Deciphering the clonal evolution of non-small cell lung cancer

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Arthur Krause

aus Deutschland

2020

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Torsten Schwede

Prof. Dr. Lukas Bubendorf

Prof. Dr. Luigi Terracciano

Basel, der 17.12.2019

Prof. Dr. Martin Spiess
Dekan

For my family…

RIP W. K.

Using no way as a way,
having no limitation as limitation
Bruce Lee

Table of contents

List of Figures

List of Abbreviations

aCGH	Array comparative genomic hybridization
APOBEC	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide
ASC	Adenosquamous carcinoma of the lung
BLCA	Bladder
BRCA	Breast
BWA	Burrows-Wheeler Aligner
CCF	Cancer cell fraction
CIN	Chromosomal instability
CNV	Chromosomal copy number variations
CRC	Colorectal
CTC	Circulating tumor cells
ctDNA	Circulating tumor DNA
CTSE	Cathepsin E
DAPI	4’,6-Diamin-2phenylindol
FACETS	Fraction and Allele-Specific Copy Number Estimates from Tumor Sequencing, algorithm for copy numbers
GATK	Genome Analysis Toolkit
GBM	Glioblastoma
HIV	Human immune deficiency virus
HNSC	Head and neck squamous cell
ICI	Immune checkpoint inhibitors
IHC	Immune histochemistry
Indels	Small insertions and deletions
ITH	Intratumoral heterogeneity
KIRC	Kidney renal cell
LCC	Large cell carcinoma
LUAD	Adenocarcinoma of the lung
LUSC	Squamous cell carcinoma of the lung
NSCLC	Non-small cell lung cancer
OV	Ovary
PAH	Polycyclic aromatic hydrocarbons
pCK	Pan-cytokeratin
s-CIN	Loss or amplification of parts of a chromosome
SCLC	Small cell lung cancer
SLC5A7	Solute carrier family 5 member 7

SNV	Single nucleotide variant
SOX2	Transcription factor SRY-box 2
TERT	Gene that encodes the protein telomerase
TNSA	Tobacco-specific N-nitrosamine
TKI	Tyrosine kinase inhibitor
TTF-1	Thyroid transcription factor 1
VAF	Variant allelic fraction
UCEC	Uterine cervix
w-CIN	Loss or amplification of whole chromosome
WES	Whole-exome sequencing
WGD	Whole-genome duplication

Summary

Lung cancer is the most common and deadliest malignancy worldwide and the emergence of metastases is the main cause of lethality. One reason for this high mortality is intratumoral heterogeneity (ITH). This is the presence of multiple subclones with genetic heterogeneity as a consequence of cancer evolution, which means that tumor cells undergo molecular and phenotypic changes. Therefore, resistant or metastatic subclones can survive unnoticed for a long time and often remain undiscovered at the time of diagnosis.

The aim of this work was to shed light on the tumor evolution of three subtypes of non-small cell lung cancer (NSCLC), the most common form of lung cancer. First, the clonal relationship of primary lung adenocarcinomas (LUAD) and their matched metastases was deciphered by studying copy number variations (CNVs) and somatic mutations (SNVs) in a unique cohort of 16 patients. Lung squamous cell carcinomas (LUSC) were genomically characterized in a second cohort consisting of six patients with matched, clonally related metastases using a similar approach. And third, genomic alterations of adenosquamous carcinoma (ASC), which is a tumor entity with a mixture of LUAD and LUSC components, were investigated in three cases.

Tumor purity is a major constraint in genomic analyses. To overcome this limitation, we refined and validated a flow-sorting approach to increase the availability of tumor material with high tumor purity. This strategy increased tumor purity from 54% to 92% in LUAD and from 33% to 70% in LUSC. Using this approach, we found that more than 88% of CNVs and 80% of SNVs are shared between primary tumors and metastases in LUAD. Similarly, more than 76% of CNVs and 64% of SNVs were common between primary LUSC and the matched metastases. Both studies therefore demonstrated a close genomic relationship between primary tumors and metastases with only a limited number of subclonal genetic alterations. In addition, tumor ploidy remained stable over time, regardless if tumors were diploid or aneuploid. Together, this suggests an early accumulation of CNVs and SNVs, often before metastatic spread, with limited chromosomal instability over the disease course of both tumor entities, at least from the time the primary tumors are established.

We also found high concordance of CNVs and SNVs in the LUAD and LUSC components of all three ASCs. This unequivocally confirmed a monoclonal origin of the two distinct components in ASC. Strikingly, we found LUAD-specific mutations that are rarely observed in LUSC, which suggests a LUAD-like common ancestor cell as the cell of origin for ASC.

Our findings shed light on the genomic evolution of LUAD and LUSC and suggest that most genetic alterations can be detected in the primary tumors, thereby ITH is unlikely to significantly alter the landscape of targetable alterations due to sampling. In addition, our work is the study on the genomic landscape of ASC, a rare and highly aggressive form of NSCLC, in a comprehensive manner. Although further studies in larger cohorts are needed to confirm our findings, our studies will serve as a basis for future explorations and eventually help patients in dire need.

1. Introduction

1.1 Importance of the work

In recent years, multi-region sequencing discovered extensive ITH in several tumors. ITH results from the competition and cooperation of genetically distinct tumor subclones and explains in part how tumor cells adapt to new environment in the form of resistance mutations or metastatic development. Nevertheless, most of the studies considered only larger single tumors. Yet, metastases are the reason why cancers are so deadly and should therefore be studied in more detail. Practically in daily routine, sampling of metastases to study ITH is impossible. To complicate matters, NSCLCs consist of a relatively high proportion of non-tumor cells mixed with tumor cells, making genome analysis even more difficult. We overcame this limitation by establishing a refined nuclei flow-sorting approach, which increased tumor purity substantially.

Regarding metastasis, this study sheds light on the genomic evolution of NSCLC by exploring genomic alterations in matched pairs of primary tumors and clonally related metastases in LUAD and LUSC, the two most common subtypes of NSCLC. In both cohorts, we detected a high concordance of CNVs and SNVs among primary tumors and metastases. Most driver alterations were early events in the evolution of the tumors and present at primary and metastatic sites. This may have direct clinical implications, focusing on truncal alterations of both tumors instead of targeting on diminishing subclonal diversity.

Patients with metastatic NSCLCs are not the only ones challenging to treat. For NSCLCs with profound phenotypic heterogeneity, there is often no consensus on how they should be treated. One such example is ASC, a rare but highly aggressive NSCLC subtype characterized by a mixture of LUAD and LUSC components. Making a decision in ASC treatment is a difficult task, because an efficient therapy for one entity does not necessarily apply to the other.

In the last part of this work, we deciphered the genomic evolution of three ASCs by separately examining the genomic alterations of the LUAD and LUSC components and providing unprecedented proof for a monoclonal origin suggesting ASC as a transitional state that transdifferentiates from LUAD to LUSC. This might have not only an impact on the way, how to treat ASC, on the same line, diagnosis and subtyping of ASC might be seen earlier.

We believe that our studies are fundamental to get an understanding of the evolutionary dynamics in NSCLC that might contribute for further treatment development.

1.2 Lung cancer

Lung cancer is one of the most frequent and lethal tumor malignancies worldwide [1,2]. When both sexes are considered, it is the cancer with the highest incidence and mortality rate. In 2018, more than two million new cases and approximately 1.8 million deaths were recorded (Figure 1) [1,2]. In females and males, it ranks just after breast cancer and prostate cancer, respectively, followed by colorectal cancer in both sexes [3]. In developed countries, the lung-cancer incidence rate is higher than in less-developed countries and is predominantly related to smoking habits. The trend for lung cancer incidence has been declining in men in developed countries, probably due to tobacco control policies [4,5]. In contrast, the rates of lung cancer in women are increasing worldwide [5]. The difference in gender is the result of historical patterns of smoking and cessation, where women started to smoke later. The main risk factor for lung cancer remains tobacco consumption, which increases with the quantity and duration of smoking [3]. Besides gender, other demographic factors such as ethnicity, geography, education, income, and age play a role in the development of lung cancer and its subtypes [6].

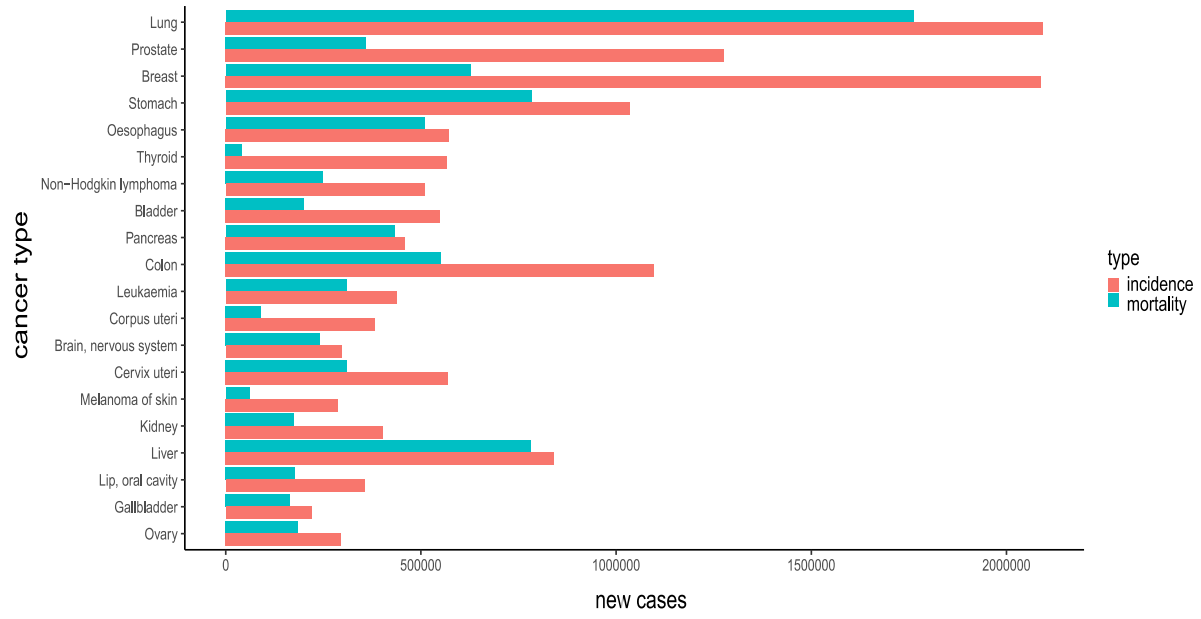


Figure 1. Estimated number of incident cases and mortality rate of the most common cancers worldwide. Lung cancer remains the most common and most mortal tumor taken together both sexes and all ages, respectively [modified from [7]].

1.3 Non-small cell lung cancer

Lung cancer survival depends predominantly on the particular stage, which describes the local extension and spreading of the tumor. Despite new approaches that detect early cancer and improved therapies, lung cancer survival of five years varies between 49% in stage 1A down to 2% in stage 4 (Figure 2). The earlier the diagnosis, the higher the survival rate [8]. Different treatment options are available to patients with localized non-small cell lung cancer (NSCLC) including tumor resection and radiation therapy with or without chemotherapy. Late-stage metastatic lung cancer is no longer treatable by surgery. Therefore, few metastases are resected to diagnose metastatic NSCLC. The fact that only small biopsies or cytology specimens are available from most patients explains the differences in understanding of the tumors on a molecular level for metastasized lung cancer.

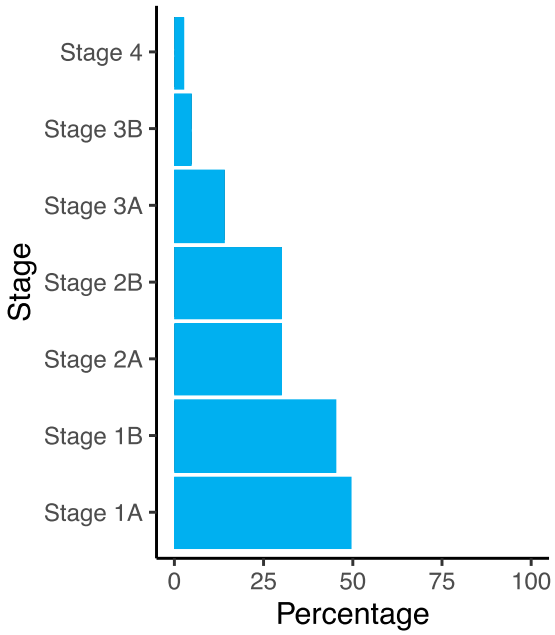


Figure 2. Lung cancer 5-year survival rates in the U.S. Lung cancer survival correlates strongly with the staging. The later the staging the poorer the survival rate [modied from [9]].

Histologically, lung cancer is classified into small cell lung cancer (SCLC) and NSCLC, which make up 15% and 85%, respectively [10]. NSCLC can be further divided by morphology into adenocarcinoma (60%, LUAD), squamous cell carcinoma (20%, LUSC), and large cell carcinoma (10%, LCC). There are also other rare but aggressive subtypes, such as

adenosquamous carcinoma (0.4% - 4%, ASC) and sarcomatoid carcinomas (0.1% - 0.4% of NSCLC, Figure 3) [11–14].

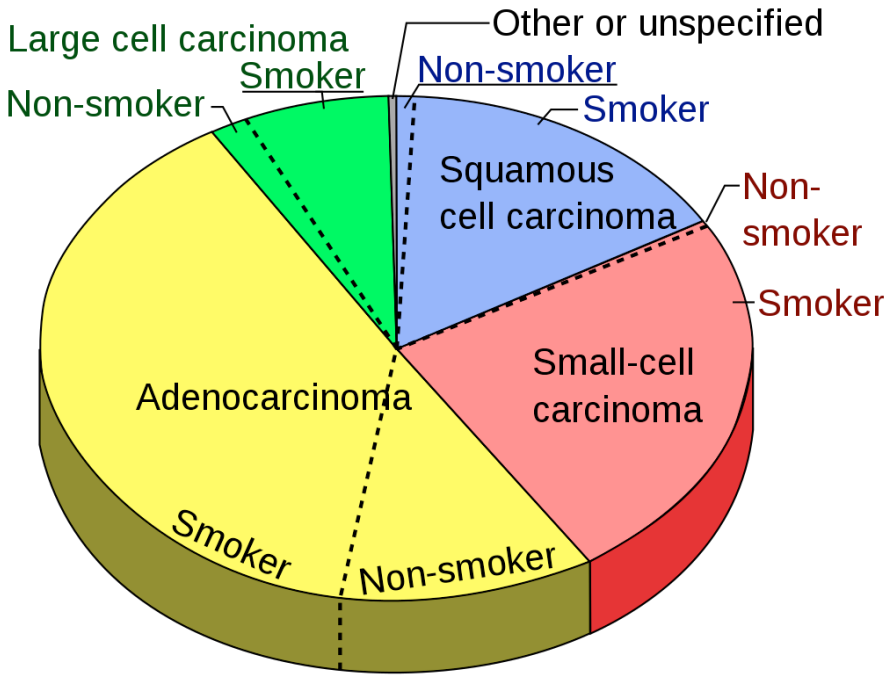


Figure 3. Histological classification of smoking and non-smoking related lung cancer. Lung cancer is subclassified into SCLC and NSCLC including LUAD, LUSC, large cell carcinoma and others such as ASC. The majority of the subtypes is strongly associated with smoking [modified from [15,16]].

1.4 Risk factors

1.4.1 Tobacco

The major risk factor contributing to lung cancer is tobacco smoking [16,17]. Already in the 1950s, the link was made between cigarette smoking and lung cancer [18,19]. Duration and quantity of smoking influence the lifetime risk of developing lung cancer. SCLC and LUSC are more closely associated with smoking than LUAD (Figure 3). Interestingly, in Western countries, the incidence of LUSC has decreased because of the augmented cessation of smoking in recent years. Conversely, the relative incidence of LUAD increased [20]. Consumption of cigars, pipes, or hookahs demonstrates similar consequences as smoking cigarettes. One single large cigar contains the same amount of tobacco as an entire pack of cigarettes [21–23]. Tobacco combustion generates at least 60 known carcinogens such as polycyclic aromatic hydrocarbons (PAH) and tobacco-specific N-nitrosamines (TNSAs) [24,25]. Carcinogens affect the DNA in several ways including the formation of DNA adducts and inducing free radical damage [26].

1.4.2 Environmental risk factors

The role of second-hand tobacco smoke in lung cancer is not to be underestimated. The smoke of so-called “passive smoking” contains PAH, TNSAs, and additional carcinogens, which are in a dose-response relationship between exposure and lung cancer risk [26,27]. Concentrations of PAH are four times higher in second-hand smoke than in the filtered mainstream cigarette smoke [28]. Damaging DNA adducts from tobacco carcinogens were also found in the urine of non-smokers due to side-stream tobacco smoke [28]. Living with a smoker spouse increases the risk of lung cancer in non-smokers by up to 20-30% [29]. Dust and surfaces of households of smokers, even if they smoke outside, are contaminated with environmental tobacco smoke [30]. Exposure to passive smoke during childhood is associated with lung cancer [31]. Other occupational exposure can have an impact on lung cancer development, such as asbestos, radon, PAH, heavy metals, air pollution, and coal burning for cooking and heating [32–34].

1.4.3 Infection

Viruses and other diseases play a role in the tumorigenesis of lung cancer. Evidence suggests an association between viral infections, such as Epstein-Barr viruses or human papillomaviruses, and the development of lung cancer [35,36]. The most common non-AIDS defining malignancy in HIV (human immunodeficiency virus) infected people is lung cancer [37]. Prior suffering of other diseases, such as chronic obstructive pulmonary disease, asthma, tuberculosis, chronic bronchitis, or emphysema are also known to increase the risk of lung cancer [38,39].

1.4.4 Genetic risk factors

Genetic predisposition can also increase the risk of lung cancer development, and a positive family history for lung cancer increases the risk of development [40–43]. Approximately 8% of lung cancers are inherited due to chromosomal altered regions, such as 5p15 and 15q25-26 [44,45]. The 5p15 region contains the telomerase reverse transcriptase (*TERT*) gene, which plays a crucial role in cell division. Studies showed a positive link between adenocarcinoma in smokers and non-smokers [46]. The region of 15q25-26 is associated with nicotine dependence and higher sensitivity for lung cancer [47].

1.4.5 Never smokers

Individuals who consume fewer than 100 cigarettes in their lifetime are regarded as never smokers [48]. Worldwide, around 25% of all diagnosed lung cancers are never smokers [49]. In recent years, the proportion of never smokers suffering from lung cancer increased [6,48]. Globally, 15% of men and 53% of women with lung cancer are never-smokers [50]. This occurrence is seen predominantly in women and younger patients. Particularly, the population in East and Southeast Asia have a higher occurrence of lung cancer in never smokers [6,48,51].

1.5 Subtypes of NSCLC

SCLC and NSCLC represent different categories of lung cancers. NSCLC is divided into its subtypes LUAD, LUSC, LCC, and more rare subtypes such as ASC. Even if tumors are categorized according to their histological morphology, immunohistological staining, and genomic profile, there is still no definite proof of the cell of origin in humans (Figure 4).

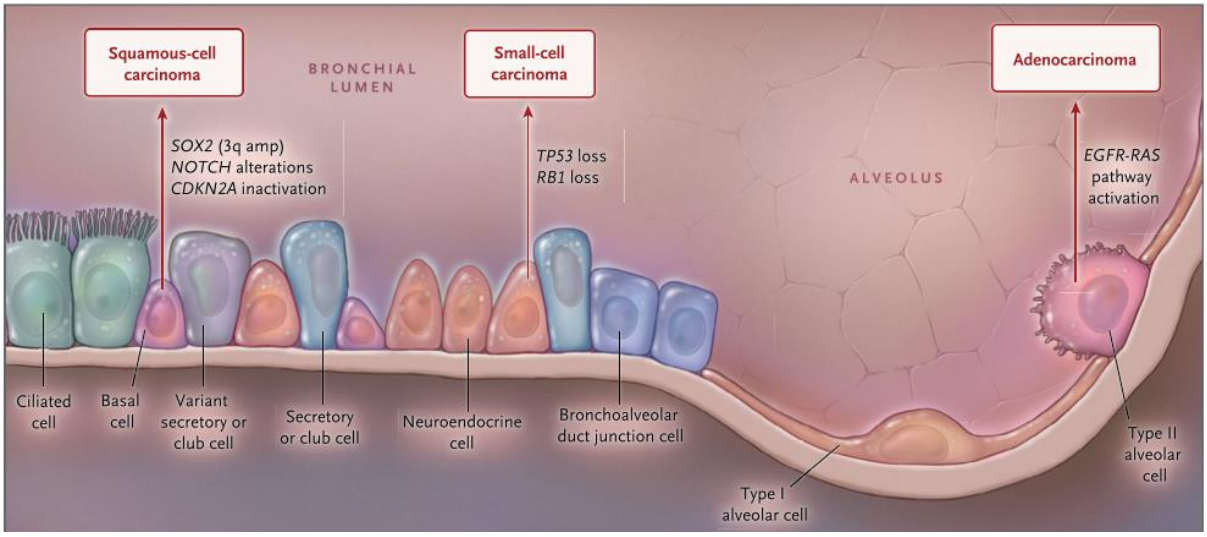


Figure 4. Possible cell of origin in lung cancer. Potential cell of origin of three frequent histologic subtypes of lung cancer. LUAD, LUSC and SCLC are illustrated [modified from [52]].

1.6 Adenocarcinoma of the lung (LUAD)

1.6.1 Histology

Adenocarcinoma is often displayed as a heterogeneous mix of different histological patterns, which includes lepidic, acinar, papillary, micropapillary, solid, and with or without mucin production (Figure 5) [34,53]. Specific biomarkers to confirm adenocarcinoma are used in pathological diagnostics comprising TTF-1 (thyroid transcription factor 1) and Napsin A [20,54–56]. TTF-1, encoded by NKX2-1, represents a lineage biomarker, which is consistent with an origin in the distal lung and Napsin A is an aspartic proteinase [20,57].

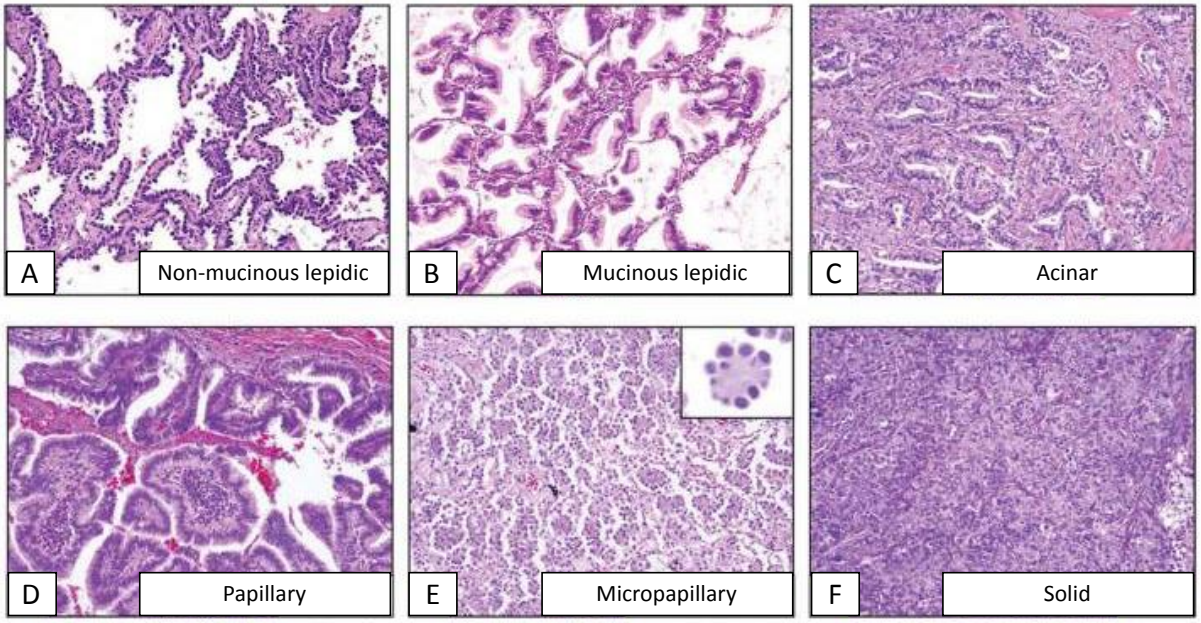


Figure 5. Histological patterns of LUAD. Different growth patterns of LUAD are illustrated as non-mucinous lepidic (A), mucinous lepidic (B), acinar (C), papillary (D), micropapillary (E) and solid (F) [modified from [58]].

1.6.2 Cell of Origin

The cells of origin of LUAD, LUSC, and ASC are still not precisely known. The lung consists of different cell types and the proportion of cell types changes along the proximal-distal axis. Recent studies demonstrated possible cell types that may give rise to a tumor, such as pneumocytes type II, bronchioalveolar duct junction cells, and club cells (former Clara cells) of the bronchoalveolar duct (Figure 4, reviewed in [52]). Older studies already proposed that club cells or pneumocytes type II could be the potential cell of origin because of TTF-1 positivity in IHC (immunohistochemistry) staining. Pneumocytes type I and type II form the smallest compartment of the lung, the alveoli, which enables gas exchange. Type II pneumocytes are

also responsible for the renewal process of both type I and type II and they also express TTF-1 [59]. Bronchioalveolar duct junction cells are regarded as stem cells functioning in a repair and self-renewal manner during lung regeneration [60]. Club cells fulfill a protective role and secrete a variety of proteins [61]. Depending on the smoking status, a different signaling pathway is active (Figure 6). Non-smokers may harbor an activating *EGFR* mutation, while mutations in *KRAS* are strongly associated with smoking-induced cancer [62].

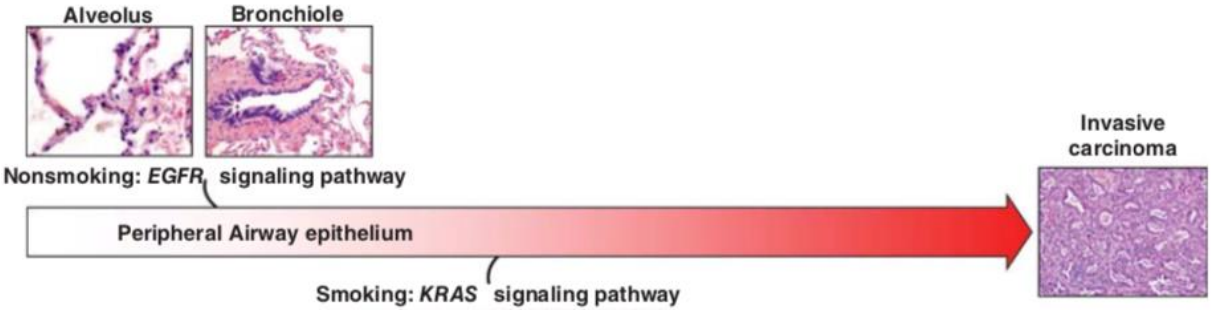


Figure 6. Molecular pathogenesis of LUAD. It is hypothesized, LUAD derives from the peripheral airway epithelium as the alveolus or brionchioles. Non-smoker and smoker tumors follow a different pathway to develop invasive carcinoma [modified from [62]].

1.7 Squamous cell carcinoma of the lung (LUSC)

1.7.1 Histology

LUSC is characterized by squamous differentiation and different layers of cells. The histological subtypes of LUSC include basaloid carcinoma, keratinized or non-keratinized LUSC cells (Figure 7) [34,55]. LUSC are defined histologically by keratinization and/or intracellular bridges, and they expressing CK5/6 (cytokeratin), SOX2 (transcription factor SRY-box 2), p63 and the truncated version p40 [20,57,63–68].

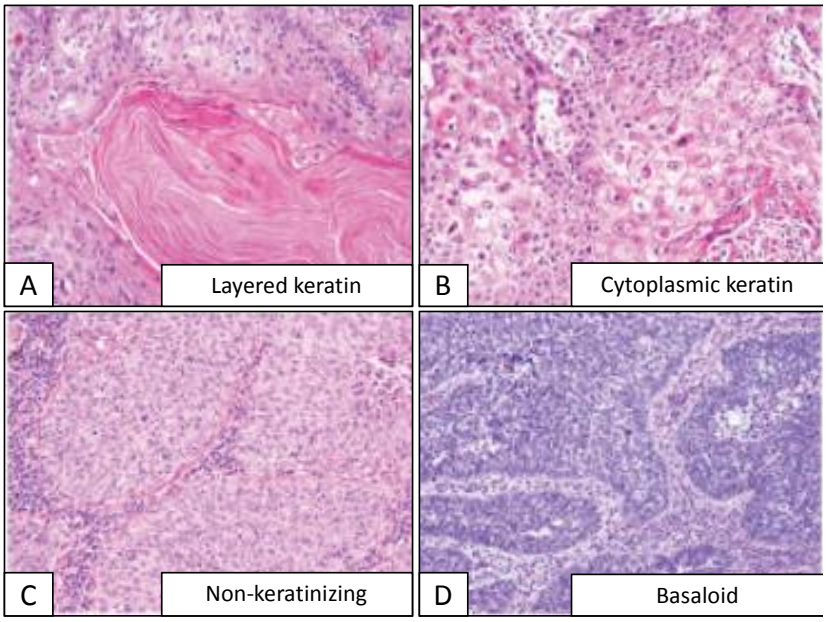


Figure 7. Histological patterns of LUSC. Different growth patterns of LUSC are displayed as layered keratin (A), cytoplasmic keratin (B), non-keratinizing (C), and basaloid (D) [modified from [69]].

1.7.2 Cell of origin

In a healthy lung, squamous cells are not present in the airways. It is postulated that normal epithelium accumulates stepwise molecular abnormalities, which leads to invasive carcinoma (Figure 8). LUSC is found in the more proximal airways and it is believed that it originates there from basal cells (Figure 4). Basal cells are involved in the developing and postnatal respiratory tract epithelium [70]. This hypothesis is supported by mouse studies using lineage tracing [52].

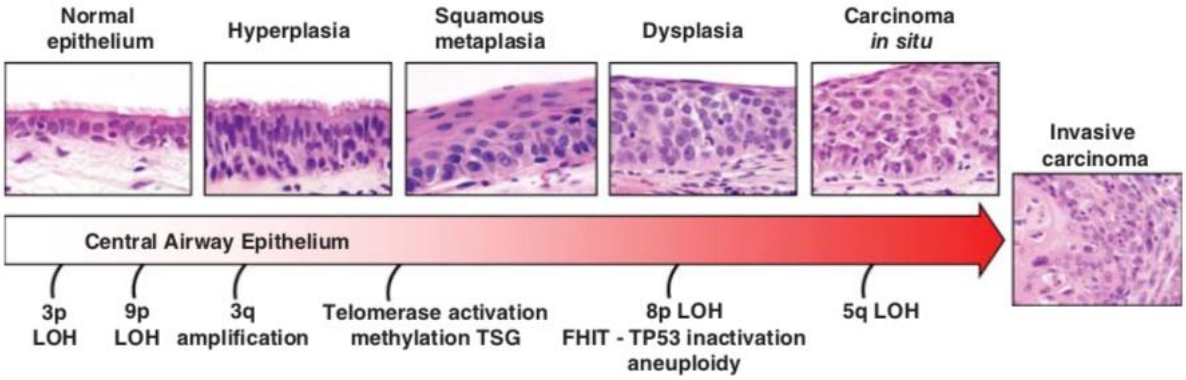


Figure 8. Molecular pathogenesis of LUSC. Normal epithelium develops sequentially to invasive carcinoma due to acquiring genomic and epigenomic abnormalities [modified from [62]].

1.8 Adenosquamous carcinoma

ASC is a rare subtype that constitutes only up to 4% of all NSCLC. By definition, it has a LUAD and LUSC morphology of at least 10% of each component [12]. ASC displays the typical histological features of both (Figure 9). The LUSC component always expresses p40, where TTF-1 is positive in approximately 80% of the LUAD components. The dichotomous tumor is more aggressive than each of its components in isolation [71,72]. The lack of large clinical trials due to the rarity of this tumor type explains why there is no well-established systemic treatment for ASC. For the same reason, not much is known about the genomic landscape of ASC [73–75]. Historically, two hypotheses were postulated. The ASC tumor derives from a monoclonal or polyclonal origin. Studies suggested a monoclonal origin due to shared single mutations in both components, which form a tumor with two entities [75–79]. Nevertheless, studies demonstrated possible transdifferentiation from LUAD to LUSC (reviewed in [80]). However, another study indicates the opposite: a transdifferentiation from LUSC to LUAD [74]. In 1981, dissected ASC cells of rat lung were isolated to single cells and subcutaneously inoculated into mice. The mice developed different manifestations of ASC, LUAD, and LUSC, and they demonstrated significant tumor cell plasticity [81]. Further mouse studies revealed, due to transdifferentiation, a transition from LUAD to a LUSC phenotype as a potential escape mechanism from drug treatment [80].

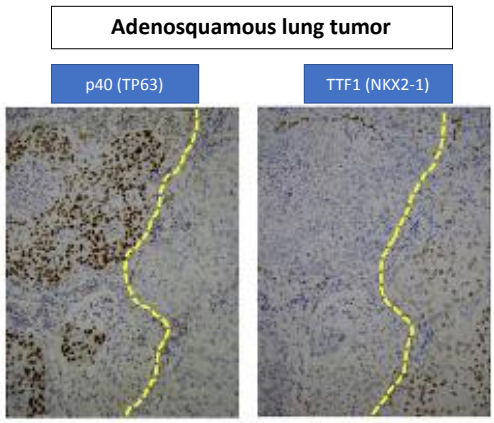


Figure 9. ASC of the lung. ASC is expressing the LUSC marker p40 and the LUAD marker TTF-1 [modified from [82]].

1.9 Treatment

Molecular heterogeneity complicates the treatment of NSCLC and the development of effective therapies. The current landscape of treatment consists of surgery, radiation, chemotherapy, targeted therapy, and immunotherapy. Surgery and radiation in a curative stage are the treatment of choice of NSCLC at an early stage, including stage I-II and in some stage III patients [83]. In the case of advanced disease (stage III and IV), systemic treatment is governed by chemotherapy, targeted treatments, and immunotherapy. Further development led to molecularly targeted therapies that aim at specific alterations in the cancer genome and, lastly, immunotherapies support the own immune system to defend against the invading cancer cells.

1.9.1 Chemotherapy

Historical use of cytotoxic therapy demonstrated no significant difference in the response rate between LUAD and LUSC; therefore, they were treated equally [84]. In 2011, a clinical trial demonstrated superior efficacy in non-squamous patients through the use of pemetrexed, a common chemotherapeutic drug [85]. Since this study, histological distinction between these two subtypes has become a necessity. Currently, chemotherapeutics are often combined to increase efficacy, such as pemetrexed and platinum-based chemotherapy in patients with non-squamous NSCLC [8].

1.9.2 Molecular targeted therapies

Targeted therapies changed the management of cancer. Tumor genotyping permits personalized therapy due to drugs that target the oncogenic genes, proteins, or tissue environment that maintains the tumor. Clonal driver mutations represent an attractive target present in all cancer cells.

The first targeted therapy was approved for use against *EGFR*-mutated NSCLC. *EGFR* is part of the receptor tyrosine kinase family and plays an important role in cell proliferation, survival, invasion, and angiogenesis [86]. Heterozygous mutations result in constitutive *EGFR* activation without the need for a ligand [87]. Most frequent *EGFR* mutations such as exon 19 deletion or missense mutation on exon 21 (L858R) are correlated with *EGFR* tyrosine kinase inhibitor (TKI) sensitivity [88]. ITH and ongoing tumor evolution accumulate mutations that are able to cause an acquired resistance. A missense mutation in exon 20 (T790M) is a common resistance mechanism [89,90]. Treatment against *EGFR* T790M-mutant tumors, osimertinib, indicates efficacy superior to other *EGFR* mutation-positive drugs in advanced NSCLC [91,92]. Most targeted therapies in NSCLC aim at LUAD, since there are more mutations that are

targetable. Furthermore, LUADs harbor more *EGFR* T790M mutations than any other NSCLC subtype [91].

Further promising results are delivered by studies with circulating tumor DNA (ctDNA). Tumor cells release ctDNA in the blood circulation by apoptosis, necrosis or active secretion [93]. Analysis of ctDNA is currently used to detect resistance mutation after targeted treatment, which includes T790M after *EGFR* TKI treatment [94]. Strikingly, single and even multiple tumor biopsies cannot represent the ITH of tumors, while ctDNA could provide a remedy in the future.

1.9.3 Immunotherapies

Tumor cells develop several mechanisms to escape the immune system. In one such method, tumor cells evade the response of the immune system by expressing PD-1 (programmed death protein 1). PD-1 interacts with PD-L1 (programmed death-ligand) of T cells and results in suppression of the immune system [95]. Tumor cells become invisible to the immune cells and can no longer be attacked. Immune checkpoint inhibitors (ICI) block the binding of PD-1 or PD-L1 and render the tumor visible so that the T cell antitumor function is restored [96]. ICI demonstrated beneficial effects in a variety of NSCLC patients [97]. The novel approach of ICI offers new hope in cancer treatment.

1.10 Genomic landscape of NSCLC

The next-generation sequencing changed the view on diseases such as cancer and, above all, the treatment options. Whole exome sequencing (WES) studies revealed peculiarities in the genomes of various cancers [98,99]. Genomes of different cancers harbor different amounts and types of mutations. Smokers with LUSC have, on average, 9.0 mutations per megabase (mutations/Mb) compared to LUAD with 6.3 mutations/Mb, or ASC with 5.4 mutations/Mb compared to non-smokers who have five to six times fewer mutations [100–102]. High tumor mutational burden and a higher number of C:G>A:T transversions are characteristic of smoke-related tumors. In contrast, never-smoker tumors have a low TMB and an increased number of C:G>T:A transitions [103,104].

LUAD and LUSC share six significantly mutated cancer genes, including *TP53, NF1, RB1, ARID1A, CDKN2A,* and *PIK3CA*. The number of common mutated genes is small. LUSC harbors seven significantly mutated cancer genes not shared with LUAD, in contrast to 21 in

LUAD (Figure 10). Prominent mutations for LUAD occur in the cancer genes *KRAS, KEAP1, EGFR, STK11, BRAF, ERBB2, MET, ATM,* and more. On the other hand, most mutated cancer genes in LUSC are *NFE2L2, PTEN, MLL2, FAT1, NOTCH1, KDM6A,* and *HRAS* [65,105–110].
Common focal chromosomal aberrations have been found in both entities as well as unique private (not shared) events. LUAD and LUSC have 13 focal amplifications in common, which include oncogenes such as *TERT, MYC, MDM2, EGFR, ERBB2, FGFR1,* and more. There are seven private amplified oncogenes in LUAD and eight in LUSC. In terms of focal deletions, LUAD is known to lose the tumor suppressor gene *SMAD4* and LUSC frequently lose *PTEN, FOXP1, FAT1, NF1,* and *KDM6A.* In both cases, 13 focal deletions are recorded, and among them, the most significant genes are *CDKN2A, PTPRD, and WWOX*.
LUSC has more mutations overall while LUAD accumulates more damaging mutations in proliferation-related pathways. On the contrary, LUSC harbors more indels (insertions and deletions). The amount of gene fusions in proliferation-related pathways is similar in both entities [111].

The frequency of oncogenic pathway alterations varies. The RTK/RAS/MAPK pathway, which is involved in proliferation, cell survival, and translation, is more often changed in LUAD (74%) than in LUSC (54%). *EGFR*, *ERBB2*, *KRAS*, *MET,* and *BRAF* are frequent drivers in LUAD, whereas *FGFR1/2/3/4* and *HRAS* alterations are common in LUSC. The cell survival and proliferation pathway of p53 are more highly mutated in LUSC (86%) than in LUAD (61%). Similar proportions are seen in the cell cycle (LUSC 79%, LUAD 56%) and PI3K (LUSC 68%, LUAD 38%) pathways. The PI3K pathway includes common mutations in both lung cancer subtypes in *PIK3CA*, as well as *STK11* inactivating mutations in LUAD and *PTEN* inactivating mutations in LUSC. Remarkably, only 1% of all cancers have Nrf2 pathway alterations, which occurred at a rate of 15% and 25% in LUAD and LUSC, respectively. Both subtypes accumulate mutations in different genes in the Nrf2 pathway, which leads to an elevated oxidative stress response. LUAD has a high incidence of *KEAP1* inactivating mutations or deletions and LUSC has an increased rate of activating mutations or amplifications in *NFE2L2* (Figure 10) [98,99,111].

Several different studies also support the differences between these two cancer types at the transcriptomic and epigenomic levels. In addition to different pathways, genes and microRNA are also differentially expressed. Different expression levels of miRNA were displayed between the two tumors. Whereas LUSC demonstrated a higher expression of miR-205, LUAD expresses miR-375 [109,112,113]. Cell metabolism genes are expressed at higher levels in LUAD, whereas structural proteins such as cytokeratins are found in an increased amount in LUSC [114,115]. In 2018, a study was published that illustrated clear differences in expression and

methylation between LUAD and LUSC. *CTSE* (cathepsin E) was overexpressed and hypomethylated in LUAD but downregulated and hypermethylated in LUSC. The opposite was seen in *SLC5A7* (solute carrier family 5 member 7) [108–110]. A further study suggests another method to distinguish LUAD and LUSC by differentially methylated genes. The genes *CDKN2A* and *MGMT* are hypomethylated and *CDH13, RUNX3,* and *APC* are hypermethylated in LUAD compared to LUSC, which allows distinct discrimination on epigenetic level [116].

1.10.1 Mutational signatures

Exogenous and endogenous processes leave characteristic imprints in the form of mutations. Different DNA damaging processes result in distinct mutational patterns. For example, smoking-related mutations are linked to C:G>A:T mutations, whereas C:G>T:A is predominantly seen in UV-exposed skin cancers [117]. Endogenous causes for mutations include a defective DNA repair pathway or a defective APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide) editing. Six possible somatic base substitutions can occur in the mutational spectra (C:G>A:T, C:G>G:C, C:G>T:A, T:A>A:T, T:A>C:G, T:A>G:C). Moreover, it was proven that neighboring nucleotides influence the mutation rate [118]. Logically, 96 possible sequence contexts (6 types of base substitution, 16 possible bases before and after the mutation site) were categorized in a catalogue of mutational signatures [119]. In the case of mutational signatures, LUAD and LUSC both have not only the smoking signature (signature 4) in common, but they also share the age-related (signature 1), APOBEC (signature 2 and signature 13), and a signature of unknown origin (signature 5). Only LUAD displays a defective DNA mismatch repair signature (signature 6) and another private signature of unknown origin (signature 17) (Figure 10) [119–123].

Significantly mutated genes

	LUAD	LUSC
KRAS	x	
KEAP1	x	
EGFR	x	
STK11	x	
SMARCA4	x	
RBM10	x	
BRAF	x	
ERBB2	x	
SETD2	x	
MGA	x	
MET	x	
ATM	x	
U2AF1	x	
RIT1	x	
ARID2	x	
SMAD4	x	
CTNNB1	x	
APC	x	
RAF1	x	
NRAS	x	
MAP2K1	x	
TP53	x	x
NF1	x	x
RB1	x	x
ARID1A	x	x
CDKN2A	x	x
PIK3CA	x	x
NFE2L2		x
PTEN		x
MLL2		x
FAT1		x
NOTCH1		x
KDM6A		x
HRAS		x

Focal amplifications

	LUAD	LUSC
NKX2-1	x	
MCL1	x	
CDK4	x	
KRAS	x	
MECOM / TERC	x	
MET	x	
TERT	x	x
MYC	x	x
MDM2	x	x
CCND1	x	x
EGFR	x	x
ERBB2	x	x
WHSC1L1	x	x
FGFR1	x	x
18q11.2	x	x
PDGFRA	x	x
KIT	x	x
KDR	x	x
19p13.2	x	x
SOX2		x
REL		x
BCL11A		x
NFE2L2		x
CDK6		x
BCL2L1		x
MYCL1		x

Focal deletions

	LUAD	LUSC
SMAD4	x	
CDKN2A	x	x
PTPRD	x	x
WWOX	x	x
B2M	x	x
PDE4D	x	x
LRP1B	x	x
RB1	x	x
CSMD1	x	x
Xp22.2	x	x
21q21.1	x	x
ZMYND11	x	x
6p22.3	x	x
4q22.1	x	x
PTEN		x
FOXP1		x
FAT1		x
NF1		x
KDM6A		x

Mutational Signature

	LUAD	LUSC
Sig. 1 Age	x	x
Sig. 2 APOBEC	x	x
Sig. 4 Smoking	x	x
Sig. 5 Unknown	x	x
Sig. 6 Def. DNA MMR	x	
Sig. 13 APOBEC	x	x
Sig. 17 Unknown	x	
Others		x

Alteration frequencies %

	LUAD	LUSC
RTK/RAS/MAPK	74	54
p53	61	86
Cell cycle	56	79
PI3K	38	68
Hippo	23	28
Myc	23	12
NOTCH	21	31
Wnt	19	18
Nrf2	15	25
TGFbeta	10	11

Figure 10. Genomic distinction between LUAD and LUSC. Differences in mutations, chromosomal aberrations, mutational signatures and pathways between LUAD and LUSC [98,99,111,119].

1.11 Tumor purity

The bulk of the tumor consists of an admixture of non-malignant cells and tumor cells. The proportion of tumor cells is known as tumor purity. Depending on the type of tumor, the purity varies. Across the major solid cancers, ovary carcinoma and glioblastoma have the highest purity, while LUAD and LUSC have the lowest purity between 60% - 70%. Strikingly, it was discovered that tumor purity is reciprocally correlated with the number of mutations (Figure 11) [124]. The lung tumors have a high mutational load mainly due to chronic smoking. The non-malignant cells are stromal cells, whereas leukocytes constitute around 30% [125]). Whereas LUAD showed a high number of inflammatory cells, LUSC presented an elevated number of cells involved in wound healing, which reflects the different biology of these two entities and the different mechanisms of the immune system [126].

2

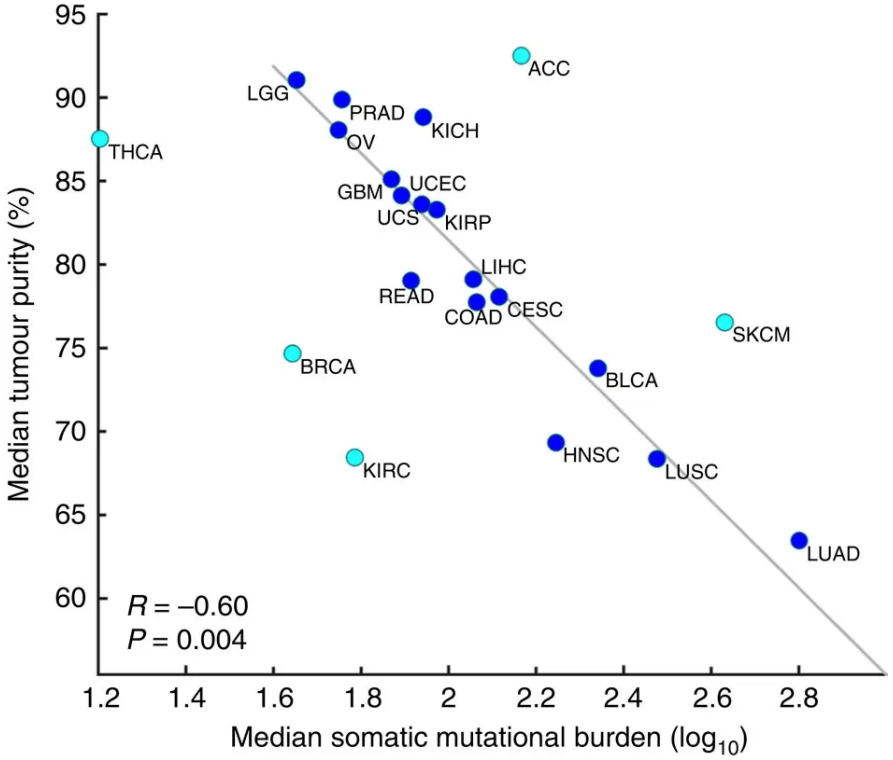


Figure 11. Tumor purity and mutational burden across 21 cancer types. Median of tumor purity and mutations per tumor sample illustrated in a scatter plot [modified from [124]]. ACC, adrenocortical; BLCA, bladder; BRCA, breast; CESC, cervical squamous cell and endocervical adenocarcinoma; COAD, colon adenocarcinoma; GBM, glioblastoma; HNSC, head and neck squamous cell; KICH, kidney chromophobe; KIRC, kidney renal cell; KIRP, kidney renal papillary cell; LGG, brain lower grade glioma; LIHC, liver hepatocellular; OV, ovary; PRAD, prostate adenocarcinoma; READ, rectum adenocarcinoma; SKCM, skin cutaneous melanoma; THCA, thyroid; UCEC, uterine corpus endometrial; UCS, uterine carcinosarcoma.

1.12 Chromosomal instability and aneuploidy

Chromosomal instability (CIN) represents a form of genomic instability that affects most human cancers [127]. It refers to an elevated missegregation rate of chromosomes during mitosis. Whole-genome duplication (WGD), or tetraploidization, can be caused in malignant cells by mitotic failure [128]. Failure rate that is too high to be repaired, results in CIN [129]. CIN assumes several forms including structural and numerical chromosomal changes over time in malignant tissues, which plays a major role in tumor initiation, development, and therapy [130–132].
It appears that tetraploid cells have a higher tolerance towards losses and gains of chromosomes, presuming that gene expression has a less significant effect compared to diploid cells [133,134]. Hypoploidy, genome sizes smaller than diploid, is less common and less reported, and this may be due to lethal effects on the cellular level, which cannot be compensated as in aneuploid tumors (Figure 12) [135,136].

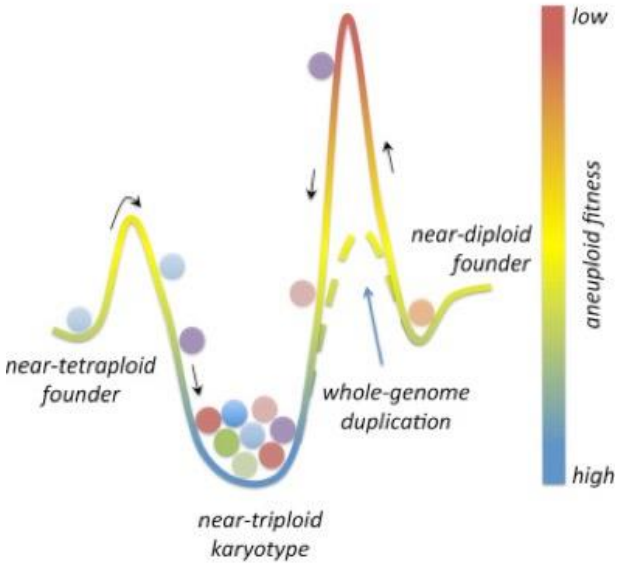


Figure 12. Tumor fitness with different genome sizes. Near-tetraploid founder cells after mitosis fall back to a near-triploid state due to chromosomal losses. Diploid tumors undergo WGD events, losing parts of their genome to backdrop into a near-triploid genome state. Both lead to a near-triploid karyotype with an increased clonal fitness [modified from [137]].

Aneuploidy is considered a hallmark of cancer and is defined by the presence of a globally abnormal number of chromosomes in a cell [138]. Healthy human cells contain 23 pairs of chromosomes (diploid), which is the normal state (euploid). Physiologically, there are exceptions such cells as hepatocytes and megakaryocytes, which may be polyploid,

possessing more than two sets of chromosomes. Historically, aneuploidy is defined as an abnormal state of DNA content in a nucleus while, in contrast, polyploidy is caused by balanced duplication of WGD. In general, the term aneuploidy means the loss or gain of a whole chromosome (w-CIN) or a structural aneuploidy, which implies a loss or amplification of parts of a chromosome arm (s-CIN) [129]. w-CIN can have an impact on the total number of chromosomes and lead to an unbalanced chromosome number, subsequently leading to the inescapable consequence of aneuploidy, which especially occurs in diploid tumors without any signs of aneuploidy [139]. Structural changes in chromosomes can alter the expression of genes or form oncogenes by gene fusions [140,141]. The first described translocation in cancer was the Philadelphia chromosome. Chromosome 9 and 22 are reciprocally translocated, which results in an oncogenic function. Overexpression of the fused *BCR-ABL1* gene fusion leads to uncontrolled proliferation [142,143]. Due to a changed number or structure of chromosomes, several genes are lost or highly amplified, which affects the cellular metabolism and leads to an increased mutation rate [144,145]. Hypothetically, this could result in a vicious circle, where CIN leads to aneuploidy and, due to changed expression, subsequent higher mutation rates drive cancer progression. Loss of tumor suppressor gene *CDKN2A* is a typical example mutated or lost in several cancers, including lung cancer. It encodes for the $p16^{INK4a}$ protein, which plays an important role in cell cycle regulation and is linked to tumor initiation [146]. On the other hand, oncogenes get amplified and overexpress proteins, as *Ras* that accelerates oncogenic signaling to malignancy [147]. Cancer genomes tend towards minimizing chromosomal regions with a higher number of tumor suppressor genes and maximizing oncogene-rich chromosomal regions leading to a near-triploid state [148,149].

All different kinds of aberrations can help cells to rapidly adapt their genome to different environments and confer enhanced fitness [150,151]. Mutations causing CIN can initiate colorectal cancer, whereas aneuploidy in the X chromosome is associated with breast cancer tumorigenesis and development [152,153]. Several clinical trials attempted to use CIN as a therapeutic approach whilst inducing CIN to a toxic level [154]. However, this also involves the danger of elevating the CIN level of low CIN tumors to an intermediate state, which worsens patient outcomes [154]. Further consequences of CIN include an increased rate of LOH events and the abnormality of the telomerase gene *TERT* [155,156]. Paradoxically, telomerase is responsible for protecting the caps at the end of the chromosomes from CIN and prolongs the cancer lifetime ad infinitum [157]. However, CIN seems to be a double-edged sword. It is generally associated with a poor outcome depending on the grade of genomic instability [158–160]. Very low and very high grades of CIN appear to improve the patient outcome. Recent results illustrated that CIN induces intratumoral heterogeneity (ITH) in NSCLC that is associated with a higher risk of recurrence and death [161,162].

CIN and aneuploidy occur in many tumors as major factors in cancer development. The following facilitates survival and adaptation to stressful conditions to the fittest cell following Darwinian selection (**Figure 13**) [163,164].

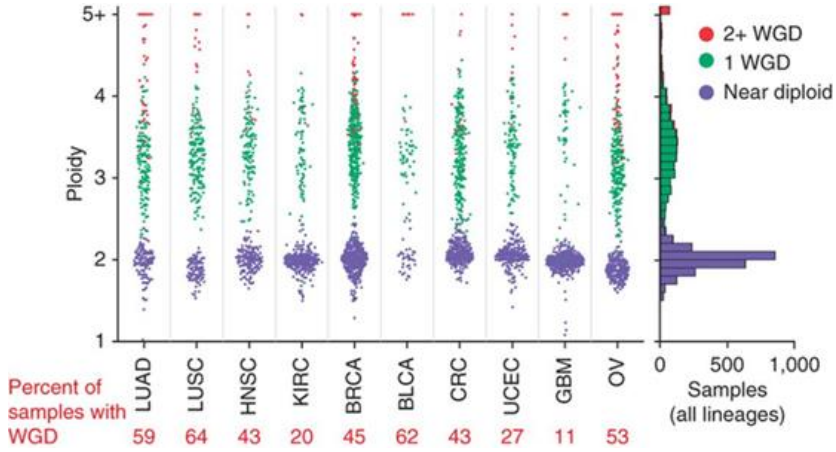


Figure 13. Aneuploidy across 26 cancer types. The ploidy differs depending if the tumor duplicated its genome. LUAD and LUSC have diploid and aneuploid tumors [modified from [165]]. WGD, whole genome duplication; HNSC, head and neck squamous cell; KIRC, kidney renal cell; BRCA, breast; BLCA, bladder; CRC, colorectal; UCEC, uterine cervix; GBM, glioblastoma; OV, ovary.

1.13 Intratumoral heterogeneity and tumor evolution

Already in the late 1800s, the pathologist Rudolf Virchow reported morphological differences between tumor cells on a microscopic level. Around a century later, 1976 Peter Nowell postulated that neoplasms arise from a single cell of origin acquiring stepwise genetic liabilities, which lays the foundation for tumor evolution and clonal diversity [166]. Non-malignant cells that convert to malignant cells through sequential acquisition of molecular alterations consequently lead to altered cell metabolism [138]. A shared ancestral mutation in all clones, is seen as a mutation in the trunk of a phylogenetic tree of tumor cells. Moreover, a subclonal altering event describes a descendant that follows a new lineage and not the trunk. Mutations that are not shared between tumors and occur only in one lineage (subclone or tumor) are called private mutations.

After malignant transformation, cancer cells continue to evolve in the bulk tumor to cells distinguishing each other through unique molecular properties, which results in heterogeneity in the tumor itself.

ITH is a continuous process on a genomic, transcriptomic and epigenomic level [161,167,168]. Different from intertumoral heterogeneity, referring to heterogeneity between tumors of the same histological type in patients, ITH depicts the genetically different subclones in one tumor. Spatial heterogeneity can be a type of ITH, where subclones leave the primary bulk tumor, metastasize to a distant organ, and develop a metastasis. Temporal heterogeneity describes the evolutionary dynamics in subpopulations in a tumor over time. Harsh environmental circumstances as cytotoxic therapy can diminish the tumor cell counts completely, except for one surviving subpopulation that is able to adapt to the selective pressure and develops resistance [169–171]. From an evolutionary perspective, there are several models. Among them, a linear and a branched evolution are the most common (Figure 14).

Linear evolution

Genomic instability leads to a stepwise accumulation of mutational processes during which cancer cells outcompete the non-malignant cells by a selective growth advantage. Due to the sequential acquisition of genetic alterations, the clone with the dominant mutation outcompetes other clones, which is termed as selective sweep. Damaging mutations with an impact on cancer development, progression and resistance are called drivers [172]. Mutations that do not contribute equally in a harmful manner are called passengers. The newly gained driver mutation increases fitness and expands clonally. The tumor depicts the linear evolution, as it is dominated by one superior clone that outcompetes the ancestral clones (Figure 14a).

Branched evolution

Branched evolution allows the co-existence of several subclones with increased fitness sharing one common ancestor. Gradually over time, new driver mutations resulting in a higher degree of heterogeneity allowing various subclones to adapt to new circumstances [161,173,174]. In this model of evolution, the amount of ITH is changing through the lifetime of the tumor. Nevertheless, colon cancer progresses stepwise and haematological cancers rather take the linear evolution, whereas most other solid cancers in breast, liver, ovarian, prostate, kidney, brain, lung, and melanoma are found more in a branched evolution (Figure 14b) [173,175–182].

Intermediate levels of ITH can have a poor clinical outcome [158]. Equally, a higher level of genomic instability can lead to unviable karyotypes and can be used for cancer treatments that increase CIN such as paclitaxel or radiotherapy, which has a better effect in tumors with higher CIN [183–185].

Even if several models exist, the truth is hard to reveal. During the course of the disease, patients cannot be biopsied several times to follow up the progression. A biopsy represents only one single snapshot. Most likely tumors undergo different evolutionary moments

depending on the stage of the development. Therefore, it is conceivable that the tumorigenesis is caused in a linear way by sequential accumulation of mutations, while the progression or resistance development is occurring rather in a subclonal lineage of a branched evolution.

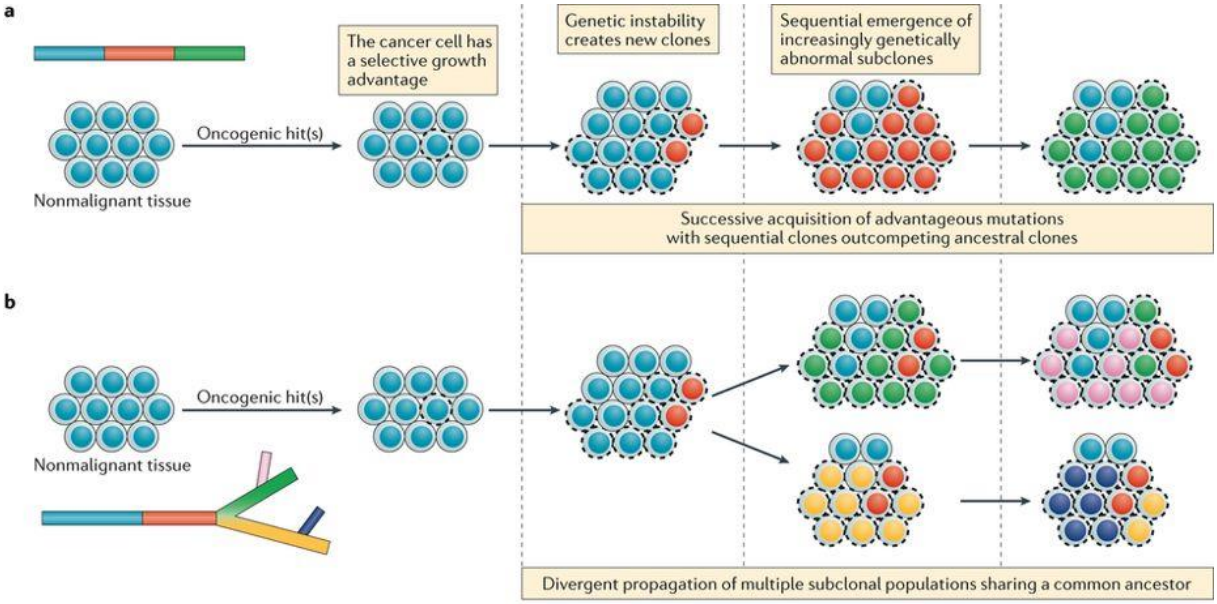


Figure 14. Differences in linear and branched tumor evolution models. The linear evolution model sketches cells accumulating genetic alterations that gain a fitness advantage (red, followed by green) compared to the other clones (**a**). The branched evolution model outlines common ancestor clones harbouring mutations that propagate into multiple genetically divergent subclonal populations (**b**) [modified from [130]].

1.14 Metastasis

Metastases are descendants from a primary malignant lesion. More than 90% of cancer-related deaths are attributed to metastases [186,187]. The emergence of metastases is a sequential process defined as “invasion-metastasis cascade”. This multistage process includes: (1) local invasion into the surrounding stroma, (2) stimulation to form new blood vessels in their local microenvironment, (3) tumor cells must survive the damaging transport caused by the immune system through the vasculature, (4) arresting at a distant organ site, (5) extravasation into the microenvironment of the distant organ site, (6) adapt to the new microenvironment and form micrometastasis and finally (7) metastatic colonization accompanied by an accumulation of genetic and epigenetic alterations [188].

After invasion into the vasculature, the tumor has access to the blood and lymphatic transport systems in the body. Circulating tumor cells (CTCs) detach from the primary as individual cells or as multi-cellular clumps [189,190]. In the bloodstream, CTCs expose themselves in a very vulnerable position, where the immune system can interact and fight against them [191]. Crucial steps are required as persistent surviving and adapting to a new environment to reach a distant organ before the metastasis can manifest. Therefore, it is not surprising that the rate of attrition is around 99.98% for metastatic colonization [192,193]. Tumor cells that leave the primary tumor to travel to a distant organ and manifest there choose a very inefficient process.

Beyond the macroscopic metastatic colonization, there is still the continuing question of how the metastatic clone became competent. The evolution of tumor cells leads to two current models of metastatic diseases, linear progression model, and parallel progression model.

Linear progression model

Primary tumors and their metastases are clonally related and originate from the same common ancestor cell (Figure 15a). The linear model depicts the idea of a dominant clone emerging late in the tumor development and disseminates prior to clinical diagnosis [192,194,195]. Genetic divergence in this model is expected to be low, for the sake of the seeding from the latest clone or subclone that evolved in the primary before dissemination.

Parallel progression model

The parallel progression model assumes early dissemination (Figure 15b). The metastatic clone or subclone evolves in parallel with the primary tumor after separation of the primary tumor. Thus, leads to a larger genetic divergence. The dissemination of metastatic clones can occur in multiple waves to several locations at the same time. Further, metastases can give rise to another metastatic lesion in the same patient [192,194].

In reality, the evolutionary progress of metastasis is much more complex. The time of dissemination cannot always be categorized as early and late events. Intermediate dissemination represents a state of genetic divergence between the two states of linear and parallel progression [194].

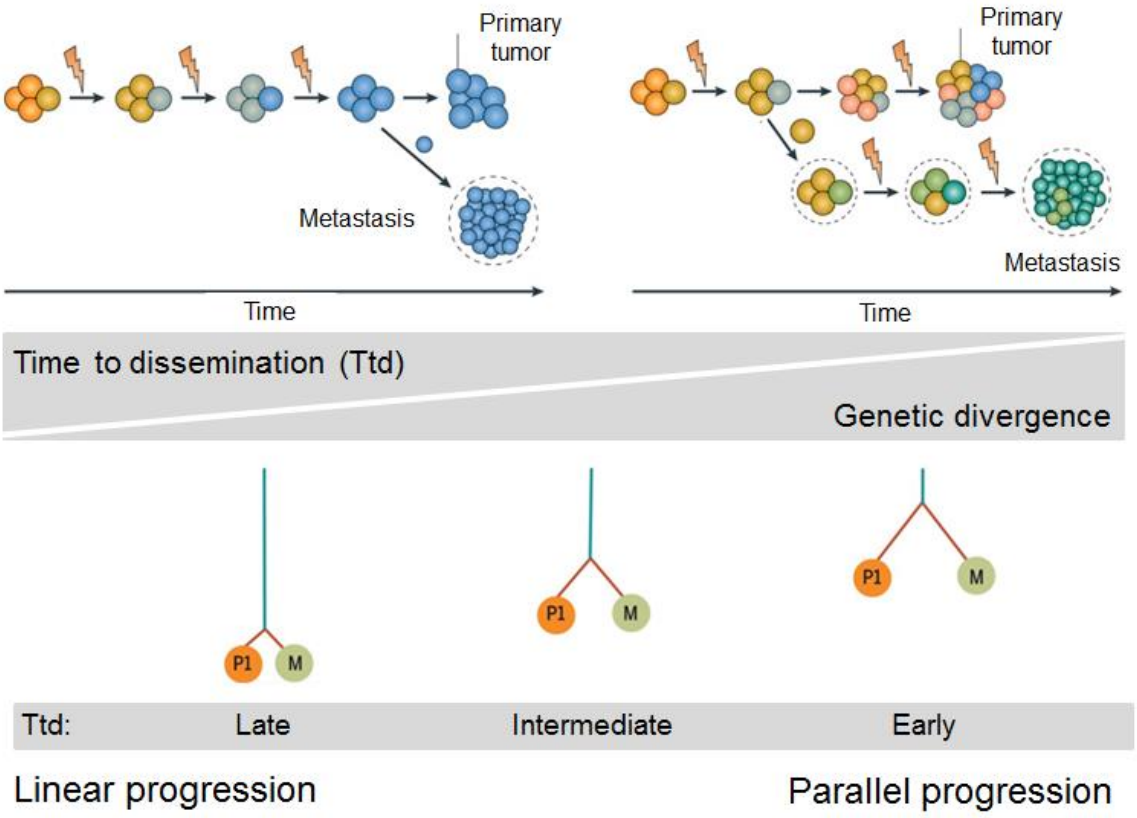


Figure 15. Model of metastatic dissemination. (**a)** Linear progression model occurs while the metastasis disseminates to a late point of time and is genetically close related to each other. (**b**) Early dissemination leads to a larger genetic divergence [modified from [194,196]].

2. Results

2.1 Part A: Exploring the spatiotemporal genetic heterogeneity in metastatic lung adenocarcinoma using a nuclei flow-sorting approach

Exploring the spatiotemporal genetic heterogeneity in metastatic lung adenocarcinoma using a nuclei flow-sorting approach

Thomas Lorber, Noemi Andor, Tanja Dietsche, Valeria Perrina, Darius Juskevicius, Karen Pereira,
Stephanie U Greer, **Arthur Krause**, David C Müller, Spasenija Savic Prince, Didier Lardinois, Michael T Barrett, Christian Ruiz andLukas Bubendorf

Journal of Pathology, 2019, 247: 199–213

For readers clarity Supplementary Materials and Methods and Supplementary Figures are attached. For further supplementary information please visit **DOI:** 10.1002/path.5183

Journal of Pathology
J Pathol 2019; **247:** 199–213
Published online 28 December 2018 in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/path.5183

ORIGINAL PAPER

Exploring the spatiotemporal genetic heterogeneity in metastatic lung adenocarcinoma using a nuclei flow-sorting approach

Thomas Lorber[1]*, Noemi Andor[2], Tanja Dietsche[1], Valeria Perrina[1], Darius Juskevicius[1], Karen Pereira[2],
Stephanie U Greer[2], Arthur Krause[1], David C Müller[1], Spasenija Savic Prince[1], Didier Lardinois[3], Michael T Barrett[4], Christian Ruiz[1†] andLukas Bubendorf[1†]

[1] *Institute for Pathology, University Hospital Basel, University of Basel, Basel, Switzerland*
[2] *Division of Oncology, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA*
[3] *Thoracic Surgery, University Hospital Basel, Basel, Switzerland*
[4] *Division of Hematology and Molecular Oncology, Mayo Clinic, Scottsdale, AZ, USA*

**Correspondence to: T Lorber, PhD, Institute for Pathology, University Hospital Basel, University of Basel, Schoenbeinstrasse 40, 4031 Basel, Switzerland. E-mail: thomas.lorber@usb.ch*

[†]These authors contributed equally to this work.

Abstract

Variable tumor cellularity can limit sensitivity and precision in comparative genomics because differences in tumor content can result in misclassifying truncal mutations as region-specific private mutations in stroma-rich regions, especially when studying tissue specimens of mediocre tumor cellularity such as lung adenocarcinomas (LUADs). To address this issue, we refined a nuclei flow-sorting approach by sorting nuclei based on ploidy and the LUAD lineage marker thyroid transcription factor 1 and applied this method to investigate genome-wide somatic copy number aberrations (SCNAs) and mutations of 409 cancer genes in 39 tumor populations obtained from 16 primary tumors and 21 matched metastases. This approach increased the mean tumor purity from 54% (range 7–89%) of unsorted material to 92% (range 79–99%) after sorting. Despite this rise in tumor purity, we detected limited genetic heterogeneity between primary tumors and their metastases. In fact, 88% of SCNAs and 80% of mutations were propagated from primary tumors to metastases and low allele frequency mutations accounted for much of the mutational heterogeneity. Even though the presence of SCNAs indicated a history of chromosomal instability
(CIN) in all tumors, metastases did not have more SCNAs than primary tumors. Moreover, tumors with biallelic TP53 or ATM mutations had high numbers of SCNAs, yet they were associated with a low interlesional genetic heterogeneity. The results of our study thus provide evidence that most macroevolutionary events occur in primary tumors before metastatic dissemination and advocate for a limited degree of CIN over time and space in this cohort of LUADs. Sampling of primary tumors thus may suffice to detect most mutations and SCNAs. In addition, metastases but not primary tumors had seeded additional metastases in three of four patients; this provides a genomic rational for surgical treatment of such oligometastatic LUADs.

Keywords: lung cancer; flow sorting; intratumoral heterogeneity; comparative genomics; metastatic tumor evolution

Received 28 June 2018; Revised 12 September 2018; Accepted 12 October 2018 *No conflicts of interest were declared.*

Introduction

Metastatic disease remains the main cause of cancer-related death, and genetic intratumoral heterogeneity (ITH) and chromosomal instability (CIN) are thought to contribute to the evolution of the metastatic cascade. While substantial ITH can challenge oncology decision-making based on single biopsies [1], studies in localized lung adenocarcinoma (LUAD) have suggested that single-region sampling may be sufficient to identify the majority of known cancer gene mutations, as carcinogen-induced tumors, such as lung cancers, are thought to have acquired most mutations before the onset of tumorigenesis [2,3].

Multi-region sampling studies have applied a wide spectrum of genomic analyses, including whole exome and/or whole genome sequencing, but they were restricted to primary LUAD tumors [3–5]. Larger studies have helped to investigate the mutational landscape of LUAD but have been limited to cohorts of unpaired primary and metastatic tumors [6,7]. LUAD studies with matched primary and corresponding metastases were restricted to a small number of patients [8,9], partly because tumors with a low tumor content were excluded [10], or used a low sequencing depth [11]. However, shallow sequencing can limit the sensitivity of genomic analyses, especially in tumors with mediocre tumor cellularity, such as LUAD with a median purity of 44% in the TCGA dataset [12,13]. Hence, further studies are needed to investigate the genetic ITH between primary LUADs and matched, i.e. clonally related, metastases.

In this study, we aimed to apply a technology that overcomes the limitation of low tumor cellularity to investigate the level of heterogeneity of somatic copy number aberrations (SCNAs) and cancer gene mutations between primary LUAD and matched metastases over time and space. We applied a refined nuclei flow-sorting approach followed by genome-wide SCNA analysis with array comparative genomic hybridization (aCGH) and deep sequencing of 409 cancer genes to a total of 37 longitudinally sampled tumors from 16 patients with metastatic LUAD.

Materials and methods

Patients and tumor samples

This retrospective study was approved by the local Ethics Committee (EKBB/EKNZ 31/12). Patients were included based on the following criteria: (1) histologically diagnosed LUAD; (2) two or more fresh-frozen tumor samples available, with one of them being the primary tumor; and (3) tumor samples differed in time point or site (supplementary material, Figure S1). Patients with clonally unrelated tumors were excluded (supplementary material, Table S1 and Figure S2). The final cohort consisted of 16 patients with primary LUAD and 21 matched metastases (Table 1 and supplementary material, Figure S3). Fresh-frozen (n=35) and formalin-fixed, paraffin-embedded (FFPE, n=2) tumor samples were obtained from the Institute for Pathology of the University Hospital Basel, Switzerland. Histology was confirmed by two experienced lung pathologists (LB, SS).

Nuclei isolation and multiparameter flow-sorting

Nuclei were isolated from fresh-frozen and FFPE tumor samples according to published protocols [14–16]. Nuclei were stained with DAPI for DNA quantification and ploidy analysis. The lineage marker thyroid transcription factor 1 (TTF-1), which is expressed in 70–80% of all LUADs [17], was used as a second parameter in TTF-1-positive tumors to discriminate tumor nuclei from stromal components. TTF-1 expression was ascertained by

immunohistochemistry (IHC) on corresponding FFPE tissue sections. In this cohort, ten of the 16 patients had TTF-1-positive tumors. Details regarding isolation of nuclei and the staining protocol are provided in the supplementary material, Supplementary materials and methods.

DNA extraction and array comparative genomic hybridization

DNA extraction, quantification, and whole genome amplification were carried out using standardized protocols, and aCGH was as previously described [14] (see supplementary material, Supplementary materials and methods).

Next-generation sequencing

We performed targeted, deep sequencing (mean coverage of 965X) of all sorted populations with the Ion Ampliseq™ Comprehensive Cancer Panel (CCP) targeting the exons of 409 oncogenes and tumor suppressor genes (TSGs) (supplementary material, Table S2). Whole exome sequencing (WES) with the Ion Ampliseq™ Exome RDY Kit (mean coverage 85X; range 68–100X) was performed only for patient 42, because of the low number of somatic mutations (n=2) in genes detected by the CCP and the complex situation revealed by aCGH in this patient's primary tumor. Mutations detected by WES were validated by ultra-deep resequencing (mean coverage 5864X; range 5203–6118X) and technical replications (re-sorting and ultra-deep resequencing) with a 400bp Ion Ampliseq™ custom panel (supplementary material, Table S3). DNA of 'diploid, CNA-neutral' populations was used as germline controls (supplementary material, Table S4). Details regarding controls, sequencing, alignment, variant calling, filtering, and annotation are provided in the supplementary material, Supplementary materials and methods.

Categorization of mutations and copy number aberrations

SCNAs and mutations were classified as 'ubiquitous' if present in all the tumor biopsies of each patient; as 'shared' if present in more than one but not all of the tumor biopsies of the four patients (patients 20, 22, 31, and 36) with oligometastatic disease, or otherwise as 'private'. Non-synonymous somatic mutations and splice-site or regulatory mutations (generally named non-silent) were discriminated from synonymous and intronic mutations (referred to as silent). Details about the functional effect prediction and interlesional genetic heterogeneity analyses are provided in the supplementary material, Supplementary materials and methods.

Results

Nuclei flow-sorting results in tumor DNA of high purity

We investigated the spatiotemporal genetic ITH in 16 patients with primary LUAD and their clonally related metastases. The detailed composition of the cohort, including clinical annotation, is summarized in Table 1 and the supplementary material, Figure S3. We subjected all 37 tumor specimens to nuclei flow-sorting, either using DAPI and TTF-1 for TTF-1-positive tumors (n=25 tumors, ten patients) or DAPI alone for TTF-1-negative tumors (n=12 tumors, six patients).

.

Table 1. Clinical characteristics of 16 patients with clonally related metastatic lung adenocarcinomas

Patient	Sex	Age at diagnosis (years)	Smoker at diagnosis	Smoking status (pack-years)	Site of primary tumor	Neoadjuvant RTX	Neoadjuvant CTX	Adjuvant RTX	Adjuvant CTX	TimePRI-MET (months)	Metastatic site(s)	Ploidy (*N*) PRI	Ploidy (*N*) MET
12	M	58	NA	NA	LLL	NA	NA	NA	NA	2	Thyroid gland	3.3	3.2
20	F	40	Active	25	LUL	No	No	No	No	41	Kidney, left	2.6	2.6
										41	Kidney, right		2.6
22	F	59	NA	NA	LUL	No	No	Yes	No	7	Brain, right (parietal)	3.4	3.4
										11	Locoregional: LLL		3.5
31	F	64	NA	NA	RUL	No	No	No	No	64	Locoregional: LLL	3.0	3.1
										64	Locoregional: RLL		3.3
34	M	60	Former	45	RUL	No	No	No	No	27	Locoregional: RLL	2.0	2.0
35	F	52	NA	NA	RLL	No	No	No	Cisplatin, vinorelbine	10	Brain, left (frontal)	3.3	3.1
36	M	51	NA	NA	LUL	No	No	No	Cisplatin, vinorelbine	11	Locoregional: RLL	2.1	2.0
										28	Locoregional: LLL		2.0
										28	Locoregional: RLL		2.1
38	M	66	Former	45	LUL	Yes	Cisplatin, docetaxel, cetuximab	No	No	14	Adrenal gland, left	3.2	3.1
39	M	78	Former	30	RLL	No	No	No	No	9	Locoregional: RUL	3.2	3.2
41	F	64	Active	40	LUL	No	No	No	No	0	Brain, left (parietal)	2.9	3.0
42	F	81	Former	1–2	RLL	No	No	No	No	7	Pleura, right	2.0 3.4 4.7	3.8
44	M	51	Active	35	LUL	No	Cisplatin, pemetrexed	No	No	1	Adrenal gland, left	5.4	5.0
46	F	74	NA	NA	RUL	No	Carboplatin, docetaxel	no	Cisplatin, pemetrexed	44	Pleura, right	3.2	2.8
48	M	67	Active	50	LUL	No	No	No	No	0	Adrenal gland, left	4.1	4.7
49	M	76	Former	20	RUL	No	No	No	No	152	Pleura, left	2.0	2.0
50	M	57	Never	0	RUL	No	Cisplatin, docetaxel	Yes	No	22	Brain, left (parietal)	3.2	3.0

CTX, chemotherapy; F, female; LLL, left lower lobe; LUL, left upper lobe; M, male; MET, metastasis; NA, unknown; PRI, primary tumor; RLL, right lower lobe; RTX, radiotherapy; RUL, right upper lobe.

This allowed the genomic characterization of biopsies with low tumor cell proportion and of diploid tumors as diploid tumor cells cannot be separated from diploid non-tumor cells by DNA content-based flow sorting alone (supplementary material, Figure S4). Sorted populations were subsequently profiled for the presence of copy number aberrations and mutations by genome-wide aCGH and deep sequencing (mean sequencing coverage 965X) of 409 cancer genes, respectively (supplementary material, Figure S5).

All tumors but one were monogenomic; i.e. they consisted of tumor cells of one single ploidy. Only the primary tumor of patient 42 was polygenomic, as evidenced by the presence of three distinct tumor populations (discussed later). This resulted in a total of 39 tumor populations to be investigated. TTF-1-positive tumor populations with diploid DNA content ($2N \pm 0.2$) were detected in four patients. The tumors of all other patients were aneuploid ($> 2.2N$). Our flow-sorting approach increased the mean tumor purity from 54% (range 7–89%) of unsorted material to 92% (range 79–99%) after sorting (supplementary material, Figure S6).

Recurrent mutations and copy number aberrations

Targeted sequencing of the 39 sorted tumor populations identified 222 non-silent mutations affecting 125 genes, of which 72% were predicted to be pathogenic. An overview of recurrent non-silent mutations and SCNAs is displayed in Figure 1. A summary of all mutations is available in the supplementary material, Tables S5 and S6 and Figure S7. The most common ubiquitous mutations, i.e. mutations present across all lesions, were found in genes previously known to be mutated in LUAD [6,19]. These included *KRAS* (six patients; 37.5%); *TP53* (5; 31.3%); *ATM*, *CSMD3*, *EGFR*, *NOTCH4*, and *STK11* (4; 31.3% each); and *EPHA7*, *GNAS*, *KEAP1*, *NF1*, *SYNE1*, and *TLR4* (3; 18.8% each). As expected, mutations in TSGs often occurred in the context of loss of heterozygosity (LOH). In fact, 40 of 62 (64.5%) mutations in the TSGs *ATM*, *KEAP1*, *NF1*, *PTPRD*, *STK11*, and *TP53* were accompanied by loss of the wild-type allele. Moreover, ubiquitous mutations in *KEAP1*, *STK11*, and *TP53* were always homozygous in nature.

Overall, we detected 2774 SCNAs in the 39 sorted tumor populations (supplementary material, Table S7, Figure S8, and Appendix S1). The median number of SCNAs per tumor was 73 (range 14–143). An average of 37% of the events were focal (< 3 Mb); 24% affected whole chromosome arms and 39% were larger than 3 Mb but smaller than a whole chromosome arm (supplementary material, Figure S9). Recurrent copy number gains were in line with previous data [6,20] and involved the chromosomal regions 1q, 5p, 7p, and 8q, as well as both arms of chr12 and chr14 (Figure 2), and mapped to genes known to be amplified in LUAD, such as *ARNT* (1q21), *TERT* (5p15), *EGFR* (7p11), *MYC* (8q24), *KRAS* (12p12), *NKX2-1*, and *FOXA1* (both 14q13). The only recurrently deleted region was 9p21, with both homozygous (seven patients; 43.8%) and heterozygous deletions (4; 25%) of *CDKN2A*. In addition to the genes located at 9p21, other known tumor suppressor genes [21–24] were found to have homozygous deletions: *RB1* (13q14.2), *PTEN* (10q23.31), *DCN* (12q21.31), *THSD4* (15q23), *HS3ST4* (16p12.1), and *WWOX* (16q23.1).

Limited interlesional heterogeneity of ploidy, copy number aberrations, and mutations

Except for hypertetraploid ($> 4N$), ploidy was stable during metastastic progression, regardless of the time interval (range 0–152months) between primary and metastatictumorresection(Figure 3A).Remarkably,the contralateral pleural effusion of patient 49 remained diploid, even though it had appeared 12.7 years after the detection of the diploid primary tumor. Overall, aneuploidy was associated with a short time to metastasis (p=0.048, univariate Cox analysis) and correlated with the total number of SCNAs (Pearson r=0.77, p<0.001; Figure 3B). Moreover, tumors with biallelic inactivation (mutation plus LOH) of *TP53* or *ATM* had a higher ploidy (p=0.006, Wilcoxon signed-rank

test) and a higher burden of SCNAs (p<0.001, Wilcoxon signed-rank test) compared with the rest of the cohort (supplementary material, Figure S10).

Importantly, metastases did not have more total or private SCNAs than primary tumors (Figure 3C and supplementary material, Figure S11A), even when divided by their *TP53/ATM* status (supplementary material, Figure S11B,C). In fact, 77% (range 47–100%) of all SCNAs were shared between primary tumors and metastases (Figure 3D). Moreover, 88% (range 68–100%) of SCNAs were passed from the primary tumors to the metastases. Likewise, 84% (range 58–100%) of SCNAs in the metastases were already present in the primary tumors. High concordance rates were also found for mutations. In total, 80% of mutations (range 50–100%) were propagated from the primary tumors to the metastases. Likewise, 80% (range 32–99%) of mutations in the metastases were already present in the primary tumors.

Regarding metastatic location, there was no difference in the ploidy or amount of private SCNAs between distant and locoregional metastases (supplementary material, Figure S12). Yet distant metastases had overall more SCNAs (p=0.02, Wilcoxon signed-rank test) than locoregional metastases, most probably because they were more often affected by biallelic inactivation of *TP53* or *ATM* (distant: 7/10 versus locoregional: 2/11; p=0.03, Fisher's exact test). In fact, only focal events (< 3 Mb) occurred more often in distant than in locoregional metastases (p=0.01, Wilcoxon signed-rank test; supplementary material, Figure S13). In line with these observations, all three diploid tumors were *TP53*/*ATM*-wild type, had the lowest number of SCNAs, and remained locoregional during metastatic progression.

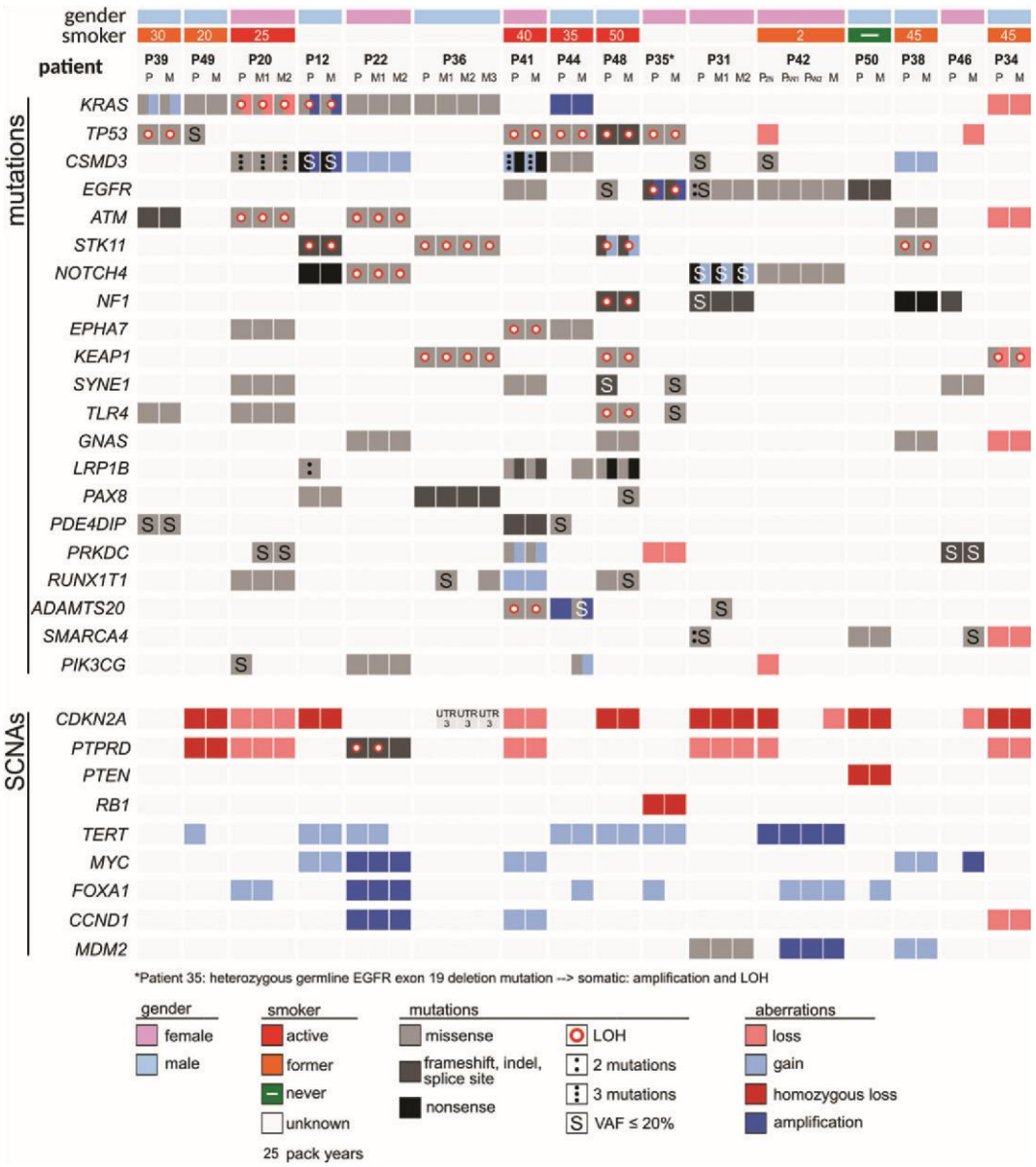


Figure 1. Overview of the genes recurrently affected by mutations and SCNAs. Only genes that are recurrently mutated, i.e. non-silent mutations in three or more patients are displayed. Overall, a high proportion of mutations and SCNAs are ubiquitously affected in all samples of individual patients. Importantly, of all 21 recurrently mutated genes, only mutations in *KRAS*, *ATM*, *NOTCH4*, *STK11*, *EPHA7*, *GNAS*, and *KEAP1* were always ubiquitous. In fact, *KEAP1* was ubiquitously mutated in three patients and always associated with LOH, which indicates a TSG function. In this study, two *EGFR* mutations and one *TP53* mutation were not propagated from primary tumors to metastases. The *TP53* mutation (P151S) occurred in the primary tumor of patient 49 with a VAF of 0.14 and was predicted to be pathogenic. Likewise, two *EGFR* mutations were private to primary tumors and not detected in the matched metastases; one occurred in exon 21 (H850Y) with a VAF of 0.04 in patient 31, co-existing with the common *EGFR* driver mutation L858R (VAF = 0.62). Double mutations of *EGFR* H850Y with other *EGFR* exon 21 mutations have been recently reported [18], but their effect on EGFR protein function is unknown. The second private *EGFR* mutation was found in patient 48. This mutation (D379N, VAF = 0.06) occurred in exon 10 and is of unknown significance. Known oncogenic driver mutations were found in all patients and included recurrent mutations of *KRAS*, *TP53*, *EGFR*, *ATM*, *STK11*, and *KEAP1*. The *MET* exon 14 splice site driver mutation in patient 46 was the only *MET* mutation in this cohort and is therefore not displayed in this figure of recurrently mutated genes. P: primary tumor; M1–3: first, second, and third metastasis. P_{2N}, P_{AN1}, P_{AN2}: diploid (2.0*N*) and two aneuploidy tumor populations (3.4*N*, 4.7*N*) in the primary tumor of patient 42. UTR: untranslated region.

Variant allele frequency as a parameter for interlesional heterogeneity

The high purity of tumor DNA in our samples enabled us to explore the distribution of variant allele frequencies (VAFs) and their temporal changes during metastatic disease with an unprecedented resolution. Interestingly, the VAFs (Pearson r=0.93, p<0.001; Figure 3E and supplementary material, Figure S14) and the copy numbers (Pearson r=0.95, p<0.001; supplementary material, Figure S15) of ubiquitous mutations remained largely unchanged between primary tumors and metastases. This resulted in density distribution patterns of VAFs and copy numbers that indicate a truncal origin for ubiquitous mutations (Figure 3F). In addition, the VAFs of ubiquitous mutations were significantly higher than the VAFs of private mutations (supplementary material, Figures S16A and S17) and subclonal analysis with Pyclone [25] identified a significantly higher cellular prevalence for ubiquitous than for private mutations (supplementary material, Figure S16B). In fact, 92% of mutations in the primary tumors with VAF < 0.1, indicating subclonality, were not spread to metastatic sites. In contrast, 91% of mutations in the primary tumors with VAF ≥ 0.1 were also detected in the metastases. Consequently, mutational heterogeneity between tumor lesions was predominated by mutations with low VAFs.

Genetic divergence between primary tumors and metastases

Metastatic dissemination follows the two general models of 'linear progression' or 'parallel progression', which are based on the relative timing of emergence and genetic divergence between the primary tumor and its metastases [26]. By measuring the mean pairwise divergence (D) in primary/metastatic (P/M divergence) pairs, we detected a broad continuum (D=0.03–0.53) of these two progression models in our LUAD cohort (Figure 4). At the lower end of the scale was the linearly progressing tumor of patient 41. Here, 97% of all SCNAs and mutations were present in both the primary tumor and the brain metastasis, which indicated a late dissemination of the metastatic founder from the primary tumor. In contrast, the highest P/M divergence was detected in patient 49, where the primary and the metastatic tumor shared only 47% of all SCNAs and mutations. However, the contralateral pleural effusion did not accumulate more private alterations than its primary tumor, which is surprising given that the effusion had emerged 12.7 years after the detection of the primary tumor. Despite a

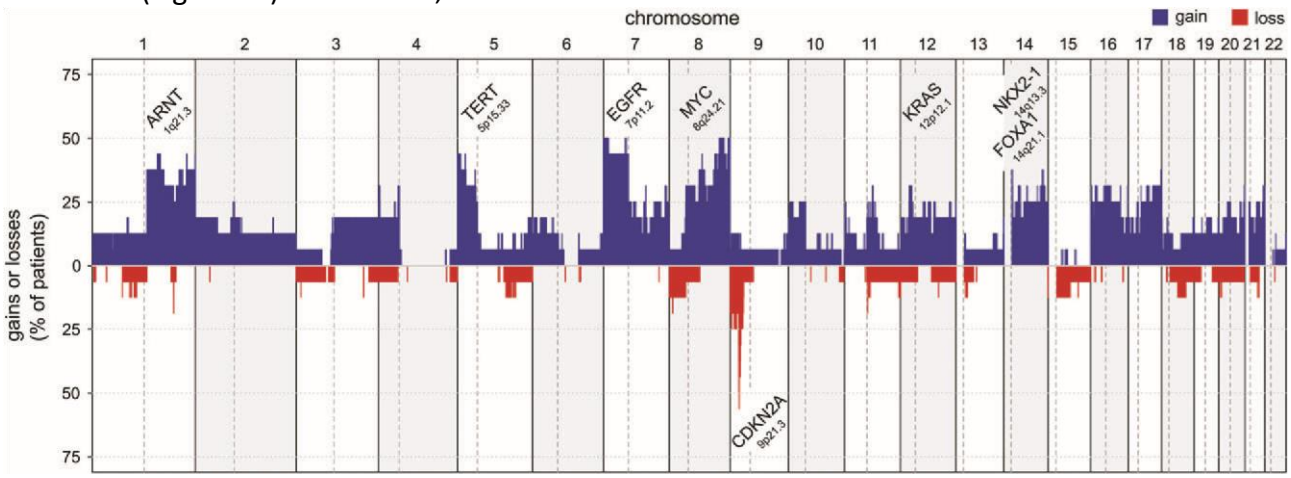


Figure 2. Somatic copy number aberrations. Overview of the recurrently amplified or deleted regions in the cohort. While six regions were amplified in ≥ 30% of the patients (1q, 5p, 7p, 8q, chr12, chr14), only one region (9p21) was recurrently (> 30% of patients) deleted in this cohort.

tendency for early appearance of metastases in tumors with low P/M divergence, we did not find a significant correlation between P/M divergence and the relative timing of metastatic tumor detection (Pearson r=0.3, p=0.28). However, we found that tumors with biallelic inactivation of *TP53* or *ATM* displayed a significantly lower P/M divergence (p=0.006, unpaired t-test) and that metastases emerged earlier in these patients (p=0.024, Wilcoxon signed-rank test) than in the rest of the cohort.

Two patterns of metastatic spread

Multiple, spatially separated metastases were available for four patients (patients 20, 22, 31, and 36), which allowed us to track tumor evolution beyond the first metastasis. We determined the clonal relationship by hierarchical clustering of Euclidean distances between the primary tumors and their metastases for each patient. The results are supportive of two patterns of metastatic evolution in our cohort (Figure 5 and supplementary material, Figure S18). First, the genomic characterization of the tumor of patient 22 suggested that two metastatic clones arose within the primary tumor, which eventually metastasized to the brain and to a different lobe in the ipsilateral lung, respectively ('polyclonal spread'). Interestingly, the intrapulmonary metastasis seemed to have disseminated earlier from the primary tumor than the brain metastasis, given that the accumulation of many private SCNAs by the intrapulmonary metastasis resulted in a higher genetic distance from the primary tumor than the genetic distance between the brain metastasis and the primary tumor. While *KRAS*, *ATM*, and *PTPRD* were mutated in all three tumor manifestations, biallelic inactivation of *PTPRD* was detected exclusively in the primary tumor and the brain metastasis but not in the intrapulmonary metastasis. Conversely, an *FLT4* mutation was shared between the primary tumor and the intrapulmonary metastasis. Additionally, the metastases had different SCNAs on chr4, but both were already present in the primary tumor (supplementary material, Appendix S1).

In contrast to patient 22, the metastases of patients 20, 31, and 36 clustered more closely to each other than to their individual primary tumors, which suggested a metastatic cascade in which the first metastasis continued to seed further metastases ('metastasis-to-metastasis spread'). This comprised contralateral spreads either intrapulmonary (patients 31 and 36) or from one kidney to the other (patient 20), with up to 5 years after the detection of the primary tumors. Known drivers (*KRAS*: patients 20 and 36; *EGFR*: patient 31) and most mutations and SCNAs were ubiquitously present and can therefore be considered early events in the evolution of these tumors. Notably, the first metastasis of patient 36 acquired a *CDKN2A* mutation that was preserved during further metastatic spreads. Likewise, the VAF of the *EGFR* L858R mutation in patient 31 was lower in both metastases than in the primary tumor ($VAF_{primary}$ =0.62, $VAF_{metastases}$ =0.40 and 0.33).

Intratumoral heterogeneity in single primary tumor

n contrast to all other monogenomic tumors in this cohort, our multi-parameter flow cytometry approach revealed that the primary tumor of patient 42 was polygenomic. It consisted of three tumor populations, one diploid (2.0N) and two aneuploid (3.4N, 4.7N) populations. All were TTF-1-positive (Figure 6A and supplementary material, Figure S19). The metastasis was monogenomic with a single aneuploid TTF-1-positive tumor population (3.7N). We confirmed the presence of these three distinct tumor populations in the primary tumor by performing IHC for the tumor suppressor p16 (Figure 6B). The homozygous deletion of *CDKN2A* (chr 9p21), which encodes p16, was exclusive to the 2.0N tumor population (Figure 6C, D). Consequently, the diploid tumor cells, but not the aneuploid cells, lacked expression of the p16 protein (Figure 6B).

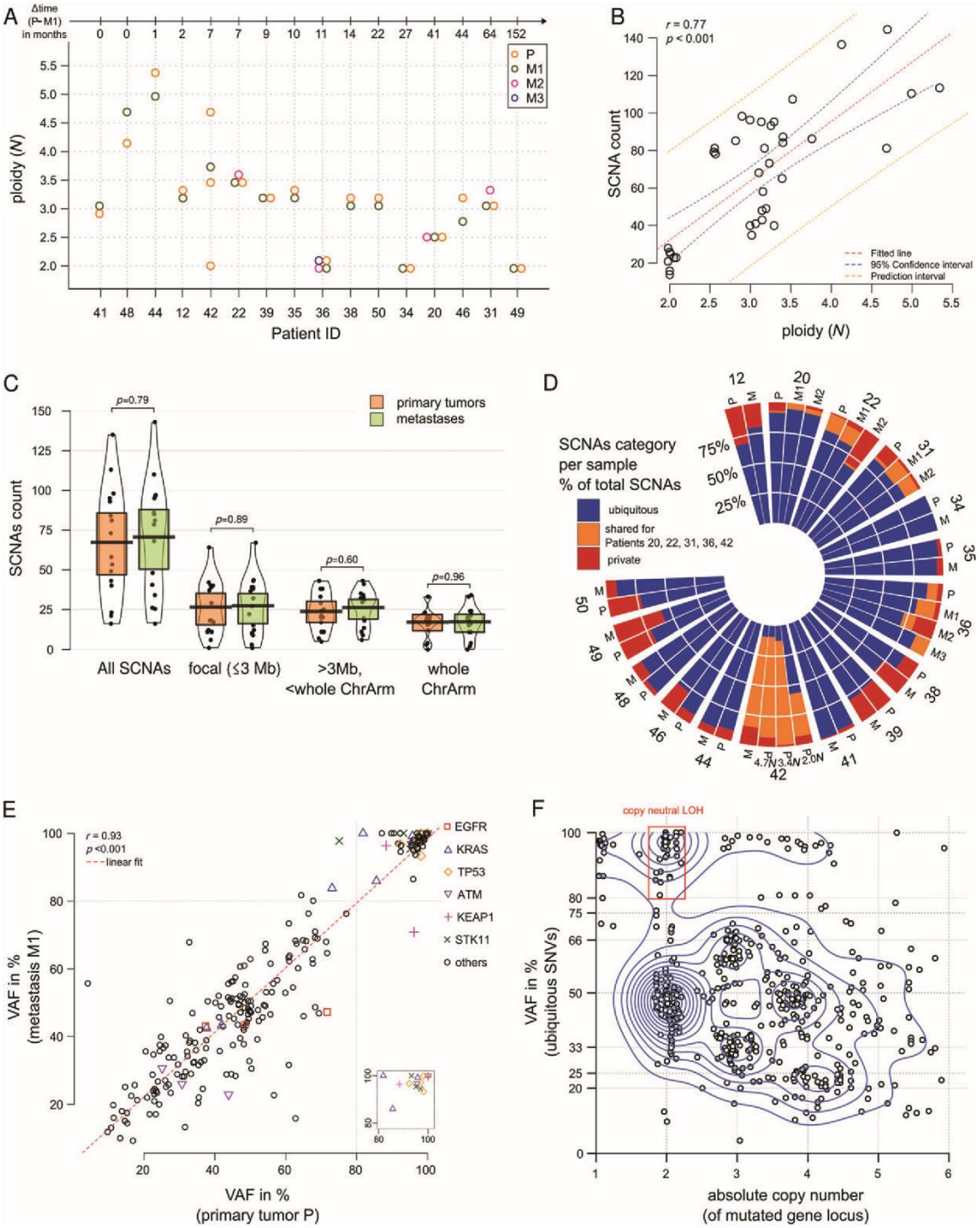


Figure 3. Comparison of the somatic copy number aberrations and mutations between primary tumors and metastases. (A) Ploidy of individual tumors did not change significantly, irrespective of the time interval between primary tumor detection and the emergence of the first metastasis (Δtime P → M1). (B) A high ploidy correlated significantly with a high burden of SCNAs. (C) Metastases did not have significantly higher amounts of SCNAs than primary tumors, despite the size of the aberration. (D) Circos plot displaying the relative amounts of SCNAs per sample that are ubiquitous (blue), shared (orange) or private (red). (E) VAF distribution of ubiquitous mutations in primary tumors (*x*-axis) and metastases (*y*-axis) showing a significant linear correlation. (F) Density plot of the VAFs (*y*-axis) and the absolute copy numbers (*x*-axis) of ubiquitous mutations, further suggesting that they are truncal. For instance, mutations with copy numbers of 3 clustered around VAFs of approximately 0.33 and 0.66, indicating the presence of one or two mutated alleles per cell, respectively. P: primary tumor; M1–3: first, second, and third metastasis.

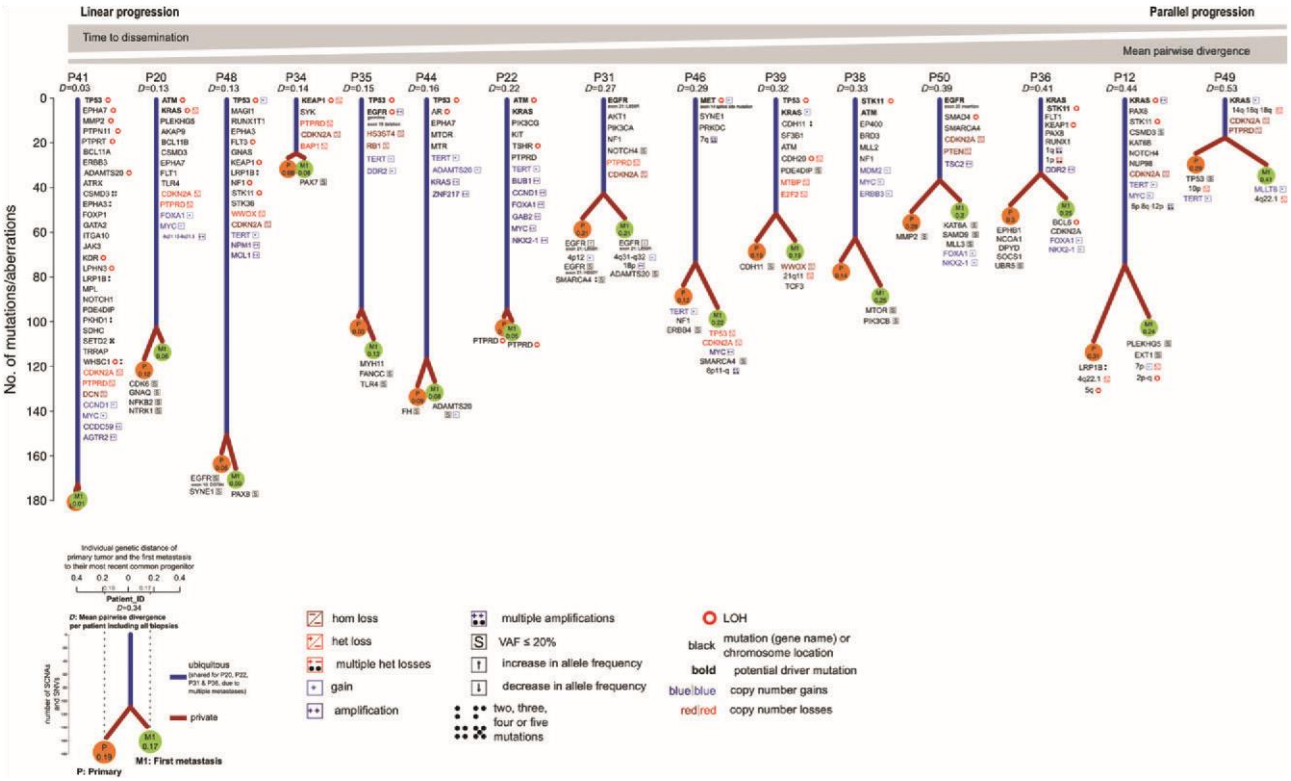


Figure 4. Genetic distances between primary tumors and metastases. A broad continuum of linear to parallel progression is present in this cohort of 16 LUADs. The genetic distances of primary tumors and metastases to their most recent common progenitor were comparable for all patients, suggesting that metastases did not accumulate many more mutational events compared with primary tumors, despite metastases having on average more time to develop. Genetic distances were calculated by the mean pairwise divergence (*D*) of all mutations and SCNAs between the primary tumor and the first biopsied metastasis. Selected SNVs and SCNAs are displayed next to the ubiquitous trunk (blue) or next to private leaves (red). The *y*-axis represents the cumulative amount of mutations and SCNAs detected in individual patient tumors. The distance of the primary tumor (P) or metastasis (M1) from the trunk (blue line) on the *x*-axis represents their individual genetic distance to their last common progenitor. Patient 42 is not shown here but is discussed in detail in Figure 6. Of note, the metastasis of patient 49 emerged 12.7 years after the primary tumor

Copy number analysis combined with WES and targeted resequencing revealed a high level of ITH. In total, only 7.6% (8/105) of all SCNAs and 27.1% (26/96) of all mutations were ubiquitous (Figure 6E). Among these, an *EGFR* exon 20 insertion mutation (p.N771delinsGY) that has been reported previously [27] (Figure 6F and supplementary material, Figure S20), an amplification of *TERT* on chr 5p, a gain of chr 7p covering the *EGFR* locus (Figure 6D), and a focal deletion of *THSD4* on chr 15q23 (supplementary material, Appendix S1) were present across all tumor populations. Interestingly, both alleles of *THSD4* were lost in the 3.4*N* population, whereas the 2.0*N* and the 4.7*N* populations in the primary tumor and the metastasis retained one copy of *THSD4*. In addition, high-level amplification of *MDM2*, an important negative regulator of the *TP53* tumor suppressor, was present in all aneuploid tumor populations but not in the diploid tumor population (Figure 6D).

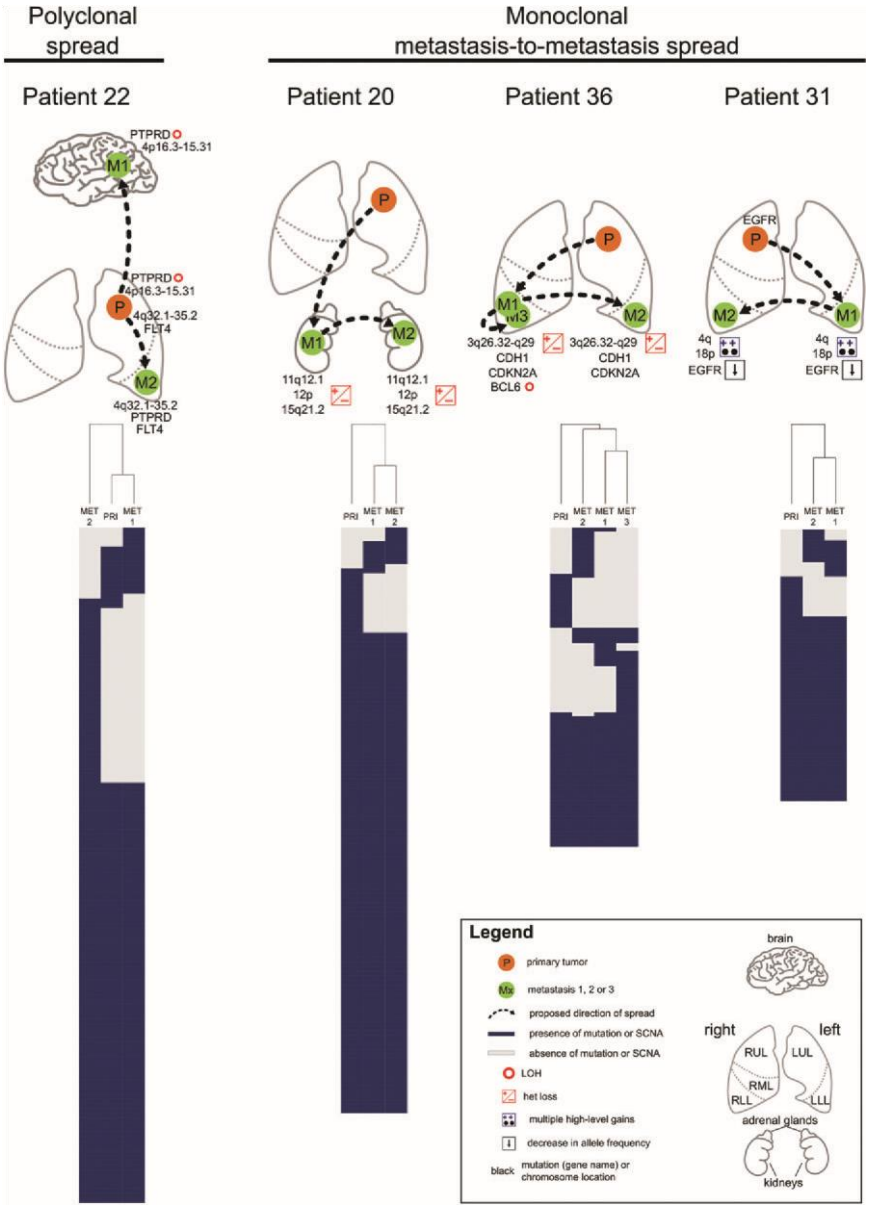


Figure 5. Two patterns of metastatic spread from patients with oligometastatic disease. Hierarchical clustering of SCNAs and mutations (including silent mutations) of sorted tumor populations of four patients with multiple tumor lesions (oligometastatic disease). While the primary tumor of patient 22 was more closely related to the brain metastasis, the metastases of the other three patients (patients 20, 31, 36) clustered more closely to each other than to the primary tumor. A detailed overview of the mutations and SCNAs in the hierarchical clustering can be found in the supplementary material, Figure l S18.

Phylogenetic relationship analysis by EXPANDS [28] at a population resolution revealed that the 4.7N population and not the 2.0N or 3.4N population of the primary tumor was most closely related to the metastasis (supplementary material, Figure S21). In fact, all SCNAs that were shared between the primary tumor and the metastasis were detected in the 4.7N population of the primary tumor (Figure 6E). However, the fact that four mutations were shared exclusively between the 2.0N tumor population and the metastasis exclude the possibility that the 4.7N population was the direct precursor of the metastasis. Interestingly, the 2.0N and 3.4N populations did not share a single SCNA or mutation except for ubiquitous alterations. Even the two regions (chr7q and chr9p) that were affected by LOH in both populations resulted from 'mirrored LOH', i.e. the paternal chromosome was lost in one and the maternal chromosome in the other population (supplementary material, Figure S22). Conversely, the 4.7N population and the metastasis retained both chromosomes. In fact,

the 4.7N population shared mutations and SCNAs either exclusively with the 2.0N or with the 3.4N population. This indicates a branched evolution with substantial subclonal diversification in this primary tumor. Despite this

Overall high ITH, the fact that 85% of mutations and 90% of SCNAs in the metastasis were already present in the primary tumor, yet in distinct populations with different ploidies, is in concordance with the rest of the cohort.

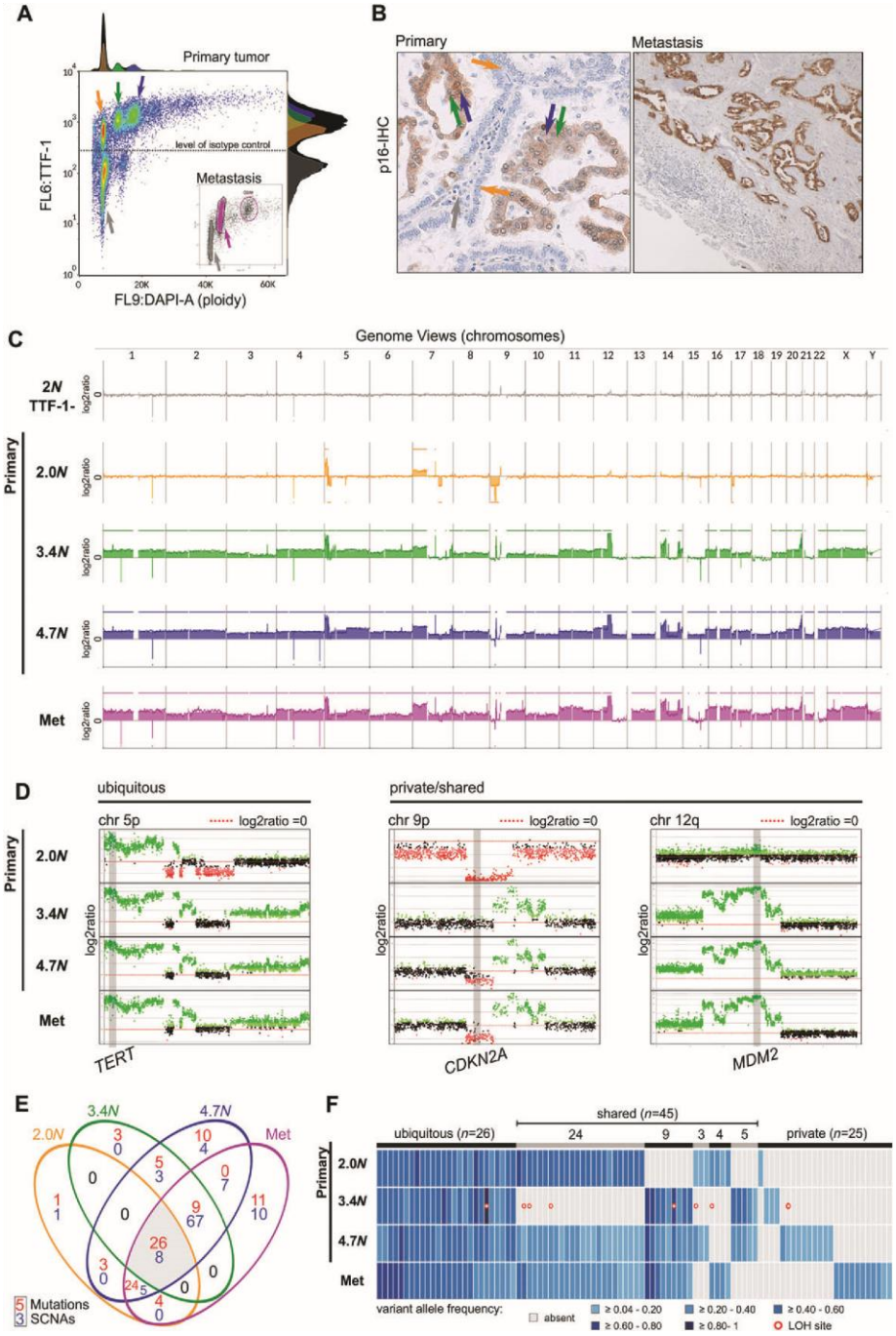


Figure 6. Legend on next page

Figure 6. Sorting reveals high intratumoral heterogeneity and complex tumor evolution in a single primary tumor. (A) The primary tumor of patient 42 consisted of three TTF-1-positive tumor populations (2.0*N*: orange; 3.4*N*: green; and 4.7*N*: blue), whereas only one tumor population (3.8*N*: pink) was present in the metastasis. The 2.0*N* population accounted for 33.3% of the tumor mass. The two aneuploid (3.4*N*, 4.7*N*) populations made up 9.5% and 36.4% of the tumor, respectively. The rest of the specimen (20.8%) consisted of diploid, TTF-1-negative CNA-neutral cells. (B) Immunohistochemistry for p16. The 2.0*N* TTF-1-positive population had a homozygous deletion of *CDKN2A*. Consequently, diploid tumor cells did not express p16. (C) All tumor populations, but not the diploid, TTF-1-negative population, displayed SCNAs. (D) Selected SCNAs that were either ubiquitous (*TERT*, chr 5p, left panel), suggesting that they occurred early in tumorigenesis; shared between the aneuploid tumor populations (*MDM2*, chr 12q, right panel); or private to the 2.0*N* population (homozygous deletion of *CDKN2A*, chr 9p, middle panel). (E) Venn diagram showing the number of SCNAs and mutations that were ubiquitous, shared or private. (F) Overview of 96 validated mutations with their VAFs and distribution among the tumor populations. Details about the mutations and their VAFs can be found in the supplementary material, Figure S20 and Table S6. Except for ubiquitous alterations, the 2.0*N* and 3.4*N* populations of the primary tumor did not share a SCNA or mutation. However, the 2.0*N* population shared 24 mutations and the breakpoint on chr 9p21 with the 4.7*N* population and with the metastasis. Likewise, the 3.4*N* population shared nine mutations and the high amplification of the *MDM2* oncogene with the 4.7*N* population and the metastasis. In fact, many mutations and SCNAs that were present in either the 2.0*N* or the 3.4*N* population were detected in the 4.7*N* population. Moreover, all SCNAs that were shared between the primary tumor (2.0*N*, 3.4*N*, and 4.7*N*) and the metastasis were present in the 4.7*N* population alone. In addition, the 2.0*N* TTF-1-positive population shared four mutations exclusively with the metastasis.

Discussion

To investigate the spatiotemporal ITH of SCNAs and cancer gene mutations and to study genome evolution in metastatic progression, we applied comprehensive genomic analyses to highly purified clonal tumor cell populations from matched primary–metastatic biopsies of 16 patients with LUAD. We observed that more than 80% of SCNAs and cancer gene mutations in the metastases were already present in the primary tumors and vice versa. While a high primary–metastatic concordance may be expected for mutations in tobacco smoking-induced lung cancers [29], it was not previously shown at the level of SCNAs in LUADs.

Chromosomal instability and aneuploidy are regarded as hallmarks of cancer [30], and inferences on the timing ofmetastatic dissemination must consider thepossibility of an inherent CIN of tumors. The fact that SCNAs were detected in all tumor populations indicates a history of CIN throughout our cohort. Although chromosomally unstable tumors would be expected to accumulate large-scale genomic alterations over time, metastases did not have more SCNAs than primary tumors in our study. This is surprising given that the median time from primary tumor detection to the occurrence of the first biopsied metastasis was almost 1 year. As recently demonstrated in breast cancer [31], punctuated evolution, defined as a short period of crisis, followed by stasis and expansion of stable clones could result in aneuploid tumors with rather stable karyotypes. However, tumors with dysregulated DNA-damage response pathways are unlikely to restore chromosomal stability. In line with previous reports [32,33], tumors with biallelic *TP53* or *ATM* mutations had more SCNAs compared with the rest of the cohort. As CIN is most likely ongoing during metastatic evolution in such tumors, it is more plausible that deleterious SCNAs are not propagated unless they provide an additional fitness advantage. Our finding is in agreement with studies reporting a high concordance of SCNAs between primary tumors and metastases in colorectal and breast cancer [34,35]. In contrast, studies in renal cell carcinoma [36] and prostate cancer [37] found that the number of SCNAs continuously increases over time, resulting in a higher burden of SCNAs at metastatic sites than at primary sites. This indicates differences in the dynamics of SCNAs during evolution and metastasis across cancer types.

Notwithstanding this cohort's overall low primary–metastatic ITH, at least two tumors showed clear signs of branched evolution and subclonal diversification, as recently reported in LUAD [4]. First, alterations unique to each of two metastases were found in the primary

tumor of patient 22, indicating that the founding metastatic subclones were present at the time of primary tumor resection. Furthermore, we observed extensive ITH in patient 42, where only 7.6% of all SCNAs and 27.1% of all mutations were present in all tumor populations. However, 85% of all mutations and 90% of all SCNAs in the metastasis were already present at the time of primary tumor resection, yet in three different tumor populations in the primary tumor. Despite the profound ITH in these two cases, they coincide with our observation that most SCNAs are established in the primary tumor before metastatic dissemination.

Our study has several limitations. Array CGH allowed only for the detection of unbalanced SCNAs. We can therefore not exclude the possibility that chromosomal abnormalities without copy number changes would result in more ITH than observed. However, NSCLCs are thought to be more dependent on SCNAs than balanced translocations or inversions [38]. Likewise, mirrored allelic imbalances, in which the maternal allele is gained or lost in one region, whereas the paternal is gained or lost in another region, are known to be outnumbered by SCNAs in LUADs [4]. Further, they would be unlikely to have identical intrachromosomal breakpoints. In addition, we were limited to study one metastatic focus in most patients, a common and unavoidable constraint in retrospective studies with
biobank material. With the caveats of a limited sample size and a moderate gene panel size fully acknowledged, this study is – to the best of our knowledge – the first to flow-sort nuclei based on both ploidy and TTF-1 expression to investigate the ITH of SCNAs and cancer gene mutations in longitudinal biopsies of matched primary and metastatic LUAD.

Our refined nuclei flow-sorting approach provides genome-wide analysis of SCNAs and deep sequencing of 409 cancer genes of tumor populations with unprecedented tumor DNA purity. This is important in comparative genomic studies because relative tumor content often varies between regions, which can result in misclassifying truncal mutations as region-specific private mutations in stroma-rich, i.e. tumor-poor, regions, especially when studying specimens of moderate tumor cellularity such as LUADs [13,32]. Importantly, our methodology enabled the genomic characterization of specimens heavily intermixed with stromal components and was suitable for both FF and FFPE tissues. Its power isexemplifiedbytheabilitytouncovermultifacete dsituations, such as in patient 42, which argues for the future use of nuclei flow-sorting approaches, ideally combining them with tumor lineage markers, such as TTF-1 in LUADs, to extend their application to the characterization of diploid tumor populations that would be missed by conventional bulk tumor analysis.

A recent multi-region sequencing study revealed a high concordance of 21 cancer gene mutations in single biopsies of primary LUADs [3]. We extended this observation to primary tumors and matched metastases for many additional cancer gene mutations and at the level of SCNAs in 16 patients with LUAD. Taken together, our data are consistent with the perception that most SCNAs and cancer gene mutations accumulate in the primary tumor before metastatic dissemination, thereby challenging the view of a high degree of CIN during metastatic evolution. Furthermore, our observation that metastases served as the source of further metastatic spread in three of the four patients with oligometastatic disease might have clinical implications, whereby the removal of the metastatic lesions has the potential to prevent further metastatic progression [39,40]. Additional approaches in larger matched cohorts are needed to better understand how CIN and ITH impact the biology of LUADs and how this interplay eventually influences cancer evolution.

Acknowledgements

This project was supported by the ‘Swiss National Science Foundation’ (SNF 138513) and the L. & Th. La Roche Foundation. We thank Bruno Grilli, Petra Hirschmann, Sabrina Rau, and

Luca Quagliata for their help in performing next-generation sequencing (NGS) and IHC investigations. The chromosomal copy number aberration data reported in this paper have been deposited at the National Center for Biotechnology Information Gene Expression Omnibus (accession No GSE119763). The sequencing data have been deposited in the Sequence Read Archive (accession No SRP158478).

Author contributions statement

TL performed research, analyzed and interpreted data, and wrote the manuscript. NA analyzed and interpreted data and revised the manuscript. VP and TD performed flow-sorting experiments. DJ provided scientific support, interpreted data, and revised the manuscript. AK performed library preparations and revised the manuscript. KP and SG performed the EXPANDS analysis. DM collected the clinical data and revised the manuscript. SS performed histological and immunohistochemical analysis. DL provided the tumor samples and clinical support. MB provided scientific support, interpreted data, and revised the manuscript. CR designed the study, analyzed and interpreted the data, and revised the manuscript. LB performed histological and immunohistochemical analysis, conceived and designed the study, analyzed and interpreted data, and revised the manuscript.

References

1. Gerlinger M, Rowan AJ, Horswell S, *et al*. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012; **366:** 883–892.
2. Molina JR, Yang P, Cassivi SD, *et al*. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc* 2008; **83:** 584–594.
3. Zhang J, Fujimoto J, Wedge DC, *et al*. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* 2014; **346:** 256–259.
4. Jamal-Hanjani M, Wilson GA, McGranahan N, *et al*. Tracking the evolution of non-small-cell lung cancer. *N Engl J Med* 2017; **376:** 2109–2121.
5. de Bruin EC, McGranahan N, Mitter R, *et al*. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* 2014; **346:** 251–256.
6. The Cancer Genome Atlas Research Network. Comprehensivemolecular profiling of lung adenocarcinoma. *Nature* 2014; **511:** 543–550.
7. Campbell JD, Alexandrov A, Kim J, *et al*. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat Genet* 2016; **48:** 607–616.
8. Tan Q, Cui J, Huang J, *et al*. Genomic alteration during metastasis of lung adenocarcinoma. *Cell Physiol Biochem* 2016; **38:** 469–486.
9. Saber A, Hiltermann TJN, Kok K, *et al*. Mutation patterns in small cell and non-small cell lung cancer patients suggest a different level of heterogeneity between primary and metastatic tumors. *Carcinogenesis* 2017; **38:** 144–151.
10. Yang B, Luo L, Luo W, *et al*. The genomic dynamics during progression of lung adenocarcinomas. *J Hum Genet* 2017; **62:** 783–788.
11. Brastianos PK, Carter SL, Santagata S, *et al*. Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer Discov* 2015; **5:** 1164–1177.
12. Sherwood J, Dearden S, Ratcliffe M, *et al*. Mutation status concordance between primary lesions and metastatic sites of advanced non-small-cell lung cancer and the impact of mutation testing methodologies: a literature review. *J Exp Clin Cancer Res* 2015; **34:** 92.
13. Carter SL, Cibulskis K, Helman E, *et al*. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* 2012; **30:** 413–421.
14. Ruiz C, Lenkiewicz E. Advancing a clinically relevant perspectiveof the clonal nature of cancer. *Proc Natl Acad Sci U S A* 2011; **108:** 12054–12059.
15. Krishan A, Dandekar PD. DAPI fluorescence in nuclei isolated fromtumors. *J Histochem Cytochem* 2005; **53:** 1033–1036.
16. Corver WE, ter Haar NT. High-resolution multiparameter DNAflow cytometry for the detection and sorting of tumor and stromal subpopulations from paraffin-embedded tissues. *Curr Protoc Cytom* 2009; Ch 6: Unit 6.27.
17. Yatabe Y, Mitsudomi T, Takahashi T. TTF-1 expression in pulmonary adenocarcinomas. *Am J Surg Pathol* 2002; **26:** 767–773.
18. Yang C-Y, Lin M-W, Chang Y-L, *et al*. Programmed cell death-ligand 1 expression is associated with a favourable immune microenvironment and better overall survival in stage I pulmonary squamous cell carcinoma. *Eur J Cancer* 2016; **57:** 91–103.
19. Ding L, Getz G, Wheeler DA, *et al*. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 2008; **455:** 1069–1075.
20. Weir BA, Woo MS, Getz G, *et al*. Characterizing the cancer genome in lung adenocarcinoma. *Nature* 2007; **450:** 893–898.
21. Knudsen ES, Wang JYJ. Targeting the RB-pathway in cancer therapy. *Clin Cancer Res* 2010; **16:** 1094–1099.
22. Nagayama K, Kohno T, Sato M, *et al*. Homozygous deletion scanning of the lung cancer genome at a 100-kb resolution. *Genes Chromosomes Cancer* 2007; **46:** 1000–1010.
23. Mlakar V, Berginc G, Volavšek M, *et al*. Presence of activating *KRAS* mutations correlates significantly with expression of tumour

suppressorgenes*DCN* and*TPM1*incolorectalcancer.*BMCCancer* 2009; **9:** 282.

24. Paige AJW, Taylor KJ, Taylor C, *et al. WWOX*: a candidate tumor suppressor gene involved in multiple tumor types. *Proc Natl Acad Sci U S A* 2001; **98:** 11417–11422.
25. Roth A, Khattra J, Yap D, *et al.* PyClone: statistical inference of clonal population structure in cancer. *Nat Methods* 2014; **11:** 396–398.
26. Turajlic S, Swanton C. Metastasis as an evolutionary process. *Science* 2016; **352:** 169–175.
27. Oxnard GR, Lo PC, Nishino M, *et al.* Natural history and molecular characteristics of lung cancers harboring EGFR exon 20 insertions. *J Thorac Oncol* 2013; **8:** 179–184.
28. Andor N, Harness JV, Müller S, *et al.* EXPANDS: expanding ploidy and allele frequency on nested subpopulations. *Bioinformatics* 2014; **30:** 50–60.
29. Govindan R, Ding L, Griffith M, *et al.* Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* 2012; **150:** 1121–1134.
30. Gordon DJ, Resio B, Pellman D. Causes and consequences ofaneuploidy in cancer. *Nat Rev Genet* 2012; **13:** 189–203.
31. Gao R, Davis A, McDonald TO, *et al.* Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat Genet* 2016; **48:** 1119–1130.
32. Zack TI, Schumacher SE, Carter SL, *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat Genet* 2013; **45:** 1134–1140.
33. Shlien A, Tabori U, Marshall CR, *et al.* Excessive genomic DNA copy number variation in the Li–Fraumeni cancer predisposition syndrome. *Proc Natl Acad Sci U S A* 2008; **105:** 11264–11269.
34. Sylvester BE, Vakiani E. Tumor evolution and intratumor heterogeneity in colorectal carcinoma: insights from comparative genomic profiling of primary tumors and matched metastases. *J Gastrointest Oncol* 2015; **6:** 668–675.
35. Navin N, Kendall J, Troge J, *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* 2011; **472:** 90–94.
36. Gerlinger M, Horswell S, Larkin J, *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat Genet* 2014; **46:** 225–233.
37. Robinson D, Van Allen EM, Wu Y-M, *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* 2015; **161:** 1215–1228.
38. Speicher MR, Petersen S, Uhrig S, *et al.* Analysis of chromosomal alterations in non-small cell lung cancer by multiplex-FISH, comparative genomic hybridization, and multicolor bar coding. *Lab Invest* 2000; **80:** 1031–1041.
39. Lanuti M. Surgical management of oligometastatic non-small celllung cancer. *Thorac Surg Clin* 2016; **26:** 287–294.
40. Guerrero E, Ahmed M. The role of stereo tactic ablative radiotherapy (SBRT) in the management of oligometastatic non small cell lung cancer. *Lung Cancer* 2016; **92:** 22–28.

*41. Juskevicius D, Dietsche T, Lorber T, *et al.* Extracavitary primary effusion lymphoma: clinical, morphological, phenotypic and cytogenetic characterization using nuclei enrichment technique. *Histopathology* 2014; **65:** 693–706.

*42. Arriola E, Lambros MBK, Jones C, *et al.* Evaluation of Phi29-based whole-genome amplification for microarray-based comparative genomic hybridisation. *Lab Invest* 2007; **87:** 75–83.

*43. Huang J, Pang J, Watanabe T, *et al.* Whole genome amplification for array comparative genomic hybridization using DNA extracted from formalin-fixed, paraffin-embedded histological sections. *J Mol Diagn* 2009; **11:** 109–116.

*44. Nilsen G, Liestol K, Lingjaerde O. copynumber: Segmentation of single- and multi-track copy number data by penalized least squares regression. Bioconductor Version 3.4; 2013. [Accessed 17 April 2017]. Available from: https://bioconductor.org/packages/release/bioc/html/copynumber.html

*45. LifeTechnologies. Ion AmpliSeq™ DNA and RNA Library Preparation. Publication Part Number MAN0006735. Revision E.0; 2017. [Accessed 26 June 2017]. Available from: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0006735_AmpliSeq_DNA_RNA_LibPrep_UG.pdf

*46. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; **38:** e164.

*47. Koboldt DC, Zhang Q, Larson DE, *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012; **22:** 568–576.

*48. Li H, Handsaker B, Wysoker A, *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* 2009; **25:** 2078–2079.

*49. Robinson JT, Thorvaldsdóttir H, Winckler W, *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011; **29:** 24–26.

*50. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003; **31:** 3812–3814.

*51. Adzhubei IA, Schmidt S, Peshkin L, *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7:** 248–249.

*52. Schwarz JM, Rödelsperger C, Schuelke M, *et al.* MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010; **7:** 575–576.

*53. Shihab HA, Gough J, Cooper DN, *et al.* Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat* 2013; **34:** 57–65.

*54. Choi Y, Sims GE, Murphy S, *et al.* Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 2012; **7:** e46688.

*55. Carter H, Chen S, Isik L, *et al.* Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. *Cancer Res* 2009; **69:** 6660–6667.

*56. Yadav VK, De S. An assessment of computational methods for estimating purity and clonality using genomic data derived from heterogeneous tumor tissue samples. *Brief Bioinform* 2015; **16:** 232–241.

*57. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 2004; **20:** 289–290.

*Cited only in supplementary material.

SUPPLEMENTARY MATERIAL

ONLINE

Supplementary materials and methods

Appendix S1. Overview of copy number aberration plots of all 39 sorted tumor populations from 37 tumor samples from 16 patients with matched primary–metastatic lung adenocarcinoma

Figure S1. Decision tree and overview of the project

Figure S2. Array-CGH of clonally unrelated cases

Figure S3. Localization of 16 primary tumors and 21 metastases for the 16 patients of the clonally related cohort

Figure S4. Overview of the multi-parameter flow-sorting approach for diploid tumors and samples with low tumor purity

Figure S5. Illustration of the multi-parameter flow-sorting approach

Figure S6. Purity estimations before and after sorting

Figure S7. Overview of mutations per patient and biopsy

Figure S8. Genome-wide overview of SCNAs per tumor population

Figure S9. Circular bar plot displaying the relative fraction of different SCNA sizes per sample in all 39 tumor populations

Figure S10. Somatic copy number aberrations in tumors with or without biallelic (mutation and LOH) inactivation of TP53 or ATM

Figure S11. Somatic copy number aberrations in primary tumors and metastases

Figure S12. Ploidy and SCNA burden in distant and locoregional metastases

Figure S13. Somatic copy number aberrations by size in distant and locoregional metastases

Figure S14. Conserved variant allele frequencies (VAFs) over the course of metastatic disease

Figure S15. Absolute copy numbers of mutations shared between the primary tumor (x-axis) and the first metastasis (y-axis)

Figure S16. Violin plots showing the distribution of variant allele frequencies (VAFs) and cellular prevalence

Figure S17. Variant allele frequencies (VAFs) and copy numbers of all cancer gene mutations detected in this cohort

Figure S18. Hierarchical clustering of mutations [e.g. chr1:12345678(C>T)] and SCNAs (e.g. 1:12345678–12345687) in patients with oligometastatic disease

Figure S19. Hematoxylin (HE) staining and IHC for TTF-1 on primary tumor of patient 42

Figure S20. Overview of 96 validated mutations with their VAFs and distribution among the tumor populations in patient 42 with a polygenomic primary tumor (tumor populations 2.0N, 3.4N, and 4.7N) and its matched metastasis (pleural effusion)

Figure S21. The phylogenetic relationship on metapopulation resolution for patient 42

Figure S22. Regions of loss of heterozygosity (LOH) in the three tumor populations from the primary tumor (2.0N, 3.4N, and 4.7N) and the metastasis of patient 42

Figure S23. Hierarchical clustering of Euclidean distances based on amplifications and deletions of all sorted tumor populations (mentioned in the supplementary material, Supplementary materials and methods)

Figure S24. Circos plot displaying the relative number of the genome affected by SCNAs in each sorted tumor population (mentioned in the supplementary material, Supplementary materials and methods)

Figure S25. Scatterplots displaying the VAFs of sequencing replicates from sorted tumor populations of the primary tumor and the metastasis of patient 42 (mentioned in the supplementary material, Supplementary materials and methods)

Figure S26. Scatterplots showing the VAFs of sorting replicates from tumor populations of the primary tumor and the metastasis of patient 42 (mentioned in the supplementary material, Supplementary materials and methods)

Figure S27. TTF-1 expression in normal cells (mentioned in the supplementary material, Supplementary materials and methods)

Table S1. Next-generation sequencing in three patients with genomically unrelated second primary lung tumors

Table S2. Names of the 409 genes sequenced with the Ion Ampliseq™ Comprehensive Cancer Panel

Table S3. Custom panel for ultra-deep resequencing of sorted populations for patient 42

Table S4. Next-generation sequencing in three patients with matched non-tumor tissue

Table S5. Non-silent and silent mutations identified across 39 sorted tumor populations from 16 patients with lung adenocarcinoma and clonally related metastases

Table S6. Per patient sequencing coverage at variant sides and allele frequencies of ubiquitous, shared, and private non-silent and silent mutations

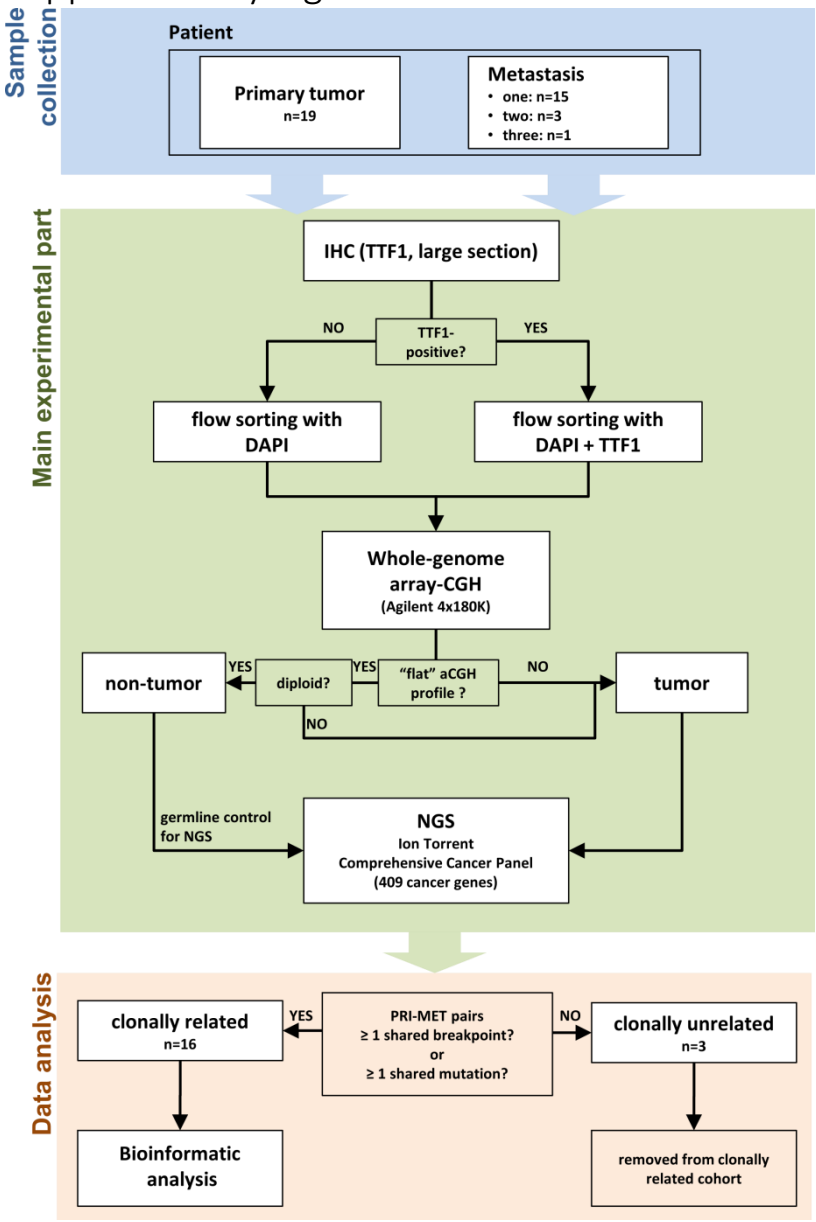
Table S7. List of SCNAs across 39 sorted tumor populations from 16 patients with lung adenocarcinoma and clonally related metastases

Table S8. Low Stringency Settings for Variant Caller for the CCP on Ion Proton (mentioned in the supplementary material, Supplementary materials and methods)

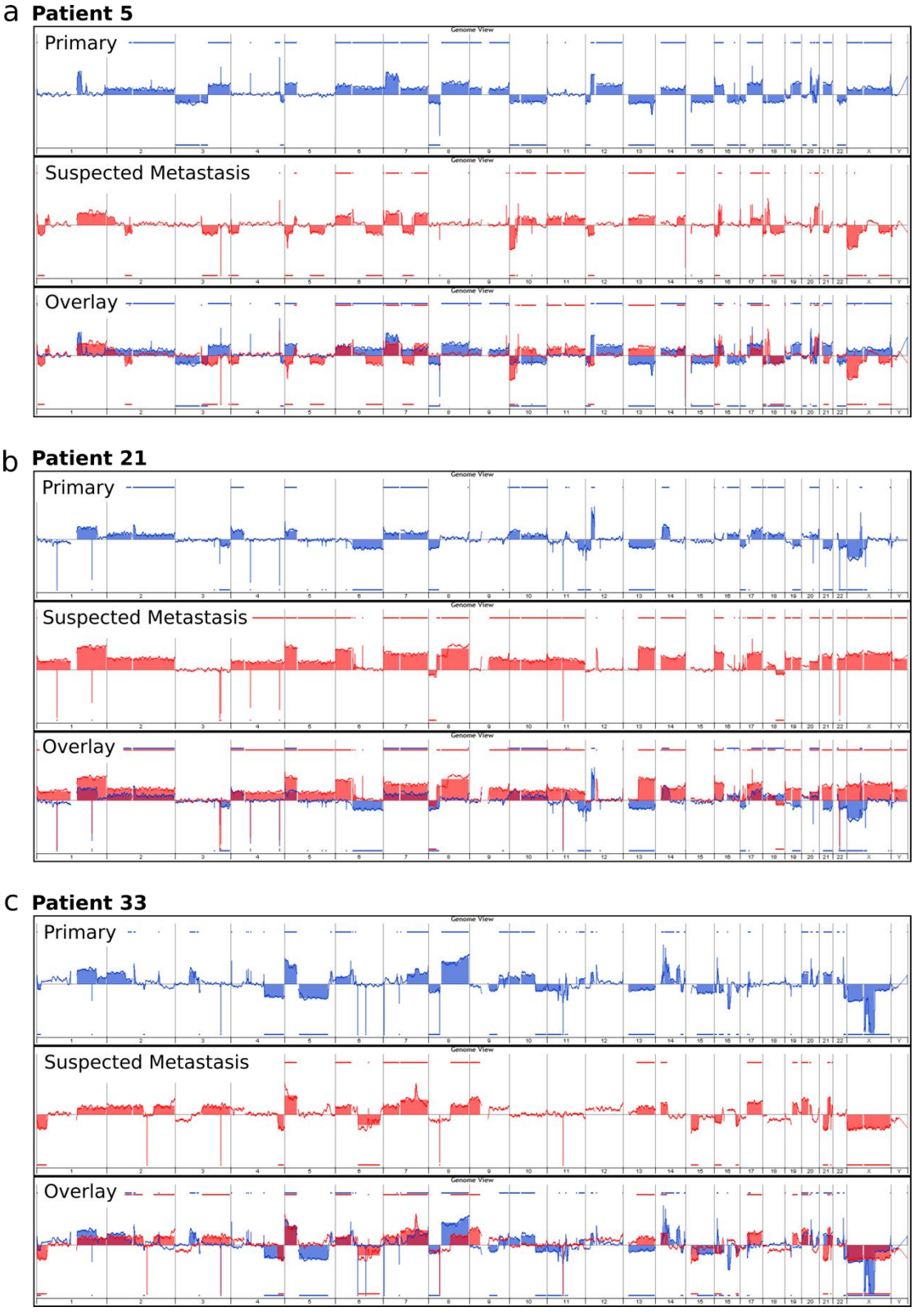
Table S9. NGS quality parameters for the Comprehensive Cancer Panel (mentioned in the supplementary material, Supplementary materials and methods)

Exploring the spatiotemporal genetic heterogeneity in metastatic lung adenocarcinoma using a nuclei flow-sorting approach
Lorber T, et al. J Pathol 2018 (DOI: 10.1002/path.5183)

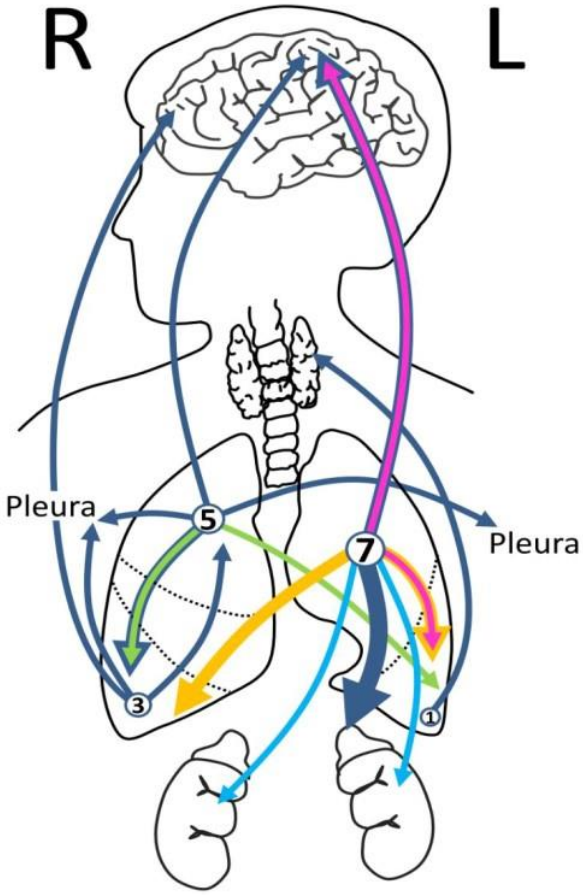
Supplementary Figures



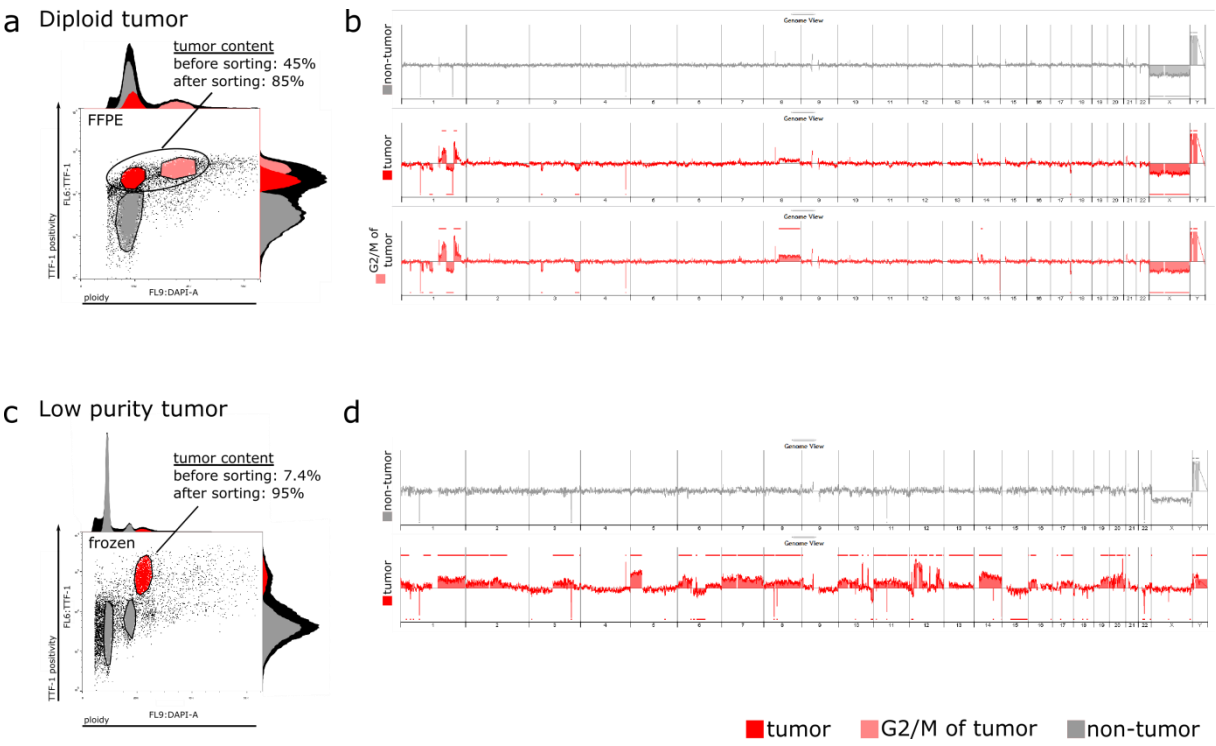
Supplementary Figure S1. Decision tree and overview of the project. Initially, 19 patients were included in this study, however in three patients the suspected metastases were clonally unrelated by both SCNAs and mutations from the primary tumors (not one breakpoint or mutation shared). These three patients were therefore excluded from the cohort of clonally related primary/metastatic lung adenocarcinomas. We performed IHC on FFPE section of all tumors to evaluate TTF-1 expression in tumor and non-tumor cells. If the tumor was TTF-1-positive, nuclei flow-sorting was performed with DAPI and TTF-1, or with DAPI only, when tumors were TTF-1-negative. Copy number aberrations of all sorted populations were assessed by aCGH. If diploid populations (in TTF-1-negative tumors) or diploid TTF-1-negative populations (in TTF-1-positive tumors) showed no copy number aberrations, we considered them to be of non-tumor origin. Likewise, if IHC indicated the presence of TTF-1-positive non-tumor cells within the biopsy (**Supplementary Figure S4**) and diploid, TTF-1- positive populations did not have any copy number aberrations in such tumors, we considered them to be of non-tumor origin.



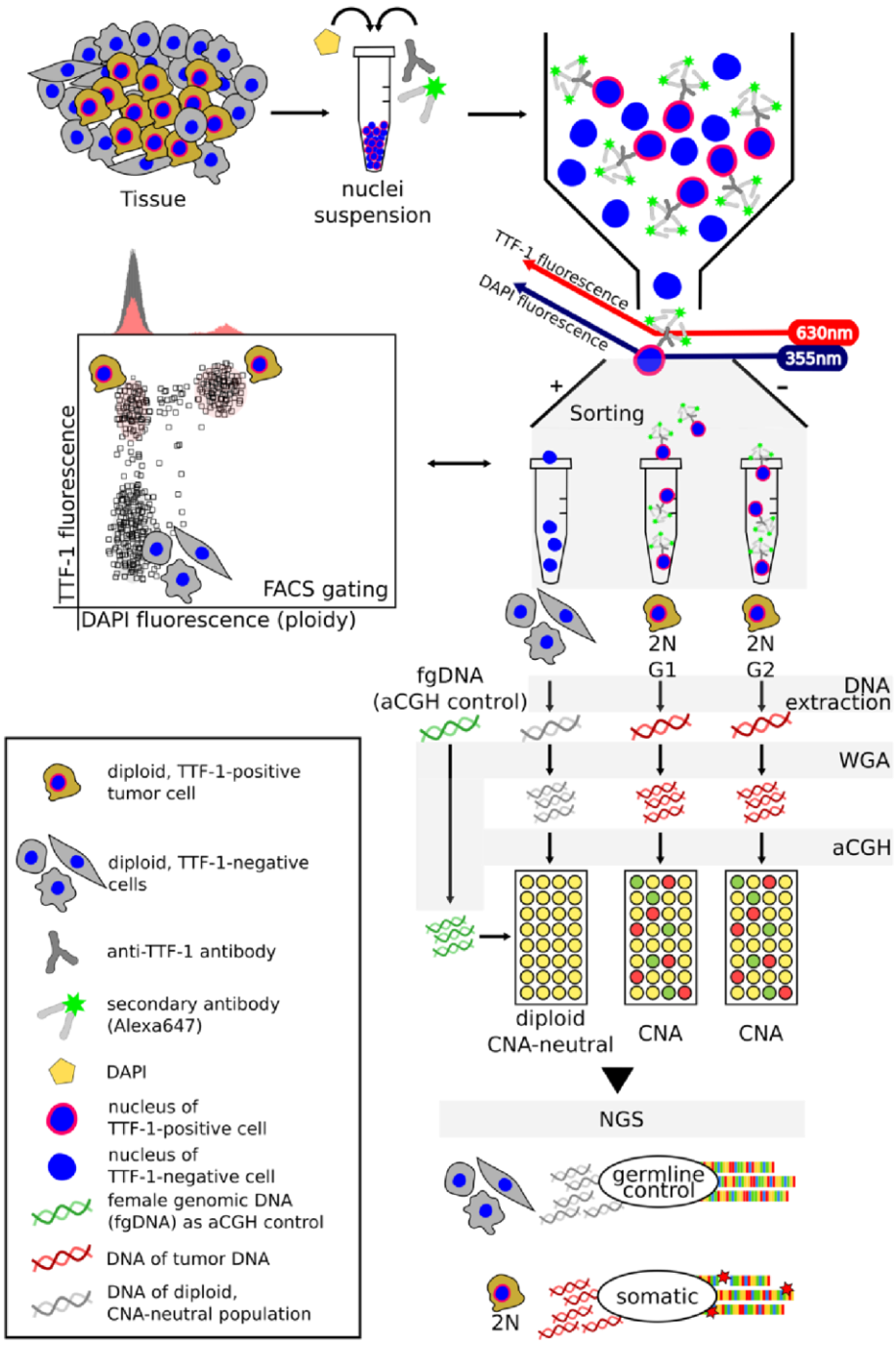
Supplementary Figure S2. Array-CGH of clonally unrelated cases. Overview of the genome-wide copy number profile of the primary tumors (blue) and the suspected metastases (red) of three patients. **A:** Patient 5, **B:** Patient 21, **C:** Patient 33. Not a single intrachromosomal SCNAs breakpoint is identical in position between the primary tumors and their suspected metastases. In addition, they do not share a single somatic mutation. (**Supplementary Table S1**). Moreover, the suspected metastases occurred in the lung. Therefore, the suspected metastases were considered to be genomically unrelated occurrences of second primary lung tumors.



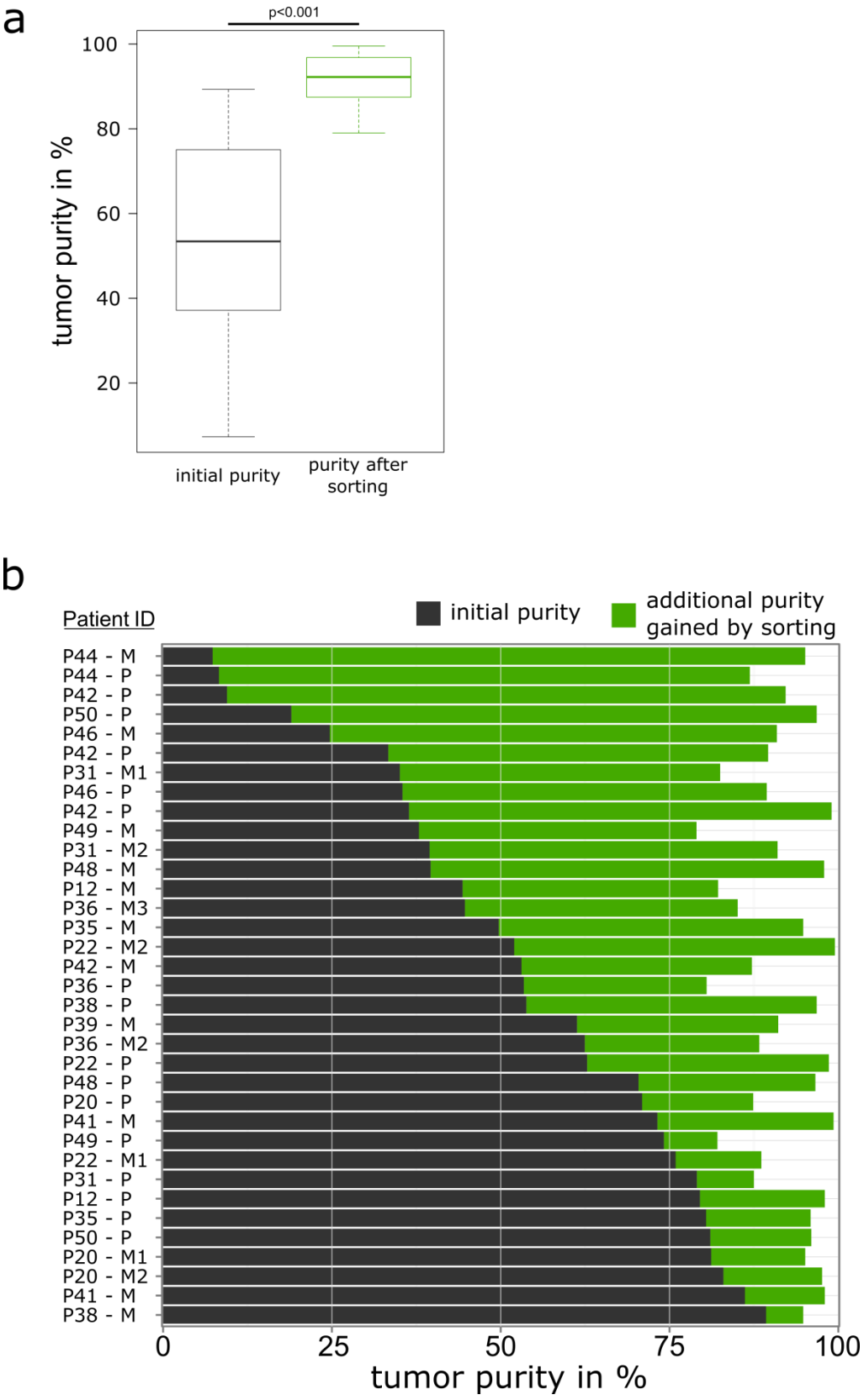
Supplementary Figure S3. Localization of 16 primary tumors and 21 metastases for the 16 patients of the clonally related cohort. The cohort included 16 patients with clonally related metastases and comprises 12 primary-metastatic pairs, three primary-metastatic-trios (Patients 20, 22 and 31) and one patient with primary tumor and three metastases (Patient 36). Metastases were either distant (n=10) or occurred locoregionally (n=11). Distant metastases affected the brain (n=4), adrenal glands (n=3), the kidneys (n=2) or the thyroid gland (n=1). Locoregional metastases occurred either ipsilateral (n=7) or contralateral (n=4) and appeared either in lung tissue (n=8) or arose as pleural effusions (n=3). The arrows in colors other than dark blue indicate the metastatic sites of the four patients with oligometastatic disease. This included three patients with two metastases (Patient 20 [bilateral kidney metastases, light blue]; Patient 22 [brain and ipsilateral lung metastasis, pink]; Patient 31 [ipsilateral and contralateral lung metastases]) and one patient with three metastases (Patient 36 [one ipsilateral and two contralateral lung metastases]). Note that arrows do not indicate a directionality of any kind. The numbers in circles indicate the number of patients (n = 16) with primary tumors in a specific lobe.



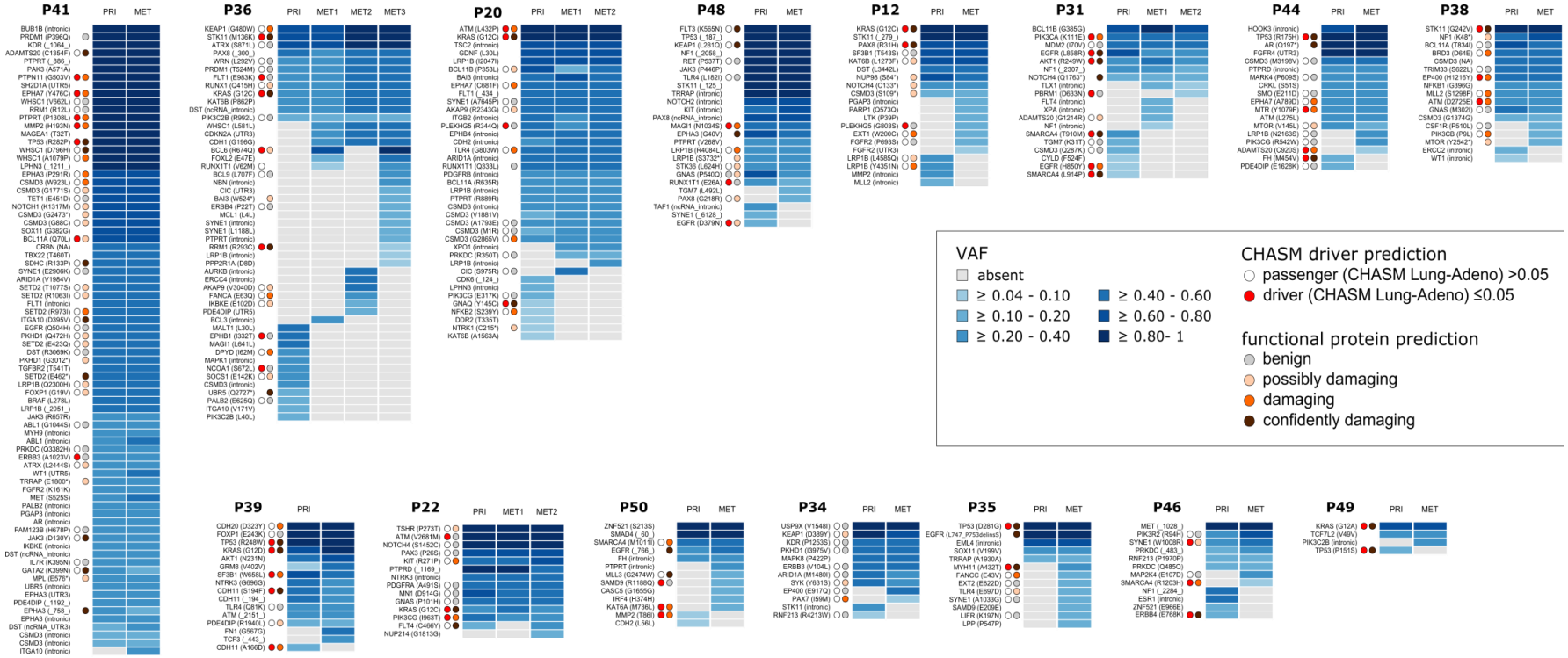
Supplementary Figure S4. Overview of the multiparameter flow sorting approach for diploid tumors and samples with low tumor purity. A: A representative example of a diploid tumor, where the diploid tumor population could be separated from diploid stroma cells only by using TTF-1 (y-axis) as a second parameter in addition to DAPI (DNA content, x-axis). **B:** Diploid TTF-1 negative cells (grey) displayed no SCNAs, while distinct SCNAs are visible for the tumor population (red). **C:** An example of an aneuploid tumor highly contaminated by stroma. Sorting with TTF-1 in addition to DAPI enriched the low purity sample from 7.4% of tumor cells to 95%. **D:** Multiple SCNAs are visible in the TTF-1 positive aneuploidy tumor population (red), whereas no SCNAs were present in the diploid, TTF-1 negative population. This workflow works with both FFPE (A, B) and FF (C, D) tissue samples.



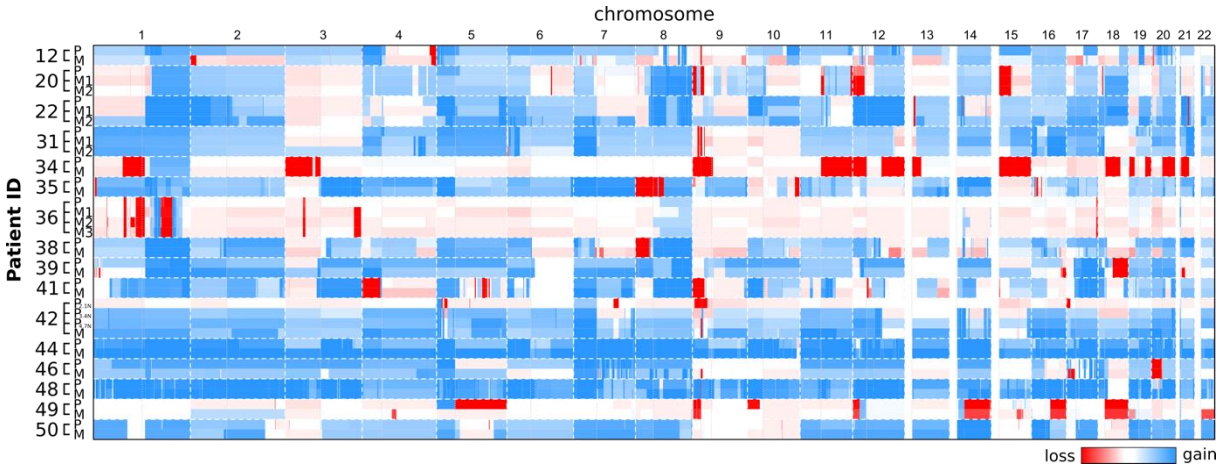
Supplementary Figure S5. Illustration of the multiparameter flow sorting approach. Here, our workflow applied to a diploid tumor is illustrated to reinforce the use of TTF-1 as a second parameter in addition to DAPI. This approach comprises the isolation of tumor nuclei from both FF and FFPE tissues, followed by a multiparameter flow-sort with DAPI for DNA content and TTF-1 for TTF-1-positive tumor cells. Subsequently DNA of sorted populations is extracted, subjected to whole genome amplification (WGA) with Phi29 followed by genome-wide detection of SCNAs (aCGH) or identification of somatic mutations in 409 cancer genes (NGS with the Ampliseq Comprehensive Cancer Panel). Nuclei of diploid tumor cells do not differ in ploidy from diploid non-tumor cells and a separation by ploidy (DAPI) is not sufficient. Staining for TTF-1 can help to sort and genomically investigate these diploid tumors. Moreover, this approach can be applied to TTF-1-positive tumors with low tumor cellularity (**Supplementary Figure S4C, D**) that are often excluded from studies because they are highly intermixed with non-tumor cells. In addition, DNA of "diploid, CNA-neutral" populations (non- tumor cells) can be utilized as germline control for sequencing analyses (**Supplementary Table S4**).



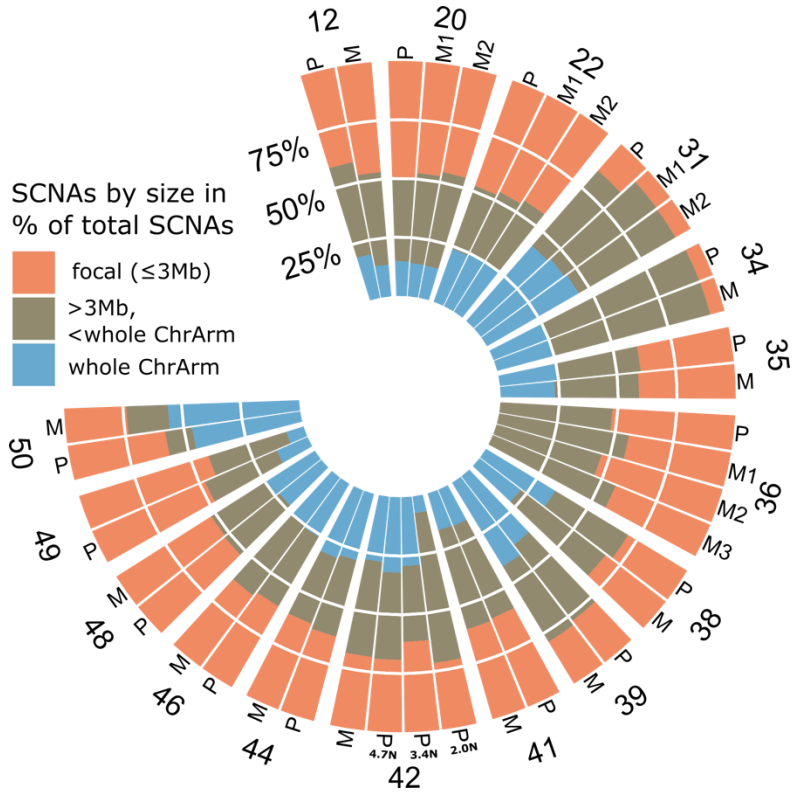
Supplementary Figure S6. **Purity estimations before and after sorting. A:** Box plots show the median, first quartile, third quartile, minimum, and maximum of purity before and after sorting. **B:** The bar plot illustrates the gained purity per sample due to the nuclei-flow sorting approach (n=35). Four of the 39 sorted tumor populations were excluded because EXPANDS could not predict purity due to a low number of SCNAs.



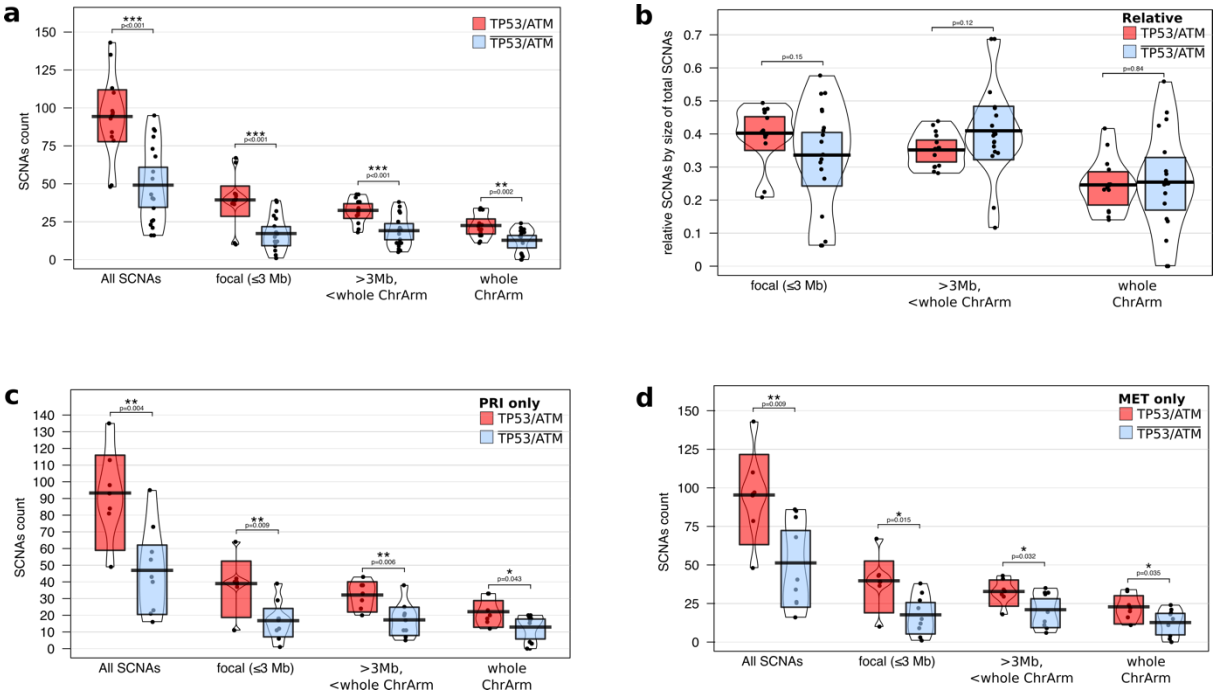
Supplementary Figure S7. Overview of mutations per patient and biopsy. P indicates the patient ID. Patient 42 is not shown here but discussed in detail in **Figure 6** and **Supplementary Figures S20 and S22.**



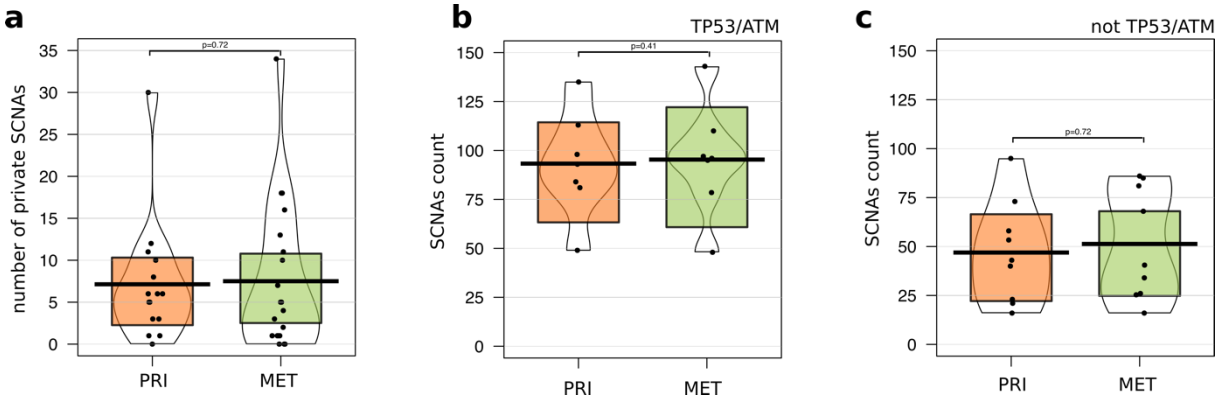
Supplementary Figure S8. Genome-wide overview of SCNAs per tumor population. P: primary tumor, M: metastasis. Red: loss, Blue: Gain.



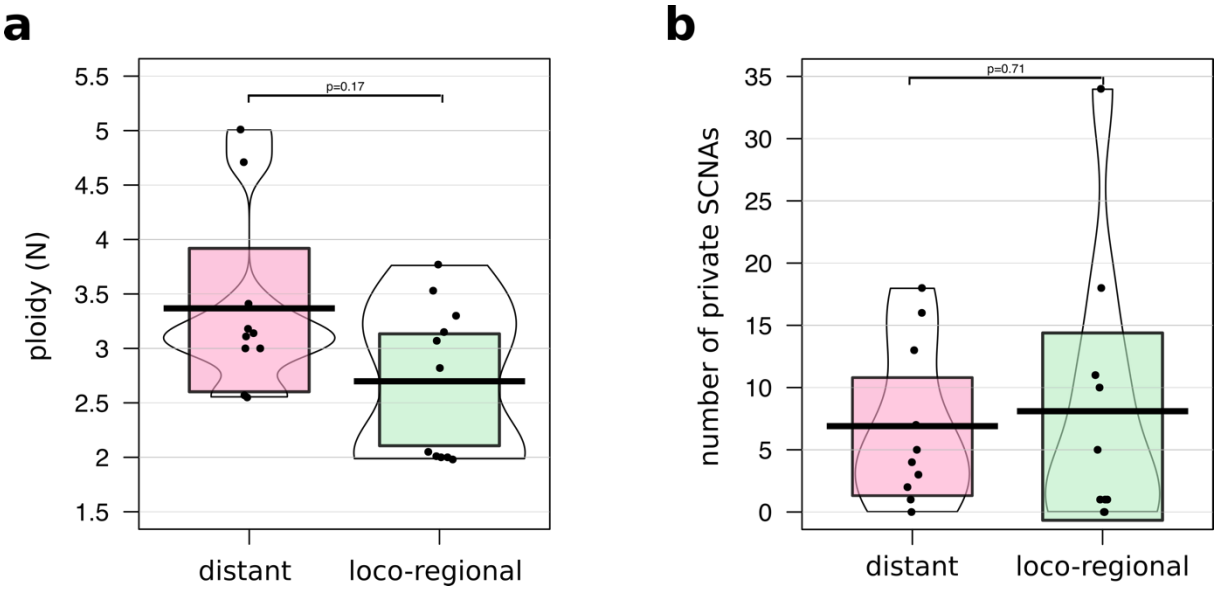
Supplementary Figure S9. Circular bar plot displaying the relative fraction of different SCNAs sizes per sample in all 39 tumor populations. P, primary tumor; M(1-3), metastases. Numbers around the circular bar plot indicate the patient number.



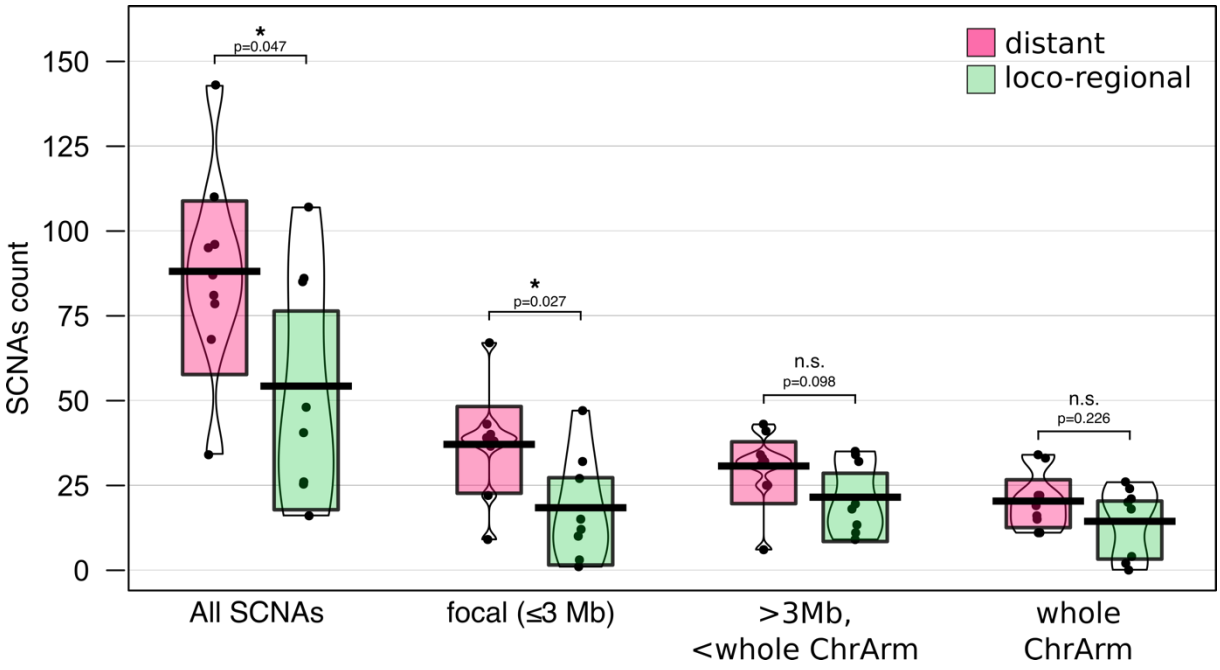
Supplementary Figure S10. Somatic copy number aberrations in tumors with or without biallelic (mutation and LOH) inactivation of TP53 or ATM. A: Absolute numbers of SCNAs by size. **B:** Relative numbers of SCNAs by size. **C:** Absolute numbers of SCNAs by size regarding primary tumors only. **D:** Absolute numbers of SCNAs by size regarding metastases only. Tumors with biallelic inactivation of TP53 or ATM (TP53/ATM, n=16) had higher numbers of SCNAs in total and across all SCNAs size categories compared to tumors without biallelic inactivation of TP53 or ATM (!"#$/&!', n=23). This difference remained significant across all sizes, even when the tumors were divided into primary tumors and metastases. Statistic test: Welch's *t*-test. ***p<0.001, **p<0.01, *p<0.05.



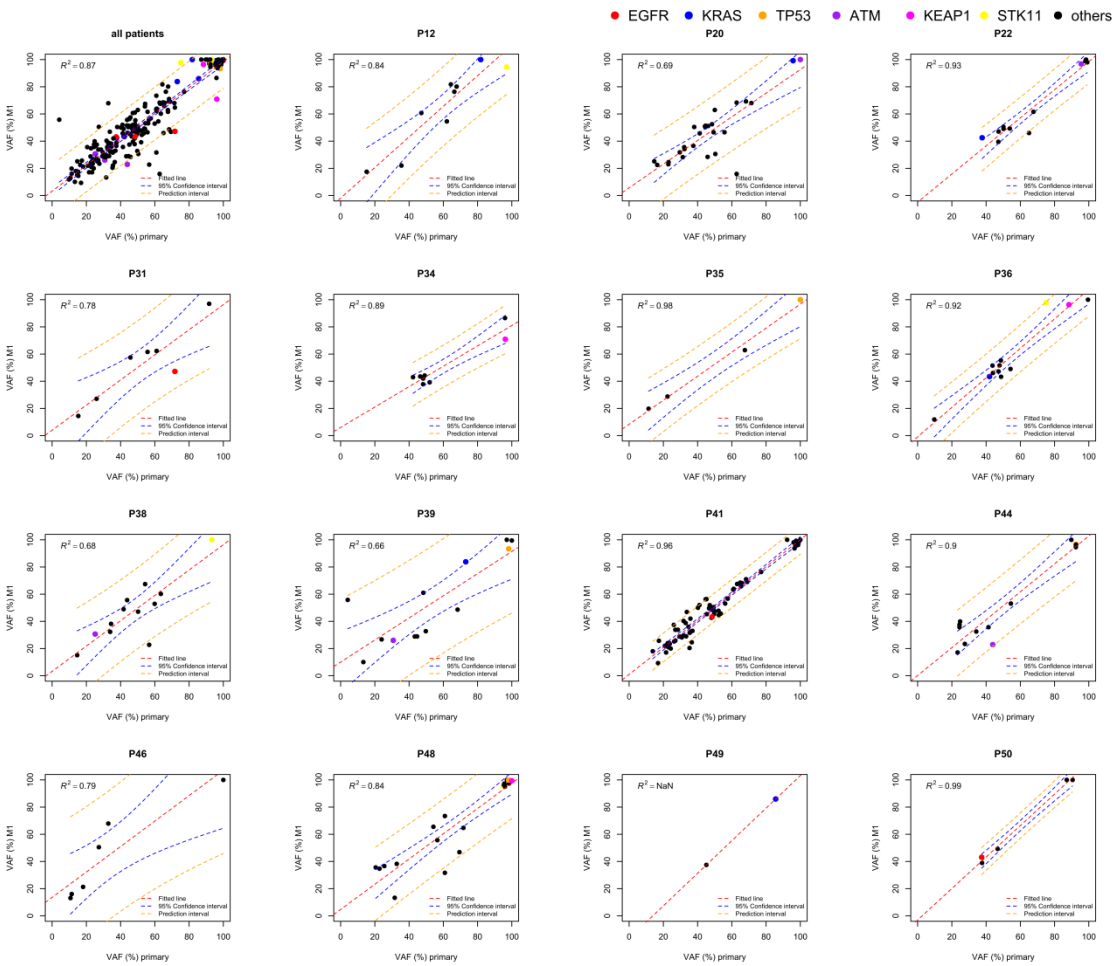
Supplementary Figure S11. Somatic copy number aberrations in primary tumors and metastases. A: Number of SCNAs private to primary tumors and metastases (Unpaired Two-Sample Wilcoxon Test). **B:** Number of SCNAs of primary tumors and metastases in tumors with biallelic inactivation of TP53 or ATM (Paired Sample T- Test). **C:** Number of SCNAs of primary tumors and metastases in tumors without biallelic inactivation of TP53 or ATM (Paired Sample T-Test). Metastases did not have more private SCNAs than primary tumors. Dividing tumors by their TP53/ATM-status also did not reveal any difference in the number of SCNAs between primary tumors and metastases.



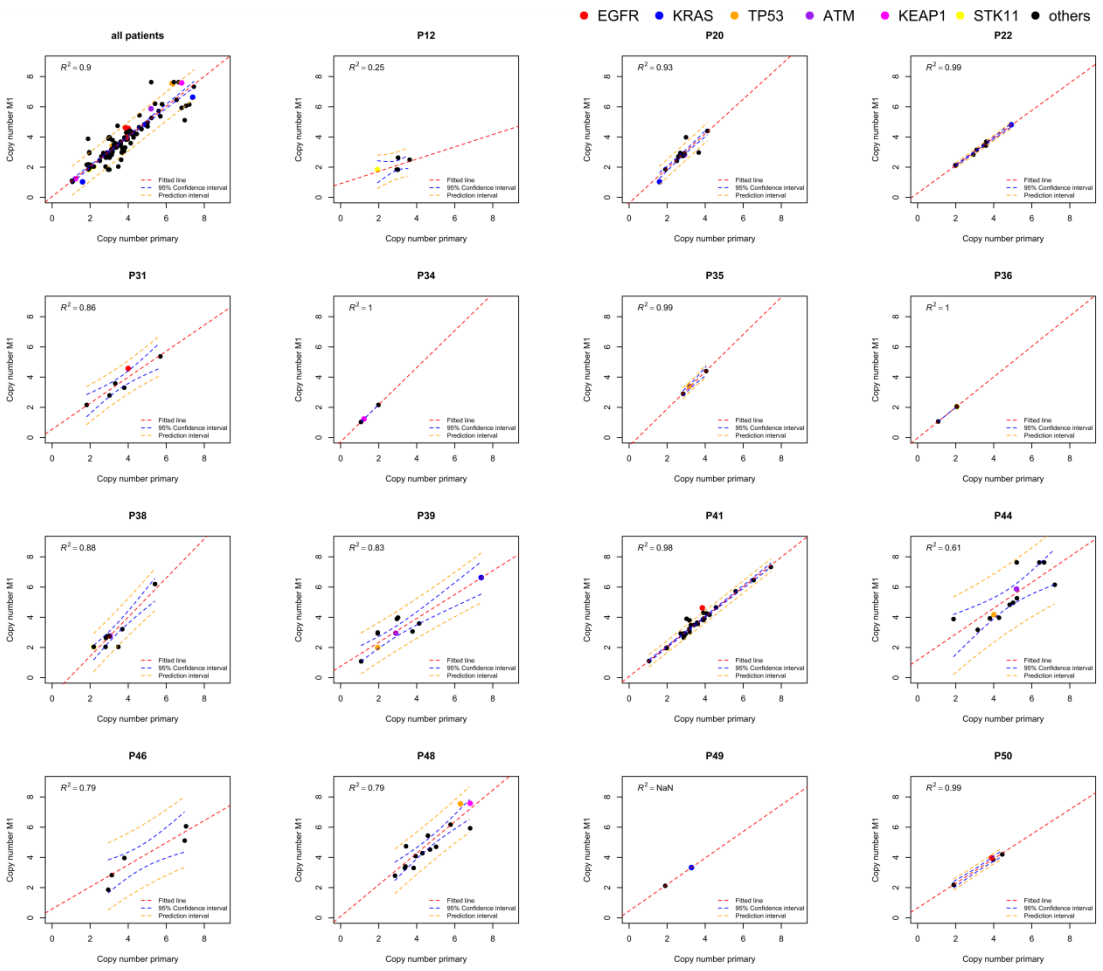
Supplementary Figure S12. Ploidy and SCNA burden in distant and loco-regional metastases. A: Violine plots displaying the distribution of ploidy between distant and loco-regional metastases (Unpaired Two-Sample Wilcoxon Test). **B:** Violine plots showing the number of SCNAs in distant and loco-regional metastases (Unpaired Two-Sample Wilcoxon Test). Neither ploidy nor the number of SCNAs was statistically significant between distant and loco-regional metastases (significance threshold p£0.05).



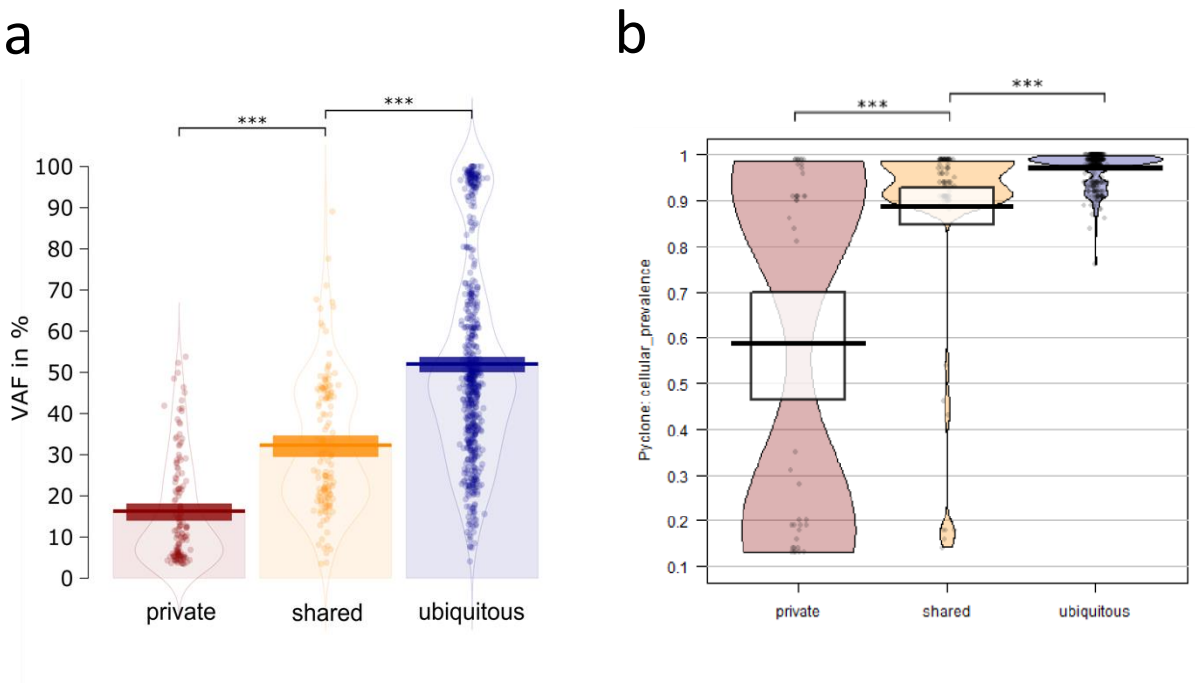
Supplementary Figure S13. Somatic copy number aberrations by size in distant and loco-regional metastases. Distant metastases had significantly more SCNAs than loco-regional metastases. In addition, focal SCNAs occurred more often in distant than in loco-regional metastases, whereas no significant difference was found for SCNAs larger than 3Mb between these two groups (Unpaired Two-Sample Wilcoxon Test, significance threshold p£0.05). n.s.: not significant.



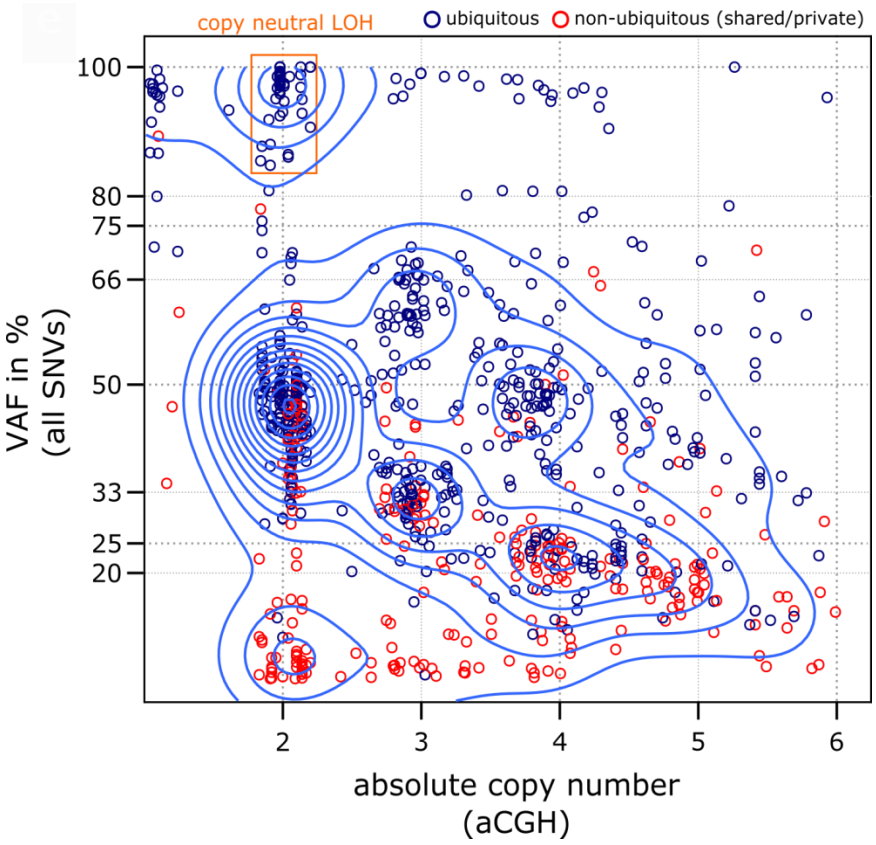
Supplementary Figure S14. Conserved variant allele frequencies (VAFs) over the course of metastatic disease. VAFs of mutations shared between the primary tumor (x-axis) and the first metastasis (y-axis). In the majority of cases this comprises ubiquitous mutations, due to the presence of only one metastasis. Patient 42 is excluded, due to the presence of multiple populations in the primary tumor. The correlation among the cohort (top left plot) is linear and significant ($p<0.001$). Little change in the VAF is evident in most mutations and patients. P: patient



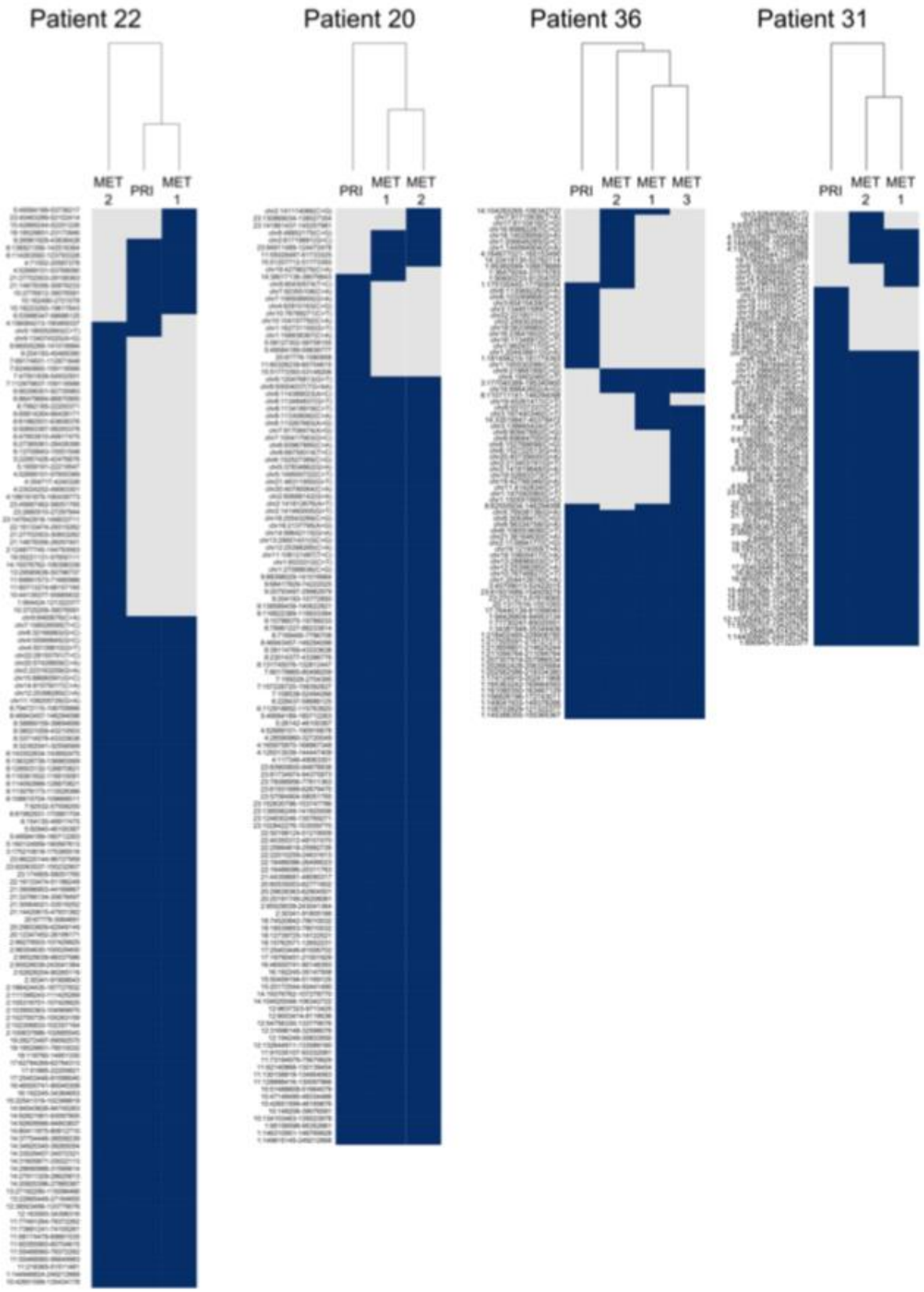
Supplementary Figure S15. Absolute copy numbers of mutations shared between the primary tumor (x-axis) and the first metastasis (y-axis). In the majority of cases, due to the presence of only one metastasis, this comprises the VAFs of ubiquitous mutations. Patient 42 is excluded, due to the presence of multiple populations in the primary tumor. The correlation among the cohort (top left plot) is linear and significant ($p<0.001$). Little difference in the absolute copy numbers is evident in the patients. P: patient



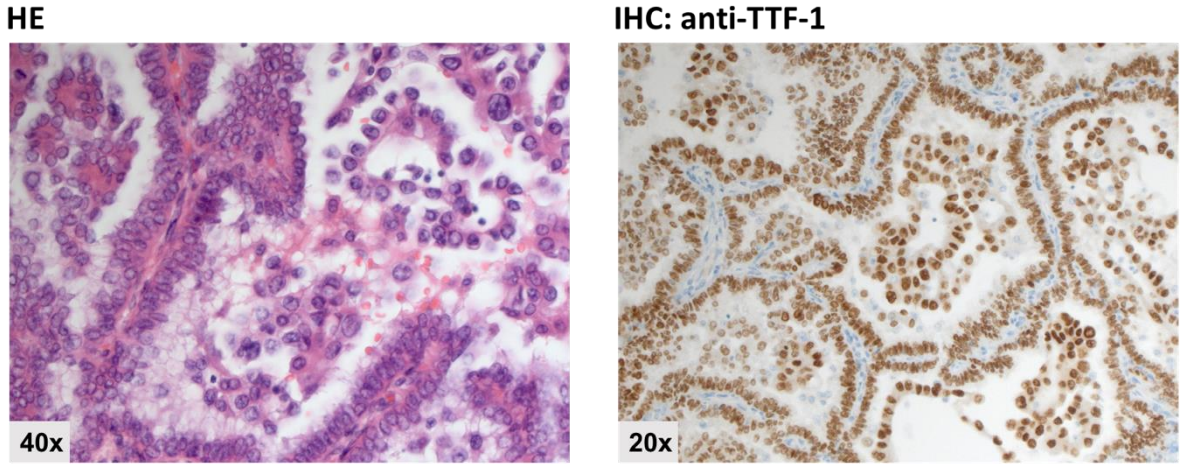
Supplementary Figure S16. Violin Plots showing the distribution of variant allele frequencies (VAFs) and cellular prevalence. A: Mutations were classified according to their presence in all tumor biopsies (ubiquitous), present in more than 1 but not all (shared, applicable for the four patients with n_{MET}>1) or present in just one. **B:** Pyclone analysis revealed that the cellular prevalence is significantly higher in ubiquitous mutations than in shared or private mutations. Pyclone needs a minimum number of mutations to reach unequivocal predictions. Therefore, only samples with a minimum of 19 mutations were included. This comprised all samples (n=15) from 5 patients (Patients 20, 36, 41, 42, 48).



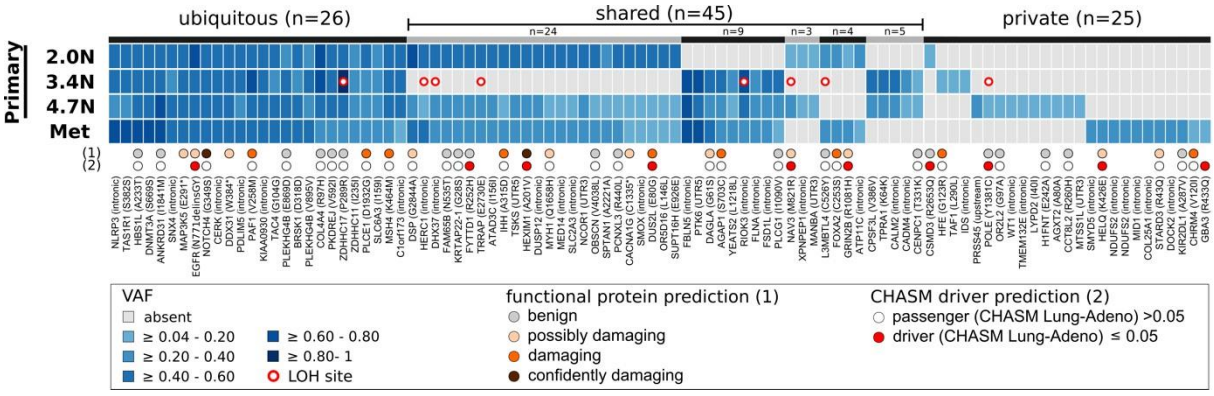
Supplementary Figure S17. Variant allele frequencies (VAF) and copy numbers of all cancer gene mutations detected in this cohort. Density plot of the absolute copy number (x-axis) and VAF (y-axis) of the mutations. Note that the majority of non-ubiquitous (shared or private) mutations accumulate at lower VAFs, indicating a non-truncal and therefore subclonal origin.



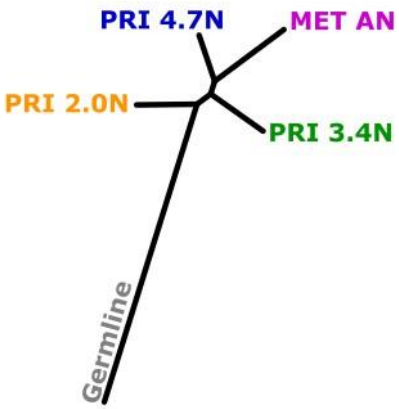
Supplementary Figure S18. Hierarchical clustering of mutations (e.g. chr1:12345678(C>T)) and SCNAs (e.g. 1:12345678-12345687) in patients with oligometastatic disease.



Supplementary Figure S19. Hematoxylin (HE) staining and IHC for TTF-1 on primary tumor of Patient 42. All tumor cells, including the micropapillary (intraluminal) and lepidic tumor cells are TTF-1-positive (brown). The TTF-1-negative (blue) cells, located intraluminal, are macrophages. Lepidic tumor cells account for the 2.0N tumor population and the micropapillary tumor cells account for the aneuploidy tumor populations (3.4N, 4.7N) in this primary tumor of Patient 42.

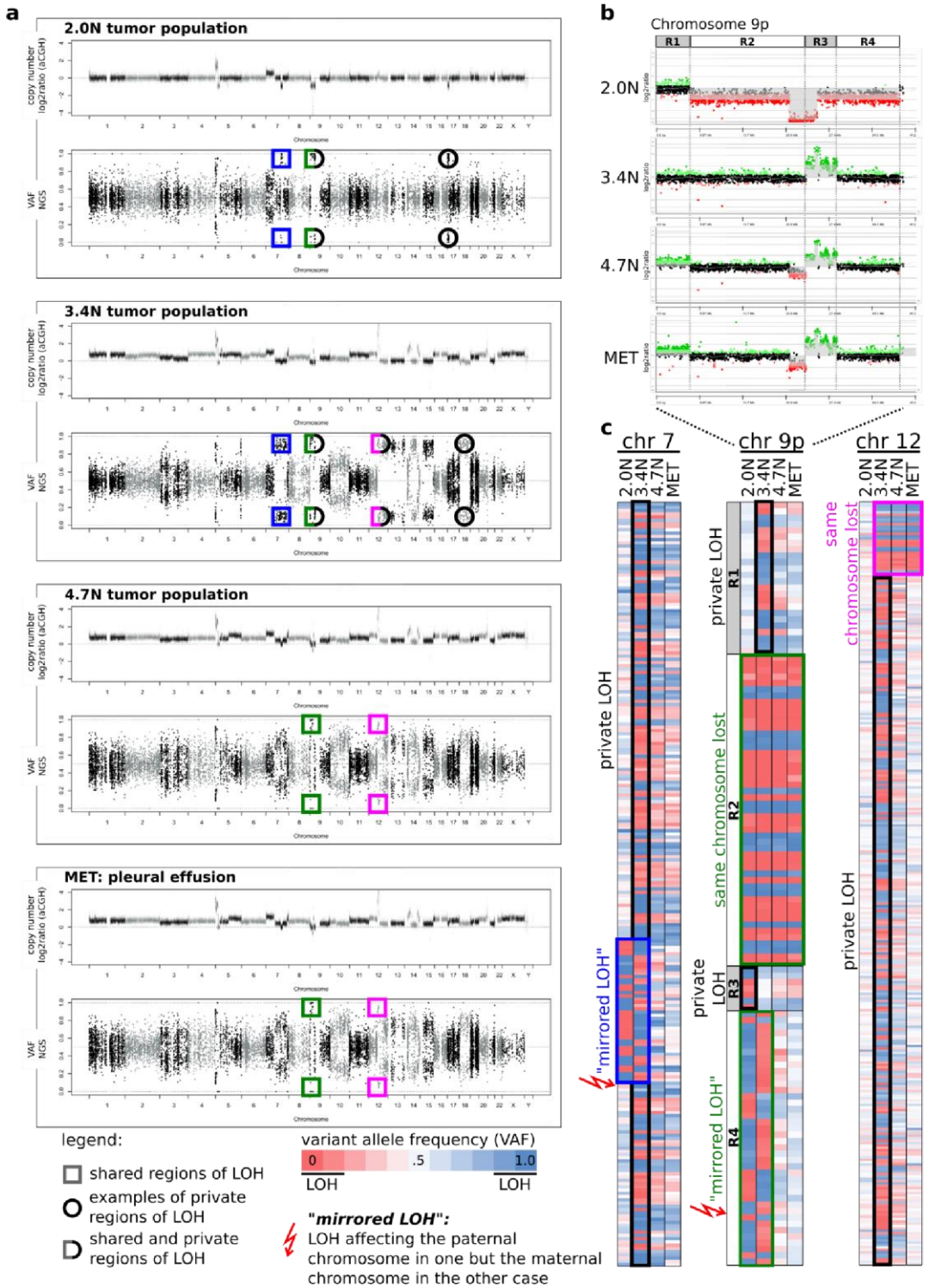


Supplementary Figure S20. Overview of 96 validated mutations with their VAFs and distribution among the tumor populations in Patient 42 with a polygenomic primary tumor (tumor populations 2.0N, 3.4N and 4.7N) and its matched metastasis (pleural effusion). Except for ubiquitous alterations, the 2.0N and 3.4N population of the primary tumor did not share a single additional SCNA or mutation. However, the 2.0N population shared 24 mutations and the breakpoint on chr 9p21 with the 4.7N population and with the metastasis. Likewise, the 3.4N population shared nine mutations and the high amplification of the *MDM2* oncogene with the 4.7N population and the metastasis. In fact, most mutations and SCNAs that were present in either the 2.0N or 3.4N population were detected in the 4.7N population. Moreover, all SCNAs that were shared between the primary tumor (2.0N, 3.4N and 4.7N) and the metastasis were present in the 4.7N population alone. In addition, the 2.0N TTF-1-positive population shared four mutations exclusively with the metastasis.

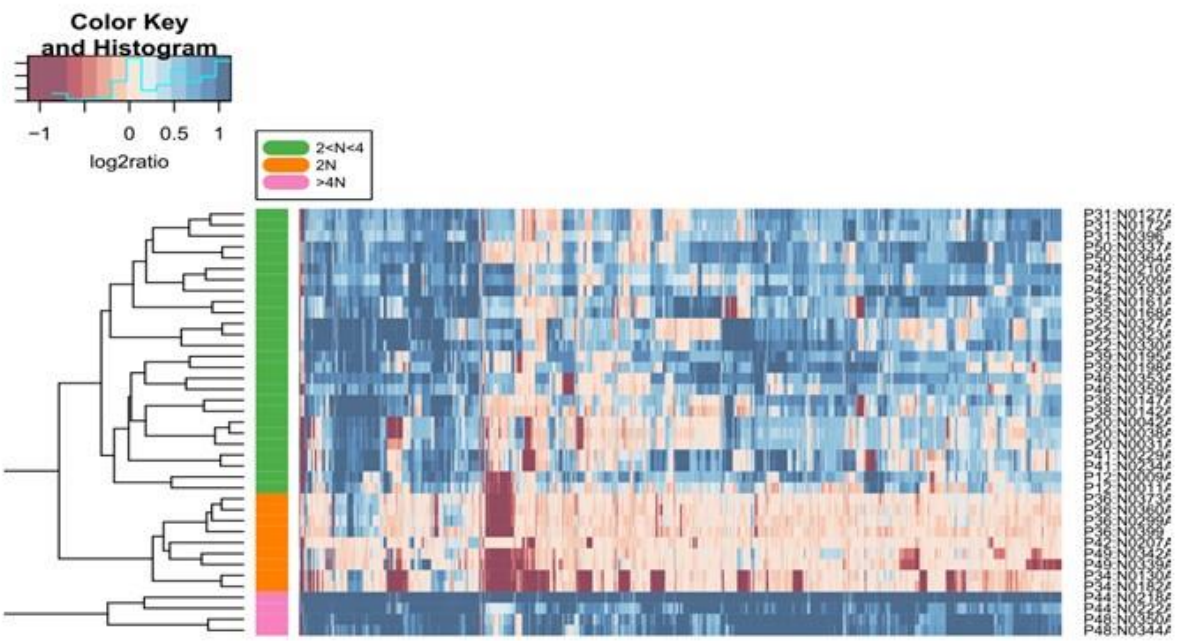


Supplementary Figure S21. The phylogenetic relationship on metapopulation resolution for Patient 42

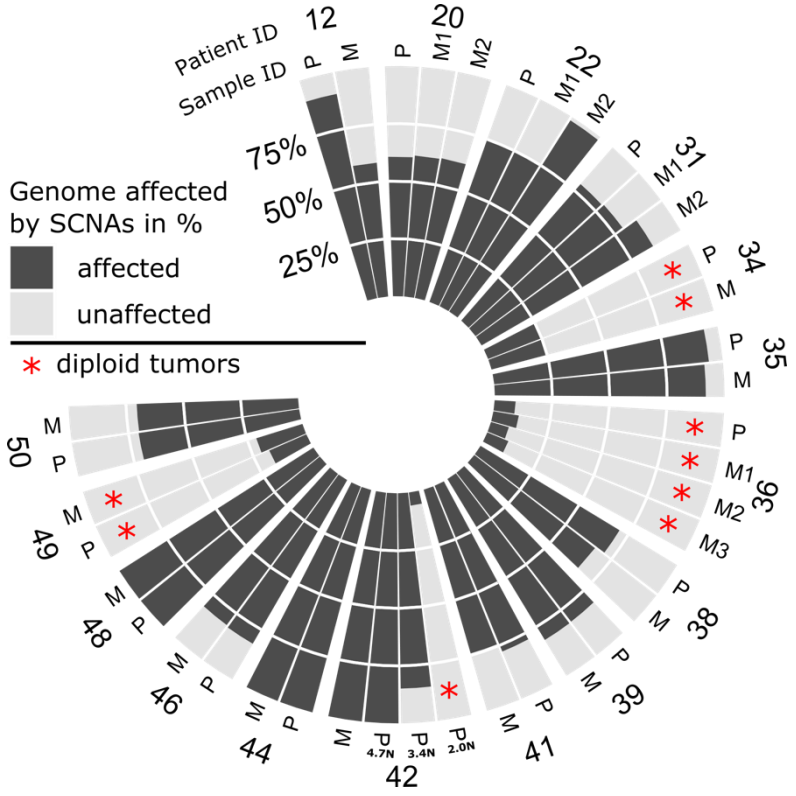
(calculated by EXPANDS based on SCNAs and mutations). The 4.7N population of the primary tumor is most closely related to the metastasis.



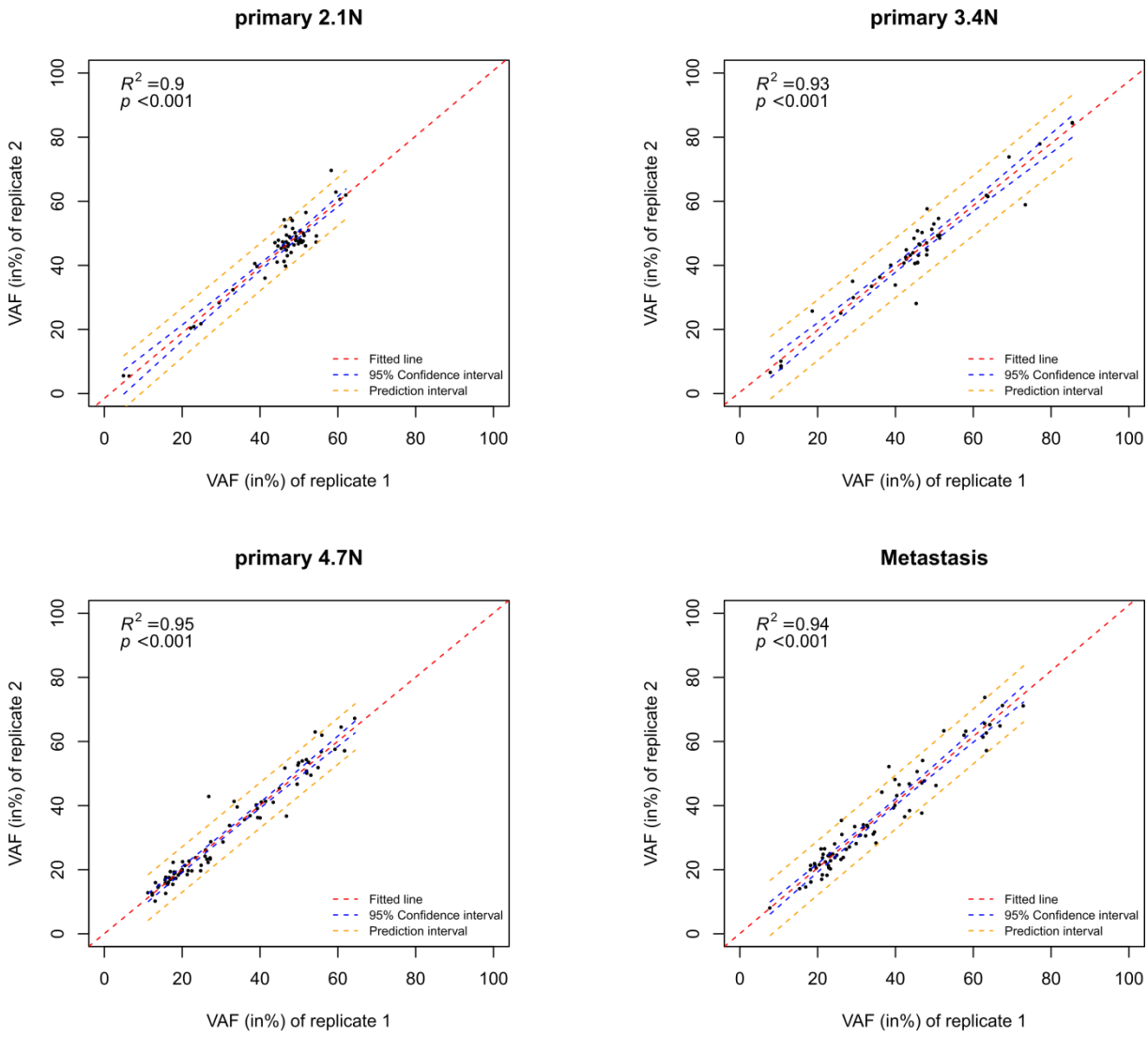
Supplementary Figure S22. Regions of loss of heterozygosity (LOH) in the three tumor populations from the primary tumor (2.0N, 3.4N and 4.7N) and the metastasis of Patient 42. A, C: Ubiquitous LOH regions (green, R2) and LOH events that are shared among the 3.4N, 4.7N population and the metastasis (pink) affect the same chromosome. In contrast, the two LOH events that are shared between the 2.0N and 3.4N population (blue, green, R4) resulted from mirrored LOH, i.e. they affected the other homologous chromosome, respectively. **B:** Detailed view of chromosome 9p. The 4.7N population and the metastasis share the same breakpoint on chromosome 9p (region R2 covering the CDKN2A locus) with the 2.0N population. However, this breakpoint is absent in the 3.4N population. Many LOH events in the 3.4N population (chr13, chr14, chr15, chr18) and the LOH in the 2.0N population (chr17) are not present in the 4.7N population or in the metastasis, indicating a branched evolution.



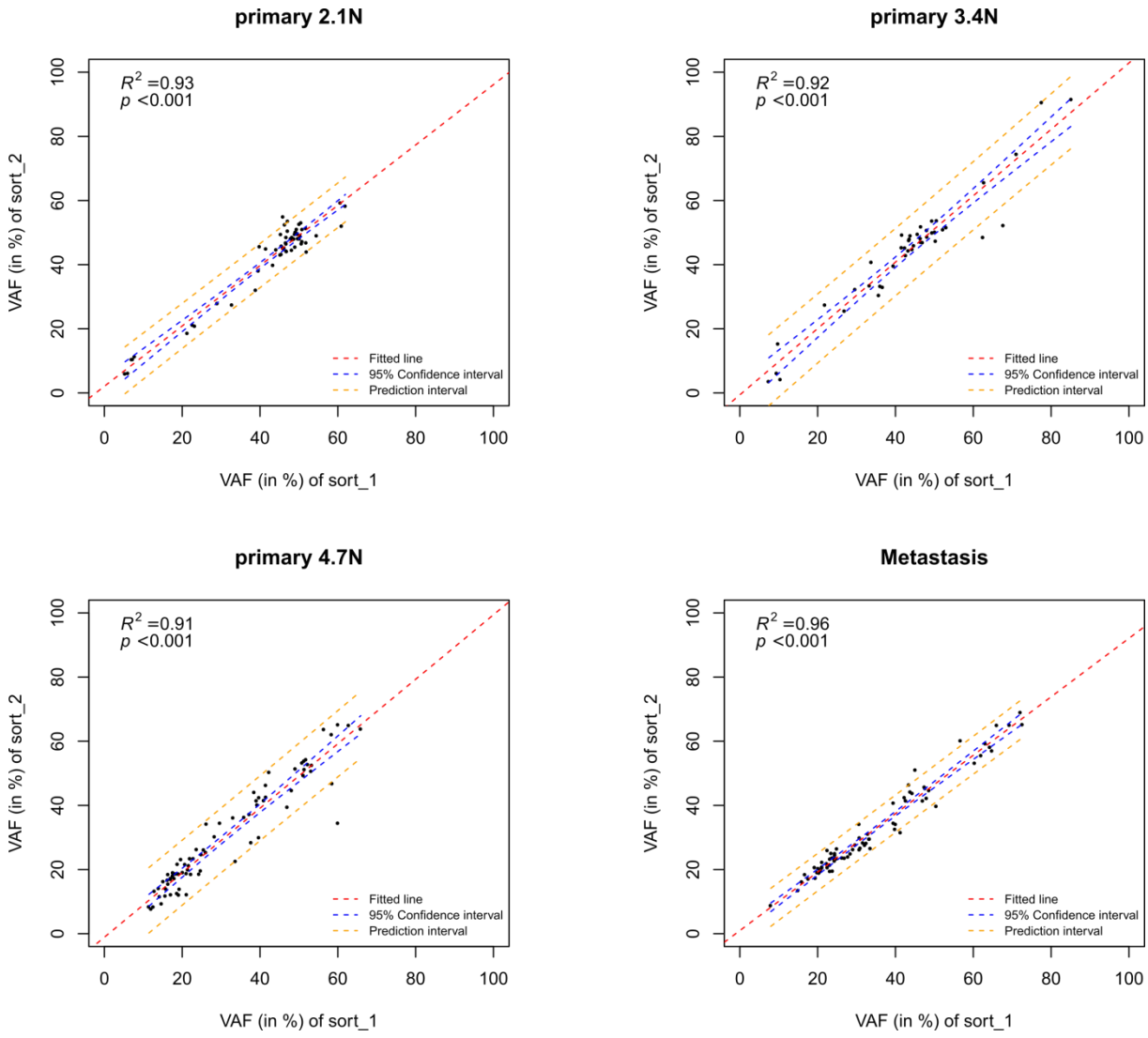
Supplementary Figure S23. Hierarchical clustering of Euclidean distances based on amplifications and deletions of all sorted tumor populations. Tumors of individiual patients (P) clustered closer together than to tumors of other patients. The only exception was the diploid tumor of Patient 42 (P42:N0207A), which clustered together with the other diploid tumors of other patients.



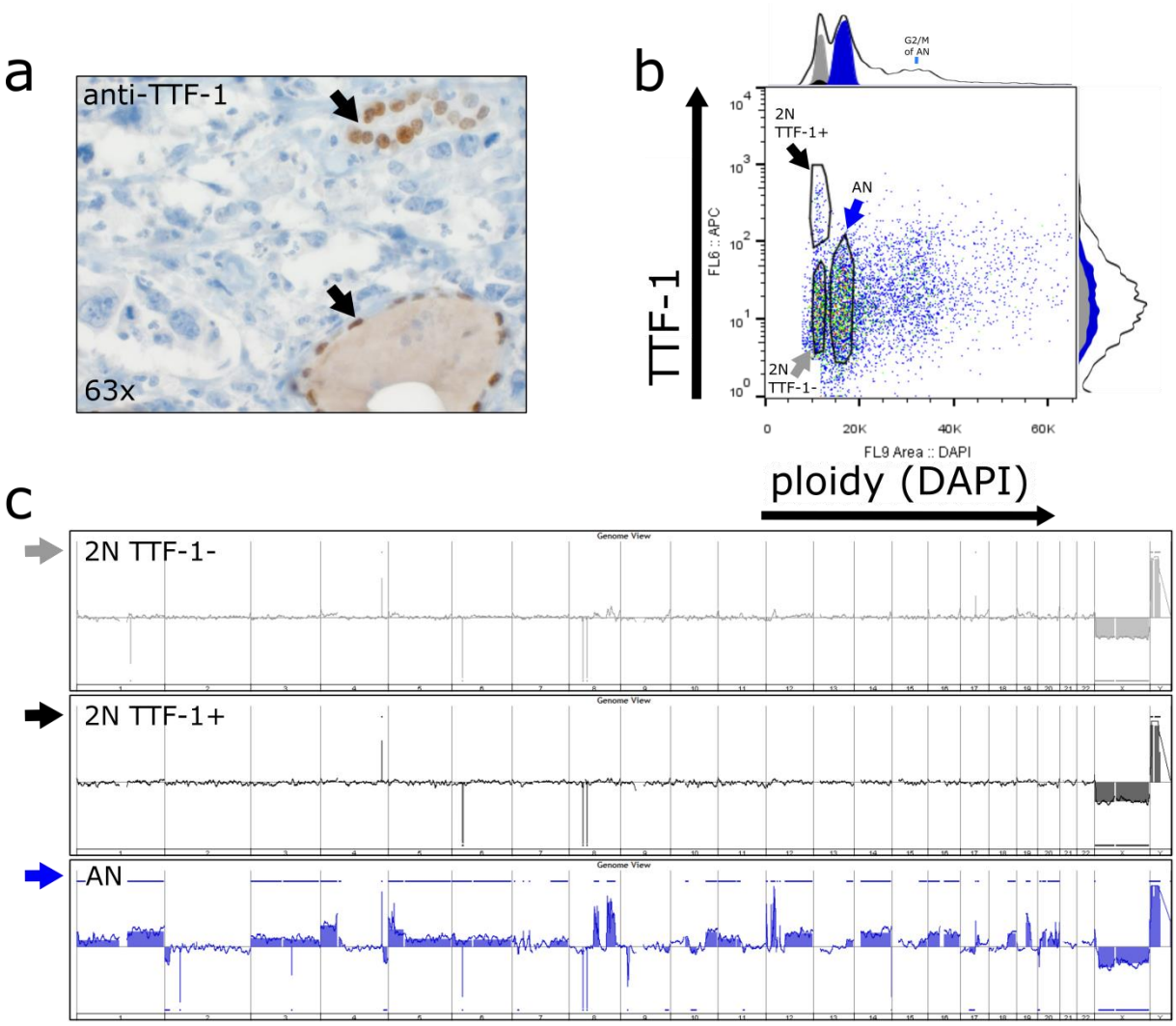
Supplementary Figure S24. Circos plot displaying the relative number of the genome affected by SCNAs in each sorted tumor population. The diploid tumor populations (red asterisks) had the highest amount of copy number neutral (absence of SCNAs) regions.



Supplementary Figure S25. Scatterplots displaying the VAFs of sequencing replicates from sorted tumor populations of the primary tumor and the metastasis of Patient 42.



Supplementary Figure S26. Scatterplots showing the VAFs of sorting replicates from tumor populations of the primary tumor and the metastasis of Patient 42.



Supplementary Figure S27. TTF-1 expression in normal cells. A: TTF-1 can be expressed in normal lung alveolar epithelial cells or as shown here in thyroid follicular cells in this case of a thyroid gland metastasis of Patient 12. When sorting with TTF-1 it is therefore necessary to check by IHC, if there are diploid, TTF-1-positive (2N TTF- 1 +) non-tumor cells within the biopsy. **B:** Three populations were sorted from this thyroid gland metastasis: a diploid, TTF-1-negative (2N TTF-1-); a diploid, TTF-1-positive (2N TTF-1+); and an aneuploid (AN) population. **C:** Array-CGH analysis validated that no SCNAs were present in the 2N TTF-1- and TTF-1+ populations, whereas the aneuploid tumor population had multiple distinct SCNAs, validating that the 2N TTF-1-positive population is thyroid follicular cells and that the tumor is aneuploid and TTF-1-negative.

Exploring the spatiotemporal genetic heterogeneity in metastatic lung adenocarcinoma using a nuclei flow-sorting approach

Lorber T, *et al*. *J Pathol* 2018 (DOI: 10.1002/path.5183)

Supplementary materials and methods

Reference numbers refer to the main text reference list

Patients and tissue samples

The final cohort consisted of 16 patients with primary LUAD and a total of 21 clonally related metastases. This comprised nine patients (56%) with distant metastases, including brain (*n* = 4), adrenal gland (*n* = 3), kidney (*n* = 1), and thyroid gland (*n* = 1) metastases, and seven patients (44%) with locoregional metastases, including intrapulmonary metastases (*n* = 4) and pleural effusions (*n* = 3). One of these patients (patient 22) presented with both a locoregional intrapulmonary metastasis and a brain metastasis. More than one metastasis was available for four (25%) of these 16 patients (patients with oligometastatic disease). Additional fresh-frozen or formalin-fixed, paraffin-embedded (FFPE) tumor samples of patients meeting these inclusion criteria were investigated, if available. Three of the 19 initial patients were excluded from this study of clonally related LUAD because their suspected locoregional metastases were genomically unrelated occurrences of second primary LUAD, as they did not share a single SCNA breakpoint or mutation with the primary tumors (supplementary material, Table S1 and Figures S1 and S2).

We excluded possible inter-patient confusion of samples by hierarchical clustering of Euclidean distances based on SCNAs (supplementary material, Figure S23). We found that matched tumors from individual patients were more similar to each other than to the tumors of other patients. One exception was the diploid primary tumor population (2.0*N*) of patient 42, which clustered together with the diploid tumor populations of the other three patients due to the overall large number of copy number neutral regions of all diploid tumors in this cohort (supplementary material, Figure S24).

Isolation of nuclei

Nuclei were isolated according to published protocols [14,15]. In brief, fresh-frozen tumors were minced in ice-cold NIM buffer [10 mM Tris–HCl (pH 7.4) containing 146 mM NaCl, 22 mM $MgCl_2$, 2 mM $CaCl_2$, 10% (v/v) DMSO, 0.05% (w/v) BSA, and 0.1% (v/v) IGEPAL CA630]. Nuclei were pelleted in a cooled centrifuge (300 × *g*, 3 min, 4°C). The nuclei pellet was resuspended in ice-cold NIM buffer and stored on ice for subsequent flow cytometric analysis for up to 8 h or at −20°C for longer storage up to 6 months.

For FFPE tissues, nuclei isolation was performed as described previously [16,41]. In brief, histological thick sections cut at 55 µm were deparaffinized in xylene and rehydrated in subsequent washings of 100%, 95%, 70%, 50%, and 30% (v/v) ethanol. Antigen retrieval was performed in citrate buffer (pH 6.0) in a heat block at 95°C for 30 min. Nuclei were obtained after digestion with collagenase III (final: 50 units/ml in 0.1 mM $CaCl_2$/PBS) in a shaking heat block (16 h, 37°C, 1000 rpm). After the nuclei were pelleted at 500 × *g* for 5 min, they were resuspended in FACS buffer [PBS (pH 7.4) containing 1 mM EDTA, 1% (v/v) FCS, and 0.1% (v/v) IGEPAL CA-630] to stop the digestion. After subsequent washings, isolated nuclei were stored in FACS buffer at 4°C for up to 1 week or at −20°C for up to 6 months.

Flow sorting of nuclei

FACS staining was performed in NIM buffer for nuclei from fresh-frozen tissues (n = 35) or in FACS buffer for nuclei derived from FFPE sections (n = 2). The composition of buffers may be found in the preceding section.

TTF-1-negative tumors

Nuclei of TTF-1-negative tumors were stained with 5 µg/ml DAPI at least 30 min before flow sorting was performed on a BD Influx™ cell sorter (Becton-Dickinson, San José, CA, USA). Doublets were discriminated and excluded from a two-parameter (DAPI-H/DAPI-A) plot, and fragmented nuclei and debris [DAPI-negative, left to the first (G_1) peak] were removed. All populations visible from a two-parameter scatterplot of side scatter (SSC) versus DAPI area (DAPI-A) were sorted into separate tubes and checked for the presence of copy number aberrations or mutations by subsequent analysis with aCGH and NGS, respectively.

TTF-1-positive tumors

Nuclei of TTF-1-positive tumors were stained with an anti-TTF-1 antibody (clone SPT24;

Novocastra™, Leica Biosystems Nussloch GmbH, Nussloch, Germany; product: NCL-LTTF1) for 4 h on ice, washed twice, and incubated for 30 min with an anti-mouse IgG-Alexa 647 secondary antibody [F(ab')2-goat anti-mouse IgG (H+L) secondary antibody, Alexa Fluor® 647 conjugate; Thermo Fisher Scientific, Waltham, MA, USA; product: A-21237]. To verify TTF-1-specific staining, an IgG_1 isotype control was used as a primary antibody control (mouse IgG1 isotype control; R&D Systems, Minneapolis MN, USA; product: MAB002). Concentrations of primary and secondary antibodies were 5 and 2 µg/ml, respectively. Stained nuclei were filtered through a 40 µm mesh, and DAPI was added to a final concentration of 5 µg/ml at least 30 min before flow sorting was performed on a BD Influx™ cell sorter (Becton-Dickinson). A gating strategy was applied to ensure single nuclei (DAPI-H/DAPI-A) and to eliminate damaged or fragmented nuclei [(DAPI-negative, left to the first (G_1) peak]. All populations visible from a two-parameter scatterplot of TTF-1 versus DAPI-A were sorted into separate tubes and checked for the presence of copy number aberrations or mutations by subsequent analysis with aCGH and NGS, respectively. The ploidy of each tumor population was calculated from the ratio of the geometric means of the DAPI signal in the tumor population versus their corresponding 'diploid, CNA-neutral' population. Flow cytometric analysis and visualization were performed with FlowJo X version 10.0.7 (FlowJo LLC, Ashland, OR, USA). All centrifugation steps were performed at 300 × *g* for 3 min in a cooled centrifuge (4°C).

DNA extraction and quantification

Genomic DNA was extracted with the Maxwell® 16 FFPE Plus LEV DNA Purification Kit (AS1135; Promega, Madison, WI, USA) and quantified with the Qubit® dsDNA HS Assay Kit (Invitrogen™, Carlsbad, CA, USA) following the manufacturer's instructions.

Whole genome amplification

DNA of sorted populations from fresh-frozen tissues was WGA with the illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare Life Sciences, Little Chalfont, Buckinghamshire, UK; product: 25660031) following the manufacturer's instructions with minor modifications in order to optimize the quality of amplified DNA and to reduce potential amplification biases for aCGH as suggested previously [42,43]: (i) 50–100 ng was used; (ii) equal amounts of test and reference DNA (46,XX; Promega, G1521) were amplified; and (iii) WGA was performed for

only 1 h, because shorter reaction times with Φ29 were shown to significantly decrease amplification biases for aCGH analyses. This whole genome amplification is based on an isothermal multiple displacement DNA amplification with the Φ29 enzyme, a high-fidelity DNA polymerase with proofreading (3'-5' exonuclease) activity. Amplified DNA was used for both aCGH and NGS. In brief, 50–100 ng of DNA from sorted populations or/and matching amounts of female reference DNA (46,XX; Promega, G1521) were SpeedVac™-ed to 1 µl and mixed with 9 µl of sample buffer. The mixture was denatured at 95°C for 3 min and immediately cooled on ice for at least 5 min. A mix of 9 µl of reaction buffer with 1 µl of the enzyme was added to the sample. Amplification was performed at 30°C for 1 h in a thermal cycler, followed by inactivation for 10 min at 75°C. Nuclease-free water was used as a negative control. Amplifications resulted in a total DNA quantity of around 4.5–5.5 µg (225– 275 ng/µl) and no DNA product in the negative water control. High-molecular-weight genomic DNA (average ≥ 10 kb) was ensured by gel electrophoresis with the FlashGel™ DNA System (Lonza, Basel, Switzerland). Amplified DNA was quantified with the Qubit® dsDNA HS Assay Kit (Invitrogen™). DNA from sorted FFPE tissues was not amplified to circumvent further DNA degradation.

Array comparative genomic hybridization

Digestion

Amplified DNA of sorted populations from fresh-frozen tissues was digested with DNaseI to reach a DNA smear from 100 to 1000 bp. Therefore, 1 µg of both amplified test and reference DNA was diluted with nuclease free water to a total volume of 7µl and mixed with

1 µl of 10´ reaction buffer [200 mM Tris–HCl (pH 7.5), 100 mM $MgCl_2$, 10 mM $CaCl_2$]. A DNaseI dilution was prepared by combining 1.25 µl of DNaseI (5000 units/ml; Thermo Fisher Scientific; Cat. # 89835) in 500 µl of 1´ reaction buffer. Two microliters of this DNaseI dilution was added to the amplified DNA to reach a total volume of 10 µl. DNA was digested in a thermal cycler for 15–18 min at 30°C, followed by DNaseI inactivation at 75°C for 10 min. Degraded DNA with a molecular weight of 100–1000 bp was assessed by gel electrophoresis using the FlashGel™ DNA System (Lonza).

DNA of sorted populations from FFPE tissues (n = 2) was not amplified and not digested. In these cases, 600 ng of unamplified reference DNA was used and digested as described above. Labelling, filtering, and hybridization procedures were not different between DNA from fresh-frozen and FFPE tissues.

Labeling and filtering

Digested test and reference DNA was labeled with Cy3-dUTP and Cy5-dUTP, respectively, using the BioPrime® Array CGH Genomic Labeling System (Thermo Fisher Scientific). Therefore, degraded DNA was incubated with 2.5´ Random Primers in a total volume of 40 µl at 95°C for 5 min, followed by the addition of 9 µl of Labeling Master Mix (5 µl of 10´ dUTP, 1 µl of Exo-Klenow, and 3 µl of Cy3-UTP or Cy5-dUTP). Labeling was performed in a thermal cycler at 37°C and stopped after 2 h by addition of 5 µl of EDTA. Labeled DNA was filtered with the Amicon Ultra 30K Filter Units (Millipore, Cat. # UFC503096). To do so, labelled DNA was mixed with 450 µl of TE buffer (10 mM Tris–HCl, 1 mM EDTA, pH 8.0) and centrifuged at 8000 ´ *g* for 10 min, followed by one wash with 500 µl of TE buffer. Finally, the DNA was eluted and successful labeling was verified by measuring the specific activity
(> 30) of each sample (pmol dye/µg DNA) using the Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific).

Hybridization, washing, and scanning

Labelled test and reference DNA was pooled and adjusted with TE buffer to a volume of 39 µl. 71 µl of Hybridization Master Mix (5 µl of Cot-1 DNA, 11 µl of Agilent 10´ Blocking Agent, 55 µl of 2´ Agilent Hybridization Buffer) was added to each test/reference sample and incubated at 98°C for 3 min, followed by another incubation at 37°C for 30 min. Subsequently, each sample was hybridized on 180K SurePrint G3 Human CGH Microarrays (Agilent Technologies, Santa Clara, CA, USA) for 24 h at 67°C in a rotating hybridization chamber. All microarray slides were washed according to the manufacturer's instructions and scanned with the Agilent 2565C DNA scanner. Images were analyzed with Agilent's Feature Extraction v10.7 using default settings. Feature-extracted aCGH data were evaluated using Agilent's CytoGenomics v3.0.1.1 software. Recentralization was performed by setting ploidy within ± 0.5 of the prior ploidy estimate, as calculated from DNA-based flow cytometry. Aberrations were called with the aberration detection algorithm ADM2 set to a threshold of 12.0, with Fuzzy Zero and GC-content (window size: 2 kb) correction. The aberration data table and raw probe $\log_2$ ratios were exported and further processed in R version 3.2.3 with the Bioconductor package 'copynumber' version 3.4 [44].

Sequencing

Controls for the detection of somatic mutations

A common difficulty associated with retrospective studies is the absence of germline controls, such as blood or other tumor-free material. While not a concern for the detection of SCNAs, due to the normal diploid state of benign cells, germline controls are important for the detection of somatic mutations by sequencing. Our flow-sorting approach allowed us to sort not only tumor populations but also 'diploid, CNA-neutral' populations, i.e. diploid populations without SCNAs. Diploid populations in TTF-1-negative tumors that were sorted with DAPI only and diploid, TTF-1-negative populations in otherwise TTF-1-positive tumors were evaluated for the presence of copy number aberrations by aCGH using the same procedure as described in the Materials and methods. All diploid or diploid, TTF-1-negative populations were 'CNA-neutral', i.e. no SCNAs were detected. We are convinced that these 'diploid, CNA-neutral' populations represented the tumor stroma because (i) they were diploid and had no SCNAs by definition; (ii) the tumor stroma is part of the bulk tumor mass and is expected to remain diploid and CNA-neutral; (iii) we validated the use of 'diploid, CNA-neutral' populations as germline substitutes in three patients, where non-tumor tissue was available (supplementary material, Table S4); and (iv) mutations that were detected in the tumor populations had low VAFs in the 'diploid, CNA-neutral' populations, which likely represent contamination with minute numbers of tumor nuclei during sorting rather than a cancerous subclone given that sorting is never 100% pure; that most tumors were aneuploid and tiny impurities with aneuploid tumor nuclei result in disproportionate increases of the VAFs in 'diploid, CNA-neutral' populations; and that in the three patients with genomically unrelated tumors these minute variant alleles were absent in the 'diploid, CNA-neutral' populations of the other tumor of the same patient, respectively (supplementary material, Table S1). Therefore, we feel confident to use the DNA of 'diploid, CNA-neutral' populations as germline controls in our cancer gene sequencing analysis pipeline.

Library preparation and sequencing

All libraries were performed according to the manufacturer's (Ion AmpliSeq™ Library Kit 2.0; Thermo Fisher Scientific; product: 4475345) instructions [45]. Template preparation and chip loading were carried out with the Ion Torrent™ Ion Chef™. CCP and whole exome sequencing

were carried out with Ion PI™ Chips on the Ion Proton™ Sequencer. Targeted resequencing with the custom panel was performed with 316™ Chips v2 on the Ion Torrent™ Ion S5™ system, which can sequence 400 bp libraries. Quality checks of all libraries were performed for both the expected library size (Agilent's High Sensitivity DNA Analysis Kit on a Bioanalyzer) and the library concentration (quantitative real-time PCR with the Ion Library Quantitation Kit). Sequence alignment to target regions of the hg19 genome was performed with the IonTorrent TorrentSuite™ software.

Variant calling and filtering

CCP panel

Variant calling was performed with the Torrent Variant Caller 5.0 plugin from the TorrentSuite (Thermo Fisher Scientific) using low stringency settings (Somatic–Proton–Low Stringency), as suggested by the manufacturer. Details are listed in the supplementary material, Table S8. Depth of coverage, coverage uniformity, and read length are summarized in the supplementary material, Table S9. Variants with a Phred-scored quality of ≥ 50 and a strand bias < 0.95 were considered but had to meet also the following thresholds: minimum coverage at variant site ≥ 10; minimum VAF ≥ 0.04; homopolymer ≤ 10; number of variant reads ≥ 5; common signal shift ≤ 0.2; and $VAF_{CNA\text{-}neutral} = 0$ or $VAF_{tumor}/VAF_{CNA\text{-}neutral} \geq 4$. We applied this last step of filtering because we used 'diploid, CNA-neutral' populations as germline controls and minute amounts of tumor DNA are inevitable during flow sorting. All variants were annotated using ANNOVAR [46].

Whole exome-sequenced tumors from patient 42

We performed a stringent variant calling and filtering for this patient's whole exomesequenced tumors, as our focus was to investigate the genomic relationships between tumor populations. Therefore, a smaller number of high confident and validated mutations were prioritized over a large number of mutations. Variant calling was performed with both the Torrent Variant Caller 5.0 plugin from the TorrentSuite and the 'somatic' tools from VarScan2 v2.3.9, as described below. To achieve a high stringency, the intersection (variants called by both algorithms) was used for further ultra-deep resequencing with an Ion Ampliseq™ custom validation panel. The intersection rather than the union of the two variant calling algorithms was used, as high confidence variants were favored over the number of

variants in respect to evolutionary analysis. Sequence alignment to target regions from the hg19 genome was performed with the Ion Torrent TorrentSuite™ software. Variant calling was performed with the Torrent Variant

Caller 5.0 plugin from the TorrentSuite using low stringency settings (Somatic–Proton–Low Stringency). A separate variant calling was performed between tumor and matched germline using the 'somatic' command from VarScan2 v2.3.9 [47].

The input for VarScan2 was the SAMtools [48] mpileup output from the combined tumor/normal samples. The 'somatic' command from VarScan2 was applied with default settings, except for the following settings: minimum coverage (--min-coverage-normal/tumor 10) was set to 10 for both tumor and normal sample. The minimum variant frequency (--minvar-freq 0.05) was set to 4%. Both tumor and normal purity (--normal/tumor-purity 0.85) were applied with a value of 85%. Additionally, variants with > 90% support on one strand only (-strand-filter 1) and with less than five supporting reads (--min-reads2 5) were filtered out.

The same settings were applied for filtering the variants identified by the Ion Torrent Variant Caller. This included variants with a Phred-scored based quality of ≥ 50, strand bias < 0.90, coverage ≥ 10, variant allele coverage ≥ 5, VAF ≥ 0.04, and variants with a common signal shift ≤ 0.2. Of these, variants were considered present if the ratio of $VAF_{tumor}/VAF_{germline}$ >= 4 or if the $VAF_{germline}$ was 0.

Finally, for the purpose of this analysis, small insertions and deletions were not evaluated and variants that were called and not filtered out by both approaches (intersection of both Variant Caller and VarScan2) were further curated by manual inspection using the Integrated Genomics Viewer IGV [49]. The variants (n = 112) passing all of these criteria were eventually used for further validation by ultra-deep resequencing (mean coverage
5864X) with an Ion Ampliseq™ custom validation panel. All variants were annotated using ANNOVAR [46].

Ion Ampliseq™ custom validation panel for patient 42

In total, 16 of the 112 mutations were either absent in all tumor regions or identified as germline variants (overall validation rate 85.7%). The mean sample specific validation rate was 90.4% (range 87.5–93.8%). For accuracy, sorted replicates (this comprises resorting of both the primary tumor and the metastasis, DNA extraction, and WGA of all sorted populations) were subjected to resequencing with this custom panel. Technical replicates (in duplicates),

starting from the library preparation step with individual barcoding, were performed for further certainty of the VAFs of the detected mutations. Regarding the value of the VAFs, we performed a linear regression analysis. Here, the mean squared correlation coefficient R^2 was 0.93 (range 0.90–0.95; supplementary material, Figure S25) for intra-sort comparison and 0.93 (range 0.91–0.96; supplementary material, Figure S26) for inter-sort comparison, including all tumor populations from the primary tumor (n = 3) and the metastasis (n = 1).

Functional effect prediction and classification of mutations

Combined scoring of five functional protein prediction algorithms was applied to non-silent mutations to distinguish pathogenic mutations from non-pathogenic mutations. All identified non-silent mutations were classified into five categories, based on the functional predictors SIFT [50], Polyphen2 [51], MutationTaster [52], FATHMM [53], and Provean [54]. Binary scoring of these predictors was performed, where 1 stands for a 'deleterious' and 0 for a 'non-deleterious' mutation.

To evaluate the level of confidence (C) of calling a mutation 'pathogenic', the predictions of all five predictor algorithms were weighted as follows:

$$C = \frac{1}{n}\sum_{i=1}^{n} X_i \ , \ X_i = \{0,1\}$$

where n is the number of predictors with available information for a specific mutation and X_i is the binary prediction of a predictor algorithm, with 0 being a 'benign' and 1 being a 'pathogenic' mutation. The impact of the mutation on the protein function was then divided into five categories, depending on the value of C: category 1: 'confidently deleterious' (C = 1); category 2: 'deleterious' ($0.66 \leq C < 1$); category 3: 'possibly deleterious' ($0.5 \leq C < 0.66$); category 4: 'possibly non-deleterious' ($0.2 \leq C < 0.5$); and category 5: 'confidently non deleterious' ($C < 0.2$).

Mutations in categories 1–3 as well as nonsense (stop-gain/truncating), splice-site, and frameshift mutations were considered 'pathogenic'. Additionally, mutations in categories 4 and 5 were considered 'pathogenic' if these were predicted to be 'driver' alterations by

CHASM [55] (Lung-Adenocarcinoma; available online under http://www.cravat.us/CRAVAT/) or 'passengers' otherwise. A default threshold of $p \leq 0.05$ [*P* value (missense)] was set for CHASM cancer driver mutations.

Interlesional genetic heterogeneity

The genetic divergence was calculated for each patient comparing the number of SCNAs and mutations that differ between two samples versus the number of SCNAs and mutations that were present in both. The mean pairwise divergence (D) was defined as

$$D = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} 0 \sum_{j=i+1}^{n} \frac{X_{i,j}}{Y_{i,j}}$$

where n is the number of samples in a patient, $X_{i,j}$ is the number of mutations and SCNAs private to either sample i or j but not present in both, and $Y_{i,j}$ is the number of mutations and SCNAs present in sample i or j.

Purity estimation and characterization of intra-tumor genetic heterogeneity

We used EXPANDS version 1.7.2 with default parameters to infer the purity of all 39 sorted tumor populations. Mutations and copy number segments were used to predict the number and size of clones that co-existed in each tumor sample. Mutations that could not be explained by a clone present in 10% or more of the sample, at ploidy six (6*N*) or less, were excluded. The size of the largest clone detected is a direct indicator for purity [27,56]. Due to the low number of somatic mutations in most cases, we used EXPANDS for purity estimations only. Purity estimations were unsuccessful for samples with insufficient numbers of SCNAs ($n = 4$).

Phylogenetic relationships between the four primary and metastatic tumor populations of patient 42 were calculated using the neighbor-joining algorithm provided by the R-package 'ape' v4.0 [57]. Hereby, pairwise distances between populations were calculated as the number of genomic alterations for which both populations have the same status, i.e. present or absent, divided by the total number of genomic alterations for which both populations have available information. Two classes of genomic alterations were included: SCNAs and mutations.

Immunohistochemistry

All immunohistochemical staining (IHC) was performed on FFPE tissue sections using an automated immunostainer (Benchmark XT; Ventana, Tucson, AZ, USA) following the manufacturer's guidelines. IHC for TTF-1 was performed to validate homogeneous staining in TTF-1 positive tumors (10/16 patients, 62.5%) prior to flow sorting. In addition, TTF-1 can be expressed in normal lung alveolar epithelial cells or in thyroid follicular cells; for example, we found TTF-1-positive thyroid follicular cells in an otherwise TTF-1-negative thyroid gland metastasis of patient 12 (supplementary material, Figure S27). Information about the presence of TTF-1-positive normal cells was used for validation of 'diploid, CNA-neutral' populations that are TTF-1-positive and of non-tumor origin. IHC for p16, one of the proteins encoded by *CDKN2A*, was performed on FFPE tissue sections of the primary tumor and the metastasis of patient 42.

2.2 Part B: Genomic evolutionary trajectory of metastatic squamous cell carcinoma of the lung

Genomic evolutionary trajectory of metastatic squamous cell carcinoma of the lung

Arthur Krause[1*], Luca Roma[1*], Thomas Lorber[1], Tanja Dietsche[1], Valeria Perrina[1], David C. Müller[1], Didier Lardinois[2], Christian Ruiz[1], Spasenija Savic Prince[1], Salvatore Piscuoglio[1,3], Charlotte KY Ng[4], Lukas Bubendorf[1]

1 Pathology, Institute of Medical Genetics and Pathology, University Hospital Basel, University of Basel, Switzerland;

2 Thoracic Surgery, University Hospital Basel, Basel, Switzerland;

3 Visceral surgery research laboratory, Department of Biomedicine; University of Basel, Basel, Switzerland;

4 Department for BioMedical Research (DBMR), University of Bern, Bern, Switzerland;

*Co-first author, contributed equally to the work

Corresponding author:

Prof. Lukas Bubendorf,

Pathology, Institute of Medical Genetics and Pathology,

University Hospital Basel,

4031 Basel, Switzerland

Tel: +41 61 265 2851

Email: Lukas.Bubendorf@usb.ch

Keywords: squamous cell carcinoma, lung cancer, metastasis, heterogeneity, tumor evolution, flow sorting, whole exome sequencing

Abstract

The extent of inter- and intratumoral genomic heterogeneity and the clonal evolution of metastatic squamous cell carcinoma of the lung (LUSC) are poorly understood. Genomic studies of LUSC are challenged by their low tumor cell content. We sought to define the genomic landscape and evolutionary trajectories of metastatic LUSC combining nuclei-flow sorting and whole exome sequencing. Tumor nuclei were sorted based on ploidy and expression of cytokeratins to enrich for tumor cells for whole exome sequencing. We investigated six primary tumors and seven matched metastases from six LUSC patients. Flow-sorting increased the mean tumor purity from 33% (range 12%-86%) to 70% (range 35%-93%). Overall, primary LUSC and their matched metastases shared more than 76% of copy number variations (CNVs) and 64% of somatic mutations in cancer genes, including mutations in common tumor suppressor genes such as *TP53.* Furthermore, the ploidy of the tumors remained unchanged between primary and metastasis in the majority of samples (5/6 patients) over time. We found differences in the mutational signatures of shared mutations compared to the private mutations in the primary or metastasis. Our results demonstrate a close genomic relationship between primary tumors and their matched metastases in this cohort of LUSC, suggesting late dissemination of the metastases from the primary tumors during tumor evolution.

Introduction

Distant metastases are more common in lung adenocarcinoma (LUAD) than in lung squamous cell carcinoma (LUSC). This is one reason why metastatic LUSC has been understudied [1]. Most published studies on metastatic LUSC are single case reports without genomic analysis of the metastasis due to a lack of sufficient material [2]. Advanced LUSC tends to affect older patients with more comorbidities than LUAD, which makes tissue sampling for research studies challenging [3]. Therefore, the mechanisms and evolutionary trajectories of metastasis in LUSC are not well understood.

Multi-region sampling and comprehensive molecular profiling of resected tumors have greatly advanced our understanding of intratumoral heterogeneity (ITH) in lung and other cancer types [4]. However, the high costs and the diagnostic reality of mostly small biopsy specimens from patients with advanced LUSC make such multi-region analyses difficult if not impossible in clinical practice. Molecular profiling of LUSC is further challenged by the often low tumor cell content of LUSC tissue with a median tumor cellularity around 60% [5]. To overcome this hurdle, we previously showed that the tumor purity of LUAD can be greatly enhanced by an advanced nuclei flow-sorting technique [6]. Briefly, cells were stained with 4',6-Diamin-2-phenylindol (DAPI) and an antibody for TTF-1 to isolate LUAD specific cells. This flow-sorting approach allowed us to distinguish and isolate aneuploid and diploid populations enabling highly informative genomic profiling [6].

Here, we applied a modified nuclei flow-sorting approach to enrich for cytokeratin-labelled tumor nuclei from tumor tissue specimens. An antibody was used for labelling cytokeratins that are highly expressed in LUSC but not in other cells. Whole exome sequencing (WES) was performed to investigate ITH and clonal evolution from six patients with LUSC and matched spatiotemporal metastases.

Materials and methods

The retrospective study comprised the following criteria for the inclusion of patients: (1) histologically confirmed LUSC with regional or distant metastases; (2) large enough specimens (i.e. surgical specimens or large biopsies) to allow for nuclei flow-sorting and WES; (3) at least two available tumor samples, fresh-frozen (FF) or formalin-fixed and paraffin-embedded (FFPE), one of which had to be the primary tumor (PT); (4) tumor samples that had been taken from different time points or body sites of the patient. Multiple regions within the primary tumor were not sampled. After WES, no shared mutations were found between the primary tumor and the presumed intrapulmonary metastasis from 4 patients. They were considered independent tumors and excluded from this study. Six patients with primary LUSC and seven matched metastases were investigated. Tumor samples, FF (n = 7) or FFPE (n = 6, Supplementary Table 1), were obtained from the archive of the Institute of Medical Genetics and Pathology of the University Hospital Basel, Switzerland. Histological slides were reviewed by a pulmonary pathologist (LB). The areas with highest tumor cell density were marked for subsequent dissection, avoiding areas rich in stromal cells or benign epithelial cells from entrapped bronchoalveolar tissue. The study was approved by the local Ethics Committee (EKBB/EKNZ 31/12).

Nuclei isolation and multiparameter flow-sorting

Nuclei isolation from FF and FFPE samples was conducted as described previously [6–9]. All nuclei were DAPI stained for sorting, DNA quantification and ploidy analysis. FF tumor nuclei were only sorted with DAPI (*n* = 7, four patients). Pan-Cytokeratin (pCK, clone MNF116, Dako, Code-Nr. M0821) was used for FFPE tumors as an additional marker for flow-sorting. The applied pCK antibody is specific against CK5, CK6, CK8, CK17 and CK19 and highly expressed in LUSC [10,11]. Cytokeratin positive cytoplasmic remnants seem to stay attached to the nuclei of FFPE tissue but are washed away in FF tissue during the procedure. This enabled refined separation of diploid/aneuploid tumor populations and benign diploid epithelial and stromal cell nuclei in FFPE tissue (*n* = 6, three patients, Supplementary Figure 1). Previously, we described the utility of sorted diploid population as germline [6]. In DAPI-sorted aneuploid tumors, the matched diploid populations were used as germline for WES (P6, P9, P18, Supplementary Figure 2). In samples sorted by DAPI and pCK, diploid pCK-negative populations were

used as germline (P103, P109, P113) [6]. Initial purity was estimated after sorting by treating all sorted nuclei as 100% and calculating all sorted populations according to their proportion.

DNA extraction and whole exome sequencing

DNA from FFPE isolated nuclei was directly extracted and quantified, whereas DNA from FF samples was first whole genome amplified using the illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare Life Sciences, Little Chalfont, Buckinghamshire, UK; product: 25660031) as previously described [6,7]. In samples that were exclusively sorted with DAPI, the diploid population was used as germline, whereas diploid pCK-negative populations were considered as germline when applicable [6]. Prior sequencing, diploid DAPI sorted samples were subjected to array comparative hybridization (aCGH) as described in detail previously [6]. Flat genomes were considered as a germline. According to manufacturer's guidelines SureSelect Human All Exon V6 Kit (Agilent) was utilized for the whole exome capturing. Paired-end 100-bp reads were generated on the Illumina NovaSeq 6000. Sequencing was conducted by CeGaT (Tübingen, Germany).

Human genome GRCh37 was used as a reference for aligning the reads. Whole exome sequencing (WES) analysis, comprising sequence processing, calling of somatic single nucleotide variants (SNVs), small insertions and deletions (indels), and allele-specific copy number variations (CNVs), detecting mutational signatures and identifying clonality, was performed using state-of-the-art bioinformatic methods, as previously described with modifications [12,13].

Briefly, SNVs and indels were called using MuTect v1.1.7 and strelka v2.9.10, respectively [14,15]. All SNVs C:G>T:A with variant allelic fractions (VAFs) less than 10% were discarded to reduce false positive results. Otherwise, SNVs or indels with VAFs < 1% or that were supported by fewer than 3 reads were not considered. If SNVs and indels were found in one tumor of a patient, a cut-off of two reads was applied to the remaining tumor/s. Variant annotation was performed using SnpEff software v4.1 [16].

The heatmap of non-synonymous mutations was generated using the R package maftools v.2.0.16 by selecting the genes in the Bailey et al. dataset that represents the significantly mutated genes in LUSC obtained from TCGA [17].

Analysis of copy number aberrations and clonality

FACETS v0.5.14 enabled the identification of allele-specific CNVs [18]. Cancer cell fraction (CCF) of each mutation was inferred using ABSOLUTE v1.0.6 [19]. A mutation was classified as clonal, if its probability of being clonal was >50% or if the lower bound of the 95% confidence interval of its CCF was >90%. Mutations were regarded as subclonal if they did not meet the mentioned criteria [20,21]. A mutation that was found in both tumors of one patient was considered as ‘trunk’. Mutations that were detected in only one tumor of the patient were considered as ‘branch’ or ‘private’.

Phylogenetic analysis and mutational signatures

As described by Murugaesu et al, a maximum parsimony tree was built for each case using binary presence/absence matrices based on the repertoire of non-synonymous and synonymous somatic mutations, gene amplifications and homozygous deletions in the biopsies of the tumors [22,23].

Decomposition of mutational signatures was conducted using the R package deconstructSigs software by selecting the mutational signatures 1, 2, 4, 5 and 13, which are based on the set of 30 mutational signatures (‘signature.cosmic’) that were observed in LUSC [24,25].

Results

Nuclei flow-sorting achieves high purity of tumor DNA

In this study, we investigated six patients with primary LUSC and their metastases to explore ITH and clonal evolution (Table 1). The cohort was comprised of one female and five male patients, three patients were former or active smokers and for three the smoking status was unknown. Our study cohort consisted of seven FF and six FFPE tissue samples (Supplementary table 1). Although DAPI stained flow-sorting allowed separating aneuploid tumor nuclei from benign stromal tissue and epithelial cells, it does not detect diploid tumors [6,7]. Previously, we refined our technology to isolate diploid tumor cells using TTF-1 and DAPI staining in LUAD [6,7]. To enable genomic profiling of diploid tumor populations in LUSC, we implemented, in addition to DAPI, pCK staining to separately select diploid ($2N \pm 0.2$) and/or aneuploid ($> 2.2N$) tumor cell populations and remove pCK-negative diploid and near tetraploid benign cells. Across the six patients, we isolated nine aneuploid tumor cell populations, seven using only DAPI and two DAPI/pCK in five patients, and four diploid tumor cell populations by DAPI/pCK staining in two patients. PT of P109 was FF and sorted as an aneuploid population by DAPI, while both diploid metastases were FFPE samples and sorted by DAPI/pCK (Supplementary table 1). Our flow-sorting approach enabled us to increase the initial mean purity from 33% (range 12% - 86%) of unsorted tumor to a mean of 70% (range 35% - 93%) after sorting (Supplementary figure 3).

Mutational landscape of primary-metastatic pairs of LUSC

We performed WES of the sorted tumor cell nuclei at a median depth of 69x (range 38x - 99x, Supplementary table 1). Median mutation rate was 10.7 mutations/Mb, which is in line with a recent study on LUSC with a median of 9 mutations/Mb [26]. PT and metastases mutation rates were similar (primary LUSC median 9.5; range 4.6 - 13.6 vs. metastases median 11.1, range 8.7 - 18.2, Supplementary table 1). The six PT tended to have fewer mutations than metastases (285; range 138 - 408 vs. 333; range 92 – 407). A median of 79% (range 25% - 96%) of mutations in the primary LUSC were also found in the corresponding metastases (Supplementary table 2). In contrast, a median of 64% (range 17% - 94%) of the mutations found in the metastases were also found in the matched primaries. The most commonly mutated cancer genes included

TP53 (9 samples, 4 patients), *FAT1* (6, 3), *FBXW7* (5, 2), *KMT2D* (5, 2) and *ARID1A* (2, 2), which are known to be mutated in a proportion of LUSC (Figure 1a) [27–30].

PT and metastatic samples of LUSC shared common mutations known to drive tumorigenesis and tumor progression. According to phylogenetic analysis, clonal architecture of PT and matched metastases revealed that mutations in driver genes, such as the tumor suppressor genes *TP53*, *FBXW7* and *KMT2D*, were predominantly clonal (Figure 1b). Clonal mutations between the PT and metastases were found in each patient, exclusively P113. PT of P103 and metastases of P9, P109 and P113 displayed private cancer gene mutations. P113 did not share any mutations in the significantly mutated cancer genes (Figure 1a) but shared 42 non-synonymous mutation of other genes. Notably, P109 indicated a common clonal trunk with its two metastases, a lymph node (Met1) and a thoracic spine metastasis (Met2).

In five out of six patients, the trunks were much longer than branches. Only P113 had a short trunk and longer branches. P109 harbored a long mutational trunk and similar number of mutations in Met1. Private mutations of P109 PT and Met2 had a much shorter branch compared to Met1 (Figure 2).

Mutational signatures were inferred to track the evolutionary differences between the PT and the metastases in LUSC (Figure 2) [25]. The most common signatures were signature 1 (CpG deamination, related to aging) and signature 4 (related to smoking) being present in all samples across all patients. Signature 5 (transcriptional strand bias, related to aging) was present in every primary-metastasis pair but not in all trunks and branches. APOBEC signatures 2 or 13 were observed in five patients. It has been proposed that APOBEC primarily contributes to mutations that occur late in the evolutionary history where it increases subclonal diversity [4,31]. They were seen in the trunk of three patients and in six branches of four patients. The APOPEC signature was variably distributed. It was identified in trunk and both branches separately of PT and metastasis in P9. In contrast it was restricted to the trunk in P109 and to the branches in P18, and only found in the PT branch of P103. In P6, it was carried in the trunk and the PT branch but not in the branch of the metastasis.

Taken together, the mutational profiles between the PT and their metastasis are closely related as previously seen in LUAD [6]. The long trunks of mutations can be primarily

attributed to the effects of aging and smoking. The pattern of mutational signatures varied between patients and during evolution of individual tumors.

Chromosomal aberrations are truncal events

CNVs of the PT were highly concordant with those in matched metastases (Figure 3a, Supplementary figure 4). In primary LUSC, a median of 82% (range 83% - 99%) of CNVs were also observed in the matched metastases, while a median of 76% (range 69% - 99%) of the aberrations in the metastases were also seen in the matched primaries. All tumor samples, PT and metastases, shared amplified regions in 3q26.1-29. Amplifications at the 3q arm are well-known in LUSC [32]. Focal amplifications in 3q26.32 contain the oncogene *PIK3CA,* telomerase component *TERC* and squamous differentiation related genes such as *TP63* and *SOX2* [33]. *TERT* amplification was truncal in three of the six patients. Overexpression of telomerase due to *TERT* amplification leads to sustained cellular self-renewal [34]. Deletions of tumor suppressor genes are the most common CNVs in lung cancer [30]. We found the following tumor suppressor genes to be lost in our cohort: *ATM* (3 samples, 2 patients)*, ARID1A* (3, 2)*, APC* (7, 4)*, CDKN2A* (9, 4)*, CSMD1* (9, 4)*, FAT1* (3, 2)*, FOXP1* (8, 4), *PTEN* (10, 5), *SETD2* (8, 4) and amplifications were detected in *SOX2* (11, 6), *MECOM* (11, 6), *TERC* (11, 6), *MYC* (4, 2) and *CCND1* (4, 2).

Both chromosomal profiles and ploidy of PT and matched metastases were highly similar in most patients (Figure 3b, Supplementary figure 4). All PT were aneuploid, including PT of P109 (3.6 *N*). Interestingly, both metastases of P109 where diploid (2*N* ± 0.2*N*).

Taken together, CNVs and ploidy status were very similar in the PT and their metastases suggesting CNVs as early events with limited heterogeneity between PT and metastasis.

Discussion

In this study, we explored the spatiotemporal heterogeneity and evolution of LUSC by WES using highly enriched nuclei from six PT and matched metastases. Intratumoral heterogeneity (ITH) and clonal evolution of primary NSCLC using comprehensive genomic profiling strategies have been analyzed in several previous studies [4,35–37]. Although metastasis is the leading cause of progression and death in NSCLC, data on matched pairs of PT and metastases remain scarce [6,38,39]. To the best of our knowledge, this study is the first to investigate spatial and temporal genomic heterogeneity of metastatic LUSC across the disease course. We demonstrate high concordance between the primary-metastasis pairs in terms of SNVs and CNVs in most of the patients. As previously shown in LUAD, most aberrations in LUSC were truncal indicating only limited heterogeneity between PT and metastasis and low subclonal diversity [6]. Our results go in line with the data of a recent pan-cancer study on metastatic solid tumors that included treatment-naive and extensively treated tumors [40]. This study revealed that 96% of driver mutations within metastases were clonal with only 4% of subloncal mutations [40]. The potential clinical impact of subclonal diversity remains poorly studied. It is possible that previous treatment diminishes subclonal diversity by selection. This question could not be adequately addressed in our study since five of the six patients had not received systemic treatment prior to tissue sampling [35].

The high concordance between primary-metastasis pairs in our study corroborates our previous data on a cohort of LUAD [6]. In this previous study, more than 80% of CNVs and cancer gene mutations were shared between the primary-metastatic LUAD specimens. Ploidy remained stable and did not change over time [6]. Here, we found the CNVs and ploidy were equally stable during metastatic spread of LUSC. This stable aneuploidy in most LUSC and LUAD patients highlights the previously proposed critical importance of a whole genome doubling (WGD) event that might select for fitness to progress and metastasize [41,42]. The only discrepant ploidy was found in P109 revealing an aneuploid PT (3.6N) and two diploid/near-diploid metastases (2N and 1.8N, respectively). Reversion from an aneuploid to a diploid state during progression is unlikely, since WGD is considered irreversible [41]. Therefore, we hypothesize that the diploid metastases had emerged from a diploid population prior to WGD of the PT in this patient. We do not know whether the diploid population had been fully replaced by

the aneuploid population after WGD, or whether we have missed a coexistent diploid population due to limited sampling of the PT. Although most LUSC and LUAD are aneuploid and metastasize after WGD, this example illustrates that stably diploid tumor cell populations can occasionally serve as the backbone of progression and metastasis in NSCLC [6]. We have also previously observed this phenomenon in prostate cancer, where a diploid tumor cell population was consistently detectable in repeated biopsies during progression giving rise to temporary aneuploid populations [7].

The pathogenesis of the vast majority NSCLC is driven by the accumulation of smoking-induced alterations. The large proportion of signature 4 in our tumor samples likely reflects the contribution of smoking-related DNA damage for early acquisition and accumulation of truncal mutations as previously proposed [4]. Mutational signatures changed during metastasis and over time revealing no obvious pattern. Despite the common trunk, the mutational signatures differed in the private branches of the PT and metastases. Furthermore, the long trunk of shared genomic alterations in most samples indicates a linear progression model in the metastatic dissemination caused by one major subclone [40,43]. In fact, a common ancestor acquired genomic alterations and remained genetically stable even after late dissemination. Nevertheless, there were substantial differences between the private branches of PT and metastases. This mutational heterogeneity might be relevant in the context of specific treatments. For example, Richard and colleagues have recently reported that different mutational signatures had an impact on the efficacy of immune checkpoint inhibitors [44].

Our study provides new insight into the genomic trajectories of genomic evolution of metastatic LUSC but did not reveal novel specific alterations beyond those that have already been reported by large-scale sequencing consortia [45].

Among CNVs, it was most striking that all tumors shared amplifications at 3q26.1-29 (30-40% in cBioPortal [45]) indicating an important role of these CNVs for tumor initiation [32,46]. *SOX2* and *TP63* are among the genes in the amplified 3p26.1-29 region and associated with squamous differentiation [47]. Notably, SOX2 and p63 are known to be amplified in at least 40% and 30%, respectively, of all LUSC but almost never in LUAD [45]. It has recently been shown that squamous differentiation can be induced by inhibiting the adenocarcinoma lineage marker TTF-1 and simultaneously overexpressing *SOX2*. Moreover, it is known that *SOX2* downregulates TTF-1 by

suppressing the TTF-1 coding gene *NKX2-1*, emphasizing an important role of *SOX2* in squamous differentiation [48,49].

Our study has several limitations. The low number of six patients precludes strong general statements on the genomic trajectories of LUSC. Nevertheless, our collection of LUSC is unique being the largest series with matched PT and metastasis pairs of LUSC so far. Due to incomplete clinical documentation of archived cases, we were unable to reconstruct the full clinicopathological history including smoking history [50]. Further confounding factors are the different tissue types (FF vs. FFPE) and sorting strategies (DAPI vs. DAPI/pCK). Using WES, we were bound to coding regions preventing any observations on translocations. Furthermore, limited by the sequencing depth and filtering it is possible that subclonal variants were missed due to the detecting limits.

Taken together, our study provides new insights into the intratumoral heterogeneity and genomic patterns of primary LUSC and matched metastases using purified tumor populations after a modified nuclei flow-sorting technique. A close clonal relationship with only limited heterogeneity was revealed in the primary-metastatic pairs, similar to what was seen in LUAD. Further studies with larger matched cohorts are required to better elucidate the critical genomic players and related molecular pathways of metastatic disease in NSCLC.

References

1. Zhou, H. *et al.* Diagnosis of Distant Metastasis of Lung Cancer: Based on Clinical and Radiomic Features. *Transl. Oncol.* **11**, 31–36 (2018).
2. Guo, Y. *et al.* Lung squamous cell carcinoma with solitary ocular metastasis and its successful treatment with thoracic surgery and chemotherapy: an interesting and rare case report. *BMC Cancer* **18**, 1004 (2018).
3. Daaboul, N., Nicholas, G. & Laurie, S. A. Algorithm for the treatment of advanced or metastatic squamous non-small-cell lung cancer: an evidence-based overview. *Current Oncology* vol. 25 77 (2018).
4. Jamal-Hanjani, M. *et al.* Tracking the Evolution of Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **376**, 2109–2121 (2017).
5. Aran, D., Sirota, M. & Butte, A. J. Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* **6**, 8971 (2015).
6. Lorber, T. *et al.* Exploring the spatiotemporal genetic heterogeneity in metastatic lung adenocarcinoma using a nuclei flow-sorting approach. *J. Pathol.* **247**, 199–213 (2019).
7. Ruiz, C. *et al.* Advancing a clinically relevant perspective of the clonal nature of cancer. *Proceedings of the National Academy of Sciences* vol. 108 12054–12059 (2011).
8. Krishan, A. & Dandekar, P. D. DAPI fluorescence in nuclei isolated from tumors. *J. Histochem. Cytochem.* **53**, 1033–1036 (2005).
9. Corver, W. E. & ter Haar, N. T. High-resolution multiparameter DNA flow cytometry for the detection and sorting of tumor and stromal subpopulations from paraffin-embedded tissues. *Curr. Protoc. Cytom.* **Chapter 7**, Unit 7.37 (2011).
10. Blobel, G. A., Moll, R., Franke, W. W. & Vogt-Moykopf, I. Cytokeratins in normal lung and lung carcinomas. *Virchows Archiv B Cell Pathology Including Molecular Pathology* vol. 45 407–429 (1984).
11. Quinlan, R. A. *et al.* Patterns of expression and organization of cytokeratin intermediate filaments. *Ann. N. Y. Acad. Sci.* **455**, 282–306 (1985).
12. Nuciforo, S. *et al.* Organoid Models of Human Liver Cancers Derived from Tumor Needle Biopsies. *Cell Reports* vol. 24 1363–1376 (2018).
13. Bertucci, F. *et al.* Genomic characterization of metastatic breast cancers. *Nature* **569**, 560–564 (2019).
14. Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).

15. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
16. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
17. Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **174**, 1034–1035 (2018).
18. Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* **44**, e131 (2016).
19. Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nature Biotechnology* vol. 30 413–421 (2012).
20. Landau, D. A. *et al.* Evolution and Impact of Subclonal Mutations in Chronic Lymphocytic Leukemia. *Cell* vol. 152 714–726 (2013).
21. Guerini-Rocco, E. *et al.* Microglandular adenosis associated with triple-negative breast cancer is a neoplastic lesion of triple-negative phenotype harbouring TP53 somatic mutations. *J. Pathol.* **238**, 677–688 (2016).
22. Murugaesu, N. *et al.* Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. *Cancer Discov.* **5**, 821–831 (2015).
23. Ng, C. K. Y. *et al.* Genetic Heterogeneity in Therapy-Naïve Synchronous Primary Breast Cancers and Their Metastases. *Clin. Cancer Res.* **23**, 4402–4415 (2017).
24. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, 31 (2016).
25. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
26. Chalmers, Z. R. *et al.* Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Medicine* vol. 9 (2017).
27. Zhang, X.-C. *et al.* Comprehensive genomic and immunological characterization of Chinese non-small cell lung cancer patients. *Nat. Commun.* **10**, 1772 (2019).
28. Li, Q., Hou, J., Hu, Z., Gu, B. & Shi, Y. Multiple mutations of lung squamous cell carcinoma shared common mechanisms. *Oncotarget* **7**, 79629–79636 (2016).

29. Network, T. C. G. A. R. & The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* vol. 489 519–525 (2012).
30. Campbell, J. D. *et al.* Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat. Genet.* **48**, 607–616 (2016).
31. Venkatesan, S. *et al.* Perspective: APOBEC mutagenesis in drug resistance and immune escape in HIV and cancer evolution. *Ann. Oncol.* **29**, 563–572 (2018).
32. Chujo, M. *et al.* Comparative genomic hybridization analysis detected frequent overrepresentation of chromosome 3q in squamous cell carcinoma of the lung. *Lung Cancer* vol. 38 23–29 (2002).
33. Campbell, J. D. *et al.* Genomic, Pathway Network, and Immunologic Features Distinguishing Squamous Carcinomas. *Cell Rep.* **23**, 194–212.e6 (2018).
34. Leão, R. *et al.* Mechanisms of human telomerase reverse transcriptase (hTERT) regulation: clinical impacts in cancer. *J. Biomed. Sci.* **25**, 22 (2018).
35. de Bruin, E. C. *et al.* Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* **346**, 251–256 (2014).
36. Zhang, L.-L. *et al.* Multiregion sequencing reveals the intratumor heterogeneity of driver mutations in TP53-driven non-small cell lung cancer. *Int. J. Cancer* **140**, 103–108 (2017).
37. Sharma, A. *et al.* Non-Genetic Intra-Tumor Heterogeneity Is a Major Predictor of Phenotypic Heterogeneity and Ongoing Evolutionary Dynamics in Lung Tumors. *Cell Rep.* **29**, 2164–2174.e5 (2019).
38. Suda, K. *et al.* Innate Genetic Evolution of Lung Cancers and Spatial Heterogeneity: Analysis of Treatment-Naïve Lesions. *J. Thorac. Oncol.* **13**, 1496–1507 (2018).
39. Roper, N. *et al.* APOBEC Mutagenesis and Copy-Number Alterations Are Drivers of Proteogenomic Tumor Evolution and Heterogeneity in Metastatic Thoracic Tumors. *Cell Rep.* **26**, 2651–2666.e6 (2019).
40. Priestley, P. *et al.* Pan-cancer whole genome analyses of metastatic solid tumors. doi:10.1101/415133.
41. Lens, S. M. A. & Medema, R. H. Cytokinesis defects and cancer. *Nat. Rev. Cancer* **19**, 32–45 (2019).
42. López, S. *et al.* Interplay between whole-genome doubling and the accumulation of deleterious alterations in cancer evolution. *Nat. Genet.* **52**, 283–293 (2020).

43. Turajlic, S. & Swanton, C. Metastasis as an evolutionary process. *Science* **352**, 169–175 (2016).
44. Richard, C. *et al.* Exome Analysis Reveals Genomic Markers Associated with Better Efficacy of Nivolumab in Lung Cancer Patients. *Clin. Cancer Res.* **25**, 957–966 (2019).
45. https//:www.cbioportal.com.
46. Teixeira, V. H. *et al.* Deciphering the genomic, epigenomic, and transcriptomic landscapes of pre-invasive lung cancer lesions. *Nat. Med.* **25**, 517–525 (2019).
47. Wilbertz, T. *et al.* SOX2 gene amplification and protein overexpression are associated with better outcome in squamous cell lung cancer. *Mod. Pathol.* **24**, 944–953 (2011).
48. Tata, P. R. *et al.* Developmental History Provides a Roadmap for the Emergence of Tumor Plasticity. *Developmental Cell* vol. 44 679–693.e5 (2018).
49. Mollaoglu, G. *et al.* The Lineage-Defining Transcription Factors SOX2 and NKX2-1 Determine Lung Cancer Cell Fate and Shape the Tumor Immune Microenvironment. *Immunity* **49**, 764–779.e9 (2018).
50. Kenfield, S. A., Wei, E. K., Stampfer, M. J., Rosner, B. A. & Colditz, G. A. Comparison of aspects of smoking among the four histological types of lung cancer. *Tob. Control* **17**, 198–204 (2008).

Figures

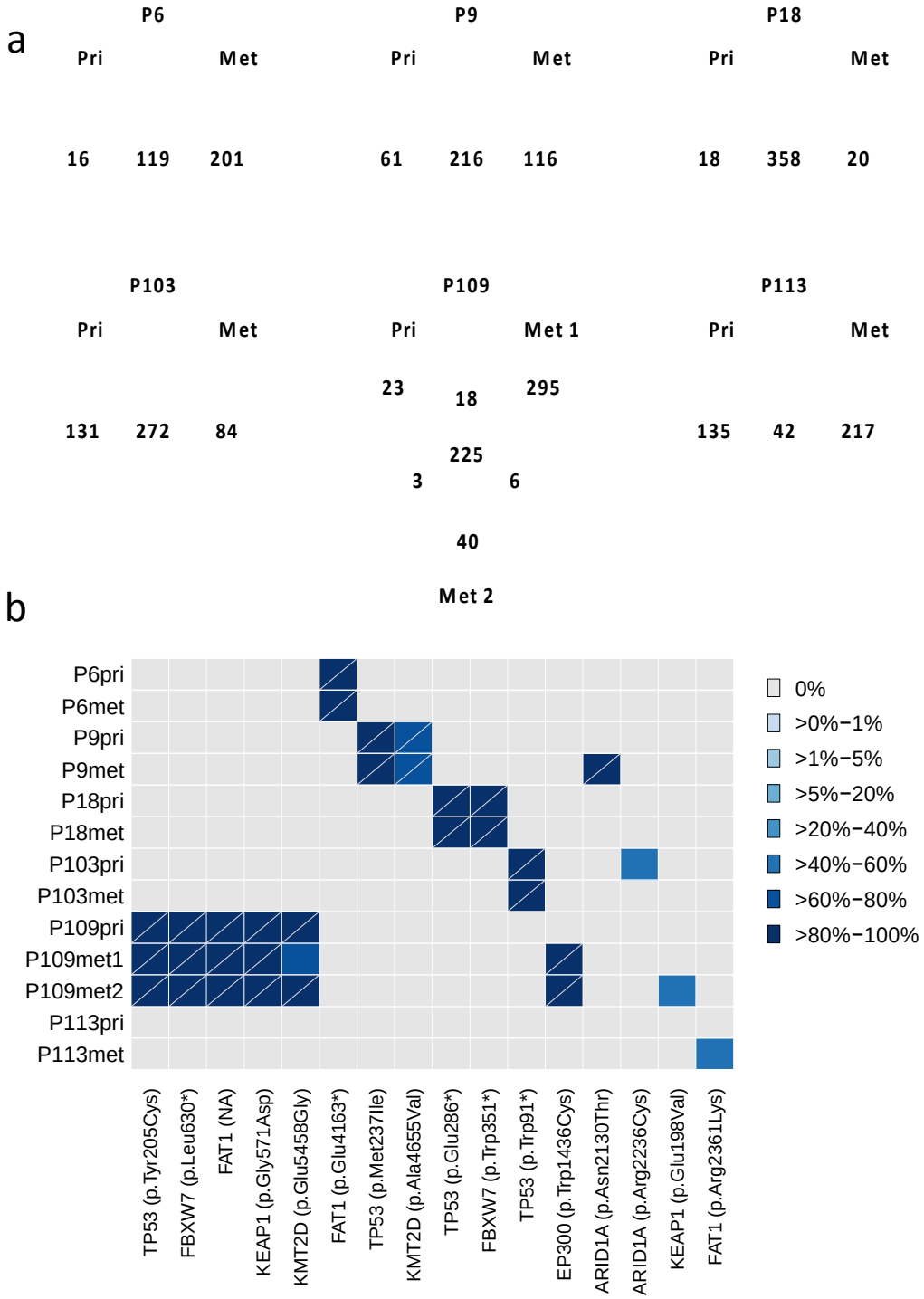


Figure 1. Clonal relationship of primary LUSC and matched metastases. (**a**) Venn diagram displays the number of non-synonymous mutations per patient that are common between the primary tumor (blue) and the metastases (red, green). (**b**) Heatmap depicts a comparison between the non-synonymous genes in the dataset of Bailey et al. that represents the most significantly mutated genes in LUSC compared to the primary tumor and the metastases of the presented patients [1]. Heatmap

illustrates the cancer cell fraction of selected mutations. Clonal mutations are illustrated with a diagonal line.

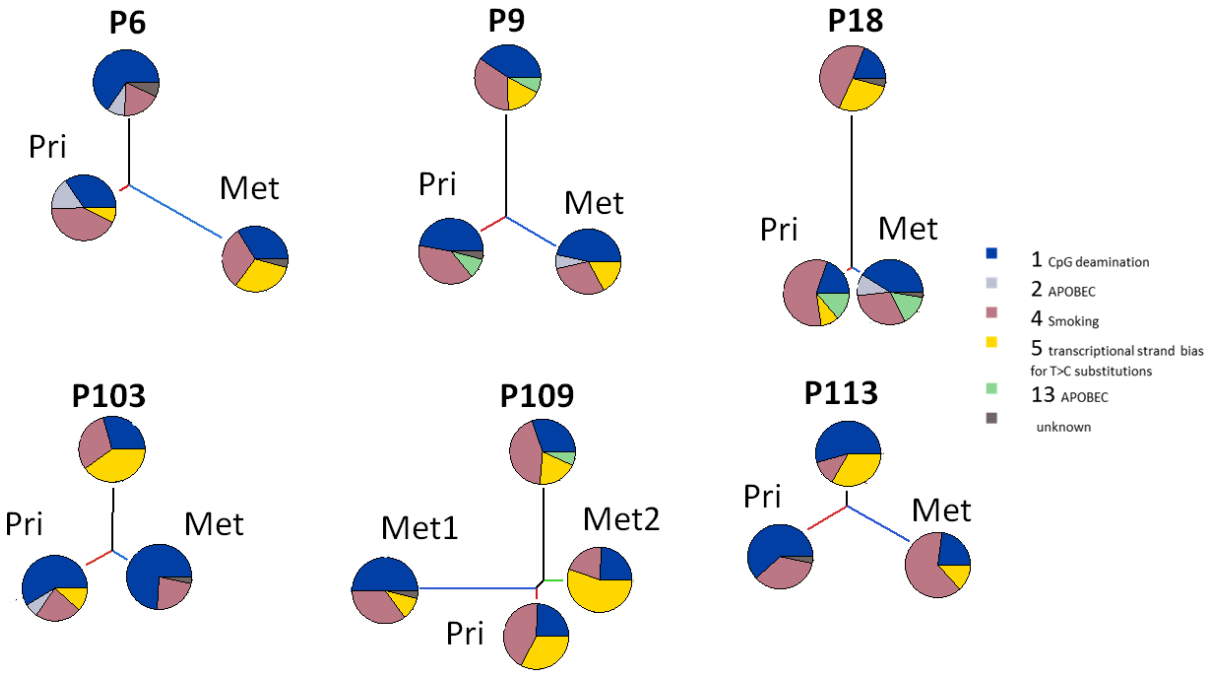


Figure 2. Evolutionary change of mutational signatures in primary tumor and metastases. Evolution of the somatic genetic alterations illustrates the changes in mutational processes. The pie chart delineates the proportion of mutational signatures. Black, blue, and red lines represent the trunk, the PT branch, and the metastasis branch, respectively. Branch lengths are proportional to the number of mutations.

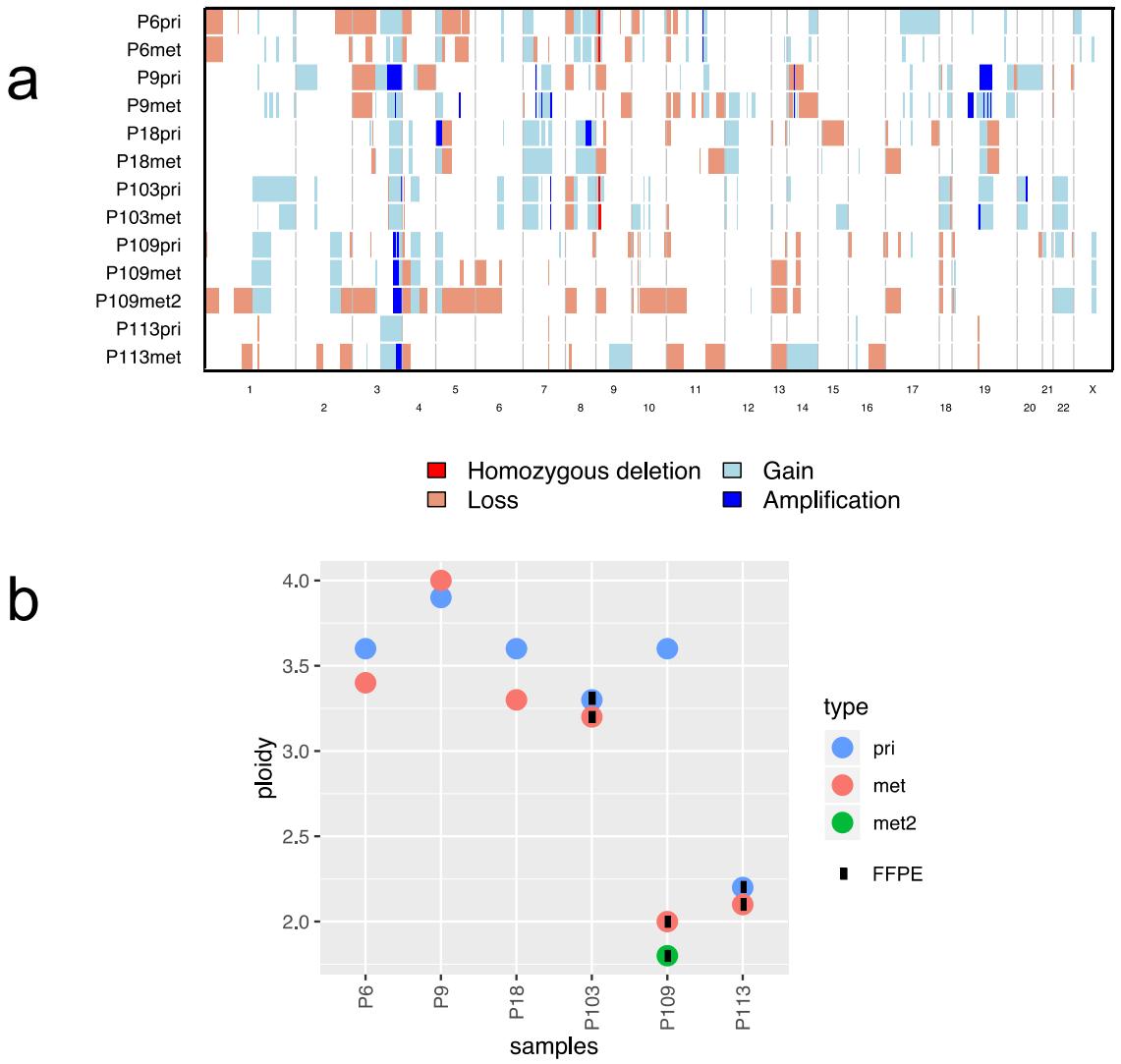


Figure 3. Similarity of primary tumors and metastases on chromosomal level. (**a**) Repertoire of copy number alterations as defined by WES. Samples represented on the *y*-axis; chromosomes are represented along the *x*-axis. Light red: copy number loss; red: homozygous deletion; light blue: copy number gain; dark blue: amplification. (**b**) Ploidy analysis measured using flow cytometry. Samples are represented on the *x*-axis; ploidy is represented in the *y*-axis. blue: primary tumor; red: first metastasis; green: second metastasis.

References

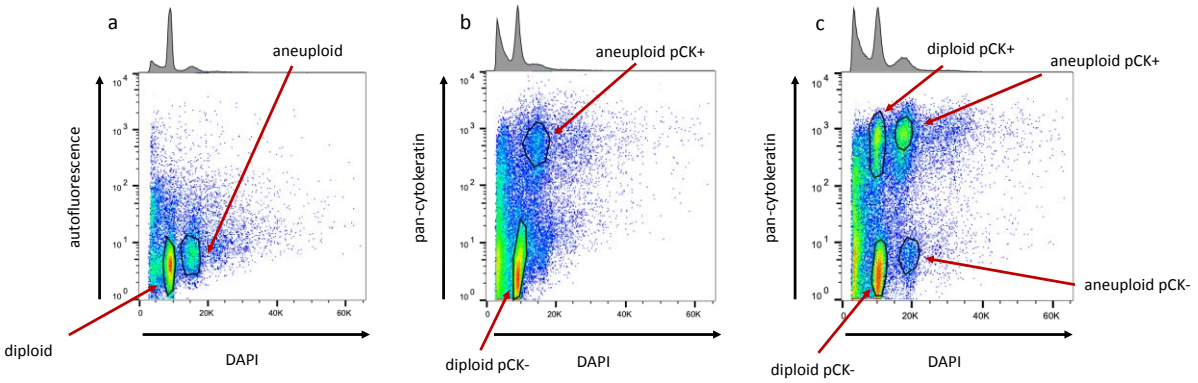
1 Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **173**, 371–385.e18 (2018).

Tables

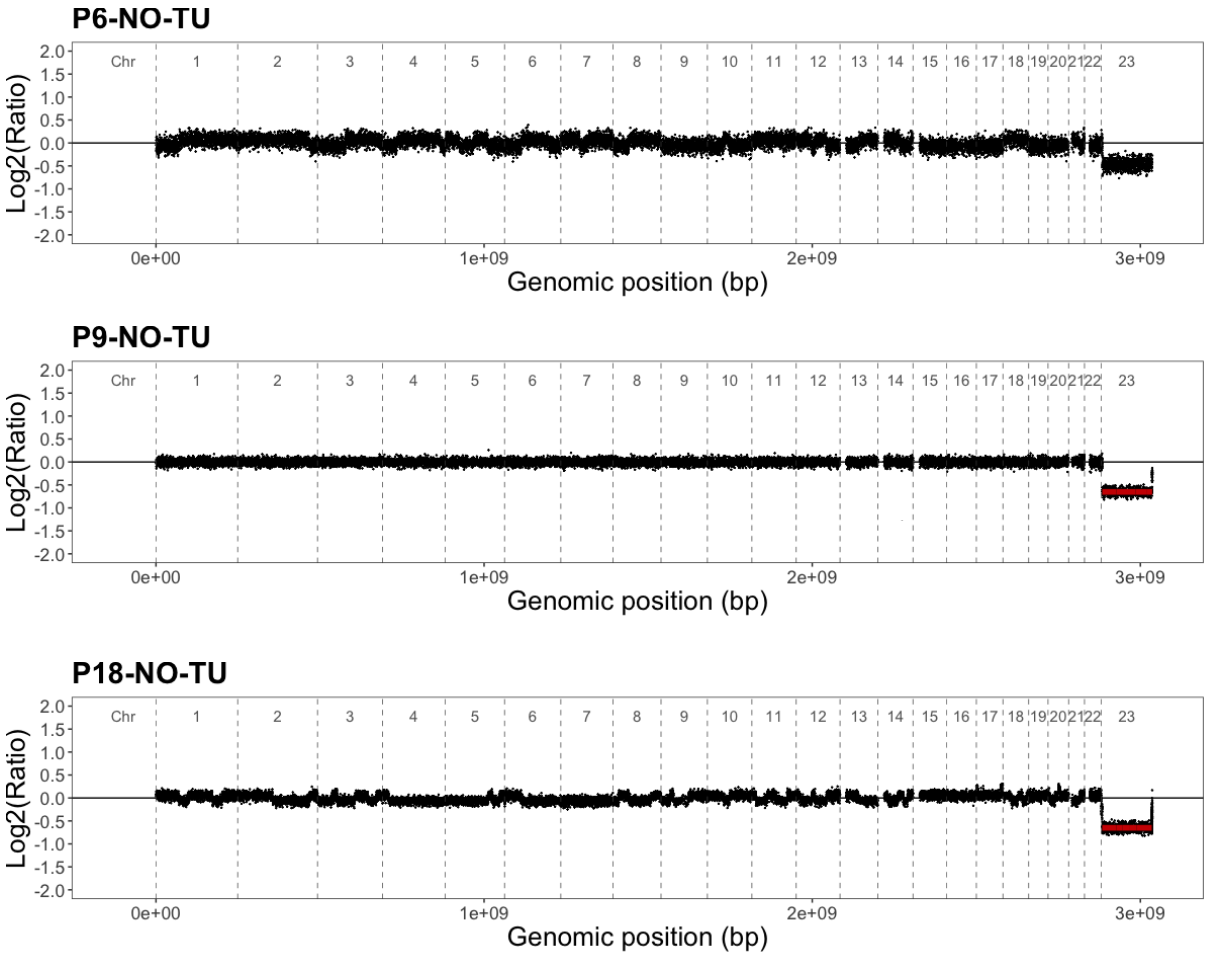
Table 1. Clinical characteristics of the cohort

Patient	6	9	18	103	109	113
Sex	m	m	m	m	m	f
Age at diagnosis (years)	54	61	64	73	68	73
Smoker at diagnosis	active	NA	former	NA	active	NA
Smoking status (pack-years)	50	NA	NA	NA	NA	NA
Radiotherapy	NA	NA	No	No	at 3rd biopsy	No
Chemotherapy	No	No	No	No	Cisplatin, Vinorelbine + Cisplatin, Gezmar	No
Time to metastasis (months)	5	8	50	1	0 (Met1) 28 (Met2)	0
Site of metastasis / recurrence	left kidney	right kidney	main bronchial carina	brain	interlobular lymph node (pN1) (Met1) thoracic spine (Met2)	subcarinal lymph node (pN2)
Ploidy (*N*) PRI	3.6	3.9	3.6	3.3	3.6	2.2
Ploidy (*N*) MET	3.4	4	3.3	3.2	2 (Met1) 1.8 (Met2)	2.1

Supplementary Figures

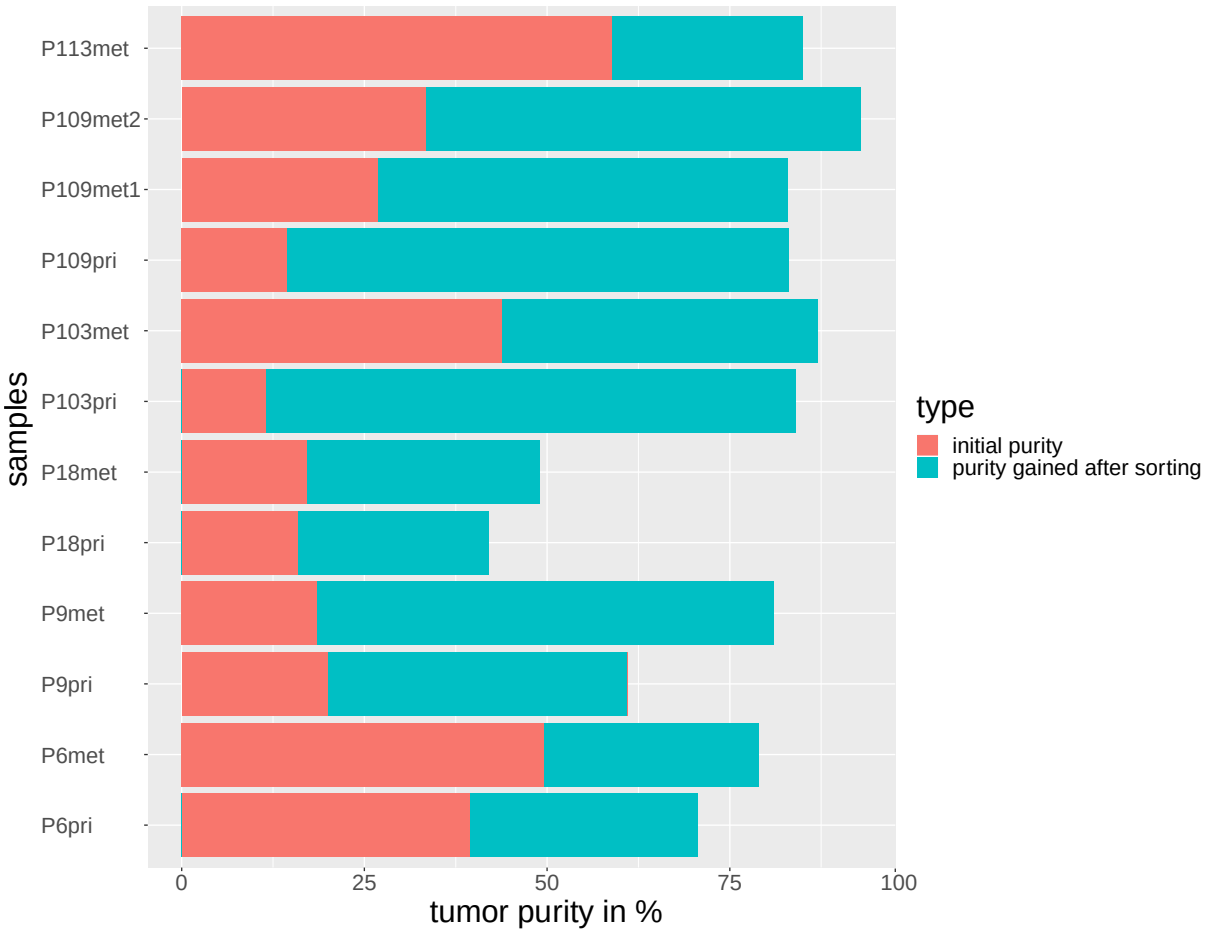


Supplementary figure 1. Nuclei flow-sorting strategy to enrich for tumor cells. Representative example for three different sorting strategies depending on the tissue type. (**a**) Fresh-frozen tissue was stained with DAPI (DNA content/ploidy) represented on the *x*-axis. Autofluorescence is depicted on the *y*-axis. Where only DAPI sorting was possible, diploid populations were considered as germline for WES analysis, as previously described in detail [1]. Aneuploid populations represent a mixture of tumor cells with proliferating G2M cells, tumor cells that underwent WGD and proliferating G2M non-tumor cells. Identification of diploid tumor cells is not possible by pure DAPI sorting. Therefore, we used array comparative hybridization to check for flat genomes prior to sequencing (Supplementary figure 2). Diploid sorted samples with flat diploid genomes were considered as germline for sequencing (**b**) To distinguish tumor cells from proliferating G2M cells, pan-cytokeratin (pCK) was used (*y*-axis) in FFPE samples. Aneuploid tumor population were pCK-positive and distinguishable from pCK-negative cells. Diploid pCK-negative populations were used as germline. (**c**) Some samples displayed four populations. Tumor population (diploid pCK+) and non-tumor population (diploid pCK-) were diploid and 'aneuploid' in the G2M phase. Therefore, the ploidy is approximately the doubled amount (aneuploid pCK- and aneuploid pCK+). In general, all populations were sorted and sequenced, if amount of DNA was sufficient.



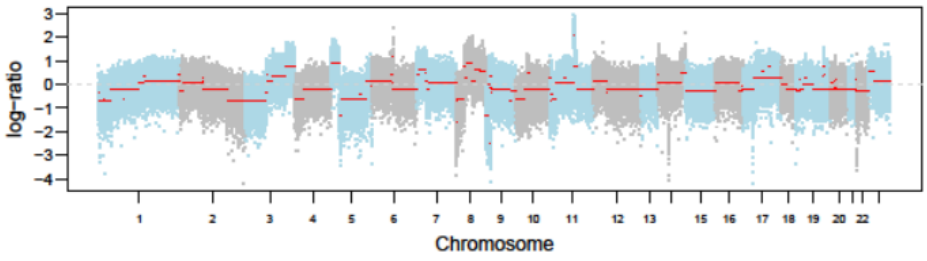
Supplementary figure 2. Copy number plots of diploid populations in samples sorted by DAPI. Illustrations show no copy number aberration in the diploid populations of P6, P9 and P18. Data was generated using array comparative hybridization [1].

.

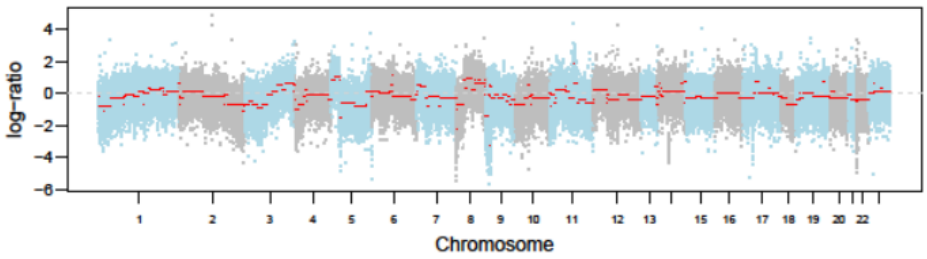


Supplementary figure 3. Purity before and after sorting. Bar plot displays the purity before sorting (red) and after sorting (blue). Purity after sorting was calculated by FACETS [2]. P113pri was excluded, because FACETS could not predict the purity. pri = primary tumor; met = metastasis.

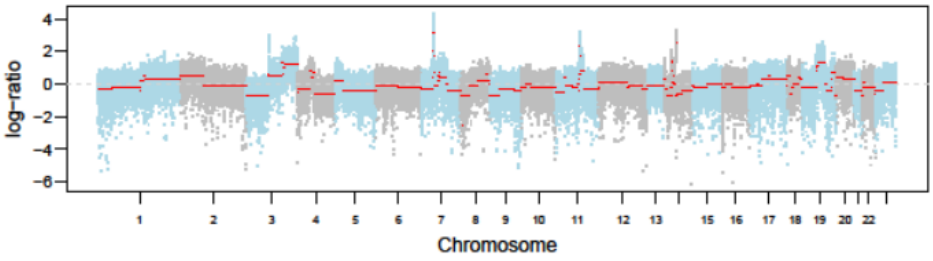
P6 primary tumor



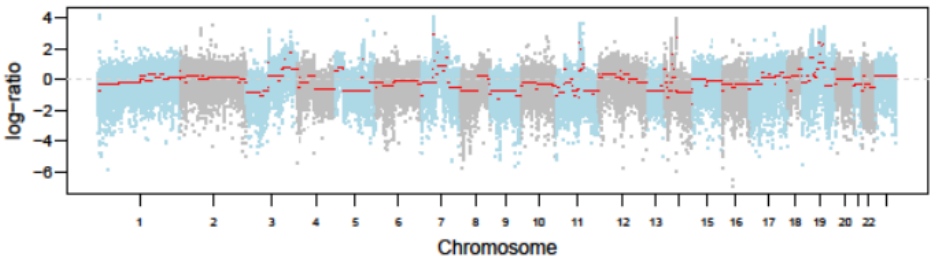
P6 metastasis



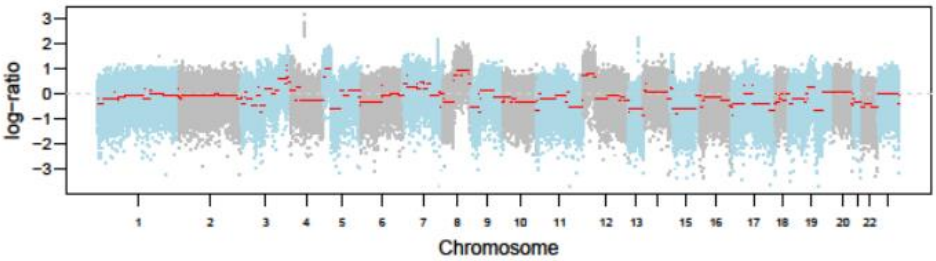
P9 primary tumor



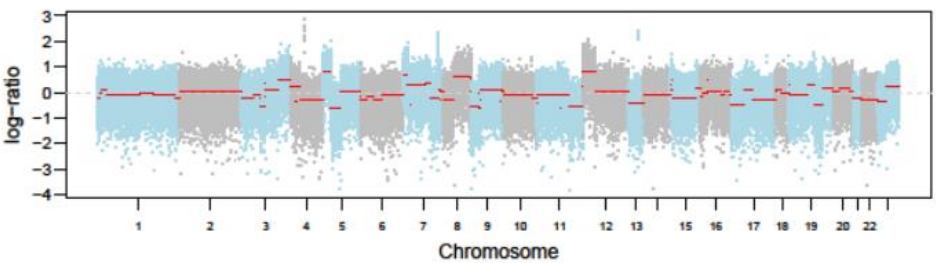
P9 metastasis



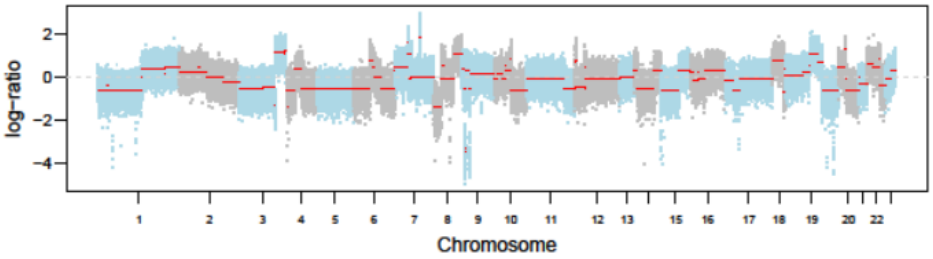
P18 primary tumor



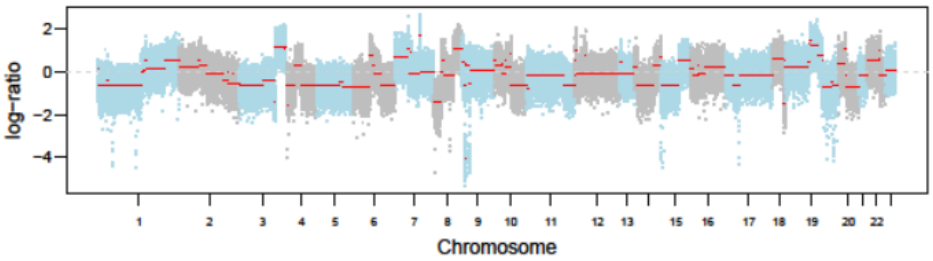
P18 metastasis



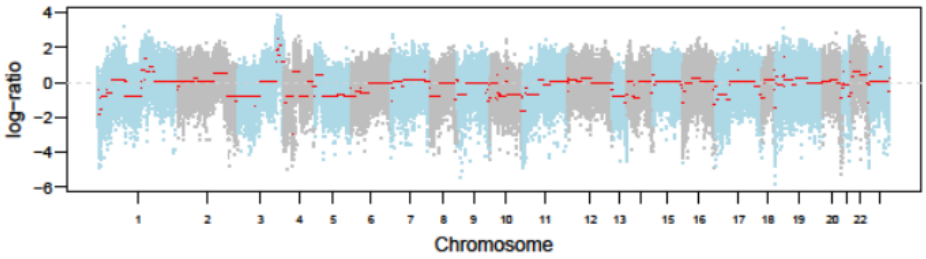
P103 primary tumor



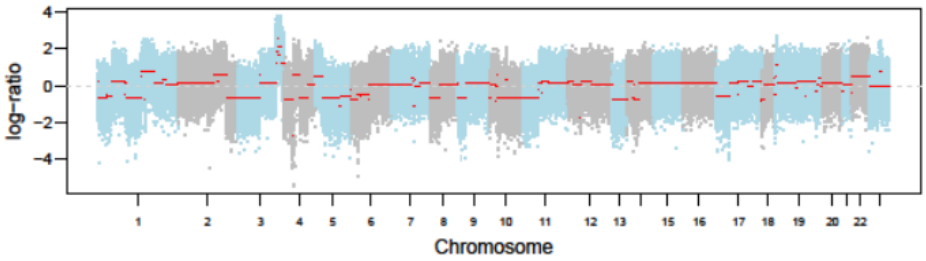
P103 metastasis



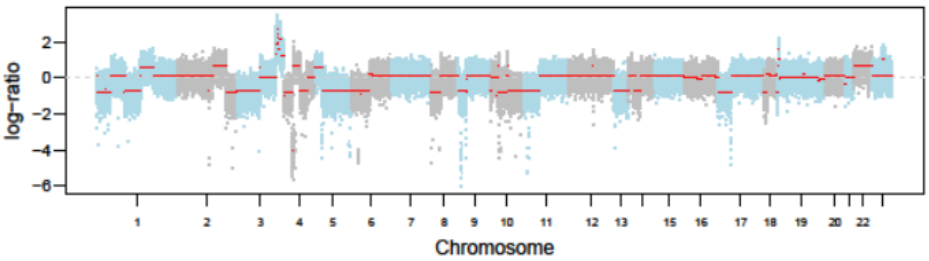
P109 primary tumor

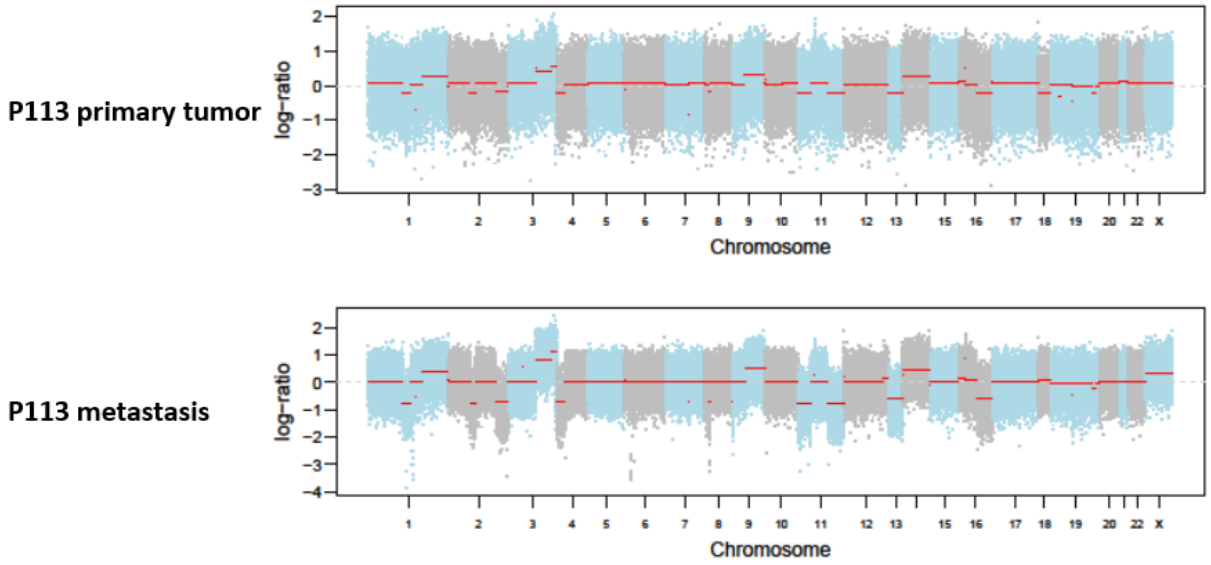


P109 metastasis 1



P109 metastasis 2





Supplementary figure 4. Copy number aberration plots of all tumors. Illustration shows copy number aberration plots of all 13 tumor samples from six patients. Plots were generated by FACETS [2].

References

1 Lorber, T. *et al.* Exploring the spatiotemporal genetic heterogeneity in metastatic lung adenocarcinoma using a nuclei flow-sorting approach. *J. Pathol.* **247**, 199–213 (2019).

2 Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* **44**, e131 (2016).

Supplementary tables

Supplementary table 1. Sorting and sequencing metrics

Patient	Tumor	Tissue	Sorted with	Ploidy	Mutations/Mb	Mean coverage	Total reads
6	primary	FF	DAPI	3.6	4.6	57.99	116651162
6	metastasis	FF	DAPI	3.4	10.7	69.02	160142677
6	non-tumor	FF	DAPI	2	-	62.17	137946635
9	primary	FF	DAPI	3.9	9.4	68.41	148484030
9	metastasis	FF	DAPI	4	11.1	73.91	150098988
9	non-tumor	FF	DAPI	2	-	62.55	132798619
18	primary	FF	DAPI	3.6	12.6	51.15	114207787
18	metastasis	FF	DAPI	3.3	12.7	73.17	127941373
18	non-tumor	FF	DAPI	2	-	77.15	152991220
103	primary	FFPE	DAPI/pCK	3.3	13.2	63.32	199580439
103	metastasis	FFPE	DAPI/pCK	3.2	12.1	75.91	177816752
103	non-tumor	FFPE	DAPI/pCK	2	-	32.34	127962411
109	primary	FF	DAPI	3.6	9.6	91.92	142302414
109	metastasis1	FFPE	DAPI/pCK	2	18.2	51.75	210698902
109	metastasis2	FFPE	DAPI/pCK	1.8	9.3	98.72	157381241
109	non-tumor	FFPE	DAPI/pCK	2	-	58.87	231972945
113	primary	FFPE	DAPI/pCK	2.2	5.9	37.55	161884161
113	metastasis	FFPE	DAPI/pCK	2.1	8.7	71.10	297508695
113	non-tumor	FFPE	DAPI/pCK	2	-	56.86	111786489

Supplementary table 2. Mutations

Supplementary table 2 is stored online.

2.3 Part C: Deciphering the clonal relationship between adenomatous and squamous components in adenosquamous carcinoma of the lung

Deciphering the clonal relationship between adenomatous and squamous components in adenosquamous carcinoma of the lung

Arthur Krause[1*], Luca Roma[1*], Thomas Lorber[1], James Habicht[2], Didier Lardinois[3], Maria Rosaria De Filippo[1,6], Spasenija Savic Prince[1], Salvatore Piscuoglio[1,4], Charlotte KY Ng[5], Lukas Bubendorf[1]

1 Institute of Molecular Genetics and Pathology, University Hospital Basel, University of Basel, Switzerland;

2 Thoracic Surgery, St. Clara Hospital, Basel, Switzerland;

3 Thoracic Surgery, University Hospital Basel, Basel, Switzerland;

4 Visceral surgery research laboratory, Clarunis, Department of Biomedicine; University of Basel, Basel, Switzerland;

5 Department for BioMedical Research (DBMR), University of Bern, Bern, Switzerland;

6 Department for BioMedical Research, Urology Research Laboratory, University of Bern, Bern, Switzerland;

*Co-first author, contributed equally to the work

Corresponding author: Prof. Lukas Bubendorf, Institute of Medical Genetics and Pathology, University Hospital Basel, Schönbeinstrasse 40, 4031 Basel, Switzerland, Tel: +41 61 265 2851, Email: Lukas.Bubendorf@usb.ch

Permanent address: Luca.Roma@usb.ch; Lukas.Bubendorf@usb.ch

Keywords: adenosquamous carcinoma, lung cancer, genomic evolution, heterogeneity, STK11, SOX2

Conflict of Interest: none

Original Publication:
The work is not being submitted for publication elsewhere at the same time and no part of the work has been published previously.

Financial Support: This work was supported by: Swiss National Foundation [SNF 320030_162781 to L.B., PZ00P3_168165 to S.P.; the Swiss Cancer League [KFS-3995-08-2016 to S.P., KFS-4543-08-2018 to C.K.Y.N.].

Word count: 3733, **Figures:** 4, **Table:** 1, **Supplementary figures:** 1, **Supplemental table:** 1, **Supplemental materials:** 1

Abstract

Adenosquamous carcinoma of the lung (ASC) is a rare subtype of non-small cell lung cancer, consisting of lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) components. ASC shows morphological characteristics of classic LUAD and LUSC but behaves more aggressively. Although ACS can serve as a model of lung cancer heterogeneity and transdifferentiation, its genomic background remains poorly understood. In this study, we sought to explore the genomic landscape of macrodissected LUAD and LUSC components of three ASC using whole exome sequencing (WES). Identified truncal mutations included the pan-cancer tumor-suppressor gene *TP53* but also *EGFR*, *BRAF,* and *MET*, which are characteristic for LUAD but uncommon in LUSC. No typical truncal LUSC driver mutations were found. Both components showed unique driver mutations that did not overlap between the three ASC. Mutational signatures of truncal mutations differed from those of the branch mutations in their descendants LUAD and LUSC. Most common signatures were related to aging (1, 5) and smoking (4). Truncal chromosomal copy number aberrations shared by all three ASC included losses of 3p, 15q and 19p, and an amplified region in 5p. Furthermore, we detected loss of *STK11* and *SOX2* amplification in ASC, which has previously been shown to drive transdifferentiation from LUAD to LUSC in preclinical mouse models. Conclusively, this is the first study using WES to elucidate the clonal evolution of ASC. It provides strong evidence that the LUAD and LUSC components of ASC share a common origin and the LUAD component appears to transform to LUSC.

Introduction

Adenosquamous carcinoma (ASC) is a rare histological subtype of lung cancer accounting for 0.4% - 4% of all lung cancers [1]. It consists of two morphologically distinct components including LUAD and LUSC [1]. From a clinical point of view, the dichotomous ASC represents a challenge. It is more aggressive than its classical single components and has been shown to be associated with worse outcome [2]. There is no specific standard treatment for ASC, and current standard of care options rely on general non-small cell lung cancer (NSCLC) guidelines. Surgical resection is the only effective mean to treat patients with ASC, mostly in conjunction with platinum-based adjuvant chemotherapy [3]. Targeted treatments can be used as first-line therapy for advanced *EGFR*-mutant or *ALK*-rearranged ASC, but there are limited data on the efficacy of EGFR- or ALK-tyrosine kinase inhibitor (TKI) in ASC due to its rarity (reviewed in [4]). From a diagnostic perspective, it is challenging to grasp the dichotomous nature of the tumor. Due to limited sampling by small biopsies, there is a high chance to miss one component so that the subsequent treatment will be based on only one of the components.

Previous studies focusing on oncogenic driver mutations revealed that ASC have similar mutation profiles and therapeutic targets as LUAD including *EGFR* mutations [5]. Yet, to the best of our knowledge, the genomic profiles of the LUAD and LUSC components within ASC and the inferred evolutionary trajectories have not been explored previously, beyond the use of single gene and gene panel testing [5]. Being defined by its distinct morphological heterogeneity, ASC represents an ideal model to study morphological transdifferentiation and clonal evolution in NSCLC. Common LUAD is considered to originate from stem cells at the bronchoalveolar junction, whereas LUSC derive from the more proximally located basal cell compartment of the bronchial epithelium [6]. Based on targeted sequencing, several mutations were detected in both components LUAD and LUSC of ASC suggesting the two entities share the same ancestor cell [5]. It has been hypothesized that they are likely to arise at the bronchoalveolar junction as LUAD and that the LUSC phenotype develops subsequently [5]. The molecular mechanisms of this adenosquamous trans-differentiation (AST) remain unknown.

This study aims at better understanding the genomic landscape and the evolutionary trajectories of ASC. To this end, we dissected the LUAD and LUSC components separately from three ASC patients for comprehensive genomic profiling of both components using whole exome sequencing (WES).

Materials and Methods

Patient cohort & Immunohistochemistry

Three cases of ASC were retrieved from the pathology archive of the Institute of Pathology and Medical Genetics, University Hospital Basel, Switzerland. Representative formalin-fixed paraffin-embedded sections were cut (4µm thick) and stained with hematoxylin and eosin. TTF-1, CK7 and p40 were used to distinguish LUAD and LUSC components. TTF-1 (Ventana catalog number 790-7456), CK7 (Ventana catalog number 790-4462) and p40 (Ventana catalog number 790-4950) staining was performed on the Benchmark Ultra (Ventana, Roche) as described previously [7]. The sections were reviewed by a pulmonary pathologist (LB), who marked distinct morphological regions for macrodissection. FISH analysis was performed according to the manufactures protocol (Zyto*Light*® SPEC SOX2/CEN3 dual color probe, catalog number Z-2127-200, Zytovision, Bremerhaven, Germany). This study has been carried out under the ethical approval EKNZ 31/12.

Macrodissection and DNA extraction

Up to ten consecutive 10µm thick unstained tissue sections were cut on glass slides. The distinct morphological components and matched normal lung tissue were identified and marked by a pathologist (LB), scratched from the glass slide using a 25g disposable syringe under a stereoscope as previously described [8]. Scratched cells were subjected to 200µl of ATL buffer and 40µl Proteinase K and incubated overnight at 56°C with 500rpm using the reagents from DNeasy Blood and Tissue Kit (Qiagen, Germantown, MD, USA) according to the manufacturer's instructions. DNA was quantified using the Qubit Fluorometer assay (Life Technologies, Carlsbad, CA, USA) as previously described [8].

Whole exome sequencing and variant annotation

Extracted DNA from the macrodissected ASC components were subjected to WES. SureSelect Human All Exon V6 Kit (Agilent) was used for the whole exome capturing according to manufacturer's guidelines. Sequencing was performed on Illumina NovaSeq 6000 using paired-end 100-bp reads and yielded a median depth of coverage

of 48x to 120x in tumor components and 49x to 122x in the corresponding germlines (Supplementary table 1). Sequencing was performed by CeGaT (Tübingen, Germany).

Reads were aligned to the reference human genome GRCh37. Somatic single nucleotide variants (SNVs) and small insertions and deletions (indels) were detected using MuTect (v.1.1.7) and Strelka (v.2.9.10), respectively [9,10]. To reduce false positive results from artifacts caused by formalin fixation, specific SNVs C:G>T:A with variant allelic fraction (VAF) less than 10% were discarded. Otherwise, SNVs or indels with a VAF < 1% or that were covered by fewer than 3 reads were discarded, if the SNVs were found in both components of one sample, a cut-off of two read was applied. Variant annotation was performed by SnpEff software v.4.1 [11].

The heatmap of non-synonymous mutations was generated using the R package maftools v.2.0.16 by selecting the genes in the Bailey et al. dataset that represents the significantly mutated genes in classic LUAD and LUSC obtained from TCGA [12].

Copy number aberration analysis and clonality analysis

Allele-specific CNAs were identified using FACETS v.0.5.14 [13]. Deletions/losses were defined with the log-ratio < -0.3 and amplifications/gains > 0.3. Cancer cell fraction (CCF) of each mutation was identified using ABSOLUTE v.1.0.6 [14]. A mutation was classified as clonal if its probability of being clonal was > 50% or if the lower bound of the 95% confidence interval of its CCF was > 90%. Mutations were considered as subclonal, if they did not meet the mentioned criteria [15,16]. A mutation that was found in both components of one biopsy was considered as ‘trunk’. Mutations that were detected in only one component of the tumor were granted as ‘branch’ or ‘private’.

Phylogenetic analysis and mutational signatures

A maximum parsimony tree was built for each case using binary presence/absence matrices based on the repertoire of non-synonymous and synonymous somatic mutations, gene amplifications and homozygous deletions in the biopsies of the tumors, as described by Murugaesu et al[17,18].

Decomposition of mutational signatures was conducted using the R package deconstructSigs by selecting mutational signatures based on the set of 30 mutational signatures (‘signature.cosmic’) that were observed in LUAD and LUSC [19,20].

Results

Adenosquamous carcinoma have different growth patterns

The main clinico-pathologic characteristics of the three patients with ASC are summarized in Table 1. All patients were >60 years old. P118 and P120 were former smokers, while the smoking status of P119 was unknown. The three tumors were at stages pT2-pT4. In all three patients, the LUAD area was larger than the LUSC area. All LUSC areas were homogeneously p40-positive, while all LUAD areas were p40-negative by immunohistochemistry (Figure 1b-c). Two LUAD were TTF-1 positive (P118 & P119). There was a predominant acinar growth pattern of the LUAD component in all three ASCs. Additional LUAD patterns were lepidic (P118 & P119) and solid (P119 & P120). Keratinization in the LUSC component was observed in P118 and P119.

LUAD and LUSC within adenosquamous carcinoma are of monoclonal origin

Overall, we found 759 non-synonymous mutations in 441 genes across the six tumor components of the three ASC. The proportion of shared mutations of LUSC and LUAD components of P118, P119 and P120 were 54%, 28% and 78%, respectively (Figure 2a, Supplementary table 2). The six separately analyzed tumor areas harbored a median of 72 (range 53 - 254) non-synonymous somatic mutations. Including synonymous mutations, the median of mutational burden was 3.5 mutations (mut)/Mb (range 2.3 - 11.6). P120 LUAD harbored 11.6 mut/Mb and LUSC 11.3 mut/Mb. The mutation rates of P118 (LUAD 3.5 mut/Mb, LUSC 2.3 mut/Mb) and P119 (LUAD 3.3 mut/Mb, LUSC 3.4 mut/Mb) were lower as compared to published studies with usual LUAD 6.3 mut/Mb, usual LUSC 9.0 mut/Mb and ASC 5.4 mut/Mb, respectively [21].

Across the samples, most of the mutations in driver genes were clonal (Figure 2b). The analysis of P118 displayed four shared cancer gene mutations, of which only *EGFR* exon 19 deletion was clonal in both. *TP53* p.Gln192* was mutated in both components and a further *TP53* p.Gly108Asp mutation was detected only in the LUAD component. P119 shared three clonal mutations *MET, BRAF* and P119 *ACVR1B* mutations. We detected shared cancer gene mutations between the components of P120. Moreover, one clonal *STRN* mutation was detected in P120 LUAD. Further, four clonal and one subclonal private mutations were found in the LUSC component.

Gene copy number data from all patients and tumor components revealed high concordance between the two components (Figure 2c). Amplifications in 5p were shared in all samples together with losses in 3p, 15q and 19p. LUAD and LUSC shared highly amplified regions such as chr 6q22.33-23.3 in P118, chr 12p11.22-11.1 and 12q14.1-21.1 in P119 and chr 8p12-11.21 and 20p13-p11.1 in P120.

Evolutionary history of genetic alterations

Truncal genomic alterations are shared in all tumor cells by definition and thus represent early events in the timeline of the tumor. Most mutations in known cancer driver genes were truncal (Figure 3). Mutated cancer genes were rather known from classic LUAD than LUSC, such as *EGFR, MET* and *BRAF.* We found only KMT2D in P120, but not any other mutations in LUSC typical driver genes, such as *NFE2L2, PTEN* and *FAT1.* Interestingly, there were fewer mutations in the branches than truncal mutations. We did not find any cancer gene mutations in the P119 LUSC component but found at least one in all other samples.

Truncal copy number variations (CNVs) might also play a role as drivers of malignancy. Across all sample's Loss of regions at chr 3p, 15q, 19p and an amplified locus at chr 5p was identified across all samples. We found late events that were restricted to only one component in one patient. In P118 LUAD, losses in regions of chr 6, 17 and 20 were found that were normal in LUSC. We detected in LUAD P119 private losses in regions of chr 19, whereas the LUSC component harbored private losses in chr 12. Further private losses in chr 12 were identified in P120 LUSC. Regions of chr 9 were found as gained in LUSC P120 but were lost in the LUAD component. Our gene copy number data suggest persistent genomic instability as chromosomal aberrations are detected even after branching of the two components.

Mutational signatures result from exogenous and endogenous mutational processes [20]. Given that truncal genomic alterations are likely acquired prior to the branching of the two components, we sought to investigate whether mutational signatures are shared or change over time (Figure 3). Lung cancer is associated with specific types of mutational signatures, depending on the subtype [20]. It was previously demonstrated that LUAD and LUSC share common signatures of type 1 (CpG deamination and aging), 2 (APOBEC), 4 (smoking), 5 (transcriptional strand bias for T>C**)** 13 (APOBEC). In addition, LUAD was found to display signatures 6 (defective DNA-repair)

and 17 (unknown origin) [20]. For all our three ASC, we found different patterns of signatures as well as commonalities.

Trunks of all three tumors displayed signatures 1 and 4. The truncal mutations of P118 and P119 were also characterized by the APOBEC signature 13. In addition, trunks of P118 and P120 carried signature 5. The signatures changed drastically in the branches. Whereas P119 LUAD was found to exhibit signatures 1, 2, 4 and 13 the LUSC component only exhibited signature 1 and 6. Signature 1 contributed to more than 50% of the mutations of all LUAD/LUSC branches, except of P118 LUAD where it was not present at all. The smoking signature was also seen in both components of P120 and in P119 LUAD, but not P119 LUSC. Several signatures were observed exclusively in a single branch, including signature 13 in P119 LUAD, signature 6 in P119 LUSC and signature 17 in P120 LUAD. Notably, signature 6 was observed in P119 LUSC, which is not common in classic LUSC, but prominent in LUAD.

Taken together, we found that most of the genomic alterations accumulated in the trunk before branching. Overall, the three mutational trunks displayed a variety of mutational signatures that changed over time in the descendant components. Aging and smoking-induced signatures 1 and 4 were the most prominent and present in almost all lineages suggesting a promoting role in tumor development and progression.

ASC has LUAD-like genomic features

Classic LUAD and LUSC have distinct genomic landscapes. LUAD frequently have mutations in *EGFR*, *MET*, *BRAF* and *STK11* while *KMT2D* or *PTEN* are frequently mutated in LUSC. *TP53* is frequently mutated in both LUAD and LUSC. We asked whether ASC would show genomic features of both classic types.

WES revealed shared mutations in *TP53* between the LUAD and LUSC components of two patients (Figure 4). *TP53* p.Gln192* and p.His179Arg mutations were observed in the trunks of P120 and P118, respectively. An additional *TP53* mutation p.Gly108Asp was found in P118 LUAD. P118 harbored a deletion in *EGFR* exon 19, a well-known driver gene, which was also detected in previous ASC studies [22]. This *EGFR* mutation is well-known in LUAD, but uncommon in usual LUSC [23]. According to TCGA, *EGFR* mutations were found in 14% in classical LUAD but only to 2% in classical LUSC (Figure 4) [24,25]. Frameshift deletions of *BRAF* and *MET*, two other

oncogenic driver genes of LUAD, were detected within the trunk of P119 [24]. *KMT2D* was the only mutation typical of LUSC that was mutated in P120 (Figure 4). Most mutations in lung cancer driver genes found were in driver genes typical of LUAD.

Arm-level deletions that occurred in all samples, such as in chr 3p are common in both classical LUAD and LUSC (described Materials and Methods section). Deletion in 15q and 19p that were found in all samples are typical of classic LUAD with an incidence of >90% in LUAD but only 20% in LUSC [26]. It was previously reported that loss of *STK11* can promote the development of ASC [27,28] Furthermore, *STK11* alterations occur in up to 14% of LUAD but in only 2% of LUSC (http://cbioportal.org) [29]**.** In fact, we found truncal of *STK11* in P118 and P120. In P120, we detected a truncal amplified region on 3q including *TP63* and *SOX2* (Figure 3). *SOX2* and *TP63* are markers for LUSC stemness and differentiation [30,31]. SOX2 amplification was confirmed using FISH (Supplementary Figure 1). P119 LUAD amplified *NKX2-1* that encodes for TTF-1, which is the lineage marker of 70-80% of LUAD [30,31].

Taken together, our results suggest that ASCs are genomically more similar to classic LUAD than to classic LUSC.

Discussion

ASC had originally been assumed to be a mixed tumor with a polyclonal origin consisting of two separate tumors until recent gene panel sequencing demonstrated shared mutations in LUAD and LUSC providing strong evidence for a monoclonal origin [5]. Here we provided unequivocal evidence for a monoclonal origin of the LUAD and LUSC components in ASC based on comprehensive mutational patterns and CNVs obtained by WES.

In our study, truncal genetic alterations included not only mutations in common cancer genes such as *TP53* but also the typical driver gene mutations of LUAD in *EGFR*, *BRAF* and deletions of *STK11* (Figure 3). This is in line with previous studies using small NGS panels or selected single gene testing of the two ASC components [5,32]. In particular, activating *EGFR* mutations have been shown to be at least as common in ASC as in classical LUAD ranging from approximately 15% in Western patients to 50% in Asian patients, but uncommon in LUSC [23,33,34]. In contrast, we found only *KMT2D* as a typical LUSC driver, but not others such as *NFE2L2 and PTEN* mutations [25,35]. Both components shared non-synonymous mutations including cancer driver genes. The clonal relatedness of the two components together with the presence of typically more LUAD than LUSC mutations suggest that ASC originates from an early LUAD ancestor.

Tumor evolution is influenced by exogenous and endogenous mutational processes that shape the spectra of different mutational signatures [36]. The type of mutational signatures may change from the trunk to branches, as previously demonstrated in breast cancer, LUAD and LUSC [18,37,38]**.** Change in mutation signatures can provide additional insights into the development and evolutionary traits of malignant tumors [39,40]. Recently, Richard and colleagues found an association between mutational signatures and clinical efficacy of nivolumab, an immune checkpoint inhibitor, in lung cancer patients [41]. The trunks of all three ASC were dominated by aging and smoking-related signatures 1 and 4, respectively, which is partly explained by a smoking history of the patients. However, the distribution of signatures changed markedly in the LUAD and LUSC branches. Notably, we did not observe a systematic accumulation of the APOBEC signature, which was previously shown to drive mutagenesis and subclonal expansion in LUAD and LUSC [39]. Further studies are needed to better understand the

biological significance and potential clinical impact of mutational signatures in lung cancer, and their association with histological transdifferentiation.

The high overall concordance between the chromosomal copy number aberrations profiles of the components support our previous data from metastatic LUAD showing that most aberrant genomic events are truncal with only limited heterogeneity between primary tumors and matched metastases [42]. Nevertheless, the two ASC components had private CNVs, which were similar to the CNVs known to be typical for their pure morphological counterparts, i.e. classical LUAD and LUSC, respectively [26]. For example, the trunk of P120 and P119 LUSC had an amplified region on 3q including *SOX2*. *SOX2* expression is known as a common feature of squamous differentiation, and *SOX2* amplification occurs in 40% of LUSC but is almost never seen in LUAD [43]. Thus, it is possible that *SOX2* amplification was at least partly responsible for the AST of truncal LUAD to LUSC in P119 and P120. This hypothesis would be in line with previous evidence from a mouse model, in which overexpression of *SOX2* induced squamous differentiation [44,45]. In fact, AST was linked to silenced TTF-1 expression in the pre-clinical mouse model, suggesting that the differential regulation of TTF-1 and *SOX2* expression might be involved in the transformation process. In P120, AST is difficult to prove. The truncal nature of *SOX2* amplification and *KMT2D* mutation could point towards a reverse sequence with a LUSC-type ancestor clone despite the truncal *STK11* loss. Deletion of *STK11,* which we found in P118 and P120, might pinpoint another mechanism of plasticity and AST. *STK11* alterations in general, occur in up to 14% in classic LUAD but in only 2% of classical LUSC (http://cbioportal.org) [29]. There is pre-clinical evidence from mouse models suggesting that *STK11* deficiency can lead to a transition from LUAD to LUSC [27,30]. *STK11* is known to suppress tumors by negatively regulating mTOR. In these models, suppression of *STK11* played a critical role in histological transformation from LUAD to LUSC [27]. Besides activating the mTOR pathway, Li et al proposed that loss of *STK11* also leads to metabolic stress by increased radical oxygen species, which in turn makes the tumor escape into squamous differentiation [27,30]. Loss of *STK11* led to a downregulation of the Polycomb repressive complex 2 that represses squamous differentiation. Consequently, transdifferentiation to a squamous phenotype was mediated [28]. Concluding, AST is a complex process that might be caused due to several different mechanisms that require further studies.

Interestingly, Park et al recently studied AST of *EGFR*-mutated LUAD to LUSC as a rare resistance mechanism to treatment with EGFR tyrosine kinase inhibitors (TKI) [46]. The three transformed LUSC cases acquired genomic alterations in the genes *PTEN, STK11, PIK3CA or RICTOR*, again pointing towards a role of *STK11* and the PI3K/AKT/mTOR pathway in the transformation to LUSC. In line with their results, we found mutations in *STK11* and *AKT* affecting the same pathway.

The main limitation of this study is the low number of patients reflecting the rarity of the tumor. In fact, this is the first study to perform comprehensive WES on the two components of ASC separately. Moreover, the usage of archived FFPE blocks and contamination with non-tumor tissue negatively affect the resolution of our analysis. Nevertheless, our CNV data of *SOX2* and *STK11* support pre-clinical evidence of genomic driven mechanisms of AST to LUSC. Although we discovered only KMT2D as a candidate mutation known to be typical but no exclusive for squamous differentiation, we cannot rule out a role of other mutations or mutational patterns. It is likely that AST in ASC is driven by both genetic and non-genetic mechanisms including transcriptional regulation and metabolic factors as discussed above. The role of epigenetic reprogramming has not yet been studied in the context of AST in ASC despite the paramount importance of epigenetics in shaping morphology [47].

Our whole-exome analysis showed compelling evidence for a clonal relationship between the LUAD and LUSC components of ASCs, and the emergence from a common LUAD-like ancestor clone. Even if the results are ambiguous, identification of *SOX2* amplification and *STK11* loss in two of the three LUSC components, each, supports previous pre-clinical evidence for a role of these genes in AST of LUAD in ASC. Further and more comprehensive studies with a larger sample size are needed to gain a better understanding of the mechanisms that drive transformation from LUAD to LUSC or LUSC to LUAD, including transcriptomics and epigenomics.

Acknowledgement

We acknowledge CeGaT (Tübingen, Germany) for the technical expertise in sequencing.

References

1. International Agency for Research on Cancer. *Who Classification of Tumours of the Lung, Pleura, Thymus and Heart*. (World Health Organization, 2015).
2. Filosso, P. L. *et al.* Adenosquamous lung carcinomas: a histologic subtype with poor prognosis. *Lung Cancer* **74**, 25–29 (2011).
3. Shi, X. *et al.* PD-L1 expression in lung adenosquamous carcinomas compared with the more common variants of non-small cell lung cancer. *Sci. Rep.* **7**, 46209 (2017).
4. Li, C. & Lu, H. Adenosquamous carcinoma of the lung. *Onco. Targets. Ther.* **11**, 4829–4835 (2018).
5. Vassella, E. *et al.* Molecular profiling of lung adenosquamous carcinoma: hybrid or genuine type? *Oncotarget* **6**, 23905–23916 (2015).
6. Kim, C. F. B. *et al.* Identification of Bronchioalveolar Stem Cells in Normal Lung and Lung Cancer. *Cell* vol. 121 823–835 (2005).
7. Blumer, T. *et al.* Hepatocellular Carcinoma Xenografts Established From Needle Biopsies Preserve the Characteristics of the Originating Tumors. *Hepatol Commun* **3**, 971–986 (2019).
8. Piscuoglio, S. *et al.* The Genomic Landscape of Male Breast Cancers. *Clin. Cancer Res.* **22**, 4045–4056 (2016).
9. Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
10. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
11. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
12. Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **173**, 371–385.e18 (2018).
13. Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* **44**, e131 (2016).
14. Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nature Biotechnology* vol. 30 413–421 (2012).

15. Landau, D. A. *et al.* Evolution and Impact of Subclonal Mutations in Chronic Lymphocytic Leukemia. *Cell* vol. 152 714–726 (2013).
16. Guerini-Rocco, E. *et al.* Microglandular adenosis associated with triple-negative breast cancer is a neoplastic lesion of triple-negative phenotype harbouring TP53 somatic mutations. *J. Pathol.* **238**, 677–688 (2016).
17. Murugaesu, N. *et al.* Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. *Cancer Discov.* **5**, 821–831 (2015).
18. Ng, C. K. Y. *et al.* Genetic Heterogeneity in Therapy-Naïve Synchronous Primary Breast Cancers and Their Metastases. *Clin. Cancer Res.* **23**, 4402–4415 (2017).
19. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, 31 (2016).
20. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
21. Chalmers, Z. R. *et al.* Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med.* **9**, 34 (2017).
22. Shi, X. *et al.* Screening for major driver oncogene alterations in adenosquamous lung carcinoma using PCR coupled with next-generation and Sanger sequencing methods. *Sci. Rep.* **6**, 22297 (2016).
23. Campbell, J. D. *et al.* Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat. Genet.* **48**, 607–616 (2016).
24. Network, T. C. G. A. R. & The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* vol. 511 543–550 (2014).
25. Network, T. C. G. A. R. & The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* vol. 489 519–525 (2012).
26. Qiu, Z.-W., Bi, J.-H., Gazdar, A. F. & Song, K. Genome-wide copy number variation pattern analysis and a classification signature for non-small cell lung cancer. *Genes, Chromosomes and Cancer* vol. 56 559–569 (2017).
27. Li, F. *et al.* LKB1 Inactivation Elicits a Redox Imbalance to Modulate Non-small Cell Lung Cancer Plasticity and Therapeutic Response. *Cancer Cell* **27**, 698–711 (2015).

28. Zhang, H. *et al.* Lkb1 inactivation drives lung cancer lineage switching governed by Polycomb Repressive Complex 2. *Nat. Commun.* **8**, 14922 (2017).
29. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
30. Hou, S. *et al.* Evidence, Mechanism, and Clinical Relevance of the Transdifferentiation from Lung Adenocarcinoma to Squamous Cell Carcinoma. *Am. J. Pathol.* **187**, 954–962 (2017).
31. Campbell, J. D. *et al.* Genomic, Pathway Network, and Immunologic Features Distinguishing Squamous Carcinomas. *Cell Rep.* **23**, 194–212.e6 (2018).
32. Kanazawa, H. *et al.* Transition from Squamous Cell Carcinoma to Adenocarcinoma in Adenosquamous Carcinoma of the Lung. *Am. J. Pathol.* **156**, 1289–1298 (2000).
33. Hsu, K.-H. *et al.* Identification of five driver gene mutations in patients with treatment-naïve lung adenocarcinoma in Taiwan. *PLoS One* **10**, e0120852 (2015).
34. Shi, Y. *et al.* A Prospective, Molecular Epidemiology Study of EGFR Mutations in Asian Patients with Advanced Non–Small-Cell Lung Cancer of Adenocarcinoma Histology (PIONEER). *Journal of Thoracic Oncology* vol. 9 154–162 (2014).
35. Zhang, X.-C. *et al.* Comprehensive genomic and immunological characterization of Chinese non-small cell lung cancer patients. *Nat. Commun.* **10**, 1772 (2019).
36. Jamal-Hanjani, M. *et al.* Tracking the Evolution of Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **376**, 2109–2121 (2017).
37. Lee, J. Y. *et al.* Lobular Carcinomas In Situ Display Intralesion Genetic Heterogeneity and Clonal Evolution in the Progression to Invasive Lobular Carcinoma. *Clinical Cancer Research* vol. 25 674–686 (2019).
38. de Bruin, E. C. *et al.* Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* **346**, 251–256 (2014).
39. Swanton, C., McGranahan, N., Starrett, G. J. & Harris, R. S. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discov.* **5**, 704–712 (2015).
40. Rogozin, I. B. *et al.* Mutational signatures and mutable motifs in cancer genomes. *Brief. Bioinform.* **19**, 1085–1101 (2018).
41. Richard, C. *et al.* Exome Analysis Reveals Genomic Markers Associated with Better Efficacy of Nivolumab in Lung Cancer Patients. *Clin. Cancer Res.* **25**, 957–966 (2019).

42. Lorber, T. *et al.* Exploring the spatiotemporal genetic heterogeneity in metastatic lung adenocarcinoma using a nuclei flow-sorting approach. *J. Pathol.* **247**, 199–213 (2019).
43. Tatsumori, T. *et al.* p40 is the best marker for diagnosing pulmonary squamous cell carcinoma: comparison with p63, cytokeratin 5/6, desmocollin-3, and sox2. *Appl. Immunohistochem. Mol. Morphol.* **22**, 377–382 (2014).
44. Tata, P. R. *et al.* Developmental History Provides a Roadmap for the Emergence of Tumor Plasticity. *Developmental Cell* vol. 44 679–693.e5 (2018).
45. Mollaoglu, G. *et al.* The Lineage-Defining Transcription Factors SOX2 and NKX2-1 Determine Lung Cancer Cell Fate and Shape the Tumor Immune Microenvironment. *Immunity* **49**, 764–779.e9 (2018).
46. Park, S. *et al.* Paired genomic analysis of squamous cell carcinoma transformed from EGFR-mutated lung adenocarcinoma. *Lung Cancer* **134**, 7–15 (2019).
47. John, R. M. & Rougeulle, C. Developmental Epigenetics: Phenotype and the Flexible Epigenome. *Front Cell Dev Biol* **6**, 130 (2018).

Tables

Table 1. Clinico-pathological features of the adenosquamous carcinomas at the time of resection

Table 1. Clinico-pathological features of the adenosquamous carcinomas at the time of resection

Patient	P118	P119	P120
Age (at resection)	62	80	66
TNM classification	pT4 pN0 cM0	pT2a pN0 cM0	pT3 pN1 cM0
Smoking status	former	unknown	former
Fraction of the adenocarcinoma component	70% (TTF1+)	60% (TTF1+)	60% (TTF1-)
Fraction of the squamous cell carcinoma component	30% (p40+)	40% (p40+)	40% (p40+)

Figures

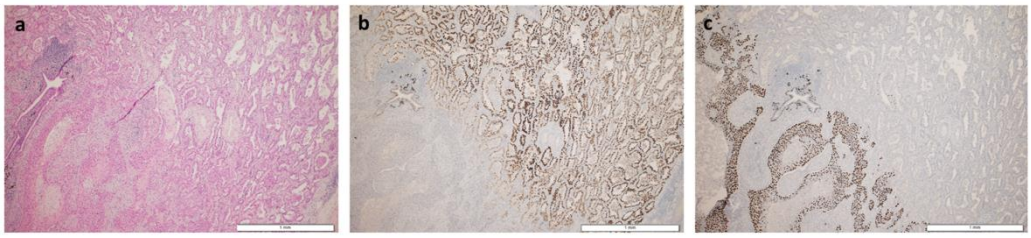


Figure 1. Adenosquamous carcinoma of the lung. Representative micrographs of a hematoxylin-eosin stained adenosquamous carcinoma (case P118) (**a**), the adenocarcinoma component immunostained with an antibody recognizing the lineage marker TTF-1 (**b**) and the squamous cell carcinoma immunostained with an antibody recognizing the specific marker p40 (**c**).

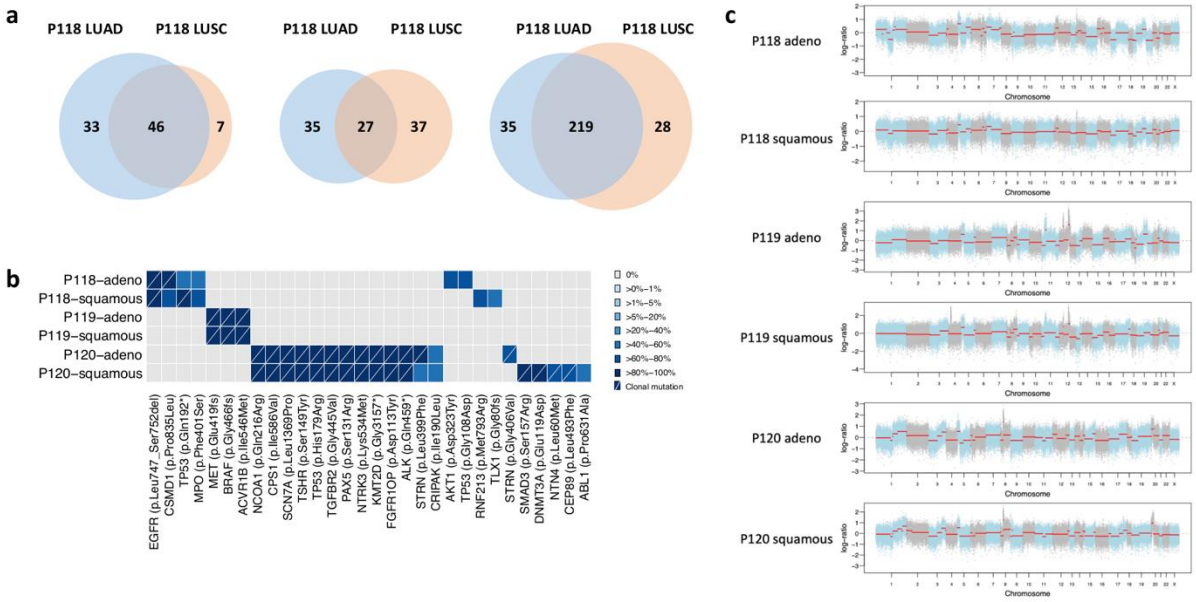


Figure 2. LUAD and LUSC within adenosquamous carcinoma are of monoclonal origin. (**a**) Venn diagrams display the number of non-synonymous mutations per patient that are common between the LUAD (blue) and LUSC (red) components. The size of the circles is proportional to the number of mutations. (**b**) Heatmap illustrates the cancer cell fraction of selected mutations. Clonal mutations are illustrated with a diagonal line. (**c)** Copy number aberrations of every sample are displayed.

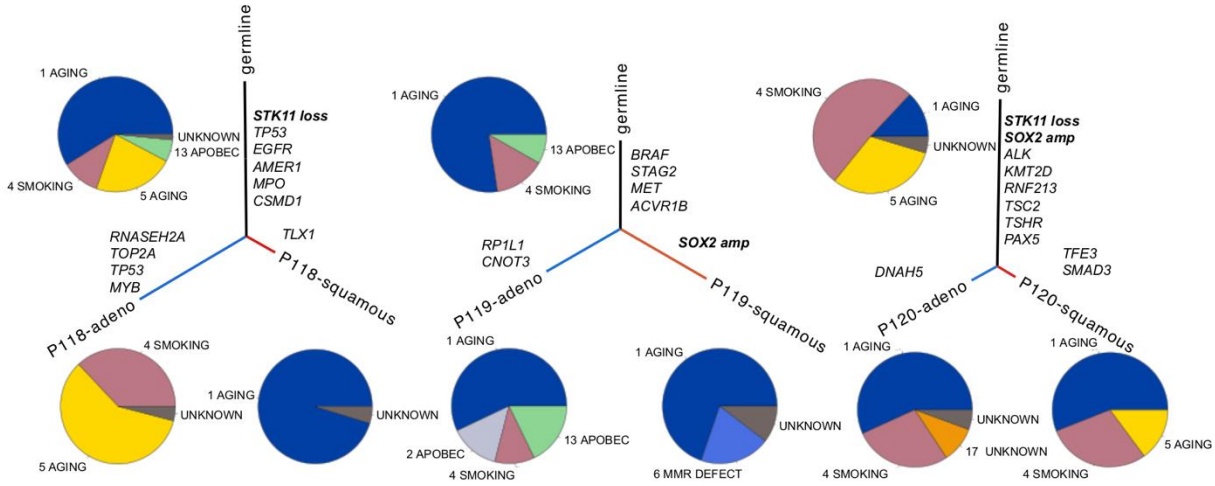


Figure 3. Mutational signatures and genomic alterations of truncal and branch mutations in LUAD and LUSC. Evolution of the somatic genetic alterations illustrates the changes in mutational processes. Pie charts delineate the proportion of mutational signatures. Black, blue, and red lines represent the trunk, the LUAD branch, and the LUSC branch, respectively. Mutations or indels in cancer genes are noted next to branches. *STK11* loss and *SOX2* amplification are noted in bold.

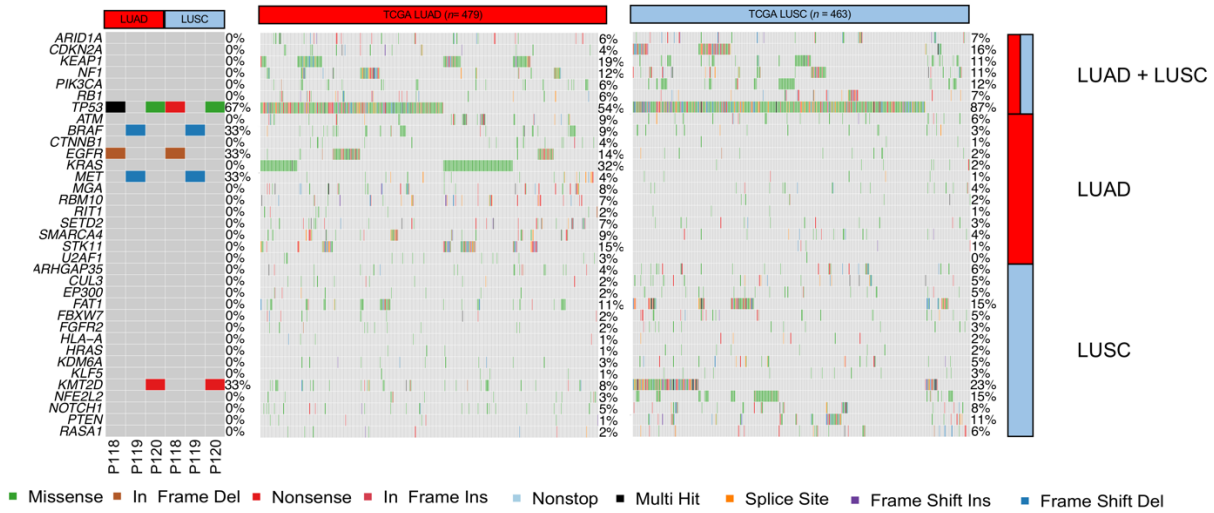
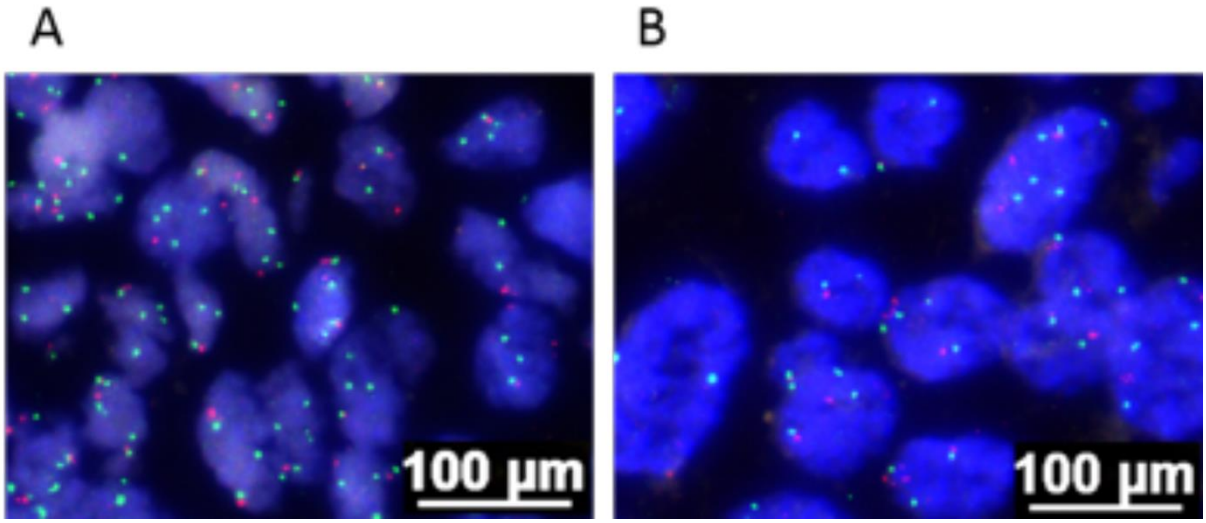


Figure 4. Non-synonymous mutation plot and comparison with TCGA LUAD and LUSC datasets. The heatmap depicts a comparison between the non-synonymous genes in the dataset of Bailey et al. that represents the most significantly mutated genes in classic LUAD (middle) and LUSC (right) [13]. The figure shows a comparison between the three ASC patients (left), TCGA LUAD (TCGA, Nature 2014) and LUSC (TCGA, Provisional) datasets from the cBioPortal (http://cbioportal.org) Left plot LUAD = adeno component of ASC; LUSC = squamous cell carcinoma component of ASC.

Order of the patients within the components of LUAD is P118, P119, P120 and in LUSC P118, P119, P120.

Supplementary Figure



Supplementary Figure 1. Amplification of *SOX2* in LUAD and LUSC components in P120. Genomic amplification of *SOX2* is pictured in both components. LUAD left; LUSC right. Centromeres red; *SOX2* green.

Supplementary Table

Supplementary Table 1. Sequencing metrics

Sample	Mean coverage	Total reads
P118-adeno	139.9	215677527
P118-squamous	136.13	209192556
P118-germline	122.28	187380110
P119-adeno	79.51	196031209
P119-squamous	47.49	123555979
P120-germline	48.54	115103389
P120-adeno	56.25	178413514
P120-squamous	73.03	141585922
P120-germline	58.06	111627453

3. Discussion

This work is composed of three research studies investigating the clonal evolution of metastatic LUAD (Results A), metastatic LUSC (Results B) and ASC (Results C). Each study has its particular focus and its own limitations. Taken together, we provide technical advancement and used state-of-the-art bioinformatics methods to ascertain beyond any doubt the clonal relationship of the different tumors and their matched metastases or their dichotomous entities.

3.1 Part A & B

Clonal tumor evolution of LUAD and LUSC

Genomic instability was defined as one of the hallmarks of cancers [138]. In recent years, many studies investigated genomic abnormalities in cancers on the whole exome or whole genome level to find cancer-type-specific mutations and copy number aberrations [98,99,197]. International consortia conducted large studies that revealed common hotspot mutations or recurrent chromosomal aberrations, which may represent potential targets [198–201]. However, most previous studies were focused on investigating primary tumors only, yet it is imperative to also investigate matched metastases, if one aims to understand tumor evolution and finding drivers of metastasis. Limited availability and the small size of metastatic samples are the main reasons why limited data are available on metastatic NSCLC. Another challenge in molecular profiling lies in the variable tumor cell population with less than 60% purity, on average, in LUAD and LUSC [124]. Low tumor cell content translates into a high presence of non-tumor cells in the tumor, which makes it difficult for bioinformatical approaches to detect somatic alterations, especially alterations that are subclonal and therefore present at low frequencies in the tumor. Therefore, lung cancer samples often must be excluded for genetic analyses. In this work, we overcame this limitation of low tumor cell content by flow cytometry. Flow cytometry allows enrichment for cells by physical features, such as size and cell granularity, but also DNA content, using DAPI, or the expression of specific markers, using antibody stainings.

Multiparameter flow-sorting increases the purity of tumor cells

Genomic characterization of tumor specimens is usually performed with whole tissue sections or laser microdissection microscopy prior to DNA extraction. Previous studies from our group and from others have established a sorting approach that used ploidy as a marker to identify aneuploid tumor populations and enrich them for downstream genomic analysis [202–205].

In this work, we further refined the nuclei flow-sorting approach to identify and enrich not only aneuploid tumors but also diploid tumors, which would otherwise be missed. Antibodies against markers expressed on tumor cells are used as a second parameter for distinguishing the tumor populations from non-malignant cells. We demonstrated the feasibility of this technique in LUAD and LUSC, where we sorted aneuploid and diploid tumor populations and sorted them at high purity. TTF-1 and p40 are both markers used in routine diagnostics to verify and distinguish LUAD from LUSC. Both are transcription factors localized in the nucleus. TTF-1 could be used to sort TTF-1-positive tumor populations in LUAD. However, we were unable to develop a valid protocol for flow-sorting by p40 for unknown reasons. Apparently, the existing p40 antibody is well established for IHC on FFPE tissue, but it is not suitable for fluorescence-based sorting. A possible explanation could be the harsh protocol of the staining that might damage the protein structure. Conversely, it was successful in LUAD with TTF-1. However, it was not possible to use p40 as an additional marker in the sorter. Therefore, we selected a pCK antibody that was previously shown to be applicable to flow-sorting other epithelial cancers, including cervical carcinoma [206]. The pCK antibody Clone MNF116 we used is specific against CK5, CK6, CK8, CK17, and CK19, which are known to be highly expressed in LUSC [207,208]. FF samples could not be sorted using pCK, while for FFPE, it was possible to use pCK. Nonetheless, FFPE tissues are fixed in 4% formaldehyde, which leads to subsequent crosslinking of proteins that might have enabled the flow-sorting procedure [209]. It appears that the CKs adhere to the nucleus after digestion of the cell membrane. This could explain why pCK is ineffective in unfixed FF samples. Where CKs are not fixed, they are washed away. It is important to mention that not only tumor cells express TTF-1 and CKs since TTF-1 is expressed naturally in, among others, alveolar cells type II and CKs, in general, in epithelial cells.

Isolating the germline control from the tumor sample

Our sorting approach permitted the distinction of several populations using TTF-1 or pCK in combination with the DNA intercalating agent DAPI. Diploid populations were sorted with DAPI alone or as a negative population in combination with TTF-1 or pCK and subjected to downstream analysis. After demonstration of copy number neutral genomic profiles in array comparative genomic hybridization (aCGH), the diploid populations were used as germline for subsequent sequencing. In addition, DNA sequencing provided further evidence of the non-tumor character of the pure DAPI diploid, diploid/pCK negative, or diploid/TTF-1 negative populations. Both projects A and B used the diploid population as germline DNA control. This turned out to be beneficial for retrospective studies since most of the samples in the archive do not have any matched normal DNA and many patients passed away over time.

Stable ploidy status during evolution

Genomically unstable tumors can accumulate mutations and aberrations that lead to increased and also decreased genome size. When the replication machinery and DNA repair mechanisms are compromised i. e. by mutations or aberrations, WGD can occur causing an abnormal karyotype [210]. DAPI intercalates the DNA and leads to a relative estimation of the ploidy of the tumor. Therefore, it is possible to isolate aneuploid tumors. In contrast diploid tumors remain undetected amid a background of benign diploid cells. Overall, we identified and isolated 39 populations from 16 patients with LUAD and 13 populations of six patients with LUSC. The LUAD and LUSC cohort had 23% (9/39) and 31% (4/13) diploid tumor populations, respectively. In general, we observed that the ploidy status was almost identical in the primary tumor and their matched metastases. There were only few exceptions as in P109 from the LUSC cohort.

The primary tumor of P109 from the LUSC cohort was a FF sample and, therefore, sorted with DAPI. The sorted tumor population was aneuploidy (3.6 *N*). The two FFPE LUSC metastases consisted of diploid pCK positive tumor populations (Met1 = 2*N*, Met2 = 1.8*N*) and displayed high genomic concordance with the aneuploid primary. This raises the question, whether the clone was diploid and undetected in the primary tumor that gave rise to the metastases. This is in line with the algorithm FACETS that calculated a ploidy of 2*N* for the primary and 1.8*N* for both metastases. The calculated ploidy of the metastases is consistent with the ploidy of the flow cytometer, whereas the primary aneuploid population (3.6*N* estimated by flow cytometer) was calculated as 2*N*. Algorithms as FACETS can not distinguish between diploid or tetraploid cells (personal discussion). Moreover, the results could indicate that the DAPI sorted tumor population in the primary tumor was in the G2M phase and therefore contained approximately the double amount of DNA. Taking this into account, it appears that ploidy did not change much in both cohorts, even after 12.7 years as we saw in patient 49 (LUAD). Ploidy represents a constant during lung tumor evolution, which is not affected by time of recurrence or spatial differences of the second tumor manifestation. In the timeline of tumor evolution, these genomic abnormalities enhancing ploidy occur early.

Limited heterogeneity in LUAD and LUSC

Except in the case of patient 42, every tumor was monogenomic with one constant ploidy. Monogenomic primary tumors disseminated a clone that was found in the metastases. After these result, one can conclude that the clones followed a linear evolutionary pattern. During the development and evolution of cancer, most genomic abnormalities are accumulated early in the tumorigenesis until it becomes invasive. Acquired plasticity of the primary tumor enables the dissemination of genetically equal clones that travel with the bloodstream and manifest in a distant organ is called distant metastasis. In this model of linear evolution, the dominating

clone leaves very late and has a long trunk that contains many mutations and CNVs. This short genetic divergence is also seen in the CNVs, which exhibit a large overlap between the primaries and their metastases in both cohorts. The WES data of our LUSC cohort allowed us to infer the mutational signatures that shape the mutational landscapes. Unsurprisingly, our results showed a significant proportion of aging signature 1 and smoking signature 4. We discovered further that the mutational signatures changed over time. Remarkably, we could compare the signature between the private mutations in the primary tumor and the metastasis, and we detected completely distinct profiles. Groundbreaking studies investigated mutational signatures and categorized them to lung cancer and its subtypes and furthermore, identified timing of occurrence in the evolution of the tumor [119,161]. Signatures 1, 4 and 5 are known to be a major part of the trunk, whereas APOBEC signatures 2 and 13 have a lower impact. As we discovered in our cohort, late mutations still display signature 1 and 5 to a large extent, but signatures 4 and 2/13 are reversed [161]. Signatures 1 and 5 were pervasive due to spontaneous mutations early and late in the tumor evolution [211]. Whereas smoking is responsible for a significant proportion of the truncal mutations that is overtaken by APOBEC over the time. APOBEC promotes late mutations in primary tumor and metastases. APOBEC is a protein complex that functions as enzyme-catalyzed deaminase, that is more likely to see in LUSC than in LUAD tumors [211]. It is not clear what triggers APOBEC, especially at an evolutionary late stage of the tumor. It induces significant subclonal diversity and contributes directly to the genomic instability [169]. Different signatures were detected in all three tumors of P109 (LUSC). Whereas the private mutations in Met1 were partially caused by a defective APOBEC system, Met2 did not show any signs of APOBEC but, instead, showed a distinct signature 5. The primary had none of the signatures. ITH is made up by the type of mutations that occur differently in the metastases than in their precursors. Strikingly, we do not see any changes in terms of mutational signatures in temporal recurrent or spatial different tumor descendants that is line with a pan-cancer metastases study [212].

The further development of the clones tends to the parallel evolution model, where several clones grow simultaneously. Similar to the distant metastatic spreading of P109, the primary tumor of patient 22 spread one metastasis to a different lobe and another one to the brain (polyclonal spread) that developed as genomically distinct. The brain metastasis was genetically closer to the primary tumor, despite the spatial difference, compared to the intrapulmonary metastasis. Interestingly, we saw the same effect in LUSC in P109, where the spine metastasis was genetically closer to the primary tumor than to the lymph node metastasis. Thus, the physical distance between the tumor manifestations appear to not correlate with genomic relationships.

Patient 42 represents an ideal case for studying the branched evolution. The polygenomic primary tumor consisted of one diploid (2.0*N*) and two aneuploid (3.4*N*, 4.7*N*) tumor populations. Only a small number of CNVs and mutations were shared between the tumor populations, which indicated very early dissemination and independent development of the tumor populations, which was proved by p16 staining. Only the diploid population had lost both copies of the *CDKN2A* gene and consequently lacked expression of p16 shown by IHC. Another metastasis-seeding mechanism was discovered in patients 22, 31, and 36, where more than one metastasis was available and investigated. However, the primary tumors gave monoclonal rise to metastasis, which in turn spread to a further metastasis (metastasis-to-metastasis spread). We detected in the first metastasis of patient 36 a *CDKN2A* mutation that was further propagating to the other metastasis. This is unequivocal evidence for the further progression of the metastases and propagation of newly acquired mutations. Not only truncal mutations were passed over.

Ongoing genomic instability in CIN-related genes leads to the misregulation of the replication and DNA repair system, which can contribute to ITH. ITH can be partly responsible for the ability to adapt rapidly to a new environment, giving rise to metastasis or the emergence of a resistant clone. This apparently endless flexibility allows cancer to be a dynamic disease that adapts rapidly to environmental changes, which explains why most cancers become resistant to targeted therapies [169]. Recently, it was demonstrated that mutational ITH was affecting the patient outcome. Strikingly, CNV ITH was associated with the patient outcome. Patients with an intermediate level of ITH of CNVs are associated with recurrence or death [161]. Despite that, it is imperative that the focus relies on truncal events as potential hot spot mutations, such as *EGFR*. Our data show a long trunk of mutations that can be used as a target in primary tumors and their metastases. Most of the mutations and aberrations are shared, which makes early truncal events an attractive target. In addition to the daily routine testing, the rate of CIN should be considered in order to predict the malignancy outcome of the tumor.

The results, taken together, demonstrate that the type of evolution is not always unambiguous. A biopsy represents a snapshot of the tumor evolution. Depending on the timing of the biopsy the type of evolution can change. A biopsy taken during tumorigenesis will perhaps find only the major clone with a long trunk driven by several exogenous and endogenous processes. Biopsies of a progressed tumor will have more subclonal diversity, which complicates the treatment. It is thinkable that the tumors begin to develop in a linear evolution but change track after dissemination and develop further independently from each other.

Limitations

Our exploratory work has several important limitations. It was previously planned to use targeted sequencing and aCGH, which were applied in the LUAD cohort. Access to WES was changed in our lab and became available only for the LUSC cohort. On the one hand, aCGH does not have such a high resolution as WES technology does and, therefore, smaller aberrations are not revealed [213]. On the other hand, copy number analysis from WES captures only the coding regions, which misses a great deal of information that is investigated in the aCGH. Further, targeted sequencing of 409 cancer genes shed light on the clonal relationship on a selected gene level. In fact, panel sequence, as we used, most likely underestimates mutational ITH, since driver genes tend to be clonal. WES analysis of the LUSC cohort was similar, but provided a much broader view of driver and passenger mutations, supporting our previous results in LUAD. WES also made it possible to study the evolution of the mutational signature and identify the largest external and internal mutation-causing factors, such as smoking and aging, respectively. However, WES is also not able to identify translocations or inversions. Flow-sorting yielded in high purity, but needed also a high amount of tissue. Therefore, several distinct tumor populations were isolated and sorted but could not be sequenced, because of the low amount of DNA. Furthermore, the limited availability of the samples, especially with matched metastases, makes it difficult to build a large cohort. The fact that LUSC is less common than LUAD is reflected by the limited size of our LUSC cohorts, which precludes definitive conclusions. Despite these limitations, the revised nuclei flow-sorting approach presented the unique opportunity to use also samples with a low tumor cell content. A disadvantage of the flow-sorting lies in the relatively large amount of tissue needed to obtain sufficient DNA.

Our qualitative approach uncovered important findings. We aimed to investigate the molecular composition of cancer cells in the primary tumor and matched metastases of LUAD and LUSC that had been collected over the years. We demonstrated a multiparameter flow-sorting technique that allowed the use of low tumor cellularity samples and sorted LUAD and LUSC to an unprecedented purity, which is applicable to any other cancer and allows further downstream analyses such as sequencing. Our work deciphered the clonal relationship between the primary tumor and metastases in NSCLC.

3.2 Part C

Studying ASC and its relevance

Part C of this thesis covers the investigation of pulmonary ASC, which comprises the two above discussed tumor entities, LUAD and LUSC. Over the past ten years, knowledge about genomics across many cancer types has increased significantly. Most large cohorts used frequent and easily accessible tumors such as those in the breast or lung [107,214]. Rare diseases such as ASC are not studied extensively for several reasons. Firstly, ASC is a rare tumor and, therefore, the availability of usable cases is very low. In general, the prevalence is higher in elderly male smokers, but still, it constitutes only a small number [215]. Secondly, a preoperative diagnosis of ASC is somewhat challenging. Biopsies are usually performed once and often contain only one of the two components LUAD or LUSC. It is reported that only 2% of ASC cases were correctly diagnosed before surgery [216]. Another study observed that up to 20% of NSCLC are misclassified due to undersampling, even with IHC [217]. It appears that more LUADs or LUSCs are dichotomous tumors than we might think, since only one part of the tumor in the biopsied sample is seen. To make a diagnosis of ASC, both components are needed. Currently, no marker can easily identify ASC on a molecular level. In 2017, a group proposed the use of serum carcinoembryonic antigen and squamous cell carcinoma antigen as a supplementary diagnostic test. They demonstrated significant differences between ASC and LUAD that could, in the future, be used as an indicator for ASC [218]. An early biomarker could help reduce misclassification of the tumors.

Lessons learnt: Polyclonal or monoclonal origin?

ASC heterogeneity is given not only by the two components but also by the morphological heterogeneity of the components. The expression of TTF-1 and p63 was demonstrated to be expressed in various distinct areas [219]. Furthermore, the growth pattern can vary depending on the proportions of the two components. Immune-deficient mice were injected with cancer-stem-like cells of patients to track the appearance of the tumors. The result was surprising since LUAD-like, LUSC-like carcinoma or ASC cells emerged [220]. This experiment goes along with a monoclonal origin and made it obvious that there must be a cell, which is able to differentiate into different entities.

Gene expression of the three tumor types LUAD, LUSC, and ASC was investigated in rats [73]. LUADs and LUSCs showed differentially expressed genes compared to each other and to ASC. Interestingly, ASC was presented as a transcriptomic hybrid of LUAD and LUSC [73]. However, it is worth mentioning that previous studies of ASC did not macrodissect the ASC tumor into its single components. Further evidence for a collision is provided by another study

that used ASC cell lines that showed distinct biological properties of the components, which suggested a polyclonal origin of ASC that once grew together. This hypothesis was refuted by several genomic studies that found several genes in common in the separated components.[75,77,79,221]. The existence of mutations occurring in two adjacent tumor entities strongly indicates a monoclonal origin. However, we are the first to present data on macrodissected components that were subjected to WES. We delivered unprecedented proof of a monoclonal origin of the two components LUAD and LUSC. We not only demonstrated that at the mutational level, but also CNVs between the two components were shared and have a common cell of origin.

ASC as a transitional state

In addition to multipotent cells that have the potential for multiple differentiation methods into different lineages by nature, other studies suggest strong plasticity due to induced transdifferentiation from LUAD to LUSC. Ji and colleagues showed that *STK11*-deficient mice demonstrated a shorter latency in tumor development and more frequent metastasis [222]. Further mouse experiments showed that *STK11*-deficiency changed the phenotype from LUAD to LUSC and upregulated the squamous differentiation gene *TP63*. Interestingly, pharmacological inhibition of *LOX* leads to promotion of the transdifferentiation, whereas overexpression results in an inhibition of the transition [223]. Another study used *LOX* inhibition in $Kras^{G12D}/Trp53^{L/L}$ mice, which had wild type *STK11* alleles, and again a transdifferentiation was observed. Notably, these mice were known to produce only LUAD [224]. In consequence, squamous differentiation was triggered without the need to lose the *STK11* gene. Evidently, *Lox* inhibition appears to decrease collagen deposition and extracellular matrix remodeling [223]. It is known that the loss of ECM boosts the release of reactive oxygen species (ROS) [225]. ROS, in turn, is again a known trigger for the transdifferentiation of LUAD to LUSC [226]. Transdifferentiation was also demonstrated by an *STK11* loss concomitant with ectopic *SOX2* expression that leads from LUAD to LUSC [227]. A prime example of tumor plasticity is demonstrated by the loss of *NKX2-1,* which leads to gut-like cell differentiation. Interestingly, *SOX2* is reactivated as soon as *NKX2-1* is lost [82]. Active *SOX2* suppresses *NKX2-1* activity, which in turn accelerates the transdifferentiation process when *NKX2-1* is lost [227]. Comparing to our data, *SOX2* was only amplified in both components of P120 LUSC and P119 LUSC was not lost. Once again it is demonstrated that mouse experiments cannot directly be transferred to humans. However, it is still possible that there is an underlying transdifferentiation mechanism, which is not only caused by genomic alterations. Recently, it was reported how epigenetic silencing of *FOXA1* and biallelic loss of *PTEN* drives squamous differentiation in bladder cancer [228]. It is possible that LUAD-to-LUSC transformation is governed by a combination of genomic, epigenomic and transcriptomic processes. Transdifferentiation does

not occur exclusively in the lung. The biological phenomenon of transdifferentiation has also been observed in colon cancer, prostate cancer, pancreatic cancer, thyroid gland carcinoma, and gastric cancer [229–233].

LUAD cannot only transdifferentiate to LUSC. Up to 10% of *EGFR*-mutant LUAD can transdifferentiate to SCLC [234]. Interestingly, 69% of the *EGFR*-mutants harbored an *EGFR* exon 19 deletion and 25% showed an L858R mutation. It was determined that all *EGFR*-mutated LUADs that transdifferentiate to SCLC lost the *RB1* gene, which suggests an important role in the cell fate decision of SCLC [235,236]. Notably, the transition from LUSC to SCLC was also observed after surgery or radiation and chemotherapy (Figure 16) [237].

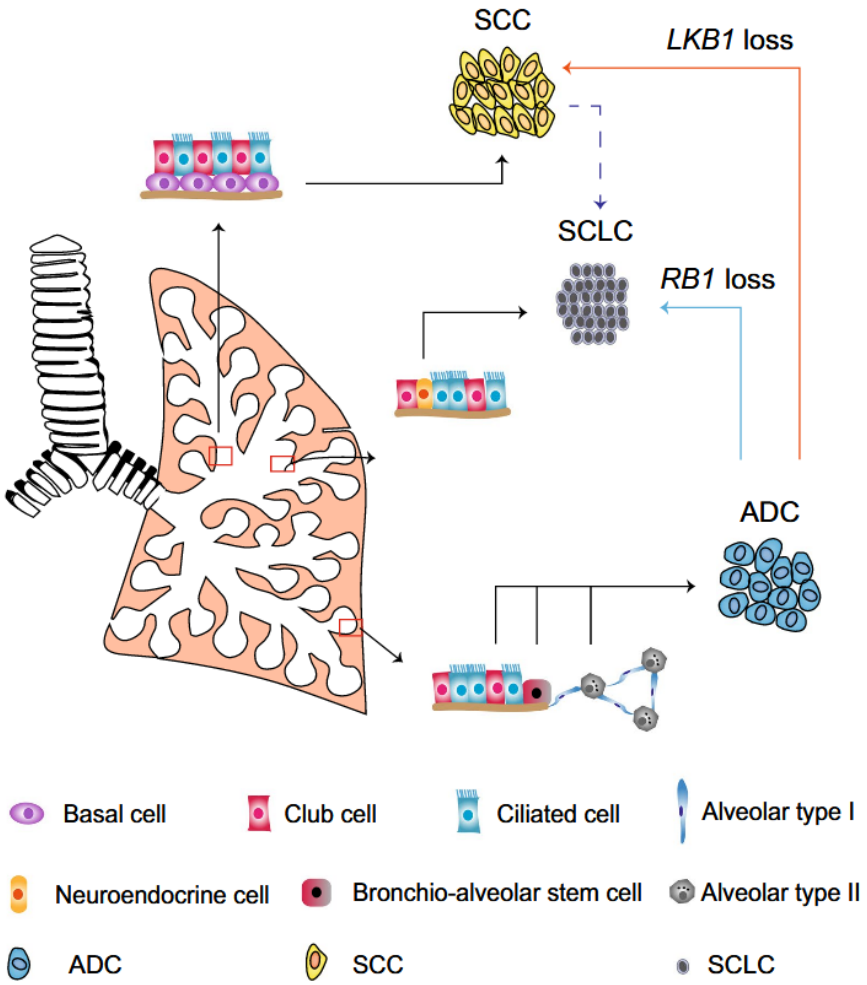


Figure 16. Transdifferentiation mechanisms of lung cancer. LUAD can transdifferentiate to SCLC while losing RB1 or to SCC while losing LKB1 that is encoded by *STK11*. SCC was also observed to change to a SCLC phenotype [modified from [238]].

The unsolved question of the right treatment

There are no evidence-based guidelines on how to treat advanced ASC, since the two entities as classical diseases show different susceptibility to treatment. In 2013, a study showed that ASC and LUAD patients showed similar characteristics regarding age, sex, and smoking habits. As in LUAD, new personalized approaches are emerging as targeted therapies that consider the mutational background. It is known that ASC has a similar prevalence of *EGFR* mutants as LUAD [239]. Therefore, as another first-line therapy, *EGFR*-TKIs can be used as in *EGFR*-mutated non-squamous lung cancer [240]. The LUSC component was similar to classical LUSC and the same applies to the LUAD component and classical LUAD [241]. Nevertheless, further research is required on larger cohorts.

Limitations

Our study has important limitations. With a study cohort of three patients, it is rather a qualitative study to investigate the genomic landscape of ASC than a quantitative approach to provide definitive conclusions. Further, it was not possible to collect all the clinicopathological criteria to conciliate it together with the obtained data. Due to limited sequencing depth, it is possible that potential subclonal ITH was overlooked. Furthermore, WES gave a resolution to study the whole exome. For discovering the genomic landscape, not only bigger studies are necessary also whole-genome sequencing approaches, to grasp all genomic changes on several levels. Finally, it is worth to mention that even with meticulate macrodissection, we cannot exclude the possibility that there was a small degree of contamination between the components.

To conclude, we macrodissected three pulmonary ASC tumors and to our knowledge, we are the first to present WES data on the single components. We provide unequivocal evidence for a common ancestor cell. Furthermore, we give strong evidence for a monoclonal relationship based on the high concordance of the genomic profiling of the two components.

However, the trunk is indicative of a progenitor population that gave rise to a LUAD entity. In this preliminary set of three ASC, we found hints that the transdifferentiation occurs from LUAD to LUSC in ASC. We identified genomic *SOX2* amplification and *STK11* loss in two of the three LUSC components, each, support previous pre-clinical evidence for a role of these genes in transdifferentiation of LUSC in ASC. Thus, it appears more likely that a change in transcriptional programs plays a role that is possibly governed by epigenetic mechanisms. Further studies are needed to resolve the mystery, if there is a transdifferentiation and how does it work.

4. General Conclusion

In this work, we investigated the evolutionary trajectory of the three different NSCLC subtypes LUAD, LUSC, and the mingled entity ASC. The presented technique enables a unique multiparameter flow-sorting approach to increase tumor purity from FF and FFPE tissue with low tumor cellularity. Our work on refining the technology to ploidy and antibody-based flow-sorting guided us to an unprecedented purity of tumor content, which allowed us to study the clonal evolution between primary tumors and matched metastases in LUAD and LUSC.

Obtained data on gene copy numbers and mutational patterns demonstrated limited heterogeneity with stable ploidy over time and space in our study cohorts. Primary tumors and metastases shared mutations and CNVs in the trunk demonstrating a clear phylogenetic lineage of a disseminating clone from the primary to manifest in a distant organ. Moreover, we found that most of the tumor-metastasis pairs underwent linear evolution. Nevertheless, we discovered that metastatic seeding is not only coming from the primary tumor. Instead, we identified metastases seeding further metastases and still shared genomic peculiarities. Parallel evolution took place, where metastases gave seeded further metastases. Here, we proved a close clonal relationship, where most of the ubiquitous genomic abnormalities happen as early events in tumorigenesis before metastatic seeding.

Similarly, we investigated the evolutionary shift in ASC. Macrodissection of the single components LUAD and LUSC followed by sequencing analysis, revealed a shared genomic landscape. Mutations and chromosomal aberrations showed a high congruence leading to the unequivocal conclusion of one common ancestor cell. Most likely LUAD differentiated to LUSC due to yet unknown mechanisms. With this work, we laid the foundation for future studies to uncover the mechanisms of transdifferentiation. Besides larger cohorts, future studies should consider not only the genomic heterogeneity of the dichotomous tumor but also investigating transcriptomic and epigenomic approaches to reveal the unknown mechanism.

5. Materials and Methods

The methods are described in each publication or manuscript (chapter Results) of this thesis. For the reader's convenience and clarity, some of the mentioned methods are expanded below.

Sample processing of Part B

Sample processing and flow-sorting was performed as described in Results Part B Material and Methods. Deviant from that Pan-Cytokeratin antibody (pCK, clone MNF116, Dako, Glostrup, Denmark, Code-Nr. M0821) was used as an additional parameter to DAPI with a concentration of 2 µg/ml. Nuclei from FF tumor specimen were sorted with DAPI (*n* = 7, four patients) only. FFPE nuclei were sorted with pCK, where it was possible (*n* = 6, three patients).

Whole exome sequencing

WES of tumor DNA and matched germline DNA was performed using the SureSelect Human All Exon V6 Kit (Agilent) according to manufacturer's guidelines. Paired-end 100-bp reads were generated from the Illumina NovaSeq 6000. Sequencing was conducted by CeGaT (Tübingen, Germany).

Bioinformatic processing for WES

State-of-the-art bioinformatic processing was performed on the high computing cluster sciCORE of the University of Basel (https://scicore.unibas.ch/). Parameter settings in below mentioned algorithms were used by default or otherwise described explicitly below (Figure 17).

Data processing

Paired-end reads in FASTQ format were aligned to the reference human genome GRCh37 applying Burrows-Wheeler Aligner (BWA, v0.7.12) [242]. Local realignment, PCR duplicate removal and base quality adjustment were performed using the Genome Analysis Toolkit (GATK, v3.6) and Picard (https://broadinstitute.github.io/picard/) [243].

Variant calling

Somatic single nucleotide variants (SNVs) and small insertions and deletions (indels) were identified using MuTect (v1.1.4) and Strelka (v1.0.15), respectively [244,245]. C:G>T:A SNVs with variant allelic fraction (VAF) less than 10% were discarded to reduce false positive results from artifacts from formalin fixation. Otherwise, SNVs or indels with VAF < 1% or that were

supported by fewer than 3 reads were discarded. If a mutation was found in one tumor in a given patient or in a component of a given tumor, a cut-off of two reads was applied. Variant annotation was performed by SNPEFF software v4.1[246].

Cancer genes and TCGA database

Cancer genes were annotated using the cancer gene lists obtained by Kandoth, Lawrence and Cancer Gene Census datasets or TCGA [98,99,247,248,249,].

CNVs and assessment of clonal composition

Deciphering allele-specific CNVs, FACETS (v0.5.5) was used [250,251]. Cancer cell fractions (CCF) were calculated using ABSOLUTE (v1.0.6) [252]. A mutation was classified as clonal, if its probability of being clonal was > 50% or if the lower bound of the 95% confidence interval of its CCF was > 90%. Mutations were considered as subclonal if they did not meet the mentioned criteria [253,254]. If a mutation was found in both tumors/components of one patient/biopsy, it was considered as 'trunk/clonal'. Mutations that were detected in only one tumor/component of the patient/tumor were considered as 'branch/subclonal'. 'Private' mutations were defined as unique mutations that are found in one tumor/component but not shared with the other tumor/component.

All synonymous and nonsynonymous mutations that passed filtering were considered for the purpose of constructing phylogenetic trees. Trees were generated applying binary presence/absence matrices built from the regional distribution of variants within the tumor. The R Bioconductor package phangorn was utilized to carry out the generation of unrooted trees with the parsimony ratchet method [255,256]. Branch lengths were determined using the acctran function.

Mutational signature

Decomposition of mutational signatures was conducted using deconstructSigs software by selecting LUAD and LUSC specific mutational signatures based on the set of 30 mutational signatures ('signature.cosmic') that were observed in LUAD and LUSC [257,258].

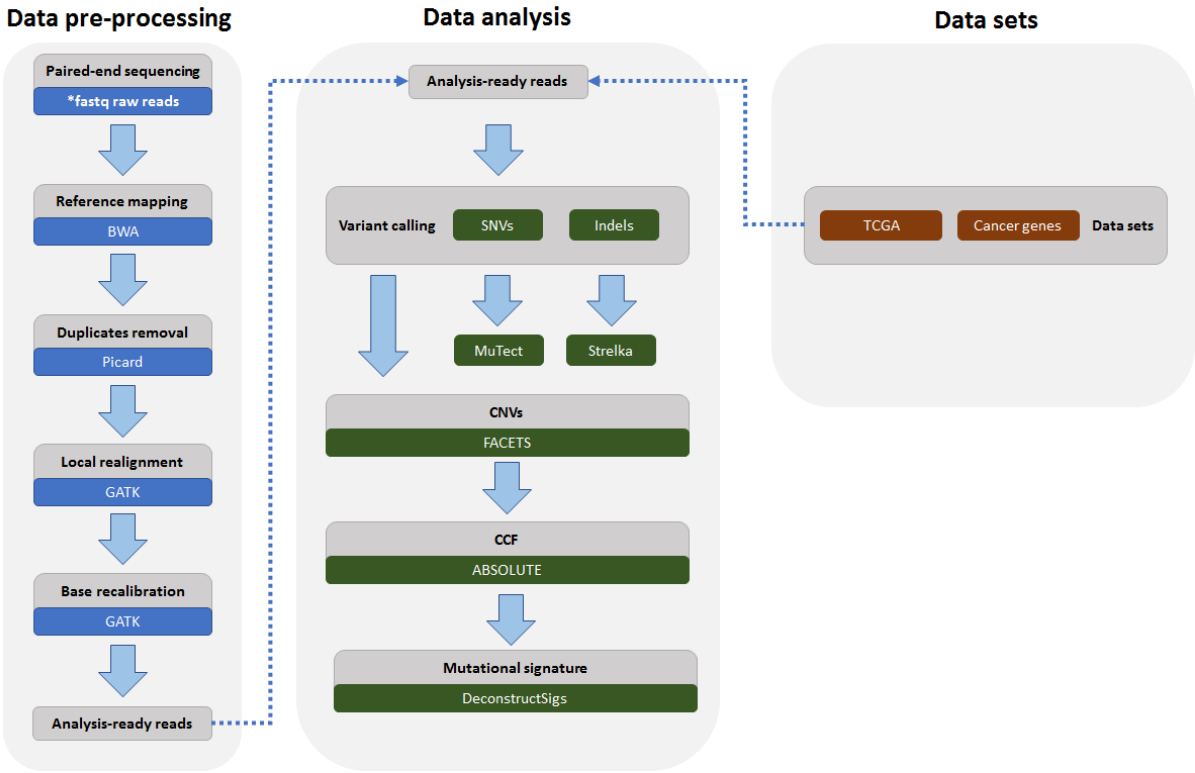


Figure 17. From raw reads to analysis results. Bioinformatic processing of WES data. Raw reads in form of base-called fastq files are pre-processed to analysis-ready reads (blue). Data analysis of processed reads to call single nucleotide variants (SNVs), insertions and deletions (indels), copy number variations (CNVs) and calculate cancer cell fractions (CCF) and mutational signatures (green). Several data sets were used for verification and or comparison (red)

6. References

1. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer* **136**, E359–E386 (2015).
2. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **68**, 394–424 (2018).
3. Tarver, T. Cancer Facts & Figures 2012. American Cancer Society (ACS). *Journal of Consumer Health On the Internet* **16**, 366–367 (2012).
4. Torre, L. A., Siegel, R. L. & Jemal, A. Lung Cancer Statistics. *Adv. Exp. Med. Biol.* **893**, 1–19 (2016).
5. Cheng, T.-Y. D. *et al.* The International Epidemiology of Lung Cancer: Latest Trends, Disparities, and Tumor Characteristics. *J. Thorac. Oncol.* **11**, 1653–1671 (2016).
6. de Groot, P. M., Wu, C. C., Carter, B. W. & Munden, R. F. The epidemiology of lung cancer. *Transl Lung Cancer Res* **7**, 220–233 (2018).
7. Available at: https://gco.iarc.fr/today/data/factsheets/populations/900-world-fact-sheets.pdf. (Accessed: 12th July 2019)
8. Reck, M. & Rabe, K. F. Precision Diagnosis and Treatment for Advanced Non–Small-Cell Lung Cancer. *New England Journal of Medicine* **377**, 849–861 (2017).
9. Website. Available at: https://www.verywellhealth.com/lung-cancer-survival-rates-by-type-and-stage-2249401. (Accessed: 28th November 2019)
10. Pikor, L. A., Ramnarine, V. R., Lam, S. & Lam, W. L. Genetic alterations defining NSCLC subtypes and their therapeutic implications. *Lung Cancer* **82**, 179–189 (2013).
11. Travis, W. D., Colby, T. V., Corrin, B., Shimosato, Y. & Brambilla, E. Histological Classification of Lung and Pleural Tumours. *Histological Typing of Lung and Pleural Tumours* 21–24 (1999). doi:10.1007/978-3-642-60049-4_2
12. International Agency for Research on Cancer. *Who Classification of Tumours of the Lung, Pleura, Thymus and Heart*. (World Health Organization, 2015).
13. Patel, M. I., Cheng, I. & Gomez, S. L. US lung cancer trends by histologic type. *Cancer* **121**, 1150–1152 (2015).
14. Zheng, M. Classification and Pathology of Lung Cancer. *Surgical Oncology Clinics of North America* **25**, 447–468 (2016).
15. Häggström, M. Medical gallery of Mikael Häggström 2014. *WikiJournal of Medicine* **1**, (2014).
16. Kenfield, S., Wei, E., Colditz, G., Stampfer, M. & Rosner, B. Comparison of aspects of Smoking Among Four Histologic Types of Lung Cancer. *American Journal of Epidemiology* **163**, S108–S108 (2006).
17. Liu, B.-Q. *et al.* Emerging tobacco hazards in China: 1. Retrospective proportional mortality study of one million deaths. *BMJ* **317**, 1411–1422 (1998).
18. Wynder, E. L. & Graham, E. A. Etiologic factors in bronchiogenic carcinoma with special reference to industrial exposures; report of eight hundred fifty-seven proved cases. *AMA Arch. Ind. Hyg. Occup. Med.* **4**, 221–235 (1951).
19. Doll, R. & Hill, A. B. Smoking and carcinoma of the lung. Preliminary report. 1950. *Bull. World Health Organ.* **77**, 84–93 (1999).

20. Langer, C. J., Besse, B., Gualberto, A., Brambilla, E. & Soria, J.-C. The Evolving Role of Histology in the Management of Advanced Non–Small-Cell Lung Cancer. *Journal of Clinical Oncology* **28**, 5311–5320 (2010).
21. National Cancer Institute (U.S.). *Cigars: health effects and trends*. (1998).
22. Higgins, I. T. T., Mahan, C. M. & Wynder, E. L. Lung cancer among cigar and pipe smokers. *Preventive Medicine* **17**, 116–128 (1988).
23. Koul, P. A. *et al.* Hookah smoking and lung cancer in the Kashmir valley of the Indian subcontinent. *Asian Pac. J. Cancer Prev.* **12**, 519–524 (2011).
24. Hoffmann, I. & Hoffmann, D. The changing cigarette: chemical studies and bioassays. *Tobacco* 93–126 (2010). doi:10.1093/acprof:oso/9780199566655.003.0006
25. Hecht, S. S. Biochemistry, biology, and carcinogenicity of tobacco-specific N-nitrosamines. *Chem. Res. Toxicol.* **11**, 559–603 (1998).
26. Hecht, S. S. Tobacco carcinogenesis: mechanisms and biomarkers. *Tobacco* 127–154 (2010). doi:10.1093/acprof:oso/9780199566655.003.0007
27. The Health Consequences of Involuntary Exposure to Tobacco Smoke: A Report of the Surgeon General. *PsycEXTRA Dataset* (2006). doi:10.1037/e601432007-001
28. Hackshaw, A. K. Lung cancer and passive smoking. *Stat. Methods Med. Res.* **7**, 119–136 (1998).
29. Öberg, M., Jaakkola, M. S., Woodward, A., Peruga, A. & Prüss-Ustün, A. Worldwide burden of disease from exposure to second-hand smoke: a retrospective analysis of data from 192 countries. *The Lancet* **377**, 139–146 (2011).
30. Matt, G. E. Households contaminated by environmental tobacco smoke: sources of infant exposures. *Tobacco Control* **13**, 29–37 (2004).
31. Vineis, P. *et al.* Environmental tobacco smoke and risk of respiratory cancer and chronic obstructive pulmonary disease in former smokers and never smokers in the EPIC prospective study. *BMJ* **330**, 277 (2005).
32. Vineis, P. *et al.* Lung cancers attributable to environmental tobacco smoke and air pollution in non-smokers in different European countries: a prospective study. *Environmental Health* **6**, (2007).
33. Spyratos, D. *et al.* Occupational exposure and lung cancer. *J. Thorac. Dis.* **5 Suppl 4**, S440–5 (2013).
34. Petersen, I. The Morphological and Molecular Diagnosis of Lung Cancer. *Deutsches Aerzteblatt Online* (2011). doi:10.3238/arztebl.2011.0525
35. Klein, F., Amin Kotb, W. F. M. & Petersen, I. Incidence of human papilloma virus in lung cancer. *Lung Cancer* **65**, 13–18 (2009).
36. Chan, T. Y. World Health Organization classification of tumours: Pathology & genetics of tumours of the urinary system and male genital organs. *Urology* **65**, 214–215 (2005).
37. Engels, E. A. *et al.* Cancer risk in people infected with human immunodeficiency virus in the United States. *Int. J. Cancer* **123**, 187–194 (2008).
38. HosgoodIII, H. D. *et al.* History of lung disease and risk of lung cancer in a population with high household fuel combustion exposures in rural China. *Lung Cancer* **81**, 343–346 (2013).
39. Brenner, D. R., McLaughlin, J. R. & Hung, R. J. Previous Lung Diseases and Lung Cancer Risk: A Systematic Review and Meta-Analysis. *PLoS ONE* **6**, e17479 (2011).

40. Wang, J. *et al.* Genetic predisposition to lung cancer: comprehensive literature integration, meta-analysis, and multiple evidence assessment of candidate-gene association studies. *Sci. Rep.* **7**, 8371 (2017).
41. Lissowska, J. *et al.* Family history and lung cancer risk: international multicentre case-control study in Eastern and Central Europe and meta-analyses. *Cancer Causes Control* **21**, 1091–1104 (2010).
42. Coté, M. L., Kardia, S. L. R., Wenzlaff, A. S., Ruckdeschel, J. C. & Schwartz, A. G. Risk of lung cancer among white and black relatives of individuals with early-onset lung cancer. *JAMA* **293**, 3036–3042 (2005).
43. Schwartz, A. G. & Ruckdeschel, J. C. Familial lung cancer: genetic susceptibility and relationship to chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **173**, 16–22 (2006).
44. Liu, P. *et al.* Familial aggregation of common sequence variants on 15q24-25.1 in lung cancer. *J. Natl. Cancer Inst.* **100**, 1326–1330 (2008).
45. Jonsson, S. *et al.* Familial risk of lung carcinoma in the Icelandic population. *JAMA* **292**, 2977–2983 (2004).
46. Landi, M. T. *et al.* A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am. J. Hum. Genet.* **85**, 679–691 (2009).
47. Thorgeirsson, T. E. *et al.* A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* **452**, 638–642 (2008).
48. Sun, S., Schiller, J. H. & Gazdar, A. F. Lung cancer in never smokers — a different disease. *Nature Reviews Cancer* **7**, 778–790 (2007).
49. Okazaki, I., Ishikawa, S., Ando, W. & Sohara, Y. Lung Adenocarcinoma in Never Smokers: Problems of Primary Prevention from Aspects of Susceptible Genes and Carcinogens. *Anticancer Research* **36**, 6207–6224 (2016).
50. Parkin, D. M., Bray, F., Ferlay, J. & Pisani, P. Global Cancer Statistics, 2002. *CA: A Cancer Journal for Clinicians* **55**, 74–108 (2005).
51. Gou, L.-Y., Niu, F.-Y., Wu, Y.-L. & Zhong, W.-Z. Differences in driver genes between smoking-related and non-smoking-related lung cancer in the Chinese population. *Cancer* **121 Suppl 17**, 3069–3079 (2015).
52. Swanton, C. & Govindan, R. Clinical Implications of Genomic Discoveries in Lung Cancer. *N. Engl. J. Med.* **374**, 1864–1873 (2016).
53. Tanoue, L. T. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma. *Yearbook of Pulmonary Disease* **2012**, 94–96 (2012).
54. Turner, B. M. *et al.* Napsin A, a new marker for lung adenocarcinoma, is complementary and more sensitive and specific than thyroid transcription factor 1 in the differential diagnosis of primary pulmonary carcinoma: evaluation of 1674 cases by tissue microarray. *Arch. Pathol. Lab. Med.* **136**, 163–171 (2012).
55. Mengoli, M. C. *et al.* The 2015 World Health Organization Classification of lung tumors: new entities since the 2004 Classification. *Pathologica* **110**, 39–67 (2018).
56. Herrmann, H., Bär, H., Kreplak, L., Strelkov, S. V. & Aebi, U. Intermediate filaments: from cell architecture to nanomechanics. *Nature Reviews Molecular Cell Biology* **8**, 562–573 (2007).

57. Davidson, M. R., Gazdar, A. F. & Clarke, B. E. The pivotal role of pathology in the management of lung cancer. *J. Thorac. Dis.* **5 Suppl 5**, S463–78 (2013).
58. Solis, L. M. *et al.* Histologic patterns and molecular characteristics of lung adenocarcinoma associated with clinical outcome. *Cancer* **118**, 2889–2899 (2012).
59. Rock, J. R. & Hogan, B. L. M. Epithelial Progenitor Cells in Lung Development, Maintenance, Repair, and Disease. *Annual Review of Cell and Developmental Biology* **27**, 493–512 (2011).
60. Liu, Q. *et al.* Lung regeneration by multipotent stem cells residing at the bronchioalveolar-duct junction. *Nat. Genet.* **51**, 728–738 (2019).
61. Atkinson, J. J., Adair-Kirk, T. L., Kelley, D. G., Demello, D. & Senior, R. M. Clara cell adhesion and migration to extracellular matrix. *Respir. Res.* **9**, 1 (2008).
62. Roth, J. A., Hong, W. K. & Komaki, R. U. *Lung Cancer*. (John Wiley & Sons, 2014).
63. Affandi, K. A., Tizen, N. M. S., Mustangin, M. & Zin, R. R. M. R. M. p40 Immunohistochemistry Is an Excellent Marker in Primary Lung Squamous Cell Carcinoma. *J Pathol Transl Med* **52**, 283–289 (2018).
64. Tatsumori, T. *et al.* p40 is the best marker for diagnosing pulmonary squamous cell carcinoma: comparison with p63, cytokeratin 5/6, desmocollin-3, and sox2. *Appl. Immunohistochem. Mol. Morphol.* **22**, 377–382 (2014).
65. Lu, Y. *et al.* Evidence that SOX2 overexpression is oncogenic in the lung. *PLoS One* **5**, e11022 (2010).
66. Campbell, J. D. *et al.* Genomic, Pathway Network, and Immunologic Features Distinguishing Squamous Carcinomas. *Cell Rep.* **23**, 194–212.e6 (2018).
67. Quinlan, R. A. *et al.* Patterns of expression and organization of cytokeratin intermediate filaments. *Ann. N. Y. Acad. Sci.* **455**, 282–306 (1985).
68. Blobel, G. A., Moll, R., Franke, W. W. & Vogt-Moykopf, I. Cytokeratins in normal lung and lung carcinomas. *Virchows Archiv B Cell Pathology Including Molecular Pathology* **45**, 407–429 (1984).
69. Kadota, K. *et al.* Comprehensive pathological analyses in lung squamous cell carcinoma: single cell invasion, nuclear diameter, and tumor budding are independent prognostic factors for worse outcomes. *J. Thorac. Oncol.* **9**, 1126–1139 (2014).
70. Yang, Y. *et al.* Spatial-Temporal Lineage Restrictions of Embryonic p63 Progenitors Establish Distinct Stem Cell Pools in Adult Airways. *Developmental Cell* **44**, 752–761.e4 (2018).
71. Cooke, D. T. *et al.* Survival comparison of adenosquamous, squamous cell, and adenocarcinoma of the lung after lobectomy. *Ann. Thorac. Surg.* **90**, 943–948 (2010).
72. Filosso, P. L. *et al.* Adenosquamous lung carcinomas: a histologic subtype with poor prognosis. *Lung Cancer* **74**, 25–29 (2011).
73. Bastide, K., Ugolin, N., Levalois, C., Bernaudin, J.-F. & Chevillard, S. Are adenosquamous lung carcinomas a simple mix of adenocarcinomas and squamous cell carcinomas, or more complex at the molecular level? *Lung Cancer* **68**, 1–9 (2010).
74. Kanazawa, H. *et al.* Transition from squamous cell carcinoma to adenocarcinoma in adenosquamous carcinoma of the lung. *Am. J. Pathol.* **156**, 1289–1298 (2000).
75. Vassella, E. *et al.* Molecular profiling of lung adenosquamous carcinoma: hybrid or genuine type? *Oncotarget* **6**, (2015).

76. Niho, S., Yokose, T., Kodama, T., Nishiwaki, Y. & Mukai, K. Clonal analysis of adenosquamous carcinoma of the lung. *Jpn. J. Cancer Res.* **90**, 1244–1247 (1999).
77. Kang, S. M. *et al.* Identical epidermal growth factor receptor mutations in adenocarcinomatous and squamous cell carcinomatous components of adenosquamous carcinoma of the lung. *Cancer* **109**, 581–587 (2007).
78. Jia, X.-L. & Chen, G. EGFR and KRAS mutations in Chinese patients with adenosquamous carcinoma of the lung. *Lung Cancer* **74**, 396–400 (2011).
79. Tochigi, N., Dacic, S., Nikiforova, M., Cieply, K. M. & Yousem, S. A. Adenosquamous carcinoma of the lung: a microdissection study of KRAS and EGFR mutational and amplification status in a western patient population. *Am. J. Clin. Pathol.* **135**, 783–789 (2011).
80. Hou, S. *et al.* Evidence, Mechanism, and Clinical Relevance of the Transdifferentiation from Lung Adenocarcinoma to Squamous Cell Carcinoma. *Am. J. Pathol.* **187**, 954–962 (2017).
81. Steele, V. E. & Nettesheim, P. Unstable cellular differentiation in adenosquamous cell carcinoma. *J. Natl. Cancer Inst.* **67**, 149–154 (1981).
82. Tata, P. R. *et al.* Developmental History Provides a Roadmap for the Emergence of Tumor Plasticity. *Dev. Cell* **44**, 679–693.e5 (2018).
83. Hirsch, F. R. *et al.* Lung cancer: current therapies and new targeted treatments. *Lancet* **389**, 299–311 (2017).
84. Schiller, J. H. *et al.* Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. *N. Engl. J. Med.* **346**, 92–98 (2002).
85. Scagliotti, G. *et al.* Treatment-by-histology interaction analyses in three phase III trials show superiority of pemetrexed in non-squamous non-small cell lung cancer. *J. Thorac. Oncol.* **6**, 64–70 (2011).
86. Wheeler, D. L., Dunn, E. F. & Harari, P. M. Understanding resistance to EGFR inhibitors-impact on future treatment strategies. *Nat. Rev. Clin. Oncol.* **7**, 493–507 (2010).
87. Lynch, T. J. *et al.* Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non–Small-Cell Lung Cancer to Gefitinib. *New England Journal of Medicine* **350**, 2129–2139 (2004).
88. Kazandjian, D. *et al.* FDA Approval Summary: Nivolumab for the Treatment of Metastatic Non-Small Cell Lung Cancer With Progression On or After Platinum-Based Chemotherapy. *Oncologist* **21**, 634–642 (2016).
89. Kobayashi, S. *et al.* EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **352**, 786–792 (2005).
90. Sequist, L. V. *et al.* Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci. Transl. Med.* **3**, 75ra26 (2011).
91. West, H. Prioritizing molecular markers to test for in the initial workup of advanced non-small cell lung cancer: wants versus needs. *Ann Transl Med* **5**, 371 (2017).
92. Soria, J.-C. *et al.* Osimertinib in Untreated EGFR-Mutated Advanced Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **378**, 113–125 (2018).
93. Elazezy, M. & Joosse, S. A. Techniques of using circulating tumor DNA as a liquid biopsy component in cancer management. *Comput. Struct. Biotechnol. J.* **16**, 370–378 (2018).

94. Oxnard, G. R. *et al.* Association Between Plasma Genotyping and Outcomes of Treatment With Osimertinib (AZD9291) in Advanced Non-Small-Cell Lung Cancer. *J. Clin. Oncol.* **34**, 3375–3382 (2016).

95. Syn, N. L., Teng, M. W. L., Mok, T. S. K. & Soo, R. A. De-novo and acquired resistance to immune checkpoint targeting. *Lancet Oncol.* **18**, e731–e741 (2017).

96. Herbst, R. S. *et al.* Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* **515**, 563–567 (2014).

97. Herbst, R. S., Morgensztern, D. & Boshoff, C. The biology and management of non-small cell lung cancer. *Nature* **553**, 446–454 (2018).

98. Network, T. C. G. A. R. & The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).

99. Network, T. C. G. A. R. & The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).

100. Chalmers, Z. R. *et al.* Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Medicine* **9**, (2017).

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

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

103. Rodin, S. N. & Rodin, A. S. On the origin of p53 G:C → T:A transversions in lung cancers. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **508**, 1–19 (2002).

104. Pfeifer, G. P. *et al.* Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene* **21**, 7435–7451 (2002).

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

106. Santos, E. *et al.* Malignant activation of a K-ras oncogene in lung carcinoma but not in normal tissue of the same patient. *Science* **223**, 661–664 (1984).

107. Campbell, J. D. *et al.* Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat. Genet.* **48**, 607–616 (2016).

108. Meng, F., Zhang, L., Ren, Y. & Ma, Q. The genomic alterations of lung adenocarcinoma and lung squamous cell carcinoma can explain the differences of their overall survival rates. *J. Cell. Physiol.* **234**, 10918–10925 (2019).

109. Sun, F. *et al.* Bioinformatics analyses of the differences between lung adenocarcinoma and squamous cell carcinoma using The Cancer Genome Atlas expression data. *Mol. Med. Rep.* **16**, 609–616 (2017).

110. Yang, Y., Wang, M. & Liu, B. Exploring and comparing of the gene expression and methylation differences between lung adenocarcinoma and squamous cell carcinoma. *J. Cell. Physiol.* **234**, 4454–4459 (2019).

111. Sanchez-Vega, F. *et al.* Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **173**, 321–337.e10 (2018).

112. Jiang, M. *et al.* Relative expressions of miR-205-5p, miR-205-3p, and miR-21 in tissues and serum of non-small cell lung cancer patients. *Mol. Cell. Biochem.* **383**, 67–75 (2013).

113. Yu, L. *et al.* Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers. *Int. J. Cancer* **127**, 2870–2878 (2010).

114. Garber, M. E. *et al.* Diversity of gene expression in adenocarcinoma of the lung. *Proceedings of the National Academy of Sciences* **98**, 13784–13789 (2001).

115. Shoshan-Barmatz, V. *et al.* A molecular signature of lung cancer: potential biomarkers for adenocarcinoma and squamous cell carcinoma. *Oncotarget* **8**, 105492–105509 (2017).

116. Huang, T. *et al.* Distinguishing Lung Adenocarcinoma from Lung Squamous Cell Carcinoma by Two Hypomethylated and Three Hypermethylated Genes: A Meta-Analysis. *PLoS One* **11**, e0149088 (2016).

117. Pfeifer, G. P. Environmental exposures and mutational patterns of cancer genomes. *Genome Med.* **2**, 54 (2010).

118. Ellegren, H., Smith, N. G. C. & Webster, M. T. Mutation rate variation in the mammalian genome. *Current Opinion in Genetics & Development* **13**, 562–568 (2003).

119. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).

120. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).

121. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3**, 246–259 (2013).

122. Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.* **15**, 585–598 (2014).

123. Alexandrov, L. B. & Stratton, M. R. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr. Opin. Genet. Dev.* **24**, 52–60 (2014).

124. Aran, D., Sirota, M. & Butte, A. J. Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* **6**, 8971 (2015).

125. Hoadley, K. A. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291–304.e6 (2018).

126. Thorsson, V. *et al.* The Immune Landscape of Cancer. *Immunity* **51**, 411–412 (2019).

127. Raynaud, F., Mina, M., Tavernari, D. & Ciriello, G. Pan-cancer inference of intra-tumor heterogeneity reveals associations with different forms of genomic instability. *PLoS Genet.* **14**, e1007669 (2018).

128. Sansregret, L. & Swanton, C. The Role of Aneuploidy in Cancer Evolution. *Cold Spring Harb. Perspect. Med.* **7**, (2017).

129. Tanaka, K. & Hirota, T. Chromosomal instability: A common feature and a therapeutic target of cancer. *Biochim. Biophys. Acta* **1866**, 64–75 (2016).

130. Sansregret, L., Vanhaesebroeck, B. & Swanton, C. Determinants and clinical implications of chromosomal instability in cancer. *Nat. Rev. Clin. Oncol.* **15**, 139–150 (2018).

131. Bakhoum, S. F. & Cantley, L. C. The Multifaceted Role of Chromosomal Instability in Cancer and Its Microenvironment. *Cell* **174**, 1347–1360 (2018).

132. Bayani, J. *et al.* Distinct patterns of structural and numerical chromosomal instability characterize sporadic ovarian cancer. *Neoplasia* **10**, 1057–1065 (2008).

133. Dewhurst, S. M. *et al.* Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution. *Cancer Discov.* **4**, 175–185 (2014).

134. Storchová, Z. *et al.* Genome-wide genetic analysis of polyploidy in yeast. *Nature* **443**, 541–547 (2006).

135. Danielsen, H. E., Pradhan, M. & Novelli, M. Revisiting tumour aneuploidy - the place of ploidy assessment in the molecular era. *Nat. Rev. Clin. Oncol.* **13**, 291–304 (2016).

136. Pfleghaar, K., Heubes, S., Cox, J., Stemmann, O. & Speicher, M. R. Securin Is Not Required for Chromosomal Stability in Human Cells. *PLoS Biology* **3**, e416 (2005).

137. Laughney, A. M., Elizalde, S., Genovese, G. & Bakhoum, S. F. Dynamics of Tumor Heterogeneity Derived from Clonal Karyotypic Evolution. *Cell Rep.* **12**, 809–820 (2015).

138. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).

139. Muleris, M. *et al.* Chromosomal instability in near-diploid colorectal cancer: a link between numbers and structure. *PLoS One* **3**, e1632 (2008).

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

141. Nambiar, M., Kari, V. & Raghavan, S. C. Chromosomal translocations in cancer. *Biochim. Biophys. Acta* **1786**, 139–152 (2008).

142. Nowell, C. The minute chromosome (Ph1) in chronic granulocytic leukemia. *Blut Zeitschrift für die Gesamte Blutforschung* **8**, 65–66 (1962).

143. Kang, Z.-J. *et al.* The Philadelphia chromosome in leukemogenesis. *Chin. J. Cancer* **35**, 48 (2016).

144. Torres, E. M. *et al.* Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* **317**, 916–924 (2007).

145. Sheltzer, J. M. *et al.* Aneuploidy drives genomic instability in yeast. *Science* **333**, 1026–1030 (2011).

146. Zhao, R., Choi, B. Y., Lee, M.-H., Bode, A. M. & Dong, Z. Implications of Genetic and Epigenetic Alterations of CDKN2A (p16 INK4a) in Cancer. *EBioMedicine* **8**, 30–39 (2016).

147. Pulciani, S., Santos, E., Long, L. K., Sorrentino, V. & Barbacid, M. ras gene Amplification and malignant transformation. *Molecular and Cellular Biology* **5**, 2836–2841 (1985).

148. Davoli, T. *et al.* Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* **155**, 948–962 (2013).

149. Bakhoum, S. F. & Landau, D. A. Chromosomal Instability as a Driver of Tumor Heterogeneity and Evolution. *Cold Spring Harb. Perspect. Med.* **7**, (2017).

150. Vargas-Rondón, N., Villegas, V. E. & Rondón-Lagos, M. The Role of Chromosomal Instability in Cancer and Therapeutic Responses. *Cancers* **10**, (2017).

151. Pavelka, N. *et al.* Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* **468**, 321–325 (2010).

152. Nowak, M. A. *et al.* The role of chromosomal instability in tumor initiation. *Proceedings of the National Academy of Sciences* **99**, 16226–16231 (2002).

153. Zhang, H. *et al.* Chromosome-wide gene dosage rebalance may benefit tumor progression. *Mol. Genet. Genomics* **293**, 895–906 (2018).

154. Gronroos, E. & López-García, C. Tolerance of Chromosomal Instability in Cancer: Mechanisms and Therapeutic Opportunities. *Cancer Res.* **78**, 6529–6535 (2018).

155. Pedersen, B. S. & De, S. Loss of heterozygosity preferentially occurs in early replicating regions in cancer genomes. *Nucleic Acids Res.* **41**, 7615–7624 (2013).

156. Artandi, S. E. & DePinho, R. A. Telomeres and telomerase in cancer. *Carcinogenesis* **31**, 9–18 (2010).

157. Bailey, S. M. Telomeres, chromosome instability and cancer. *Nucleic Acids Research* **34**, 2408–2417 (2006).

158. Roylance, R. *et al.* Relationship of Extreme Chromosomal Instability with Long-term Survival in a Retrospective Analysis of Primary Breast Cancer. *Cancer Epidemiology Biomarkers & Prevention* **20**, 2183–2194 (2011).

159. Birkbak, N. J. *et al.* Paradoxical relationship between chromosomal instability and survival outcome in cancer. *Cancer Res.* **71**, 3447–3452 (2011).

160. Andor, N. *et al.* Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat. Med.* **22**, 105–113 (2016).

161. Jamal-Hanjani, M. *et al.* Tracking the Evolution of Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **376**, 2109–2121 (2017).

162. McGranahan, N., Burrell, R. A., Endesfelder, D., Novelli, M. R. & Swanton, C. Cancer chromosomal instability: therapeutic and diagnostic challenges. *EMBO Rep.* **13**, 528–538 (2012).

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

164. Selmecki, A. M. *et al.* Polyploidy can drive rapid adaptation in yeast. *Nature* **519**, 349–352 (2015).

165. Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).

166. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**, 23–28 (1976).

167. Biswas, D. *et al.* A clonal expression biomarker associates with lung cancer mortality. *Nat. Med.* (2019). doi:10.1038/s41591-019-0595-z

168. Assenov, Y., Brocks, D. & Gerhäuser, C. Intratumor heterogeneity in epigenetic patterns. *Semin. Cancer Biol.* **51**, 12–21 (2018).

169. Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nature Reviews Clinical Oncology* **15**, 81–94 (2018).

170. Ramirez, M. *et al.* Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells. *Nature Communications* **7**, (2016).

171. Hata, A. N. *et al.* Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. *Nat. Med.* **22**, 262–269 (2016).

172. Davis, A., Gao, R. & Navin, N. Tumor evolution: Linear, branching, neutral or punctuated? *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1867**, 151–161 (2017).

173. Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* **46**, 225–233 (2014).

174. Hiley, C., de Bruin, E. C., McGranahan, N. & Swanton, C. Deciphering intratumor heterogeneity and temporal acquisition of driver events to refine precision medicine. *Genome Biol.* **15**, 453 (2014).

175. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506–510 (2012).

176. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).

177. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).

178. McPherson, A. *et al.* Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nat. Genet.* **48**, 758–767 (2016).

179. Boutros, P. C. *et al.* Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat. Genet.* **47**, 736–745 (2015).

180. Ling, S. et al. Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. Proceedings of the National Academy of Sciences **112**, E6496-E6505

181. Harbst, K. et al. Multiregion Whole-Exome Sequencing Uncovers the Genetic Evolution and Mutational Heterogeneity of Early-Stage Metastatic Melanoma. Cancer Res. **76**, 4765–4774 (2016).

182. Sottoriva, A. *et al.* Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 4009–4014 (2013).

183. Bakhoum, S. F. *et al.* Numerical chromosomal instability mediates susceptibility to radiation treatment. *Nat. Commun.* **6**, 5990 (2015).

184. Zasadil, L. M. *et al.* Cytotoxicity of paclitaxel in breast cancer is due to chromosome missegregation on multipolar spindles. *Sci. Transl. Med.* **6**, 229ra43 (2014).

185. Cahill, D. P., Kinzler, K. W., Vogelstein, B. & Lengauer, C. Genetic instability and darwinian selection in tumours. *Trends in Genetics* **15**, M57–M60 (1999).

186. Cagan, R. Faculty of 1000 evaluation for Mediators of vascular remodelling co-opted for sequential steps in lung metastasis. *F1000 - Post-publication peer review of the biomedical literature* (2007). doi:10.3410/f.1080882.533819

187. Steeg, P. S. Tumor metastasis: mechanistic insights and clinical challenges. *Nat. Med.* **12**, 895–904 (2006).

188. Valastyan, S. & Weinberg, R. A. Tumor metastasis: molecular insights and evolving paradigms. *Cell* **147**, 275–292 (2011).

189. Aceto, N. *et al.* Circulating Tumor Cell Clusters Are Oligoclonal Precursors of Breast Cancer Metastasis. *Cell* **158**, 1110–1122 (2014).

190. Aceto, N., Toner, M., Maheswaran, S. & Haber, D. A. En Route to Metastasis: Circulating Tumor Cell Clusters and Epithelial-to-Mesenchymal Transition. *Trends Cancer Res.* **1**, 44–52 (2015).

191. Lambert, A. W., Pattabiraman, D. R. & Weinberg, R. A. Emerging Biological Principles of Metastasis. *Cell* **168**, 670–691 (2017).

192. Luzzi, K. J. *et al.* Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *Am. J. Pathol.* **153**, 865–873 (1998).

193. Chambers, A. F., Groom, A. C. & MacDonald, I. C. Dissemination and growth of cancer cells in metastatic sites. *Nat. Rev. Cancer* **2**, 563–572 (2002).

194. Turajlic, S. & Swanton, C. Metastasis as an evolutionary process. *Science* **352**, 169–175 (2016).

195. Gundem, G. *et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* **520**, 353–357 (2015).

196. Hunter, K. W., Amin, R., Deasy, S., Ha, N.-H. & Wakefield, L. Genetic insights into the morass of metastatic heterogeneity. *Nat. Rev. Cancer* **18**, 211–223 (2018).

197. Nakagawa, H. & Fujita, M. Whole genome sequencing analysis for cancer genomics and precision medicine. *Cancer Sci.* **109**, 513–522 (2018).

198. Valabrega, G., Montemurro, F. & Aglietta, M. Trastuzumab: mechanism of action, resistance and future perspectives in HER2-overexpressing breast cancer. *Ann. Oncol.* **18**, 977–984 (2007).

199. Druker, B. J. *et al.* Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr–Abl positive cells. *Nature Medicine* **2**, 561–566 (1996).

200. Druker, B. J. *et al.* Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *N. Engl. J. Med.* **355**, 2408–2417 (2006).

201. Grigoriu, B., Berghmans, T. & Meert, A.-P. Management of EGFR mutated nonsmall cell lung carcinoma patients. *Eur. Respir. J.* **45**, 1132–1141 (2015).

202. Ruiz, C. *et al.* Advancing a clinically relevant perspective of the clonal nature of cancer. *Proceedings of the National Academy of Sciences* **108**, 12054–12059 (2011).

203. Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011).

204. Holley, T. *et al.* Deep clonal profiling of formalin fixed paraffin embedded clinical samples. *PLoS One* **7**, e50586 (2012).

205. Lorber, T. *et al.* Exploring the spatiotemporal genetic heterogeneity in metastatic lung adenocarcinoma using a nuclei flow-sorting approach. *J. Pathol.* **247**, 199–213 (2019).

206. Corver, W. E. *et al.* High-resolution multi-parameter DNA flow cytometry enables detection of tumour and stromal cell subpopulations in paraffin-embedded tissues. *J. Pathol.* **206**, 233–241 (2005).

207. Blobel, G. A., Moll, R., Franke, W. W. & Vogt-Moykopf, I. Cytokeratins in normal lung and lung carcinomas. *Virchows Archiv B Cell Pathology Including Molecular Pathology* **45**, 407–429 (1984).

208. Quinlan, R. A. *et al.* Patterns of expression and organization of cytokeratin intermediate filaments. *Ann. N. Y. Acad. Sci.* **455**, 282–306 (1985).

209. O'Rourke, M. B. & Padula, M. P. Analysis of formalin-fixed, paraffin-embedded (FFPE) tissue via proteomic techniques and misconceptions of antigen retrieval. *Biotechniques* **60**, 229–238 (2016).

210. Davoli, T. & de Lange, T. The causes and consequences of polyploidy in normal development and cancer. *Annu. Rev. Cell Dev. Biol.* **27**, 585–610 (2011).

211. Swanton, C., McGranahan, N., Starrett, G. J. & Harris, R. S. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discov.* **5**, 704–712 (2015).

212. Priestley, P. *et al.* Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* **575**, 210–216 (2019).

213. Kallioniemi, A. *et al.* Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* **258**, 818–821 (1992).

214. Razavi, P. *et al.* The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. *Cancer Cell* **34**, 427–438.e6 (2018).

215. Li, C. & Lu, H. Adenosquamous carcinoma of the lung. *OncoTargets and Therapy* **11**, 4829–4835 (2018).

216. Mordant, P. *et al.* Adenosquamous carcinoma of the lung: surgical management, pathologic characteristics, and prognostic implications. *Ann. Thorac. Surg.* **95**, 1189–1195 (2013).

217. Zachara-Szczakowski, S., Verdun, T. & Churg, A. Accuracy of classifying poorly differentiated non–small cell lung carcinoma biopsies with commonly used lung carcinoma markers. *Human Pathology* **46**, 776–782 (2015).

218. Jin, X., Xu, X., Xu, H., Lv, L. & Lu, H. The Diagnostic Value of Carcinoembryonic Antigen and Squamous Cell Carcinoma Antigen in Lung Adenosquamous Carcinoma. *Clin. Lab.* **63**, 801–808 (2017).

219. Shimoji, M. *et al.* A clinicopathological and immunohistological re-evaluation of adenosquamous carcinoma of the lung. *Pathol. Int.* **61**, 717–722 (2011).

220. Mather, J. P. *et al.* Isolation of cancer stem like cells from human adenosquamous carcinoma of the lung supports a monoclonal origin from a multipotential tissue stem cell. *PLoS One* **8**, e79456 (2013).

221. Toyooka, S. *et al.* Mutations of epidermal growth factor receptor and K-ras genes in adenosquamous carcinoma of the lung. *Int. J. Cancer* **118**, 1588–1590 (2006).

222. Ji, H. *et al.* LKB1 modulates lung cancer differentiation and metastasis. *Nature* **448**, 807–810 (2007).

223. Han, X. *et al.* Transdifferentiation of lung adenocarcinoma in mice with Lkb1 deficiency to squamous cell carcinoma. *Nature Communications* **5**, (2014).

224. Yao, S. *et al.* Lysyl oxidase inhibition drives the transdifferentiation from lung adenocarcinoma to squamous cell carcinoma in mice. doi:10.1101/314393

225. Schafer, Z. T. *et al.* Antioxidant and oncogene rescue of metabolic defects caused by loss of matrix attachment. *Nature* **461**, 109–113 (2009).

226. Li, F. *et al.* LKB1 Inactivation Elicits a Redox Imbalance to Modulate Non-small Cell Lung Cancer Plasticity and Therapeutic Response. *Cancer Cell* **27**, 698–711 (2015).

227. Mollaoglu, G. *et al.* The Lineage-Defining Transcription Factors SOX2 and NKX2-1 Determine Lung Cancer Cell Fate and Shape the Tumor Immune Microenvironment. *Immunity* **49**, 764–779.e9 (2018).

228. Osei-Amponsa, V. *et al.* Hypermethylation of FOXA1 and allelic loss of PTEN drive squamous differentiation and promote heterogeneity in bladder cancer. *Oncogene* (2019). doi:10.1038/s41388-019-1063-4

229. Kang, D. B., Oh, J. T., Jo, H. J. & Park, W. C. Primary adenosquamous carcinoma of the colon. *J. Korean Surg. Soc.* **80 Suppl 1**, S31–5 (2011).

230. Mishra, S. *et al.* Primary Adenosquamous Carcinoma of the Prostate: A Rare Aggressive Tumor. *Clinical Genitourinary Cancer* **12**, e29–e31 (2014).

231. Regi, P. *et al.* Clinicopathological features of adenosquamous pancreatic cancer. *Langenbeck's Archives of Surgery* **396**, 217–222 (2011).

232. Rausch, T., Benhattar, J., Sutter, M. & Andrejevic-Blant, S. Thyroid carcinoma with papillary and squamous features: Report of a case with histogenetic considerations. *Pathology - Research and Practice* **206**, 263–269 (2010).

233. Ahn, S., Bae, G. E. & Kim, K.-M. Exuberant squamous metaplasia of the gastric mucosa in a patient with gastric adenocarcinoma. *Diagnostic Pathology* **10**, (2015).

234. Marcoux, N. *et al.* EGFR-Mutant Adenocarcinomas That Transform to Small-Cell Lung Cancer and Other Neuroendocrine Carcinomas: Clinical Outcomes. *J. Clin. Oncol.* **37**, 278–285 (2019).

235. Niederst, M. J. *et al.* RB loss in resistant EGFR mutant lung adenocarcinomas that transform to small-cell lung cancer. *Nat. Commun.* **6**, 6377 (2015).

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

237. Ahmed, T. *et al.* Non-small cell lung cancer transdifferentiation into small cell lung cancer: A case series. *Lung Cancer* **122**, 220–223 (2018).

238. Chen, Y., Tang, W. Y., Tong, X. & Ji, H. Pathological transition as the arising mechanism for drug resistance in lung cancer. *Cancer Communications* **39**, (2019).

239. Wang, R. *et al.* Analysis of major known driver mutations and prognosis in resected adenosquamous lung carcinomas. *J. Thorac. Oncol.* **9**, 760–768 (2014).

240. Rosell, R. *et al.* Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol.* **13**, 239–246 (2012).

241. Shi, X. *et al.* PD-L1 expression in lung adenosquamous carcinomas compared with the more common variants of non-small cell lung cancer. *Sci. Rep.* **7**, 46209 (2017).

242. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

243. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

244. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).

245. Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).

246. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).

247. Futreal, P. A. *et al.* A census of human cancer genes. *Nature Reviews Cancer* **4**, 177–183 (2004).

248. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).

249. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).

250. Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* **44**, e131 (2016).

251. Piscuoglio, S. *et al.* The Genomic Landscape of Male Breast Cancers. *Clin. Cancer Res.* **22**, 4045–4056 (2016).

252. Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).

253. Landau, D. A. *et al.* Evolution and Impact of Subclonal Mutations in Chronic Lymphocytic Leukemia. *Cell* **152**, 714–726 (2013).

254. Guerini-Rocco, E. *et al.* Microglandular adenosis associated with triple-negative breast cancer is a neoplastic lesion of triple-negative phenotype harbouring TP53 somatic mutations. *J. Pathol.* **238**, 677–688 (2016).

255. Nixon, K. C. The Parsimony Ratchet, a New Method for Rapid Parsimony Analysis. *Cladistics* **15**, 407–414 (1999).

256. Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).

257. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, 31 (2016).

258. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).

259. Kalemkerian, G. P. *et al.* Molecular Testing Guideline for the Selection of Patients With Lung Cancer for Treatment With Targeted Tyrosine Kinase Inhibitors: American Society of Clinical Oncology Endorsement of the College of American Pathologists/International Association for the Study of Lung Cancer/Association for Molecular Pathology Clinical Practice Guideline Update. *J. Clin. Oncol.* **36**, 911–919 (2018).

260. Oliver, T. G., Patel, J. & Akerley, W. Squamous Non–small Cell Lung Cancer as a Distinct Clinical Entity. *American Journal of Clinical Oncology* **38**, 220–226 (2015).

261. Cao, Y., Bryan, T. M. & Reddel, R. R. Increased copy number of the TERT and TERC telomerase subunit genes in cancer cells. *Cancer Science* **99**, 1092–1099 (2008).