Thesis for the Master's degree in Biology

Main field of study in toxicology

A mechanistic study of functional polymorphisms in the *MDM2* gene

Monica Hongrø Solbakken

Department of Biology

Faculty of mathematics and natural sciences

UNIVERSITY OF OSLO 12/2008

Acknowledgements

This project was carried out at the National Institute of Occupational Health and the master degree obtained at the University of Oslo. Supervisors were Dr.philos Shan Zienolddiny and Professor Dr.philos Steinar Øvrebø.

First of all I would like to thank my supervisors, Steinar for opening my eyes to toxicology and Shan for encouraging me to try to stand on my own two "researcher" legs. This has been an experience I would not be without. Further I would like to express my happiness to have had the chance to get to know and work with Nina Elisabeth Landvik, Elín Einarsdóttir Thornèr, Kent Hart, Kine Martinsen, Tove Andreasen and Kristine Haugen. I would also like to give my thanks to the rest of the Tox group at NIOH for making me feel welcome.

To all of you on the 5th floor; Siw Basmoen, Tove Igeland, Anders Rasmushaugen, Fang-Chin Lin, Yu-Chiang Lai, Hanne Karlsen and Siri Bråthen – I would not have survived without our good laughs!

To all of my friends: thank you for enduring my ups and downs as well as my absent-mindedness. I promise I will make it up to you! You rock!

A special great thanks to Karène Jacques Jensen for the infinite number of times you supported me and gave me a hug, for all the help when I was writing my thesis and for all the great concerts we have been to. You are the best!

I would like to thank my family; my mom Lise Hongrø and my dad Frank Solbakken for believing in me, my brother Tommi Hongrø Andersen for supporting me and the rest of my family for all the good times. I love you all!

Lastly, I would like to express my gratitude to my fiancé Lars Erik Hamre. You are my solid ground, love and "Monster". Thank you for being who you are.

Oslo, December 2008

Monica Hongrø Solbakken

Abstract

The MDM2 protein, a p53 inhibitor, is involved in cell cycle control and apoptosis. *MDM2* may promote tumorigenesis when over expressed, amplified or mutated. The expression of *MDM2* has been shown to be modified by some functional single nucleotide polymorphisms (SNPs). SNP309, a T → G transversion, is localized to position 309 in the *MDM2* P2 promoter. An association between SNP309 and lung cancer has been reported. However, there is no functional data for this SNP in human lung cells. In this project the *MDM2* P2 promoter activity with T/T or G/G SNP309 genotypes in human lung cells was characterized. Two luciferase reporter vectors containing the SNP309/P2 region (SNP309-T/G-Luc) were constructed and transfected into human lung cells. Since *MDM2* has been associated with an increased risk of cancer in women and has been reported to interact with estrogen receptors, SNP309 estrogen responsiveness was also investigated. This was done by exposure of the SNP309-T/G-Luc-transfected cells with 17ß-estradiol.

SNP309-T/G-Luc transcriptional activity was significantly higher with the T/T genotype in lung cell lines. To rule out a tissue specific effect, cell lines derived from breast, colon and cervix were investigated. Consistently, the SNP309-T/T-Luc reporter vector had a higher transcriptional activity. Estrogen exposure in 4 human lung cell lines showed no significant effect on transcriptional activity. Still, the T/T genotype showed a higher transcriptional activity, except in one cell line where there was no difference between the two genotypes. A trend indicating a decreased transcriptional activity for G/G in the female lung cancer cell line NCI-H2009 was observed. For SNP309 further elucidation of interacting factors and possible linkage with other SNPs, need to be investigated.

A second functional SNP, C1797G, in the *MDM2* P1 promoter was recently identified and shown to affect *MDM2* mRNA levels. An association between C1797G and risk for bladder cancer has been observed. The role of this polymorphism in susceptibility to lung cancer has not been investigated. In this project its association with lung cancer in a panel of lung cancer patients and healthy controls from Norway was investigated. The results showed a possible protective effect for lung cancer in C1797G heterozygote subjects after adjustment for age, sex and smoking. This effect was only seen for female smokers and was statistically significant indicating a gender-specific effect. Functional studies of this SNP should be performed to investigate the underlying mechanisms for the observed protective effect.

Abbrevations

Amp^R ..... Ampicillin resistance

ANOVA ..... Analysis of variance

Apaf-1..... Apoptosis activating factor 1

ARF ..... Alternative reading frame

ATM ..... Ataxia telangiectasia mutated

ATP ..... Adenosine triphosphate

B(a)P..... Benzo(a)pyrene

Cdk ..... Cyclin dependent kinase

cDNA ..... Complementary DNA

ChiP ..... Chromatin immunoprecipitation

CI ..... Confidence interval

CYP ..... Cytochrome P450 oxidase

Cyt c ..... Cytochrome c

ddNTP ..... Dideoxy ribonucleotid 5'-triphosphate

DNA ..... Deoxyribonucleic acid

dNTP ..... Deoxy ribonucleotide 5'-triphosphate

dsDNA……. Double stranded DNA

E.coli ..... *Escherichia coli*

EMSA..... Electro mobility shift assay

ERα..... Estrogen receptor alpha

ERβ..... Estrogen receptor beta

EtBr ..... Ethidium bromide

G1 ..... Gap 1 phase

G2 ..... Gap 2 phase

HAUSP..Herpes virus associated ubiquitin spesific protease

IARC .....International Agency for Research on Cancer

LAF ..... Laminar air flow

LB.....Luria-Bertani

LB-amp ..... Luria-Bertani with Ampicillin

Li Fraumeni..... Individuals with germ line mutations in *p53*

M……..Molar

M phase .....Mitotic phase

MDM2.....Murine minute double 2

ml……Milliliters

mmol……Millimol

mRNA .....Messenger ribonucleic acid

NaOH .....Sodium hydroxide

NCBI .....National Center for Biotechnology Information

Neoplastic .....Abnormal growth/accumulation of cells

NES ..... Nuclear export signal

NIOH……..National Institute for Occupational Health

NLS .....Nuclear localization signal

NoLS .....Nucleolar localization signal

NSCLC..... Non-small cell lung cancer

OR .....Odds ratio

P1 ..... *MDM2* promoter 1

P2 ..... *MDM2* promoter 2

P3……*MDM2* promoter 3

PAH..... Polycyclic aromatic hydrocarbon

PCR ..... Polymerase chain reaction

Rb ..... Retinoblastoma

RFLP .....Restriction fragment length polymorphism

RNA ..... Ribonucleic acid

rpm .....Revolutions per minute

S.....S phase

SCLC.....Small cell lung cancer

SDS.....Sodium dodecyl sulphate

SE ..... Standard error

SNP..... Single nucleotide polymorphism

ssDNA .....single stranded DNA

SUMO ..... Small ubiquitin-like modifier

SV40.....Simvian virus 40

UV .....Ultraviolet

µg..... Microgram

Contents

1. Introduction

1.1 Lung cancer

Lung cancer is one of the leading causes of cancer death and is often associated with exposure to carcinogenic substances [1]. Lung cancer can be divided into two major categories: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). NSCLC can be further divided into squamous cell carcinoma, adenocarcinoma and large cell carcinoma [1].

Both in men and women, lung cancer is now the third most common cancer form in Norway. In females, lung cancer incidence has increased over the last years with 7,3 % in 1978-1982 to 21,8 % in 2002-2006 but the incidence in men has gradually decreased since the 1990s. In 2004, there were 10 500 cancer deaths in Norway where lung, colorectal, prostate and female breast cancers were responsible for about half of these [2].

Lung cancer etiology is well established and thus lung cancer is a useful model when studying gene-environment interactions. Furthermore it is a good model to study the molecular mechanisms underlying carcinogen-induced tumors [3].

1.2 Lung function and physiology

The principal functions of the lungs are ventilation, perfusion (delivery of arterial blood to the capillary bed in the lungs) and diffusion. The gas exchange takes place in the alveolar space with diffusion of oxygen and carbon dioxide over the epithelium. Lungs are exposed to toxic substances by two routes; inhalation or by the bloodstream [4].

Inhaled particles are deposited in the airways according to size, shape, density and water solubility (also for gases). Highly water soluble gases/particles deposit early in the respiratory system and become relatively non-toxic. Insoluble gases/particles such as O_3 and NO_2 penetrate deep into the alveolar space where they can elicit a toxic effect. If the gas/particle is extremely water insoluble it can cross the epithelial surface and possibly be taken up in the bloodstream to exert its toxicity in other locations. Some effects of toxic substances can be bronchoconstriction, increased proliferation of defense cells, fibrosis, asthma and lung cancer. There are many substances known to cause both

acute and chronic damage to the lungs. Some of these are metals, fibers (i.e. asbestos), coal and cotton dust, inorganic compound such as ammonia and hydrogen fluoride, nitrogen oxides and ozone. These give rise to oxidative damage, DNA adducts and direct physical damage [4].

1.2.1 Some lung damaging substances

Asbestos is a silica fiber and comes in several forms. Exposure occurs in mining and construction industries where asbestos once was widely used due to its insulating and fireproofing properties. It can give rise to 3 types of lung disease: asbestosis (similar to fibrosis), lung cancer and malignant mesothelioma. Macrophages release inflammatory mediators such as cytokines and interleukins in response to the fibers. This leads to additional stress on the lung with additional collagen production and damage caused by reactive intermediates which can lead to cancer [5].

Free radicals are made by loosing or gaining an electron or by homolytic fission of a covalent bond. Several chemicals, for example Paraquat, can give rise to free radicals and then often by transferring electrons to molecular oxygen, forming radicals and then restoring the parent xenobiotic (redox cycling). These radicals are also produced endogenously as intermediates during enzymatic activity, during detoxication of xenobiotics and during inflammation. Free radicals have been shown to be involved in chemical carcinogenesis and induction of inflammation. In addition they may interact directly with DNA to give rise to several structural changes in the nucleotides or DNA strand [5,6].

Polycyclic aromatic hydrocarbons (PAHs) are found in tar and coal deposits as well as being formed during incomplete combustion of organic matter and fossil fuels. Humans are exposed mainly by pollution from traffic, burning of waste, burning of fuel/gas and by consumption of grilled/smoked food. PAHs are a group of structurally related compounds consisting of aromatic, benzene like, carbon rings. PAHs can also contain unsaturated rings with 4, 5 or 6 hydrocarbons. These rings may be substituted with different chemical groups such as alkyl, amino, halogen or nitro groups which will alter the chemical and toxic properties of the compound [7,8].

One of the biomarkers for exposure to PAHs is the concentration of benzo(a)pyrene (B(a)P) in an organism [9]. As an example of detoxication, B(a)P metabolism is shown in figure 1-1.

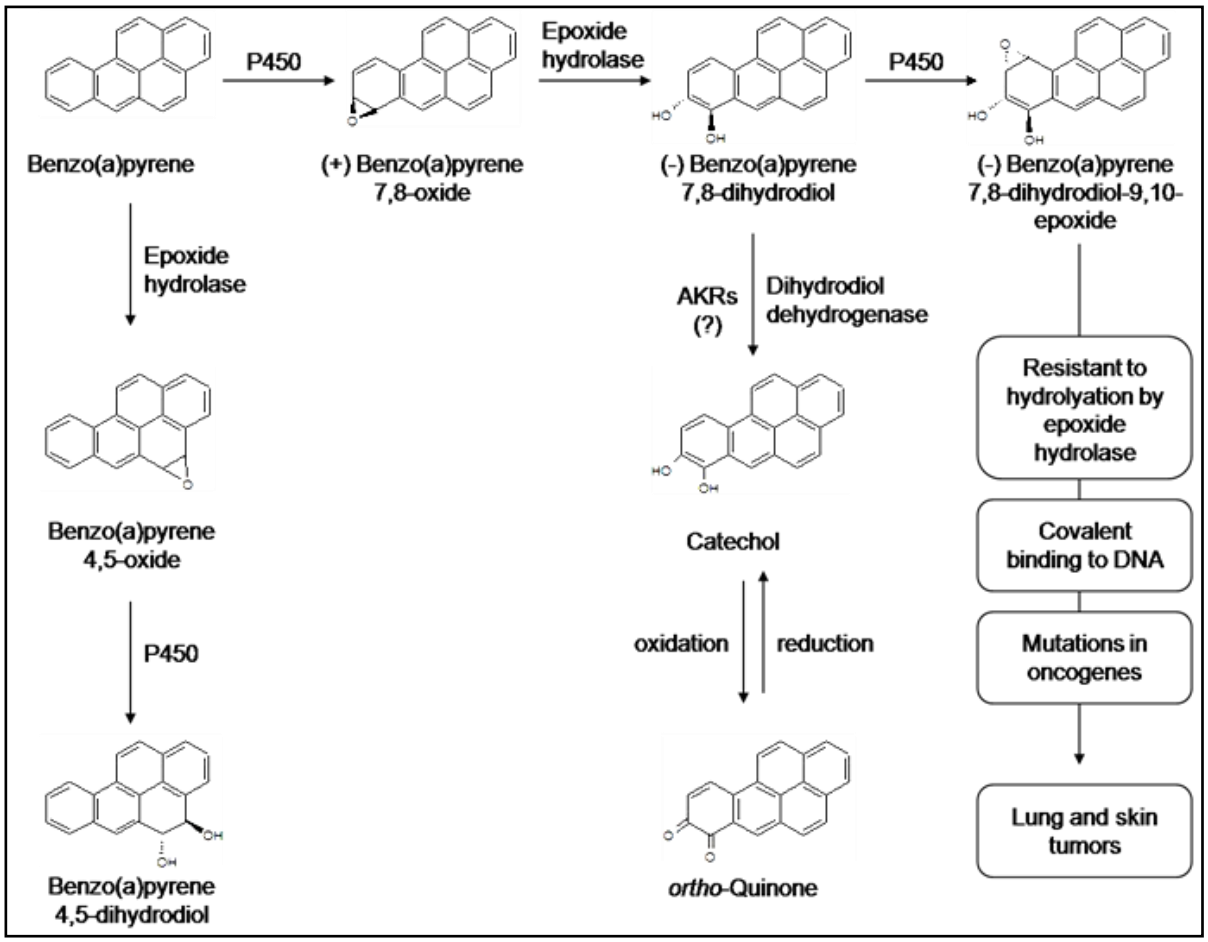


Figure 1-1: The biotransformation of benzo(a)pyrene showing different metabolic outcomes depending on positioning of the initial hydroxyl group. AKR; aldoketo reductase. P450; Cytochrome P450 oxidase. Modified after Casarett and Doull's Basic science of Poisons [5].

B(a)P is a good example of how intermediates can become carcinogenic and that there are several outcomes of the detoxifying process [5].

Exposure to cigarette smoke is one well known factor that often leads to lung damage and cancer. The smoke emerging from the mouthpiece of a cigarette contains approximately 4800 compounds of which 60 have been classified as carcinogens by the International Agency for Research on Cancer (IARC). These include, among others, PAHs, nitrosamines, nickel, cadmium, ethylene oxide and all are shown to induce lung tumors in at least one animal species. Cigarette smoke also contains free radicals and induces oxidative damage in the lung. In addition, NO and other oxidants from cigarette smoke contribute to oxidative damage [10]. In figure 1-2 an overview of cancer development in smokers is shown.

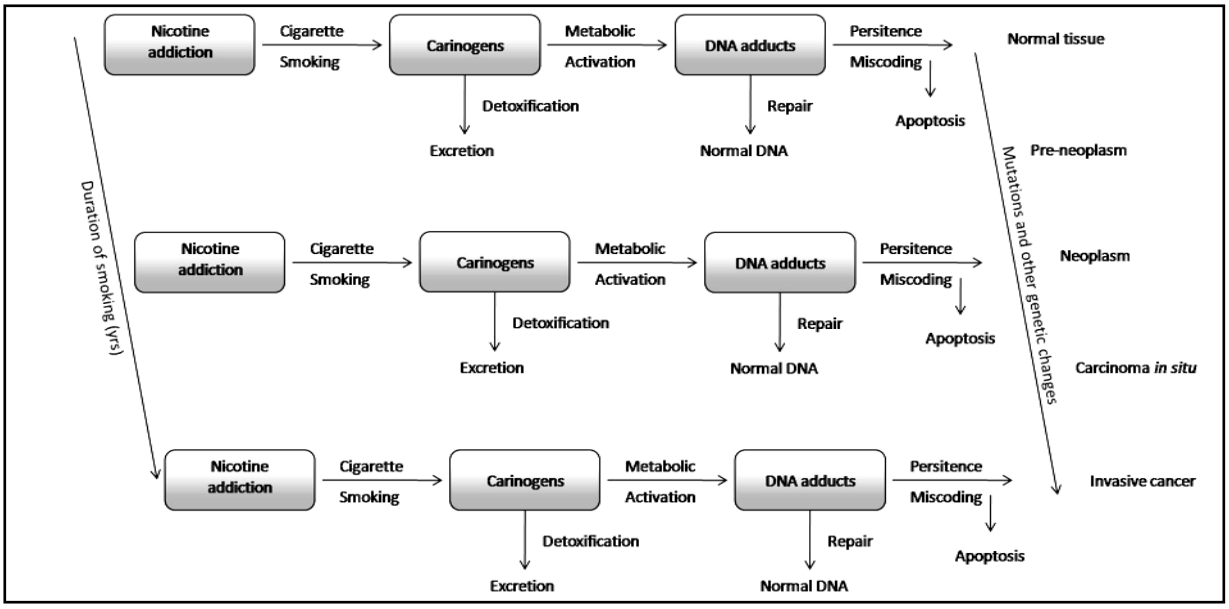


Figure 1-2: Cancer development during chronic exposure to cigarette smoke. The different detoxifying steps and how the pre-neoplastic lesion develops during prolonged cigarette smoke exposure. Modified after Pfeifer et al [10].

Tobacco smoke carcinogens, and other carcinogens, can be metabolically activated by detoxifying enzymes (usually cytochrome P450 oxidase (CYP)). Metabolites can interact with DNA to form DNA adducts. These adducts may lead to miscoding, mutations and possibly cancer if the adducts escapes the repair mechanisms [11]. Lung cancer is a disease where 80 % of the cases are attributed to smoking. In contrast only about 15 % of smokers get this illness [12]. Both primary and second hand cigarette smoke exposure is associated with lung cancer, though with different relative risk. The risk of developing lung cancer decrease after smoking cessation [11].

1.2.2 Phase I and II detoxifying enzymes

To detoxify xenobiotics cells have evolved phase I and II metabolizing enzymes. Phase I biotransforming enzymes are oxidases where CYP is one of the most active and have numerous substrates. CYP is also involved in the metabolism and possible activation of xenobiotics and other carcinogenic compounds. Endogenously, CYP is amongst other involved in the biosynthesis of steroid hormones, bile acids, fat-soluble vitamins and fatty acids. CYP catalyzes the addition of an oxygen atom to the toxic compound to increase its water solubility and simplify excretion. This detoxication and excretion process is, if needed, further aided by phase II conjugating enzymes. Phase II enzymes attach conjugates that will further detoxify the compound. An example of phase II enzymes is glutathione transferases which are located in the cytoplasm in most cells.

Glutathione transferases catalyses the addition of glutathione (a tripeptide) to an electrophilic atom (O, N, S or C) in the xenobiotic molecule. The negative side of the detoxifying process is that some of the intermediates produced are reactive and usually have electrophilic centers. These can interact with nucleophilic DNA and proteins leading to adduct formation and cell damage [5,10,13].

1.3 Estrogen and lung cancer

An increase in lung cancer incidence in females has been seen. This has lead to an increase in the focus on estrogen involvements in lung cancer development. Estrogen affects the growth, differentiation and function of several cell types in both males and females. NSCLC cells have been stimulated to proliferation *in vitro* by estrogen. It has been suggested that estrogens are direct carcinogens after metabolic activation to catechol (possible mutagenic and DNA adduct formation) and may be directly involved in carcinogenesis by promoting cell proliferation and altering the metabolic activation of carcinogens [1,14,15].

Estrogen has two receptors in humans: estrogen receptor alpha (ERα) and estrogen receptor beta (ERβ). The receptors have different tissue distribution and ERβ is the predominant receptor found in lung tissue. ERα and ERβ have been suggested to have opposite functions, proliferative and non-proliferative respectively. ERα is more likely to be expressed in lung tumors in females than in men [14-17]. When estrogen binds to ERs the resulting complex functions as a transcriptional activator of specific estrogen responsive genes. The ligand-ER complex can also interact with other transcription factors. Upon binding their ligand ERs go through conformational changes that will engage the transcription machinery and/or locally affect the chromatin structure to regulate transcription [5,14].

ERs may interact with several genes involved in carcinogenesis. An association between the p53 inhibitor *MDM2* (minute double murine) over-expression and *ERα* expression has been reported. Further, Saji *et al* has shown that MDM2 enhances the transcriptional activity of *ERα* [15].

1.4 Gentic suceptibility and risk of lung cancer

Different individuals have different susceptibilities to disease. Some of these differences are believed to be the cause of heritable traits that modify the effect of environmental exposures [12]. Possible host factors involved in susceptibility are the balance between phase I and phase II

detoxication of carcinogens and the individual's capacity to repair DNA damage. In addition there are multiple common low-penetrance (low risk) gene alleles in the general population which are associated with small increases in individual risk. These contribute considerably to overall lung cancer risk in the population.

More than 10 million genetic variants exist and they are divided into 2 major groups: tandem repeats and single nucleotide polymorphisms (SNP). Tandem repeats are a collection of successive repeats of various lengths. A SNP is where the same nucleotide position can have several different genotypes in the population. The less common genotype variation must be present in 1 % or more of the population to be characterized as a SNP. SNPs can alter the gene expression by affecting promoter activity and alter the protein product by affecting RNA splicing and the amino acid composition [18]. These alterations may lead to different metabolism and different responsiveness to drugs and carcinogens. SNPs can also result in inter-individual responses when these mutations are present in cell cycle regulatory genes, DNA repair genes, inflammatory genes etc [19].

1.5 Carcinogenesis

Carcinogenesis is the progression of a normal cell to a malignant cell. To become malignant the cell must overcome several check points and ignore regulatory signals for cell proliferation. In addition it must avoid apoptosis (programmed cell death) and any other programmed limitation to proliferation such as senescence and differentiation [20].

In general carcinogenesis can be divided into 3 steps: initiation, promotion and progression. Initiation is when a hidden genetic alteration arises and is fixed in the genome by several rounds of cell division. This alteration may arise naturally, after exposure to a genotoxic chemical or after physical damage. The initiation step is irreversible. An initiated cell can be promoted by physical damage or by repeated exposure promoting substances. Promotion alters the genome expression of the initiated cell. It also enhances the proliferation signals and leads to excessive cell division. This gives rise to several benign pre-neoplastic lesions, but these will regress if the promoting agent is removed. Progression is the transition from a pre-neoplastic lesion to a malignant neoplastic lesion by major karyotypic alterations in the already genetically unstable cell. Progression is spontaneous and occurs with a low frequency [5,20,21].

Cancer-critical genes are highly conserved and are categorized in two groups: proto-oncogenes and tumor suppressor genes. Mutations in proto-

oncogene will lead to gain of function (oncogene) whereas in the tumor suppressor gene it will lead to loss of function. An oncogene can be identified as a gene that, when introduced into a normal cell, will result in cancer-like characteristics in the cell. In a normal cell the proto-oncogenes are transiently transcribed to regulate the cell cycle, but if they are permanently up-regulated they favor cell transition to a neoplastic lesion [5,20,22].

1.6 Cell cycle control

Cell division is necessary to replace dead and damaged cells in tissues and maintain epithelial turnover in for example the skin and the gastrointestinal tract. The immune system also uses this ability to mount an expanding adaptive immune response. If a damaged cell continues to divide and proliferate it may give rise to cancer. To help prevent this, the cell has a complex regulatory system to control the cell cycle [20,23,24].

The cell cycle consists of two phases: the interphase and the mitotic phase (M phase). The interphase can further be divided into gap 1 (G1), DNA synthesis (S) and gap 2 (G2) phases which are shown in figure 1-3.

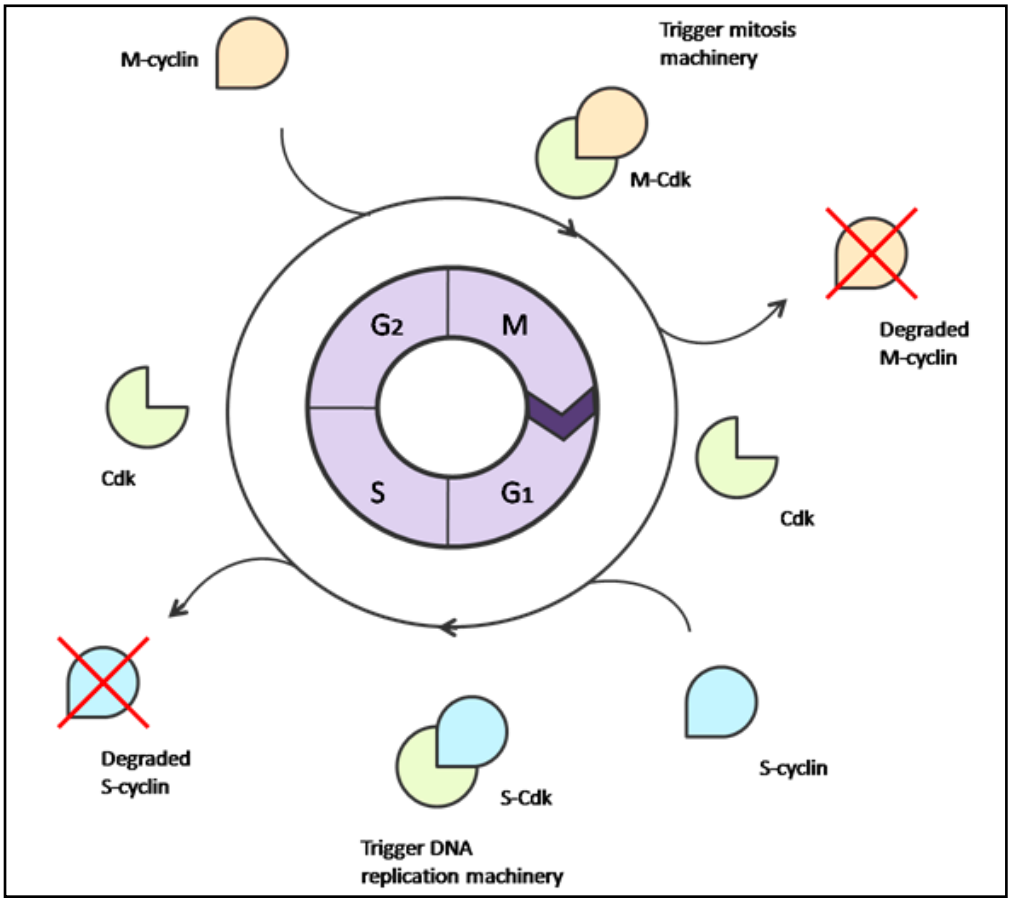


Figure 1-3: Overview of the cell cycle phases showing the involvement of different cyclin-dependent kinases and cyclins in addition to which cyclin-Cdk complexes initiate processes for entry into the next cell cycle phase. Modified after The Cell 4th edition [20].

G1, M and G2 provide cell cycle checkpoints and allow the cell to increase the amount of cytoplasm and number of organelles before continuing with either DNA synthesis or mitosis. The G1 checkpoint is dependent on environmental stimuli and if there is favorable conditions the S-phase is initiated. In late G2 the DNA must be fully replicated and the conditions favorable for the cell to enter M phase. Check point ensures that damaged /incomplete DNA does not become distributed to possible daughter cells [20,23].

The cell cycle is mainly controlled by two groups of proteins; cyclins and cyclin-dependent kinases (Cdk). The cyclin-Cdk system is regulated both on gene level by regulating transcriptional activity and on protein level by phosphorylation, inhibitory proteins and proteasomal degradation [20,25]. It is often the G1 progression and the S phase initiation that are disrupted in cancer cells. The kinase activity during G1/S transition is mediated by a transcription factor called E2F. E2F binds to promoters in genes involved in S phase entry. E2F is controlled by the retinoblastoma protein (Rb) which is a cell cycle inhibitor. When the cell receives external

proliferation signals the amount of active cyclin-Cdk exceeds the amount of Rb and Rb is phosphorylated by the active Cdk. This reduces Rb's affinity for E2F and it dissociates. E2F is then free to activate S-phase gene expression. Another important factor involved in the G1 checkpoint is the transcription factor and tumor suppressor gene product p53. Activation of p53 is often mediated through DNA damage and results in activation of DNA repair mechanisms, controlled cell death and/or cell cycle arrest [20,24,26].

Steroid hormones are involved in regulation of the cell cycle by modulating the effects of cyclin and Cdk inhibitory proteins. The effect of estrogens on proliferation is modulated by ERs. Estrogen-ER complexes functions as transcription factors or interacts with other transcription factors. Estrogens stimulate the G1 phase of the cell cycle by increasing the amount of G1 specific cyclins. This event is again regulated by Cdk inhibitors and activators [25].

1.7 Apoptosis

Apoptosis, or programmed cell death, is initiated in cells that are not longer needed i.e. during tissue formation, in damaged and in infected cells [27]. In contrast to necrosis, apoptosis does not affect the neighboring cells. An apoptotic cell will shrink and condense at the same time as the cell displays factors that cause it to be phagocytosed before any leakage of cell contents can occur. During necrosis cell content enters the extra cellular environment and induce inflammation [20]. Programmed cell death by apoptosis is characterized by the activation of a family of cysteine proteases called caspases. Caspases are tightly regulated by a number of factors such as Bcl-2, BAX and p53. Functional mutations in these will lead to dysregulation of apoptosis and possibly cancer [20,27].

Examples of incidences that can induce the apoptotic program are disruption of the cytoskeleton, DNA damage, disruption of the endoplasmatic reticulum and ligand binding to death receptors (extrinsic pathway). All apoptotic pathways converge on one factor: cytochrome c (cyt c). Cyt c is found in the intermediate compartment in mitochondria. Released cyt c in the cytosol binds to the apoptosis activating factor 1 (Apaf-1). This complex will activate caspase 9 which sequentially will activate caspases 3 and 7 leading to cleavage of downstream substrates. This mediates the breakdown of the cytoskeleton and nuclear envelope, condensation of DNA and signaling to phagocytic neighboring cells [20,23,27].

DNA damage frequently induces apoptosis, but the cellular response to the apoptosis signals varies with cell type, cell cycle status and differentiation state. The outcome is dependent on different expression patterns of pro- and anti-apoptotic factors [27].

1.8 p53 - pathways and functions

p53 is a tumor suppressor gene and its product is the transcription factor p53 which is involved in cell cycle control, apoptosis and maintenance of genetic stability. *p53* is frequently mutated in cancers and the mutations usually lead to loss of p53 function [28,29]. In normal cells its activity is kept low to avoid cell cycle disruption and untimely death [29]. p53 is activated by factors disrupting its binding to the p53 inhibitor MDM2. In response to DNA damage p53 is phosphorylated by kinases. It can also be activated by a transcription factor called c-Myc which will induce the transcription of an MDM2 inhibitor called *p14ARF*. When p53 is activated it binds to several control regions for genes involved in maintenance of genetic stability, inhibition of the cell cycle and apoptosis [27,29]. p53 induces or maintains growth arrest through expression of cell cycle regulators [30].

Activation of the p53 pathway through DNA damage leads to several possible outcomes [31,32]. These outcomes are shown in figure 1-4.

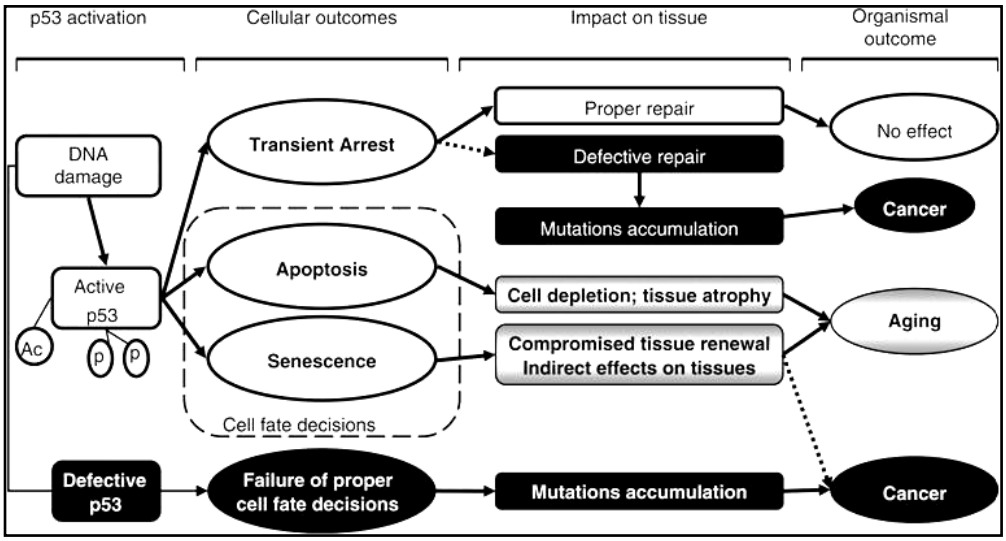


Figure 1-4: Effects of the p53 pathway showing different outcomes where the cell is able to repair/not repair the damage and subsequent tissue effects. Courtesy of Rodier et al [31].

When cells detect genotoxic stress they activate the DNA damage response. This response is a complex interaction between DNA repair factors and cell cycle regulators. Loss of p53 affects this response and

compromises proper DNA repair and cell fate decisions. The different types of DNA damage (base damage, adducts or strand breaks) recruit factors such as ataxia teleangiectasia mutated (ATM) protein kinase which will phosphorylate p53. This reduces binding to MDM2 and leads to stable activation of p53. In the case of severe damage, p53 will translocate to mitochondria and induce apoptosis or induce transcription of death receptors. If less severe the DNA repair machinery will be activated [31].

1.9 Murine Double Minute 2 (MDM2)

The MDM2 protein, a p53 inhibitor, is involved in cell cycle control, regulation of p53 activity and functions as an E3 ubiquitin ligase. *MDM2* is a proto-oncogene that will promote tumorigenesis when over expressed or mutated [33].

The structure of the *MDM2* gene is showed in figure 1-5.

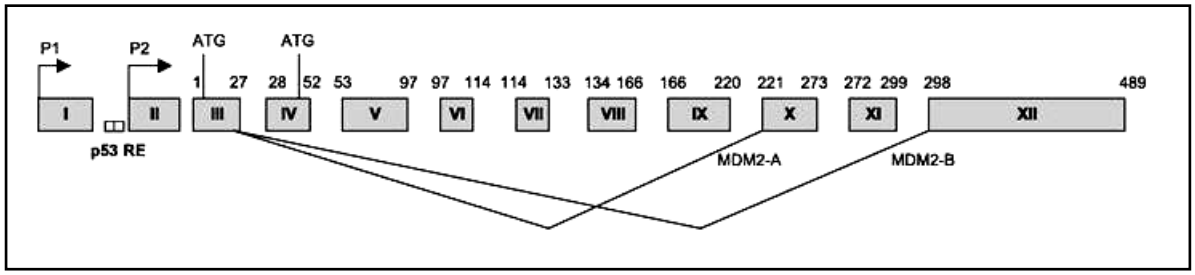


Figure 1-5: An overview of the MDM2 gene showing exons 1 – 12, start codons and p53 response elements. The two major splice variants are also shown as MDM2-A and MDM2-B. From Iwakuma et al [33].

The human *MDM2* gene is 34 kb in size, consists of 12 exons and is located on chromosome 12q13-14. The gene has 2 promoters: P1, P2 and there are also reports indicating a third promoter (P3) in the third intron of *MDM2.* The first start codon after P3 is located at the 50th amino acid in the full-length MDM2 protein which means that this version will be truncated by 49 amino acids. This will give a protein without a p53 binding domain [34]. P1 is likely to regulate the MDM2 levels in non-stressed cells. P2, which is p53 and radiation responsive, is located to the first intron and it is likely to regulate MDM2 levels in stressed cells [33,35]. The two promoters P1 and P2 can give rise to two different mRNA (messenger ribonucleic acid) products. These mRNA products give rise to the full length p90 (start codon in exon 3) and the shorter protein p76 (start codon in exon 4). p76 is missing a part of the p53 binding domain, which makes it unable to interact with p53. It can also act as a dominant negative inhibitor of p90. Since p90 is inhibited by p76 it cannot interact with p53 and p53 levels will rise [33,35,36]. The consequence of

transcription from either of P1 or P2 is still unknown since the start codon for the full length protein is located in exon 3. It is thought that the RNA (ribonucleic acid) translation of products from P1 is less efficient than translation of products from P2 [35].

Alternative splicing (different exon combinations) of *MDM2* has been identified in different cancer types as well as in normal tissues. In figure 1-6 some alternative splice variant proteins are shown.

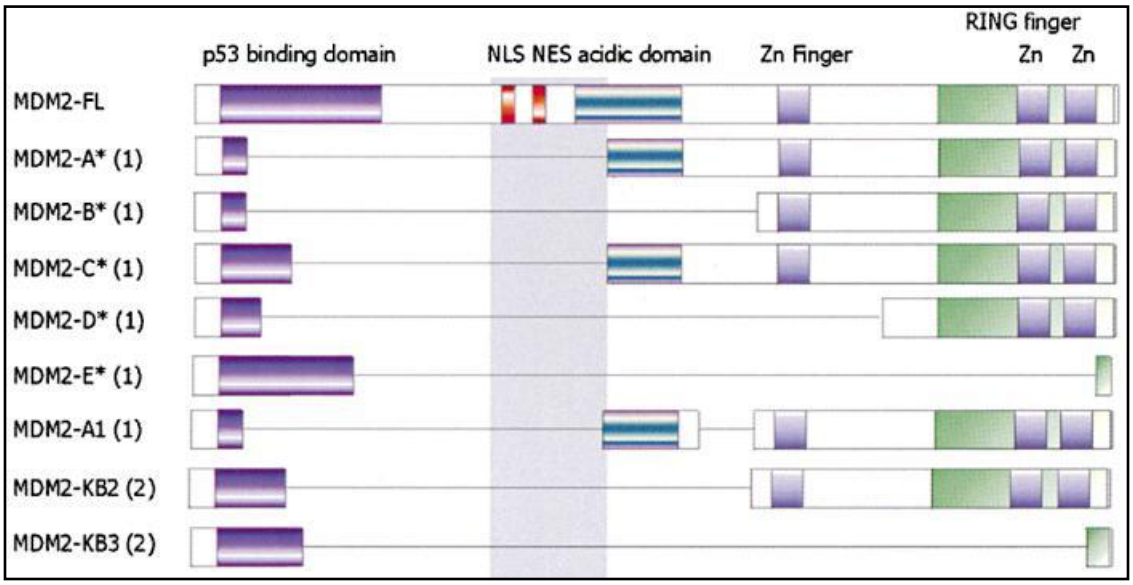


Figure 1-6: Examples of alternative splice variant proteins of MDM2 where many of them have no/truncated p53 binding domains. The RING domain is largely conserved as well as the zinc fingers. There are more than 40 known splice variants of MDM2. From Bartel et al [37].

Of the more than 40 splice variants found, the majority lacks a complete p53 interaction domain. *In vitro* studies have confirmed that 4 of these are unable to interact with p53 at all. Splice variant proteins are usually expressed alongside the full length MDM2 and it has been shown that they can regulate each other as exemplified by p76 and p90 *MDM2* transcripts [37]. Splice variants have been detected in both normal lung tissue and NSCLC tissue. Significantly higher amounts of alternative MDM2 variants have been seen in smokers versus non-smokers and males show more splice variants than women [38]. Some splice variants are only found in one or a few tumor types and suggest an association with tumorigenesis and contribution to tumor characteristics [39].

The MDM2 protein contains several domains that are important for its different functions. The protein and an overview of the different domains are shown in figure 1-7.

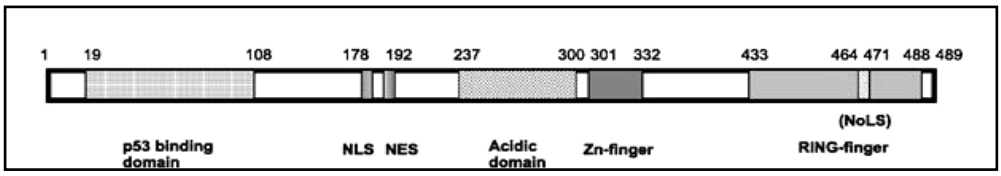


Figure 1-7: Protein overview of MDM2 showing the p53 binding domain, the localization signals, the RING-domain, the zinc finger and the acidic domain. From Iwakuma et al [33].

MDM2 has three localization motifs; a nucleolar (NoLS) and a nuclear (NLS) import signal in addition to a nuclear export signal (NES). It also has several interaction domains; a RING domain, an acidic domain, a zinc finger and a p53 domain. The p53 interaction domain enables *MDM2* to interfere with p53 transcriptional activity even when MDM2 is not expressed. This is accomplished when p53 binds to the p53 transactivation domain on *MDM2.* This sequesters p53 and leads to expression of MDM2 to further bind p53. All the different localization signals lead the shuffling of MDM2 between the cytosol, nucleus and the nucleolar space. The RING domain is essential for the ligase function of MDM2. The acidic domain is involved in the binding of p300/CBP, whereas the function of the zink finger domain is unknown [33].

MDM2 activity is regulated and there is a balance between when MDM2 ubiquitinates itself or ubiquitinates other proteins. This balance is controlled by posttranslational modifications on the MDM2 protein such as sumoylation (SUMO stands for small ubiquitin-like modifier) and phosphorylation. When MDM2 is sumoylated its E3 ligase activity is shifted towards p53. MDM2 sumoylation is stimulated by p14ARF. Phosphorylation of MDM2 will attenuate the degradation of p53 as well. It has been suggested that acetylation of certain lysine residues in MDM2 also will decrease the E3 activity [41].

1.9.1 MDM2 and cell cycle control

MDM2 interacts with several growth inhibitory factors such as p53, Rb and p14Arf. p14ARF is an alternative reading frame (ARF) of the CDKN2A locus and a growth inhibitor [42,43]. Retinoblastoma protein (Rb) is a tumor suppressor with functions similar to p53 in cell cycle regulation and cell death. MDM2 regulates its activity by binding Rb and disturbing Rb mediated G1 arrest [42]. MDM2 also interferes with Rb ability to inhibit the E2F transcription factor and the cell cycle will progress through G1 [33]. The MDM2-Rb interaction impairs the formation of a complex with p53 thereby overcoming MDM2's ability to inhibit p53 mediated apoptosis [42]. MDM2 also interacts with the transcription factor Sp1. Binding prevents Sp1 to bind to its specific DNA

sequences and thus block transcription. This interaction is challenged by the Rb protein which will reactivate Sp1 by replacing Sp1 in the MDM2-Sp1 complex. Since Sp1 is a general transcription factor the outcome of prevented transcription activation remains unknown [33,44].

In addition MDM2 has three growth inhibitory domains and is thus believed to be involved in cell cycle control. Over-expression of the MDM2 full length version will lead to G1 cell cycle arrest in a normal diploid human cell. MDM2 also seem to harbor a tumorigenic domain in its amino terminal and can therefore become oncogenic in a cell that does not sense the MDM2 mediated cell cycle arrest. It is also possible that several of the MDM2 splice variants code for this tumorigenic protein without functionally growth inhibitory domains [45]. MDM2 is often over expressed in cancers possibly inhibiting the p53 apoptotic pathway and the DNA damage response. Over-expression of MDM2 has been associated with both poor and favorable prognosis [39,40].

1.9.2 MDM2 and E3 ligase function

MDM2 functions as an E3 ubiquitin ligase as part of an E1, E2 and E3 ubiquitination complex. The E1 enzyme binds ubiquitin, which is a 76 amino acid long protein, and activates it with the use of adenosine triphosphate (ATP). The activated ubiquitin is transferred to E2 which is a conjugation enzyme. E2 transfers ubiquitin to MDM2 (E3) which will ligate it to the target, for example p53. This ubiquitin is covalently bonded to a lysine residue in the target. MDM2 can also ubiquitinate itself [20,33]. The RING motif in MDM2 is a common E3 ligase motif responsible for the ligase activity in many proteins [33]. p14ARF binds to the RING domain responsible for the E3 ligase function of MDM2 to regulate its activity [43].

A protein can be ubiquitinated in two ways; polyubiquitination on one lysine residue or monoubiquitination on several lysine residues, and these will have different outcomes. A polyubiquitinated lysine will function as a signal for proteasomal degradation and when MDM2 ubiquitinates itself this is what will happen [20,33]. [33]. p53 is monoubiquitinated by MDM2 and this means that other factors are involved in MDM2 mediated p53 proteasomal degradation. At least 3 different domains in MDM2 seem to be necessary for this function: the p53 interaction domain to locate p53, the RING domain and the p300/CBP domain to completely mark p53 for degradation (see figure 1-7) [40,43]. p14ARF is a growth suppressor and can interact with both p53 and MDM2. It is shown that binding of p14ARF to p53 inhibits ubiquitination by MDM2 and thus inhibits p53 degradation. P300/CBP (CREB binding protein) prevents p14ARF from interacting with MDM2 [33,46],[40].

The deubiquitination enzyme HAUSP (herpes virus-associated ubiquitin-specific protease) is a direct antagonist of MDM2 activity by deubiquitinating p53 after stimulation by DNA damage. HAUSP thus protects p53 from MDM2 mediated degradation [43]. HAUSP can also deubiquitinate MDM2 and has been shown to be involved in apoptosis [47]. Cells without HAUSP have an unstable MDM2 which leads to elevated levels of p53 [48],[49].

1.9.3 MDM2 and the tumor suppressor p53

MDM2 is a negative regulator of p53. MDM2 regulates p53 in several ways; i) MDM2 binds to the DNA binding site on p53 and prevents p53-DNA interaction and subsequent transcription activation. ii) MDM2 ubiquitinates p53 as a signal for proteasomal degradation. iii) MDM2 functions as an export signal to the cytosol of the MDM2-p53 complex [29,33]. *MDM2* is one of the genes regulated by p53 as a transcription factor, but since MDM2 is a negative regulator of p53 this gives rise to a negative feedback loop [33]. Phosphorylation of MDM2 and p53 reduces the affinity for the p53-MDM2 interaction and p53 is then free to activate the DNA damage response [43].

MDM2 contains several localization signals and the nuclear export signal in MDM2 is critical for p53 degradation which takes place in the cytosol [33,40]. Later this notion has been challenged by the findings that p53 is degraded both in the cytosol as well as in the nucleus. p53 has two NES and the monoubiquitination by MDM2 reveals these [43]. Human p14ARF sequesters MDM2 in the nucleolus and blocks export of p53 to the cytosol [42].

1.9.4 Other members of the MDM2 family

MDM4, also called MDMX, is an MDM2 related protein. The proteins have the highest homology at the amino terminal where the p53 binding domain is located. Residues involved in p53 interaction are conserved in both proteins as well as the RING domain. In additional they share a zinc finger in the central region of the proteins [33,50,51].

Some studies have shown that MDM4 interacts with MDM2 via the RING domain to stabilize MDM2 by preventing autoubiquitinylation [46]. It binds to p53 with the same requirements as MDM2 and they cooperate in the regulation of p53. MDM4 does not function as an E3 ligase and has no NLS or NES. In the absence of MDM4, MDM2 is relatively inefficient at degrading p53 due to its very short half-life. Following DNA damage MDM2 marks MDM4 for degradation and the amount of MDM4 declines. MDM4 appears to mask the transcriptional activation domain on p53 and

thus inhibits p53. It also prevents p53 degradation by MDM2 [50]. The MDM2-MDM4-p53 interaction is summarized in figure 1-8.

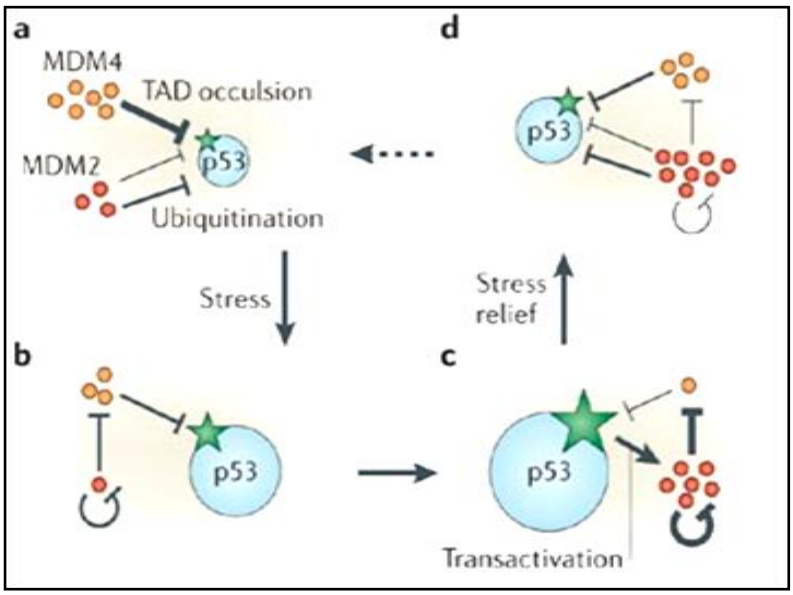


Figure 1-8: How MDM2, MDM4 and p53 interact. A: In an unstressed cell, p53 is kept at low levels due to MDM2-mediated ubiquitinylation. p53 is also kept inactive due to MDM4-mediated transactivation domain (TAD) occlusion. B: After stress, MDM2 degrades itself and MDM4, this leads to the accumulation and activation of p53. C: As activated p53 transactivates MDM2, the increasing amount of MDM2 degrades MDM4 more efficiently, enabling full p53 activation D: Following stress relief, the accumulated MDM2 targets p53 again and p53 levels decrease. In addition as MDM4 levels increase, p53 activity also decreases. Modified from Toledo et al [32].

1.9.5 MDM2 and estrogen

Studies have shown that MDM2 expression increase when the cells are exposed to estrogen. This is most likely an effect of ERα mediated transcriptional activation of *MDM2* when estrogen is present [52]. MDM2 has been shown to interact with ERα and can ubiquitinate ERα for degradation. Duong *et al* has shown that MDM2 is involved in both ligand-dependent and independent decrease of ERα stability. The ligand-dependent decrease is executed in a complex with p53. The ligand-independent pathway is a result of MDM2 over-expression where the interaction with ERα becomes p53 independent [14].

It has also been seen a correlation between increased *MDM2* expression and ERα in breast cancer. Here MDM2 interacted with ERα resulting in increased transcription of estrogen responsive genes [15].

1.9.6 MDM2 and other interacting factors

There are several more proteins interacting with MDM2 and these are briefly overviewed in table 1-1 [33,41,43].

Table 1-1: Factors interacting with MDM2..

Interacting protein	Function and effect
C-abl	Product of a proto-oncogene and is a kinase. Involved in growth/survival signaling. Involved in mediating apoptosis most likely through MDM2 and p53 by accumulating p53. MDM2 is phosphorylated and p53 interaction is blocked.
Androgen receptor	Binds to the RING domain and is marked for degradation.
Cyclin G	Regulates a phosphatase that dephosphorylates MDM2 and thus activates it and p53 is degraded.
L11	Ribosomal protein that binds MDM2 and sequesters it to the nucleoleus. This leads to stabilization of p53. Localization and L11 dependent.
Kinases	Several kinases interact with MDM2, among others ATM and phosphatidylinositol 3-OH-kinase (PI3) . They both phosphorylate specific serine residues that are important for the translocation of MDM2 to the nucleus from the cytosol. This reduces cellular levels of p53 and p53 trancriptional activity.

1.10 Polymorphisms in *MDM2*

There are at least 253 SNPs located in *MDM2* reported to the NCBI (National Center for Biotechnology Information) SNP database where the majority of these are in introns [53].

1.10.1 SNP309

SNP309 was reported in a study by Bond *et al* where they investigated the possibility of naturally occurring genetic variations in important components of the p53 pathway. SNP309 is localized at position 309 in the *MDM2* intronic promoter P2. This region is used by p53 to activate transcription of *MDM2* [54]. An overview of the P2 region with SNP309, transcription factor binding sites and response elements is shown in figure 1-9.

Figure 1-9: SNP309 localization in intron 1. Binding sites for transcription factors Ets and AP1 are shown with p53 response elements. From Bond et al [54].

SNP309 is a T → G transversion and in the NCBI SNP database it has the ID rs2279744 (sequence shown in appendix I). Bond *et al* also found that cells homozygous for SNP309 had a higher endogenous level of MDM2, and that there was a putative transcription factor binding site for Sp1 in the SNP309 region. They observed that the putative Sp1 binding site was extended by one nucleotide when the G allele was present. The *in vitro* elevated G/G promoter activity suggested that Sp1 can activate *MDM2* transcription. This was verified by electro mobility shift assay (EMSA) with the G probe having higher affinity for purified recombinant human Sp1. Similar results obtained with nuclear cell extract. The *MDM2* promoter – Sp1 interaction was verified with the chromatin immunoprecipitation assay (ChiP) [36,54]. The elevated MDM2 levels were further shown to calm the p53 DNA damage response after exposure to the chemotherapeutic drug Etopside in that a lower death-rate in the cell cultures were observed [36,54].

The increased in *MDM2* transcription can be partly mediated by ER binding to the *MDM2* P2 promoter. This in addition to the extended Sp1 transcription site with the SNP309 G/G genotype may indicate that estrogen can influence the SNP309 effect on MDM2 transcription [52]. Bond et al proposed a possible chain of events shown in figure 1-10.

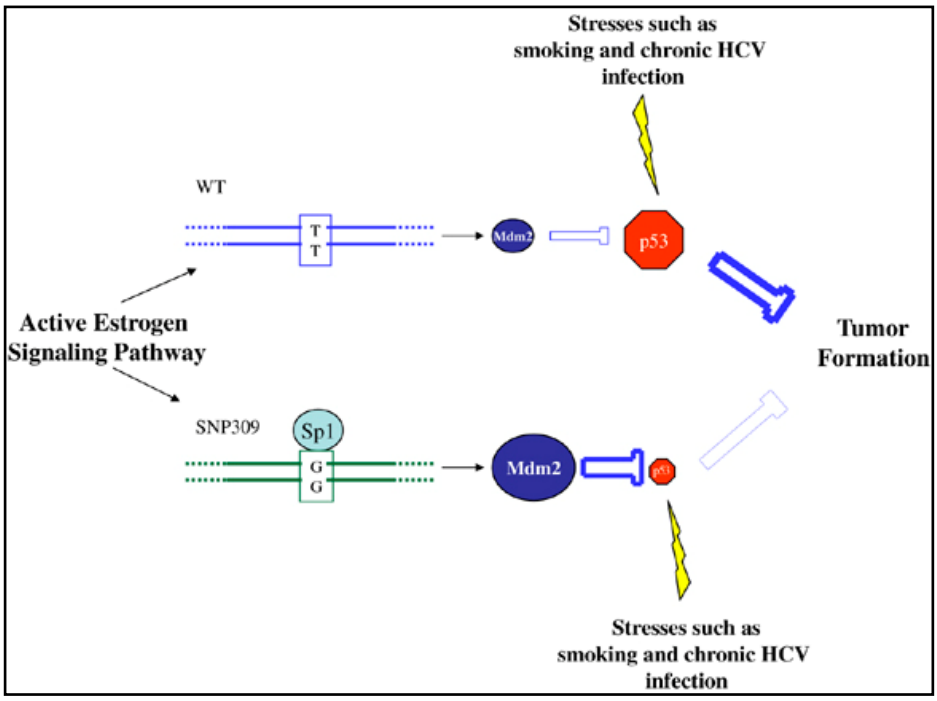


Figure 1-10: The model proposed by Bond et al for how the interaction between SNP309/gender/hormones and stress affects MDM2, p53 and tumor formation. In the presence of an active estrogen signaling pathway the G allele is thought to increase the MDM2 levels with the help of the extended Sp1 binding site. This will then attenuate the p53 apoptotic pathway. From Bond et al [52].

Further, ERα with estrogen present has been shown to be a potent activator of Sp1-driven transcription [55].

SNP309 has later been both associated and not associated with several cancer forms, among others lung cancer, prostate cancer, colon cancer and endometrial cancer. A general overview of articles and their results are presented in appendix V. An overview can also be found in an article by Wilkening *et al* [56] where a combined analysis of the case-control studies available is described. A total of 4276 lung cancer cases and 5318 controls, from different ethnic populations, were analyzed and the results are shown in figure 1-11.

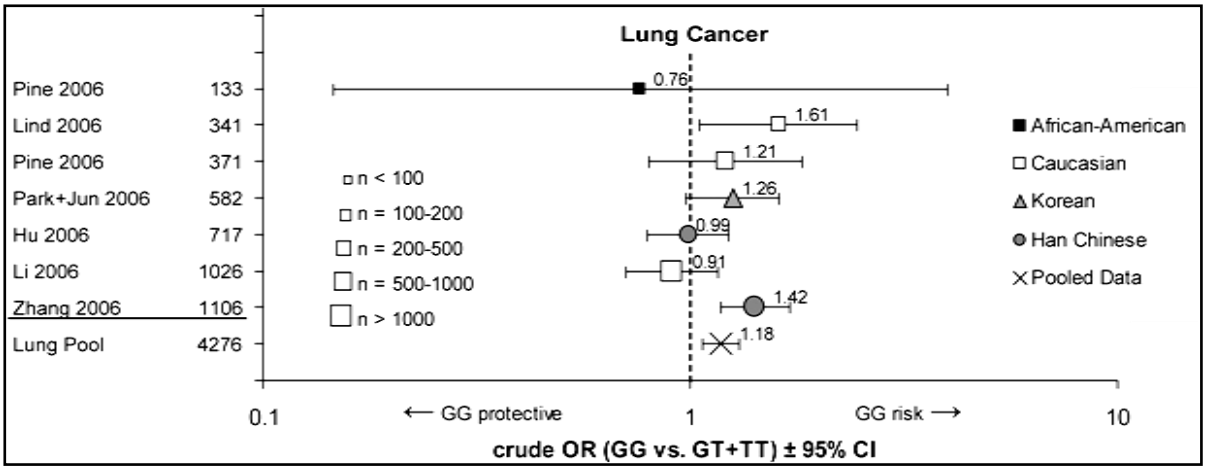


Figure 1-11: The combined analysis of lung cancer risk and SNP309 by Wilkening et al. The individual articles included in this analysis are shown on the left whereas the population studied is shown on the right. From Wilkening et al [56]

For lung cancer an association with SNP309 was found, but the individual studies show some variability in their results. The genotype distribution in different ethnic groups varies and this can be the reason that some population does not give rise to risk association where others do [56].

1.10.2 C1797G

A second functional SNP in *MDM2* was recently reported by Wang *et al* [53]. This SNP is a C → G transition located in the P1 promoter of *MDM2*. In the NCBI SNP database it is given the ID rs937282 (sequence shown in appendix I) and is located ~780 bp upstream of SNP309.

The genotype distribution varies in the different populations as shown in table 1-2.

Table 1-2: Genotype distribution in different populations for SNP C1797G. From the NCBI SNP database.

Population	C/C	C/G	G/G
European	0,217	0,478	0,304
African-American	0,087	0,348	0,565
Asian	0,174	0,739	0,087
Global	0,299	0,379	0,322

An association between the SNP and risk for bladder cancer was also reported [53]. Odds ratio (OR) for G/G homozygous genotype adjusted for age, sex, smoking and alcohol from a logistic regression model was 2,45 (95 % CI 1,02-5,72). The authors also saw that the G/G genotype increased a putative binding site for the transcription factor C/EBPα and increased

the *in vitro* expression of *MDM2.* The C/EBPα binding was confirmed by EMSA. In 22 bladder tumor samples from Chinese individuals with different genotypes, carriers of the G/G genotype had higher *MDM2* mRNA levels as well as protein levels [53].

1.11 Project aims and hypothesis

Even though lung cancer is mainly correlated with exposure to cigarette smoke and other carcinogens, many studies indicate the involvement of genetic polymorphisms. Several polymorphisms in the MDM2/p53 pathway have been identified and many of these have been associated with risk of lung cancer. Lind *et al* showed that SNP309 G/G was associated with NSCLC in a Norwegian population. The G/G genotype was associated with higher risk (OR 1,62) and in women this risk was 3 times as high (OR 4,1). The higher risk in women may indicate involvement of sex-specific hormones [57]. Estrogens have the ability to induce maturation and differentiation of the ling at the same time that elevated estrogen levels is a known risk factor for lung cancer [58]. In addition there has been observed earlier age of onset in women with SNP309. The increased risk has been related to a higher transcriptional activity of *MDM2* having the SNP309 G/G genotype [54,59]. It has been shown that SNP309 has an extended Sp1 binding site [54].

The aim of this study was to characterize the promoter activity with the different genotypes and to see if these are consistent with the case-control study done by Lind *et al*. In addition this SNP was thought to be estrogen responsive and a higher transcriptional activity under estrogen exposure was expected.

To investigate this two SNP309 reporter constructs were made. These contained approximately 500 bp of the P2 promoter, with SNP309 T/T or G/G genotypes, upstream of the reporter gene *Firefly luciferase* in the pGL3 basic reporter vector. These were transiently transfect them into human lung cell lines to characterize basal transcriptional activity. Since *MDM2* is thought to be estrogen responsive transfected cells were also exposed to 17β-estradiol to characterize SNP309 estrogen responsiveness.

Recently a new SNP called C1797G was found in the P1 promoter of *MDM2* [53]. This SNP was found to affect *MDM2* expression and it was associated with risk of bladder cancer in a Chinese population. To elucidate if there was any risk with C1797G and NSCLC in the Norwegian population, a series of lung cancer patients and matched healthy controls were genotyped and a risk association study was performed.

The work performed in this project is shown in figure 1-11.

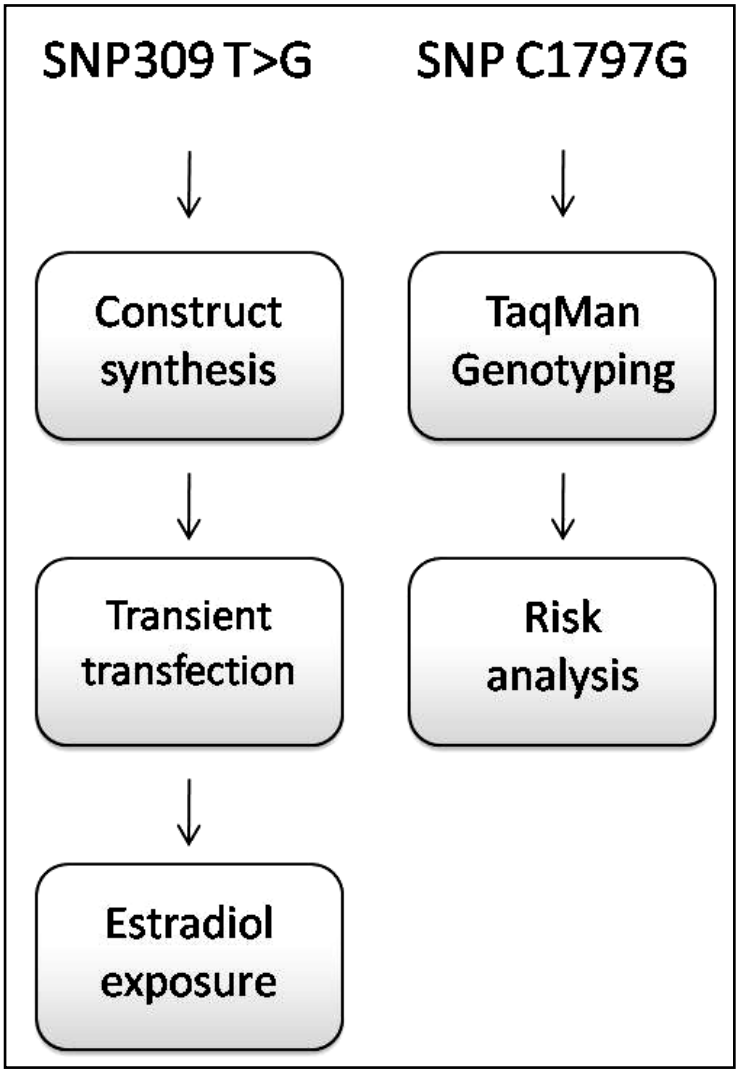


Figure 1-12: Overview of the work done in this project.

2. Methods

2.1 Synthesis of *MDM2* SNP309-T/G-Luc constructs

The basis for the SNP309 construct synthesis is 6 DNA samples from the Biobank (Appendix IV) with known genotype for SNP309. The samples used in the construct synthesis originated from lung cancer patients and the DNA was isolated from the surrounding normal lung tissue or whole blood samples.

2.1.1 Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR) is an *in vitro* method used to amplify specific DNA sequences. The reaction can be divided into three steps: denaturation, annealing and polymerization. These steps are illustrated in figure 2-1.

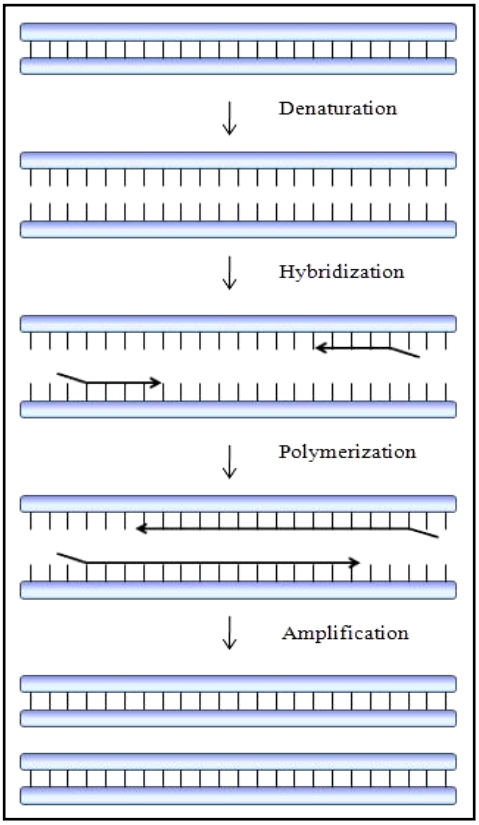


Figure 2-1: The three steps in PCR.

1. The denaturation step is usually done at 95 °C and removes undesired structures such as double stranded DNA (dsDNA) and DNA-primer complexes by disturbing the non-covalent interactions between the two strands.

2. Hybridization, also called annealing, is when the primer binds its complementary sequence in the template. The optimal temperature for this reaction is dependent on primer length and GC-content.

3. During the polymerization (elongation) step the DNA polymerase elongates the primer to produce the complementary DNA strand (cDNA).

Optimal conditions must be established for each reaction. A general reaction mixture would consist of: a forward and a reverse primer, 2'-deoxy 5'-triphosphates (dNTPs), a heat-stable polymerase, reaction buffer, divalent cations and a DNA template. Steps one through three makes up one cycle and is repeated between 25 to 35 times. The reaction will terminate if one or more of the components become depleted.

DNA samples, which were known to have either the SNP309 G/G or T/T genotype, were chosen and are shown in table 2-1.

Table 2-1: DB samples used in PCR.

SNP 309 G/G	SNP 309 T/T
DB 539	DB 503
DB 507	DB 499
DB 495	DB 501

A primer pair was designed to amplify a sequence of 471 bp containing SNP309. The fragment produced would function as a promoter when inserted upstream of the *Firefly luciferase* in the pGL3 basic luciferase vector. To facilitate the ligation into pGL3 basic the primers were designed to contain two different restriction sites (MluI and BglII) compatible with the multiple cloning region in pGL3 basic. The designed primers with added restriction sites (in bold) are shown in figure 2-2.

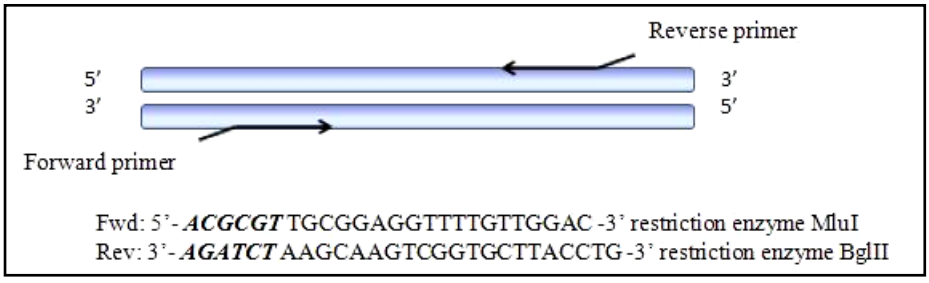


Figure 2-2: Primers specific for the P2 region containing additional restriction sites that were incorporated in the final PCR product (shown in bold).

The PCR reaction was set up in a total volume of 20 µl (shown in table 2-2) and the PCR program is shown in table 2-3.

Table 2-2: Reaction mixture for the PCR.

	In each reaction:	Stock solutions:
Reverse primer	0,5 µl	10 pmol/µl
Forward primer	0,5 µl	10 pmol/µl
Reaction buffer	2 µl	10 X
Hot Fire polymerase	0,5 µl	5 U/µl
dNTP	2 µl	2 mM
dH_2O	13 µl	
$MgCl_2$	1 µl	50 mM
Template	0,5 µl	100 ng/µl

Table 2-3: PCR program.

Temperature	Time	
94 °C	3 min	
94 °C	40 sec	X 30
55 °C	40 sec	
72 °C	60 sec	
72 °C	10 min	
4 °C	∞	

2.1.2 Agarose gel electrophoresis

Agarose gel electrophoresis is a common technique used to separate, identify and/or purify DNA fragments according to size. Agarose is a linear polymer that is composed of alternating D- and L-galactose residues linked by glycosidic bonds. When the gel is placed under an electric field it is able to separate DNA from 50 bp (base pair) to several mega bases in size depending on the agarose concentration. The separated DNA is visualized by staining with fluorescent intercalating dyes, such as ethidium bromide (EtBr), which fluoresce when exposed to UV-light. PCR-samples are mixed with a loading buffer which function is to visualize how far the gel has run (predictable rate). It also contains glycerol that weighs down the samples so that they collect in the bottom of the well.

1. To make a small 2 % gel 50 ml of 1 x TAE buffer or 0,5 x TBE buffer was mixed with 1 g of SeaKem agarose and boiled in a microwave oven to completely melt the agarose (approximately 2 minutes).

2. The agarose was cooled to about 50 °C before adding EtBr to a concentration of 0,2 µg/ml (at higher temperatures the EtBr will break down) and the agarose was left to polymerize for 20 minutes.

3. 5 µl of the PCR-sample was mixed with 2 µl of loading buffer and loaded on to the gel. A suitable DNA standard was loaded in a separate lane as a size reference.

4. The gel was run for 45 minutes at 100 V.

2.1.3 Vectors

pGL3 basic is a reporter vector containing the *Firefly luciferase* and the luciferase gene is cloned from *Photinus pyralis* (the firefly). It has no eukaryotic promoter or enhancer sequence and transcription from the *Firefly luciferase* is dependent on an insert in the multiple cloning site. One can also insert other possibly regulatory sequences in the enhancer region upstream of the multiple cloning site.

pGL3 control is a reporter vector containing the SV40 promoter and enhancer sequences to ensure a high stable expression of Firefly luciferase. In this thesis this has been used as a transfection control.

pRL-TK is an internal control vector containing the *Renilla luciferase* cloned from *Renilla reniformis* (the sea pansy). This gene is controlled by the herpes simplex virus thymidine kinase promoter which gives low to moderate levels of Renilla luciferase. This works as a baseline comparison and gives more stable results when co-transfected with *Firefly luciferase* vectors.

All vector luciferase vector products are dependent on a substrate to generate luminescence since luciferases are enzymes. The substrates are given through buffers during the experiment (see chapter 2.3.5).

All vector maps are shown in appendix II.

2.1.4 Restriction digestion

Restriction enzymes are a group of proteins derived from bacteria that have recognition sequences in DNA, usually palindromic 4-8 bp long sequences, to which they bind and cut. Restriction digestion can be used for genotyping and as a part of cloning.

To construct the luciferase expression vector containing the *MDM2* promoter (SNP309-T/G-Luc constructs), the PCR product and pGL3 basic were digested with MluI and BglII restriction enzymes (double digest) to make sticky overhangs. This ensures insertion in the correction orientation in the subsequent ligation reaction. The double digest setup is shown in table 2-4.

Table 2-4: Restriction digestion of PCR product and vector.

	Insert	Vector
PCR product	14 µl	0
pGL3 vector	0	5 µl
NEB buffer 3	2 µl	2 µl
Bgl II	1 µl (10 U)	1 µl (10 U)
Mlu I	1 µl (10 U)	1 µl (10 U)
dH20	2 µl	11 µl
Total volume	20 µl	20 µl

Both reactions were incubated for 2 hours at 37 °C, separated on a 2 % agarose gel to confirm that the digestion products sizes.

2.1.5 Purification of restriction digestion products

To remove any unused components from the digestion mixture that might interfere with downstream reactions, the products were cleaned up using the QIAquick nucleotide removal kit.

http://www1.qiagen.com/Products/DnaCleanup/GelPcrSiCleanupSystems/QIAquickNucleotideRemovalKit.aspx

This method is based on the use of resin columns that will bind DNA so that other residues can be washed away. After washing, DNA was eluted from the column in 150 µl of elution buffer. The purified products were further up-concentrated using Microcon® YM-30 Centrifugal Filter Units from Millipore.

http://www.millipore.com/catalogue/item/42422

This filter retained any DNA fragments larger than 50 bp dsDNA or 60 bp single stranded DNA (ssDNA) and the samples were finally eluted in 30 µl dH_2O.

2.1.6 Ligation of *MDM2* insert and pGL3 vector

A ligation reaction will glue together two pieces of DNA. This is done by an enzyme called DNA ligase which synthesizes phosphodiester bonds. The DNA ends either have complementary sequence overhangs or blunt ends, which will give either a specific or unspecific ligation, respectively. The enzyme used here is the T4 DNA ligase derived from the T4 bacteriophage. It will join two cohesive DNA ends that have the 3' hydroxyl termini and 5' phosphate termini respectively. The ligation reaction setup is shown in table 2-5.

Table 2-5: Ligation reaction.

Digested PCR product	7 µl
Digested pGL3 diluted 1:10	1 µl
T4 DNA ligase	1 µl
T4 DNA ligase buffer	1 µl

The ligation reaction was run over night for 12 hours at 16 °C followed by 4 hours at 10 °C.

2.2 Work with Eschericia coli (E.coli).

All work with bacteria was performed in a laminator air-flow bench (LAF) and basic sterile techniques were followed. All equipment, media and solutions were autoclaved or sterile filtered before use. The *Escherichia coli* (*E.coli)* strain used was 5-alpha F'Iq Competent *E.coli* (New England Biolabs).This strain has high transformation efficiency for non-methylated DNA derived from *in vitro* synthesis.

2.2.1 Uptake of *E.coli* from long term storage

1. A glycerol stock of the desired *E.coli* clone was taken from the -80 °C freezer and put on ice.
2. A sterile inoculation pin was used to transfer a small amount of stock to 5 ml of LB-media in a Falcon tube.
3. The culture was grown over night at 37 °C and 250 rpm. The next day the culture was ready for use.

2.2.2 Long term storage of *E.coli*

Storage of *E.coli* was done by freezing them to –80 °C in LB-media containing 15 % glycerol. The glycerol content prevented lysis of the bacterial cell wall during freezing.

1. Colonies were grown in 5 ml of LB-media to early lag-phase.
2. 750 µl of the culture was mixed with 250 µl of 60 % glycerol and immediately put in liquid nitrogen.

The glycerol stocks were stored at –80 °C.

2.2.3 Transformation of *E.coli* and cloning of SNP309-T/G-Luc constructs

Transformation is the uptake of foreign DNA from the external environment into bacteria. This process will take place if the bacteria are competent (able to take up DNA), either by induced mechanical stress such as electroporation or chemically induced with for example calcium dichloride ($CaCl_2$) treatment. Bacteria can also be naturally competent, which is a genetic trait. DNA binding proteins on the bacterial surface binds the DNA and transports it through the cell wall. The foreign DNA can either be incorporated into the genome or it can reside in the cytoplasm and replicate independently of the bacterial genome. When a plasmid is taken up by a bacterium and replicated to make several copies we refer to this as cloning.

The transformation was performed as follows:

1. 2 µl of the ligation product (from table 2-6) was added to 50 µl of competent *E.coli* cells.

2. The mixture was incubated on ice for 30 minutes, followed by a heat-shock for 30 seconds at 42 °C and left on ice for 5 minutes.

3. 950 µl of SOC medium (see appendix I) was added and the bacteria was incubated for 1 hour at 37 °C and 250 rpm.

4. The transformation mixture was plated out on Luria-Bertani (LB) and LB-Ampicillin agar plates (LB-amp) (see appendix I) with the dilutions 1:10, 1:20 and the remaining transformation mix on a last plate. The plates were incubated over night at 37 °C.

The following day Ampicillin resistant colonies (due to pGL3 *beta-lactamase* gene) were picked and screened by PCR. For this PCR (table 2-6), primers that bind in the vector sequence just outside the multiple cloning region, were used. These primers were RV3 and GL2 (see appendix II).

Table 2-6: PCR reaction for screening of colonies.

dH_2O	17 µl
Reaction buffer 10X	2,5 µl
dNTP 2mM	2
RV3 10 pmol	0,5 µl
GL2 10 pmol	0,5 µl
Hot Fire polymerase 5 U/µl	0,5 µl
$MgCl_2$ 50 mM	1 µl
Template	picked colony

The fragment size produced was expected to be 619 bp long.

2.2.4 Plasmid isolation

The positive colonies were picked and the plasmid isolated using the Mini- or Maxiprep using kits from Qiagen.

Miniprep and Maxiprep

These methods are based on the alkaline lysis of bacterial cells with sodium hydroxide (NaOH) followed by adsorption of the plasmid DNA on a silica gel in the presence of salt. After adding the NaOH the lysate is neutralized and SDS (sodium dodecyl sulphate) precipitates the chromosomal DNA and cell debris. The chromosomal DNA attaches to SDS, but the plasmid DNA renaturates and stays in suspension. The precipitate is removed by centrifugation and the supernatant is applied to the silica membrane. The silica membrane retains the plasmid DNA during all wash steps until elution. For protocols see URL below.

http://www1.qiagen.com/Products/Plasmid/QIAprepMiniprepSystem/QIAprepSpinMiniprepKit.aspx

The output yield for Miniprep is estimated to be between 5 and 15 µg and for Maxiprep 20 µg to 500 µg. All isolated plasmids were checked with PCR containing GL2 and RV3 primers and restriction digestion with MluI and BglII enzymes as described in chapter 2.1.4.

2.2.5 Sequencing

There are two well-known and used methods for sequencing: the Sanger method and the Maxam-Gilbert method. The Sanger method is used in this thesis and is based on the generation of fragments that terminate at specific points along the sequence. Termination is the result of added dideoxy 5'-triphosphates (ddNTPs) in the sequencing mixture. The DNA template is mixed with a DNA polymerase, one sequencing primer, buffer

and both dNTPs (2'-deoxy 5'-triphosphate) and ddNTPs. The ddNTPs are labeled with non-radioactive fluorescent dyes and are incorporated randomly in the sequence. They will terminate the elongation because they lack the 3' hydroxyl groups that would otherwise be a part of a phosphodiester bond. At the end of the reaction one will have fragments differing one nucleotide in size because of this random termination. The reaction mix is then run through a capillary which contains a polymer. A laser detects the ddNTPs signals and generates the sequence with the help of compatible software.

The kit used was BigDye® Terminator v1.1 Cycle Sequencing Kit from Applied Biosystems. The kit contained a ready reaction mix with ddNTPs, dNTPs, buffer and polymerase as well as a positive control DNA. The sequencing was preformed with Applied Biosystems ABI Prism® 310 Genetic Analyzer. The preceding single primer PCR reaction (table 2-7) was set up according to the enclosed protocol.

Table 2-7: Single primer PCR reaction also containing ddNTPs.

Terminator Ready reaction mix	4 µl
Template construct	100 ng
Sequencing primer (GL2 or RV3)	1,6 pmol
dH_2O	3 µl
Total volume	10 µl

The reaction was run on a thermal cycler with the program shown in table 2-8.

Table 2-8: Single primer PCR program.

Temperature	Time	
96 °C	60 sec	
96 °C	10 sec	X 35
50 °C	5 sec	
60 °C	4 min	
4 °C	∞	

The reaction products were cleaned up with DyeEx 2.0 Spin kit from Qiagen to remove any unincorporated ddNTPs that would otherwise lead false positive detection of non-existent bases. The kit is based on a resin gel, through which the DNA is filtered. The resin columns were vortexed to evenly disperse the resin before use and excessive fluid was removed by centrifugation at 2800 rpm for 3 minutes. The single primer PCR sample was added, spun down at the same speed/time and collected in an Eppendorf tube.

http://www1.qiagen.com/Products/DnaCleanup/DyeExDye-TerminatorRemovalSystem/DyeEx2_0SpinKit.aspx

The sample was freeze-dried and resolved in HiDi formamide in a total volume of 20 µl, before being mounted on the 310 Genetic Analyzer. Instruments were operated according to the manual with the use of POP-6™ Polymer and running buffer, both from Applied Biosystems. Colonies with the correct plasmid insert sequence were cultured and the plasmid was isolated with the Maxiprep kit (described earlier in chapter 2.2.4).

2.2.6 Quantification of DNA

The isolated SNP309-T/G-Luc constructs were quantified by using both a spectrophotometer (Eppendorf BioPhotometer) or by using Invitrogen's Quant-iT™ PicoGreen® dsDNA Reagent.

http://probes.invitrogen.com/media/pis/mp07581.pdf

PicoGreen is a fluorophore that upon binding to dsDNA fluoresce. This nucleic acid stain is used to quantify the amount of dsDNA in a solution without being disturbed by any RNA or free nucleotides in the sample. The samples are mixed with the PicoGreen reagent and analyzed on a spectrofluorometer. Concentrations are calculated based on the standard curve made from the enclosed Lambda DNA and are shown in table 2-9.

Table 2-9: PicoGreen standard curve dilution series.

Standard	Concentration	ng of DNA in 200 µl well
1	2 µg/ml	200 ng
2	1 µg/ml	100 ng
3	0,2 µg/ml	20 ng
4	0,04 µg/ml	2 ng
5	0,008 µg/ml	0,2 ng
6	0 µg/ml	0 ng

The standard and the samples were read using a Packard Fusion microplate analyzer.

The results from the standard dilution series were plotted and linear regression was done to the standard curve. By the use of linear regression the standard curve function was obtained. Sample concentrations were found by reading of the standard curve and calculation backwards by considering the dilutions. An example of a PicoGreen standard curve is shown in figure 2-3.

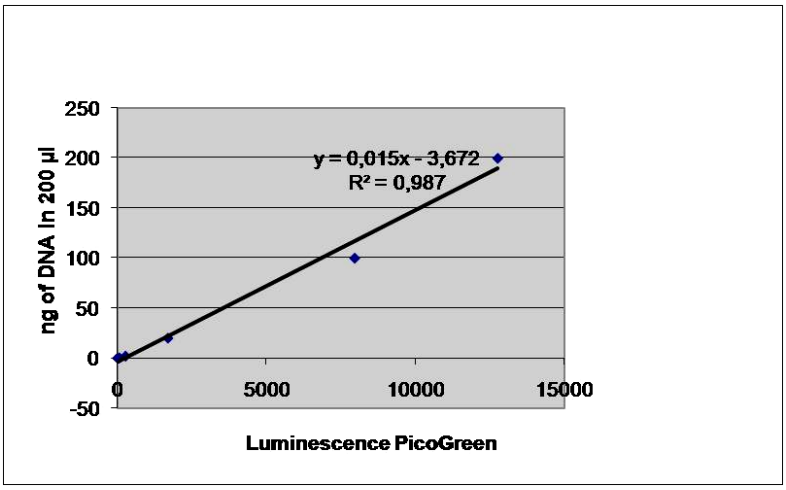


Figure 2-3: PicoGreen standard curve showing both the curve function and the R2 value.

The sample concentration was found by reading of the standard curve from the sample luminescence value and calculation backwards considering the dilutions.

2.3 Human cell culture experiments

All techniques involving cell cultures were performed in a LAF bench and basic sterile techniques were followed. All equipment, media and solutions were autoclaved or sterile filtered before use. The cells used were both human cancerous and immortalized cells (see appendix I).

2.3.1 Thawing of cells from liquid nitrogen storage

During uptake of a desired cell line from liquid nitrogen storage the following protocol was followed:

1. A vial containing the desired cell line was taken out of liquid nitrogen storage and put on ice.

2. The vial was transferred to a beaker of 70 % ethanol pre-warmed to 37 °C.

3. The cell culture was thawed about ¾, transferred to a 15 ml Falcon tube where 5 ml of appropriate media was added slowly (about 2 minutes).

4. The suspension was spun down for 4 minutes at 1000 rpm and the supernatant was discarded to remove all DMSO (dimethyl sulphoxide).

5. Cells where resuspended in 8 ml of complete media and transferred to a 9 mm Petri dish or a 15 cm^2 flask. Cell culture media was changed the next day.

2.3.2 Maintaining epithelial and suspension cell cultures

After reaching confluency the cells were replated with the use of a trypsin solution (PET). Trypsin degrades the cell adhesion molecules anchored to the plastic dish allowing us to replate the cells and maintaining the cell culture. The technique is done as follows:

1. Cells were washed with 3-5 ml of phosphate buffered saline (PBS).

2. 1 ml of PET containing 0,02 % trypsin was added and the petri dish was placed in the incubator until the cells were in suspension.

3. The trypsination was stopped by adding 3 ml of complete media containing serum.

4. Cells in suspension were transferred a Falcon tube and spun down for 4 minutes at 4 °C and 1000 rpm.

5. The supernatant was discarded and the cells were resuspended in 3-5 ml of media. If the cells were to be used in an experiment they were counted by using a Bürker counting chamber and distributed to appropriate dishes/flasks in known numbers. If they were not to be used in an experiment they were distributed 1:4 to 1:10 depending on the cell line.

When working with suspension cells the culture was spun down directly. When used in an experiment the cells were stained with Trypan blue 0,4 % (50µl cell culture in 450 µl Trypan blue) before counting. Trypan blue is used to distinguish between viable and dead cells where dead cells became blue because of permeated cell membranes.

2.3.3 Storage of human cells in liquid nitrogen

Human cell lines are stored in liquid nitrogen suspended in a DMSO (dimethyl sulphoxide)/antibiotic/media solution where the DMSO prevents crystallization and lysis of cell membranes. To freeze the cells the following was done:

1. When the cells were grown to confluency they were trypsinated according to protocol 2.3.2, spun down and resuspended in antibiotic freeze media (AF, 500 µl per vial to be used).

2. The cells were distributed to vials and 500 µl of DMSO 8% was added to each vial.

The freezing process was done gradually to –80 °C before being stored in the liquid nitrogen tank.

2.3.4 Transient transfection

Transfection is the delivery of cloned DNA into a eukaryotic cell and can be achieved by three means; biochemical transfection, physical transfection and virus-mediated transduction. The choice of method is based on the cells ability to survive the stress that comes with chosen transfection protocol and efficiency requirements. Transfection can be divided into two groups; stable and transient transfection. Stable transfection is used to make a clonal cell line where the transfected DNA/gene is integrate into the chromosomal DNA. Here, if it is a gene, it will be transcribed and produce moderate amounts of protein. The DNA integrated into the genome needs to contain a selection marker (for example antibiotic resistance) so that transfected clones can be selected from the non-transfected cells. The efficiency of stable transfection is 1-2 orders of magnitude lower than for transient transfection. During transient transfection the recombinant DNA is introduced into cells, but temporarily. This gives a high level of expression of the target gene, but the DNA is not necessarily incorporated into the chromosomal DNA. Other functions, such as promoter activity, can also be analyzed. Transient transfection is a desired method when working with many samples and results are obtained within fairly short time.

In this thesis a transient biochemical transfection is used. A cationic lipid solution is mixed with the negatively charged DNA to form DNA-lipid vesicles (liposomes). These liposomes, where the lipids are positively charged, fuse with the negatively charged cell membrane to transfer DNA across the lipid bilayer.

Three different transfection reagents were used during optimalization of the transfection. In the end the preferred reagent was NEB TransPass™ D1. For all transfection reagents each replicate contained 200 000 cells (all except for BEP-2D and BEAS-2B which had 250 000 cells) which were distributed to one 3,5 cm dish. Each individual experiment consisted of three replicates per condition/parameter and the following day the cells were approximately 50-80 % confluent. In every experiment a positive control and negative control was included. The positive control was the pGL3 control vector containing the SV40 promoter and enhancers to give a strong positive control signal. The negative control was a transfection without DNA.

Set 1 – Lipofectamin and Plus reagents (optimalization only).

. Transient transfection was carried out as follows:

1. To 100 µl of transfection media 1 µg of the SNP309-T/T-Luc or SNP309-G/G-Luc construct and 0,1 µg of Renilla vector was added and mixed.

2. 4 µl of Plus reagent was added and after mixing the reaction was incubated for 15 minutes at room temperature.

3. To 100 µl of new transfection media 5 µl of Lipofectamin was added. The solution was mixed and added to the DNA/Plus reagent solution. Incubated for 15 minutes at room temperature.

4. The Lipofectamin-DNA-Plus reagent mix (200 µl) was diluted in 1 ml of transfection media and distributed to the petri dish. The cells were incubated for 3 hours at 37 °C.

5. After incubation the cells were washed with 1 ml of PBS before adding 2 ml of complete media.

Set 2 – FuGENE 6 transfection reagent (optimalization only).

1. To 100 µl of serum free media 3 µl of FuGENE reagent was added.

2. After mixing, 1 µg of the SNP309-T/T-Luc or SNP309-G/G-Luc construct and 0,1 µg of Renilla vector was added. The solution was mixed and incubated for 15 minutes at room temperature.

3. The transfection mixture was added directly to the complete media in the dish and the cells were incubated over night at 37 °C.

4. The following day the cells were washed with 1 ml of PBS and complete media was added.

Set 3 – NEB TransPass™ D1.

1. To 1 ml of serum free media 3 µl of transfection reagent was added.

2. After mixing, DNA was added and the solution was incubated for 20-30 minutes at room temperature. A total of 1,1 µg or 1,05 µg DNA was used, the latter for H2009 cells only, to give either a 1:10 or 1:20 ratio.

3. The cells were washed with 1 ml of PBS before adding the transfection mixture. and incubated for 3 hours at 37 °C.

4. After incubation the cells were again washed with PBS and 2 ml complete media was added. The cells were incubated for approximately 24 hours at 37 °C.

2.3.5 Dual-Luciferase® Reporter assay system

The pGL3 basic vector used in the synthesis of the SNP309-T/G-Luc constructs contains a *Firefly luciferase* gene downstream from the multiple cloning site. Introduction of the *MDM2* promoter will therefore control the activity of this downstream reporter gene. The SNP309-T/G-Luc constructs were transiently co-transfected with the Renilla luciferase vector. The two different luciferases are activated by different buffers containing specific substrates and thus the co-transfected Renilla vector will give a baseline for comparison to the Firefly luciferase activity. When using the ratio between the two luciferases one will achieve more stable results. Because the *Firefly luciferase* gene is regulated by the promoter insert, the ratio between the Firefly and the Renilla vector will give an indication of the *MDM2* promoter activity compared with the different SNP309 genotypes.

After the completed transient transfection the Luciferase/Renilla ratio was read with the use of Promega's Dual-Luciferase® Reporter assay kit.

1. 200 µl of 1 x Passive Lysis Buffer was added to the transfected cells and the dish was placed on a gyrotray for 15 – 30 minutes to achieve cell lysis and freeing the luciferase enzymes.

2. To each well to be used on a 96 well plate 100 µl of Luciferase Assay Buffer II was added. This buffer will provide substrate the Firefly luciferase produced by the SNP309-T/G-Luc constructs.

3. 20 µl of sample was added to each well before mixing. The luminescence from the Firefly was read for 1 second on the luminometer.

4. To the same well 100 µl of Stop & Glo buffer was added before mixing. This buffer contains a substrate for the Renilla luciferase produced by the pRL-TK as well as a quencher for the Firefly luciferase activity. The luminescence from Renilla luciferase was read as for the Firefly luciferase. The ratio Firefly/Renilla luminescence was then calculated.

2.3.6 Exposure of cells to 17β-estradiol

Transient transfected cells were also exposed to 17β-estradiol immediately after transfection. This was done with three concentrations of 17β-estradiol (0 nM, 1nM and 10nM) with 3 replicates for each concentration.

1. Cells were plated and transiently transfected as described in chapter 2.3.4 with NEB TransPass™ D1.

2. In 2 ml of complete media, 17β-estradiol solved in DMSO was diluted to the concentration of 10 μM.

3. For each dish to be exposed 2 ml of media was to be used and 17β-estradiol from the 10 μM stock was diluted to 10 nM and 1 nM. DMSO content in the final cell culture media was never above 1 ‰.

4. 17β-estradiol was added immediately after the 3 hours incubation step in transient transfection after washing with 3-5 ml of PBS.

5. Incubated for 24 hours and the luciferase and renilla luminescence were read as described earlier with the DLR™ assay (chapter 2.3.5).

2.4 Genotyping

TaqMan is a PCR based method where there are specially designed primers and a SNP specific set of reporter probes (oligonucleotides) involved. The probes are labeled at the 5' end with a fluorescent group (here FAM for G/G and VIC for C/C). At their 3' end there is a non-fluorescent quencher that will eliminate background noise and make the assay more sensitive. The quencher absorbs any fluorescence from the reporter group in an intact probe due to close proximity between the two. During the PCR there will be generated more sequence specific DNA that the probes will bind to, but when the DNA polymerase replicates the template its 5'-3' exonuclease activity will cleave the fluorophore from the probe. The quencher can no longer absorb the fluorescence from the fluorophore and a signal can be detected. The signal intensity increase as the PCR progresses as a result of more probe cleavage. Since the two probes are specific for either one of the SNP genotypes, one or both fluorophores will dominate in a reaction giving either a homozygous or heterozygous genotype. Several SNPs can be genotyped in a single reaction as long as the probes have different fluorophores.

TaqMan assays from Applied Biosystems have been improved by using a Minor Grove Binder. This contacts the DNA helix in the minor groove and

improves probe-template specificity and interaction stability. In figure 2-4 an overview of the TaqMan PCR reaction from Applied Biosystems is shown.

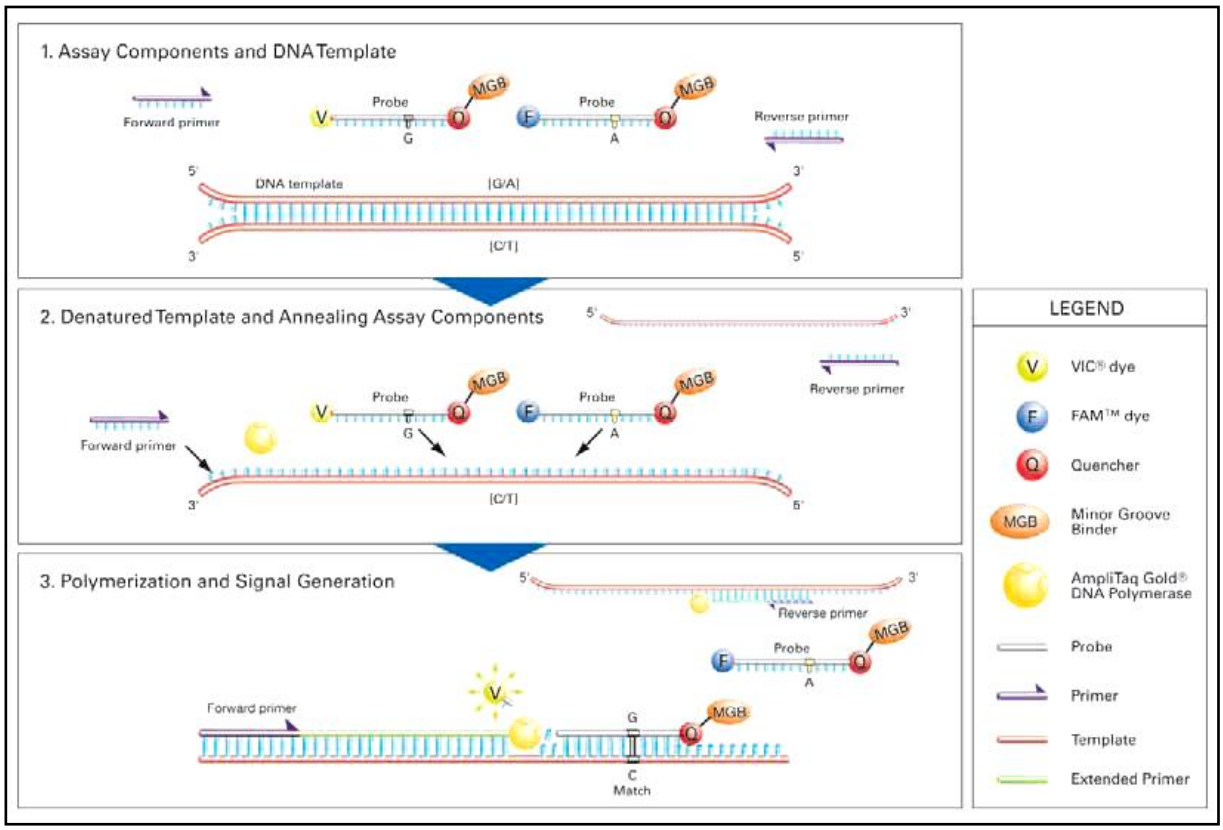


Figure 2-4: Overview of the TaqMan method. From Applied Biosystems.

2.4.1 Restriction Fragment Length Polymorphism (RFLP)

This is another genotyping method based on PCR and restriction digestion. It is dependent on that a site-specific restriction enzyme site is present in the SNP sequence and that it is disrupted when the other SNP genotype is present. The PCR primers are designed so that digestion fragments can be readily separated on an agarose gel. The PCR is run with the specific primers and run on an agarose gel to ensure the presence of a PCR product. The PCR product is then digested with the chosen enzyme. The restriction reaction is loaded onto a high concentration agarose gel and run.

Both SNP309 and C1797G could be genotyped this way. Cell lines used in transient transfection with SNP309 constructs where genotyped to elucidate endogenous SNP309 status. Biobank samples for the C1797G - lung cancer association, that were not genotyped with TaqMan, were also genotyped this way.

SNP309

A PCR was run as described in chapter 2.1.1. This PCR was the digested with MspA1I restriction enzyme. The SNP309 G allele gave rise to an additional restriction site for this enzyme. The restriction digestion was set up as described in table 2-10 and separated on a 4 % agarose gel with EtBr.

Table 2-10: RFLP restriction digestion for SNP309.

PCR product	7 µl
NEB buffer 2	2 µl
MspA1I	0,5 µl (5 U)
dH_2O	10,5 µl
Total volume	20 µl

C1797G

The primers designed for C1797G are shown in appendix II. This SNP gives rise to a site for the restriction enzyme Hpy188I. A PCR reaction was set up as shown in table 2-11 which was further digested with Hpy188I as shown in table 2-12 to be separated on a 4 % agarose gel with EtBr.

Table 2-11: Preliminary PCR for subsequent RFLP genotyping of SNP C1797G.

	In each reaction:	**Stock solutions:**
Reverse primer	0,5 µl	10 pmol/µl
Forward primer	0,5 µl	10 pmol/µl
Reaction buffer	3 µl	10 X
Hot Fire polymerase	0,5 µl	5 U/µl
dNTP	2 µl	2 mM
dH_2O	22 µl	
$MgCl_2$	1 µl	50 mM
Template	0,5 µl	100 ng/µl

Table 2-12: RFLP restriction digestion for SNP C1797G.

PCR product	15 µl
NEB buffer 4	3 µl
Hpy188I	0,5 µl (5 U)
dH_2O	11,5 µl
Total volume	30 µl

2.4.2 TaqMan genotyping of biobank for SNP C1797G

The TaqMan reaction was run on 384 well plates covered with optical film. In table 2-13 an overview of the reaction contents is shown. The TaqMan PCR program is shown in table 2-14.

Table 2-13: TaqMan reaction contents.

Master PCR mix	1,5 µl
40 X SNP assay mix	0,1 µl
Template DNA 20 ng/µl	0,5 µl
Filtered water	3,46 µl
Total volume	5 µl

Table 2-14: TaqMan PCR program.

Temperature	Time	
50 °C	2 min	
95 °C	10 min	
92 °C	15 sec	X 58
60 °C	60 sec	
4 °C	∞	

1. The SNP assay mix was thawed on ice while the DNA samples were mixed and spun down.

2. All components of the reaction, except DNA template, were mixed and distributed in every other row on the 384 well plate.

3. DNA template was added and the plate was covered with optical film, spun down and placed on the thermal cycler.

4. The plate was read and the data analyzed with the Applied Biosystems 7900 HT and supplementary SDS 2.0 software.

2.5 Statistics

A description of the statistical methods used in this thesis is described in this chapter. For a thorough description of the details behind the statistical methods please see appendix VI.

One assumes that *in vitro* cell culture experiments gives normally distributed results, but before any statistical analysis was performed the data was checked for normally distribution with the use of SigmaPlot 10.0. SigmaPlot uses the Kolmogrov-Smirnov test to see if the data provided is normally distributed. This test standardizes the data and

compares them to a standard normal distribution The normality test was passed before continuing with the selected T-test or ANOVA (analysis of variance) test.

2.5.1 P-values and odds ratios (OR)

P-values

The p-value is a measurement of the probability of achieving a result that is alike or more extreme than what is observed. This value is based on true null hypothesis so one can also describe it as the amount of evidence you have against the null hypothesis. One rejects the null hypothesis if the p-value is less than the level set as significant (usually 0,05). If the null hypothesis is true, a p-value of 0,05 indicates that there is only a 5 % likeliness that the results you have are like those observed [60].

Odds ratio

Odds ratio can be defined as the ratio between the odds of a certain happening in one group and the odds for the same happening in another group. The groups must be dichotomous (from the same whole, but not overlap). An odds ratio of 1 indicates that the odds are the same in both groups. An odds ratio > 1 indicates that the event is more likely in the first groups as a ratio < 1 indicates more likeliness in the second group. The odds ratio must be ≥ 0 [60].

2.5.2 Linear regression

A regression analysis will study the relationship between 2 variables where one is dependent and the other is explanatory/independent. The relationship is assumed to be linear, the observations independent and that the errors have the same variability. Linear regression gives the regression line function as shown in equation 1.

$$Eq.1: \quad y = a + bx$$

Here b describes how much the y-value will change with one unit difference in x.

To find the regression line the method of least squares is employed. It calculates the smallest deviation from a straight line for all the data points present.

Also the R^2 value can be calculated and this is the Pearson's coefficient of correlation. Pearson's coefficient will be used with normally distributed data and will vary between -1 and 1 for negative and positive correlation, respectively. It describes the strength and direction of the linear

relationship between the two variables and how much variation there is around the regression line. For an increasing regression line the value should be as close as possible to 1 because of the positive correlation. A value closer to 0 will indicate no linear relationship between the variables [60].

2.5.3 T-test

The T-test (Student T-test) is used to study the differences in means or medians in a group of observations. There are two types of this test: paired T-test and T-test. The T-test can be used to conclude on observations in one group or in two groups. Two groups of data can be analyzed as one (paired T-test) but the data must then be related and the difference between observations must be calculated. An example where the paired T-test can be used is when measuring the blood pressure in one patient before and after treatment. The T-test can be used when comparing the same parameter in two different groups.

The paired T-test has been chosen in this thesis when comparing the activity of SNP309 G/G and T/T constructs transiently transfected into human cell lines. This choice is based on that the constructs are alike except for that one SNP, that the experiment conditions are otherwise the same and that a cell culture is derived from one larger culture and thus all cultures are assumed to be identical.

2.5.4 ANOVA (Analysis of variance)

ANOVA, which is a parametrical statistical test, is an alternative regression analysis where the dependent variable is continuous, like in regression analysis, and the explanatory variable is categorical. In these situations one can choose between regression analysis and ANOVA. ANOVA is used when there are more groups than 2 for comparison and the T-test cannot be used. The data needs to be normally distributed, independent and equal variance in the groups. The principle is to use the variance instead of standard deviation to analyze a sample group's variability [60].

2.5.5 Logistic regression analysis (SNP association study)

In a case-control study the higher the number of samples the more statistical conclusions one can make based on the test results and they will be more accurate. The confidence interval calculated from the observations in the study will give a measure of the accuracy of the mean calculated. This interval will, with a certain degree of accuracy, show the lowest and highest value that the true population might have. In a risk-

association one can study, for example, a genotype distribution in both healthy controls and disease cases, possibly identifying a risk allele. The control group is often matched to the case group to ensure that the two are as similar as possible. This is especially important in consideration of risk factors for the disease or other outcome investigated. Risk factors can be age, smoking habits, gender, diet etc [60].

Logistic regression is a model where the probability of an incidence is found by plotting the data to a logistic curve. The variables may be either categorical or numerical. This test is often used when testing for associations between various genotypes and disease. The genotype distribution in controls should be in Hardy-Weinberg equilibrium. This equilibrium is based on that various genotypes in a randomly mating population will have a certain distribution. This is not so important for cases since we then are selection certain individuals from the population. When using logistic regression one can include correction for confounding factors such as age, smoking, gender etc to obtain a more accurate modulation of the possible risk that will accompany the factor one is analyzing [60].

3. Results

3.1 Construction of *MDM2* SNP309-T/G-Luc expression vectors

3.1.1 Cloning

In order to insert the *MDM2* P2 promoter, with a G or T at the SNP309 position, into the reporter vector pGL3 basic, primers specific for the *MDM2* P2 region were used. The forward primer contained a restriction site for the endonuclease MluI and the reverse primer contained a restriction site for BglII. DNA from 3 individuals, harboring either the G/G or T/T genotype, was used as template. The PCR product would have the expected size of 471 bp which was confirmed by agarose gel electrophoresis (figure 3-1). Ladders are shown in appendix I.

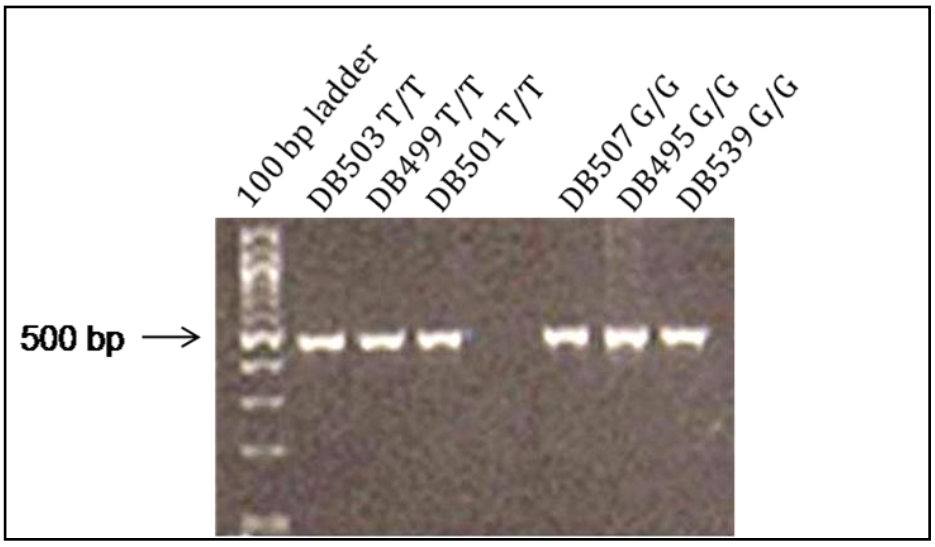


Figure 3-1: MDM2 P2 PCR products of DNA samples shown with a 100 bp ladder. Samples DB503, DB 499 and DB 501 contained the T/T SNP309 genotype. Samples DB507, DB495 and DB539 contained the G/G SNP309 genotype. All samples had the expected size of approximately 500 bp.

Both the PCR products and the pGL3 basic vector were double digested with MluI/BglII and ligated together. The finished expression vector comprising of pGL3 basic and the inserted *MDM2* P2 promoter region is shown in figure 3-2.

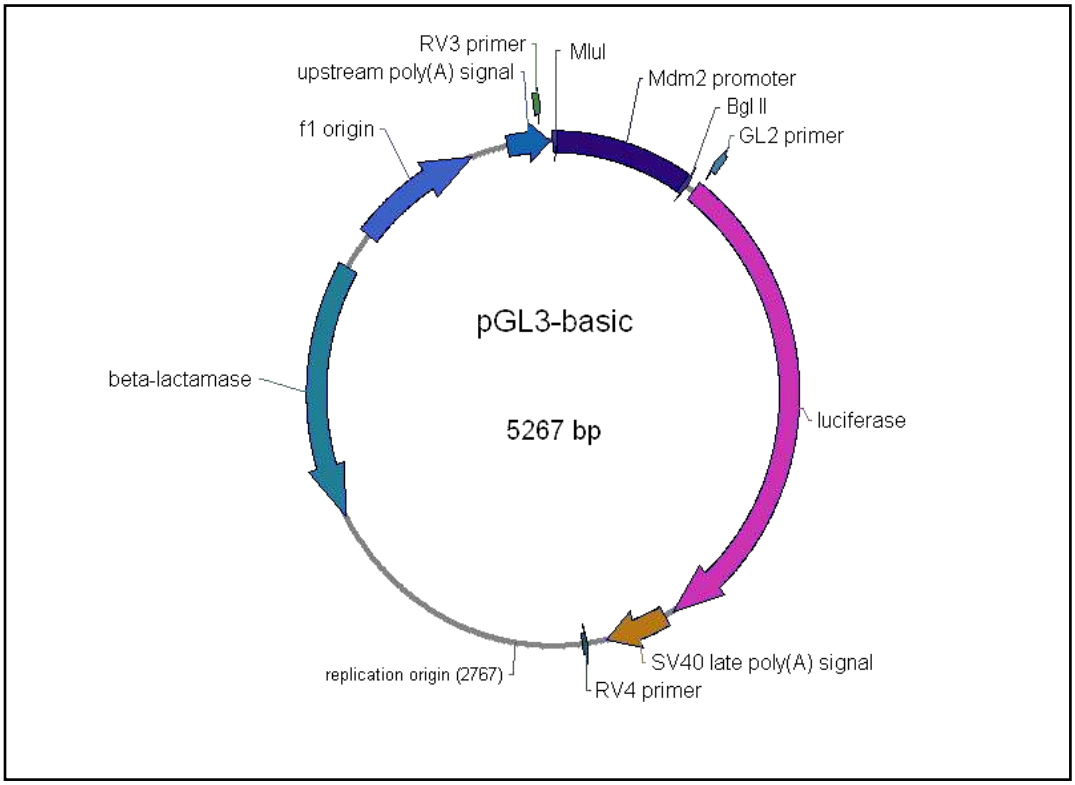


Figure 3-2: Vector map of the SNP309-T/G-Luc construct with the MDM2 P2 promoter inserted upstream for the luciferase gene (reporter). The forward primer RV3 and the reverse primer GL2 are marked.

The SNP309-T/G-Luc expression vectors were transformed into competent *E.coli.* After transformation, Amp^R (Ampicillin resistant) colonies were checked with a control PCR containing primers RV3 and GL2 which bind in the vector sequence. Primer locations are shown in figure 3-3. Colonies containing the expected 613 bp fragment were sequenced to verify SNP309 genotype and to ensure integrity of the DNA sequence. Sequencing results from an *E.coli* colony containing either SNP309-G/G-Luc or SNP309-T/T-Luc, respectively, is shown in figure 3-3. Complementary nucleotides were detected since the forward primer was used and synthesizes the complementary strand. See appendix II for more details.

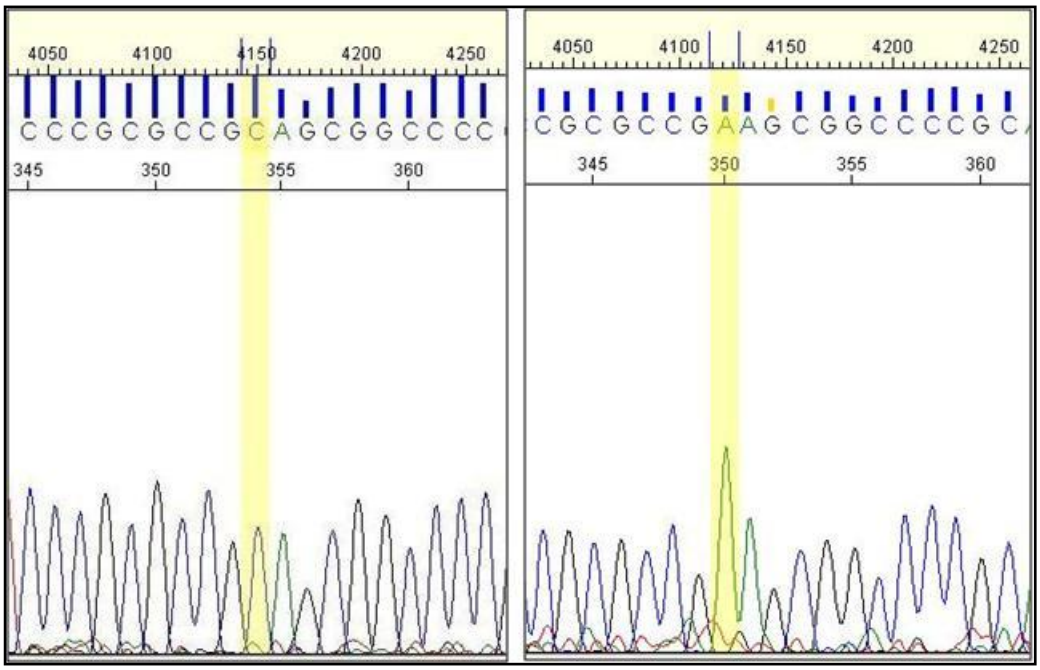


Figure 3-3: Sequencing results of SNP309-T/G-Luc constructs, G and T respectively. Reverse primer used and therefore complementary nucleotides are shown.

Two colonies containing the correct SNP309-Luc construct for either G/G or T/T genotype were selected, amplified and the plasmids were isolated by Maxiprep. A new round of PCR and restriction digestion was performed to further verify the correct plasmid contents before use in cell culture experiments.

3.2 Human cell culture experiments

The SNP309 constructs were transiently co-transfected with pRL-TK (Renilla) into human lung epithelial cell lines and later also epithelial cell lines from cervix, breast and colon (shown in table 3-1 and 3-2).

For each cell line at least three individual experiments were performed. In each of these experiments three replicates were used. The mean and standard error (SE) from the replicates was used in the paired t-test to test for differences transcriptional activity. The results are shown as means plotted with SE in histograms. Asterix (*) indicate statistical significance $p < 0,05$ with * and $p < 0,01$ with **.

To elucidate if SNP309 was estrogen responsive four lung epithelial cell lines were exposed to 17β-estradiol after transient transfection. A comparison of unexposed (0 nM) and exposed cells (1 nM and 10 nM 17β-estradiol) was performed with the ANOVA test.

All raw data from transient transfection are shown in appendix III.

3.2.1 Transfection results from human lung cell lines

The SNP309-T/G-Luc constructs were transiently transfected in both immortalized normal and cancerous human lung cell lines to study SNP309 effect on the *MDM2* P2 promoter transcriptional activity in cell culture systems. The lung cell line details are shown in table 3-1.

Table 3-1: Lung epithelial cell lines used in transient transfection.

Cell line	Type	Tissue	Gender
A549	Carcinoma	Lung	Male
BEAS-2B	Normal, SV40 immortalized	Lung	Male
BEP-2D	Normal, HPV immortalized	Lung	Male
NCI-H2009	Adenocarcinoma	Lung	Female
NCI-H460	Large cell	Lung	Male

Cell lines were co-transfected with either SNP309-T/T-Luc or SNP309-G/G-Luc together with pRL-TK. The relative transcriptional activity was calculated (Firefly/Renilla ratio) and a paired t-test was performed. In figure 3-4 A-E the results from all transfected human lung cell lines are shown.

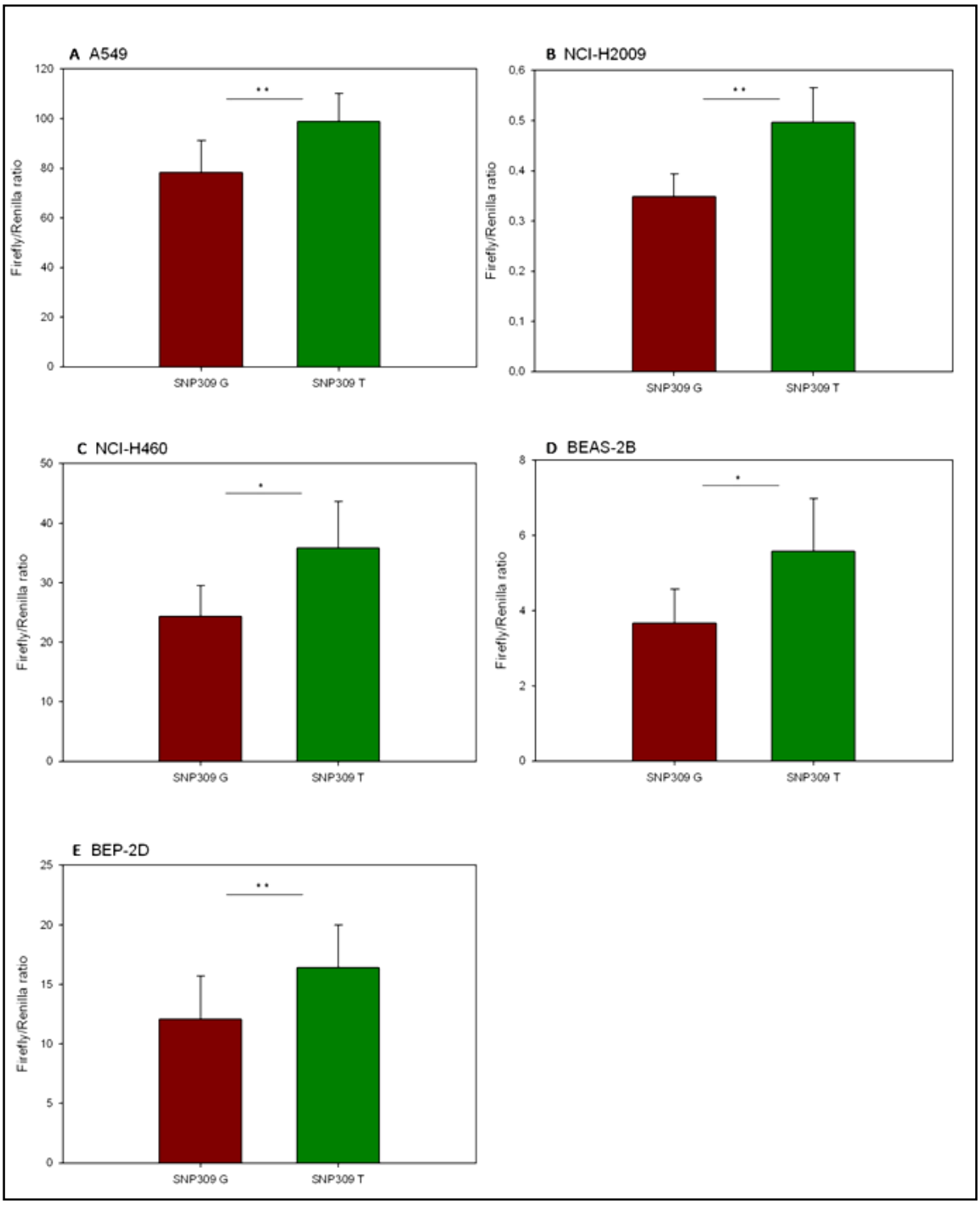


Figure 3-4 A-E: Relative transcriptional activity from SNP309-T/T-Luc (green) and SNP309-G/G-Luc (red) vectors. Statistical analysis performed with the paired t-test. Error bars = SE.

All lung cell lines, both immortalized and cancerous, showed increased transcriptional activity with the SNP309-T/T-Luc expression vector. NCI-H2009, the only female lung cell line, showed inverted ratios with the Renilla activity being higher than the Firefly activity. A549 was the cell line showing the most transcription from the both vectors whereas the

immortalized cell lines BEAS-2B and BEP-2D had the lowest. The different cell lines had various ratios indicating diverse inter cell line luciferase construct transcription.

To investigate if the higher transcriptional activity from SNP309-T/T-Luc was tissue specific, 3 additional cell lines originating from other tissues were transfected. These cell lines are listed in table 3-2.

Table 3-2: Non-lung cell lines used in transient transfection.

Cell line	Type	Tissue	Gender
HeLa	Carcinoma	Cervix	Female
MCF-7	Adenocarinoma	Breast	Female
HCT-116	Carcinoma	Colon	Male

These cell lines were treated the same way as the lung cell lines from table 3-1. In figure 3-5 A-C the results from all transfected human lung cell lines are shown.

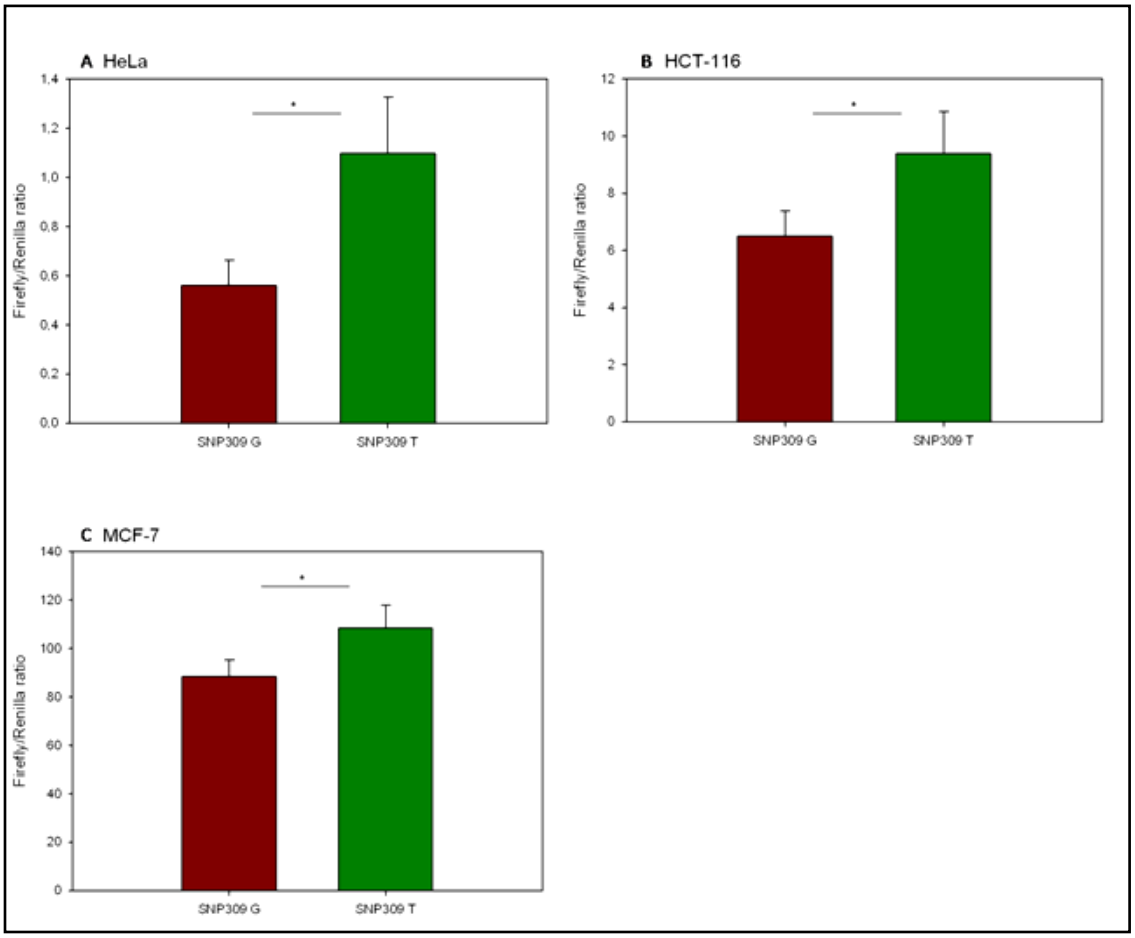


Figure 3-5 A-C: Relative transcriptional activity from SNP309-T/T-Luc (green) and SNP309-G/G-Luc (red) vectors in cell lines originating from cervix, colon and breast. Statistical analysis performed with the paired t-test. Error bars = SE.

Also in these cell lines the result was increased transcriptional activity from the SNP309-T/T-Luc vector. MCF-7 was the cell line inducing the highest transcriptional activity from the SNP309-T/G-Luc constructs. HeLa (also a female cell line) showed inverted ratios with Renilla activity higher than Firefly activity for SNP309-G/G-Luc. Again the different cell lines had various ratios indicating diverse inter cell line luciferase construct transcription.

3.2.2 Transient transfection with subsequent 17β-estradiol exposure

There are reports that *MDM2* is estrogen responsive (see chapter 1.9.5).Therefore to characterize possible estrogen responsiveness for SNP309, transiently transfected cells were exposed to 17β-estradiol for 24 hours. For these experiments the cell lines A549, NCI-H2009, BEAS-2B and BEP-2B, were chosen. Cell lines were co-transfected with SNP309-T/T-Luc or SNP309-G/G-Luc with pRL-TK with following exposure to 17β-estradiol (1 nM and 10 nM) for 24 hours. Relative transcriptional activity was calculated and exposed replicates were compared with an unexposed control. The results are presented in figure 3-6.

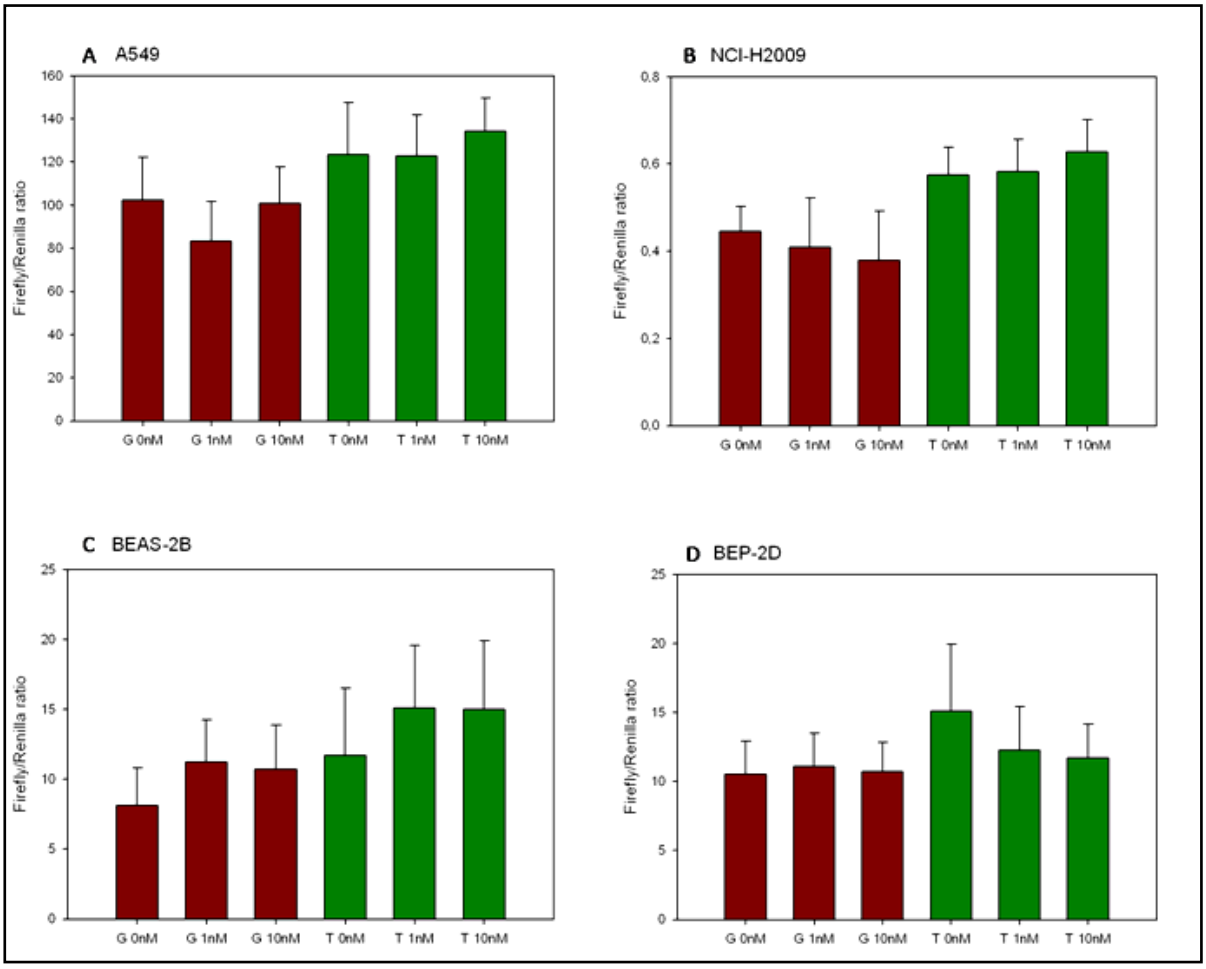


Figure 3-6 A-D: Relative transcriptional activity from SNP309-T/T-Luc and SNP309-G/G-Luc with increasing concentrations of 17β-estradiol. Statistical analysis performed with ANOVA. Error bars = SE.

In figure 3-7 it is evident that the SNP309-T/T-Luc construct still has a higher transcriptional activity. In figure 3-7 A and B the A549 and NCI-H2009 tumor cell lines show a slightly increased transcriptional activity for the T/T genotype with 10 nM of 17β-estradiol present. NCI-H2009, a female cell line, shows a trend towards decreased transcriptional activity with the G/G genotype with 10 nM 17β-estradiol present in contrast to A549. Both immortalized cell lines (figure 3-7 C and D) show no specific trend after 17β-estradiol exposure. For the cell line BEAS-2B exposed cultures have a higher ratio than non-exposed cultures. BEP-2D seems have the same transcriptional activity from both constructs. None of the results shown in figure 3-7 were statistical significant with the ANOVA test. In table 3-3 ER status and endogenous SNP309 genotype status is shown for the cell lines exposed to 17β-estradiol.

Table 3-3: Overview of ER status in cell lines exposed to 17β-estradiol.

Cell line	ER alpha	ER beta
A459	Negative	Positive
NCI-H2009	Negative	Positive
BEAS-2B	Partly positive	Positive
BEP-2D	Partly positive	Positive

In tables 3-4 and 3-5 a summary of the cell culture experiments are shown with P-values and more transcriptionally active genotype.

Table 3-4: Summary of transient transfection experiments.

Cell line	More active genotype	P-value
A549	T/T	0,008
BEAS-2B	T/T	0,0395
BEP-2D	T/T	0,015
NCI-H2009	T/T	0,002
NCI-H460	T/T	0,047
HeLa	T/T	0,025
MCF-7	T/T	0,05
HCT-116	T/T	0,016

Table 3-5: Summary of 17β-estradiol exposure experiments.

Cell line	Exposed to 17β-estradiol	P-value G/G	P-value T/T
A549	T/T dominant, indication for increased promoter activity for both genotypes.	0,72	0.90
BEAS-2B	T/T dominant, indication for increased promoter activity for both genotypes.	0,78	0.87
BEP-2D	T/T slightly dominant, non-responsive towards exposure.	0,98	0,78
NCI-H2009	T/T dominant, indication for increased promoter activity for T/T and decreased activity for G/G when exposed.	0,9	0,85

3.3 Verification of SNP genotypes

Biobank samples were genotypes for both SNP309 and C1797G. Genotyping and SNP association studies with SNP309 was done before

this project was started [57]. To characterize the endogenous SNP309 genotype status the all cell lines were genotyped as shown in chapter 2.4.1.

3.3.1 Genotyping of SNP309 with RFLP

For SNP309, the cell lines used were genotyped with RFLP. When digesting the PCR fragment (from *MDM2* specific primers) with MspA1I the T/T genotype would give 233 + 113 + 93 + 31 bp fragments and the G/G genotype 187 + 113 + 93 + 46 + 31 bp fragments, respectively. Samples with the T/G genotype would have 233 + 187 + 113 + 93 + 46 and 31 bp fragments. In figure 3-7 an agarose gel picture of PCR products digested with MspA1I is shown with some of the samples genotyped.

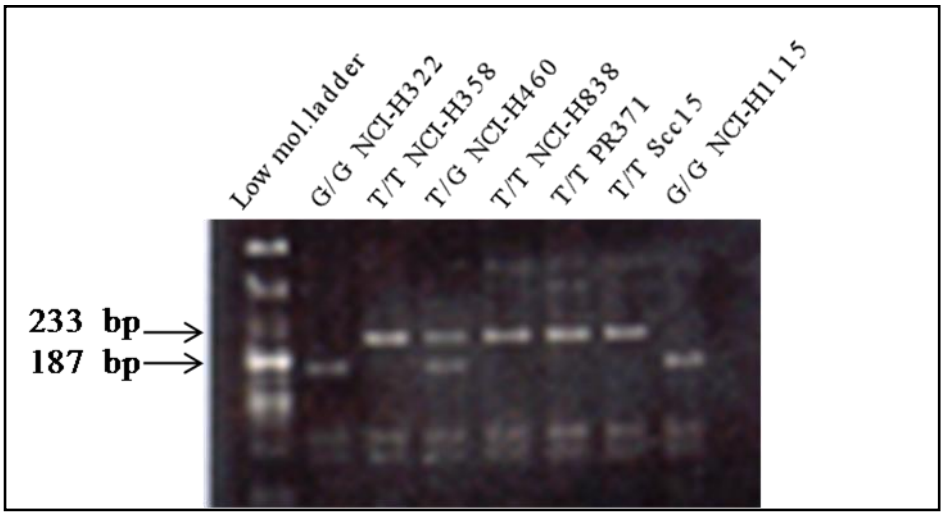


Figure 3-7: Genotyping of SNP309 with RFLP in cell lines. The two fragments 233 bp and 187 bp (indicated by arrows) distinguish the different genotypes.

The endogenous SNP3009 genotype status in all cell lines is shown in table 3-6.

Table 3-6: SNP309 genotype in cell lines.

Cell line	Genotype	Cell line	Genotype	Cell line	Genotype	Cell line	Genotype	Cell line	Genotype
A427	T/T	BL1672	T/T	DLD-1	T/G	H460	T/G	MCF-7	T/G
A459	T/T	BL193	T/T	H1155	G/G	H838	T/T	MDA-231	T/G
BEAS-2B	T/G	BL2009	T/G	H2009	T/G	HCT-116	T/T	PR371	T/T
BEP-2D	T/G	Calu-1	T/T	H322	G/G	HeLa	T/T	Scc15	T/T
BL15	T/T	DB354	T/T	H358	T/T	LOVO	G/G	THP-1	G/G

3.3.2 C1797G genotyping of lung cancer paitens and controls

The Biobank contains DNA samples isolated from lung cancer patients as well as controls which has been isolated from blood or normal adjacent

lung tissue. The samples used here are matched on age, smoking and gender. Samples were genotyped for SNP C1797G located to the P1 promoter in *MDM2* with the TaqMan method. This assay consists of two probes specific for each allele of the SNP. The two probes are labeled with different reporter fluorophores (VIC and FAM) that can be distinguished in solution. The TaqMan instrument reads of the amount of the two markers in each well and presents the results as a text file and a diagram. In figure 3-8 an example of a TaqMan diagram from C1797G genotyping is shown.

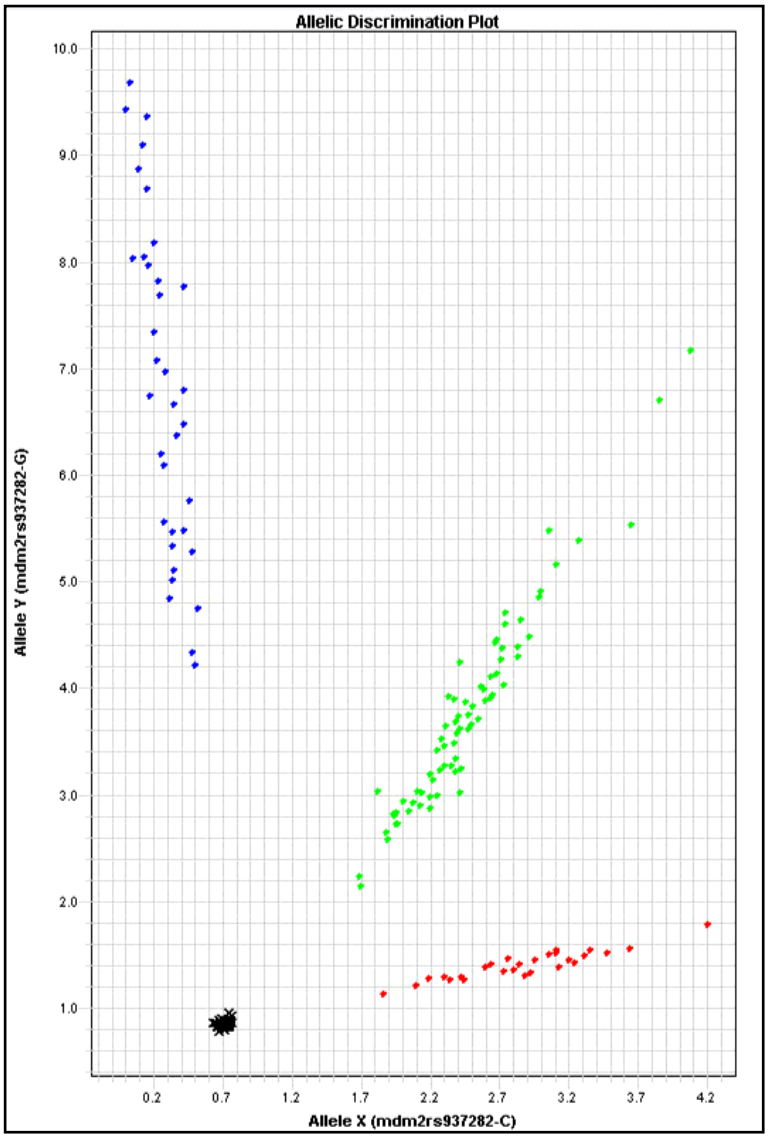


Figure 3-8: Example of TaqMan genotyping diagram output. The G allele signal is shown on the Y axis and the C allele signal on the X axis. Blue = G/G, red = C/C and green = C/G. Black = non-DNA samples. The axis labels are related to signal strengths emitted from the different wells.

A 10 % repetition was performed and all samples that were not genotyped by this method were further genotyped with the RFLP method.

3.3.3 Genotyping of C1797G with RFLP

To verify TaqMan data some samples were genotyped with RFLP. This was also done for samples not successfully genotyped with TaqMan. PCR was run on the samples using specific C1797G primers to produce a 540 bp fragment. This fragment was digested with the restriction endonuclease Hpy188I and the fragments made were separated on an agarose gel. The C/C genotype does not have a HPY188I restriction site and only shows one fragment. The G/G genotype gives rise to 2 fragments of 402 and 122 bp. In figure 3-9 an example of an agarose gel with different C1797G genotypes is shown.

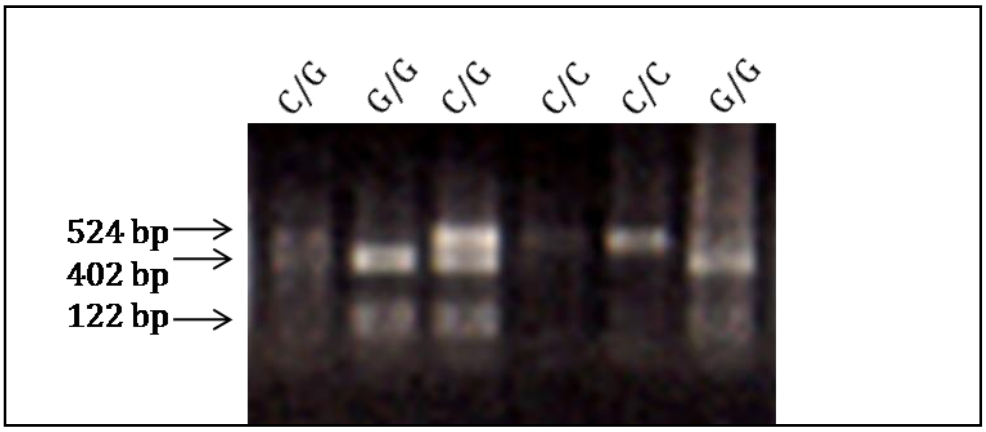


Figure 3-9: RFLP of C1797G. Arrows indicating bands to distinguish between the different genotypes at 524 bp, 402 bp and 122 bp, respectively.

3.3.4 Risk association for lung cancer with SNP C1797G

To elucidate any risk association between C1797G and lung cancer a series of lung cancer cases and matched controls were genotyped for SNP C1797G (chapter 2.4, 3.3.2 and 3.3.3). The risk associated for ach genotype was calculated with a logistic regression analysis. Characteristic of lung cancer cases and healthy controls are shown in table 3-9. The frequency of genotypes in the populations and the results from the logistic regression are shown in table 3-10.

Table 3-7: Characteristics of lung cancer cases and controls, all smokers.

Parameter	**Cases** N = 440	**Controls** N = 442
Median age (min - max)	66 (31 - 85)	60 (50 - 83)
Sex (male/female)	281/84	335/105
Number of cigarettes per day		
Mean ± SD	14,72 ± 8,71	14,59 ± 6,29
Median (min-max)	14,00 (2-60)	15,00 (3-40)
Total smoking years		
Mean ± SD	39,53 ± 14,12	42,3 ± 8,45
Median (min-max)	40,50 (2-69)	41,00 (15-65)
Total pack years		
Mean ± SD	29,78 ± 18,49	31,18 ± 15,07
Median (min-max)	27,00 (1-113)	28,50, (5-84)

Table 3-8: Association of C1797G polymorphism with lung cancer. ORs are adjusted for smoking, age and sex.

Gene/polymorphism	**Controls**	**NSCLC**	**Odds ratio (95 % CI)**
C1797G (rs937282)			
Total			
G/G	88	87	1,0
G/C	249	152	0,62 (0,43 - 0,90) P = 0,01
C/C	95	97	1,02 (0,68 - 1,55) P = 0,92
Men			
G/G	71	66	1,0
G/C	178	114	0,69 (0,46 - 1,03) P = 0,07
C/C	79	76	1,00 (0,63 - 1,59) P = 0,10
Women			
G/G	17	21	1,0
G/C	71	38	0,44 (0,20 - 0,96) P = 0,04
C/C	16	21	1,28 (0,49 - 3,35) P = 0,61

In this population the C/C genotype is more frequent and the G/G genotype rare. There is an overall protective effect of the heterozygous genotype C/G towards lung cancer in smokers. This effect is not significant in men, but significant in women.

3.4 Bioinformatic analysis for Putative transcription factor binding sites for SNP309 and C1797G

For both SNPs putative transcription factor binding sites were elucidated in databases PROMO/Alggen and AliBaba [61,62]. This was done to find any possible transcription factor binding sites in the SNPs sequence region. The function of the different transcription factors are described in appendix IV.

Since Bond *et al* had used AliBaba this database was also used. All putative transcription factor binding sites from the PROMO and AliBaba databases are shown in figure 3-10.

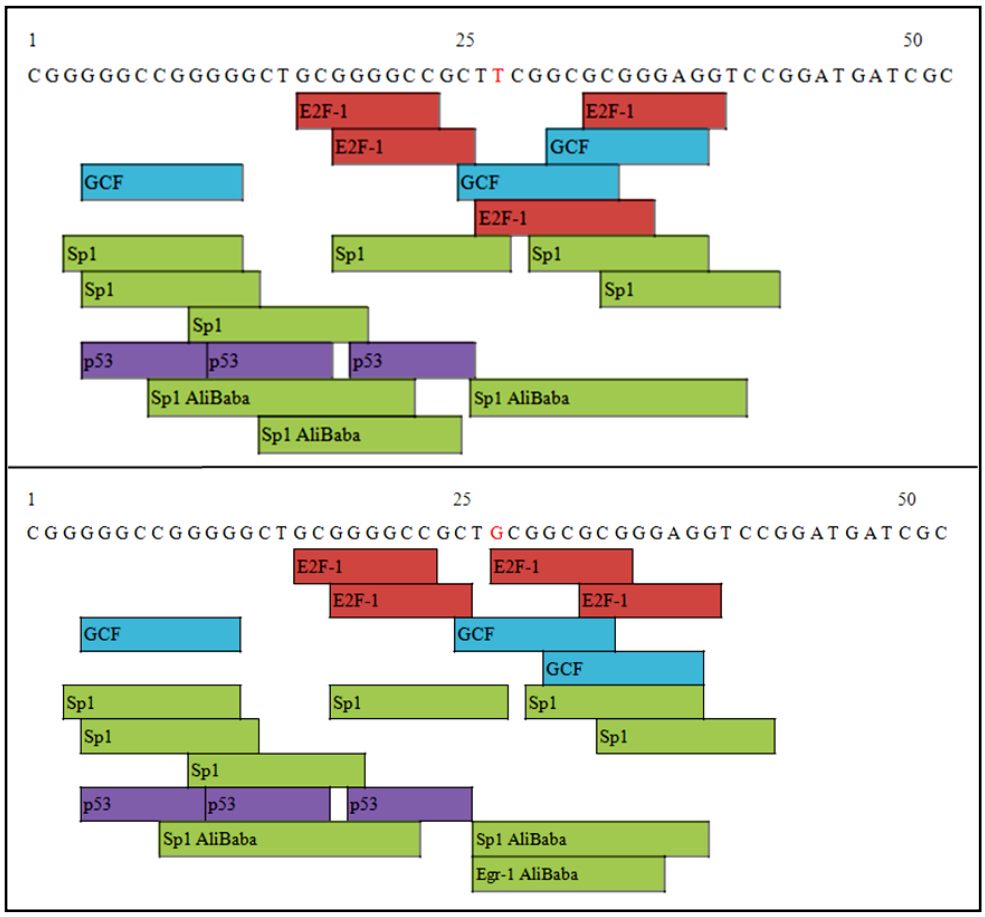


Figure 3-10: All putative transcription factor binding sites for the SNP309 region. Those from AliBaba are marked. SNP309 shown in red.

SNP309 sites found in PROMO and their dissimilarities are shown in table 3-7.

Table 3-9: Putative transcription factor binding to SNP309 region. Factors are showed with dissimilarities from the original binding sites and current binding sites.

Genotype	Transcription factor	Dissimilarity	Binding site
T/T	GCF	6,99 %	CTTCGGCGC
	Sp1	19,66 %	GGGGCCGCTT
G/G	GCF	6,99 %	CTTCGGCGC
	E2F-1	12,32 %	GCGGCGCG
	Sp1	19,43 %	GGGGCCGCTG

The difference between the two SNP309 genotypes is the addition of the putative Egr-1 binding site over SNP309 G/G. G/G also has a slightly smaller dissimilarity for Sp1 binding. T/T elongates the binding site for E2F-1. Also for T/T one Sp1 binding site from AliBaba is extended by 2 nucleotides downstream from the SNP in addition to obtaining a third Sp1 binding site.

SNP C1797G was also analyzed in the PROMO database. Since Wang *et al* had used AliBaba this database was also included. All putative transcription factor binding sites from both databases are shown in figure 3-11.

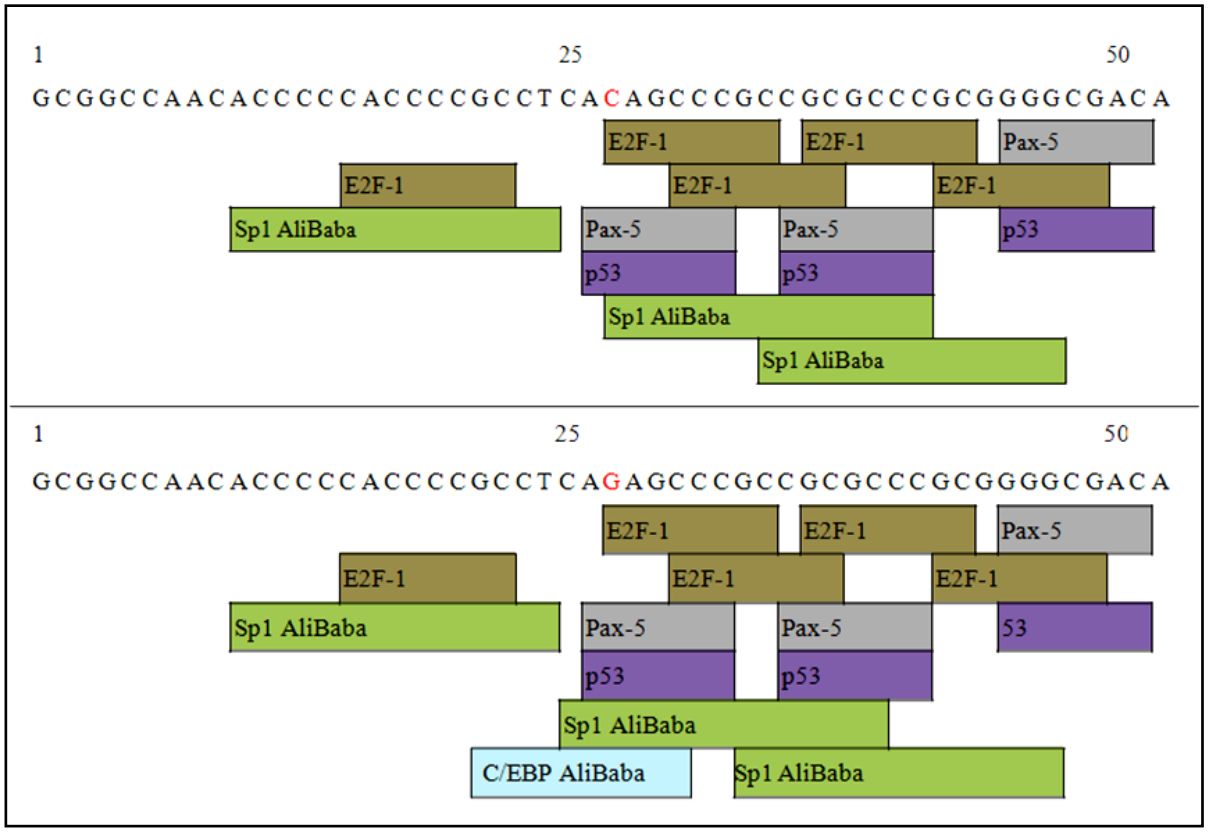


Figure 3-11: All putative transcription factor binding sites in the SNP C1797G region. Those from AliBaba are marked. SNP marked in red.

For SNP C1797G sites found in PROMO are shown in table 3-8 with their respective dissimilarities.

Table 3-10: Putative transcription factor binding sites from PROMO binding to C1797G. Factors are shown with dissimilarities and binding sites.

Genotype	Transcription factor	Dissimilarity	Binding site
C/C	E2F-1	11,89 %	CAGCCCGC
	Pax5	4,01 %	ACAGCCC
	p53	7,27 %	ACAGCCC
G/G	E2F-1	11,89 %	GAGCCCGC
	Pax5	4,01 %	AGAGCCC
	p53	8,54 %	AGAGCCC

The two C1797G genotypes differ in the putative binding of the C/EBP transcription factor. The G/G genotype has C/EBP binding and an extended Sp1 binding site. C/C has a slightly smaller dissimilarity for p53 binding.

4. Discussion

MDM2 SNP309 has been associated with several cancer forms, the results are ambiguous and vary with cancer type and patient/control ethnicity [56,63-72]. The first article describing a lung cancer - SNP309 association was published in 2006 by Hu *et al* where the authors showed no association with increased lung cancer risk in a Chinese population [73]. Later Lind *et al* published a paper where they showed an association between NSCLC and SNP309. When they analyzed for gender a strong association in women was discovered [57]. Two meta-analyses have shown risk for lung cancer with SNP309. Also, a stronger association was seen for the Asian population [56,74].

In this thesis the aim of the study was to further elucidate the results from Lind *et al* with construction of SNP309-T/G-Luc vectors to investigate the relative transcriptional activity with T/T or G/G genotypes.

4.1 Functional studies of SNP309

Bond *et al* (who discovered SNP309) transfected the HeLa cell line with their SNP309 constructs and observed that the G/G construct was more transcriptionally active than T/T [54]. This goes against the results obtained in this thesis for the HeLa cell line. Here the T/T genotype leads to a more transcriptionally active *Firefly luciferase* gene. This more active T/T genotype trend was seen in all transfection experiments in lung cell lines (figure 3-4 A-E) and all were statistically significant. To rule out any tissue specific increase in transcription activity with T/T in lung cell lines, 3 cell lines origination from cervix, breast and colon were also transfected. These also showed a higher transcriptional activity with the T/T genotype (figure 3-5 A-C) and these results were statistically significant. This excludes tissue specific effects on the SNP309-T/G-Luc constructs. When comparing expression in the different cell lines to endogenous SNP309 genotype, gender and tissue origin there was no apparent pattern.

The difference between the methods used by Bond *et al* and the present study are different expression vectors (pGL3 vs. pGL2). Bond *et al* do not describe which of the pGL2 vectors that has been used, but when assuming the use of pGL2 basic the two vectors differs in containing a synthetic poly A signal or not. The manufacturer (Promega) has redesigned pGL2 and created the pGL3 to obtain higher expression through the inclusion of the poly A signal. The other difference between the two projects is the length of the SNP containing region inserted into the expression vectors. Bond *et al* has used a SNP309 sequence available

from NCBI (showed in appendix II, slightly modified and 51 bp) in contrast to the 471 bp segment used in this project. There is a SNP located ~20 bp upstream of SNP309 (ID rs7484572) which is included in the SNP309-T/G-Luc constructs used in this project, whereas Bond *et al* has excluded this SNP from their constructs. This neighboring SNP may affect the promoter activity and thus give rise to the different experimental results.

The results obtained from transient transfection in the present study are not comparable to any other studies. One study has indicated that the G/G genotype might be protective in women and another indicates that T/T gives longer patient survival [75,76].

4.2 The response of SNP309 genotypes to 17ß-estradiol

It has for some time been a focus on women and that they might have higher risk for lung cancer and other pulmonary diseases than men. Also there has been seen an increase in lung cancer incidences in women which might be related to a survival bias because of less cardiovascular disease in women compared to men [52,58,77].

Since women have been shown to have a higher lung cancer risk a possible role for estrogen was plausible. The aim of the study was to characterize SNP309s response to 17ß-estradiol in the form of changed SNP309-T/G-Luc transcription activity. After 24 hours exposure to 17ß-estradiol the results showed no significant effect (figure 3-6A-D). As in only transfected cells, the T/T genotype still lead to a higher transcriptional activity compared to G/G. There might be a trend towards increased transcriptional activity from both constructs in the presence of 10 nM 17ß-estradiol. The female cell line NCI-H2009 shows a slight decrease in transcriptional activity with the G/G genotype with 10nM 17ß-estradiol. The immortalized cell lines BEAS-2B and BEP-2D showed no specific change and the BEP-2D cell line showed no difference in transcriptional activity between the two genotypes. All cell lines express ERß, but A549 and NCI-H2009 are ERα negative whereas BEAS-2B and BEP-2D are partly positive [17].

Estrogen is a known inducer of cell proliferation and differentiation in the lung and its receptor is a transcription factor. It is therefore likely that estrogen is involved in lung tumorigenesis [58]. The presence of ERα has been associated with phase I enzymes such as CYP, CYP1B1 can metabolize Estradiol to a potentially mutagenic compound in addition to bioactivation of other carcinogens. Fasco *et al* has shown that ERα is more expressed in female lung tumors and in normal tissue than in male tissue

[78]. Dougherty *et al* showed that lung adenocarcinoma cell lines proliferate in response to Estradiol [16]. There are studies indicating that ERα is a more potent activator of estrogen responsive elements than its family member ERß. It has also been observed that ERα is a better activator of Sp1 driven genes and this might be the interaction that elevates the expression. Still it is unlikely that such few factors comprises the whole picture when investigating the regulation of transcriptional activity from a gene [55].

Both ERα and ERß physically interacts with the transcription factor Sp1[79]. It has been shown that MDM2 physically interacts with ERα. In an *in vitro* study it was observed that ERα and MDM2 form a complex which induces transcription from estrogen responsive genes [15]. There are evidences that suggest that MDM2 might be the factor involved in ERα turnover in the presence of Estradiol. ERα interacts with MDM2 in a complex with p53 and these two proteins are required for Estradiol dependent down-regulation of ERα [14]. In addition to this, the levels of p53 and MDM2 have been shown to increase with estrogen present. It has been suggested that when MDM2 is over-expressed it is involved in a ligand-independent turnover of ERα and that MDM2 ubiquitinates ERα [80].

Bond *et al* [81] has shown that SNP309 G/G women have about 13 years earlier age of onset for diffuse large B cell lymphoma (DLCBL). Estrogen has been proposed as the interacting factor since women who have been exposed to estrogens has altered risks for DLCBL in addition to that many DLCBLs are ER positive. Bond *et al* therefore investigated if pre-menopausal women had lower age of onset due to SNP309 effect on the estrogen pathway. They found that no T/T women were diagnosed with DLCBL before the age of 55 in contrast to G/G women where half of them already had a DLCBL diagnosis at the same age. The authors concluded that estrogen may be the factor that allows SNP309 to accelerate DLCBL tumor formation. This same effect has been documented for soft tissue sarcomas and breast carcinomas in women with a 14 year earlier age at diagnosis and that tumors were ER positive. In breast carcinomas only the patients with ER positive tumors showed an association with earlier age at diagnosis. Bond *et al* concluded that an active estrogen signaling pathway affects the G allele of SNP309 to accelerate tumor formation in women [81].

There might be an estrogen effect in the female cell line NCI-H2009, but this cannot be due to ERα since the cell line does not express this subtype of ER. The immortalized cell lines, which are partly positive for ERα show, a reduced difference (BEAS-2B) and no difference (BEP-2D) in

expression from the two constructs and this might be the effect of ERα on the SNP309-G/G-Luc promoter region.

4.3 Endogenous MDM2 levels

Bond *et al* investigated the endogenous levels of MDM2 in cell lines genotyped to be homozygous T/T or G/G and observed lower MDM2 levels in the T/T cell lines. They have genotyped the NCI-H460 cell line to be homozygous T/T. Genotyping data in this project contradicts this and show that this cell line is heterozygous (figure 3-7). To confirm the genotype, the cell line DNA was sequenced and the T/G genotype verified (data not shown). In the experiment where endogenous Sp1 is reduced by treatment with the RNA synthesis inhibitor Mithramycin A, the NCI-H460 cell line deviates from the other T/T cell line and shows little reduction in MDM2 levels. This may be due to the incorrect genotyping of this cell line [54].

4.4 Transcription factor binding sites and SNP309

When examining the sequence area around SNP309 Bond *et al* [54] found a putative Sp1 transcription factor binding site which in the presence of the G/G genotype was elongated by one nucleotide. This was found in the AliBaba transcription factor database. I was unable to replicate this finding in the PROMO/Transfac database, but observed a slightly decreased dissimilarity for binding with the G/G genotype present (figure 3-10). When trying the AliBaba database the putative Sp1 binding site was not extended by one nucleotide to include the G/G genotype of SNP309, but extended with 2 nucleotides in the other direction with the T/T genotype present. The AliBaba database revealed an additional binding site with G/G present for the transcription factor Egr-1. Bond *et al* verified their Sp1 binding site results with EMSA. This showed more binding of Sp1 to G/G both with purified Sp1 and protein extract from the cell line HeLa. They further verified this by ChiP and detected bound Sp1 to the P2 promoter in a SNP309 G/G cell line but this was not compared to the Sp1 amount bound in a T/T cell line.

Bond *et al* further wanted to elucidate Sp1 involvement in P2 activity. They used their luciferase expression vector with the SNP309 G/G or T/T. These were co-transfected with a Sp1 expression vector in a Sp1 deficient cell line. Both SNP309 constructs were stimulated with increasing amounts of Sp1, but G/G more than T/T. The SNP309 sequences used by Bond *et al* contained several Sp1 binding sites and even though they observed increased binding of Sp1 to SNP309 G/G this cannot be

attributed to an extended Sp1 binding site since this cannot be replicated in the AliBaba database. Furthermore the same database showed more binding sites for Sp1 with the T/T genotype. It should be mentioned that these transcription factor databases change continuously and the inability to replicate Bond *et al* findings of the elongated Sp1 site may be due to updates of these databases.

The neighboring SNP to SNP309 is a C/G transition. This SNP also has different putative transcription factor binding sites as shown in figure 4-1.

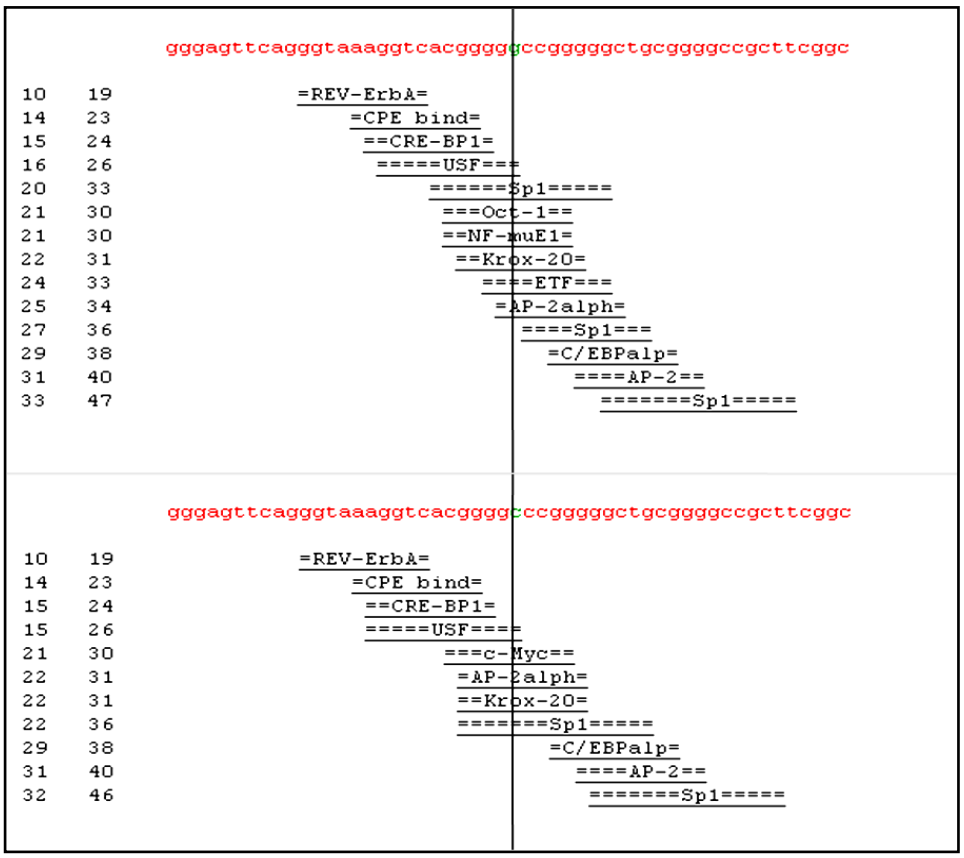


Figure 4-1: Transcription factor binding sites for SNP ~20 bp upstream of SNP309. From the AliBaba database.

With the C/C genotype the Sp1 binding site is extended, but for the G/G genotype several other transcription factor binding sites appear. Oct-1 and ETF are transcription factors not directly involved in growth and differentiation. Transcription factors c-Myc, AP-2alph and NF-muE1 are involved in growth and differentiation (appendix IV). In the constructs made in the present study this SNP has the G/G genotype (sequencing results appendix II). If these are linked (SNP309T/T to SNP G/G) the additional transcription factors might contribute to the elevated transcriptional activity from the SNP309-T/T-Luc construct.

4.5 Association of the C1797G SNP with lung cancer

Wang *et al* [53] recently identified a new SNP in MDM2 P1 which was shown to be functional. They also found an association between this SNP and risk of developing bladder cancer in a Chinese population. Since this SNP is also located in a *MDM2* promoter region it can possibly affect *MDM2* transcription. Therefore it was desirable to investigate its effect on lung cancer risk. The genotype distribution in this Norwegian population showed that the C/C and G/G genotypes were only slightly different with a higher frequency of C/C individuals. There was no pattern seen regarding difference in genotype distribution between cases and controls. According to the NCBI database the genotype distribution varies between the different populations (table 1-2). The G allele (rare) was shown by Wang *et al* to be associated with higher *MDM2* mRNA levels.

Individuals carrying the heterozygous genotype in the present study showed a low OR, especially in women and this was statistical significant. This might indicate a lung cancer protective genotype, particularly for smoking women.

Wang *et al* also showed that the C/EBP transcription factor binds the G/G genotype of C1797G. By bioinformatic analysis this finding could be replicated, but in addition an extended Sp1 binding site was discovered (figure 3-11). The putative binding site for p53 had a slight decreased dissimilarity when binding to the C/C genotype of C1797G. These putative transcription factor binding sites for G/G may explain the increased risk of cancer with an increased *MDM2* transcriptional activity which again leads to p53 attenuation

5. Conclusion and future work

The persistent higher SNP309-T/T-Luc activity is the opposite from the hypothesis stating that the expected was a more transcriptionally active G/G genotype. This can be the result of the neighboring SNP or from the additional E2F and elongated Sp1 factor binding site with the T/T genotype. The G/G genotype shows the estimation of another binding site for Egr-1 and shorter E2F-1/Sp1 binding sites. The two SNPs separated by ~20 bp are most likely linked and contribute together to regulate P2 activity. To characterize the individual SNPs effect on the MDM2 P2 promoter new constructs should be made. The results obtained here must be further elucidated with determination of *in vitro* transcription factor binding for example by performing EMSA to investigate transcription factor binding to the SNP309 region. Characterization of endogenous MDM2 levels compared to SNP309 genotype for cell lines located at NIOH (National Institute of Occupational Health) might also help to map the SNP309 effect on MDM2.

In regard to estrogen responsiveness there was no significant effect observed. To ensure that the lack of ERα in the female lung cancer cell line NCI-H2009 is not the reason that estrogen responsiveness was not detected additional female ERα positive cell lines needs to be studied. It also would be useful to investigate the possible Sp1-ERα-SNP309 interaction since Sp1 and ERα has been observed in complex and that both ERα and Sp1 interacts with *MDM2*.

The new C1797G SNP was discovered at the end of this project. Since this also is a functional SNP in the *MDM2* promoter region it was included in the project. There was not found an increased risk for lung cancer in this population, but the heterozygous genotype is likely to be protective in smoking women. Functional studies with this SNP needs to be performed.

Appendix I

Enzymes

Enzyme	Manufacturer	Buffer
HotFirePol DNA polymerase	Solis BioDyne	B2 10X reaction buffer
T4 DNA ligase	New England Biolabs	T4 DNA ligase buffer

Restriction enzymes

Enzyme	Manufacturer	Buffer
BglII	New England Biolabs	NEB buffer 3
Hpy188I	New England Biolabs	NEB buffer 4
MluI	New England Biolabs	NEB buffer 3
MspA1I	New England Biolabs	NEB buffer 4 / BSA

Bacteria

E.coli DH5α F'Iq from New England Biolabs with following gene specifications:

F' *proA⁺B⁺ lacI^q Δ(lacZ)M15 zzf::Tn10* (Tet^R)*/fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17.*

Cell lines

Cell line	Media	Tissue origin	Gender	Distributor
BL15, BL78, BL193, BL1672, BL2009.	RPMI 1640 w/ 5 % FBS	Blood lymphocytes?	Same person as Hcc (15, 78, 193) and H2009. BL1672 is male.	ATCC
A549	DMEM/F12 w/10 % FBS	Lung adeno	Male	ATCC
BEAS-2B	LHC-9 w/ 4 % BSA	Lung, SV40 immortalized	Male	ATCC
BEP-2D	LHC-9 w/ 4 % BSA	Lung, HPV immortalized	Male	ATCC
Hcc78 (Scc78), Hcc193 (Scc193)	HITES w/ 5 % FBS	Lung adeno	Male	ATCC

HCT116	McCoy w/ 10 % FBS	Colon	Male	ATCC
HeLa	DMEM/F12 w/ 10 % FBS	Cervix	Female	ATCC
NCI-H2009	HITES w/ 5 % FBS	Lung adeno	Female	ATCC
NCI-H460	RPMI 1640 w/ 5 % FBS	Lung large cell	Male	ATCC
MCF-7	RPMI 1640 ++ w/ 10 % FBS	Breast carcinoma	Female	ATCC
MDA-MB-231	RPMI 1640 ++ w/ 10 % FBS	Breast carcinoma	Female	ATCC
PC3	RPMI 1640 + w/ 10 % FBS	Prostate	Male	ATCC
THP-1	RPMI 1640 +++ w/ 10 % FBS	Acute monocytic leukemia	Male	ATCC

Kits

Applied Biosystems

Sequencing reaction components supplied in the BigDye® Terminator v1.1 Cycle Sequencing Kit.

Invitrogen

Quantification of dsDNA with Invitrogen's Quant-iT™ PicoGreen® dsDNA Reagent.

MilliPore:

Up-concentration of purified PCR fragments with Microcon® YM-30 Centrifugal Filter Units.

Qiagen:

PCR reaction cleanup with QIAquick nucleotide removal kit.

Plasmid isolation with QIAprep Spin Miniprep kit and QIAprep Endofree Maxiprep kit.

Removal of unincorporated ddNTPs with DyeEx 2.0 Spin kit

Instruments

Luminometers	Turner biosystems Modulus microplate.
	Packard Fusion TM plate analyzer.
Sequencing instrument	ABI prism 310 genetic analyzer
Spectrophotometer	Eppendorf biophotometer
TaqMan instrument	ABI prism 7900 HT sequence detection system
Thermal cyclers	Peltier Thermal Cycler DYAD™ DNA engine 96 wells.

	Peltier Thermal Gradient Cycler PTC-200 384 wells

Chemicals

Acetic acid (glacial)	Merck	Ham's F12	Gibco
Agarose (SeaKem GTG)	Lonza	Hydrocortisone	Sigma
Ampicillin sodium salt	Kebolab	HCl	Merck
Bacto™-agar	BD	HEPES	Sigma
Bacto™-tryptone	BD	Insulin	Sigma
Boric acid	Sigma	KCl	Merck
Bovine serum albumin	Sigma	Leibovitz's L15 media	Invitrogen
Bromophenol blue	Chroma-Gesellschaft	Ladders	New England Biolabs
Collagen	Vitrogen	LHC-9	Invitrogen
Dubecco's Modified Eagle Medium	Gibco	Lipofectamin reagent	Invitrogen
Dimethyl sulphoxide 99%	Sigma	McCoy 5A media	Invitrogen
dNTPs	Quanta	NaCl	Merck
EDTA (Triplex III)	Merck	NaHCO3	Merck
Etyleneglycol bis NNNN-tetra acetrate	Sigma	$NaHPO4 \cdot H_2O$	Merck
Epidermal Growth Factor	Sigma	NEB D1 transpass	New England Biolabs
Ethidium bromide	Biorad	Phenol red	Sigma
Ethanol (rectified, absolute)	Arcus	Plus reagent	Invitrogen
Fetal bovine serum	Gibco	Propane-2-ol (isopropanol)	Merck
Fugene transfection reagent	Roche	Penicillin-Streptomyocin	Gibco
Fungizone	Gibco	RPMI 1640	Gibco
Gentamicin	Invitrogen	Selenium (Na_2SeO_3)	Sigma
Glucose	Sigma	Sodium puryvate	Sigma
L-Glutamine	Sigma	Transferrin	Sigma
Glycerol 99,5 %	Aldrich	Tris-base 7-9®	Sigma
		Trypsin	Sigma
		Yeast extract	DIFCO

Solutions

All solutions are made with ddH_2O and sterile filtered before use where needed.

AF (antibiotic freeze media)

76 % L15 media 2 % 1M HEPES, 2 % PS, 20 % FBS.

BSA (bovine serum albumin) 1 mg/ml stock

100 mg BSA, 100 ml HBS.

DMSO (dimethyl sulphoxide) for cell culture storage

50 % L15 media, 2 % 1 M HEPES, 8 % DMSO, 40 % FBS. EDTA

EDTA 0,5 M disodium dihydrate (ethylenediaminetetraacetic acid) 50 ml

9,3 g EDTA, 50 ml H_2O, pH adjusted to 8.

EGF (epidermal growht factor) 10 µg/ml

200 µg EGF, 2 ml BSA solution, 18 ml HBS.

EGTA (etyleneglycol bis NNNN-tetra acetrate) 2 %

200 mg EGTA, 100 ml HBS.

FBS (fetal bovine serum)

Heat inactivated at 56 ºC for 45 minutes.

Glucose 200 mg/ml

20 g glucose, 100 ml H_2O.

Glutamine 200 mM

2,9 g glutamine, 100 ml H_2O.

HBS (HEPES buffered saline)

4,76 g HEPES, 7,07 g NaCl, 0,20 g KCl, 1,70 g glucose, 1,94 g $NaHPO4 \cdot H_2O$, 1,04 ml 0,12 % phenol red, H_2O to 1 l.

HC (hydrocortisone) 1,0 mM

7,2 mg HC, 20 ml rectified ethanol.

Insulin 4 mg/ml

100 mg insulin, 25 ml 4 mM HCl.

$NaHCO^3$ (carbonate) 7,5 %

7,5 g $NaHCO_3$, 100 ml H_2O.

PBS (phosphate buffered saline)

7,07 g NaCl, 0,20 g KCl, 1,94 g $NaHPO4 \cdot H_2O$, 1,04 ml 0,12 % phenol red, H_2O to 1 l.

PET (PVP, EGTA, Trypsin stock)

1 % PVP, 1 % EGTA, 0,2 % Trypsin stock, 7,8 % HBS.

Phenol red 125 mg/ml

125 phenol red, 360 µl 1N NaOH, 100 ml H_2O.

PVP (Polyvinyl pyrrolodion) 10 %

10 g PVP, 100 ml HBS.

Selenium (Na_2SeO_3) 0,1 M, 100 ml

1,73 g Na_2SeO_3, 100 ml H_2O.

TAE buffer (Tris-acetate-EDTA buffer) 50 X, 1L

242 g Tris base, 57,1 ml glacial acetic acid, 100 ml 0,5 M EDTA pH 8.

TBE buffer (Tris-borate-EDTA buffer) 5 X , 1L

54 g Tris base, 27,5 g boric acid, 20 ml 0,5 M EDTA pH 8.

TE buffer (Tris-EDTA buffer) 10 X, pH 8

100 med mer Tris-Cl pH 8, 10 med mer EDTA pH 8.

TF (transferrin) 5 mg/ml

500 mg transferrin, 10 ml BSA stock, 90 ml HBS.

Tris-Cl 1M pH 8, 1L

121,1 g Tris base, 42 ml HCl, H_2O to 1L.

Trypsin 1 %

50 mg trypsin, 5 ml HBS.

Collagen solution for petri dish coating 0,03 mg/ml

1 % Collagen solution 3,13 mg/ml, 99 % HBS.

Bacteria culture media and agar:

Luria-Bertani LB 1L

10 g tryptone, 5 g yeast extract, 10 g NaCl, H_2O to 1L. Autoclaved.

LB-agar

10 g tryptone, 5 g yeast extract, 10 g NaCl, 15 g Bactoagar, H_2O to 1L. Autoclaved.

LB-Ampicillin

LB-media and Ampicillin to the concentration of 100 µg/ml added after autoclavation and cooling to 50 °C.

SOC 1L

20 g tryptone, 5 g yeast extract, 0,5 g NaCl, 20 ml 1M glucose added after autoclavation.

Cell culture media

After 2 months, 5 ml of L-glutamine was added to the media.

DMEM/F12 w/ 10 % FBS

- 500 ml DMEM/F12
- 7,5 ml HEPES
- 625 µl Insulin
- 50 µl HC
- 500 µl EGF
- 145 µl Selenium
- 5 ml PS
- 500 µl TF
- 50 ml FBS (10 %)

HITES w/ 5 % FBS

- 500 ml RPMI 1640
- 25 ml FBS
- 5 ml PS
- 5 µl HC
- 625 µl Insulin
- 1 ml TF
- 300 µl Selenium
- 5 ml HEPES

RPMI 1640 w/ 5 % FBS

- 500 ml RPMI 1640
- 500 µl EGF
- 250 µl HC
- 500 µl Insulin
- 5 ml PS
- 25 ml FBS

RPMI 1640 + w/ 10 % FBS PC3

- 500 ml RPMI 1640
- 5 ml PS
- 50 ml FBS

RPMI 1640 ++ w/ 10 % FBS MCF-7

- 500 ml RPMI 1640
- 50 ml FBS
- 5 ml Natriumpyruvat
- 5 ml PS

RPMI 1640 +++ w / 10 % FBS THP-1

- 500 ml RPMI 1640
- 5,5 ml HEPES
- 55 mg Sodium pyruvate
- 550 µl Gentamicin
- 50 ml FBS

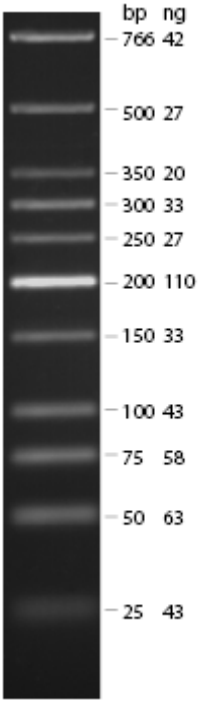
LHC-9

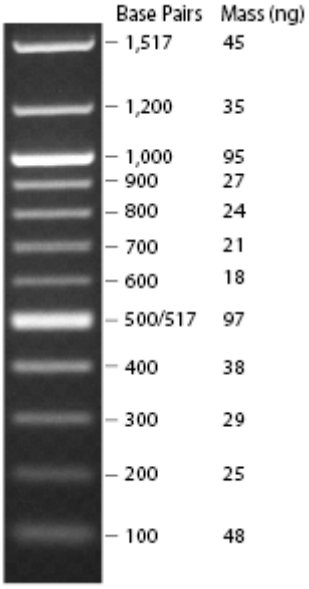
- 500 ml LHC-9
- 20 ml BSA
- (2,5 ml Fungizone)

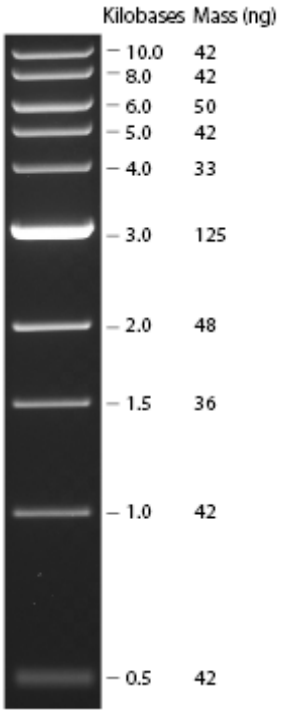
McCoy

- 500 ml McCoy
- 50 ml FBS

DNA ladders Low molecular, 100 bp and 1kb ladders, respectively.

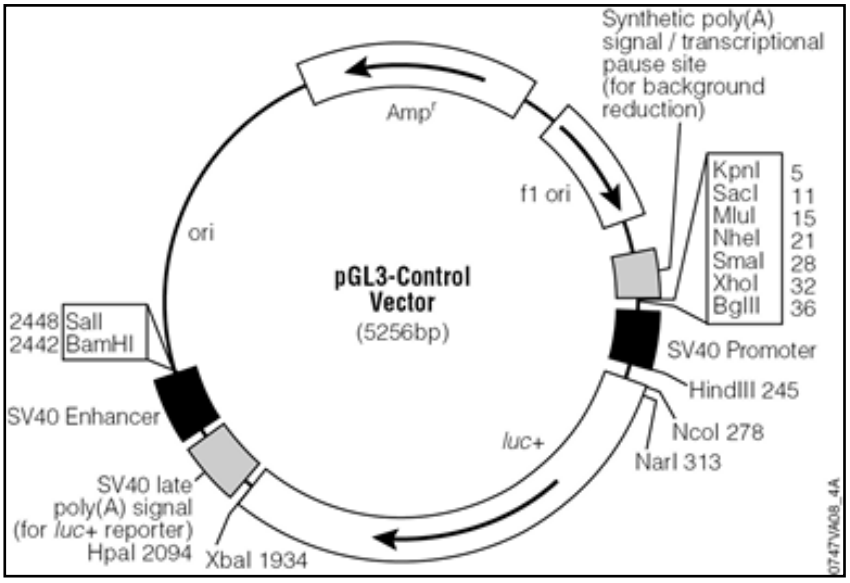


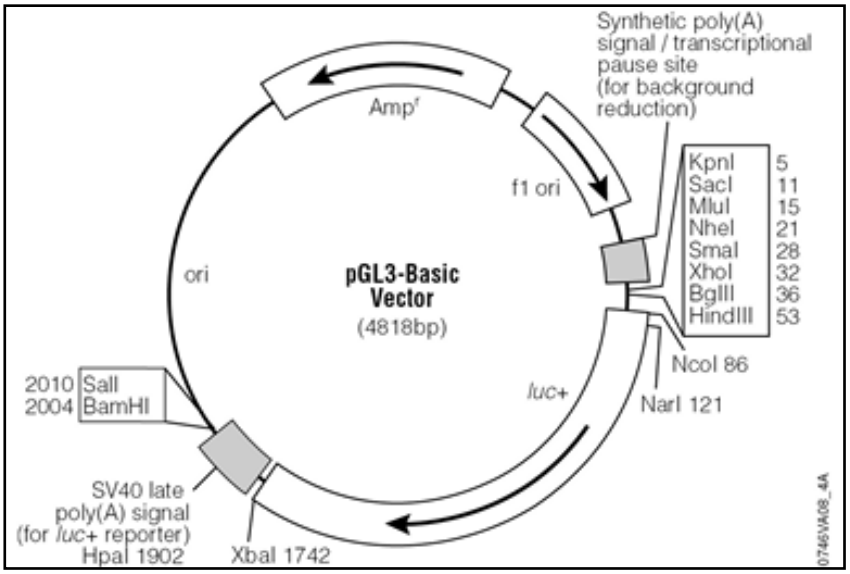


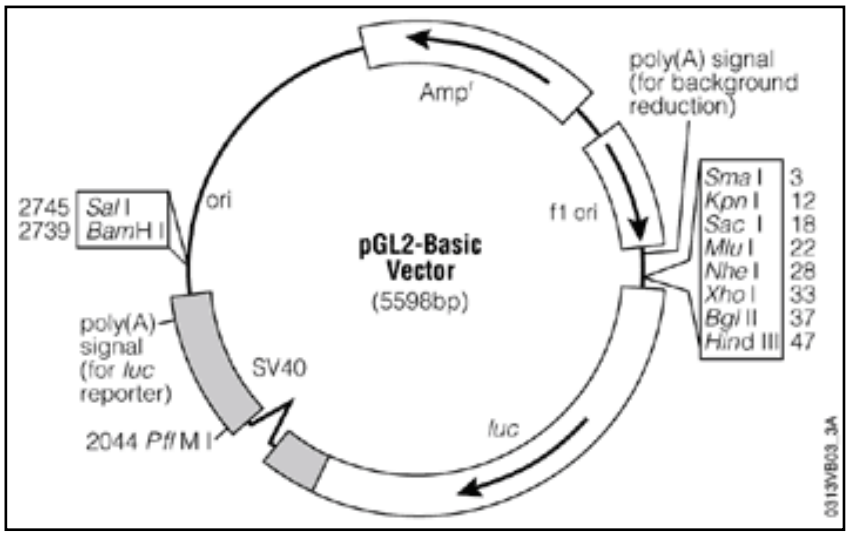


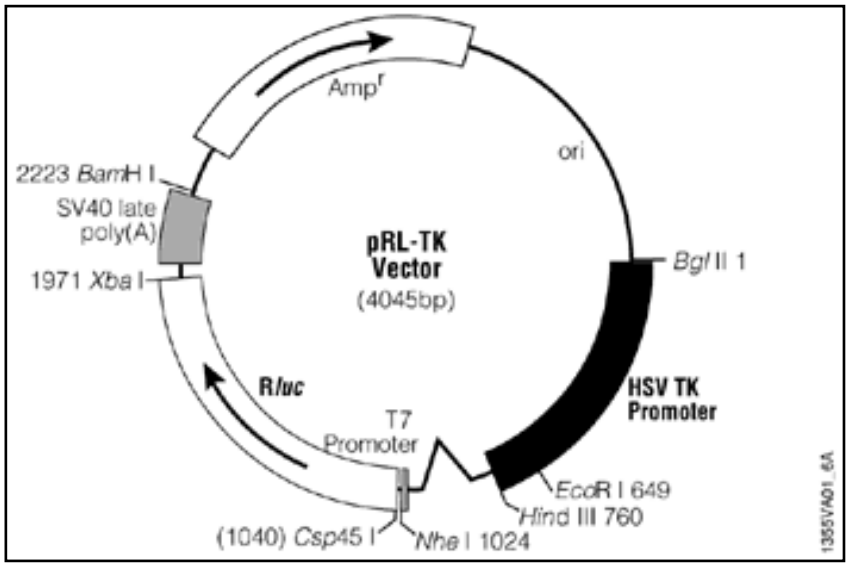
Appendix II

Vector maps









Primers

SNP309 forward: 5'-ACGCGTTGCGGAGGTTTTGTTGGAC-3'

SNP309 reverse: 5'-AGATCTAAGCAAGTCGGTGCTTACCTG-3'

RV3: 5´-CTAGCAAAATAGGCTGTCCC-3´

GL2: 5´-CTTTATGTTTTTGGCGTCTTCCA-3´

C1797G forward: 5-GCACGCGTCTAAAAGCGCAGAGTAACCG-3'

C1796G reverse: 5'-GCAGATCTCATCCGGGCATTTGTG-3'

SNP sequences:

SNP309

GTGTGCCTGTGTGCACTTGGGTCTGT[C/T]GGATGACATTGGGTCACTGTGTCTG

C1797G

GCGGCCAACACCCCCACCCCGCCTCA[C/G]AGCCCGCCGCGCCCGCGGGGCGACA

Neighboring SNP to SNP309

GGGAGTTCAGGGTAAAGGTCACGGGG[C/G]CCGGGGGCTGCGGGGCCGCTTCGGC

Appendix III

Sequencing

Primers are marked in pink/blue and SNP309 is marked in green. Neighbor SNP is marked in grey.

SNP309-G/G-Luc complement sequence.

AGGGSTMYAARGGTACCGGGAATGCCAAGCTTACTTAGATCGCAGATCTAAGCAAGTCGGT
GCTTACCTGGATCAGCWKAGAAAAAGTGGCGTGCATCCGTGCCCACAGGTCTACCCTCCAA
TCGCCACTGAACACAGCTGGGAAAATGCATGGTTTAAATAGCCCCAGCTGGAGACAAGTCA
GGACTTAACTCCTTTTACTGCAGTTTCGGAACGTGTCTGAACTTGACCAGCTCAAGAGGAAA
AGCTGAGTCAACCTGCCCACTGAACCGGCCCAATCCCGCCCAGACTACGCGCAGCGTTCAC
ACTAGTGACCCGACAGGCACCTGCGATCATCCGGACCTCCCGCGCCGCAGCGGCCCCGCAG
CCCCCGGCCCCCGTGACCTTTACCCTGAACTCCCGCGGAGACCTCCGAACCACCCCCACCC
CCACCGCCGCGAGAGCCGTCCGAAATCCCGCCCTCCTCCCTGGCGGCGACTGCCTAGCCCC
AGTCCAACAAAACCTCCGCAACGCGTAAGAGCTCGGTACCTATCGATAGAGAAATGTTCTG
GCACCTGCACTTGCACTGGGGACAGCCTATTTTGCTAGTTTGTTTTGTTTCGTTTTGTTTTGA
TGGAGAGCGTATGTTAGTACTATCGATTCACACAAAAAACCAACACACAGATGTAATGAAAA
TAAAGATATTTTATTGCGGCCGCTCCAAGTACCTCCCGTACCTTAATATTACTTACTTATCAT
GGTAGCTTGGGCTGGCGTAATARCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAKTTGC
GCAGCCTGAATGGCGAATGCAAATTGTAAGCGTTAATATTTTGTAAAATTCGCGTAAATTTT
TGTTAATCAGCTCATTTTTTAACCAATAGGCGAAATCGGSAAATCCCTTAWAATMAAAGAAT
AGMYYGAKATAGKGKTGAGTGTTGTTCMAGTTTGAACAAGAGTCCACTATTAAARAWSKG
AMTYCAACGTYAAACGGCGAAAAACGTYTATCWGGYGATGCCCWCTAYCTGACCMTYMCC
TCATYAMGTTTTTGGGGTCMRGKGKCTAATYACTAMACKGAAWCCTTAASGGASCCYCGW
YTCAGCTTKAKGGGWAATKCTGCTACTTGTGTGASAYGAAAGGAWAWAGTMWARGWACK
SGCTYAGGYKYTGCTWYKTMYKGACACCSCWCSWACWYMCMYAYCYCYCT

SNP309-T/T-Luc complement sequence.

TGGGYMWMAMRKTACCGGATGCCAGCTTACTTAGATCGCAGATCTAAGCAAGTCGGTGCT
TACCTGGATCMSWWGTKAAAAAGTGGCGTGCGTCCGTGCCCACAGGTCTACCCTCCAATC
GCCACTGAACACAGCTGGGAAAATGCATGGTTTAAATAGCCCCAGCTGGAGACAAGTCAGG
ACTTAACTCCTTTTACTGCAKTTTCGGAACGTGTCTGAACTTGACCAGCTCAAGAGGAAAAG
CTGAGTCAACCTGCCCACTGAACCGGCCCAATCCCGCCCAGACTACGCGCAGCGTTCACAC
TAGTGACCCGACAGGCACCTGCGATCATCCGGACCTCCCGCGCCGAAGCGGCCCCGCAGCC
CCCGGCCCCCGTGACCTTTACCCTGAACTCCCGCGGAGACCTCCGAACCACCCCCACCCCC
ACCGCCGCGAGAGCCGTCCGAAATCCCGCCCTCCTCCCTGGCGGCGACTGCCTAGCCCCAG
TCCAACAAAACCTCCGCAACGCGTAAGAGCTCGGTACCTATCGATAGAGAAATGTTCTGGC
ACCTGCACTTGYACTGGGRACAGCCTATTTTGCTAGTTTGTTTTGTTTCGTTTTGTTTTGATG
GASAGCGTATGTTAGTACTATCGATTCACACWAAAAACCAACACACAGATGYAATGAAAATA
AAGATATTTTTATTGYRGCCGCTCCAAGTACCTCCCGTAYCTTAATATTACTTACTTATCWT
GKTAKCTTGGGGTGGCGTAATAGCGAKKGGCCCGCWCGATCGCCCTTCYCACAGTTGYGC
AGCCTGAATGRCGAATGCMAATTGTASWGTTAWTATTTTKGTTAAMTTCSCTTAAATTTTGT
WAATCARSTYAWTTTTTSMCARTAGGYCWAATSGCAAAATCYCTTWTAMTCWWAAGAATW
ARMCRAATMGGKTSAATGTGTTCWRTTGGAACAMRAATCCMCWATWAGRAMKGGGACTC
WACGYMAAKGGCGAYARCGTCWTCWGGSRAWYGKCCWTWASKGWAACWCTMWCCTAA
MMRKTTTTTKSGKCCAAGKGYCKTAAGSAMTWWATTSKAATCCTMSAGGARYCCCCWAYT
TWWAMTTAYGSGYAAASCCKKTAACTSGKC

Transfection

A549

	Luciferase	Renilla	Ratio	Luciferase	Renilla	Ratio
G	414000,00	3931,00	105,32	68740,00	1100,00	62,49
G	429400,00	4344,00	98,85	73180,00	1058,00	69,17
G	399900,00	3757,00	106,44	59580,00	1116,00	53,39
Gj.snitt	**414433,33**	**4010,67**	**103,54**	**67166,67**	**1091,33**	**61,68**
T	409000,00	3497,00	116,96	86400,00	1102,00	78,40
T	441300,00	3660,00	120,57	93070,00	1144,00	81,35
T	430900,00	3452,00	124,83	91570,00	1043,00	87,79
Gj.snitt	**427066,67**	**3536,33**	**120,79**	**90346,67**	**1096,33**	**82,52**

	Luciferase	**Renilla**	**Ratio**
G	88840,00	1537,00	57,80
G	116500,00	1595,00	73,04
G	103700,00	1339,00	77,45
Gj.snitt	**103013,33**	**1490,33**	**69,43**
T	127100,00	1485,00	85,59
T	139600,00	1479,00	94,39
T	134700,00	1354,00	99,48
Gj.snitt	**133800,00**	**1439,33**	**93,15**

NCI-H2009

	Luciferase	Renilla	Ratio	Luciferase	Renilla	Ratio
G	702	2100	0,33428571	595	1124	0,52935943
G	604	2270	0,2660793	600	1284	0,46728972
G	213	6290	0,03386328	671	1278	0,52503912
Gj.snitt	**506,333333**	**3553,33333**	**0,21140943**	**622**	**1228,66667**	**0,50722942**
T	440	1537	0,28627196	1155	1454	0,79436039
T	380	1226	0,30995106	1085	1587	0,6836799
T	202	661	0,30559758	1153	1435	0,80348432
Gj.snitt	**340,666667**	**1141,33333**	**0,30060687**	**1131**	**1492**	**0,7605082**

	Luciferase	Renilla	Ratio	Luciferase	Renilla	Ratio
G	730	2070	0,352657	519	1230	0,42195122
G	719	2040	0,35245098	558	1290	0,43255814
G	480	1494	0,32128514	643	1367	0,47037308
Gj.snitt	**643**	**1868**	**0,34213104**	**573,333333**	**1295,66667**	**0,44162748**
T	768	1495	0,51371237	701	1267	0,55327545
T	939	1844	0,50921909	851	1376	0,6184593
T	468	961	0,48699272	828	1238	0,66882068
Gj.snitt	**725**	**1433,33333**	**0,50330806**	**793,333333**	**1293,66667**	**0,61351848**

	Luciferase	Renilla	Ratio	Luciferase	Renilla	Ratio
G	583	2030	0,28719212	508	1672	0,30382775
G	527	1832	0,28766376	501	1639	0,30567419
G	406	1589	0,25550661	406	1212	0,3349835
Gj.snitt	**505,333333**	**1817**	**0,27678749**	**471,666667**	**1507,66667**	**0,31482848**
T	367	956	0,38389121	540	1100	0,49090909
T	234	659	0,35508346	596	1757	0,33921457
T	410	1001	0,40959041	476	1105	0,43076923
Gj.snitt	**337**	**872**	**0,38285503**	**537,333333**	**1320,66667**	**0,42029763**

BEAS-2B

	Luciferase	Renilla	Ratio	Luciferase	Renilla	Ratio
G	3842	711	5,40365682	1972	576	3,42361111
G	4355	728	5,98214286	1682	509	3,30451866
G	4928	856	5,75700935	2233	544	4,10477941
Gj.snitt	**4375**	**765**	**5,71895425**	**1962,33333**	**543**	**3,61387354**
T	8044	1055	7,62464455	3457	832	4,15504808
T	6788	876	7,74885845	3075	717	4,28870293
T	8527	914	9,32932166	3967	512	7,74804688
Gj.snitt	**7786,33333**	**948,333333**	**8,21054482**	**3499,66667**	**687**	**5,09412906**

	Luciferase	Renilla	Ratio	Luciferase	Renilla	Ratio
G	2125	1497	1,41950568	2600	525	4,95238095
G	1681	1392	1,20761494	1900	525	3,61904762
G	1952	1472	1,32608696	2050	579	3,54058722
Gj.snitt	**1919,33333**	**1453,66667**	**1,32033937**	**2183,33333**	**543**	**4,0208717**
T	3487	1908	1,82756813	3898	556	7,01079137
T	3864	2454	1,57457213	4456	567	7,85890653
T	4057	2202	1,84241599	3197	534	5,98689139
Gj.snitt	**3802,66667**	**2188**	**1,73796466**	**3850,33333**	**552,333333**	**6,97103199**

BEP-2D

	Luciferase	Renilla	Ratio	Luciferase	Renilla	Ratio
G	4332	1597	2,7125861	7736	299	25,8729097
G	5980	1615	3,70278638	10710	378	28,3333333
G	3682	1394	2,64131994	11160	505	22,0990099
Gj.snitt	**4664,66667**	**1535,33333**	**3,01889747**	**9868,66667**	**394**	**25,4350843**
T	6783	1679	4,03990471	13110	440	29,7954545
T	7986	1691	4,72264932	13230	543	24,3646409
T	9721	2243	4,33392778	11730	415	28,2650602
Gj.snitt	**8163,33333**	**1871**	**4,36549393**	**12690**	**466**	**27,4750519**

	Luciferase	Renilla	Ratio	Luciferase	Renilla	Ratio
G	3406	642	5,30529595	3757	265	14,1773585
G	2695	585	4,60683761	589	90	6,54444444
G	2781	575	4,83652174	1061	96	11,0520833
Gj.snitt	**2960,66667**	**600,666667**	**4,91621843**	**1802,33333**	**150,333333**	**10,5912954**
T	5424	710	7,63943662	3362	197	17,0659898
T	4906	708	6,92937853	2790	161	17,3291925
T	4619	543	8,50644567	2070	147	14,0816327
Gj.snitt	**4983**	**653,666667**	**7,69175361**	**2740,66667**	**168,333333**	**16,1589383**

	Luciferase	Renilla	Ratio
G	10780	949	11,3593256
G	13140	1017	12,920354
G	12110	880	13,7613636
Gj.snitt	**12010**	**948,666667**	**12,6803477**
T	20290	1098	18,4790528
T	23010	1217	18,9071487
T	26430	1098	24,0710383
Gj.snitt	**23243,3333**	**1137,66667**	**20,4857466**

NCI-H460

	Luciferase	Renilla	Ratio	Luciferase	Renilla	Ratio
G	13460	526	25,5893536	16930	659	25,6904401
G	11680	522	22,3754789	11470	476	24,0966387
G	8435	494	17,0748988	14290	514	27,8015564
Gj.snitt	**11191,6667**	**514**	**21,7736706**	**14230**	**549,666667**	**25,8884172**
T	17180	753	22,815405	23070	501	46,0479042
T	17000	727	23,3837689	26360	637	41,3814757
T	17590	612	28,7418301	25380	558	45,483871
Gj.snitt	**17256,6667**	**697,333333**	**24,7466539**	**24936,6667**	**565,333333**	**44,1096698**

	Luciferase	Renilla	Ratio	Luciferase	Renilla	Ratio
G	7055	676	10,4363905	28260	753	37,5298805
G	7096	429	16,5407925	26790	686	39,0524781
G	4149	399	10,3984962	26510	748	35,4411765
Gj.snitt	**6100**	**501,333333**	**12,1675532**	**27186,6667**	**729**	**37,2930956**
T	13120	609	21,543514	62320	1131	55,1016799
T	17760	779	22,7984596	63600	1104	57,6086957
T	13600	798	17,0426065	71250	1494	47,6907631
Gj.snitt	**14826,6667**	**728,666667**	**20,347667**	**65723,3333**	**1243**	**52,8747654**

HeLa

	Luciferase	Renilla	Ratio	Luciferase	Renilla	Ratio
G	2254	4262	0,52885969	422	681	0,61967695
G	2295	3548	0,64684329	803	984	0,81605691
G	1444	2498	0,57806245	678	976	0,69467213
Gj.snitt	**1997,66667**	**3436**	**0,58139309**	**634,333333**	**880,333333**	**0,72056039**
T	4148	3781	1,09706427	1850	1360	1,36029412
T	4776	4596	1,03916449	2085	1398	1,49141631
T	4064	4163	0,97621907	1769	1284	1,37772586
Gj.snitt	**4329,33333**	**4180**	**1,03572568**	**1901,33333**	**1347,33333**	**1,41118258**

				Luciferase	Renilla	Ratio
G	677	3070	0,22052117	857	1368	0,62646199
G	1023	3689	0,27731092	920	1386	0,66378066
G	784	2757	0,28436707	1140	1513	0,75346993
Gj.snitt	**828**	**3172**	**0,26103405**	**972,333333**	**1422,33333**	**0,68361847**
T	2239	6595	0,33949962	2714	2294	1,18308631
T	2841	5743	0,49468919	3309	2141	1,54553947
T	3151	5371	0,58666915	3799	2260	1,68097345
Gj.snitt	**2743,66667**	**5903**	**0,46479191**	**3274**	**2231,66667**	**1,46706497**

MCF-7

	Luciferase	Renilla	Ratio	Luciferase	Renilla	Ratio
G	150400	1518	99,0777339	407100	5580	72,9569892
G	130900	1488	87,9704301	399200	5478	72,8733114
G	89420	991	90,2320888	254400	3575	71,1608392
Gj.snitt	**123573,333**	**1332,33333**	**92,4267509**	**353566,667**	**4877,66667**	**72,3303799**
T	200700	1957	102,554931	467800	3988	117,301906
T	200100	1779	112,478921	444600	3732	119,131833
T	165900	1744	95,1261468	535000	4729	113,13174
Gj.snitt	**188900**	**1826,66667**	**103,386666**	**482466,667**	**4149,66667**	**116,521826**

	Luciferase	Renilla	Ratio	Luciferase	Renilla	Ratio
G	115000	1431	80,3633823	187600	2369	79,1895314
G	149400	1840	81,1956522	242500	2871	84,4653431
G	157100	1703	92,2489724	230000	2958	77,75524
Gj.snitt	**140500**	**1658**	**84,6026689**	**220033,333**	**2732,66667**	**80,4700382**
T	85720	906	94,6136865	313600	3468	90,4267589
T	76320	903	84,5182724	308000	3418	90,1111761
T	44880	465	96,516129	292100	3383	86,3434821
Gj.snitt	**68973,3333**	**758**	**91,882696**	**304566,667**	**3423**	**88,9604724**

HCT-116

	Luciferase	Renilla	Ratio	Luciferase	Renilla	Ratio
G	483500	54260	8,91079985	359000	87050	4,12406663
G	500500	57560	8,6952745	336400	69610	4,83263899
G	479700	55810	8,59523383	321200	66430	4,83516484
Gj.snitt	**487900**	**55876,6667**	**8,73376939**	**338866,667**	**74363,3333**	**4,59729015**
T	490300	41500	11,8144578	434700	75160	5,78366152
T	759700	52480	14,4759909	467600	77650	6,02189311
T	640200	52480	12,1989329	433800	72230	6,00581476
Gj.snitt	**630066,667**	**48820**	**12,8297939**	**445366,667**	**75013,3333**	**5,93712313**

	Luciferase	Renilla	Ratio	Luciferase	Renilla	Ratio
G	145900	24960	5,84535256	330600	49550	6,67204844
G	157700	28820	5,47189452	423200	63180	6,69832225
G	164400	29450	5,58234295	443800	58180	7,62805088
Gj.snitt	**156000**	**27743,3333**	**5,63319668**	**399200**	**56970**	**6,99947386**
T	214900	27730	7,74972953	516200	50850	10,1514258
T	242000	30050	8,05324459	526600	48180	10,9298464
T	244100	26050	9,37044146	561400	55360	10,140896
Gj.snitt	**233666,667**	**27943,3333**	**8,39113853**	**534733,333**	**51463,3333**	**10,4073894**

Estrogen exposure

A549

G 0 nM	68740	1100	62,4909091	T 0 nM	86400	1102	78,4029038
G 0 nM	73180	1058	69,168242	T 0 nM	93070	1144	81,3548951
G 0 nM	59580	1116	53,3870968	T 0 nM	91570	1043	87,7948226
Gj.snitt	**67166,6667**	**1091,33333**	**61,6820826**	**Gj.snitt**	**90346,6667**	**1096,33333**	**82,5175405**
G 1 nM	9987	348	28,6982759	T 1 nM	31670	404	78,3910891
G 1 nM	9178	271	33,8671587	T 1 nM	23770	314	75,7006369
G 1 nM	12550	441	28,4580499	T 1 nM	22940	346	66,300578
Gj.snitt	**10571,6667**	**353,333333**	**30,3411615**	**Gj.snitt**	**26126,6667**	**354,666667**	**73,4641014**
G 10 nM	9512	200	47,56	T 10 nM	22700	167	135,928144
G 10 nM	17080	311	54,9196141	T 10 nM	13490	109	123,761468
G 10 nM	28620	381	75,1181102	T 10 nM	11230	107	104,953271
Gj.snitt	**18404**	**297,333333**	**59,1992415**	**Gj.snitt**	**15806,6667**	**127,666667**	**121,547628**
G 0 nM	34600	294	117,687075	T 0 nM	23920	281	85,1245552
G 0 nM	21390	245	87,3061224	T 0 nM	30580	289	105,813149
G 0 nM	26490	252	105,119048	T 0 nM	25100	203	123,64532
Gj.snitt	**27493,3333**	**263,666667**	**103,370748**	**Gj.snitt**	**26533,3333**	**257,666667**	**104,861008**
G 1 nM	110700	1509	73,359841	T 1 nM	157200	1198	131,218698
G 1 nM	118100	1350	87,4814815	T 1 nM	168600	1723	97,8525827
G 1 nM	154200	2460	62,6829268	T 1 nM	284500	2447	116,264814
Gj.snitt	**127666,667**	**1773**	**74,5080831**	**Gj.snitt**	**203433,333**	**1789,33333**	**115,112032**
G 10 nM	168500	1959	86,0132721	T 10 nM	279200	2910	95,9450172
G 10 nM	441700	5110	86,4383562	T 10 nM	365200	2537	143,949547
G 10 nM	221400	2652	83,4841629	T 10 nM	204900	2039	100,490436
Gj.snitt	**277200**	**3240,33333**	**85,3119304**	**Gj.snitt**	**283100**	**2495,33333**	**113,461667**
G 0 nM	88840	1537	57,8009109	T 0 nM	127100	1485	85,5892256
G 0 nM	116500	1595	73,0407524	T 0 nM	139600	1479	94,3881001
G 0 nM	103700	1339	77,4458551	T 0 nM	134700	1354	99,4830133
Gj.snitt	**103013,333**	**1490,33333**	**69,4291728**	**Gj.snitt**	**133800**	**1439,33333**	**93,1534463**
G 1 nM	53180	974	54,5995893	T 1 nM	144000	1574	91,4866582
G 1 nM	110200	1426	77,2791024	T 1 nM	123900	1066	116,228893
G 1 nM	41820	695	60,1726619	T 1 nM	126600	1458	86,8312757
Gj.snitt	**68400**	**1031,66667**	**64,0171179**	**Gj.snitt**	**131500**	**1366**	**98,1822757**
G 10 nM	142300	2093	67,9885332	T 10 nM	289500	2492	116,17175
G 10 nM	182400	2075	87,9036145	T 10 nM	206300	2145	96,1771562
G 10 nM	146000	1679	86,9565217	T 10 nM	121600	1312	92,6829268
Gj.snitt	**156900**	**1949**	**80,9495565**	**Gj.snitt**	**205800**	**1983**	**101,677278**
G 0 nM	414000	3931	105,316713	T 0 nM	409000	3497	116,957392
G 0 nM	429400	4344	98,8489871	T 0 nM	441300	3660	120,57377
G 0 nM	399900	3757	106,44131	T 0 nM	430900	3452	124,826188
Gj.snitt	**414433,333**	**4010,66667**	**103,53567**	**Gj.snitt**	**427066,667**	**3536,33333**	**120,785783**
G 1 nM	452300	3902	115,914915	T 1 nM	440400	3006	146,506986
G 1 nM	332900	3073	108,330622	T 1 nM	483600	3060	158,039216
G 1 nM	259600	2300	112,869565	T 1 nM	228000	1698	134,275618
Gj.snitt	**348266,667**	**3091,66667**	**112,371701**	**Gj.snitt**	**384000**	**2588**	**146,27394**
G 10 nM	79130	646	122,49226	T 10 nM	83080	603	137,777778
G 10 nM	70570	532	132,650376	T 10 nM	143800	916	156,9869
G 10 nM	158000	1292	122,291022	T 10 nM	129400	864	149,768519

Gj.snitt	**102566,667**	**823,333333**	**125,811219**	**Gj.snitt**	**118760**	**794,333333**	**148,177732**
G 0 nM	186200	1025	181,658537	T 0 nM	131300	648	202,623457
G 0 nM	163300	963	169,574247	T 0 nM	164000	640	256,25
G 0 nM	149000	866	172,055427	T 0 nM	51740	272	190,220588
Gj.snitt	**166166,667**	**951,333333**	**174,429404**	**Gj.snitt**	**115680**	**520**	**216,364682**
G 1 nM	158300	988	160,222672	T 1 nM	96170	464	207,262931
G 1 nM	121800	686	177,55102	T 1 nM	103200	542	190,405904
G 1 nM	30900	461	67,0281996	T 1 nM	88900	604	147,18543
Gj.snitt	**103666,667**	**711,666667**	**134,933964**	**Gj.snitt**	**96090**	**536,666667**	**181,618089**
G 10 nM	84440	561	150,516934	T 10 nM	165100	804	205,348259
G 10 nM	32890	247	133,157895	T 10 nM	74790	365	204,90411
G 10 nM	67870	388	174,92268	T 10 nM	45950	305	150,655738
Gj.snitt	**61733,3333**	**398,666667**	**152,865836**	**Gj.snitt**	**95280**	**491,333333**	**186,969369**

NCI-H2009

	Luciferase	Renilla	Ratio		Luciferase	Renilla	Ratio
G 0 nM	730	2070	0,352657	T 0 nm	768	1495	0,51371237
G 0 nM	719	2040	0,35245098	T 0 nm	939	1844	0,50921909
G 0 nM	480	1494	0,32128514	T 0 nm	468	961	0,48699272
Gj.snitt	**643**	**1868**	**0,34213104**	**Gj.snitt**	**725**	**1433,33333**	**0,50330806**
G 1 nM	274	1043	0,26270374	T 1 nM	871	1885	0,46206897
G 1 nM	280	1219	0,22969647	T 1 nM	932	1998	0,46646647
G 1 nM	158	1146	0,13787086	T 1 nM	815	1825	0,44657534
Gj.snitt	**237,333333**	**1136**	**0,21009036**	**Gj.snitt**	**872,666667**	**1902,66667**	**0,45837026**
G 10 nM	153	651	0,23502304	T 10 nM	600	1230	0,48780488
G 10 nM	99	885	0,11186441	T 10 nM	744	1332	0,55855856
G 10 nM	88	616	0,14285714	T 10 nM	604	1139	0,53028973
Gj.snitt	**113,333333**	**717,333333**	**0,15799257**	**Gj.snitt**	**649,333333**	**1233,66667**	**0,52555105**

	Luciferase	Renilla	Ratio		Luciferase	Renilla	Ratio
G 0 nM	481	1122	0,42869875	T 0 nM	369	758	0,48680739
G 0 nM	613	1495	0,41003344	T 0 nM	391	689	0,56748911
G 0 nM	671	1315	0,51026616	T 0 nM	413	820	0,50365854
Gj.snitt	**588,333333**	**1310,66667**	**0,44966612**	**Gj.snitt**	**391**	**755,666667**	**0,51931835**
G 1 nM	323	813	0,39729397	T 1 nM	357	734	0,48637602
G 1 nM	323	807	0,40024783	T 1 nM	409	629	0,65023847
G 1 nM	463	1048	0,44179389	T 1 nM	475	828	0,5736715
Gj.snitt	**369,666667**	**889,333333**	**0,4131119**	**Gj.snitt**	**413,666667**	**730,333333**	**0,57009533**
G 10 nM	236	622	0,37942122	T 10 nM	385	617	0,62398703
G 10 nM	345	769	0,44863459	T 10 nM	469	923	0,50812568
G 10 nM	407	943	0,43160127	T 10 nM	408	643	0,63452566
Gj.snitt	**329,333333**	**778**	**0,42330763**	**Gj.snitt**	**420,666667**	**727,666667**	**0,58887946**

	Luciferase	Renilla	Ratio		Luciferase	Renilla	Ratio
G 0 nM	1025	1941	0,52807831	T 0 nM	867	1363	0,63609685
G 0 nM	918	1667	0,55068986	T 0 nM	1242	1809	0,68656716
G 0 nM	807	1462	0,55198358	T 0 nM	1163	1483	0,78422117
Gj.snitt	**916,666667**	**1690**	**0,54358392**	**Gj.snitt**	**1090,66667**	**1551,66667**	**0,70229506**
G 1 nM	746	1397	0,53400143	T 1 nM	1231	1625	0,75753846
G 1 nM	716	1301	0,55034589	T 1 nM	1220	1480	0,82432432

G 1 nM	796	1101	0,72297911	T 1 nM	1060	1856	0,57112069
Gj.snitt	**752,666667**	**1266,33333**	**0,60244214**	**Gj.snitt**	**1170,33333**	**1653,66667**	**0,71766116**
G 10 nM	610	1051	0,58039962	T 10 nM	755	1033	0,73088093
G 10 nM	848	1594	0,53199498	T 10 nM	1028	1339	0,76773712
G 10 nM	769	1428	0,53851541	T 10 nM	1197	1471	0,81373215
Gj.snitt	**742,333333**	**1357,66667**	**0,54677142**	**Gj.snitt**	**993,333333**	**1281**	**0,7707834**

BEAS-2B

	Luciferase	Renilla	Ratio		Luciferase	Renilla	Ratio
G 1 nM	18070	912	19,8135965	**T 1 nM**	10900	637	17,11146
G 1 nM	16260	991	16,407669	**T 1 nM**	14650	503	29,1252485
G 1 nM	14390	898	16,0244989	**T 1 nM**	22050	942	23,4076433
Gj.snitt	**16240**	**933,666667**	**17,4152548**	**Gj.snitt**	**15866,6667**	**694**	**23,2147839**
G 10 nM	10930	592	18,4628378	**T 10 nM**	24110	821	29,3666261
G 10 nM	14210	765	18,5751634	**T 10 nM**	26090	1590	16,408805
G 10 nM	12900	788	16,3705584	**T 10 nM**	19570	837	23,3811231
Gj.snitt	**12680**	**715**	**17,8028532**	**Gj.snitt**	**23256,6667**	**1082,66667**	**23,0521847**

	Luciferase	Renilla	Ratio		Luciferase	Renilla	Ratio
G 0	10040	766	13,1070496	**T 0**	12370	648	19,0895062
G 0	11120	847	13,1286895	**T 0**	11940	596	20,033557
G 0	10570	742	14,245283	**T 0**	12670	514	24,6498054
Gj.snitt	**10576,6667**	**785**	**13,4734607**	**Gj.snitt**	**12326,6667**	**586**	**21,2576229**
G 1 nM	10000	623	16,0513644	**T 1 nM**	11000	411	26,7639903
G 1 nM	9742	616	15,8149351	**T 1 nM**	11680	574	20,3484321
G 1 nM	9639	692	13,9291908	**T 1 nM**	14280	705	20,2553191
Gj.snitt	**9793,66667**	**643,666667**	**15,2651634**	**Gj.snitt**	**12320**	**563,333333**	**22,4559138**
G 10 nM	6023	428	14,0724299	**T 10 nM**	9826	385	25,5220779
G 10 nM	8922	601	14,8452579	**T 10 nM**	12370	554	22,3285199
G 10 nM	11060	792	13,9646465	**T 10 nM**	11480	480	23,9166667
Gj.snitt	**8668,33333**	**607**	**14,2941114**	**Gj.snitt**	**11225,3333**	**473**	**23,9224215**

	Luciferase	Renilla	Ratio		Luciferase	Renilla	Ratio
G 0	16630	2528	6,57832278	**T 0**	13580	2018	6,72943508
G 0	16530	3558	4,64586847	**T 0**	13090	1454	9,00275103
G 0	22420	3305	6,78366112	**T 0**	8364	1142	7,32399299
Gj.snitt	**18526,6667**	**3130,33333**	**5,91843254**	**Gj.snitt**	**11678**	**1538**	**7,68539304**
G 1 nM	12280	2057	5,96985902	**T 1 nM**	11330	1347	8,41128434
G 1 nM	14800	2002	7,39260739	**T 1 nM**	15610	1549	10,0774693
G 1 nM	18470	2228	8,28994614	**T 1 nM**	10110	1398	7,23175966
Gj.snitt	**15183,3333**	**2095,66667**	**7,21747085**	**Gj.snitt**	**12350**	**1431,33333**	**8,57350444**
G 10 nM	5384	581	9,26678141	**T 10 nM**	3702	1172	3,15870307
G 10 nM	3028	984	3,07723577	**T 10 nM**	6768	758	8,92875989
G 10 nM	5515	820	6,72560976	**T 10 nM**	5067	631	8,03011094
Gj.snitt	**4642,33333**	**795**	**6,35654231**	**Gj.snitt**	**5179**	**853,666667**	**6,70585797**

	Luciferase	Renilla	Ratio		Luciferase	Renilla	Ratio
G 0	1220	286	4,26573427	**T 0**	907	178	5,09550562
G 0	1497	259	5,77992278	**T 0**	689	145	4,75172414
G 0	3159	636	4,96698113	**T 0**	969	110	8,80909091
Gj.snitt	**1958,66667**	**393,666667**	**4,97544454**	**Gj.snitt**	**855**	**144,333333**	**6,21877355**

G 1 nM	8510	1635	5,20489297	**T 1 nM**	8631	1256	6,87181529
G 1 nM	7606	1520	5,00394737	**T 1 nM**	6510	1209	5,38461538
G 1 nM	1216	320	3,8	**T 1 nM**	6921	1123	6,16295637
Gj.snitt	**5777,33333**	**1158,33333**	**4,66961344**	**Gj.snitt**	**7354**	**1196**	**6,13979568**
G 10 nM	5816	1437	4,04732081	**T 10 nM**	7646	1250	6,1168
G 10 nM	2771	762	3,63648294	**T 10 nM**	5808	790	7,35189873
G 10 nM	1012	189	5,35449735	**T 10 nM**	4879	918	5,31481481
Gj.snitt	**3199,66667**	**796**	**4,34610037**	**Gj.snitt**	**6111**	**986**	**6,26117118**

BEP-2D

	Luciferase	Renilla	Ratio		Luciferase	Renilla	Ratio
G 0 nM	4188	520	8,05384615	T 0 nM	1595	227	7,02643172
G 0 nM	2980	757	3,93659181	T 0 nM	1124	190	5,91578947
G 0 nM	2526	449	5,62583519	T 0 nM	1413	239	5,91213389
Gj.snitt	**3231,33333**	**575,333333**	**5,87209105**	**Gj.snitt**	**1377,33333**	**218,666667**	**6,28478503**
G 1 nM	3172	433	7,3256351	T 1 nM	2783	466	5,972103
G 1 nM	3007	463	6,49460043	T 1 nM	3250	433	7,50577367
G 1 nM	2528	430	5,87906977	T 1 nM	3640	616	5,90909091
Gj.snitt	**2902,33333**	**442**	**6,5664351**	**Gj.snitt**	**3224,33333**	**505**	**6,46232253**
G 10 nM	3346	544	6,15073529	T 10 nM	3103	413	7,51331719
G 10 nM	3450	488	7,06967213	T 10 nM	2957	406	7,28325123
G 10 nM	3078	483	6,37267081	T 10 nM	2907	408	7,125
Gj.snitt	**3291,33333**	**505**	**6,53102608**	**Gj.snitt**	**2989**	**409**	**7,30718947**
G 0	6399	553	11,5714286	T 0	7022	447	15,7091723
G 0	5754	487	11,8151951	T 0	7348	486	15,1193416
G 0	5781	511	11,3131115	T 0	7747	446	17,3699552
Gj.snitt	**5978**	**517**	**11,5665784**	**Gj.snitt**	**7372,33333**	**459,666667**	**16,0661563**
G 1 nM	6215	501	12,4051896	T 1 nM	4939	363	13,6060606
G 1 nM	6753	521	12,9616123	T 1 nM	6432	468	13,7435897
G 1 nM	6191	538	11,5074349	T 1 nM	4377	369	11,8617886
Gj.snitt	**6386,33333**	**520**	**12,2914123**	**Gj.snitt**	**5249,33333**	**400**	**13,0704797**
G 10 nM	5710	479	11,9206681	T 10 nM	1952	222	8,79279279
G 10 nM	5735	468	12,2542735	T 10 nM	7198	472	15,25
G 10 nM	6336	411	15,4160584	T 10 nM	2706	236	11,4661017
Gj.snitt	**5927**	**452,666667**	**13,197**	**Gj.snitt**	**3952**	**310**	**11,8362982**
G 0	10720	819	13,0891331	T 0	12590	709	17,7574048
G 0	10460	721	14,5076283	T 0	18090	580	31,1896552
G 0	10600	721	14,7018031	T 0	14620	732	19,9726776
Gj.snitt	**10593,3333**	**753,666667**	**14,0995215**	**Gj.snitt**	**15100**	**673,666667**	**22,9732459**
G 1 nM	8086	554	14,5956679	T 1 nM	8510	498	17,0883534
G 1 nM	9061	558	16,2383513	T 1 nM	8933	579	15,4283247
G 1 nM	7817	617	12,6693679	T 1 nM	7084	366	19,3551913
Gj.snitt	**8321,33333**	**576,333333**	**14,501129**	**Gj.snitt**	**8175,66667**	**481**	**17,2906231**
G 10 nM	6527	558	11,6971326	T 10 nM	1011	107	9,44859813
G 10 nM	5453	447	12,1991051	T 10 nM	1496	95	15,7473684
G 10 nM	7142	533	13,3996248	T 10 nM	4215	187	22,540107
Gj.snitt	**6374**	**512,666667**	**12,4319542**	**Gj.snitt**	**2240,66667**	**129,666667**	**15,9120245**

Appendix IV

The Biobank located at NIOH for lung cancer studies.

This project is based on the case-control study done (Lind *et al*) with samples from the Biobank located at the National Institute for Occupational Health (NIOH). The cases were 341 Caucasians born in Norway and of Norwegian origin with newly diagnosed NSCLC which were admitted for surgery at the university hospitals in Oslo or Bergen between 1986 and 2001. Controls were recruited from a general health survey of the general population conducted by the National Health Surveys in the Oslo area (HUBRO) in 2000-2001. The controls in this study were selected from this survey based on the following criteria: (i) they were ≥59 years of age; (ii) had smoked >5 cigarettes per day; and (iii) were current smokers or quit smoking for <5 years. Among them, 412 smokers without any known history of cancer were randomly selected and frequency matched with the cases on age, smoking dose (pack-years) and male:female ratio. From these individuals, blood samples and lung tissue samples have been obtained and used for DNA isolation. This DNA has been the basis for the SNP association study by Lind *et al* as well as the C1797G study in this thesis.

Transcription factors:

E2F-1	Involved in cell cycle control (G1-S progression), may regulate the action of tumor suppressor proteins, binds Rb, can mediate p53 dependent apoptosis and proliferation.
C/EBP	Regulates the cell cycle by inducing cell cycle arrest, involved in weight homeostasis, regulates genes involved in inflammation and immune responses example the IL-1 responsive element on IL-6, can stimulate the expression of collagen type I.
p53	Involved in cell cycle control, apoptosis and genetic stability.
Sp1	A general transcription factor. May be involved in the axis development (anterior-posterior etc), organogenesis, differentiation and tissue growth.
Pax-5	Regulates early development and alterations in regulation promotes neoplastic lesions.
GCF	Repressor that binds GC-rich sequences. May regulate the expression of tumor necrosis factors and epidermal growth factor receptors.
Egr-1	Early growth response-1 is involved in differentiation and mitogenesis. Might be a cancer suppressor.
ETF	Participates in catalyzing the initial step of the mitochondrial fatty acid beta-oxidation. It shuttles electrons between primary flavoprotein dehydrogenases and the membrane-bound electron transfer flavoprotein ubiquinone oxidoreductase.
Oct-1	Transcription factor that binds to the octamer motif (5'-ATTTGCAT-3') and activates the promoters of the genes for some small nuclear RNAs (snRNA) and of genes such as those for histone H2B and immunoglobulins.
NF-muE1	Multifunctional transcription factor that exhibits positive and negative control on a large number of cellular and viral genes by binding to sites overlapping the transcription start site. May play an important role in development and differentiation.
c-Myc	Participates in the regulation of gene transcription. Binds DNA both in a non-specific manner and also specifically to recognizes the core sequence 5'-

	CAC[GA]TG-3'. Seems to activate the transcription of growth-related genes.
AP-2alph	Sequence-specific DNA-binding protein that interacts with inducible viral and cellular enhancer elements to regulate transcription of selected genes. AP-2 factors bind to the consensus sequence 5'-GCCNNNGGC-3' and activate genes involved in a large spectrum of important biological functions including proper eye, face, body wall, limb and neural tube development.

Appendix V

SNP309 association studies and their outcomes.

Authors	Tissue	Other findings	Risk association with SNP309	Population
K. Mittelstrass *et al [82]*	Lung	Possibly earlier age of onset for women.	No	Caucasian
X. Zhang *et al [83]*	Lung	Also association with SNP p53 Arg72Pro	Yes	Han Chinese
Z. Hu *et al [73]*	Lung		No	Chinese
S. Pine *et al [84]*	Lung	Also no association with MDM2 SNP354	No	African-American Caucasian
Lind *et al [57]*	Lung NSCLC	Especially association with women. Mutations in p53 associated with higher age of diagnosis	Yes	Caucasian
H. J. Jun *et al [85]*	Lung	P73 SNP G4C14 – A4T14 also associated with risk. In addition an additive effect between the SNPs.	Yes	Korean
G. Liu *et al [86]*	Lung NSCLC		Yes in never/light smokers	Caucasian
G. Li *et al [76]*	Lung	Protective function G allele? Higher risk with TT genotype (especially males).	No	Caucasian
J. Han *et al [75]*	Lung NSCLC	Clinical outcome of NSCLC. Shorter overall survival time. SNP p53 codon 72 gave a worse clinical outcome.	Yes	Korean
M. Ruijs *et al [87]*	Any primary tumor	Li Fraumeni patients. Earlier age of onset.	Yes	Dutch Finnish
V. Galic *et al [66]*	Ovarian	SNP p53 codon 72 alone associated with shorter survival.	No	Caucasian
	Peritoneal	SNP p53 codon 72 alone associated with shorter	No	Caucasian

		survival.		
G. Bond *et al [81]*	Diffuse large B-cell lymphoma	Earlier age of onset for ER positive tumor patients.	No	Caucasian (Ashkenazi Jewish)
	Breast (invasive ductal)	Earlier age of onset in patients with ER positive tumors.	No	Caucasian (Ashkenazi Jewish)
	Soft tissue sarcoma	Earlier age of onset.	No	Caucasian
R. Stoehr *et al [88]*	Prostate		No	German
B. Talseth *et al [70]*	Colorectal	Age of onset in individuals with hereditary nonpolyposis colorectal cancer.	No	Australian Polish
G. Bond *et al [65]*	Colorectal	Earlier age of onset in women.	No	Italian Finnish
D. Krekac *et al [67]*	Breast		No	Czech
C. Walsh *et al [72]*	Endometrial		Yes	Caucasian African-American Hispanic Asian
P. Alhopuro *et al [64]*	Uterine		No	Finnish
	Colorectal	Earlier age of onset in women.	No	
	Head and neck		No	
N. Dharel *et al [89]*	Liver	Patients with hepatitis C.	Yes	Japanese
H. Tsuiki *et al [90]*	Glioma		No	Japanese

Appendix VI

Odds ratio

Odds ratio can be defined as the ratio between the odds of a certain happening in one group and the odds for the same happening in another group [60].

The equation for odds ratio calculation is shown below.

$$Eq.1: \quad \frac{P/(1-P)}{Q/(1-Q)} = \frac{P/(1-Q)}{Q/(1-P)}$$

Linear regression

A regression analysis will study the relationship between 2 variables where one is dependent and the other is explanatory/independent. Linear regression gives the regression line function as shown in equation 1.

$$Eq.2: \quad y = a + bx$$

Here b describes how much the y-value will change with one unit difference in x. This function is not accurately fulfilled, but an estimated function is satisfied and shown in equation 2.

$$Eq.3: \quad y = a + bx + \varepsilon$$

Here ε indicates the individual variation and is the difference between the estimated line value and the actual observation.

To find the regression line the method of least squares is employed. For the number of individual x_i and y_i values the means are calculated (x and y) and put into equation 3.

$$Eq.4: \quad a = y - bx$$

The b value can be further calculated shown in equation 4.

$$Eq.5: \quad b = \frac{\sum_{i=1}^{n}(x_i - x)(y_i - y)}{\sum_{i=1}^{n}(x_i - x)^2}$$

Here b is called the regression coefficient. These estimated values are then used in equation 1.

Also the R^2 value can be calculated and this is the Pearson's coefficient of regression. This coefficient is closely related to the regression coefficient b described in equation 4. Pearson's coefficient will be used with normally distributed data and will vary between -1 and 1 for negative and positive correlation, respectively. It describes the strength and direction of the linear relationship between the two variables and how much variation there is around the regression line. For an increasing regression line the value should be as close as possible to 1 because of the positive correlation. A value closer to 0 will indicate no linear relationship between the variables [60].

T-test

The T-test (Student T-test) is used to study the differences in means or medians in a group of observations. There are two types of this test: paired T-test and T-test. The T-test can be used to conclude on observations in one group or in two groups. Two groups of data can be analyzed as

one (paired T-test) but the data must then be related and the difference between observations must be calculated.

An example where the paired T-test can be used is when measuring blood pressure before and after treatment in the same patient. The difference between reading 1 and reading 2 is calculated. If deviation from 0 is found in addition to a p-value below 0,05 the null hypothesis is rejected. The equation used is shown in equation 5.

$$Eq.6: \quad t = \frac{d}{s/\sqrt{n}}$$

In equation 5 d is the mean of the calculated differences in the measurements, s is the standard deviation in the differences calculated and n is the number of subjects used.

An example where the T-test is used is during the comparison of two independent groups of data. The data must be normally distributed in addition to the assumption that the standard deviations in the two groups are approximately the same. The null hypothesis is that there is no difference between the two groups. The means of the two groups are calculated and put into equation 6.

$$Eq.7: \quad t = \frac{X_1 - X_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Here X is the mean of group 1 and 2 respectively, s the standard deviation and n the number of observations in group 1 and 2 respectively.

The standard deviation, assumed alike in the two groups, is calculated from equation 7.

$$Eq.8: \quad s = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

This test can be used for example when analyzing the same parameter, but in different subject groups. Again, if the t-value deviates from 0 and the p-value is below 0,05 the null hypothesis is rejected [60].

In this thesis standard error has been chosen to be included in the graphs. Standard error is the standard deviation of the different measurements deviation from the expected value. It is often used with the T-test to provide a confidence interval for an estimated mean or difference of means. It can also provide an indication of the uncertainty and its size. Standard deviation is a measurement of the distribution of a datasample's values, the deviation from the value's mean, an estimate of the variability of the population from which the sample was drawn. When the data is normally distributes 95 % of the individuals will have values within 2 standard deviations from the mean. The standard error of the sample mean depends on both the standard deviation and the sample size and can be described as SE = SD/√(sample size). The standard error falls as the sample size increases since the chance variation is reduced. By contrast the standard deviation will not tend to change as we increase the size of our sample [91].

Logistic regression analysis (SNP association study)

The logistic regression function follows the equation 9 where z may be any number both positive and negative infinitely whereas the f(z) varies between 0 and 1.

$$Eq.9: \quad f(z) = \frac{1}{1 + e^{-z}}$$

z represents the exposure to risk factors and f(z) is the probability of a certain outcome.

$$Eq.10: \quad z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

β_k is the regression coefficient of x_k. A positive regression coefficient indicates that the risk factor increases the probability of the outcome.

The application of a logistic regression may be illustrated using a fictitious example of death from heart disease. This simplified model uses only three risk factors (age, sex, and blood cholesterol level) to predict the 10-year risk of death from heart disease. This is the model that we fit:

$\beta_0 = -5.0$ (the intercept)

$\beta_1 = +2.0$

$\beta_2 = -1.0$

$\beta_3 = +1.2$

x_1 = age in decades, less 5.0

x_2 = sex, where 0 is male and 1 is female

x_3 = cholesterol level, in mmol/L less 5.0

Which means the model is:

$$Eq.11: \quad risk\ of\ death = \frac{1}{1+e^{-z}}, where\ z = -5{,}0 + 2{,}0x_1 - 1{,}0x_2 + 1{,}2x_3$$

In this model, increasing age is associated with an increasing risk of death from heart disease (z goes up by 2.0 for every 10 years over the age of 50), female sex is associated with a decreased risk of death from heart disease (z goes down by 1.0 if the patient is female), and increasing cholesterol is associated with an increasing risk of death (z goes up by 1.2 for each 1 mmol/L increase in cholesterol above 5mmol/L) [60].

The outcome of a logistic regression is the odds ratio which helps to elucidate the factors with the greatest importance in the outcome one is studying. The confidence interval indicate the statistical significance of the predicted odds ratio [92].

References

1. Sun, S., Schiller, J.H. and Gazdar, A.F. (2007) Lung cancer in never smokers--a different disease. *Nature Reviews Cancer*, **7**, 778-90.
2. Bray, F. Cancer Registry of Norway - Cancer in Norway 2006.
3. Haugen, A., Ryberg, D., Mollerup, S., Zienolddiny, S., Skaug, V. and Svendsrud, D.H. (2000) Gene-environment interactions in human lung cancer. *Toxicology Letters*, **112-113**, 233-7.
4. Klaassen, C.D. and Watkins, J.B. (2003) *Casarett and Doull's Essentials of Toxicology*. The McGraw-Hill Companies, Inc.
5. Klaassen, C.D. (2001) Casarett and Doull's Toxicology - The basic science of poisons.
6. Hussain, S.P. and Harris, C.C. (2007) Inflammation and cancer: an ancient link with novel potentials. *International Journal of Cancer*, **121**, 2373-80.
7. Bostrom, C.E., Gerde, P., Hanberg, A., Jernstrom, B., Johansson, C., Kyrklund, T., Rannug, A., Tornqvist, M., Victorin, K. and Westerholm, R. (2002) Cancer risk assessment, indicators, and guidelines for polycyclic aromatic hydrocarbons in the ambient air. *Environmental Health Perspectives*, **110 Suppl 3**, 451-88.
8. Holme, J.A., Gorria, M., Arlt, V.M., Ovrebo, S., Solhaug, A., Tekpli, X., Landvik, N.E., Huc, L., Fardel, O. and Lagadic-Gossmann, D. (2007) Different mechanisms involved in apoptosis following exposure to benzo[a]pyrene in F258 and Hepa1c1c7 cells. *Chemico-Biological Interactions*, **167**, 41-55.
9. Hylland, K. (2006) Polycyclic aromatic hydrocarbon (PAH) ecotoxicology in marine ecosystems. *Journal of Toxicology and Environmental Health part A*, **69**, 109-23.
10. Pfeifer, G.P., Denissenko, M.F., Olivier, M., Tretyakova, N., Hecht, S.S. and Hainaut, P. (2002) Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene*, **21**, 7435-51.
11. Hecht, S.S. (1999) Tobacco smoke carcinogens and lung cancer. *Journal of the National Cancer Institute*, **91**, 1194-210.
12. Spitz, M.R., Wei, Q., Dong, Q., Amos, C.I. and Wu, X. (2003) Genetic susceptibility to lung cancer: the role of DNA damage and repair. *Cancer Epidemiology, Biomarkers & Prevention* **12**, 689-98.
13. Mollerup, S., Berge, G., Baera, R., Skaug, V., Hewer, A., Phillips, D.H., Stangeland, L. and Haugen, A. (2006) Sex differences in risk of lung cancer: Expression of genes in the PAH bioactivation pathway in relation to smoking and bulky DNA adducts. *International Journal of Cancer*, **119**, 741-4.
14. Duong, V., Boulle, N., Daujat, S., Chauvet, J., Bonnet, S., Neel, H. and Cavailles, V. (2007) Differential regulation of estrogen receptor alpha turnover and transactivation by Mdm2 and stress-inducing agents. *Cancer Research*, **67**, 5513-21.
15. Saji, S., Okumura, N., Eguchi, H., Nakashima, S., Suzuki, A., Toi, M., Nozawa, Y. and Hayashi, S. (2001) MDM2 enhances the function of estrogen receptor alpha in human breast cancer cells. *Biochemical and Biophysical Research Communications*, **281**, 259-65.
16. Dougherty, S.M., Mazhawidza, W., Bohn, A.R., Robinson, K.A., Mattingly, K.A., Blankenship, K.A., Huff, M.O., McGregor, W.G. and Klinge, C.M. (2006) Gender difference in the activity but not expression of estrogen receptors alpha and beta in human lung adenocarcinoma cells. *Endocrine-Related Cancer*, **13**, 113-34.
17. Mollerup, S., Jorgensen, K., Berge, G. and Haugen, A. (2002) Expression of estrogen receptors alpha and beta in human lung tissue and cell lines. *Lung Cancer*, **37**, 153-9.
18. Kiyohara, C., Yoshimasu, K., Takayama, K. and Nakanishi, Y. (2007) Lung cancer susceptibility: are we on our way to identifying a high-risk group? *Future Oncology*, **3**, 617-27.
19. Nebert, D.W., Zhang, G. and Vesell, E.S. (2008) From human genetics and genomics to pharmacogenetics and pharmacogenomics: past lessons, future directions. *Drug Metabolism Reviews*, **40**, 187-224.
20. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002) *The Molecular Biology of the Cell*. Garland Science, London
New York.

21. Pitot, H.C. (1993) The molecular biology of carcinogenesis. *Cancer*, **72**, 962-70.
22. Weinberg, R.A. (1994) Oncogenes and tumor suppressor genes. *CA; Cancer Journal for Clinicians*, **44**, 160-70.
23. Okada, H. and Mak, T.W. (2004) Pathways of apoptotic and non-apoptotic death in tumour cells. *Nature Reviews Cancer*, **4**, 592-603.
24. Luch, A. (2002) Cell cycle control and cell division: implications for chemically induced carcinogenesis. *Chembiochem*, **3**, 506-16.
25. Pestell, R.G., Albanese, C., Reutens, A.T., Segall, J.E., Lee, R.J. and Arnold, A. (1999) The cyclins and cyclin-dependent kinase inhibitors in hormonal regulation of proliferation and differentiation. *Endocrine Reviews*, **20**, 501-34.
26. Sun, Y. (2006) p53 and its downstream proteins as molecular targets of cancer. *Molecular Carcinogenesis*, **45**, 409-15.
27. Bratton, S.B. and Cohen, G.M. (2001) Apoptotic death sensor: an organelle's alter ego? *Trends in Pharmacological Sciences*, **22**, 306-15.
28. Hussain, S.P. and Harris, C.C. (2000) Molecular epidemiology and carcinogenesis: endogenous and exogenous carcinogens. *Mutation Research*, **462**, 311-22.
29. Bose, I. and Ghosh, B. (2007) The p53-MDM2 network: from oscillations to apoptosis. *Journal of Biosciences*, **32**, 991-7.
30. Norbury, C.J. and Zhivotovsky, B. (2004) DNA damage-induced apoptosis. *Oncogene*, **23**, 2797-808.
31. Rodier, F., Campisi, J. and Bhaumik, D. (2007) Two faces of p53: aging and tumor suppression. *Nucleic Acids Research*, **35**, 7475-84.
32. Toledo, F. and Wahl, G.M. (2006) Regulating the p53 pathway: in vitro hypotheses, in vivo veritas. *Nature Reviews Cancer*, **6**, 909-23.
33. Iwakuma, T. and Lozano, G. (2003) MDM2, an introduction. *Molecular Cancer Research*, **1**, 993-1000.
34. Liang, H. and Lunec, J. (2005) Characterisation of a novel p53 down-regulated promoter in intron 3 of the human MDM2 oncogene. *Gene*, **361**, 112-8.
35. Perry, M.E. (2004) Mdm2 in the response to radiation. *Molecular Cancer Research*, **2**, 9-19.
36. Bond, G.L., Hu, W. and Levine, A. (2005) A single nucleotide polymorphism in the MDM2 gene: from a molecular and cellular explanation to clinical effect. *Cancer Research*, **65**, 5481-4.
37. Bartel, F., Harris, L.C., Wurl, P. and Taubert, H. (2004) MDM2 and its splice variant messenger RNAs: expression in tumors and down-regulation using antisense oligonucleotides. *Molecular Cancer Research*, **2**, 29-35.
38. Weng, M.W., Lai, J.C., Hsu, C.P., Yu, K.Y., Chen, C.Y., Lin, T.S., Lai, W.W., Lee, H. and Ko, J.L. (2005) Alternative splicing of MDM2 mRNA in lung carcinomas and lung cell lines. *Environonmental and Molecular Mutagenesis*, **46**, 1-11.
39. Bartel, F., Taubert, H. and Harris, L.C. (2002) Alternative and aberrant splicing of MDM2 mRNA in human cancer. *Cancer Cell*, **2**, 9-15.
40. Deb, S.P. (2003) Cell cycle regulatory functions of the human oncoprotein MDM2. *Molecular Cancer Research*, **1**, 1009-16.
41. Meek, D.W. and Knippschild, U. (2003) Posttranslational modification of MDM2. *Molecular Cancer Research*, **1**, 1017-26.
42. Ganguli, G. and Wasylyk, B. (2003) p53-independent functions of MDM2. *Molecular Cancer Research*, **1**, 1027-35.
43. Moll, U.M. and Petrenko, O. (2003) The MDM2-p53 interaction. *Molecular Cancer Research*, **1**, 1001-8.
44. Johnson-Pais, T., Degnin, C. and Thayer, M.J. (2001) pRB induces Sp1 activity by relieving inhibition mediated by MDM2. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 2211-6.
45. Brown, D.R., Thomas, C.A. and Deb, S.P. (1998) The human oncoprotein MDM2 arrests the cell cycle: elimination of its cell-cycle-inhibitory function induces tumorigenesis. *The EMBO journal (European Molecular Biology Organization)*, **17**, 2513-25.
46. Marine, J.C. and Jochemsen, A.G. (2005) Mdmx as an essential regulator of p53 activity. *Biochemical and Biophysical Research Communications*, **331**, 750-60.

47. Brooks, C.L., Li, M., Hu, M., Shi, Y. and Gu, W. (2007) The p53--Mdm2--HAUSP complex is involved in p53 stabilization by HAUSP. *Oncogene*, **26**, 7262-6.
48. Li, M., Brooks, C.L., Kon, N. and Gu, W. (2004) A dynamic role of HAUSP in the p53-Mdm2 pathway. *Molecular Cell*, **13**, 879-86.
49. Stevenson, L.F., Sparks, A., Allende-Vega, N., Xirodimas, D.P., Lane, D.P. and Saville, M.K. (2007) The deubiquitinating enzyme USP2a regulates the p53 pathway by targeting Mdm2. *The EMBO journal (European Molecular Biology Organization)*, **26**, 976-86.
50. Marine, J.C., Dyer, M.A. and Jochemsen, A.G. (2007) MDMX: from bench to bedside. *Journal of Cell Science*, **120**, 371-8.
51. Toledo, F. and Wahl, G.M. (2007) MDM2 and MDM4: p53 regulators as targets in anticancer therapy. *International Journal of Biochemistry and Cell biology*, **39**, 1476-82.
52. Bond, G.L. and Levine, A.J. (2007) A single nucleotide polymorphism in the p53 pathway interacts with gender, environmental stresses and tumor genetics to influence cancer in humans. *Oncogene*, **26**, 1317-23.
53. Wang, M., Zhang, Z., Zhu, H., Fu, G., Wang, S., Wu, D., Zhou, J. and Wei, Q. (2008) A novel functional polymorphism C1797G in the MDM2 promoter is associated with risk of bladder cancer in a Chinese population. *Clinical Cancer Research*, **14**, 3633-40.
54. Bond, G.L., Hu, W., Bond, E.E., Robins, H., Lutzker, S.G., Arva, N.C., Bargonetti, J., Bartel, F., Taubert, H., Wuerl, P., Onel, K., Yip, L., Hwang, S.J., Strong, L.C., Lozano, G. and Levine, A.J. (2004) A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans. *Cell*, **119**, 591-602.
55. Schultz, J.R., Petz, L.N. and Nardulli, A.M. (2005) Cell- and ligand-specific regulation of promoters containing activator protein-1 and Sp1 sites by estrogen receptors alpha and beta. *Journal of Biological Chemistry*, **280**, 347-54.
56. Wilkening, S., Bermejo, J.L. and Hemminki, K. (2007) MDM2 SNP309 and cancer risk: a combined analysis. *Carcinogenesis*, **28**, 2262-7.
57. Lind, H., Zienolddiny, S., Ekstrom, P.O., Skaug, V. and Haugen, A. (2006) Association of a functional polymorphism in the promoter of the MDM2 gene with risk of nonsmall cell lung cancer. *International Journal of Cancer*, **119**, 718-21.
58. Ben-Zaken Cohen, S., Pare, P.D., Man, S.F. and Sin, D.D. (2007) The growing burden of chronic obstructive pulmonary disease and lung cancer in women: examining sex differences in cigarette smoke metabolism. *American Journal of Respiratory and Critical Care Medicine*, **176**, 113-20.
59. Arva, N.C., Gopen, T.R., Talbott, K.E., Campbell, L.E., Chicas, A., White, D.E., Bond, G.L., Levine, A.J. and Bargonetti, J. (2005) A chromatin-associated and transcriptionally inactive p53-Mdm2 complex occurs in mdm2 SNP309 homozygous cells. *The Journal of Biological Chemistry*, **280**, 26776-87.
60. Benestad, H.B. and Laake, P. (2005) *Forskningsmetode i medisin og biofag*. Gyldendal Norsk Forlag, Oslo.
61. Farre, D., Roset, R., Huerta, M., Adsuara, J.E., Rosello, L., Alba, M.M. and Messeguer, X. (2003) Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Research*, **31**, 3651-3.
62. Messeguer, X., Escudero, R., Farre, D., Nunez, O., Martinez, J. and Alba, M.M. (2002) PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics*, **18**, 333-4.
63. Campbell, I.G., Eccles, D.M. and Choong, D.Y. (2006) No association of the MDM2 SNP309 polymorphism with risk of breast or ovarian cancer. *Cancer Letters*, **240**, 195-7.
64. Alhopuro, P., Ylisaukko-Oja, S.K., Koskinen, W.J., Bono, P., Arola, J., Jarvinen, H.J., Mecklin, J.P., Atula, T., Kontio, R., Makitie, A.A., Suominen, S., Leivo, I., Vahteristo, P., Aaltonen, L.M. and Aaltonen, L.A. (2005) The MDM2 promoter polymorphism SNP309T->G and the risk of uterine leiomyosarcoma, colorectal cancer, and squamous cell carcinoma of the head and neck. *Journal of Medical Genetics*, **42**, 694-8.
65. Bond, G.L., Menin, C., Bertorelle, R., Alhopuro, P., Aaltonen, L.A. and Levine, A.J. (2006) MDM2 SNP309 accelerates colorectal tumour formation in women. *Journal of Medical Genetics*, **43**, 950-2.

66. Galic, V., Willner, J., Wollan, M., Garg, R., Garcia, R., Goff, B.A., Gray, H.J. and Swisher, E.M. (2007) Common polymorphisms in TP53 and MDM2 and the relationship to TP53 mutations and clinical outcomes in women with ovarian and peritoneal carcinomas. *Genes, Chromosomes & Cancer*, **46**, 239-47.
67. Krekac, D., Brozkova, K., Knoflickova, D., Hrstka, R., Muller, P., Nenutil, R. and Vojtesek, B. (2008) MDM2SNP309 does not associate with elevated MDM2 protein expression or breast cancer risk. *Oncology*, **74**, 84-7.
68. Menin, C., Scaini, M.C., De Salvo, G.L., Biscuola, M., Quaggio, M., Esposito, G., Belluco, C., Montagna, M., Agata, S., D'Andrea, E., Nitti, D., Amadori, A. and Bertorelle, R. (2006) Association between MDM2-SNP309 and age at colorectal cancer diagnosis according to p53 mutation status. *Journal of the National Cancer Institute*, **98**, 285-8.
69. Schmidt, M.K., Reincke, S., Broeks, A., Braaf, L.M., Hogervorst, F.B., Tollenaar, R.A., Johnson, N., Fletcher, O., Peto, J., Tommiska, J., Blomqvist, C., Nevanlinna, H.A., Healey, C.S., Dunning, A.M., Pharoah, P.D., Easton, D.F., Dork, T. and Van't Veer, L.J. (2007) Do MDM2 SNP309 and TP53 R72P interact in breast cancer susceptibility? A large pooled series from the breast cancer association consortium. *Cancer Research*, **67**, 9584-90.
70. Talseth, B.A., Meldrum, C., Suchy, J., Kurzawski, G., Lubinski, J. and Scott, R.J. (2007) MDM2 SNP309 T>G alone or in combination with the TP53 R72P polymorphism does not appear to influence disease expression and age of diagnosis of colorectal cancer in HNPCC patients. *International Journal of Cancer*, **120**, 563-5.
71. Terry, K., McGrath, M., Lee, I.M., Buring, J. and De Vivo, I. (2008) MDM2 SNP309 is associated with endometrial cancer risk. *Cancer Epidemiology, Biomarkers & Prevention*, **17**, 983-6.
72. Walsh, C.S., Miller, C.W., Karlan, B.Y. and Koeffler, H.P. (2007) Association between a functional single nucleotide polymorphism in the MDM2 gene and sporadic endometrial cancer risk. *Gynecologic Oncology*, **104**, 660-4.
73. Hu, Z., Ma, H., Lu, D., Qian, J., Zhou, J., Chen, Y., Xu, L., Wang, X., Wei, Q. and Shen, H. (2006) Genetic variants in the MDM2 promoter and lung cancer risk in a Chinese population. *International Journal of Cancer*, **118**, 1275-8.
74. Hu, Z., Jin, G., Wang, L., Chen, F., Wang, X. and Shen, H. (2007) MDM2 Promoter Polymorphism SNP309 Contributes to Tumor Susceptibility: Evidence from 21 Case-Control Studies. *Cancer Epidemiology, Biomarkers and Prevention*, **16**, 2717-23.
75. Han, J.Y., Lee, G.K., Jang, D.H., Lee, S.Y. and Lee, J.S. (2008) Association of p53 codon 72 polymorphism and MDM2 SNP309 with clinical outcome of advanced nonsmall cell lung cancer. *Cancer*, **113**, 799-807.
76. Li, G., Zhai, X., Zhang, Z., Chamberlain, R.M., Spitz, M.R. and Wei, Q. (2006) MDM2 gene promoter polymorphisms and risk of lung cancer: a case-control analysis. *Carcinogenesis*, **27**, 2028-33.
77. Mollerup, S., Ryberg, D., Hewer, A., Phillips, D.H. and Haugen, A. (1999) Sex differences in lung CYP1A1 expression and DNA adduct levels among lung cancer patients. *Cancer Research*, **59**, 3317-20.
78. Fasco, M.J., Hurteau, G.J. and Spivack, S.D. (2002) Gender-dependent expression of alpha and beta estrogen receptors in human nontumor and tumor lung tissue. *Molecular and Cellular Endocrinology*, **188**, 125-40.
79. Berge, G., Mollerup, S., S, O.V., Hewer, A., Phillips, D.H., Eilertsen, E. and Haugen, A. (2004) Role of estrogen receptor in regulation of polycyclic aromatic hydrocarbon metabolic activation in lung. *Lung Cancer*, **45**, 289-97.
80. Kinyamu, H.K. and Archer, T.K. (2003) Estrogen receptor-dependent proteasomal degradation of the glucocorticoid receptor is coupled to an increase in mdm2 protein expression. *Molecular and Cellular Biology*, **23**, 5867-81.
81. Bond, G.L., Hirshfield, K.M., Kirchhoff, T., Alexe, G., Bond, E.E., Robins, H., Bartel, F., Taubert, H., Wuerl, P., Hait, W., Toppmeyer, D., Offit, K. and Levine, A.J. (2006) MDM2 SNP309 accelerates tumor formation in a gender-specific and hormone-dependent manner. *Cancer Research*, **66**, 5104-10.
82. Mittelstrass, K., Sauter, W., Rosenberger, A., Illig, T., Timofeeva, M., Klopp, N., Dienemann, H., Meese, E., Sybrecht, G., Woelke, G., Cebulla, M., Degen, M., Morr, H., Drings, P., Groeschel, A., Kreymborg, K.G., Haeussinger, K., Hoeffken, G., Schmidt, C.,

Jilge, B., Schmidt, W., Ko, Y.D., Taeuscher, D., Chang-Claude, J., Wichmann, H.E., Bickeboeller, H. and Risch, A. (2008) Early onset lung cancer, cigarette smoking and the SNP309 of the murine double minute-2 (MDM2) gene. *BMC Cancer*, **8**, 113.

83. Zhang, X., Miao, X., Guo, Y., Tan, W., Zhou, Y., Sun, T., Wang, Y. and Lin, D. (2006) Genetic polymorphisms in cell cycle regulatory genes MDM2 and TP53 are associated with susceptibility to lung cancer. *Human mutation*, **27**, 110-7.
84. Pine, S.R., Mechanic, L.E., Bowman, E.D., Welsh, J.A., Chanock, S.C., Shields, P.G. and Harris, C.C. (2006) MDM2 SNP309 and SNP354 are not associated with lung cancer risk. *Cancer epidemiology, biomarkers & prevention*, **15**, 1559-61.
85. Jun, H.J., Park, S.H., Lee, W.K., Choi, J.E., Jang, J.S., Kim, E.J., Cha, S.I., Kim, D.S., Kam, S., Kim, C.H., Kang, Y.M., Jung, T.H. and Park, J.Y. (2007) Combined effects of p73 and MDM2 polymorphisms on the risk of lung cancer. *Molecular Carcinogenesis*, **46**, 100-5.
86. Liu, G., Wheatley-Price, P., Zhou, W., Park, S., Heist, R.S., Asomaning, K., Wain, J.C., Lynch, T.J., Su, L. and Christiani, D.C. (2008) Genetic polymorphisms of MDM2, cumulative cigarette smoking and nonsmall cell lung cancer risk. *International Journal of Cancer*, **122**, 915-8.
87. Ruijs, M.W., Schmidt, M.K., Nevanlinna, H., Tommiska, J., Aittomaki, K., Pruntel, R., Verhoef, S. and Van't Veer, L.J. (2007) The single-nucleotide polymorphism 309 in the MDM2 gene contributes to the Li-Fraumeni syndrome and related phenotypes. *European Journal of Human Genetics*, **15**, 110-4.
88. Stoehr, R., Hitzenbichler, F., Kneitz, B., Hammerschmied, C.G., Burger, M., Tannapfel, A. and Hartmann, A. (2008) Mdm2-SNP309 polymorphism in prostate cancer: no evidence for association with increased risk or histopathological tumour characteristics. *British Journal of Cancer*, **99**, 78-82.
89. Dharel, N., Kato, N., Muroyama, R., Moriyama, M., Shao, R.X., Kawabe, T. and Omata, M. (2006) MDM2 promoter SNP309 is associated with the risk of hepatocellular carcinoma in patients with chronic hepatitis C. *Clinical Cancer Research*, **12**, 4867-71.
90. Tsuiki, H., Nishi, T., Takeshima, H., Yano, S., Nakamura, H., Makino, K. and Kuratsu, J. (2007) Single nucleotide polymorphism 309 affects murin-double-minute 2 protein expression but not glioma tumorigenesis. *Neurologia medico-chirurgica* **47**, 203-8; discussion 208-9.
91. Altman, D.G. and Bland, J.M. (2005) Standard deviations and standard errors. *BMJ*, **331**, 903.
92. Worster, A., Fan, J. and Ismaila, A. (2007) Understanding linear and logistic regression analyses. *Canadian journal of emergency medical care*, **9**, 111-3.