Genomic profile of tumorgrafts identifies *B2M* as a novel tumor suppressor gene in lung cancer

Carolina Ruivo Pereira

TESI DOCTORAL UPF / 2016

DIRECTORA DE LA TESI

Dra. Montse Sánchez Céspedes

PROGRAMA D'EPIGENÈTICA I BIOLOGIA DEL CÀNCER

upf. Universitat Pompeu Fabra Barcelona

À minha adorada Mãe, a mulher da minha vida.

Ao meu Pai, pela infelicidade de saber melhor do que eu o que é o cancro.

Agradecimentos / Agradecimientos

[PT] Durante a longa jornada que culminou no desenvolvimento desta tese, muitas foram as pessoas que me prestaram apoio, e que sem as quais este trabalho não teria sido levado a bom porto.

[ES] A lo largo del viaje que culminó en el desarrollo de esta tesis, muchas fueron las personas que me apoyaron, y que sin las cuales este trabajo no hubiese llegado a buen término.

En primer lugar, mis más sinceros agradecimientos a Montse, por acogerme en su laboratório y poder realizar la tesis doctoral bajo su dirección. Gracias por haberme facilitado los medios necesarios para llevar a cabo los experimentos, por la ayuda a interpretar datos y por la disponibilidad para discutir ideas y resultados. Pero ante todo, agradezco la puerta del despacho siempre abierta, la motivación transmitida cuando hizo falta y la paciencia para escucharme hablar de mis frustraciones científicas *"Carol, que ya te saldrán las cosas!"*. Crecí mucho no solamente como investigadora, pero también como persona. Por último, mi inmensa gratitud por la humanidad y la comprensión demonstrada en estos últimos meses, que han sido particularmente difíciles para mí.

Agradezco a mis compañeros de laboratorio, por las conversaciones interminables sobre el cáncer, por echarme una mano con los experimentos y por los *meetings* informales para discutir resultados y ayudarme a solucionar los desafíos que afronté con *B2M*. Gracias por aguantarme a mí y a mis manías.

Pero como en todo, existen personas que por una razón u otra se merecen una atención especial. A Eva, mi gratitud por el enorme apoyo y la ayuda prestada a cambio de nada, por la paciencia inagotable de su parte para responder a mis dudas, y por la cercanía y los consejos de como sacar el mejor partido del doctorado. A Ester, estoy agradecida por el interés constante en mi trabajo, por tener siempre respuestas a mis inquietudes y por enseñarme a ver el lado positivo que hay en todo lo que uno hace. A Patri, mil gracias por ser mucho más que una compañera de trabajo: más que pedir primers y hacer cortes histológicos, fue la mejor amiga que yo podría pedir. Estar (literalmente) a su lado durante tanto tiempo no tiene precio y hizo que todo fuera mucho más llano. Gracias por las risas, por la cumplicidad sin precedentes y por ser incansable dándome ánimo y fuerza. Mi gratitud eterna por haberme escuchado siempre que necesité hablar y por no preguntar nada siempre que quise mantenerme callada.

Agradezco también el aporte de todos los que estuvieron directamente implicados en este trabajo. A Toni, Sebas y al equipo del CNIO por solucionar con un *script* informático lo que yo tardaría años; a Anna por el soporte estadístico; a María José por el laborioso trabajo clasificando imunohistoquímicas y por los análisis de supervivencia; a Alberto por ser el maestro de los ratones; a Enric Condom por la ayuda con la histopatología; a todos los que cedieron muestras de tumores de pulmón para que yo pudiera incluirlas en este estudio (Enriqueta Felip, Jun Yokota, Ernest Nadal, Julián Carretero, Luis Montuenga y Luca Roz).

Deixo para último a mais profunda e sentida gratidão à minha família. Aos meus pais, pelo amor incondicional e pelo esforço de uma vida inteira. Obrigada pelas conversas diárias que sempre me fizeram sentir acompanhada mesmo estando longe, e pelas visitas a Barcelona com o tio Luís e a Luísa, que me permitiam fazer um descanso da tese. Ao meu pai, estou grata por me ter incutido o valor da preserverança e por me incentivar sempre a dar o melhor de mim. A ele peço-lhe desculpa por a distância ter impossibilitado que o acompanhasse da forma que eu gostaria na triste batalha contra o cancro. À minha mãe, meu céu na terra, obrigada pela cumplicidade infinita e por me ter ensinado a ser mais tolerante e a dar importância ao que vale a pena. A ela, agradeço tudo o que não se pode descrever em palavras.

Abstract

Next-generation sequencing technologies provide invaluable information to untangle the genetic complexity underlying human tumor development. Nonetheless, discovering cancer genes in low-purity tumors constitute a major hurdle.

Therefore, whole-exome and transcriptome sequencing were performed over lung tumorgrafts. Along with alterations detected in well-established lung cancer genes, changes in genes implicated in various biological functions like epigenetic regulation and immunity were also uncovered.

Here, *B2M* (encoding the β2-microglobulin), a component of the major histocompatibility complex class I (MHC-I), exhibited tumor suppressive features. Mutations in *B2M* emerged in 5% of lung tumors, most of which hindered the correct complex assembly. The loss of MHC-I (identified in one third of the cases) was correlated to lower cytotoxic T-cell intratumoral infiltration, denoting impairments in immune-mediated tumor eradication.

Trends between such deficiency and reduced survival in patients treated with anti-PD-1/PD-L1 agents suggests that β2-microglobulin may act as a biomarker, predicting response to such immunologically-based therapies in lung cancer.

Resumo

As tecnologias de sequenciação de nova geração proporcionam informação valiosa para desvendar a complexidade genética inerente ao desenvolvimento de tumores humanos. No entanto, descobrir genes de cancro em tumores com baixo grau de pureza constitui um obstáculo considerável.

Por esse motivo, levou-se a cabo a sequenciação de exomas e transcriptomas de xenotransplantes de pulmão. Além da detecção de alterações em genes de cancro de pulmão previamente estabelecidos, foram descobertos outros cambios em genes implicados em várias funções biológicas, como são a regulação epigenética e a imunidade.

Neste estudo, *B2M* que codifica a β2-microglobulina, uma proteina que compõe o complexo principal de histocompatibilidade de classe I (MHC-I), apresentou características típicas de gene supressor tumoral. Mutações em *B2M* atingiram 5% dos tumores de pulmão, sendo que a maioria impossibilitou a correcta formação do complexo. A perda do MHC-I (identificada num terço dos casos) relacionou-se com baixos níveis de infiltração intratumoral por linfócitos T citotóxicos, indicando possíveis impedimentos na erradicação de tumores mediada pelo sistema imune.

A tendência observada entre a deficiência de MHC-I e a redução da sobrevivência de pacientes tratados com inibidores de PD-1/PD-L1 denota o potencial uso da β2-microglobulina como biomarcador de resposta a estas imunoterapias em carcinomas de pulmão.

Prologue

Lung cancer is the leading cause of cancer-related mortality in the world. Despite the advances in standard treatments, the prognosis remains dismal. The arising of high-throughput sequencing revolutionized the cancer research and had a profound impact on the understanding of the molecular biology that underpins lung pathogenesis. The discovery of numerous cancer-related genes paved the way to the development of effective anticancer agents that are currently being administered to target the activity of mutated genes.

Therefore, the conceptual framework of this dissertation lays on genome-wide approaches to characterize lung tumors and find genes potentially relevant within the clinical context.

The current thesis describes the application of lung patient-derived xenografts as a reliable source of material to find genetic alterations using next-generation sequencing approaches. The purity degree achieved during the grafting process improved the detection of genetic abnormalities, being particularly valuable in uncovering homozygous deletions and gene amplifications.

Within the multitude of altered genes found in exome and transcriptome data, *B2M* (encoding the β2-microglobulin) was selected for in-depth investigations. Exploring the role of this gene in lung cancer pathogenesis constituted one of the central points of the present work. Mutations in *B2M* were identified in 5% of lung cancers and its inactivation prevented the correct assembly of the major histocompatibility complex class I (MHC-I). Alterations in other members of the antigen presentation machinery were also reported,

and are likely to result in a similar outcome. As a consequence, the complex was not able to reach the cell surface, which would probably hamper the tumor recognition and elimination by the $CD8^+$ T-cells. The loss of MHC-I, which constituted a common event in lung tumors, was significantly correlated to lower levels of cytotoxic T-cell intratumoral infiltration.

These findings denote that *B2M* exhibits tumor suppressive functions by allowing lung tumors to evade immune surveillance. Moreover, preliminary data demonstrated that high levels of MHC-I, in association with strong intratumoral immune infiltration, conferred a better response to anti-PD1/PD-L1 therapeutic drugs. Consequently, these observations illustrate the potential use of β2-microglobulin as a biomarker to predict treatment response to these immunotherapies in lung cancer patients.

Abbreviations

AD	Adenocarcinoma
APM	Antigen presentation molecule
β2m	β2-microglobulin
CDKN2A	Cyclin dependent kinase inhibitor 2A
CRC	Colorectal cancer
CTL	Cytotoxic T-lymphocyte
CTLA-4	Cytotoxic T lymphocyte-associated antigen 4
EGFR	Epidermal growth factor receptor
ER	Endoplasmic reticulum
G6PD	Glucose-6-phosphate dehydrogenase
GA	Gene amplification
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HD	Homozygous deletion
HGP	Human Genome Project
HPV	Human papillomavirus
IFNα	Interferon alpha
IFNγ	Interferon gamma
IHC	Immunohistochemistry
LC	Lung cancer
LCC	Large cell carcinoma
LFS	Li-Fraumeni syndrome

LOH	Loss of heterozygosity
mAb	Monoclonal antibody
MHC-I	Major histocompatibility complex class I
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NGS	Next-generation sequencing
NNK	Nicotine-derived nitrosamine ketone
NSCLC	Non-small cell lung cancer
PAH	Polycyclic aromatic hydrocarbon
PD-1	Programmed cell death 1
PD-L1	Programmed cell death ligand 1
PDX	Patient-derived xenograft
RCC	Renal cell carcinoma
SCC	Squamous cell carcinoma
SCLC	Small cell lung cancer
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variation
TCGA	The Cancer Genome Atlas
TK	Tyrosin kinase
TMA	Tissue microarray
TSG	Tumor suppressor gene
VEGF	Vascular epithelial growth factor
WES	Whole-exome sequencing
WHO	World Health Organization

Table of contents

1. INTRODUCTION

1.1. On the biology of cancer

Cancer is defined as a set of diseases sharing the common feature of uncontrolled cell growth, disregarding the molecular rules that regulate cell division. While healthy counterparts follow these instructions, cancer cells develop the autonomy to overcome the standard signals, and thus, become able to invade the surrounding tissues and spread to other parts of the body.

a) The clonal origin of human tumors

In 1976, Peter Nowell wrote a convincing article proposing that cancer was a dynamic process originated in a single abnormal cell. In

his publication, he postulated that tumor development was mutation-driven, progressing through a sequential series of accumulation and selection of genetic changes. Similarly to the Darwinian concept of evolution, from these successive rounds of mutation and selective expansion, clones harboring growth advantageous characteristics would emerge and proliferate, producing cell populations with higher aggressive phenotypes[1].

This revolutionary theory, now known as the *cancer clonal evolution model*, was supported by previous work on the X-linked polymorphic glucose-6-phosphate dehydrogenase (G6PD) locus in leiomyoma[2]. Due to the random inactivation of the X chromosome, women heterozygous for G6PD have two different cell populations, each one expressing one of the two forms – A or B. Each individual cell carries only one form, and consequently, healthy tissues are a mosaic. In contrast, leiomyomas were found to harbor either A or B phenotype, demonstrating the unicellular origin of the tumors (Figure 1).

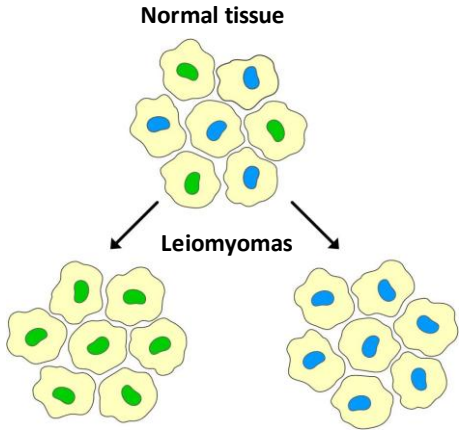


Figure 1. Cancer arises from a single cell. In heterozygous women, healthy tissues are a mosaic, in contrast to leiomyomas, which present only one form of G6PD.

b) Etiology

Cancer is the foremost cause of death in the world, killing about 8.2 million people every year[3], and the global burden is expected to rise in the future. With a complex epidemiology, this humankind plague emphasizes the population growth and aging, along with the increased adoption of risk behaviors and exposure to predisposing agents.

Longevity

Given that 60% of new cases affect individuals aged 65 or older[4], cancer has been repeatedly regarded as an age-related disease. Although the reasons are still unclear, three major hypotheses were postulated to explain the incremented incidence of cancer as people age. The first theory proposes that longevity provides cells time enough to accumulate alterations required to induce neoplastic lesions. In other words, elderly are more likely to get cancer due to more prolonged contact to carcinogens[5]. The second premise suggests that progressive shifts in the internal homeostasis (particularly those regarding the immune and the endocrine systems) offer a more suitable environment for tumor onset[6]. Finally, the third argument states that the typical cancer-prone phenotype of older individuals represents the cumulative effects of the mutational burden, increased epigenetic gene silencing, telomere dysfunction and changes in stromal environment[7].

Environment and lifestyle

Along with longevity, compelling epidemiologic data show that environmental exposure to carcinogens and lifestyle changes towards

unhealthy behaviors (poor diet, tobacco and alcohol consumption, obesity and lack of physical exercise) have profound effects in determining tumor development[3,8]. In fact, up to 90% of the most common cancers are caused by extrinsic processes[9] (external factors that influence mutation rates).

This undeniable correlation highlights that a large proportion of cases could actually be prevented. Examples include colorectal cancer, which is widely considered to be an environmental disease, with 75% of all sporadic cases being driven by unhealthy dietary habits[10]. Likewise, about 90% of melanomas are attributable to excessive exposure to solar radiation[11]. Tobacco is not only the utmost responsible for 85% of all lung tumors, but is strongly implicated in the development of several other cancers, like those from the upper aero-digestive tract, bladder, kidney, liver and pancreas as well[12].

Infection

It is also well established that pathogens substantially increase the risk of developing cancer. Estimations pinpoint that 17.8% of tumors worldwide are triggered by infectious agents[13].

In 2008, around 2 million newly diagnosed cancers were attributable to bacterial or viral infection, of which 90% were induced by the Gram-negative *Helicobacter pylori* bacteria, by the human papillomavirus (HPV) and by the hepatitis B (HBV) and C viruses (HCV)[14]. These pathogens have been largely implicated in gastric, cervix and liver cancers, respectively[15–17] (Figure 2).

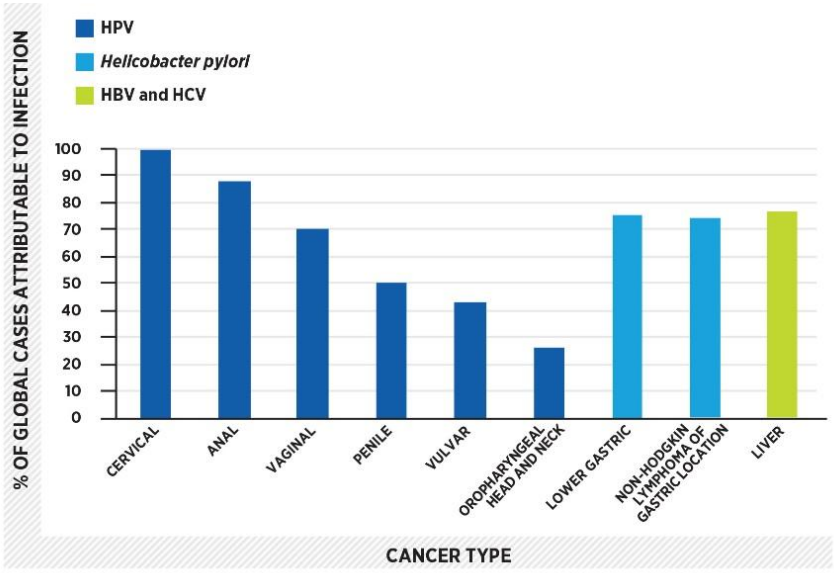


Figure 2. Infection by HPV, HBV and HCV viruses, together with *Helicobacter pylori* are the major cancer-caused pathogens. Each one of them is linked to a specific type of cancer (data from 2008). From AACR Cancer Progress Report 2015.

Inheritance

Hereditary cancer syndromes are thought to account for merely 5 to 10% of all cancers[18], and are usually inherited in an autosomal dominant fashion. Examples include Li-Fraumeni syndrome (LFS), initially reported in 1969 in families harboring soft-tissue sarcomas, breast cancer and other neoplasms in young children and adults[19]. The lifetime penetrance for the disorder is almost complete, with 80% of the individuals developing cancer at age 50[20]. Genetic predisposition is tightly linked to germline mutations in *TP53*, which occur in about 80% of the families meeting the standard criteria for LFS[21]. Similarly, inherited mutations in DNA mismatch repair genes (mainly *MLH1* and *MSH2*) are associated with Lynch syndrome, increasing the lifetime risk of carriers to develop colorectal cancer in up to 80%[22].

1.2. Principles of cancer genetics

The initiation and progression of cancer reflects the complex interaction between genetics and environmental agents, which act together to induce abnormal cell behavior and tumor development. Elementary speaking, despite the role of the environment in tumorigenesis, cancer arises by the accumulation of somatic mutations during lifetime (Figure 3). Most mutations found in cancer cells are called "passengers", often distributed in a random fashion across the genome and not contributing to the oncogenic process. Conversely, those conferring growth advantage to the cells constitute the subset of "drivers", and are causally implicated in cancer onset[23].

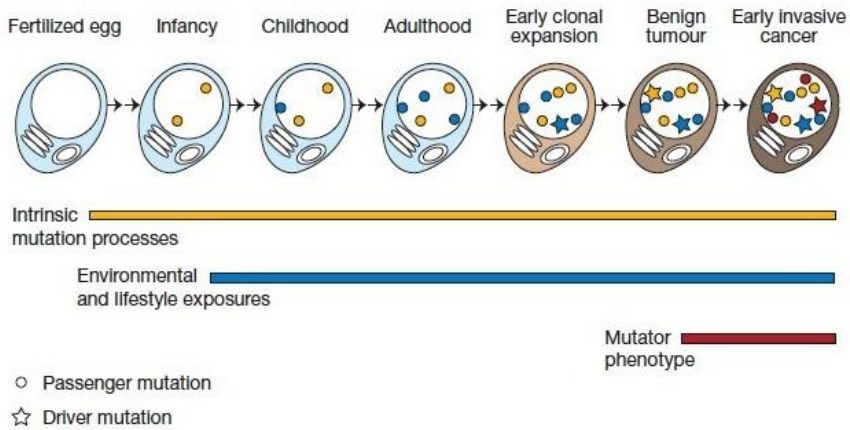


Figure 3. Genetic alterations in cancer cells are acquired throughout lifetime. Environmental and lifestyle habits largely contribute to the accumulation of both passenger and driver mutations, which will eventually result in cancer. The presence of mutator phenotypes caused by defects in the DNA repair machinery during the development of cancer increase the overall mutational burden. Modified from Stratton *et al.*, 2009[23].

a) Cancer genes

By definition, driver mutations are typically clustered in cancer genes, encoding proteins that act mainly as regulators of cell proliferation, cell differentiation and cell death[24]. Oncogenes and tumor suppressor genes (TSGs) represent two major categories of cancer genes; finding out who they are and how they promote tumorigenesis has been one of the toughest endeavors in cancer research.

"Turning on" oncogenes

Gain-of-function mutations convert proto-oncogenes into oncogenes. The malignant conversion requires the acquisition of activating point mutations, resulting in constitutively activated proteins. Most human tumors harbor single nucleotide substitutions in key regulatory genes, including those from the Ras family (*KRAS*, *HRAS* and *NRAS*) and *BRAF*[25,26]. Oncogene transformation can also occur by virtue of chromosomal rearrangement or gene amplification, leading to higher-than-normal levels of expression, as commonly observed in *ERBB2*, *BCR-ABL1* and *MYC* family genes (*MYC*, *MYCN* and *MYCL1*)[27,28]. Such genetic alterations are dominant by nature, and consequently, only a single mutated allele is sufficient to instigate carcinogenesis[29] (Figure 4).

Typically, oncogenes evade apoptotic signals to inhibit cell death, induce unrestrained cell proliferation by mimicking normal growth signals and surpassing cell cycle checkpoints, promote cellular migration and enable tumor cells to survive without substratum anchorage[29,30]. Oncogenes, mainly those coding for kinase proteins,

have been successfully used to develop targeted anticancer drugs to improve the treatment of various oncologic malignancies.

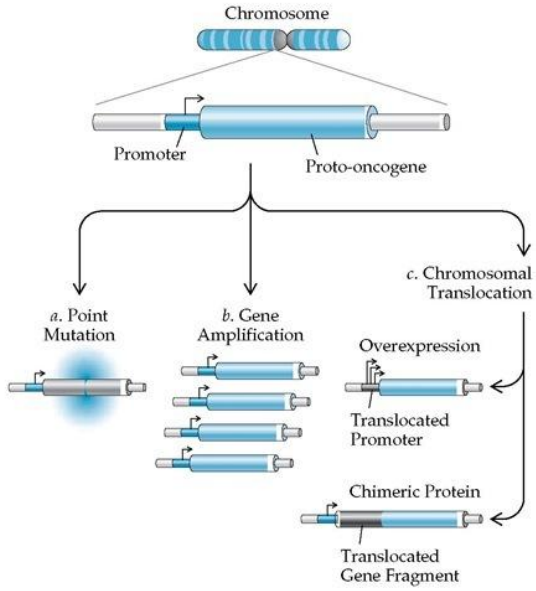


Figure 4. Molecular processes that convert proto-oncogenes into oncogenes. Oncogene activation may be triggered by somatic point mutation, gene amplification and chromosomal rearrangements. From ACP Medicine, 3rd edition, 2007.

Silencing tumor suppressors

TSGs encode checkpoint proteins that inhibit cell growth and maintain the cell cycle regulation system under control. Unlike oncogenes, the neoplastic properties of these cancer genes arise through loss-of-function mutations that lead to gene inactivation. Relevant TSGs include *TP53*, *RB1*, *PTEN*, *BRCA1* and *BRCA2*, which can be turned-off by several mechanisms (insertions, deletions and nonsense mutations), reducing gene activity or resulting in truncated proteins[31].

In general, TSGs are recessive at the cellular level, and thus, in order to get malignant physiological effects, the inactivation of both maternal and paternal alleles is required. This concept emerged from Knudson's analysis about the incidence of heritable retinoblastoma in young infants[32], which is on the basis of the "two-hit" model of tumorigenesis[33] (Figure 5).

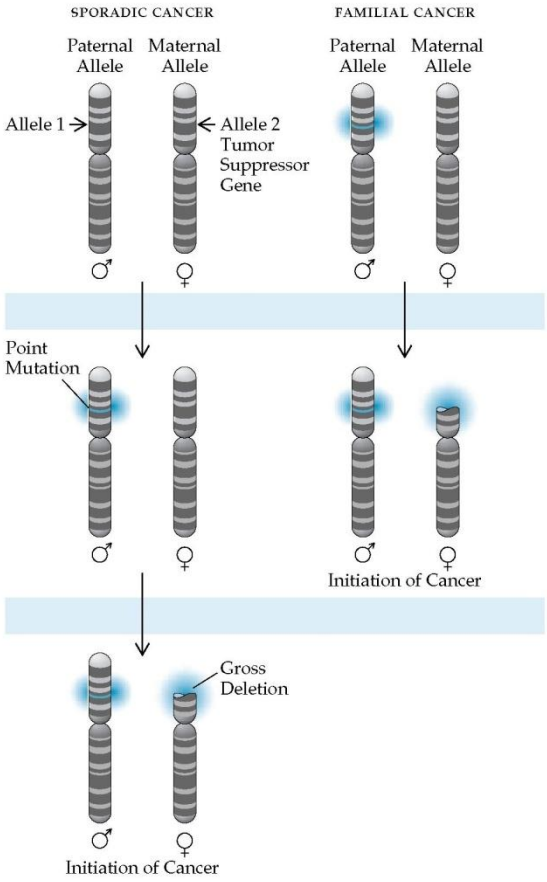


Figure 5. Knudson's two-hit model of tumorigenesis. The inactivation of both alleles of a TSG is required to trigger malignancy. While sporadic cancers need two independent somatic events, familial cancers arise from a single somatic mutation. Modified from ACP Medicine, 3rd edition, 2007.

b) From cancer genes to cancer genomes

Cancer pathogenesis relies on inherited genetic alterations and acquired somatic mutations. "*If we wish to learn more about cancer, we must now concentrate on the cellular genome. (…) The classification of the genes will facilitate the identification of those involved in progression.*"[34]. As foreseen by Dulbecco in 1986, the Human Genome Project (HGP) reached a milestone in 2003, when the first reference sequence of the human genome was released[35], providing a comprehensive scaffold to accurately infer an entirely new landscape of cancer-associated genes. Therefore, the knowledge of the basic mechanisms underlying cancer development accelerated considerably, and by 2004, an inventory of mutational data from the literature reported 291 genes as causally involved in oncogenesis[24].

The realm of next-generation sequencing

The arising of next-generation sequencing (NGS) technologies represents a tremendous paradigm shift that revolutionized the understanding of cancer biology. Before the launched of Roche 454, the first parallel sequencing platform to become commercially available in 2004, cancer research relied on the automated Sanger system. Such method was used in the HGP, a US$3.8 billion project that took an international consortium and lasted more than 10 years[35].

The first whole cancer genome was successfully sequenced using NGS in 2008[36] and by 2015, The Cancer Genome Atlas (TCGA) Research Network was gathering large-scale genomic data provided by 10,000 tumors from more than 25 cancer types, including a catalogue of 10

million mutations[37]. The widespread use of massive parallel sequencing strategies became a practical mode of generating unprecedented amounts of data in a fast and cost-effective way. Yet, the easiness of sequencing whole human genomes is still challenging in terms of bioinformatics analysis and interpretation of genetic variants.

Most cancer-causing mutations are exonic, and therefore, lay on protein-coding genes, which account for approximately 2% of the whole genome. Thus, exome-sequencing emerged as a faster and simpler alternative to provide new insights into tumor complexity and heterogeneity. In sum, NGS methods brought endless opportunities to discover active oncogenic pathways, aberrant proteins, tumor-specific mutations and genes that contribute to cancer progression and metastasis.

c) Cancer and its hallmarks

Defects in the regulatory circuits of cells caused by the accumulation of multiple DNA changes are the basis of cancer onset. In 2000, Hanahan and Weinberg proposed the "hallmarks of cancer", a unifying framework that illustrates the six common changes in cell physiology that collectively dictate the malignant transformation of normal cells[30]. In this model, tumors are seen as complex tissues, in which cancer cells interact and depend on their surrounding healthy counterparts, in order to survive and proliferate. The six features proposed to be virtually shared by most of human cancers are indicated next.

Self-sufficiency in growth signals

In order to move into an active proliferation state, normal cells require mitogenic growth signals to initiate cell division. However, malignant cells exhibit a reduced dependency on exogenous stimuli, by synthesizing stimulatory growth factors, like platelet-derived growth factor (PDGF) and transforming growth factor alpha (TGFα) on their own.

Insensitivity to anti-growth signals

Healthy tissues maintain homeostasis by means of anti-proliferative mechanisms; such inhibitory signals may force cells to be trapped into the quiescent phase or to induce a permanent post-mitotic state. Tumor cells are able to circumvent this barrier by disrupting important anti-proliferative pathways, of which the retinoblastoma protein (pRb) circuit represents a master example[38].

Limitless replicative potential

Unlike normal cells which can pass only through a finite number of duplications before getting in senescence, cancer cells achieve a state of immortality. Shutting down the intracellular program that controls cell division maintains functional telomerases, avoiding telomeres to become shorter.

Evading apoptosis

In addition to high proliferation levels, tumor cell population size is also dictated by the rate of cell death. The apoptotic machinery is constantly monitoring cell integrity, which can be triggered whenever

damage or malfunction is detected. By regulating anti- and pro-apoptotic factors, like members of the Bcl-2 protein family, tumor cells surpass these cellular circuits, avoiding apoptosis[39]. Other important players include the activation of *PIK3CA* and the inactivation of *PTEN*[40,41].

Sustained angiogenesis

On their way to malignant transformation, cells depend on blood vessels to ensure sufficient amounts of nutrients and oxygen. Angiogenesis, which is only transiently turned on in normal cells, is almost constitutively activated in tumors through an "angiogenic switch" developed by cancer cells and mediated by the release of vascular epithelial growth factor (VEGF)[42], fibroblast growth factors (FGF-1/2) and other pro-angiogenic elements[43].

Tissue invasion and metastasis

Turning on invasion-metastasis cascades allow tumors to migrate and survive in different environments, where nutrients and space are unrestricted. Modifications in the cell physical attachment to the microenvironment and the activation of extracellular proteases are strategies to promote invasion. This is accomplished by changes in the expression levels of molecules from several families, mainly immunoglobulins, cadherins and integrins.

The six aforementioned capabilities, represented in Figure 6, are critical to enable tumors to grow and metastasize. Yet, advances in cancer research underline other mechanistic processes that were not mentioned in Hanahan and Weinberg's first publication.

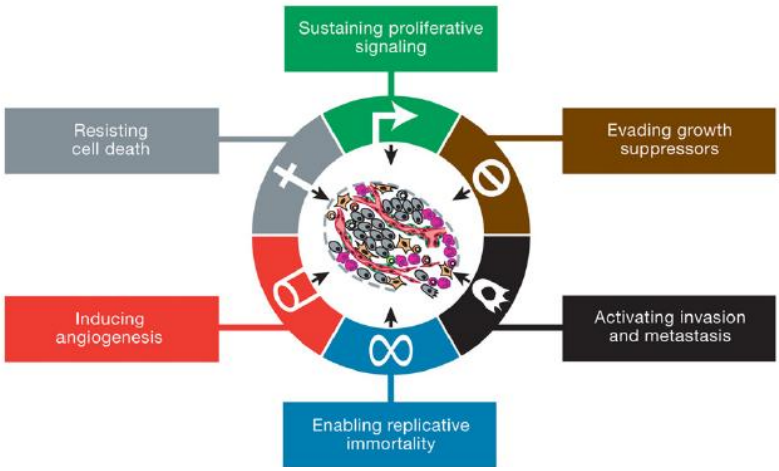


Figure 6. The hallmarks of cancer, as originally proposed by Hannahan and Weinberg in 2000. From Hannahan & Weinberg, 2011[30].

Thus, a decade later, the conceptual model was revisited and the original foundations were updated, adding two enabling characteristics pivotal to the carcinogenic process: genome instability and mutation and tumor-promoting inflammation[44]. Plus, two novel hallmarks (detailed below) were also incorporated into the framework.

Reprogramming energy metabolism

Cancer cells require huge amounts of energy to fuel their growth. However, the tumor microenvironment is frequently hypoxic and therefore, cells cannot rely on Krebs cycle for ATP production, which depends on normal oxygenation. Remarkably, tumor cells rather prefer the aerobic glycolysis state, even in the presence of oxygen. Although less efficient when compared to the mitochondrial oxidative phosphorylation, the glycolytic switch favors the synthesis of macromolecules and organelles necessary to generate new cancer cells.

Evading immune destruction

The immune system represents an effective barrier against cancer, by recognizing and eliminating transformed cells long before they can form tumor masses. Nonetheless, the constant selective pressure exerted by a competent immune system leads to the emergence of clones that are capable to avoid immune-mediated destruction, a process known as cancer immunoediting.

1.3. Immunology meets cancer

The primary task of the immune system is to protect the host against pathogenic invasion and tumors. To do so, an entire set of organs, cells and soluble components are organized in a fine-tuned network with the amazing ability to discriminate between "self", "altered-self" and "non-self".

The roots of cancer immunology lay in the hypothesis established by Paul Ehrlich in 1909[45], predicting that the immune system would play a significant role on scanning and eradicating transformed cells in the body. About 50 years later, Burnet introduced the concept of immune surveillance, which postulates that tumors – if they are clinically manifested – bypassed T-cell recognition by losing immunogenicity[46]. Nevertheless, the immune surveillance theory has long been immersed in controversy. Despite strong evidence supporting the effectiveness of the immune system in suppressing tumorigenesis[47,48], other studies have shown otherwise: it can also sculpt the emergence of primary tumors capable to escape from immune destruction[47,49]. The dual

opposite roles of immunity constitute the theoretical principle behind the cancer immunoediting model, while simultaneously incorporates and broadens the immunosurveillance concept.

a) The three E's of cancer immunoediting

Cancer immunoediting is a dynamic process, in which extrinsic immune pressure can act in two ways, either preventing tumor outgrowth or promoting neoplasia, by shaping tumor immunogenicity or restraining the host-protecting immune response. The process encompasses three distinct phases: elimination, equilibrium and escape[50].

During the elimination phase, the innate and the adaptive immunity work in synergy to detect and destroy any sign of neoplasia before its clinical manifestation. Cellular effectors such as T-lymphocytes and natural killer (NK) cells become activated, induced by inflammatory signals released by tumor cells and stromal components and consequently, malignant cells are destroyed and the host remains free of cancer.

However, if some variants survive, they may enter the equilibrium phase, in which the constant selective pressure from adaptive immunity maintains the tumor in a functional dormancy state. As an outcome of the Darwinian selection exerted on genetically unstable tumor cells, editing of tumor immunogenicity occurs and new variants emerge, carrying mutations that confer additional resistance to immune attacks.

Eventually, this outgrowth can no longer be blocked and surviving tumor cells that acquired insensitivity to immunologic elimination enter the escape phase and expand in an uncontrolled manner, resulting in clinically apparent disease[51,52] (Figure 7).

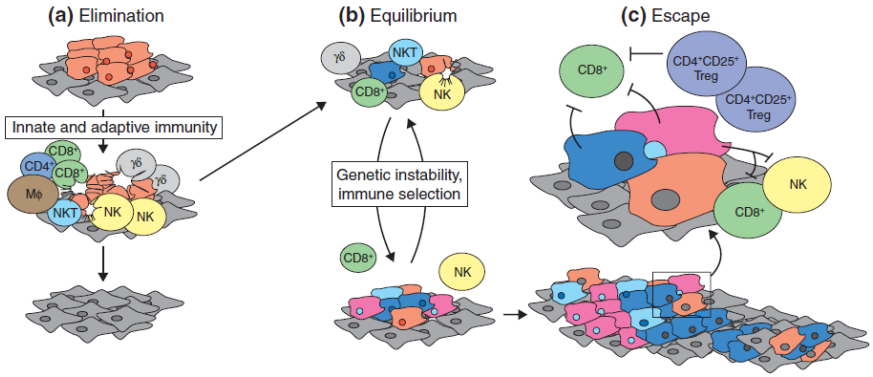


Figure 7. The cancer immunoediting process. (a) Members of the innate and the adaptive immunity may eliminate malignant cells. (b) If elimination is not achieved, cancer cells reach an equilibrium, in which genetic unstable clones may be selected. (c) Eventually, such clones can escape the action of the immune system and proliferate unchecked. From Strausberg, 2005[53].

b) MHC class I molecules and immune evasion

Malignant transformation goes along with increasing rates of somatic mutations, and thus, the generation of abnormal self proteins. The major histocompatibility complex class I (MHC-I) is present in virtually all nucleated cells of the body and has a specific task: to assemble and present intracellular peptides or antigens to the immune cells. Accordingly, transformed cells displaying aberrant antigens on the cell surface would eventually trigger their own destruction, mediated by $CD8^+$ cytotoxic T-lymphocytes (CTLs)[54,55]. Nevertheless, impaired antigen presentation through MHC-I loss or down-

regulation is frequently observed in human cancers, and represents an important strategy to evade immune destruction[56–59].

Garrido and colleagues classified the molecular alterations in MHC-I that ultimately hamper the correct presentation of tumor antigens in two main classes: reversible regulatory defects and irreversible structural defects (Figure 8)[60]. Reversible failures occur at the transcriptional level and may often be repaired by cytokine treatment, such as interferon gamma (IFNγ)[61]. *TAP1*, *TAP2*, *TAPBP*, *LMP2* and *LMP7* constitute some of the antigen presentation molecules (APMs) reported to be down-regulated in several cancer types, including colon, renal and breast cancer[62–65]. Some HLA-I specific locus (*HLA-A*, *-B* or *–C*) were also found to be weakly expressed in tissues derived from melanoma, breast carcinoma and bladder cancer[59,61,66]. Yet, even though HLA-I loss is one of the most widely described phenotypes, it is rather caused by irreversible damages, precluding the use of any immunomodulation treatment. Mutations and chromosomal defects underlie the total or partial absence (in case of some haplotypes or alleles) of HLA-I, which implicates not only HLA-I genes, but also *B2M*, encoding the β2-microglobulin (β2m). Loss of heterozygosity (LOH) at chromosome 6p21 containing *HLA-ABC* genes and at chromosome 15q21 encompassing *B2M* genomic location is frequently observed in oncologic malignancies, along with insertions, deletions and nonsense mutations[62,67–71]. Mutations in other APMs seem to be rare, assuming that regulatory shifts rather than genetic defects are on the origin of abnormal expression levels.

	Irreversible structural defects
Chromosome 6	**Loss of heterozygosity (LOH)**
	HLA haplotype loss
	Mutations of MHC class I heavy chain genes
	HLA allelic loss
Chromosome 15	**Loss of heterozygosity (LOH)**
	β2m gene
	Mutations/deletions
	β2m gene
IFN transduction pathway	Jack-STAT pathway blockade
	Reversible regulatory defects
Transcriptional downregulation	Coordinated downregulation of HLA A genes
	Coordinated downregulation of HLA B genes
	Coordinated downregulation of HLA C genes
	Coordinated downregulation of heavy chain, β2m and APM molecules
	Coordinated downregulation of antigen-presenting machinery (APM)
Hypermethylation	of MHC class I genes
Oncogene activation [c-myc, her-2/neu, oncogenic adenovirus 12 (Ad12)]	Downregulation of HLA-B genes
	Downregulation of HLA class I and APM genes
	Inhibition of the post-transcriptional processing of MHC class I mRNA

Figure 8. Alterations in MHC-I expression levels can be caused by genetic, epigenetic or transcriptional abnormalities, which represent either reversible regulatory defects or structural changes that cannot be reversed. Modified from Garrido *et al.*, 2010[60].

c) Immunotherapy to defeat cancer

The immune system is designed to perceive and eliminate anomalous cells. The genetic abnormalities that accompany malignant transformation lead to the expression of atypical antigens, differentiating tumor cells from their normal counterparts. This concept constitutes the foundation of anticancer immunotherapy, designed to boost or restore the host's immune function to kill abnormal cells, while keeping healthy cells unaffected[72]. In the last three decades, several immunotherapeutic strategies were developed to be employed in treating cancer (Figure 9). Below, the most well-known regimens are addressed in detail.

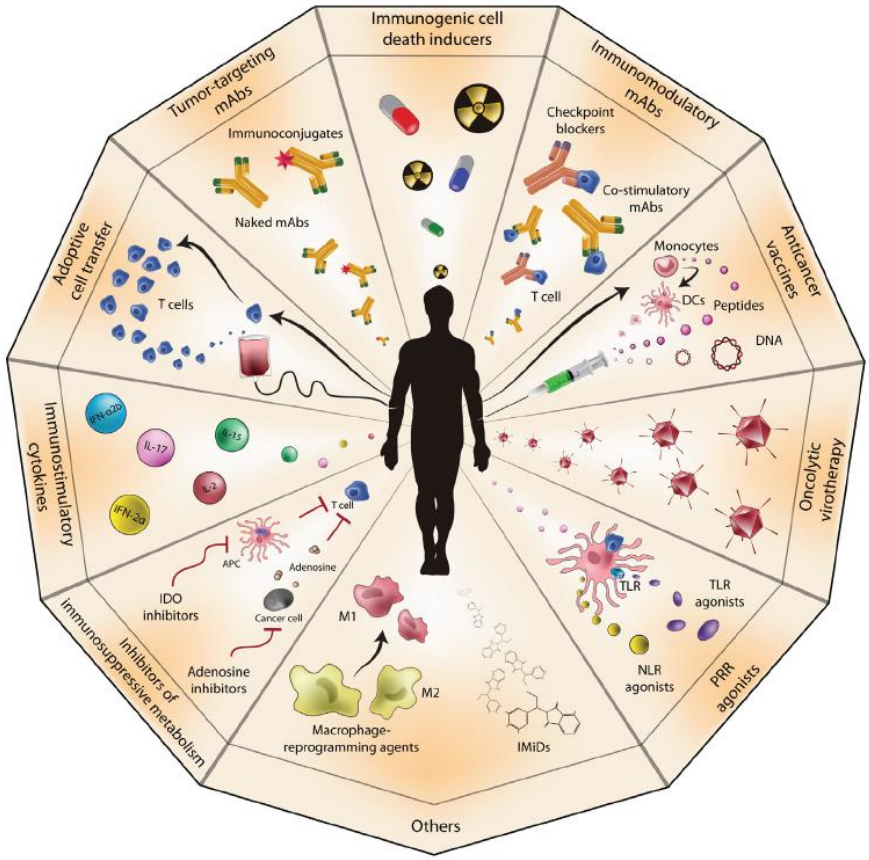


Figure 9. Representation of the immunotherapeutic strategies that were developed in the last decades to treat oncologic malignancies. From Galluzzi *et al.,* 2014[73].

Cancer vaccines

Estimations indicate that 10 to 15% of human cancers are viral-induced[74]. Preventive vaccines with well-defined viral epitopes such as those targeting serotypes of the HPV demonstrated protection against anogenital diseases, and were likely to do so against HPV-associated cervical cancer[75,76]. Apart from viral, other tumor antigens generated by self proteins can also elicit specific immune responses, like tissue differentiation antigens (Melan-A and CEA) and cancer/testis antigens

(MAGE and NY-ESO-1)[77]. Despite most cancer vaccines offered little or no effectiveness in inducing clinical response, Sipuleucel-T was the first cell-based vaccine approved for administration for its efficiency in prolonging overall survival in patients with metastatic prostate cancer[78].

Monoclonal antibodies

Currently, the concept of monoclonal antibody (mAb) is at the vanguard of anticancer immunotherapy, holding vast potential for its high efficacy and specificity. In this regard, rituximab, a chimeric anti-CD20 antibody which promptly activates the complement system, showed significant benefits against chronic lymphocytic leukemia and non-Hodgkin lymphoma[79,80]. For solid tumors, cetuximab is currently being administered to improve the treatment of head and neck carcinoma and colorectal cancer (CRC), through the blockade of EGFR signaling pathway. Likewise, the anticancer mAb trastuzumab induced responses in 25 to 30% of breast cancer patients over-expressing the HER2 protein[81]. The VEGF inhibitor bevacizumab was the first anticancer mAb to be approved that did not targeted tumor cells specifically. However, its large anti-angiogenic activity was observed in multiple cancer types, including CRC, glioblastoma multiforme, cervical carcinoma, renal cell carcinoma (RCC) and lung cancer (LC)[73].

Modulators of immune checkpoints

Checkpoint-blocking mAbs represent an established and the most promising paradigm of active immunotherapy, designed to (re)-

activate the host immune system to fight against malignant cells. The remarkable success of checkpoint molecules has been reflected by long-lasting clinical benefits accomplished in cohorts of patients suffering from a variety of solid tumors (reviewed by Melero *et al.*, 2013[82] and by Pardoll, 2012[83]). Ipilimumab, a mAb targeting the cytotoxic T lymphocyte-associated antigen 4 (CTLA-4), was the first to prolong survival in a randomized trial of metastatic melanoma, in which 18% of the ipilimumab-treated patients survived beyond two years (versus 5% receiving gp100 peptide vaccine alone)[84]. This breakthrough accelerated its final approval in 2011, and ever since, ipilimumab has been administrated in individuals with unresectable or advanced melanoma. Blocking the interaction of the programmed cell death 1 (PD-1) expressed by activated T-cells with its ligand programmed cell death ligand 1 (PD-L1) in tumor cells was found to induce tumor regression in large proportions of patients with non-small cell lung cancer (NSCLC), melanoma, bladder cancer and RCC[85,86]. As a consequence, followed on the heels of anti-CTLA-4, the PD-1/PD-L1-pathway blockers nivolumab and pembrolizumab were both released in 2014 for treating stage IV melanoma patients. In 2015, nivolumab was also licensed to be used in patients with advanced NSCLC, based on the robust results provided by a phase III clinical trial, reporting improvements on the overall survival, comparing with docetaxel treatment alone[87]. Similarly, the consecutive success of pembrolizumab in decreasing tumor size across different clinical trials (reviewed in Bagley *et al.*, 2015[88]) resulted in its approval for NSCLC treatment that same year.

1.4. Lung cancer

Lung carcinomas arise as a consequence of progressive pathological changes in the epithelial cells lining the bronchi, bronchioles and alveoli. Throughout their development, tumors acquire the ability to spread to other parts of the respiratory system and beyond, invading distant organs and tissues throughout the body.

a) General statistics

LC is the most frequently diagnosed and represents the leading cause of cancer-related deaths worldwide, killing 1.4 million people every year[89]. The incidence rate is higher in developed countries, particularly in North America and Europe, and lower in developing nations of Africa and South-America[90], as displayed in Figure 10. Globally, men are diagnosed more often than women (1,241,600 new annual cases versus 583,100)[3].

Differences in geographic areas and genders are generally attributed to trends in population prevalence of tobacco consumption. Despite the therapeutic advances accomplished during the last decades, the 5-year relative survival is 17%, being one of the cancers with worst outcomes. Nonetheless, survival rates are largely dependent on the stage at diagnosis, ranging from 55% for localized tumors (within the lungs) to 4% for metastatic conditions. The poor prognosis is also shaped by the absence of signs and symptoms, and hence, merely 16% of all LC cases are prone to be prematurely detected[91].

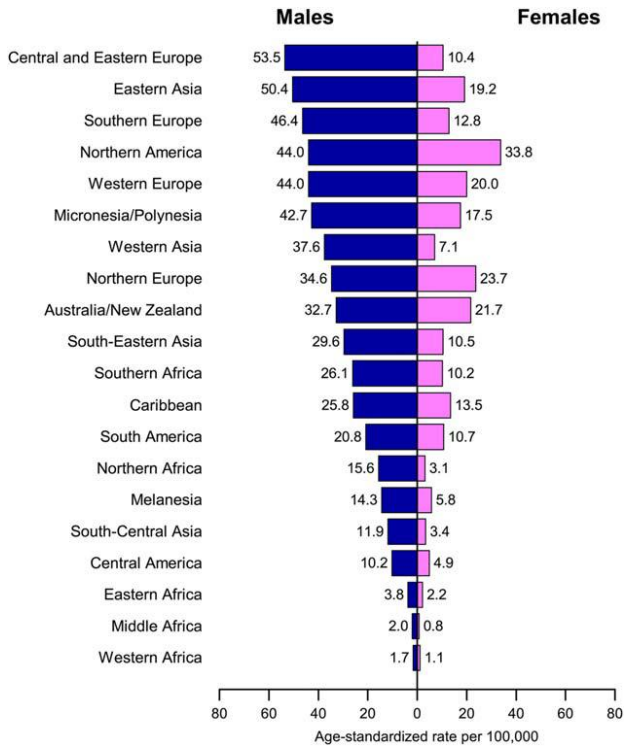


Figure 10. Age standardized LC incidence rates among men and women by world region. From Jemal *et al.*, 2011[90].

b) The prime cause: tobacco

Given that tobacco is by far the supreme responsible for tumor onset, LC is constantly labeled as one of the most preventable oncologic diseases. Actually, 85% of all LC cases are directly attributed to cigarette smoking[92,93], and the risk increases with the duration and intensity of smoking, initiation age and the amount of packs smoked per year. The link between the smoking status and the histological type of lung cancer is particularly strong in squamous cell carcinoma (SCC) and small cell lung cancer (SCLC), being less evident in lung adenocarcinoma (AD)[92].

According to the American Cancer Society (www.cancer.org), tobacco smoke is made up of more than 7,000 chemical compounds, of which more than 70 are potent carcinogens. Polycyclic aromatic hydrocarbons (PAHs) are a class of such mutagens that confer a specific molecular signature in smoking-associated LCs, inducing G to T transversions in key cellular regulators like *TP53*[94,95]. Similarly, tobacco-specific N-nitrosamines correspond to another major carcinogenic group, of which the nicotine-derived nitrosamine ketone (NNK) is of particular concern. In normal conditions, every time smoking-associated carcinogens bind the DNA and create adducts, the DNA repair machinery is able to remove them or alternatively, cells undergo apoptosis. However, failures in repair mechanisms can drive to permanent mutations in oncogenes and TSGs that eventually end up in lung tumorigenesis (Figure 11)[96].

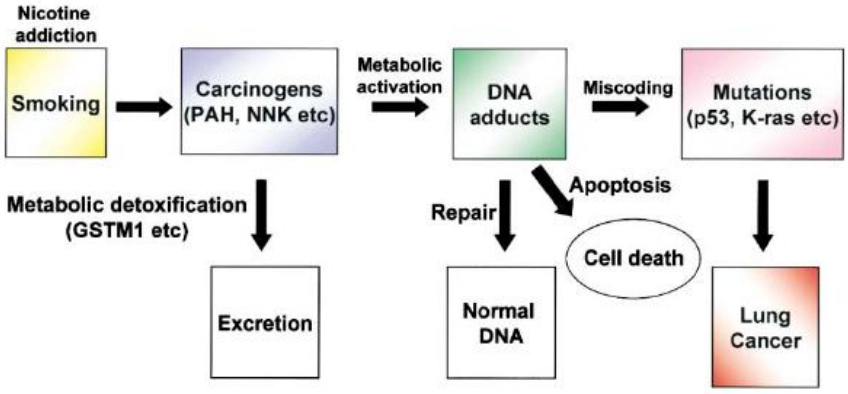


Figure 11. The pathogenesis of smoking-driven lung carcinomas. Carcinogens in tobacco smoke can be either excreted or can bind the DNA and form adducts. If such adducts persist, mutations arise and cancer emerges. From Minna *et al.*, 2002[97].

c) Other risk factors

Even though the magnitude of tobacco consumption far outweighs the consequences of any other risk factor, the remaining 15% of LC cases arising in never smokers must be caused by different components[98]. As an illustration, 77% of women living in some parts of Asia who are lifelong never smokers ultimately develop LC[99].

The exposure to asbestos, arsenic and radon results in genetic alterations, as they are able to generate reactive oxygen species and to disrupt critical cell signaling pathways. Changes in microRNA expression and DNA methylation patterns, as well as modifications at the histone level represent other important means by which such carcinogens promote the development of lung tumors[100]. Moreover, estimations point out passive smoking (by both exhaled and sidestream smoke) as the responsible for 17% of LC cases in never smokers[101]. In fact, after exposing five non-smoker males to sidestream cigarette smoke produced by machine smoking, Hecht and coworkers were able to detect several carcinogenic metabolites in their urine[102].

d) Histological classification

The correct histological classification of lung tumors is imperative in clinical practice, as it provides valuable information that will shape diagnosis, prognosis and treatment decisions. The malignant transformation occurring in the lungs can lead to a variety of tumor types, reflecting the diversity on cell populations that compose the respiratory tissues.

The major guidelines to categorize lung tumors were released by the World Health Organization (WHO) in 2004[103], which were followed by several updates that incorporated additional changes, translating the scientific advances in LC genetics and personalized therapies. Thus, these tumors are typically classified in four major histological types: adenocarcinomas, squamous cell carcinomas and large cell carcinomas (LCC) – collectively referred to as NSCLC – and small cell lung carcinomas. NSCLCs comprises the majority of LC cases, and account for approximately 85% of all lung tumors[104].

Adenocarcinoma

AD is the most commonly diagnosed histological type (about 39% of NSCLC cases[105]) and is predominant in individuals who never smoked, particularly women. The development of ADs take place in cells harboring glandular or secretory properties and present acinar, papillary, bronchioalveolar or solid with mucin growth patterns[103]. Up to 75% of pulmonary ADs express biomarkers consistent with an origin in the distal region of the lung, such as the thyroid transcription factor 1 (TTF-1)[106].

Squamous cell carcinoma

Accounting for 20% of NSCLC cases[105], SCC arises from bronchial epithelial cells and shows keratinization, pearl formation or intercellular bridges, depending on the differentiation degree. The huge correlation with cigarette smoking translates the fact that over 90% of the cases are diagnosed in smokers[103]. SCCs can be

distinguished from other histological subtypes by the expression of the squamous marker, p63[107].

Large cell carcinoma

LCC is the less represented subtype, summing up only 3% of NSCLC cases[105]. Most LCCs lack specific discriminating cytological features, presented in cellular aggregates with indistinct cellular boarders and irregular chromatin distribution. Being poorly differentiated tumors by definition, the diagnosis of LCC is based on exclusion, after ruling out that cells do not appear morphologically glandular or squamous and do not express AD or SCC biomarkers[103].

Small cell lung carcinoma

SCLCs account for 15% of all LC cases[105] and constitute the most aggressive form of the disease, having greater potential to metastasize than any other histological type. Holding neuroendocrine features, SCLC is composed by small round-shaped cells with narrow cytoplasm, poorly defined borders and finely granular nuclear chromatin[103]. Extensive necrosis is frequently observed, as well as high mitotic rates.

e) Mutational spectrum of lung tumors

Similarly to other cancer types, tumors of the lung arise from the sequential accumulation of genetic mutations throughout life. As a consequence, the balance between proliferation and quiescence is disrupted, leading to the formation of tumor masses[30]. Even though

cigarette smoking is the utmost responsible for LC, only 10% of smokers develop the disease[108], suggesting that the genetic background has a crucial role in carcinogenesis.

The genetic diversity is immense, and due to the chronic mutagenic burden of tobacco, LC genomes exhibit the highest number of somatic mutations, only surpassed by melanoma[109]. While some genes were already known to be mutated before the release of the human genome, technological advances during the last decade, especially after the arising of NGS platforms, have bolstered the annotation of altered genes and cellular pathways that induce and sustain tumorigenesis (Figure 12).

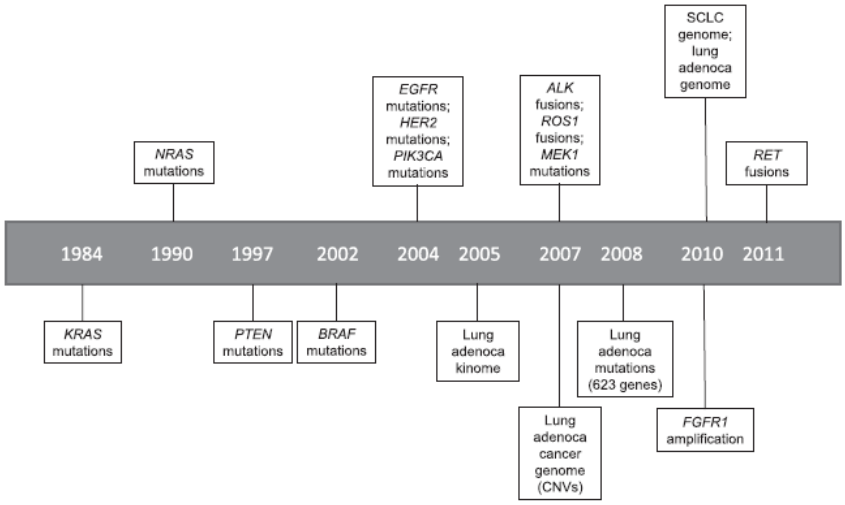


Figure 12. Timeline of the significant genetic alterations discovered in lung tumors. The advent of NGS technologies led to a profound acceleration regarding the knowledge about the molecular features of LCs. From Levy *et al.*, 2012[110].

Thanks to high-throughput methods, the genomic landscape of thousands of lung carcinomas encompassing the major subtypes was unraveled[111–113]. This milestone defined an important paradigm change, since the traditional characterization of LC relied merely on the

histological features. Today, lung carcinomas are seen as well-defined molecular entities, harboring mutations and expression patterns that are vital for diagnosis and prognosis purposes, empowering treatment decisions towards a more personalized therapy (Figure 13).

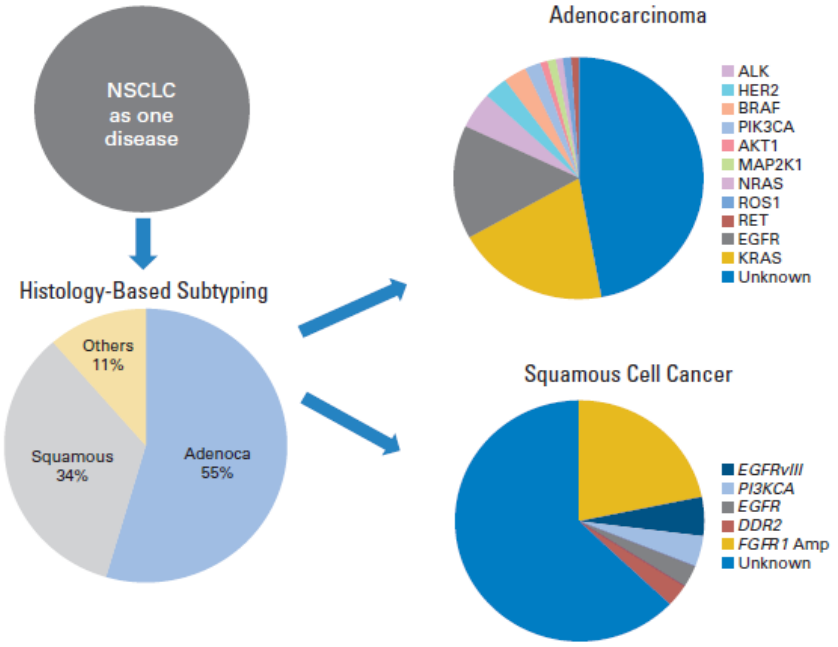


Figure 13. The conventional classification of lung tumors based on histological features was to some extent replaced by a molecular view, with a profound impact in clinics. From Li *et al.*, 2013[114].

Below, the most notable genes and biological pathways altered in lung carcinomas are addressed in detail.

EGFR gene family members

The epidermal growth factor receptor (*EGFR*) belongs to the ErbB family of surface receptor tyrosin kinase (TK), together with *HER2*, *HER3* and *HER4*. Mutations in *EGFR* are observed in 10 to 40% of NSCLCs, depending on the ethnic background, being particularly

associated with ADs (over 50%) presenting bronchioloalveolar features from women with no history of smoking[115,116].

Signal transduction stimulated by extracellular growth factors leads to cell proliferation, angiogenesis, invasion and metastasis, through the hyper-activation of downstream PI3K/AKT/mTOR, JAK/STAT and RAS/RAF/MEK/ERK signaling pathways[117]. In 80% of the cases, *EGFR* activating mutations target exons 18 to 21 (small in-frame deletions and point mutations), encoding part of the TK domain and are clustered next to the ATP-binding pocket[115]. Although to a lesser extent, over-expression by virtue of gene amplification and copy number alterations are also reported in LC patients. A great number of mutations in the gene increase the sensitivity and the clinical responses to targeted agents, which are now commonly administered in clinics[116,118]. Available drugs include EGFR TK inhibitors (erlotinib and gefitinib) and the anti-EGFR monoclonal antibody cetuximab. Genetic alterations in *HER2* account for 4% of NSCLC patients, and seem to be exclusive of ADs (present in about 10%), with mutations laying mostly on the TK domain[119].

RAS/RAF/MEK pathway

The RAS/RAF/MEK cascade plays a key role in lung carcinogenesis, with at least one mutation in the pathway identified in 70% of lung ADs[120], of which the most affected gene is *KRAS*. Actually, mutations in *KRAS* constitute one of the commonest oncogenic alterations in such histological subtype, totalizing 25 to 30% of the cases, and presenting a strong association with cigarette smoking. Typically, lung ADs harbor a single amino acid substitution located in codon 12,

which tend to be mutually exclusive with *EGFR* mutations. As a consequence, patients harboring *KRAS*-driven lung tumors are resistant to TK inhibitors[121,122]. Downstream in the pathway, mutant *BRAF* was also shown to be implicated in the pathogenesis of NSCLC, indentified in near 5% of the ADs, and rarely on SCCs. V600E hotspot mutation is the most recurrently found (57% of the cases) and significantly correlates with poor prognosis[123]. The BRAF-specific inhibitor vemurafenib is approved for clinical use in V600E-mediated metastatic melanoma and may constitute a promising drug to target NSCLC in the near future.

PI3K/AKT/mTOR pathway

PI3Ks are lipid kinases involved in the regulation of cell growth, proliferation and apoptosis. The pathway becomes activated through a variety of membrane TK receptors, such as EGFR, insulin-like growth factor 1 (IGF-1) receptor and VEGF receptor[124]. Aberrant signaling through mutations in *PIK3CA*, encoding the catalytic subunit p110α, occurs in less than 3% of NSCLCs, with higher prevalence in the SCC subtype. Such alterations affect the catalytic domain and unlike most oncogenic driver mutations, they are found concomitantly with *EGFR* and *KRAS*[125]. Although rare (1.9% in NSCLCs), mutations in the downstream gene *AKT* constitute another mechanism to dysregulate the pathway[126]. *STK11* (also known as *LKB1*) is a serine-threonine kinase that acts as a TSG in NSCLC by inhibiting mTOR. Genetic loss is reported in a very high proportion (30% of LCs, particularly in ADs from smokers), and arise via somatic mutations or deletions[127]. Strikingly, *STK11* alterations emerge concurrently with other driver genes, such as *KRAS*, *TP53* and *SMARCA4*[128]. Another relevant TSG

with the ability of inhibiting mTOR is *PTEN*, encoding a lipid and protein phosphatase. Alterations in *PTEN* are relatively common in LC (around 5%), mainly induced by LOH, and are particularly associated with SCCs derived from ever smokers[129].

TP53 and p16^{INK4a}/RB pathway

TP53 encodes the p53 protein, a master regulator that acts in cell cycle arrest or in inducing apoptosis upon DNA damage or carcinogenic stress. *TP53* genetic loss constitutes the most significant event in LC, emerging virtually in all SCLCs[113] and occurring in 81% of SCCs and to a lesser extent in ADs (50%)[130]. Mutations strike mostly the DNA-binding domain, but the spectrum is distinct, depending on the smoking background, with a significant higher frequency of G to T transversions in smokers, induced by PAHs[131].

The p16^{INK4a}/RB pathway regulates progression from G1 to S phase of the cell cycle. *RB* was the first gene described to be mutated in LC, and constitutes an universal event in SCLC malignancies[113]. In NSCLC, the pathway is frequently switched off by cyclin dependent kinase inhibitor 2A (*CDKN2A*), a gene altered in 60% of NSCLCs[132]. As foreseeable, mutations in *RB* and *CDKN2A* are mutually exclusive.

Oncogenic fusions

Rearrangements of *ALK*, a membrane TK receptor, are usually associated with *EML4* and occur in 7% of LC cases[133]. Such oncogenic fusion proteins hold a constitutively activated kinase and emerge through small inversions at chromosome 2. Depending on the breakpoint in *EML4*, multiple variants have been described. Still, even

though to a lesser degree, *ALK* can also combine itself with other partners[134]. Gene translocations are preferentially found in light- or never-smokers of young age[135]. Surprisingly, *ALK-EML4* positive lung tumors are sensitive to crizotinib, a drug initially developed as a MET inhibitor, which demonstrated to improve survival[136]. Similarly to *ALK*, *ROS1* is another TK receptor that is translocated in a small subset of LC cases (less than 2%), specifically in ADs from young never-smokers[137]. Likewise, *ROS1*-rearranged NSCLC patients also had a confirmed response to the TK inhibitor, crizotinib[138].

2. OBJECTIVES

Lung cancer is the deadliest form of cancer worldwide and the limited effectiveness of the conventional treatments represents a tremendous challenge to overcome its poor prognosis.

In recent years, the large-scale genomic profiling of lung tumors has fueled the development of efficient anticancer agents that are currently being administered to target the activity of altered genes. Yet, directed therapies are scarce and can only benefit a small subset of patients. Discovering genes that may act as biomarkers to improve therapeutic responses or to empower treatment decisions is, thus, an imperative issue.

Therefore, the central goal of this thesis is to investigate novel genes implicated in pulmonary pathogenesis that may be helpful within the clinical context.

In particular:

- Discovering mutated genes in lung carcinomas using high-throughput sequencing approaches over tumorgraft models.
- Determining the type and mutation rate of such alterations across lung tumors and possible correlations with clinical and histopathological parameters.
- Restoring the expression of the selected genes in gene-deficient LC cell lines as a way to establish models to study their implication in carcinogenesis.
- Investigating the effects of gene restitution in cancer-related processes, while exploring the molecular pathways in which they are directly involved.
- Addressing their potential role as molecular markers under the clinical scope.

3. MATERIAL AND METHODS

3.1. Cell lines and tumor samples

a) Cell lines

Most cell lines used in this study were purchased from the American Type Culture Collection (Rockville, MD, USA). Cells were maintained at 37ºC in a 5% CO_2 humid atmosphere and grown in the culture medium recommended by the manufacturer, supplemented with 1% penicillin-streptomycin solution and heat-inactivated fetal bovine serum (for details, see Table 1). For subcultivation purposes, cells growing in monolayer were detached from the surface using a Trypsin-EDTA solution. Before every experiment, cell lines were

tested to discard any possible contamination by *Mycoplasma* and were often authenticated by genotyping for specific mutations (*TP53*, *SMARCA4*, *STK11* or any other relevant gene) through Sanger sequencing.

Table 1. Histology and growth conditions of the main LC cell lines used in this study.

Cell line	Histology	Growth	Culture medium
H2009	AD	Adherent	HITES + 5% FBS
H2135	NSCLC – n/a	Mixed	HITES + 5% FBS
H2342	AD	Adherent	HITES + 5% FBS

AD: adenocarcinoma; NSCLC: non-small cell lung cancer; n/a: not available; FBS: fetal bovine serum.

b) Tumor samples

Tumor specimens were collected from patients with NSCLC at the time of surgical resection. Tumors and the corresponding healthy samples were obtained from patients with appropriate consent from each Institutional Review Board. Normal tissue consisted of blood components or adjacent non-tumorigenic lung tissue located at least 2cm away from the tumor. Available histological sections were reviewed and classified according to WHO criteria.

Freshly frozen or paraffin-embedded tumors used for genetic screening and immunohistochemistry (IHC) were acquired from the Tumor Bank of the Spanish National Cancer Research Center (CNIO, Madrid, Spain) and from the Center for Applied Medical Research (CIMA, Navarra, Spain). Biological samples used to generate each patient-derived xenograft (PDX) and matched normal tissues were obtained from the HUB-ICO-IDIBELL Biobank (Barcelona, Spain)

(PDXs 1 to 8), the IRCCS Foundation, National Cancer Institute (Turin, Italy) (PDXs 9 to 12) and the La Ribera University Hospital (Valencia, Spain) (PDXs 13 and 14). Samples were extracted from patients who received no prior treatment. Paraffin-embedded tumors from patients treated with anti-PD-1 or anti-PD-L1 mAbs were obtained from the Vall d'Hebron University Hospital (Barcelona, Spain).

3.2. Generation of lung tumorgrafts

PDX models were generated in athymic mice male *nu/nu* aged 4-5 weeks (Harlan-Laboratories Inc., Indianapolis, IN, USA), which were maintained in a sterile environment. The design of experiments and protocols were approved by the Institutional Animal Care and Use Committee. Shortly after surgical resection, tumors were suspended in standard culture media, dissociated in similar pieces of 2-4mm^3 with a scalpel, and subsequently engrafted in the back of isoflurane-anesthetized mice through skin incision. Some of the PDXs were orthotopically implanted, as described elsewhere[139]. Briefly, mice were subjected to thoracotomy and tumor fragments were placed between the second and the third lung lobes. Once grown, specimens were harvested and cut into small pieces to be transplanted for multiple serial passages. Engraftment samples were cryopreserved before each transplantation series for subsequent applications, namely for whole-exome and transcriptome sequencing.

3.3. High-throughput sequencing

a) Nucleic acid preparation

To extract DNA, PDX samples and their corresponding non-tumorigenic tissue were kept overnight at 58ºC under a 1% SDS-proteinase K (10mg/ml) solution. Digested tissues were subjected to phenol-chloroform extraction and ethanol precipitation, following standard protocols. DNA quality was checked by visualization in agarose and quantification was assessed by NanoDrop and PicoGreen.

RNA isolation was performed using the TRIzol® Plus RNA Purification Kit (Purelink RNA Mini Kit, Thermo Fisher Scientific, Waltham, MA, USA). RNA quantification and integrity (RIN) was evaluated on the Agilent 2100 Bioanalyzer.

b) Whole-exome and transcriptome sequencing

The SureSelect Kit (Agilent Technologies, Santa Clara, CA, USA) and the TruSeq RNA Library Preparation Kit (Illumina, San Diego, CA) were used for exome capture and RNA library preparation, respectively. Tumorgrafts and normal samples underwent sequencing on HiSeq Illumina Analyzers (Illumina) using the paired-end 2x100bp read option for whole-exome sequencing (WES) and the 2x75 option for RNA-seq. Both analysis were carried out at the Spanish National Genome Analysis Center (CNAG, Barcelona, Spain). Data were analyzed at the Spanish National Cancer Centre (CNIO, Madrid, Spain) and at the CNAG. Standard pipelines were applied for

trimming and assembling sequencing reads, before being mapped against the GRCh37/hg19 version of the human genome. Mouse-derived reads were discarded using the GEM-mapper pipeline, contrasting the data with the mouse mm10 reference genome.

c) Bioinformatics analysis

The detection of PDX-specific structural and somatic modifications was achieved by comparing the unique variants of each normal-tumor pair. Single nucleotide variants (SNVs) and short indels were called using the SAMtools (version 0.1.19) package with default settings, once the following cutoffs were applied to filter out variants before generating the final lists: i) genomic regions with low mappability; ii) read depth >10 at least in one sample; iii) strand bias $P<0.001$ and tail distance bias $P<0.05$ at least in one sample. Variant annotation was done using dbSNP (version 137) and dbNSFP (version 1.3light) databases and effect predictions were carried out with snpEff. Genes were scored and ranked by functional categories, based on the corresponding somatic mutation type.

Finding regions holding gene amplifications (GAs) and homozygous deletions (HDs) was done by calculating the normalized number of reads for each exon and establishing certain thresholds. Thus, to select GAs, the ratio of the normalized reads in tumor *versus* normal matched DNA had to be greater than 15. Likewise, for homozygous deleted regions, that ratio had to be less than 0.05 (20-fold) in at least two consecutive exons.

For RNA-seq data analysis, the Multicov command from Bedtools v2.17.0 was used to obtain raw read counts and the number of reads per kilobase per million mapped reads (RPKM) was calculated for every exon. The integration of the copy number data from WES with the gene expression data from RNA-seq empowered and validated the GAs. Accordingly, as expected, selected GAs resulted in gene over-expression (defined as a 10-fold increase in the expression levels on the target tumorgraft, with respect to the mean expression of the remainder tumors).

Finally, fusion transcripts were found applying TOPHAT 2.0.10 and BOWTIE 2.0.6 algorithms with default settings.

3.4. Searching for lung cancer-related genes

a) Selecting candidate oncogenes

To detect putative oncogenes, genes affected by strong and focal GA were selected. Moreover, those carrying missense and in-frame indel mutations (with consequent mRNA expression) were also included in the final list. Additional criteria were taken into account, electing only those presenting mutational clusters in cancers (meaning about 20% of the reported missense mutations or in-frame indels were placed in hotspots[140]) or that the same mutation found in the present work was recurrent in at least two other tumors. All the filtering was performed according to the data gathered in cBioPortal[141].

b) Selecting candidate tumor suppressors

For TSG detection, only genes harboring homozygous frameshift indels, nonsense mutations and HDs were selected. The criteria followed consisted, with few exceptions, that at least 10% of the reported mutations for a given candidate were truncating. Data was compared with the one available in COSMIC[142,143] and cBioPortal[141].

3.5. Loss of heterozygosity

To test for LOH in tumor samples, four previously described microsatellite markers distributed upstream and downstream of *B2M* were examined: D15S214, D15S641, D15S659 and D15S123. Similarly, assessing tumor cell purity of PDX samples was done using microsatellites D19S886 y D19S221, which enclose *STK11* and *SMARCA4*, respectively. Fluorescent-labeled primers were used to amplify DNA fragments from tumors and normal tissues (Table 2).

Table 2. Microsatellite markers employed to test for LOH in tumor samples.

Marker	Motif	Sequence (5' - 3')	Dye	Expected size (bp)
D15S214	(CA)n	F: GGAGGGCACTTCCTGAG	FAM	264
		R: GCCTGGCATCACGACT		
D15S641	(TAA)n	F: AACAAAGGGAGACCCTCATC	FAM	105
		R: GACACCCCAGTAGCAATGAG		
D15S659	(GATA)n	F: GTGGATAGACACATGACAGATAGG	HEX	184
		R: TATTTGGCAAGGATAGATACAGG		
D15S123	(CA)n	F: AGCTGAACCCAATGGACT	HEX	191
		R: TTTCATGCCACCAACAAA		
D19S886	(CA)n	F: TGGATCTACACTCCGGC	FAM	158
		R: ATTTTACTGGCTGGCACTTG		
D19S221	(CA)n	F: GCAAGACTCTGACTCAACAAAA	FAM	174
		R: CATAGAGATCAATGGCATGAAA		

The informative peaks were analyzed by comparing the size of the two alleles in normal and tumor tissue, applying the equation $\frac{T1 \times N2}{T2 \times N1}$, where *T* is tumor, *N* is normal, and 1 and 2 correspond to the peak height of the shorter and longer alleles, respectively. LOH was considered positive for ratios greater than 1.5 or less than 0.66.

3.6. Genetic screening of *B2M* and *CALR*

To determine the frequency of mutations in ***B2M*** and ***CALR*** genes, exons 1 to 3 of ***B2M*** (NM_004048.2) and exon 9 of ***CALR*** (NM_004343) were PCR-amplified in a panel of LC cell lines and lung primary tumors representing several histological types (see Table 3 for additional information).

Table 3. Primer pairs used to amplify *B2M* and *CALR* genes.

Target	Sequence (5' - 3')	Product size (bp)	Annealing temperature
B2M exon 1	CTGGCTTGGAGACAGGTGAC	481	60ºC
	GACGCTTATCGACGCCCTAA		
B2M exon 2	TGTTGGGAAGGTGGAAGCTC	555	60ºC
	GGGATGGGACTCATTCAGGG		
B2M exon 3	TGGGTAGGAACAGCAGCCTA	389	60ºC
	GCAGTTCCTTTGCCCTCTCT		
CALR exon 9	GAGTTTGGCAACGAGACGTG	323	59ºC
	GCCTCTCTACAGCTCGTCCT		

PCR products were cleaned up with a mixture of exonuclease I and antarctic phosphatase to remove unconsumed leftovers, followed by direct sequencing using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Foster City, CA, USA). Mutation surveyor v3.25 software (SoftGenetics) was used to analyze sequences and

annotate genetic alterations. HDs were considered when product amplification was recurrently negative in multiplexed PCR reactions with *GAPDH* as a control.

3.7. Cloning and cell line infections

a) Cloning *B2M* and *B2M*-mutants

All fragments were inserted on the pLVX-IRES-ZsGreen1 plasmid (Clontech, Mountain View, CA, USA) for cloning (Figure 14). The complete *B2M* cDNA (from start to stop codons) was PCR-amplified from the previously retrotranscribed human RNA pool (Agilent Technologies, Santa Clara, CA, USA), using forward and reverse primers harboring 5' *Not*I and *BamH*I overhanging restriction endonucleases, respectively. The construct was transformed into DH5α competent bacteria to grow in LB agar dishes supplemented with 100μg/ml of ampicillin. Plasmids were purified from grown colonies with the NucleoSpin Plasmid DNA Purification Kit (Macherey-Nagel, Düren, Germany) and inserts were confirmed by Sanger sequencing to be wild type for *B2M*. Constructs bearing c.238T>G, c.199G>C and c.95G>A point mutations (sometimes referred to as B2M*W80G, B2M*E67Q and B2M*R32H, respectively) were generated by targeted mutagenesis. PCR amplification was performed with the Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific, Waltham, MA), employing 5'-phosphorylated

primers specifically designed to anneal back to back and to incorporate the desired nucleotide changes (Table 4).

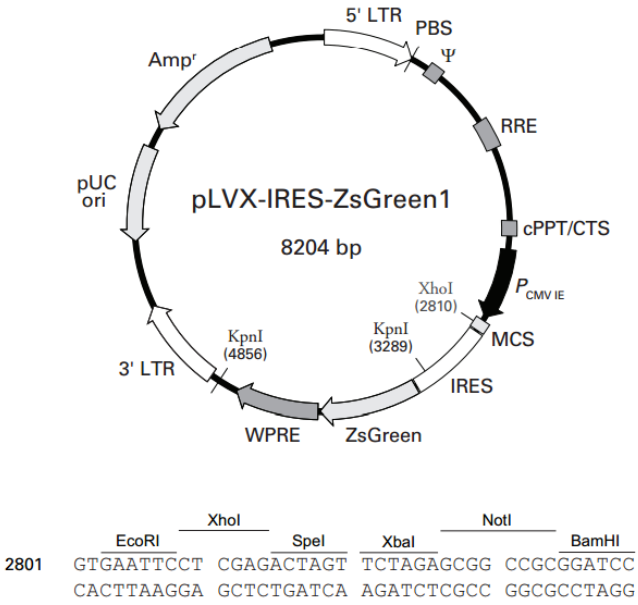


Figure 14. Map of the pLVX-IRES-ZsGreen1 vector. Below, the Multiple Cloning Site (MCS).

Table 4. Primers used to generate all mutant forms of *B2M*.

Tumor	Sequence (5' - 3')	Annealing temperature	Mutation
SCC-1	[P]AGAGAATGGAAAGTCAAATTTCCTG	67ºC	c.95G>A
	[P]GCTGGATGATGTGAGTAAACCTG		
SCC-2	[P]CATTCAGACTTGTCTTTCAGCAAG	66ºC	c.199G>C
	[P]CTCCACTTTTTGAATTCTCTCTCC		
SCC-3	[P]TTGTACTACACTGAATTCACCCCC	68ºC	c.238T>G
	[P]GAGATAGAAAGACCCGTCCTTGC		

[P]: phosphorylation

b) Generation of viral particles and infection

Viral particles were produced by transiently transfect the HEK-293T cell line. Shortly, cells were seeded in 100mm petri dishes in antibiotic-

free DMEM supplemented with 10% fetal bovine serum, in order to achieve 70% confluence at the moment of transfection. About 10μg of the lentiviral constructs (empty vector or mock, *B2M* and *B2M*-mutant forms B2M*W80G, B2M*E67Q and B2M*R32H) were combined with 2.5μg of pMD2.G and 7.5 μg of psPAX2 (Addgene, Cambridge, MA, USA) in Opti-MEM (Thermo Fisher Scientific, Waltham, MA, USA) and added to the cells, along with the polyethylenimine (PEI) transfection agent. After 8 hours, medium was changed and 2 days later, the viral supernatant was harvested and filtered.

Infecting LC cell lines of interest with viral particles was performed using polybrene. Due to the ZsGreen protein of the vector, positively infected cells were seen in green under a fluorescence microscope. Still, the ectopic expression of all constructs in infected cells was always confirmed by western blot before every experiment.

3.8. Antibodies, treatments and immunoassays

a) Antibodies and treatments

Anti-B2M (BBM.1, Santa Cruz Biotechnology and D8P1H, Cell Signaling Technology), anti-HLA class I (EMR8-5, Abcam), anti-PD-L1 (E1L3N, Cell Signaling Technology), anti-CD8 (ab4055, Abcam) and anti-beta tubulin HRP (ab21085, Abcam) antibodies were used for western blot, immunofluorescence, immunohistochemistry and flow cytometry.

Cell stimulation with human interferon gamma (IFNγ) and alpha (IFNα), purchased from PeproTech (London, UK), was performed by direct addition of 30ng/ml to the culture media.

b) Western blot

For western blot, cells were collected in lysis buffer (2% SDS, 10% glycerol and 50mM TRIS-HCl, pH 8.0) and 30μg of total protein was separated by SDS-PAGE, transferred into polyvinylidene difluoride membranes and blotted according to the antibody manufacturer's suggestions. Membranes were washed with TBS 0.1% Tween-20 and incubated with HRP-conjugated polyclonal goat anti-rabbit (P0448) or anti-mouse (P0447) antibodies (both purchased from Dako, Glostrup, Denmark) diluted at 1:5,000 in skim milk. Immunoreactive bands were developed using the Immobilon Western Chemiluminescent HRP Substrate (Merck Millipore, Darmstadt, Germany).

c) Immunofluorescence

To perform immunofluorescence of β2m and HLA-I proteins, cells were seeded on circular coverslips for 24h in order to achieve 70% confluence at the time of the experiment. For H2135 cell line, which is a mixed culture, glasses were previously treated with poly-L-lysine (Sigma Aldrich, St. Louis, MO, USA) to enhance surface attachment. Cells were fixed with a 4% formaldehyde solution (Electron Microscopy Sciences, Hatfield, PA, USA) and blocked with 10% goat serum. Coverslips were kept with primary antibodies overnight at 4ºC, washed and incubated with 1:500 diluted Alexa Fluor-594

fluorochrome-conjugated secondary antibody (Life Technologies, Gaithersburg, MD, USA) for 1h. Cell nuclei were stained with DAPI (0.1μg/μl) and glasses were mounted with Mowiol 4-88 reagent (Merck Millipore, Darmstadt, Germany) before being examined under a confocal microscope. All rinse steps were carried out using PBS alone to avoid cell permeabilization.

d) Flow cytometry

For flow cytometry, cells of interest were infected with constructs 2 days before the experiment. Cells were fixed with 4% formaldehyde and washed with PBS, before blocking with 10% goat serum. About 500,000 cells were labeled with 300μl of primary antibody (anti-HLA-I at 1μg/μl). Cell suspensions were centrifuged, rinsed and incubated with 1:500 dilution of Alexa Fluor-647 fluorochrome-conjugated antibody (Life Technologies, Gaithersburg, MD, USA). Finally, samples were transferred to cytometry tubes and analyzed in a FACSCanto II (BD Biosciences), in which 10,000 events were collected from viable GFP-expressing cells. Appropriate controls were included in the run to provide thresholds to establish the settings and the staining background. Results were analyzed using the FACSDiva software (version 6.1.3).

e) Immunohistochemistry

For IHC, 4μm-thick paraffin-embedded sections of lung primary tumor samples (in some cases in the form of tissue microarrays (TMAs)) were deparaffinized overnight at 62ºC followed by xylene

immersion. Samples were rehydrated in a decreasing ethanol series (100% – 96% – 70%) ending in distilled water. After microwaving with Tris/EDTA pH 9.0 for antigen retrieval, endogenous peroxidase was inhibited with a 3% hydrogen peroxide solution, blocked in 10% goat serum and incubated with primary antibodies overnight at 4ºC (all anti- B2M, HLA-I, CD8 and PD-L1 antibodies at 1:100 dilution). HRP-conjugated polyclonal goat (anti-mouse or anti-rabbit) secondary antibodies (Dako, Glostrup, Denmark) were used in 1h incubations at room temperature. Labeling detection was done applying the ImmPACT DAB Peroxidase (HRP) Substrate kit (Vector Laboratories, Burlingame, CA, USA) and tissue sections were counterstained with hematoxylin for 1min. Once dehydrated in an ethanol battery and cleared in xylene for 1h, samples were mounted with coverslips with DPX mounting medium (Merck Millipore, Darmstadt, Germany). Washes were carried out with TBS 0.01% Triton solution. Sections were evaluated under a Leica DM1000 microscope by two independent observers without any previous knowledge of the clinicopathological data.

β2m, HLA-I and PD-L1 staining evaluation

To evaluate staining degrees of β2m, HLA-I and PD-L1, expression was defined as: i) strong, when tumor cell membranes were intensely marked, presenting the same staining level as stromal lymphocytes or endothelial cells; ii) moderate, when membrane staining was weaker than the surrounding normal cells; iii) negative, when there was no staining at all.

$CD8^+$ T-cell intratumoral infiltration degree

To determine the grade of CTL intratumoral infiltration, five serial snapshots representative of each tissue section were taken under the 10X objective of the microscope. Images were analyzed using the FIJI-ImageJ software, applying the color deconvolution tool to separate hematoxylin, DAB and background staining. Thus, once the tumor region was delimited in the image, the area occupied by the CTLs was able to be estimated. The overall infiltration was indicated as the mean area of the five sections. Samples displaying mean areas >1,500μm^2 were considered to be strongly infiltrated. Those between 1,000 and 1,500μm^2 were moderately infiltrated and the ones having CTL invasion areas <1,000μm^2 were evaluated as negative.

3.9. Cell proliferation assays

Cell lines infected with either the empty vector or the *B2M*-carrying vector were seeded in 96-well plates (cell density of 1,000 cells/100μL medium/well). Over 9 days, 10μl of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) reagent (5mg/ml) were daily added to the cells, followed by a 3h incubation at 37ºC. Finally, formazan crystals were dissolved with 200μl of a solution containing 50% N-N-dimethylformamide, 20% SDS, 2.5% glacial acetic acid and 2.1% HCl 1M. Absorbance was assessed in a microplate reader at 595nm wavelength, to calculate cell viability.

3.10. Gene expression profiles

About 100ng of total RNA from *B2M*-deficient cell lines H2009 and H2135, infected with either the empty vector or *B2M* were selected to carry out a gene expression microarray. RNA integrity values ranged from 9.0 to 10.0 when examined by Lab-on-a-chip technology on an Agilent 2100 bioanalyzer. The Genomics Unit from the Center for Genomic Regulation (CRG, Barcelona) performed the expression arrays and the preliminary data analysis. Briefly, the commercial One-Color Microarray-Based Gene Expression Analysis kit was used for labeling. Hybridization took place on the SurePrint G3 Human Gene Expression v2 microarray 8x60K (Agilent microarray design ID 014850, P/N G4112F), which detects 60,000 transcripts, and a G2505B DNA microarray scanner was used for scanning purposes. Images were quantified through the Agilent Feature Extraction software and fluorescence intensity of each array element was subtracted from the background.

The selection of genes that constituted the *B2M*-expression signature was based on the following criteria: i) 1.15-fold induction or repression in the cells with the wild type *B2M*, comparing with mock-carrying cells; ii) $P<0.05$. Furthermore, only genes for which these premises occurred simultaneously in both cell lines were considered.

4. RESULTS

4.1. Discovering lung cancer genes

Every tumor is unique and the course of its evolution is largely shaped by the intrinsic genetic background. The continuous growing knowledge about the molecular events that drive oncogenesis has led to the discovery of new LC-related genes with great clinical power as targets to directed therapies.

a) Genetic landscape of lung tumorgrafts

One of the major goals of this thesis was to characterize lung tumorgrafts using high-throughput approaches to discover novel genes directly implicated in LC development. The growth of human

tumors in different species requires immunodeficiency of the host animal to prevent rejection of the transplanted foreign tissues. Accordingly, to generate PDXs, athymic nude mice were employed to engraft 14 freshly excised primary NSCLCs, either subcutaneously or orthotopically. Details about the patients and tumorgraft features are summarized in Table 5.

Table 5. Clinical features of patients and characteristics of PDX samples. Passage refers to the number of serial transplants performed before sequencing.

Sample	Passage	Histology	Gender	Age	Smoking	Grade	Stage	TNM
PDX-1	2	SCC	Male	63	Current	G4	IIB	pT3pN0
PDX-2	2	LCC (NE)	Male	71	Former	G4	IIIA	pT3pN1
PDX-3	2	SCC	Male	76	Former	G2	IB	pT2apN0
PDX-4	2	AD	Male	54	Former	G4	IIB	pT3pN0
PDX-5	2	SCC	Male	74	Current	G3	IB	pT2apN0
PDX-6	2	SCC	Male	70	Former	G2	IIIA	pT4pN1
PDX-7	2	AD	Male	52	Current	na	IIIA	pT3N2
PDX-8	2	SCC	Male	74	Former	na	IIB	pT2pN1
PDX-9	48	AD	Female	73	Former	G3	IIIA	T1aN2M0
PDX-10	17	AD	Male	74	Current	G3	IIIA	T2bN2M0
PDX-11	3	AD	Male	67	Former	G3	IA	T1aN0M0
PDX-12	3	AD	Male	68	Current	na	IIIA	T3N2M0
PDX-13	1	AD	Male	61	Former	G2	na	pT2aNxMx
PDX-14	1	SCC	Male	61	Current	G2	IA	pT1aN0Mx

SCC: squamous cell carcinoma; AD: adenocarcinoma; LCC (NE): large cell neuroendocrine carcinoma; na: not available.

As a result of tumor implantation, the human stroma was gradually replaced by murine components. This phenomenon constituted an enormous advantage before the NGS analysis, as the purity degree of the tumors increased with the drift of human normal tissue. As seen in Figure 15, the enrichment on malignant cells was clearly noticed by LOH analysis using highly polymorphic microsatellite markers and by direct sequencing of *TP53*.

Driver mutations are commonly clustered in protein-coding regions, which account for less than 2% of the whole genome[144]. Thus, all 14 tumorgrafts including seven ADs, six SCCs and one LCC of neuroendocrine origin underwent WES. Furthermore, for 10 of them (all but PDXs 9, 12, 13 and 14), transcriptome sequencing was also performed.

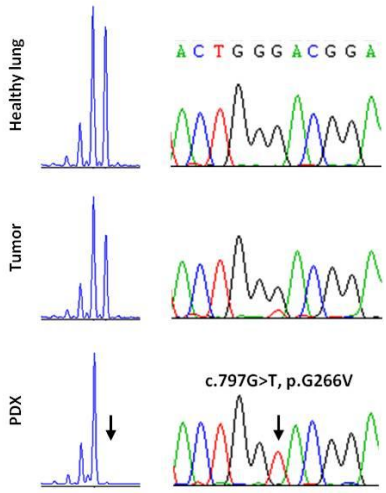


Figure 15. Example of the purity levels of PDX-7, as evidenced by microsatellite analysis (D19S886 marker) and by Sanger sequencing. The enrichment of primary tumors in malignant cells is clearly noted by the complete loss of one of the alleles and by the disappearance of the G nucleotide in *TP53* gene, as pinpointed by the black arrows.

WES of lung tumorgrafts resulted in 77X mean target coverage, with above 90% of bases covered at least 10X. Mouse-derived reads were discarded by mapping against the mouse reference genome (mm10), and the proportion of foreign contamination was able to be estimated, which ranged from 1 to 9%, depending on the sample. Information about WES statistics is presented on the supplementary Table S1.

Sequencing the matched healthy tissue in parallel with PDX samples facilitated the subtraction of the genetic variants from the non-

cancerous genome, and thereby, single nucleotide polymorphisms (SNPs) could be easily recognized and removed. Consequently, tumor-specific alterations including SNVs, indels, HDs and GAs were finally annotated for each sample. The global landscape comprising the somatic mutations found in the protein-coding regions of the tumorgrafts is represented in Figure 16.

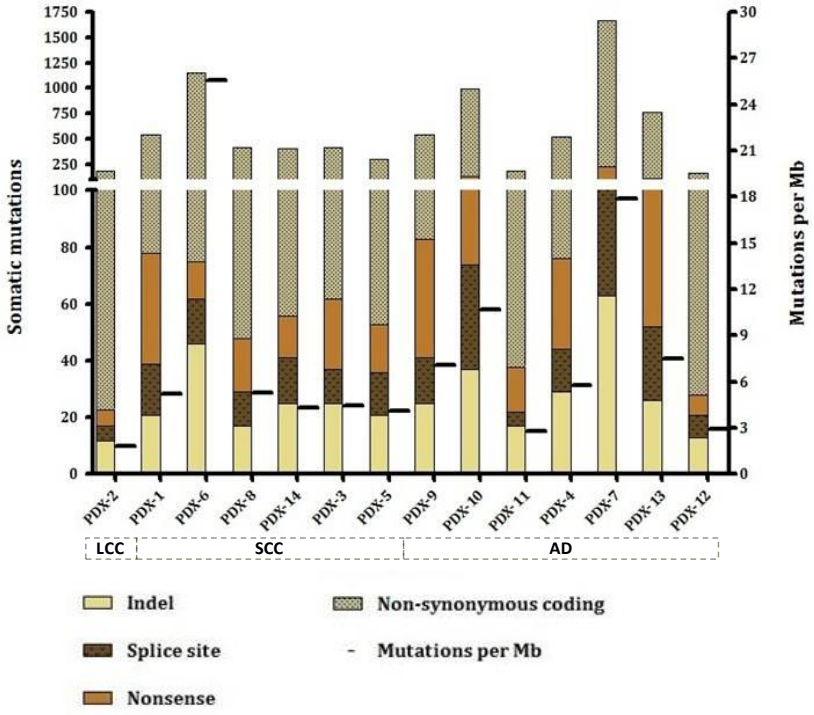


Figure 16. Global landscape of non-synonymous somatic mutations detected by WES in protein-coding exons across PDX samples. The number of mutations per Mb, indicated by the black line, was calculated based on the total amount of alterations (comprising also synonymous-coding). LCC: large cell carcinoma; SCC: squamous cell carcinoma; AD: adenocarcinoma.

Overall, 8,259 tumor-specific non-synonymous alterations were obtained across the entire dataset, comprising 7,167 missense, 475 nonsense, 240 splice site and 377 indels (in which 285 frameshift mutations are included) – see Table 6 for more detailed information.

Table 6. Somatic mutations detected by WES in the coding regions of PDXs. The calculated number of mutations per Mb is also indicated.

Sample	Indels	Splice site	Nonsense	NSyn	Syn	Total	Mutations per Mb
PDX-1	21	18	39	469	162	709	5.2
PDX-2	12	5	6	160	83	266	1.8
PDX-3	25	12	25	353	145	560	4.4
PDX-4	29	15	32	449	193	718	5.8
PDX-5	21	15	17	253	140	446	4.2
PDX-6	46	16	13	1079	1165	2319	25.5
PDX-7	63	39	124	1434	837	2497	17.9
PDX-8	17	12	19	373	201	622	5.3
PDX-9	25	16	42	459	287	829	7.1
PDX-10	37	37	63	855	370	1362	10.7
PDX-11	17	5	16	147	96	281	2.8
PDX-12	13	8	7	138	63	229	2.9
PDX-13	26	26	57	651	322	1082	7.5
PDX-14	25	16	15	347	172	575	4.3

NSyn: non-synonymous; Syn: synonymous.

Additionally, 62 genes were found to be affected by GAs, encompassing 9 distinct chromosomal regions. Such regions are indicated in supplementary Table S2, as well as the candidate genes that registered the highest values of mRNA expression. Amplified regions were annotated by searching for genomic locations presenting a 10-fold higher read count in PDXs, comparing to the corresponding normal tissue.

Likewise, exomes of tumor-normal pairs were scrutinized in order to find zones in which sequencing reads were absent in the tumorgrafts, while present in the matched healthy lung. Thus, applying such read-count strategy, 74 genes distributed across 26 different chromosomal regions were discovered to be homozygous deleted. Further details about the candidate TSGs are included in supplementary Table S3. Of

note, some of the genes were amplified or deleted in more than one PDX sample.

Except for PDX-6, the number of non-synonymous coding mutations was always higher than the synonymous events, indicating an elevated passenger mutation load (Table 6). Tumorgrafts presented distinct values of mutations per Mb in coding regions, ranging from 2 to 26 (mean = 7.5). To calculate the normalized mutation rate, the following formula was applied[145]: $\frac{\Sigma \text{ somatic mutations}}{\Sigma \text{ length of exome targets x number of samples}}$, being 51Mb the capacity of the Agilent's exome capture kit used in this study and 14 the number of samples analyzed.

The information acquired by WES was complemented with the one provided through RNA-seq. The transcriptome analysis of most PDXs was helpful to evaluate how mutations found in exomes were manifested at the mRNA level. In this regard, trends of a negative correlation between mutation events and gene expression levels were noted, meaning that a large subset of alterations affected non-transcribed genes, a phenomenon already described elsewhere[109,146].

WES demonstrated that the mutation rates of genes that are widely known by their carcinogenic role were higher than those formerly reported for lung primary tumors[111,112,147]. This fact may underscore the high purity degree provided by PDXs, which facilitated the detection of homozygous events, as opposed to what happens when tumor specimens are directly sequenced. One example concerns the biallelic mutations detected in *TP53*, which were present in virtually all tumors (93%) in the current study (Figure 17). Despite the small size of the sample cohort analyzed in this work, the mutational profile of the

tumorgrafts was consistent with the general genetic signature associated with lung tumors.

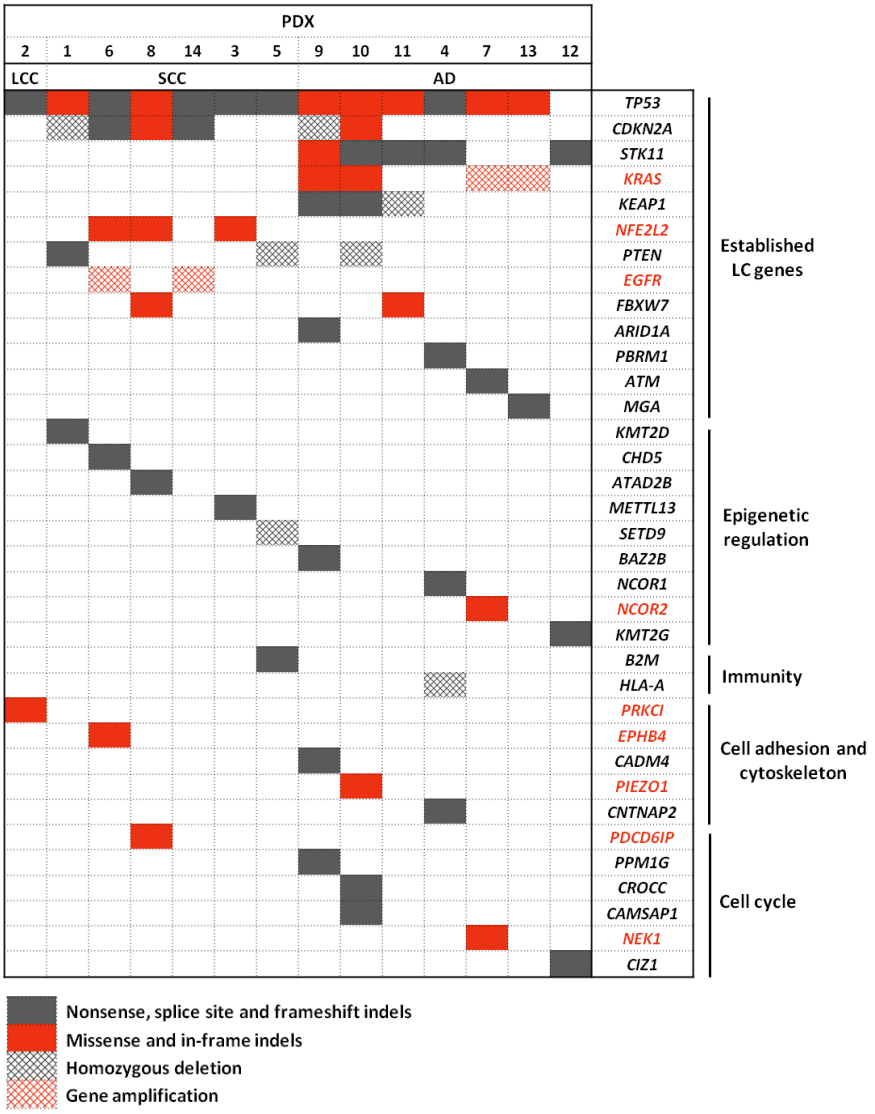


Figure 17. Oncoplot of selected genetic alterations identified in PDX samples. Here were included not only known LC-related genes, but also novel candidates, grouped according to their biological function. Oncogenes and TSGs are highlighted in red and dark gray, respectively.

As denoted in Figure 17, alterations in *KEAP1*, *KRAS* and *STK11* were specific of ADs (the latter affecting 71% of the tumors), whereas defects in *NFE2L2* were detected only in SCCs. In contrast, mutations in *FBXW7*, *PTEN* and *CDKN2A* were found uniformly distributed across all histological types. Some of the abovementioned genes (*KEAP1*, *PTEN* and *CDKN2A*) were inactivated by HDs, which would have been hardly detected if there was admixture between tumoral and non-malignant DNA. Furthermore, genetic modifications in *ARID1A* and *PBRM1* and in *ATM* and *MGA* were detected separately in distinct tumorgrafts (PDXs 9, 4, 7 and 13, respectively). The mutually exclusiveness of such changes is in agreement with an aberrant SWI/SNF-MYC/MAX network, as once described[148]. *ATM* was inactivated in one of the tumors. In this case, there were two different heterozygous DNA changes, most likely affecting distinct alleles as double hit mutations. This observation was verified by RNA-seq, in which the transcript was not expressed in the *ATM*-mutant tumor (PDX-7), while abundantly transcribed in the remainder PDXs.

Moreover, most lung ADs (five out of seven) carried wild type forms of known oncogenes that act as growth factor receptors or that are associated with signal transduction pathways, such as *ALK*, *BRAF*, *EGFR*, genes from the *RAS* family, *MET*, *PIK3CA*, *RET* and *ROS*. Two of these tumors (PDXs 7 and 13) carried *KRAS* amplifications, which were also corroborated by transcript over-expression detected in the two transcriptomes. Similarly, *EGFR* presented GA coupled with very high expression levels in two SCCs (PDXs 6 and 14). Plus, PDX-7 held an *EGFR* mutation (E529K) of unknown significance,

located outside the TK domain that was neither previously described nor included in public databases.

b) Identifying novel lung cancer-related genes

One important hurdle regarding the use of NGS technologies concerns the massive amount of data generated per sample. Indeed, given the enormous number of mutations obtained from WES, specific criteria had to be applied to facilitate the identification of LC gene candidates.

To be considered as tumor suppressors, mutated genes had to fulfill all the following requirements: i) hold HDs, frameshift or nonsense mutations in homozygous state; ii) be widely expressed in the non-mutant tumorgrafts; iii) present high rates of truncating mutations (above 20%) reported in public databases, namely COSMIC, CCLE and cBioPortal[141,143,149]. In a similar fashion, the scrutiny of candidate oncogenes was performed, based on some guidelines: i) harbor amino acid changes, in-frame indels or strong and focal GAs; ii) the mutated transcripts had to be expressed – or over-expressed in the case of amplified genes; iii) present evidence of mutation hotspots (for missense changes and in-frame indels) reported in databases.

The aforementioned filters were established taking into account the common features exhibited by cancer genes. The final high-confidence candidates are indicated in supplementary Tables S2, S3, S4 and S5. Compared to the initial list of mutation-harboring genes, the number of candidate LC-genes was dramatically reduced, rendering a total of 77 TSGs (47 and 30, altered by nucleotide changes and HDs,

respectively) and 67 oncogenes (38 and 29, carrying nucleotide changes and GAs, respectively). Even though various genes on the list corresponded to well-known LC-genes, they satisfied the established criteria and so, were included. Plus, several were found to be mutated through distinct mechanisms, depending on the sample, as evidenced in Figure 17. The selected genes encoded proteins implicated in a wide variety of biological functions, such as epigenetic regulation, intercellular adhesion and cell polarity, cell cycle control and immunity.

Epigenetic regulatory genes included the methyltransferases *KMT2D* (also known as *MLL2*), *METTL13*, *SETD9* and *KMT2G* (or *SETD1B*) and proteins containing bromodomains (*BAZ2B*) and chromodomains (*CHD5* and *ATAD2B*). It was noticeable that abnormalities in these genes hit predominantly SCCs and were found to be mutually exclusive with each other, but not with alterations in members of the SWI/SNF complex (see Figure 17).

Apart from revealing how genetic defects are manifested at the mRNA level, transcriptome sequencing was critical to uncover gene fusions in human lung tumors. In this study, 56 chimeric transcripts were detected among all samples – for further details see supplementary Table S6. The gene fusion involving *MYH7* and *NGDN* identified in PDX-3 was particularly interesting. These two genes are placed next to each other and back to back on chromosome 14 (14q11.2), as shown in Figure 18A, and the fusion found here translates a genomic inversion that took place in this region. The breaking points, determined at the DNA level, were located in the 5'UTR of *MYH7* and in the third intron of *NGDN*. As a result, two in-frame abundant

protein isoforms were detected, including one longer form containing the full-length *MYH7* transcript (Figure 18B).

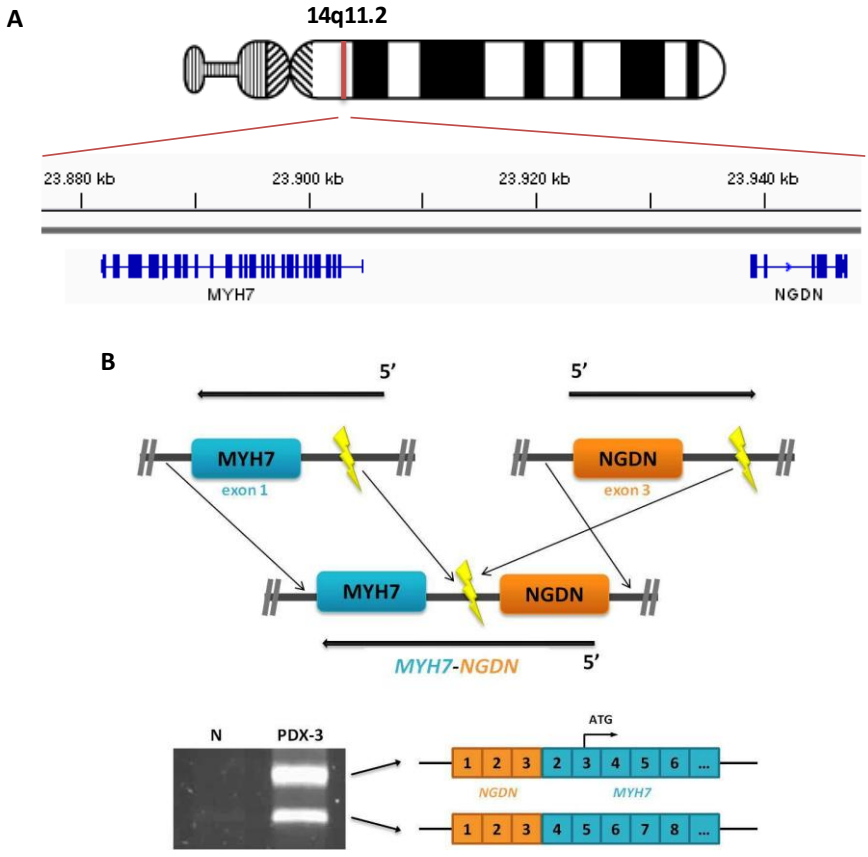


Figure 18. Diagram of the gene fusion found in PDX-3. **A.** Genomic location of *MYH7* and *NGDN*. **B.** Upper panel: structural representation of the inversion, in which breaking points are indicated by lightening marks. Bottom panel: agarose gel showing PCR products of the most abundant transcripts of *NGDN-MYH7* with the corresponding schematic illustration at the mRNA level. N: normal tissue.

MYH7 codes for the myosin heavy chain, which is not typically expressed in lung tissue. However, the gene became over-expressed in the tumorgraft that carried the inversion. An *in-silico* search on the GEO repository (http://www.ncbi.nlm.nih.gov/gds) revealed that high expression levels of MYH7 are not quite common in LC.

However, records regarding a large panel of lung cell lines revealed that H82 cells, which are SCLC in origin, presented MYH7 over-expression (Figure 19). In fact, such chimeric protein that was responsible for its over-expression in H82 was described before, even though the fusion involved a distinct partner, *PVT1*[150].

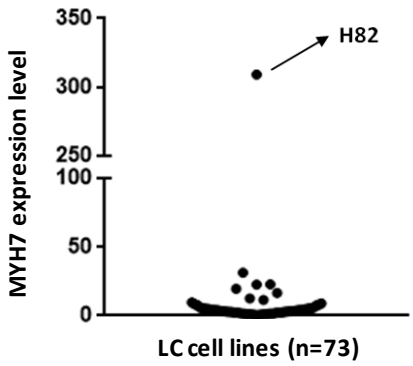


Figure 19. Expression levels of MYH7 in a cohort of LC cell lines accessed in the GEO repository (dataset GSE4824).

4.2. A novel tumor suppressor: *B2M*

The application of stringent criteria to discover new candidate genes implicated in LC reduced the original set of mutated genes by almost 100-fold.

In this regard, aiming to identify tumor suppressors in particular, the emphasis laid on the 78 selected TSGs. The type and mutation frequency of those genes in cancer was carefully examined with searches in publically available databases. Among the candidates, *B2M* stood out for presenting high rates of inactivating mutations. Hence, the gene was chosen for further in-depth studies. *B2M* encodes the

β2-microglobulin, the invariable light chain that stabilizes the MHC-I, and therefore, is involved in the correct presentation of intracellular antigens across the cell surface.

a) *B2M* is altered in tumors of the lung

With the arising of NGS, information from thousands of cancer genomes became easily accessible online. Such massive amounts of data are tremendously valuable, and were crucial to choose *B2M* for further investigation. The biallelic inactivation of *B2M* was detected in PDX-5, caused by a nonsense mutation in the codon 67, located within the exon 2 (Figure 20). *B2M* is located in the long arm of chromosome 15 (15q21.1) and spans approximately 7Kb of genomic DNA. The gene presents an immunoglobulin C1-set domain and comprises four exons, of which three are protein-coding.

In the present study, the conceptual principles shared by Vogelstein and collaborators were followed to find driver genes in cancer[140]. These authors postulated that one should attend primarily to the non-random mutation patterns, and less to the mutation frequency. Knowing that TSGs tend to harbor protein-truncating alterations, genes would then be classified as tumor suppressors if at least 20% of the recorded mutations were inactivating. Considering their definition, *B2M* fulfilled such requirement by far.

In the archives of COSMIC database for tumor samples (accessed in May 19, 2016)[142,143], nonsense and frameshift mutations totalized almost 50% of all DNA changes in *B2M*. Comparable information was obtained by data gathered in the cBioPortal[141], as evidenced in

Figure 21, where the distribution of genetic abnormalities reported in human cancers is displayed.

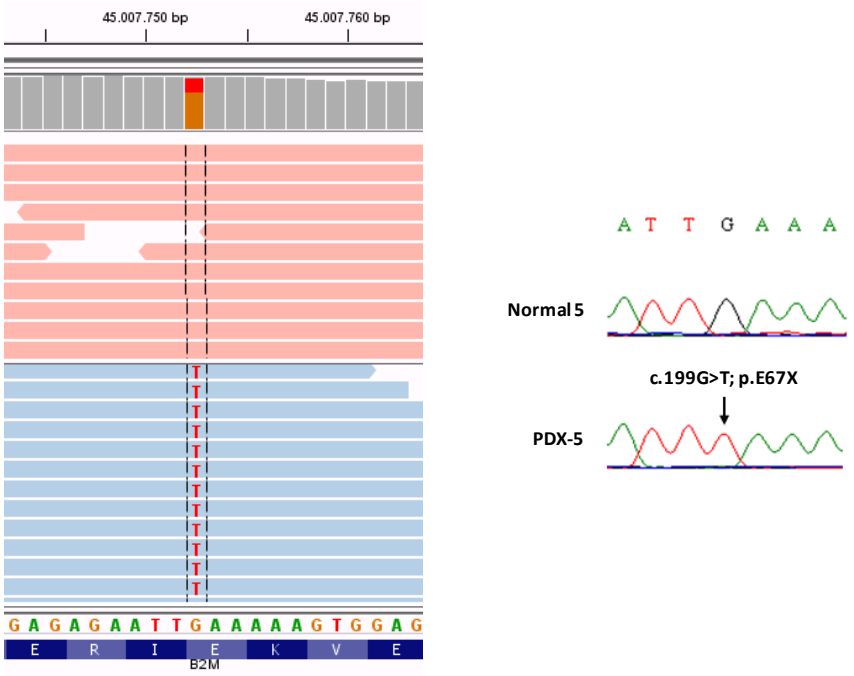


Figure 20. Somatic mutation detected in PDX-5, leading to a premature stop codon at residue 67 of *B2M*. Left: snapshot of the nonsense mutation, as displayed by the Integrative Genomics Viewer software. Reads shown in red and blue correspond to the healthy tissue and the PDX, respectively. Right: the mutation was validated by Sanger sequencing, in which a single clean peak is exhibited, highlighting the complete abrogation of normal tissue obtained from the engraftment process.

B2M, which accumulated 73% (122 out of 166) of truncating modifications, constituted the gene presenting the highest rate of inactivating mutations in databases within the list of TSG candidates (excluding the well-established LC genes) – see supplementary Table S3 and Table S4.

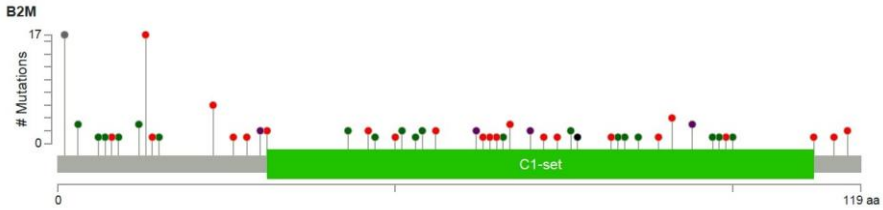


Figure 21. Distribution of somatic mutations in *B2M* reported in cBioPortal for human cancers (accessed in May 19, 2016). Circles are colored according to the distinct categories – green: missense; red: truncating mutations including nonsense, frameshift indels and splice site; purple: residues affected by multiple mutation types; gray: other types of mutations; black: in-frame indels. In case of different kinds of mutations at a single position, the color of the circle corresponds to the most frequent mutation category.

Thereafter, in order to determine the mutation frequency of *B2M*, a genetic screening covering the entire coding region of the gene was performed over a large panel of LC samples. Exons 1 to 3 were PCR-amplified and sequenced using the standard Sanger method. The sample set comprised 79 LC cell lines and 148 lung primary tumors from diverse histopathological types, including 64 SCCs, 52 ADs and 32 SCLCs.

The screening led to the identification of various genetic changes that were distributed throughout the gene, with the exception of exon 3, as depicted in Figure 22. Additional details regarding the detected DNA changes are indicated in Table 7. The overall mutation rate was determined to be 5%, with no association with any particular histopathological type (in tumors, affected four SCCs (PDX-5 plus SCC-1 to -3), one AD (AD-1) and two SCLCs (SCLC-1 and -2).

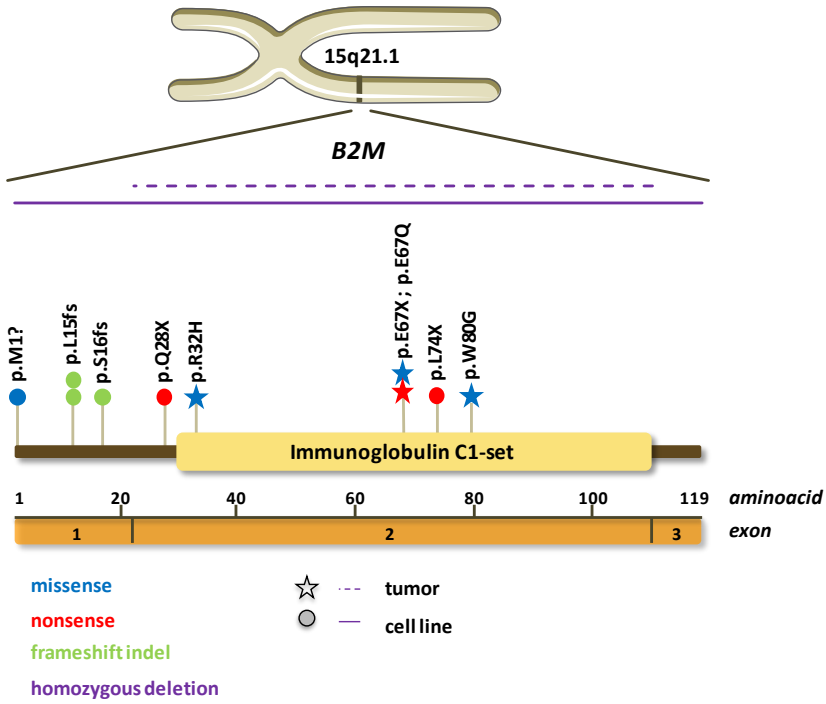


Figure 22. *B2M* mutations found in LC cell lines and primary tumors represented over the cDNA. Alterations are indicated according to their predicted effect and positioned with respect to the reference sequence NM_004048.2.

Table 7. Information about the genetic changes of *B2M* identified in LC cell lines and primary tumors (reference mRNA sequence: NM_004048.2).

LC cell lines (n=79)				
Sample	**DNA change**	**Predicted effect**	**Status**	**Somatic**
H2135	c.45_48*del*TTCT	p.S16fs	Homozygous	NA
H2342	Complete HD	p.0	Homozygous	NA
H2009	c.[1A>G]; [82C>T]	p.[M1?]; [p.Q28X]	Heterozygous	NA
H1417	c.43_44*del*CT	p.L15fs	Homozygous	NA
Lung primary tumors (N = 148)				
Sample	**DNA change**	**Predicted effect**	**Status**	**Somatic**
SCC-1	c.95G>A	p.R32H	NA	Yes
SCC-2	c.199G>C	p.E67Q	NA	Yes
PDX-5	c.199G>T	p.E67X	Homozygous	Yes
SCC-3	c.238T>G	p.W80G	Homozygous	Yes
AD-1	c.43_44*del*CT	p.L15fs	NA	Yes
SCLC-1	HD of exon 2	p.0	NA	NA
SCLC-2	c.221T>A	p.L74X	NA	NA

NA: not addressed / not applicable.

The chromatograms reporting the alterations found in cell lines are displayed in Figure 23A, B and C. DNA changes included frameshift deletions identified in H1417 and H2135, as well as a start codon loss plus a nonsense mutation in H2009. The complete deletion of *B2M* in H2342 is also shown in the agarose gel (Figure 23D).

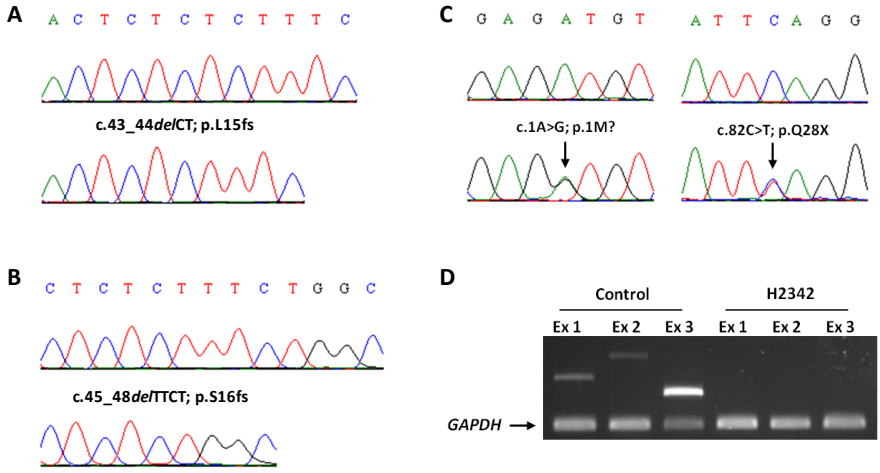


Figure 23. *B2M* mutations detected in LC cell lines. **A.** H1417 and **B.** H2135 harbored homozygous frameshift deletions in exon 1. **C.** H2009 carried a missense mutation in the start codon and a nonsense mutation in exon 2, both heterozygous. **D.** *B2M* was entirely deleted in H2342.

Encoded by *B2M*, the β2-microglobulin has a molecular weight of 11.6kDa and is composed by 119 amino acids, including 20 residues on the N-terminus corresponding to the signal peptide. Every mutation identified in LC cell lines predicted the complete absence of β2m or the expression of truncating forms of the protein (see Table 7), which made them undetectable by western blot (Figure 24). Except for the H2009, which carried two different inactivating mutations, the genetic modifications harbored by the remainder models were homozygous by LOH.

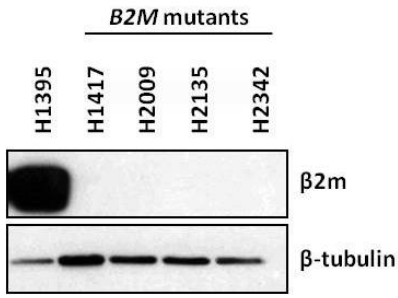


Figure 24. β2m status in *B2M*-mutant cell lines. Mutations resulted in the complete abrogation of the protein. H1395: positive control.

The mutational analysis of *B2M* over lung primary tumors led to the identification of three missense mutations in SCCs: c.95G>A (p.R32H), c.199G>T (p.E67Q) and c.238T>G (p.W80G). In addition, three other inactivating mutations comprehending a deletion of exon 2 (SCLC-1), a frameshift deletion (AD-1) and a nonsense mutation (SCLC-2) were also discovered (Figure 25).

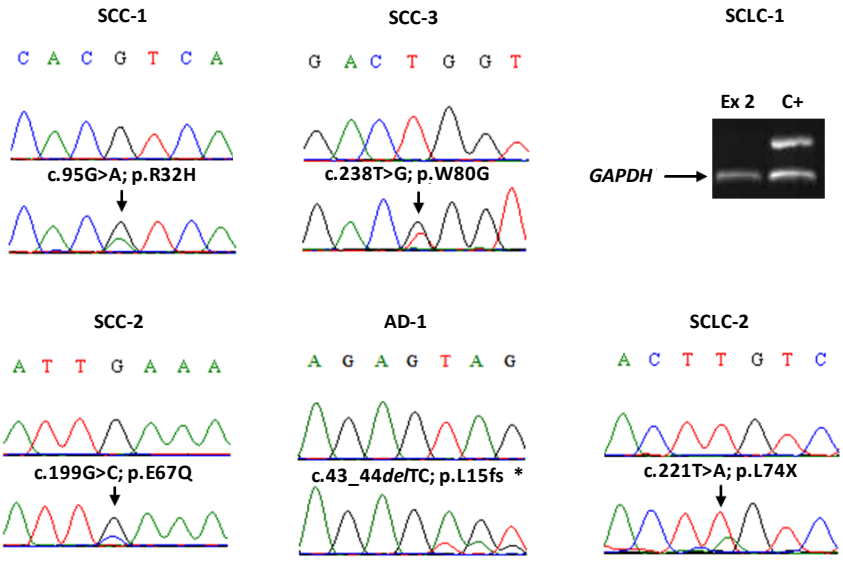


Figure 25. Alterations in *B2M* uncovered in lung primary tumors. Sanger sequencing identified missense, nonsense and frameshift mutations. The deletion of exon 2 in the SCLC-1 sample was seen by agarose gel. *: reverse sequence.

Given the lack of available matched normal DNA or the high contamination levels with healthy material in lung tumor specimens, the zygosity status of the genetic modifications could only be assessed in one sample (SCC-3). Testing for LOH was done using polymorphic microsatellite markers located upstream (D15S214 and D15S641) and downstream (D15S659 and D15S123) of *B2M*. In this tumor, the analysis confirmed that the c.238T>A mutation was homozygous, as depicted in Figure 26.

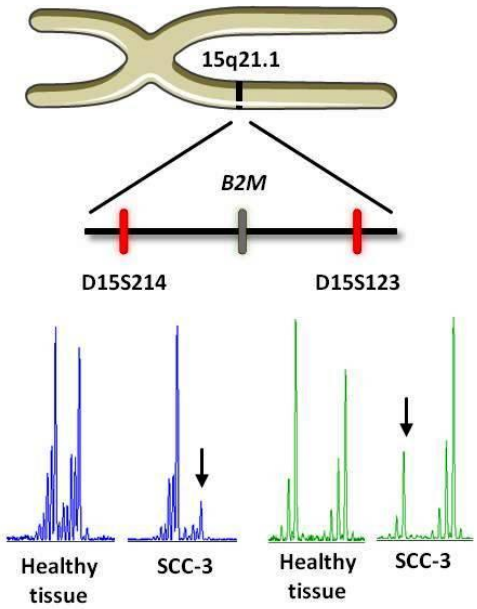


Figure 26. LOH analysis for the SCC-3 tumor. Microsatellites flanking *B2M* were amplified to check the zygosity status of the c.238T>A (p.W80G) mutation. Allele losses are indicated by black arrows.

The c.238T>A predicts a tryptophan-to-glycine change (p.W80G) at the protein level. This specific tryptophan is very well conserved across vertebrates (Figure 27) and the p.W80G modification is predicted to be "probably damaging", according to the PolyPhen-2[151], a tool that measures the possible impact of amino acid changes on the

structure and function of human proteins using straightforward physical and comparative considerations. Conversely, the two other residue substitutions were predicted to be less harmful, p.R32H as "possibly damaging" and p.E67Q as "benign".

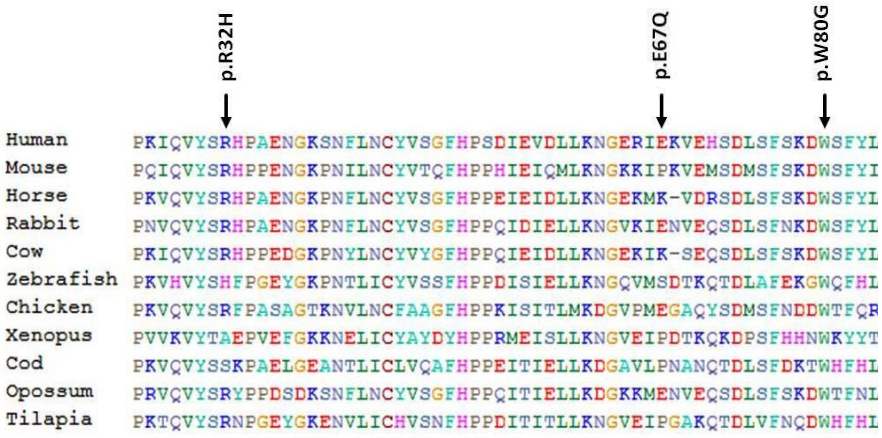


Figure 27. Amino acid conservation of β2m across vertebrates. Multiple sequence alignment, indicating the residue changes discovered in this study.

b) β2m in lung cancer: being or not being

The relatively high mutation rate in lung tumors and the mutational patterns exhibited by *B2M* supported the hypothesis that the gene could be a TSG. The encoded protein, β2m, is a crucial component of the MHC-I. Once coupled to the HLA-I (classically encoded by *HLA-A*, *-B* and *–C* genes), the heterodimer migrates to the cell surface and ensures the correct display of intracellular antigens to the immune cells (Figure 28).

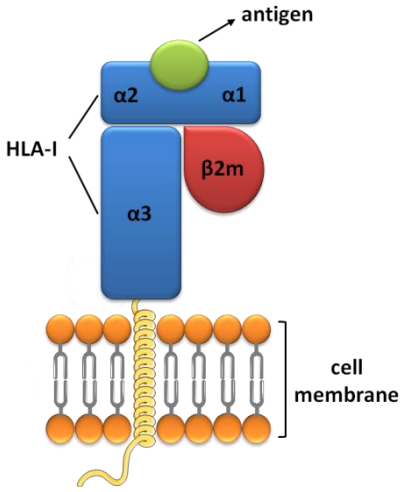


Figure 28. Structure of MHC-I molecules. The MHC-I consists of the heavy chain, HLA-I, which is composed by three domains (α1, α2 and α3) and a light chain, β2m. The complex gets complete after the loading of intracellular antigens into the peptide binding groove.

Aiming to investigate the status of the MHC-I complex in lung carcinomas, western blot with antibodies targeting both proteins were performed over protein extracts of LC cell lines. As noted in Figure 29, β2m was often absent or down-regulated, particularly in the SCLC subset, even in the absence of *B2M* genetic modifications. MHC-I down-regulation in SCLCs has been known for a long time, and is likely to reflect the neural-type of differentiation attributed to the cells of origin[152].

To explore whether this phenomenon was also verified in lung primary tumors, IHC was performed to infer the levels of both β2m and HLA-I proteins in a panel of 420 lung ADs and SCCs organized in TMAs. As specified in Figure 30, the status of the complex was classified into three categories, depending on the staining intensity: i) strong, if tumor cell membranes presented the same staining level as the stromal lymphocytes or the endothelial cells; ii) moderate, if the

membrane staining grade was weaker than the surrounding normal cells; iii) negative, if there was no staining at all. Since β2m and HLA-I have ubiquitous expression in virtually all nucleated cells[153], the surrounding healthy lung tissue presented high expression levels, and thus, it functioned as an internal experimental positive control, favoring the evaluation process.

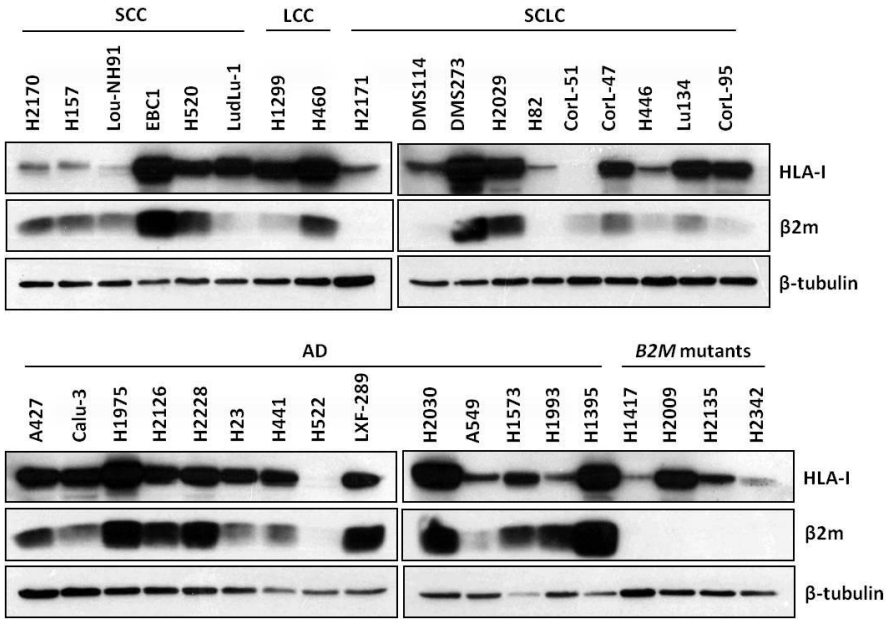


Figure 29. Western blot of β2m and HLA-I in LC cell lines. SCC: squamous cell carcinoma; AD: adenocarcinoma; LCC: large cell carcinoma; SCLC: small cell lung cancer.

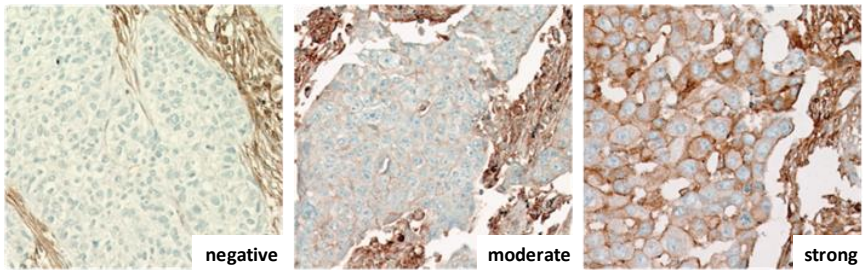
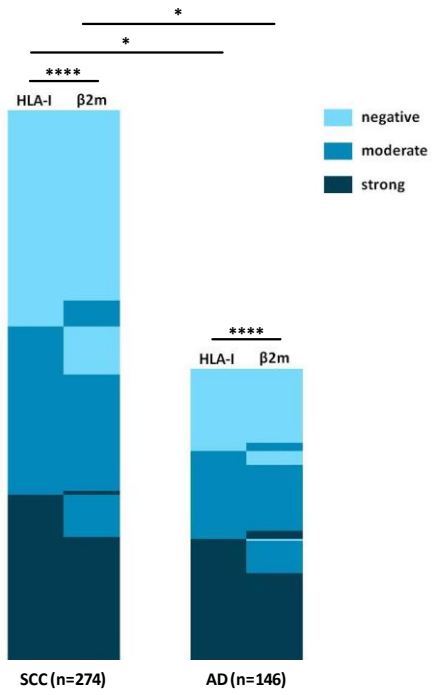


Figure 30. Grades established to evaluate IHC assays in tumors. Samples were classified according to the staining intensity as negative, moderate or strong.

Figure 31 illustrates the allocation of samples, according to the three established groups. Roughly, the IHC analysis for β2m and HLA-I resulted in a strong positive signal for 25% of the samples (23% of SCCs and 30% of ADs). Moreover, about one third of the panel exhibited complete absence of both proteins, suggesting that the loss (or intense down-regulation) of the MHC-I constitutes a common event in lung carcinomas.



Squamous cell carcinoma

		β2m		
		Neg	Mod	Str
HLA-I	Neg	94	13	0
	Mod	24	58	2
	Str	0	21	62

Adenocarcinoma

		β2m		
		Neg	Mod	Str
HLA-I	Neg	37	4	0
	Mod	7	33	4
	Str	1	16	44

Figure 31. Distribution of the LC samples by histopathological type, according to the β2m and HLA-I staining degrees. Chi-square test. * $P<0.05$; **** $P<0.0001$. SCC: squamous cell carcinoma; AD: adenocarcinoma; Neg: negative; Mod: moderate; Str: strong.

As expected, lung tumors that possessed genetic abnormalities in *B2M* predicting truncating proteins (namely the frameshift on AD-1 and the nonsense mutation found in PDX-5) did not present β2m or HLA-I in the cell surface (Figure 32). However, since *B2M* mutations only

affect a small proportion of LCs, genetic alterations striking other genes implicated in the antigen processing and presentation machinery may underlie the absence of MHC-I in the cell surface of these tumors.

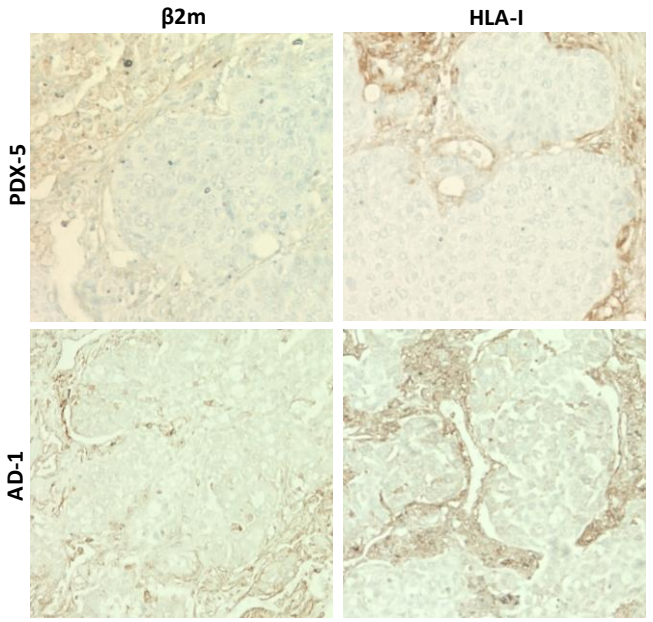


Figure 32. IHC analysis of β2m and HLA-I proteins in tumors harboring truncating mutations in *B2M*. Original magnification: 200X.

As denoted in Figure 31, there was a very significant direct association among the levels of β2m and HLA-I ($P<0.0001$; chi-square test), emphasizing the strong degree of interdependency with each other. Regarding the connection of MHC-I with the distinct histological types of NSCLC, it was observed that the AD subgroup included more specimens with elevated staining degrees for both proteins than the SCC ($P<0.05$; chi-square test).

Details about some clinicopathological features were available for a subset of LC patients. Therefore, taking advantage of these data,

potential associations between the status of the MHC-I and the clinical or pathological traits were able to be investigated. The results, which are summarized in Table 8, revealed that, apart from the histological properties, no other correlations were found between MHC-I and the evaluated parameters (tumor size, lymph node involvement, grade or stage).

Table 8. Categorization of β2m and HLA-I immunostaining, according to the clinical and pathological characteristics of the LC patients. Chi-square test.

	β2m				HLA-I			
	Neg	**Mod**	**Str**		**Neg**	**Mod**	**Str**	
Age (years)								
<60	12	14	19		12	10	26	
60-70	27	30	25	*P=NS*	23	25	37	*P=NS*
>70	17	14	21		20	9	23	
Sex								
Male	55	50	60	*P=NS*	56	38	76	*P=NS*
Female	7	10	11		6	8	15	
Stage								
I-II	52	46	64	*P=NS*	50	39	76	*P=NS*
III-IV	7	13	6		9	6	13	
T stage								
T1	12	22	22		13	17	27	
T2	39	28	40	*P=NS*	38	23	49	*P=NS*
T3-T4	10	9	8		10	5	14	
Lymph node								
N-	40	34	48	*P=NS*	38	31	55	*P=NS*
N+	9	16	14		11	7	25	
Grade								
Well	17	11	10		18	10	10	
Moderate	9	4	8	*P=NS*	10	3	8	*P=NS*
Poor	8	4	5		10	2	5	
Histology								
SCC	119	90	63	*P<0.05*	108	84	82	*P<0.05*
AD	44	51	48		41	44	61	

SCC: squamous cell carcinoma; AD: adenocarcinoma; Neg: negative; Mod: moderate; Str: strong; NS: not significant.

When testing for patient survival, the outcome was quite similar. As shown in the Kaplan-Meier curves of Figure 33, no differences were observed in relapse-free or overall survival of the LC patients, upon

stratification by tumor levels of β2m and HLA-I proteins (for further information see supplementary Figure S1).

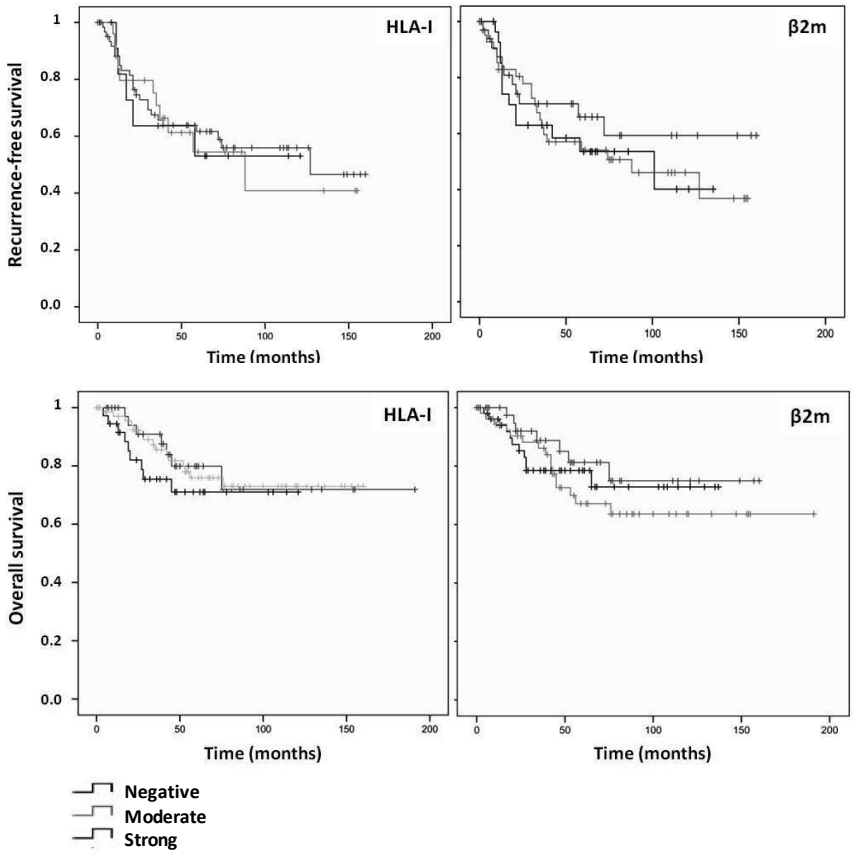


Figure 33. Kaplan–Meier curves of recurrence-free and overall survival in LC patients, with respect to the negative, moderate or strong immunostaining categories for MHC-I proteins.

c) Restituting *B2M*: effects and consequences

β2-microglobulin is primarily known for its function as a chaperone in stabilizing the MHC-I complex. Nonetheless, little or no information exists regarding its putative role in proliferation. This question was addressed by taking advantage of three *B2M*-deficient LC cell lines

H2009, H2135 and H2342, identified over the course of this study. To do so, the 120 codons (from start to stop) comprising the protein-coding region of *B2M* were PCR-amplified and cloned on a lentiviral expression vector. Constructs loaded with either *B2M* (wtB2M) or the empty vector (mock) were introduced in the three cell line models, restoring the expression of the wild type form of β2m in a stable fashion. MTT assays, designed to quantify cell viability (and indirectly proliferation), revealed no differences amongst the cells carrying the wtB2M or the mock for any cell line model, as denoted in Figure 34.

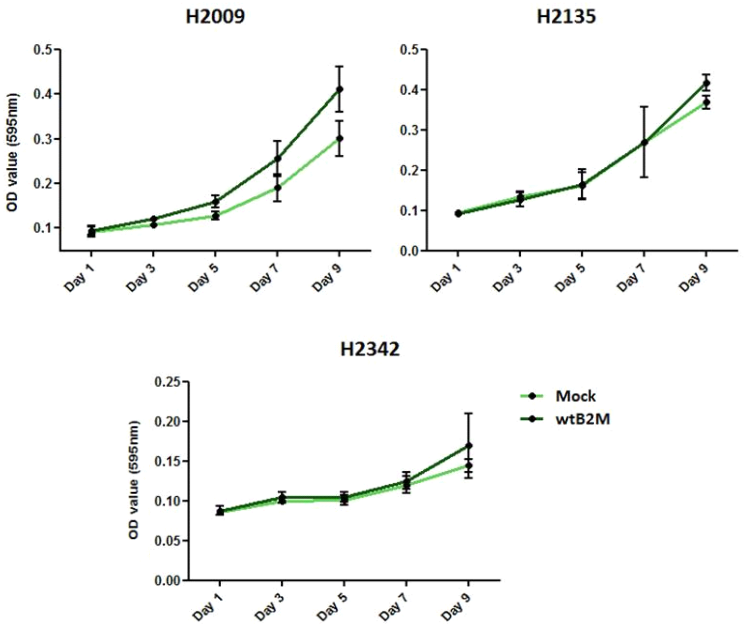


Figure 34. MTT assays for *B2M*-deficient cell lines infected with *B2M* or mock-carrying vectors. Wilcoxon matched-pairs signed rank test.

Aiming to explore the outcomes concerning the intracellular location of MHC-I upon *B2M* restoration, the two abovementioned constructs were used to infect those same cell lines (H2009, H2135 and H2342). Immunofluorescence assays clearly showed that the ectopic expression

of β2m resulted in the successful migration of HLA-I to the cell surface in the H2009 and H2135 cells, as seen in Figure 35. Conversely, this phenomenon was not so intensely noted in H2342, which may reflect some intrinsic feature, making them less prone to be efficiently infected, comparing to their counterparts.

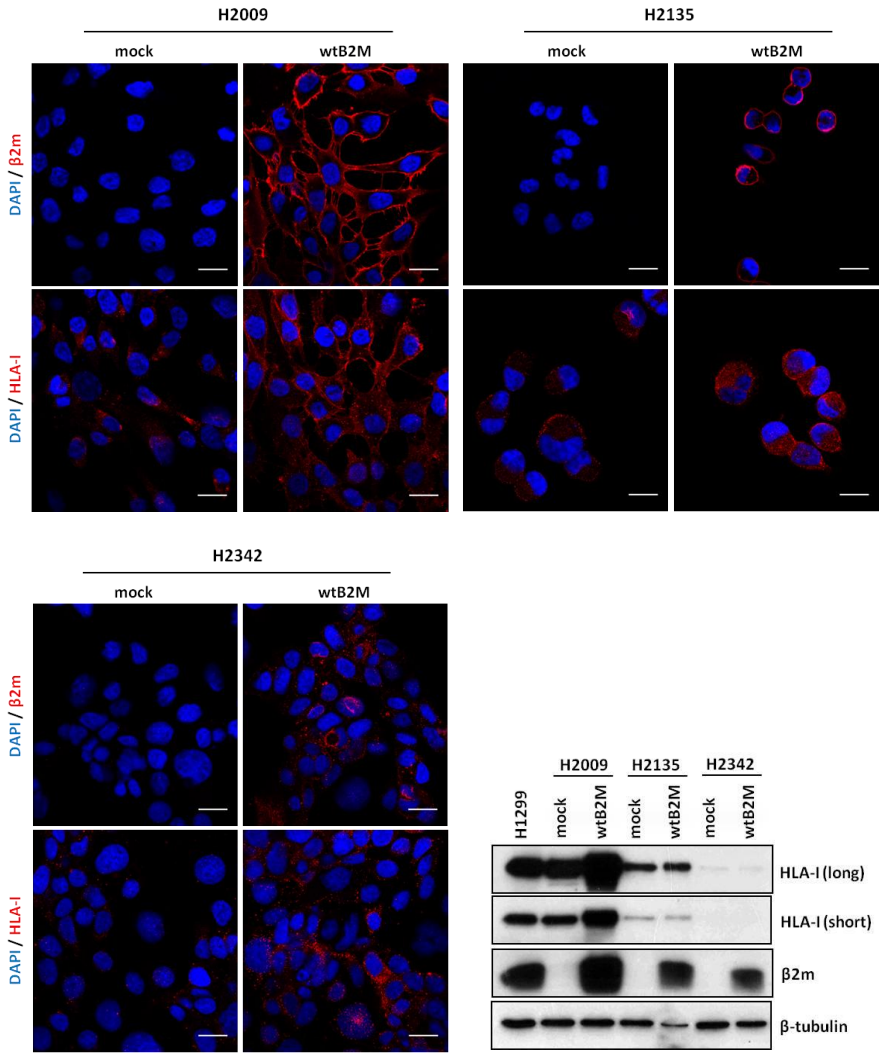


Figure 35. Immunofluorescence of β2m and HLA-I for H2009, H2135 and H2342 after the ectopic and stable expression of wtB2M. Nuclei were stained with DAPI. Representative fluorescent images are shown. Scale bar, 20μm. Bottom right panel: western blot of MHC-I proteins upon *B2M* restitution. H1299: positive control.

As noticed in the western blot panel of Figure 35, all but the H2342 cell line model expressed HLA-I. Plus, in the specific case of H2009, recovering *B2M* led to a slight increase in HLA-I levels. The immunofluorescence evidenced that, prior to the wtB2M restoration, HLA-I accumulated as punctuate clusters along the cytoplasm. The correct display of MHC-I complex across the plasma membrane upon the recovery of the β2m expression was also validated by flow cytometry. The number of cells with HLA-I in the cell surface after gene restitution augmented in the H2009 and H2135 cells (Figure 36).

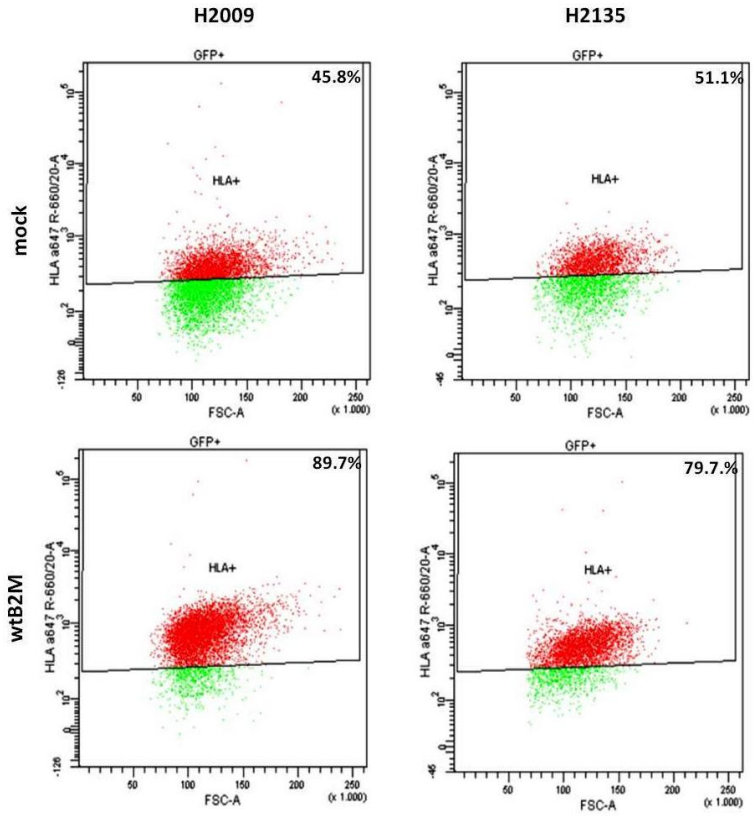


Figure 36. Flow cytometry plots to detect HLA-I in H2009 and H2135 cell populations infected with either the empty vector (mock) or the wtB2M-expressing vector. Plot built over 10,000 counts. The percentages of HLA-I-positive cells (within the GFP-expressing population) are indicated for each panel.

As exhibited in the plot, H2009 registered the largest increase (from 45.8% to 89.7%), followed by the H2135 cells, from 51.1% to 79.7%. As previously observed by western blot (Figure 35) the levels of HLA-I in H2342 cells were almost undetectable. Therefore, it was not surprising that no differences were registered between the H2342 cells bearing the mock or the wild type form of *B2M* by flow cytometry (Figure 37).

Considering the continuous hurdles regarding the lentiviral infection, which constantly led to undetectable levels of HLA-I, the H2342 cell line model was discarded from further functional studies.

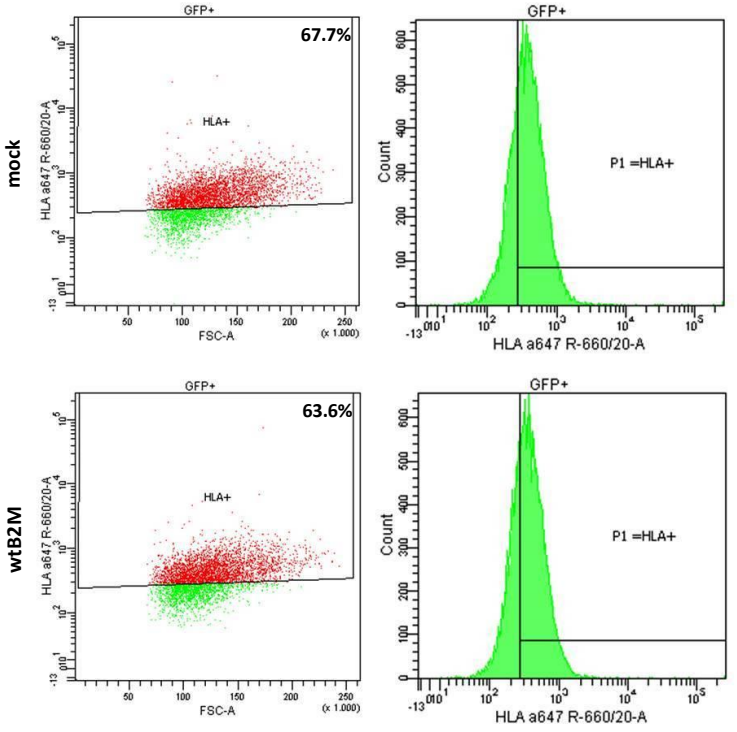


Figure 37. Flow cytometry plots of HLA-I in H2342 cell populations infected with either the mock or the wtB2M. Plot built over 10,000 counts. The percentages of cells exhibiting HLA-I on the surface (within the GFP-positive population) are indicated in each panel.

However, it has been widely reported that the administration of interferon (either IFNγ or alpha (IFNα)) up-regulates the expression of most members of the antigen presentation machinery[60,154]. Hence, one last test was made on H2342 cells, to determine whether this premise could be validated in this cell line. Infected cells underwent a treatment with 30ng/ml of IFNγ for 30 hours, and as observed in Figure 38, the intracellular levels of HLA-I were only slightly boosted in a *B2M*-independent manner, since the same effect was shared by mock cells (see the western blot).

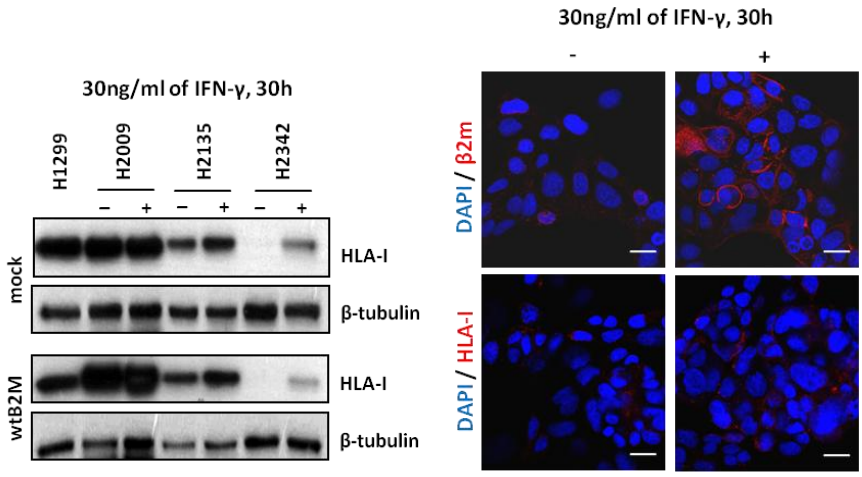


Figure 38. Cell treatments with IFNγ for 30h. Left: western blot showing HLA-I levels before (-) and after (+) treatment with IFNγ in the indicated *B2M*-mutant cells infected with the mock or the wtB2M-carrying vector. H1299: positive control. Right: immunofluorescence of β2m and HLA-I in H2342 cells after ectopic expression of the wtB2M before (-) and after (+) treatment with IFNγ. Nuclei were stained with DAPI. Representative fluorescent images are shown. Scale bar: 20μm.

This concept of interferon response was transposed to LC cell lines harboring the wild type *B2M* and presumably without genetic defects in any other members involved in the MHC-I-dependent antigen processing and presentation machinery. As demonstrated in Figure 39,

the levels of β2m and HLA-I increased upon treatment with 3ng/ml of IFNα in LC cell lines.

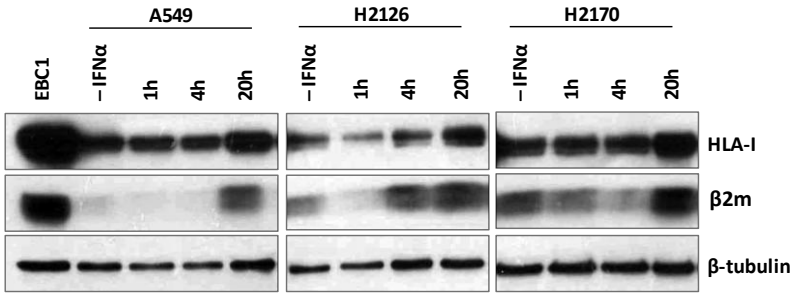


Figure 39. Expression levels of β2m and HLA-I during a 20h-time period upon 3ng/ml of IFNα treatment of several LC cell lines.

To provide new insights about the consequences of β2m deficiency in LC, the gene expression profiles of H2009 and H2135 after wtB2M restitution were determined. In this regard, the microarray data analysis resulted in a list of about 160 genes differentially expressed in wtB2M cells (concurrently in both cell line models), compared to their control counterparts ($P<0.05$ and fold change >1.15) – supplementary Table S7 and S8. Overall, changes in gene expression were very subtle, with few genes showing up- or down-regulation above 1.5-fold. Taking into account that β2m is primarily known to play a role in immunity, such outcome was to some extent expected, since cells involved in this analysis were not subject to any treatment with immune stimuli that would probably remark the expression signature of *B2M*. Nevertheless, genes that changed their expression pattern were associated with several biological functions, such as interleukin signaling, immunity and glycosyltransferases and carboxylic acid biosynthesis. In addition, some of the differentially expressed

transcripts were known targets of IFNγ (*LILRB1* and *C1QTNF3*) and alpha (*ST3GAL6*) – Figure 40.

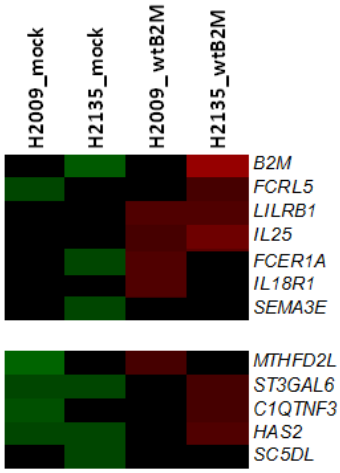


Figure 40. Heatmap of selected up- and down-regulated genes upon wtB2M restitution in the H2009 and H2135 cell lines. Genes were grouped in two major categories. Top: interleukin and immunity-related; bottom: glycosyltransferases and carboxylic acid biosynthesis.

d) *B2M*-mutant forms and MHC-I function

As pointed out before, among all the genes that were pre-selected to be putative tumor suppressors, *B2M* harbored the highest rate of truncating mutations. Several genetic changes identified in the present work through the exhaustive genetic screening in LC cell lines and primary tumors had been earlier reported in public databases, some of which were placed in hotspots (Figure 41). Nevertheless, even though most led to gene inactivation, three lung tumor samples were found with *B2M* missense mutations of unknown functional relevance. So, it was convenient to verify whether these amino acid substitutions impaired the assembly and the right positioning of the MHC-I.

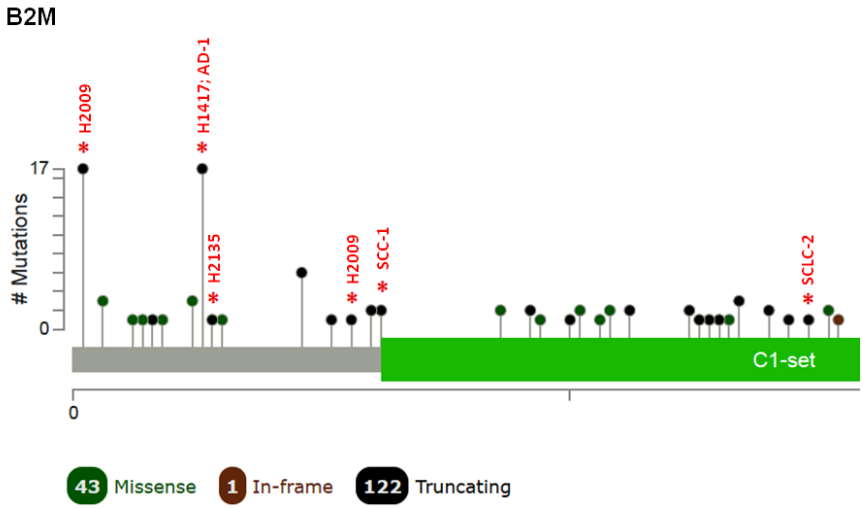


Figure 41. Partial snapshot of *B2M*, as displayed in cBioPortal. Some genetic changes in *B2M* discovered in the present study were recurrent (accessed in July 2, 2016) and are indicated in red. Mutation diagram circles are colored with respect to the alteration types (missense, in-frame and truncating). In case of distinct mutation types at a single position, the color of the circle translates the most frequent.

Therefore, directed mutagenesis was performed over the wild type form of *B2M* to generate the mutants c.238T>G, c.199G>C and c.95G>A (from now on named as B2M*W80G, B2M*E67Q and B2M*R32H, respectively). These abnormal versions of *B2M* were cloned in a lentiviral vector to subsequently infect the H2009 and H2135 LC cell lines. The different β2m-mutant proteins were then stably expressed and the functional consequences were addressed by immunofluorescence. As demonstrated in Figure 42, cells expressing B2M*E67Q and B2M*R32H exhibited β2m and HLA-I proteins properly positioned, denoting the ability of the complex to successfully migrate towards the cell surface. Still, despite the apparent lack of effect in cell positioning, the possibility that such mutants will not interfere with the appropriate functioning of MHC-I complex cannot be fully discarded.

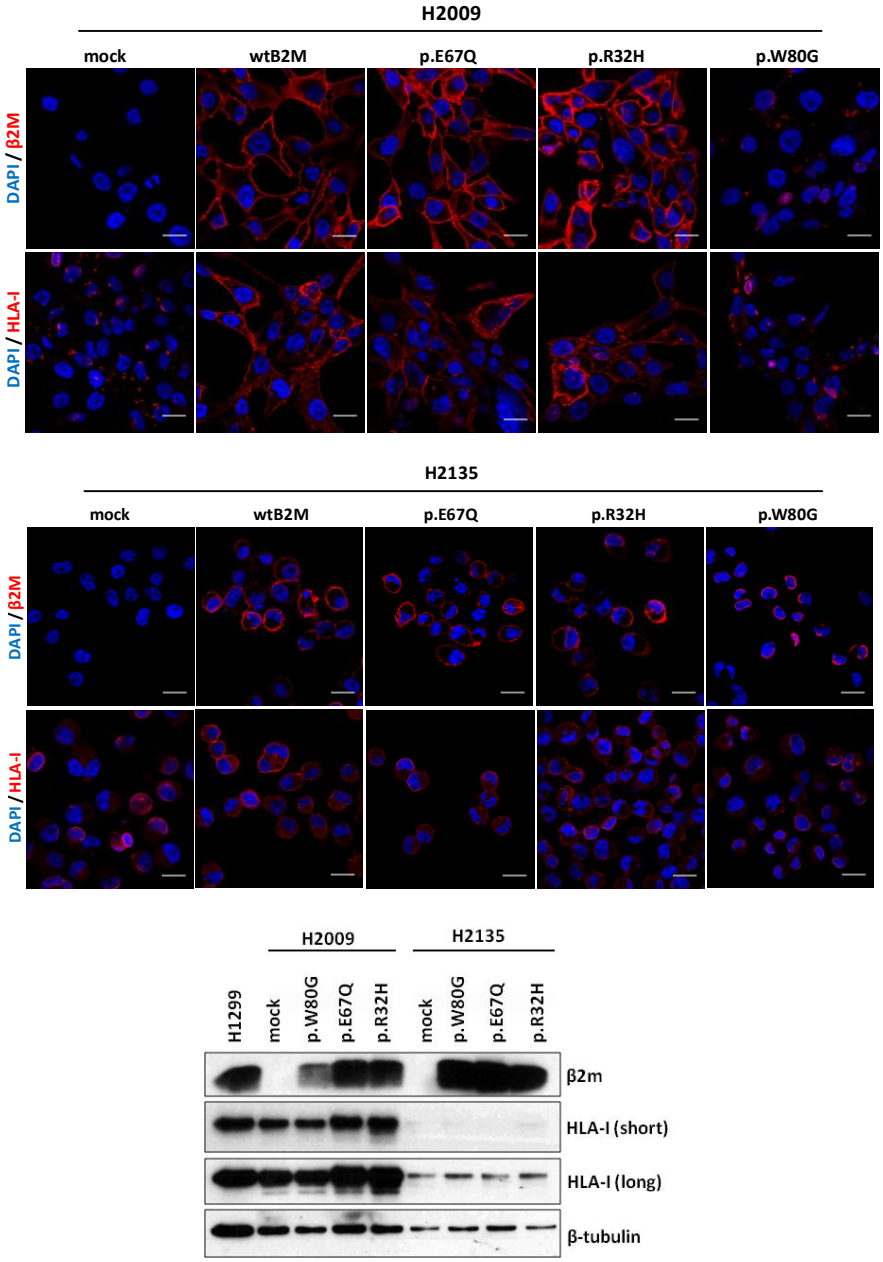


Figure 42. Immunofluorescence targeting β2m and HLA-I on cell lines expressing mutant forms of *B2M*. Upper panel: H2009; bottom panel: H2135. Nuclei were stained with DAPI. Representative images are shown. Scale bar: 20μm. Western blot showing MHC-I protein levels of the indicated stably expressed *B2M*-mutants. H1299: positive control.

In contrast, in cells holding the B2M*W80G mutant, the MHC-I was not able to reach the cell surface, and both proteins were retained in the cytosol (somewhere surrounding the nuclei). As previously mentioned, the p.W80G substitution was predicted to be deleterious, and indeed, these results suggest an aberrant function of *B2M*, impairing the proper assembly of MHC-I. It is also interesting to note that, unlike the other mutant forms and the wild type protein, the B2M*W80G mutant did not triggered an augment of HLA-I levels in the H2009 cells (see the western blot panel of Figure 42).

4.3. Dysfunctions in the MHC class I pathway

The presentation of tumor antigens throughout the cell surface depends on the correct maturation of the MHC-I, which takes places in the endoplasmic reticulum (ER). This fine-tuned process is mediated by several proteins of the antigen processing and presentation machinery that work in synchrony to fold and assembly the complex. Key members encompass calnexin (*CANX*), calreticulin (*CALR*), the isomerase ERp57 (also known as *PDIA3*), the transporters *TAP1*/*TAP2* and tapasin (*TAPBP*)[55].

Figure 43 depicts a schematic overview of the MHC-I-mediated antigen processing and presentation pathway. Briefly, intracellular proteins are degraded by the proteasome in the cytosol, generating peptides optimized for MHC-I binding. Transporter associated with antigen processing TAP1 and TAP2 translocate the antigenic peptides into the ER-lumen to be loaded onto the newly synthesized MHC-I heterodimers.

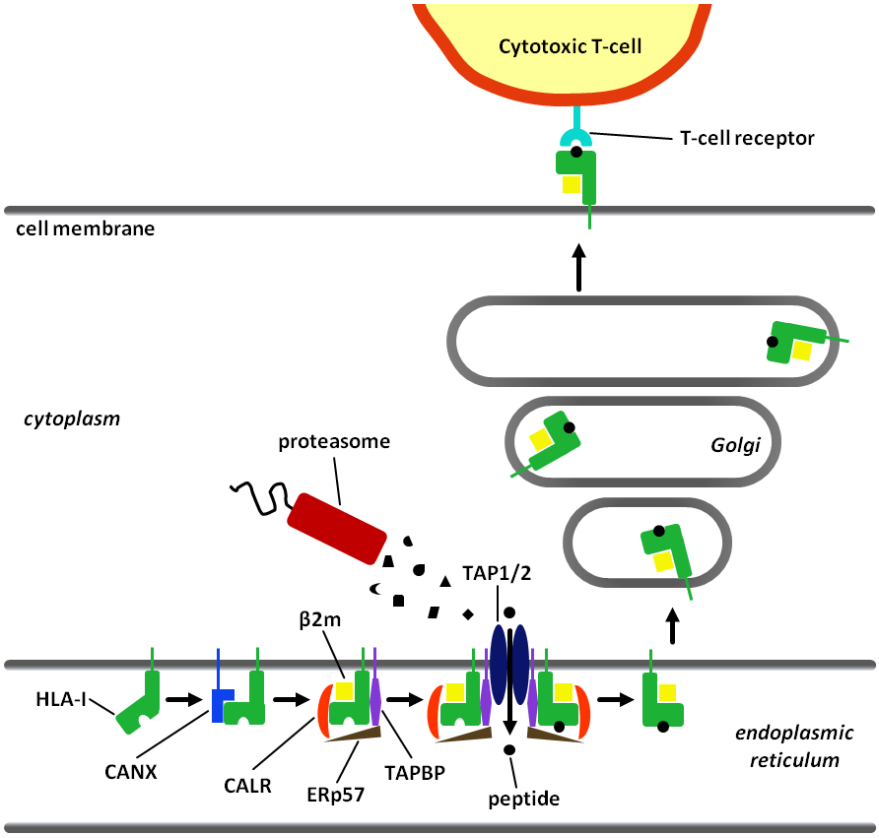


Figure 43. Illustration of the MHC-I-mediated antigen presentation. Intracellular proteins are cleaved into peptides by the proteasome. The transporter for antigen processing (TAP1/2) translocates peptides into the lumen of the ER, where they are awaited by the MHC-I heterodimers. These were previously synthesized with the assistance of the chaperones calnexin, calreticulin, ERp57 and tapasin. Peptide binding triggers the release of MHC-I molecules from the ER towards the cell surface, where they can be presented to the cytotoxic T-cells.

The synthesis of the complex begins with the stabilization of the heavy chain by calnexin, a chaperone that promotes the assembly of β2m. The recruitment of ERp57 and tapasin is triggered and the interaction with TAP1/2 ensures peptides to be appropriately loaded in the binding groove of MHC-I molecules. Antigen binding is required to the correct folding, which results in a conformational

change that releases the chaperones. Consequently, the MHC-I complex leaves the ER and is transported through the Golgi apparatus to the cell surface. Antigenic peptides can finally interact with the TCRs of CTLs, inducing immune tolerance in case of self-antigens or immune destruction, if derived from non-native or transformed proteins.

a) Beyond *B2M*: defects in other members

The results presented in this thesis underline that the loss of MHC-I is a recurrent event in LC. Hence, it is fair thinking that defects on any proteins (other than β2m and HLA-I) implicated in the antigen processing and presentation machinery would probably hamper the proper functioning of the pathway. An *in-silico* comprehensive search was performed to ascertain how often these genes were inactivated in LC.

As indicated in Figure 44, truncating mutations on *CALR*, *PDIA3* and *TAP1* genes had been reported in the H446, H1155 and HCC-366 LC cell lines, respectively. The validation by Sanger sequencing revealed that nonsense mutations in *CALR* and *PDIA3* were heterozygous, and therefore, these genes were likely to maintain their activity, at least partially. The lack of genetic material precluded the corroboration of the *B2M* change in H2347. Similarly, the *HLA-A* frameshift indel could not be addressed by failures in PCR amplification, possibly due to the extreme polymorphic properties of *HLA-I* genes, making the primer design almost impracticable.

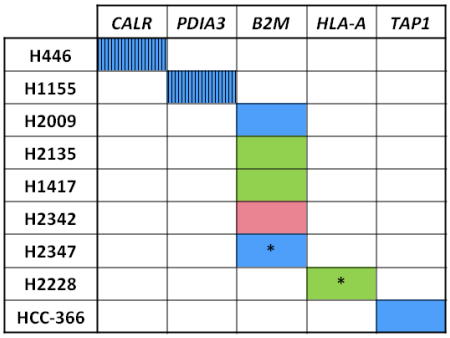


Figure 44. Inactivating mutations in genes of the MHC-I-dependent antigen presentation pathway for LC cell lines. Data combines reports from public databases and those provided by the present study. Blue: nonsense; green: frameshift; pink: HD. *: changes that could not be validated; vertical lines: heterozygous mutations.

Inactivation of *HLA-A* was also detected in one of the tumorgrafts (PDX-4) analyzed in the present study (see Figure 17), which held a tumor-specific HD of the entire gene. Intriguingly, a manually inspection of the WES files for *HLA-A* revealed that seven samples presented none or very few reads in the normal DNA. This detail reinforces and reflects the drawbacks of addressing the status of the *HLA-I* genes. Being the most polymorphic genes in the genome, aligning sequencing reads to the human reference sequence may have been infeasible, and so, they might have been automatically discarded. Thereby, evaluating the role of *HLA-I* alterations in human cancers is an issue of concern, due to the technical challenges.

The alteration in *CALR* gene in H446 was located in exon 9, a region that is recurrently mutated in myeloproliferative neoplasias harboring wild type *JAK2*[155,156]. Since abnormalities striking *JAK2* in LC patients occur at very low frequencies (in about 1%)[157], a genetic screening over that specific exon was done in a series of 79 lung primary tumors and 62 LC cell lines. Still, no additional mutations were found, suggesting that DNA changes in this specific exon are a rare event in

LC (although the existence of other *CALR* mutations located elsewhere in the gene cannot be ruled out).

Taken together, these observations indicate that the common lack of MHC-I in the surface of lung malignant cells may be beyond the malfunction of β2m. Abnormalities in other proteins of the MHC-I-dependent antigen presentation machinery would drive to comparable outcomes, contributing to the arising of immune evasion phenotypes. In fact, the markedly absence or down-regulation of MHC-I in the SCLC histological subtype referred in the literature and corroborated in the current work was also noticed by digging in GEO datasets for a wide variety of APMs (Figure 45).

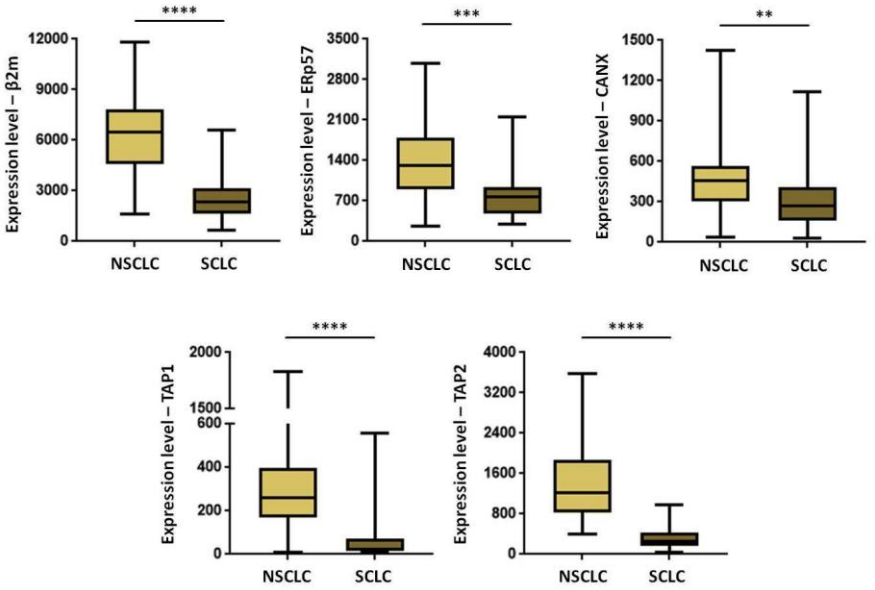


Figure 45. Expression levels of proteins involved in the MHC-I antigen presentation pathway for LC cell lines: 51 NSCLCs and 22 SCLCs. Plotted data derived from GEO (GSE4824 dataset). Whiskers indicate the lowest and highest values. Mann-Whitney test. ** $P<0.01$; *** $P<0.001$; **** $P<0.0001$. NSCLC: non-small cell lung cancer; SCLC: small cell lung cancer.

As plotted in such figure, data gathered from gene expression arrays performed over a large set of LC cell lines (GSE4824), shows that besides β2m, the transcript levels of ERp57, TAP1, TAP2 and CANX were also significantly reduced in SCLCs, comparing with the NSCLC group. This underpins the evidence of a global down-regulation of the entire pathway, and may explain the extraordinary aggressiveness and the intrinsic metastatic features exhibited by these lung tumors.

4.4. MHC class I and tumor immune escape

The structural alteration of the tissues induced by neoplastic transformation stimulates immune responses that eventually result in the eradication of the malignant cells. Thus, failures on tumor antigen presentation due to defects in the correct location of MHC-I can compromise the antitumor response triggered by CTLs.

a) MHC-I modulates the antitumor immunity

Here, we tested whether the presence of β2m and HLA-I could predict the intratumoral infiltration by CTLs. IHC analysis were performed to detect CTLs in a subset of 25 LC samples (13 SCCs and 12 ADs) exhibiting either strong (n=7) or negative (n=18) staining for both MHC-I components. To determine the grade of CTL invasion in tumor nests, snapshots were taken from every sample and analyzed in order to isolate the DAB staining. The tumor region was delimited and the area occupied by stained immune cells was automatically calculated, as specified in Figure 46.

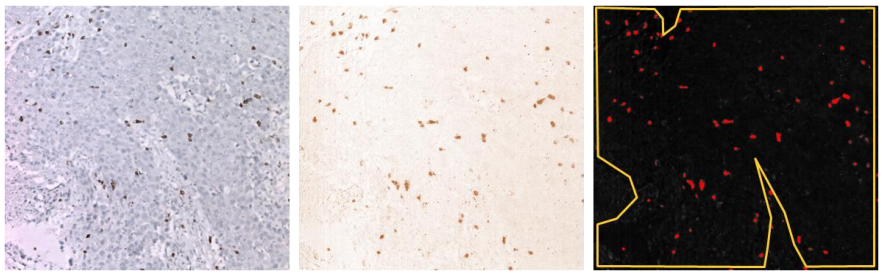


Figure 46. Estimation of CTL infiltration degree in tumor samples. Left panel: original image (magnification 100X). Middle panel: isolated DAB staining. Right panel: delimitation of the tumor nests (yellow line) to calculate the area occupied by CTLs (matching the red spots).

The results obtained are plotted in Figure 47, in which is notorious that the highest intratumoral infiltration grades were associated with those samples exhibiting the strongest levels of β2m and HLA-I expression. Further details are summarized in supplementary Table S9. These findings seem to underscore that the presence of MHC-I-mediated tumor antigens on the cell surface modulates the recruitment of tumor-infiltrating $CD8^+$ T-cells.

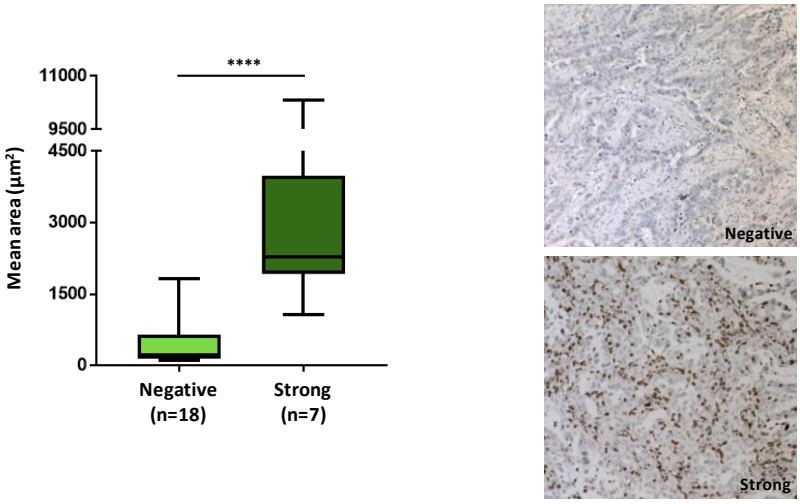


Figure 47. Distribution of tumors that presented negative or strong levels of MHC-I, according to the area occupied by infiltrated CTLs. Whiskers delimit the highest and lowest values. Mann-Whitney test; **** $P<0.0001$. Right panel: example of negative and strong tumor CD8 T-cell infiltration.

However, presenting tumor epitopes via MHC-I is not sufficient to induce CTL activation and an appropriate immune response, which is dependent of a refined balance between stimulatory and inhibitory signals. Thus, tumor cells displaying inhibitory molecules, like PD-L1 will enhance immune tolerance and avoid their own elimination. Inhibiting these immunosuppressive elements is on the basis of current anticancer immunotherapy using monoclonal antibodies to block PD-L1 or PD-1, its co-receptor in T-cells[158].

Here, aiming to get some insights about the relationship between MHC-I complex (β2m and HLA-I) and PD-L1 in LC, the status of such ligand was evaluated by IHC in a group of more than 200 lung tumors. In a comparable manner with other markers, specimens were classified based on their staining grade (negative, moderate or strong). The results are represented in Figure 48, in which an obvious association was observed between PD-L1 and MHC-I. In this regard, high expression levels of MHC-I were concomitant to stronger PD-L1 staining degrees.

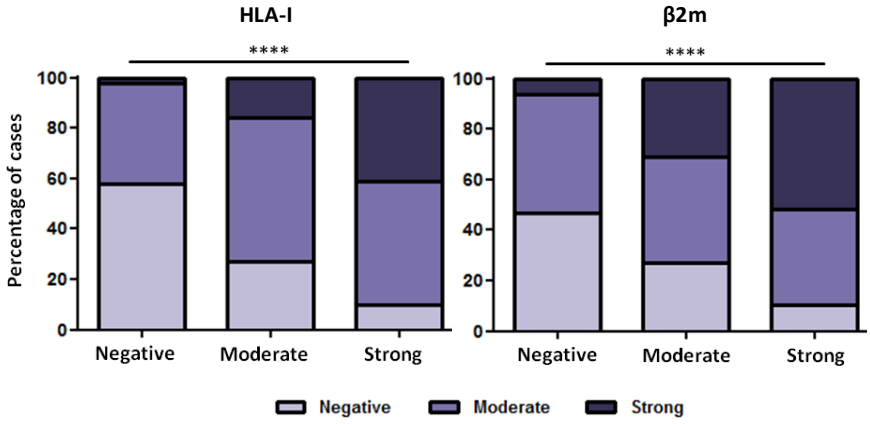


Figure 48. Distribution of the distinct PD-L1 immunostaining categories, according to the expression levels of β2m and HLA-I. Chi-square test. **** $P<0.0001$.

b) β2m as a biomarker in lung cancer therapy

Monoclonal antibodies against PD-1 and PD-L1 have emerged as a promising anticancer therapy for NSCLC, and multiple agents are currently being used with significant improvements in patient survival (reviewed in Bagley *et al.*, 2015[88]). Theoretically, the presentation of tumor antigens by the MHC-I complex is a mandatory step to trigger CTL activation, which in turn, may be inhibited by the PD-1/PD-L1 interaction.

In this study, it seemed convenient to untangle whether increased levels of MHC-I and/or higher CTL intratumor infiltration could act as predictive biomarkers of therapeutic efficiency. Thus, tumor tissues from 15 NSCLC patients treated with distinct immunotherapeutic agents were tested for β2m and CTL infiltration (Table 9). Patients were selected within the scope of a clinical trial for bearing PD-L1-positive tumors: five underwent anti-PD-1 drug administration (either pembrolizumab or nivolumab) and ten were treated with the anti-PD-L1 antibody, atezolizumab. Staining intensities for β2m and CTL were determined, and samples were stratified based on the combined evaluation of both markers. The effects on survival were accounted from the date they started to receive immunotherapy.

According to the Kaplan-Meier curve in Figure 49, patients categorized globally as high and moderate had longer life expectancies after receiving the treatment than those included in the negative category. Nonetheless, the relatively small cohort of analyzed patients and the heterogeneity attributed to the treatment strategy may underlie the lack of statistical significance (P=0.077; log rank test).

Table 9. NSCLC patients and their response to anti-PD-1 or anti-PD-L1 treatments. The degrees of β2m expression and CTL infiltration are specified, and constituted the basis to get the "overall" column, summarizing the data.

Case	Type	Age	Sex	Drug	β2m	CTL	Overall	Response
P1	AD	53	M	Nivol	Mod	Mod	Mod	SD
P2	SCC	57	M	Atezo	Neg	Mod	Neg	SD
P3	AD	58	F	Atezo	Str	Str	High	SD
P4	SCC	78	M	Atezo	Str	Str	High	PR
P5	SCC	61	M	Atezo	Mod	Str	Mod	SD
P6	SCC	54	M	Atezo	Str	Str	High	PD
P7	AD	59	M	Atezo	Str	Str	High	CR
P8	SCC	76	M	Atezo	Mod	Mod	Mod	PR
P9	SCC	78	M	Atezo	Neg	Neg	Neg	PR
P10	SCC	69	M	Pembr	Neg	Neg	Neg	PD
P11	AD	62	M	Pembr	Neg	Mod	Neg	PR
P12	AD	67	F	Atezo	Str	Str	High	SD
P13	AD	58	F	Atezo	Mod	Str	Mod	PR
P14	AD	60	F	Nivol	Str	Str	High	PD
P15	AD	63	M	Nivol	Mod	Mod	Mod	PR

Neg: negative; Mod: moderate; Str: strong; Nivol: Nivolumab; Atezo: Atezolizumab; Pembr: Pembrolizumab; AD: adenocarcinoma; SCC: squamous cell carcinoma; M: male; F: female; SD: stable disease; PR: partial response; CR: complete response; PD: progressive disease.

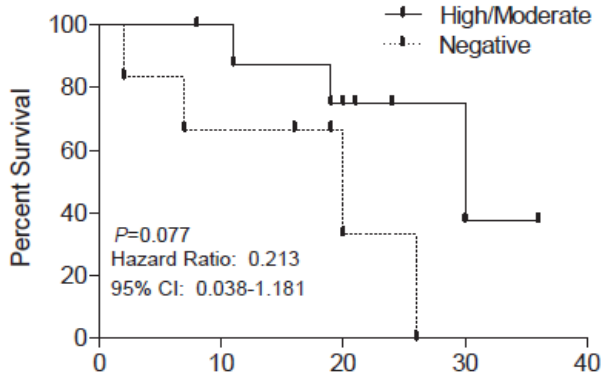


Figure 49. Kaplan–Meier curve showing survival after initiation of treatment with immunotherapeutic agents, in regard to β2m staining and CTL infiltration parameters combined. Probabilities are those associated with the differences between the two groups revealed by the log-rank test.

Yet, it is worth mentioning that a complete response was achieved by one patient, who presented intense staining for β2m and substantial CTL intratumoral infiltration. Altogether, trends observed in these experiments suggest that levels of β2m (and consequently of MHC-I) and cytotoxic T-cell invasiveness in lung tumors constitute potential biomarkers to predict treatment response to anti-PD-1/PD-L1 mAbs.

5. DISCUSSION

The central goal of the current dissertation was to apply genome-wide approaches to discover novel genes causally implicated in LC, taking in mind their potential use in clinics.

The extensive use of massive NGS technologies has become a routine to generate massive amounts of data in a rapid and cost-efficient way. In this regard, international consortiums have been created to unravel the genomic landscape of LC by sequencing hundreds of tumor samples. The major histopathological types were already subject to such comprehensive analysis, which were crucial to uncover key biological processes and to identify cancer drivers of clinical relevance[111,113,157,159]. However, these studies share one particular detail in common: samples were directly sequenced after surgical resection. This means that a fraction of the analyzed tumor sample corresponded

to adjacent non-malignant tissue. Consequently, the admixture between normal and tumor cells would affect the sensitivity of the sequencing analysis, precluding the detection of homozygous somatic alterations and impairing the recognition of GAs and HDs. Knowing that oncogenes are frequently activated by GAs[29] and that homozygous modifications (either derived from nonsense and frameshift mutations or by HDs) are responsible for TSG inactivation[31], tumor purity is, indeed, a fundamental issue.

To overcome the abovementioned hurdles, the genomic signature of lung tumors was determined through WES and transcriptome sequencing, using 14 PDXs as a source of tumor material. Once engrafted in immunodeficient mice, human stromal components were drifted, and tumorgrafts became enriched in malignant cells. PDXs are recognized by their ability to preserve not only the histopathological features, but also the genetic background and the expression profile of the original tumors[160–162]. The purity degree achieved upon serial rounds of transplantation constituted the main reason why PDXs proved to be suitable models to improve the identification of genetic abnormalities. Moreover, sequencing the corresponding normal tissue facilitated the detection of tumor-specific mutations, given that SNPs could be automatically discarded.

Driver mutations are generally clustered in protein-coding genes, which accounts for less than 2% of the human genome[144]. WES has lower costs when compared with sequencing entire genomes, and originates smaller amounts of output data, which are more manageable to work with. Plus, protein-coding variants are naturally more easily interpreted than their non-coding counterparts. Therefore, coupling

WES with RNA-sequencing enabled to evaluate how mutations found in exomes were manifested at the mRNA level.

Given that the typical mutation patterns of oncogenes and TSGs display non-random properties, rigorous criteria were applied to select LC-gene candidates. The selection method was largely complemented with the information provided by public databases, in which NGS data from hundreds of human cancers could be readily assessed. Digging in databases is especially important when the sample size is relatively small, as in this case, once the odds of finding recurrently mutated genes are very low.

Hence, the 8,259 non-synonymous alterations initially listed for the 14 PDXs were able to be greatly reduced to 77 TSGs and 67 oncogenes. The final list also comprehends well-established LC-genes, some of which were found to be mutated in rates higher than regularly reported. Examples include alterations in *STK11* (71%) and *KEAP1* (43%), identified solely in lung ADs, in contrast with the rates determined elsewhere (17% for both genes in a panel of 230 tumors)[112]. Likewise, abnormalities in *NFE2L2* were only found in SCCs (3 out of 6), also in higher proportions than those previously described, and occurred in mutual exclusiveness with *KEAP1* alterations[111,147], as expected for proteins that share the same functional pathway. *TP53* inactivation is thought to be a mandatory requirement to the development of SCLCs[113]. Although *TP53* mutations are reported to affect NSCLCs to a lesser extent (40 to 80%)[111,120], this study demonstrated otherwise, attesting its universal inactivation in LC. Here, defects in *TP53* were noticed in 13 out of 14 PDXs, which may partially evoke the tobacco history of the patients, once they were

all current or former smokers[163,164]. Nonetheless, the overestimation of the biallelic inactivation frequencies uncovered in the current work, which went far beyond *TP53*, is likely to translate the high purity levels of the tumorgrafts. Without the absence of normal DNA, discern the zygosity status of genetic mutations would be impracticable. Similarly, discovering HDs (as in *PTEN* and *CDKN2A*) would constitute a very arduous task. Within the group of the well-known LC genes found to be mutated were members of the SWI/SNF-MYC/MAX network (*ARID1A*, *PBRM1*, *ATM* and *MGA*), which were altered in a mutual exclusive fashion, as described elsewhere[148]. Indeed, SWI/SNF factors play an important role in cancer formation, as they often exhibit biallelic gene inactivation, consistent with tumor suppressive functions[165–167]. DNA changes in chromatin remodeling genes were more evidently noted in SCCs. Not long ago, Gonzalez-Perez and coworkers performed a comprehensive genomic study to determine the mutational scenario of chromatin regulatory factors in over 4,500 human cancers[168]. Interestingly, some of the genes that fulfilled the requirements to be considered as LC drivers in the present work are in accordance with those pinpointed by these authors as "likely drivers" (*NCOR1* and *KMT2D*, for example). Using a multiprotein approach, they demonstrated that chromatin regulatory complexes, rather than individual genes, are subject to positive selection during tumorigenesis, which is in agreement to what was discussed above for SWI/SNF members.

As denoted, the LC gene candidates discovered in this work included genes involved in several biological functions, such as epigenetic regulation, immunity, cell cycle and cell adhesion. All of these cellular

processes are encompassed within the framework of the "hallmarks of cancer" that dictate the malignant transformation of healthy cells, as proposed by Hanahan and Weinberg[30,44]. Yet, even though the application of very stringent criteria increases the chances of them being actually LC genes, studies to validate their exact role in cancer are still imperative.

The mean number of mutations per Mb was 7.5, which is in accordance with prior data for lung tumors[111]. Yet, the heterogeneity exhibited between samples possibly reflects not only the individual smoking habits, but also the type of genes that became mutated during the carcinogenic process – PDXs with dysfunctions in DNA repair, for instance, are more prone have higher mutational burdens[169].

Integrating analysis of WES and RNA-seq data showed a negative correlation trend between gene expression levels and mutation frequency, which is consistent with the transcription-coupled repair. This repair system detects DNA adducts and acts on reducing the interference with the transcription caused by DNA damage. As a result, the strands that are transcribed are more efficiently repaired, leading to a mutation prevalence bias towards the non-transcribed DNA strands[109]. This effect is frequently observed, mostly in UV light or tobacco-driven cancers, by the higher propensity of these carcinogenic agents to induce the formation of DNA adducts[144,170].

RNA-sequencing was also a reliable source to identify structural alterations. The identification of the *MYH7-NGDN* fusion transcript in PDX-3 deserved a special attention, as the reading frame of the resulting protein was preserved. The over-expression of *MYH7* triggered by genetic fusions had been previously reported in H82, a

LC cell line (fused with *PVT1*, located in chromosome 8)[150]. Apart from this cell line, fusions implicating *PVT1* were also reported in several cancer types[171,172], including in SCLC[169]. Thus, to some extent, the gene within the H82 inversion is believed to be *PVT1*. Nonetheless, despite occurring in low rates, the harmful consequences of *MYH7* over-expression in LC cannot be ruled out, and should be investigated in the future.

Taking into account the importance of validating candidate cancer genes identified by high-throughput methods, *B2M* was selected for further investigation. Data gathered in databases evidenced a very high rate of truncating mutations of *B2M* in human tumors, providing crucial support about *B2M* constituting a TSG. Furthermore, exhaustive genomic studies had been listed *B2M* as a statistically significant mutated gene in a wide variety of solid tumors (gastric, uterine, colorectal and breast cancers[173,174]). However, little was known about its status in lung carcinomas.

The results obtained in the present thesis revealed that *B2M* is altered in about 5% of lung tumors from all major histopathological groups. It is worth mentioning that this rate may be underestimated, since the genetic screening performed in primary tumors did not include specific approaches to detect HDs. In conformity with previous publications, most mutations induced the biallelic inactivation of the gene, as a consequence of LOH and truncating mutations[69–71,175,176].

B2M encodes the invariant chain of the MHC-I, β2-microglobulin. Along with HLA-I, the complex migrates to the plasma membrane and displays intracellular antigens to the $CD8^+$ T-cells. Loss of MHC-I in the cell surface occurs in 40 to 90% of human malignancies[177], and

here, it was noted in one third of the lung tumors. This phenomenon was particularly evident in SCLCs, and may reflect the neural origin of this kind of tumors[152,178]. Surpassing immune surveillance through MHC-I repression may also explain the remarkable aggressiveness and metastatic ability exhibited by SCLCs.

Even though the deficient expression of MHC-I has been occasionally associated with tumor grade and stage, and with disease progression and survival in some malignant tumors (in which lung cancer is included)[153,179,180], we could not find correlations with any clinical feature in our sample set. Nevertheless, other research groups had similar findings[66], indicating that these inconsistencies may reflect different sensitivity degrees of antibodies or distinct criteria during the staining evaluation process ("intermediate" groups may increase the discrepancies across studies). The only significant association noted here regarded the histological type, in which the expression of MHC-I was stronger among the AD subtype, remarking previous documented observations[153].

Similar to what happens in other cancer cell types[181,182], restituting the wild type form of *B2M* in gene-deficient LC cell lines recovered the expression and the correct positioning of MHC-I in the cell surface. The same outcome was achieved upon the restoration of two *B2M*-mutants (B2M*E67Q and B2M*R32H) carrying the amino acid substitutions identified in the genetic screening performed in lung tumors. In contrast, the mutation corresponding to the tryptophan-to-glycine residue change (p.W80G) impaired the migration of MHC-I to the plasma membrane. Immunofluorescence assays showed that the complex was retained somewhere around the nucleus, most likely in

the ER lumen. This specific residue, which is highly conserved across vertebrates, is thought to play a critical role in promoting the link between β2m and the alpha-1 chain of HLA-I[183]. Actually, binding tests revealed that the W80G mutant presented a 30-fold lower affinity to interact with HLA-I, due to the loss of complex stability[184].

The presentation of tumor antigens to the CTLs can be precluded by irreversible defects (permanent structural anomalies[60], which are usually triggered by LOH and other genetic alterations in *HLA-ABC* genes or in *B2M* itself[67,68,70]. Besides, it is documented that MHC-I absence is more likely originated by regulatory modifications, rather than genetic changes[185–187], and thus, it could be easily reversed by cytokine treatment with interferon. This may function in some cases, as evidenced here, with the IFNγ and IFNα treatments performed over LC cell lines. However, apart from the inactivating mutations in *B2M*, the presence of truncating alterations in other members of the MHC-I-mediated antigen processing and presentation machinery was also evidenced in this thesis. The HD in *HLA-A* (found in PDX-4 sample) and mutations in *PDIA3*, *TAP1* and *CALR* in LC cell lines underscore the role played by APMs that are usually neglected and less investigated. This suggests that the extension of the genetic alterations affecting this pathway may be underestimated, a fact that is particularly relevant in the case of HLA-I genes, as they constitute the most polymorphic genetic entities in the genome. Assessing their status in cancer is technically impracticable by NGS methods, due to the short read alignment, as unveiled by Brandt and collaborators[188]. These authors postulated that boosting sequencing coverage would not solve the problem, and contrasting polymorphic regions against multiple

reference sequences (such as the IMGT/HLA database that comprises all known HLA-I alleles) would rather be preferred. Fortunately, efforts have been made to solve these hurdles and novel pipelines had successfully enlightened the mutational frame of the HLA-I genes[189]. Mutations in the exon 9 of *CALR* were found to be rare in LC. Yet, mutations clustered in this specific spot are quite common among myeloproliferative neoplasms presenting wild type forms of *JAK2* and *MPL* genes[155,156]. Collectively, these preliminary data underline the importance of carefully screen these genes across human cancers.

As mentioned earlier, the major function of MHC-I is to display intracellular peptides to the $CD8^+$ T-lymphocytes, so they can recognize and initiate tumor cell destruction. Accordingly, this study confirmed that lung tumors harboring positive staining for MHC-I had the highest levels of CTL intratumoral infiltration. Loss of β2m had previously been associated with lower CTL infiltration in melanoma[71], in papillary thyroid cancer[56], in ovarian carcinoma[190] and even in NSCLC[153]. These facts highlight that displaying tumor epitopes across the plasma membrane triggers the recruitment of tumor-infiltrating $CD8^+$ lymphocytes. Unlike CTLs, NK cell-mediating killing is independent on MHC-I molecules – in fact, they serve as inhibitory signals, since NK cells recognize the "missing self". Therefore, one could speculate that NK-based therapies could be employed in tumors harboring dysfunctions in MHC-I. Nevertheless, despite the initial promising anticancer applications, the therapeutic activity of NKs in clinical trials has proven to be very limited[191,192].

The full activation of CTLs relies on a fine-tuned balance between stimulatory (through MHC-I-dependent peptide presentation) and

inhibitory signals (molecules that promote immune tolerance). One of those inhibitory stimuli is given by PD-1/PD-L1 interaction, and the pathway blockage through the administration of mAbs represents the foundation of current anticancer immunotherapy[158]. The importance of anti-PD-1/PD-L1 drugs in LC is supported by the prevalence of PD-L1 expression in lung tumors, as denoted in this work and elsewhere[193,194]. Numerous studies correlated PD-L1 tumor positivity with increased response rates to these agents[85,87,195,196]. However, the use of PD-L1 expression as a predictive biomarker has becoming a matter of controversy, once patients with PD-L1 negative tumors started driving benefit from PD-1/PD-L1 mAbs, with response rates of 10 to 14%[195,197]. The growing evidence indicates that tumor PD-L1 expression may not be able to predict the overall benefit of such immunotherapy in LC patients.

Here, the display of MHC-I across the plasma membrane occurred concomitantly not only with higher intratumoral CTL infiltration, but also with positive PD-L1 staining. These interconnections are on the line of their known biological functions and are thought to be critical in forecasting immune response to anti-PD1/PD-L1 blocking drugs. In fact, even in the case of PD-L1 positive tumors, in the absence of tumor antigens on the surface, $CD8^+$ T-cells would not be able to recognize cancer cells and modest or no response rates would be expected in patients treated with these agents.

Although performed in a small cohort of PD-L1 positive patients, the results obtained here showed that exhibiting MHC-I in the membrane (which was coupled with high intratumoral CTL infiltration grades) conferred better response to anti-PD1/PD-L1 immunotherapy. CTLs

constitute the utmost effectors of killing during PD-1/PD-L1 blockade[82], and thus, evaluating tumor infiltration degrees should be implemented, in order to improve patient selectivity. Furthermore, the acquired resistance to immunotherapy caused by *B2M* loss was reported in melanoma[198,199], emphasizing that the significance attributed to β2m (and consequently to the MHC-I) may be certainly underestimated within the immunotherapeutic context. Indeed, despite the undeniable role of MHC-I in mediating tumor cell eradication, its effect on predicting response rates to anti-PD1/PD-L1 agents had not been tested before.

Hopefully, the observations reported in this thesis will prompt clinical trials with large cohorts of LC patients to ultimately validate the use of β2m and CTL infiltration as potent biomarkers in predicting response to these promising therapies.

6. CONCLUSIONS

The results presented in the current thesis support the establishment of the following conclusions:

- Tumorgrafts are very convenient models to use in genomic profiling, as they favor the increase of tumor purity, avoiding the admixture between malignant and normal cells.

- The genetic signature of PDXs identified recurrent alterations in genes widely known for its role in lung pathogenesis. For some of the genes, mutation frequencies were higher than those previously reported, denoting that genetic modifications are more efficiently detected by sequencing tumorgrafts, rather than primary tumors.

- The combination of WES with RNA-sequencing reinforces the reliability of the results, smoothes the selection of putative oncogenes and tumor suppressors, and assists in discarding passenger mutations. Moreover, transcriptomes offer additional data on chimeric proteins with possible implications in LC.

- The candidate gene *B2M* was biallelically inactivated in 5% of LCs from several histopathological types, a pattern that indicates its tumor suppressive nature.

- Although without any detectable outcomes on proliferation, the restitution of the wild type *B2M* in gene-deficient LC cell lines restored the correct positioning of MHC-I in the cell surface. These findings emphasize that the inactivation of *B2M* constitutes a mechanism that promotes tumor progression, in which malignant cells are allowed to escape from immune surveillance.

- Recovering the expression of the *B2M*-mutant form carrying the p.W80G residue change impairs the appropriate binding to the heavy chain (HLA-I), precluding the migration of the MHC-I to the plasma membrane. This fact underpins the deleterious properties of the c.238T>G mutation.

- The absence or down-regulation of β2m and HLA-I proteins occurs in one third of lung carcinomas. However, the MHC-I loss in the cell membrane may also be driven by genetic abnormalities in other genes directly involved in the antigen processing and presentation, such as *CALR*, *HLA-A*, *TAP1* and *PDIA3*.

- Preliminary findings suggest that higher levels of MHC-I expression and of $CD8^+$ T-cell infiltration confer better response to PD-1/PD-L1

checkpoint-blocking treatments. Consequently, these two concepts should be included within the clinical context as powerful biomarkers in foretelling responses to immune-based therapies in LC patients.

7. BIBLIOGRAPHY

1. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194,** 23–8 (1976).

2. Linder, D. & Gartler, S. M. Glucose-6-phosphate dehydrogenase mosaicism: utilization as a cell marker in the study of leiomyomas. *Science* **150,** 67–9 (1965).

3. Torre, L. A. *et al.* Global Cancer Statistics, 2012. *CA Cancer J Clin* **65,** 87–108 (2015).

4. Yancik, R. Cancer burden in the aged: An epidemiologic and demographic overview. *Cancer* **80,** 1273–1283 (1997).

5. Peto, R., Roe, F. J., Lee, P. N., Levy, L. & Clack, J. Cancer and ageing in mice and men. *Br. J. Cancer* **32,** 411–426 (1975).

6. Anisimov, V. N. Carcinogenesis and aging. *Adv. Cancer Res.* **40,** 365–424 (1983).

7. DePinho, R. a. The age of cancer. *Nature* **408,** 248–54 (2000).

8. Anand, P. *et al.* Cancer is a preventable disease that requires major lifestyle changes. *Pharm. Res.* **25,** 2097–116 (2008).

9. Wu, S., Powers, S., Zhu, W. & Hannun, Y. A. Substantial contribution of extrinsic risk factors to cancer development. *Nature* **529,** 43–47 (2016).

10. Johnson, I. T. & Lund, E. K. Review article: Nutrition, obesity and colorectal cancer. *Aliment. Pharmacol. Ther.* **26,** 161–181 (2007).

11. Parkin, D. M. Cancers attributable to exposure to hormones in the UK in 2010. *Br. J. Cancer* **105,** S42–8 (2011).

12. Vineis, P. *et al.* Tobacco and cancer: recent epidemiological evidence. *J. Natl. Cancer Inst.* **96,** 99–106 (2004).

13. Parkin, D. M. The global health burden of infection-associated cancers in the year 2002. *Int. J. Cancer* **118,** 3030–3044 (2006).

14. De Martel, C. *et al.* Global burden of cancers attributable to infections in 2008: A review and synthetic analysis. *Lancet Oncol.* **13,** 607–615 (2012).

15. Uemura, N. *et al.* Helicobacter pylori Infection and the Development of Gastric Cancer. *N. Engl. J. Med.* **345,** 784–789 (2001).

16. Munoz N, Bosch FX, de Sanjose S, Herrero R, Castellsague X,

Shah KV, Snijders PJF, M. C. Epidemiologic Classification of Human Papillomavirus Types Associated with Cervical Cancer. *N. Engl. J. Med.* **348,** 518–527 (2003).

17. El-Serag, H. B. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology* **142,** 1264–1273.e1 (2012).

18. Garber, J. E. Hereditary Cancer Predisposition Syndromes. *J. Clin. Oncol.* **23,** 276–292 (2005).

19. Li, F. P. & Fraumeni, J. F. Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? *Ann. Intern. Med.* **71,** 747–752 (1969).

20. Hwang, S.-J., Lozano, G., Amos, C. I. & Strong, L. C. Germline p53 mutations in a cohort with childhood sarcoma: sex differences in cancer risk. *Am. J. Hum. Genet.* **72,** 975–83 (2003).

21. Varley, J. M. Germline TP53 mutations and Li-Fraumeni syndrome. *Hum. Mutat.* **21,** 313–320 (2003).

22. Vasen, H. F. *et al.* Cancer risk in families with hereditary nonpolyposis colorectal cancer diagnosed by mutation analysis [published erratum appears in Gastroenterology 1996 Nov;111(5):1402]. *Gastroenterology* **110,** 1020–1027 (1996).

23. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458,** 719–724 (2009).

24. Futreal, P. A. *et al.* A census of human cancer genes. *Nat. Rev. Cancer* **4,** 177–183 (2004).

25. Shaw, R. J. & Cantley, L. C. Ras, PI(3)K and mTOR signalling controls tumour cell growth. *Nature* **441,** 424–430 (2006).

26. Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417,** 949–954 (2002).

27. Santarius, T., Shipley, J., Brewer, D., Stratton, M. R. & Cooper, C. S. A census of amplified and overexpressed human cancer genes. *Nat. Rev. Cancer* **10,** 59–64 (2010).

28. Mitelman, F., Johansson, B. & Mertens, F. The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer* **7,** 233–245 (2007).

29. Croce, C. M. Oncogenes and Cancer. *N. Engl. J. Med.* **358,** 502–511 (2008).

30. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100,** 57–70 (2000).

31. Macleod, K. Tumor suppressor genes. *Curr. Opin. Genet. Dev.* **10,** 81–93 (2000).

32. Knudson, A. & G. Mutation and Cancer: Statistical Study of Retinoblastoma. *Proc. Natl. Acad. Sci. U. S. A.* **68,** 820–823 (1971).

33. Knudson, A. G. Cancer genetics. *Am. J. Med. Genet.* **111,** 96–102 (2002).

34. Dulbecco, R. A turning point in cancer research: sequencing the human genome. *Science (80-.).* **231,** 1055–1056 (1986).

35. International Human Genome Sequencing Consortium.

Finishing the euchromatic sequence of the human genome. *Nature* **431,** 931–45 (2004).

36. Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456,** 66–72 (2008).

37. O'Keefe, B. The future of cancer genomics. *Nat. Med.* **21,** 99–99 (2015).

38. Weinberg, R. A. The retinoblastoma protein and cell cycle control. *Cell* **81,** 323–30 (1995).

39. Adams, J. M. & Cory, S. The Bcl-2 apoptotic switch in cancer development and therapy. *Oncogene* **26,** 1324–37 (2007).

40. Samuels, Y. *et al.* Mutant PIK3CA promotes cell growth and invasion of human cancer cells. *Cancer Cell* **7,** 561–573 (2005).

41. Cantley, L. C. & Neel, B. G. New insights into tumor suppression: PTEN suppresses tumor formation by restraining the phosphoinositide 3-kinase/AKT pathway. *Proc. Natl. Acad. Sci.* **96,** 4240–4245 (1999).

42. Veikkola, T. & Alitalo, K. VEGFs, receptors and angiogenesis. *Semin. Cancer Biol.* **9,** 211–220 (1999).

43. Liekens, S., De Clercq, E. & Neyts, J. Angiogenesis: regulators and clinical applications. *Biochem. Pharmacol.* **61,** 253–70 (2001).

44. Hanahan, D. & Weinberg, R. a. Hallmarks of cancer: The next generation. *Cell* **144,** 646–674 (2011).

45. Ehrlich, P. Über Den Jetzigen Stand Der Karzinomforschung. *Beiträge zur experimentellen Pathologie und Chemotherapie* 117–164

(1909).

46. Burnet, M. Cancer - A biological approach. *Br. Med. J.* **1,** 841–847 (1957).

47. Shankaran, V. *et al.* IFNγ and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature* **410,** 1107–1111 (2001).

48. Street, S. E. a, Trapani, J. a, MacGregor, D. & Smyth, M. J. Suppression of lymphoma and epithelial malignancies effected by interferon gamma. *J. Exp. Med.* **196,** 129–134 (2002).

49. Willimsky, G. & Blankenstein, T. Sporadic immunogenic tumours avoid destruction by inducing T-cell tolerance. *Nature* **437,** 141–146 (2005).

50. Dunn, G. P., Old, L. J. & Schreiber, R. D. The three Es of cancer immunoediting. *Annu. Rev. Immunol.* **22,** 329–60 (2004).

51. Dunn, G. P., Old, L. J. & Schreiber, R. D. The immunobiology of cancer immunosurveillance and immunoediting. *Immunity* **21,** 137–48 (2004).

52. Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J. & Schreiber, R. D. Cancer immunoediting: from immunosurveillance to tumor escape. *Nat. Immunol.* **3,** 991–998 (2002).

53. Strausberg, R. L. Tumor microenvironments, the immune system and cancer survival. *Genome Biol.* **6,** 211 (2005).

54. Groothuis, T. A. M., Griekspoor, A. C., Neijssen, J. J., Herberts, C. A. & Neefjes, J. J. MHC class I alleles and their

exploration of the antigen-processing machinery. *Immunol. Rev.* **207,** 60–76 (2005).

55. Raghavan, M., Del Cid, N., Rizvi, S. M. & Peters, L. R. MHC class I assembly: out and about. *Trends Immunol.* **29,** 436–443 (2008).

56. Angell, T. E., Lechner, M. G., Jang, J. K., LoPresti, J. S. & Epstein, A. L. MHC class I loss is a frequent mechanism of immune escape in papillary thyroid cancer that is reversed by interferon and selumetinib treatment in vitro. *Clin. Cancer Res.* **20,** 6034–44 (2014).

57. N.P., R. *et al.* Identification of human cancers deficient in antigen processing. *J. Exp. Med.* **177,** 265–272 (1993).

58. Mizukami, Y. *et al.* Downregulation of HLA Class I molecules in the tumour is associated with a poor prognosis in patients with oesophageal squamous cell carcinoma. *Br. J. Cancer* **99,** 1462–1467 (2008).

59. Cabrera, T. *et al.* High frequency of altered HLA class I phenotypes in invasive breast carcinomas. *Hum. Immunol.* **50,** 127–134 (1996).

60. Garrido, F., Cabrera, T. & Aptsiauri, N. 'Hard' and 'soft' lesions underlying the HLA class I alterations in cancer cells: Implications for immunotherapy. *Int. J. Cancer* **127,** 249–256 (2010).

61. Méndez, R. *et al.* Characterization of HLA class I altered phenotypes in a panel of human melanoma cell lines. *Cancer*

Immunol. Immunother. **57,** 719–729 (2008).

62. Cabrera, C. M., López-Nevot, M. Á., Jiménez, P. & Garrido, F. Involvement of the chaperone tapasin in HLA-B44 allelic losses in colorectal tumors. *Int. J. Cancer* **113,** 611–618 (2005).

63. Seliger, B. Molecular mechanisms of MHC class I abnormalities and APM components in human tumors. *Cancer Immunol. Immunother.* **57,** 1719–1726 (2008).

64. Seliger, B. *et al.* Analysis of the major histocompatibility complex class I antigen presentation machinery in normal and malignant renal cells: evidence for deficiencies associated with transformation and progression. *Cancer Res.* **56,** 1756–60 (1996).

65. Vitale, M. *et al.* HLA class I antigen and transporter associated with antigen processing (TAP1 and TAP2) down-regulation in high-grade primary breast carcinoma lesions. *Cancer Res.* **58,** 737–42 (1998).

66. Romero, J. M. *et al.* Coordinated downregulation of the antigen presentation machinery and HLA class I/β2-microglobulin complex is responsible for HLA-ABC loss in bladder cancer. *Int. J. Cancer* **113,** 605–610 (2005).

67. Maleno, I., Lopez-Nevot, M., Cabrera, T., Salinero, J. & Garrido, F. Multiple mechanisms generate HLA class I altered phenotypes in laryngeal carcinomas: High frequency of HLA haplotype loss associated with loss of heterozygosity in chromosome region 6p21. *Cancer Immunol. Immunother.* **51,** 389–396 (2002).

68. Maleno, I. *et al.* LOH at 6p21.3 region and HLA class altered phenotypes in bladder carcinomas. *Immunogenetics* **58,** 503–510 (2006).

69. Benitez, R. *et al.* Mutations of the beta2-microglobulin gene result in a lack of HLA class I molecules on melanoma cells of two patients immunized with MAGE peptides. *Tissue Antigens* **52,** 520–529 (1998).

70. Maleno, I. *et al.* Frequent loss of heterozygosity in the β2-microglobulin region of chromosome 15 in primary human tumors. *Immunogenetics* **63,** 65–71 (2011).

71. Del Campo, A. B. *et al.* Immune escape of cancer cells with beta2-microglobulin loss over the course of metastatic melanoma. *Int. J. Cancer* **134,** 102–113 (2014).

72. Lesterhuis, W. J., Haanen, J. B. A. G. & Punt, C. J. A. Cancer immunotherapy--revisited. *Nat. Rev. Drug Discov.* **10,** 591–600 (2011).

73. Galluzzi, L. *et al.* Classification of current anticancer immunotherapies. *Oncotarget* **5,** 12472–508 (2014).

74. Moore, P. S. & Chang, Y. Why do viruses cause cancer? Highlights of the first century of human tumour virology. *Nat. Rev. Cancer* **10,** 878–889 (2010).

75. Garland, S. M. *et al.* Quadrivalent Vaccine against Human Papillomavirus to Prevent Anogenital Diseases. *N. Engl. J. Med.* **356,** 1928–1943 (2007).

76. Paavonen, J. *et al.* Efficacy of human papillomavirus (HPV)-

16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): final analysis of a double-blind, randomised study in young women. *Lancet* **374,** 301–314 (2009).

77. Melero, I. *et al.* Therapeutic vaccines for cancer: an overview of clinical trials. *Nat. Rev. Clin. Oncol.* **11,** 509–524 (2014).

78. Kantoff, P. W. *et al.* Sipuleucel-T Immunotherapy for Castration-Resistant Prostate Cancer. *N. Engl. J. Med.* **363,** 411–422 (2010).

79. Robak, T. Rituximab for chronic lymphocytic leukemia. *Expert Opin. Biol. Ther.* **12,** 503–515 (2012).

80. Scott, S. D. Rituximab: A new therapeutic monoclonal antibody for non-Hodgkin's lymphoma. *Cancer Practice* **6,** 195–197 (1998).

81. Slamon, D. J. *et al.* Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.* **344,** 783–92 (2001).

82. Pardoll, D. M. The blockade of immune checkpoints in cancer immunotherapy. *Nat. Rev. Cancer* **12,** 252–264 (2012).

83. Melero, I., Grimaldi, A. M., Perez-Gracia, J. L. & Ascierto, P. a. Clinical development of immunostimulatory monoclonal antibodies and opportunities for combination. *Clin. Cancer Res.* **19,** 997–1008 (2013).

84. Hodi, F. S. *et al.* Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. *N. Engl. J. Med.* **363,** 711–723 (2010).

85. Topalian, S. L. *et al.* Safety, Activity, and Immune Correlates of Anti–PD-1 Antibody in Cancer. *N. Engl. J. Med.* **366,** 2443–2454 (2012).

86. Powles, T. *et al.* MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer. *Nature* **515,** 558–62 (2014).

87. Borghaei, H. *et al.* Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **373,** 1627–39 (2015).

88. Bagley, S. J., Bauml, J. M. & Langer, C. J. PD-1/PD-L1 Immune Checkpoint Blockade in Non–Small Cell Lung Cancer. *Clin. Adv. Hematol. Oncol.* **13,** 676–683 (2015).

89. Ferlay, J. *et al.* Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* **127,** 2893–2917 (2010).

90. Jemal, A., Bray, F. & Ferlay, J. Global Cancer Statistics: 2011. *CA Cancer J Clin* **49,** 69–90 (2011).

91. American Cancer Society. Cancer facts & figures 2016. (2016).

92. Pesch, B. *et al.* Cigarette smoking and lung cancer—relative risk estimates for the major histological types from a pooled analysis of case–control studies. *Int. J. Cancer* **131,** 1210–1219 (2012).

93. Shopland, D. R. Tobacco use and its contribution to early cancer mortality with a special emphasis on cigarette smoking. *Environ. Health Perspect.* **103,** 131–142 (1995).

94. Hainaut, P. & Pfeifer, G. P. Patterns of p53 G-->T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke. *Carcinogenesis* **22,** 367–74 (2001).

95. Smith, L. E. *et al.* Targeting of lung cancer mutational hotspots by polycyclic aromatic hydrocarbons. *J. Natl. Cancer Inst.* **92,** 803–811 (2000).

96. Akopyan, G. & Bonavida, B. Understanding tobacco smoke carcinogen NNK and lung tumorigenesis. *Int. J. Oncol.* **29,** 745–52 (2006).

97. Minna, J. D., Roth, J. a. & Gazdar, A. F. Focus on lung cancer. *Cancer Cell* **1,** 49–52 (2002).

98. Thun, M. J. *et al.* Lung cancer death rates in lifelong nonsmokers. *J. Natl. Cancer Inst.* **98,** 691–699 (2006).

99. Badar, F. *et al.* Characteristics of lung cancer patients - the Shaukat Khanum memorial experience. *Asian Pacific J. Cancer Prev.* **7,** 245–248 (2006).

100. Hubaux, R. *et al.* Arsenic, asbestos and radon: emerging players in lung tumorigenesis. *Environ. Health* **11,** 89 (2012).

101. Janerich, D. T. *et al.* Lung Cancer and Exposure to Tobacco Smoke in the Household. *N. Engl. J. Med.* **323,** 632–636 (1990).

102. Hecht, S. S. *et al.* A tobacco-specific lung carcinogen in the urine of men exposed to cigarette smoke. *N. Engl. J. Med.* **329,** 1543–6 (1993).

103. Travis, W. D., Brambilla, E., Müller-Hermelink, H. K. & Harris, C. World Health Organization classification of tumours; tumours of lung, pleura, thymus and heart. *World Heal. Organ. Classif. tumours; tumours lung, pleura, thymus Hear.* 9–122 (2004).

104. Herbst, R. S., Heymach, J. V. & Lippman, S. M. Lung Cancer. *N. Engl. J. Med.* **359,** 1367–1380 (2008).

105. Howlader, N. *et al.* SEER Cancer Statistics Review 1975-2008 National Cancer Institute SEER Cancer Statistics Review 1975-2008 National Cancer Institute. *Cancer* **2006,** 1992–2008 (2011).

106. Zamecnik, J. & Kodet, R. Value of thyroid transcription factor-1 and surfactant apoprotein A in the differential diagnosis of pulmonary carcinomas: A study of 109 cases. *Virchows Arch.* **440,** 353–361 (2002).

107. Travis, W. D., Brambilla, E. & Riely, G. J. New Pathologic Classification of Lung Cancer: Relevance for Clinical Practice and Clinical Trials. *J. Clin. Oncol.* **31,** 992–1001 (2013).

108. Peto, R. *et al.* Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. *BMJ* **321,** 323–329 (2000).

109. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–21 (2013).

110. Levy, M. a., Lovly, C. M. & Pao, W. Translating genomic information into clinical medicine: Lung cancer as a paradigm. *Genome Res.* **22,** 2101–2108 (2012).

111. Hammerman, P. *et al.* Comprehensive genomic characterization

of squamous cell lung cancers. *Nature* **489,** 519–525 (2012).

112. Collisson, E. A. *et al.* Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511,** 543–550 (2014).

113. George, J. *et al.* Comprehensive genomic profiles of small cell lung cancer. *Nature* **524,** 47–53 (2015).

114. Li, T., Kung, H.-J., Mack, P. C. & Gandara, D. R. Genotyping and Genomic Profiling of Non–Small-Cell Lung Cancer: Implications for Current and Future Therapies. *J. Clin. Oncol.* **31,** 1039–1049 (2013).

115. Sharma, S. V, Bell, D. W., Settleman, J. & Haber, D. a. Epidermal growth factor receptor mutations in lung cancer. *Nat. Rev. Cancer* **7,** 169–181 (2007).

116. Pao, W. *et al.* EGF receptor gene mutations are common in lung cancers from 'never smokers' and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 13306–13311 (2004).

117. Yarden, Y. & Sliwkowski, M. X. Untangling the ErbB signalling network. *Nat. Rev. Mol. Cell Biol.* **2,** 127–137 (2001).

118. Maemondo, M. *et al.* Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *N. Engl. J. Med.* **362,** 2380–2388 (2010).

119. Stephens, P. *et al.* Lung cancer: intragenic ERBB2 kinase mutations in tumours. *Nature* **431,** 525–6 (2004).

120. Ding, L. *et al.* Somatic mutations affect key pathways in lung

adenocarcinoma. *Nature* **455,** 1069–1075 (2008).

121. Riely, G. J. *et al.* Frequency and distinctive spectrum of KRAS mutations in never smokers with lung adenocarcinoma. *Clin. Cancer Res.* **14,** 5731–5734 (2008).

122. Rodenhuis, S. & Slebos, R. J. Clinical significance of ras oncogene activation in human lung cancer. *Cancer Res.* **52,** 2665s–2669s (1992).

123. Marchetti, A. *et al.* Clinical features and outcome of patients with non-small-cell lung cancer harboring BRAF mutations. *J Clin Oncol* **29,** 3574–3579 (2011).

124. Vivanco, I. & Sawyers, C. L. The phosphatidylinositol 3-Kinase-AKT pathway in human cancer. *Nat Rev Cancer* **2,** 489–501 (2002).

125. Yamamoto, H. *et al.* PIK3CA mutations and copy number gains in human lung cancers. *Cancer Res.* **68,** 6913–6921 (2008).

126. Malanga, D. *et al.* Activating E17K mutation in the gene encoding the protein kinase AKT in a subset of squamous cell carcinoma of the lung. *Cell Cycle* **7,** 665–669 (2008).

127. Matsumoto, S. *et al.* Prevalence and specificity of LKB1 genetic alterations in lung cancers. *Oncogene* **26,** 5911–8 (2007).

128. Sanchez-Cespedes, M. The role of LKB1 in lung cancer. *Fam. Cancer* **10,** 447–453 (2011).

129. Jin, G. *et al.* PTEN mutations and relationship to EGFR, ERBB2, KRAS, and TP53 mutations in non-small cell lung

cancers. *Lung Cancer* **69,** 279–283 (2010).

130. Imielinski, M. *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150,** 1107–1120 (2012).

131. Husgafvel-Pursiainen, K. *et al.* P53 Mutations and Exposure To Environmental Tobacco Smoke in a Multicenter Study on Lung Cancer. *Cancer Res.* **60,** 2906–2911 (2000).

132. Blanco, R. *et al.* A gene-alteration profile of human lung cancer cell lines. *Hum. Mutat.* **30,** 1199–1206 (2009).

133. Soda, M. *et al.* Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* **448,** 561–566 (2007).

134. Rikova, K. *et al.* Global Survey of Phosphotyrosine Signaling Identifies Oncogenic Kinases in Lung Cancer. *Cell* **131,** 1190–1203 (2007).

135. Wong, D. W. S. *et al.* The EML4-ALK fusion gene is involved in various histologic types of lung cancers from nonsmokers with wild-type EGFR and KRAS. *Cancer* **115,** 1723–1733 (2009).

136. Shaw, A. T. *et al.* Effect of crizotinib on overall survival in patients with advanced non-small-cell lung cancer harbouring ALK gene rearrangement: A retrospective analysis. *Lancet Oncol.* **12,** 1004–1012 (2011).

137. Bergethon, K. *et al.* ROS1 rearrangements define a unique molecular class of lung cancers. *J. Clin. Oncol.* **30,** 863–870

(2012).

138. D'Arcangelo, M., Wynes, M. W. & Hirsch, F. R. The role of anaplastic lymphoma kinase inhibitors in the treatment of advanced nonsmall cell lung cancer. *Curr. Opin. Oncol.* **25,** 121–9 (2013).

139. Ambrogio, C. *et al.* Modeling lung cancer evolution and preclinical response by orthotopic mouse allografts. *Cancer Res.* **74,** 5978–5988 (2014).

140. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339,** 1546–58 (2013).

141. Cerami, E. *et al.* The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2,** 401–404 (2012).

142. Forbes, S. A. *et al.* COSMIC: Mining complete cancer genomes in the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **39,** 945–950 (2011).

143. Forbes, S. A. *et al.* COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43,** D805–D811 (2015).

144. Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463,** 191–196 (2010).

145. Wang, L. *et al.* Whole-exome sequencing of human pancreatic cancers and characterization of genomic instability caused by MLH1 haploinsufficiency and complete deficiency. *Genome Res.*

22, 208–219 (2012).

146. Stratton, M. R. Exploring the genomes of cancer cells: progress and promise. *Science* **331,** 1553–1558 (2011).

147. The Clinical Lung Cancer Genome Project (CLCGP) and Network Genomic Medicine (NGM), „. A genomics-based classification of human lung tumors. *Sci. Transl. Med.* **5,** 209ra153 (2013).

148. Romero, O. A. *et al.* MAX inactivation in small cell lung cancer disrupts MYC-SWI/SNF programs and is synthetic lethal with BRG1. *Cancer Discov.* **4,** 292–303 (2014).

149. Barretina, J., Caponigro, G. & Stransky, N. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483,** 603–607 (2012).

150. Iwakawa, R. *et al.* Genome-wide identification of genes with amplification and/or fusion in small cell lung cancer. *Genes, Chromosom. Cancer* **52,** 802–816 (2013).

151. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7,** 248–249 (2010).

152. Doyle, A. *et al.* Markedly decreased expression of class I histocompatibility antigens, protein, and mRNA in human small-cell lung cancer. *J. Exp. Med.* **161,** 1135–51 (1985).

153. Kikuchi, E. *et al.* HLA class I antigen expression is associated with a favorable prognosis in early stage non-small cell lung cancer. *Cancer Sci.* **98,** 1424–1430 (2007).

154. Yang, I. *et al.* Modulation of major histocompatibility complex Class I molecules and major histocompatibility complex-bound immunogenic peptides induced by interferon-alpha and interferon-gamma treatment of human glioblastoma multiforme. *J. Neurosurg.* **100,** 310–319 (2004).

155. Nangalia, J. *et al.* Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N. Engl. J. Med.* **369,** 2391–405 (2013).

156. Klampfl, T. *et al.* Somatic mutations of calreticulin in myeloproliferative neoplasms. *N. Engl. J. Med.* **369,** 2379–90 (2013).

157. Govindan, R. *et al.* Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* **150,** 1121–1134 (2012).

158. Topalian, S. L., Drake, C. G. & Pardoll, D. M. Targeting the PD-1/B7-H1(PD-L1) pathway to activate anti-tumor immunity. *Curr. Opin. Immunol.* **24,** 207–212 (2012).

159. Weir, B. a *et al.* Characterizing the cancer genome in lung adenocarcinoma. *Nature* **450,** 893–898 (2007).

160. DeRose, Y. S. *et al.* Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. *Nat. Med.* **17,** 1514–1520 (2011).

161. Daniel, V. C. *et al.* A primary xenograft model of small-cell lung cancer reveals irreversible changes in gene expression imposed

by culture in vitro. *Cancer Res.* **69,** 3364–3373 (2009).

162. Reyal, F. *et al.* Molecular profiling of patient-derived breast cancer xenografts. *Breast Cancer Res.* **14,** R11 (2012).

163. Peifer, M. *et al.* Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nat. Genet.* **44,** 1104–1110 (2012).

164. Halvorsen, A. R. *et al.* TP53 Mutation Spectrum in Smokers and Never Smoking Lung Cancer Patients. *Front. Genet.* **7,** 85 (2016).

165. Medina, P. P. & Sanchez Cespedes, M. Involvement of the chromatin-remodeling factor BRG1/SMARCA4 in human cancer. *Epigenetics* **3,** 64–68 (2008).

166. Wilson, B. G. & Roberts, C. W. SWI/SNF nucleosome remodellers and cancer. *Nat Rev Cancer* **11,** 481–492 (2011).

167. Romero, O. a & Sanchez-Cespedes, M. The SWI/SNF genetic blockade: effects in cell differentiation, cancer and developmental diseases. *Oncogene* **33,** 2681–9 (2014).

168. Gonzalez-Perez, A., Jene-Sanz, A. & Lopez-Bigas, N. The mutational landscape of chromatin regulatory factors across 4,623 tumor samples. *Genome Biol.* **14,** r106 (2013).

169. Pleasance, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463,** 184–190 (2010).

170. Lee, W. *et al.* The mutation spectrum revealed by paired

genome sequences from a lung cancer patient. *Nature* **465,** 473–477 (2010).

171. Tseng, Y.-Y. *et al.* PVT1 dependence in cancer with MYC copy-number increase. *Nature* **512,** 82–86 (2014).

172. L'Abbate, A. *et al.* Genomic organization and evolution of double minutes/homogeneously staining regions with MYC amplification in human cancer. *Nucleic Acids Res.* **42,** 9131–9145 (2014).

173. Bass, A. J. *et al.* Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513,** 202–9 (2014).

174. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160,** 48–61 (2015).

175. Paschen, A. *et al.* The coincidence of chromosome 15 aberrations and beta2-microglobulin gene mutations is causative for the total loss of human leukocyte antigen class I expression in melanoma. *Clin. Cancer Res.* **12,** 3297–305 (2006).

176. Feenstra, M. *et al.* HLA class I expression and chromosomal deletions at 6p and 15q in head and neck squamous cell carcinomas. *Tissue Antigens* **54,** 235–45 (1999).

177. Bubeník, J. MHC class I down-regulation: tumour escape from immune surveillance? (review). *International journal of oncology* **25,** 487–491 (2004).

178. Funa, K., Gazdar, A. F., Minna, J. D. & Linnoila, R. I. Paucity of (beta)2-microglobulin expression on small cell lung cancer,

bronchial carcinoids, and certain other neuroendocrine tumors. *Lab. Investig.* **55,** 186–193 (1986).

179. Kageshita, T., Hirai, S., Ono, T., Hicklin, D. J. & Ferrone, S. Down-Regulation of HLA Class I Antigen-Processing Molecules in Malignant Melanoma. *Am. J. Pathol.* **154,** 745–754 (1999).

180. Tsukahara, T. *et al.* Prognostic significance of HLA class I expression in osteosarcoma defined by anti-pan HLA class I monoclonal antibody, EMR8-5. *Cancer Sci.* **97,** 1374–1380 (2006).

181. Imanishi, T. *et al.* Correlation between expression of major histocompatibility complex class I and that of antigen presenting machineries in carcinoma cell lines of the pancreas, biliary tract and colon. *Kobe J. Med. Sci.* **52,** 85–95 (2006).

182. Del Campo, A. B. *et al.* Adenovirus expressing β2-microglobulin recovers HLA class I expression and antitumor immunity by increasing T-cell recognition. *Cancer Gene Ther.* **21,** 317–32 (2014).

183. Natalello, A. *et al.* Wild type beta-2 microglobulin and DE loop mutants display a common fibrillar architecture. *PLoS One* **10,** 1–14 (2015).

184. Esposito, G. *et al.* The Controlling Roles of Trp60 and Trp95 in β2-Microglobulin Function, Folding and Amyloid Aggregation Properties. *J. Mol. Biol.* **378,** 885–895 (2008).

185. Delp, K., Momburg, F., Hilmes, C., Huber, C. & Seliger, B.

Functional deficiencies of components of the MHC class I antigen pathway in human tumors of epithelial origin. *Bone Marrow Transplant.* **25 Suppl 2,** S88–95 (2000).

186. Campoli, M. & Ferrone, S. HLA antigen changes in malignant cells: epigenetic mechanisms and biologic significance. *Oncogene* **27,** 5869–85 (2008).

187. Paulson, K. G. *et al.* Downregulation of MHC-I Expression Is Prevalent but Reversible in Merkel Cell Carcinoma. *Cancer Immunol. Res.* **2,** 1071–1079 (2014).

188. Brandt, D. Y. C. *et al.* Mapping Bias Overestimates Reference Allele Frequencies at the HLA Genes in the 1000 Genomes Project Phase I Data. *G3 (Bethesda).* **5,** 931–41 (2015).

189. Shukla, S. A. *et al.* Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* **33,** 1152–1158 (2015).

190. Han, L. *et al.* HLA class i antigen processing machinery component expression and intratumoral T-Cell infiltrate as independent prognostic markers in ovarian carcinoma. *Clin. Cancer Res.* **14,** 3372–3379 (2008).

191. Dahlberg, C. I. M., Sarhan, D., Chrobok, M., Duru, A. D. & Alici, E. Natural killer cell-based therapies targeting cancer: Possible strategies to gain and sustain anti-tumor activity. *Front. Immunol.* **6,** (2015).

192. Cheng, M., Chen, Y., Xiao, W., Sun, R. & Tian, Z. NK cell-based immunotherapy for malignant diseases. *Cell. Mol.*

Immunol. **10,** 230–52 (2013).

193. Taube, J. M. Unleashing the immune system: PD-1 and PD-Ls in the pre-treatment tumor microenvironment and correlation with response to PD-1/PD-L1 blockade. *Oncoimmunology* **3,** e963413 (2014).

194. Pan, Z.-K., Ye, F., Wu, X., An, H.-X. & Wu, J.-X. Clinicopathological and prognostic significance of programmed cell death ligand1 (PD-L1) expression in patients with non-small cell lung cancer: a meta-analysis. *J. Thorac. Dis.* **7,** 462–70 (2015).

195. Passiglia, F. *et al.* PD-L1 expression as predictive biomarker in patients with NSCLC: a pooled analysis. *Oncotarget* **7,** (2016).

196. Garon, E. B. *et al.* Pembrolizumab for the treatment of non-small-cell lung cancer. *N. Engl. J. Med.* **372,** 2018–28 (2015).

197. Rizvi, N. A. *et al.* Activity and safety of nivolumab, an anti-PD-1 immune checkpoint inhibitor, for patients with advanced, refractory squamous non-small-cell lung cancer (CheckMate 063): A phase 2, single-arm trial. *Lancet Oncol.* **16,** 257–265 (2015).

198. Restifo, N. P. *et al.* Loss of Functional Beta2-Microglobulin in Metastatic Melanomas From Five Patients Receiving Immunotherapy. *JNCI J. Natl. Cancer Inst.* **88,** 100–108 (1996).

199. Zaretsky, J. M. *et al.* Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma. *N. Engl. J. Med.* (2016).

8. SUPPLEMENTARY MATERIAL

Additional information to support the results obtained in the current work is gathered in this section. Tables and figures are displayed by citation order.

Table S1. Exome sequencing statistics.

Sample	Million read-pairs	% uniquely mapping	% unmapped	Mean coverage	Coverage at depth 10	% mouse reads
Normal 1	27,394	91,98	0,14	60,17	92,10	0,66
PDX-1	24,480	92,16	0,48	51,57	90,00	
Normal 2	23,458	92,44	0,16	50,50	88,77	0,87
PDX-2	20,858	92,64	0,29	48,52	90,60	
Normal 3	27,219	92,81	0,16	60,35	94,90	1,01
PDX-3	27,119	91,88	0,31	57,82	90,50	
Normal 4	26,767	92,70	0,18	59,22	94,10	0,95
PDX-4	26,972	91,59	0,47	55,56	90,70	
Normal 5	45,868	92,76	0,21	113,96	92,72	2,09
PDX-5	39,085	93,06	0,23	100,72	89,27	
Normal 6	45,485	93,83	0,17	116,54	93,30	1,52
PDX-6	47,442	94,59	0,18	124,88	92,46	
Normal 7	39,526	94,09	0,16	100,83	92,21	9,38
PDX-7	42,645	91,31	0,40	108,43	89,49	
Normal 8	33,175	90,53	2,28	74,92	93,59	3,24
PDX-8	28,646	90,18	2,45	65,70	89,79	
Normal 9	41,276	90,64	2,51	95,15	95,87	8,77
PDX-9	31,221	89,52	2,44	73,22	92,86	
Normal 10	42,286	89,97	1,89	100,80	95,95	3,62
PDX-10	33,008	89,84	1,85	76,51	92,46	
Normal 11	31,586	90,95	2,17	71,95	93,38	2,54
PDX-11	34,195	90,79	2,59	81,03	93,58	
Normal 12	37,195	90,87	2,21	86,42	95,18	2,47
PDX-12	38,892	89,99	2,01	92,34	94,98	
Normal 13	27,311	92,49	0,24	62,84	94,50	3,49
PDX-13	26,784	88,57	1,36	53,03	89,80	
Normal 14	27,263	92,63	0,30	61,21	94,30	2,14
PDX-14	27,764	89,76	0,91	56,50	90,50	

Table S2. Regions with high amplification levels and the corresponding embedded genes. Candidate (and known) oncogenes are displayed in the right column. Genes labeled in red were found amplified in more than one sample.

Sample	Region	Genes within the amplified region	Candidates with the highest expression levels
PDX-13	chr12:24970942-26492330	*AC026310.1, C12orf77, LRMP, CASC1, LYRM5, KRAS, AC087239.1, IFLTD1, RASSF8, BHLHE41, SSPN*	ND
PDX-14	chr6:17898411-19837986	*NHLRC1, TPMT, KDM1B, DEK, RNF144B*	ND
PDX-14	chr7:54610424-56082864	*VSTM2A, SEC61G, EGFR, LANCL2, VOPP1, SEPT14, MRPS17, ZNF713, GBAS*	ND
PDX-3	chr2:47202112-48822441	*C2orf61, CALM2, EPCAM, MSH2, KCNK12, AC138655.1, MSH6, FBXO11, FOXN2, PPP1R21, AC073082.1, STON1, STON1-GTF2A1L*	*CALM2, EPCAM, MSH2, KCNK12, MSH6, FBXO11, FOXN2, PPP1R21*
PDX-3	chr3:178525198-179754387	*KCNMB2, ZMAT3, PIK3CA, KCNMB3, ZNF639, MFN1, GNB4, ACTL6A, MRPL47, NDUFB5, USP13, PEX5L*	*ZMAT3, PIK3CA, ZNF639, MFN1, ACTL6A, MRPL47, NDUFB5*
PDX-5	chr19:34663668-34973546	*CTB-133G6.1, CTD-2207O23.3, ARHGEF18, PEX11G, C19orf45, ZNF358, MCOLN1, PNPLA6*	*ARHGEF18, PEX11G, C19orf45, ZNF358, MCOLN1, PNPLA6*
PDX-6	chr7:54610424-55272949	*VSTM2A, SEC61G, EGFR*	*SEC61G, EGFR*
PDX-6	chr11:82444771-82443491	*ANO1, FADD, PPFIA1, CTTN*	*ANO1, FADD, PPFIA1, CTTN*
PDX-7	chr12:25261482-25679033	*CASC1, LYRM5, KRAS*	*LYRM5, KRAS*

ND: no data.

Table S3. Homozygous deleted regions detected in PDX samples**.** Genes labeled in red correspond to well-established or candidate TSGs.

Sample	Region	Genes withing the deleted region and TSG candidates
PDX-10	chr10: 89544227-90122482	*ATAD1* (E1-5), *KLLN* (all) *PTEN* (all), *RNLS* (E5-7)
PDX-5	chr22:36960398-36983595	*CACNG2* (E3-4)
PDX-5	chr5: 55753716-59189449	*CTC-236F12.4* (all), *AC022431.2* (all), *MAP3K1* (all), *SETD9* (all), *MIER3* (all), *GPBP1* (all), *ACTBL2* (all), *PLK2* (all), *GAPT* (all), *CTD-2117L12.1* (all), *RAB3C* (all), *PDE4D* (all)
PDX-4	chr21:18931294-22910278	*CXADR* (E4-7), *BTG3* (all), *C21orf91* (all), *CHODL* (all), *TMPRSS15* (all), *NCAM2* (all)
PDX-6	chr1: 98039316-98060722	*DPYD* (E9-11)
PDX-10	chr2: 11591777-11932132	*E2F6* (E1-4), *GREB1* (all), *NTSR2* (all), *LPIN1* (E1-14)
PDX-9	chr19:4233126-4237085	*EBI3* (E3-5)
PDX-9	chr6:139206643-139206967	*ECT2L* (E16-17)
PDX-7	chr5:49695107-50687275	*EMB* (E1-9), *PARP8* (all), *ISL1* (E1-5)
PDX-6	ch3:71096093-71179834	*FOXP1* (E1-5)
PDX-4	chr6: 29910534-29913232	*HLA-A* (all)
PDX-9	chr9: 21440507-22029516	*IFNA1* (all), *IFNE* (all), *MTAP* (all), *C9orf53* (all), *CDKN2A* (all), *CDKN2B* (all), *RP11-145E5.5* (all), *DMRTA1* (all)
PDX-1	chr9:21440507-22029516	*IFNW1* (all), *IFNA21* , *IFNA4* (all), *IFNA7* (all), *IFNA10* (all), *IFNA16* (all), *IFNA17* (all), *IFNA14* (all), *IFNA5* (all), *KLHL9* (all), *IFNA13* (all), *IFNA2* (all), *IFNA1* (all), *IFNE* (all), *MTAP* (all), *C9orf53* (all), *CDKN2A* (all), *CDKN2B* (all), *RP11-145E5.5* (all), *DMRTA1* (all), *ELAVL2* (E3-7)
PDX-11	chr19:10597328-10610709	*KEAP1* (all)
PDX-4	chr21:15481314-15755440	*LIPI* (all), *RBM11* (all), *HSPA13* (all)
PDX-1	chr2: 141762904-142012210	*LRP1B* (E4-15)
PDX-9	chr2:141762904-141816623	*LRP1B* (E9-15)
PDX-7	chr2:135075100-135076310	*MGAT5* (E4-5)
PDX-10	chr8:15967594-16035497	*MSR1* (E5-10)
PDX-5	chr13:30072555-30088720	*MTUS2* (E12-14), *SLC7A1* (E13)
PDX-10	chr20:8352029-8770908	*PLCB1* (E3-31)
PDX-5	chr10: 89653782-89653866	*PTEN* (E2)
PDX-3	chr9:77112893-77761657	*RORB* (all), *TRPM6* (all), *C9orf40* (all), *C9orf41* (all), *NMRK1* (all), *OSTF1* (all)
PDX-6	chr6: 152541977-152542694	*SYNE1* (E118-119)
PDX-5	chr13:19748003-19755887	*TUBA3C* (all), *RP11-408E5.4* (all)
PDX-3	chr19:22255611-22256366	*ZNF257* (E2-3)

Table S4. List of the known TSGs and TSG candidates. Genes in red were found by manually inspecting the .bam sequencing files.

Gene	Protein	Sample	Position	
ARHGEF10L	Rho guanine nucleotide exchange factor 10-like protein	PDX-14	chr1	18014174
ARID1A	AT-rich interactive domain-containing protein 1A	PDX-9	chr1	27099479
ATAD2B	ATPase family AAA domain-containing protein 2B	PDX-8	chr2	24042732
ATG4D	Cysteine protease ATG4D	PDX-14	chr19	10662763
ATM	Serine-protein kinase ATM	PDX-7	chr11	108203488
ATM	Serine-protein kinase ATM	PDX-7	chr11	108099967
B2M	Beta-2-microglobulin	PDX-5	chr15	45007752
BAZ2B	Bromodomain adjacent to zinc finger domain protein 2B	PDX-9	chr2	160239091
CADM4	Cell adhesion molecule 4	PDX-9	chr19	44131912
CAMSAP1	Calmodulin-regulated spectrin-associated protein 1	PDX-10	chr9	138713929
CDKN2A	Cyclin-dependent kinase inhibitor 2A	PDX6	chr9	21971051
CDKN2A	Cyclin-dependent kinase inhibitor 2A	PDX8	chr9	21971108
CDKN2A	Cyclin-dependent kinase inhibitor 2A	PDX10	chr9	21971110
CDKN2A	Cyclin-dependent kinase inhibitor 2A	PDX14	chr9	
CHD5	Chromodomain-helicase-DNA-binding protein 5	PDX-6	chr1	6214762
CIZ1	Cip1-interacting zinc finger protein	PDX-12	chr9	130947844
CNOT10	CCR4-NOT transcription complex subunit 10	PDX-3	chr3	32811377
CNTNAP2	Contactin-associated protein-like 2	PDX-4	chr7	147183134
CROCC	Rootletin	PDX-10	chr1	17285153
CTSL2	Cathepsin L2	PDX-1	chr9	99795310
DDX41	Probable ATP-dependent RNA helicase DDX41	PDX-5	chr5	176943196
DENND2C	DENN domain-containing protein 2C	PDX-10	chr1	115137093
DNAJC10	DnaJ homolog subfamily C member 10	PDX-13	chr2	183597240
FAM83G/SLC5A10	Sodium/glucose cotransporter 5	PDX-10	chr17	18882980
FBXW7	F-box/WD repeat-containing protein 7	PDX-8	chr4	153247244
FBXW7	F-box/WD repeat-containing protein 7	PDX-11	chr4	153249393
GALNT2	Polypeptide N-acetylgalactosaminyltransferase 2	PDX-3	chr1	230410196
GM2A	Ganglioside GM2 activator	PDX-7	chr5	150646307
GRWD1	Glutamate-rich WD repeat-containing protein 1	PDX-10	chr19	48956209
INSL4	Early placenta insulin-like peptide	PDX-1	chr9	5233851
KEAP1	Kelch-like ECH-associated protein 1	PDX-9	chr19	10602328
KEAP1	Kelch-like ECH-associated protein 1	PDX-10	chr19	10610136
LTBP2	Latent-transforming growth factor beta-binding protein 2	PDX-14	chr14	74975424
LTBP2	Latent-transforming growth factor beta-binding protein 2	PDX-14	chr14	97873937
METTL13	Methyltransferase-like protein 13	PDX-3	chr1	171761374

Table S4. *Continuation*

Gene	Protein	Sample	Position	
MGA	MAX gene-associated protein	PDX-13	chr15	42042185
MLL2/KMT2D	Histone-lysine N-methyltransferase 2D	PDX-1	chr12	49434923
NCOR1	Nuclear receptor corepressor 1	PDX-4	chr17	15995208
NUP214	Nuclear pore complex protein Nup214	PDX-7	chr9	134000999
PBRM1	Protein polybromo-1	PDX-4	chr3	52661289
PHF23	PHD finger protein 23	PDX-1	chr17	7139711
PPM1G	Protein phosphatase 1G	PDX-9	chr2	27632256
PTEN	Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN	PDX-1	chr10	
RLF	Zinc finger protein Rlf	PDX-7	chr1	40702895
SBF1	Myotubularin-related protein 5	PDX-7	chr22	50894661
SCAF1	Splicing factor, arginine/serine-rich 19	PDX-5	chr19	50154694
SETD1B/KMT2G	Histone-lysine N-methyltransferase SETD1B	PDX-12	chr12	122248588
STK11	Serine/threonine-protein kinase STK11	PDX-9	chr19	1218423
STK11	Serine/threonine-protein kinase STK11	PDX-10	chr19	1222007
STK11	Serine/threonine-protein kinase STK11	PDX-11	chr19	1218478
STK11	Serine/threonine-protein kinase STK11	PDX-12	chr19	1220487
SUCLG2	Succinyl-CoA ligase [GDP-forming] subunit beta, mitochondrial	PDX-7	chr3	67451193
TARBP1	Probable methyltransferase TARBP1	PDX-13	chr1	234563440
TP53	Cellular tumor antigen p53	PDX-1	chr17	7577535
TP53	Cellular tumor antigen p53	PDX-2	chr17	Splice
TP53	Cellular tumor antigen p53	PDX-3	chr17	7577057
TP53	Cellular tumor antigen p53	PDX-4	chr17	7578290
TP53	Cellular tumor antigen p53	PDX-5	chr17	
TP53	Cellular tumor antigen p53	PDX-6	chr17	7578370
TP53	Cellular tumor antigen p53	PDX-7	chr17	7577141
TP53	Cellular tumor antigen p53	PDX-8	chr17	7578442
TP53	Cellular tumor antigen p53	PDX-9	chr17	7577551
TP53	Cellular tumor antigen p53	PDX-10	chr17	7578395
TP53	Cellular tumor antigen p53	PDX-11	chr17	7576857
TP53	Cellular tumor antigen p53	PDX-13	chr17	7578526
TP53	Cellular tumor antigen p53	PDX-14	chr17	7578229
ZIC1	Zinc finger protein ZIC 1	PDX-7	chr3	147128837
ZNF268	Zinc finger protein 268	PDX-7	chr12	133780433
ZNF845	Zinc finger protein 845	PDX-1	chr19	53854652

Table S5. List of the well-known and candidate oncogenes.

Gene	Protein	Sample	Position	
ACSL1	Long-chain-fatty-acid--CoA ligase 1	PDX-7	chr4	185681557
ALG10	Dol-P-Glc:Glc(2)Man(9)GlcNAc(2)-PP-Dol alpha-1,2-glucosyltransferase	PDX-7	chr12	34179753
ALG12	Dol-P-Man:Man(7)GlcNAc(2)-PP-Dol alpha-1,6-mannosyltransferase	PDX-10	chr22	50298024
ARNTL2	Aryl hydrocarbon receptor nuclear translocator-like protein 2	PDX-7	chr12	27543092
ATP13A3	Probable cation-transporting ATPase 13A3	PDX-1	chr3	194158134
ATP9A	Probable phospholipid-transporting ATPase IIA	PDX-7	chr20	50287726
BRSK1	Serine/threonine-protein kinase BRSK1	PDX-7	chr19	55800943
DDX10	Probable ATP-dependent RNA helicase DDX10	PDX-13	chr11	108546412
ECHDC1	Ethylmalonyl-CoA decarboxylase	PDX-10	chr6	127648182
EGFR	Epidermal Growth Factor Receptor	PDX-7	chr7	55229248
EPHB4	Ephrin type-B receptor 4	PDX-6	chr7	100405112
GBE1	1,4-alpha-glucan-branching enzyme	PDX-10	chr3	81548278
GLYR1	Putative oxidoreductase GLYR1	PDX-10	chr16	4862235
KAT7	Histone acetyltransferase KAT7	PDX-8	chr17	47875855
KMT5B	Histone-lysine N-methyltransferase KMT5B	PDX-9	chr11	67941364
KRAS	GTPase KRas	PDX-9	chr12	25398285
KRAS	GTPase KRas	PDX-10	chr12	25398285
LARP7	La-related protein 7	PDX-7	chr4	113565914
LUM	Lumican	PDX-13	chr12	91498030
MKRN3	Probable E3 ubiquitin-protein ligase makorin-3	PDX-13	chr15	23811531
MLXIP	MLX-interacting protein	PDX-2	chr12	122617989
NCOR2	Nuclear receptor corepressor 2	PDX-7	chr12	124886981
NEK1	Serine/threonine-protein kinase Nek1	PDX-7	chr4	170482682
NFE2L2	Nuclear factor erythroid 2-related factor 2	PDX-6	chr2	178096379
NFE2L2	Nuclear factor erythroid 2-related factor 2	PDX-6	chr2	178098810
NFE2L2	Nuclear factor erythroid 2-related factor 2	PDX-8	chr2	178098953
NFE2L2	Nuclear factor erythroid 2-related factor 2	PDX-3	chr2	178098945
NOS1AP	Carboxyl-terminal PDZ ligand of neuronal nitric oxide synthase protein	PDX-9	chr1	162326869

Table S5. *Continuation*

Gene	Protein	Sample	Position	
PDCD6IP	Programmed cell death 6-interacting protein	PDX-8	chr3	33866819
PIEZO1	Piezo-type mechanosensitive ion channel	PDX-10	chr16	88786310
PRKCI	Protein kinase C iota type	PDX-2	chr3	169981200
RAI1	Retinoic acid-induced protein 1	PDX-13	chr17	17698048
RBM19	Probable RNA-binding protein 19	PDX-9	chr12	114392996
RNF103	E3 ubiquitin-protein ligase RNF103	PDX-3	chr2	86831179-80
SCARA3	Scavenger receptor class A member 3	PDX-8	chr8	27516466
SDHA	Succinate dehydrogenase [ubiquinone] flavoprotein	PDX-13	chr5	233701
SERPINB6	Serpin B6	PDX-4	chr6	2948804
SPTY2D1	Protein SPT2 homolog	PDX-1	chr11	18636974
TBC1D4	TBC1 domain family member 4	PDX-14	chr13	75923344
TNRC6A	Trinucleotide repeat-containing gene 6A protein	PDX-14	chr16	24788506
TTYH2	Protein tweety homolog 2	PDX-7	chr17	72233511
VEZF1	Vascular endothelial zinc finger 1	PDX-14	chr17	56056656

Table S6. Chimeric transcripts detected through RNA-seq in PDX samples.

Sample	Gene 1	Position of gene 1		Gene 2	Position of gene 2	
PDX-10	*TFEC*	chr7	115777203	*CAPZA2*	chr7	116538888
PDX-10	*NOTCH2*	chr1	120572527	*AKNAD1*	chr1	109369922
PDX-11	*AC016700.5*	chr2	70329652	*MRPL36*	chr5	1799905
PDX-3	*NGDN*	chr14	23940185	*MYH7*	chr14	23902435
PDX-4	*HMOX1*	chr22	35779098	*MCM5*	chr22	35813848
PDX-4	*HSPA5*	chr9	128000419	*MAPKAP1*	chr9	128348005
PDX-5	*E4F1*	chr16	2279675	*LOC653786*	chr16	22584542
PDX-5	*E4F1*	chr16	2279675	*LOC653786*	chr16	22587932
PDX-5	*E4F1*	chr16	2282575	*LOC653786*	chr16	22587931
PDX-5	*E4F1*	chr16	2279675	*OTOA*	chr16	21768402
PDX-5	*E4F1*	chr16	2279675	*OTOA*	chr16	21771790
PDX-5	*E4F1*	chr16	2282575	*OTOA*	chr16	21771789
PDX-5	*ELF1*	chr13	41556117	*KIR3DX1*	chr19	55044262
PDX-5	*FUZ*	chr19	50316586	*MED25*	chr19	50322552
PDX-5	*FUZ*	chr19	50316556	*MED25*	chr19	50322923
PDX-5	*LOC653786*	chr16	22585713	*RAB26*	chr16	2203151
PDX-5	*OTOA*	chr16	21769573	*RAB26*	chr16	2203151
PDX-6	*AKR1CL1*	chr10	5202774	*AKR1C7P*	chr10	5324860
PDX-6	*AC016700.5*	chr2	70329652	*MRPL36*	chr5	1799905
PDX-6	*POLA1*	chrX	24906238	*PTPRS*	chr19	5258137
PDX-6	*PTENP1*	chr9	33676487	*LINC00865*	chr10	91594955
PDX-6	*RECK*	chr9	36083812	*YES1*	chr18	796919
PDX-6	*ZNF37BP*	chr10	43047024	*ZNF37A*	chr10	38384671
PDX-6	*ZNF681*	chr19	23940433	*RPSAP58*	chr19	23945934
PDX-7	*ACER3*	chr11	76670056	*TANC1*	chr2	159922482
PDX-7	*ANKRD13A*	chr12	110437832	*DDX52*	chr17	35984389
PDX-7	*ATL2*	chr2	38570409	*AC016995.3*	chr2	38723619
PDX-7	*BAP1*	chr3	52440267	*STAB1*	chr3	52538806
PDX-7	*CHCHD8*	chr11	73587801	*TESK2*	chr1	45920950
PDX-7	*DDX60*	chr4	169216833	*PALLD*	chr4	169633073
PDX-7	*DDX60*	chr4	169225295	*PALLD*	chr4	169633073
PDX-7	*DENR*	chr12	123253612	*KNTC1*	chr12	123086082
PDX-7	*DHPS*	chr19	12790272	*WDR83*	chr19	12783658

Table S6. *Continuation*

Sample	Gene 1	Position of gene 1		Gene 2	Position of gene 2	
PDX-7	*DUSP16*	chr12	12707469	*GPR19*	chr12	12815403
PDX-7	*DUSP16*	chr12	12715050	*GPR19*	chr12	12815403
PDX-7	*EIF2AK1*	chr7	6080522	*ZNF12*	chr7	6737441
PDX-7	*EIF2S2*	chr20	32684508	*RP4-785G19.5*	chr20	32837189
PDX-7	*FAM59A*	chr18	29890155	*LRP3*	chr19	33687635
PDX-7	*LTBP2*	chr14	75016565	*YLPM1*	chr14	75296031
PDX-7	*MICAL2*	chr11	12226003	*RP13-631K18.2*	chr11	12074769
PDX-7	*MTMR12*	chr5	32233878	*ZFR*	chr5	32379315
PDX-7	*PAIP1*	chr5	43533839	*NNT*	chr5	43623642
PDX-7	*PITPNM1*	chr11	67259539	*CHCHD8*	chr11	73586038
PDX-7	*PPP1CB*	chr2	28975040	*TRMT61B*	chr2	29087985
PDX-7	*PPP1CB*	chr2	28975041	*TRMT61B*	chr2	29075364
PDX-7	*PPP1CB*	chr2	28975041	*TRMT61B*	chr2	29084173
PDX-7	*RAB7A*	chr3	128445201	*SMG5*	chr1	156230416
PDX-7	*RAD21*	chr8	117878820	*LOC100499467*	chr17	70427360
PDX-7	*RAD21*	chr8	117878822	*LOC100499467*	chr17	70424436
PDX-7	*RGS12*	chr4	3344779	*HTT*	chr4	3122954
PDX-7	*RIMKLB*	chr12	8850892	*SETD4*	chr21	37422462
PDX-7	*ROD1*	chr9	114989618	*PBX1*	chr1	164619749
PDX-7	*ROD1*	chr9	114989690	*PBX1*	chr1	164648417
PDX-7	*ROD1*	chr9	114989690	*PBX1*	chr1	164761730
PDX-7	*ZDHHC3*	chr3	44981852	*CCR3*	chr3	46236890
PDX-8	*PLOD2*	chr3	145820541	*IGHV7-81*	chr14	107283095

RECURRENCE-FREE SURVIVAL

				Censored	
		Total	Events	N	%
HLA-I	Negative	13	5	8	61.5
	Moderate	25	11	14	56.0
	Strong	63	25	38	60.3
	Global	101	41	60	59.4

				Censored	
		Total	Events	N	%
β2m	Negative	29	13	16	55.2
	Moderate	42	21	21	50.0
	Strong	34	11	23	67.6
	Global	105	45	60	57.1

OVERALL SURVIVAL

				Censored	
		Total	Events	N	%
HLA-I	Negative	38	9	29	76.3
	Moderate	38	7	31	81.6
	Strong	72	15	57	79.2
	Global	148	31	117	79.1

				Censored	
		Total	Events	N	%
β2m	Negative	53	11	42	79.2
	Moderate	54	16	38	70.4
	Strong	45	7	38	84.4
	Global	152	34	118	77.6

Figure S1. Details about recurrence-free and overall survival of LC patients, with respect to the staining classification of β2m and HLA-I (negative, moderate and strong). The event was considered to be the death.

Table S7. List of down-regulated genes upon *B2M* restoration in gene-deficient cells (H2009 and H2135) presenting fold change <-1.15 and *P*<0.05 for both models simultaneously.

	H2009-wtB2M vs H2009-mock		H2135-wtB2M vs H2135-mock	
Gene	**Fold change**	**P value**	**Fold change**	**P value**
CSAD	-1,19	0,04	-1,18	0,05
MMP25	-1,38	0,05	-1,36	0,05
KCNC3	-1,31	0,00	-1,17	0,05
NHLH1	-1,25	0,05	-1,25	0,05
PAX6	-1,34	0,02	-1,30	0,04
CCDC33	-1,37	0,01	-1,29	0,04
BSX	-1,57	0,01	-1,45	0,04
ANKRD56	-1,54	0,00	-1,25	0,03
IGF2-AS	-1,32	0,00	-1,22	0,03
ND5	-1,33	0,02	-1,31	0,03
CA7	-1,18	0,02	-1,16	0,03
KLK5	-1,34	0,00	-1,20	0,03
SPRY4	-1,22	0,04	-1,24	0,02
RTDR1	-1,46	0,00	-1,35	0,02
ZNF677	-1,25	0,02	-1,25	0,02
CYTL1	-1,15	0,05	-1,18	0,02
FAM50B	-1,29	0,01	-1,24	0,02
FCGR1B	-1,42	0,01	-1,36	0,02
FAM27L	-1,34	0,02	-1,34	0,02
FGF8	-1,26	0,04	-1,32	0,01
SSC5D	-1,19	0,04	-1,24	0,01
CACNB2	-1,26	0,01	-1,24	0,01
NPPC	-1,58	0,00	-1,44	0,01
ATP2B2	-1,37	0,00	-1,31	0,01
TRIM3	-1,20	0,02	-1,23	0,01
SPAG11B	-1,40	0,01	-1,42	0,01
TCERG1L	-1,22	0,04	-1,30	0,01
SPG20OS	-1,15	0,04	-1,21	0,01
USP17L2	-1,16	0,03	-1,22	0,00
WDR69	-1,33	0,03	-1,46	0,00
PROC	-1,29	0,02	-1,41	0,00
IL29	-1,26	0,00	-1,26	0,00
KRT25	-1,29	0,02	-1,41	0,00
KRT33B	-1,17	0,02	-1,26	0,00
FLT4	-1,19	0,02	-1,30	0,00
APLNR	-1,29	0,05	-1,56	0,00
ATP2B3	-1,22	0,02	-1,39	0,00
ZNF236	-1,24	0,00	-1,30	0,00
RAI2	-1,33	0,02	-1,66	0,00
PRKCG	-1,26	0,02	-1,63	0,00
LRRC4	-1,17	0,05	-1,65	0,00

Table S8. List of up-regulated genes upon *B2M* restitution in H2009 and H2135 cell lines (fold change >1.15 and P<0.05 in both models, simultaneously).

	H2009-wtB2M vs H2009-mock		H2135-wtB2M vs H2135-mock	
Gene	**Fold change**	**P value**	**Fold change**	**P value**
GAS7	1,26	0,01	1,20	0,05
PCDHB4	1,36	0,01	1,25	0,05
FRAS1	1,26	0,04	1,25	0,05
AMN1	1,25	0,01	1,18	0,05
RFESD	1,28	0,01	1,19	0,05
ST3GAL6	1,21	0,05	1,21	0,05
IL18R1	1,29	0,00	1,19	0,05
TRIM55	1,29	0,00	1,17	0,05
LOH12CR2	1,26	0,03	1,23	0,04
LILRB1	1,34	0,00	1,18	0,04
STX16	1,20	0,02	1,17	0,04
MGAT1	1,19	0,03	1,18	0,04
TTC39B	1,29	0,00	1,20	0,04
CNTLN	1,23	0,03	1,21	0,04
SOX18	1,18	0,03	1,16	0,04
KLRF1	1,32	0,00	1,21	0,04
PSG6	1,18	0,02	1,16	0,04
PLG	1,22	0,03	1,20	0,04
FPGT-TNNI3K	1,31	0,00	1,20	0,04
C1QTNF3	1,48	0,00	1,29	0,04
GPR21	1,20	0,01	1,16	0,04
UTS2D	1,33	0,02	1,27	0,04
GDAP1L1	1,48	0,00	1,29	0,04
ZNF391	1,25	0,03	1,24	0,04
HAS2	1,27	0,01	1,21	0,04
GLIPR1L2	1,20	0,03	1,19	0,04
ZFP64	1,23	0,02	1,19	0,04
PCDHB1	1,23	0,01	1,18	0,04
TET1	1,33	0,01	1,23	0,04
MYPN	1,18	0,04	1,19	0,04
FAM75A2	1,19	0,03	1,18	0,04
ZCWPW2	1,23	0,04	1,23	0,04
ADAMTS3	1,22	0,03	1,21	0,04
SNORD42A	1,18	0,05	1,20	0,03
AOAH	1,17	0,02	1,16	0,03
ZNF790	1,19	0,04	1,20	0,03
CHL1	1,22	0,03	1,21	0,03
ARNT2	1,18	0,04	1,19	0,03
IL25	1,40	0,00	1,15	0,03

Table S8. *Continuation*

	H2009-wtB2M vs H2009-mock		H2135-wtB2M vs H2135-mock	
Gene	**Fold change**	**P value**	**Fold change**	**P value**
SLC25A31	1,16	0,04	1,16	0,03
OR51B5	1,24	0,03	1,24	0,03
FCER1A	1,21	0,02	1,19	0,03
TIAL1	1,44	0,02	1,39	0,03
GNGT2	1,22	0,02	1,21	0,03
SC5DL	1,24	0,01	1,20	0,03
ZNF507	1,20	0,01	1,17	0,03
FGR	1,30	0,00	1,21	0,03
FGL2	1,21	0,00	1,15	0,03
GNGT1	1,34	0,00	1,17	0,03
TRIML2	1,20	0,02	1,18	0,03
PION	1,23	0,05	1,27	0,02
TFAP4	1,22	0,02	1,22	0,02
WDPCP	1,20	0,02	1,19	0,02
GRM3	1,27	0,02	1,25	0,02
ENTPD5	1,20	0,02	1,20	0,02
OR52A1	1,21	0,04	1,23	0,02
CLEC4C	1,27	0,01	1,25	0,02
ZNF836	1,23	0,01	1,20	0,02
GOLGA3	1,26	0,04	1,31	0,02
TFPI	1,26	0,02	1,25	0,02
ADAT2	1,23	0,01	1,21	0,02
ZNF876P	1,32	0,03	1,35	0,02
PRDM2	1,23	0,03	1,26	0,02
LACE1	1,24	0,01	1,20	0,02
THAP2	1,44	0,00	1,28	0,02
DICER1-AS	1,23	0,01	1,22	0,02
FAM69A	1,17	0,04	1,21	0,02
MTTP	1,22	0,01	1,22	0,02
SEMA3E	1,32	0,00	1,23	0,01
ZNF461	1,17	0,02	1,18	0,01
RMND5A	1,22	0,03	1,26	0,01
KRT37	1,16	0,03	1,18	0,01
POU6F2	1,27	0,04	1,34	0,01
ATP2B4	1,25	0,02	1,27	0,01
MED12L	1,24	0,02	1,26	0,01
PKD1L1	1,17	0,03	1,20	0,01
TACR3	1,15	0,03	1,18	0,01
NRG4	1,24	0,02	1,28	0,01

Table S8. *Continuation*

	H2009-wtB2M vs H2009-mock		H2135-wtB2M vs H2135-mock	
Gene	**Fold change**	**P value**	**Fold change**	**P value**
FOXO4	1,24	0,01	1,22	0,01
INTS6	1,25	0,03	1,30	0,01
MLNR	1,28	0,01	1,28	0,01
ZFP36L1	1,19	0,03	1,24	0,01
MTHFD2L	1,28	0,01	1,29	0,01
NXT2	1,34	0,01	1,31	0,01
SP2	1,24	0,04	1,32	0,01
RNASE9	1,38	0,01	1,36	0,01
DDI2	1,27	0,02	1,31	0,01
SESTD1	1,27	0,01	1,26	0,01
CDC14A	1,21	0,04	1,30	0,01
AMBRA1	1,35	0,00	1,31	0,01
FIGN	1,33	0,01	1,35	0,01
CA5B	1,53	0,00	1,33	0,01
MSMP	1,28	0,04	1,40	0,01
SCAI	1,20	0,02	1,25	0,00
PMFBP1	1,26	0,01	1,27	0,00
TIFA	1,22	0,05	1,35	0,00
PPARA	1,22	0,03	1,32	0,00
MS4A5	1,24	0,04	1,36	0,00
CAMTA1	1,30	0,00	1,26	0,00
MYH7	1,22	0,03	1,32	0,00
AK7	1,35	0,01	1,38	0,00
ZC3H6	1,42	0,00	1,36	0,00
PDIK1L	1,25	0,01	1,29	0,00
GSTA5	1,24	0,02	1,35	0,00
FCRL5	1,22	0,05	1,37	0,00
ADAM21	1,29	0,01	1,41	0,00
ZNF678	1,27	0,01	1,34	0,00
SCN4B	1,23	0,01	1,30	0,00
GPR155	1,29	0,01	1,43	0,00
TMPRSS7	1,28	0,00	1,34	0,00
CUL3	1,20	0,05	1,40	0,00
CSMD2	1,18	0,02	1,31	0,00
SERPINA13	1,45	0,00	1,62	0,00
GGTA1P	1,23	0,03	1,45	0,00
MSL3	1,16	0,05	1,35	0,00
PLA2R1	1,17	0,04	1,53	0,00
B2M	1,39	0,00	3,96	0,00

Table S9. CTL infiltration measurements in tumor samples.

Sample	Histology	MHC-I	Mean area (μm^2)
1	AD	negative	166.2
2	SCC	negative	1811
3	AD	strong	1067.2
4	SCC	negative	136.8
5	SCC	negative	136.2
6	AD	strong	2114.8
7	SCC	strong	2803.6
8	AD	strong	10317.6
9	AD	negative	615
10	AD	negative	836.8
11	AD	strong	3982.2
12	AD	negative	159.2
13	SCC	negative	227.2
14	SCC	negative	146.2
15	SCC	negative	140
16	SCC	strong	1923,6
17	SCC	strong	2277
18	SCC	negative	562.6
19	AD	negative	753.6
20	AD	negative	336.2
21	AD	negative	753.6
22	AD	negative	336.2
23	SCC	negative	158.6
24	AD	negative	109.2
25	SCC	negative	219.2

AD: adenocarcinoma; SCC: squamous cell carcinoma.