M 2016

U.PORTO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

DNA METHYLATION PROFILE AS A TOOL FOR LUNG CANCER DIAGNOSIS & SUBTYPING

FRANCISCA OLIVEIRA DE MESQUITA DINIZ
DISSERTAÇÃO DE MESTRADO APRESENTADA
AO INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
DA UNIVERSIDADE DO PORTO EM
ONCOLOGIA – ESPECIALIZAÇÃO EM ONCOLOGIA MOLECULAR

Francisca Oliveira de Mesquita Diniz

DNA METHYLATION PROFILE AS A TOOL FOR LUNG CANCER DIAGNOSIS & SUBTYPING

Dissertação de Candidatura ao grau de Mestre em Oncologia – Especialização em Oncologia Molecular submetida ao Instituto de Ciências Biomédicas de Abel Salazar da Universidade do Porto.

Orientador

Professor Doutor Rui Manuel Ferreira Henrique

Professor Catedrático Convidado

Departamento de Patologia e Imunologia Molecular

Instituto de Ciências Biomédicas Abel Salazar

Universidade do Porto

Diretor do Serviço de Anatomia Patológica

Investigador Sénior do Grupo de Epigenética e Biologia do Cancro

Centro de Investigação

Instituto Português de Oncologia – Porto

Coorientadora

Professora Doutora Carmen de Lurdes Fonseca Jerónimo

Professora Associada convidada com Agregação

Departamento de Patologia e Imunologia Molecular

Instituto de Ciências Biomédicas Abel Salazar

Universidade do Porto

Investigadora Auxiliar e coordenadora do Grupo de

Epigenética e Biologia do Cancro

Centro de Investigação

Instituto Português de Oncologia – Porto

“If you can’t fly, then run,

if you can’t run, then walk,

if you can’t walk, then crawl,

but whatever you do,

you have to keep moving forward.”

-*Martin Luther King Jr.*

AGRADECIMENTOS

Seguir em frente quando estamos perante dificuldades torna-se muito mais fácil quando temos pessoas ao nosso lado dispostas a ajudar e a apoiar, é a elas que segue o meu agradecimento.

Antes de mais, ao Professor Rui Henrique e à Professora Carmen Jerónimo, meu orientador e coorientadora respetivamente, um obrigada por me terem recebido no Grupo de Epigenética e Biologia do Cancro, por todas as oportunidades, ensinamentos, horas dedicadas, críticas, conselhos e acima de tudo pela paciência. Obrigada por terem confiado em mim.

Ao Professor Manuel Teixeira, Diretor do Centro de Investigação do IPO-Porto, por ter autorizado a realização desta tese no Laboratório 3.

À Professora Berta Silva, Diretora do Mestrado em Oncologia, obrigada por me ter aceite no referido ciclo.

Agradeço ainda às Dras. Marta Soares e Ana Vítor Silva da Clinica de Pulmão, à Dra. Ana Luísa Cunha e à Técnica Sofia Paulino do Serviço de Anatomia Patológica, bem como ao Engº Luís Antunes do Serviço de Epidemiologia por toda a dedicação, paciência e contribuição para que a concretização desta dissertação.

Não poderia esquecer a Dona Luísa, amiga e sempre atenta ao que fazia falta no laboratório para que tudo corresse na normalidade e não nos faltasse nada. Obrigada pelos docinhos e pelos carinhos.

Ao João Carvalho, Inês Graça, Ana Luís e Sara Reis (a velha guarda), um sincero obrigado por terem estado sempre lá, atentos, dedicados e sempre com uma palavra de alento! Sem dúvida que com vocês aprendi muito mais do que poderia imaginar e nada teria sido igual sem o vosso apoio incondicional e os vossos conhecimentos. Além de tudo isto ainda me deram a vossa amizade e carinho que vou levar sempre no coração!

Aos meus colegas do dia-a-dia na luta no laboratório: Salta (a amiga a quem vou dever notas de 5€ até ao fim, obrigada por me apoiares desde as coisas mais difíceis até às mais básicas, obrigada por tudo o que me transmitiste e acima de tudo pela paciência!), Soraia (tantas saudades do teu bullying), Catarina B. (obrigada pelas conversas, teorias, apoio, discussões e paciência), Lameirinhas (raspas lâminas como ninguém), Daniela (sempre pronta a mandar uma piada perversa com duplo sentido), Micaela (nunca esquecerei a tua voz de comando nem o apoio nos dias de maior angústia), Nuno (és e serás sempre insubstituível, nunca conheci ninguém como tu, obrigada por nunca falhares), Catarina Araújo (obrigada pela tua loucura e lado guna que me animou sempre), Diana (a minha amiga chique, grito histérico e mega sorrisão, obrigada pelos

conselhos e brincadeiras) David, Maria, Ângela e Ana laura (aos 4 mais novos desejo tudo de bom e que o ano de trabalho que se avizinha seja fácil de levar).
Eva e Maria João, minhas compinchas do mestrado, do laboratório e da vida! Com vocês partilhei 2 anos incríveis e sei que o próximo ano irá trazer coisas muito boas para as 3 e que vão ser vividas juntas, isso deixa-me extremamente feliz! Obrigada por tudo! Por estarem nos dias bons, nos dias menos bons, na saúde e literalmente na doença, de dia e de noite, nos sorrisos e nas lágrimas! Sois a melhor surpresa com que este mestrado me presenteou.
Tenho que agradecer também às minhas irmãs que trago comigo há 5 anos, desde o primeiro ano de licenciatura. Marta, Ju, Sara e Renata, a vocês é impossível enumerar tudo o que de bom me deram nestes 5 anos académicos, mas sei que sem o vosso apoio incondicional há muito que teria desistido. Obrigada pelas horas de gargalhadas, de discussões, de gritos de felicidade, pelos abraços e pelo vosso sorriso. Sei que por muito que nos desencontremos, acharemos sempre um caminho que nos traga de volta!
Aos meus colegas de licenciatura, Mário Gil e Vanessa, obrigada por não deixarem que a distância e mudanças alterassem a nossa amizade e por estarem sempre por perto.
Agradeço ainda aos meus amigos de Monção, que felizmente por mais meses que passemos afastados nunca falhou a vossa amizade e os nossos reencontros mostram que os anos não passam.
Ao Lucas e Mateus, os meus irmãos, obrigada por nunca perderem o vosso lado de pestes e pelo carinho que conseguem demonstrar mesmo estando constantemente a chatear-vos a cabeça! Quero sempre que o nosso espirito de javardice e união se mantenha. Obrigada por me porem a sorrir nos dias em que a vontade para isso não era assim tanta.
Aos meus avós, Eduardo e Ju, um obrigada por todas as palavras, abraços, ensinamentos e acima de tudo um obrigada pela coragem que sempre me motivaram a ter. Um obrigada cheio de saudade à minha avó Ofélia, da qual guardo as seguintes palavras: “Não deixes morrer em ti a esperança e continua a gostar da vida”.
Por fim, mas os últimos são os primeiros, um profundo agradecimento aos meus pais. Sempre lutaram para me incutir os valores que achavam certos, motivaram-me a correr pelo que eu queria alcançar sem nunca atalhar pelos facilitismos e fizeram de tudo para ter tudo o que acharam necessário. Obrigada por me deixarem sonhar sem nunca cair na ilusão do mundo perfeito! Obrigada por me ampararem nas quedas e as que falharam me ajudarem a levantar É devido a vocês que sou como sou e que consegui chegar onde cheguei. Obrigada por terem dito sim mas também por terem dito não. Obrigada! Simplesmente isso.

É aos meus ídolos a quem dedico todo o trabalho desta tese, os meus pais, Maria João e Pedro.

This study was funded by a grant of the Research Centre of Portuguese Oncology Institute of Porto (CI-IPOP-74-2016).

SUMMARY

BACKGROUND: Lung cancer (LCa) is the most common cancer in both men and women, being responsible for more deaths than any other malignancy. It is well known the importance of early diagnosis of LCa and personalized therapy according to disease genomic characteristics. Thus, discrimination between the LCa subtypes becomes a key to reduce the mortality rate because allows for more specific treatments. LCa are broadly classified into two groups: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), the latter being divided into adenocarcinoma, squamous cell carcinoma (SCC) and large cell carcinoma (LCC), among other less frequent subtypes.

AIMS: The main goal of this dissertation was to evaluate the methylation profile of the major LCa subtypes, with a panel of genes previously reported to be hypermethylated in LCa: *APC*, *HOXA9*, *RARβ2*, *RASSF1A*, *TFPI2* and *SHOX2*. In particular, we aimed to discriminate adenocarcinoma, the most prevalent subtype, from the other LCa subtypes. Moreover, we evaluated the association between the gene-panel methylation levels and standard clinicopathological parameters as well as determined the prognostic value of the same gene-panel.

MATERIAL AND METHODS: Methylation levels of *APC, RASSF1A, RARβ2*, *HOXA9*, *SHOX2* and *TFPI2* were assessed using real-time quantitative MSP in bisulfite-modified DNA extracted from formalin-fixed paraffin embedded tissue samples from 152 LCa and 22 normal lung parenchyma (NL) from individuals with other neoplasias. Survival analyses were conducted to evaluate its prognostic value.

RESULTS: Methylation levels of *APC*, *HOXA9*, RARβ2 and *RASSF1A* discriminated the major subtypes, NSCLC from SCLC ($P < 0.001$; $P = 0.021$; $P < 0.001$; $P < 0.001$; respectively). *APC* and *RASSF1A* distinguished SCC and Adenocarcinoma from SCLC ($P < 0.001$; $P < 0.001$; respectively), whereas. *RARβ2* discriminated all subtypes of NSCLC from SCLC (Adenocarcinoma vs SCLC, $P < 0.001$; SCC vs SCLC, $P < 0.001$; LCC vs SCLC, $P = 0.036$). *HOXA9* also differentiated Adenocarcinoma from SCLC ($P < 0.001$), and it was the only gene that discriminated Adenocarcinoma from SCC ($P = 0.024$).. Low *APC*, *HOXA9*, *RARβ2* and *RASSF1A* promoter methylation levels associated with poorer disease specific survival, although not independently as it was dependent of poor tumor differentiation. Low *RASSF1A* promoter methylation levels also predicted poor disease-free survival in univariable analysis but due to its association with tumor differentiation, it did not retain independent prognostic significance in multivariable analysis.

CONCLUSIONS: Assessment of *RARβ2* and *HOXA9* promoter methylation levels using qMSP is able to to discriminate among major LCa subtypes in tissue samples. The clinical usefulness of these biomarkers in plasma will be tested in the near future.

RESUMO

INTRODUÇÃO: O cancro do pulmão é mais comum tanto em homens como em mulheres, sendo, ainda, o principal responsável pela mortalidade associada a cancro. Está bem estabelecida a importância de um diagnóstico precoce de cancro do pulmão bem como a instituição de terapia personalizada, a qual é realizada de acordo com as características da neoplasia. Portanto, descriminar precocemente os principais subtipos de cancro de pulmão torna-se determinante para reduzir a taxa de mortalidade, uma vez que permite melhor especificar as estratégias terapêuticas. O cancro do pulmão é habitualmente classificado em dois grupos: carcinoma de pequenas células e carcinoma de não pequenas células. Este último subdivide-se em adenocarcinoma, carcinoma epidermoide e carcinoma de grandes células, para além de outros subtipos menos expressivos em termos de frequência.

OBJECTIVOS: O objetivo principal desta dissertação de mestrado foi avaliar os perfis de metilação de diferentes subtipos de cancro do pulmão com um painel de genes previamente descritos na literatura - *APC*, *HOXA9*, *RARβ2*, *RASSF1A*, *TFPI2* e *SHOX2*. Mais especificamente, pretendeu-se descriminar os adenocarcinomas, que representam o subtipo mais prevalente, dos restantes subtipos de cancro do pulmão. Adicionalmente, foi analisada a associação entre os níveis de metilação do painel de genes e as características clinico-patológicas, bem como o valor prognóstico.

MATERIAL E MÉTODOS: Os níveis de metilação de *APC, RASSF1A, RARβ2*, *HOXA9*, *SHOX2* e *TFPI2* foram determinados através de PCR quantitativo de metilação em tempo real, utilizando DNA modificado por bissulfito de sódio extraído de amostras de tecido fixado em formol e incluído em parafina de 152 cancros do pulmão e 22 amostras de parênquima pulmonar normal proveniente de indivíduos com outras neoplasias. A análise de sobrevivência foi realizada para avaliar o valor prognóstico dos genes do painel.

RESULTADOS: Os níveis de metilação dos genes *APC*, *HOXA9*, RARβ2 e *RASSF1A* descriminaram os subtipos principais de cancro de pulmão (P < 0.001; P = 0.021; P < 0.001; P < 0.001; respetivamente). *APC* e *RASSF1A* diferenciaram os carcinomas epidermoides e adenocarcinomas dos carcinomas de pequenas células (P < 0.001; P < 0.001; respetivamente), enquanto que *RARβ2* descriminou todos os subtipos pertencentes aos carcinomas de não pequenas células dos carcinomas de pequenas células (Adenocarcinoma vs Carcinoma de pequenas células, P < 0.001; Carcinoma epidermoide vs Carcinoma de pequenas células, P < 0.001; Carcinoma de grandes células vs Carcinoma de pequenas células, P = 0.036). *HOXA9* também diferenciou os adenocarcinomas dos carcinomas de pequenas células (P < 0.001), sendo, ainda, o

único gene a descriminar os adenocarcinomas dos carcinomas epidermoides (P = 0.024). Por outro lado, *SHOX2* e *TFPI2* não mostraram diferenças estatisticamente significativas entre nenhum dos subtipos. Níveis baixos de metilação do promotor dos genes *APC*, *HOXA9*, *RARβ2* e *RASSF1A* associaram-se a pior sobrevivência específica de doença, mas dependente do grau de diferenciação, enquanto que baixos níveis de metilação do gene *RASSF1A* se associaram a pior sobrevivência livre de doença, mas também dependente do grau de diferenciação.

CONCLUSÕES: A avaliação dos níveis de metilação dos promotores dos genes *RARβ2* e *HOXA9* podem ser úteis para descriminar os subtipos de cancro do pulmão em amostras de tecidos parafinado. A utilidade clínica destes genes como biomarcadores em plasma será avaliada num futuro próximo.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND SYMBOLS xv

1. INTRODUCTION 6

1.1 Lung Cancer 2

1.1.1 Epidemiology and causes 2

1.1.1.1 Risk Factors 4

1.1.1.1.1 Cigarettes 4

1.1.1.1.2 Exposure to Other Carcinogens 4

1.1.1.1.3 Family History 5

1.1.2 Lung Cancer Subtypes 5

1.1.2.1 SCLC 6

1.1.2.2 NSCLC 6

1.1.2.2.1 Adenocarcinoma 6

1.1.2.2.2 Squamous Cell Carcinoma (SCC) 6

1.1.2.2.3 Large Cell Carcinoma (LCC) 7

1.1.3 Diagnosis 7

1.1.4 Staging 8

1.1.5 Prognosis 9

1.1.6 Treatment 10

1.1.6.1 Surgery 10

1.1.6.2 Radiotherapy (RT) 10

1.1.6.3 Chemotherapy (ChT) 11

1.1.6.4 Targeted Therapy 11

1.2 Epigenetics 15

1.2.1 Epigenetic Mechanisms 15

1.2.1.1 Non-coding RNAs 16

1.2.1.2 Histone Post-translational Modifications and Variants 16

1.2.1.3 DNA methylation 16

1.2.2 DNA methylation in Lung Cancer 18

1.2.2.1 Hypermethylated genes in Lung Cancer 19

1.2.2.1.1 APC 20

1.2.2.1.2 HOXA9 21

1.2.2.1.3 RARβ2 21
1.2.2.1.4 RASSF1A 21
1.2.2.1.5 SHOX2 22
1.2.2.1.6 TFPI2 22
2. AIMS 2
3. MATERIALS AND METHODS 24
3.1 Study cohort – Patients and Samples 26
3.2 DNA Extraction From Formalin-fixed Paraffin-embedded Tissues (FFPE) 26
3.3 Bisulfite treatment of DNA, quantitative methylation- specific polymerase chain reaction (qMSP) 26
3.4 Statistical analysis 28
4. RESULTS 30
4.1 Clinical Samples 30
4.2 Assessment of aberrant promoter methylation levels in LCa and controls 31
4.3 Association between quantitative promoter methylation and clinicopathological parameters 32
4.4 Distribution of methylation levels according to major LCa subtypes 32
4.5 Distribution of methylation levels according to LCa histological subtypes 33
4.6 Survival analyses 37
4.6.1 Disease- Specific Survival 37
4.6.2 Disease-Free Survival 42
5. DISCUSSION 30
6. CONCLUSIONS AND FUTURE PERSPECTIVES 46
7. REFERENCES 52
APPENDICES 54
APPENDIX I II
APPENDIX II III
APPENDIX III IV
APPENDIX IV V

FIGURES INDEX

Figure 1 - (A) Estimated worldwide cancer incidence rates for both genders; (B) Estimated worldwide cancer mortality rates for both genders. Adapted from Ferlay, 2012[3]. 2
Figure 2 - Histological patterns of lung cancer. (A) Small cell lung cancer; (B) Adenocarcinoma, (C) Squamous cell carcinoma; (D) Large cell carcinoma. 5
Figure 3 - Staging of lung cancer according to the TNM system[13]. 9
Figure 4 - Molecular diagnosis and treatment of lung cancer at different stages [10]. 10
Figure 5 - Proportion of known driver mutations in Non-small cell lung cancer(NSCLC)[16]. 12
***Figure 6* -** Overview of molecular pathways and potential targets in Non-small cell lung cancer (NSCLC)[16]. 14
Figure 7 - Four distinct mechanisms of epigenetic regulation [19]..................................... 15
Figure 8 - Conversion of cytosine to 5-methylcytosine by DNA methyltransferase (DNMT). The methyl group (CH_3) is transferred from S-adenosylmethionine (SAM) to 5-carbon position of cytosine by DNMT[25]. 17
Figure 9 - DNA methylation in normal and cancer cells. In normal cells, promoter Cpg islands are unmethylated while in cancer cells, they have acquired aberrant DNA methylation, and consequently, transcriptional silencing. Adapted from Baylin, 2015 [29]. ..18
Figure 10 - Cancer is a result of the interaction between permanent genetic mutations and dynamic epigenetic alterations. During cancer formation, a large number of epigenetic and genetic alterations lead to abnormal gene expression which evoke genome instability. Adapted from Mehta, 2015; Chen, 2014 [10, 31]. 19
Figure 11 – DNA modification by sodium bisulfite. Following bisulfite conversion, methylated cytosines remain unchanged, while unmethylated cytosines are deaminated to uracils. The level of DNA promoter hypermethylation its quantified by quantitative methylation-specific PCR (qMSP). 27
Figure 12 - Boxplots of (A) RASSF1A, (B) RARβ2, (C) APC and (D) HOXA9 promoter methylation levels between Lung Cancer (LCa) and normal lung (NL) samples. 31
Figure 13 - Boxplots of the methylation levels of (A) APC, (B) RARβ2 and (C) RASSF1A in the different stages (Mann Whitney Test, **P < 0.010; P*** < 0.01). 32
Figure 14 – Boxplots of (A) RASSF1A, (B) RARβ2, (C) APC and (D) HOXA9 promoter methylation levels between the major subtypes of Lung Cancer (LCa) samples. (NSCLC: Non-small cell lung cancer) (Mann Whitney Test, *P < 0.001; P*** < 0.01) 33
***Figure 15* –** Boxplots of (A) RASSF1A, (B) RARβ2, (C) APC and (D) HOXA9 promoter methylation levels between subtypes of Lung Cancer (LCa) samples. (SCC: Squamous

cell carcinoma; Ade: Adenocarcinoma; LCC: Large cell carcinoma; SCLC: Small cell lung cancer) (Mann Whitney Test, *P < 0.001; P*** < 0.01) 35

Figure 16 – Schematic representation of the association between genes promoter methylation and discrimination of major and minor Lung Cancer (LCa) subtypes. (NSCLC: Non-small cell lung cancer; SCLC: small cell lung cancer)..36

Figure 17 - Disease-Specific Survival according to (A) RASSF1A, (B) RARβ2, (C) APC and (D) HOXA9 methylation levels..37

Figure 18 – Disease-Specific Survival according to (A) Differentiation grade, (B) Stage and (C) Histological subtypes. (SCC: Squamous cell carcinoma; Ade: Adenocarcinoma; LCC: Large cell carcinoma; SCLC: Small cell lung cancer)...38

Figure 19 - Disease-Free Survival according to (A) Differentiation grade, (B) RASSF1A methylation level. 42

Figure 20 - Boxplots of (A) TFPI2 and (B) SHOX2 promoter methylation levels between Lung Cancer (LCa) and normal lung (NL) samples. IV

Figure 21 - Boxplots of (A) TFPI2 and (B) SHOX2 promoter methylation levels between the major subtypes of Lung Cancer (LCa) samples. (NSCLC: Non-small cell lung cancer)V

Figure 22 - Boxplots of (A) TFPI2 and (B) SHOX2 promoter methylation levels between subtypes of Lung Cancer (LCa) samples. (SCC: Squamous cell carcinoma; Ade: Adenocarcinoma; LCC: Large cell carcinoma; SCLC: Small cell lung cancer) VI

TABLES INDEX

Table 1 - Panel of genes hypermethylated in lung cancer. 20
Table 2 - Primers sequences used and qMSP conditions for each of the tested genes. ..28
Table 3 - Clinical and Histopathological characteristics of patients with Lung cancer and Normal pulmonary parenchyma. 30
Table 4 - Kruskal Wallis and Mann Whitney tests analyze of APC, HOXA9, RARβ2, RASSF1A, SHOX2 and TFPI2 promoter methylation levels between subtypes of Lung Cancer (LCa) samples. 34
Table 5 - Distribution of promoter methylation levels of cancer-related genes in Lung Cancer (LCa) samples measured by qMSP. 34
Table 6 – Univariate Cox regression analyses assessing the potential of clinical and epigenetic variables in the prediction of disease-specific survival for 152 LCa patients....39
Table 7 - Multivariate Cox regression analyses assessing the potential of clinical and epigenetic variables in the prediction of disease-specific survival for 152 LCa patients....40
Table 8 - Multivariate Cox regression analyses assessing the potential of clinical and epigenetic variables in the prediction of disease-specific survival for 152 LCa patients....40
Table 9 - Multivariate Cox regression analyses assessing the potential of clinical and epigenetic variables in the prediction of disease-specific survival for 152 LCa patients....41
Table 10 - Multivariate Cox regression analyses assessing the potential of clinical and epigenetic variables in the prediction of disease-specific survival for 152 LCa patients....41
Table 11 – Univariate Cox regression analyses assessing the potential of clinical and epigenetic variables in the prediction of disease-free survival for 152 LCa patients. 43
Table 12 - Multivariate Cox regression analyses assessing the potential of clinical and epigenetic variables in the prediction of disease-free survival for 152 LCa patients 44

LIST OF ABBREVIATIONS AND SYMBOLS

***ACTβ* -** *β*- Actina

Ade - Adenocarcinoma

ALK - Anaplastic lymphoma kinase

***APC* -** Adenomatous polyposis coli

ChT - Chemotherapy

CpG - Cytosine-phosphate-Guanine

CT - Computerized topographies

cTNM - Clinical TNM

DFS - Disease-Free Survival

DNMT - DNA methyltransferase

DSS - Disease-Specific Survival

EGFR - Epidermal growth factor receptor

EML4 - Echinoderm microtubule-associated protein like-4

FFPE - Formalin-fixed paraffin-embedded

LCa - Lung Cancer

LCC - Large cell carcinoma

miRNA - microRNA

mRNA - messenger RNA

MSP - Methylation Specific PCR

NcRNA - Non-coding RNA

NL - Normal Lung

NLST - National Lung Screening Trial

NSCLC - Non-small cell lung cancer

PCR - Polymerase chain reaction

PET - Positron emission tomography

pTNM - Pathological TNM

qMSP - Quantitative methylation specific PCR

***RARβ2* -** Retinoic acid receptor β2

***RASSF1A* -** Ras association domain family 1 isoform A

RT - Radiotherapy

SAM - S-adenosymethionine

SCC - Squamous cell carcinoma

SCLC - Small cell lung cancer

***SHOX2* -** Short stature homeobox 2

TCGA - The cancer genome atlas

***TFPI2* -** Tissue factor pathway inhibitor 2

TKI - Tyrosine kinase inhibitor

TSG - Tumor suppressor genes

5mC - 5-methylcytosine

1. INTRODUCTION

1.1 Lung Cancer

1.1.1 Epidemiology and causes

Lung cancer (LCa) has been considered the most commonly diagnosed cancer in the world for several decades. As the leading cause of cancer related death in the world, LCa is currently a public health problem of enormous magnitude[1]. In 2012, LCa was estimated to be the most common cancer worldwide (12.9% of total diagnosed cases) and the leading cause of cancer- related deaths worldwide (19.4% of total cancer cases). In fact it represents more than one-fifth of all cancer related deaths, which is higher than breast, colon and prostate cancer combined (Figure 1). Despite the incidence rate is lower in women compared to men, its remains the main cause of death by cancer for both genders[2].

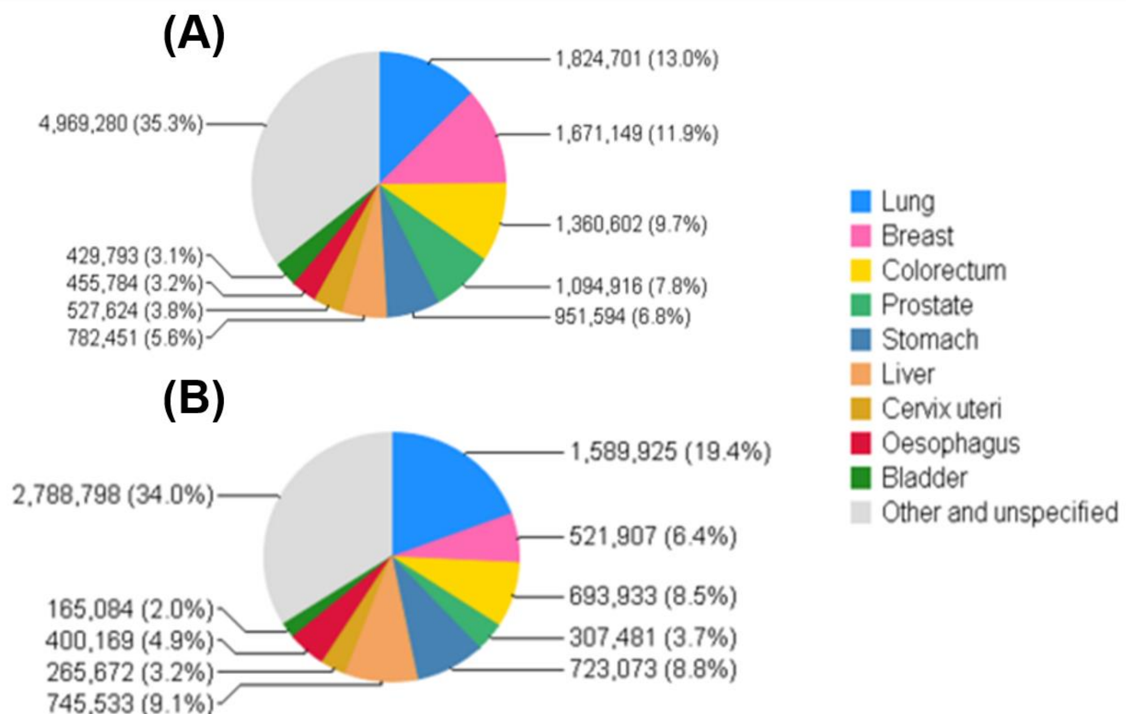

Figure 1 - (A) Estimated worldwide cancer incidence rates for both genders; (B) Estimated worldwide cancer mortality rates for both genders. *Adapted from Ferlay, 2012[3].*

Incidence and mortality rates of LCa in Europe are slightly different to those that characterize worldwide distribution. Regarding the incidence rate, is the fourth more incident, representing 11.9% of total diagnosed cases, more specifically the second more frequent in men (15.9%) and the third more common at women (7.4%). In terms of mortality it is considered the most frequent cause of cancer related deaths in Europe (one fifth of the total), being the most common cause of cancer death in men (26.1%), and the third in women (12.7%)[4].

In Portugal, LCa is the fourth most frequent malignant neoplasia and is the second most mortal, following closely colorectal cancer (8.5%) [4].

Importantly in LCa, the mortality rate parallels the incidence rate mainly due to persistently low patient survival. Despite the development of clinical diagnosis techniques and treatment, the overall 5-year survival rate remains extremely low (10%). This poor outcome is attributable not only to the fact that almost two thirds of cases are diagnosed at advanced stages but also to the high rate of recurrence after surgical resection [5, 6].

LCa tends to be most incident in developed countries, especially in North America and Europe, and less common in developing countries, particularly in Africa and South America[1].

At older age groups, both mortality and incidence rates continue to increase for both genders. However the increasing rates are decelerating more in men than in women. Regarding younger age groups, the rates of LCa are decreasing, for both genders, being more evident in men than women[1].

The hypothesis that women might have a greater LCa risk than men with the same smoking habits has been suggested. Nevertheless, several other studies that compared the relative risk of a specific degree of smoking history for men and women demonstrate very similar risks. Interesting differences in LCa characteristics between men and women have been noted. First, women with LCa present a better prognosis than men. Second, estrogens may augment lung cancer risk. Third, among never smokers, women have higher percentage of adenocarcinomas and higher prevalence of EGFR mutations than men. These observations suggest that distinct gender differences in lung carcinogenesis might potentially be clinically important[1].

Nonetheless, there are other factors considered to be possible risk factors, such as asbestos, pulmonary chronic disease, environmental pollution or family history[7].

1.1.1.1 Risk Factors

Several risk factors contribute for development of LCa including tobacco smoking, asbestos, radon, environmental pollution or family history[7].

1.1.1.1.1 Cigarettes

Cigarette smoking is by far the leading cause of LCa[8]. About 85% of LCa patients presents a tobacco-smoking history and approximately 50% were former smokers[9]. The risk increase with duration of smoking and the number of cigarettes smoked daily[1]. Patients with a smoking history of at least 20 to 30 pack-years present a substantially increased risk to develop LCa. Smoking cessation is associated with a gradual reduction, however it does not reach that of a never smoker.[9] Smoking confers an approximately 25-fold increased risk for lung cancer in current smokers[1]. Tobacco smoke is characterized by a complexity of compounds that promote damage in lung cells and clearly contributes to the accumulation of genetic alterations in lung cancer[8]. Most of lung cancer cases in men (85%) and nearly half of lung cancer cases in women are estimated as being the consequence of tobacco smoking[7].

Passive exposure to cigarette is another risk factor that contributes to nearly 1% of all cases of LCa[9]. Passive smokers inhale a complex mixture of smoke, which is now widely referred to as "environmental tobacco smoke". Passive smoking is more weakly associated with LCa than is active smoking. This fact is due to the lower doses of carcinogens received by the nonsmoker compared with the smoker. Marriage to a smoker has been associated with about a 20% risk increase and exposure in the workplace has been associated with an increased risk of 24% to a twofold increase at the highest levels of exposure[1].

1.1.1.1.2 Exposure to Other Carcinogens

Occupational exposure to lung carcinogens have been estimated to account for about 9% to 15% of LCa cases. Cigarette smoking potentiates the effect of some of the known occupational lung carcinogens[1]. Asbestos, is a well-established occupational carcinogen which acts synergistically with smoke and increase the risk to LCa. Occupational exposure to asbestos leads to an estimated 4-fold higher risk for LCa[9]. Specifically, a person who smokes and has been exposed to asbestos has a greater than 50-fold elevated risk for LCa than does a nonsmoker with no asbestos exposure[1]. Radon exposure has also been implicated in the development of 5 to 8% of lung cancer cases[9].

1.1.1.1.3 Family History

A positive family history of lung cancer is a clinically useful risk indicator[1]. Patients with a family history of early lung cancer (before 60 years old) accounts for an approximately 2.5-fold increased risk. Genetic susceptibility may be seen with rare autosomal dominant genes that explain only few cases of early-onset LCa. Contrarily, common genetic variants or polymorphisms are more likely to affect LCa risk[8].

Nevertheless, as with smoking, not all who are exposed to these environmental factors go on to develop lung cancer[1].

1.1.2 Lung Cancer Subtypes

Lung cancer may present multiple histologic types as classified by conventional light microscopy. There are two main histological groups: small cell lung cancer (SCLC) comprising approximately 20% of LCa cases and non-small cell lung cancer (NSCLC), which represents the remaining lung tumors (Figure 2). Histologically, NSCLC include three major histological subtypes: adenocarcinoma – the most prevalent form (40%) –, squamous cell carcinoma (25%) and large-cell carcinoma (10%)[10].

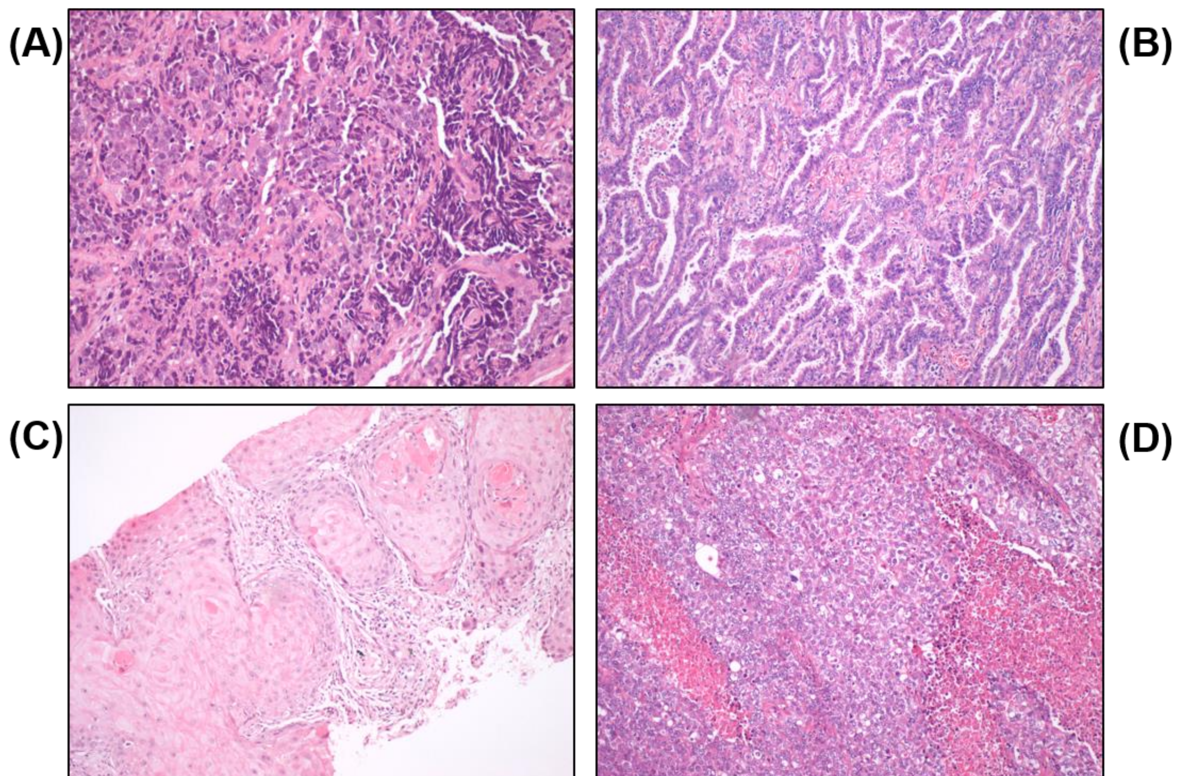

Figure 2 **-** Histological patterns of lung cancer. (A) Small cell lung cancer; (B) Adenocarcinoma, (C) Squamous cell carcinoma; (D) Large cell carcinoma.

1.1.2.1 SCLC

SCLC is one of the most aggressive and rapidly growing types of LCa. This subtype is characterized by a poor prognosis due to a propensity for early hematogenous dissemination[7]. Cigarette smoking has a strong connection with this type of cancer, being 98% of all SCLC cases caused by tobacco smoke. Clinically and biologically is considered different from NSCLC. Pathologic diagnosis can be challenging because of an abundance of necrotic tissue but is established by characteristic features such dense sheets of small cells with scant cytoplasm, finely granular nuclear chromatin, high degree of mitoses, necrosis and inconspicuous or absent nucleoli (Figure 2A)[9]. SCLC has a dismal prognosis, with a 2 year survival rate of only 10% with metastatic disease and a 5 year survival rate of approximately of 25% with is no metastatic involvement. Younger age, female gender and surgery for limited disease are favorable features. Contrarily, continued smoking is a strong adverse prognostic factor. It is frequently identified by chest imaging, and more specifically, in lung parenchyma, that may spread along bronchi in a subepithelial and radial pattern, also involving lymphatic vessels[7].

1.1.2.2 NSCLC

1.1.2.2.1 Adenocarcinoma

In the last two decades, adenocarcinoma incidence has been rising and it is now the most predominant histological subtype, surpassing squamous cell cancer[9]. This might be due to the changes in the design and in the characteristics of manufactured cigarettes which might have increase the puff volume, causing a shift from more central deposition of tobacco smoke to more peripheral deposition. This is particularly relevant since this type of LCa usually originate in peripheral airways[7]. Moreover, malignant lesions in this region may be present for a long time before symptoms manifestation, being mostly diagnosed in advanced stages. Histologically adenocarcinomas are characterized by glandular differentiation with mucin production (Figure 2B)[11]. Generally, this histological type is diagnosed in women, non-smokers and in Asians. However, never-smokers and women are favorable prognostic factors. There are several subtypes of adenocarcinoma, however, the majority are histologically heterogeneous and thus classified as mixed[7]. Compared with squamous cell, this subtype is prone to develop distant metastasis. Invasive adenocarcinoma represents nearly 90% of all cases of adenocarcinoma[7, 9].

1.1.2.2.2 Squamous Cell Carcinoma (SCC)

Squamous cell carcinoma, also known as epidermoid carcinoma, represents the second most incident subtype[7]. Histologically it is characterized as a malignant epithelial

tumor that shows keratinization and/or intercellular bridges (Figure 2C)[12]. This type of LCa grows commonly in central areas around major bronchi in a stratified or pseudo ductal arrangement[9]. Commonly is has a slow development, increasing the probability of finding it in early stages compared to other types of LCa. However, SCC has a tendency to be locally aggressive, involving adjacent structures through direct invasion. This subtype is more common in men and smokers, when compared with other histological subtypes[7].

1.1.2.2.3 Large Cell Carcinoma (LCC)

Large cell carcinomas have been classified as poorly differentiated carcinomas that lack any squamous or adenocarcinoma differentiation (Figure 2D). Gene expression profiling has shown evidence of epithelial-mesenchymal transition as a frequent finding in large cell carcinomas, reflecting their poor differentiation compared to other NSCLC. Only when additional staining is negative, unclear, or not available the diagnosis of large cell carcinoma is made. However, their incidence is decreasing as a reflection of alteration in the approach of in pathologists' diagnostic which is mainly due to the introduction of immunohistochemistry for glandular and squamous markers. LCC lesions are typically localized on peripheral solid masses that are usually large, circumscribed, commonly with necrosis, but rarely with cavitation[7]. LCC, commonly has a rapid growth associated with a vast capacity to spread. This subtype is often associated with an aggressive clinical course and poor survival rates, even when it is found in the setting of early-stage disease[9]. Classification as large cell carcinoma requires morphological and immunohistochemical exclusion of other tumor types, as both cytological appearances can occur in other types of NSCLC[7]. Since this histologic subtype is often difficult to accurately diagnose owing to an abundance of necrotic tissue and poor degree of differentiation, diagnosis requires an adequate tissue sampling. Most of LCC patients are smokers[9].

1.1.3 Diagnosis

As LCa symptoms are similar to those of common several disease they are sometimes disregarded and the diagnosis is often delayed. There are several symptoms connected with the presence of LCa depending on the degree of tumor development[11]. Some symptoms that should raise suspicion of LCa are coughing up blood (hemoptysis), chest and bone pain, breathing problems, weakness or loss of sensation in body parts. It is imperative, when this symptoms are detected, to determine whether these alterations are due LCa or other respiratory disease[11, 12].

The majority of LCa patients have other tobacco-related cardiopulmonary diseases, therefore these overlapping symptoms often result in a late diagnosis of malignant disease[9].

Moreover, at diagnosis, only 15% of patients with LCa are asymptomatic. Accurate diagnostic characterization of lung cancer is essential, since the status of mediastinal nodal metastases is crucial for determining prognosis, assessing resectability, and selecting the appropriate treatment strategy for primary LCa[9].

Early stage LCa is often manifested as pulmonary nodules, defined as “rounded opacity, well or poorly defined, measuring up to 3 cm in diameter”. Pulmonary nodules may often be due to current or prior infection, although they also may be the manifestation of early cancer. National Lung Screening Trial (NLST) demonstrated that more than 95% of all detected nodules were false positives and noncancerous[9].

Nowadays, the major detection tools are evaluation of clinical history, bronchoscopy (to allow evaluation of the extent of the disease in the tracheobronchial tree), blood tests joined with physical exams (to examine the general signs of health), chest x-ray (to evaluate the presence and size of tumors or abnormal fluid in the chest), computerized topographies (CT) scan (to examine the disease extent or the presence of pulmonary nodules) and biopsy (to allow tumor identification). However, the majority of diagnosis are made incidentally on a chest radiography[9, 12].

1.1.4 Staging

When a tumoral mass is detected during diagnosis, LCa staging is essential for selection of the most appropriate treatment. Patients are staged according to the TNM classification for malignant tumors. This classification accounts the location and extension of the tumor, which might be organ confined or disseminated (lymph nodes, bones, liver and adrenal gland)[12, 13].

The letter **T** describes the size and degree of locoregional invasion of the primary tumor. The letter **N** indicates the extent of regional lymph node involvement and the letter **M** shows the presence of distant metastases (Figure 3)[13].

TNM can be based on clinical diagnostic examinations (cTNM) or based on surgical/pathological material (pTNM). Clinical classification is based on the evidence acquired before treatment, including physical examination, imaging studies, laboratory tests and staging procedures (bronchoscopy for example). Pathological classification uses the evidence acquired before treatment, supplemented or modified by the additional evidence acquired during and after surgery (particularly from pathologic examination).

TNM is essential to treatment planning, evaluation of treatment outcomes and prognosis determination[14].

The staging process is more efficient when several specific examinations are combined, as like blood tests, biopsies, surgical evaluation, CT scan or positron emission tomography (PET). Preoperative biopsies are less invasive and allows for an adequate sampling[13]. Furthermore, the recommended biopsy procedures for screen-detected suspicious pulmonary nodules resulted in a low intervention rate for benign nodules[15]. CT scan with contrast injection is the most requested staging technique since it allows visualization of metastasis in several organs. This exam has two major limitations namely the lack of ability to detect microscopic metastatic disease and the high rate of benign nodules' detection. PET scan has a great sensitivity, allowing visualization of the metabolic activity of malignant disease. Moreover, it has the ability to characterize LCa nodular stage[12, 13, 15].

Supraclavicular	Scalene	Mediastinal Contra-	Mediastinal Ipsi-	Subcarinal	Hilar Contra-	Hilar Ipsi-	Peribronchial (ipsilateral)	Lymph Node (N)
+	+	+			+			N3
–	–	–	+ &/ +		–			N2
–	–	–	–	–	–	+ &/ +		N1
–	–	–	–	–	–	–	–	N0

Lymph Node (N)	T1a	T1b	T2a	T2b	T3	T4
M1a or M1b	Stage IV (Metastatic: M1a or M1b, any T, any N)					
N3	Stage IIIB	Stage IIIB	Stage IIIB	Stage IIIB	Stage IIIB	Stage IIIB
N2	Stage IIIA	Stage IIIA	Stage IIIA	Stage IIIA	Stage IIIA	Stage IIIB
N1	Stage IIA	Stage IIA	Stage IIA	Stage IIB	Stage IIIA	Stage IIIA
N0	Stage IA	Stage IA	Stage IB	Stage IIA	Stage IIB	Stage IIIA

T1a	T1b	T2a	T2b	T3	T4	Primary Tumor (T)
≤2cm	>2cm but ≤3cm	>3cm but ≤5cm	>5cm but ≤7cm	>7cm	Any	a. Size
No invasion proximal to lobar bronchus		Main bronchus (≥2cm distal to the carina)		Main bronchus (<2cm distal to the carina)	–	b. Endo-bronchial location
Surrounded by lung or visceral pleura		Visceral pleura		Chest wall/diaphragm/ mediastinal pleura/ parietal pericardium	Mediastinum/trachea/heart/ great vessels/esophagus/ vertebral body/carina	c. Local Invasion
		Atelectasis/obstructive pneumonitis that extends to the hilar region but does not involve the entire lung		Atelectasis/obstructive pneumonitis of entire lung; separate tumor nodule(s) in ipsilateral primary tumor lobe	Separate tumor nodule(s) within the ipsilateral lung but different lobe as the primary mass	d. Other

Metastatic (M):

M1a:
Local intrathoracic spread:
- Malignant pleural/pericardial effusion
- Separate tumor nodule(s) in the contralateral lung

M1b:
Disseminated (extrathoracic) disease:
Liver, bone, brain, adrenal gland, etc.

Figure 3 - Staging of lung cancer according to the TNM system[13].

1.1.5 Prognosis

This disease is often asymptomatic in early stages. Thus, at diagnosis most of the patients present locally advanced or metastatic disease and therefore a worse prognosis. Patients with LCa have a cure rate of only 16%, even in the most advanced Western health systems. The prognosis of LCa is highly dependent on disease stage, being never-smoking status and female sex favorable prognostic factors.[9].

Nevertheless, the main criteria are performance status and disease extension at diagnosis (TNM stage), with the advanced stages displaying the worst prognosis[7].

1.1.6 Treatment

LCa treatment depends on histopathological diagnosis, disease stage and patient's general condition. There are several ways to treat LCa, including surgery, chemotherapy, radiation and targeted therapy (Figure 4)[10].

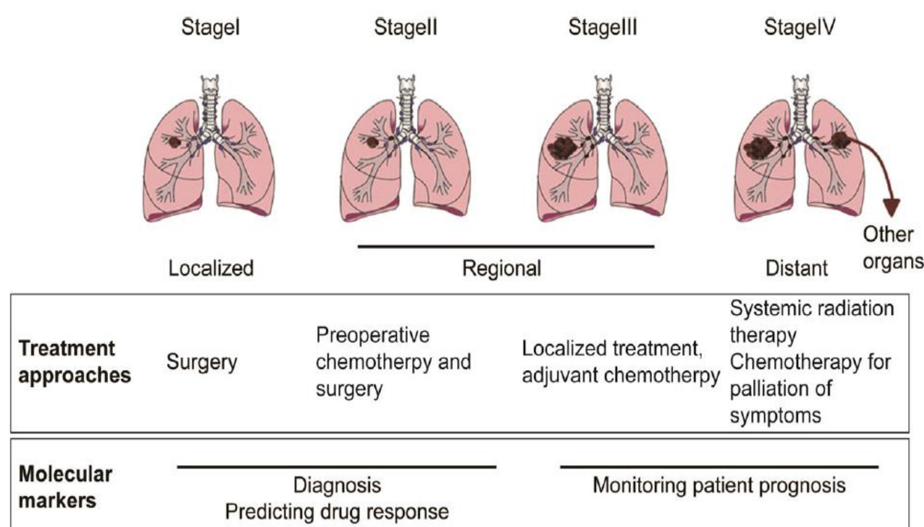

Figure 4 - Molecular diagnosis and treatment of lung cancer at different stages [10].

1.1.6.1 Surgery

Tumor resection by surgery remains the best and the most successful treatment approach for patients with early stage disease (stage I and II and selected patients with stage IIIA), whose LCa are limited to the hemithorax and can be totally encompassed by excision. Tumors can be removed by anatomic segmentectomy, pneumonectomy or lobectomy. Lobectomy is currently the standard care that will result in complete resection of the tumor mass. However the great majority of patients present at diagnosis time inoperable tumors[12].

1.1.6.2 Radiotherapy (RT)

Radiotherapy (RT) is performed in patients with resectable tumors that are medically unfit or refuse to undergo surgical resection. In these cases RT is used to control primary tumor growth and regional lymphatic dissemination. Therapeutic doses of radiation must be delivered to the target site, minimizing incidental irradiation of surrounding normal tissues. This process typically requires a planning CT scan with the

patient in treatment position. The radiation oncologist defines the target and surrounding normal tissues on the CT images using special treatment planning software. RT is also used as adjuvant therapy for patients with incomplete resection or node-positive disease and as palliative therapy, controlling symptoms and improving life quality [12].

1.1.6.3 Chemotherapy (ChT)

Chemotherapy has become the standard care for treating SCLC and unselected advanced NSCLC, and has also been advocated as an integral part of combined modality approaches to disease earlier stages. Initially it was used in patients with advanced metastatic disease as a palliative measure. Currently is used with curative intent alone or combined with others therapies. It was demonstrated that induction chemotherapy followed by RT prolongs the overall survival of patients with unresectable stage III disease compared with patients receiving RT alone. Therefore, chemotherapy has an emerging role in stage IIIA (N2) disease. The use of induction chemotherapy in patients (stage IIIA) alone or in conjunction with RT, results in a 5-year survival of 20 to 30 % compared with 5 to 10 % with surgery alone.[9]

1.1.6.4 Targeted Therapy

The identification of new potential biomarkers led to a novel strategy, named targeted therapy. In the last years has been improved mainly due to the information from molecular studies that identify specific alterations in groups of LCa patients. Contrarily to other LCa treatments, which act directly against cancer cells or tumor, immunotherapy, is a more sophisticated method that stimulates the patient's immune system to target cancer cells. This therapy in the majority of the cases presents less severe side effects. Several agents that target various molecular pathways are being studied (Figure 5)[16, 17]. Nowadays, the agents used for LCa treatment include: inhibitors and antibodies of epidermal growth factor receptor (EGFR); inhibitor of EML4-ALK inhibitors[17].

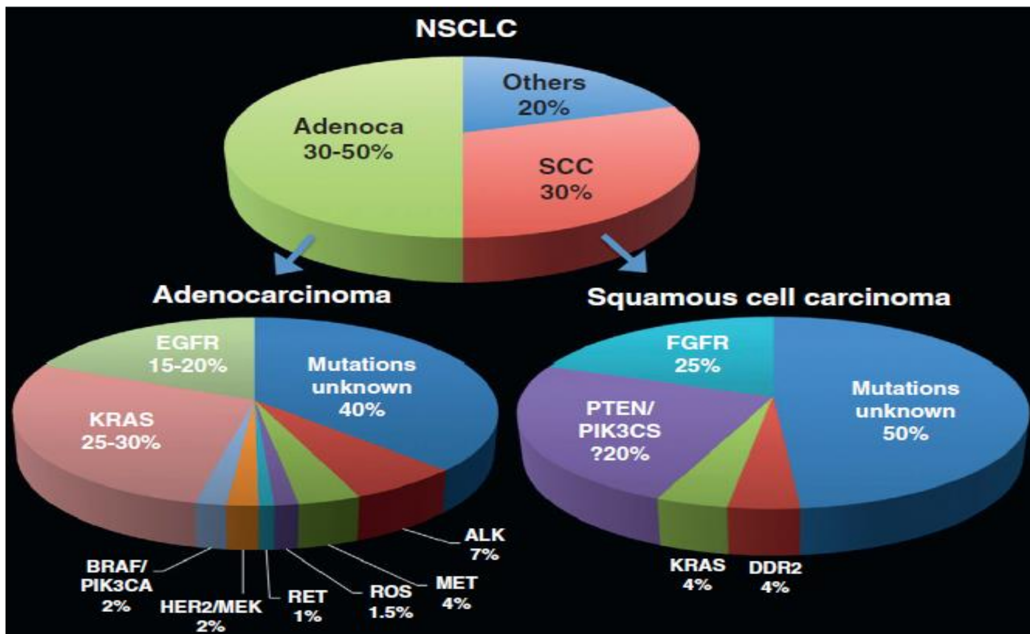

Figure 5 - Proportion of known driver mutations in Non-small cell lung cancer(NSCLC)[16].

- ***EGFR***

Mutations in tyrosine kinases receptors, such as epidermal growth factor receptor (*EGFR*) are well known cancer predictive biomarker. When EGFR is constitutively activated by mutations there are several inhibitors that can be used namely: gefitinib, erlotinib, lapatinib and cetuximab. *EGFR*-targeted inhibitors include monoclonal antibodies, that target *EGFR* extracellular domain, and tyrosine kinase inhibitors (TKIs), which are small molecules that inhibit intracellular tyrosine kinase activity of *EGFR*. The somatic mutations at the kinase domain of *EGFR* strongly correlates with sensitivity to *EGFR* inhibitors, being observed in roughly 10-20% of cases of lung adenocarcinomas, from patients of European descent and in roughly 50% of cases from patients of East Asian descent. These proportions can be explained to local smoking rates (areas with high smoking rates have lower rates of EGFR- mutated cancers). These mutations preferentially affect patients with adenocarcinoma subtype who never smoked, females and East Asian ethnicity. *EGFR* mutation is not only a predictive biomarker to EGFR tyrosine kinase inhibitors but also a prognostic factor. Therefore, the presence and the type of *EGFR* mutations is indicative which of patients will respond to therapy with EGFR inhibitors (Figure 6)[7, 8, 12, 16, 17].

- **EML4-ALK**

The inversion of two closely located genes on chromosome 2p, fusion of PTK echinoderm microtubule-associated protein like-4 (EML4) with anaplastic lymphoma kinase (ALK) yields the EML4-ALK fusion protein. The EML4 - ALK fused oncogene is

present in up to 3-7% of NSCLC and promotes malignant growth and proliferation. Similarly to EGFR alterations, *ALK* rearrangements are more likely to be seen in specific populations. Thus, young patients with adenocarcinoma subtype (mostly associated with an acinar pattern) who are light or never-smokers, males and frequent signet ring cells seen on histology are the main subset of patients with ALK alterations and benefit from treatment with the *ALK* inhibitor crizotinib. Clinical testing guidelines for *ALK* fusion detection in lung adenocarcinoma is already standard care. Moreover, immunohistochemistry is also sensible and specific tool for *ALK* rearrangements detection (Figure 6)[7, 8, 12, 16, 17].

There are other potential biomarkers with therapeutic value, but without targeted therapies, yet (e.g. *KRAS*)[16].

- ***KRAS***

KRAS mutations are the most common oncogenic driver alteration at the tyrosine kinase receptor pathway of lung adenocarcinomas in Caucasian populations. In fact a mutation rate of roughly 30% has been described in these population compared to only 10% in East Asian population. This mutation is associated with tobacco smoking which might explain this high percentage. KRAS mutation has been associated with poor prognosis, and importantly predicts chemotherapy and EGFR TKIs resistance. Although KRAS was one of the first described oncogenic drivers in NSCLC, effective targeting of this alteration remains a therapeutic challenge and no effective treatments for KRAS-mutant lung adenocarcinomas have been discovered so far. Direct RAS inhibition with salirasib was been proved unsuccessful; hence novel approaches are currently tested to inhibit downstream molecules in the RAS/RAF/MEK/ERK and PI3K/AKT/mTOR pathways (Figure 6)[7, 8, 12, 16, 17].

Targeted therapy for SCC is now a major focus of research. Recent discoveries from the cancer genome atlas (TCGA) about the molecular pathology of SCC have identified several important signaling pathways (Figure 5). Although these pathways can be inhibited, clinically meaningful benefits were not achieved yet. Ongoing work should hopefully see the identification of targeted agents for SCC in the near future[16].

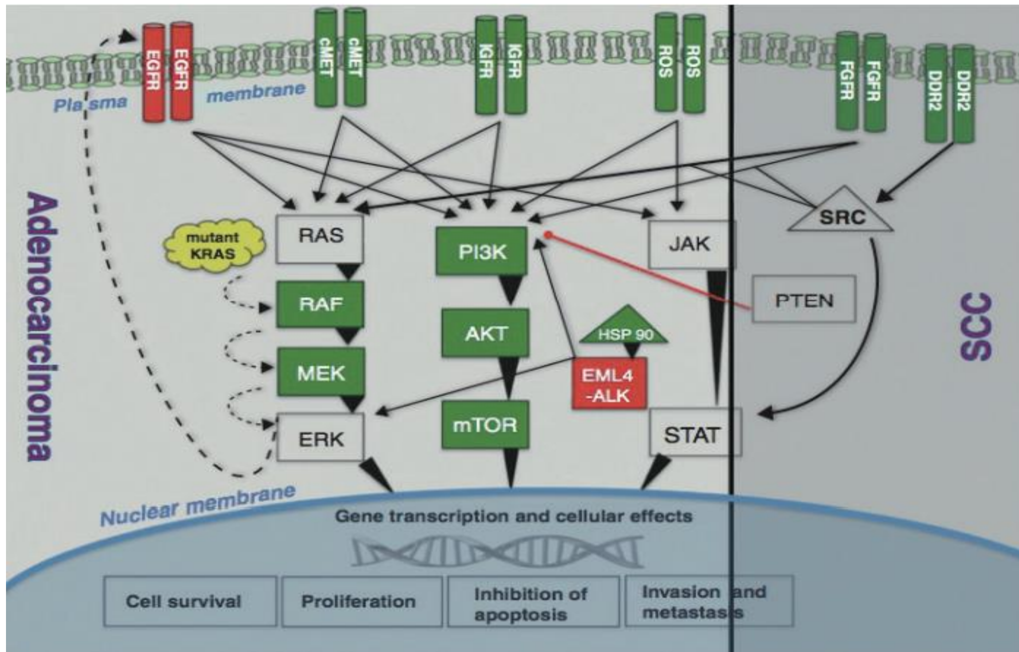

***Figure 6* -** Overview of molecular pathways and potential targets in Non-small cell lung cancer (NSCLC)[16]**.**

1.2 Epigenetics

The term epigenetics derived from the Greek prefix *epi-* meaning "what stays beyond" *–genetics.* The original definition by Conrad Waddington (1941), epigenetics referred to all molecular pathways modulating the expression of a genotype into a particular phenotype. However, with the rapid growth of genetics, the meaning of the word has gradually narrowed. Epigenetics today is generally defined as ''the study of heritable changes in gene function and that do not alter the primary DNA sequence''.[18]

1.2.1 Epigenetic Mechanisms

Epigenetic mechanisms can be grouped into at least four major types of modifications: DNA methylation, non-coding RNAs, histone post-translational modifications and histone variants (Figure 7).[19] These mechanisms are essential for normal development and maintenance of tissue-specific gene expression patterns in mammals. Abnormal epigenetic modifications were shown to contribute to common human diseases, including cancer [20]. Indeed, deregulation of epigenetic mechanisms are present in all types of tumors, contributing to its development and progression.[21] The reversible nature of epigenetic aberrations has led to the emergence of the promising field of epigenetic therapy.[20]

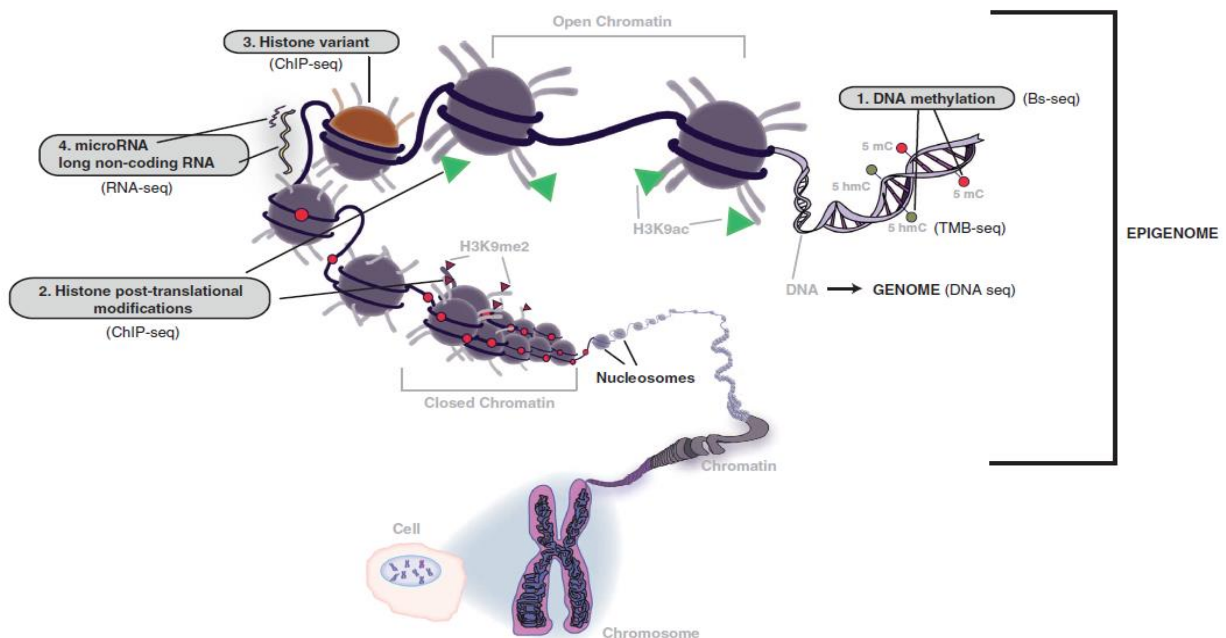

Figure 7 - Four distinct mechanisms of epigenetic regulation [19].

1.2.1.1 Non-coding RNAs

Non-coding RNAs (NcRNAs) are a class of RNA sequences that do not encode for proteins but are transcribed and biologically active. They are involved in a wide range of cellular functions, as chromosome dynamic control, splicing, RNA editing, translation inhibition and mRNA degradation. NcRNAs are composed by transcribed ultraconserved regions, small nucleolar RNAs, Piwi- interacting RNAs, large intergenic NcRNAs, long NcRNAs and microRNAs (miRNAs)[22]. MiRNAs are without doubt the best studied class. They are small non-coding RNA molecules that can negatively regulate the expression of up to hundreds of messenger RNA (mRNA) targets[23]. In normal cells, microRNAs are responsible for the fine-tuning of homeostatic gene expression and help to confer robustness to cellular processes, which is required for inducing and keeping cell fate decisions. MicroRNAs have also been implicated in the oncogenic transformation and their expression is altered at early stages of lung cancer[10].

1.2.1.2 Histone Post-translational Modifications and Variants

In eukaryotic cells, chromatin is composed by DNA and histones, and it is in this context that transcription takes place. Histones are dynamic regulators of gene activity that undergo a wide variety of post-translational modifications influencing chromatin structure and recruitment of proteins complexes to DNA. Eight histones, one pair of each H2A, H2B, H3 and H4 constitute the basic unit of chromatin, the nucleosome. Histone H1 binds to the DNA between the nucleosomes. Post-translational modifications of histones are an epigenetic mechanism for the establishment and maintenance of gene activity, and consequently, regulate a wide range of cellular processes. The best characterized post-translational modifications are methylation, acetylation and phosphorylation[24, 25]

1.2.1.3 DNA methylation

DNA methylation is the best studied epigenetic modification being the major alteration that takes place during aging, embryogenesis and carcinogenesis[26].

The DNA methylation consists in the addition of methyl group (CH_3) to the 5'carbon of a cytosine nucleotide preceding a guanine, originating 5-methylcytosine (5mC). This enzymatic addition is a normal process within cells[25, 27]. DNA methylation is catalyzed by a series of sophisticated enzymes called DNA methyltransferases (DNMTs) that use *S*-adenosylmethionine (SAM) as a methyl donor group (Figure 8). Methylation in mammals primarily occurs in CpGs dinucleotides and only occasionally in non-CpG sites [17, 28, 29].

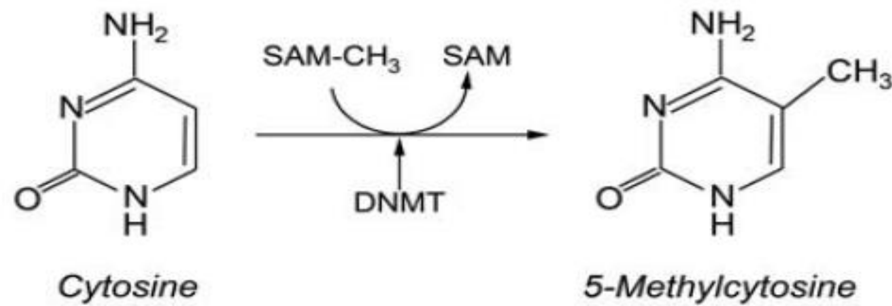

Figure 8 - Conversion of cytosine to 5-methylcytosine by DNA methyltransferase (DNMT). The methyl group (CH_3) is transferred from S-adenosylmethionine (SAM) to 5-carbon position of cytosine by DNMT[25].

There are four known biologically active DNMTs in mammals: DNMT1, DNMT2, DNMT3a and DNMT3b.[26] DNMIT1 is responsible for maintaining DNA methylation and copies pre-existing methylation pattern onto the newly synthetized strand immediately after DNA replication. DNMT1' function is to ensure that the methylation pattern of the parental cells is identically reproduced in each daughter cell. There is considerable evidence indicating an upregulation of DNMT1 in cancer. DNMT2 just appears to be involved in methylation of RNA and has shown only weak DNA methylation ability *in vitro.* DNMT3a and DNMT3b are the enzymes responsible for *de novo* methylation at CpG sites during embryogenesis targeting unmethylated CpG dinucleotides. Even though DNMT1 appear to be responsible for most of DNA-methylating capacity in cancer cells, specially at promoter regions, recent studies suggest an interaction between DNMT1 and DNMT3b to ensure propagation of methylation patterns during DNA replication in cancer cells[21, 24-26]

In normal mammalian cells, CpG islands are proximal to gene promoter regions (Figure 9). These regions are largely protected from DNA methylation and reside in restricted regions of open chromatin, or euchromatic states, which are favorable to gene transcription. In contrast, for most regions of the genome, such as gene bodies, repeat elements and pericentromeric regions of the genome, cytosines in CpG dinucleotides are methylated (Figure 8). This pattern of DNA methylation is common to the bulk of the human genome, which is packed as closed unfavorable for transcription[29].

Global DNA hypomethylation occurs in cancer cells, which results in chromosomal instability and activation of proto-oncogenes. Concomitantly, abnormal methylation of gene promoter regions (hypermethylation) leads to tumor suppressor silencing (Figure 9). CpG islands, the major targets of DNA methyltransferases, are associated with the transcription start sites of almost half of human genes[8]. CG dinucleotides occur at a high frequency in tumor suppressor genes (TSG) promoters, and these CpG islands are usually unmethylated or hypomethylated in normal cells, allowing the initiation of transcription. However, during malignant transformation, CpG islands became methylated

or hypermethylated, leading to repression of TSG transcription and potentiating oncogenesis[30]. Despite CpG islands cover approximately 1% of the total human genome, they are present in >50% of human gene promoters which indicates their functional importance in transcriptional control[24]. Approximately 75% of all CpG dinucleotides in normal cells are methylated in the human genome[21].

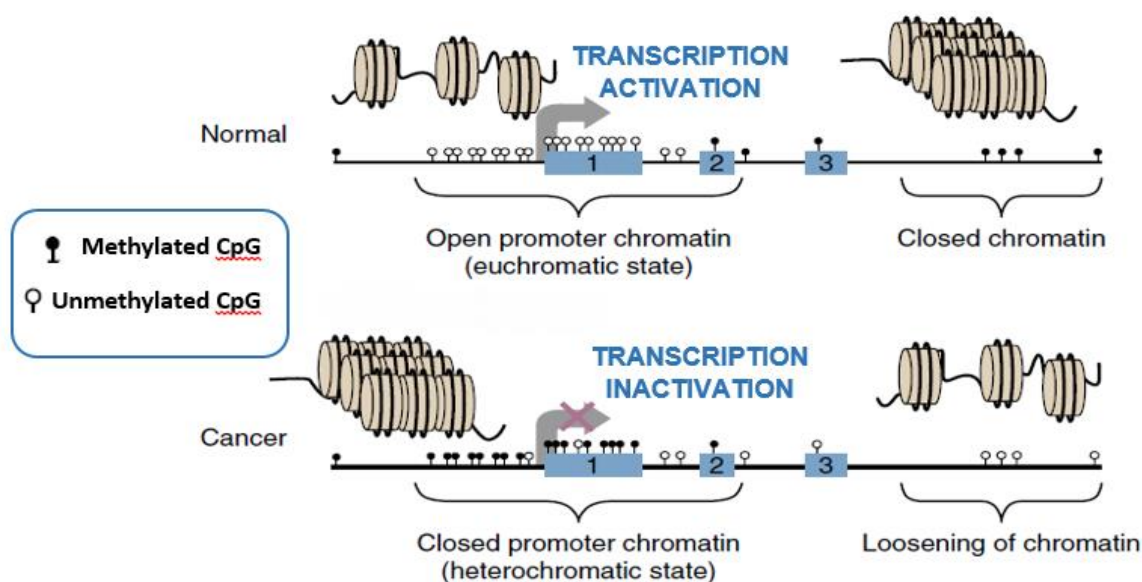

Figure 9 - DNA methylation in normal and cancer cells. In normal cells, promoter Cpg islands are unmethylated while in cancer cells, they have acquired aberrant DNA methylation, and consequently, transcriptional silencing. *Adapted from Baylin, 2015* [29].

The common occurrence of DNA hypermethylation in all types of cancer makes it an ideal biomarker, one that has been extensively investigated. DNA methylation is an inherently ideal substrate for cancer biomarker development for several key reasons. An advantage of DNA methylation over protein-based markers is that it is readily amplifiable and easily detectable using PCR-based approaches. Furthermore and, contrarily to cancer-specific DNA mutations, cancer-specific DNA hypermethylation occurs in defined regions, usually in or near the promoter of genes[27]. Moreover, the prevalence of CpG methylation changes at literally hundreds of genes in a given tumor affords a vast number of possible tumor-specific targets for assay development. Finally, its association with gene silencing allows DNA methylation to serve as a substitute marker for gene expression, effectively providing a positive reading for negative expression[21].

1.2.2 DNA methylation in Lung Cancer

Lung cancer develops through a multistage process involving permanent genetic alterations, dynamic epigenetic changes and environmental factors (Figure 10)[10].

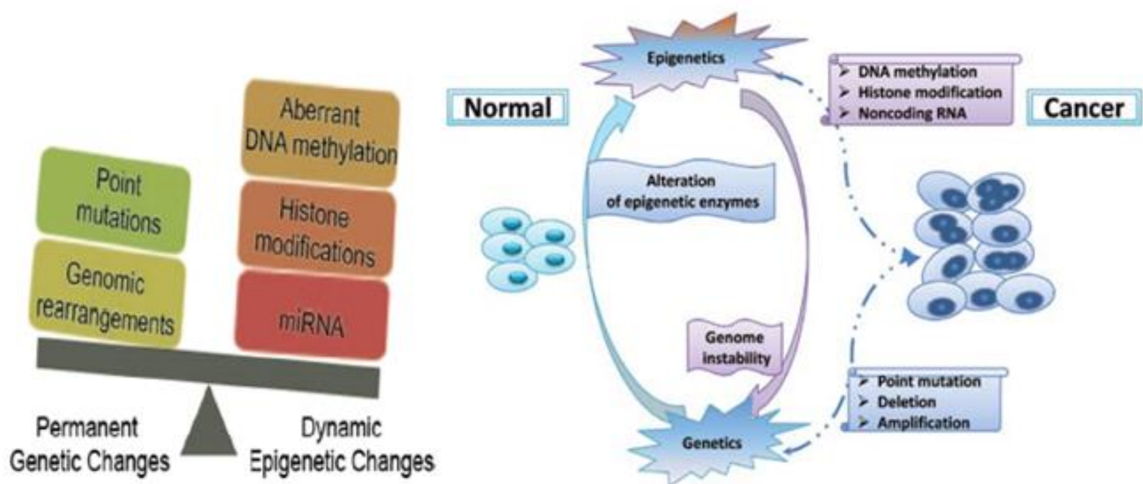

Figure 10 - Cancer is a result of the interaction between permanent genetic mutations and dynamic epigenetic alterations. During cancer formation, a large number of epigenetic and genetic alterations lead to abnormal gene expression which evoke genome instability. *Adapted from Mehta, 2015; Chen, 2014* [10, 31].

DNA methylation plays a critical role in repressing gene activity of several TSG and maintaining genome stability. It has been demonstrated that two major changes in methylation status occur during carcinogenesis: regional promoter hypermethylation and genome wide hypomethylation. These methylation changes are critically associated with transcriptional silencing of the involved genes[26].

It was suggested that LCa harbors a CpG island methylator phenotype, in other words a tumor phenotype characterized by widespread hypermethylation. This is not totally surprising, given the well-known upregulation of DNMTs in NSCLC[23].

DNA methylation plays an important role in the etiology of LCa, therefore it might have a potential value as diagnostic and prognostic biomarkers. Expanding our understanding of how epigenetic events contribute to the genesis of LCa and how they can be translated into clinical relevant biomarkers and therapeutic targets will enhance our capacity to manage LCa patients and consequently reduce the heavy global burden of this critical disease[23]. Research on epigenetic has provided new insights of early cancer development and progression, allowing increased knowledge of early stages of the disease and therapeutic interventions[24].

1.2.2.1 Hypermethylated genes in Lung Cancer

Currently, there are many genes described as hypermethylated in LCa. Some of the most studied in this context and more informative include: *APC*, *HOXA9*, *RARβ2*, *RASSF1A*, *SHOX2* and *TFPI2* (Table 1) [32].

Table 1 - Panel of genes hypermethylated in lung cancer.

Gene	Locus	Gene function	References
APC	5q22.2	Adenomatous polyposis coli: TSG that acts as a negative regulator of Wnt and also is involved in cell migration and adhesion, transcriptional activation, and apoptosis	[33]
HOXA9	7p15-7p14.2	*HOX* genes encode transcription factors that play essential roles in regulation of embryonic morphogenesis in animals	[34]
RARB2	9p24.2	Retinoic acid receptor beta is involved in cell growth and differentiation.	[5]
RASSF1A	3p21.31	Ras association (RalGDS/AF-6) domain family member 1 is a putative TSG involved in apoptosis and cell cycle control	[35]
SHOX2	3q25.32	Homeobox family gene is involved in gene transcription with putative involvement in cell growth and differentiation	[36, 37]
TFPI2	7q22	*TFPI2* decreases activation of metalloproteinases	[38, 39]

1.2.2.1.1 APC

The *adenomatous polyposis coli* (*APC*) gene encodes for a cytoplasmic protein involved in cell signaling through Wnt pathway which plays an important role in cell-cycle regulation and apoptosis (Table 1)[40, 41]. The APC binds to β-catenin, axin and glycogen synthase kinase 3β to form a large protein complex, in which β-catenin is phosphorylated and broken down, resulting in negative regulation of the Wnt signaling pathway[40]. An impaired function of *APC* is often attributable to mutations within the coding sequence of the gene. This in turn leads to lack of degradation and nuclear accumulation of β-catenin which acts as a transcriptional activator, causing loss of cell growth control[41]. In addition, *APC* is involved in cell motility through its association with microtubules and it also stimulate guanine nucleotide exchange factor[40]. *APC* is considered a tumor suppressor gene and high *APC* promoter methylation is significantly associated with a decrease in survival at LCa. Therefore *APC* is promise a biomarker of biologically aggressive NSCLC[42].

1.2.2.1.2 HOXA9

HOX genes encode transcription factors that are critical in the regulation of embryonic morphogenesis in animals (Table 1)[43]. HOX proteins are essential switches of development stage-specific and cell-specific gene regulation. Thus, HOX proteins are key determinants of cell identity and potential targets during tumorigenesis[44]. Most of the *HOXA* promoters contain highly dense CpG islands, and its methylation is integral to the control of *HOXA9* gene expression[34]. In LCa, *HOXA9* displays higher methylation levels in tumor tissues than normal tissues. Therefore, detection of aberrant *HOXA9* gene hypermethylation might be useful as biomarker for the early diagnosis of primary LCa[45].

1.2.2.1.3 RARβ2

Retinoic acid is known to interact with nuclear retinoic acid receptors and retinoic X receptors. Both receptors have three subtypes (alpha, beta and gamma) which have distinct functions[46]. Receptors of the RAR family are differentially expressed during development and in adults life. There is strong evidence that *RARβ* plays a central role in epithelial cells growth regulation and in tumorigenesis[47]. The *RARβ2* gene has two different promoters and transcripts which are produced by alternative splicing. Most human cells express *RARβ2* as predominant form. This isoform plays a central role in mediation of growth inhibition of different types of cancer cells and is responsible for coding vitamin A nuclear receptor which is required for normal cell growth and differentiation (Table 1)[48-50]. *RARβ2* expression is not only lost or reduced in a large percentage of LCa patients but also in a people with high risk to development LCa. Approximately 40% of NSCLC present loss or reduced *RARβ2*[46, 48].It was also described that *RARβ2* hypermethylation might be associated with short recurrence-free survival in never-smokers adenocarcinoma's patients [46].

1.2.2.1.4 RASSF1A

Ras association domain family 1 isoform A (*RASSF1A*) is a tumor suppressor gene whose inactivation is implicated in the development of many human cancers (Table 1). It is termed *RASSF1A* because the protein contains a putative Ras association domain[35]. The RASSF1A protein, encoded by one of the 8 splicing isoforms, termed 1A to 1H, is expressed in all normal human tissues, and carries several domains mediating protein-protein interactions with multiple partners[51]. *RASSF1A* modulates a broad range of essential cellular functions for normal growth control, such cell motility, invasion, cell cycle and apoptosis, regulation of microtubules and maintenance of genomic stability[35]. Besides this, it was also suggested that *RASSF1A* plays a role in tumor cell adhesion and motility[51]. This gene appears to suffer frequent transcriptional inactivation in tumor cells

due to aberrant promoter methylation[35]. It has been reported that *RASSF1A* gene is frequently inactivated in primary LCa by the *de novo* methylation of CpG islands in the promoter region[52]. Due to this, *RASSF1A* represents an important potential diagnostic target[35].

1.2.2.1.5 SHOX2

The human *Short Stature Homeobox 2* (*SHOX2*) has been identified as highly homologous to the short stature homeobox gene *SHOX* (Table 1). Homeobox genes code for proteins harboring specific DNA-binding homeodomains (homeoproteins), which play fundamental roles in vertebrate development and differentiation by acting as transcriptional regulators. *SHOX2* is a known regulator of chondrocyte hypertrophy and act in skeleton development and embryogenic pattern formation[36, 37]. Genomic gain of chromosome 3q involving the *SHOX2* gene has been recognized as one of the most prevalent and significant chromosomal rearrangements in LCa[36]. *SHOX2* hypermethylation was shown to be a useful biomarker for detecting SCC and SCLC with high specificity and sensitiviy[36, 53]. An *in vitro* diagnostic test for *SHOX2* methylation has recently become commercially available in Europe, and it was demonstrated that it helped pathologists in the diagnosis of LCa with sensitivity of 68% and 95% of specificity[54].

1.2.2.1.6 TFPI2

The human *Tissue Factor Pathway Inhibitor 2* (*TFPI2*) is a potential inhibitor of the plasmin within the extracellular matrix. Degradation of this protein was strongly associated with the progression of LCa (Table 1)[38]. *TFPI2* is synthesized and secreted by endothelial, mesenchymal and epithelial cells, monocytes/macrophages and the syncytiotrophoblast[39]. *TFPI2* decreases activation of metalloproteinases (MMP1, MMP3, MMP9 and MMP13) which inhibit plasmin and trypsin leading to a reduction of tumor invasion and metastasis[39, 55]. Downregulation of *TFPI2* promote migration and invasion of LCa lines. Thus *TFPI2* is considered a TSG in LCa and aberrant *TFPI2* promoter hypermethylation may be a valuable prognostic marker[38].

2. AIMS

Lung cancer, a complex disease involving both genetic and epigenetic changes, is the leading cause of cancer related deaths worldwide for both genders. The high mortality rate of this disease is mainly due to the high incidence coupled with its dismal 5-year survival rate of only 10%, despite the development of novel clinical diagnosis techniques and chemotherapy. The prognosis and treatment of LCa varies depending on subtype. Therefore, improved tools for early detection and discrimination of the different subtypes of LCa are urgently needed, in order to have a faster, efficient and targeted treatment, increasing in turn the survival rate of this disease.

Several epigenetic alterations are involved in LCa development and progression. Since DNA methylation markers are stable and amenable to be easily assessed by PCR based measurement and due to their early onset, into LCa they might have potential in diagnosis and prognosis of this malignancy. Thus, the main objective of this master thesis was to evaluate the methylation profile of the different LCa subtypes, with a panel of previously described genes for LCa - *APC*, *HOXA9*, *RARβ2*, *RASSF1A*, *TFPI2* and *SHOX2*. Particularly we aimed to discriminate adenocarcinoma, the most prevalent subtype, from the other major LCa subtypes.

Thus, the specific aims of this master dissertation were:

- Validate the previously identified DNA-methylation based markers (*APC*, *HOXA9*, *RARβ2*, *RASSF1A*, *TFPI2* and *SHOX2*) in lung cancer tissue samples.
- Identify a methylation profile of lung cancer major subtypes;
- Evaluate the association between the gene-panel methylation levels and standard clinicopathological parameters
- Determine the prognostic value of gene-panel methylation levels.

3. MATERIALS AND METHODS

3.1 Study cohort – Patients and Samples

For this study, 152 LCa samples, including 63 squamous cell carcinomas (SCC), 58 adenocarcinomas (Ade), six large cell carcinomas (LCC) and 25 small cell carcinomas (SCLC), were obtained from the archives of the Department of Pathology. All the tumor tissue samples were obtained from tumorectomy specimens of patients diagnosed and treated at the Portuguese Oncology Institute of Porto with no previous history of lung cancer. Tissues were routinely fixed and paraffin-embedded for standard pathologic examination, allowing for tumor classification and World Health Organization[7]/AJCC Cancer Staging Manual[14] grading and staging. Additionally, an independent set of 22 paraffin-embedded normal pulmonary parenchyma collected from individuals with other neoplasias was used as controls. Relevant clinical data were collected from clinical charts. This study was approved by institutional ethics review board (CES-IPOPFG-EPE 120/015).

3.2 DNA Extraction From Formalin-fixed Paraffin-embedded Tissues (FFPE)

For each case, the slides, that were previously stained with Hematoxylin & Eosin were delimited the area with tumor cells. Tumor areas were macrodissected from the eight-micrometer thick tissue sections to maximize the proportion of malignant cells (>70%), and subsequently deparaffinized and rehydrated using xylene and 100% ethanol. Then samples were digested with proteinase K (20 mg/ml, 60µl). DNA was extracted using standard phenol-chloroform, ethanol, ammonium chloride and glycogen protocol. After elution, DNA concentration was measured by NanoDrop Lite Spectrophotometer (NanoDrop Technologies, USA) and stored at -20°C (Appendix I).

3.3 Bisulfite treatment of DNA, quantitative methylation-specific polymerase chain reaction (qMSP)

Bisulfite treatment of DNA samples was performed using EZ DNA methylation – Gold kit (Zymo Research, Orange, CA), and the converted DNA was eluted in 36-60µl of distilled water and stored at -80ºC (Appendix II). This method allows for the assessment of the methylation status of individual CpG islands in genomic DNA. The major advantage of sodium bisulfite-based assays is that they require very small amounts of DNA and consequently, are compatible with DNA obtained from macrodissected paraffin-embedded tissue samples[56].

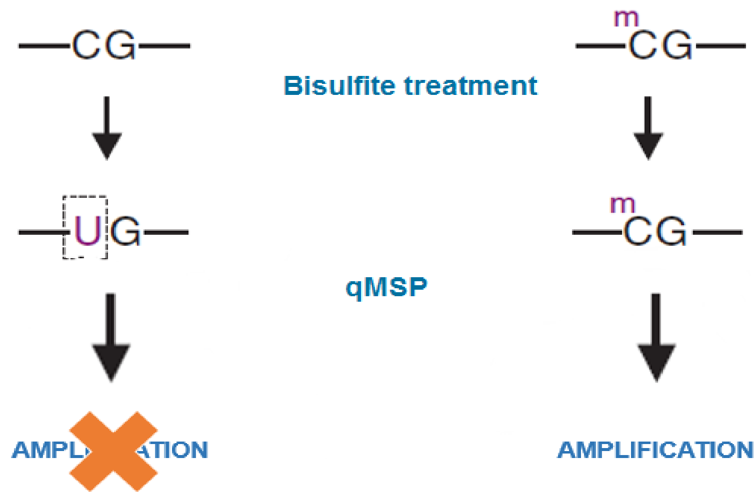

Figure 11 – DNA modification by sodium bisulfite. Following bisulfite conversion, methylated cytosines remain unchanged, while unmethylated cytosines are deaminated to uracils. The level of DNA promoter hypermethylation its quantified by quantitative methylation-specific PCR (qMSP).

Quantitative methylation-specific PCR (qMSP) was performed using the bisulfite-modified DNA as the template. Reactions were carried out in 384-well plates using LightCycler 480 (Roche, Germany). Briefly, per each well 2 µL of modified DNA and 5 µL of KAPA SYBR® FAST qPCR Master. The volume of primers used varied (according to Table 2) and sterile distilled water was added in order to total 10 µL of reaction volume (Appendix III). The thermocycler conditions were as follows: initial denaturation at 95ºC for 3min; followed by 40 cycles of 95ºC for 3s, 30s for annealing, extension and data acquisition (temperature specified in Table 2).

The β-Actin (*ACTβ*) gene was used for normalization and control of the quantity of DNA. The relative level of methylated DNA for each gene in each sample was determined by the comparison between values obtained for each target gene and values of the internal reference gene. The ratio was then multiplied by 1000 for easier tabulation ([ML = (target gene/ACTβ) x 1000]). Experiments were performed in triplicate, with water blanks as negative controls and five serial dilutions (dilution factor of 5) of a fully methylated bisulfite modified universal DNA control (*in vitro* methylated human DNA, Chemicon). PCR primers are described in Table 2.

Table 2 - Primers sequences used and qMSP conditions for each of the tested genes.

Gene	Forward (F)	Reverse (R)	Annealing T⁰C	Concentration per reaction (F + R)
β-Actin	TGGTGATGG AGGAGGTTT AGTAAGT	AACCAATAAA ACCTACTCC TCCCTTAA	60	400 nM
APC	TGTGTTTTAT TGCGGAGTG C	CACATATCG ATCACGTAC GC	62	300 nM
HOXA9	TATTTAGTCG GTATTCGC	ACCTCGAAC GCTTCCAT	60	300 nM
RARβ2	TCGAGAACG CGAGCGATT	GACCAATCC AACCGAAAC	60	300 nM
RASSF1A	GGGTTTTGC GAGAGCGCG	GCTAACAAA CGCGAACCG	60	300 nM
SHOX2	ATTCGTATTT GGTCGCGTA C	CTACTACGA CCGCCACTA CC	62	300 nM
TFPI2	GGCGGGGT GATAGTTTTC	TACTCCAAA CGACCCGAA T	62	300 nM

3.4 Statistical analysis

Differences in methylation levels of tested genes among the histological subtypes were assessed by the non-parametric Kruskall-Wallis test, followed by pairwise comparisons through Mann-Whitney *U* Test.

The Kaplan-Meier method was employed to estimate the probability of survival as a function of time and survival differences were analyzed by the log-rank test.

A Cox-regression model comprising clinicopathological variables was computed to assess the relative contribution of each variable to the follow-up status.

Two-tailed P-values were derived from statistical tests, using a computer assisted program (SPSS Version 22.0, Chicago, IL), and results were considered statistically significant at $P < 0.05$, with Bonferroni´s correction for multiple tests, when applicable.

4. RESULTS

4.1 Clinical Samples

A total of 250 patients were initially enrolled in this study. However, because most lung cancers are not surgically treated, especially LCC, tissue availability is restricted to small biopsies from which good quality DNA is difficult to obtain, due to these limitations, only 152 tumors were selected for testing. These patients were consecutively diagnosed from 2001 to 2015,. Of these patients, 113 (74.3%) were male and 39 (25.6%) were females, with a median age of 64 years (range, 45 – 83) (Table 3).

Table 3 - Clinical and Histopathological characteristics of patients with Lung cancer and Normal pulmonary parenchyma.

CLINICOPATHOLOGIC CHARACTERISTICS	LUNG CANCER PATIENTS (LCa)	NORMAL LUNG (NL)
Patients, *n*	152	22
Gender, *n*		
Male	113	18
Female	39	4
Age median, years (range)	64 (45 – 83)	47 (2 – 75)
Histological Subtype, *n / (%)*		
Adenocarcinoma	58 (28 Male; 30 Female) / 39%	n.a.
Squamous Cell Carcinoma	63 (60 Male; 3 Female) / 41%	n.a.
Large Cell Carcinoma	6 (6 Male) / 4%	n.a.
Small Cell Lung Cancer	25 (19 Male; 6 Female) / 16%	n.a.
Pathological stage, *n*		
Stage I	74	n.a.
Stage II	33	n.a.
Stage III	24	n.a.
Stage IV	21	n.a.
Differentiation		
Well	22	n.a.
Moderate	74	n.a.
Poor	25	n.a.
Not available	31	22

n.a. Not applicable

Clinical characteristics of all patients enrolled in this study are summarized in Table 3. The majority of patients did not acknowledge smoking habits and, thus, we excluded this variable from analysis.

4.2 Assessment of aberrant promoter methylation levels in LCa and controls

RASSF1A methylation levels were significantly higher in LCa compared to controls (P < 0.001), whereas for *APC, RARβ2* and *HOXA9* significantly higher methylation levels were found in controls (P < 0.001; P < 0.001; P < 0.001; respectively). However ROC curve analysis was not performed because in controls no RASSF1A promoter methylation was found (Figure 12).

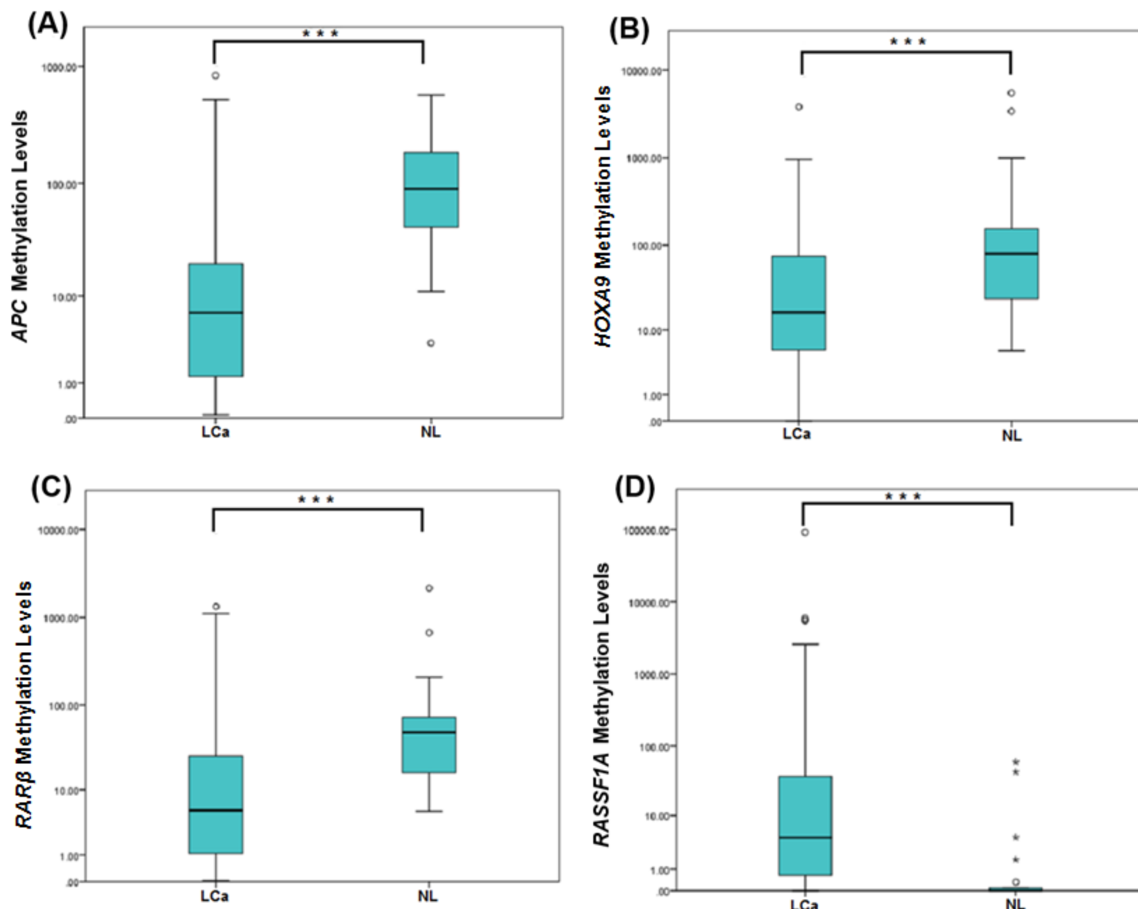

Figure 12 - Boxplots of (A) *APC*, (B) *HOXA9*, (C) *RARβ2* and (D) *RASSF1A* promoter methylation levels between Lung Cancer (LCa) and normal lung (NL) samples.

Moreover, no statistically significant differences were depicted for SHOX2 and TFPI2 promoter methylation between LCa and control samples (Appendix III).

4.3 Association between quantitative promoter methylation and clinicopathological parameters

No significant associations were found between promoter methylation levels and patients' age, gender or tumor differentiation. However, a significant association was found between methylation levels of *APC*, *RARβ2* and *RASSF1A* and advanced pathological stage [*APC* (P < 0.001), *RARβ2* (P < 0.001) and *RASSF1A* (P= < 0.001), (Figure 13)].

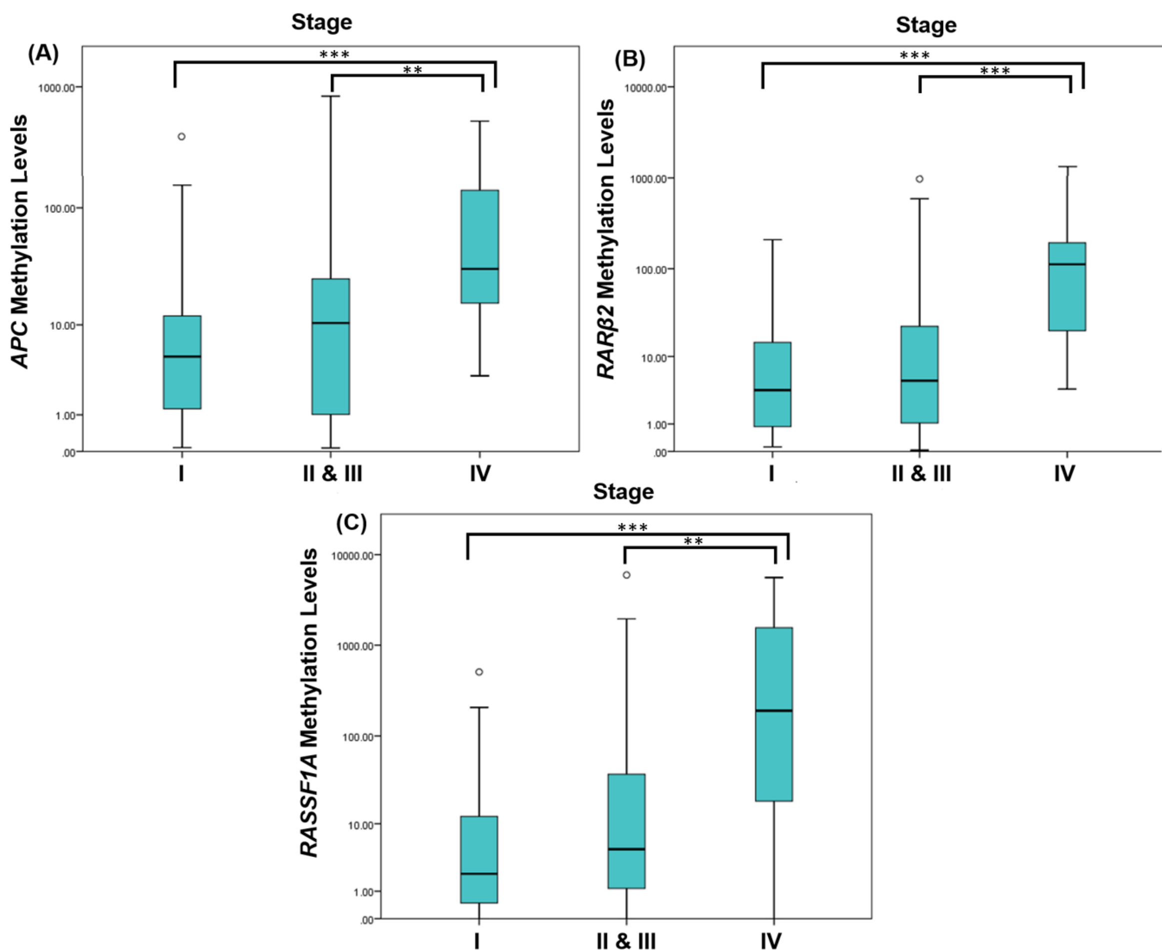

Figure 13 - Boxplots of the methylation levels of (A) *APC*, (B) *RARβ2* and (C) *RASSF1A* in the different stages (Mann Whitney Test, **P < 0.010; P*** < 0.01).

4.4 Distribution of methylation levels according to major LCa subtypes

Overall, *APC*, *HOXA9*, RARβ2 and *RASSF1A* methylation levels were significantly different between the two major LCa subtypes (NSCLC and SCLC) (***P < 0.001; *P =

0.021; ***P < 0.001; ***P < 0.001; respectively) (Appendix IV). Specifically, methylation levels of the four genes were higher in SCLC than in NSCLC (Figure 14).

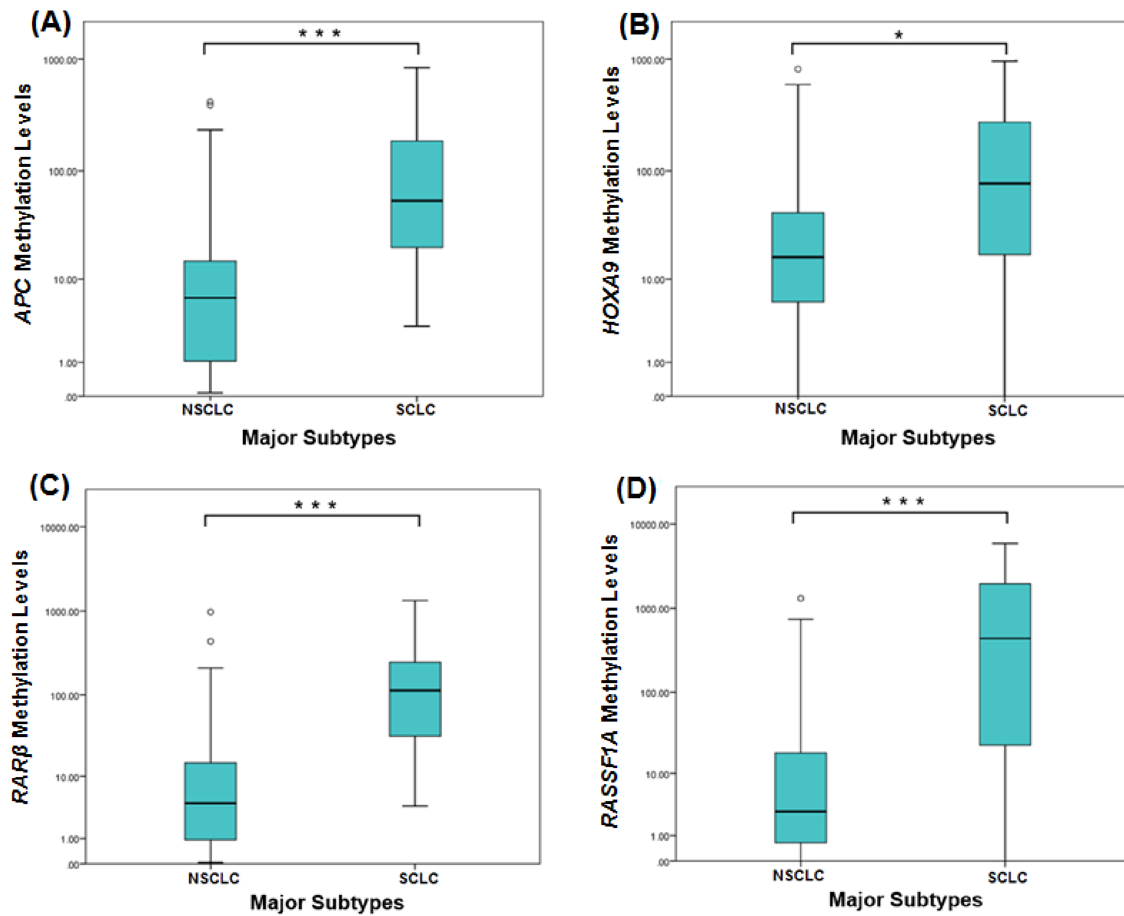

Figure 14 – Boxplots of (A) *APC*, (B*) HOXA9*, (C) *RARβ2* and (D) *RASSF1A* promoter methylation levels between the major subtypes of Lung Cancer (LCa) samples. (NSCLC: Non-small cell lung cancer) (Mann Whitney Test, *P < 0.001; P*** < 0.01)

4.5 Distribution of methylation levels according to LCa histological subtypes

Methylation levels of all genes, except for *SHOX2* and *TFPI2* differed significantly among the four LCa subtypes (p < 0.001 for all, Kruskal-Wallis test). Pair-wise comparisons are shown in Table 4 and graphically illustrated in Figure 15 (Appendix V).

Table 4 - Kruskal Wallis and Mann Whitney tests analyze of *APC*, *HOXA9*, *RARβ2*, *RASSF1A*, *SHOX2* and *TFPI2* promoter methylation levels between subtypes of Lung Cancer (LCa) samples.

	APC	***HOXA9***	***RARβ2***	***RASSF1A***	***SHOX2***	***TFPI2***
Ade vs SCLC	**0.000**	**0.006**	**0.000**	**0.000**	0.714	0.104
SCC vs SCLC	**0.000**	0.096	**0.000**	**0.000**	0.694	0.112
LCC vs SCLC	0.069	0.441	**0.009**	0.095	0.976	0.555
Ade vs SCC	0.632	**0.000**	0.697	0.259	0.409	0.993
Ade vs LCC	0.634	0.140	0.133	0.226	0.880	0.991
SCC vs LCC	0.593	0.976	0.174	0.097	0.485	0.976
P *value* [1]	**0.000**	**0.000**	**0.000**	**0.000**	0.828	0.423

[1]Kruskal Wallis test; Ade – Adenocarcinoma; SCLC – Small Cell Lung Cancer; SCC – Squamous Cell Carcinoma; LCC- Large Cell Carcinoma

Generally, SCLC showed the highest methylation levels, significantly differing from Ade for the four genes (*APC, HOXA9*, *RARβ2* and *RASSF1A)* and from SCC in three genes (*APC, RARβ2* and *RASSF1A),* whereas only differed from LCC for *RARβ2* methylation (Table 5 and Figure 15). Interestingly, Adenocarcinomas and SCC only differed for *HOXA9* methylation levels.

Table 5 - Distribution of promoter methylation levels of cancer-related genes in Lung Cancer (LCa) samples measured by qMSP.

	NSCLC Median (IQR)			**SCLC Median (IQR)**
	SCC	**Ade**	**LCC**	
APC	5.758 (0.954-16.563)	6.079 (1.379-14.482)	10.784 (0.705-80.732)	53.867 (19.824-211.923)
HOXA9	17.902 (10.476-78.753)	7.348 (2.666-22.343)	23.704 (7.570-174.509)	77.104 (11.193-300.859)
RARβ2	4.233 (0.941-13.982)	3.002 (0.818-15.012)	9.941 (4.445-37.177)	113.573 (28.928-314.484)
RASSF1A	2.230 (0.191-15.362)	3.269 (0.757-19.102)	6.240 (3.549-445.682)	438.113 (20.601-2037-204)
SHOX2	11.226 (1.942-36.750)	6.198 (1.014-43.109)	6.736 (1.581-29.858)	13.377 (0.000-675.465)
TFPI2	0.000 (0.000-2.429)	0.000 (0.000-1.163)	0.000 (0.000-538.651)	0.000 (0.000-0.000)

Ade – Adenocarcinoma; SCLC – Small Cell Lung Cancer; SCC – Squamous Cell Carcinoma; LCC- Large Cell Carcinoma; IQR- Interquartile range

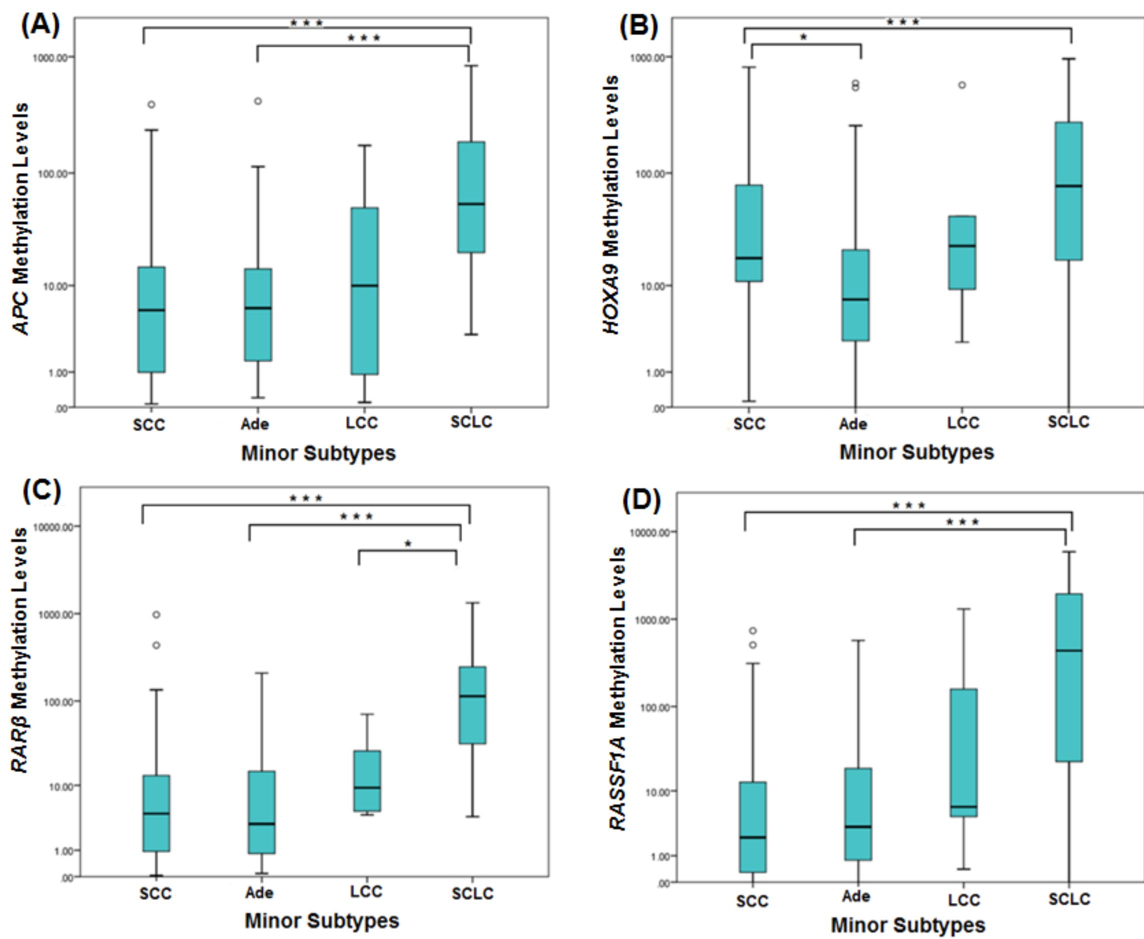

Figure 15 – Boxplots of (A) APC, (B) HOXA9, (C) RARβ2 and (D) RASSF1A promoter methylation levels between subtypes of Lung Cancer (LCa) samples. (SCC: Squamous cell carcinoma; Ade: Adenocarcinoma; LCC: Large cell carcinoma; SCLC: Small cell lung cancer) (Mann Whitney Test, *P < 0.001; P*** < 0.01)

Therefore, the analyzed genes are able to discriminate among different LCa subtypes (Figure 16).

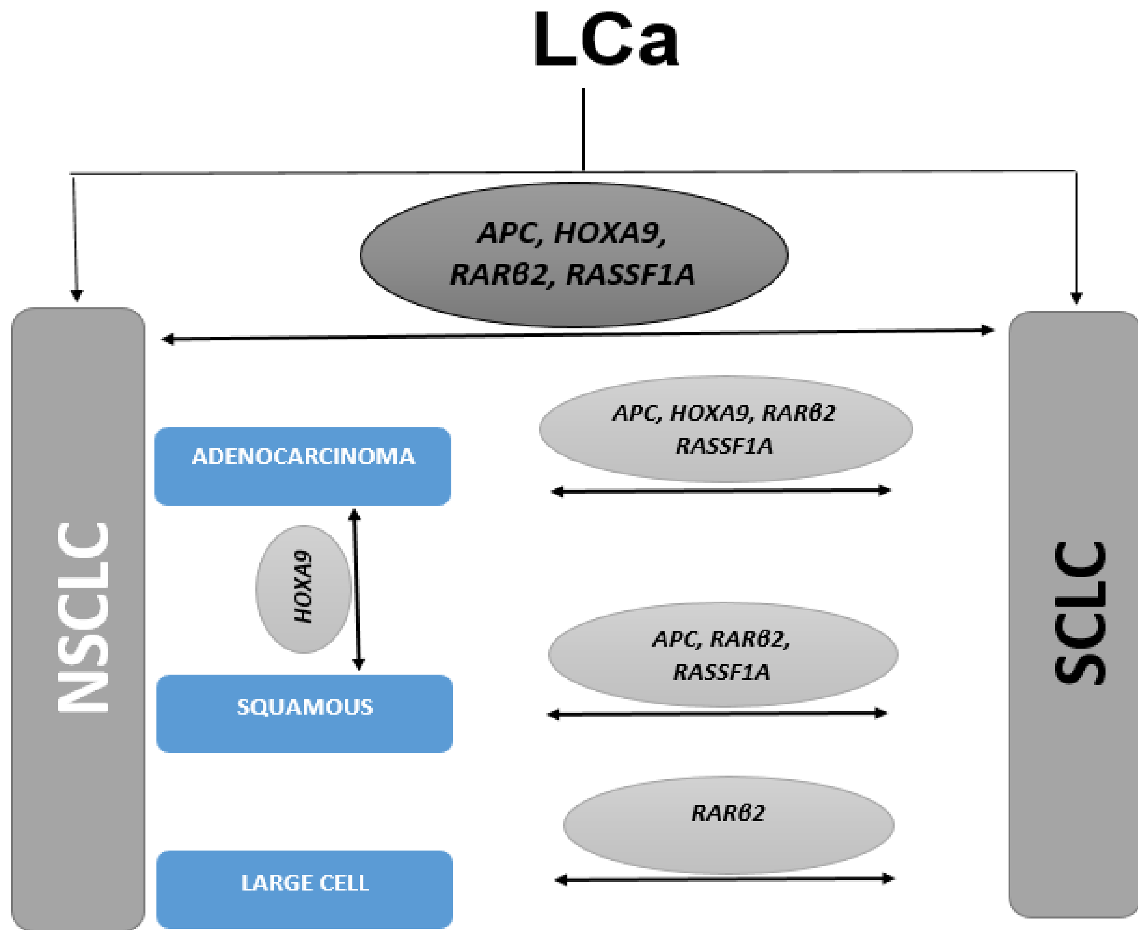

Figure 16 **–** Schematic representation of the association between genes promoter methylation and discrimination of major and minor Lung Cancer (LCa) subtypes. (NSCLC: Non-small cell lung cancer; SCLC: small cell lung cancer)

4.6 Survival analyses

The median follow-up of this LCa patient cohort was 41 months (range: 1-174 months). At the time of last follow-up, 80 patients were alive with no evidence of cancer, 8 patients were alive with cancer progression and 64 patients had deceased, 52 of which due to LCa.

4.6.1 Disease- Specific Survival

For statistical purposes, *RASSF1A, RARβ2*, *APC* and *HOXA9* methylation levels were dichotomized using the percentile 75, whereas for *SHOX2* and *TFPI2* the percentile 50 was used as threshold value. DSS analysis showed that patients with higher *RASSF1A*, *RARβ2*, *APC* and *HOXA9* promoter methylation levels had a significantly shorter survival (P < 0.001; P < 0.001; P < 0.001; P = 0.001; respectively; Figure 17; Table 6). This was not observed for *SHOX2* and *TFPI2* promoter methylation, thus we only further analyzed *RASSF1A*, *RARβ2*, *APC* and *HOXA9* methylation.

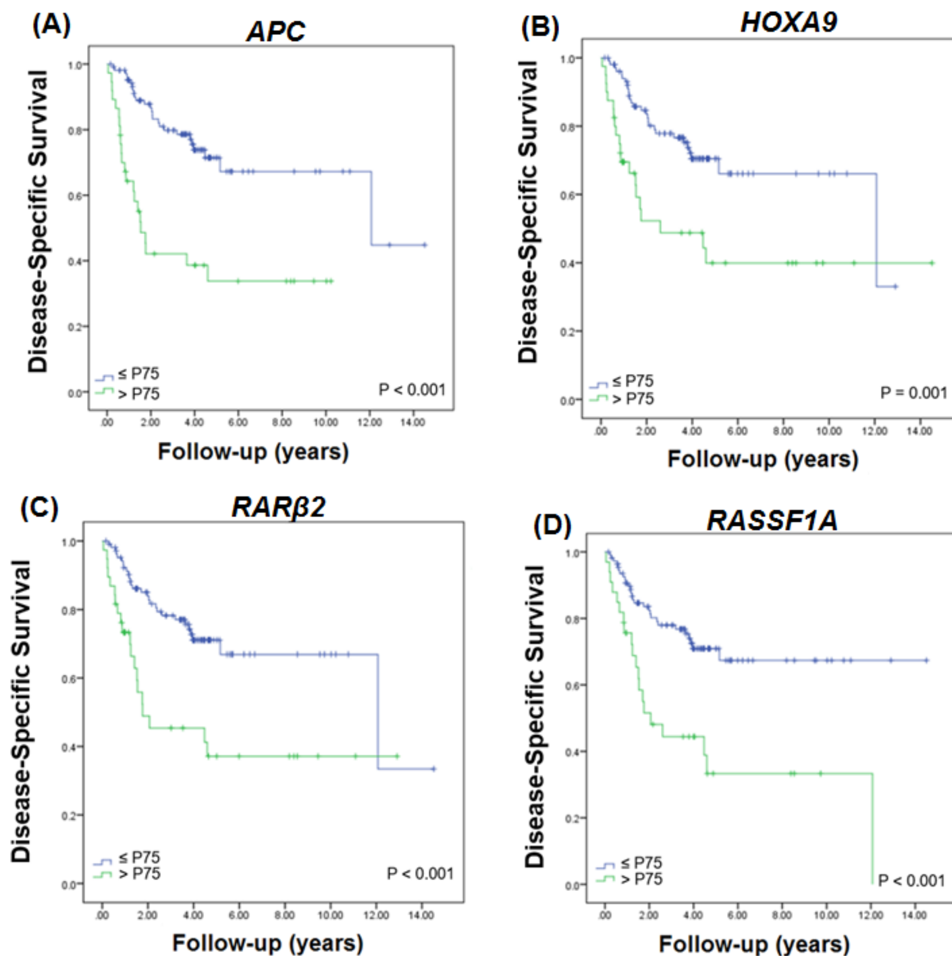

Figure 17 - Disease-Specific Survival according to (A) *APC*, (B*) HOXA9*, (C) *RARβ2* and (D) *RASSF1A* methylation levels.

Concerning clinicopathological variables, poor differentiation, higher pathological stage and LCC and SCLC subtypes were significantly associated with worse prognosis (Figure 18 and Table 6). Thus we only further analyzed the significant variables.

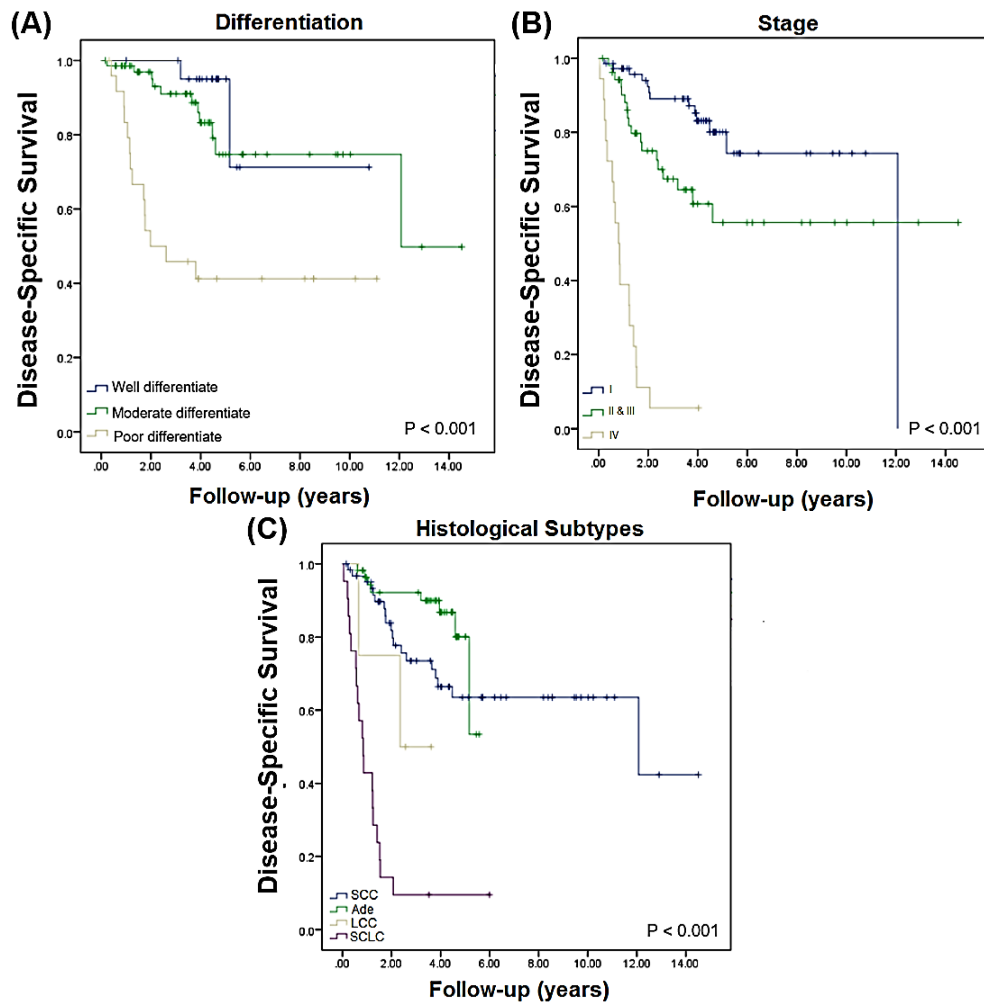

Figure 18 – Disease-Specific Survival according to (A) Differentiation grade, (B) Stage and (C) Histological subtypes. (SCC: Squamous cell carcinoma; Ade: Adenocarcinoma; LCC: Large cell carcinoma; SCLC: Small cell lung cancer)

Table 6 – Univariate Cox regression analyses assessing the potential of clinical and epigenetic variables in the prediction of disease-specific survival for 152 LCa patients.

DISEASE-SPECIFIC SURVIVAL	VARIABLES	HAZARD RATIO (HR)	95 % CI for OR	P value
UNIVARIABLE	**Age diagnosis**	1.02	(0.99 – 1.06)	0.177
	Gender	2.20	(0.99 – 4.91)	0.054
	Differentiation			
	Well *vs* Moderate	2.06	(0.45 - 9.43)	0.350
	Well *vs* Poor	9.28	(2.10 – 41.00)	**0.003**
	Stage			
	I *vs* II & III	2.33	(1.14 – 4.79)	**0.021**
	I *vs* IV	18.57	(8.57 – 40.25)	**0.000**
	Histological Subtype			
	Ade *vs* SCC	2.08	(0.90 – 4.81)	0.085
	Ade *vs* LCC	4.89	(1.03 – 23.29)	**0.046**
	Ade *vs* SCLC	17.85	(7.60 – 41.95)	**0.000**
	APC methylation ≥ p75	3.71	(2.08 – 6.60)	**0.000**
	HOXA9 methylation ≥ p75	2.57	(1.44 – 4.58)	**0.001**
	RARβ2 methylation ≥ p75	2.77	(1.55 – 4.91)	**0.000**
	RASSF1A methylation ≥ p75	2.96	(1.66 – 5.26)	**0.000**
	SHOX2 methylation ≥ p50	1.41	(0.79 – 2.52)	0.246
	TFPI2 methylation ≥ p50	0.75	(0.42 – 1.36)	0.344

Ade: Adenocarcinoma; SCC: Squamous Cell Carcinoma; LCC: Large Cell Carcinoma; SCLC: Small Cell Lung Cancer

To identify which independent factors jointly had a significant influence on overall survival, the Cox proportional hazards modeling technique was applied. Hence, we introduced in a Cox-regression model for DSS all statistically significant variables (stage, histological subtypes, differentiation, *APC, HOXA9*, *RARβ2 and RASSF1A*), that were selected in the final model as independent predictors of outcome. The degrees of differentiation are not applicable to SCLC and LCC, thus, these subtypes were excluded from analysis by Cox regression model.

The multivariate models for *APC, HOXA9*, *RARβ2 and RASSF1A* demonstrated that shorter DSS was dependent of poor differentiation (Table 7, 8, 9 and 10). Patients

with poorly differentiated tumors and high *RASSF1A*, *RARβ2*, *APC* and *HOXA9* promoter methylation levels had an 8.19-fold, 10.34-fold, 8.98-fold and 8.10-fold increased likelihood of dying from LCa, respectively. Therefore, none of the molecular variables retained independent prognostic value in multivariable analysis.

Table 7 - Multivariate Cox regression analyses assessing the potential of clinical and epigenetic variables in the prediction of disease-specific survival for 152 LCa patients.

DISEASE-SPECIFIC SURVIVAL	VARIABLES	HAZARD RATIO (HR)	95 % CI for OR	P value
MULTIVARIABLE	**Differentiation**			
	Well *vs* Moderate	1.84	(0.36 – 9.26)	0.462
	Well *vs* Poor	8.10	(1.53 – 42.81)	**0.014**
	Stage			
	I *vs* II & III	1.60	(0.72 – 3.54)	0.249
	I *vs* IV	0.00	(0.00 – 0.00)	0.981
	Histological Subtype			
	Ade *vs* SCC	1.02	(0.39 – 2.66)	0.976
	APC methylation ≥ p75	1.09	(0.44 – 2.73)	0.853

Ade: Adenocarcinoma; SCC: Squamous Cell Carcinoma

Table 8 - Multivariate Cox regression analyses assessing the potential of clinical and epigenetic variables in the prediction of disease-specific survival for 152 LCa patients.

DISEASE-SPECIFIC SURVIVAL	VARIABLES	HAZARD RATIO (HR)	95 % CI for OR	P value
MULTIVARIABLE	**Differentiation**			
	Well *vs* Moderate	1.79	(0.36 – 9.10)	0.478
	Well *vs* Poor	8.98	(1.73 – 46.70)	**0.009**
	Stage			
	I *vs* II & III	1.89	(0.82 – 4.39)	0.136
	I *vs* IV	0.00	(0.00 – 0.00)	0.981
	Histological Subtype			
	Ade *vs* SCC	1.12	(0.43 – 2.92)	0.819
	HOXA9 methylation ≥ p75	0.601	(0.22 – 1.63)	0.316

Ade: Adenocarcinoma; SCC: Squamous Cell Carcinoma

Table 9 - Multivariate Cox regression analyses assessing the potential of clinical and epigenetic variables in the prediction of disease-specific survival for 152 LCa patients.

DISEASE-SPECIFIC SURVIVAL	VARIABLES	HAZARD RATIO (HR)	95 % CI for OR	P value
MULTIVARIABLE	**Differentiation**			
	Well *vs* Moderate	1.93	(0.39 – 9.89)	0.419
	Well *vs* Poor	10.34	(1.92 – 54.71)	**0.007**
	Stage			
	I *vs* II & III	1.92	(0.85 – 4.24)	0.116
	I *vs* IV	0.00	(0.00 – 0.00)	0.981
	Histological Subtype			
	Ade *vs* SCC	1.02	(0.39 – 2.65)	0.969
	RARβ2 methylation ≥ p75	0.48	(1.64 – 1.38)	0.173

Ade: Adenocarcinoma; SCC: Squamous Cell Carcinoma

Table 10 - Multivariate Cox regression analyses assessing the potential of clinical and epigenetic variables in the prediction of disease-specific survival for 152 LCa patients.

DISEASE-SPECIFIC SURVIVAL	VARIABLES	HAZARD RATIO (HR)	95 % CI for OR	P value
MULTIVARIABLE	**Differentiation**			
	Well *vs* Moderate	1.84	(0.37 – 9.25)	0.461
	Well *vs* Poor	8.19	(1.56 – 42.96)	**0.013**
	Stage			
	I *vs* II & III	1.58	(0.72 – 3.49)	0.256
	I *vs* IV	0.00	(0.00 – 0.00)	0.987
	Histological Subtype			
	Ade *vs* SCC	2.08	(0.39 – 2.64)	0.981
	RASSF1A methylation ≥ p75	1.21	(0.48 – 3.05)	0.687

Ade: Adenocarcinoma; SCC: Squamous Cell Carcinoma

4.6.2 Disease-Free Survival

Recurrence is an important endpoint in LCa, thus we tested the prognostic value of clinicopathological variables and gene promoter methylation levels in this setting.

For statistical purposes, *RASSF1A and TFPI2* methylation levels were dichotomized using percentile 75, *HOXA9* and *SHOX2* using percentile 50, and *RARβ2* as well as *APC* using percentile 25, as threshold values.

Only differentiation grade and *RASSF1A* methylation levels were statistically associated with shorter DFS in univariate analysis (P = 0.005; P = 0.034; respectively) (Figure 19; Table 11). Thus, we only analyzed the association between these variables and DFS in multivariable model.

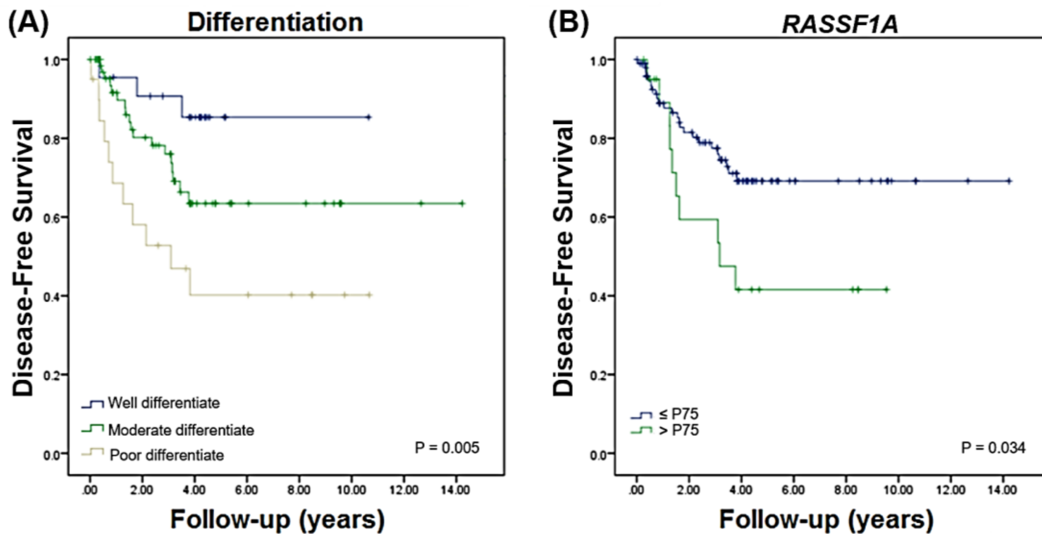

Figure 19 - Disease-Free Survival according to (A) Differentiation grade, (B) *RASSF1A* methylation level.

Table 11 – Univariate Cox regression analyses assessing the potential of clinical and epigenetic variables in the prediction of disease-free survival for 152 LCa patients.

DISEASE-FREE SURVIVAL	VARIABLES	HAZARD RATIO (HR)	95 % CI for OR	P value
UNIVARIABLE	**Age diagnosis**	0.972	(0.93 – 1.01)	0.164
	Gender	0.521	(0.22 – 1.26)	0.147
	Differentiation			
	Well *vs* Moderate	2.71	(0.80 - 9.24)	0.110
	Well *vs* Poor	5.71	(1.60 – 20.49)	**0.008**
	Stage			
	I *vs* II & III	1.83	(0.93 – 3.60)	0.08
	I *vs* IV	0.00	(0.00 – 0.00)	0.98
	Histological Subtype			
	Ade *vs* SCC	1.64	(0.80 – 3.36)	0.176
	Ade *vs* LCC	0.98	(0.13 – 7.56)	0.985
	Ade *vs* SCLC	2.28	(0.296 – 17.58)	0.428
	APC methylation ≥ p25	1.61	(0.819 – 3.17)	0.167
	HOXA9 methylation ≥ p50	1.66	(0.69 – 4.01)	0.260
	RARβ2 methylation ≥ p25	2.18	(0.90 – 5.28)	0.083
	RASSF1A methylation ≥ p75	2.18	(1.04 – 4.57)	**0.039**
	SHOX2 methylation ≥ p50	1.15	(0.59 – 2.26)	0.679
	TFPI2 methylation ≥ p75	0.65	(0.27 – 1.57)	0.332

Ade: Adenocarcinoma; SCC: Squamous Cell Carcinoma; LCC: Large Cell Carcinoma; SCLC: Small Cell Lung Cancer

Similar to DSS, patients with poorly differentiated tumors and concomitant high *RASSF1A* promoter methylation levels had shorter DFS and a 5.39-fold increased hazard ratio (Table 12). Thus, high *RASSF1A* methylation levels does not independently predict for shorter DFS.

Table 12 - Multivariate Cox regression analyses assessing the potential of clinical and epigenetic variables in the prediction of disease-free survival for 152 LCa patients

DISEASE-FREE SURVIVAL	VARIABLES	HAZARD RATIO (HR)	95 % CI for OR	P value
MULTIVARIABLE	**Differentiation**			
	Well *vs* Moderate	2.59	(0.76 – 8.81)	0.129
	Well *vs* Poor	5.39	(1.50 – 19.42)	**0.010**
	RASSF1A methylation ≥ p75	2.08	(0.96 – 4.51)	0.064

5. DISCUSSION

Lung cancer is the most common malignancy in both genders and the most frequent cause of cancer-related death. Currently, no screening test is available for LCa, thus, patients suspected of having LCa are diagnosed because of symptoms, routine exams or secondarily to other clinical imaging investigations. The importance of an early diagnosis along with personalized therapy is currently acknowledged as major issues in LCa management[57]. Thus, accurate LCa subtype discrimination is critical to reduce mortality rate through improvement of therapeutic strategies.

Using aberrantly methylated genes, we attempted to discriminate LCa from normal tissues. *RASSF1A* methylated levels discriminated LCa from controls. Nevertheless, the normal tissues used showed higher *APC, RARβ2* and *HOXA9* methylation levels comparing to LCa. This might be related with the origin of the control tissues. These were procured from patients with lung metastasis from non-pulmonary neoplasms, to avoid a possible "field-effect phenomenon" in lung tissues from lung cancer patients. Because the tumors involving the lung in those cases were metastatic, it was assumed that pulmonary parenchyma was normal. Thus, finding high promoter methylation levels for some genes might either represent an effect of aging, alterations due to previous therapy or contamination with tumor cells at distance. Concerning *SHOX2* and *TFPI2* aberrant methylation, no significant differences were disclosed between LCa and NL samples, contrarily to other reports on *SHOX2* promoter methylation in lung cancer [61, 37, 54] . Consequently we focused mostly in LCa subtypes' discrimination.

Interestingly, *APC*, *RARβ2* and *RASSF1A* promoter methylation levels significantly associated with advanced pathological stage. Nevertheless, Kim *et al.*[52] and Endoh *et al.*[62] did not find significant associations between *RASSF1A* promoter methylation levels and pathological stage and Usadel *et al*[41] also showed no association between *APC* methylation levels and LCa stage. Contrarily, Ponomaryova *et al.*[49] reported that increased *RARβ2* methylation levels in cell-free DNA were associated with tumor stage. These discrepancies might be explained by the subtype of the analyzed tumors and the clinical samples, as well as differences in the methylation assessment methodologies. It should be recalled that most of the cases in this series derive from surgical specimens, thus representing mostly LCa that are clinically amenable to curative-intent excision, which correspond to early stage tumors.

Concerning the major LCa subtypes (NSCLC and SCLC), our results indicate that *APC, RASSF1A, RARβ2* and *HOXA9* methylation levels can discriminate these subtypes, however the same was not observed for *SHOX2* and *TFPI2* methylation levels. Furthermore higher *APC, RASSF1A, RARβ2* and *HOXA9* methylation levels were found in

SCLC compared to NSCLC. Our findings contradict those of Virmani *et al* [33], which found higher *APC* methylation levels in NSCLC cell lines. This might be due to the different methodological approach, since we have used quantitative whereas they have used qualitative methylation-specific PCR. Indeed, quantitative MSP is a sensitive and specific methodology that requires only minute amounts of DNA. It is able to differentiate between methylated and unmethylated DNA and consequently quantify the methylation level of sample using oligonucleotides whose 3'-ends match the methylation status of specific CpG sites in a bisulfite-treated template [63 64]. However, MSP it is not quantitative, and may, thus, lead to false positive results when the PCR conditions used are not optimal. Moreover specificity of MSP is significantly lower compared to that of qMSP[63, 65, 66].

Conversely, the distribution of both *RASSF1A* and *RARβ2* methylation levels were similar to those of previous reports [47, 62, 67-69]. Regarding *SHOX2* and *TFPI2* the same was not observed. Both genes were found to be highly methylated in SCLC [37, 38, 54, 55], in contrast our results whereas *TFPI2* showed basal methylation levels in both subtypes. Concerning *HOXA9,* high *HOXA9* methylation levels were described in NSCLC [34, 45, 70], however at our knowledge no data were published so far concerning SCLC, probably due to scarcity of specimens available compared with NSCLC, a much more frequent malignancy despite being less aggressive [55].

The originality of this dissertation lies in the assessment of methylation levels of cancer-related genes to discriminate LCa subtypes. Therefore we observed that *APC, HOXA9*, *RARβ2 and RASSF1A* might discriminate among some minor subtypes, but the same was not observed for *SHOX2* and *TFPI2*.

APC, HOXA9, RARβ2, and *RASSF1A* promoter methylation levels were able to distinguish adenocarcinomas from SCLC. Nonetheless, Guo *et al.*[71] reported that APC methylation status strongly associated with NSCLC, especially with adenocarcinoma, but did not discriminate this subtype from others. Our results also showed that *APC, RARβ2* and *RASSF1A* methylation levels discriminate SCC from SCLC. Moreover, we demonstrated that *RARβ2* methylation levels are able to discriminate LCC from SCLC, whereas *HOXA9* distinguished adenocarcinoma from SCC in agreement with results reported by Hwang *et al.*[45]. However, contrarily to Kneip *et al.*[54] no significant differences were reported among minor subtypes relatively to *SHOX2* promoter methylation levels. Different assay sensitivity and small sample size might be the reasons for these discrepancies. Virmani *et al.*[47] found that frequencies in *RARβ2* gene promoter methylation did not differ significantly between adenocarcinomas (37%) and SCC (54%), which further confirms our observations. On other hand, Ponomaryova *et al.*[49] described that in SCC patients, *RARβ2* methylation levels in the cell surface blood of cell free DNA

was higher compared with patients with adenocarcinoma. There are a number of possible reasons for this divergence, including specific physiological characteristics in the progression of each tumor or the quantity and quality of DNA template extracted from FFPE tissues and other factors.

The last goal of this dissertation consisted on the determination of the prognostic value of candidate methylated genes.

In univariate analysis, most standard clinicopathological parameters associated with DSS (differentiation grade, stage, subtype) or DFS (tumor differentiation). Moreover, higher *APC, RASSF1A, RARβ2* and *HOXA9* methylation levels associated with shorter DSS, whereas higher *RASSF1A* methylation levels also associated with shorter DFS. To verify these correlations, multivariate regression models were established.

However, in multivariate analysis only tumor differentiation retained independent prognostic value, both for DSS and DFS. Several studies verified a significant association between genes' promoter hypermethylation and poorly differentiated LCa[72-74], which is in line with our results. However, contrarily to other studies no associations were found between *APC*, *TFPI2*, *HOXA9* and *RARβ2* methylation levels and poor prognosis[34, 38, 41, 45, 49, 75], which could be partially attributable to our choice of samples and methodology.

The data regarding *RASSF1A* methylation is rather controversial, although it has been suggested as a useful tool for LCa diagnosis[49, 62, 76, 77], associations between *RASSF1A* methylation level and poor prognostics [49, 78, 79], were also reported, contrarily to our observations. However, Drilon *et al.*[72] demonstrated that *RASSF1A* promoter methylation was not prognostic for early tumor recurrence in their study with resected NSCLC. Similarly to our results, an association between gene hypermethylation and poorly differentiated histology, was also demonstrated.

Thus, *APC, RASSF1A, RARβ2* and *HOXA9* promoter methylation levels are not independently indicative of more clinically aggressive LCa, because that is dependent of tumor differentiation.

This exploratory study aimed to differentiate LCa subtypes with a panel of six hypermethylated genes, and evaluate the prognostic value of those molecular alterations. Thus far no other study has attempted the same, which might explain lack of studies to compare our results. Importantly, the most promising candidate biomarkers might be tested in cell free DNA from plasma samples of individuals suspected of carrying LCa. Such test might not only speed-up the diagnostic process but also help discriminate the LCa subtype to assist in clinical decision-making.

6. CONCLUSIONS AND FUTURE PERSPECTIVES

In conclusion, the evaluation of *RARβ2* and *HOXA9* promoter methylation levels using qMSP appears to be useful to discriminate among major and minor LCa subtypes in FFPE tissues. High *APC*, *HOXA9*, RARβ2 and *RASSF1A* promoter methylation levels associate with poor disease specific survival owing to its association with tumor differentiation. The same holds true for high *RASSF1A* promoter methylation levels and poor disease free survival. The clinical usefulness of these results requires validation in a set of plasma samples in the near future. Additional studies are necessary to optimize the ability to discriminate LCa subtypes using *RARβ2* and *HOXA9* promoter methylation levels.

Validation in plasma it is an important aim because a sensitive detection method could enable early diagnosis and improve survival of LCa patients. It is believed that plasma DNA is of tumor origin because the genetic alterations are similar to those found in the corresponding primary tumors[80]. Furthermore, many investigators have reported that microsatellite alterations and gene mutations could be identified in the plasma DNA of various cancer patients, which must derive from cancer cells. Thus, circulating tumor-derived DNA might be used as a source for tumor detection[76], and quantification of cell-free DNA in plasma and characterization of specific molecular changes could be very useful for the management of LCa[80].

The major advantage of plasma is that it may be used as minimally invasive approach for early diagnosis and screening[80]. However, plasma samples are more challenging, as the total amount of lung-derived DNA and the fraction of tumor DNA are expected to be significantly lower, moreover, it is reported that sensitivity and specificity are slightly lower when using plasma compared e.g. with bronchial aspirates. Furthermore, blood plasma contains a complex mixture of DNA originating potentially from any part of the body. Hence, the analytical performance requirements for analyzing these body fluids are higher and the markers need to be specific for lung tumor DNA to assure a high specificity of the test[54].

7. REFERENCES

1. Alberg AJ, Brock MV, Samet JM. 52 - Epidemiology of Lung Cancer A2 - Broaddus, V. Courtney. In: Mason RJ, Ernst JD, King TE, et al., eds. *Murray and Nadel's Textbook of Respiratory Medicine (Sixth Edition)*. Philadelphia: W.B. Saunders; 2016:927-939.e925.

2. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer* 2015;136:E359-E386.

3. Ferlay J SI, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet].

Lyon, France: International Agency for Research on Cancer; 2013. *globocaniarcfr, accessed on 19/7/2016*.

4. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *European journal of cancer* 2013;49:1374-1403.

5. Kim Y, Jin D, Lee BB, et al. RARβ2 hypermethylation is associated with poor recurrence-free survival in never-smokers with adenocarcinoma of the lung. *Clinical epigenetics* 2015;7:1.

6. Brambilla C, Fievet F, Jeanmart M, et al. Early detection of lung cancer: role of biomarkers. *European Respiratory Journal* 2003;21:36s-44s.

7. Travis WD, Brambilla E, Burke A, et al. *WHO classification of tumours of the lung, pleura, thymus and heart*. International Agency for Research on Cancer; 2015.

8. Massion PP, Sequist LV, Pao W. 51 - Biology of Lung Cancer A2 - Broaddus, V. Courtney. In: Mason RJ, Ernst JD, King TE, et al., eds. *Murray and Nadel's Textbook of Respiratory Medicine (Sixth Edition)*. Philadelphia: W.B. Saunders; 2016:912-926.e916.

9. Khuri FR. Lung Cancer and Other Pulmonary Neoplasms. In: Crow MK, Doroshow JH, Drazen JM, et al., eds. *Goldman-Cecil Medicine, twenty-fifth edition*. 2016.

10. Mehta A, Dobersch S, Romero-Olmedo AJ, et al. Epigenetics in lung cancer diagnosis and therapy. *Cancer and Metastasis Reviews* 2015:1-13.

11. Board PATE. Non-Small Cell Lung Cancer Treatment (PDQ®). 2016.

12. Schrump DS, Carter D, Kelsey CR, et al. Non-Small Cell Lung Cancer. In: DeVilta VT, Lawrence TS, Rosenberg SA, eds. *Cancer Principles and Practice of Oncology*. 2011:2800.

13. UyBico SJ, Wu CC, Suh RD, et al. Lung Cancer Staging Essentials: The New TNM Staging System and Potential Imaging Pitfalls 1. *Radiographics* 2010;30:1163-1181.

14. Frederick L, Page DL, Fleming ID, et al. *AJCC cancer staging manual, 7th Edition*. Springer Science & Business Media; 2010.

15. Wagnetz U, Menezes RJ, Boerner S, et al. CT screening for lung cancer: implication of lung biopsy recommendations. *American Journal of Roentgenology* 2012;198:351-358.

16. Alamgeer M, Ganju V, Watkins DN. Novel therapeutic targets in non-small cell lung cancer. *Current opinion in pharmacology* 2013;13:394-401.

17. Larsen JE, Govindan R, Minna JD. 32 - Molecular Basis of Lung Cancer. *The Molecular Basis of Cancer (Fourth Edition)*. Philadelphia: Content Repository Only!; 2015:475-490.e471.

18. Dupont C, Armant DR, Brenner CA. Epigenetics: definition, mechanisms and clinical perspective. *Seminars in reproductive medicine*. NIH Public Access; 2009;27:351.

19. Valdes-Mora F, Clark S. Prostate cancer epigenetic biomarkers: next-generation technologies. *Oncogene* 2015;34:1609-1618.

20. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. *Carcinogenesis* 2010;31:27-36.

21. Walter K, Holcomb T, Januario T, et al. Discovery and development of DNA methylation-based biomarkers for lung cancer. *Epigenomics* 2014;6:59-72.

22. Ramalho-Carvalho J, Fromm B, Henrique R, et al. Deciphering the function of non-coding RNAs in prostate cancer. *Cancer and Metastasis Reviews* 2016:1-28.

23. Langevin SM, Kratzke RA, Kelsey KT. Epigenetics of lung cancer. *Translational Research* 2015;165:74-90.

24. Liloglou T, Bediaga NG, Brown BR, et al. Epigenetic biomarkers in lung cancer. *Cancer letters* 2014;342:200-212.

25. Gibney E, Nolan C. Epigenetics and gene expression. *Heredity* 2010;105:4-13.

26. Yang B. Chapter 9 - Epigenetics: DNA Hypermethylation in Cancer A2 - Tubbs, Raymond R. In: Stoler MH, ed. *Cell and Tissue Based Molecular Pathology*. Philadelphia: Churchill Livingstone; 2009:84-90.

27. Anglim PP, Alonzo TA, Laird-Offringa IA. DNA methylation-based biomarkers for early detection of non-small cell lung cancer: an update. *Mol Cancer* 2008;7:81.

28. Garcia-Gimenez J, Sanchis-Gomar F, Lippi G, et al. Epigenetic biomarkers: A new perspective in laboratory diagnostics. *Clinica Chimica Acta* 2012;413:1576-1582.

29. Baylin SB. 5 - Epigenetics and Cancer A2 - Mendelsohn, John. In: Gray JW, Howley PM, Israel MA, et al., eds. *The Molecular Basis of Cancer (Fourth Edition)*. Philadelphia: Content Repository Only!; 2015:67-78.e63.

30. Forde PM, Brahmer JR, Kelly RJ. New Strategies in Lung Cancer: Epigenetic Therapy for Non–Small Cell Lung Cancer. *Clinical Cancer Research* 2014;20:2244-2248.

31. Chen Q, Zhu X, Li Y, et al. Epigenetic regulation and cancer (Review). *Oncology reports* 2014;31:523-532.

32. Costa-Pinheiro P, Montezuma D, Henrique R, et al. Diagnostic and prognostic epigenetic biomarkers in cancer. *Epigenomics* 2015;7:1003-1015.

33. Virmani AK, Rathi A, Sathyanarayana UG, et al. Aberrant methylation of the adenomatous polyposis coli (APC) gene promoter 1A in breast and lung carcinomas. *Clinical Cancer Research* 2001;7:1998-2004.

34. Hwang JA, Lee BB, Kim Y, et al. HOXA9 inhibits migration of lung cancer cells and its hypermethylation is associated with recurrence in non-small cell lung cancer. *Molecular carcinogenesis* 2015;54.

35. Donninger H, Vos MD, Clark GJ. The RASSF1A tumor suppressor. *Journal of cell science* 2007;120:3163-3172.

36. Schneider KU, Dietrich D, Fleischhacker M, et al. Correlation of SHOX2 gene amplification and DNA methylation in lung cancer tumors. *BMC cancer* 2011;11:102.

37. 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.

38. Wu D, Xiong L, Wu S, et al. TFPI-2 methylation predicts poor prognosis in non-small cell lung cancer. *Lung Cancer* 2012;76:106-111.

39. Hube F, Reverdiau P, Iochmann S, et al. Characterization and functional analysis of TFPI-2 gene promoter in a human choriocarcinoma cell line. *Thrombosis research* 2003;109:207-215.

40. Senda T, Shimomura A, Iizuka-Kogo A. Adenomatous polyposis coli (Apc) tumor suppressor gene as a multifunctional gene. *Anatomical science international* 2005;80:121-131.

41. Usadel H, Brabender J, Danenberg KD, et al. Quantitative adenomatous polyposis coli promoter methylation analysis in tumor tissue, serum, and plasma DNA of patients with lung cancer. *Cancer research* 2002;62:371-375.

42. Brabender J, Usadel H, Danenberg KD, et al. Adenomatous polyposis coli gene promoter hypermethylation in non-small cell lung cancer is associated with survival. *Oncogene* 2001;20:3528-3532.

43. Garcia-Fernàndez J. The genesis and evolution of homeobox gene clusters. *Nature Reviews Genetics* 2005;6:881-892.

44. Rauch T, Wang Z, Zhang X, et al. Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. *Proceedings of the National Academy of Sciences* 2007;104:5527-5532.

45. Hwang S-H, Kim KU, Kim J-E, et al. Detection of HOXA9 gene methylation in tumor tissues and induced sputum samples from primary lung cancer patients. *Clinical Chemistry and Laboratory Medicine* 2011;49:699-704.
46. Kim Y, Jin D, Lee BB, et al. RARβ2 hypermethylation is associated with poor recurrence-free survival in never-smokers with adenocarcinoma of the lung. *Clinical epigenetics* 2015;7:32.
47. Virmani AK, Rathi A, Zöchbauer-Müller S, et al. Promoter methylation and silencing of the retinoic acid receptor-β gene in lung carcinomas. *Journal of the National Cancer Institute* 2000;92:1303-1307.
48. Kim JS, Lee H, Kim H, et al. Promoter Methylation of Retinoic Acid Receptor Beta 2 and the Development of Second Primary Lung Cancers in Non–Small-Cell Lung Cancer. *Journal of clinical oncology* 2004;22:3443-3450.
49. Ponomaryova AA, Rykova EY, Cherdyntseva NV, et al. RARβ2 gene methylation level in the circulating DNA from blood of patients with lung cancer. *European Journal of Cancer Prevention* 2011;20:453-455.
50. Nakayama T, Watanabe M, Yamanaka M, et al. The Role of Epigenetic Modifications in Retinoic Acid Receptor &bgr; 2 Gene Expression in Human Prostate Cancers. *Laboratory Investigation* 2001;81:1049-1057.
51. Dubois F, Keller M, Calvayrac O, et al. RASSF1A Suppresses the Invasion and Metastatic Potential of Human Non–Small Cell Lung Cancer Cells by Inhibiting YAP Activation through the GEF-H1/RhoB Pathway. *Cancer research* 2016;76:1627-1640.
52. Kim D-H, Kim JS, Ji Y-I, et al. Hypermethylation of RASSF1A promoter is associated with the age at starting smoking and a poor prognosis in primary non-small cell lung cancer. *Cancer research* 2003;63:3743-3746.
53. Ilse P, Biesterfeld S, Pomjanski N, et al. SHOX2 DNA methylation is a tumour marker in pleural effusions. *Cancer Genomics-Proteomics* 2013;10:217-223.
54. Kneip C, Schmidt B, Seegebarth A, et al. SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer in plasma. *Journal of Thoracic Oncology* 2011;6:1632-1638.
55. Lavergne M, Jourdan M-L, Blechet C, et al. Beneficial role of overexpression of TFPI-2 on tumour progression in human small cell lung cancer. *FEBS open bio* 2013;3:291-301.
56. Eads CA, Danenberg KD, Kawakami K, et al. MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic acids research* 2000;28:e32-00.
57. Dietrich D, Kneip C, Raji O, et al. Performance evaluation of the DNA methylation biomarker SHOX2 for the aid in diagnosis of lung cancer based on the analysis of bronchial aspirates. *International journal of oncology* 2012;40:825.

58. Reinert T, Borre M, Christiansen A, et al. Diagnosis of bladder cancer recurrence based on urinary levels of EOMES, HOXA9, POU4F2, TWIST1, VIM, and ZNF154 hypermethylation. *PloS one* 2012;7:e46297.
59. Lee BB, Lee EJ, Jung EH, et al. Aberrant methylation of APC, MGMT, RASSF2A, and Wif-1 genes in plasma as a biomarker for early detection of colorectal cancer. *Clinical Cancer Research* 2009;15:6185-6191.
60. Liu L, Kron KJ, Pethe VV, et al. Association of tissue promoter methylation levels of APC, TGFβ2, HOXD3 and RASSF1A with prostate cancer progression. *International Journal of Cancer* 2011;129:2454-2462.
61. Zhao Q-T, Guo T, Wang H-E, et al. Diagnostic value of SHOX2 Dna methylation in lung cancer: a meta-analysis. *OncoTargets and therapy* 2015;8:3433.
62. Endoh H, Yatabe Y, Shimizu S, et al. RASSF1A gene inactivation in non-small cell lung cancer and its clinical implication. *International journal of cancer* 2003;106:45-51.
63. Shames DS, Minna JD, Gazdar AF. Methods for detecting DNA methylation in tumors: from bench to bedside. *Cancer letters* 2007;251:187-198.
64. Husseiny MI, Kuroda A, Kaye AN, et al. Development of a quantitative methylation-specific polymerase chain reaction method for monitoring beta cell death in type 1 diabetes. *PloS one* 2012;7:e47942.
65. Parrella P, la Torre A, Copetti M, et al. High specificity of quantitative methylation-specific PCR analysis for MGMT promoter hypermethylation detection in gliomas. *BioMed Research International* 2009;2009.
66. Stuopelytė K, Daniūnaitė K, Laurinavičienė A, et al. High-resolution melting-based quantitative analysis of RASSF1 methylation in breast cancer. *Medicina (Kaunas)* 2013;49:78-83.
67. Dammann R, Takahashi T, Pfeifer GP. The CpG island of the novel tumor suppressor gene RASSF1A is intensely methylated in primary small cell lung carcinomas. *Oncogene* 2001;20:3563-3567.
68. Burbee DG, Forgacs E, Zöchbauer-Müller S, et al. Epigenetic inactivation of RASSF1A in lung and breast cancers and malignant phenotype suppression. *Journal of the National Cancer Institute* 2001;93:691-699.
69. Zhao Y, Zhou H, Ma K, et al. Abnormal methylation of seven genes and their associations with clinical characteristics in early stage non-small cell lung cancer. *Oncology letters* 2013;5:1211-1218.
70. Plowright L, Harrington K, Pandha H, et al. HOX transcription factors are potential therapeutic targets in non-small-cell lung cancer (targeting HOX genes in lung cancer). *British journal of cancer* 2009;100:470-475.

71. Guo S, Tan L, Pu W, et al. Quantitative assessment of the diagnostic role of APC promoter methylation in non-small cell lung cancer. *Clinical epigenetics* 2014;6:1.

72. Drilon A, Sugita H, Sima CS, et al. A Prospective Study of Tumor Suppressor Gene Methylation as a Prognostic Biomarker in Surgically Resected Stage I to IIIA Non–Small-Cell Lung Cancers. *Journal of Thoracic Oncology* 2014;9:1272-1277.

73. Choi N, Son DS, Song I, et al. RASSF1A is not appropriate as an early detection marker or a prognostic marker for non-small cell lung cancer. *International journal of cancer* 2005;115:575-581.

74. Wang J, Lee JJ, Wang L, et al. Value of p16INK4a and RASSF1A promoter hypermethylation in prognosis of patients with resectable non–small cell lung cancer. *Clinical cancer research* 2004;10:6119-6125.

75. Muñiz-Hernández S, Huerta-Yepez S, Hernández-Pedro N, et al. Association between nuclear expression of retinoic acid receptor alpha and beta and clinicopathological features and prognosis of advanced non-small cell lung cancer. *International Journal of Clinical Oncology* 2016:1-11.

76. Gao L, Xie E, Yu T, et al. Methylated APC and RASSF1A in multiple specimens contribute to the differential diagnosis of patients with undetermined solitary pulmonary nodules. *Journal of thoracic disease* 2015;7:422-432.

77. Hubers AJ, van der Drift MA, Prinsen CF, et al. Methylation analysis in spontaneous sputum for lung cancer diagnosis. *Lung Cancer* 2014;84:127-133.

78. Ko E, Lee BB, Kim Y, et al. Association of RASSF1A and p63 with Poor Recurrence-Free Survival in Node-Negative Stage I–II Non–Small Cell Lung Cancer. *Clinical Cancer Research* 2013;19:1204-1212.

79. Yanagawa N, Tamura G, Oizumi H, et al. Promoter hypermethylation of RASSF1A and RUNX3 genes as an independent prognostic prediction marker in surgically resected non-small cell lung cancers. *Lung cancer* 2007;58:131-138.

80. Pathak AK, Bhutani M, Kumar S, et al. Circulating cell-free DNA in plasma/serum of lung cancer patients as a potential screening and prognostic tool. *Clinical Chemistry* 2006;52:1833-1842.

APPENDICES

APPENDIX I

DNA Extraction From Formalin-fixed Paraffin-embedded Tissues

From each case, a representative paraffin block was selected and an experienced pathologist delimited the area of tumor to be macrodissected. A set of 12 serial tissue sections, with 8μm of thickness, were cut from corresponding paraffin block and placed on glass slides. A disposable sterile scalpel blade was used to macrodissect the tumor areas which were subsequently placed in labeled 1,5mL tubes.

Tissue samples were then desparaffinized using Xilol and Ethanol 100%, 90%, 70% and 50%, and digested in 1000μL of digestion buffer, composed by Tris-HCl 1M, EDTA 0,1M, Tween 20 and sterile bidistilled water (B.Braun, Melsungen, Germany), plus proteinase K (20mg/ml, 60μL) (Sigma-Aldrich®, Germany), by incubation for 1 to 2 days in a water-bath at 55ºC, until total digestion was accomplished.

DNA was extracted from tissues samples by the standard phenolchloroform procedure, using 500 μL of phenol-chloroform solution at pH 8 (Sigma-Aldrich®, Germany; Merck, Germany) in Phase Lock Gel Light tubes (5 PRIME, Germany). After centrifuging the tubes for 15 min at 13,000rpm, the upper aqueous phase containing DNA was transferred to a new tube, and then precipitated at -20ºC overnight using chilled Ethanol 100% (2 volumes of original amount of this phase), Ammonium Acetate 7,5M (1/3 volume) (Sigma-Aldrich®, Germany) and glycogen (2μL).

Posteriorly, samples were washed in ethanol 70%, the pellets air dried and then eluted in 10 μL of sterile distilled water. DNA concentration and purity were assessed using NanoDrop Lite Spectrophotometer (Nanodrop Technologies, USA) and stored at -20°C until further use.

APPENDIX II

Bisulfite Treatment of DNA

Bisulfite treatment of DNA samples was performed using EZ DNA methylation – Gold kit (Zymo Research, Orange, CA) after extraction and quantification of DNA.

Before beginning the procedure, we have to calculate the DNA volume that we will use, accordingly to its concentration and the quantity that we want to have (1000ng), and then we have to add sterile distilled water to the calculated DNA volume of each sample, until we reach the final volume of 20µL.

To each tube was added 130 µL of CT Conversion Reagent and then incubated in a GeneAmp PCR System 2700 thermal cycler (Applied Biosystems, USA) at 98°C for 10 minutes and then at 64°C for 3 cycles of 60 minutes each.

Once finished the incubation, samples were transferred to a Zymo-Spin IC column with 600 µL of M-binding buffer and centrifuged at 10,000 rpm for 30 seconds. After being washed with 100µL of M-Wash buffer and again centrifuged at 10,000 rpm for 30 seconds, desulphonation was achieved with an incubation at room temperature with 200 µL of M-Desulphonation buffer for 20 minutes.

After the incubation, the columns were centrifuged at 10,000 rpm for 30 seconds followed by two washing steps with 200 µL of M-Wash buffer and centrifugations at 10,000 rpm for 30 seconds.

Finally, the column was removed from the collection tube and placed in a 1.5 mL tube. The modified DNA was eluted by incubating the column with 30 µL of sterile distilled water for 5 minutes at room temperature followed by a centrifugation at 12,000 rpm for 30 seconds. This last step was repeated allowing a final volume of 60 µL of modified DNA for each sample. The modified DNA was stored at -80°C until further use.

CpGenome^{TM} Universal Methylated DNA (Merck Millipore, Germany) was also modified, using the guidelines described above and eluted in a total of 20 µL of sterile distilled water.

APPENDIX III

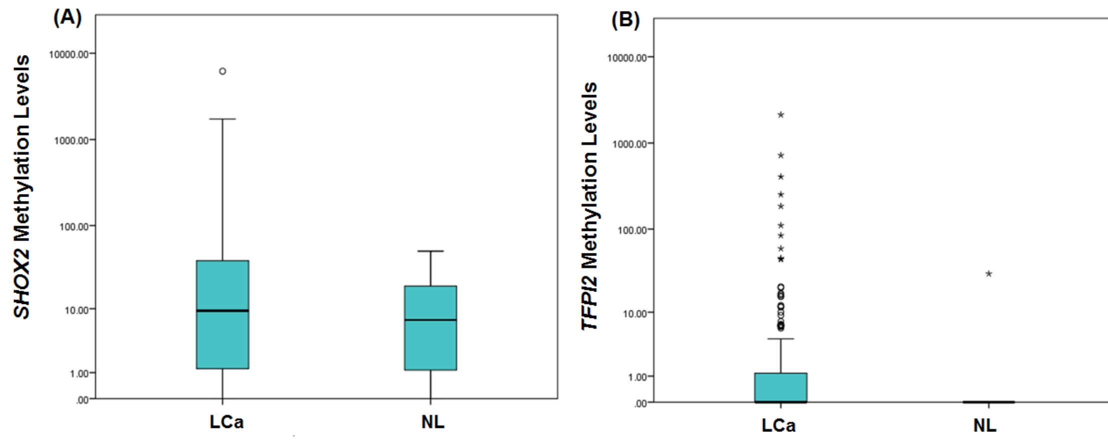

Figure 20 - Boxplots of (A) *SHOX2* and (B*) TFPI2* promoter methylation levels between Lung Cancer (LCa) and normal lung (NL) samples.

APPENDIX IV

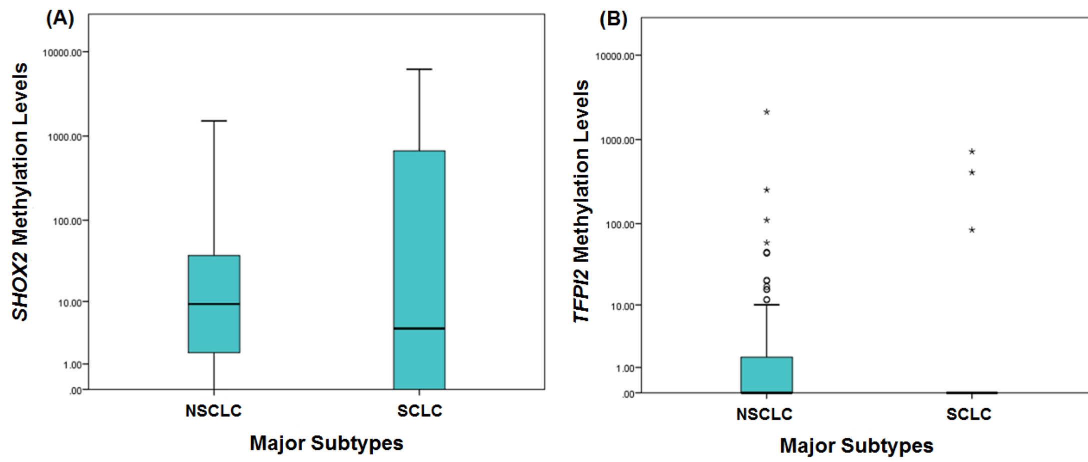

Figure 21 - Boxplots of (A) *SHOX2* and (B*) TFPI2* promoter methylation levels between the major subtypes of Lung Cancer (LCa) samples. (NSCLC: Non-small cell lung cancer)

APPENDIX V

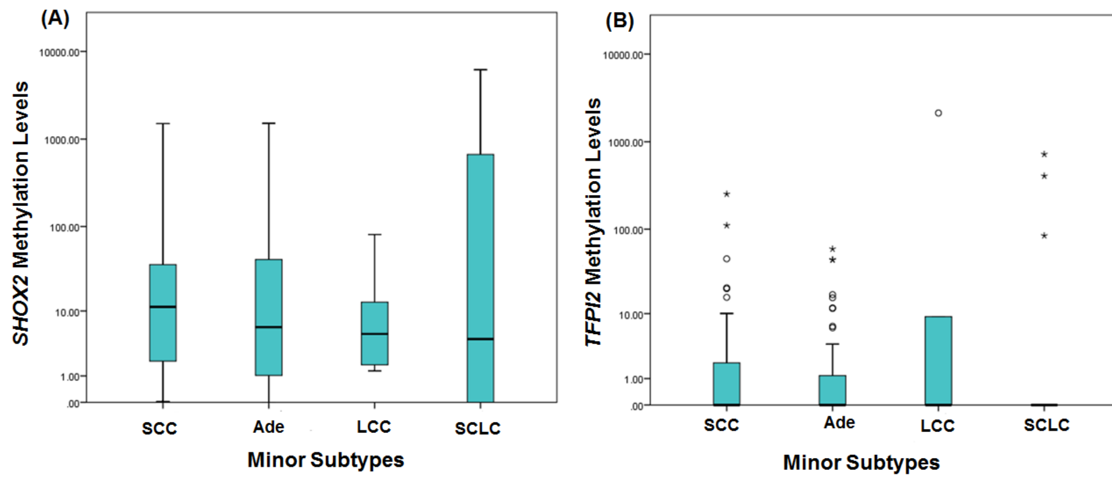

Figure 22 - Boxplots of (A) *SHOX2* and (B*) TFPI2* promoter methylation levels between subtypes of Lung Cancer (LCa) samples. (SCC: Squamous cell carcinoma; Ade: Adenocarcinoma; LCC: Large cell carcinoma; SCLC: Small cell lung cancer)