revealing the methylation pattern of that particular sequence.

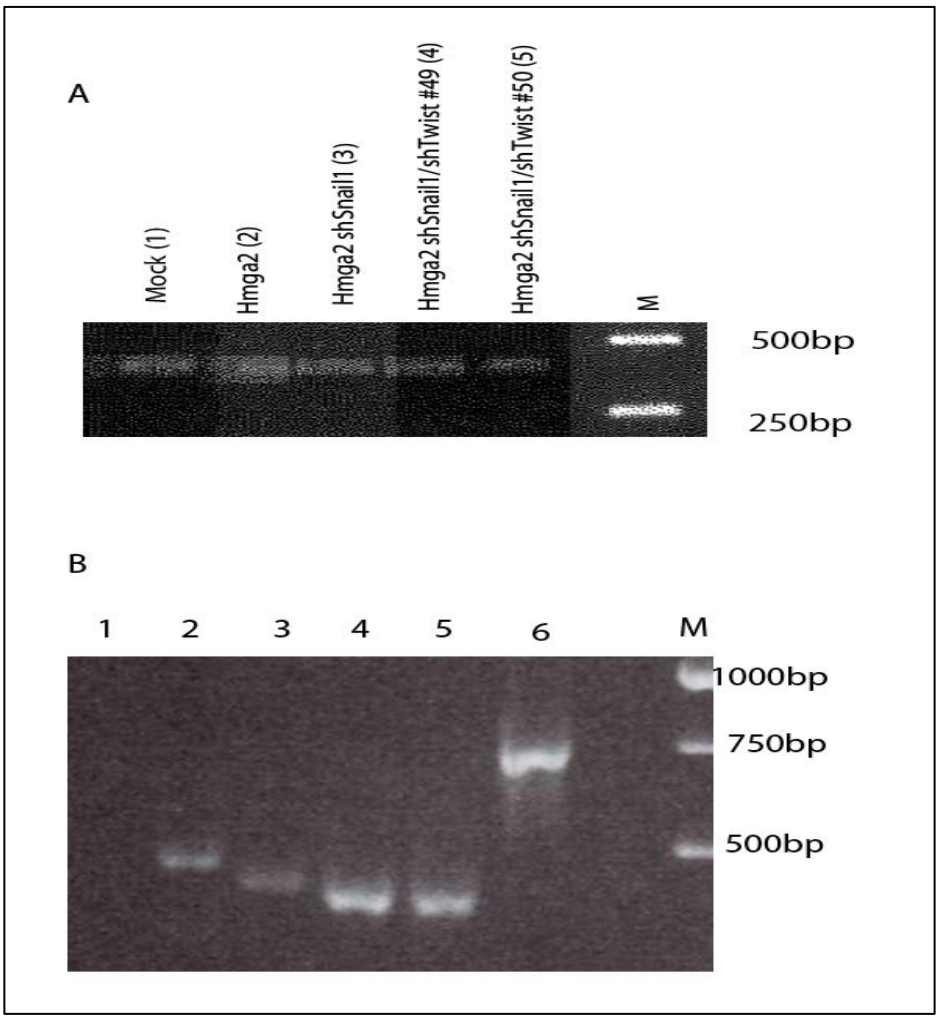

Figure 3. ***A****, Amplified Cdh1 promoter region (416 bp) from the 5 indicated NMuMG clones.* ***B****, The amplified Cdh1 promoter region extracted from E. coli (416 bp), lanes 1-5 corresponds to each of the NMuMG cell clones shown in panel A. Lane 6 is a vector control (750 bp). In panels A and B, M indicates the molecular size ladder and sizes in bp are shown on the right of each gel.*

2.12 GST-fusion protein preparation

Buffers

GST purification buffer:

	Volume
1X PBS	376 ml
1 M Tris pH 8.0	20 ml
10% Triton X-100	4 ml
1 M $MgCl_2$	0.2 ml

Lysis buffer:

	Volume
GST purification buffer	20 ml
100mg/ml Lysozyme	200 μl
1 M DTT (dithiothreitol)	100 μl
0.1 M PMSF (phenyl-methyl-sulfoxide)	100 μl
Aprotinin	200 μl

Hmga2-constructs:

Full length
Deleted C-terminal
C-terminal
N-terminal + one AT-hook
N-terminal

2.12.1 Growth of *E.coli* containing the Hmga2 constructs

A frozen *E.*coli glycerol stock containing pGEX plasmid with either construct was scraped using a pipette tip, end the pipette tip inoculated in 5 ml LB media containing 100μl/ml ampicillin. The same procedure was followed for each construct (5 samples in total). The *E.coli* samples were incubated for 16 hours in a shaker at 37°C.

The next day, the cultures were diluted (1:10 dilution) with 45 ml LB media containing 100 μl/ml ampicillin. The cultures were incubated in 100 ml flasks in a shaker at 37°C until the bacterial density reached A_{600nm}= 0.5-0.8 (~2 hours). When the right A_{600nm}-value was reached, 0.3 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) were added and the incubation continued in a shaker at room temperature for 6 hours in order to induce transcription of the constructs and synthesis of the corresponding proteins. The culture was transferred to 50 ml tubes and spun for 15 minutes at 4,000 rpm, 4°C. The supernatant was discarded and the pellet stored (if needed, storage at -20°C). The cell pellet was resuspended in lysis buffer (5 ml/sample. The resuspended bacteria were incubated on ice for 15 minutes, until the solution became more viscous. The samples were sonicated (30% power) for 30 seconds, 1-minute pulse while keeping the samples on ice, and repeat the sonication 3 times for each sample. The samples were centrifuged at 4,000 rpm for 30 minutes at 4°C. From each sample, 5 μl were taken as total cell lysate (TCL), and 1 μl 6X SDS loading buffer was added to each TCL sample prior to gel electrophoresis.

2.12.2 Preparation of GST-protein beads

A bed volume of Glutathione sepharose 4B slurry (300μl) was washed with GST purification buffer in a chromatography column. Once the buffer passed through the column, the

sonicated bacterial cell samples were applied in respective columns and were allowed to flow through at 4°C. When the cell lysate went through, the columns were washed twice with 5 ml GST purification buffer. The beads were collected in 1 ml 1X PBS, spun down at 2,000 rpm for 2 minutes at 4°C. The supernatant was removed and 300 μl 1X PBS + 0.02% sodium azide were added and stored at 4°C.

2.12.3 Polyacrylamide gel electrophoresis of GST-fusion proteins

A bed volume of 10 μl of bed volume from each sample was used to electrophorese through a 10% polyacrylamide gel along with the TCL samples and bovine serum albumin (BSA) standards (1 μg, 5 μg and 20 μg). The gel was then treated with Coomassie brilliant blue solution (20% (v/v) acetic acid, 0.025% (w/v) Coomassie blue G-250) for 30 minutes at room temperature, followed by treatment with de-staining solution (40% (v/v) methanol, 10% (v/v) acetic acid) for 30 minutes.

From the Coomassie stained gel the GST-protein yield could be estimated.

2.13 GST protein pull-down

Buffers

Protein lysis buffer:

	Volume
1 M Hepes pH 7.5	10 ml
5 M NaCl	24 ml
1 M $MgCl_2$	400 μl
0.5 M EGTA	0.8 ml
1 M β-glycerol phosphate	8 ml
0.5 M NaF	8 ml
0.1 M Sodium vanadate	4 ml
Glycerol	2 ml
20% Triton-X	10 ml
1X PBS	334 ml

Before use: add fresh DTT (final conc: 1M), protease inhibitor (final con: 1X), aprotinin (final conc: 1X) and PMSF (final conc: 0.1mM).

2.13.1 Cell culture and harvesting

NMuMG cells were plated on 6X Ø15 cm plates, incubated in a 37°C, 5% CO_2 incubator until the cells reached 80% confluency. At that point the cells were trypsinized and collected in 2 ml complete media (per plate) and pooled in a 15 ml falcon tube. The tube was centrifuged at 1,000 rpm for 3 minutes. The supernatant was removed and the cell pellet

washed with 1X PBS, 5 ml, centrifuged again at 1,000 rpm for 3 minutes. The supernatant was removed and 3 ml protein lysis buffer (=500 μl/sample) was added to the cell pellet. The pellet was resuspended in the buffer and incubated on ice for 10 minutes and passed through a 20G and a 27G syringe 5-6 times. The lysate was centrifuged at 14,000 rpm for 10 minutes at 4°C. From the supernatant 50 μl lysate was collected and mixed with 10 μl 6X SDS loading buffer, boiled for 5minutes and stored at -20°C; this would later be used as TCL when electrophoresing the samples on a polyacrylamide gel.

2.13.2 GST-fusion bead incubation with mammalian cell lysate

A bed volume corresponding to 10 μg total GST-fusion protein was collected from each GST-fusion construct sample. These beads were washed twice by adding 500 μl GST purification buffer and centrifuging them at 2000 rpm for 2 minutes at 4°C. The cell lysate was the divided equally into the tubes with GST-constructs (~ 490 μl/sample). The GST-constructs were incubated with the cell lysate in a rotor at 4°C for 2 hours. The tubes were then centrifuged at 2,000 rpm for 2 minutes at 4°C. Supernatant was removed and the beads were washed with 1 ml GST purification buffer, centrifuged, 800 μl supernatant was removed and the beads were resuspended and transferred to new tubes and washed two more times. Then, the supernatant was removed thoroughly with a 30G syringe. The beads were then resuspended in 40μl 1X SDS loading buffer and boiled for 5 minutes.

The samples were electrophoresed on a 10% polyacrylamide gel together with the TCL, then transferred to a nitrocellulose membrane, which was blotted for CTCF (1:1000, anti-CTCF rabbit antibody, Millipore) (Fig. 10B).

3. Results and Discussion

3.1 TGFβ induces EMT in NMuMG cells

When stimulating NMuMG cells with TGFβ (5ng/ml) for 0, 24 and 48 hours E-cadherin expression decreased at both the mRNA and protein levels (Fig. 5A), these results are expected and consistent with previous studies (Vogelmann *et al.* 2005, Schumpelick *et al.* 2009 and Valcourt *et al.* 2005). The EMT TF Snail1 increased upon TGFβ stimulation and seems to peak around 24 hours of stimulation at both the mRNA and protein levels (Fig. 5A). When looking at the mRNA and protein expression level for the mesenchymal marker fibronectin, we can see that it showed a strong induction in response to TGFβ (Fig. 5A). When looking through a phase contrast microscope, the cells gradually lost their aggregated, cobblestone morphology and became more sparse and elongated upon TGFβ stimulation (Fig. 5A). Altogether, these data are consistent with previous studies showing that TGFβ induces EMT in NMuMG cells (Choi *et al.* 2007 and Valcourt *et al.* 2005).

3.2 TGFβ induced EMT is reversible in NMuMG cells

When stimulating NMuMG cells with TGFβ for longer time points (0, 3, 5 and 21 days), the same trend is observed as before (Fig. 5B). Epithelial proteins are down regulated and mesenchymal proteins are upregulated in response to TGFβ. An exciting result is observed when TGFβ is present for 21 days and then removed for 14 days (WD (withdrawal) sample). Here we can see that the mesenchymal features (strong expression of fibronectin and N-cadherin) that the cells feature after 21 days of TGFβ stimulation disappears after removal of TGFβ, and the epithelial expression pattern returns (E-cadherin expression is upregulated) (Fig. 5B). This is an example of mesenchymal to epithelial transition (MET) and shows that EMT/MET is a very dynamic system, a very important ability for tumors that spreads to new tissues through EMT and needs to establish secondary tumors using MET. This result further indicates that although cells undergo EMT, and completely loses their epithelial morphology, they can rescue that phenotype if the microenvironment changes. These results nicely recapitulate experiments performed in our laboratory using NMuMG cell cultures with chronic treatment with TGFβ (Gal *et al.* 2008).

3.3 TGFβ represses *Cdh1* promoter activity

In order to see how the *Cdh1* promoter activity was affected by increasing TGFβ stimulation a luciferase reporter assay was performed in NMuMG cells (Fig. 6B). As seen in figure 6B, the promoter activity is gradually reduced by TGFβ stimulation. This shows that the signal cascade initiated by TGFβ, resulted in repression of the E-cadherin promoter, which we know from earlier work to be correct (Valcourt *et al.* 2005) and which nicely correlates with the downregulation of *E-cadherin* mRNA and protein levels observed at the endogenous level (Fig. 5A).

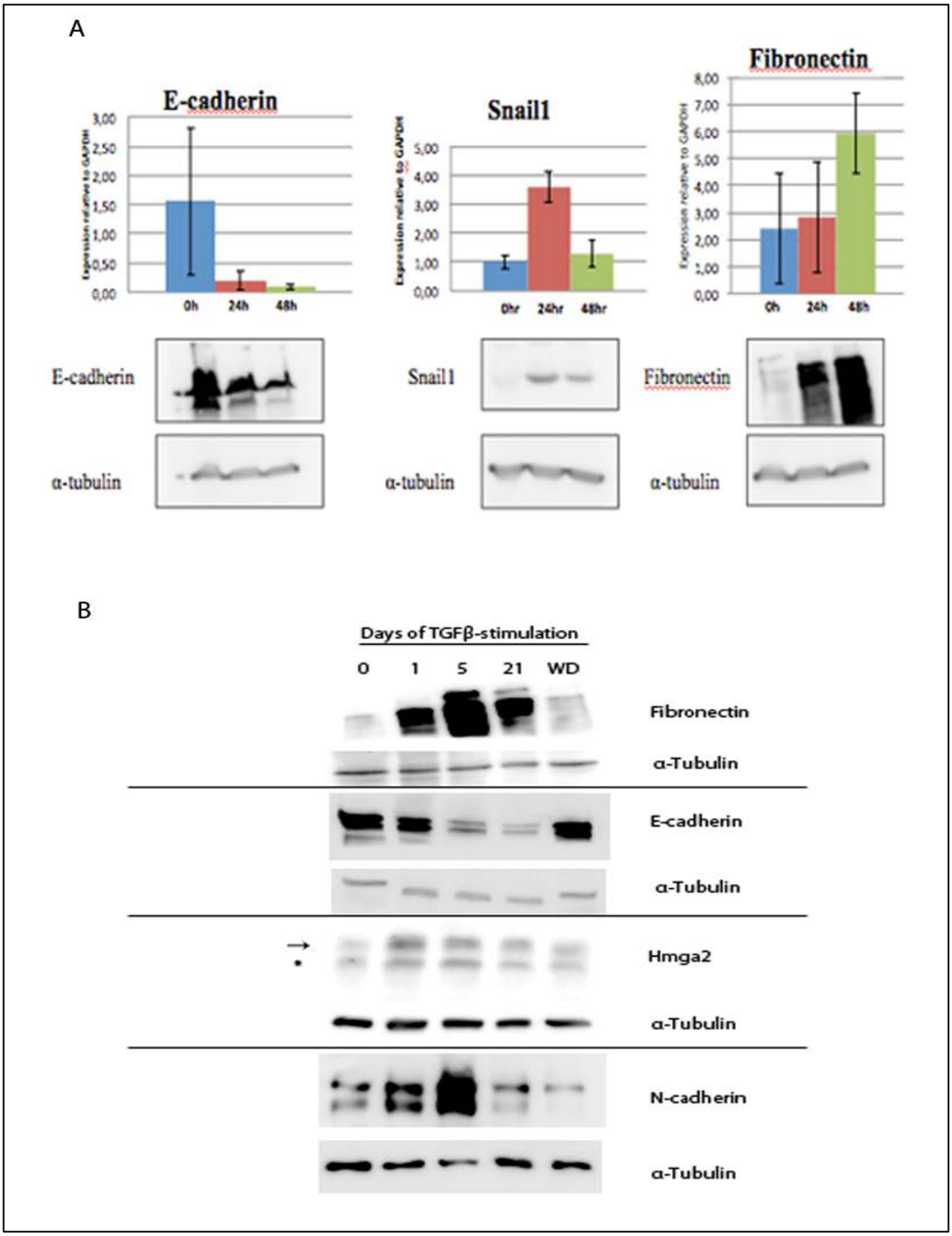

Figure 5. ***A****, Real-time RT-PCR analysis of E-cadherin, Snail1 and fibronectin mRNA (top graphs) expression after stimulation of NMuMG cells with TGFβ1 for the indicated time points. Expression of each mRNA is expressed relative to the housekeeping gene Gapdh. Average determinations from triplicate samples with corresponding error bars are plotted. Corresponding immunoblots of the same proteins in the same time course experiment are shown below and α-tubulin is used as loading control.* ***B****, Immunoblot analysis of the indicated proteins during a time course experiment of TGFβ stimulation and withdrawal (WD) in NMuMG cells. An arrow shows the specific Hmga2 protein band and a star indicates a non-specific protein band.*

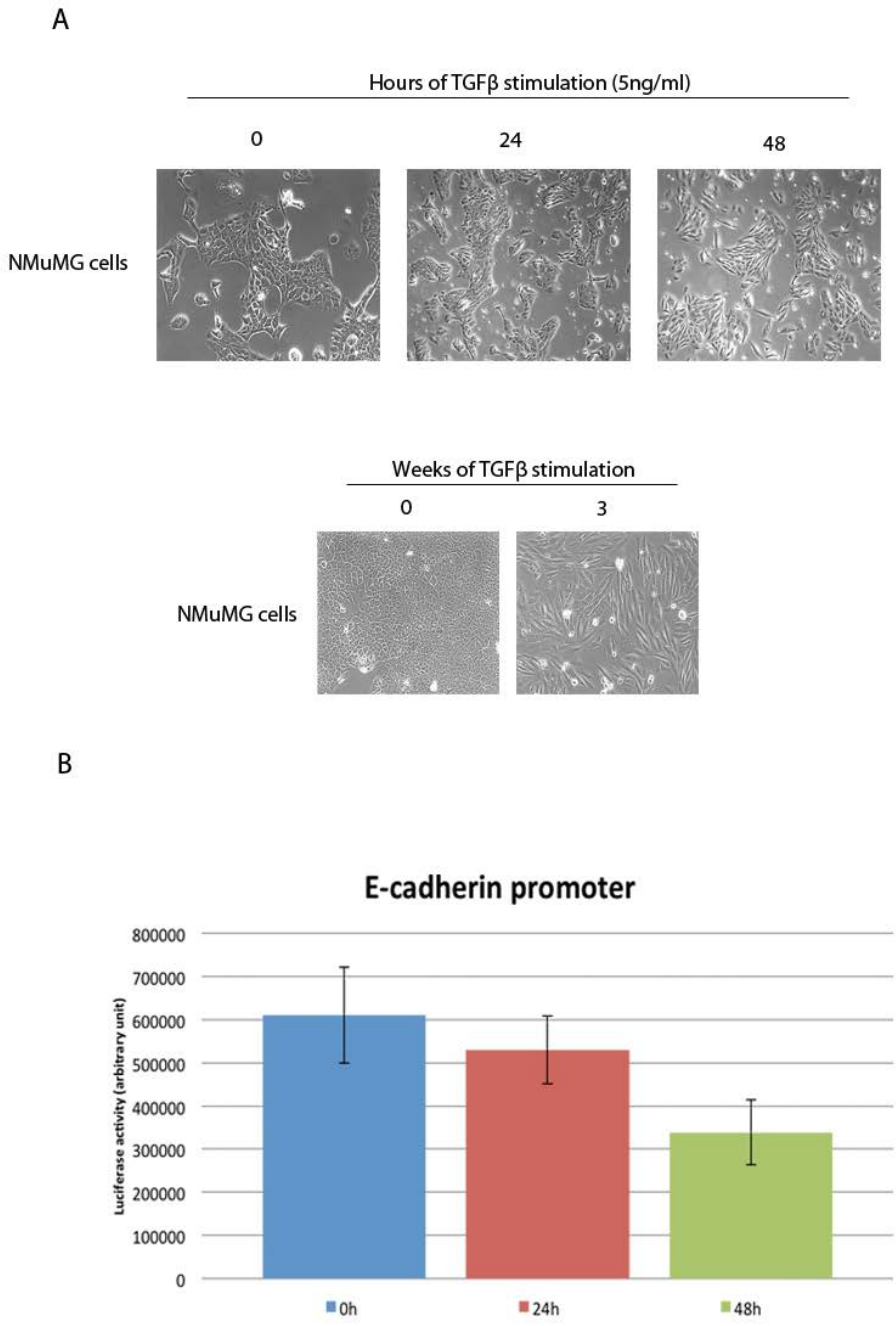

Figure 6. ***A****, Phase contrast microscopy of NMuMG cells stimulated with TGFβ1 as indicated; characteristic epithelial (tight cobblestoned) or mesenchymal (spindle-like elongated) morphology is obvious.* ***B****, E-cadherin promoter-luciferase activity graph from transfected NmuMG cells after stimulation with TGFβ1 for the indicated time points. The dat are expressed as reltove luciferase activity units normalized to corresponding β-galactosidase units. Average determinations from triplicate samples with corresponding error bars are plotted.*

3.4 Hmga2 plays an essential role in TGFβ induced EMT

Earlier work from Thuault *et al.* (2006) has shown that Hmga2 is important for the TGFβ induced EMT response. This made us decide to focus on this factor to test whether it is involved in the repression of the *Cdh1* promoter region.

By analyzing the E-cadherin expression in Hmga2 overexpressing NMuMG clones we could see that a fully functional Hmga2 expression is needed to successfully repress E-cadherin expression (Fig. 7A). The wild type Hmga2 overexpressing clone had an efficient repression of the E-cadherin expression compared to mock cells, both at protein level and when looking at immunofluorescence images. But when a mutated Hmga2 (Hmga2 with a deleted C-terminal) was overexpressed, E-cadherin expression could not be reduced, neither was fibronectin expression induced.

Furthermore, the wild type Hmga2 overexpressing clone was the only clone that was able to induce Snail1 expression at mRNA level (Fig. 7B). Most probably it is the lack of Snail1 induction in the mutated Hmga2 clone that explains its inefficient E-cadherin repression and fibronectin induction.

Although the induction of Snail1 (and other EMT TFs) by wild type Hmga2 is the major factors involved in E-cadherin repression, Tan *et al.* (2012) showed that E-cadherin expression could not be rescued in Hmga2 overexpression clones although Snail1 and Twist were knocked down with short hairpin RNAs (shRNA) (Fig. 8). This result was a bit surprising because after knock down of the *Cdh1* repressors, E-cadherin expression should have returned. This led to the hypothesis that Hmga2 might directly repress E-cadherin. Since hyper methylation is a common feature in cancer cells, and the E-cadherin promoter has been shown to be heavily methylated in cancer cells, it was hypothesized that perhaps Hmga2 is somehow involved in this methylation.

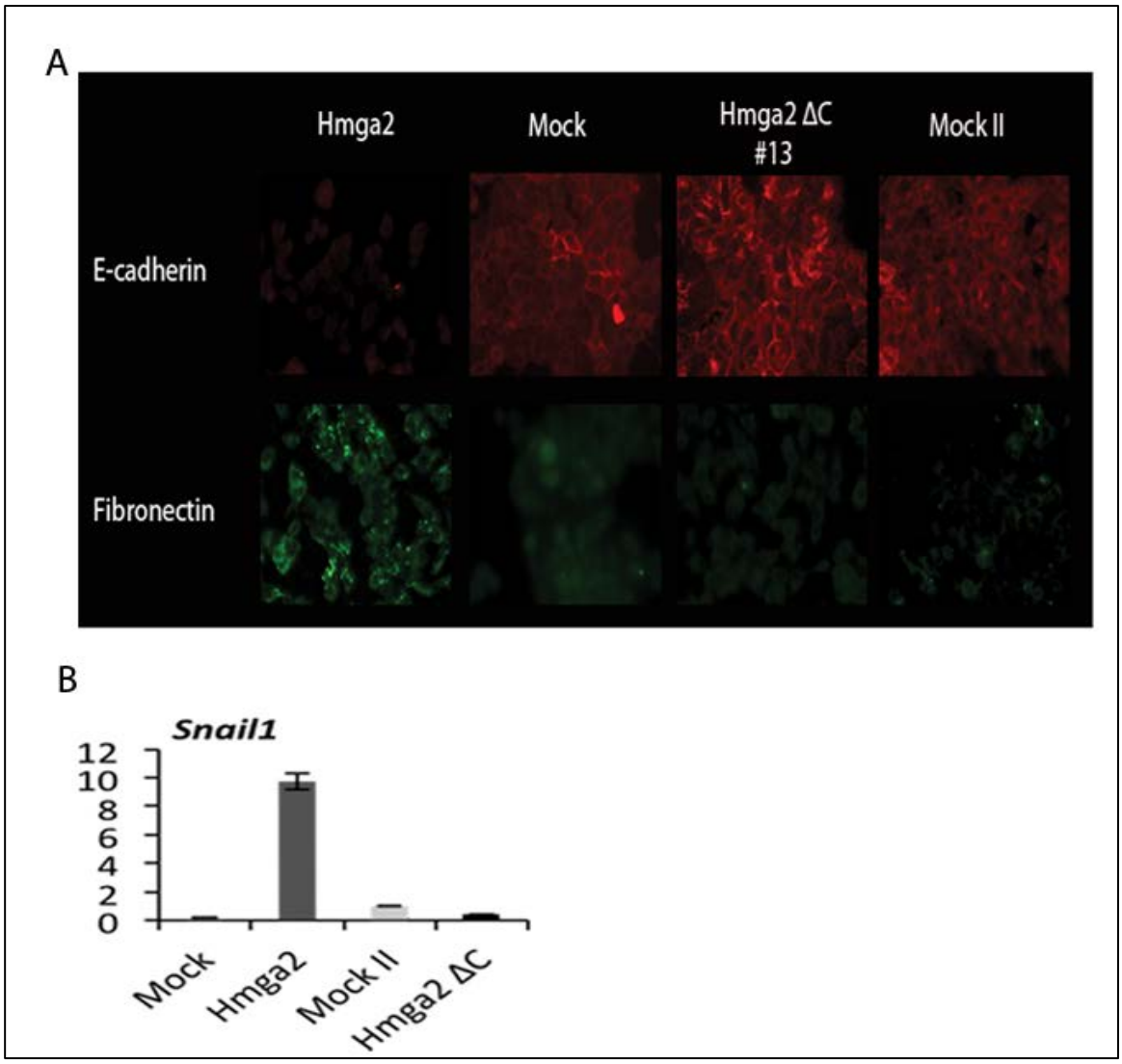

Figure 7. ***A****, Hmga2 (wild type and dC mutant) transfected NMuMG clones along with the mock transfected NMuMG clones. Immunofluorescent staining for either E-cadherin or fibronectin.* ***B****, Snail1 mRNA expression in the corresponding NMuMG clones from fig.4A (Work done by E-Jean Tan).*

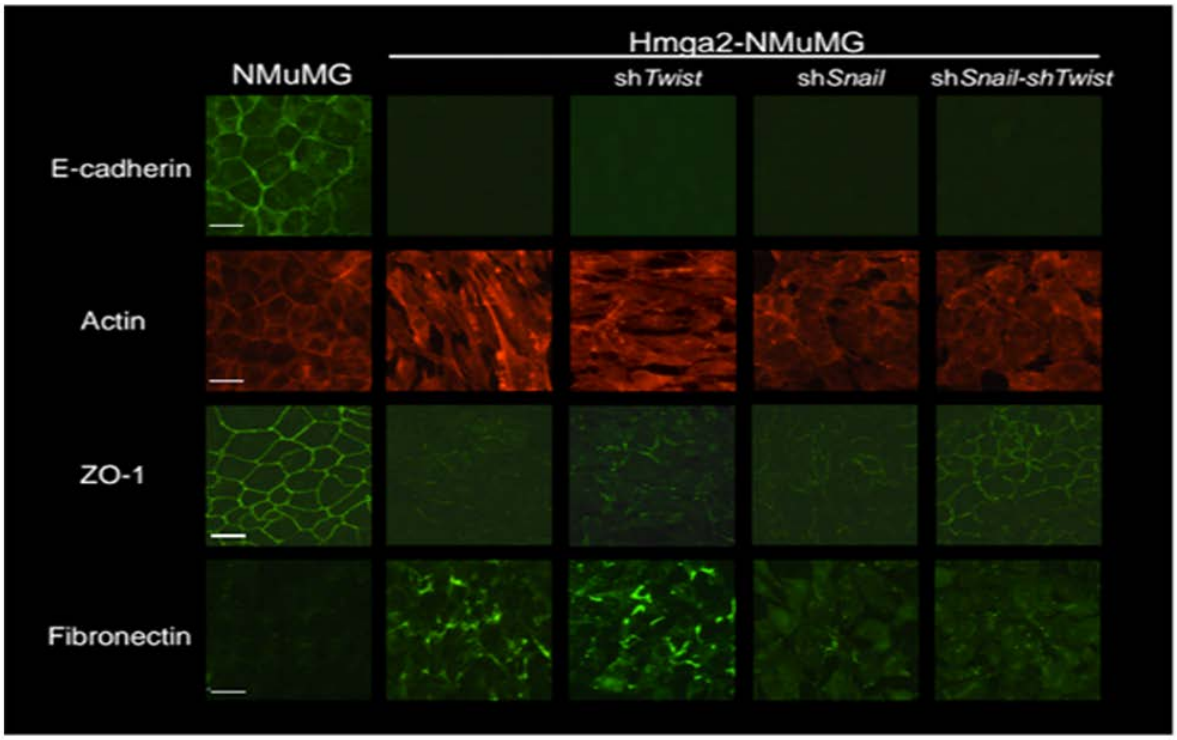

Figure 8. Hmga2 overexpressing NMuMG clones with Twist, Snail1, Twist and Snail1 or neither transcription factor silenced, untransfected NMuMG clone to the left. Immunofluorescent staining for E-cadherin, Actin, ZO-1 and Fibronectin (Work done by E-Jean Tan).

3.5 DNA methylation pattern at the Cdh1 promoter region in NMuMG clones overexpressing Hmga2

To test whether the *Cdh1* promoter was methylated in the Hmga2 clones (Mock, Hmga2, Hmga2+shSnail, Hmga2+shSnail+shTwist#49 and Hmga2+shSnail+shTwist#50) a DNA methylation analysis was done, using a methylation sensitive restriction enzyme (MspI/HpaII) (work done by my supervisor, E-Jean Tan) (Fig. 9A). This analysis proved that the *Cdh1* promoters in all of the Hmga2 overexpressing clones were methylated, even when Snail1 and Twist were knocked down by shRNA. To proceed and find out to what degree the promoters were methylated and if there was any differences among the clones, we followed up the restriction analysis with bisulfite sequencing.

The bisulfite treatment essentially converts unmethylated cytosine to uracil without altering the methylated cytosine. The uracils will be replaced by guanines after PCR amplification, which makes it possible to reveal the methylation pattern by sequencing and comparing the bisulfite treated DNA sequenced with the original DNA sequence (Bock *et al.* 2005). This method enabled us to look at approximately 416 bp of the methylation pattern at a single-nucleotide resolution. In our case we looked at the promoter region of the *Cdh1* gene. The sequence alignment was done using BiQ Analyzer software. In figure 9B we can see the

results. It is very clear that the *Cdh1* promoter region is methylated in the Hmga2 clones, to a much higher degree compared to mock-transfected clones. When comparing among the Hmga2 overexpressing clones that had control silencing or specific silencing of Snail1 or Twist, no significant difference could be found, indicating that all these cell clones had equally high methylation of their *Cdh1* promoters. This methylation pattern nicely correlates with the absence of expression of E-cadherin in all the NMuMG-Hmag2 clones analyzed in this study.

This result shows that Hmga2 is involved in methylation of the *E-cadherin* gene, which keeps it repressed in a long-term perspective, even when Snail1 and Twist are knocked down. The methylation seems to be independent of Snail1 and Twist, since no difference in degree of methylation can be seen between the Hmga2 overexpressing clones and after silencing Snail1 and Twist. The exact molecular machinery behind this promoter methylation still remains unclear.

Since the DNA methylation is carried out by DNMTs, it is reasonable to think that Hmga2 is involved in the recruitment of the right DNMT member, or perhaps that Hmga2 enables the DNMT to bind to the DNA by bending the DNA.

3.6 Involvement of DNMTs in EMT

The DNMT family members are known to be involved in methylation of DNA (Kumar *et al.* 1994), and methylation clearly plays an essential role in the epigenetic silencing of E-cadherin in human cancer. In order to explore the involvement of DNMTs in the general TGFβ induced EMT response, a series of knockdown experiments were done. Dnmt1 and Dnmt3a were knocked down separately in NMuMG cells and the effect this knockdown had on EMT was analyzed by measuring fibronectin and E-cadherin protein levels in response to TGFβ (5ng/ml, 48h) was analyzed (Fig. 10A). Since the cells were stimulated with TGFβ for 48 hours, this experiment only measures the fast EMT response to TGFβ and not the slower silencing effects that has been associated with DNMT function in other studies (Oda *et al.* 2006).

The result shows that E-cadherin expression is still reduced in response to TGFβ in the same manner as in the control, which indicates that neither Dnmt1 or Dnmt3a are involved in the fast TGFβ induced repression of E-cadherin. As for fibronectin, the induction is not quite as strong upon TGFβ stimulation when Dnmt3a is knocked down. This result indicates that Dnmt3a is involved in the immediate induction of fibronectin in response to TGFβ. The way this would work at a molecular level still remains unclear. However, we can firmly conclude that during the onset of EMT, TGFβ utilizes Hmga2, Snail1 and Twist to downregulate E-cadherin and induce fibronectin. Furthermore, we can see that the DNMT function and corresponding methylation of gene sequences may start much later and during prolonged stimulation with agents such as TGFβ.

3.7 Hmga2 interacts with CTCF

CTCF is an eleven-zinc finger protein that was discovered by Lobanenkov *et al.* 1990. It can form protective bounadries around certain DNA regions, often tumor suppressor genes, blocking repression of these genes. At the same time, it has been shown to block enhancer activity in vertebrates (Bell *et al.* 1999). CTCFs ability to block both induction and repression of transcription makes it to one of the few genomic insulators known today. CTCF plays an important role in developmental processes, where it helps orchestrating the delicate interplay of onset and turn off between genes (Herold *et al.* 2012).

It has been shown that CTCF binding to *Cdh1* protects it from repression, which keeps the gene active (Witcher *et al.* 2010). When DNA is methylated, CTCF loses affinity (Filippova *et al.* 2001 and Hark *et al.* 2000). This fact affects *Cdh1* activity negatively.

Since this report has implicated Hmga2 to be involved in the methylation of *Cdh1* promoter, a GST-pulldown experiment was done with Hmga2 constructs to analyze the interaction between Hmga2 and CTCF (Fig. 10B). The constructs in this GST-pulldown were used in order to explore if CTCF and Hmga2 interact, and if they do, to determine which part(s) of Hmga2 interact with CTCF (Fig. 10C). As we can see in figure 10B, Hmga2 full length (FL) and CTCF are indeed interacting, but the strongest interaction is observed in the Hmga2 constructs lacking the C-terminal region and containing at least one AT-hook (dC and N1). This result indicates that the C-terminal region somehow disrupts the Hmga2-CTCF interaction and that the Hmga2 constructs lacking this part provides a better binding epitope for CTCF. In addition, when comparing the constructs N1 and N2 it seems like having at least one AT-hook is needed for a successful interaction (Fig. 10B).

Although these results indicate that Hmga2 interacts with CTCF and this interaction could in theory be responsible for interfering with CTCFs protective boundary around *Cdh1*, more studies are needed to prove this model. Another possible theory is that Hmga2 recruits the epigenetic machinery to *Cdh1* and the methylation of the promoter region creates an unfavorable CTCF binding environment.

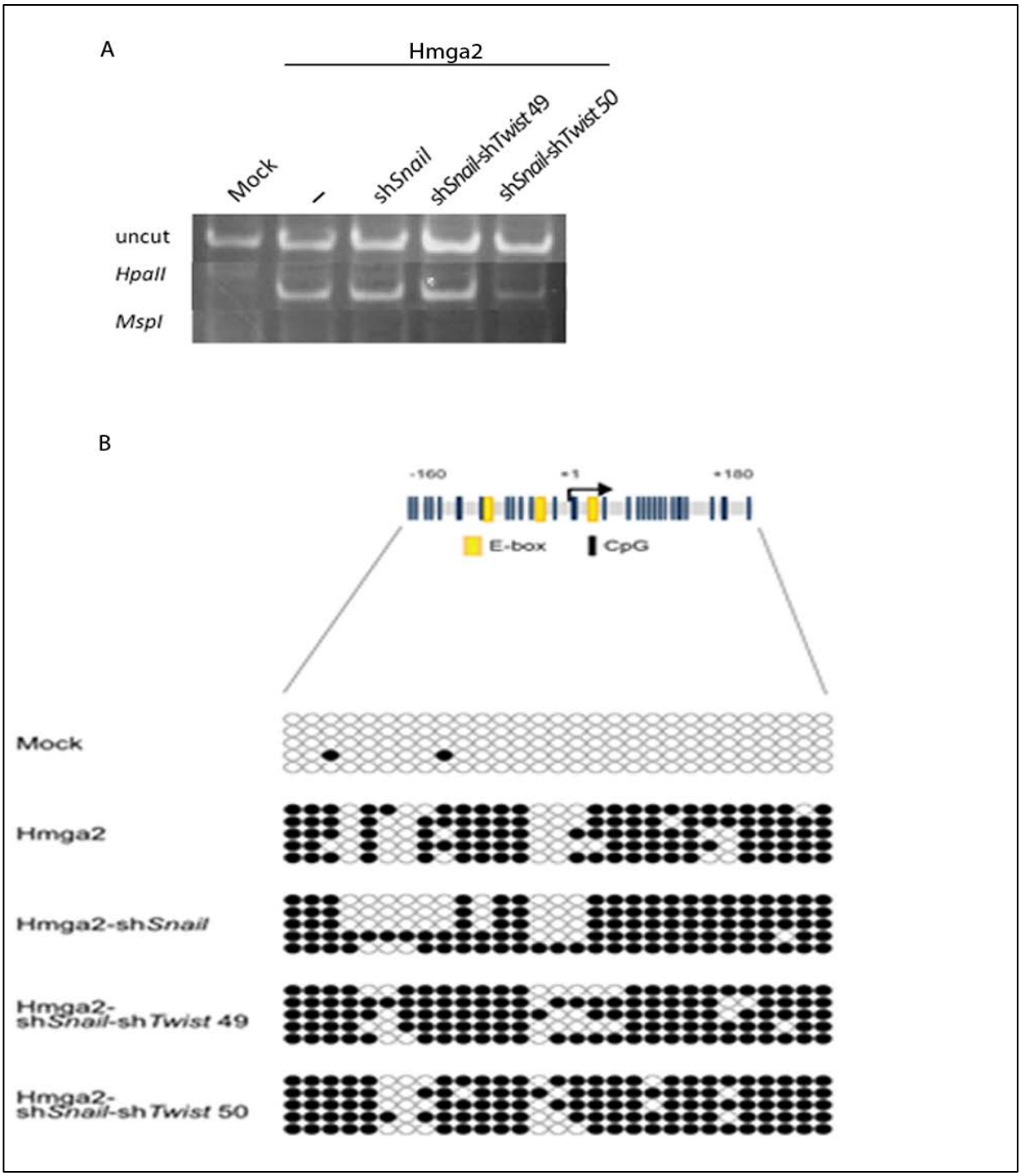

Figure 9. ***A****, Methylation analysis of the Cdh1 promoter in Hmga2 overexpressing NMuMG clones.* ***B****, Pattern of the methylated CpG islands in proximity of the Cdh1 promoter in Hmga2 overexpressing NMuMG clones.*

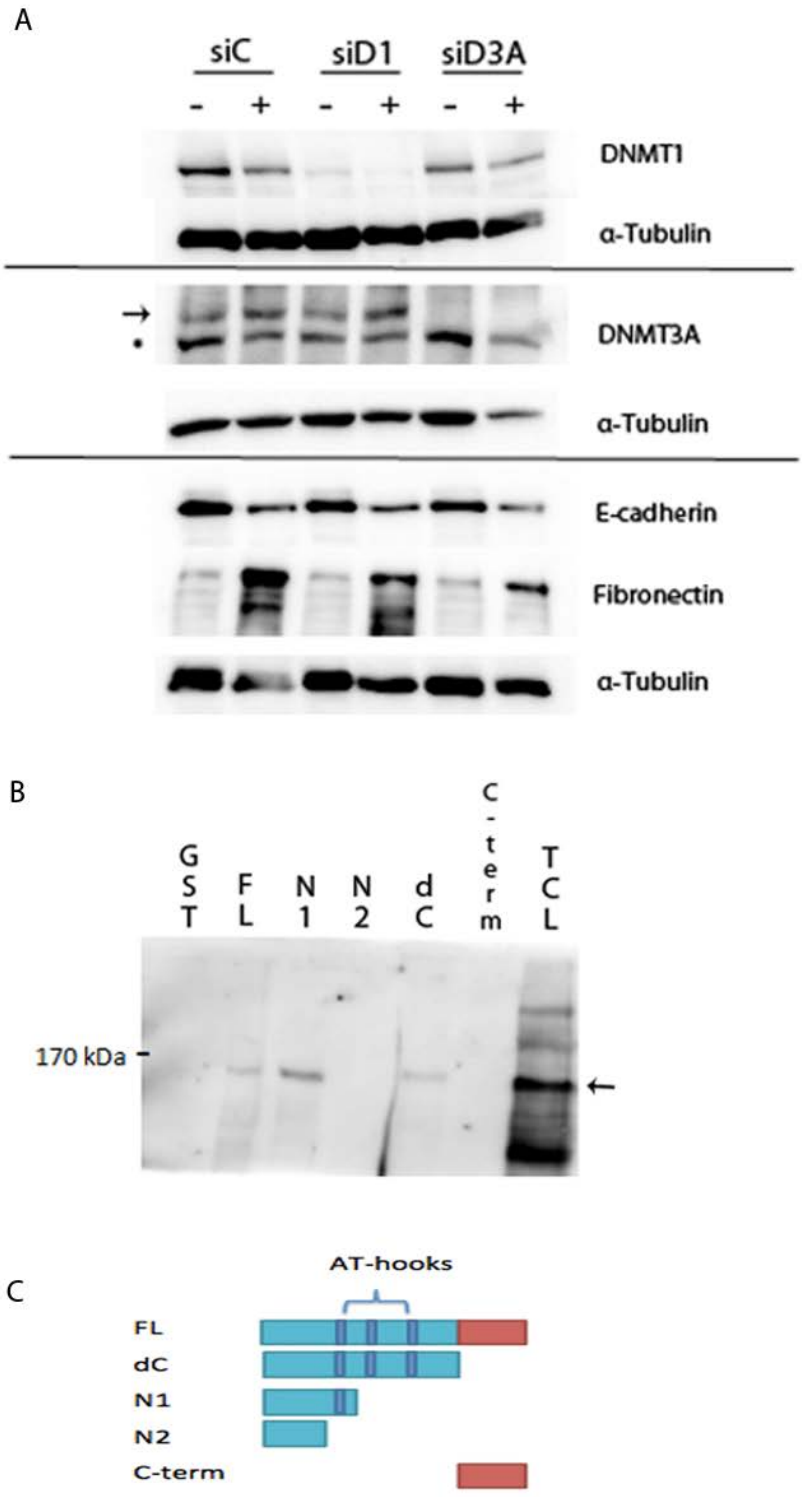

Figure 10. ***A****, Immunoblot analysis of the indicated proteins upon silencing of DNMT1, DNMT3A or Control with or without TGFβ stimulation (48 hours). An arrow shows the specific DNMT3A protein band and a star indicates a non-specific protein band.* ***B****, Immunoblot analysis of GST-pulldown with Hmga2 constructs, blotting for CTCF (fig. 11B). An arrow shows the specific CTCF band.* ***C****, Schematic figure of the Hmga2 constructs.*

3.8 Conclusions and future perspectives

The focus of this project has been to further assess the effects of TGFβ stimulation on the EMT response in NMuMG cells. By looking at the typical markers for both epithelial and mesenchymal cell phenotypes, and how the mRNA and protein expression of these markers alters during the presence, absence and withdrawal of TGFβ, the extraordinary plasticity of the EMT/MET response has been further characterized.

In addition to the analysis of the EMT response in NMuMG cells, the epigenetic regulation of the epithelial marker protein E-cadherin has been evaluated. It has been shown that extensive methylation of CpG islands in proximity to the promoter region of the E-cadherin *Cdh1* is repressing the E-cadherin expression in Hmga2 overexpressing NMuMG clones, even when its transcriptional repressors (Snail1 and Twist) are absent. These results, together with the GST-pulldown interaction studies, indicate that Hmga2 is involved in the methylation of *Cdh1*, although the exact mechanism still remains unclear.

It would be intersesting to further evaluate if Hmga2 is somehow involved in the recruitment of DNMTs, which are responsible for DNA methylation. If it would be proved that Hmga2 is directly involved in the DNMT recruitment, it would be obvious that Hmga2 is not only involved but also essential for the methylation and silencing of the *Cdh1* promoter. In addition it would be interesting to see if Hmga2 is involved in the silencing of any other epithelial genes in a similar fashion as in the case of the E-cadherin gene.

4. Acknowledgments

This master project was carried out at the Uppsala branch of Ludwig institute for cancer research.

First and foremost I would like to thank E-Jean Tan for the excellent supervision during this project. Thank you for learning me the tricks, supporting me and letting me succeed and fail at my own pace. Thank you for always take time for my questions, I really appreciate that.

Secondly, a big thanks to Aris for allowing me to work under his supervision in the lab, and for his support throughout the project. You have given me important feedback and showed me how to approach problems in a scientific way. I am also impressed of how you lead the lab in a friendly and pedagogical way, a great role model for any group leader.

I would also like to thank everybody else at the Ludwig institute for creating such a friendly environment to work in and for helping me in the everyday work. It has been a real pleasure to meet and work with you all, and I hope we will stay in touch and meet again in the future.

5. References

Bell, A. C., A. G. West, et al. (1999). "The Protein CTCF Is Required for the Enhancer Blocking Activity of Vertebrate Insulators." Cell **98**(3): 387-396.

Bird, A. P. (1986). "CpG-rich islands and the function of DNA methylation." Nature **321**(6067): 209-213.

Bock, C., S. Reither, T. Mikeska, M. Paulsen, J. Walter and T. Lengauer (2005). "BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing." Bioinformatics 21(21): 4067-8.

Choi, J., S. Y. Park, et al. (2007). "Transforming Growth Factor-Œ≤1 Represses E-Cadherin Production via Slug Expression in Lens Epithelial Cells." Investigative Ophthalmology & Visual Science **48**(6): 2708-2718.

de Martel, C., J. Ferlay, et al. (2012). "Global burden of cancers attributable to infections in 2008: a review and synthetic analysis." The Lancet Oncology **13**(6): 607-615.

Feinberg, A. P. (1983). "Hypomethylation distinguishes genes of some human cancers from their normal counterparts." Nature **301**(5895): 89-92.

Filippova, G. N., C. P. Thienes, et al. (2001). "CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus." Nature genetics **28**(4): 335-343.

Fusco, A. (2007). "Roles of HMGA proteins in cancer." Nature reviews cancer **7**(12): 899-910.

Guarino, M., B. Rubino, et al. (2007). "The role of epithelial-mesenchymal transition in cancer pathology." Pathology **39**(3): 305-318.

Guarino, M., B. Rubino, et al. (2007). "The role of epithelial-mesenchymal transition in cancer pathology." Pathology **39**(3): 305-318.

Görisch, S. M., M. Wachsmuth, et al. (2005). "Histone acetylation increases chromatin accessibility." Journal of Cell Science **118**(24): 5825-5834.

Hark, A. T., C. J. Schoenherr, et al. (2000). "CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus." Nature **405**(6785): 486-489.

Heldin, C.-H., M. Vanlandewijck, et al. (2012). "Regulation of EMT by TGFŒ≤ in cancer." FEBS Letters **586**(14): 1959-1970.

Herold, M., M. Bartkuhn, et al. "CTCF: insights into insulator function during development." Development **139**(6): 1045-1057.

Iwatsuki, M., K. Mimori, et al. (2010). "Epithelial–mesenchymal transition in cancer development and its clinical significance." Cancer Science **101**(2): 293-299.

Mauviel, D. J. a. A. (2004). "Mammalian Transforming growth factor betas: Smad signaling and physio-pathological roles." The international journal of biochemistry & cell biology **36**(7): 1161-1165.

Mazzio, E. A. and K. F. A. Soliman (2012). "Basic concepts of epigenetics: Impact of environmental signals on gene expression." Epigenetics **7**(2): 119-130.

Miettinen, P. J., R. Ebner, et al. (1994). "TGF-beta induced transdifferentiation of mammary epithelial cells to mesenchymal cells: involvement of type I receptors." The Journal of Cell Biology **127**(6): 2021-2036.

Nieto, M. A. (2002). "The snail superfamily of zinc-finger transcription factors." Nat Rev Mol Cell Biol **3**(3): 155-166.

Oda, M., A. Yamagiwa, et al. (2006). "DNA methylation regulates long-range gene silencing of an X-linked homeobox gene cluster in a lineage-specific manner." Genes & Development **20**(24): 3382-3394.

Onder, T. T., P. B. Gupta, et al. (2008). "Loss of E-Cadherin Promotes Metastasis via Multiple Downstream Transcriptional Pathways." Cancer Research **68**(10): 3645-3654.

Piek, E., A. Moustakas, et al. (1999). "TGF-(beta) type I receptor/ALK-5 and Smad proteins mediate epithelial to mesenchymal transdifferentiation in NMuMG breast epithelial cells." Journal of Cell Science **112**(24): 4557-4568.

S Kumar, X. C., S Klimasauskas, S Mi, J Posfai, R J Roberts, and G G Wilson (1994). "The DNA (cytosine-5) methyltransferases." Nucleic acid research **22**(1): 1-10.

Schumpelick, V., H. P. Bruch, et al. (2009). Epithelial-Mesenchymal-Transition (EMT): The role of E-cadherin transcription regulators SIP1, Twist and Snail in colorectal adenomas. Chirurgisches Forum und DGAV Forum 2009, Springer Berlin Heidelberg. **38:** 245-246.

Stewart, M. D., J. Li, et al. (2005). "Relationship between Histone H3 Lysine 9 Methylation, Transcription Repression, and Heterochromatin Protein 1 Recruitment." Molecular and cell biology **25**(7): 2525-2538.

Suvá, M. L., N. Riggi, et al. (2013). "Epigenetic Reprogramming in Cancer." Science **339**(6127): 1567-1570.

Tan, E.-J., S. Thuault, et al. (2012). "Regulation of Transcription Factor Twist Expression by the DNA Architectural Protein High Mobility Group A2 during Epithelial-to-Mesenchymal Transition." Journal of Biological Chemistry **287**(10): 7134-7145.

Ten Dijke, P., M.-J. Goumans, et al. (2002). "Regulation of cell proliferation by Smad proteins." Journal of Cellular Physiology **191**(1): 1-16.

Thiery, J. P., H. Acloque, et al. (2009). "Epithelial-Mesenchymal Transitions in Development and Disease." Cell **139**(5): 871-890.

Thuault, S., E.-J. Tan, et al. (2008). "HMGA2 and Smads Co-regulate SNAIL1 Expression during Induction of Epithelial-to-Mesenchymal Transition." Journal of Biological Chemistry **283**(48): 33437-33446.

Thuault, S., U. Valcourt, et al. (2006). "Transforming growth factor-β employs HMGA2 to elicit epithelial‚ÄìÄìmesenchymal transition." The Journal of Cell Biology **174(**2): 175-183.

V V Lobanenkov, R. H. N., V V Adler, H Paterson, E M Klenova, A V Polotskaja, G H Goodwin (1990). "A novel sequence-specific DNA binding protein which interacts

with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene." Oncogene **5**(12): 1743-1753.

Valcourt, U., M. Kowanetz, et al. (2005). "TGF-β and the Smad Signaling Pathway Support Transcriptomic Reprogramming during Epithelial-Mesenchymal Cell Transition." Molecular Biology of the Cell **16**(4): 1987-2002.

Vogelmann, R., M.-D. Nguyen-tat, et al. (2005). "TGFŒ≤-induced downregulation of E-cadherin-based cell-cell adhesion depends on PI3-kinase and PTEN." Journal of Cell Science **118**(20): 4901-4912.

Witcher, M. and M. B, Emerson (2010). "Epigenetic Silencing of the p16INK4a Tumor Suppressor is Associated with Loss of CTCF Binding and a Chromatin Boundary." Molecular cell **34**(3): 271-284.

Yoshiura, K., Y. Kanai, et al. (1995). "Silencing of the E-cadherin invasion-suppressor gene by CpG methylation in human carcinomas." Proceedings of the National Academy of Sciences **92**(16): 7416-7419.

6. Supplementary data

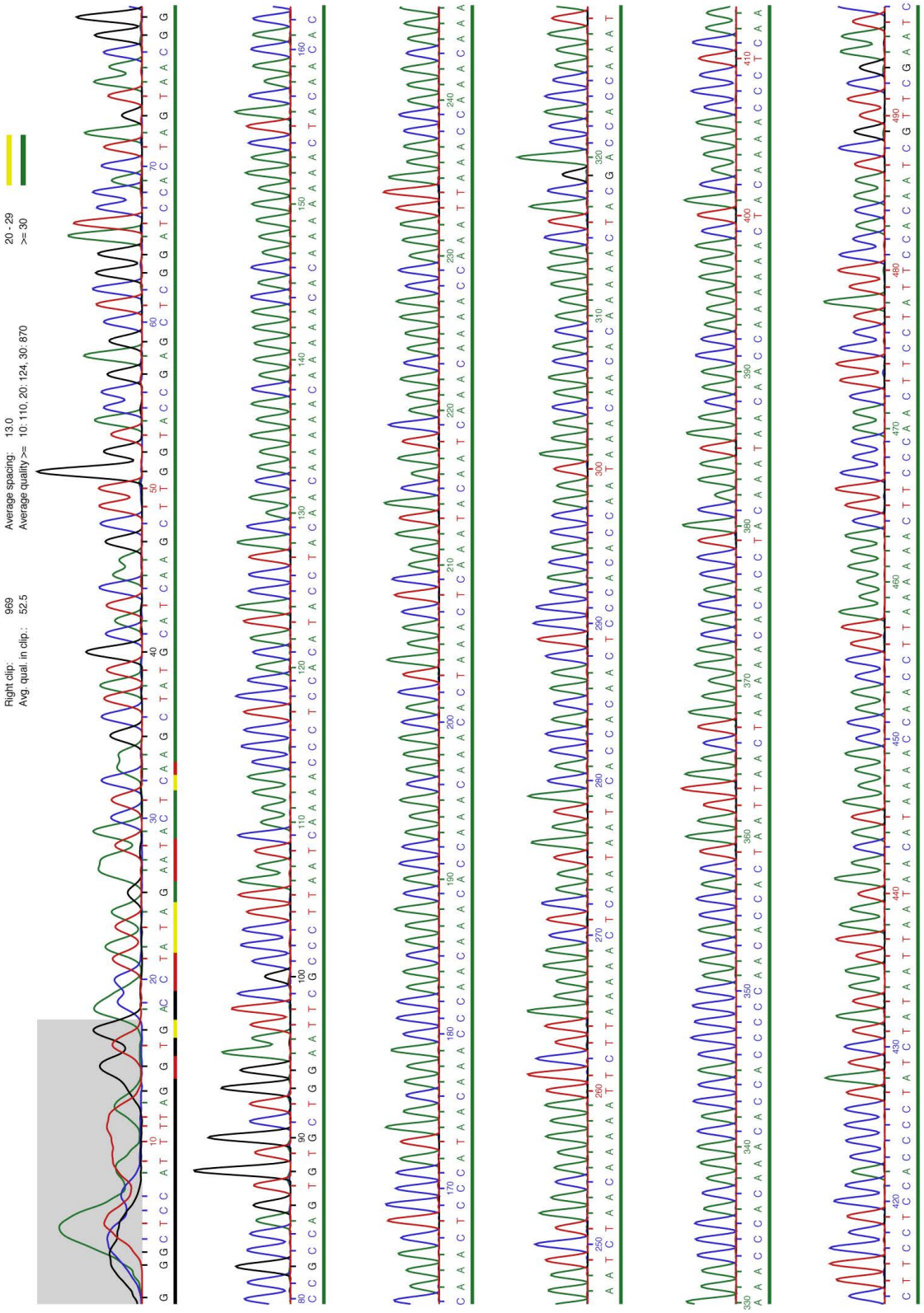

Figure 11 Raw sequence data from the Hmga2 shSnail1 (3) clone (lane 3 in figure 3B). Sequence underlined with green is unambiguous, as proven by the equel height of each peak in the chromatogram of the sequencing instrument. Sequence shown in this figure is not the complete analyzed sequence.