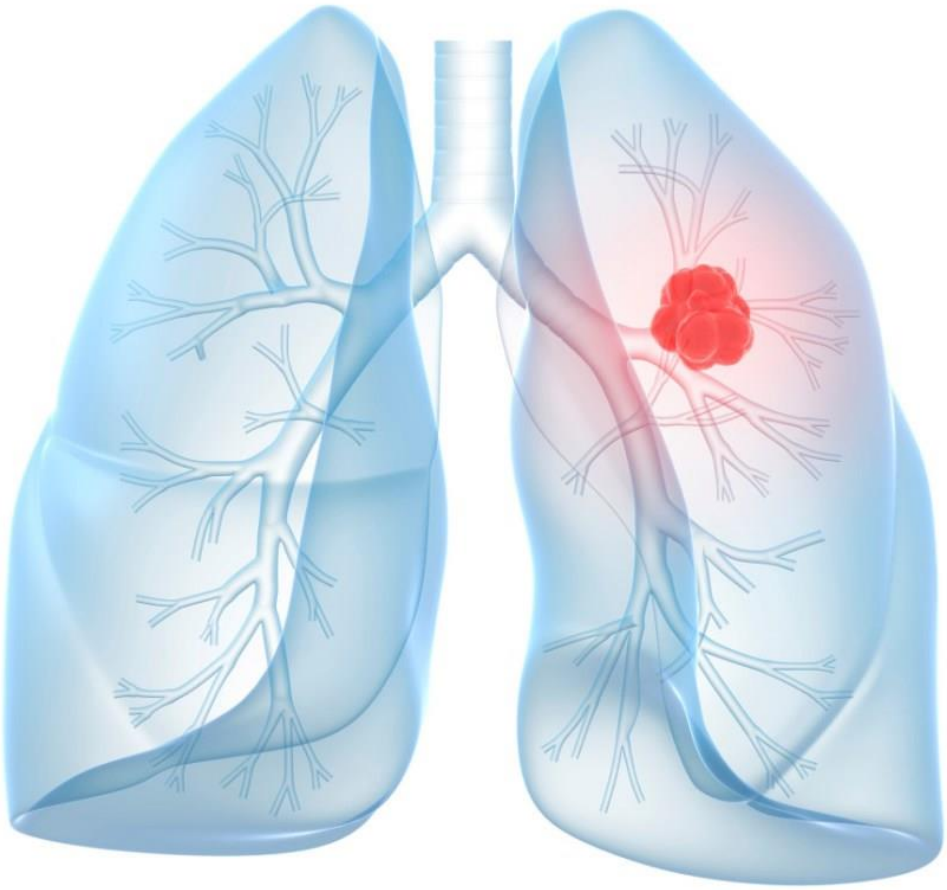
Mutations and aneuploidy in lung cancer

A thesis to find out whether there are significantly mutated genes and karyotypes found within lung cancer and to see if there is correlation between these two genomic alterations.



Frits Deiman

27-02-2017

Biomedical Sciences

Floris Foijer

Abstract

The deadliest type of cancer in this world is lung cancer. This is a cancer type characterized by growth of a malignant tumor on pulmonary tissue. This thesis focuses on finding the significantly mutated genes and karyotypes found in lung squamous cell carcinoma, lung adenocarcinoma and small cell lung cancer and to furthermore see whether there is correlation between these two genomic aberrations. In order to establish the significantly mutated genes found in lung cancer, the cancer genome atlas and its derived articles were used. Various significantly mutated genes were found that could be involved in lung cancer. There were also subtype specific genes found that could contribute to this specific type of lung cancer. The karyotypes found within the various types of lung cancer were determined using the Mitelman database of chromosome aberrations and gene fusions in cancer. With the help of this database the rates of missing and extra copies of chromosomes per subtype were established. By comparing the chromosomal locations of the significantly mutated genes and the karyotypes we were able to establish that there definitely is some degree of correlation between these two genomic aberrations. Studying these genomic aberrations can help us find therapeutic targets, which can be used to create a therapy for this deadly disease.

Contents

Introduction

Cancer is an uncontrolled growth of cells in the body usually caused by oncogenes, malfunctioning tumor suppressor genes and defective DNA repair pathways[1]. Cancer is a malignant growth of cells that is characterized by a few capabilities. They have the ability to evade apoptosis, are self-sufficient in growth signals, insensitive to anti-growth signals, capable of invading tissues and metastasis, have limitless replicative potential and sustained angiogenesis[2]. The deadliest type of cancer in this world is lung cancer[3]. This type of cancer is characterized by growth of a malignant tumor in pulmonary tissue[4].

Lung cancer can be divided into two major groups: Small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). NSCLC is further subdivided into adenocarcinoma, squamous-cell carcinoma and large-cell carcinoma[5]. SCLC is described as a malignant epithelial tumor with small cells. These small cells are characterized by little cytoplasm, unclear membranes and an aberrant nuclei[6]. Unlike SCLC, NSCLC is a type of lung cancer that consists of cells that aren't smaller than normal cells in lung tissue. The subtype of NSCLC is distinguished based on the cell type that the tumor develops at. Squamous cell carcinoma is one of the three main NSCLC that commonly develops in the inner parts of the lung at the squamous cells. Adenocarcinoma is the most common lung cancer type that develops on the outer parts of the lungs, the adenomatous cells. Large-cell carcinoma can arise in any cell type that is part of the lung. This subtype is difficult to treat as it tends to grow and metastasize at a fast pace[7].

Lung cancer has a lot of shared genomic imbalances or molecular alterations. These alterations can be mutations or numeric (aneuploidy) or structural changes (inversions, deletions and translocations). However, every type of lung cancer also has his own unique alterations. There have been large-scale genomic analyses of these varying lung cancer types to determine the shared and distinct alterations. These genomic analyses of cancerous tumours can help us identify new potential therapeutic targets which can be used to treat patients suffering from this deadly condition[8 9].

Studies have been recorded in The Cancer Genome Atlas (TCGA), which is a database that enables researchers to analyze cancer genomic data sets[10,11]. These recorded studies have been executed to determine the genomic alterations in cancer using various methods. The methods discussed later in this thesis that were used in other studies to analyze the mutations in lung cancer are whole genome sequencing (used for SCLC) and whole exome sequencing (used for adenocarcinomas and squamous cell carcinomas). The difference between these two sequencing techniques is that whole exome sequencing is aimed at sequencing the protein coding regions, which reduces the cost and increases the feasibility of having a higher number of samples. Whole genome sequencing on the other hand is a more expensive sequencing method which gives more information about the complete genome as it sequences both coding and non-coding regions[12]. In this thesis the primary focus will be on SCLC, adenocarcinomas and squamous cell carcinomas as there hasn't been a large scale study about the comprehensive genomic character of large cell carcinomas due to the low prevalence of 10%[7].

Another relevant feature of lung cancer that can be used to study lung cancer is the aneuploidy found within these cancers. Aneuploidy is an abnormal numeric change in the amount of chromosomes of a cell and can be found in several diseases but is also an indicator of cancer. In this thesis we will try to asses if there is correlation between the karyotypes of the varying lung cancers[13].

Firstly, how can one measure aneuploidy? Between all the techniques that are used to assess aneuploidy it is important to know whether the experimental cells are dividing (proliferating) or non-dividing (post mitotic cells) as some techniques require proliferating cells, whilst others don't. Techniques where proliferating cells are obligatory are for example; Fluorescence In Situ Hybridization (FISH). This is a technique that can establish chromosomal abnormalities. During FISH proliferating cells are fixed during the metaphase of the mitosis, followed by hybridization using a specific labeled probe on a target of your choice. After hybridization the unbound probes will be washed away and the hybridized cells signal will be measured using fluorescence[14 15]. There are also other FISH-adapted protocols like Multiplex FISH (M-FISH), combined binary ratio labeling (COBRA-FISH), multiplex multicolor banding (mMCB) and spectral karyotyping (SKY), which are more complex and expensive FISH methods that require more or different probes, but are worth the results due to them coloring whole chromosomes or specific chromosome arms or -regions to visualize chromosomal alterations[15 16]. All these FISH-adapted protocols require dividing cells, but there is also a FISH-adapted protocol that doesn't require dividing cells to assess aneuploidy. This protocol is called interphase FISH (iFISH). This method doesn't require proliferating cells as it enables probes to hybridize to uncondensed chromosomes in the interphase[15].

Other methods that do not require proliferating cells to assess the aneuploidy are arrays like comparative genome hybridization (CGH) or single nucleotide polymorphism (SNP) arrays. CGH arrays are based on the comparison of target and control DNA which are labeled in different colors. This is followed by hybridization of the DNA on a specific gene panel. Then with the use of fluorescence one can determine if there is a difference in gene dosage between the target and control DNA[15 17]. The same principle applies to SNP arrays. The difference between SNP arrays and CGH arrays is that CGH uses specific oligonucleotides or a gene panel and a reference genome. SNP arrays on the other hand make use SNPs, which are frequently found types of variation in the human genome[15 18].

Lastly, to quickly establish the ploidy of non-dividing cells flow cytometry can be used. This is a technique that focuses on the counting and measuring of the physical and chemical characteristics of cells through a fluid stream. In aneuploidy research, the DNA of a cell can be fluorescently labeled, when going through the flowcytometer the cell will emit fluorescence which can be detected. In this way the amount of DNA can be determined in comparison to reference cells and ploidy can be determined[15,19,20].

The database used in this thesis to determine the aneuploidy rates of the varying lung cancers is the Mitelman database of chromosome aberrations and gene fusions in cancer. This is a database created to relate chromosomal aberrations to the characteristics of tumors. Within this database 164 cases of lung adenocarcinoma, 147 cases of squamous cell carcinoma and 49 cases of small cell lung cancer are documented[21].

In the results section of this thesis the methods discussed that were used in other studies to determine mutations were whole genome and whole exome sequencing. What is the value of sequencing a whole genome and not taking intratumor heterogeneity into account in comparison to a technique like single cell sequencing where the intratumor heterogeneity can be determined? This Intratumor heterogeneity is a phenomenon within tumors where subclone of a tumor has a different genomic make-up than the other. It can be caused by the microenvironment[22] or genomic instabilities such as endogenous point-mutations and errors in chromosome segregation[23].

Intratumor heterogeneity can have a lot of influence on cancer treatment as it determines if a therapy can work or not. This is because the intratumor heterogeneity can cause that a specific cure works on one cell of a tumor and not work on the next, due to resistance caused by the heterogeneity. This is called treatment resistance and it is a recurring problem within cancer treatment. When a potential therapeutic target within a tumor is found and there is a resistant subclone, this subclone will survive and proliferate, leading to the relapse of this tumor[24,25].

Single cell sequencing can help us gain insight in the heterogeneity of a tumor, and possibly in the future predict whether the tumor cells will respond to therapy. This can't be done with whole genomic analysis strategies executed on tumor samples as a whole as there will be looked at an average. Instead one would have to look at every cell independently to make sure that the therapeutic target is relevant to all cells within the tumor. Techniques like single cell sequencing are methods which can be used to karyotype all chromosomes per cell. This karyotyping can give us a prediction to the tumors response to therapy and can help us in finding a cure[25–27].

The goal of this thesis is to give an overview of the recurrent genomic alterations; aneuploidy and mutations found in multiple lung cancer types. These types are squamous cell carcinoma, adenocarcinoma and small cell lung cancer and to furthermore determine if there is correlation between the significantly mutated genes in lung cancer and their karyotypes.

Results

Significantly mutated genes found in genomic alteration analysis SCLC

A whole genome sequencing study of SCLC published in Nature analyzed 110 SCLC samples. The analysis led to the establishment of figure 1. This figure shows that there are 16 significant mutated genes for SCLC (highlighted in bold). They also found that nearly all tumors have a mutation in the TP53 and RB1. The mutations found in the significant mutated genes can differ from; missense, nonsense, splice site, in-frame inversion/deletion, rearrangement or a frameshift mutation. The other significantly mutated genes found in SCLC in a lesser ratio are KIAA1211, COL22A1, RGS8, FPR1, EP300, CREBBP, ASPM, ALMS1, PDE4DIP, XRN, PTGFRN, TP73, FMN2 and NOTCH1[8].

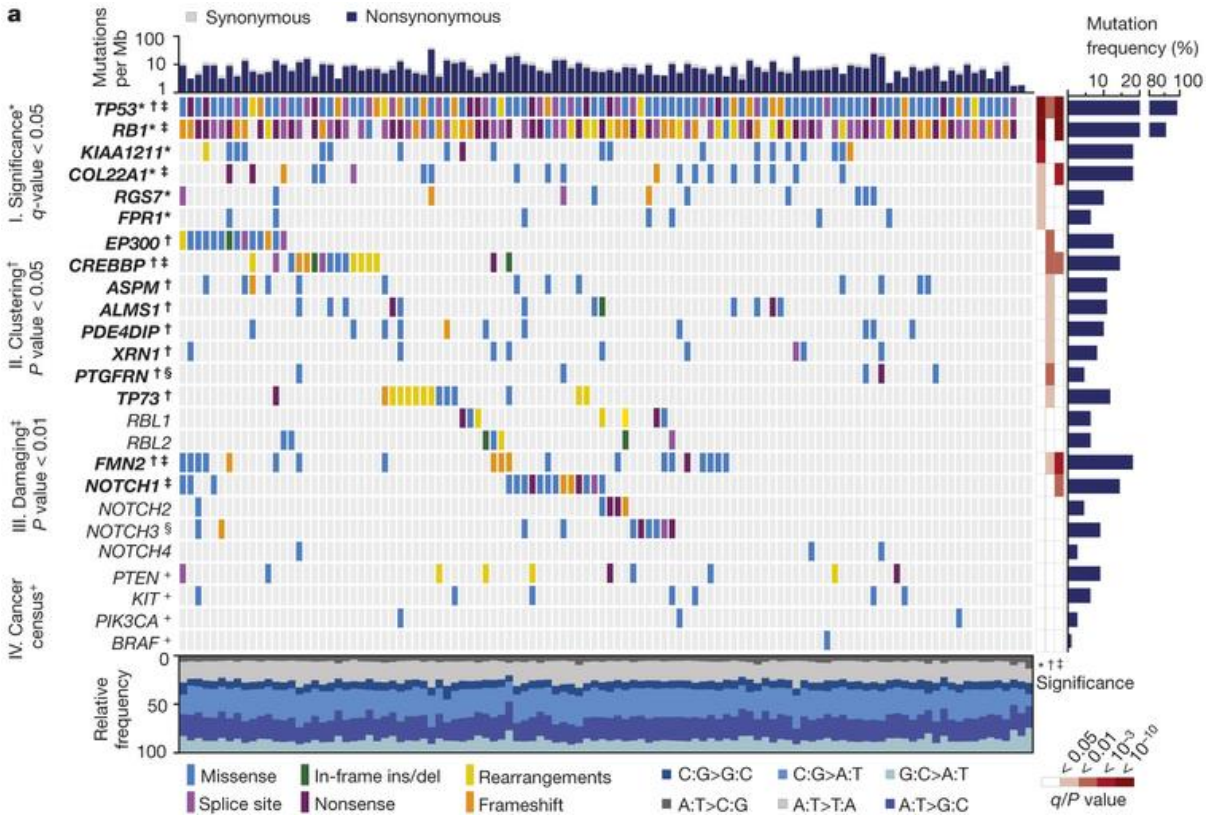


Figure 1 Genomic alterations found in SCLC[8]

Significantly mutated genes found in genomic alteration analysis lung squamous cell carcinoma

A whole exome sequencing study of squamous cell carcinoma published in Nature analyzed 178 lung squamous cell carcinoma samples. The analysis led to the establishment of figure 2. This figure shows that there are 10 significant mutated genes for squamous cell carcinoma. They also found that nearly all tumors have a mutation in the TP53 (81%). Other significantly mutated genes were CDKN2A, PTEN, PIK3CA, KEAP1, MLL2, HLA-A, NFE2L2, NOTCH1 and RB1. These mutations can differ from; silent (synonymous), missense, nonsense, splice site, in-frame inversions/deletions, frameshift mutations[28].

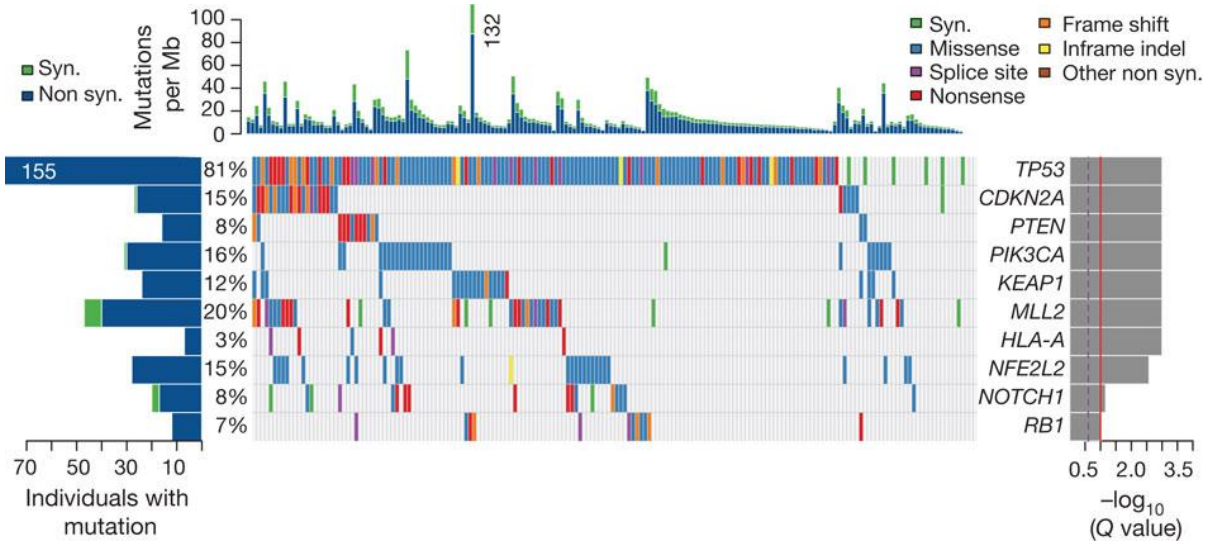


Figure 2 Genomic alterations found in squamous cell carcinoma[28]

Significantly mutated genes found in genomic alteration analysis lung adenocarcinoma

A whole exome sequencing study of lung adenocarcinoma published in Nature analyzed 230 lung adenocarcinoma samples. The analysis led to the establishment of figure 3. Figure 3 shows that there are 18 significantly mutated genes for this type of lung cancer. From the candidate genes the most common mutation found was TP53 (46%), secondly KRAS (33%). The other significantly mutated genes found in adenocarcinoma in a lesser ratio were KEAP1, STK11, EGFR, NF1, BRAF, SETD2, RBM10, MGA, MET, ARID1A, PIK3CA, SMARCA4, RB1, CDKN2A, U2AF1, RIT1. These mutations can differ from; missense, nonsense, splice site, in-frame inversions/deletions and frameshift mutations[29].

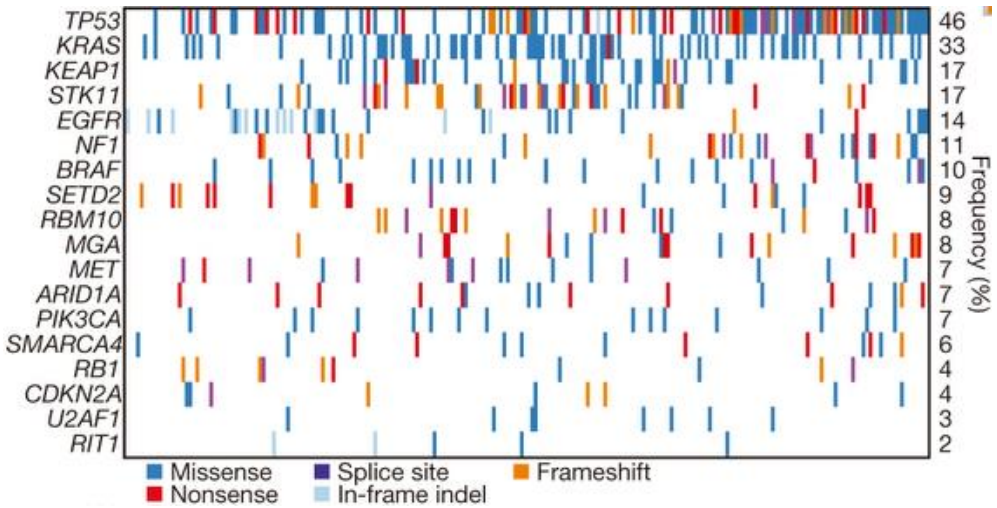


Figure 3 Genomic alterations found in lung adenocarcinoma[29]

Recurrent molecular alterations found in SCLC, squamous-cell carcinoma and adenocarcinoma

The figures 1 to 3 presented on the previous pages are studies which have been executed by various laboratories. The cancer genome atlas is a data system that combines all data from 3 small cell lung cancer (190 combined samples), 5 lung adenocarcinoma (1132 combined samples) and 2 lung squamous cell carcinoma studies (682 combined samples) and processes them. All these studies combined showed recurrent molecular alterations in various genes, these are displayed in table 1. In this table the mutations and commonly altered pathways that have been recurrently found within each examined lung cancer are documented[30]. We can see that the only mutual mutation found within each lung cancer examined is a mutation in TP53. Other mutated genes that are involved in the cell-cycle are CDKN2A and RBI1. A lot of genes of the RTK/PI3K-MTOR signaling have also been found to be recurrently mutated of which PIK3CA and PTEN have been found in more than one type of lung cancer. Other mutations have also been found in oxidative stress response (of which the KEAP1 gene in both NSCLC examined), MYC pathway, epigenetic deregulation, aberrant splicing, squamous differentiation and neuroendocrine differentiation genes.

Table 1 The cancer genome atlas data; recurrent molecular alterations found within lung cancer (adapted from [30])

Type of alterations	*Adenocarcinoma*	*Squamous-Cell Carcinoma*	*Small cell lung cancer*
Cell-cycle mutations	TP53 (46%), CDKN2A (4%)	TP53 (91%), CDKN2A (17%), RB1(7%)	TP53 (92%), RB1 (75%)
	RKT/PI3K-MTOR signaling:	RKT/PI3K-MTOR signaling:	RKT/PI3K-MTOR signaling:
	KRAS (33%), EFGR (14%), BRAF (10%), STK11 (17%), MET (8%), NF1 (11%), PIK3CA (7%), RIT1 (2%)	PIK3CA (16%), PTEN (8%), HRAS (3%)	PTEN (5%)
Other mutations	Oxidative stress response: KEAP1 (17%)	Oxidative stress response: CUL3 (6%), KEAP1 (12%), NFE2L2 (15%)	Epigenetic deregulation: EP300 (11%), CREBBP (10%)
	MYC pathway: MGA (8%)	Squamous differentiation: NOTCH1 (8%), ASCL4 (3%), NOTCH 2 (5%)	Neuroendocrine differentiation: NOTCH1 (15%)
	Aberrant splicing: U2AF1 (3%), RBM10 (8%)		
Commonly altered pathways	MAPK and PI3K signaling, oxidative stress response, cell-cycle progression, RNA splicing and processing, nucleosome remodeling	Squamous-cell differentiation, oxidative stress response, MAPK and PI3K signaling	Cell-cycle regulation, PI3K signaling, regulation of nucleosome transcriptional and remodeling, NOTCH signaling and neuroendocrine differentiation

Many varying numeric aberrations found within karyotypes of lung cancer cells

With use of the Mitelmans database the rates of numeric aberrations within the varying types of lung cancer were determined. Within the Mitelmans database 49 SCLC samples, 147 lung squamous cell carcinomas (SC) and 164 lung adenocarcinomas (AC) are documented. The rates of missing and extra copies were determined and organized into table 2. This table visualizes the extra and missing copies rates found within lung cancer per chromosome. A bright red shade means that within this cancer the aberration rate is low, while a bright green shade means that the chromosomal aberration rate is high. The highest chromosomal aberration rate was found in SC where 48.3 % of the samples obtained from male test subjects showed a missing copy of the Y chromosome. The lowest chromosomal aberrations were found in SCLC where there were no missing copies found in chromosome 1, 3. On the other hand there weren't any extra copies documented in chromosomes 5, 11, 13, 22 and Y from SCLC. The table shows that in SC and AC there is a high prevalence of missing copies in every chromosome except for chromosome number 7, which has a high rate of extra copies. In SCLC the numeric aberrations with the highest prevalence are a missing copy of chromosome 4 and 15[21].

Table 2 Numeric aberrations rates found in SCLC, SC and AC using the Mitelmans database

Chromosome	SCLC extra copy (%)	SCLC missing copies (%)	SC extra copies (%)	SC missing copies (%)	AC extra copies (%)	AC missing copies (%)
1	6.1	0.0	6.8	12.9	6.7	9.8
2	8.2	12.2	5.4	8.2	7.9	6.1
3	10.2	0.0	5.4	12.9	7.9	12.8
4	4.1	20.4	2.7	22.4	3.0	15.2
5	0.0	10.2	4.8	17.7	10.4	13.4
6	8.2	6.1	4.1	17.0	5.5	15.9
7	6.1	6.1	20.4	6.8	25.6	3.7
8	8.2	8.2	4.1	19.0	6.7	20.1
9	4.1	10.2	7.5	22.4	4.3	29.9
10	4.1	16.3	3.4	23.1	4.3	19.5
11	0.0	8.2	4.1	16.3	7.9	9.1
12	8.2	10.2	7.5	10.9	9.8	12.8
13	0.0	12.2	4.8	24.5	4.9	31.1
14	4.1	6.1	2.7	22.4	4.9	25.6
15	4.1	20.4	4.1	28.6	4.3	21.3
16	6.1	10.2	4.1	21.1	6.7	17.1
17	4.1	16.3	6.1	18.4	8.5	18.3
18	8.2	12.2	2.0	24.5	4.3	27.4
19	12.2	6.1	5.4	15.6	3.7	25.6
20	12.2	6.1	11.6	17.0	15.9	14.0
21	4.1	12.2	2.0	25.2	7.3	27.4
22	0.0	18.4	4.8	21.8	3.0	32.9
X	8.2	12.2	4.8	17.0	5.5	27.4
Y	0.0	14.3	4.1	48.3	3.0	32.9

Correlation between aneuploidy and location of tumor suppressor genes found.

The expectation is that oncogenes will be present on chromosomes that have extra copies and tumor suppressor genes will be present on chromosomes that have missing copies. To assess whether this is true, table 3 was created using table 1 and 2, in which the genes from table 1 with a prevalence higher than 5% have been picked. These genes were linked to their respective cancer type and chromosomal location. Furthermore, the genes were also categorized as either an oncogene or a tumor suppressor gene. Thereafter the aneuploidy rate of the chromosome was determined using table 2. Table 3 visualizes whether the chromosomal location of genes and the aneuploidy on this respective chromosome is expected or not. Green means it is in accordance to the expectation. Red means it is not and yellow means that the rates of extra or missing copies were similar. 90% of the significantly mutated tumor suppressor genes were present on chromosomes that had a high rate missing copies. 10% were present on chromosomes with a balanced aneuploidy rate. On the other hand, 43% of the significantly mutated oncogenes were present on chromosomes that have a high rate of extra copies. 57% were present on chromosomes with a high rate of both missing or on chromosomes with a balanced rate.

Table 3 Significantly mutated genes found in lung cancer on their respective chromosomes and their aneuploidy rate

Cancer type	*Gene*	*Gene type*	*Chromosome*	*Aneuploidy rate*
AC	KRAS	Oncogene	12	Balance
AC	EFGR	Oncogene	7	Extra copies
AC	BRAF	Oncogene	7	Extra copies
AC	MET	Oncogene	7	Extra copies
NSCLC	PIK3CA	Oncogene	3	Missing copies
SC	NFE2L2	Oncogene	2	Balance
SCLC	CREBBP	Oncogene	16	Missing copies
ALL	TP53	Tumor Suppressor	17	Missing copies
NSCLC	CDKN2A	Tumor Suppressor	9	Missing copies
SC, SCLC	RB1	Tumor Suppressor	13	Missing copies
AC	STK11	Tumor Suppressor	19	Missing copies
AC	NF1	Tumor Suppressor	17	Missing copies
SC, SCLC	PTEN	Tumor Suppressor	10	Missing copies
NSCLC	KEAP1	Tumor Suppressor	19	Missing copies
SC	CUL3	Tumor Suppressor	2	Balance
SCLC	EP300	Tumor Suppressor	22	Missing copies
SC, SCLC	NOTCH1	Tumor Suppressor	9	Missing copies
AC	MGA	Tumor Suppressor	15	Missing copies

Conclusion and discussion

The goal of this thesis was to give an overview of the recurrent genomic alterations aneuploidy and mutations found in the lung cancer types: Squamous cell carcinoma, adenocarcinoma and small cell lung cancers and to furthermore determine if there is correlation between the significantly mutated genes in lung cancer and their karyotypes.

The most frequently mutated gene in all lung cancers is the TP53 gene. The others cell-cycle mutations that aren't found in all lung cancers are in the RB1 (for SC and SCLC) and CDKN2A (for NSCLC) gene. These three genes are tumor suppressor genes. These are genes that inhibit cell proliferation and tumor development. Tumor suppressor genes are genes that are often lost or inactivated in cancers, because when these are lost or inactivated the cell loses control of its cell growth, which can lead to uncontrollable growth of cells, a characteristic of cancer[31].

Pathways found to be mutated in more than one subtype of lung cancer are the MAPK and PI3K signaling pathways. The MAPK pathway is a pathway that is frequently dysregulated within cancers as it controls cell growth, proliferation, differentiation, migration and apoptosis[32]. This is because the mutations in this pathway can contribute to cancer cell migration and development of resistance to apoptosis, which are hallmarks of tumorigenesis[33]. The most frequently mutated gene within this pathway in cancer cells is the KRAS gene. This is an oncogene gene that encodes the K-Ras protein, which is a protein that regulates cell division[34]. The other pathway is the PI3K signaling pathway. This is a pathway, that is like MAPK often dysregulated within cancer, The PI3K pathway revolves around a second messenger: PIP3, which is a compound that is able to recruit AKT to activate growth, proliferation and apoptosis[35]. The most frequently mutated gene within this pathway is the PIK3CA gene, which causes constitutive kinase activity. Which is a nonstop functional state that may contribute to initiation or progression of cancer[35,36]. Another interesting significantly mutated gene found within this pathway is PTEN, which is a tumor suppressor gene like the previously mentioned TP53, RB1 and CDKN2A gene. The other mutated genes within this pathway generally cause growth, proliferation or anti-apoptosis.

The mutated oxidative stress response genes in both NSCLC subtypes was the KEAP1 gene. KEAP1 is a protein that functions as a tumor suppressor[37] and interacts with Nrf2. This is a transcription factor that regulates the expression of antioxidant proteins to protect the cell against oxidative damage. Due to the mutation KEAP1 will get inactivated or have a reduced expression leading to upregulation of Nrf2[38,39]. This has been shown to lead to chemoresistance. Overexpression of Nrf2 has been found in many cancers which gives cancer cells an advantage in survival and growth[40]. The final significantly mutated gene found in squamous cell carcinomas involved in oxidative stress response is the CUL3 gene. This gene codes for a protein that plays an important role in the ubiquitination of specific proteins. One of these proteins is the previously mentioned Nrf2[41]. Interestingly all significantly mutated oxidative stress response genes found seem to be involved with Nrf2.

Other pathways containing mutations found within each exclusive lung cancer is the MYC pathway in lung adenocarcinoma. This is a pathway that is involved in cell proliferation, differentiation and apoptosis[42]. Therefore, it isn't strange for a mutation in this pathway to escalate into uncontrolled cell proliferation. Others mutations found in adenocarcinoma were mutations in the genes U2AF1 and RBM10. These are genes involved in aberrant splicing. Aberrant splicing can contribute to cancer due to the formation of aberrant mRNA's which encode aberrant proteins that have unique properties that can influence growth and differentiation of a cancer cell[43].

Mutations exclusively found in small cell lung cancer were epigenetic deregulation genes: EP300 and CREBBP. Both EP300 and CREBBP code for a histone acetyltransferase protein. These are proteins that regulate transcription and are important in cell proliferation and differentiation[40,44]. EP300 has been suggested to behave as a tumor suppressor gene[45]. The CREBBP gene on the other hand has been shown promote tumor growth in lung cancer cells by interacting and acetylating CPSF4 which promotes hTERT. This is the gene that encodes telomerase reverse transcriptase (TERT). TERT is an enzyme that maintains the ends of telomeres, which is a trait often found in oncogenesis[44,46].

The final group of mutations found within lung cancers are within differentiation genes, NOTCH1 in SCLC and SC and NOTCH2 and ASCL4 in exclusively SC. The NOTCH genes belong to a pathway that is involved in the normal development of many tissues and has been found to potentially contribute to cancer development as it is able to influence cellular transformation. Researchers have found that a potential role of notch signaling in SCLC is the downregulation of the Notch pathway[47]. The Notch pathway appears to be a double edged sword within cancer, as also a gain of function mutation and ligand-mediated activation of the Notch pathway can contribute to the formation of cancer[48]. ASCL4 unlike NOTCH isn't involved in the notch signaling pathway. ASCL4 however is a transcription factor essential for tissue development and differentiation. The exact role of ASCL4 is unknown, but as it is a gene involved cell proliferation it isn't strange that it is found in a mutated state in cancer cells[28,49].

Generally, it is found that oncogenes and tumor suppressor genes are of the essence in aneuploidy in karyotypes of cancers. For example; trisomy 21 appears to have a tumor suppressant effect in solid tumors due to overexpression of 2 genes found on this chromosome that inhibit tumors angiogenesis, which is a capability of a malignant tumor[50]. And so, for a cancer cell it would be more beneficial to not have trisomy 21. This is in accordance to the karyotypes that were found in the Mitelman database as table 2 tells us that there is a high rate of missing copies and a low rate of extra copies of chromosome 21 in lung cancer. The same goes for chromosome Y in NSCLC, which has an incredible rate of missing copies. The highest found within lung cancer, this is also thought because this deletion of the chromosome gives loss to a tumor suppressor gene, which can play a role in tumor progression[51,52].

On the other side of this spectrum there is for example chromosome 7, which has a high rate of extra copies in NSCLC. This is due to the oncogenes present on this chromosome. These genes are EGFR, MET, which have been determined as a significantly mutated gene in NSCLC in figure 3. Other suspected oncogenes on chromosome 7 are FTSJ2, NUDT1, TAF6 and POLR2J[53,54]. Due to all these oncogenes present on chromosome 7, it would be very beneficial for a cancer cell to have an extra copy of this chromosome. This can also be found in the karyotypes uploaded to the Mitelman database as a high rate of extra copies chromosome 7 were determined within NSCLC.

Generally, there were more missing then extra copies found within lung cancer, which would imply that losing various chromosomes and thus losing tumor suppressor genes present on these chromosomes is more beneficial to the lung cancer then gaining a chromosome, which can amplify the amount of oncogenes present in cancer cells. Carcinogenesis depends on both activating of proto-oncogenes and deactivation of tumor suppressor genes. It also depends on the composition of the chromosome. One chromosome can be more enriched in oncogenes and therefore it would be beneficial for a cancer cell to have this chromosome amplified. Another chromosome can be more enriched in tumor suppressor genes and therefore be suppressed in cancer cells[55].

In table 3 we were able to establish that there is a high rate of correlation between significantly mutated tumor suppressor genes found in lung cancer and loss of chromosomes on which these are localized. There is a bias within aneuploidy. Chromosome loss in cancer is more common than chromosome gain[55]. Another general pattern of aneuploidy in cancer is that chromosomes 7 and 20 are often gained and chromosomes like 13 and 22 are often lost[55]. This is also found back in table 2. We were able to establish that chromosome 7 contains significantly mutated oncogenes found in NSCLC. Finding other extra copies of chromosomes can imply that genes encoded on this chromosome can contribute to cancer when overexpressed[56], just like severely deleted chromosomes can have tumor suppressor genes encoded. For example, the Y chromosome in squamous cell carcinoma, which has a missing copy rate of 48.3%. However, as one can see in table 3, there hasn't been a significantly mutated gene found on this chromosome within the research performed. This can be caused by the fact that a cancer cell doesn't have much use for the Y chromosome or due to a hidden tumor suppressor gene located on this chromosome.

The significantly mutated oncogenes found in lung cancer have a lower correlation between gene location and aneuploidy then tumor suppressor genes. It appears that chromosome loss in cancer is more common than chromosome gain. Firstly we have to ask ourselves: What is gain and what is loss of a chromosome? It has been reported that some cancer cells tend to show a trisomic character[57]. With the molecular techniques used to determine the ploidy of a cell there is no reference used. Researchers are looking at the relative expression of a chromosome within a sample. When a cell is more tri- then disomic, a disomic pair of chromosomes can be interpreted as a loss of chromosomes, while in reality every other chromosome has a gained pair, but due to this relative expression the disomic pair will be marked as a loss of a chromosome. Based on the information gathered one would expect that chromosomes that have the significantly mutated oncogenes encoded will have extra copies as this would be beneficial to the cancer cell, but this was only found with chromosome 7 and not in chromosomes 2, 3, 12 & 16 (at which the other significantly mutated oncogenes are located). Chromosome 2 and 3 rarely show copy number changes within cancers[55], therefore it's not strange that the correlation between the mutations and aneuploidy doesn't hold with these chromosomes.

In total there definitely is correlation between the recurrent genomic alterations; aneuploidy and mutations found in lung cancer and its subtypes. It was found that chromosomes that contain significantly mutated tumor suppressor genes in lung cancer have a very high rate of missing copies. The same applies to chromosomes containing significantly mutated oncogenes as these tend to have extra copies, which can contribute to the overexpression of these genes and thus to carcinogenesis. The pathways these genes are involved in can be used as therapeutic targets when the tumor is homogenous. In heterogeneous tumors one would have to find a therapeutic target present in all tumor cells, in order to find this target it is useful to know what kind of karyotypes and mutations are common in every varying cancer in order to make a rapid diagnosis and establish a proper specific therapy for each individual patient[25].

References

1. (US), N. C. for B. I. Cancers. (1998).
2. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100,** 57–70 (2000).
3. Cancer Research UK. Worldwide cancer mortality statistics. (2014). Available at: http://www.cancerresearchuk.org/health-professional/cancer-statistics/worldwide-cancer/mortality#heading-One. (Accessed: 16th March 2017)
4. Gezondheidsplein. Longkanker. (2011). Available at: https://www.gezondheidsplein.nl/aandoeningen/longkanker/item36780. (Accessed: 14th February 2017)
5. Travis, W. D. *et al.* The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J. Thorac. Oncol.* **10,** 1243–1260 (2015).
6. van Meerbeeck, J. P., Fennell, D. A. & De Ruysscher, D. K. Small-cell lung cancer. *Lancet* **378,** 1741–1755 (2011).
7. Kanker.nl. Niet-kleincellige longkanker. Available at: https://www.kanker.nl/bibliotheek/niet-kleincellige-longkanker/wat-is/461-niet-kleincellige-longkanker. (Accessed: 14th February 2017)
8. George, J. *et al.* Comprehensive genomic profiles of small cell lung cancer. *Nature* **524,** 47–53 (2015).
9. Masuda, A. & Takahashi, T. Chromosome instability in human lung cancers: possible underlying mechanisms and potential consequences in the pathogenesis. *Oncogene* **21,** 6884–6897 (2002).
10. National Cancer Institute, N. H. G. R. I. Home - The Cancer Genome Atlas - Cancer Genome - TCGA. Available at: https://cancergenome.nih.gov/. (Accessed: 15th February 2017)
11. Gao et al. Sci. Signal. 2013 & Cerami et al. Cancer Discov. 2012. cBioPortal for Cancer Genomics. Available at: http://www.cbioportal.org/. (Accessed: 15th February 2017)
12. Genohub. Whole Genome Sequencing (WGS) vs. Whole Exome Sequencing (WES). (2015). Available at: https://blog.genohub.com/2015/02/21/whole-genome-sequencing-wgs-vs-whole-exome-sequencing-wes/. (Accessed: 14th February 2017)
13. Rajagopalan, H. & Lengauer, C. Aneuploidy and cancer. *Nature* **432,** 338–341 (2004).
14. Silva. FISH protocols. (2010). Available at: https://www.arb-silva.de/fish-probes/fish-protocols/. (Accessed: 14th February 2017)
15. Bakker, B., van den Bos, H., Lansdorp, P. M. & Foijer, F. How to count chromosomes in a cell: An overview of current and novel technologies. *BioEssays* **37,** 570–577 (2015).
16. Anderson, R. in *Methods in molecular biology (Clifton, N.J.)* **659,** 83–97 (2010).
17. Pinkel, D. *et al.* High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* **20,** 207–211 (1998).
18. Wang, D. G. *et al.* Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280,** 1077–82 (1998).

19. Bogh, L. D. & Duling, T. A. Flow cytometry instrumentation in research and clinical laboratories. *Clin. Lab. Sci.* **6,** 167–73

20. Givan, A. L. in *Methods in molecular biology (Clifton, N.J.)* **699,** 1–29 (2011).

21. Mitelman F, J. B. and M. F. (Eds. . Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer. (2017). Available at: https://cgap.nci.nih.gov/Chromosomes/Mitelman. (Accessed: 14th February 2017)

22. Junttila, M. R. & de Sauvage, F. J. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature* **501,** 346–354 (2013).

23. Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501,** 338–345 (2013).

24. Marusyk, A. & Polyak, K. Tumor heterogeneity: Causes and consequences. *Biochim. Biophys. Acta - Rev. Cancer* **1805,** 105–117 (2010).

25. Venkatesan, S. & Swanton, C. Tumor Evolutionary Principles: How Intratumor Heterogeneity Influences Cancer Treatment and Outcome. *Am. Soc. Clin. Oncol. Educ. B.* **36,** e141–e149 (2016).

26. Swanton, C. Intratumor heterogeneity: evolution through space and time. *Cancer Res.* **72,** 4875–82 (2012).

27. van den Bos, H. *et al.* Single-cell whole genome sequencing reveals no evidence for common aneuploidy in normal and Alzheimer's disease neurons. *Genome Biol.* **17,** 116 (2016).

28. Hammerman, P. S. *et al.* Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489,** 519–525 (2012).

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

30. Swanton, C. & Govindan, R. Clinical Implications of Genomic Discoveries in Lung Cancer. *N. Engl. J. Med.* **374,** 1864–1873 (2016).

31. Sarkozy, C., Traverse-Glehen, A. & Coiffier, B. Double-hit and double-protein-expression lymphomas: aggressive and refractory lymphomas. *Lancet Oncol.* **16,** e555–e567 (2015).

32. Dhillon, A. S., Hagan, S., Rath, O. & Kolch, W. MAP kinase signalling pathways in cancer. *Oncogene* **26,** 3279–3290 (2007).

33. Kim, E. K. & Choi, E.-J. Pathological roles of MAPK signaling pathways in human diseases. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1802,** 396–405 (2010).

34. Karachaliou, N. *et al.* KRAS Mutations in Lung Cancer. *Clin. Lung Cancer* **14,** 205–214 (2013).

35. Yuan, T. L. & Cantley, L. C. PI3K pathway alterations in cancer: variations on a theme. *Oncogene* **27,** 5497–5510 (2008).

36. Paul, M. K. & Mukhopadhyay, A. K. Tyrosine kinase - Role and significance in Cancer. *Int. J. Med. Sci.* **1,** 101–115 (2004).

37. Hast, B. E. *et al.* Cancer-Derived Mutations in KEAP1 Impair NRF2 Degradation but not Ubiquitination. *Cancer Res.* **74,** 808–817 (2014).

38. Ellis, E. F. Pharmacologic therapy of asthma. *Postgrad. Med.* **59,** 127–33 (1976).

39. Reuter, S., Gupta, S. C., Chaturvedi, M. M. & Aggarwal, B. B. Oxidative stress, inflammation, and cancer: how are they linked? *Free Radic. Biol. Med.* **49,** 1603–16 (2010).

40. LAU, A., VILLENEUVE, N., SUN, Z., WONG, P. & ZHANG, D. Dual roles of Nrf2 in cancer. *Pharmacol. Res.* **58,** 262–270 (2008).

41. Chen, H.-Y. & Chen, R.-H. Cullin 3 Ubiquitin Ligases in Cancer Biology: Functions and Therapeutic Implications. *Front. Oncol.* **6,** 113 (2016).

42. Wagner, A. J., Le Beau, M. M., Diaz, M. O. & Hay, N. Expression, regulation, and chromosomal localization of the Max gene. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 3111–5 (1992).

43. Fackenthal, J. D. & Godley, L. A. Aberrant RNA splicing and its functional consequences in cancer cells. *Dis. Model. Mech.* **1,** 37–42 (2008).

44. Tang, Z. *et al.* CREB-binding protein regulates lung cancer growth by targeting MAPK and CPSF4 signaling pathway. *Mol. Oncol.* **10,** 317–329 (2016).

45. Gayther, S. A. *et al.* Mutations truncating the EP300 acetylase in human cancers. *Nat. Genet.* **24,** 300–303 (2000).

46. Nakamura, T. M. *et al.* Telomerase catalytic subunit homologs from fission yeast and human. *Science* **277,** 955–9 (1997).

47. Allenspach, E. J., Maillard, I., Aster, J. C. & Pear, W. S. Notch signaling in cancer. *Cancer Biol. Ther.* **1,** 466–76

48. Koch, U. & Radtke, F. Notch and cancer: a double-edged sword. *Cell. Mol. Life Sci.* **64,** 2746–2762 (2007).

49. HUGO Gene Nomenclature Committee. ASCL4 Symbol Report. Available at: http://www.genenames.org/cgi-bin/gene_symbol_report?hgnc_id=HGNC:24311. (Accessed: 22nd February 2017)

50. Gordon, D. J., Resio, B. & Pellman, D. Causes and consequences of aneuploidy in cancer. *Nat. Rev. Genet.* **13,** 189 (2012).

51. Center, R., Lukeis, R., Vrazas, V. & Garson, O. M. Y chromosome loss and rearrangement in non-small-cell lung cancer. *Int. J. cancer* **55,** 390–3 (1993).

52. Wong, H. Y. *et al.* TMSB4Y is a candidate tumor suppressor on the Y chromosome and is deleted in male breast cancer. *Oncotarget* **6,** 44927–40 (2015).

53. Bell, J. *et al.* Integrative genomic and gene expression analysis of chromosome 7 identified novel oncogene loci in non-small cell lung cancer. *Genome* **51,** 1032–1039 (2008).

54. Engelman, J. A. *et al.* MET Amplification Leads to Gefitinib Resistance in Lung Cancer by Activating ERBB3 Signaling. *Science (80-.).* **316,** (2007).

55. Knouse, K. A., Davoli, T., Elledge, S. J. & Amon, A. Aneuploidy in Cancer: Seq-ing Answers to Old Questions. *Annu. Rev. Cancer Biol* **11820,** 1–18 (2017).

56. Hodgson, J. G., Chin, K., Collins, C. & Gray, J. W. Genome amplification of chromosome 20 in breast cancer. *Breast Cancer Res. Treat.* **78,** 337–45 (2003).

57. Gisselsson, D. *et al.* Generation of trisomies in cancer cells by multipolar mitosis and incomplete cytokinesis. *Proc. Natl. Acad. Sci.* **107,** 20489–20493 (2010).