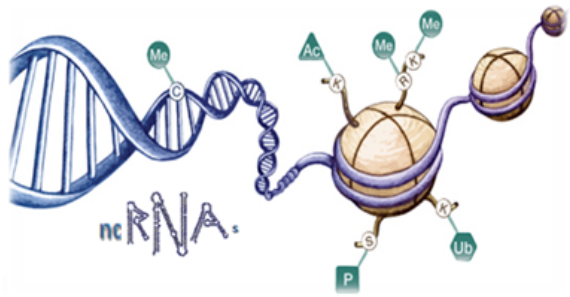
Μεταπτυχιακό Πρόγραμμα Σπουδών στην
"**Κυτταρική & Γενετική Αιτιολογία, Διαγνωστική & Θεραπευτική των Ασθενειών του Ανθρώπου**"
Τ.Θ. 2208, 71003 Ηράκλειο, Κρήτη

Graduate Program in
"**The Molecular Basis of Human Disease**"
P.O.Box 2208, 71003 Heraklion-Crete, Greece

Epigenetic Deregulation in Lung Cancer

Επιγενετική Απορρύθμιση στον Καρκίνο του Πνεύμονα



Paschalia Pantazi

MSc Dissertation

Submitted to the *Department of Medicine, University of Crete*

Supervisor: Prof. Aristides Eliopoulos

Herakleion, July 2015

Contents

Abstract

Despite the great advances in cancer research, lung cancer still remains the number one killer type of cancer worldwide, mainly due to lack of efficiency in early diagnosis. Typically, lung cancer bears both genetic and epigenetic abnormalities. Non-coding RNAs (ncRNAs) are epigenetic regulators frequently distorted in cancer. The aim of this project was dual: to examine the expression profile of previously identified targets and to perform a microarray-based profiling of lung cancer samples.

We assessed hsa-miR-7150 expression in Non Small Cell Lung Cancer (NSCLC). No difference in the levels of hsa-miR-7150 expression between tumour and normal samples was detected. However, hsa-miR-7150 expression was associated with histology (p=0.006) and tumour stage (p=0.038). We also investigated EGFR (Epidermal Growth Factor Receptor) mRNA expression in lung cancer and its potential regulation by 2 lncRNAs (LOC102723622 and EGFR-AS1) located within the EGFR locus. The expression of EGFR, EGFR-AS1 and LOC102723622 was assessed by real time qPCR in lung tumour and adjacent normal tissue samples. In addition, we analyzed the methylation status of EGFR gene promoter by bisulfite pyrosequencing. Only LOC102723622 showed significant elevated expression in tumours compared to normal lung tissue (median fold change = 3.5, p=0.019). Expression of EGFR was correlated to EGFR-AS1 expression (Rho=0.43, p=0.005). EGFR promoter was unmethylated in both tumour and normal tissues. Finally, we constructed a lncRNA microarray experiment. Analysis of the results pointed out some promising lncRNA biomarker candidates for further validation.

Overall, in this study we clarified that hsa-miR-7150 is not an ideal biomarker for lung cancer. Moreover, for the first time we analysed the methylation status of EGFR promoter and demonstrated a unique potential epigenetic regulation of EGFR through an intragenic lncRNA. However, the significance of LOC102723622 overexpression must be further investigated to ascertain its functional importance in lung cancer development.

Σύνοψη

Παρά τα σπουδαία επιτεύγματα στην έρευνα για τον καρκίνο, ο καρκίνος του πνεύμονα παραμένει ο νούμερο ένα θανατηφόρος τύπος καρκίνου παγκοσμίως, κυρίως λόγω της έλλειψης αποτελεσματικότητας για έγκαιρη διάγνωση. Τυπικά, ο καρκίνος του πνεύμονα φέρει τόσο γενετικές όσο και επιγενετικές ανωμαλίες. Τα μη κωδικά ριβονουκλεικά οξέα (ncRNAs) είναι επιγενετικοί ρυθμιστές, συχνά απορυθμισμένοι στον καρκίνο. Ο στόχος αυτής της εργασίας ήταν διττός: να εξετάσει την έκφραση συγκεκριμένων μορίων και να χαρακτηρίσει το μοριακο προφίλ δειγμάτων καρκίνου του πνεύμονα χρησιμοποιώντας μικροσυστοιχίες.

Εξετάσαμε την έκφραση του hsa-miR-7150 στον μη μικροκυτταρικό καρκίνο του πνέυμονα (NSCLC). Δεν εντοπίστηκε διαφορά στα επίπεδα έκφρασης του hsa-miR-7150 μεταξύ καρκινικών και φυσιολογικών δειγμάτων. Ωστόσο, παρατηρήθηκε συσχέτιση της έκφρασης του hsa-miR-7150 με τον ιστολογικό τύπο ($p = 0.006$) και το στάδιο του καρκίνου ($p = 0.038$). Διερευνήσαμε επίσης την έκφραση του υποδοχέα του επιδερμικού αυξητικού παράγοντα (EGFR) στον καρκίνο του πνεύμονα καθώς και την πιθανότητα ρύθμισής του από 2 μακρά μη κωδικά ριβονουκλεικά οξέα (lncRNAs - LOC102723622 και EGFR-AS1) που βρίσκονται εντός του γενετικού τόπου του EGFR. Η έκφραση των EGFR, EGFR-AS1 και LOC102723622 εκτιμήθηκε με αλυσιδωτή αντίδραση πολυμεράσης πραγματικού χρόνου (real time qPCR) σε καρκινικά και φυσιολογικά δείγματα ιστού του πνεύμονα. Μόνο το γονίδιο LOC102723622 έδειξε σημαντικά αυξημένη έκφραση στα καρκινικά σε σύγκριση με τα φυσιολογικά δείγματα (διάμεση μεταβολή = 3,5 φορές, $p = 0.019$). Παρατηρήθηκε μικρή συσχέτηση της έκφρασης του EGFR με την έκφραση του EGFR-AS1 ($Rho = 0.43$, $\rho = 0.005$). Επιπλέον, αναλύσαμε τα ποσοστά μεθυλίωσης του υποκινητή του γονιδίου του EGFR με πυραλληλούχηση (pyrosequencing). Ο υποκινητής του EGFR βρέθηκε μη-μεθυλιωμένος τόσο στα καρκινικά όσο και στα φυσιολογικά δείγματα ιστού. Τέλος, πραγματοποιήσαμε ένα πείραμα μικροσυστοιχιών βασισμένο στην έκφραση μακρών μη κωδικών ριβονουκλεικών οξέων. Η ανάλυση των αποτελεσμάτων ανέδειξε κάποια υποψήφια μακρά μη κωδικά ριβονουκλεικά οξέα ως βιοδείκτες για περαιτέρω ανάλυση .

Συνολικά, στην παρούσα μελέτη έχουμε απορρίψει το hsa-miR-7150 ως βιοδείκτη για τον καρκίνο του πνεύμονα, για πρώτη φορά, αναλύσαμε τα ποσοστά μεθυλίωσης του υποκινητή του γονιδίου του EGFR και προτίνουμε έναν τρόπο επιγενετικής ρύθμισης του EGFR μέσω ενδογονιδιακών μακρών μη κωδικών ριβονουκλεικών οξέων. Ωστόσο, η υπερέκφραση του LOC102723622 που εντοπίστηκε πρέπει να διερευνηθεί περαιτέρω ώστε να εξακριβωθεί η λειτουργική σημασία της στην ανάπτυξη του καρκίνου του πνεύμονα.

1 Introduction

1.1 Lung cancer

1.1.1 Epidemiology

Lung cancer is one of the most frequently diagnosed cancers in Europe[1] (**Figure 1.**) and the US[2], accounting for 13% of all new cases of cancer worldwide. Although lung cancer incidence rates began to decline after 1990s as a result of reduced smoking prevalence, it still remains the leading cause of cancer-related mortality worldwide. 1.75 million deaths from cancer were reported in Europe in 2012, 353,000 of them were due to lung cancer.[1] Less than 10% of people with lung cancer survive the disease for at least five years after diagnosis.[3] This high mortality rate can be explained, to an extent, as lung cancer symptoms become apparent only in late stages of the disease and chronic treatment is followed by a fast onset of chemoresistance. Therefore, early detection and curative treatment are hard to achieve.[4]

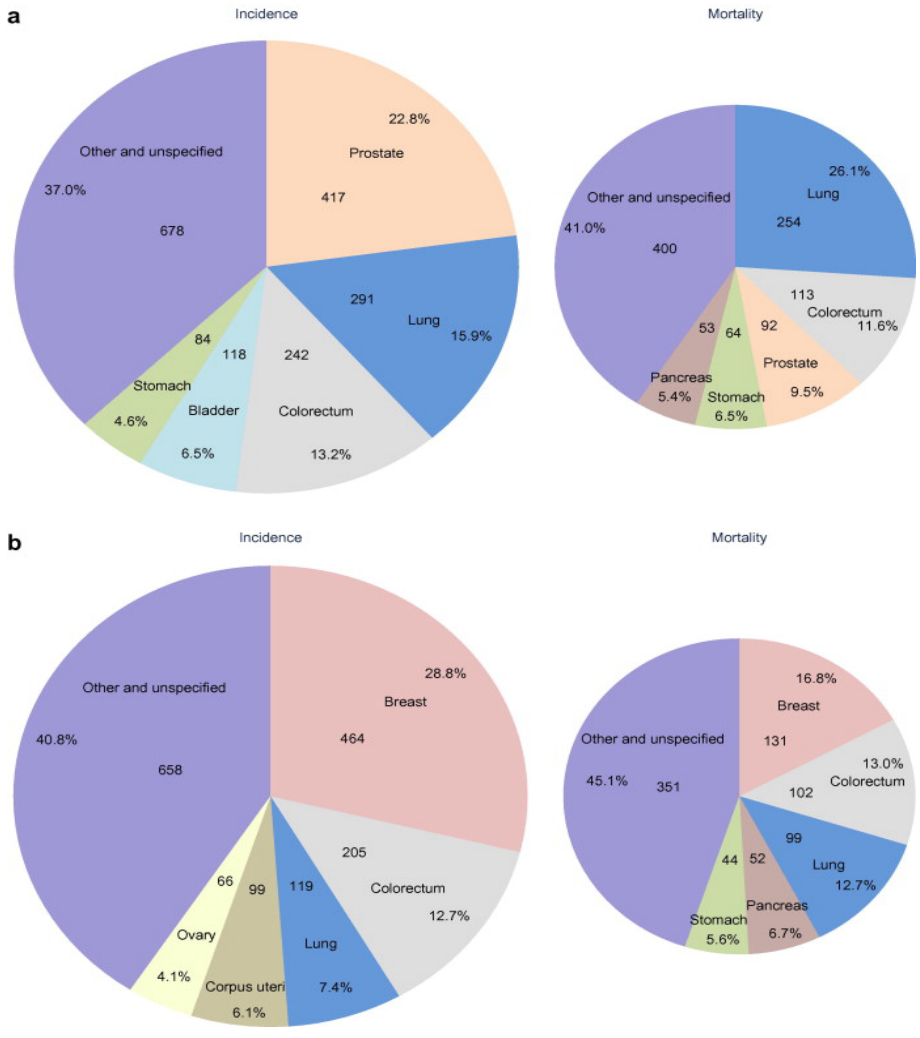


Figure 1. Distribution of the expected cases and deaths for the 5 most common cancers in Europe 2012 in males (a) and females (b) (total number of cases or deaths is given as well). Adapted from Ferlay et al.

1.1.2 Risk factors

Cancer is a multifactorial disease, resulting from the combined influence of genetic factors acting in concert with environmental insults. Lung cancer incidence increases with age, as mutations and gene defects accumulate. 89% of lung cancer cases each year in the UK are linked to major lifestyle and other risk factors.[3]

Smoking

Cancers caused by smoking are, to a high degree, preventable cancers. Tobacco smoking is responsible for approximately 90% of lung cancer cases in the UK.[3] A single cigarette contains a mixture of thousands of compounds, most of which are well-established carcinogens. These carcinogens are metabolised in the body and the metabolites that are produced are cleared up through a detoxification process. Generally, people with a higher activation and lower detoxification capacity are at higher risk for smoking-related cancers. The metabolic activation of carcinogens results in the formation of DNA adducts, which are absolutely central to the carcinogenic process. However, some carcinogens can directly form DNA adducts without metabolic activation.[5]

Since smoking was first linked to lung cancer in 1950,[6] many studies have shown that it is mainly associated with squamous cell carcinoma of the lung rather than other types of lung cancer, suggesting that different cancer types might arise due to different aetiology.[7,8]

Passive smoking is a great issue especially when it comes to children. Sidestream smoke, which comes from the burning tip of the cigarette, is about 4 times more toxic than mainstream smoke (smoke directly inhaled through the mouth end of the cigarette), although people inhale it in a more diluted form. This is because sidestream smoke contains much higher levels of many of the poisons and cancer-causing chemicals in cigarettes.[3]

Electronic cigarettes are the new trend in smoking cessation. Their safety though is still questioned. They are safer than conventional cigarettes, and therefore acceptable to use as a means of quitting smoking, however their popularity increases especially in people of young age, raising an alert.[9]

Radon

Decay of uranium, thorium, and radium in rocks and soil results in radon release. Radon is an invisible and odourless gas that seeps up through the ground and diffuses into the air. Depending on local geology in some places, radon dissolves into ground water. Radon decays quickly, giving off tiny radioactive particles which when inhaled, can damage the lung.[10] Long-term exposure to radon can

lead to lung cancer. An estimated 9% of lung cancer deaths in Europe are linked to indoor radon exposure.[11]

Occupational Exposures

Occupational risk factors play significant roles among the possible causes of lung cancer and are potentially preventable. Asbestos, arsenic, silica, radon and many other industrial agents are reported to be carcinogenic to humans. Various professions have reported increased lung cancer incidence (**Table 1.**).[12]

Table 1. Occupations and industries known to present an excess risk of lung cancer. Adapted from De Matteis et al.

Industry	Occupation/Process/Chemicals
Agriculture	Vineyard workers using arsenical insecticides (before 1970)
Mining and quarrying	Arsenic, uranium, iron-ore, granite, and asbestos mining; talc mining/milling
Granite production	Cutting, polishing, etc., of granites stones
Ceramic and refractory brick	Ceramic and pottery workers
Asbestos production	Insulating material production
Metals (iron and steel basic industries)	Iron and steel founding
Metals (non-ferrous basic industries)	Copper, zinc, cadmium, alluminium, nickel, chromates, beryllium
Shipbuilding, motor vehicle, railroad equipment manufacturing	Shipyard and dockyard, motor vehicle, railroad manufacture workers
Gas	Coke plant workers and gas production workers
Construction	Insulators and pipe coverers, roofers, asphalt workers
Other	Painters (construction, automotive industry, and other users)

Health Issues

Many diseases have been associated with increased lung cancer incidence. For example, lung cancer risk is higher in rheumatoid arthritis patients and up to double in people with systemic lupus erythematosus, compared to the general population.[13] COPD (Chronic Obstructive Pulmonary Disease),[14] Diabetes,[15] HIV (Human Immunodeficiency Virus) infection,[16] Pneumonia[14] and Obesity[17] have also been associated with high risk of lung cancer. On the other hand, people suffering from Alzheimer's[18], Parkinson's[19] or coeliac disease[20] have lower chances of developing lung cancer. Several studies showed that higher levels of physical activity may lead to reduction in lung cancer risk. This includes activity at work, in the household, and leisure activity. At the same time, consumption of cruciferous vegetables, citrus fruits and foods with a high total carotenoid content and vitamin C lowers the risk for lung cancer especially in women. Consumption of red meat and poultry, though, has been associated with high risk of cancer.[3]

Genetic make-up

Cancer is a multifactorial disease and involves deregulation of many genes, in addition to environmental influences. A person has a genetic predisposition or susceptibility for a disease when he/she while not born with a disease may be at high risk of acquiring it. The genetic susceptibility to cancer is mainly due to the presence of one or more gene mutations, and/or a combination of alleles responsible for cell multiplication and repair.[21]

Having a first-degree relative with lung cancer increases the risk of having the disease by 51%. The risk is even greater if a brother/sister has lung cancer.[3] A cohort study in Japanese middle-aged population suggested that people with a family history of lung cancer are more likely to acquire the disease themselves. Moreover, hereditary predisposition seemed to affect more female than male population. According to the same study, having a first-degree relative with lung cancer, significantly increases the risk of lung cancer incidence by 95%.[22]

1.1.3 Histological Classification

Lung cancer is divided into two main categories according to the histological characterization of the tumours: Small Cell Lung Cancer (SCLC) and Non-Small Cell Lung Cancer (NSCLC)(**Figure 2.**). SCLC is the rarer type, accounting for only one in 8 lung cancers. It is a neuroendocrine tumour type with cells that are smaller in size than most other cancer cells. It is a fast-growing cancer with a highly metastatic potential. The most common, NSCLC type can be further grouped into Squamous Cell Carcinomas (SqCCL), adenocarcinomas (AdenoCa) and Large Cell Carcinomas. SqCCL begins in the thin, flat cells that line the passages of the respiratory tract and makes up approximately 30% of lung cancer diagnoses. Adenocarcinoma begins in the cells that form the lining of the lungs, has gland-like properties and makes up just over 30% of lung cancer diagnoses. Large cell carcinoma is poorly differentiated, grows fast and makes up about 9% of lung cancer diagnoses.[23,24]

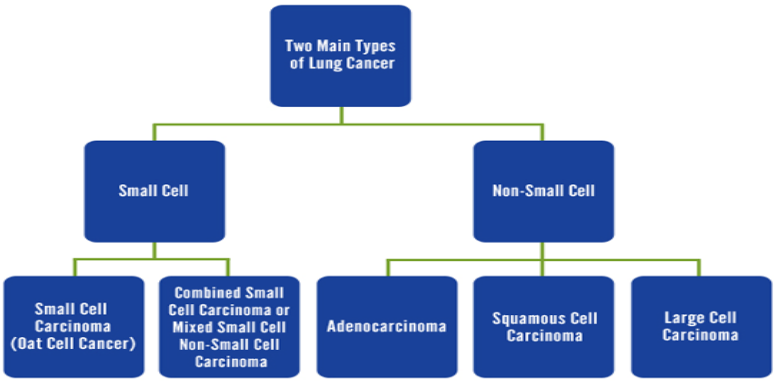


Figure 2. Histological classification of lung cancer. Adapted from the American Lung Association.

1.1.4 Lung cancer staging

Staging is the process of describing the extent to which cancer has spread from its site of origin. It is used to assess a patient's prognosis and determine the choice of therapy. There are different ways of staging lung cancer.[25,26] The most commonly used staging systems are (**Table 2.**):

a) *the number staging system*, according to which lung cancer is described by a number, zero (0) through four (IV):
   - Stage 1 – cancer is small and localised in one area of the lung (localized)
   - Stage 2 and 3 – cancer is larger and may have grown into the surrounding tissues and there may be cancer cells in the lymph nodes (locally advanced)
   - Stage 4 – cancer has spread to another part of the body (secondary or metastatic cancer)

b) *the TNM staging system* that utilizes the letters T, N and M to assess tumours by:
   - the size of the primary tumour (T);
   - the degree to which regional lymph nodes (N) are involved and
   - the absence or presence of distant metastases (M) - cancer that has spread from the original (primary) tumour to distant organs or distant lymph nodes.

Table 2. Staging of lung cancer using both number and TNM staging systems. Adapted from http://www.cancernews.com/data/Article/265.asp.

		Tumor (T)				Mets (M)
		T1	T2	T3	T4	M1
Nodes (N)	N0	IA	IB	IIB	IIIB	IV
	N1	IIA	IIB	IIIA	IIIB	IV
	N2	IIIA	IIIA	IIIA	IIIB	IV
	N3	IIIB	IIIB	IIIB	IIIB	IV
Mets (M)	M1	IV	IV	IV	IV	IV

1.2 Epigenetics

The word 'epigenotype' was first introduced by Waddington in 1942 to describe the interim between genotypes and phenotypes.[27] Nowadays, the term "Epigenetics" is used to define heritable changes in phenotype or gene expression that do not affect the primary DNA sequence. Epigenetic modifications can be grouped into 4 major categories: DNA (de)methylation, histone modifications, nucleosome remodelling, and non-coding RNAs (ncRNAs) (**Figure 3.**).[28] These mechanisms interact with each other and follow a precise molecular orchestration in order to maintain or change epigenetic marks throughout the genome; this crosstalk seems to be pivotal for various processes like normal development and differentiation, maintenance of tissue specific gene expression patterns in mammals, X-chromosome inactivation, genomic imprinting and suppression of transposable elements.[29]

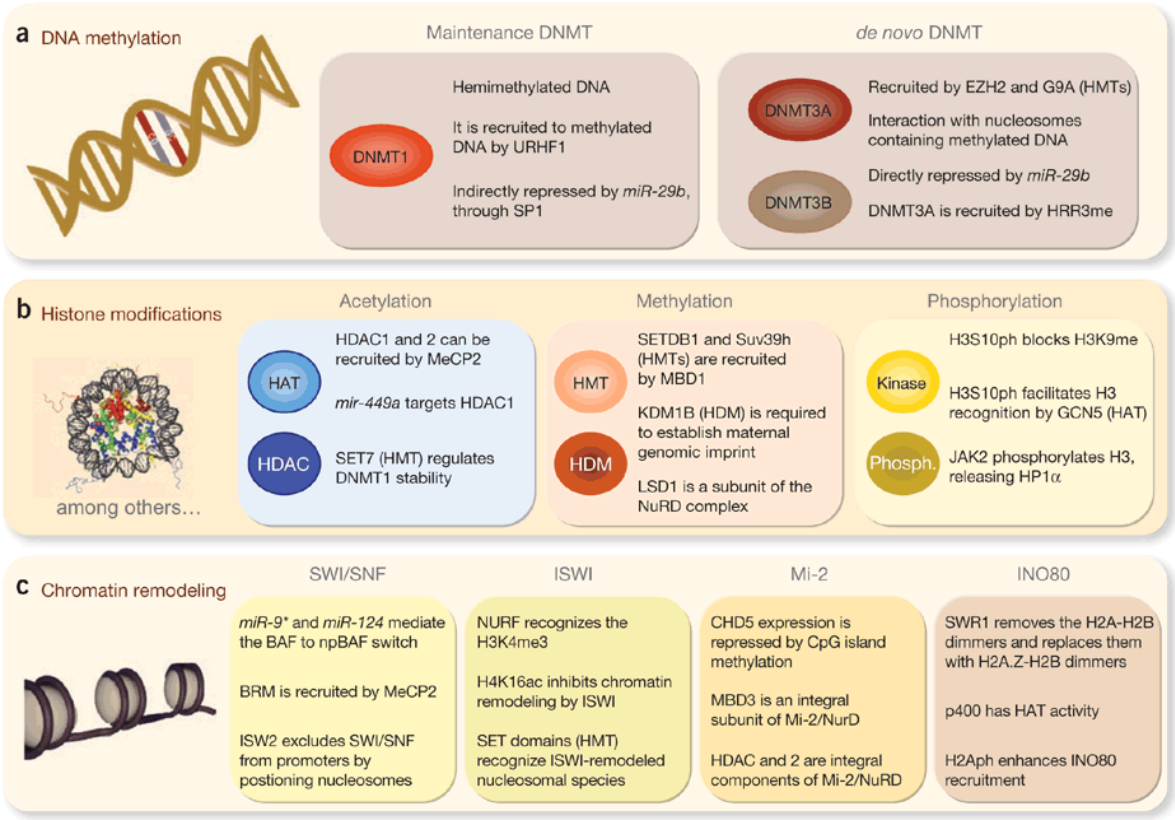


Figure 3. Epigenetic marks are catalysed by different epigenetic complexes, whose principal families are illustrated here. (a–c) Epigenetic regulation depends on the interplay among the different players: DNA methylation (a), histone marks (b) nucleosome positioning (**c**) and non-coding RNAs (not shown). The interaction among the different factors brings about the final outcome. This figure illustrates selected examples of the possible interrelations among the various epigenetic players. Adapted from Portela et al.

1.2.1 DNA methylation

DNA methylation is the most stable and most thoroughly studied epigenetic modification in humans. It is defined as the covalent addition of a methyl group (CH3) on the 5'position of cytosine within CpG (deoxycytidine-phosphate-deoxyguanosine) dinucleotides leading to the formation of 5-methylcytosine (5-mC). Distribution of CpG dinucleotides is asymmetric in the human genome, with CpG rich and CpG poor regions and although they make up less than 1% of genomic DNA (gDNA), CpGs are present in approximately 60% of human genes.[30] Recently, DNA methylation has also been described in CpG island surrounding regions termed as "CpG shores" and "CpG shelves" (up to 2kb and 4kb from the CpG island, respectively).[31]

It is widely approved that DNA methylation is associated with condensed chromatin structure and gene silencing. DNA methylation-associated proteins read the methylation patterns and in turn facilitate the recruitment of histone modifiers and chromatin-remodelling complexes.[32] DNA methylation is catalysed by DNA methyltransferases (DNMTs); de novo methylation is supported by DNMT3a and 3b[33] while DNMT1[34] is responsible for maintenance of tissue specific methylation patterns. Demethylation can occur passively, when maintenance of methylation cannot be achieved or actively through a multistage process. So far the existence of a DNA demethylase is still unknown but several enzymes have been shown to be involved in catalysing the conversion of 5mC to cytosine through the 5-hydroxymethylcytosine (5-hmC) intermediate compound. The first oxidation of 5-mC to 5-hmC is carried out by TET1 (Ten-eleven translocation methyl-cytosine dioxygenase 1) while several ways of thereafter transforming 5-hydroxymethyl cytosine to cytosine have been suggested and involve both nuclear and base excision repair pathways.[35]

Recent publications report the presence of adenine methylation alongside with cytosine methylation in mammals, indicating a major field broadening in the near future.[36]

1.2.2 Histone Modifications

Histones are basic proteins that associate with DNA in the nucleus and help condense it into chromatin. There are five major families of histones: H1/H5, H2A, H2B, H3 and H4. Histones H2A, H2B, H3 and H4 are known as the core histones, while histones H1 and H5 are the linker histones. The core histones form octamers, around which DNA is wrapped forming the nucleosome. Histones contain a globular C-terminal domain and an unstructured N-terminal tail. The latter undergoes a number of post-translational covalent modifications including acetylation, methylation, phosphorylation, deamination, ADP ribosylation, ubiquitination, SUMOylation and others. These modifications are implicated in essential physiological processes such as transcription, DNA

replication, DNA repair and chromosome condensation. Histone modifications affect these processes by changing the accessibility of chromatin.[37]

For example, histone acetylation is one of the most well-investigated histone modifications that tightly regulate chromatin accessibility and it consists of the addition of an acetyl-group on the amino group of lysine (K) residues of histones. In this way the positive charge of histones is neutralized, thus decreasing their interaction with the negatively charged phosphate groups of DNA. As a result, chromatin acquires a more relaxed structure that allows transcription factors and other enzymes to initiate transcription. Acetylation levels are controlled by Histone Acetylases (HATs) and Histone Deacetylases (HDACs).[38]

1.2.3 Nucleosome Remodelling

The combinatory effect of epigenetic mechanisms yields the final outcome of open or closed chromatin, thus allowing or disallowing factors to bind DNA and drive transcription. In general, active genes are frequently found in nucleosomal depleted regions (NDRs), flanked by nucleosomes containing the histone variant H2A.Z, instead of the core histone protein H2A, and trimethylated H3K4. Inactivation of gene expression is achieved via the action of Polycomb proteins and repressive histone modifications like trimethylation of H3K27 (H3K27me3). In this context, gene promoters are methylated and are occupied by nucleosomes with trimethylated histones H3 at the residues of Lys 9 (H3K9me3), and HDACs are attracted to the region. Transcription factor binding is disallowed as the recognized DNA sequence has changed due to DNA methylation or increased density of nucleosome particles themselves.[39,40] It has been shown that long non-coding RNAs also regulate the interaction between chromatin and remodelling complexes such as PRC2 (Polycomb Repressor Complex 2) by bringing them to specific genomic loci and shifting their position when necessary.[41]

1.2.4 Non-coding RNAs

Non-coding RNAs are functional RNA molecules that are not translated into proteins. They are roughly divided into two major categories; short ncRNAs (<200 nt) and long ncRNAS (>200 nt). Short ncRNAs include microRNAs (miRNAs), piwi-interacting RNAs (piRNAs), small nucleolar RNAs (snoRNAs) and small nuclear RNAs (snRNAs), among others. Long ncRNAs (lncRNAs) may act as signals, guides, scaffolds or decoys in order to regulate gene expression in various ways and include ribosomal ncRNA (rRNA), XIST RNA and others. Both groups are known to play a role in heterochromatin formation, histone modification, DNA methylation targeting, and gene silencing.[42] In this study we will mainly focus on miRNAs and lncRNAs.

MicroRNAs

MicroRNAs are non-coding RNA molecules, about 21 - 25 nucleotides in length. They are partially complementary to one or more messenger RNA (mRNA) molecules, and their main function is to control gene expression in a variety of manners, including translational repression, mRNA cleavage, and deadenylation (**Figure 4.**).[43] Since the first miRNA was found in C.Elegans 12 years ago,[44] the list of miRNAs identified through random cloning and sequencing or computational prediction in the human genome and their potential target genes is growing rapidly.[45]

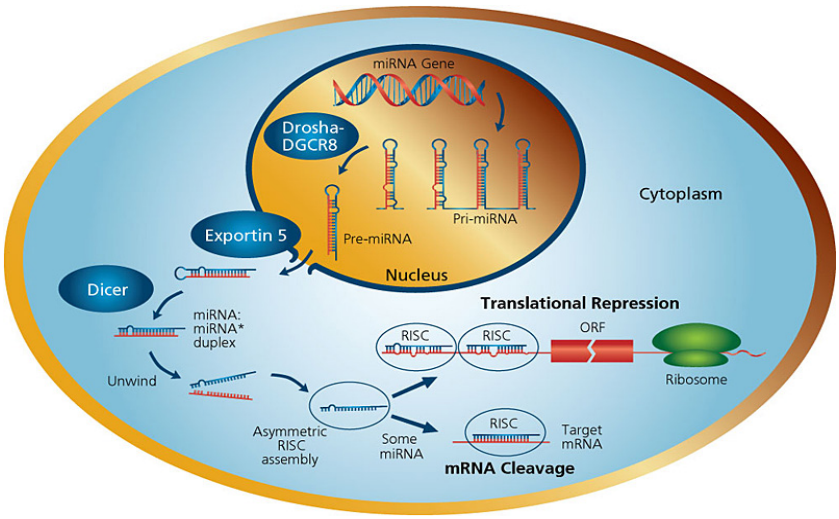


Figure 4. miRNA biogenesis and functions. Adapted from http://www.sigmaaldrich.com/life-science/functional-genomics-and-rnai/mirna/learning-center/mirna-introduction.html

MicroRNAs can be found in the genome as individual genes or clusters of several miRNAs that have common regulation. They can be found inside protein coding gene introns or exons or in non-protein coding RNAs. Their biogenesis starts when Drosha (RNAse type III enzyme) processes the long double-stranded primary transcripts (pri-miRNA), to form a hairpin baring precursor (pre-miRNA) (**Figure 4.**). The pre-miRNA is exported to the cytoplasm where it undergoes an additional processing step by the RNAse III enzyme Dicer generating the miRNA, a double-stranded RNA approximately 22 nt in length. The strand with lower stability base pairing of the 2 - 4 nt at the 5' end of the duplex preferentially associates with the RNA-induced silencing complex (RISC) and thus becomes the active miRNA. The active miRNA is situated in the RISC, which facilitates degradation of the target mRNA through direct cleavage or inhibition of protein synthesis. It has been suggested that transcripts may be regulated by multiple miRNAs and an individual miRNA may target numerous transcripts.[43]

In this study we will mainly focus on the hsa-miR-7150.

hsa-miR-7150

Oulas et al.[46] have introduced a highly accurate tool called SSCprofiler, which can be used to identify novel miRNA gene candidates in the human genome. Combining both analytical and experimental

techniques, they revealed the mature sequence of a novel miRNA gene within a cancer associated genomic region. In spite of the fact that hsa-miR-7150 existence was reported in 2009, still its complete characterization or role in disease is missing from the literature.

Long non-coding RNAs

LncRNAs constitute a diverse group of ncRNAs, with various functions and limited conservation amongst species. The majority of lncRNAs are 5' capped, 3' polyadenylated, baring multiple exons and are subjected to transcriptional regulation as coding mRNAs. They are frequently transcribed by RNA polymerase II, although transcription by RNA polymerase III has also been reported and they can show complex splicing patterns. LncRNAs can be localized in sense or antisense orientation to protein-coding genes, within introns or in intergenic regions of the genome.

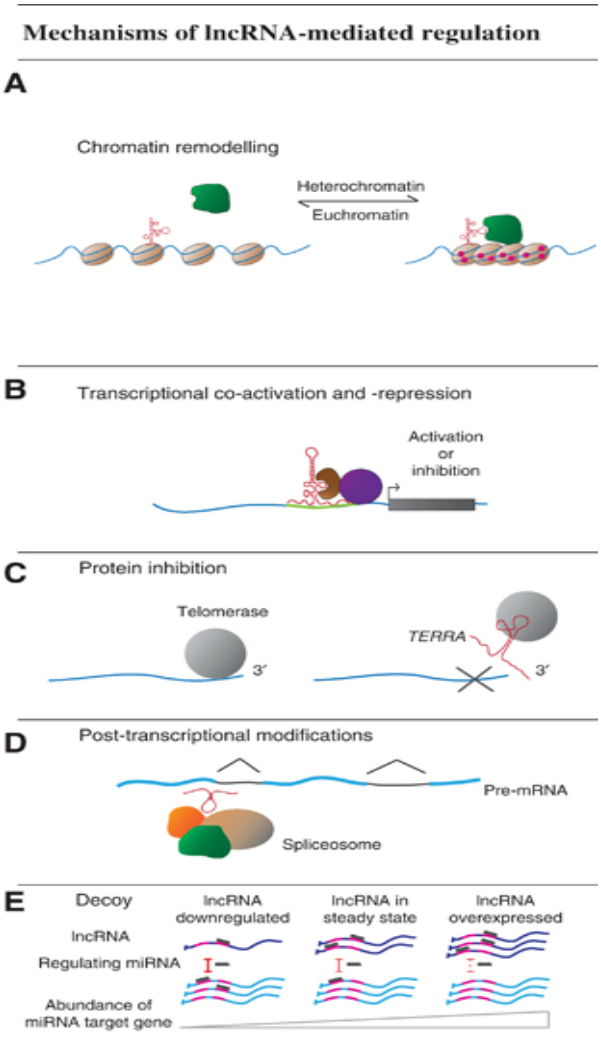


Figure 5. Generalised mechanisms of lncRNAs functions. LncRNAs act through a variety of mechanisms such as remodelling of chromatin (**A**), transcriptional co-activation or -repression (**B**), protein inhibition (**C**), as post-transcriptional modifiers (**D**) or decoy elements (**E**). Adapted by Cheetham et al.

LncRNAs appear to be able to regulate gene expression through a diverse group of mechanisms (**Figure 5.**). While many lncRNAs may function in cis (through the act of their transcription), a significant proportion of lncRNAs have intrinsic RNA-mediated functions in trans. There are lncRNAs that can perform epigenetic silencing in cis: lncRNA transcripts such as Xist downregulate their expression by becoming inaccessible to transcription factors. The same lncRNAs can also recruit chromatin remodelling proteins to the region. Epigenetic silencing in trans: lncRNAs such as HOTAIR can interact with chromatin modifying proteins to epigenetically silence genes at another locus. lncRNAs can activate or repress transcription. The act of lncRNA transcription itself opens chromatin of a genetic locus and permits access to the transcription machinery to transcribe other protein coding genes. Furthermore, lncRNA transcripts can activate transcription factors via allosteric interactions. On the other

hand, transcription of lncRNAs near protein coding genes can repress expression because the lncRNAs "occupy" the machinery and physically prevent transcription of the protein coding gene. Moreover, if a lncRNA sequence overlaps with a transcription factor binding site, the lncRNA transcript can hybridize to this site to prevent binding. Finally, various lncRNAs may be processed into short ncRNAs such as siRNAs that can downregulate gene expression by degrading the mRNA transcripts. [47-51]

In this study we will mainly focus on two lncRNAs: EGFR-AS1 and LOC102723622.

EGFR-AS1 (EGFR AntiSense RNA 1)

Hillier et al. used mRNA isolated from tissues representing various developmental states in order to obtain tag sequences (ESTs) from different genes and to deposit them in the publicly accessible Data Base for Expressed Sequence Tags. In this attempt, they identified EGFR-AS1, a lncRNA encoded from the complement strand of EGFR (Epidermal Growth Factor Receptor) gene (**Figure 6.**). Despite its promising location (EGFR is a major player in lung cancer development) no further studies exist on EGFR-AS1.

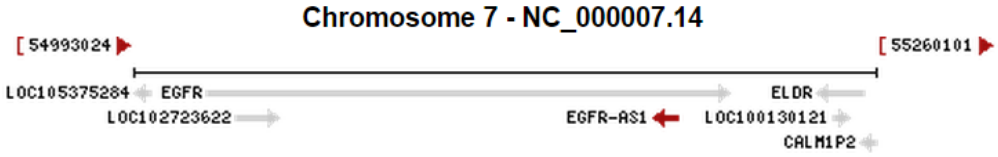


Figure 6. EGFR-AS1 genomic locus. Adapted from http://www.ncbi.nlm.nih.gov/gene/100507500.

LOC102723622

LOC102723622 is a lncRNA within the first intron of EGFR gene (**Figure 7.**). It consists of two exons. No studies have been performed on LOC102723622 so far.

Figure 7. LOC102723622 is located in the first exon of EGFR gene. EGFR transcripts are depicted with green colour. Adapted from http://www.ncbi.nlm.nih.gov/gene/?term=LOC102723622.

1.3 Epigenetic Modifications in Lung Cancer

The epigenome of a cell is maintained through cell divisions. Failure to do so, can potentially lead to activation or inhibition of signalling pathways resulting in catastrophic events such as carcinogenesis. Cancer cells are characterized by global changes in DNA methylation, histone modification patterns, chromatin-modifying enzyme expression and non coding RNA profiles. Aberrant epigenetic landscape has been documented in lung cancer. Multiple studies have focused on the development of epigenetic biomarkers for early detection, metastatic risk assessment, and therapeutic stratification of patients.

The high frequency of DNA methylation aberrations in lung cancer has been reported in a large number of studies. Among the most thoroughly studied genes for DNA methylation in lung cancer are RASSF1, RARβ, p16 (CDKN2A), MGMT, SHOX2, APC, BRCA1, and DAPK, providing a fair list of candidates with biomarker potential. DNA methylation has been assessed for miRNAs as well; hsa-miR-34a, hsa-miR-9-2, hsa-miR-9-3 and hsa-miR-193 were reported to be hypermethylated in NSCLC tissues.[52] Global genome hypomethylation has also been shown in lung cancer contributing to genomic instability, associating with poor outcome and triggering overexpression of aberrant transcripts.[53] The only approved DNA methylation marker for lung cancer diagnosis is the SHOX2 gene methylation assay, which has been used in clinical practice. SHOX2 methylation was identified as a biomarker capable of reliably differentiating between lung tumour and normal tissues.[54]

The most prominent alteration in histone modification in cancer cells is a global reduction of H4K16ac. The main class of HDACs implicated in this process is the Sirtuin family of proteins that seem to be either mutated or overexpressed in cancer.[29] Histone modification analysis has been suggested to be able to discriminate tumour samples according to their histology, providing promising diagnostic markers for lung cancer. Changes in the overall levels of histone modifications can also influence prognosis. Loss of H4K20me3 predicts lower survival in patients with stage I adenocarcinoma.[55] The epigenetic pattern of H3K4me2, H2AK5ac and H3K9ac influences the clinical outcome of NSCLC patients, especially in early-stage tumours.[56] The cellular levels of H3K4me2 and H3K18ac have also been reported as predictors of clinical outcome in lung and other adenocarcinomas.[57]

Transcription activator BRG1 is part of the large ATP-dependent chromatin remodelling complex SWI/SNF, which is required for transcriptional activation of genes normally repressed by chromatin. In vivo studies have shown that inactivation of BRG1 enhances the development of lung adenomas

in presence of carcinogens.[58] In humans, loss of heterozygosity and point mutations of BRG1 genomic loci have been described, conferring reduced survival to lung cancer patients.[59,60]

While our understanding of miRNAs expression patterns and function in normal physiology and disease has just started to emerge, a growing number of miRNAs has already been found aberrantly expressed in lung cancer.[61] Several attempts to develop miRNA based biomarker panels for lung cancer have been made. Bediaga et al. have recently developed a prediction algorithm based on miRNA expression that can assist clinical diagnosis of lung cancer in minimal biopsy material to improve clinical management. The panel of miRNAs expression signature had very high sensitivity (97.5%) and specificity (96.3%), providing a future promising tool for the clinic.[62] Lately, miRNAs have been isolated from patients' body fluids such as blood, urine, sputum etc. and a considerable effort has been devoted to decode the information carried by circulating miRNAs. Body fluids represent an attractive clinical material because they can be collected using non-invasive or minimally invasive procedures. The first comprehensive analysis of circulating miRNAs of patients with NSCLC was performed by Chen et al. when they sequenced all serum miRNAs of healthy individuals and cancer patients. Two NSCLC-specific miRNAs—miR-25 and miR-223 were present at significantly higher levels in the sera from cancer patients than controls. However, diligent validation is needed in order for their diagnostic value to be determined.[63]

An increasing number of lncRNAs has been associated with lung cancer (**Table 3.**). The best studied lncRNAs in lung cancer are HOTAIR (HOX transcript antisense RNA) and MALAT1 (metastasis-associated lung adenocarcinoma transcript 1). Elevated expression of HOTAIR is correlated with invasion, metastasis, and resistance to cisplatin in patients with lung cancer. In cancer, elevated expression of HOTAIR shifts PRC2-mediated gene repression from oncogenes to tumour-suppressors. Therapeutically, HOTAIR inhibition combined with traditional chemotherapy could potentially sensitize drug resistant tumours.[64] Beyond its value as a metastasis prediction marker, MALAT1 has an active role in lung cancer metastasis by tuning up the expression of metastasis-associated genes. Studies have shown that targeting MALAT1 expression in human xenograft tumours with antisense oligonucleotides drastically reduces lung cancer metastasis in vivo and raises MALAT1 as a potential therapeutic target in lung cancer.[65] BCYRN1 (Brain CYtoplasmic RNA 1)[66], HNF1A-AS1 (HNF1a AntiSense RNA 1)[67], ANRIL (Antisense Non-coding RNA in the INK4 Locus)[68] and a continuous list of newly described lncRNAs have been associated with lung cancer metastasis and exhibit great potential as therapeutic targets in lung cancer management.

Table 3. Lung cancer associated lncRNAs, the cell process they are involved in and associated clinical features. Adapted from Loewen et al.

lncRNA	Intersecting molecules and pathways	Cell processes	Associated clinical features
AK126698	Reduces NKD2, activates β-catenin	Anti-apoptosis, resistance to cisplatin	Unknown
CARLo-5	Unknown	Cell cycle, proliferation, invasion, EMT	↑ in NSCLC, lymph node metastasis, poor survival
CCAT2	Unknown	Proliferation, migration, invasion	↑ in LAC, lymph node metastasis
H19	Induced by cigarette smoke	Unknown	↑ in NSCLC , poor survival
HOTAIR	Induced by Col-1 . Affects expression of gelatinases . Represses cell-adhesion genes , $p21^{waf1}$, and HOXA5	Proliferation, migration invasion resistance to cisplatin in vitro & in vivo	↑ in NSCLC, lymph node and brain metastasis, poor survival . ↑ in cisplatin-refractory LAC . ↑ in SCLC, lymphatic invasion, relapse
LCAL1	Unknown	Proliferation	↑ in NSCLC
MALAT1	Affects expression of Bcl-2 and metastasis related genes	EMT tumor growth in vivo survival	↑ in NSCLC, brain metastasis, poor survival . ↑in periphery blood of NSCLC
MVH	Affects expression of MMP-2/-9	Proliferation & invasion	↑ in LAC and LSCC, advanced TNM stage, lymph node metastasis, poor prognosis
SCAL1	Induced by cigarette smoke and NRF2	Protection against oxidative stress	↑ NSCLC
SOX2ot	Affects expression of EZH2	Cell cycle, proliferation	↑ in LSCC, poor survival
ZXF1	Antisense to ACTA2	Migration & invasion	↑ in LAC, lymph node metastasis, advanced TNM stage, poor survival
BANCR	Inhibits the expression of EMT markers	Induces apoptosis, inhibits EMT, migration, invasion, metastasis in vivo	↓ in LAC and LSCC, lymph node metastasis, advanced TNM stage, poor survival
GAS6-AS1	Antisense to and represses expression of GAS6	Unknown	↓ in NSCLC, advanced TNM stage, poor survival
MEG3	Induces p53	Inhibits proliferation & growth in vivo, pro-apoptosis	↓ in NSCLC, advanced TNM stage, poor survival
SPRY4-IT1	Intronic to SPRY4, silenced by EZH2	Inhibits invasion, growth & metastasis in vivo, induces apoptosis	↓ in NSCLC, pathological stage, lymph node metastasis
TARID	Activates TCF21 via GADD45A	Unknown	↓ in LAC and LSCC
TUG1	Induced by p53, represses HOXB7 via PRC2	Inhibits proliferation & growth in vivo	↓ in NSCLC, advanced TNM stage, poor survival

1.4 Aims and objectives

The aim of this study was to ascertain the level of deregulation of ncRNAs in lung cancer in both an agonistic and a hypothesis based fashion.

Specific objectives:

- to determine the role of hsa-miR-7150 in NSCLC
- to study the expression profile of EGFR-related lncRNAs in NSCLC
- to perform a microarray-based lncRNA expression profiling of NSCLC

2 Materials and Methods

2.1 Primers

Primer design for expression and sequencing analysis

We designed primers and probes for real time PCR using the OLIGO 7 software (Molecular Biology Insights)(**Figure 8.**). For each primer or primer set, OLIGO 7 displays a multitude of important parameters that one needs to take into account when designing primers, such as DNA secondary structure, dimer formation, false priming and homology, internal stability, composition and physical properties. Although OLIGO 7 can automatically select primers we customized the search protocols (scoring system) in detail, in order to optimize the results according to our specific needs every time.[69] Oligonucleotides were ordered from MWG (**Table 5.**), dissolved in TEG solution (10mM Tris-HCl ph 8.0, 1 mM EDTA, 50% Glycerol) and stored at -20°C. After optimization optimal primer/probe concentrations as well as thermal profiles were determined to ensure maximum sensitivity, specificity and reproducibility of the assay.

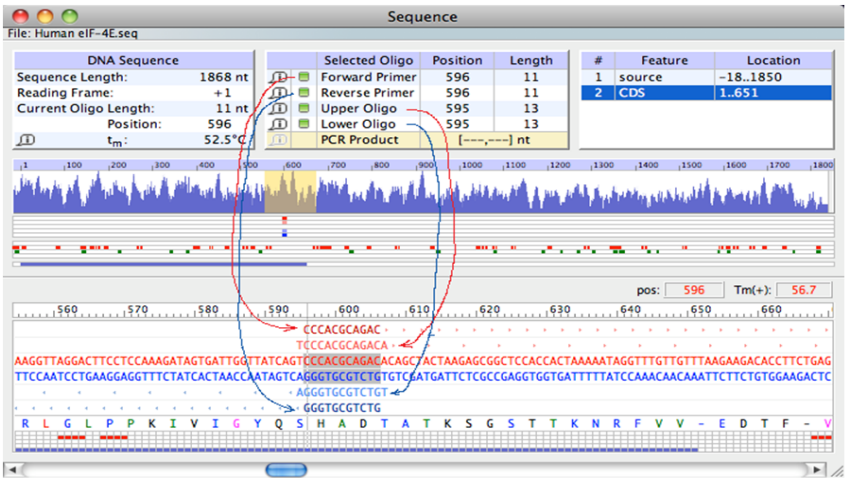


Figure 8. The Oligo 7 software. Adapted from http://www.oligo.net/.

We also used off the self assays bought by Life Technologies that comprise already optimized primer set and probe (**Table 4.**).

Table 4. Off the shelf TaqMan Expression Assays used in our experiments.

Assay	Part Number
EGFR	HS01076086_M1
hsa-miR-7150	PN4398987
RNU48	PN4448887

Table 5. Primers and probes designed with OLIGO 7 for our study.

Primer/Probe Name	Sequence 5'→3'	Modification	Tm (°C) @ 200 nM
ACTBexp-F	GGCACCCAGCACAATGAAG	none	58.7
ACTBexp-R	CATACTCCTGCTTGCTGATCCA	none	58.9
ACTBexp -P	CTCCTCCTGAGCGCAAGTACTCCGTG	5' TAMRA-3' NFQ	68.8
TBPexp-F	GGGGAGCTGTGATGTGAAGTTT	none	59.1
TBPexp-R	AAACCAGGAAATAACTCTGGCTCA	none	59.2
TBPexp-P	AAGGCCTTGTGCTCACCCACCAAC	5' TAMRA-3' NFQ	68.0
EGFR_AS1_F	TCTGGAATTCAGGTCACCTTTTAC	none	50.0
EGFR_AS1_R	TCAGTCCTGCCGAGTCATTC	none	50.1
EGFR_AS1_P	AGGCGCTCACCCAA	5' FAM-3' MBG	38.6
L1027-F	ACAACATGAAGAATCAGTTAATGA	none	52.1
L1027-R	AAGTGCAACTGGTCAGGAA	none	52.7
L1027-P	TTGCCTGGACTTTCTAAGTGCCC	5' FAM-3' NFQ	62.5
TREGseq-F	GTACCACTTCCTACCCTCGTAA	none	55.9
TREGseq-R	TATGGCTGATTATGATTATGATCCTCTG	none	61.3

*F : Forward primer, R: Reverse primer, P: Probe

*FAM, TAMRA: fluorescent dyes

*NFQ (Non Fluorescent Quencher), MBG (Minor Groove Binder): non fluorescent quenchers

Primer design for methylation analysis

Analysis of DNA methylation by pyrosequencing needs a good PCR amplification yield, making assay design crucial. Pyromark Assay Design 2.0 (QIAGEN) was used to design primers for methylation analysis (**Figure 9.**). The sequence of interest was downloaded from Genebank (NCBI) and an *in silico* bisulfite conversion was carried out by the software. An important issue is the inclusion of bisulfite control sites within the

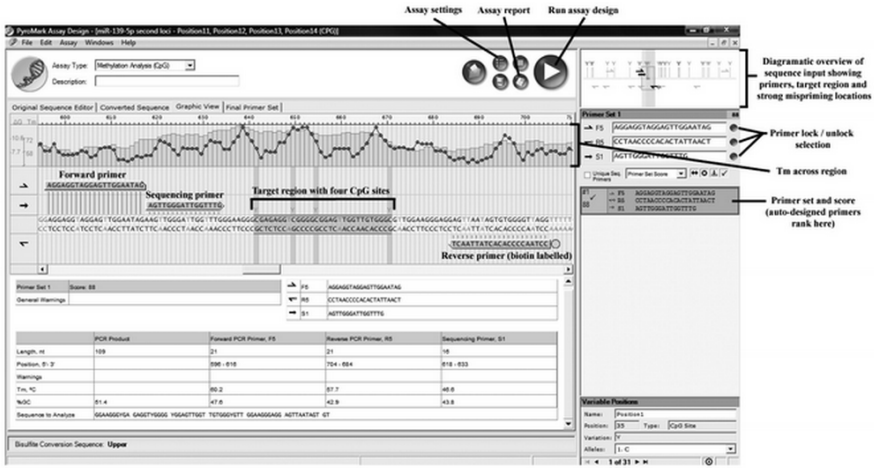


Figure 9. Pyromark Assay Design 2.0 software interface. Adapted from Colyer et al.

sequenced region. These are preferably ACA and GCA (converted to ATA and GTA accordingly) trinucleotides. Incorporation of C at these sites during pyrosequencing indicates poor bisulfite conversion. The software ranks primer sets for the target sequence. A primer set score of >85 is considered to be excellent, however, primer sets of >70 can work very well.[70] Oligonucleotides were also ordered from MWG (**Table 6.**), dissolved in TEG solution and stored at -20°C.

Table 6. Primers for methylation analysis designed with Pyromark Assay Design 2.0 software.

Primer Name	Sequence 5'→3'	Modification	Tm (°C) @ 200 nM
EGFRmeth-F	GGAGAGAGTTTTTAGGGGAGTAG	none	55.5
EGFRmeth-Rb	CTTTCTTTTCCTCCAAAACC	5'-biotinylation	54.1
EGFRmeth-S	GAGTTTTTAGGGGAGTA	none	42.6

*F : Forward primer, R: Reverse primer, S: Sequencing primer

2.2 RNA Extraction

Figure 10. Precellys 2 ml Soft Tissue Homogenizing Ceramic Beads Kit and the homogenizer (Cayman Chemical) (up) and QIAshredder columns (QIAGEN) (down).

For fresh frozen tissue homogenization we added 700µl TRI Reagent (Zymo Research) in the frozen sections. Then, we either transferred the mixture in the Precellys tubes (Precellys 2 mL Soft Tissue Homogenizing Ceramic Beads Kit, Cayman Chemical) and quickly vortexed them for 30 seconds, or passed them through QIAshredder columns (QIAGEN) (**Figure 10.**). For the cell lines, we added 700µl TRI Reagent in an approximately 80% confluent flask, we scraped the bottom of the flask to detach cells, collected them in a new 1.5ml tube and vortexed.

RNA extraction was performed using two kits; miRNeasy Mini Kit (QIAGEN) or Direct-zol™ RNA MiniPrep kit (Zymo Research).

The miRNeasy Mini Kit (QIAGEN) allows for small RNAs isolation (has previously been used in the lab for miRNAs isolation) and the protocol we used was the following:

1. The tubes containing the homogenate were left on the bench top at room temperature (15–25°C) for 5 minutes.
2. 140µl chloroform were added to the tubes containing the homogenate and capped securely. Vigorous shake for 15 seconds.

3. The tubes were left at room temperature for 2–3 minutes on the bench top.
4. Centrifugation for 15 minutes at 12,000 x g at 4°C.
5. The upper aqueous phase was transferred into a new collection tube and 1.5 volumes (usually 525µl) of 96% ethanol were added and mixed thoroughly by pipetting up and down several times.
6. Up to 700µl of the sample were pipetted, including any precipitate that might have formed, into an miRNeasy Mini spin column in a 2ml collection tube. Centrifugation at 12000 x g for 1 minute at room temperature (15–25°C). Flow-through was discarded. We repeated this step with the remainder of the sample.
7. 700µl Buffer RWT were added to the miRNeasy Mini spin column. Centrifugation for 1 minute at 12000 x g to wash the column.
8. 500µl Buffer RPE was added into the miRNeasy Mini spin column. Centrifugation for 1 minute at 12000 x g to wash the column. We repeated this step.
9. Dry spin; the miRNeasy Mini spin column was placed on a new Collection Tube and centrifuged for 3 minutes at 12000 x g.
10. The miRNeasy Mini spin column was placed on a new 1.5 ml collection tube and 30µl RNase-free water were added directly onto the miRNeasy Mini spin column membrane. Centrifugation for 3 minutes at 12000 x g to elute RNA.

RNA extraction from fresh frozen tissue and cell lines for gene and lncRNA expression experiments was isolated using the Direct-zol™ RNA MiniPrep kit (Zymo Research) as follows:

1. One volume of 96% ethanol (approximately 750µl) was added directly to one volume sample homogenate (1:1) in TRI Reagent and mixed well by vortexing.
2. The mixture was loaded into a Zymo-Spin™ IIC Column and centrifuged (14000 x g) for 1 minute. The column was then transferred into a new Collection Tube and the Collection Tube containing the flow-through was discarded.
3. At this point, we treated the samples with DNase I in-column:
    - The column was washed with 400µl RNA Wash Buffer and centrifuged (14000 x g) for 1 minute and flow-through was discarded.
    - 80µl DNase I Reaction Mix were added directly to the column matrix:
        - DNase I 5 µl (6 U/µl)
        - DNA Digestion Buffer 75 µl
    - The column was incubated at 20-30°C for 15 minutes.

4. Then 400µl Direct-zol™ RNA PreWash were added to the column. Centrifugation (14000 x g) for 1 minute. Flow-through was discarded and this step was repeated.
5. 700µl RNA Wash Buffer were added to the column. Centrifugation (14000 x g) for 1 minute. Flow-through and Collection Tube were discarded.
6. The column was carefully transferred into a new Collection Tube and centrifuged (14000 x g) for 3 minutes (dry spin).
7. Finally, the column was transferred in a new 1.5ml tube and 50µl DNase/RNase-Free Water were added directly to the column matrix. We let it stand at RT for 2 minutes and centrifuged (14000 x g) for 2 minutes.
   RNA was kept in -80°C or -20°C for short storage.

2.3 RNA quantification and quality control

Agilent Bioanalyser 2100 - Agilent RNA 6000 Nano (Agilent Technologies)

This system is essentially a miniaturized version of gels used to separate nucleic acids or proteins for analysis. Briefly, samples are combined with a fluorescent dye and injected into wells in the chip. The samples move through a gel matrix in the microchannels and are separated by electrophoresis. The samples then are detected by fluorescence, and electropherograms and gel-like images are created by the data analysis software for sizing and quantification (**Figure 11.**). RNA integrity and quantity are measured. To determine RIN values (RNA Integrity Number), the instrument software uses an algorithm that takes into account the electrophoretic trace of RNA. The RIN scale ranges from 0 to 10, with 10 indicating maximum RNA integrity.

Protocol:

We allowed all reagents to equilibrate to room temperature for 30 minutes before use.

1. 550 µl of Agilent RNA 6000 Nano gel matrix (red) were placed into the top receptacle of a spin filter.
2. Spin for 10 minutes at 1500 x g.
3. The RNA 6000 Nano dye concentrate (blue) was vortexed for 10 seconds and spun down.
4. 1 µl of RNA 6000 Nano dye concentrate (blue) was added to 65 µl of filtered gel, vortexed thoroughly and spun for 10 minutes at 14000 x g.

Figure 11. Agilent Priming Station and Bioanalyser.

5. A new RNA Nano chip was placed on the chip priming station.
6. 9 µl of the gel- dye mix were added at the bottom of the well marked "**G**".
7. The plunger was positioned at 1 ml and then the chip priming station was closed. The lock of the latch clicks when the priming station is closed correctly.
8. The plunger of the syringe was pressed down held by the clip.
9. We waited for exactly 30 seconds and then released the plunger with the clip release mechanism.
10. The chip priming station was opened and 9 µl of the gel- dye mix was pipetted in the other wells marked "G".
11. 5 µl of the RNA 6000 Nano marker (green) were added into the well marked with the ladder symbol and each of the 12 sample wells.
12. Ladder and samples were denatured (70 °C for 2 minutes and then ice).
13. 1 µl of the RNA ladder was added into the well marked with the ladder symbol and 1 µl of each sample into each of the 12 sample wells.
14. The chip was then placed in the IKA vortex mixer and vortexed for 60 seconds at 2400 rpm.
15. Finally the chip was inserted in the Agilent 2100 Bioanalyzer and the appropriate assay was ran (Eukaryote Total RNA nano).

Nanodrop 2000 (Thermo Scientific)

Spectrophotometer for minimal sample quantity (**Figure 12.**). The patented sample retention system used by NanoDrop 2000 (Thermo Scientific) allows for the analysis of 0.5 µl - 2.0 µl samples, without the need for cuvettes or capillaries.

1. With the arm opened, 0.5µl of sample or solvent (blank) was pipetted directly onto the pedestal.
2. Arm was closed and a sample column was formed.
3. The pedestal then moved to automatically adjust for an optimal path length (0.05 mm - 1 mm).
4. When the measurement was complete, the surfaces are simply wiped with a lint-free lab wipe before going on to the next sample.

Figure 12. Nanodrop 2000 (Thermo Scientific).

2.4 Microarray

44 pairs of tumour and adjacent normal lung tissue with RIN values greater than 7.5 were normalized at 70ng/µl and sent to CGR (Centre for Genomic Regulation, University of Liverpool) in dry ice. A mix of SqCCL and adenocarcinoma samples was present. The tumours were mostly tumour stage 2 (T2), nodal stage from 0 to 2 (N0-N2), and both genders were represented equally. Some of the patients had repeated data (duplicate samples as controls).

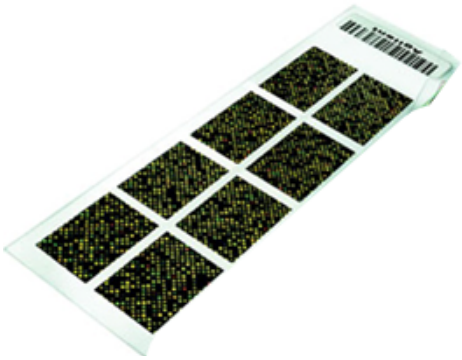


Figure 13. 6 Agilent microarray chips were used in our experiment.

Array data was generated by the CGR personnel. 6 Agilent slides each containing 8 arrays were used to identify tumour-associated lncRNA expression signatures (**Figure 13.**). RNA from tumours and adjacent normal lung tissue was applied to two-colour arrays. Each array had tumour samples in the Cy3 channel and normal in the Cy5 channel. The array type used was Agilent Technologies design 047718 which was generated by Gencode consortium (the v15 human lncRNAs annotations, see http://www.gencodegenes.org/lncrna_microarray.html). Each lncRNA transcript had 2 probes and there were 22001 targeted transcripts and 17535 randomly selected protein coding targets on the array. The samples were labelled from 1-88 with odd numbers being normal tissue and even numbers being tumour tissue. The array contained probes for mRNAs, lncRNAs, anti-sense transcripts etc. The differentially expressed genes were checked against publications in NCBI (The National Center for Biotechnology Information).

2.5 Reverse Transcription

Reverse transcription is the synthesis of single-stranded DNA (cDNA) using single-stranded RNA as a template, mediated by the enzyme reverse transcriptase. There are three main RNA priming strategies: a) random oligo-mers, b) oligo(dT), c) specific primers. Both oligo(dT)s and random oligo-mers allow many different targets to be studied from the same cDNA pool, while specific primers reverse transcribe a specific target. Reverse transcription for lncRNAs and genes expression was performed using anchored oligo(dT)s, an advanced type of oligo(dT)s that apart from T residues they carry combinations of other residues that help the annealing at the 3´UTR/polyA junction. Reverse transcription of miRNAs was performed using customized assays by Life technologies (technique is described below).

All reactions were run in a Bench-top thermal cycler GENEAMP SYSTEM 9700 (Applied Biosystems).

Reverse transcription for lncRNAs and genes expression was performed using the High Capacity cDNA Reverse Transcription Kit (Life Technologies) as follows:

1. 300ng of each RNA sample along with nuclease-free water up to 10μl total volume, were placed in strips, heated up at 70°C for 5 minutes and then placed on ice for at least 2 minutes. This would allow denaturation of RNA and would ease primer annealing.
2. The following mix was prepared (**Table 7.**), mixed and spun down. 10μl of the mix were added in each RNA sample from step 1 and the thermal profile (**Table 7.**)was performed.

Table 7. Reverse transcription for lncRNAs and genes expression was performed using the High Capacity cDNA Reverse Transcription Kit (Life Technologies).

Component	Quantity per sample
dNTPs (100mM)	0.8 μl
MultiScribe™ Reverse Transcriptase, (50 U/μL)	1 μl
10X Reverse Transcription Buffer	2 μl
anchored oligo(dT) (100 μM)	0.5 μl
Nuclease-free water	5.7 μl

Step	Time	Temperature
Hold	10 min	25°C
Hold	120 min	37°C
Hold	5 min	85°C
Hold	∞	4°C

3. 30 μl of ddH_2O (double distilled water) were added to the cDNAs. Samples were stored at -20°C.

We used off the shelf TaqMan® MicroRNA Assay (Life Technologies) to evaluate the expression of hsa-miR-7150 and RNU48 (small nucleolar RNA used as endogenous control).

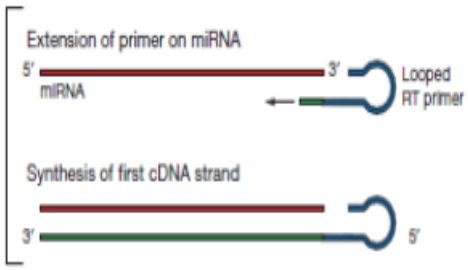


Figure 14. TaqMan® MicroRNA Assays technology for reverse transcription of miRNAs.

miRNAs are very short sequences to facilitate the primer set and probe needed in standard TaqMan® Assay-based real-time PCR (e.g. for gene expression). To address this challenge, TaqMan® MicroRNA Assays employ a target-specific stem-loop reverse transcription of the short length mature miRNA (**Figure 14.**). The primer extends the 3' prime end of the target to produce an adequate template where primers and probes can be used as in standard TaqMan® Assay-based real-time PCR. What is more, the stem-loop structure in the tail of the primer allows for specific detection of the mature, biologically active miRNA.

We prepared 15µl reactions as shown in **Table 8**. Denaturation of RNA here is not necessary due to the miRNAs' limited length. The conditions given (**Table 8.**) are the optimized conditions for both assays (hsa-miR-7150 and RNU48).

Table 8. Conditions for reverse transcription of miRNAs using TaqMan® MicroRNA Assays.

Component	Quantity
100mM dNTPs	0.15 µl
MultiScribe™ Reverse Transcriptase, 50 U/µL	1 µl
10X Reverse Transcription Buffer	1.5 µl
RNase Inhibitor, 20 U/µL	0.19 µl
5X RT primer hsa-miR-7150	3 µl
5X RT primer RNU48	3 µl
Nuclease-free water	Up to 15 µl
RNA	50 ng

Step	Time	Temperature
Hold	30 min	16 °C
Hold	30 min	42 °C
Hold	5 min	85 °C
Hold	∞	4

60 µl of ddH_2O were added to the samples after reverse transcription. Samples were stored at -20°C.

2.6 Real Time qPCR

The Real-time PCR method allows detection of PCR amplification during early phases of the reaction. Accumulation of the amplicon is detected during the reaction and data is measured at the exponential phase. TaqMan Assay-based real-time PCR utilizes a set of primers, a probe, polymerase with exonuclease activity, dNTPs and DNA template (**Figure 15.**). The probe consists of two types of fluorophores. While the probe is intact, the quencher (Q) is close to and reduces the fluorescence from the reporter fluorophore (R) (Fluorescence Resonance Energy Transfer (FRET)). Taq polymerase then amplifies the product and removes (5' exonuclease activity) the TaqMan probe from the template DNA. This separates the quencher from the reporter, and allows the reporter to give off its energy and the emitted light is captured by a camera. The more times the denaturing and annealing takes place, the more opportunities there are for the TaqMan probe to bind and, in turn, the more emitted light is detected. In addition, TaqMan probes allow multiple genes to be measured at the same sample

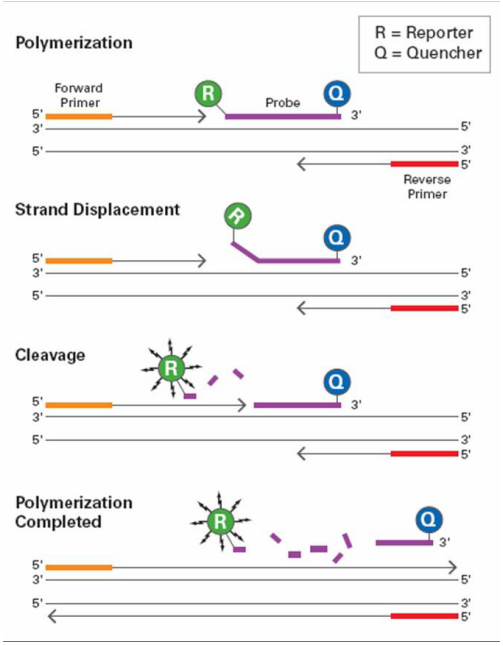


Figure 15. TaqMan® Assay-based real-time PCR.

(multiplex PCR), since fluorescent dyes with different emission spectra may be attached to different probes. Multiplex PCR allows internal controls to be co-amplified in a single tube.

All reactions were run in StepOnePlus, 7500 Fast or ViiA7 Real-Time PCR Systems (Life Technologies). The TaqMan Gene Expression Master mix II (Life Technologies) and QuantiTect (Qiagen) kits were used. TaqMan assays were either bought as 20x probe-primer mixes from Life Technologies (EGFR) or probe mixes were prepared as shown in **Table 9**.

Table 9. Probe mixes used in our expression analysis experiments.

20x ACTB probe mix	
Component	**Quantity**
ACTBexp-F	15 µM
ACTBexp-R	15 µM
ACTBexp-P	5 µM

20x TBP probe mix	
Component	**Quantity**
TBPexp-F	18 µM
TBPexp-R	18 µM
TBPexp-P	5 µM

20x EGFR-AS1 probe mix	
Component	**Quantity**
EGFR-AS1-F	18 µM
EGFR-AS1-R	18 µM
EGFR-AS1-P	5 µM

20x LOC102723622 probe mix	
Component	**Quantity**
L1027-F	18 µM
L1027-R	18 µM
L1027-P	5 µM

F : Forward primer, R: Reverse primer, P: Probe

EGFR expression

Table 10. Real Time PCR conditions used in EGFR expression experiments.

Component	**Quantity per sample**
TaqMan Gene Expression Master Mix II (2x)	10 µl
20x EGFR TaqMan Assay	1 µl
20x ACTB probe mix	1 µl
ddH_2O	6 µl
cDNA	2 µl

Step	**Time**	**Temperature**
Hold	2 min	50°C
Hold	10 min	95°C
40 cycles	15 sec	95°C
	1 min	60°C*

*Measurements were taken at this stage

EGFR-AS1 expression

Table 11. Real Time PCR conditions used in EGFR-AS1 expression experiments.

Component	**Quantity per sample**
QuantiTect master mix (2x)	10 µl
20x EGFR-AS1 probe mix	1 µl
20x TBP probe mix	1 µl
ddH_2O	4 µl
cDNA	4 µl

Step	**Time**	**Temperature**
Hold	15 min	95°C
45 cycles	15 sec	95°C
	30 sec	57°C
	1 min	60°C*

*Measurements were taken at this stage

LOC102723622 expression

Table 12. Real Time PCR conditions used in LOC102723622 expression experiments.

Component	Quantity per sample
QuantiTect master mix (2x)	10 µl
20x LOC102723622 probe mix	1 µl
20x TBP probe mix	1 µl
ddH_2O	4 µl
cDNA	4 µl

Step	Time	Temperature
Hold	15 min	95°C
45 cycles	15 sec	95°C
	30 sec	52°C
	1 min	60°C*

*Measurements were taken at this stage

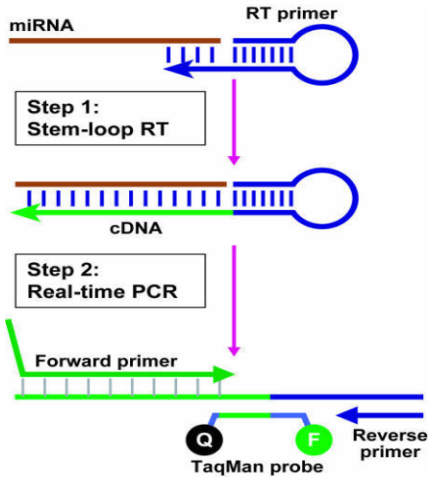


Figure 16. Taqman MicroRNA Assays technology.

TaqMan MicroRNA Assays employ a target-specific stem-loop reverse transcription primer to address the challenge of the short length of mature miRNA (as already mentioned in the Reverse Transcription section) (**Figure 16.**). The primer extends the 3' end of the target to produce a template that can facilitate the annealing of primers and probe as in standard TaqMan Assay-based real-time PCR. The chemistry behind this method was described above. For the miRNA expression we used off the self assays bought from Life Technologies. Multiplex reaction cannot be achieved as miRNAs are all labelled with the same fluorescent molecule. Therefore, separate reactions for hsa-miR-7150 and the endogenous control (normalizer) RNU48 were performed as shown in **Table 13** and **Table 14**.

hsa-miR-7150 expression

Table 13. Real Time PCR conditions used in hsa-miR-7150 expression experiments.

Component	Quantity per sample
TaqMan Gene Expression Master Mix II (2x)	7.5 µl
20x hsa-miR-7150 TaqMan Assay	0.5 µl
ddH_2O	6 µl
cDNA	3 µl

Step	Time	Temperature
Hold	2 min	50°C
Hold	10 min	95°C
45 cycles	15 sec	95°C
	1 min	62°C*

*Measurements were taken at this stage

RNU48 expression

Table 14. Real Time PCR conditions used in RNU48 expression experiments.

Component	Quantity per sample
TaqMan Gene Expression Master Mix II (2x)	7.5 µl
20x RNU48 TaqMan Assay	0.5 µl
ddH_2O	6 µl
cDNA	3 µl

Step	Time	Temperature
Hold	2 min	50°C
Hold	10 min	95°C
40 cycles	15 sec	95°C
	1 min	60°C*

*Measurements were taken at this stage

Data analysis

The ΔΔCt method was used to analyse all expression experiments. With this method, Cts (Cycle thresholds) for the gene of interest (GOI) in both the test sample and calibrator sample are adjusted in relation to a normalizer (norm) gene Ct from the same two samples. The resulting ΔΔCt (Δ represents the difference) value is incorporated to determine the fold difference in expression.

$$\text{Fold difference} = 2^{-\Delta\Delta Ct}$$

$$\Delta Ct \text{ sample} - \Delta Ct \text{ calibrator} = \Delta\Delta Ct$$

$$Ct\ GOI^{sample} - Ct\ norm^{sample} = \Delta Ct \text{ sample}$$

$$Ct\ GOI^{calibrator} - Ct\ norm^{calibrator} = \Delta Ct \text{ calibrator}$$

2.7 DNA Extraction

1. The sample (fresh frozen lung tissue) was briefly centrifuged.
2. The appropriate amount of lysis buffer (400 mM Tris-HCl pH 8.0, 10 mM EDTA, 150 mM NaCl, 1% SDS) was added, depending on the amount of sample.
3. Proteinase K (10mg/ml) was added to final concentration of 100µg/ml followed by overnight incubation at 50°C.
4. If the sample was lysed, RNase A (25mg/ml) was added to final concentration of 50µl/ml. Incubation at 37°C for 1 hour.
5. Proteinase K (10mg/ml) was added to final concentration of 100µg/ml. Incubation at 50°C for 2-3 hours.
6. Subsequently the samples were left at room temperature to cool down.

7. An equivalent volume of phenol-chloroform in 1:1 ratio was added. Gently mixing. Centrifugation at 3900 x g for 5 min. The upper, aqueous layer was then transferred into a new tube.
8. Again an equivalent volume of phenol-chloroform in 1:1 ratio was added. Gently mixing. Centrifugation at 3900 x g for 5 min. The upper, aqueous layer was then transferred into a new tube.
9. Isopropanol in 1:1 ratio was added into the tube with the aqueous phase. Gently mixing. Incubation at 20^{o}C overnight.
10. Centrifugation in a cooling centrifuge at 3900 x g for 15 min. The supernatant was poured off.
11. The samples were vacuum dried for 1 hour.
12. 500µl of 0.5 x TE were added. The samples were either left at room temperature for 1 hour to dissolve or stored at 4^{o}C for future use.

2.8 Agarose Gel Electrophoresis

In order to check and identify the PCR products agarose gel electrophoresis was performed.

1. In a conical flask, 2 or 1 % agarose was added into 0.5x TBE.
2. The mixture was placed in the microwave until it was boiled resulting in a crystal clear solution.
3. Then, 5% SafeView (3 mg/ml, NBS Biologicals) was added and mixed well mounted on an orbital shaker for about 4 minutes in order to cool down.
4. The gel mix was transferred to the casting tray and a comb or two were placed on it.
5. The gel was left to set.
6. The combs were removed from the set gel and the gel along with the tray was placed into an electrophoresis tank.
7. 0.5x TBE was added into the tank to cover the gel.
8. 5 µl PCR product mixed with 2 µl loading buffer (usually CoralLoad buffer, Qiagen) and loaded into the wells.
9. The gel ran at 80 V for approximately 45 min.
10. After the run was completed, the bands were visualized on a UV transilluminator at 365nm.

2.9 Methylation Analysis

2.9.1 Bisulfite Treatment

When DNA is treated with sodium bisulfite, unmethylated cytosines are converted to uracils and consequently to thymines after PCR, while methylated cytosines remain unaffected (**Figure 17.**).[71] Therefore, bisulfite treatment provides single nucleotide resolution information dependent on the methylation status of a segment of DNA.

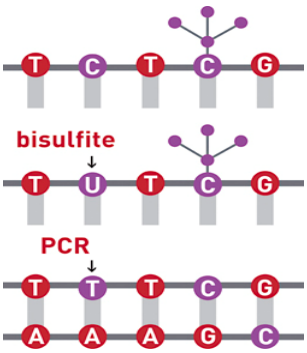


Figure 17. When DNA is treated with sodium bisulfite, 5'-C are converted into U and consequently into T after PCR, while 5'-meC remains unconverted. Adapted from http://www.diagenode.com/en/applications/bisulfite-conversion.php

The EZ DNA Methylation-Gold ™ Kit (Zymo Research) was used. The protocol was the following:

1. Firstly, the CT Conversion Reagent was prepared by mixing for 10 min 900 µl of water, 300µl of M-Dilution Buffer and 50µl of M-Dissolving Buffer to a tube of CT Conversion Reagent.
2. Ethanol was added into the M-Wash Buffer.
3. 130µl of the CT Conversion Reagent were added to 20 µl (1 µg) of the DNA sample in a PCR tube.
4. The sample tube was then placed in a thermal cycler, where the following steps were followed:
    - 98°C for 10 min
    - 64°C for 2.5 hours
    - 4°C storage up to 20 hours.
5. The Zymo-Spin ™ IC Columns were placed into the provided Collection Tubes. The DNA with the 130µl CT Conversion Reagent was mixed with 600µl M-Binding Buffer and the final mixture was loaded into the columns.
6. Centrifugation at 14000 x g for 1 min. The flow through was discarded.
7. 100µl of M-Wash Buffer were added to the column. Centrifugation at 14000 x g for 1 min.
8. 400µl of M-Desulphonation Buffer were added to the column. Incubation at room temperature for 20 min. Centrifugation at 14000 x g for 1 min.
9. 200µl of M-Wash Buffer were added to the column. Centrifugation at 14000 x g for 1 min. Step 9 was performed twice.
10. Subsequently, the column was placed into a 1.5 ml tube. 50µl of M-Elution Buffer were added directly to the column matrix. Centrifugation for 1 min at 14000 x g (elution of the DNA).
11. The DNA samples were stored at -20°C.

2.9.2 PCR for Pyrosequencing

The bisulfite treated DNA locus that we wanted to analyse for methylation was amplified. The PyroMark PCR kit (Qiagen) was used and the conditions are described in **Table 15** and **Table 16**.

Table 15. Primer mix used in PCR for EGFR methylation analysis.

EGFRmeth primer mix	
Component	**Quantity**
EGFRmeth-F (100µM)	7 µl
EGFRmeth-Rb (100µM)	3.5 µl
ddH_2O	89.5 µl

F: Forward primer, Rb: Reverse biotinylated primer

Table 16. PCR conditions for EGFR methylation analysis.

Component	Quantity per sample
PyroMark PCR mix (2x)	12.5 µl
EGFRmeth primer mix	1 µl
CoralLoad Concentrate (10x)	2.5 µl
ddH_2O	7 µl
Bisulfite treated DNA (20ng/µl)	2 µl

Step	Time	Temperature
Hold	15 min	95°C
40 cycles	30 sec	94°C
	30 sec	55°C
	30 sec	72°C
Hold	∞	4°C

PCR reactions were run in a Bench-top thermal cycler GENEAMP SYSTEM 9700 (Applied Biosystems). The PCR products were loaded and run in a 2% agarose gel and visualized so that integrity and density of the bands were assured.

2.9.3 Pyrosequencing for Methylation Analysis

To evaluate the methylation status of EGFR promoter after PCR, a pyrosequencing assay was performed.

1. First, the binding premix was prepared (quantities per sample):
    - Binding Buffer 50 µl
    - Streptavidin beads 3 µl
    - ddH_2O 22 µl
2. 75µl of binding mix were added to 25µl of PCR product and transferred to a 96 well plate (round bottom).

3. The plate was sealed with adhesive film and placed on a vortex for approximately 20 minutes at 350rpm.
4. The annealing mix was prepared and placed in a soft 96 well plate (quantities per sample):
    - Annealing Buffer 43.5µl
    - EGFRmeth-S (10µM) 1.5µl
5. The required volume of substrate, enzyme and nucleotides was added in the cartridge according to the volume information given by the PyroMark Q96 ID Software (**Figure 18.**)
6. The pyrosequencing station was filled with 70% Ethanol, Denaturation Solution (0.2M NaOH), ammonium acetate solution (pH 7.5) and high purity water (**Figure 19.**).
7. The Vacuum Prep Tool was dipped into dH_20 to get wet, then into the 96 well plate where the samples were placed
8. The vacuum then was switched on.
9. The Vacuum Prep Tool was dipped into 70% Ethanol, NaOH solution and ammonium acetate solution respectively.
10. The tool was placed directly onto the 96-well plate containing the annealing mix and after switching the vacuum off, the tool was dipped into the plate.
11. The plate was placed onto a thermoblock for 2 min at 80°C.
12. The plate was allowed to stay at room temperature for 2 min and finally placed into the pyrosequencer to start the run.

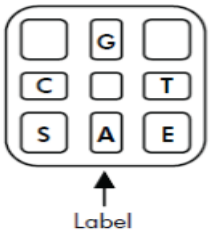


Figure 18. PyroMark Q96 Cartridge. Enzyme mixture (E), substrate mixture (S) and nucleotides (A, C, G, T) are loaded in the illustrated pattern. Adapted from Qiagen.

Figure 19. The PyroMark Vacuum Prep Workstation consists of a hand-held Vacuum Prep Tool and a Vacuum Prep Worktable, connected to a vacuum source, containing four solution troughs. Adapted from Qiagen.

2.10 Tissue Culture

Lung cell lines used: A549, CORL23, CRL5802, IMR90, LUDLU1, LUNG14, NHBE, HBEC_3KT, SKMES, SKLU1, H358, LUDLU1.

Reculturing from frozen stock

1. All cell lines were defrosted from liquid nitrogen, by being immersed in a 37^{o}C water bath for 1 min.
2. Once thawed, cell suspension was transferred in a universal tube containing 10ml cold DMEM (Dulbecco's Modified Eagle Medium) or PBS (Phosphate Buffered Saline).
3. Centrifugation at 200 x g for 5 minutes. The supernatant was discarded.
4. The pellet was resuspended in 10ml DMEM+5% FBS (Fetal Bovine Serum) and transferred into a 75cm^2 flask.
5. Flask was placed in the incubator (37^{o}C and 5%CO_2).

Cell maintenance

1. Cells were washed with 10ml PBS.
2. 3ml trypsin were added in the flask and left in the incubator for approximately 3 minutes.
3. Once the cells were detached, 5ml of DMEM+5% FBS were added in the flask to inactivate trypsin.
4. The flask content was transferred into a universal tube and centrifuged for 5 minutes at 200 x g.
5. The supernatant was discarded.
6. The pellet was resuspended either in 2ml DMEM+5% FBS and half the quantity was then seeded in a new flask containing 9 ml DMEM+5% FBS, or in 1ml cell freezing medium and placed in a cryogenic vial as stock in the liquid nitrogen.

2.11 Cloning

The Tet-Express Systems are inducible gene expression systems for mammalian cells. Target cells that contain the gene of interest under the control of a TRE3G promoter (PTRE3G) will express high levels of the gene, when cultured in the presence of Tet-Express, a tetracycline transactivator with protein transduction properties. We used the pTRE3G-BI-mCherry vector (Clontech Laboratories, Inc.) to clone LOC102723622 (**Figure 20.**).

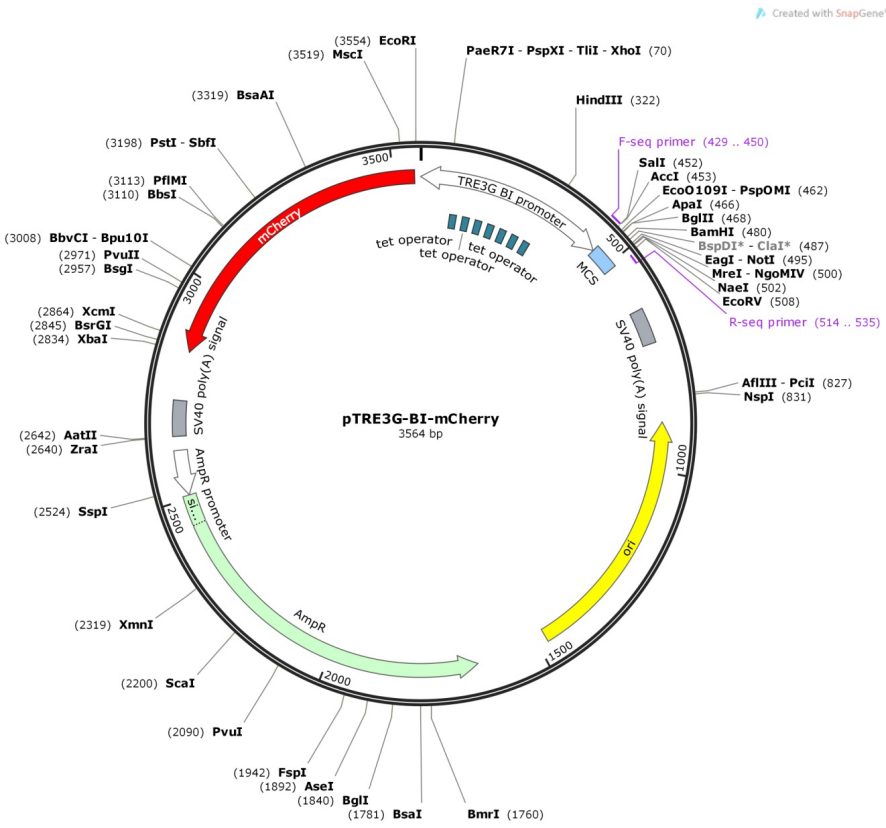


Figure 20. The pTRE3G-BI-mCherry vector map. The MCS (Multiple Cloning Site) where recombination occurred as well as the sequencing primers are depicted.

Procedure:

1. The LOC102723622 intronless sequence was bought by Life Technologies. The lyophilised DNA fragment was dissolved in ddH_2O to obtain a final concentration of 20ng/µl.
2. Vector linearization:
    - The following mix was prepared:
        - pTRE3G-BI-mCherry vector (40ng/µl) 25 µl
        - Buffer 3.1 (New England Biolabs) 3 µl
        - BamH I (New England Biolabs) 2 µl

- The mix was incubated at 37°C for 2 hours and then at 65°C for 5 minutes.

3. Clean-up of the linearized vector using the Nucleospin Gel and PCR clean-up kit (Macherey-Nagel):
    - 60 µl of NT1 buffer were added to the tube of linearized vector (ratio 2:1).
    - The mixture was then transferred into a column and the column was placed on the vacuum system.
    - The column was then washed with 700 µl of NT3 buffer.
    - Dry spin for 3 minutes at 14000 x g.
    - 25 µl of pre-warmed NE buffer was added to the centre of the column and let astand for 5 minutes.
    - Spin for 4 minutes at 14000 x g.
    - Eluate (linearised vector) was nanodroped and 4µl were run on a 1% agarose gel.
4. Preparation of the media:
    - For the Liquid cultures (600ml):
        - Tryptone 6 gr
        - NaCl 3 gr
        - Yeast extract 3 gr
        - dH2O 600 ml
    - For the Solid cultures (300ml):
        - Tryptone 3 gr
        - NaCl 1.5 gr
        - Yeast extract 1.5 gr
        - Agar 4.5 gr
        - dH2O 300 ml
    - The media were autoclaved at 121°C for 20 minutes.
    - When not very hot, 100µg/ml ampicillin (stock 100mg/ml) were added to the media.
    - Next, 2 sterile 145/20mm dishes (greiner bio-one) were carefully filled with 15ml of the medium for the solid cultures in a class II cabinet. When set, the dishes were placed in the fridge upside down until used.
5. Insertion of the LOC102723622 fragment into the pTRE3G-BI-mCherry vector through recombination, using the In-Fusion HD Cloning Kit (Clontech Laboratories, Inc.):
    - The In-Fusion Recombination mix was prepared:
        - Linearized vector (3.5kb, 33.7ng/µl) 0.5 µl
        - LOC102723622 fragment (1.5kb, 20ng/µl) 0.7 µl

- 5x In-Fusion HD Enzyme Premix 1 µl
- ddH_2O 2.8 µl

- The mix was incubated at 50°C for 15 minutes.

6. Bacteria Transformation:
   - 1µl of the In-Fusion Recombination mix (from step 5) was added to 40µl of Stellar Competent Cells (In-Fusion HD Plus Kit (Clontech Laboratories, Inc.)).
   - The cells were then incubated on ice for 30 minutes.
   - Afterwards, the cells were heatshocked at 42°C for 45 seconds and immediately on ice for 2 minutes.
   - 500µl of pre-warmed (at 37°C) SOC medium (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose) were added to the cells.
   - Incubation at 37°C, 60 rpm for 1 hour.
7. Seeding the transformed bacteria on the dishes (work performed in a class II cabinet):
   - 2 serial 100-fold dilutions of the transformed cells were made.
   - 250µl of the dilutions were spread on each dish with agar (from step 4).
   - The dishes were placed in the incubator (37°C) upside down overnight.
8. Colonies selection and production of liquid cultures (work performed in a class II cabinet):
   - The colonies that were easy to get, were marked on the dish (**Figure 21.**)
   - 6ml of medium for liquid cultures was added to 10 vacuette collection tubes (greiner bio-one).
   - Each chosen colony was picked up with a microbiology loop and was then rinsed into one of the vacuette tubes containing the medium.
   - The liquid cultures were left overnight at 37°C shaking horizontally.

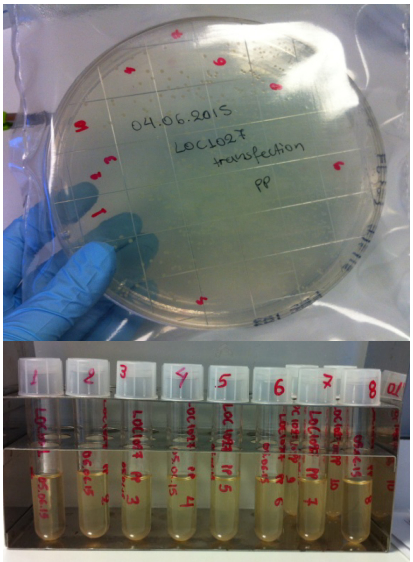


Figure 21. Solid culture (up) and liquid cultures (down). The marked colonies on the dish were used to produce the liquid cultures.

9. Testing the insert:
   - 1ml of each of the 10 liquid cultures was transferred into a 1.5ml tube and spun down at 14000 x g for 5 minutes.

- The supernatant was discarded and the pellet was resuspended in 100μl mild lysis buffer (1xTE, 0.001x TritonX-100).
- Incubation at 95^{o}C for 5 minutes.
- Centrifugation at 14000 x g for 5 minutes.
- 50μl of the supernatant were transferred into a new 0.6ml tube.
- qPCR was performed for LOC102723622 as shown in **Table 17**.

Table 17. Real Time PCR conditions for LOC102723622 insert verification.

Component	Quantity per sample
QuantiTect master mix (2x)	12.5 μl
20x LOC102723622 probe mix[+]	1.25 μl
ddH_2O	10.25 μl
supernatant	1 μl

[+] 20x LOC102723622 probe mix was described in the Real Time PCR section.

Step	Time	Temperature
Hold	15 min	95°C
45 cycles	15 sec	95°C
	30 sec	52°C
	1 min	60°C*

*Measurements were taken at this stage.

10. The positive colony was grown further and 800μl of the culture were mixed with 200μl sterilized glycerol to prepare glycerol stocks of LOC102723622 transformed bacteria. The 2 stocks that were produced were stored at -80^{o}C.
11. The rest of the liquid culture (5ml) was kept to isolate the plasmid using the Zyppy Plasmid Miniprep kit (Zymo Research):
    - First the OD of the culture was measured in a common spectrophotometer and found to be 0.3.
    - The culture was placed into a 15ml falcon tube and spun down at 3000 rpm for 10 minutes (Heraeus Megafuge 2.0).
    - The supernatant was discarded and the pellet was resuspended in 600μl ddH_2O.
    - 100μl of 7x lysis buffer were added to the tube and mixed gently for exactly 2 minutes.
    - 350μl of cold neutralization buffer were then added to the tube and mixed gently.
    - The tube was centrifuged at 3000 rpm for 5 minutes.
    - The supernatant was transferred in the column.
    - The columns were placed on the vacuum station.
    - 200μl of Endo-Wash Buffer were added on the column.
    - The column was then washed with 400μl Zyppy Wash Buffer twice.
    - The column was dry spun for 3 minutes at 14000 x g.
    - 30μl of pre-warmed Elution Buffer were added on the column and let stand at room temperature for 15 minutes.

- The column was then centrifuged at 14000 x g for 4 minutes.
- The eluate was nanodroped and stored in the fridge.

12. The insert was then cut with HindIII:
    - The following reaction was prepared in a PCR tube:
        - Buffer 2 (New England Biolabs) 5µl
        - HindIII (New England Biolabs) 3µl
        - ddH_2O 2µl
    - The tube was incubated at 37°C for 3 hours and then at 65°C for 5 minutes. It was stored in the fridge until used.
13. Clean-up of the cut insert using the QIAquick PCR Purification Kit (QIAGEN) :
    - 250µl of PB Buffer were added to the sample.
    - The mixture was transferred to a QIAquick column and the column was transferred on the vacuum station.
    - The column was washed with 750µl PE Buffer twice.
    - Dry spin at 14000 x g for 3 minutes.
    - 30µl of pre-warmed Elution Buffer were added on the column and let stand at room temperature for 5 minutes.
    - Spin at 14000 x g for 3 minutes.
    - The eluate was nanodroped and left in the vacuum overnight to evaporate.
14. Sequencing reaction. We used the BigDye Terminator v3.1 kit (Life Technologies):
    - We resuspended the dried insert (from step 13) into 14µl ddH_2O (70ng/µl) and let stand at room temperature until dissolved.
    - Two mixes (one for each primer) were prepared and the conditions shown in **Table 18** were used.

Table 18. Sequencing Reaction conditions for insert verification.

Component	Quantity per sample
BiqDye	2 µl
Sequencing Buffer	1 µl
TREGseq-F /TREGseq-R primer (3.2µM)	0.5 µl
Insert	6.5 µl

Step	Time	Temperature
Hold	1 min	96°C
27 cycles	20 sec	95°C
	10 sec	55°C
	4 min	60°C
Hold	4 min	60°C
Hold	∞	4°C

15. Sequencing reaction clean-up, using CENTRI-SEP spin columns (Princeton Separations, Inc.):
    - The desired number of strips was separated by cutting the foil between the strips.

- The well outlets were opened on each strip by cutting through the integral bottom seal constriction.
- The top foil was removed and each strip was inserted in a suitable reservoir for the collection of the interstitial liquid.
- The strips were spun for 2 min at 750 x g in order for the liquid to be removed.
- 20 µl of sample were added to the centre of each well of the Centri-Sep 8 strip.
- The loaded Centri-Sep 8 strip was placed into an 8-well PCR strip and spun for 2 min at 750 x g.
- The samples were then placed in a vacuum chamber until dried out.

16. Sequencing using the 3130 Genetic Analyzer (Applied Biosystems):
    - Firstly, the buffer for the electrophoresis was prepared and placed in the appropriate containers in the 3130 Genetic Analyzer:
        - 3730 Buffer (10x) 4 ml
        - ddH2O 36 ml
    - The samples were dissolved in 10µl of formamide, placed at 95^{o}C for 5 minutes and then immediately on ice for 2 minutes.
    - The samples were then transferred in the sequencing plate, and the plate was placed in the Analyzer.
    - The appropriate program was run (POP-7 polymer in a 36cm array).

2.12 Statistical analysis

For the statistical analysis of the results, IBM SPSS Statistics V22.0 package was used. Non-parametric tests were performed as our data did not follow the Gaussian distribution. To assess the difference in expression between normal and tumour samples, Wilcoxon Signed Ranks test for paired analysis was performed. To examine association between the expression and clinicopathological characteristics, the Mann-Whitney test or its equivalent for multiple comparisons Kruskal-Wallis test were performed. The Spearman's correlation coefficient was used to evaluate correlations among the expression of various genes. Statistical significance was considered at the level of 0.05. For the microarray data analysis, R statistical package was used.

3 Results

3.1 hsa-miR-7150 in NSCLC

3.1.1 Optimization

Although off the shelf assays come with the certainty of optimal efficiency and quality controls performed, we decided to test our assays. Firstly, we performed five 5-fold serial dilutions of three lung cell lines and performed RT qPCR for hsa-miR-7150 and RNU48 (**Figure 22.**).

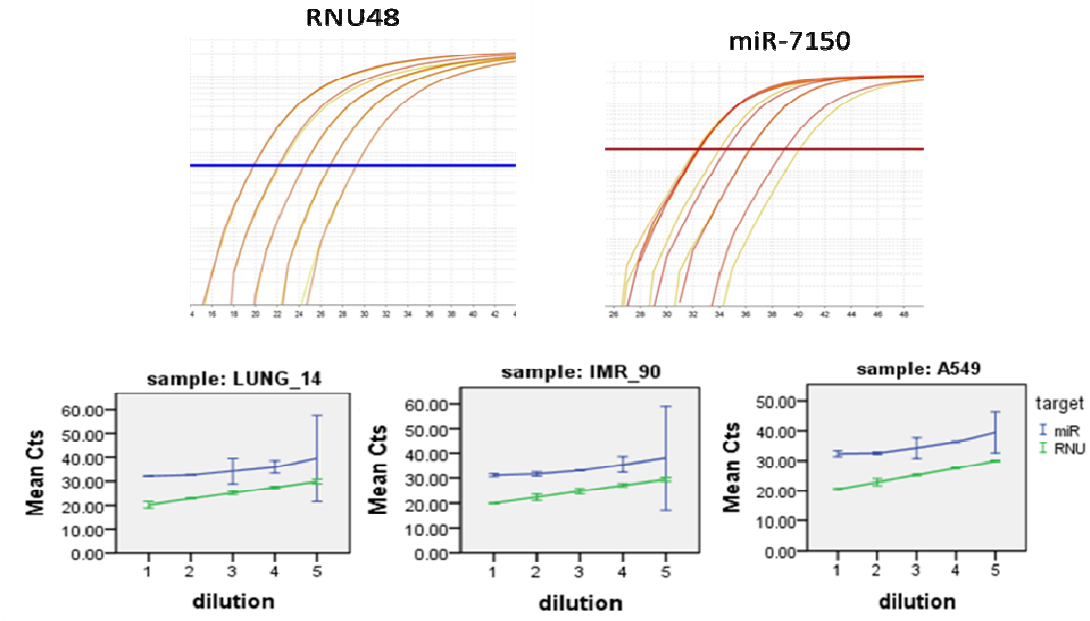


Figure 22. RT qPCR for hsa-miR7150 and RNU48 in 5-fold serial dilutions of three lung cell lines (LUNG_14, IMR_90 and A549).

Both assays, were working pretty consistently when the concentration of the sample was changing. RNU48 Ct was increasing exactly 2.3 cycles in each next dilution. hsa-miR-7150 assay did not perform as good as RNU48, but still the results were reproducible. To further test the assays, we performed the real time qPCR reaction in a gradient of annealing/extension temperatures (60-64°C), although the proposed temperature is 60°C (**Figure 23.**). The optimal annealing/extension temperature is the highest temperature in which, the reaction works. Here we selected 62°C as the optimal annealing/extension

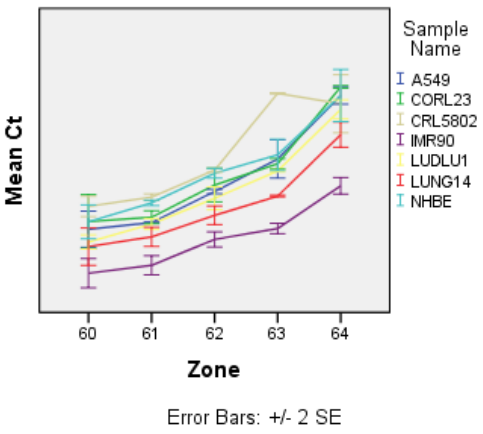


Figure 23. RT qPCR for hsa-miR7150 and RNU48 in 7 lung cell lines (A549, CORL23, CRL5802, IMR90, LUDLU1, LUNG14, NHBE) in a gradient of annealing/extension temperatures.

temperature for the ongoing experiments. In higher temperatures the reaction decays, while lower temperatures carry the risk of non-specific product amplification.

3.1.2 Expression of hsa-miR-7150 in NSCLC

Overall the expression of hsa-miR-7150 was measured in 160 paired tumour and normal samples from NSCLC cancer patients. No significant difference between the expression of hsa-miR-7150 between tumour and normal samples was detected (**Figure 24.**). hsa-miR-7150 expression was associated with histology and tumour stage but not with any other clinicopathological characteristics like nodal status or gender (**Table 19., Figure 25.** and **Figure 26.**).

Table 19. Main clinicopathological characteristics of patients' samples used for hsa-miR-7150 expression analysis.

		Number of patients (80 total)	**Mean expression**	**95% CI**		**p value**
				Lower bound	**Upper bound**	
Gender	males	43	38.317	10.855	65.780	0.404
	females	37	16.191	9.705	22.677	
Histology	AdenoCa	33	15.079	5.484	24.673	0.006
	SqCCL	47	37.215	12.540	61.889	
T stage	T1	14	6.542	2.753	10.332	0.038
	T2	58	33.105	12.793	53.416	
	T3	8	29.378	-0.823	59.580	
	T4	0	N/A	N/A	N/A	
N stage	N0	46	31.679	6.495	56.863	0.603
	N1	23	18.081	9.448	26.714	
	N2	11	33.962	2.369	65.556	

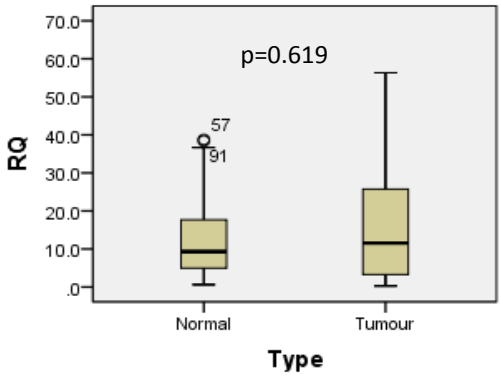


Figure 24. Box plot comparing the expression levels of hsa-miR-7150 between tumour and adjacent normal samples.

Ranks

		N	Mean Rank	Sum of Ranks
RQ_T - RQ_N	Negative Ranks	45[a]	35.42	1594.00
	Positive Ranks	37[b]	48.89	1809.00
	Ties	0[c]		
	Total	82		

a. RQ_T < RQ_N
b. RQ_T > RQ_N
c. RQ_T = RQ_N

Test Statistics[a]

	RQ_T - RQ_N
Z	-.497[b]
Asymp. Sig. (2-tailed)	.619

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.

Figure 25. Wilcoxon Signed Ranks test for paired analysis of has-miR-7150 expression between tumour and adjacent tumour samples. No difference was detected (p=0.619).

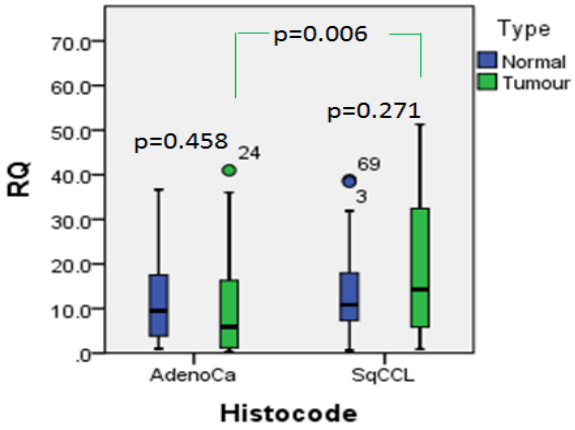


Figure 26. Box plot showing that although expression of hsa-miR-7150 is significantly elevated in SqCCL compared to adenocarcinoma samples, differences between tumours and normals within each histotype were not observed.

3.1.3 hsa-miR-7150 targets – *in silico* analysis

We searched for possible targets of hsa-miR-7150 using the online tool ComiR (Combinatorial miRNA target prediction tool) (**Figure 27.**). ComiR runs a number of algorithms to predict whether a given mRNA is targeted by a set of miRNAs or a given miRNA targets a set of mRNAs.[72]

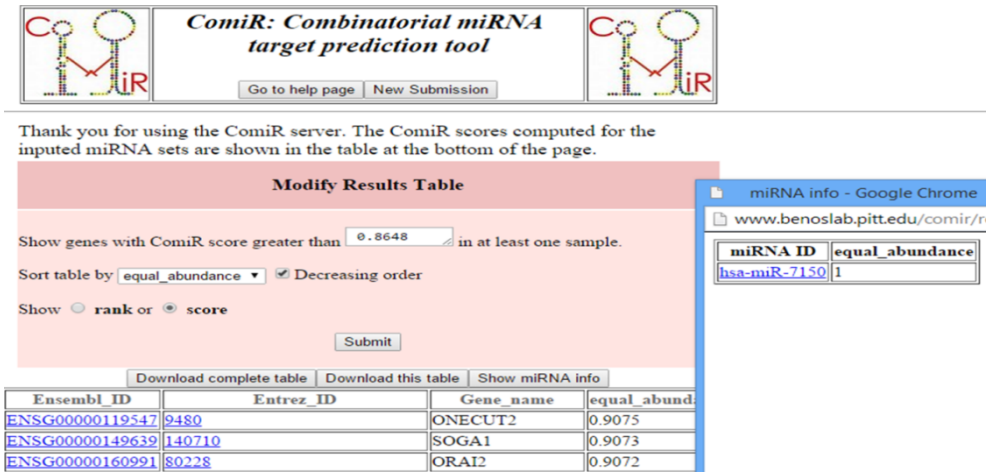


Figure 27. Online search for potential targets of hsa-miR-7150 using ComiR.
(http://www.benoslab.pitt.edu/comir/index2.php)

We selected the hits with ComiR score greater than 0.8648 (default) and tried to cluster this large list of genes for better management and functional interpretation. Therefore, we used DAVID (the Database for Annotation, Visualization and Integrated Discovery) [73] version 6.7 to perform gene ontology analysis according to the biological process our predicted targets were involved in. The cluster with the greatest p value included genes that played a role in cell morphogenesis, while the cluster with the highest amount of genes encompassed regulators of transcription (**Figure 28.**).

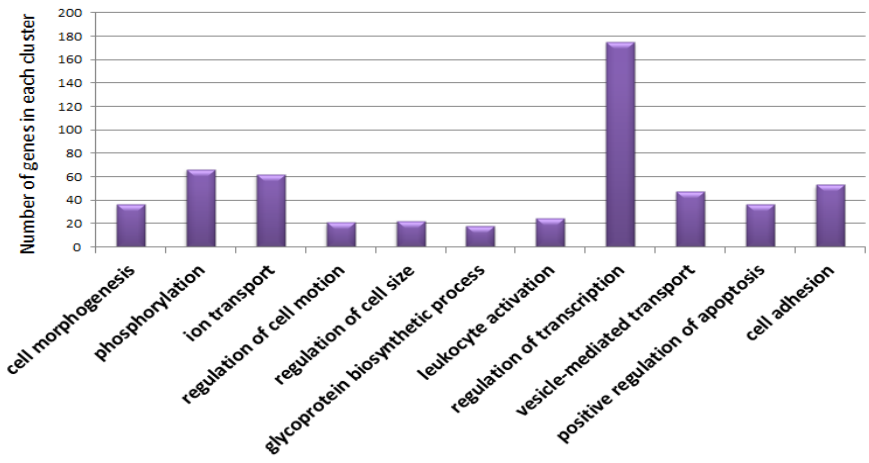


Figure 28. Bar chart demonstrating the distribution of the most important groups describing the function of the genes that are predicted targets of hsa-miR-7150. The x axis from left to right represents decreasing p values of the clusters formed (all p values greater than 0.05). Gene ontology was performed based on biological function using DAVID (http://david.abcc.ncifcrf.gov/).

3.2 Involvement of EGFR-AS1 and LOC102723622 in EGFR expression regulation in NSCLC

3.2.1 EGFR expression and methylation analysis

EGFR expression was measured in 40 paired tumour and normal samples from NSCLC patients (20 pairs in total). Data did not follow a normal distribution, therefore, non-parametric tests were applied. No difference in the expression of EGFR was detected between tumour and adjacent normal samples (**Figure 29.**). Gender was the only clinicopathological characteristic that seemed to be associated with EGFR expression (**Table 20.** and **Table 21.**).

Table 20. Main clinicopathological characteristics of patients' samples used for EGFR expression analysis.

		Number of patients (20 total)	Mean expression	95% CI		p value
				Lower bound	Upper bound	
Gender	males	11	0.448	0.040	0.856	0.025
	females	9	0.107	0.047	0.167	
Histology	AdenoCa	7	0.211	0.091	0.332	0.438
	SqCCL	13	0.339	-0.015	0.694	
T stage	T1	2	0.225	-1.903	2.354	0.239
	T2	16	0.311	0.314	0.591	
	T3	1	N/A	N/A	N/A	
	T4	1	N/A	N/A	N/A	
N stage	N0	9	0.261	0.850	0.436	0.189
	N1	8	0.338	-0.279	0.955	
	N2	3	0.281	-0.146	0.708	
Differentiation degree	Good	6	0.198	0.050	0.347	0.429
	Good/ Moderate	1	N/A	N/A	N/A	
	Moderate	11	0.338	-0.773	0.753	
	Poor	2	0.101	-0.438	0.640	

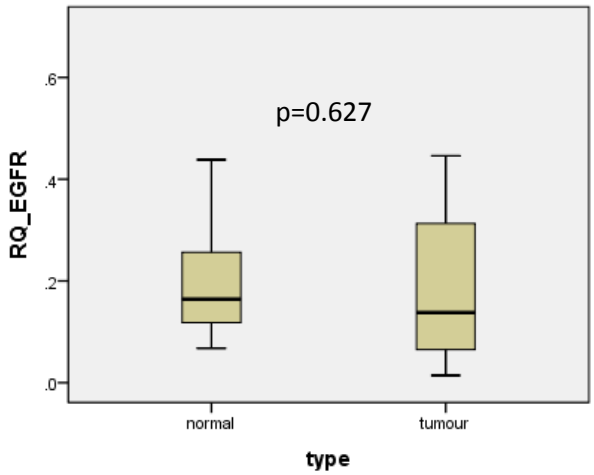


Figure 29. Box plot comparing the expression of EGFR in tumour and normal samples. Wilcoxon Signed Ranks test for paired analysis of EGFR expression between tumour and adjacent tumour samples was performed. No difference was detected (p=0.627).

Table 21. EGFR expression was higher in males than females. Mann Whitney test was performed (p=0.25).

Ranks

Gender		N	Mean Rank	Sum of Ranks
RQ_T_EGFR	male	11	13.18	145.00
	female	9	7.22	65.00
	Total	20		

	RQ_T_EGFR
Mann-Whitney U	20.000
Wilcoxon W	65.000
Z	-2.241
Asymp. Sig. (2-tailed)	.025
Exact Sig. [2*(1-tailed Sig.)]	.025[b]

Next, we wanted to assess whether EGFR was epigenetically regulated through methylation and whether methylation patterns differ between cancer and normal states. 20 pairs of tumour and normal lung samples were bisulfite treated. A part of EGFR promoter was amplified (**Figure 30.**) and pyrosequencing analysis was performed (**Figure 31.**). No difference in the methylation levels among normal and tumour samples was detected. EGFR promoter was unmethylated in both tumour and normal samples (mean methylation levels were below 6% for both types of tissue) (**Figure 32.**).

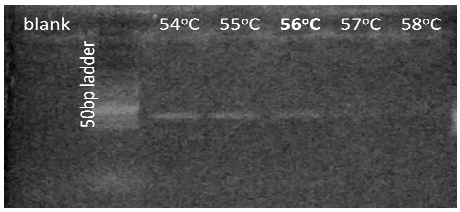


Figure 30. A gradient of annealing temperatures was used during EGFR PCR optimization for methylation analysis. 55°C was selected as the optimal annealing temperature for the consequent PCR reactions.

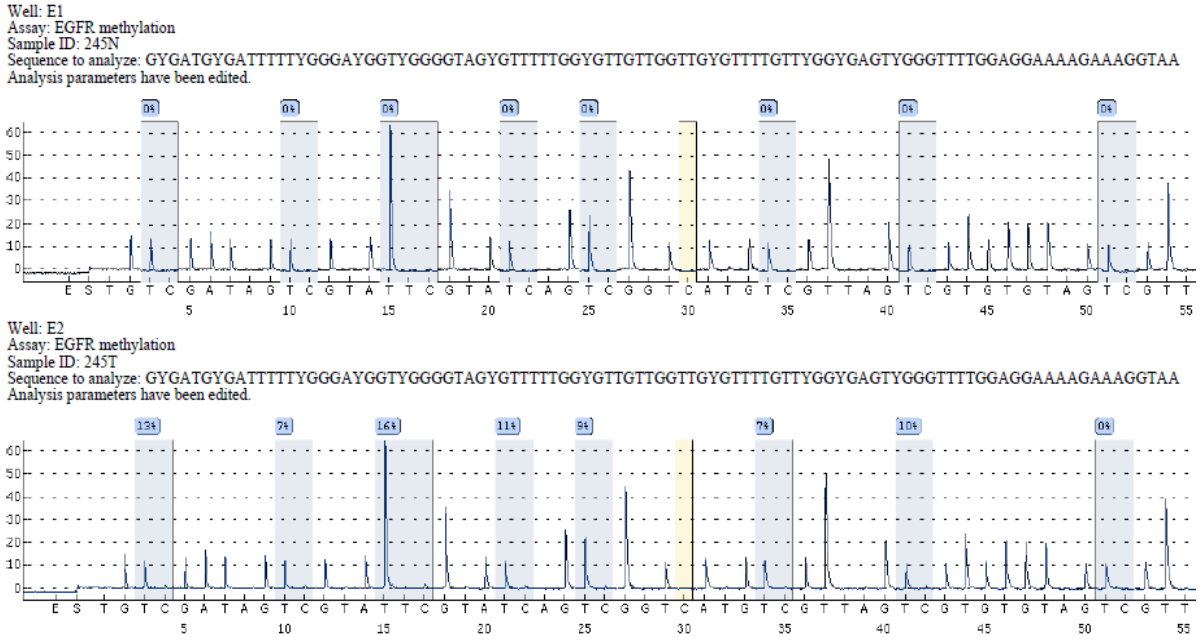


Figure 31. Pyrosequencing analysis for EGFR methylation. An example of pyrograms for a pair of samples.

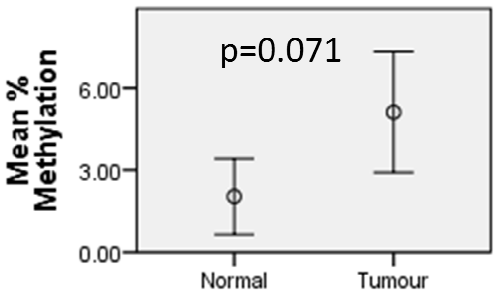


Figure 32. Difference in the mean methylation level observed in tumour and normal samples for EGFR. No statistically significant differences were observed when Mann Whitney test were performed (p-value=0.071).

3.2.2 EGFR-AS1 expression analysis

To assess EGFR-AS1 expression we designed the appropriate primers and probe for real-time PCR. To optimize EGFR-AS1 reaction conditions, we tested the assay in a gradient of annealing temperatures and selected the optimal annealing temperature as the highest temperature in which, the reaction worked. We also selected TBP as the most appropriate endogenous control, as TBP is a low expressed gene (as EGFR-AS1) and also multiplexing EGFR-AS1 and TBP would not give unwanted effects.

EGFR-AS1 expression was measured in 88 paired tumour and normal samples from NSCLC patients (44 pairs in total). Data were not normally distributed, therefore, non-parametric tests were applied. No difference in the expression of EGFR-AS1 was detected among tumour and adjacent normal samples (**Figure 33.**). No association with any clinicopathological characteristics like histology, differentiation, nodal status or gender was observed (**Table 22.**).

Table 22. Main clinicopathological characteristics of patients' samples used for EGFR-AS1 expression analysis.

		Number of patients (44 total)	**Mean expression**	**95% CI**		**p value**
				Lower bound	**Upper bound**	
Gender	males	23	6.762	3.699	9.826	0.124
	females	21	9.421	5.471	13.372	
Histology	AdenoCa	18	4.812	0.939	8.685	0.504
	SqCCL	26	4.719	1.437	8.000	
T stage	T1	3	10.285	-25.877	46.448	0.431
	T2	39	8.104	5.645	10.564	
	T3	1	N/A	N/A	N/A	
	T4	1	N/A	N/A	N/A	
N stage	N0	24	8.634	5.101	12.167	0.282
	N1	15	6.378	2.194	10.563	
	N2	5	10.100	1.172	19.028	
Differentiation degree	Good	12	5.159	1.547	8.771	0.531
	Good/ Moderate	5	10.542	-2.007	23.091	
	Moderate	23	9.538	5.743	13.332	
	Poor	4	4.849	2.584	7.113	

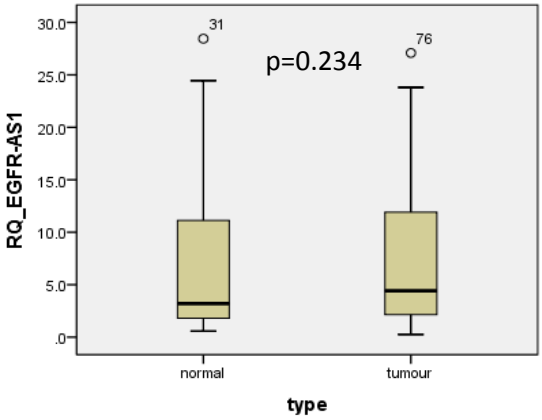


Figure 33. Box plot comparing the expression of EGFR-AS1 in tumour and normal samples. Wilcoxon Signed Ranks test for paired analysis of EGFR expression between tumour and adjacent tumour samples was performed. No difference was detected (p=0.234).

3.2.3 LOC102723622 expression analysis

LOC102723622 expression was measured in 68 paired tumour and normal samples from NSCLC patients (34 pairs in total). Data were not normally distributed, therefore, non-parametric tests were applied. LOC102723622 expression was significantly elevated (median fold change = 3.5, p=0.019) in tumour compared to normal samples (**Figure 34.**). No association with any clinicopathological characteristics like histology, differentiation, nodal status or gender was observed (**Table. 23**).

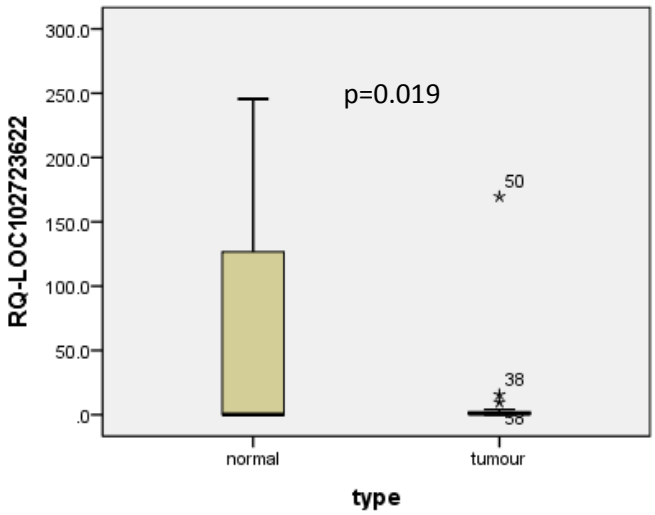


Figure 34. Box plot comparing the expression of LOC102723622 in tumour and normal samples. Wilcoxon Signed Ranks test for paired analysis of LOC102723622 expression between tumour and adjacent tumour samples was performed. A statistically significant difference was observed (p=0.019).

Table 23. Main clinicopathological characteristics of patients' samples used for LOC102723622 expression analysis.

		Number of patients (34 total)	Mean expression	95% CI		p value
				Lower bound	Upper bound	
Gender	males	18	308.013	-318.025	934.052	0.721
	females	16	451.288	-415.380	1317.956	
Histology	AdenoCa	13	502.337	-589.537	1594.211	0.917
	SqCCL	21	296.879	-234.744	828.503	
T stage	T1	3	1.224	-1.229	3.676	0.778
	T2	30	425.320	-138.686	989.327	
	T3	0	N/A	N/A	N/A	
	T4	1	N/A	N/A	N/A	
N stage	N0	19	634.223	-270.803	1539.255	0.222
	N1	11	64.262	-75.940	204.464	
	N2	4	1.916	-0.397	4.229	
Differentiation degree	Good	11	593.420	-726.196	1913.037	0.901
	Good/ Moderate	2	84.820	-992.561	1162.200	
	Moderate	18	336.985	-290.396	964.367	
	Poor	3	0.616	-0.848	2.081	

3.2.4 Correlation of EGFR, EGFR-AS1 and LOC102723622 expression

20 pairs of tumour and normal NSCLC samples were assessed for the expression of EGFR, EGFR-AS1 and LOC102723622 by real time qPCR. Data were not normally distributed, therefore, non-parametric correlation tests (Spearman's correlation) were performed. Data were split according to their type (tumour or normal). The relationship among the expression of the three genes was assessed.

Table 24. Correlations among the expression of EGFR, EGFR-AS1 and LOC102723622 in tumour samples.

Correlations

			RQ_EGFR	RQ_egfr_AS1	RQ_LOC
Spearman's rho	RQ_T_EGFR	Correlation Coefficient	1.000		
		Sig. (2-tailed)	.		
		N	20		
	RQ_T_egfr_AS1	Correlation Coefficient	.457*	1.000	
		Sig. (2-tailed)	.043	.	
		N	20	20	
	RQ_T_LOC102723622	Correlation Coefficient	.525*	.600**	1.000
		Sig. (2-tailed)	.018	.005	.
		N	20	20	20

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

In cancer, the expression of all three genes was significantly correlated ($p<0.05$), but correlation coefficients were very low (Rho< 0.6) (**Table 24.)** Assessment of the normal samples alone showed that there is a weak correlation (Rho = 0.5, p=0.025) between the expression of EGFR and EGFR-AS1 (**Table 25.**).

Table 25. Correlations among the expression of EGFR, EGFR-AS1 and LOC102723622 in normal samples.

Correlations

			RQ_EGFR	RQ_egfr_AS1	RQ_LOC
Spearman's rho	RQ_N_EGFR	Correlation Coefficient	1.000		
		Sig. (2-tailed)	.		
		N	20		
	RQ_N_egfr_AS1	Correlation Coefficient	.499*	1.000	
		Sig. (2-tailed)	.025	.	
		N	20	20	
	RQ_N_LOC102723622	Correlation Coefficient	.004	.211	1.000
		Sig. (2-tailed)	.987	.371	.
		N	20	20	20

*. Correlation is significant at the 0.05 level (2-tailed).

3.2.5 LOC102723622 cloning

In order to evaluate the role of LOC102723622 in carcinogenesis, we proceeded in functional experiments. Our aim was to overexpress LOC102723622 in cell lines that normally express it in very low levels. Therefore, we bought the LOC102723622 sequence (without the introns) and inserted it into a vector. Consequently, we transformed competent bacteria and seeded them on a solid medium covered dish. We randomly selected a number of colonies and tested for the insert using the expression primers for real time PCR (**Figure 35.**).

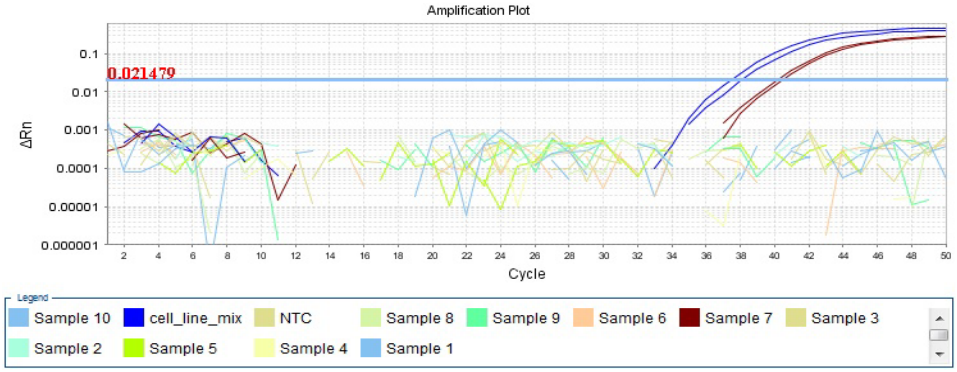


Figure 35. 10 randomly selected colonies were tested for the insert using the expression primers for real time PCR. Only colony 7 (sample 7 –red) and the positive control (cell_line_mix –blue) showed amplification.

One colony was positive for the insert and thus amplified further. Then, we extracted and cleaned the plasmid and performed sequencing in order to validate the insert. Sequencing results were analysed using the Sequencing Analysis Software v5.3.1 (Applied Biosystems) (**Figure 36.**). We verified that the insert in the vector was the desired one (LOC102723622 intronless fragment), although we could only read approximately 200 bases after each primer. Finally, we made glycerol stocks of the transformed bacteria and stored them at -80^{o}C. Further details in the Materials and Methods section. Due to time limitations, we did not proceed to knock in experiments in human cell lines.

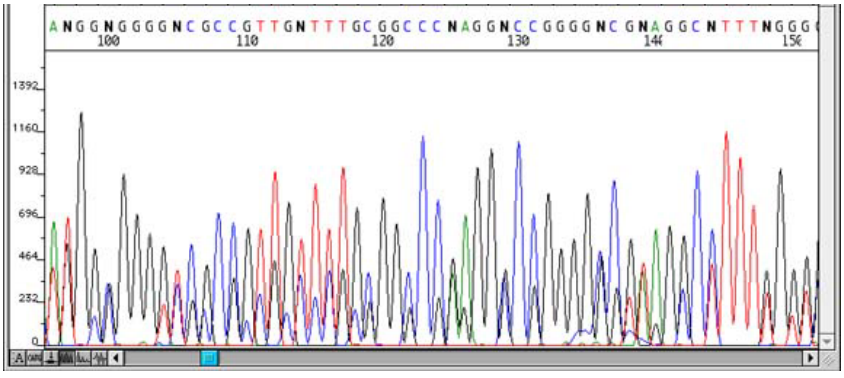


Figure 36. Sequencing of the LOC102723622 insert in the 3130 Genetic Analyzer.

3.3 Long non coding RNA Microarray

The main aim of the microarray experiment was to identify lncRNAs that were differentially expressed in tumour versus normal samples, or in adenocarcinoma versus SqCCL. The selected molecules would be further analysed primarily for their potential as biomarkers or for their involvement in lung cancer pathogenesis.

The main clinicopathological characteristics of the patients' samples used for the lncRNA microarrays are shown in **Table 26**.

Table 26. Main clinicopathological characteristics of patients' samples used for the lncRNA microarray.

		Number of patients (44 total)
Gender	males	31
	females	13
Histology	AdenoCa	21
	SqCCL	23
T stage	T1	2
	T2	32
	T3	8
	T4	2
N stage	N0	22
	N1	12
	N2	11
Differentiation degree	Good	16
	Good/Moderate	2
	Moderate	22
	Poor	4

We searched for lncRNA sets that were:

- highly upregulated in lung tumours (or specific independent subsets of the tumours: e.g., adenocarcinoma specific / SqCCL specific)
- reproducible in external data (especially in non-microarray data)
- independent of confounding factors (eg, stromal / blood contamination / array level biases)

- not induced in COPD / other lung diseases
- technically independent (not due to highly similar probes or contig expression)
- not due to the noisy tail
- both biologically and statistically complementary
- robustly induced at an early tumour stage
- biologically reasonable

Background correction and other adjustments like array position and experimental batches were taken into account. With adjusted p-value of 0.001 and $\log_2$-fold change of 0.5, we found a large number of probes differentially expressed between tumour and normal samples, and/or between adenocarcinomas and SqCCL (**Table 27. and Figure 37.**). This includes multiple probes for the same transcript and multiple transcripts for the same gene.

Table 27. Top induced and top repressed lncRNA probes found in tumour samples compared to normal ones. (Gene Names are not given in full – unpublished data)

Top induced lncRNA probes in tumour samples		
Gene Name	**Fold Change**	**p value**
ENST00000533xxx	2.869562	0.000152794
ENST00000554xxx	4.94323	0.000274203
ENST00000580xxx	3.607866	0.000319646
ENST00000592xxx	1.392158	0.000559057
ENST00000530xxx	1.202435	0.003858094
ENST00000471xxx	1.978691	0.005410147
Top repressed lncRNA probes in tumour samples		
Gene Name	**Fold Change**	**p value**
ENST00000439xxx	-1.664139	1.76E-06
ENST00000499xxx	-3.633362	4.94E-06
ENST00000515xxx	-1.394228	5.49E-06
ENST00000409xxx	-1.776927	5.83E-06
ENST00000566xxx	-5.151783	1.49E-05
ENST00000594xxx	-3.483867	1.68E-05

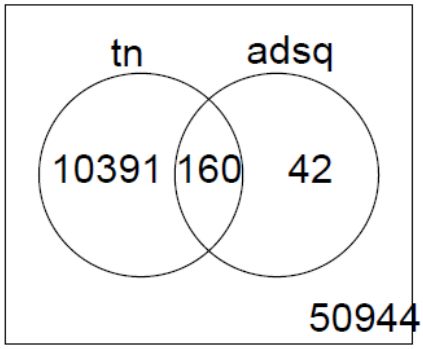


Figure 37. Venn diagram of the number of probe positions that were classified as differentially expressed at p < 0.001, with a fold-change > 2 in either of the contrasts (tn: Tumour vs Normal; adsq: Adeno vs Squamous).

The next step was to do a complete analysis of the ncRNA data, in a gene-centred manner, and generate two separate lists of genes for further follow up. The two lists corresponded to genes that were of biological interest and genes that might have a use as biomarkers. Genes that were downregulated in cancer were not considered in the latter list. Firstly, we identified all genes present on the expression data, with a view to keep only information about ncRNAs. We defined the well-annotated, multiple-probe ncRNA genes to be those genes that were represented on the microarray, for which at least 3 probes are present, a gene ID existed, and none of its transcripts had a 'protein-coding' annotation, but at least one of its transcripts had some annotation (be it a lncRNA, pseudogene, snRNA etc). With this restricted choice of genes, we had 1810 genes and 14558 probes on the array.

The best candidates were those with p-value for tumour/normal comparison < 0.01 and consistent behaviour of most probes (z-score p-value < 0.05 across all probes). Less promising candidates either had t/n p-value < 0.01 or had both a t/n p value <0.05 for at least one probe and a z-score p value < 0.05. At least a 2-fold induction was seen for all markers, for at least one probe on the array.

Overall, we determined some good biomarker candidates that were upregulated in tumours compared to normal samples as well as some good candidate ncRNAs that were downregulated in tumours. Additionally, we were able to identify some ncRNAs that were modestly upregulated in tumours, and which showed diferential expression between adenocarcinoma and SqCCL. Finally, there were some other ncRNAs that appeared to be differentially expressed between adenocarcinomas and SqCCL, but were not upregulated in tumours on average.

4 Discussion

Although raising awareness about lung cancer screening and prevention has been a great move in many countries, lung cancer still remains the number one killer type of cancer worldwide. For years people have participated in successful screening programs for breast cancer, cervical cancer and prostate cancer, but for lung cancer, conventional approaches for detecting the disease such as cytological examination or CT scan have failed to diagnose all cases. On the contrary, X-ray examination on its own could overload a patient. This indicates that specific and sensitive molecular approaches that could detect minimum disease burden could assist early diagnosis of lung cancer.

Many studies have moved towards this direction, trying to identify biomarkers for early detection of lung cancer. The term "biomarker" in this sense is used to describe a characteristic that is objectively measured and evaluated as an indicator of pathogenic processes.[74] Emerging studies have revealed that ncRNA detection in body fluid may serve as a potential "tool" to detect NSCLC in a non-invasive way, making screening more approachable. Although the diagnostic accuracy of ncRNA detection in body fluids is still under question, cutting edge technology, such as digital PCR, promise to detect low abundance cancer associated molecules in patients' blood.[75]

To date, apart from SHOX2 gene methylation assay for lung cancer diagnosis, biomarkers for lung cancer that are used in the clinic, mainly define and/or monitor therapy (e.g., EGFR, KRAS, or ALK mutations).[76] The established evidence regarding the involvement of epigenetic deregulation in lung cancer sets research of epigenetic markers at very high priority in lung cancer diagnostics.[53] ncRNAs and especially miRNAs have been favourably studied as biomarkers.[62,77] miRNA expression profiles are quite reproducible and as molecules, miRNAs are very stable. These characteristics of miRNAs make them ideal candidates. Lately, lncRNAs have gained ground in the field of cancer research due to their diverse roles in cell physiology. They are regulators of key pathways that govern cell growth, apoptosis, and metastasis and are commonly distorted in cancer. lncRNAs appear as promising biomarkers and a novel class of potential drug targets.

In this study we investigated the potential of hsa-miR-7150 as a lung cancer biomarker. We also assessed the expression of EGFR-AS1 and LOC102723622 in association with EGFR regulation in lung cancer samples. Finally, we tried to identify differentially expressed lncRNAs in NSCLC that could serve as biomarkers.

hsa-miR-7150

hsa-miR-7150 was introduced by Oulas et al.[46] while developing a bioinformatic program for novel miRNA identification in the human genome. No further literature regarding its role or its expression exists. hsa-miR-7150 is located in a cancer associated genomic locus, thus we hypothesised that it might be deregulated in lung cancer.

We used a customized TaqMan Assay from Life Technologies, which we tested and optimized for our final experiments. The specificity of off the self assays can be questioned as the sequences of primers and probes are not known. Blasting the miRNA sequence, showed that hsa-miR-7150 shared a high degree of homology with other miRNAs. Therefore, for our real time PCR experiments, we adopted as the optimal annealing/extension temperature the highest temperature in which, the reaction worked without decaying, considering that the higher the temperature the more specific the primer annealing.

We compared hsa-miR-7150 expression levels in 80 pairs of lung tumour and normal tissues, but we did not find any difference. Unfortunately, this result obliterates the potential of this molecule as a biomarker. However, we did find an association between hsa-miR-7150 expression and histology. It is worth mentioning that no difference between tumours and normals in each histological type separately was observed. However, hsa-miR-expression was slightly lower (but not significantly) in adenocarcinoma tumour samples compared to normals, while hsa-miR-7150 expression was slightly higher (but not significantly) in SqCCL tumour samples compared to normal ones (**Figure 26.**). This differences overall might have accounted for the statistically significant difference of hsa-miR-7150 expression in adenocarcinomas compared to SqCCL. Furthermore, a borderline significant increased expression of hsa-miR-7150 was observed in T2 stage tumours compared to earlier (T1) stage tumours. This could be explained by the fact that more advanced tumours suffer from greater deregulation, but still the comparable levels of hsa-miR-7150 expression between tumour and normal samples led us to the conclusion that hsa-miR-7150 is not an ideal biomarker candidate for lung cancer.

Finally, we searched for hsa-miR-7150 possible targets and clustered them according to the biological processes they were involved in. Non cancer related targets, would further support the conclusion that hsa-miR-7150 is not a cancer indicator. Among the most important clusters, there were cell morphogenesis genes, phosphorylation related genes and ion transport related genes. Although these data do not directly point to cancer, no direct conclusions can be drawn.

Involvement of EGFR-AS1 and LOC102723622 in EGFR regulation in NSCLC

EGFR is a tyrosine kinase receptor that belongs to the HER/Erb-B family and is involved in cell proliferation, differentiation, survival, angiogenesis, and migration.[78] Aberrant regulation of EGFR in NSCLC has been reported more than 3 decades ago.[79] To date, the discovery of EGFR mutations in NSCLC has managed to identify the patients who can benefit from molecularly targeted therapies based on tyrosine kinase inhibitors such as gefitinib, erlotinib, afatinib, dacomitinib and icotinib.[80]

The majority of studies have focused on EGFR mutations and/or therapeutic stratification of patients accordingly. However, expression of EGFR in NSCLC has also been reported in studies mainly using immunohistochemical methods for expression analysis. Frequencies for EGFR overexpression were ranging between 32% and 47%., while significantly higher EGFR expression in SqCCL compared to adenocarcinomas and large cell carcinomas has been observed.[81-83] A controversy in the prognostic significance of EGFR expression exists in the literature. Some studies found positive correlations between EGFR overexpression and tumour invasiveness,[84] or poorer survival,[85] whereas others showed no correlation at all.[86]

In this study we assessed the mRNA expression of EGFR in 20 pairs of tumour and adjacent normal tissues. We did not manage to detect a difference between tumour and normal samples or any association between EGFR expression and clinicopathological characteristics, apart from gender, coming in opposition with most of the existing literature. This could be explained in various ways. For example, the size of our sample set was too small for a difference to be detected. The power of the study should be calculated before moving on from these pilot data. Moreover, in the literature, overexpression of EGFR is mainly observed on the protein level. A very limited number of studies have used mRNA as the main source to evaluate EGFR expression. Brabender et al.[87] used TaqMan Assays to assess EGFR mRNA expression, however to analyse their results, they used the ratio between mRNA expression in normal and tumour tissue as a measure of the degree of gene expression, which differed from the ΔΔCt method we applied. Here, it becomes clear that elevated protein expression does not necessarily mean elevated mRNA expression and viceversa. Moreover, different ways of data analysis might also contribute to differences in the final outcome.

We also assessed EGFR promoter methylation levels. A borderline significance (p=0.071) between methylation levels of tumour and normal samples was observed, however, in this case statistical significance did not coincide with biological significance, as in both types of tissues mean percentage methylation was below 6%. To the best of our knowledge, this is the first time that EGFR methylation has been assessed.

Unfortunately, no literature on EGFR-AS1 or LOC102723622 exists. Both lncRNAs are located in the genetic locus of EGFR. Hence we hypothesised that they might be involved in EGFR regulation or they might be regulated in a common way. We assessed the expression of the 2 lncRNAs in NSCLC samples and adjacent normal tissues. Only LOC102723622 showed significant elevated expression in tumours compared to normal lung tissue (median fold change = 3.5, p=0.019). No association among the expression of EGFR-AS1 or LOC102723622 and any clinicopathological characteristics was found.

Then, we wanted to evaluate the relationship among the expression of the three genes (EGFR, EGFR-AS1 and LOC102723622). 20 pairs of tumour and normal lung tissues were commonly used in all 3 genes expression experiments. We performed correlation analysis of the 3 genes expression data in these 20 pairs. Firstly, we split the data according to their type. In tumour samples, the expression of all three genes was significantly correlated ($p<0.05$), but correlation coefficients were very low. In the normal samples, a weak correlation (Rho = 0.5, p=0.025) between the expression of EGFR and EGFR-AS1 was reported. The correlations in the tumour samples could be explained by the general deregulation than governs cancer cells. As unfortunate as it might appear, weak correlations of biologically unrelated genes can arise frequently when studying cancer. According to Hanahan and Weinberg,[88] the macroscopic physiological changes must also correlate with global alterations of the molecular profiles of gene transcription. On the other hand, the weak correlation between EGFR and EGFR-AS1 expression in normal samples might be a sign of common regulation or interaction in normal physiology. As opposed to miRNAs, lncRNAs play a plethora of roles. What we would ideally expect would be a negative correlation, in order to empower the hypothesis that EGFR expression can be silenced by EGFR-AS1. However, even this result, does not exclude the possibility of direct or indirect interaction between the two, especially when ncRNAs are there to tune up, rather than regulate transcription; transcription regulator proteins are the main basic players here.

LOC102723622 was overexpressed in NSCLC. To assess the importance of this result, we proceeded in functional experiments. Our aim was firstly to overexpress LOC102723622 in cell lines that express it in low levels and then to knock it down in cell lines that overexpress it in order to detect the changes that LOC102723622 expression confers. These experiments would lead to the elucidation of LOC102723622 role in cancer and normal physiology. However, due to time limitations we were only able to produce the construct carrying the LOC102723622 sequence and make bacteria stocks for future use.

lncRNA microarray

Gene expression profiles can indicate cellular functions, biochemical pathways and regulatory mechanisms. In case of diseased cells or tissues, gene expression profiles compared to normal controls, may reveal the disease pathology and identify new therapeutic points, assist diagnosis and evaluate prognosis. Microarray technology has enabled the analysis of the whole transcriptome of a cell. Several studies have used the microarray technology in order to elucidate the changes in the expression profiles in different types of malignancies. Expression profiles can be compared in tumour versus normal samples, treated versus untreated patients' samples or among the various stages of cancer.[89]

In our microarray experiment we used the tumour versus control approach, in which the tumour gene expression profile was compared with its corresponding control sample in order to measure the differences between both phenotypes. Our aim was to identify lncRNAs that were differentially expressed in tumour versus normal samples. We considered mainly the well-annotated multiple-probe lncRNA genes and divided them in two categories: a) genes that were of biological interest and b) genes that might have a use as biomarkers. Genes that were downregulated in cancer were not considered as good biomarker candidates, due to their powerlessness to be actually measured in a clinical setting. We also took into account the histology of the samples tested. After applying strict analysis criteria, we finally ended up with a list of 13 interesting targets for further study (either induced in tumour versus normal, or differentially expressed between adenocarcinomas and SqCCL with induction in one of the two). Due to time limitations, we did not manage to further validate these targets. Future work will include real time PCR for the targets in the training set (samples used in the microarray set-up) and then in a separate validation cohort of 150 paired NSCLC samples. Our future goal is to identify a lncRNA expression signature that would be indicative for early stage NSCLC. Ideally, this signature would be detectable in body fluids as well, to ease lung cancer screening (no radiation, minimally invasive technique) of high risk individuals.

Biomarker qualification involves four major steps: 1) discovery phase, 2) experimental validation, 3) preclinical validation, and 4) clinical validation.[74] Numerous studies describe new biomarkers and their potential use in the clinic; however, very few of them manage to reach clinical validation. The main obstacles are the poor reproducibility of the studies, sample heterogeneity, methodological biases, quality control implementation, and significant interlaboratory variability in experimental performance and data analysis. This raises the importance of diligent validation of the results obtained.[90]

Searching the literature, we found two studies that have utilized lncRNA based microarrays for lung cancer research. Xu et al.[91] used microarrays to compare the lncRNA and mRNA expression profiles in lung adenocarcinoma versus normal tissue samples. They initially identified a number of interesting candidate lncRNAs and further validated them with real time PCR. We detected no common targets with Xu's team, but this is most probably due to the absence of SqCCL samples (in Xu's study cohort) or due to different stringency criteria used during the analysis. Yang et al.[92] used microarrays to assess the expression profiles of mRNAs, lncRNA and miRNA in A549 cells versus cisplatin resistant A549 cells. Differentially expressed RNAs were then verified by real time PCR and analysed further. They found that numerous ncRNAs were differently expressed between cisplatin resistant and sensitive cells. They also revealed that AK126698 lncRNA regulated A549 cells cisplatin resistance partly through the canonical Wnt pathway, suggesting a possible therapeutic intervention for cisplatin resistant NSCLC patients.

5 Conclusions

- No difference in the expression of hsa-miR-7150 between lung tumour and adjacent normal samples was detected. hsa-miR-7150 expression was associated with the histological type and stage of the tumour.
- No difference in the expression of EGFR between lung tumour and adjacent normal samples was detected. EGFR expression was associated with gender but not with any other clinicopathological characteristics. EGFR promoter was unmethylated in both tumour and normal samples.
- No difference in the expression of EGFR-AS1 between lung tumour and adjacent normal samples was detected. No association between EGFR-AS1 expression and any clinicopathological characteristics was observed.
- A statistically significant increase in the expression of LOC102723622 in tumour compared to normal samples was detected. No association between LOC102723622 expression and any clinicopathological characteristics was observed.
- EGFR-AS1 expression was weakly correlated with EGFR expression in normal lung samples.
- Microarray analysis designated 13 lncRNAs, either induced in tumour versus normal or differentially expressed between adenocarcinomas and SqCCL, for further analysis.

6 Appendix: Abbreviations used in the text and their definitions.

Abbreviation	Definitions
SCLC	Small Cell Lung Cancer
NSCLC	Non-Small Cell Lung Cancer
SqCCL	Squamous Cell Carcinoma of the Lung
AdenoCa	Adenocarcinoma
COPD	Chronic Obstructive Pulmonary Disease
HIV	Human Immunodeficiency Virus
DNA	DeoxyriboNucleic Acid
RNA	RiboNucleic Acid
ncRNA	Non-coding RNA
CpG	deoxyCytidine-phosphate-deoxyGuanosine
5-mC	5-methylCytosine
5-hmC	5-hydroxymethylCytosine
TET1	Ten-Eleven Translocation 1 protein
DNMT	DNA MethylTransferases
HAT	Histone Acetyl Transferase
HDAC	Histone DeACetylase
TSS	Transcription Start Site
NDR	Nucleosome Depleted Region
gDNA	genomic DNA
miRNA	microRNA
piRNA	piwi-interacting RNA
snoRNA	small nucleolar RNA
snRNA	Small nuclear RNA
lncRNA	long non-coding RNA
rRNA	ribosomal RNA
mRNA	messenger RNA
siRNA	Short interfering RNA
nt	nucleotide(s)
RISC	RNA-induced silencing complex
PRC	Polycomb Repressor Complex
EST	Expressed Sequence Tag
EGFR	Epidermal Growth Factor Receptor
BRG1	Transcription activator BRG1 or ATP-dependent helicase SMARCA4
HOTAIR	HOX Transcript Antisense RNA
MALAT	Metastasis Associated Lung Adenocarcinoma Transcript 1
BCYRN1	Brain CYtoplasmic RNA 1
HNF1A-AS1	HNF1a AntiSense RNA 1
ANRIL	Antisense Non-coding RNA in the INK4 Locus
EMT	Epithelial to Mesenchymal Transition
ml	milli litre
µl	microlitre
cDNA	complementary DNA
min	minute(s)
sec	second(s)

dNTP	deoxy-Nucleotide-Tri-Phosphate
FRET	Fluorescence Resonance Energy Transfer
CoMIR	Combinatorial miRNA target prediction tool
DAVID	the Database for Annotation, Visualization and Integrated Discovery
NCBI	The National Center for Biotechnology Information
rpm	revolutions per minute
x g	gravitational force
CO_2	Carbon diOxide
dH_2O	distilled water
ddH_2O	double distilled water
DMEM	Dulbecco's Modified Eagle Medium
PBS	Phosphate Buffered Saline
Ct	Threshold Cycle
ΔCt	delta Ct
ΔΔCt	delta delta Ct
norm	normaliser or else calibrator
N/A	Non Applicable
CI	Confidence Interval
CT scan	Computed Tomography scan

7 Acknowledgements

I wish to acknowledge my supervisor, Prof. Aristides Eliopoulos for his continuous guidance and support in accomplishing this thesis. I would also like to thank our collaborator Dr. Triantafillos Liloglou for the significant contribution to this thesis, the constant help, encouragement and advice he has provided throughout my MSc course. I must also express my gratitude to the 3 member committee –Prof. Eliopoulos, Dr. Liloglou and Dr. Mavroudis- for evaluating my thesis.

I am also grateful to Dr. Amelia Acha Sagredo and Dr. Georgios Nikolaidis, who were so willing to cooperate with me and helped me accomplish all the lab work. A big thank you to Dr. Russell Hyde for the microarray analysis.

I would also like to thank the Greek State Scholarship Foundation (I.K.Y.) for the financial support by offering me the IKY Scholarship of Excellence for master studies in Greece – the Siemens program.

Lastly, I would like to thank my family and friends for their full support during my studies.

8 References

1. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer.* Apr 2013;49(6):1374-1403.
2. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA: a cancer journal for clinicians.* Jan-Feb 2014;64(1):9-29.
3. CRUK. http://www.cancerresearchuk.org/health-professional/cancer-statistics.
4. Kanne JP. Screening for lung cancer: what have we learned? *AJR. American journal of roentgenology.* Mar 2014;202(3):530-535.
5. General ARotS. *How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease*: Centers for Disease Control and Prevention (US); National Center for Chronic Disease Prevention and Health Promotion (US); Office on Smoking and Health (US).
6. Doll R, Hill AB. Smoking and carcinoma of the lung; preliminary report. *British medical journal.* Sep 30 1950;2(4682):739-748.
7. Khuder SA. Effect of cigarette smoking on major histological types of lung cancer: a meta-analysis. *Lung Cancer.* Feb-Mar 2001;31(2-3):139-148.
8. Pesch B, Kendzia B, Gustavsson P, et al. Cigarette smoking and lung cancer--relative risk estimates for the major histological types from a pooled analysis of case-control studies. *International journal of cancer. Journal international du cancer.* Sep 1 2012;131(5):1210-1219.
9. Peralta AR, Guntur VP. Safety and efficacy of electronic cigarettes: a review. *Missouri medicine.* May-Jun 2014;111(3):238-244.
10. Robertson A, Allen J, Laney R, Curnow A. The cellular and molecular carcinogenic effects of radon exposure: a review. *International journal of molecular sciences.* 2013;14(7):14024-14063.
11. Darby S, Hill D, Auvinen A, et al. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. *BMJ.* Jan 29 2005;330(7485):223.
12. De Matteis S, Consonni D, Bertazzi PA. Exposure to occupational carcinogens and lung cancer risk. Evolution of epidemiological estimates of attributable fraction. *Acta bio-medica : Atenei Parmensis.* 2008;79 Suppl 1:34-42.
13. Archontogeorgis K, Steiropoulos P, Tzouvelekis A, Nena E, Bouros D. Lung cancer and interstitial lung diseases: a systematic review. *Pulmonary medicine.* 2012;2012:315918.
14. Brenner DR, McLaughlin JR, Hung RJ. Previous lung diseases and lung cancer risk: a systematic review and meta-analysis. *PloS one.* 2011;6(3):e17479.
15. Lee JY, Jeon I, Lee JM, Yoon JM, Park SM. Diabetes mellitus as an independent risk factor for lung cancer: a meta-analysis of observational studies. *Eur J Cancer.* Jul 2013;49(10):2411-2423.
16. Kirk GD, Merlo CA, Lung HIVS. HIV infection in the etiology of lung cancer: confounding, causality, and consequences. *Proceedings of the American Thoracic Society.* Jun 2011;8(3):326-332.
17. Yang Y, Dong J, Sun K, et al. Obesity and incidence of lung cancer: a meta-analysis. *International journal of cancer. Journal international du cancer.* Mar 1 2013;132(5):1162-1169.
18. Shi HB, Tang B, Liu YW, Wang XF, Chen GJ. Alzheimer disease and cancer risk: a meta-analysis. *Journal of cancer research and clinical oncology.* Mar 2015;141(3):485-494.
19. Bajaj A, Driver JA, Schernhammer ES. Parkinson's disease and cancer risk: a systematic review and meta-analysis. *Cancer causes & control : CCC.* May 2010;21(5):697-707.
20. Ilus T, Kaukinen K, Virta LJ, Pukkala E, Collin P. Incidence of malignancies in diagnosed celiac patients: a population-based estimate. *The American journal of gastroenterology.* Sep 2014;109(9):1471-1477.
21. WHO. http://www.who.int/genomics/public/geneticdiseases/en/index3.html.
22. Nitadori J, Inoue M, Iwasaki M, et al. Association between lung cancer incidence and family history of lung cancer: data from a large-scale population-based cohort study, the JPHC study. *Chest.* Oct 2006;130(4):968-975.
23. WHO. *Lung Cancer Histologies.* http://www.lung.org/lung-disease/lung-cancer/learning-more-about-lung-cancer/understanding-lung-cancer/knowing-the-basics.html.
24. Beasley MB, Brambilla E, Travis WD. The 2004 World Health Organization classification of lung tumors. *Seminars in roentgenology.* Apr 2005;40(2):90-97.
25. Tanoue LT, Detterbeck FC. New TNM classification for non-small-cell lung cancer. *Expert review of anticancer therapy.* Apr 2009;9(4):413-423.
26. AJCC. https://cancerstaging.org/references-tools/quickreferences/Documents/LungMedium.pdf.
27. Waddington CH. The epigenotype. 1942. *International journal of epidemiology.* Feb 2012;41(1):10-13.

28. Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A. An operational definition of epigenetics. *Genes & development.* Apr 1 2009;23(7):781-783.

29. Portela A, Esteller M. Epigenetic modifications and human disease. *Nature biotechnology.* Oct 2010;28(10):1057-1068.

30. Jones PA, Liang G. Rethinking how DNA methylation patterns are maintained. *Nature reviews. Genetics.* Nov 2009;10(11):805-811.

31. Sandoval J, Heyn H, Moran S, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics : official journal of the DNA Methylation Society.* Jun 2011;6(6):692-702.

32. Thomson JP, Skene PJ, Selfridge J, et al. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature.* Apr 15 2010;464(7291):1082-1086.

33. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell.* Oct 29 1999;99(3):247-257.

34. Robert MF, Morin S, Beaulieu N, et al. DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells. *Nature genetics.* Jan 2003;33(1):61-65.

35. Globisch D, Munzel M, Muller M, et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PloS one.* 2010;5(12):e15367.

36. Heyn H, Esteller M. An Adenine Code for DNA: A Second Life for N6-Methyladenine. *Cell.* May 7 2015;161(4):710-713.

37. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell research.* Mar 2011;21(3):381-395.

38. Verdone L, Caserta M, Di Mauro E. Role of histone acetylation in the control of gene expression. *Biochemistry and cell biology = Biochimie et biologie cellulaire.* Jun 2005;83(3):344-353.

39. Fraga MF, Ballestar E, Villar-Garea A, et al. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nature genetics.* Apr 2005;37(4):391-400.

40. Gabrowski P. *RNA processing*. Croatia2011.

41. Khalil AM, Guttman M, Huarte M, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America.* Jul 14 2009;106(28):11667-11672.

42. Esteller M. Non-coding RNAs in human disease. *Nature reviews. Genetics.* Dec 2011;12(12):861-874.

43. Hammond SM. An overview of microRNAs. *Advanced drug delivery reviews.* May 12 2015.

44. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell.* Dec 3 1993;75(5):843-854.

45. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. *Carcinogenesis.* Jan 2010;31(1):27-36.

46. Oulas A, Boutla A, Gkirtzou K, Reczko M, Kalantidis K, Poirazi P. Prediction of novel microRNA genes in cancer-associated genomic regions--a combined computational and experimental approach. *Nucleic acids research.* Jun 2009;37(10):3276-3287.

47. Angrand PO, Vennin C, Le Bourhis X, Adriaenssens E. The role of long non-coding RNAs in genome formatting and expression. *Frontiers in genetics.* 2015;6:165.

48. Guttman M, Donaghey J, Carey BW, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature.* Sep 15 2011;477(7364):295-300.

49. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annual review of biochemistry.* 2012;81:145-166.

50. Cheetham SW, Gruhl F, Mattick JS, Dinger ME. Long noncoding RNAs and the genetics of cancer. *British journal of cancer.* Jun 25 2013;108(12):2419-2425.

51. McManus. http://mcmanuslab.ucsf.edu/node/251.

52. Pantazi P, Acha-Sagredo A, Papaioannou A, Nikolaidis G, Liloglou T. DNA methylation: its role in transcriptional regulation and association with lung cancer. *Research and Reports in Biochemistry.* 2015;2015:5:11–30.

53. Liloglou T, Bediaga NG, Brown BR, Field JK, Davies MP. Epigenetic biomarkers in lung cancer. *Cancer letters.* Jan 28 2014;342(2):200-212.

54. Schmidt B, Liebenberg V, Dietrich D, et al. SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer based on bronchial aspirates. *BMC cancer.* 2010;10:600.

55. Van Den Broeck A, Brambilla E, Moro-Sibilot D, et al. Loss of histone H4K20 trimethylation occurs in preneoplasia and influences prognosis of non-small cell lung cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research.* Nov 15 2008;14(22):7237-7245.

56. Barlesi F, Giaccone G, Gallegos-Ruiz MI, et al. Global histone modifications predict prognosis of resected non small-cell lung cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* Oct 1 2007;25(28):4358-4364.

57. Seligson DB, Horvath S, McBrian MA, et al. Global levels of histone modifications predict prognosis in different cancers. *The American journal of pathology.* May 2009;174(5):1619-1628.

58. Glaros S, Cirrincione GM, Palanca A, Metzger D, Reisman D. Targeted knockout of BRG1 potentiates lung cancer development. *Cancer research.* May 15 2008;68(10):3689-3696.

59. Medina PP, Carretero J, Fraga MF, Esteller M, Sidransky D, Sanchez-Cespedes M. Genetic and epigenetic screening for gene alterations of the chromatin-remodeling factor, SMARCA4/BRG1, in lung tumors. *Genes, chromosomes & cancer.* Oct 2004;41(2):170-177.

60. Reisman DN, Sciarrotta J, Wang W, Funkhouser WK, Weissman BE. Loss of BRG1/BRM in human lung cancer cell lines and primary lung cancers: correlation with poor prognosis. *Cancer research.* Feb 1 2003;63(3):560-566.

61. Feng B, Zhang K, Wang R, Chen L. Non-small-cell lung cancer and miRNAs: novel biomarkers and promising tools for treatment. *Clin Sci (Lond).* May 1 2015;128(10):619-634.

62. Bediaga NG, Davies MP, Acha-Sagredo A, et al. A microRNA-based prediction algorithm for diagnosis of non-small lung cell carcinoma in minimal biopsy material. *British journal of cancer.* Oct 29 2013;109(9):2404-2411.

63. Chen X, Ba Y, Ma L, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell research.* Oct 2008;18(10):997-1006.

64. Loewen G, Jayawickramarajah J, Zhuo Y, Shan B. Functions of lncRNA HOTAIR in lung cancer. *Journal of hematology & oncology.* 2014;7(1):90.

65. Gutschner T, Hammerle M, Eissmann M, et al. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer research.* Feb 1 2013;73(3):1180-1189.

66. Hu T, Lu YR. BCYRN1, a c-MYC-activated long non-coding RNA, regulates cell metastasis of non-small-cell lung cancer. *Cancer cell international.* 2015;15:36.

67. Wu Y, Liu H, Shi X, Yao Y, Yang W, Song Y. The long non-coding RNA HNF1A-AS1 regulates proliferation and metastasis in lung adenocarcinoma. *Oncotarget.* Apr 20 2015;6(11):9160-9172.

68. Lin L, Gu ZT, Chen WH, Cao KJ. Increased expression of the long non-coding RNA ANRIL promotes lung cancer cell metastasis and correlates with poor prognosis. *Diagnostic pathology.* Mar 27 2015;10(1):14.

69. OLIGO7. http://www.oligo.net/.

70. Colyer HA, Armstrong RN, Sharpe DJ, Mills KI. Detection and analysis of DNA methylation by pyrosequencing. *Methods Mol Biol.* 2012;863:281-292.

71. Frommer M, McDonald LE, Millar DS, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America.* Mar 1 1992;89(5):1827-1831.

72. Coronnello C, Benos PV. ComiR: Combinatorial microRNA target prediction tool. *Nucleic acids research.* Jul 2013;41(Web Server issue):W159-164.

73. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols.* 2009;4(1):44-57.

74. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. *PLoS medicine.* 2012;9(5):e1001216.

75. Ferracin M, Lupini L, Salamon I, et al. Absolute quantification of cell-free microRNAs in cancer patients. *Oncotarget.* May 2 2015.

76. Korpanty GJ, Graham DM, Vincent MD, Leighl NB. Biomarkers That Currently Affect Clinical Practice in Lung Cancer: EGFR, ALK, MET, ROS-1, and KRAS. *Frontiers in oncology.* 2014;4:204.

77. Powrozek T, Krawczyk P, Kowalski DM, Winiarczyk K, Olszyna-Serementa M, Milanowski J. Plasma circulating microRNA-944 and microRNA-3662 as potential histologic type-specific early lung cancer biomarkers. *Translational research : the journal of laboratory and clinical medicine.* May 29 2015.

78. Janmaat ML, Giaccone G. The epidermal growth factor receptor pathway and its inhibition as anticancer therapy. *Drugs Today (Barc).* 2003;39 Suppl C:61-80.

79. Hendler FJ, Ozanne BW. Human squamous cell lung cancers express increased epidermal growth factor receptors. *The Journal of clinical investigation.* Aug 1984;74(2):647-651.

80. Ellis PM, Coakley N, Feld R, Kuruvilla S, Ung YC. Use of the epidermal growth factor receptor inhibitors gefitinib, erlotinib, afatinib, dacomitinib, and icotinib in the treatment of non-small-cell lung cancer: a systematic review. *Curr Oncol.* Jun 2015;22(3):e183-215.

81. Rusch V, Baselga J, Cordon-Cardo C, et al. Differential expression of the epidermal growth factor receptor and its ligands in primary non-small cell lung cancers and adjacent benign lung. *Cancer research.* May 15 1993;53(10 Suppl):2379-2385.

82. Pfeiffer P, Clausen PP, Andersen K, Rose C. Lack of prognostic significance of epidermal growth factor receptor and the oncoprotein p185HER-2 in patients with systemically untreated non-small-cell lung cancer: an immunohistochemical study on cryosections. *British journal of cancer.* Jul 1996;74(1):86-91.

83. Ettinger DS. Clinical implications of EGFR expression in the development and progression of solid tumors: focus on non-small cell lung cancer. *The oncologist.* Apr 2006;11(4):358-373.

84. Pavelic K, Banjac Z, Pavelic J, Spaventi S. Evidence for a role of EGF receptor in the progression of human lung carcinoma. *Anticancer research.* Jul-Aug 1993;13(4):1133-1137.

85. Veale D, Kerr N, Gibson GJ, Kelly PJ, Harris AL. The relationship of quantitative epidermal growth factor receptor expression in non-small cell lung cancer to long term survival. *British journal of cancer.* Jul 1993;68(1):162-165.

86. Rusch V, Klimstra D, Venkatraman E, Pisters PW, Langenfeld J, Dmitrovsky E. Overexpression of the epidermal growth factor receptor and its ligand transforming growth factor alpha is frequent in resectable non-small cell lung cancer but does not predict tumor progression. *Clinical cancer research : an official journal of the American Association for Cancer Research.* Apr 1997;3(4):515-522.

87. Brabender J, Danenberg KD, Metzger R, et al. Epidermal growth factor receptor and HER2-neu mRNA expression in non-small cell lung cancer Is correlated with survival. *Clinical cancer research : an official journal of the American Association for Cancer Research.* Jul 2001;7(7):1850-1855.

88. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* Mar 4 2011;144(5):646-674.

89. Russo G, Zegar C, Giordano A. Advantages and limitations of microarray technology in human cancer. *Oncogene.* Sep 29 2003;22(42):6497-6507.

90. Sandoval J, Peiro-Chova L, Pallardo FV, Garcia-Gimenez JL. Epigenetic biomarkers in laboratory diagnostics: emerging approaches and opportunities. *Expert review of molecular diagnostics.* Jun 2013;13(5):457-471.

91. Xu G, Chen J, Pan Q, et al. Long noncoding RNA expression profiles of lung adenocarcinoma ascertained by microarray analysis. *PloS one.* 2014;9(8):e104044.

92. Yang Y, Li H, Hou S, Hu B, Liu J, Wang J. The noncoding RNA expression profile and the effect of lncRNA AK126698 on cisplatin resistance in non-small-cell lung cancer cell. *PloS one.* 2013;8(5):e65309.