SAPIENZA
UNIVERSITÀ DI ROMA

Ph.D. COURSE IN CELL AND DEVELOPMENTAL BIOLOGY
XXXV Cycle (A.Y. 2021/2022)

In-depth characterization of a novel aggressive subtype of lung cancer and dissection of exosomal-miRNAs as key-players for early identification

Ph.D. student:
Francesco Mazzarelli

Supervisor:
Dr. Fabrizio Bianchi
Internal tutor:
Pr. Rodolfo Negri
Coordinator:
Pr. Giulia De Lorenzo

Summary of the whole work

Lung adenocarcinoma (LUAD) is the most common type of non-small-cell lung cancer, and a subset of patients with early-stage LUAD experience relapses and adverse prognosis. Here we identified by a 10-gene prognostic signature an aggressive LUAD subtype, named C1, that is characterized by loss of lung cell lineage and gain of stem cell-like characteristics, along with mutator and immune evasion phenotypes.

In addition, with a multi-tiered approach using coupled RNA-seq and miRNA-seq data analysis of a large cohort of lung cancer patients, we found high accuracy in scoring aggressive disease and identified a minimal signature of 7-miRNAs, which we validated in a cohort of FFPE lung adenocarcinoma samples. The results demonstrate the reliability of miRNA-based biomarkers for lung cancer prognostication and make a step forward to the application of miRNA biomarkers in the clinical routine.

In line with this, we finally identified a 6-exo-miRNA signature derived from LUAD patients exosomes that is a reliable biomarker to identify the C1-LUAD subtype. Moreover, we also showed that miR-223-3p was significantly enriched in exosomes derived from C1-like cells and acted as a modulator of immune response.

Overall, we provide new insights into the molecular and biological features of aggressive C1-LUAD, providing new biomarkers for the early identification of aggressive subtypes of lung cancer.

Synopsis of the whole work

Lung cancer accounts for ~2 million of new diagnoses and ~1.8 millions of deaths worldwide every year and lung adenocarcinoma (LUAD) is the most common subtype accounting ~50% of all lung cancer cases. Despite the improvements made in lung cancer early detection and prognosis, a large fraction of patients (~30–40%) still experiences poor prognosis due to cancer progression. The molecular heterogeneity of lung cancer is often associated with the metastatic dissemination of tumor cells and chemoresistance. Therefore, the discovery of lung cancer molecular subtypes can help to improve diagnosis, prognostication, and identification of key druggable targets. The main objective of this thesis was to characterize a new aggressive molecular subtype of lung cancer, named C1-LUAD, and to discover circulating biomarkers for its early identification. We addressed this aim by using two distinct approaches: 1) multi-tiered approach to comprehensively characterize the molecular profile of C1-LUAD; 2) analysis of exosomal miRNAs derived from patients to generate a predictive signature. In order to characterize the C1-LUAD subtype, we dissected the transcriptional profile of three independent cohorts of LUAD patients (N=1183), classifying patients in C1 and nonC1 based on the 10-gene prognostic signature. We uncovered that in all the cohorts, C1-LUAD is associated to worst prognosis also when disease is diagnosed at stage I. Importantly, we observed that C1-LUAD revealed a significant increase in the expression of some "Esophagus" and "Midgut/Hindgut" markers along with a significant downregulation of lineage-specifying genes associated with lung cell identity (e.g. NKX2-1, SFTPC). Moreover, gene set enrichment analysis (GSEA) highlighted an overrepresentation of molecular signatures associated to epithelial-to-mesenchymal transition (EMT)/stem cells, the immune system and hypoxia, which are hallmarks of metastatic cancer. Next, we performed a phenotypic screening in 12 LUAD cell lines. Combining proliferation, migration, invasion and spheres forming efficiency parameters in a phenotypic scoring algorithm, C1-like cell lines showed an overall higher aggressive phenotype as compared to the nonC1-like cell lines. We also investigated the mutational profile showing that the tumor mutation burden was significantly higher in C1-LUAD. Consistently, we found higher rates of mutations in C1-LUAD of known tumor suppressor genes, oncogenes, or genes described to have both tumor suppressor and oncogene activities. Finally, we investigated the tumor-infiltrating immune composition of C1-LUAD. Overall, Th2(CD4+) cells were predominant, whereas pro-inflammatory T helper Th1 and Th17 cells were depleted in C1-LUAD. Interestingly, GSEA and CIBERSORTx confirmed a significant higher fraction of IRF4+ Tregs in C1-LUAD and a reduction of the IRF4− Tregs population. Taken together these data demonstrated that loss of lung cell lineage with acquirement of stem cell-like features, along with the occurrence of mutator and immune evasion phenotype, characterize the aggressiveness of C1-LUAD

In order to discover new biomarkers to early identify C1-LUAD, we performed miRNA expression profile of the TCGA-LUAD cohort, showing 200 miRNAs differentially expressed. Lasso regularization was then applied to identify two signatures of 14- and 19-miRNA, which displayed a high accuracy in C1/non-C1 cancer patients stratification.
Next, we explored the miRNA-mRNA interactome characterizing C1 tumors by performing ARACNe (Algorithm for the Reconstruction of Accurate Cellular Networks) using the set of 200 miRNA, and a set of 2900 mRNA genes found significantly regulated in C1-LUAD.
We found a set of interacting networks with 7 miRNA as "HUBs" which derived from both the 19-miRNA and 14-miRNA signatures. Hierarchical clustering analysis of this 7-miRNA signature showed an overall increased expression in C1-LUAD. Importantly, the 7-miRNA signature had a cross-validated AUC of 0.79 in C1/non-C1 patients' stratification. Finally, we validated the 7 miRNA-signature in an external cohort of 44 lung adenocarcinoma tissues, which was collected at the IRCCS Casa Sollievo della Sofferenza Hospital (CSS). We used the 10-gene signature and calculated relative risk-score to stratify the cohort into C1 (n = 16) and non-C1 (n = 28) groups as a control. Next, using logistic regression, we found that the 7-miRNA model stratified C1 from non-C1 tumors with an AUC of 0.76 and with significant difference in C1 predicted probability. To investigate exosomal miRNAs as circulating biomarkers for C1-LUAD identification, we isolated exosomes from 32 plasma samples of the CSS Cohort, divided in 15 C1 and 17 nonC1. We then performed qRT-PCR on a subset of 23-miRNAs derived from the previous analysis, highlighting a total 13-miRNAs differentially expressed in C1 and nonC1 exosomes. Using unconditional logistic regression analysis, we found that the 13-exo-miRNA signature identified correctly C1 patients with a bootstrap corrected AUC=0.64. Then we reduced the numbers of predictors to 6 exo-miRNAs, which stratified C1 from non-C1 tumors with a corrected AUC of 0.73 and with significant difference in C1 predicted probability. Then, we started to investigate the role of exosomes and exo-miRNA in aggressiveness using the 12 LUAD cell lines previously stratified in C1-like and nonC1-like. We demonstrated that C1-like cells are more prone to internalize self and exogenous exosomes then nonC1-like cells. Coherently, we performed the exo-miRNAs analysis as for the patients, showing 9 exo-miRNAs differentially expressed. Of such miRNAs, exo-miR-223-3p levels of C1-like cells were in accord with levels founded in C1-patients. Of note, miR-223-3p was significantly associated with immune, migration and proliferation gene-pathways in a miRPathDb analysis, suggesting a key-role for C1-like cells. Surprisingly, LUAD cells had low levels of miR-223-3p and C1-like cells overexpressing miR-223-3p showe significative reduction of migration and invasion. Consistently with high levels of miR-223-3p in C1-like exosome, we observed significantly lower production of Interferon gamma associated with lower vitality and proliferation of Jurkat activated cells conditioned with media derived from C1-like cells overexpressing miR-223-3p. These data confirmed a role of exo-miR-223-p in producing an immune evasion phenotype.

In conclusion here we demonstrated that the characterization of new molecular subtypes of lung cancer is crucial to better understand the heterogeneity of this disease. Moreover, we generated a circulating miRNA signature to early identify C1-LUAD subtype, highlighting the role of exosomes and exo-miRNA as reliable biomarkers and key-players in metastatic phenotype.

Chapter I:
Aim of the thesis

I.1 Aim of the thesis

The principal aim of my PhD project was to demonstrate the presence of new molecular subtypes of lung cancers which impact the diagnosis and prognosis by means of miRNA-based expression signatures. This thesis will go through the description of experimental strategies and results focused on three main aims:

Aim I: Characterization of an aggressive molecular subtype of lung adenocarcinoma identified by a gene expression (mRNA) signature.

Aim II: Identification of a miRNA-based expression signature which capture the molecular complexity of this LUAD aggressive molecular subtype.

Aim III: Dissecting exosomal-miRNA as key-players for the early-identification of aggressive lung adenocarcinoma using plasma samples.

Chapter II:
Characterization of an aggressive molecular subtype of lung adenocarcinoma identified by a gene expression (mRNA) signature

This Chapter answer to the Aim I. Results here contained are also part of the following publication:

Aggressive early-stage lung adenocarcinoma is characterized by epithelial cell plasticity with acquirement of stem-like traits and immune evasion phenotype

Valentina Melocchi†, Elisa Dama†, **Francesco Mazzarelli**, Roberto Cuttano, Tommaso Colangelo, Leonarda Di Candia, Enrico Lugli, Giulia Veronesi, Giuseppe Pelosi, Gian Maria Ferretti, Marco Taurchini, Paolo Graziano and Fabrizio Bianchi *

PMID: 34172935 DOI: 10.1038/s41388-021-01909-z

II.1 Summary

Lung adenocarcinoma (LUAD) is the most common type of non-small-cell lung cancer, accounting for about 40-50% of all cases. Despite improvements in early detection and personalized medicine, many patients with early-stage LUAD still experience disease relapses and poor outcomes. Previous studies have identified molecular subtypes of LUAD with distinct gene expression and mutational profiles, which are associated with prognosis, but their underlying biological and molecular features are not well understood. Therefore, it is not clear what causes aggressive LUAD.

In this study, we used a multi-tiered approach to investigate the molecular and functional characteristics of LUAD in multiple patient cohorts and LUAD cell lines. We analyzed the tumor transcriptome and the mutational and immune gene expression profiles, and we conducted cancer cell phenotypic screening using LUAD cell lines. Our findings suggest that a specific subset of LUAD, which we called C1-LUAD, has aggressive behavior due to the loss of lung cell lineage and gain of stem cell-like characteristics, along with mutator and immune evasion phenotypes. This subset can be identified using a 10-gene prognostic signature, and it is found in both early-stage and advanced-stage disease. Patients with this subtype have a poor prognosis. Our data suggest that transformed lung stem/progenitor cells and/or reprogrammed epithelial cells with CSC characteristics are the underlying causes of this aggressive disease. These findings may lead to new, more aggressive, therapeutic strategies for early-stage C1-LUAD.

II.2 Introduction

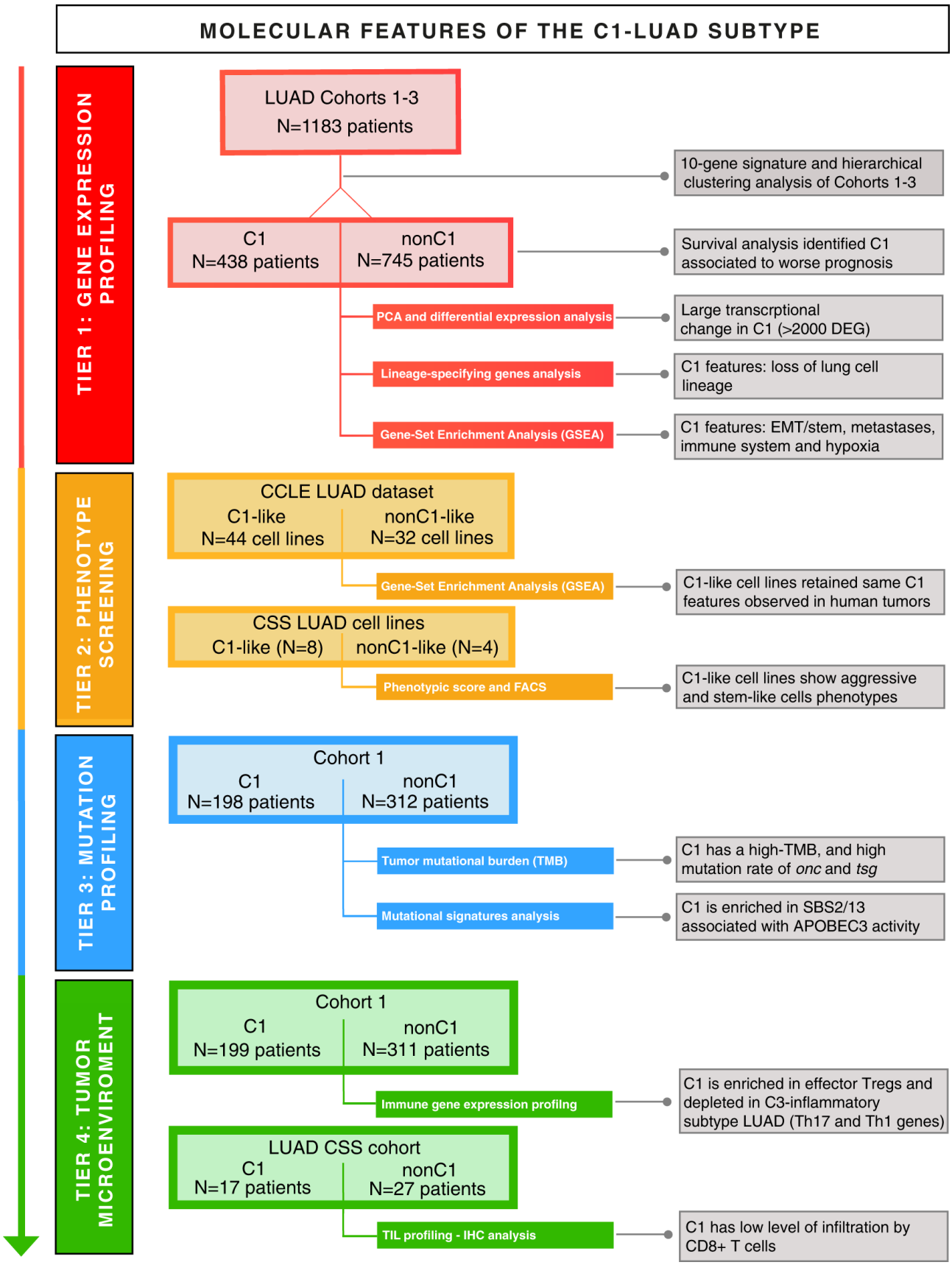
Every year, approximately 2 million new cases of lung cancer are diagnosed worldwide, resulting in around 1.8 million deaths [1]. The most common type of lung cancer is non-small-cell lung cancer (NSCLC), which comprises various tumor subtypes, including adenocarcinoma, squamous cell carcinoma, and large cell carcinoma [2]. Despite advancements in early detection and prognosis, around 30-40% of patients with early-stage NSCLC who undergo surgery still experience poor outcomes due to cancer progression [3-5]. The molecular heterogeneity of lung cancer plays a crucial role in metastatic dissemination and chemoresistance [6]. This cancer is characterized by high tumor mutational burden and genetic heterogeneity, which is influenced by smoking, the primary risk factor for lung cancer development [5]. Studies have identified a mutational signature induced by tobacco smoke carcinogens [9]. Genome instability fuels molecular heterogeneity and tumor evolution, leading to cell subsets with inheritable metastatic potential [8, 10]. The tumor transcriptome can be a valuable tool in studying genetic heterogeneity, as exemplified by recent research on lung cancer molecular subtypes [7, 11]. The discovery of these subtypes can help improve lung cancer diagnosis, prognostication, and identify targets for treatment. For instance, the TRU subtype has a lower tumor

mutational burden and ploidy, is enriched in mutations in the epidermal growth factor receptor and has a more favorable prognosis [7]. On the other hand, tumors of the PI subtype have an unfavorable prognosis and are enriched in NF1 and TP53 co-mutations, while the PP subtype has frequent mutations in KEAP1 and STK11 genes [7]. The expression of certain lineage-specifying genes, such as NKX2-1 and SOX2, can trigger alternative developmental programs leading to distinct molecular subtypes, each dominated by different gene expression profiles and actionable genetic alterations. A previous study identified a robust prognostic model for lung adenocarcinoma based on a 10-gene expression signature [15]. The loss of lung cell lineage and the acquisition of stem cell-like features, along with the occurrence of mutator and immune evasion phenotypes, are characteristic of an aggressive subtype of early-stage lung adenocarcinoma, which was named C1-LUAD and identified using the 10-gene prognostic signature [15]. The molecular diagnosis of C1-LUAD has greater prognostic significance than other biomarkers and previous molecular classifications, providing new possibilities for more aggressive treatment strategies for patients with early-stage C1-LUAD.

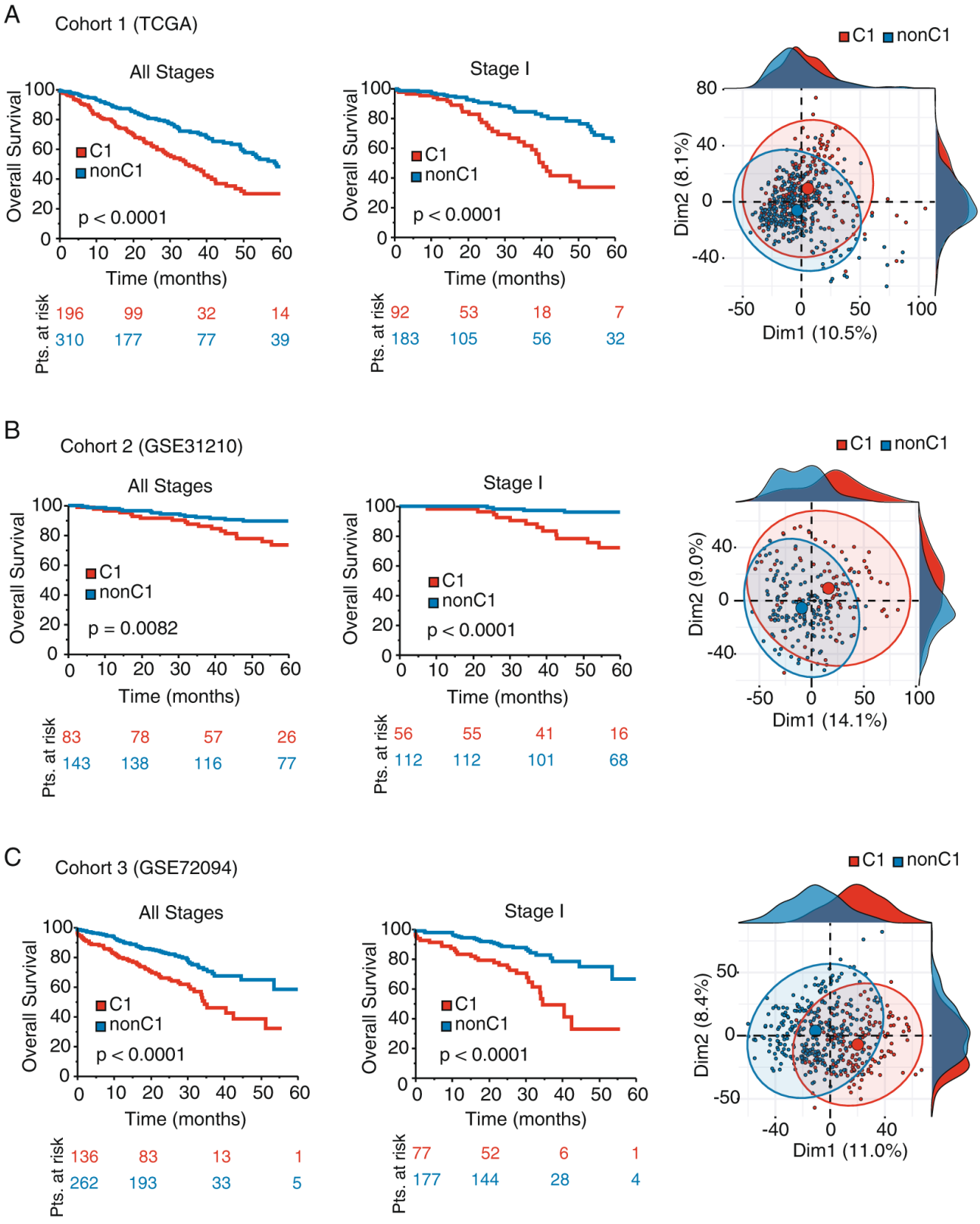
II.3 Results

C1-LUAD has the worst prognosis and is characterized by loss of lung cell lineage

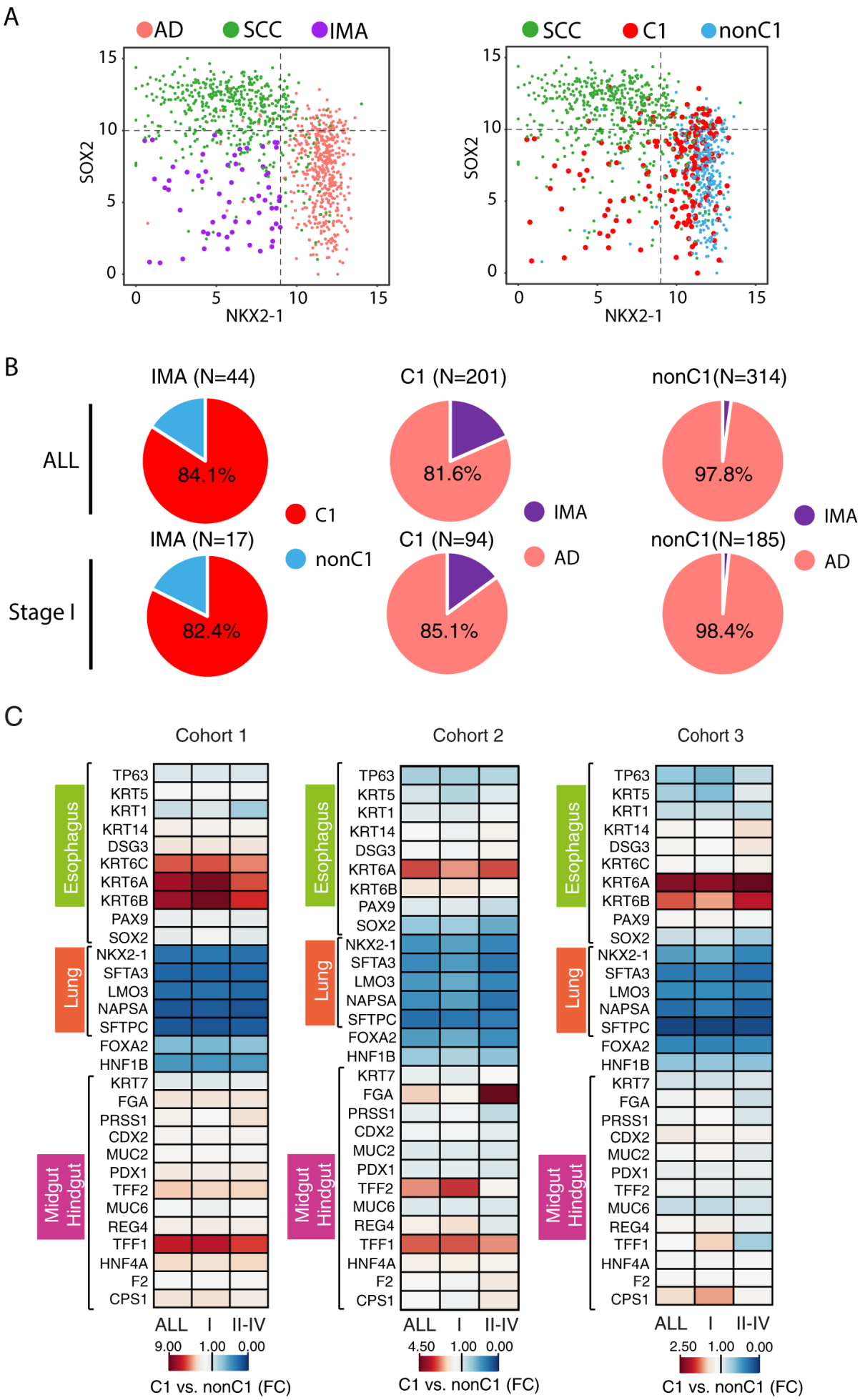
To gain a thorough understanding of the molecular profile of LUAD, we used a multi-tiered strategy (Fig.1). Initially, we analyzed the transcriptome of 1183 LUAD patients from three separate cohorts (Cohorts 1–3). Using a 10-gene expression signature [16] and hierarchical clustering analysis, useful to identify molecular subtype [7, 11, 17,18,19,20], we identify a previously described subset of LUAD known as C1-LUAD (also referred to as C1) [15, 16], which was present in all three cohorts and consisted of 37% of the patients (N = 438). We found that even patients with stage I disease had an unfavorable prognosis if they had C1-LUAD (Fig. 2). We then used principal component analysis (Fig.2) to investigate the extent of transcriptional change in C1- vs. nonC1-LUAD, which revealed different transcriptional profiles characterizing C1- and nonC1-LUAD. Out of 338 commonly regulated genes across Cohorts 1–3, 330 had the same trend of regulation (107 upregulated and 223 downregulated). Recently, Tata et al. [24] reported that invasive mucinous adenocarcinoma (IMA), also known as MAD, has a poor prognosis and is characterized by the loss of expression of NKX2-1 and SOX2. In our study, we found that 84% of the IMAs originally classified in Tata et al. were included in C1-LUAD (Fig. 3A,B), which had an increase in the expression of some "Esophagus" and "Midgut/Hindgut" markers (Fig. 3C) and a significant downregulation of lineage-specifying genes associated with lung cell identity (such as NKX2-1 and SFTPC, as shown in Fig. 3C).



Chapter II.3 Figure 1: Multi-tiered analysis of molecular features of the C1-LUAD subtype.
Study flow-chart with data sets, analysis, and main results.



Chapter II.3 Figure 2: Gene expression analysis of 1183 patients with lung adenocarcinoma.
Overall survival of patients with C1- and nonC1-LUAD tumors stratified by the 10-gene signature is shown for **A** Cohort 1 (TCGA), **B** Cohort 2 (GSE31210), and **C** in Cohort 3 (GSE72094). Numbers of patients at risk at 0, 20, 40, and 60 months are reported underneath the graphs. Log-rank *p* value is also shown. Color codes are as per the legend. On the right, in **A**–**C**, the first two dimensions (Dim1 and Dim2) of PCA are shown for every cohort analyzed as well as the relative distribution plots. The percentage of variability captured by Dim1 and Dim2 is shown in *Y*- and *X*-axes, respectively. Color codes are as per the legend.


A
AD
SCC
IMA
SCC
C1
nonC1
SOX2
NKX2-1
B
IMA (N=44)
C1 (N=201)
nonC1(N=314)
ALL
84.1%
81.6%
97.8%
C1
IMA
nonC1
AD
IMA (N=17)
C1 (N=94)
nonC1(N=185)
Stage I
82.4%
85.1%
98.4%
C
Cohort 1
Cohort 2
Cohort 3
Esophagus
Lung
Midgut Hindgut
TP63
KRT5
KRT1
KRT14
DSG3
KRT6C
KRT6A
KRT6B
PAX9
SOX2
NKX2-1
SFTA3
LMO3
NAPSA
SFTPC
FOXA2
HNF1B
KRT7
FGA
PRSS1
CDX2
MUC2
PDX1
TFF2
MUC6
REG4
TFF1
HNF4A
F2
CPS1
ALL
I
II-IV
9.00 1.00 0.00
4.50 1.00 0.00
2.50 1.00 0.00
C1 vs. nonC1 (FC)

Chapter II.3 Figure 3: Lineage-specifying gene expression profile of C1- and nonC1-LUAD samples.
A Scatterplot analysis of the *NKX2-1* (*x*-axes) and the *SOX2* (*y*-axes) expression (Log_2 RSEM normalized counts) of the TCGA lung cancer samples (Cohort 1, and squamous cell carcinoma samples (LUSC, *N* = 499), for a total of 1014 samples) grouped by different histological and molecular subtypes: AD adenocarcinoma, SCC squamous cell carcinoma, IMA invasive mucinous carcinoma; and C1- and nonC1-LUAD. **B** On the left, pie charts showing the percentage of C1- and nonC1-LUAD overlapping with IMA. Middle and right, pie charts showing the percentage of IMA or adenocarcinoma (AD) in C1- or in nonC1-LUAD samples. Results are shown for all tumor stages or limited to stage I tumors. The number of samples in each category is also shown. **C** Heatmaps of differentially expressed lineage-specifying genes in C1- vs. nonC1-LUAD in the three cohorts analyzed (Cohorts 1–3). Genes are grouped based on the developmental lineage specification (esophagus, lung, midgut, and hindgut). ALL all tumors, I stage I tumors, II–IV stages II–IV tumors. Color codes are as per the legend (FC fold change differences).

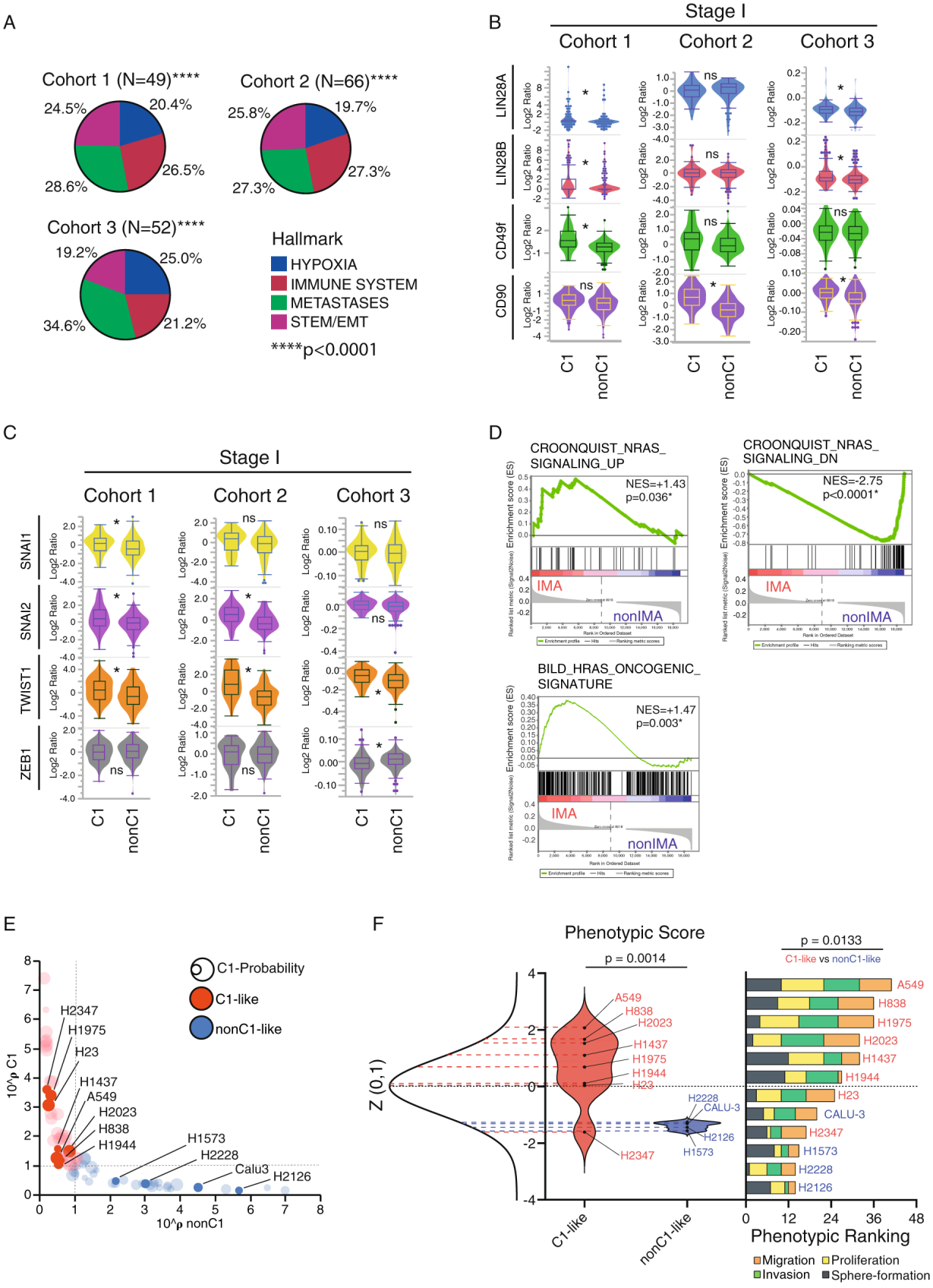
C1-LUAD is enriched in molecular signatures characteristic of undifferentiated cancer

To delve further into the molecular characteristics of C1-LUAD, we conducted gene set enrichment analysis (GSEA) using a large collection of curated gene sets (N = 4738) that virtually represent all known molecular mechanisms and cellular phenotypes. By comparing the transcriptional profiles of C1-LUAD and nonC1-LUAD, the GSEA identifies the overrepresentation of gene sets in a ranked list of genes. We discovered that among the positively and significantly enriched gene sets in stage I C1-LUAD, there was an overrepresentation (>20%, in at least two out of three cohorts; $p < 0.0001$, Z-test) of molecular signatures that are associated with metastatic cancer, such as epithelial-to-mesenchymal transition (EMT)/stem cells, metastases, the immune system, and hypoxia (Fig. 4A) [25]. Notably, upon analyzing the expression profile of genes relevant to stem cell biology and EMT, we identified a significant upregulation of several markers in stage I C1-LUAD (adjusted $p < 0.05$; Fig. 4B, C). In contrast, we observed an enrichment of gene sets related to RAS signaling in IMA samples, which drives differentiation toward mucinous adenocarcinoma, as previously reported (Fig. 4D) [24, 26]. Moreover, the enrichment of signatures related to stemness, poor survival, and metastasis was exclusively retained in "non-IMA" C1-LUAD.

C1-like LUAD cells have are characterized by aggressive phenotype

Then, we investigated whether C1-LUAD was functionally linked to biologically aggressive phenotypes. We utilized a large collection of LUAD cell lines from the Cancer Cell Line Encyclopedia (CCLE, N = 76, [29]), which we binarily categorized into C1-like or nonC1-like based on their 10-gene expression profile (Fig. 4E). We then conducted a phenotypic screening using a subset of 12 cell lines (8 C1-like and 4 nonC1-like) that were already present in our laboratory to measure various parameters that are related to established cancer phenotypes. These parameters included: (1) doubling time to measure cell proliferation, (C1-like = 32 h, nonC1-like = 50 h; median values); (2) wound healing percentage to measure cell migration (C1-like = 89%, nonC1-like = 29%; median values); (3) the number of cells in transwell assay to measure cell invasion (C1-like = 82 Cells/Field, nonC1-like = 22 Cells/Field; median values) and the efficiency to generate tumor spheres for cancer stem cell (CSC) enrichment (C1-like = 12%, nonC1-like = 8%, median values) (see Fig. S1 and Table S1) [30]. These parameters were then combined to create a phenotypic score and ranking, which provided a comprehensive index of tumor cell aggressiveness. As

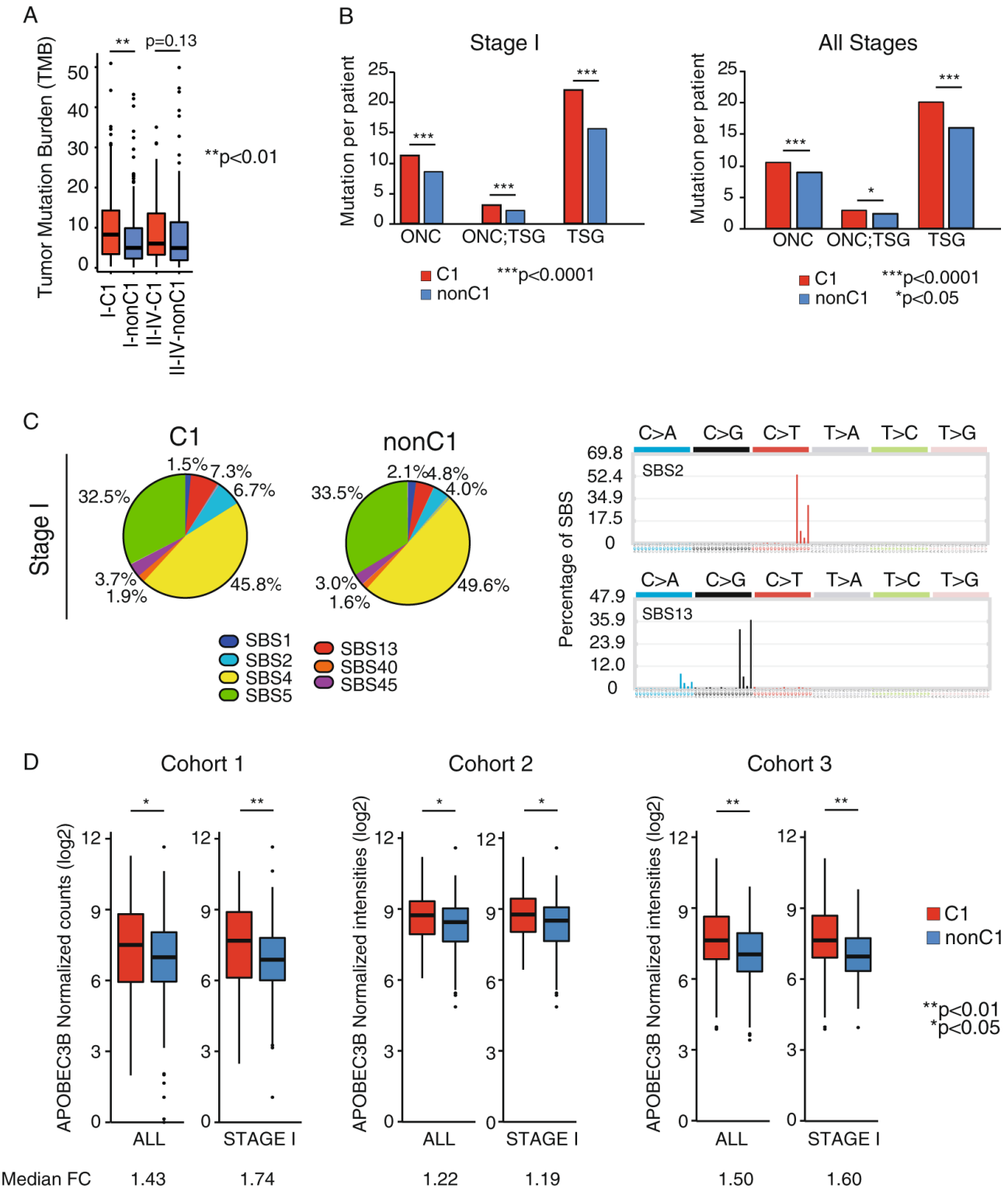
hypothesized, the C1-like cell lines exhibited an overall higher phenotypic score and ranking ($p = 0.0014$ and $p = 0.0133$, respectively, Wilcoxon test; Fig. 4F) compared to the nonC1-like cell lines. All these observations are likely to contribute to the clinical aggressiveness of C1-LUAD.



Chapter II.3 Figure 4: Functional analysis of C1-LUAD.
A Gene set enrichment analysis (GSEA) of curated gene sets (*N*=4 738) in C1- vs. nonC1-LUAD, in the Cohorts 1–3. Pie charts represent the percentage of gene sets mapping to hypoxia, immune system, metastases, and stem/EMT that we found overrepresented (>20%) among the top C2 curated gene sets enriched in C1-LUAD tumors (FDR<25%). *N* total number of gene sets overrepresented (>20%). B Gene expression profile analysis of stem cell and C EMT markers. Violin plots and box plots represent the distribution of log2 ratio (median centered) expression values in C1- and nonC1-LUAD, in the three cohorts of samples analyzed (Cohorts 1–3). Asterisks indicate significant differential expression (FDR<0.05; Benjamini–Hochberg correction). NS not significant (FDR>0.05). D Gene set enrichment analysis (GSEA) of RAS associated pathways in IMA samples (*N*=44) vs. C1-LUAD (non-IMA, *N*=164). E Classification in C1-like or nonC1-like of the Cancer Cell Line Encyclopedia (CCLE) lung adenocarcinoma cells (*N*=76). *Y*-axes, correlation coefficient (power of ten) of CCLE 10-gene expression profile with centroid values calculated in C1-LUAD samples (see "Methods"). *X*-axes, correlation coefficient (power of ten) of CCLE 10-gene expression profile with centroid values calculated in nonC1-LUAD samples (see "Methods"). The size of the filled circles corresponds to the probability of being a C1-LUAD (see "Methods"). The indicated cell lines correspond to the ones available in the lab and considered for further experiments. F On the left, violin plot of the phenotypic scores (*z*-scores) of the 12 cell lines analyzed. Bell curve for expected distribution of data. Cell lines are labeled in red (C1-like) or blue (nonC1-like). On the right, the stacked bar chart of the phenotypic rankings obtained by the indicated assay using the 12 adenocarcinoma cell lines. Cell lines are labeled in red (C1-like) or blue (nonC1-like). *p* value for phenotypic score was calculated using *t*-test; *p* value for phenotypic ranking was calculated using Wilcoxon test.

The mutational profile of C1-LUAD

Overall, the tumor mutation burden (TMB) was significantly higher in stage I C1-LUAD samples and there was also a trend towards significance for stages II-IV C1-LUAD samples (Fig. 5A). We observed higher rates of mutations in C1-LUAD for genes known to be involved in tumor suppression (tsg), oncogenes (onc), or genes that have both tumor suppression and oncogene activities (onc;tsg) (Fig.5B). However, the enrichment of tumor mutational signatures defined by single base substitution (SBS) signatures in the COSMIC database (cancer.sanger.ac.uk [34]) confirmed that tobacco smoking associated SBS4 and SBS5 signatures [34] were the most common in both C1-LUAD and nonC1-LUAD tumors (>40% and >30%, respectively) (Fig. 5C). Notably, we found that SBS2 and SBS13 signatures were significantly higher ($p < 0.0001$; Wilcoxon test) in C1-LUAD compared to nonC1-LUAD tumors (Fig. 5C). These two signatures are associated with the activity of APOBEC3 family members (APOBEC3 A-H), which normally function in RNA editing to protect cells from pathogens, but in cancer, they increase mutagenesis and promote subclonal evolution [35]. Specifically, APOBEC3B expression is upregulated in multiple cancer types and is linked to poor prognosis and resistance to treatment [35]. We also observed a significant increase ($p < 0.05$; Wilcoxon test) in APOBEC3B expression in C1-LUAD samples from all cohorts analyzed (Fig. 5D). Taken together, these findings suggest an elevated mutator phenotype in C1-LUAD that requires further investigation.

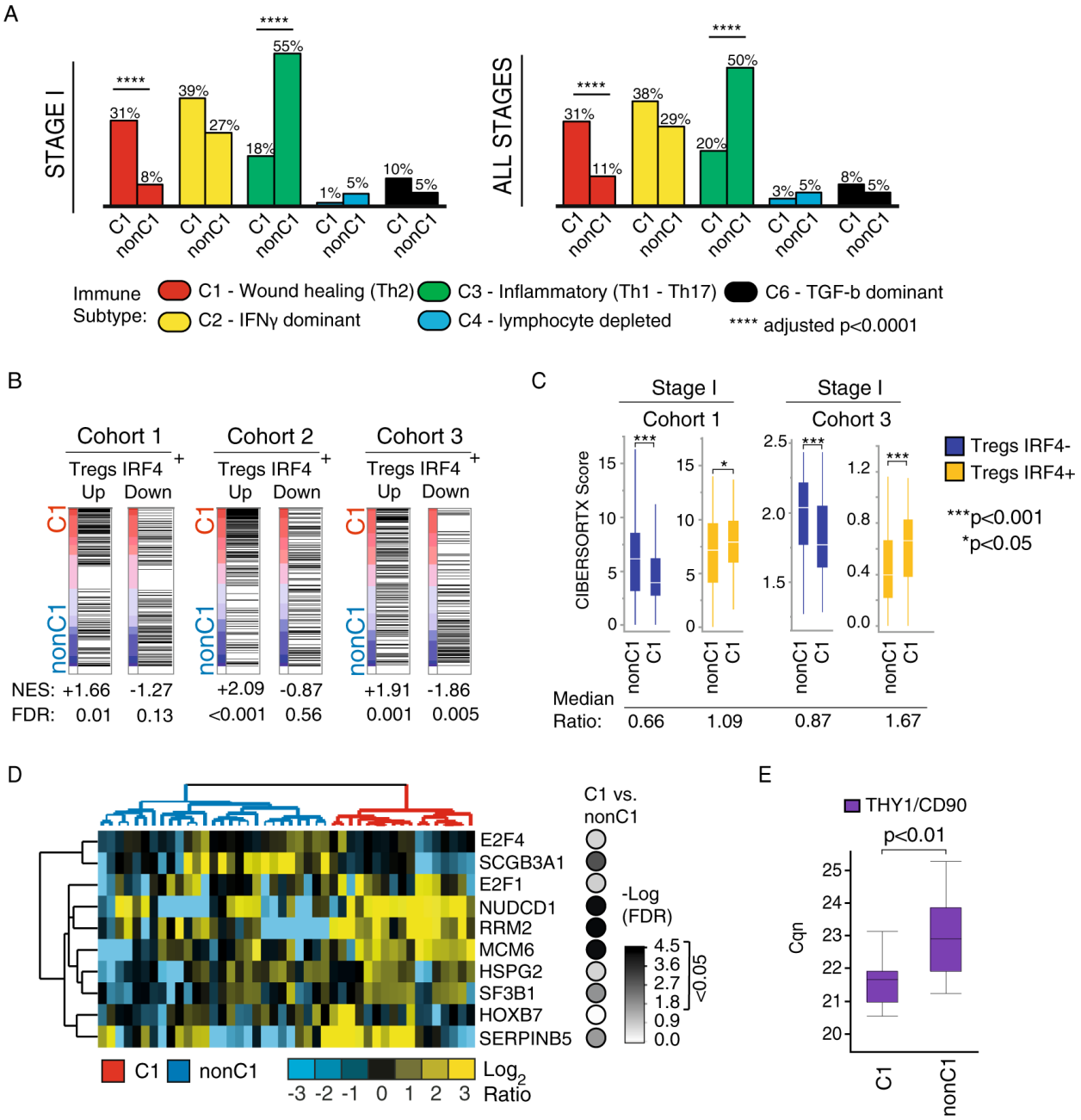


Chapter II.3 Figure 5: Mutational profile of C1- and nonC1-LUADs.
A Box plot of tumor mutation burden (TMB) of stage I C1-LUAD, stage I nonC1-LUAD and stages II–IV C1- and nonC1-LUAD samples. *p* values were computed by the Wilcoxon test. **B** Bar plot of tumor suppressor genes (*tsg*), oncogenes (*onc*), or of genes with both tumor suppressor and oncogene activities (*onc; tsg*), of stage I C1- and nonC1-LUAD (on the left), or considering all stages C1- and nonC1-LUAD samples (on the right). *p* values were computed by the Wilcoxon test. **C** Pie chart of the single base substitution (SBS) signatures distribution in stage I C1-LUAD and stage I nonC1-LUAD samples and, on the right, the graphical representation of the distribution of percentage of SBS2 and SBS13 signatures in Cohort 1. Adjusted *p* values were computed by Wilcoxon test (Benjamini–Hochberg correction). **D** Box plot of APOBEC3B gene expression in C1- and nonC1-LUAD samples (stage I, and ALL: all tumor stages) in Cohorts 1–3. *p* value was computed by the Wilcoxon test. Median fold change (FC) is also shown for all comparisons.

Exploiting the immune-infiltration of C1-LUAD

Finally, we investigated whether the immune composition [36] of the tumor microenvironment infiltrating C1-LUAD contributes to its aggressive behavior. Our initial analysis identified significant enrichment of several signatures associated with the immune system in C1-LUAD (as demonstrated in Fig. 4A). To predict the hierarchy of the immune tumor microenvironment in C1- and nonC1-LUAD stage I or all-stage disease, we analyzed the distribution of different immune cell subsets using the "Immune Subtypes" characterized by Thorsson et al. [37] in the TCGA-LUAD dataset (Cohort 1) (Fig. 6A). The wound healing subtype (C1), which is enriched in Th2(CD4+) cells, was predominantly present in C1-LUAD (31% in C1-LUAD vs. 11% in nonC1-LUAD; $p < 0.0001$; post hoc test following chi-square, Bonferroni correction). In contrast, the inflammatory subtype (C3), which is enriched in pro-inflammatory T helper Th1 and Th17 cells that enhance CD8+ T cells cytotoxicity, was depleted in C1-LUAD (20% in C1-LUAD vs. 50% in nonC1-LUAD; $p < 0.0001$; post hoc test following chi-square, Bonferroni correction) (as shown in Fig. 6A). Recently, a subset of CD4+ regulatory T cells, which are activated by interferon regulatory factor 4 (IRF4) transcription factor (IRF4+ Tregs), have been identified in various human cancers, including non-small cell lung cancer (NSCLC). These cells have been found to possess a superior capacity to suppress the antitumor immune response and are associated with the accumulation of exhausted CD8+ T cells [38]. To assess the abundance of IRF4+ Tregs in C1- and nonC1-LUAD samples, we employed the IRF4+ Treg specific transcriptional signature (consisting of 382 genes) developed by Alvisi et al. [38] and utilized GSEA and CIBERSORTx. Interestingly, the GSEA revealed that the IRF4+ Treg signature was significantly enriched in C1-LUAD (Fig. 6B). CIBERSORTx analysis confirmed a significant (in Cohorts 1 and 3) higher fraction of IRF4+ Tregs in C1-LUAD (Fig. 6C) and a reduction of the IRF4− Tregs population (Fig. 6C). While no significant differences were found in Cohort 2, where most of the samples (198 out of 226; 88%) were negative for IRF4+ Tregs (data not shown). These findings indicate that the more aggressive C1-LUAD subtype is characterized by a strong immunosuppressive microenvironment, in line with previous data on NSCLC [38]. Finally, we conducted an immunohistochemistry (IHC) analysis on an independent cohort of LUAD patients ($N = 85$) admitted to the Casa Sollievo della Sofferenza Research Hospital (CSS) to evaluate the levels of CD4+ and CD8+ T cells and PD-L1 expression [39]. We performed the IHC and qRT-PCR analysis on formalin-fixed paraffin-embedded (FFPE) tumor tissue samples. Out of the 85 samples, 44 samples were suitable for qRT-PCR analysis of the 10-gene signatures. Using hierarchical clustering analysis, we observed that the clusters of samples corresponded to C1- or nonC1-LUAD, as we had previously observed in Cohorts 1-3 (Fig. 6D). Remarkably, the expression of a cancer stem cell (CSC) associated biomarker, specifically THY1/CD90, was discovered to be significantly upregulated in C1-LUAD ($p = 0.002$, Wilcoxon test; Fig. 6E), confirming previous analyses. The frequency of samples exhibiting high levels of CD4+ cell infiltration

was comparable between C1-LUAD (10/17, 59%) and nonC1-LUAD samples (17/27, 63%) (p = 0.78; Fig. 6F). Conversely, there was a tendency towards lower CD8+ lymphocyte infiltration in the C1-LUAD samples (3/17, 18%) compared to nonC1-LUAD (12/27, 44%) (p = 0.06; Fig. 6F), which is consistent with our previous findings from immune expression profiling. Lastly, PD-L1 expression was found to be comparable between C1/nonC1-LUAD samples (Fig. 6F).



F

	C1 (N=17)		nonC1 (N=27)		P-value[a]
CD4+					
high	10	58.8	17	63.0	0.78
low	7	41.2	10	37.0	
CD8+					
high	3	17.6	12	44.4	0.06
low	14	82.4	15	55.6	
PD-L1					
high	4	23.5	3	11.1	0.23
low	0	0.0	2	7.4	
neg	13	76.5	22	81.5	

[a] Likelihood-Ratio test

Chapter II.3 Figure 6: Immunoprofile analysis of C1- and nonC1-LUADs.
A Bar plots represent distribution (percentage; *Y*-axes) of patients according to the immune subtype classes defined by Thorsson et al. in all stages and stage I C1- and nonC1-LUAD samples of the Cohort 1. Adjusted *p* values were computed by post hoc test following chi-square (Bonferroni correction). The C5 subtype, immunologically quiet and dominated by M2 macrophages, was not reported within the Cohort 1 samples. **B** GSEA of the immunoexpression signatures specific for $CD4^+$ $IRF4^+$ effector Tregs (Treg Up upregulated genes in Treg, Treg Down downregulated genes in Treg) in C1-LUAD vs. nonC1-LUAD tumors. NES normalized enrichment score, FDR false-discovery rate based on 1000 random samples permutations. **C** CIBERSORTx analysis of abundances of $CD4^+$ $IRF4^+$ effector Tregs in stage I LUAD (C1- and nonC1-LUAD samples) in Cohorts 1 and 3 samples. *Y*-axes, CIBERSORTx score representing an estimation of cell fraction of each specific subpopulation in each tumor sample. Significant *p* values were computed by the Wilcoxon test. Median Ratio, ratios of median CIBERSORTx scores in C1 vs. nonC1. **D** Hierarchical clustering analysis of the 10-gene expression signature in the CSS cohort (*N*=44). C1- and nonC1-LUAD samples are colored as per the legend. The false-discovery rates (FDR; expressed as –Log) of 10 genes differential expression analysis of C1 vs. nonC1 are also shown and colored as per the legend. **E** Box plots of THY1/CD90 gene expression measured by qRT-PCR in the CSS cohort (*N*=44). *Y*-axes, normalized (GUSB, HPRT1, and ESD1 genes used as housekeeping) quantification cycle (Cqn). **F** Immunohistochemistry analysis of $CD4^+$ and $CD8^+$ immune cells, and PD-L1 expression in C1- (*N*=17) and nonC1-LUAD (*N*=27) samples of CSS cohort. Significant *p* values were computed by the Likelihood-ratio test.

II.4 Discussion

In this study, we presented proof of a molecular subtype of LUAD, called C1-LUAD, which we identified through the use of a prognostic 10-gene signature. This 10-gene signature comprises genes involved in divergent mechanisms but that are synergistic in the progression of cancer, including (1) cell proliferation genes (E2F1, E2F4, RRM2, MCM6), (2) genes involved in development/stemness (HOXB7 [40, 41], SF3B1 [42]), cell migration and invasion (SERPINB5 [43], NUDCD1 [44], HSPG2 [45]), and (3) genes involved in epithelial airway differentiation (SCGB3A1 [46,47,48]). Overall, we have observed that the expression changes in a substantial proportion of these genes indicate the aggressive behavior of C1-LUAD. Gene expression profile analysis revealed approximately 3000 genes that were differentially regulated between C1- and nonC1-LUAD, with enriched gene expression signatures linked to the activation of molecular mechanisms related to tumor dedifferentiation and progression (Fig. 4A). The C1-LUAD displayed an increased expression of cancer stem cell (CSC) and epithelial-mesenchymal transition (EMT) markers and a reduced expression of genes related to lung cell lineage. This is in line with our previous observations in which the overexpression of HOXB7, a member of the 10-gene signature, promotes cell reprogramming by inducing the LIN28B, a pluripotency factor (Fig. 4B [40]). Interestingly, a sizable fraction (approximately 15-18%) of these C1-LUAD patients featured invasive mucinous adenocarcinoma (IMA) [24], which are characterized by missing lineage-specifying factors in the lung, such as the NKX2-1 (TTF-1) gene [13] and the SOX2 gene [14], and gain of KRAS(G12D) mutation, which induce the lineage conversion to gut differentiation (e.g., CDX2, TFF1, HNF4a, PDX1, MUC2, MUC6 overexpression) [24, 26]. Furthermore, the IMA subtype, but not the other C1-LUAD, was identified as having an enrichment in RAS-related signatures by the GSEA analysis (Fig. 4D), leading to the retention of stem cell-related signatures. Activation of RAS signaling pathway may trigger the conversion of C1-LUAD tumors into gut-like differentiated entities, an area that necessitates further research. The elevated expression of keratin 6A (KRT6A; Fig. 3C) supports the less differentiated state of C1-LUAD. Keratin 6A is a marker of p63+/KRT5+ distal airway stem cells that act as stem/progenitor cells for the regeneration of damaged

alveoli. Notably, there was significant enrichment of an expression signature of a human proximal stem cell (HPSC) in LUAD cells [28]. HPSC cells are characterized by low expression of NKX2-1 and mixed cellular identity and are thought to serve as a precursor to the EMT state, which drives the metastatic capacity in primary LUAD [28]. Furthermore, C1-LUAD tumors demonstrated an increased occurrence of genetic mutations, indicating an elevated mutator phenotype, as demonstrated by a higher tumor mutational burden (TMB) and an enriched presence of APOBEC-related substitution signatures, along with an upregulated expression of APOBEC3B. It is noteworthy that the expansion of subpopulations of CSCs can increase the genetic and non-genetic intra-tumor heterogeneity (ITH) [55]. The ITH and CSCs have been shown to impact the cytotoxic activity of tumor-infiltrating lymphocytes [56] which plays a crucial role in tumor progression. The cancer cells can reprogram a range of host cells, such as resident and infiltrating cells, by interacting with membrane receptors, the release of cytokines and microvesicles, to escape immunosurveillance during metastatic dissemination [57]. Our immune gene expression profiling analyses revealed heterogeneity in lung adenocarcinoma, with C1-LUAD exhibiting a more permissive immune environment than other LUAD subtypes. C1-LUAD exhibited a specific immune molecular subtype pattern, as originally described by Thorsson et al. [37], correlating with the infiltration of immunosuppressive Th2 cells involved in cancer progression [57] and with the depletion of cytotoxic Th1 and Th17 cells. The presence of a more favorable immune environment in C1-LUAD was further reinforced by the higher abundance of Tregs IRF4+. Infiltrating Tregs IRF4+ cells in tumors were found to have a negative correlation with cytotoxic CD8+ T cells and to co-occur with exhausted T cells (CD8, C5) [38]. Smoking tobacco was found to suppress the immune response and to associate with tumor immune suppressive phenotypes [60]. Interestingly, our analysis of IHC data in our cohort of samples (CSS) confirmed a reduction in cytotoxic T cells (CD8+) in C1-LUAD samples with a moderate trend towards significance (Fig. 6F).

A limitation of present study is the unavailability of molecular data (beyond gene expression data) in Cohorts 2 and 3. However, Cohort 1 is actually considered the benchmark for lung cancer molecular analysis with matched RNA, DNA, protein profile data, and molecular (i.e., TRU, non-TRU) immune subtypes characterization. However, Cohorts 2 and 3 are both monocentric consecutive series, with accurate follow-up information and previously used for survival analysis [61, 62].

Despite previous research has suggested that different molecular subtypes of lung adenocarcinoma exist, their biological and molecular features have not been thoroughly investigated. In this study, we demonstrate that the C1-LUAD molecular subtype is characterized by the loss of lung cell identity and the acquisition of stem cell-like properties, as well as mutator and immune evasion phenotypes, which contribute to its aggressive behavior. Additional research is required to investigate the underlying mechanisms

responsible for the development of C1-LUAD, which may lead to the identification of new and effective therapeutic targets for LUAD patients.

II.5 Material and Methods

Cell culture

All LUAD cell lines were obtained from ATCC and maintained in RPMI (Gibco; Thermo Fisher Scientific, Inc., Waltham, MA, USA) with 10% FBS, 1% penicillin/streptomycin (Gibco; Thermo Fisher Scientific, Inc., Waltham, MA, USA). Cells were growth at 37 °C in a humidified incubator with 5% CO2. Mycoplasma testing was routinely performed during all experimental procedures.

Real-time PCR gene expression profiling

The total mRNA was extracted from all 85 FFPE blocks using the RNeasy FFPE Kit (Qiagen, Inc.; Hilden, Germany). The qRT-PCR analysis was performed using the QuantStudio 12K Flex Real-Time PCR System. The expression of the GUSB gene was used as mRNA quality control. Further details can be found in the Supplementary Information.

Immunohistochemical analysis

The IHC analysis was performed by an automated Autostainer Link 48 (Agilent DAKO; Carpinteria, CA, USA) platform according to the manufacturer's instructions. The histopathological evaluation of tumor tissues was performed by a board-certified pathologist (PG). Further details can be found in the Supplementary Information.

Gene expression analysis

Details on cohorts of samples, data sets availability, data processing and analysis can be found in the Supplementary Information.

II.6 References

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer J Clinicians. 2018;68:394–424.

[2] Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma. J Thorac Oncol. 2011;6:244–85.

[3] Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365:395–409.

[4] de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial. N Engl J Med. 2020;382:503–13.

[5] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70:7–30.

[6] Pribluda A, de la Cruz CC, Jackson EL. Intratumoral heterogeneity: from diversity comes resistance. Clin Cancer Res. 2015;21:2916–23.

[7] Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014;511:543–50.

[8] Zhang J, Fujimoto J, Wedge DC, Song X, Seth S, Chow CW, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. Science. 2014;346:256–9.

[9] Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, PCAWG Mutational Signatures Working Group, PCAWG Consortium, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578:94–101.

[10] de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science. 2014;346:251–6.

[11] Chen F, Zhang Y, Parra E, Rodriguez J, Behrens C, Akbani R, et al. Multiplatform-based molecular subtypes of non-small-cell lung cancer. Oncogene. 2017;36:1384–93.

[12] Devarakonda S, Morgensztern D, Govindan R. Genomic alterations in lung adenocarcinoma. Lancet Oncol. 2015;16:e342–51.

[13] Yatabe Y, Mitsudomi T, Takahashi T. TTF-1 expression in pulmonary adenocarcinomas. Am J Surg Pathol. 2002;26:767–73.

[14] Ferone G, Song J-Y, Sutherland KD, Bhaskaran R, Monkhorst K, Lambooij J-P, et al. SOX2 is the determining oncogenic switch in promoting lung squamous cell carcinoma from different cells of origin. Cancer Cell. 2016;30:519–32.

[15] Dama E, Melocchi V, Dezi F, Pirroni S, Carletti RM, Brambilla D, et al. An aggressive subtype of stage I lung adenocarcinoma with molecular and prognostic characteristics typical of advanced lung cancers. Clin Cancer Res: Off J Am Assoc Cancer Res. 2017;23:62–72.

[16] Bianchi F, Nuciforo P, Vecchi M, Bernard L, Tizzoni L, Marchetti A, et al. Survival prediction of stage I lung adenocarcinomas by expression of 10 genes. J Clin Invest. 2007;117:3436–44.

[17] Takeuchi T, Tomida S, Yatabe Y, Kosaka T, Osada H, Yanagisawa K, et al. Expression profile–defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. JCO. 2006;24:1679–88.

[18] Bryant CM, Albertus DL, Kim S, Chen G, Brambilla C, Guedj M, et al. Clinically relevant characterization of lung adenocarcinoma subtypes based on cellular pathways: an international validation study. PLoS ONE. 2010;5:e11712.

[19] Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci USA. 2001;98:13790–5.

[20] Beer DG, Kardia SLR, Huang C-C, Giordano TJ, Levin AM, Misek DE, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med. 2002;8:816–24.

[21] Nazarov PV, Muller A, Kaoma T, Nicot N, Maximo C, Birembaut P, et al. RNA sequencing and transcriptome arrays analyses show opposing results for alternative splicing in patient derived samples. BMC Genomics. 2017;18:443.

[22] Cheung WKC, Nguyen DX. Lineage factors and differentiation states in lung cancer progression. Oncogene. 2015;34:5771–80.

[23] Moisés J, Navarro A, Santasusagna S, Viñolas N, Molins L, Ramirez J, et al. NKX2–1 expression as a prognostic marker in early-stage non-small-cell lung cancer. BMC Pulm Med. 2017;17:197.

[24] Tata PR, Chow RD, Saladi SV, Tata A, Konkimalla A, Bara A, et al. Developmental history provides a roadmap for the emergence of tumor plasticity. Developmental Cell. 2018;44:679–93.

[25] Fares J, Fares MY, Khachfe HH, Salhab HA, Fares Y. Molecular principles of metastasis: a hallmark of cancer revisited. Sig Transduct Target Ther. 2020;5:28.

[26] Sonzogni A, Bianchi F, Fabbri A, Cossa M, Rossi G, Cavazza A, et al. Pulmonary adenocarcinoma with mucin production modulates phenotype according to common genetic traits: a reappraisal of mucinous adenocarcinoma and colloid adenocarcinoma: Reappraising mucinous and colloid adenocarcinoma of the lung. J Path: Clin Res. 2017;3:139–51.

[27] Tammela T, Sanchez-Rivera FJ, Cetinbas NM, Wu K, Joshi NS, Helenius K, et al. A Wnt-producing niche drives proliferative potential and progression in lung adenocarcinoma. Nature. 2017;545:355–9.

[28] Marjanovic ND, Hofree M, Chan JE, Canner D, Wu K, Trakala M, et al. Emergence of a high-plasticity cell state during lung cancer evolution. Cancer Cell. 2020;38:229–46.

[29] Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER, et al. Next-generation characterization of the cancer cell line encyclopedia. Nature. 2019;569:503–8.

[30] Dontu G. In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev. 2003;17:1253–70.

[31] Lu H, Clauser KR, Tam WL, Frose J, Ye X, Eaton EN, et al. A breast cancer stem cell niche supported by juxtacrine signalling from monocytes and macrophages. Nat Cell Biol. 2014;16:1105–17.

[32] Yan X, Luo H, Zhou X, Zhu B, Wang Y, Bian X. Identification of CD90 as a marker for lung cancer stem cells in A549 and H446 cell lines. Oncol Rep. 2013;30:2733–40.

[33] Tokheim C, Karchin R. CHASMplus reveals the scope of somatic missense mutations driving human cancers. Cell Syst. 2019;9:9–23.

[34] Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. Nucleic Acids Res. 2019;47:D941–7.

[35] Swanton C, McGranahan N, Starrett GJ, Harris RS. APOBEC enzymes: mutagenic fuel for cancer evolution and heterogeneity. Cancer Discov. 2015;5:704–12.

[36] Bremnes RM, Busund L-T, Kilvær TL, Andersen S, Richardsen E, Paulsen EE, et al. The role of tumor-infiltrating lymphocytes in development, progression, and prognosis of non–small cell lung cancer. J Thorac Oncol. 2016;11:789–800.

[37] Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang T-H, et al. The immune landscape of cancer. Immunity. 2018;48:812–30.

[38] Alvisi G, Brummelman J, Puccio S, Mazza EMC, Tomada EP, Losurdo A, et al. IRF4 instructs effector Treg differentiation and immune suppression in human cancer. J Clin Investig. 2020;130:3137–50.

[39] Yu H, Boyle TA, Zhou C, Rimm DL, Hirsch FR. PD-L1 expression in lung cancer. J Thorac Oncol. 2016;11:964–75.

[40] Monterisi S, Lo Riso P, Russo K, Bertalot G, Vecchi M, Testa G, et al. HOXB7 overexpression in lung cancer is a hallmark of acquired stem-like phenotype. Oncogene. 2018;37:3575–88.

[41] Care A, Valtieri M, Mattia G, Meccia E, Masella B, Luchetti L, et al. Enforced expression of HOXB7 promotes hematopoietic stem cell proliferation and myeloid-restricted progenitor differentiation. Oncogene. 1999;18:1993–2001.

[42] Dolatshad H, Pellagatti A, Fernandez-Mercado M, Yip BH, Malcovati L, Attwood M, et al. Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. Leukemia. 2015;29:1092–103.

[43] Tian C, Öhlund D, Rickelt S, Lidström T, Huang Y, Hao L, et al. Cancer cell–derived matrisome proteins promote metastasis in pancreatic ductal adenocarcinoma. Cancer Res. 2020;80:1461–74.

[44] Han B, Zhang Y-Y, Xu K, Bai Y, Wan L-H, Miao S-K, et al. NUDCD1 promotes metastasis through inducing EMT and inhibiting apoptosis in colorectal cancer. Am J Cancer Res. 2018;8:810–23.

[45] Grindel BJ, Martinez JR, Tellman TV, Harrington DA, Zafar H, Nakhleh L, et al. Matrilysin/MMP-7 cleavage of perlecan/HSPG2 complexed with semaphorin 3A supports FAK-mediated stromal invasion by prostate cancer cells. Sci Rep. 2018;8:7262.

[46] McCauley KB, Alysandratos K-D, Jacob A, Hawkins F, Caballero IS, Vedaie M, et al. Single-cell transcriptomic profiling of pluripotent stem cell-derived SCGB3A2+ airway epithelium. Stem Cell Rep. 2018;10:1579–95.

[47] Naizhen X, Kido T, Yokoyama S, Linnoila RI, Kimura S. Spatiotemporal expression of three secretoglobin proteins, SCGB1A1, SCGB3A1, and SCGB3A2, in mouse airway epithelia. J Histochem Cytochem. 2019;67:453–63.

[48] Reynolds SD, Reynolds PR, Pryhuber GS, Finder JD, Stripp BR. Secretoglobins SCGB3A1 and SCGB3A2 define secretory cell subsets in mouse and human airways. Am J Respir Crit Care Med. 2002;166:1498–509.

[49] Kumar PA, Hu Y, Yamamoto Y, Hoe NB, Wei TS, Mu D, et al. Distal airway stem cells yield alveoli in vitro and during lung regeneration following H1N1 influenza infection. Cell. 2011;147:525–38.

[50] Zuo W, Zhang T, Wu DZ, Guan SP, Liew A-A, Yamamoto Y, et al. p63+Krt5+ distal airway stem cells are essential for lung regeneration. Nature. 2015;517:616–20.

[51] Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the evolution of non–small-cell lung cancer. N Engl J Med. 2017;376:2109–21.

[52] Wolf J, Seto T, Han J-Y, Reguart N, Garon EB, Groen HJM, et al. Capmatinib in *MET* Exon 14–mutated or *MET*-amplified non-small-cell lung cancer. N Engl J Med. 2020;383:944–57.

[53] Orkin SH, Hochedlinger K. Chromatin connections to pluripotency and cellular reprogramming. Cell. 2011;145:835–50.

[54] Schoenfeld AJ, Bandlamudi C, Lavery JA, Montecalvo J, Namakydoust A, Rizvi H, et al. The genomic landscape of *SMARCA4* alterations and associations with outcomes in patients with lung cancer. Clin Cancer Res. 2020;26:5701–8.

[55] Prasetyanti PR, Medema JP. Intra-tumor heterogeneity from a cancer stem cell perspective. Mol Cancer. 2017;16:41.

[56] Miranda A, Hamilton PT, Zhang AW, Pattnaik S, Becht E, Mezheyeuski A, et al. Cancer stemness, intratumoral heterogeneity, and immune response across cancers. Proc Natl Acad Sci USA. 2019;116:9020–9.

[57] Burkholder B, Huang R-Y, Burgess R, Luo S, Jones VS, Zhang W, et al. Tumor-induced perturbations of cytokines and immune cell networks. Biochim Biophys Acta. 2014;1845:182–201.

[58] Richards CH, Mohammed Z, Qayyum T, Horgan PG, McMillan DC. The prognostic value of histological tumor necrosis in solid organ malignant disease: a systematic review. Future Oncol. 2011;7:1223–35.

[59] Jaillon S, Ponzetta A, Di Mitri D, Santoni A, Bonecchi R, Mantovani A. Neutrophil diversity and plasticity in tumour progression and therapy. Nat Rev Cancer. 2020;20:485–503.

[60] Stämpfli MR, Anderson GP. How cigarette smoke skews immune responses to promote infection, lung disease and cancer. Nat Rev Immunol. 2009;9:377–84.

[61] Schabath MB, Welsh EA, Fulp WJ, Chen L, Teer JK, Thompson ZJ, et al. Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma. Oncogene. 2016;35:3209–16.

[62] Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. Cancer Res. 2012;72:100–11.

II.7 Supplemental Information

Supplemental Methods

Gene expression profile analysis (RNAseq and Affymetrix) of LUAD datasets

From the cBIO data portal (https://www.cbioportal.org/), we downloaded gene expression data of Cohort 1 (i.e., TCGA-LUAD) and clinical information of 515 patients with lung adenocarcinoma as well gene expression data of TCGA lung squamous cell carcinoma (SCC) samples (N=499, originally used by Tata et al.). Cohort 2 (GSE31210) and Cohort 3 (GSE72094) gene expression data and clinical information of 226 and 442 patients with lung adenocarcinoma, respectively, were downloaded from the GEO Database (https://www.ncbi.nlm.nih.gov/geo/). The CCLE reads per kilobase of transcript per million reads mapped (RPKM) data were downloaded from the Depmap portal (https://depmap.org/portal/; 18Q2 release). From the original file, we extracted only lung adenocarcinoma cell lines, resulting in a total of 76 cell lines. Details on datasets normalization, gene expression profile analysis, mutation and CNV analysis and immune expression profiling, are reported in the Supplementary Information. All statistical tests were two-sided. We used RSEM normalized counts from the Cohort 1 (TCGA-LUAD) and from the TCGA-LUSC (N=499) cohort, while we used log2 normalized data available in GEO database (https://www.ncbi.nlm.nih.gov/geo/) for Cohort 2 (accession #: GSE31210) and Cohort 3 (accession #: GSE72094) datasets. Since Cohort 2 and Cohort 3 datasets derive from Affymetrix microarray analysis, multiple probes mapping to unique genes were averaged in order to obtain a single expression value for each gene. Median centered data were used for all analyses. For differential gene expression (DEG) analysis, we selected most

informative genes in terms of their variability of gene expression across samples, i.e. those belonging to the upper 25% (standard deviation > Q3) of the standard deviation distribution. A total of 4899 genes for Cohort 1, 5877 genes for Cohort 2 and 5530 genes for Cohort 3, were retain for subsequent DEG analyses. We then applied DESeq2 R package (Cohort 1) and Wilcoxon test (Cohort 2 and 3), followed by Benjamini-Hochberg adjustment to identify differentially expressed genes. For hierarchical clustering analysis, after median centering data, the uncentered correlation and the centroid linkage methods were used to cluster gene expression data in Cluster 3.0 for Mac OS X (http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm). Hierarchical clustering structures and heatmaps were generated using Java TreeView 1.1.1 (http://jtreeview.sourceforge.net). JMP 15 (JMP®, Version 15. SAS Institute Inc., Cary, NC, 1989-2019) generated Kaplan-Meier survival plots based on clustering analysis together with calculation of p-values based on the log-rank test.

Dataset selection criteria (GEO database)

In GEO database (homo sapiens), we first searched for datasets (high throughput sequencing, or array-based) with >200 samples by using the term "lung adenocarcinomas". A total of 99 datasets were identified: i) 94 datasets were excluded because total samples were less than 200; ii) 1 dataset was excluded because only expression data for 8 out of 10 genes (10-gene signature) were available (GSE68465); iii) 3 datasets were excluded because contain samples derived from tumor cell lines (GSE83744, GSE5949, GSE57083); iv) 1 dataset was then selected for our analyses i.e., the Cohort 2 (GSE31210). Furthermore, we searched for datasets (high throughput sequencing, or array-based) with >200 samples by using the term "lung adenocarcinoma" and "survival". A total of 73 datasets were identified: i) 70 datasets were excluded because total samples were less than 200; ii) 1 dataset was excluded because only expression data for 8 out of 10 genes (10-gene signature) were available (GSE68465); iii) 1 dataset was excluded because samples were derived from non-tumor lung tissue (GSE71181); iv) 1 dataset was then selected for our analyses i.e., the Cohort 3 (GSE72094).

CSS lung adenocarcinoma cohort

Between January 2017 and April 2020, 85 patients with lung adenocarcinoma underwent surgery at the Casa Sollievo della Sofferenza Research Hospital (CSS, San Giovanni Rotondo, Italy). Informed consent for the experimentation was obtained from all patients involved in the study. None of these patients received preoperative chemotherapy. Clinical information was obtained through review of medical records. Vital status was assessed through the Vital Records Offices of the patients' towns of residence or by contacting directly the patients or their families.

Survival analysis

Overall survival was defined as the time from the date of tumor resection until death from any cause and estimated by the Kaplan-Meier method. Follow-up was truncated at 3 years in order to reduce the potential overestimation of overall mortality with respect to lung-cancer specific mortality. Univariate and multivariable Cox regression models were used to relate survival time to clinical-pathological and molecular information. All survival analyses were performed using the SAS software, version 9.4 (SAS Institute, Inc., Cary, NC, USA) and R 3.3.1 (R Core Team, R Foundation for Statistical Computing, Vienna, Austria 2016). P-values lower than 0.05 were considered statistically significant.

Principal Component Analysis (PCA)

PCA was performed using most informative genes in terms of variability (standard deviation>Q3) as previously described. We used the prcomp R function and the R library cowplot for plotting. PCA was limited to lower-order components (Dim1 and Dim2) due to previously reported low dimensionality of gene expression data.

Lineage analysis

To be consistent with the classification of Cohort 1 patients into LUAD and IMA reported in Tata et al. we used the very same data to analyze the distribution of C1- and nonC1-LUAD patients according to lineage specific markers (SOX2 and NKX2-1). IMA (aka MAD)/AD classification was retrieved from the supplemental data available in Tata et al. We then retrieved gene expression data of genes preferentially expressed in the lung, esophagus and midgut/hindgut for all the 3 datasets we analyzed. Using this list of genes, we calculated the C1- vs. nonC1-LUAD fold change (FC) for stage I patients as the difference of the means using R and we plot the results using JMP 15.

Gene Set Enrichment Analysis (GSEA)

GSEA (GSEA, https://www.gsea-msigdb.org/gsea/index.jsp) was performed using Signal2Noise metric, 1000 random sample sets permutation, and median gene expression values for class comparison. Significant Gene Sets were considered as those with a false-discovery rate (q-value) less then 25% (default GSEA setting). Among all the enriched signatures we identified those regarding stem/EMT phenotypes, hypoxia status, immune system differentiation and metastasis as hallmarks of aggressiveness. The complete results are available online as 'Supplementary Dataset SD2-7'. For each dataset, we compared the number of observed enriched signatures in the aforementioned hallmarks of aggressiveness to the expected distribution (Z-test) derived from 100 GSEA analyses (comparing C1 vs nonC1) of 100 random datasets.

CCLE lung cancer cell lines classification

Bulk gene expression profile characterizing C1 and non-C1 tumor subtypes is the result of complex interaction of tumors and their microenvironment. Therefore, in order to classify CCLE lung adenocarcinoma cell lines (growth in 2D, without any microenvironment) in C1-like or nonC1-like subtypes, we implemented centroid correlation where the bulk 10-gene expression profile of C1- and nonC1-LUAD was used as reference for calculating correlation scores with CCLE 10-gene expression profiles. SCGB3A1 and SERPINB5 of the 10-gene signature were barely detected in CCLE datasets (~95% of cell lines were negative to SCGB3A1, and ~40% of cell lines scored negative to SERPINB5), therefore they were excluded from further analyses. Next, Z-score values of each gene of the 10-gene signature were computed in all samples of Cohort 1. We then calculated the averages of the 10-gene signature Z-scores in the cluster of C1-LUAD patients with unfavorable prognosis (N=201) and in the cluster of 'C4' patients (N=177) previously reported with most favorable prognosis. These Z-score averages were used as C1-LUAD and nonC1-LUAD centroids, to calculate their expression correlation (Pearson, **ρ**) to that of the CCLE dataset. Next, we transformed **ρ** values to power of ten (i.e., 10^**ρ**) to stabilize variance to get normally distributed data, and we used these values to stratify CCLE cell lines. A cell line was classified as C1-like if its $10^{\rho_{C1}}$ was greater than its $10^{\rho_{nonC1}}$, and vice versa. CCLE cell lines were finally stratified in in C1-like (N=41) or nonC1-like cells (N=35). To calculate the probability of a cell line to be classified as C1- or nonC1-like we used ridge-penalized unconditional logistic regression that was applied in the Cohort 1(N=378) to model the odds of being in the C1-LUAD cluster (vs. nonC1-LUAD cluster) as a function of the 8 genes detected in CCLE dataset. The expression of each gene was dichotomized according to the median. Cross-validated (10-fold) log-likelihood with optimization (50 simulations) of the tuning penalty parameter was applied. Probability of being in the C1-LUAD cluster was estimated, and model performance was assessed using the area under the receiver operating curve (AUC=0.83).

Proliferation Assay

Cell proliferation (i.e., doubling time) was evaluated with the CyQUANT Cell Proliferation Assay Kit (Molecular Probes, Thermo Fisher Scientific, Inc., Waltham, MA, USA), accordingly with manufactures instruction. Briefly, lung adenocarcinoma cell lines were seeded onto 96-well plates in exponential phase and CyQuant was added in a ratio of 1:10 directly in cell media at 24h and 96h. Fluorescence was measured using a microplate spectrofluorometer (Synergy HTX; BioTek, Agilent, Carpinteria, CA, USA) at excitation and emission wavelengths of 480 and 520 nm, respectively. Then, the growth rate (gr) was computed as follow:

$$gr = \frac{\ln\left(\frac{N(t)}{N(0)}\right)}{t}$$

- $N(t)$ = number of cells at time t
- $N(0)$ = number of cells at time 0
- gr = growth rate
- t = time in hours

Finally, the doubling time was obtained as follow:

$$doubling\ time = \frac{\ln(2)}{growth\ rate}$$

Migration (wound-healing) assay

Cells were seeded onto a 6-well plate and left to almost reach confluency (90-95% confluence) at 37°C in a humidified incubator with 5% CO_2. Once obtained the desired confluence, cells were starved over-night using RPMI with 1% FBS and 1% penicillin/streptomycin. The day after, a scratch was made on the cell monolayer by using a sterile P-200 pipette tip. Cells were then washed with 1 ml of sterile 1X PBS to remove floating cells and debris, followed by the addition of 2 ml of complete growth media (10% FBS). Ten images were acquired at 0h time point with a 10X magnification using an Eclipse TE300 fluorescence microscope (Nikon; Tokyo, Japan). Cells were incubated (at 37°C in a humidified incubator with 5% CO_2) and other ten images were acquired after 24h. Fiji software (ImageJ) was used to measure the wound area at 0h and 24h to calculate the percentage of healing. Representative images can be found in Supplemental Figure S1.

Invasion Assay

Invasion assays were performed using trans-well chambers with 8.0-μm pore polycarbonate membrane (Falcon; Corning, NY, USA) coated with Matrigel growth factor reduced (Corning; Corning, NY, USA, Lot 4132005) at a concentration of 150 μg/ml in a 24-well plate. Cells were suspended in RPMI medium (0% FBS) and seeded at a density of $1x10^5$ into the upper chamber of the trans-well. In the lower chamber, cell culture RPMI medium supplemented with 20% of FBS was added and used as chemoattractant. After 24h of incubation at 37°C in a humidified incubator with 5% CO_2, trans-wells were fixed using 4% PFA and cells nuclei stained using DAPI (1ng/ml working solution; Sigma-Aldrich; St. Louis, MO, USA). Cells remained in the upper chamber were removed by using cotton swabs and washed 2 times with 1ml of 1X PBS. After DAPI staining (1ng/ml working solution; Sigma-Aldrich), ten images were acquired at 10X magnification for each cell line using Eclipse TE300 Microscope (Nikon). Invaded cells were evaluated counting nuclei with Cell Count tool of Fiji software (ImageJ). Representative images can be found in Supplemental Figure S1.

Tumor Spheres Formation Assay

Spheroid formation was performed in a 12-well plates coated (48h) with poly-HEMA (4.8mg/well; Sigma Aldrich). Cells in serum-free RPMI were passed through a 50μm strainer and a total of 500 cells were admixed to Matrigel (growth factor reduced; Corning, Lot 4132005) in a 1:1 ratio and left to solidify for 1h at 37°C in a humidified incubator with 5% CO_2. Spheroids were growth in a DMEM-F12 (Sigma Aldrich) supplemented with 2 mM Glu, 100 U/ml, 5 μg/ml insulin, 0.5 μg/ml hydrocortisone, 2% B27 (Invitrogen; Thermo Fisher Scientific, Inc., Waltham, MA, USA), 20 ng/ml EGF and human b-FGF, and 4 μg/ml heparin. After Matrigel solidification, 1ml of pre-warmed (37°C) spheroid media was slowly added in the center of the well to avoid Matrigel disruption. After 20 days, spheres were counted and images acquired at 10X magnification using the Eclipse TE300 Microscope (Nikon). Images were analyzed using Fiji software (ImageJ) to measure sphere diameter by setting the cut-off at 50μm.

Phenotypic score

Each parameter was transformed into z-scores and a phenotypic score was defined as the sum over the four z-scores and divided by 2, obtaining a variable still following the standard distribution (mean of 0 and standard deviation of 1). Cell lines (N=12) were also ordered by descending proliferation and increasing migration, invasion, and tumor spheres formation, and four ranks (from 1 to 12) were assigned to each cell line, for each assay. Next, we obtained a phenotypic ranking of the 12 cell lines by summing the ranks obtained (higher values correspond to more aggressive cell line). T-test and Wilcoxon test were used to compare C1-like/non-C1-like cell lines phenotypic score and phenotypic ranking distributions, respectively.

Mutation analysis

The Mutect maf data available for 510 out of 515 patients of Cohort 1 was downloaded from TCGA portal (https://portal.gdc.cancer.gov/) and was used in our analysis. Silent mutations were discarded. Tumor mutation burden (TMB) was calculated as the number of mutations over the total Mb sequenced (34.8 Mbases). We downloaded COSMIC Single Base Substitutions (SBS) signatures corresponding to each Cohort 1 patient from synapse.org website (accession ID: syn11801497) and we calculated the average of the number of mutations belonging to each SBS across C1- and nonC1-LUAD patients, both in all stages and stage I. Somatic mutations were also annotated with tumor suppressor genes (TSG), retrieved from the tumor suppressor gene database, and oncogenes (ONC), retrieved from oncogene database. TSG and ONC rate for stage I C1- and nonC1-LUAD subtypes were calculated as the number of mutations in TSG/ONC genes over the number of mutated patients.

Copy Number Alteration analysis

Copy number values for DNA repair genes (N=18) were downloaded from cBIO data portal (https://www.cbioportal.org/). We calculated, for each gene, the number of amplification (GISTIC value = 2 or 1) and deletions (GISTIC value = -2 or -1) for patients of C1- and nonC1-LUAD subset, both in all stages and stage I, as the total number of patients with amplifications/deletions over the total number of patients.

Thorsson dataset analysis

We retrieved data regarding the immune subtypes defined by Thorsson et al. [37] and the subclonal genome fraction for 510 Cohort 1 samples from supplementary data of Thorsson et al. [37] Each sample was assigned to a specific immune subtype class according to the prevalence of different expression traits of the immune system as described in Thorsson et al. [37]

Real Time PCR gene expression profiling

Samples (N=41) with GUSB housekeeping expression higher than 30 CT (low quality RNA) were excluded from further analyses. The CD90/THY1 expression was also detected in 41 out of 44 (CTn<30) samples.

CIBERSORTx analysis. CIBERSORTx was run using the online web tool (https://cibersortx.stanford.edu/index.php) and following developers' instructions. For Cohort 1, i.e. RNAseq level data, the CIBERSORTx analysis was conducted using the following settings: disabled quantile normalization, absolute mode running and 100 permutations. For Cohort 2 and Cohort 3, i.e. Affymetrix level data, we used the following settings: batch correction (B-mode), absolute mode running and 100 permutations. CIBERSORTx score is an estimation of cell fraction of each specific subpopulation in each tumor sample.

Immunohistochemical analysis

Briefly, the 3μm thick FFPE tissue sections were deparaffinized in xylene, rehydrated in graded alcohols, washed in double-distilled water, and pretreated with the DAKO solution (EnVision FLEX Target Retrieval Solution, High pH) at 97°C. The slides were incubated with primary monoclonal antibodies against CD4 (clone 4B12, Ready-touse, Agilent DAKO), CD8 (clone C8/144B diluted 1:25, Agilent DAKO), for 30 min. The Antigen–antibody reaction was visualized using the EnVision FLEX kit with diaminobenzidine as chromogen. Finally, the slides were counterstained with hematoxylin and covered.

A semi-quantitative scoring method of intratumoral lymphocytes on each representative lung adenocarcinoma full section was adopted. Necrosis, artifacts and macrophages were

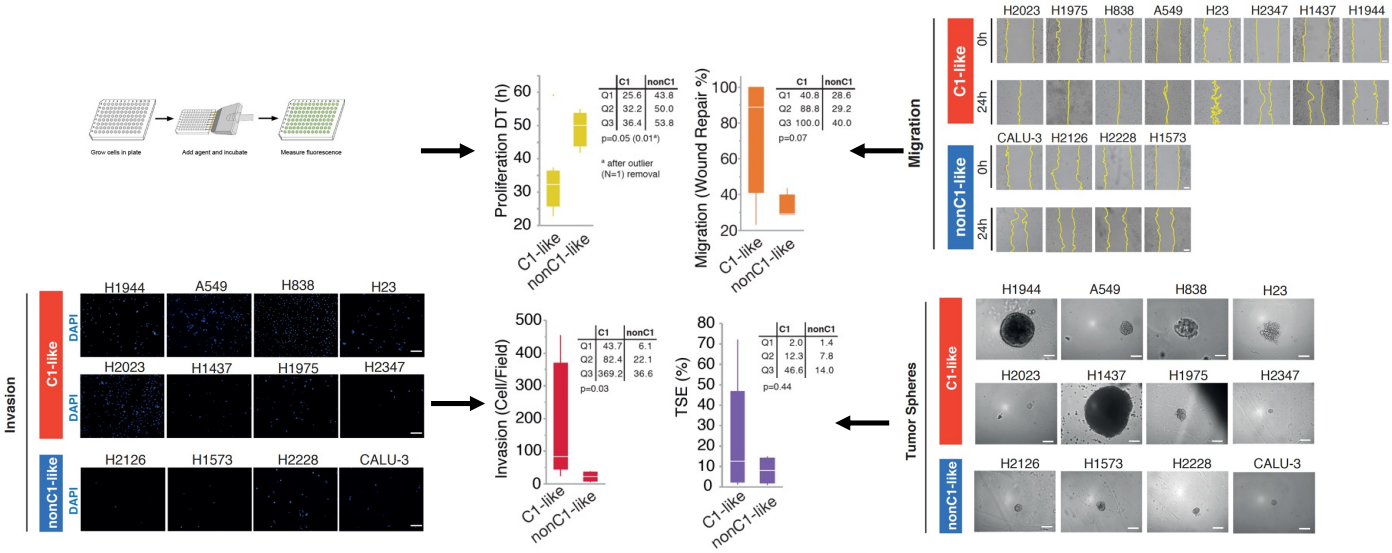
excluded. CD4 and CD8 immunoreactivity was respectively evaluated as a score low and high (low: 1-49% and high :>50%) as the percentage of total number of lymphocytes within tumor areas. Programmed death-ligand 1 (PD-L1) immunohistochemical testing was performed using companion diagnostic 22C3 pharmDx kit (Dako, Carpintera, CA, USA) on Dako Autostainer. PD-L1 protein expression was assessed using tumor proportion score (TPS), i.e. the percentage of viable tumor cells showing partial or complete membrane staining at any intensity. PD-L1 status was clustered into three categories, negative (TPS < 1%), low (1-49%) and high (≥50%), respective.

II.8 Supplemental Tables and Figures

Cell Line	Cluster CCLE	Tumor Sphere Efficency (%)		Proliferation Doubling Time (h)		Invasion (Cell/Field)		Migration (Wound Repair %)	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE
A549	*C1-like*	33.8	5.37	22.72	0.81	297.33	16.61	94.54	2.65
H1437	*C1-like*	72	26.87	26.36	1.13	34.47	1.56	39.83	3.00
H1944	*C1-like*	50.9	4.38	37.36	2.18	92.2	2.91	22.87	0.49
H1975	*C1-like*	15.7	4.10	31.75	0.55	72.67	4.43	100.00	0.00
H2023	*C1-like*	0.8	1.41	32.59	2.01	453.97	22.17	100.00	0.00
H23	*C1-like*	1.8	5.66	33.67	1.50	71.53	4.39	83.02	1.81
H2347	*C1-like*	8.9	0.71	59.34	2.52	22.65	2.79	43.74	4.19
H838	*C1-like*	2.5	0.99	25.31	1.03	393.1	12.44	100.00	0.00
CALU-3	*nonC1-like*	3.7	0.42	50.23	0.51	37.6	2.76	43.49	1.15
H1573	*nonC1-like*	14.7	0.42	54.97	2.97	10.6	0.69	28.89	3.35
H2126	*nonC1-like*	11.9	2.40	49.69	2.44	4.65	0.67	28.56	6.10
H2228	*nonC1-like*	0.6	0.00	41.81	1.47	33.67	1.73	29.42	4.99

SE, standard error of mean

Chapter II.8 Table S1: Phenotypic characterization of the 12 lung cancer cell lines.
Results of proliferation, migration, invasion and tumor sphere efficiency, are reported with relative means and standard error (SE). Sphere forming efficiency was performed in biological duplicate while all other experiments were biological triplicates.



Chapter II.8 Figure S1: Phenotypic Screening assay in C1-like and nonC1-like cells
From left to right, box plots representing: i) the proliferation rate expressed as hours (h) for cell doubling time (DT); ii) the percentage area wound healed (expressed as percentage); iii) the number of invaded cells/field (absolute number); iv) the tumor sphere efficiency (TSE) expressed as percentage versus number of cells plated. Tables within the plots represent the first (Q1), second (Q2, median), and third (Q3) quartile of data distribution considering C1-like (C1) or nonC1-like (nonC1) cell lines. Representative images of migration, invasion assays and tumor spheroids were reported. White bars within the pictures correspond to 100um size. P-values were calculated by Wilcoxon test.

Chapter III:
Identification of a miRNA-based expression signature which capture the molecular complexity of C-LUAD aggressive molecular subtype

This Chapter answer to the Aim II. Results here contained are also part of the following publication:

Non-Coding RNAs as Prognostic Biomarkers: A miRNA Signature Specific for Aggressive Early-Stage Lung Adenocarcinomas

Elisa Damat, Valentina Melocchit, **Francesco Mazzarelli**, Tommaso Colangelo, Roberto Cuttano, Leonarda Di Candia, Gian Maria Ferretti, Marco Taurchini, Paolo Graziano and Fabrizio Bianchi *

PMID: 33333738 PMCID: PMC7768474 DOI: 10.3390/ncrna6040048

III.1 Summary

The burden of lung cancer can be alleviated through primary and secondary prevention strategies, including anti-smoking campaigns and low-dose CT screening for high-risk individuals (those over the age of 50 and who have a smoking history of over 30 packs/year). Trials of CT screening have demonstrated a stage-shift towards earlier detection of lung cancer and a decrease in mortality by approximately 20%. However, a significant proportion of early-stage patients (30-50%) still experience relapse and poor prognosis, emphasizing the need for effective prognostic biomarkers in stage I lung cancer. In this study, we employed a multi-tiered approach that involved analyzing RNA-seq and miRNA-seq data from a large cohort of lung cancer patients (TCGA-LUAD, n = 510) to identify prognostic miRNA signatures in stage I lung adenocarcinoma. The identified signatures demonstrated high accuracy (AUC ranging between 0.79 and 0.85) in predicting aggressive disease. Furthermore, using a network-based approach, we identified a minimal signature of 7 miRNAs that was validated in a cohort of FFPE lung adenocarcinoma samples (CSS, n = 44) and controlled a variety of genes associated with cancer-relevant pathways. Our results highlight the potential of miRNA-based biomarkers for prognostication in lung cancer and represent a promising step towards the clinical application of such biomarkers.

III.2 Introduction

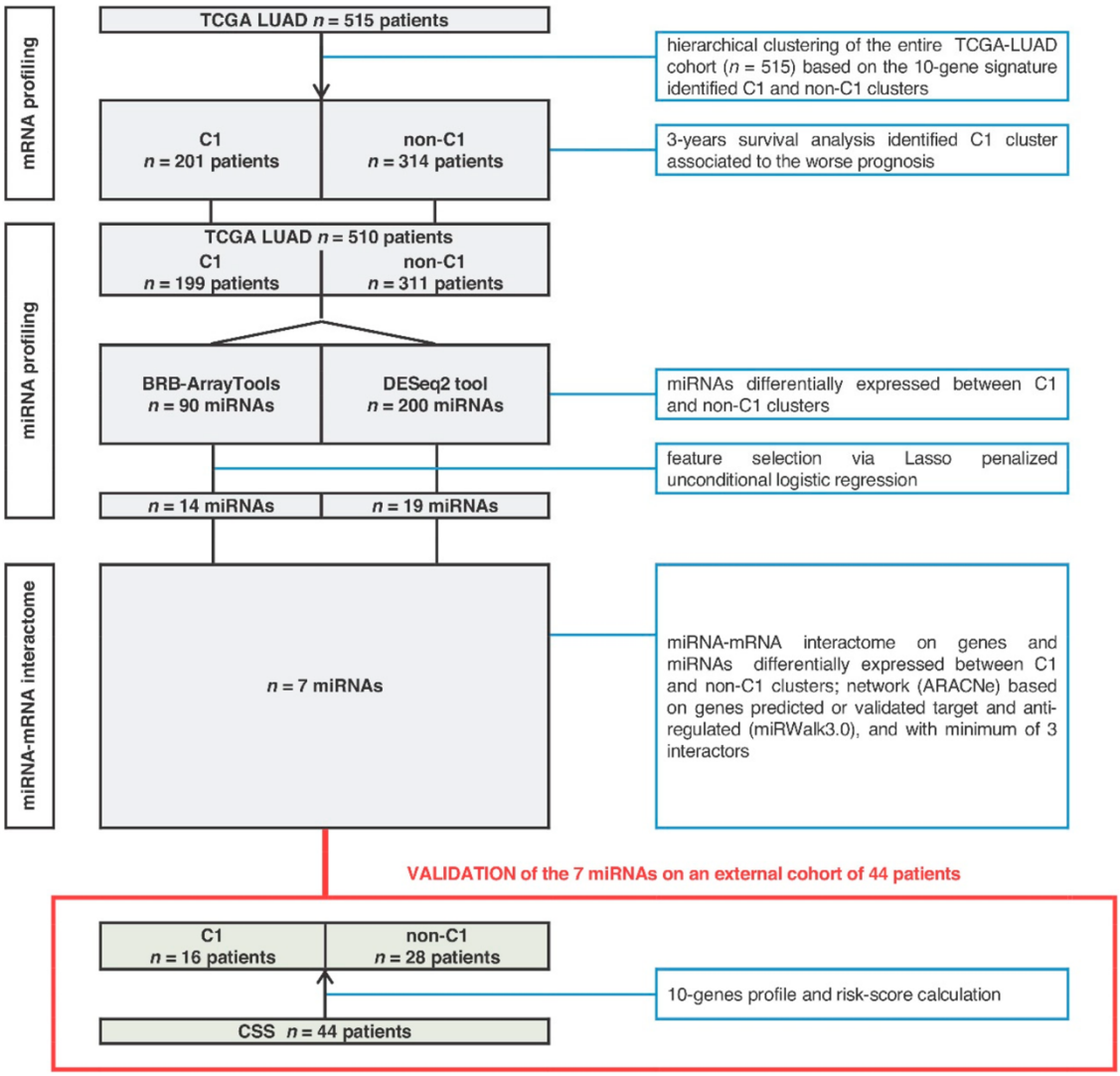
The most recent worldwide lung cancer data reveals a burden of 2.09 million new cases and 1.76 million fatalities in 2018 [1]. Non-Small-Cell Lung Cancer (NSCLC) represents the primary type of lung cancer (80-85%) comprising several heterogeneous tumor subtypes, including lung adenocarcinoma (ADC) (~40% of lung cancers), squamous cell carcinoma (SqCC) (~25% of lung cancers), and large cell carcinoma (LCC) (~10% of lung cancers) [2]. Although advances in lung cancer treatment, such as stereotactic ablative radiotherapy (SABR), targeted therapy, and immunotherapy, have been significant in the past few decades [3,4,5], global lung cancer mortality rates have remained relatively unchanged in the last 40 years, with some variability worldwide due to different lifestyle, environmental, and occupational exposures [4,6]. However, the use of primary and secondary prevention strategies, such as anti-smoking campaigns and large CT screening programs, have led to a reduction of lung cancer mortality by approximately 20% in enrolled patients, and an increase in the progression-free survival of lung cancer patients [7,8]. Nonetheless, the high degree of molecular heterogeneity in lung adenocarcinoma (ADC) contributes to the dissemination of a substantial fraction of early stage tumors (~30–50%) and poses significant challenges for effective therapeutic interventions [9]. In-depth molecular and functional profiling of ADC could provide valuable insights into specific molecular subtypes, thereby identifying alternative therapeutic options. Recently, we developed a 10-gene prognostic signature for stage I ADC that can identify a subset of tumors known as C1-ADC [10,11]. This subtype exhibits distinctive gene/protein expression patterns and genetic alterations

that resemble more advanced cancer, and the signature can be accurately measured using qRT-PCR, Affymetrix or RNA-seq, in both fresh-frozen and formalin-fixed, paraffin-embedded (FFPE) specimens [11]. In order to facilitate the clinical implementation of the 10-gene signature, we introduce a miRNA signature that serves as a substitute for the 10 genes to predict the prognostic risk of ADC. A miRNA-based prognostic signature would mitigate the issues of utilizing low-quality mRNA extracted from FFPE samples that are frequently utilized for diagnostic purposes. Shorter non-coding RNA molecules, such as miRNA, are more resilient to adverse conditions and suitable for most of the expression profiling methods, including quantitative real-time polymerase chain reaction (qRT-PCR), due to their stability and compatibility [12,13].

III.3 Results

Identification of miRNA-signature

We devised a multi-tiered strategy as illustrated in Fig. 1 to identify a surrogate miRNA-based prognostic signature for patients with ADC. Initially, we conducted a gene expression analysis of 515 ADC patients from the TCGA-LUAD cohort with mRNA data. Table 1 provides details about patients characteristics. Stage I tumors accounted for 54% of the cohort, and 71% had a smoking history. The median follow-up duration for survivors was 2.1 years. Hierarchical clustering analysis of the TCGA-LUAD cohort (n = 515) patients utilizing the 10-gene signature resulted in four main clusters, namely C1 (n = 201), C2 (n = 98), C3 (n = 39), and C4 (n = 177) (Fig. 2a), which is in accordance with prior studies [11]. Analysis of 3-year overall survival showed no significant differences between C2, C3, and C4 clusters (log-rank test p-value = 0.90 and p-value = 0.48 in stage I and advanced stages, respectively), which were thus combined into non-C1 clusters. C1 cluster had the worst prognosis in both stage I (p-value = 0.0010) and more advanced stages (p-value = 0.0061) (Fig. 2b). Additionally, the C1 cluster had a significantly higher proportion of male patients, more advanced lung cancer, and nearly significant higher percentage of smokers, in agreement with the worse prognosis previously reported [9].

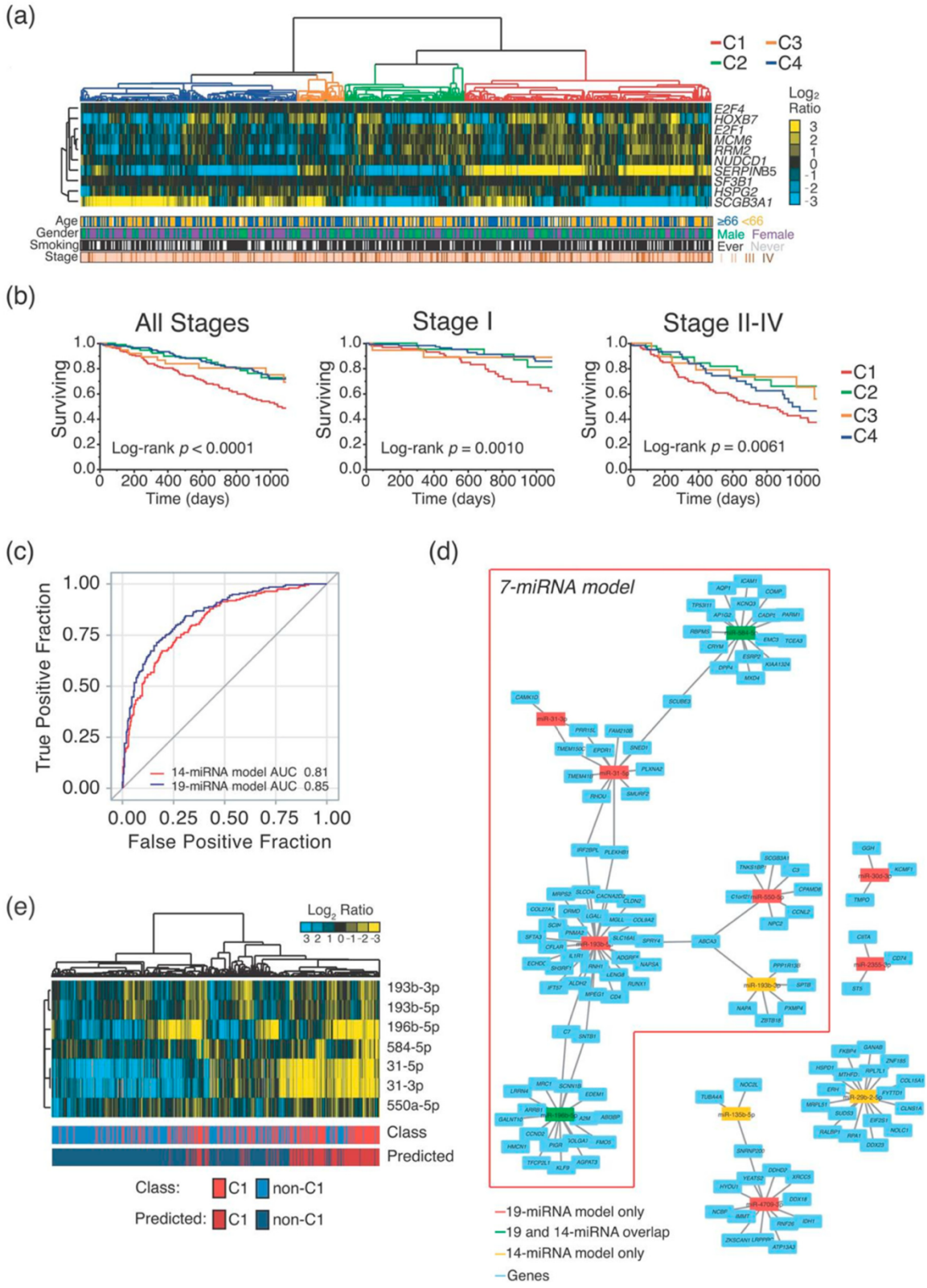


Chapter III.3 Figure 1: Flow chart of study design with data sets and analysis.

Table 1. Patients and tumors characteristics.

	TCGA-LUAD Cohort *n* = 515	CSS Cohort *n* = 44
Age [years]		
Median (Q1; Q3)	66 (59;73) [1]	73 (67;77)
Gender		
Male	238 (46.2%)	27 (61.4%)
Female	277 (53.8%)	17 (38.6%)
Smoking status		
Current/former smoker	367 (71.3%)	20 (45.5%)
Never smoker	63 (12.2%)	11 (25.0%)
Missing smoking status	85 (16.5%)	13 (29.5%)
Stage		
Stage I	279 (54.2%)	31 (70.5%) [2]
Stage II	124 (24.1%)	6 (13.6%)
Stage III	84 (16.3%)	6 (13.6%)
Stage IV	27 (5.2%)	1 (2.3%)
Missing stage	1 (0.2)	-
Follow-up [3]		
Survivors length of follow-up		
<1 year	52 (10.3%)	13 (31.7%)
1–2 years	128 (25.3%)	11 (26.8%)
2–3 years	56 (11.1%)	10 (24.4%)
>3 years	133 (26.3%)	5 (12.2%)
Deaths within 3 years	137 (27.1%)	2 (4.9%) [4]

Percentages could not add up to 100 due to rounding; [1] 19 patients with missing information on age; [2] 1 patient with adenocarcinoma in situ; [3] 9 patients with missing follow-up in the TCGA-LUAD cohort; [4] 3 deaths were excluded: 1 without date of death, and 2 within 30 days from surgery.



Chapter III.3 Figure 2: mRNA and miRNA expression profile analysis of the TCGA-LUAD cohort.

(**a**) Hierarchical clustering analysis of the 10-gene expression signature. C1-C4 clusters are colored as per the legend. Age, gender, smoking status and stage are colored as per the legend; Unavailable information is colored in white. (**b**) Kaplan–Meier curves for 3-years overall survival stratified by C1–C4 clusters. Log-rank *p*-values are shown for C1 vs. non-C1 clusters (C2-C4) comparison. (**c**) Receiver operating characteristic (ROC) curves showing the False Positive Fraction and True Positive Fraction of the 19- (in blue) and 14-miRNA (in red) models. The areas under curve (AUC) are reported. (**d**) Networks of miRNA derived from 19-, 14- and 7-miRNAs model and corresponding target genes. Light blue rectangles represent genes; red rectangles represent miRNA from 19-miRNA model; yellow rectangles represent miRNA from 14-miRNA model; green rectangles represent miRNA from both 14- and 19-miRNA model. (**e**) Hierarchical clustering of 7-miRNAs in the TCGA-LUAD cohort. C1 and non-C1 tumors (defined according to the 10-gene signature) are colored as per the legend. Predicted C1 and non-C1 tumors (defined according the 7-miRNA logistic model) are colored as per the legend.

Next, we proceeded with miRNA expression profiling of 510 ADC patients out of the TCGA-LUAD cohort, whose miRNA data were available. We used both the DESeq2 R package and BRB-ArrayTools as alternative statistical methods to identify differentially expressed miRNAs in C1 and non-C1 ADC clusters. We analyzed a total of 382 miRNAs and found 200 to be differentially expressed by DESeq2 and 90 by BRB-ArrayTools. Among these, 87 miRNAs were found to overlap in both sets. We then applied Lasso regularization to identify optimized miRNA-based signatures that could stratify C1 from non-C1 tumors. We derived two signatures consisting of 14 miRNAs (from the 90 miRNA-set) and 19 miRNAs (from the 200 miRNA-set), respectively, of which 5 miRNAs overlapped (Table 2). These signatures demonstrated high accuracy in stratifying C1/non-C1 cancer patients (cross-validated AUC = 0.81 and AUC = 0.85, respectively; Fig. 2c)

.

Table 2. TCGA-LUAD cohort. Differentially expressed miRNAs composing the three signatures with 19, 14 and 7 miRNAs.

miRNA	Accession	Signature	TCGA-LUAD Cohort—C1 vs. non-C1 Cluster		
			FC	*p*-value [1]	C1 Trend
hsa-miR-193b-5p	MIMAT0004767	19- and 7-miRNA	1.5	3.3×10^{-7}	↑
hsa-miR-31-3p	MIMAT0004504	19- and 7-miRNA	3.2	1.9×10^{-20}	↑
hsa-miR-31-5p	MIMAT0000089	19- and 7-miRNA	3.1	1.7×10^{-18}	↑
hsa-miR-550a-5p	MIMAT0004800	19- and 7-miRNA	1.5	6.0×10^{-9}	↑
hsa-miR-196b-5p	MIMAT0001080	19-, 14-miRNA and 7-miRNA	3.2	9.8×10^{-21}	↑
hsa-miR-584-5p	MIMAT0003249	19-, 14-miRNA and 7-miRNA	2.8	1.2×10^{-40}	↑
hsa-miR-30d-5p	MIMAT0000245	19- and 14-miRNA	0.6	4.8×10^{-16}	↓
hsa-miR-582-3p	MIMAT0004797	19- and 14-miRNA	2.2	2.5×10^{-18}	↑
hsa-miR-9-5p	MIMAT0000441	19 and 14-miRNA	1.8	1.7×10^{-6}	↑
hsa-let-7c-3p	MIMAT0026472	19-miRNA	0.8	1.9×10^{-2}	↓
hsa-miR-138-5p	MIMAT0000430	19-miRNA	1.9	1.2×10^{-10}	↑
hsa-miR-196a-5p	MIMAT0000226	19-miRNA	1.4	2.7×10^{-2}	↑
hsa-miR-203a-3p	MIMAT0000264	19-miRNA	1.4	3.1×10^{-4}	↑
hsa-miR-215-5p	MIMAT0000272	19-miRNA	5.0	1.2×10^{-37}	↑
hsa-miR-2355-3p	MIMAT0017950	19-miRNA	1.3	5.4×10^{-5}	↑
hsa-miR-30d-3p	MIMAT0004551	19-miRNA	0.6	2.5×10^{-15}	↓
hsa-miR-4709-3p	MIMAT0019812	19-miRNA	0.5	1.3×10^{-19}	↓
hsa-miR-548b-3p	MIMAT0003254	19-miRNA	0.6	7.2×10^{-10}	↓
hsa-miR-675-3p	MIMAT0006790	19-miRNA	2.1	1.5×10^{-8}	↑
hsa-miR-193b-3p	MIMAT0002819	14- and 7-miRNA	1.4	8.6×10^{-6}	↑
hsa-miR-135b-5p	MIMAT0000758	14-miRNA	0.7	3.7×10^{-6}	↓
hsa-miR-187-3p	MIMAT0000262	14-miRNA	0.6	2.3×10^{-4}	↓
hsa-miR-192-5p	MIMAT0000222	14-miRNA	3.1	9.8×10^{-21}	↑
hsa-miR-210-3p	MIMAT0000267	14-miRNA	1.2	6.4×10^{-2}	↑
hsa-miR-29b-2-5p	MIMAT0004515	14-miRNA	0.7	1.2×10^{-7}	↓
hsa-miR-3065-3p	MIMAT0015378	14-miRNA	0.7	4.2×10^{-5}	↓
hsa-miR-375-3p	MIMAT0000728	14-miRNA	1.2	1.7×10^{-1}	↑
hsa-miR-708-5p	MIMAT0004926	14-miRNA	1.3	2.7×10^{-3}	↑

[1] Wald test adjusted (Benjamini–Hochberg method) from DESeq2 tool. Accession, miRbase mature miRNA accession number. FC, fold change. C1 trend, expression trend in C1 samples versus non-C1 samples.

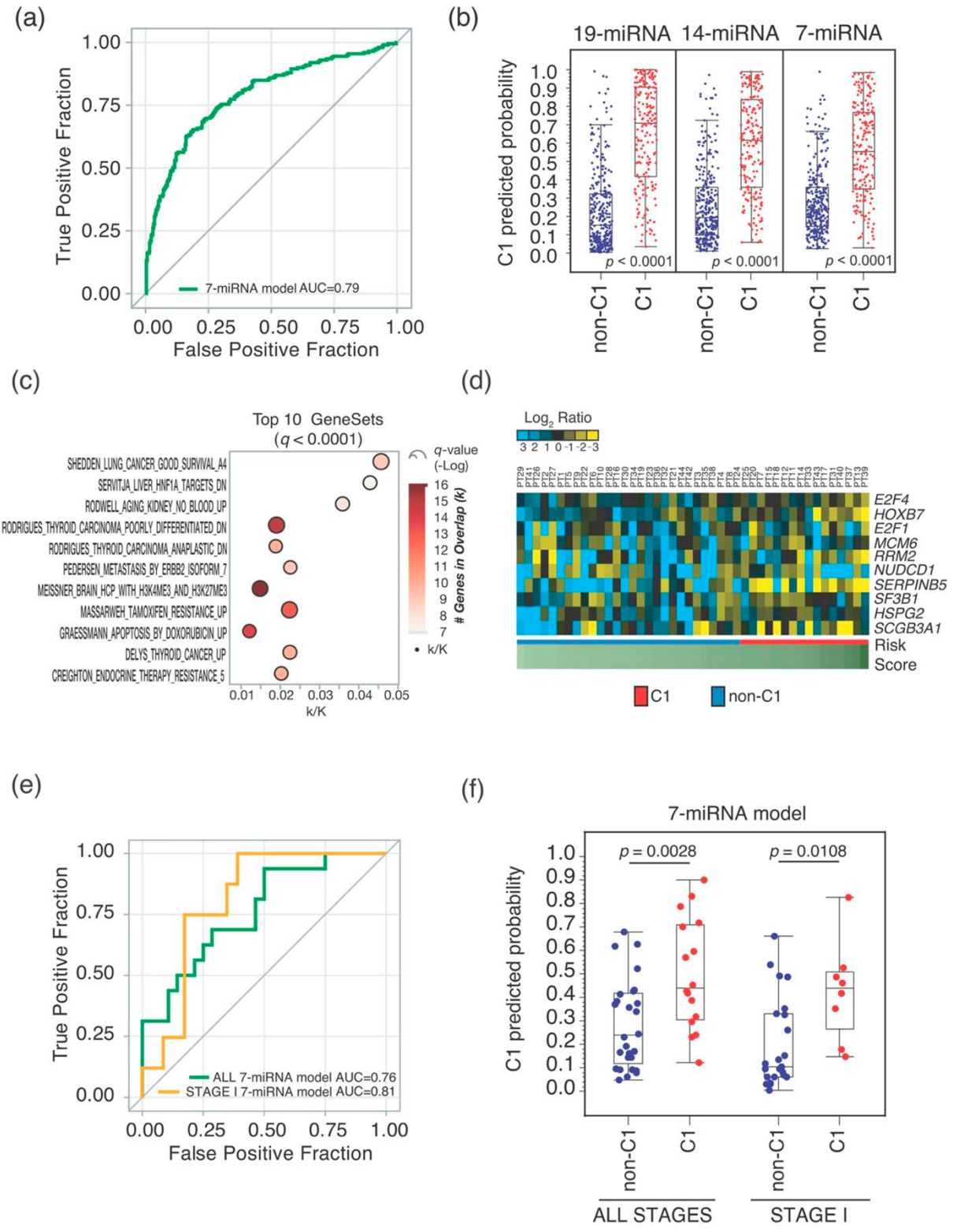
In order to simplify the miRNA-based biomarkers, we aimed to identify a minimal set of miRNAs that could accurately identify C1 aggressive disease, similar to the 14- and 19-miRNA signatures. Our approach was based on two assumptions: (i) the function of a miRNA is dependent on the network of targeted mRNAs, which in this case are those differentially expressed in C1/non-C1 tumors, and (ii) a prognostic biomarker should be

functionally linked to mechanisms involved in tumor progression. Therefore, we investigated the miRNA-mRNA interactome in C1 tumors using ARACNe (a computational algorithm for the reconstruction of accurate cellular networks) using a set of 200 miRNAs and 2900 mRNA genes found to be significantly regulated in C1-ADC (with a p-value < 0.05) by DESeq2. To minimize technical variability, we restricted our analysis only to genes identified by DESeq2, and we applied the following criteria to rewire the C1 miRNA-mRNA interactome: (1) selecting miRNA-mRNA pairs unique to C1 tumors and specific, but not exclusive, to stage I,(n = 2858); (2) selecting miRNA that are predicted to target C1 genes (n = 1787, using miRWalk3.0), and (3) those with an opposite expression trend to C1 genes (n = 598); (4) selecting miRNA that interact with at least three C1 genes (n = 528). We identified a set of interacting networks with seven miRNAs as "HUBs" derived from both the 19-miRNA and 14-miRNA signatures (Fig. 2d) among the miRNA-mRNA networks. Hierarchical clustering analysis of this 7-miRNA signature revealed an overall increased expression in more aggressive C1 tumors (Fig. 2e). Significantly, the 7-miRNA signature exhibited a cross-validated area under the curve (AUC) of 0.79 in stratifying C1 and non-C1 patients, which is comparable to the other two signatures (Fig. 3a), even when differences in C1 predicted probability were considered (Fig. 3b). The predicted C1 class from all three signatures (7-, 14-, and 19-miRNA) presented a significantly increased hazard of death at 3 years in patients of all stages, with an increased risk equivalent to that of C1 patients identified by using the 10-gene signature (Table 3). However, focusing the analysis on stage I ADC patients revealed that the 7-miRNA signature was the best risk-stratification method, with a two-fold increased risk of death for C1 patients (HR = 2.11; 95% confidence interval: 1.11-4.00; p-value = 0.0223) (Table 3). Interestingly, the networks of genes targeted by these 7 miRNAs were significantly enriched (q-value < 0.0001) in gene sets representing molecular mechanisms related to cancer progression, thus confirming our initial hypotheses (Fig. 3c)

Table 3. TCGA-LUAD cohort. Univariate and multivariable Cox regression analyses for 3-years overall survival in patients of all stages and stratified by stage.

		Univariate Analysis		Multivariable Analysis [1]	
	n (*n* Deaths)	HR (95% CI)	Wald Test *p*-value	HR (95% CI)	Wald Test *p*-value
ALL STAGES	**501 (135)** [2]				
10-gene	194 (75)	2.21 (1.57–3.10)	<0.0001	2.03 (1.43–2.87)	<0.0001
19-miRNA	169 (66)	2.13 (1.52–2.99)	<0.0001	1.85 (1.31–2.61)	0.0005
14-miRNA	165 (67)	2.17 (1.55–3.04)	<0.0001	2.06 (1.46–2.91)	<0.0001
7-miRNA	146 (67)	2.90 (2.07–4.06)	<0.0001	2.69 (1.91–3.78)	<0.0001
STAGE I	**274 (40)**				
10-gene	92 (23)	2.86 (1.53–5.36)	0.0010	2.96 (1.55–5.65)	0.0010
19-miRNA	73 (11)	1.07 (0.54–2.15)	0.8462	1.12 (0.55–2.26)	0.7529
14-miRNA	79 (17)	1.90 (1.01–3.56)	0.0451	1.99 (1.05–3.79)	0.0359
7-miRNA	65 (15)	2.11 (1.11–4.00)	0.0223	2.14 (1.11–4.12)	0.0235
STAGE II-IV	**226 (95)**				
10-gene	101 (52)	1.69 (1.13–2.54)	0.0108	1.64 (1.08–2.49)	0.0207
19-miRNA	95 (55)	2.27 (1.51–3.41)	<0.0001	2.18 (1.43–3.31)	0.0003
14-miRNA	86 (50)	2.00 (1.34–3.00)	0.0007	2.04 (1.35–3.08)	0.0007
7-miRNA	80 (52)	2.89 (1.93–4.33)	<0.0001	2.91 (1.93–4.39)	<0.0001

[1] all stages analyses were adjusted for age, sex, smoking status and stage; analyses stratified by stage were adjusted for age, sex and smoking status; [2] 1 patient with missing stage and 9 patients with missing follow-up.



Chapter III.3 Figure 3: Validation of the 7-miRNA model.

(**a**) ROC curve showing the False Positive Fraction and True Positive Fraction of the 7-miRNA model. The AUC is reported. (**b**) Box-plot for C1 predicted probability in C1 and non-C1 patients. Predicted probabilities are calculated through the 19-, 14- and 7-miRNA models. Wilcoxon–Mann–Whitney test *p*-values are reported. (**c**) Bubble plot of top 10 GeneSets found significantly overlapping with gene networks targeted by the 7-miRNA signature. Bubbles size is proportional to statistical significance (-Log of *q*-value) and color codes refer to number of genes found in the overlap. In *X*-axis, ratios (k/K) of overlap of the query set of genes (k) with overlapping GeneSet size (K). (**d**) Heatmap of the 10-gene expression of CSS cohort. C1 and non-C1 tumors are colored as per the legend. Risk scores are calculated based on the 10-gene risk model. (**e**) ROC curves showing the False Positive Fraction and True Positive Fraction of the 7-miRNA model in the CSS cohort, for all stages (in green) or only stage I tumors (in orange). The AUC are reported. (**f**) Box-plot for C1 predicted probability in C1 and non-C1 tumors in CSS cohort, for all-stages tumors and stage I tumors. Predicted probabilities are calculated through the 7-miRNA model. Wilcoxon–Mann–Whitney test *p*-values are reported.

Seven-miRNA-Signature Validation

To validate our findings, we evaluated the 7 miRNA-signature in a separate group of 44 patients with lung adenocarcinoma at the IRCCS Casa Sollievo della Sofferenza Hospital. The clinical pathological characteristics of this cohort are described in Table 1, and it was found to have a higher proportion of stage I tumors (70%) than the TCGA-LUAD cohort (54%). We analyzed FFPE samples using qRT-PCR and calculated relative risk-score to stratifiy the cohort into C1 (n = 16) and non-C1 (n = 28) groups. We then used Low-Density Taqman miRNA Arrays to profile the 7-miRNA signature in the same cohort and derived a model based on its expression profile using logistic regression. The 7-miRNA model accurately stratified C1 from non-C1 tumors with an AUC of 0.76 (Fig. 3e) and significant differences (p-value = 0.0028) in C1 predicted probability (Fig. 3f). Remarkably, when we focused on stage I tumors, we obtained even better results, with an AUC of 0.81 (Fig. 3e) and a significant difference (p-value = 0.0108) in C1 predicted probability (Fig. 3f).

III.4 Discussion

Recent advancements in early diagnosis of lung cancer through large-scale low-dose CT screening trials have resulted in a shift towards diagnosing patients at earlier stages, leading to a decrease in mortality as demonstrated by the NELSON and NLST trials [7,8]. Consequently, there is a pressing need for prognostic biomarkers that can identify stage I lung cancer patients who could benefit from systemic adjuvant chemotherapy (platinum-based) instead of molecular targeted/immunotherapeutic approaches in cases of aggressive disease. With the development of more sophisticated and precise bioinformatic tools and computational approaches, novel diagnostic and prognostic cancer biomarkers can be identified quickly and accurately, focusing on proteins, coding and non-coding transcripts, and epigenetic modifications [14,15]. In the field of lung cancer, the differential expression of miRNAs has shown promise for the development of innovative cancer biomarkers [16]. In this study, we present miRNA signatures that can serve as surrogates for a previously reported 10-gene prognostic signature in stage I ADC [11]. Our analysis revealed that the 7-, 14-, and 19-miRNA signatures were all effective in identifying aggressive C1-ADC disease (AUC = 0.79–0.85). It is worth noting that 6 out of 7 miRNAs of the 7-miRNA signature were well-detected in FFPE samples (median Ct < 30), which confirmed the higher stability of miRNAs in low-quality RNA [17]. We adopted a network-rewiring strategy in our approach by specifically selecting miRNA-mRNA pairs that characterize aggressive stage I tumors (C1). This allowed us to select a core of 7 miRNAs that can stratify C1 from non-C1 samples with an accuracy comparable to the 14- and 19-miRNA models (Fig. 2c and Fig. 3a). Importantly, these miRNAs interacted with the C1 tumor transcriptome, indicating their potential involvement in molecular mechanisms associated with tumor progression. Our analysis revealed a significant overlap between the "7-miRNA network" and several gene

sets representing cancer-relevant pathways (Fig. 3c). Further experiments are needed to investigate the functional relevance of this 7-miRNA network in lung cancer progression.
The 7-miRNA signature is composed by hsa-miR-31-5p, hsa-miR-31-3p, hsa-miR-193b-3p, hsa-miR-193b-5p, hsa-miR-196b-5p, hsa-miR-550a-5p and hsa-miR-584-5p. Some of these miRNAs were described to be altered in cancer and also with a functional role. Alterations in the expression of miR-31-5p and miR-31-3p were associated with a variety of cancers including lung cancer and have both oncogenic and onco-suppressor behavior [18-24]. Furthermore, hsa-miR-193b-3p was found associated to a tumor-suppressor phenotype (by targeting STMN1) in hepatocellular carcinoma [25] and colorectal cancer [26]. Remarkably, hsa-miR-193 in prostate cancer was shown to target FOXM1 and RRM2 [27], the last being one of the genes composing the prognostic 10-gene signature. The hsa-miR-196b-5p was found to promote tumor progression in non-small cell lung cancer when up-regulated [28]. Interestingly, hsa-miR-196b-5p was also found to molecularly interact with HOXB7 and GALNT5 and inhibit their expression in colorectal cancer [29]. Again, HOXB7 is one of the genes composing the 10-gene signature, which we also recently showed to promote a stem cell-like phenotype in lung adenocarcinoma [30]. On the other hand, very little is known about hsa-miR-550a-5p and hsa-miR-584-5p biological functions. However, hsa-miR-550a-5p was found to be a prognostic factor in ADC in association with other 3 miRNAs indicating its possible oncogenic role [31]. Lastly, in gastric cancer, hsa-miR-584-5p was found to induce apoptosis and inhibits proliferation [32], while in hepatocellular carcinoma hsa-miR-584-5p was shown to have an oncogenic role [33].
In conclusion, our study established a 7-miRNA signature that effectively identifies aggressive stage I lung adenocarcinoma. These miRNAs can be quantified using standard qRT-PCR, and FFPE samples. Nonetheless, a limitation of our research is the modest size of the external validation cohort (CSS) and the brief follow-up period (1.6 median years in survivors, and three deaths recorded within 3 years), which precludes us from accurately determining the extent of excess mortality risk in patients with predicted aggressive tumors.

III.5 Material and Methods

TCGA-LUAD Cohort

We selected the cohort of 515 patients with lung adenocarcinomas from the TCGA data portal (https://portal.gdc.cancer.gov/) at 2018. A total of 510 tumors were profiled for both gene and miRNA expression. Log2 read counts were used for expression analysis. Patients follow-up information was used for survival analysis: overall survival was defined as the time from the date of tumor resection until death from any cause. Follow-up was truncated at 3 years to reduce the potential overestimation of overall mortality with respect to lung cancer–specific mortality.

The CSS Cohort

We selected a cohort of 44 patients with lung adenocarcinoma underwent surgery between February 2017 and February 2020 at the CSS. Written informed consent was obtained from all study patients. None of these patients received preoperative chemotherapy. Clinical information was obtained through review of medical records. Vital status was assessed through the Vital Records Offices of the patients' towns of residence or by contacting directly the patients or their families.

Gene Expression Analysis of the TCGA-LUAD Cohort

Hierarchical clustering analysis was performed on the 10-gene signature for the entire cohort of 510 patients. Clustering was done by using Cluster 3.0 for Mac OS X (C Clustering Library 1.56, Tokyo, Japan) with uncentered correlation and centroid linkage, and Java TreeView software environment (version 1.1.6r4; http://jtreeview.sourceforge.net). A total of four main branches were selected to build clusters. Kaplan–Meier survival curves were stratified by clusters and log-rank test p-values were calculated. C1 cluster was associated to the worse prognosis, and all other clusters were pooled together (non-C1 clusters).

To retain most informative expression data (i.e., transcripts detected in most of tumor samples), we considered miRNAs with raw counts >0 in at least the 50% of patients either in C1 or non-C1, identifying a total of 382 miRNAs. This allowed us to reduce also the complexity of the TCGA-LUAD dataset (2237 miRNAs). We applied the complexity reduction also to genes and we selected the most varying across all samples (standard deviation in the top 25%), identifying a total of 4899 genes. Using DESeq2 R package (R Core Team, R Foundation for Statistical Computing, Vienna, Austria) [34], we identified a total of 2900 differentially expressed genes between C1 and non-C1 tumors.

BRB-ArrayTools [35] and DESeq2 (R package) [34] tools were used for class prediction (C1 cluster vs. non-C1 clusters) according to miRNA expression. BRB-ArrayTools uses statistics based on two-sample T-test with multivariate permutations test (1000 random permutations); confidence level of false discovery rate assessment, 80%; maximum allowed proportion of false-positive genes, 0.05. DESeq2 is based on Wald test statistics to identify differentially expressed transcripts. Lists of miRNAs differentially expressed obtained from BRB-ArrayTools and DESeq2 tools were subsequently reduced via Lasso regularization. In details, a penalized unconditional logistic regression was applied considering cluster as discrete outcome (C1 cluster vs. non-C1 clusters) and miRNA expressions as explanatory variables. Cross-validated (10-fold) log-likelihood with optimization (50 simulations) of the tuning penalty parameter was used to control for potential overfitting.

Starting from differentially expressed genes (identified with DESeq2) and miRNAs (identified with both DESeq2 and BRB-ArrayTools), we used ARACNe [36] with 1000 bootstraps to infer direct regulatory relationships between transcriptional regulators (i.e., miRNAs) and target genes. ARACNe was performed using all patients, stage I patients and stage II-IV patients. miRNA target genes were retrieved using miRWalk 3.0 [37].

Probability of being in the C1 cluster was estimated using the unconditional logistic regression for the three signatures of 19, 14 and 7 miRNAs. Model performance was assessed using the cross-validated area under the receiver operating curve, and assessing the difference in C1 predicted probability between C1 and non-C1 patients (Wilcoxon–Mann–Whitney test). Cox regression model was used to evaluate the prognostic role of these miRNA signatures and their ability to recapitulate the risk-stratification of the original 10-genes signature.

To receive insights into the biology of the 7-miRNA model, we verified the enrichment of cancer-relevant pathways associated to their target genes. We investigated the Molecular Signature Database (MSigDB; v7.2) (https://www.gsea-msigdb.org/gsea/msigdb/annotate.jsp) using the list of 87 targeted genes by interrogating the CGP (chemical and genetic perturbations, 3358 gene sets). Bubble plot analysis was performed using JMP 15.2.1 (SAS Institute, Inc., Cary, NC, USA) software.

Hierarchical clustering analysis was performed on the 7-miRNA signature for 510 patients, those with available miRNA expression data. Clustering was completed by using Cluster 3.0 for Mac OS X (C Clustering Library 1.56) with uncentered correlation and centroid linkage, and Java TreeView software environment (version 1.1.6r4; http://jtreeview.sourceforge.net).

RNA Extraction and qRT-PCR Analysis and Data Interpretation

One tissue core (1.5 mm in diameter) from FFPE blocks, in representative tumor areas with adequate tumor cellularity (>60%) selected by a pathologist, was processed for total RNA extraction. The AllPrep DNA/RNA FFPE kit (QIAGEN, Inc., Hilden, Germany) was used for total RNA extraction. Quantitative real-time PCR (qRT-PCR) was performed to analyze the 10-genes signature as described in Dama et al. [11]. Briefly, RNA was quantified using Nanodrop ND-10000 Spectrophotometer and a total of 200 ng was retro-transcribed using SuperScript VILO cDNA Synthesis Kit (Thermo Fisher Scientific, Inc., Waltham, MA, USA) (Thermo Fisher Scientific) and pre-amplified for 10 cycles with PreAmp Master Mix Kit (Thermo Fisher Scientific), following manufacturer's instructions. qRT-PCR analysis was performed starting from 1:10 diluted pre-amplified cDNA, using the TaqMan Fast Advance Master Mix and hydrolysis probes (Thermo Fisher Scientific; for primers see Dama et al. [11]), in a QuantStudio 12k Flex (Thermo Fisher Scientific). Thermal cycling amplification was performed with an initial incubation at 95 °C for 30 s, followed by 45 cycles of 95 °C for 5 s and 60 °C for 30 s. For miRNA expression analysis, a total of 10 ng RNA was reverse-transcribed using the TaqMan Advanced miRNA cDNA Synthesis Kit (Thermo Fisher Scientific). Poly(A) tailing, adapter ligation, RT reaction and miR-Amp were performed following manufacturer's instructions. qRT-PCR was performed following manufacturer's instructions (i.e., 95 °C for 30 s, 45 cycles of 95 °C for 5 s, and 60 °C for 30 s) using a Card Custom Advance (Thermo Fisher Scientific; see chapter IV Table S2 for the list of primers)

in a QuantStudio 12k Flex (Thermo Fisher Scientific). The hsa-miR-16-5p was used as standard reference for CT normalization using a previously described methodology [38]. Briefly, the normalized CT of each miRNA (i) of each sample (j) was calculated as the difference between the raw CTij and a scaling factor (SF) specific for each sample (j); the SFj represented the difference between the raw CT of the miRNA "hsa-miR-16-5p" used as a reference in the sample (j) and a constant equal to 21.87. Notably, hsa-miR-16-5p expression profile analysis in both TCGA-LUAD and CSS cohorts revealed a comparable expression in C1 and non-C1 tumors subsets.

Risk-scores were assigned to each patient based to the 10-gene risk model described in Dama et al. [11]. Before applying the risk-model, data were rescaled (q1-q3 normalization). Patients with risk-scores higher than the 66th percentiles [11] were classified as C1 tumors. Next, unconditional logistic regression (C1 vs. non-C1 tumors) with 7 miRNAs as explanatory variables was applied, and the area under the receiver operating curve was calculated. Difference in C1 predicted probability between C1 and non-C1 patients was evaluated through Wilcoxon–Mann–Whitney test.

All statistical analyses were performed using SAS software, version 9.4 (SAS Institute, Inc., Cary, NC, USA) and R 3.3.1 (R Core Team, 2016) and JMP 15 (SAS). p-values less than 0.05 were considered statistically significant.

III.6 References

[1] Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424.

[2] Travis, W.D.; Colby, T.V.; Corrin, B.; Shimosato, Y.; Brambilla, E. *Histological Typing of Lung and Pleural Tumours*, 3rd ed.; WHO. World Health Organization. International Histological Classification of Tumours; Springer: Berlin/Heidelberg, Germany, 1999; ISBN 978-3-540-65219-9.

[3] Jones, G.S.; Baldwin, D.R. Recent advances in the management of lung cancer. *Clin. Med.* **2018**, *18*, s41–s46.

[4] Falzone, L.; Salomone, S.; Libra, M. Evolution of Cancer Pharmacological Treatments at the Turn of the Third Millennium. *Front. Pharmacol.* **2018**, *9*, 1300.

[5] McCoach, C.E. A Cautionary Analysis of Immunotherapy Prior to Targeted Therapy. *J. Thorac. Oncol.* **2019**, *14*, 8–10.

[6] Barta, J.A.; Powell, C.A.; Wisnivesky, J.P. Global Epidemiology of Lung Cancer. *Ann. Glob. Health* **2019**, *85*, 8.

[7] The National Lung Screening Trial Research Team. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *N. Engl. J. Med.* **2011**, *365*, 395–409.

[8] De Koning, H.J.; van der Aalst, C.M.; de Jong, P.A.; Scholten, E.T.; Nackaerts, K.; Heuvelmans, M.A.; Lammers, J.-W.J.; Weenink, C.; Yousaf-Khan, U.; Horeweg, N.; et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. *N. Engl. J. Med.* **2020**, *382*, 503–513.

[9] Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2020. *CA Cancer J. Clin.* **2020**, *70*, 7–30.

[10] Bianchi, F.; Nuciforo, P.; Vecchi, M.; Bernard, L.; Tizzoni, L.; Marchetti, A.; Buttitta, F.; Felicioni, L.; Nicassio, F.; Fiore, P.P.D. Survival prediction of stage I lung adenocarcinomas by expression of 10 genes. *J. Clin. Investig.* **2007**, *117*, 3436–3444.

[11] Dama, E.; Melocchi, V.; Dezi, F.; Pirroni, S.; Carletti, R.M.; Brambilla, D.; Bertalot, G.; Casiraghi, M.; Maisonneuve, P.; Barberis, M.; et al. An Aggressive Subtype of Stage I Lung Adenocarcinoma with Molecular and Prognostic Characteristics Typical of Advanced Lung Cancers. *Clin. Cancer Res.* **2017**, *23*, 62–72.

[12] Li, J.; Smyth, P.; Flavin, R.; Cahill, S.; Denning, K.; Aherne, S.; Guenther, S.M.; O'Leary, J.J.; Sheils, O. Comparison of miRNA expression patterns using total RNA extracted from matched samples of formalin-fixed paraffin-embedded (FFPE) cells and snap frozen cells. *BMC Biotechnol.* **2007**, *7*, 36.

[13] Hall, J.S.; Taylor, J.; Valentine, H.R.; Irlam, J.J.; Eustace, A.; Hoskin, P.J.; Miller, C.J.; West, C.M.L. Enhanced stability of microRNA expression facilitates classification of FFPE tumour samples exhibiting near total mRNA degradation. *Br. J. Cancer* **2012**, *107*, 684–694.

[14] Dijkstra, K.K.; Voabil, P.; Schumacher, T.N.; Voest, E.E. Genomics- and Transcriptomics-Based Patient Selection for Cancer Treatment with Immune Checkpoint Inhibitors: A Review. *JAMA Oncol.* **2016**, *2*, 1490–1495.

[15] Beane, J.; Campbell, J.D.; Lel, J.; Vick, J.; Spira, A. Genomic approaches to accelerate cancer interception. *Lancet Oncol.* **2017**, *18*, e494–e502.

[16] Dama, E.; Melocchi, V.; Colangelo, T.; Cuttano, R.; Bianchi, F. Deciphering the Molecular Profile of Lung Cancer: New Strategies for the Early Detection and Prognostic Stratification. *J. Clin. Med.* **2019**, *8*, 108.

[17] Jung, M.; Schaefer, A.; Steiner, I.; Kempkensteffen, C.; Stephan, C.; Erbersdobler, A.; Jung, K. Robust MicroRNA Stability in Degraded RNA Preparations from Human Tissue and Cell Samples. *Clin. Chem.* **2010**, *56*, 998–1006.

[18] Creighton, C.J.; Fountain, M.D.; Yu, Z.; Nagaraja, A.K.; Zhu, H.; Khan, M.; Olokpa, E.; Zariff, A.; Gunaratne, P.H.; Matzuk, M.M.; et al. Molecular Profiling Uncovers a p53-Associated Role for MicroRNA-31 in Inhibiting the Proliferation of Serous Ovarian Carcinomas and Other Cancers. *Cancer Res.* **2010**, *70*, 1906–1915.

[19] Manceau, G.; Imbeaud, S.; Thiébaut, R.; Liébaert, F.; Fontaine, K.; Rousseau, F.; Génin, B.; Corre, D.L.; Didelot, A.; Vincent, M.; et al. Hsa-miR-31-3p Expression Is Linked to Progression-free Survival in Patients with KRAS Wild-type Metastatic Colorectal Cancer Treated with Anti-EGFR Therapy. *Clin. Cancer Res.* **2014**, *20*, 3338–3347.

[20] Sun, D.; Yu, F.; Ma, Y.; Zhao, R.; Chen, X.; Zhu, J.; Zhang, C.-Y.; Chen, J.; Zhang, J. MicroRNA-31 Activates the RAS Pathway and Functions as an Oncogenic MicroRNA in Human Colorectal Cancer by Repressing RAS p21 GTPase Activating Protein 1 (RASA1). *J. Biol. Chem.* **2013**, *288*, 9508–9518.

[21] Angius, A.; Pira, G.; Scanu, A.M.; Uva, P.; Sotgiu, G.; Saderi, L.; Manca, A.; Serra, C.; Uleri, E.; Piu, C.; et al. MicroRNA-425-5p Expression Affects BRAF/RAS/MAPK Pathways In Colorectal Cancers. *Int. J. Med. Sci.* **2019**, *16*, 1480–1491.

[22] Liu, X.; Sempere, L.F.; Ouyang, H.; Memoli, V.A.; Andrew, A.S.; Luo, Y.; Demidenko, E.; Korc, M.; Shi, W.; Preis, M.; et al. MicroRNA-31 functions as an oncogenic microRNA in mouse and human lung cancer cells by repressing specific tumor suppressors. *J. Clin. Investig.* **2010**, *120*, 1298–1309.

[23] Edmonds, M.D.; Boyd, K.L.; Moyo, T.; Mitra, R.; Duszynski, R.; Arrate, M.P.; Chen, X.; Zhao, Z.; Blackwell, T.S.; Andl, T.; et al. MicroRNA-31 initiates lung tumorigenesis and promotes mutant KRAS-driven lung cancer. *J. Clin. Investig.* **2016**, *126*, 349–364.

[24] Meng, W.; Ye, Z.; Cui, R.; Perry, J.; Dedousi-Huebner, V.; Huebner, A.; Wang, Y.; Li, B.; Volinia, S.; Nakanishi, H.; et al. MicroRNA-31 Predicts the Presence of Lymph Node Metastases and Survival in Patients with Lung Adenocarcinoma. *Clin. Cancer Res.* **2013**, *19*, 5423–5433.

[25] Xu, C.; Liu, S.; Fu, H.; Li, S.; Tie, Y.; Zhu, J.; Xing, R.; Jin, Y.; Sun, Z.; Zheng, X. MicroRNA-193b regulates proliferation, migration and invasion in human hepatocellular carcinoma cells. *Eur. J. Cancer* **2010**, *46*, 2828–2836.

[26] Guo, F.; Luo, Y.; Mu, Y.-F.; Qin, S.-L.; Qi, Y.; Qiu, Y.-E.; Zhong, M. miR-193b directly targets STMN1 and inhibits the malignant phenotype in colorectal cancer. *Am. J. Cancer Res.* **2016**, *6*, 2463–2475.

[27] Mazzu, Y.Z.; Yoshikawa, Y.; Nandakumar, S.; Chakraborty, G.; Armenia, J.; Jehane, L.E.; Lee, G.-S.M.; Kantoff, P.W. Methylation-associated miR-193b silencing activates master drivers of aggressive prostate cancer. *Mol. Oncol.* **2019**, *13*, 1944–1958.

[28] Liang, G.; Meng, W.; Huang, X.; Zhu, W.; Yin, C.; Wang, C.; Fassan, M.; Yu, Y.; Kudo, M.; Xiao, S.; et al. miR-196b-5p–mediated downregulation of TSPAN12 and GATA6 promotes tumor progression in non-small cell lung cancer. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 4347–4357.

[29] Stiegelbauer, V.; Vychytilova-Faltejskova, P.; Karbiener, M.; Pehserl, A.-M.; Reicher, A.; Resel, M.; Heitzer, E.; Ivan, C.; Bullock, M.; Ling, H.; et al. miR-196b-5p Regulates Colorectal Cancer Cell Migration and Metastases through Interaction with HOXB7 and GALNT5. *Clin. Cancer Res.* **2017**, *23*, 5255–5266.

[30] Monterisi, S.; Lo Riso, P.; Russo, K.; Bertalot, G.; Vecchi, M.; Testa, G.; Di Fiore, P.P.; Bianchi, F. HOXB7 overexpression in lung cancer is a hallmark of acquired stem-like phenotype. *Oncogene* **2018**, *37*, 3575–3588.

[31] Lin, Y.; Lv, Y.; Liang, R.; Yuan, C.; Zhang, J.; He, D.; Zheng, X.; Zhang, J. Four-miRNA signature as a prognostic tool for lung adenocarcinoma. *Onco Targets Ther.* **2018**, *11*, 29–36.

[32] Li, Q.; Li, Z.; Wei, S.; Wang, W.; Chen, Z.; Zhang, L.; Chen, L.; Li, B.; Sun, G.; Xu, J.; et al. Overexpression of miR-584-5p inhibits proliferation and induces apoptosis by targeting WW domain-containing E3 ubiquitin protein ligase 1 in gastric cancer. *J. Exp. Clin. Cancer Res.* **2017**, *36*, 59.

[33] Wei, H.; Wang, J.; Xu, Z.; Lu, Y.; Wu, X.; Zhuo, C.; Tan, C.; Tang, Q.; Pu, J. miR-584-5p regulates hepatocellular carcinoma cell migration and invasion through targeting KCNE2. *Mol. Genet. Genom. Med.* **2019**, *7*, e702.

[34] Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550.

[35] Simon, R.; Lam, A.; Li, M.-C.; Ngan, M.; Menenzes, S.; Zhao, Y. Analysis of Gene Expression Data Using BRB-Array Tools. *Cancer Inform.* **2007**, *3*, 11–17.

[36] Lachmann, A.; Giorgi, F.M.; Lopez, G.; Califano, A. ARACNe-AP: Gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics* **2016**, *32*, 2233–2235.
[37] Sticht, C.; Torre, C.D.L.; Parveen, A.; Gretz, N. miRWalk: An online resource for prediction of microRNA binding sites. *PLoS ONE* **2018**, *13*, e0206239.
[38] Montani, F.; Marzi, M.J.; Dezi, F.; Dama, E.; Carletti, R.M.; Bonizzi, G.; Bertolotti, R.; Bellomi, M.; Rampinelli, C.; Maisonneuve, P.; et al. miR-Test: A Blood Test for Lung Cancer Early Detection. *JNCI J. Natl. Cancer Inst.* **2015**, *107*, djv063.

III.7 Supplemental Information

The following are available online at https://www.mdpi.com/2311-553X/6/4/48/s1.

Chapter IV:
Dissecting exosomal-miRNAs as key-players for the early identification of an aggressive subtype of lung adenocarcinoma

This Chapter answer to the Aim III

IV.1 Summary

Recently, tumor cells were found to release exosomes that are nanosized vesicles of 40-150nm in diameter, which can contribute to EMT and metastasis by delivering their molecular cargo to other targeted tumor cells. Exosomes cargo can also contain microRNAs (i.e., exo-miR) which mediate cancer-relevant phenotypes such as tumor invasiveness, EMT, immuno-escaping, stromal activation and pre-metastatic niche formation.

Here, we demonstrated that a 6-exo-miRNA signature derived from LUAD patients is a reliable biomarker to identify the C1-LUAD subtype. Moreover, we investigated the biological activity of exosomes, showing that C1-like cells were more prone to internalize both self and exogenous exosomes. Finally, we showed that miR-223-3p was significantly more expressed in C1-like exosomes and acting as a modulator of immune response. Overall, this study provides insights into the role of exo-miRNAs in lung cancer progression and identifies a potential biomarker for the early identification of aggressive subtypes of lung cancer.

IV.2 Introduction

Lung cancer is the main cause of cancer related deaths in industrialized world with approximately 1.8 million death every year. Non-Small Cell Lung cancer (NSCLC) is the most frequently diagnosed lung cancer type (80-90% of all cases) and lung adenocarcinoma (LUAD) is the major subtype.[1] Early diagnosis by low-dose computed tomography (LDCT) was recently shown to be effective in reducing lung cancer mortality (~20% at 5 years)[2]. However, the management of lung cancer patients can be difficult due to pathological and genetic heterogeneity of the disease. Indeed, a significant fraction of patients underwent surgery (~30%) still experience local and distant recurrence, which ultimately makes relatively unpredictable the prognosis of patients.[1] The presence of subpopulation of cancer cells in the primary lesion with stem-like features contribute to metastatic spreading of the disease ultimately leading to local and distant recurrence.[3] Stem-like cancer cells expand under the pressure of an aberrant activation of the epithelial-to-mesenchymal-transition (EMT) pathway. Indeed, epithelial cancer cells can be reprogrammed in cells with mesenchymal traits under the action of a group of transcription factors (e.g., SLUG, SNAIL, TWIST, ZEB1 etc.) frequently overexpressed in metastatic cancer.[3] In line with this events, we previously discovered a molecular subtype of lung adenocarcinoma by a 10-genes signature, namely C1-LUAD, which correlates with worst prognosis and have peculiar gene expression characteristic of mesenchymal and stem-like cells.[4] Yet, we recently identified also a signature of 7 microRNAs as a surrogates of the 10-genes to distinguish C1-LUAD subtype and which targets a set of genes differentially expressed in C1-LUAD thus suggesting a relevant role of these miRNAs as hubs of mechanisms for tumor progression [5]. Importantly, miRNAs can be secreted in body fluids (e.g., serum, plasma) via exosomes i.e. extracellular vesicles of 30-150nm in diameter.[6],[7] Exosomes were shown to have crucial role in tumor progression, immuno-escaping, stromal activation and pre-metastatic niche formation, although mechanisms of action were only partially unveiled.[8], [9] Tumor exosomes has also drawn increasing attention in the last years because are stable and accessible biomarkers for diagnosis and prognosis of cancer.[6] Actually, there is a lack of non-invasive biomarkers for the identification of aggressive molecular subtype like the C1-LUAD. For such reasons, here we propose a comprehensive exo-miRNA profile in plasma samples of lung adenocarcinoma patients with the C1-LUAD subtype, to develop biomarkers for the early-identification of this aggressive subtype as well as to investigate the role of exo-miRNAs as key-players in the acquirement of the aggressive features of C1-LUAD.

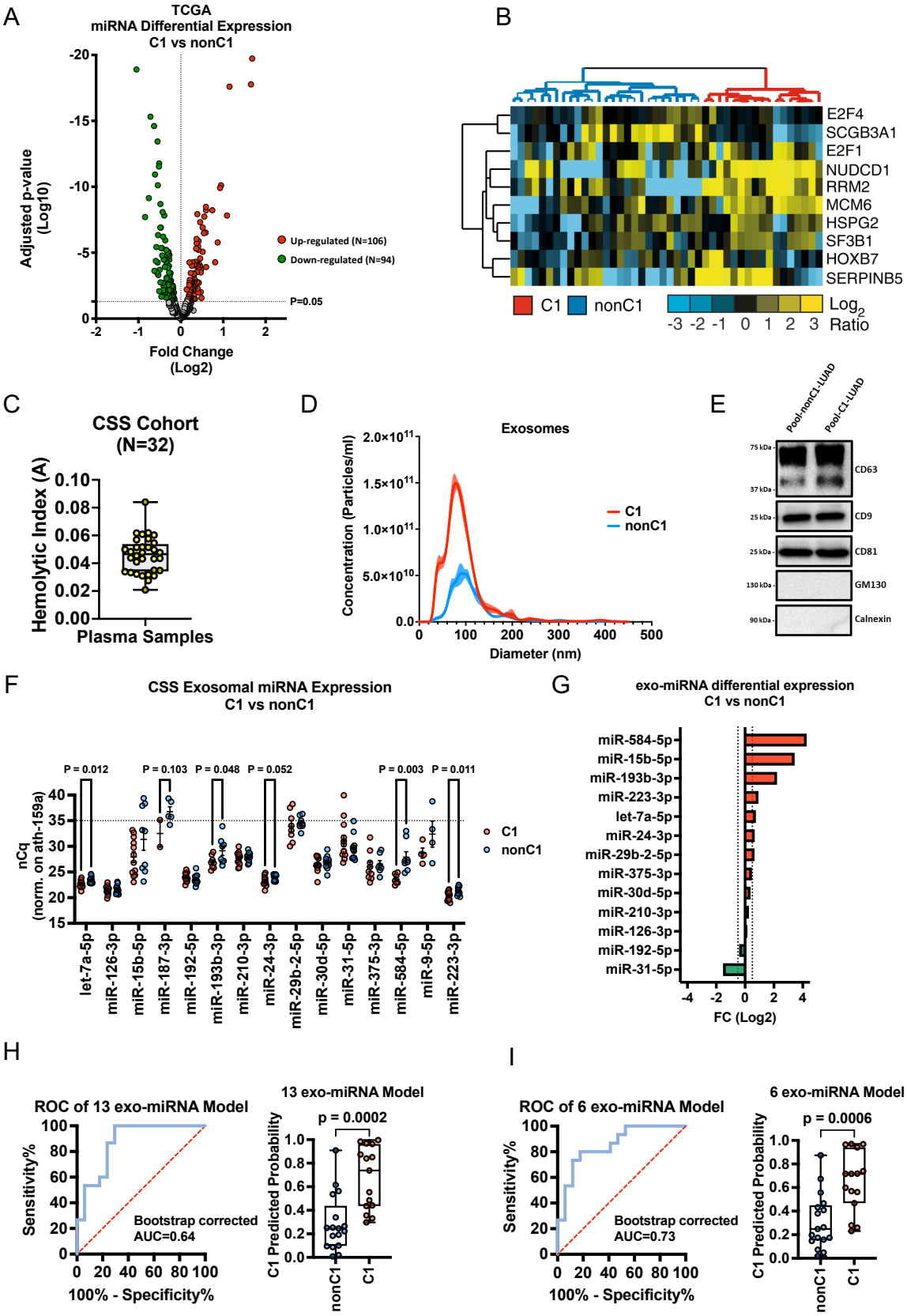
IV.3 Results

C1-patients have a peculiar plasma exosomal miRNA profile able identify the aggressive molecular subtype

We first analyzed the miRNA expression profile of C1- and nonC1-LUAD using the TCGA cohort (see chapter III)[5]. We found 200 miRNAs resulted significantly differentially expressed (Fig. 1 A) which targeted network of genes related to aggressiveness (see also Chapter III)[4], [5]. Of these, we selected a total 23 miRNAs (overlapping also the stepwise miRNA identified in the tissue signature, see chapter III [5])to be further investigated in exosomes purified from plasma samples by qRT-PCR. First, we took advantage of our CSS FFPE LUAD Cohort (N=44) already stratified in C1- and nonC1- using both 10-genes signature (Fig. 1 B, see also Chapter II).

Then, we collected 32 plasma samples in our institute (i.e., the CSS Cohort), divided in 14 C1- and 17 nonC1-LUAD. We performed a quality control evaluating hemolytic index (ref, Fig.1 C eventually discarding samples with a value >0,2) and we proceeded with the isolation of exosomes using the kit ExoQuick[10], [11] which is the gold standard for exosomes purification. Before RNA extraction, we also evaluated the quality and purity of our exosomes following the ISEV guidelines[5], [12]. Exosomes were analyzed for diameter using Nanosight (Fig.1D) that confirmed a diameter ranging between 30-150nm in line with exosomes-like particles; then we also performed a western blot on tetraspanins (CD9, CD81, CD63) which are transmembrane proteins decorating exosome surface, and other non-exosome proteins such as GM130 and Calnexin used as negative controls markers, which confirmed the purity of our samples (Fig1. E). Finally, we extracted RNA from exosomes and performed expression screening of the subset of 23 miRNAs by qRT-PCR. Fifteen out of 23 miRNAs were detected in exosomes (Fig. 1F). Of these, let-7a-5p, miR-193b-3p, miR-584-5p and miR-223-3p had significant higher expression in C1 exosomes, whereas miR-24-3p had a strong trend of higher expression in C1 exosomes (Fig. 1 F). Next, we performed a differential expression analysis on a total 13-miRNAs that were expressed in at least the 75% of the patients, showing 8 miRNAs enriched and 2 miRNAs depleted in C1 exosomes (Fig. 1 G). Based on these data, we used unconditional logistic regression analysis to build a risk model based on this 13-miRNA exosomal signature (13-exo-miRNA) to diagnose C1-/nonC1-LUAD subtype. Performance of the model was evaluated by calculating the Area Under the Curve (AUC). We found that the 13-exo-miRNA signature identified correctly C1 patients with a bootstrap corrected AUC of 0.64 (Fig. 1 H, left), and significantly higher predicted probability in C1- vs. nonC1-LUAD (Fig. 1 H, right). Then we evaluated a reduced model composed by 6 miRNAs to reduce complexity of the test (let-7a-5p, miR-126-3p, miR-584-5p, miR-15b-5p, miR-30d-5p and miR-31-5p). We obtained a bootstrap corrected AUC of 0.76 (Fig1. H, left), with significantly higher predictive probability value in C1 respect to nonC1 LUAD (Fig1. H, right). These data showed a peculiar exosomal miRNA signature in

plasma samples which can identify C1 subtype which holds the potential to early-identify C1 aggressive subtype using small aliquots of plasma (~0.5mL) samples.

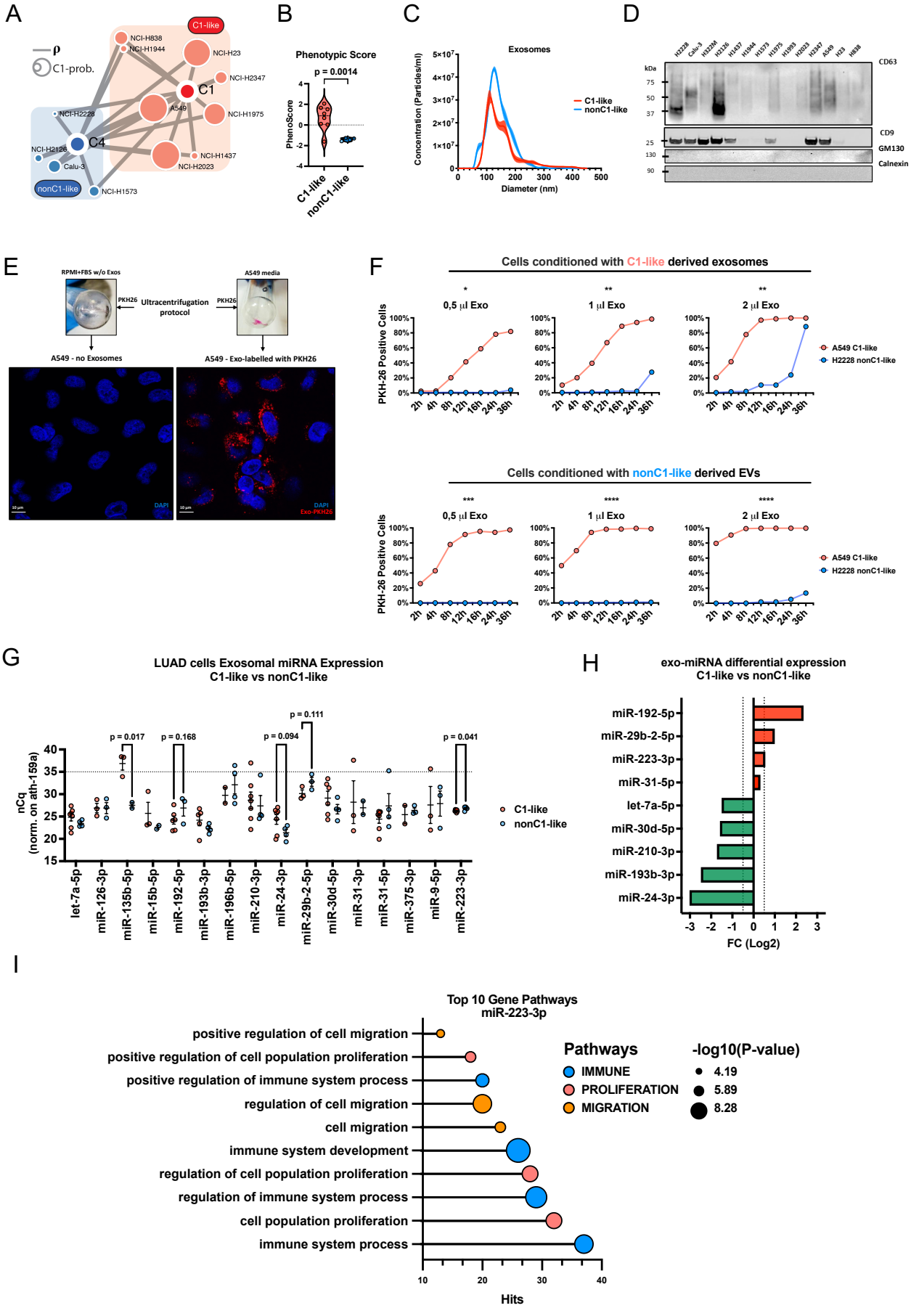


Chapter IV.3 Figure 1: Plasma derived exosomal miRNA signatures to identify C1-LUAD.
A. Volcano plot showing differentially expressed miRNAs analyzed by DESeq2 in C1-LUAD vs. nonC1-LUAD patients. Gray dot, not significant; green dot, down-regulated ($p<0.05$); red dot, upregulated ($p<0.05$); statistical significance was calculated using the Wald test with Benjamin-Hochberg adjustment. **B.** Hierarchical clustering analysis of the 10-gene expression signature in the CSS cohort ($N=44$). C1- and nonC1-LUAD samples are colored as per the legend. **C.** Box plot representing the Haemolytic Index in plasma samples of CSS cohort (N=32). On the y-axes, Absorbance levels (A) derived from the formula of Hemolytic Index (see methods) were represented. **D.** Representative Nanosight analysis of exosomes isolated from C1 and nonC1 plasma. Most abundant peaks are comprised between 30-150nm as expected. **E.** Western Blot representing CD63 (Up, 37-75 kDA, stripped bands due to the glycosylated state of the protein), CD9 (Middle, 25 kDA), CD81 (Down, 25 kDA) of exosomes pool-C1 and pool-nonC1 isolated from plasma. GM130 and Calnexin are used as control markers of other vesicles. **F.** Dot-plot of exosomal-miRNAs expression as normalized Cq (nCq) in C1 and nonC1 patients. Ath-159a spike-in was used as normalizer. On y-axes, dot-line represents the cut-off of miRNA expression (nCq>35). Statistical significance ($p<0.05$) was calculated using Student t-test. P-values <0.2 are reported in the graph. **G.** Differential expression analysis of exo-miRNAs represented as Fold Change values (FC; log2 of $2^{-\Delta\Delta Cq}$) in C1 vs nonC1 patients. Green bar, down-regulated; red bar, up-regulated. Dot lines on x-axes (-0.55 and+0,55) represent the FC cut-off. **H.** ROC curve of the 13-exo-miRNA model and box-plot for C1 predicted probability in C1 and non-C1 patients using the 13-exo-miRNA model. Bootstrap corrected AUC and Student t-test p-value are reported. **I.** ROC curve of the reduced 6-exo-miRNA model and box-plot for C1 predicted probability in C1 and non-C1 patients using the 6-exo-miRNA model. Bootstrap corrected AUC and Student t-test p-value are reported.

C1-like cells internalize high quantity of exosomes and showed exo-miRNAs differential expression

Next, we further investigated the role of C1-exo-miRNA in LUAD aggressiveness. We took the advantage of a panel of LUAD cell lines already stratified in C1-like (N=8) and nonC1-like (N=4) by the 10-genes signature and validated by phenotypic screening analysis (Fig.1 A and B; see also Chapter II)[5]. Next, we isolated exosomes using by an optimized differential centrifugation protocol we already published[13] and we performed quality control evaluation as previously described. As showed in Figure 2 C and D, exosomes derived from LUAD cells had a diameter comprise in 30-150nm and were enriched in tetraspanins. Then, we evaluated biologically activity of exosomes purified. To this, we labelled exosomes derived from a C1-like LUAD cell line (i.e., A549) with PKH-26 (a lipophilic dye) and visualized their cell internalization by confocal microscopy (Fig. 2 E). Next, we labelled exosomes released by a nonC1-like cell lines (i.e. H2228) and we conducted a combined time-course analysis to evaluate kinetic up-take of C1-/nonC1-exosomes in C1-like cells or in non-C1-like cells by FACS analysis. As showed in Fig. 2 F, C1-like cells and nonC1-like cells have a kinetic up-take which is time and dose dependent. Moreover C1-like cells are more prone to internalize self and exogenous exosomes then nonC1-like cells (Fig. 2 F). Of note, nonC1-like cells have higher up-take of C1-like exosomes in respect of self-exosomes, but they remain less efficient in kinetic up-take compared with C1-like cells.

Finally, we evaluated the expression of these 23 miRNAs in C1-/nonC1-like cells. Coherently, we found expressed 16 out of 23 miRNAs in both C1-like and nonC1-like derived exosomes (Fig 2 F). In agreement with data previously obtained from CSS cohort of LUAD patients, Exo-miRNAs were found differentially expressed in C1-like and nonC1-like exosomes. Of these, miR-15b-5p, miR-192-5p and miR-29b-5p had higher expression trend in exosomes derived from C1-like cells, although only miR-223-3p was significantly more expressed in C1-like exosomes (Fig. 2 F).


A
C1-like
NCI-H838
NCI-H1944
NCI-H23
NCI-H2347
C1
A549
NCI-H1975
NCI-H2228
C4
NCI-H2126
Calu-3
NCI-H1437
NCI-H2023
nonC1-like
NCI-H1573
ρ
C1-prob.
B
Phenotypic Score
p = 0.0014
PhenoScore
C1-like
nonC1-like
C
Exosomes
Concentration (Particles/ml)
Diameter (nm)
C1-like
nonC1-like
D
kDa
CD63
CD9
GM130
Calnexin
E
RPMI+FBS w/o Exos
PKH26
Ultracentrifugation protocol
A549 media
A549 - no Exosomes
A549 - Exo-labelled with PKH26
DAPI
Exo-PKH26
F
Cells conditioned with C1-like derived exosomes
0,5 µl Exo
1 µl Exo
2 µl Exo
PKH-26 Positive Cells
A549 C1-like
H2228 nonC1-like
Cells conditioned with nonC1-like derived EVs
G
LUAD cells Exosomal miRNA Expression
C1-like vs nonC1-like
nCq (norm. on ath-159a)
p = 0.017
p = 0.168
p = 0.094
p = 0.111
p = 0.041
let-7a-5p
miR-126-3p
miR-135b-5p
miR-15b-5p
miR-192-5p
miR-193b-3p
miR-196b-5p
miR-210-3p
miR-24-3p
miR-29b-2-5p
miR-30d-5p
miR-31-3p
miR-31-5p
miR-375-3p
miR-9-5p
miR-223-3p
C1-like
nonC1-like
H
exo-miRNA differential expression
C1-like vs nonC1-like
miR-192-5p
miR-29b-2-5p
miR-223-3p
miR-31-5p
let-7a-5p
miR-30d-5p
miR-210-3p
miR-193b-3p
miR-24-3p
FC (Log2)
I
Top 10 Gene Pathways
miR-223-3p
positive regulation of cell migration
positive regulation of cell population proliferation
positive regulation of immune system process
regulation of cell migration
cell migration
immune system development
regulation of cell population proliferation
regulation of immune system process
cell population proliferation
immune system process
Hits
Pathways
IMMUNE
PROLIFERATION
MIGRATION
-log10(P-value)
4.19
5.89
8.28

Chapter IV.3 Figure 2: Biological activity evaluation of LUAD cells exosomes and characteristic exo-miRNA profile of C1-like and nonC1-like cells.

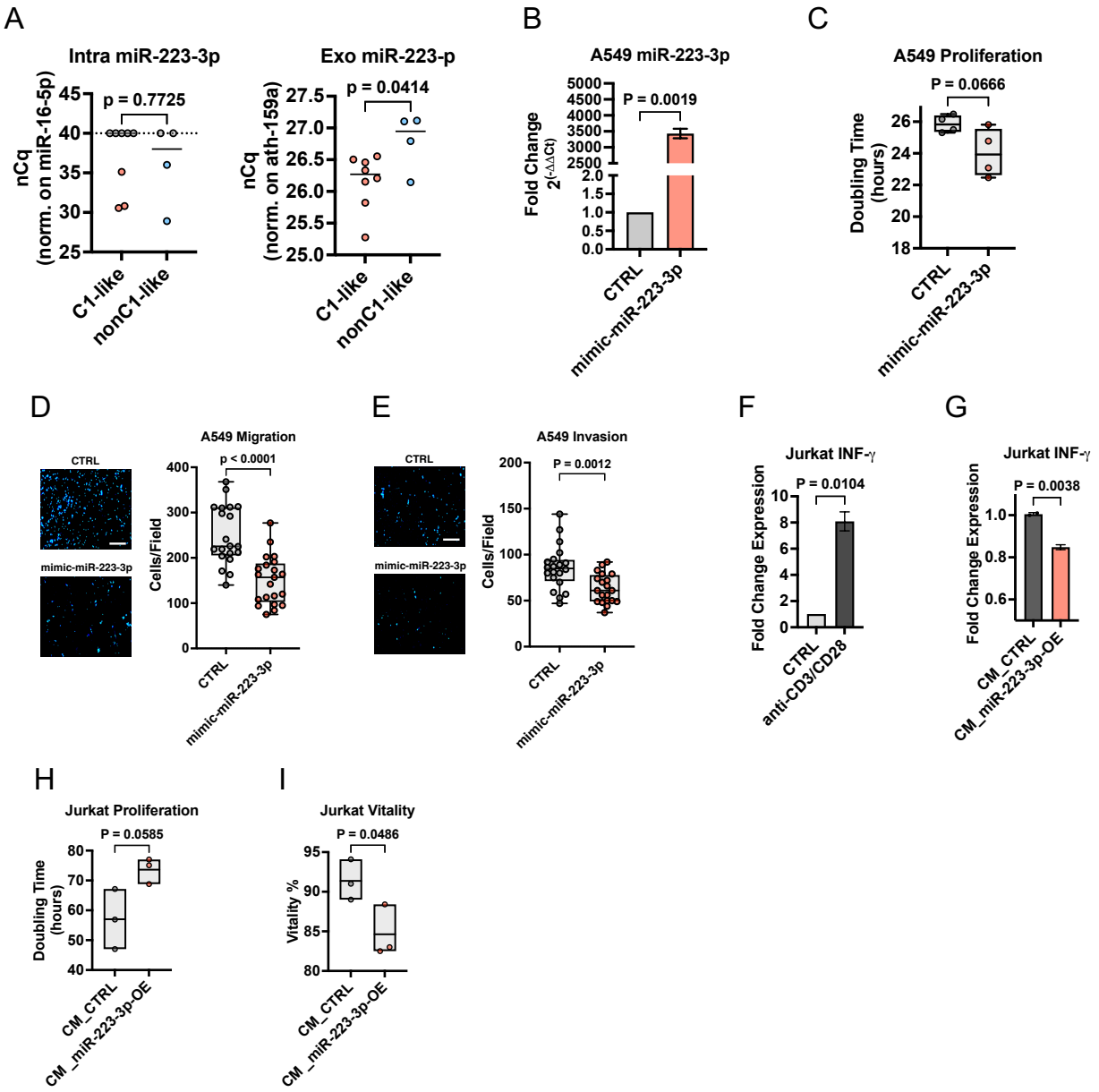
A. Network of CCLE lung adenocarcinoma cell lines (N=12) based on 10-gene expression correlation (edges) and probability of being a C1-like (node size) (see Methods and Chapter II) **B.** Violin plot of the phenotypic scores (*z*-scores) for the 12 cell lines analyzed. P-value (Student t-test) are reported. **C.** Representative Nanosight analysis of exosomes isolated from C1-like and nonC1-like cells. Most abundant peaks are comprised between 30-150nm as expected **D.** Western Blot representing CD63 (Up, 37-75 kDA, stripped bands due to the glycosylated state of the protein), CD9 (Middle, 25 kDA) of exosomes derived from a panel of 14 LUAD cells, comprising the 8 C1-like and the 4 nonC1-like. GM130 and Calnexin are used as control markers of other vesicles. **E.** PKH-26 labelled exosomes internalization in A549 visualized with magnitude 63X plus digital zoom by confocal microscope. Control: media with FBS exosomes depleted. Scale bar are reported. **F.** Flow cytometric analysis of exosomes-PKH26 kinetic up-take by C1-like and nonC1 like cells. Cells were conditioned with exosomes derived from both C1-like cells (A549, Up) and non C1-like cells (H2228, Down). P-values calculated with student t-test.*p<0.05; **p<0.005; ***p<0.0005; ****p<0.00001 **G.** Dot-plot representative of exosomal-miRNAs expression as normalized Cq (nC1) in C1-like and nonC1-like cells. Ath-159a spike-in was used as normalizer. On y-axes, dot-line represent the cut-off of miRNA expression (nCq>35). Statistical significance ($p<0.05$) was calculated using Student t-test. P-values <0.2 are also reported. **H.** Differential expression analysis of exo-miRNAs represented as Fold Change values (FC; log2 of $2^{-\Delta\Delta Cq}$) in C1-like vs nonC1-like cells. Green bar, down-regulated; red bar, up-regulated. Dot lines on x-axes (-0.55 and+0,55) represent the FC cut-off. **I.** Lollipop graph representing analysis performed with miRPathDB[14] of the top 10 gene-pathways related to miR-223-3p derived from experimentally validated Gene Ontology Cell Biology Processes. On y-axes Genes Pathways, colored as per the legend. On x-axes number of genes involved in the pathway (Hits). Size of the bubbles represent p-value expressed as adjusted FDR -Log10 p-values.

Conversely, miR-24-3p showed a lower expression in C1-like exosomes with only miR-135b-5p significantly down-regulated (Fig. 2 F). To focus our attention only on reliable exo-miRNAs, we performed a differential expression analysis on a total 9-miRNAs expressed in at least the 75% of the samples, showing 4 miRNAs up-regulated and 5 miRNAs down-regulated in C1-like exosomes (Fig. 2 G). Interestingly, miR-223-3p was significantly enriched in C1-like exosomes as previously showed in human plasma and included also in the 13-exo-miRNA model (Fig. 1 G, H). For such reasons, we performed a miR-223-3p/gene-pathways analysis using miRPathDB[14], selecting the top 10 gene-pathways related to miR-223-3p experimentally validated in Gene Ontology. We found that miR-223-3p strongly correlated with Immune, Migration and Proliferation gene-pathways (Fig. 2 I). These data suggest us that exo-miR-223-3p could be a key-player in the aggressiveness of C1-LUAD.

miR-223-3p overexpressed in C1-like reduce aggressive phenotype, but his secretion modulates cytotoxic T-cells activity

To better understand how miR-223-3p can be functional in C1-like cells enhanced aggressiveness, we first evaluated its intracellular expression by qRT-PCR and found in both C1-like and nonC1-like cells a very low expression of this miRNA which was in contrast with the results obtained in exosomes (Fig. 3 A). Then, we transiently overexpressed miR-223-3p (Fig. 3 B) in C1-like A549 cells to study phenotypic changes in terms of proliferation, migration and invasion (Fig. 3 C to E). Despite proliferation was augmented (Fig. 3 C), we found a significantly lower migration (Fig. 3 D, left) and invasion (Fig. 3 E, left) in overexpressing miR-223-3p cells relative to control condition (i.e. using a scramble oligo), suggesting a tumor suppressive role of miR-223-3p in C1-like cells, in line with its endogenous low expression level in intracellular compartment. Our previous analysis using miRPathDB (Fig. 2 I) disclosed a potential role of miR-223-3p in immune related pathways, we thus tested how the conditioned media derived from the A549 C1-like

cells over-expressing miR-223-3p (CM-miR-223-3p-OE) could impact on cytotoxic T-cells, using Jurkat leukemic T cell line as model system [15]. First, we treated Jurkat cells with anti-CD3/CD28 for 72h and we observed the Interferon gamma (INFG) production by FACS, taking advantage of an inhibitory cocktail which retain secretion factors in the intracellular compartment (Fig. 3 F). Next, because of INFG produced by cytotoxic T-cells could negatively impact lung cancer cell proliferation[16], we evaluated if CM-miR-223-3p-OE modulates levels of INFG production by FACS and indeed we confirmed a significantly lower production after 72h of growth conditioning (Fig. 3 G). Finally, we verified the number of live cells by trypan blue count, showing a significantly lower percentage of viable cells in activated Jurkat with CM-miR-223-3p-OE compared to control CM (Fig. 3 I). In line with this, the doubling time of activated Jurkat cells was higher thus indicating a lower proliferation rate. These results highlighted a dual role for miR-223-3p which intracellularly can reduce cancer cell migration and invasion while its secretion through exosomes can inactivate T-cells and reduce INFG production thus contributing to elicit an immune evasive phenotype which we previously found to hallmark C1 tumors [4].



Chapter IV.3 Figure 3: miR-223-3p phenotypic investigation in C1-like cells
A. Intracellular (left) and exosomal (right) miR-223-3p expression is shown as normalized Cq (nCq; Intracellular expression was normalized using miR-16-5p; expression in exosomes was normalized using the ath-159a) in C1-like and nonC1-like cells. Grey dot: Undetermined miRNA (Cq>40); the dot-line represents the Cq cutoff (40 Cq). P-values calculated with student t-test were reported. **B.** qRT-PCR differential expression analysis of miR-223-3p in A549 transfected with mimic-miR-223-3p. CTRL: siRNA_CTRL. P-value calculated with student t-test was reported. **C.** Proliferation evaluation by Doubling Time in A549 transfected with mimic-miR-223-3p. CTRL: siRNA_CTRL. P-value calculated with student t-test was reported. **D.** Representative images and quantification of cells migration by trans-well assay in A549 transfected with mimic-miR-223-3p. CTRL: siRNA_CTRL. P-value calculated with student t-test was reported. Scale bar: 10 um **E.** Representative images and quantification of cells invasion by trans-well assay in A549 transfected with mimic-miR-223-3p. CTRL: siRNA_CTRL. P-value calculated with student t-test was reported. Scale bar: 10 um **F.** Interferon-γ production evaluation by FACS in Jurkat cell line, comparing control vs. activated cells with anti-CD3/CD28 (see methods) **G.** Differential expression analysis of Interferon-γ production in Jurkat activated cells conditioned with the C1-like A549 CM_miR-223-3p-OE. CM_CTRL: Conditioned media derived from siRNA CTRL transfected A549. Student t-test p-value reported in the graph **H.** Proliferation evaluation by Doubling Time in Jurkat activated cells conditioned with the C1-like A549 CM_miR-223-3p-OE. CM_CTRL: Conditioned media derived from siRNA CTRL transfected A549. Student t-test p-value reported in the graph. **I.** Vitality evaluation by Doubling Time in Jurkat activated cells conditioned with the C1-like A549 CM_miR-223-3p-OE. CM_CTRL: Conditioned media derived from siRNA CTRL transfected A549. Student t-test p-value reported in the graph.

IV.4 Discussion

Here we showed that the analysis of exosomal-miRNAs in plasma samples of LUAD patients can yield reliable biomarkers for the identification of an aggressive subtype of lung cancer, namely the C1-LUAD. We demonstrated that a model of 6-exo-miRNA can identify C1 subtype with an AUC=0.73. In literature there are several studies about circulating miRNAs signature diagnostic for lung cancer. However, an accurate prediction of prognosis is still difficult due to the high heterogeny of lung cancer [17]–[19]. Recently, extracellular vesicles and exosomes have been deeply investigated as sources for diagnostic and prognostic biomarkers[20]–[26], due to their higher stability and good representation of intrinsic molecular characteristics of the tumor[24], [27], [28]. Notably, our 6-exo-miRNA signature can distinguish aggressive lung tumors with peculiar molecular characteristics which can be exploited for developing next-generation therapeutics. Moreover, this signature is composed by miRNAs with an already established role in lung cancer. Indeed, let-7a-5p, miR-126-3p and miR-584-5p were reported to be tumor suppressor in lung cancer.[29]–[31] Yet, miR-30d-5p and miR-31-5p are very well-known biomarkers of lung cancer, with mechanisms involved in tumor progression.[32]–[35] Of note, miR-31-5p was also already reported in lung cancer exosomes with a role in promoting metastasis formation, whereas miR-126-3p was found to be secreted in exosomes but with a controversial role[36]–[38]. Interestingly, miR-126-3p was enriched in exosomes of high-risk lung cancer patients and it has a pro-tumorigenic effect on stromal cells [37], which is in line with our findings showing high metastatic potential of C1-LUAD also when disease was diagnosed at stage I (see also chapter II)[4], [39]. Instead, the roles of miR-15b-5p are largely unknown in lung cancer though already involved in tumor chemoresistance[40]–[44].

Exosomes are fundamental regulator of cancer traits and metastasis propagation. However, the mechanisms involved in their secretion and internalization remain poorly understood [45]–[47]. Interestingly, we demonstrated that C1-like cells are more prone to internalize

both self and exogenous exosomes compared with nonC1-like cells (Fig. 2 F) thus suggesting a differential tropism of up-take. In details, despite C1-like cells internalize high amount of self-exosomes, they drastically prefer exogenous one, reaching the 100% of up-take in just 8 hours of growth conditioning with the highest quantity of nonC1-like exosomes (Fig. 2F). Accordingly, similar result was observed for nonC1-like cells, which can achieve the 90% of up-take at 36 hours with the highest quantity of C1-like exosomes, compared to less than 20% with their self-exosomes (Fig. 2 F). These preliminary findings, suggest how lung cancer exosomes can be transferred among tumor cells characterized by a different aggressive behavior (i.e. C1- vs. nonC1-like). It has been demonstrated that lung cancer derived exosomes drive invasion and tumorigenic initiation modulating E-cadherin, a well-known marker of EMT [48]. Notably, CD9, a tetraspanins highly enriched in membrane exosomes, was showed to mediate up-take of exosomes in pancreatic cancer providing aggressiveness [49]. Similarly, we also observed expression of CD9 in LUAD exosomes, otherwise, a more quantitative analysis is needed to understand if there are differences in CD9 levels between C1-like and nonC1-like exosomes. [49]–[51].

C1-like cells showed also a peculiar exo-miRNAs profile, with miR-223-3p being the top significantly differentially expressed and highly enriched in C1-like exosomes. Upon miR-223-3p overexpression in C1-like cells, we observed a reduction in migration and in invasion which hallmark the behavior of metastatic cells. This data is in line with literature were miR-223-3p act as a tumor suppressor miRNA[41], [42], [44]. Otherwise, the higher secretion of miR-223-3p in exosomes of C1-like cells, demonstrated a role in reducing, vitality, proliferation and IFNG production of activated Jurkat cells. This effect could be linked to the role of miR-223-3p in the modulation of immune system [50]–[54]. Indeed, miR-223-3p is involved in modulating immune responses by regulating the activation of immune cells such as macrophages, neutrophils, and dendritic cells[51]. Furthermore, miR-223-3p serves as a negative feedback mechanism controlling excessive innate immune responses in the maintenance of myeloid cell homeostasis[52], and regulates the expression of several genes involved in the immune response, including TLR2, TLR4, and STAT3[55]. Our data suggest that C1-tumors prefer to secrete miR-223-3p via exosomes in the microenvironment resulting in immunoevasion phenotype, which is in line with the strong immune-modulative role of the miR-223-3p. Of note, this hypothesis could exploit the already demonstrated role of exosomes in modulating immune infiltration in cancer[9], [56], [57], suggesting a crucial role of the exo-miR-223-3p. Furthermore, the immune evasion phenotype is a characteristic trait of C1-LUAD, confirming indirectly the role of high exo-miR-223-3p in this aggressive cancer (see also Chapter II).[4]

In conclusion, our study provides important insights into the potential of exosomal miRNAs as biomarkers for the identification of aggressive subtypes of lung cancer, such as the C1-

LUAD. Moreover, our findings highlight the requirement to better understand the mechanisms that regulate exosome internalization in aggressive lung cancer, as well as the potential of exosomal miRNAs in modulating cancer phenotypes. Future studies are needed to further validate these findings in larger cohorts of LUAD patients.

IV.5 Material and Methods

CSS Cohort

We selected a cohort of 44 patients with lung adenocarcinoma underwent surgery between February 2017 and February 2020 at the CSS. Written informed consent was obtained from all study patients. None of these patients received preoperative chemotherapy. Clinical information was obtained through review of medical records. Vital status was assessed through the Vital Records Offices of the patients' towns of residence or by contacting directly the patients or their families.

miRNA Expression Analysis of the TCGA-LUAD Cohort

We used TCGA-LUAD cohort (N=510; https://portal.gdc.cancer.gov/) with matched gene and miRNA expression data available in Log2 read count and previously analyzed in *Dama et al*. [5]. Patients were previously stratified in C1 and nonC1 based on 10-gene signature[5]. To focus on the most informative data, we only considered miRNAs with raw counts >0 in at least 50% of patients in either C1 or non-C1, resulting in 382 miRNAs.

FFPE RNA Extraction and qRT-PCR for 10-gene signature

Up to 5 tissue sections (5 μm thick) were prepared from each specimen according to tumor size and/or adequate tumor cellularity (>60%) selected by the pathologist. RNA from FFPE tissues was extracted using AllPrep DNA/RNA FFPE Kit (Qiagen) following the manufacturer's instructions. Briefly, samples were deparaffinized using Xylene and then washed twice in Ethanol. Next, Proteinase K was added to degraded proteins and supernatant was collected and processed in the RNA binding mini-spin column. RNA was then quantified using the Nanodrop ND-1000 Spectrophotometer; a total of 200ng were used for the retro-transcription reaction (SuperScript In Vilo IV Reverse Transcriptase, Invitrogen, ThermoFisher Sicentific) and pre-amplified using the PreAmp Master Mix Kit (Thermo Fisher Scientific) for 10 cycles, following the manufacturer's instructions. The resulting cDNA was then diluted 1:10 prior to PCR analysis (2 ng of cDNA/reaction). Finally, for the analysis of 10-genes were used TaqMan gene expression probes (ThermoFisher Scientific) and the TaqMan Fast Advance Master Mix (ThermoFisher Scientific). qRT-PCR reactions were run on the QuantStudio 12K Flex with a 96-block (ThermoFisher Scientific). For primers list see Supplemental Table 1.

Plasma samples processing

The matched plasma samples were obtained from patients of the Thoracic Surgery Department at CSS, before surgery and in fasting conditions that had signed the informed consent. Plasma samples were collected in BD Vacutainer PPT Tubes. Plasma were centrifugated two times at 1300 x g for 10 min. After the centrifugation plasma samples were aliquoted into 0.5 mL barcoded tubes for long-storage at -80°C. To evaluate the Heamolytic Index of plasma samples, haemolysis was measured by use of different lipemia independent measurement of haemolysis (haemolytic index), a method based on the absorption spectrum of oxyhemoglobin (A=414 nm), haemoglobin (A=385 nm) and lipemia-independent haemolysis score (coefficient a=0.16) following the equation:

H.I.=A414-A385+0.16*A385

following the method of *Marzi et al.*[18]. Briefly, 70 ul of plasma was dispensed in a 96-well PCR plate, centrifuged at 1000 x g for 5 minutes at 4°C. Then 50ul were transferred to a clean well of a 96-well flat-bottomed clear plate, avoiding touching the bottom of the well. The same volume (50ul) of PBS was likewise dispensed into a clean well and used as a blank reference. Ultraviolet-visible (UV-Vis) absorbance measurements are carried out using an Epoch Spectrophotometer (Bioteck) at a wavelength of 414 nm and 385 nm. After normalization of data to the blank measurement the 'H.I.' The absorbance peak at 414 nm and 385 nm, and the H.I. are then annotated in the database. Samples with H.I. > 0.2 were flagged as haemolytic and discarded, because of the higher presence of oxohaemoglobin, haemoglobin and lipemia, that change miRNAs expression.[58][18]

Exosomes isolation from plasma and cell lines

Exosomes were isolated from plasma samples using ExoQuick (System Bioscience), following manufacturer recommendations. The pellet obtained was resuspended in 1X PBS for further analysis or lysed with RIPA buffer (50 mM Tris-HCl, pH 7.4, 150 mM sodium chloride, 1 mM ethylenediaminetetraacetic acid (EDTA), 1% NP-40 1% Sodium Deoxycholic acid, 0.1% sodium dodecylsulfate (SDS)). Exosomes from cell culture media were isolated following our optimized protocol consisting in four successive rounds of centrifugation and one filtration step. First, the medium collected from cell cultures after 72 h of incubation was submitted to 10 min low speed centrifugation at 300 x g at 4° C. The supernatant was transferred in a new tube and centrifuge at 2000 x g, 4°C for 10min. The supernatant was then filtered at 0.22 μm to remove larger size vesicles. To generate the exosome pellet, the supernatant was transferred to ultraclear ultracentrifuge tubes (Beckman Coulter) and centrifuged at 100,000 x g for 90 min at 4°C, using an Optima™ L-90K ultracentrifuge (Beckman Coulter) with a SW 41 Ti swinging-bucket rotor (Beckman Coulter). The resulting exosome pellet was resuspended in PBS 1X and washed with another ultracentrifugation

step in the same conditions; the final pellet was lysed in RIPA Buffer for Western Blot analysis or resuspended in 1X PBS for other purpose. For fluorescent exosomes labelling, after the first ultracentrifugation we added a lipophilic fluorescent dye (PKH26; Sigma-Aldrich) to label exosomes followed by other two ultracentrifuges to wash contaminants and dye excess. A red colored exosomes pellet was obtained. Labelled exosomes were resuspended in 50 μl 1X PBS and stored at 4°C.

Nanoparticle Tracking Analysis (NTA)

Exosomes number and size distribution was determined by nanoparticle tracking analysis (NTA) using a NanoSight NS300 system (Malvern Technologies, Malvern, UK). The NTA was configured with a 488 nm laser and a high sensitivity scientific CMOS camera. Samples were analysed under constant flow conditions (flow rate =30) at 25 °C and 5 × 60 s videos were captured with a camera level ranging between 14 and 16, based on the sample. Data were analyzed using NTA 3.4 Build 3.4.4 software with a detection threshold ranging between 4 and 6, based on the sample.

Western Blot

The exosome protein samples were diluted with Laemli Pierce™ LDS Sample Buffer 4X (ThermoFisher Scientific<9 in Non-Reducing condition for tetraspanins (CD63, CD9, CD81) or in Reducing condition for negative exosomes markers (GM130, Calnexin), and heated to 90°C for 10 minutes. 20 μl of exosomes lysate was then loaded in a 12% gel. Gels were run at 130 volts for 1:30 hour and a Trans-Blot® Turbo™ Transfer System was used for the transfer of the proteins from the gel to a polyvinylidene difluoride (PVDF) membrane. After the transfer, the membranes were blocked 1 hour with a 5% milk or BSA blocking solution. The antibodies used for exosomes characterization were anti-CD9 (CBL162, MM2/57 monoclonal-mouse, Merck Millipore), anti-CD63 (CBL553, RFAC4 monoclonal-mouse, Merck Millipore), anti-CD81 (Monoclonal Mouse IgG2B Clone # 454720; Bio-Techne), anti-GM130 (D6B1 XP Rabbit mAb #12480; Cell Signaling) and anti-Calnexin (C5C9 Rabbit mAb #2679; Cell Signaling), all diluted in ratio 1:1000 in BSA 5%. All the primary antibodies have been incubated overnight at 4°C in gently agitation. All the secondary antibodies were diluted in ratio 1:3000 and were incubated for 1 hour at RT, in gently agitation and subsequently washed three times for 10 min in TBS-Tween 0,1%. Detection was aided by the reporter enzyme horseradish peroxidase (HRP), which is conjugated to the secondary antibody. In the presence of hydrogen peroxide, HRP oxidizes the compound luminol in a chemical reaction that emits light (chemiluminescence). Pico PLUS Chemiluminescent Substrate (ThermoFisher Scientific) or SuperSignal™ West Femto Maximum Sensitivity Substrate (ThermoFisher Scientific) were used to the detection of chemiluminescent signal using the ChemiDoc Imaging System (BioRad) and ImageLab software was used to analyze the images (BioRad).

CCLE lung cancer cell lines classification.

In order to classify CCLE lung adenocarcinoma cell lines in C1-like or nonC1-like subtypes we followed our method already published [4]. Briefly, ee implemented centroid correlation where the bulk 10-gene expression profile of C1- and nonC1-LUAD was used as reference for calculating correlation scores with CCLE 10-gene expression profiles. We excluded from the analysis SCGB3A1 and SERPINB5 because they are principally expressed. Next, Z-score values of each gene of the 10-gene signature were computed in all samples of TCGA-LUAD. We then calculated the averages of the 10-gene signature Z-scores in the cluster of C1-LUAD patients with unfavorable prognosis) and in the cluster of 'C4' patients. These Z-score averages were used as C1-LUAD and nonC1-LUAD centroids, to calculate their expression correlation (Pearson, **ρ**) to that of the CCLE dataset. Next, we transformed **ρ** values to power of ten (i.e., 10^**ρ**) to stabilize variance to get normally distributed data, and we used these values to stratify CCLE cell lines. A cell line was classified as C1-like if its $10^{\boldsymbol{\rho}_{C1}}$ was greater than its $10^{\boldsymbol{\rho}_{nonC1}}$, and vice versa. To calculate the probability of a cell line to be classified as C1- or nonC1-like we used ridge-penalized unconditional logistic regression that was applied in the TCGA-LUAD to model the odds of being in the C1-LUAD cluster (vs. nonC1-LUAD cluster) as a function of the 8 genes detected in CCLE dataset. The expression of each gene was dichotomized according to the median. Cross-validated (10-fold) log-likelihood with optimization (50 simulations) of the tuning penalty parameter was applied. Probability of being in the C1-LUAD cluster was estimated, and model performance was assessed using the area under the receiver operating curve (AUC=0.83).

Phenotypic score

Each parameter of phenotypic assay was transformed into z-scores and a phenotypic score was defined as the sum over the four z-scores and divided by 2, obtaining a variable still following the standard distribution (mean of 0 and standard deviation of 1). Cell lines (N=12) were also ordered by descending proliferation and increasing migration, invasion, and tumor spheres formation, and four ranks (from 1 to 12) were assigned to each cell line, for each assay. T-test test was used to compare C1-like/non-C1-like cell lines phenotypic score distributions.

Cell Culture

H2023, H1975, A549, H23, H2347, H838, H1944, H1437, H1573, CALU-3, H2228, H2126 and H1993 lung adenocarcinoma cell lines were obtained from ATCC. H322M lung adenocarcinoma and Jurkat human T lymphocyte cell lines were gently granted from IFOM. All cell lines were maintained in RPMI (Gibco; Invitrogen) with 10% FBS, and 1% penicillin/streptomycin (Gibco;Invitrogen) at 37°C in a humidified incubator with 5% CO2. Every cell line was plated at different concentration, based on their growth curves, so that

all lines can be processed at a confluence of about 80%. One day after plating for the experiments, media of all cells was changed with another one containing the same composition and concentration, but with exosome's depleted FBS (Euroclone).

Cell transfection experiments

Transfection experiments on A549 cell line were carried out by performing reverse transfection with Lipofectamine RNAiMAX (ThermoFisher Scientific) according to the manufacturer's instructions. Briefly, A549 150.000 cells were plated in a 6-well with 0,5 nM of miR-223-3p mimic (MSY0007230, Qiagen) or recommended 0,5 nM All Stars negative control siRNA (cat. 1027281, Qiagen). After 24h cell media was changed with complete RPMI with FBS w/o exosomes. Cells were detached and lysed after 72h post-transfection and conditioned media was collected and submitted to two steps of centrifugation (first at 300g 10 min; second at 2000g 10min) to eliminate cell debris and apoptotic bodies. Conditioned media was stored at 4°C and used for further experiment.

Activation of Jurkat and Interferon gamma staining

Jurkat cells were plated at a density of $1x10^6$/ml and were treated for 72h with ImmunoCult™ Human CD3/CD28 T Cell Activator (Stemcell Technologies) and with eBioscience™ Protein Transport Inhibitor Cocktail (500X; ThermoFisher Scientific; 00-4980-03), following manufacturer's instructions. Moreover, Jurkat growth was with media derived from A549 transfected with negative control siRNA or with mimic-miR-223-3p. After 72h, cells were pelleted at 300g 5min and fixed with PFA 0,4%. Then, cells were permeabilized using permeabilization buffer (Invitrogen, ThermoFisher Scientific) and stained with anti-IFN-gamma (Monoclonal Antibody (4S.B3), PE, eBioscience™) for flow cytometry analysis.

Flow Cytometry

Cellular internalization of PKH26-positive exosomes in LUAD cell lines and Interferon-gamma staining in Jurkat cells were measured by MoFlo Astrios cell sorter (Beckman Coulter). From the initial forward scatter vs. side scatter plot (FSC/SSC), the cell population was gated. A second plot showed the specific fluorescence. DAPI was used to stain nuclei of viable cells. On each experiment unstained control samples were used. Final analyses were performed using FlowJo (BD Biosciences).

Cells and exosomes RNA extraction

Total RNA purification, including miRNAs, from exosome and cells is based on lysis with guanidinium thiocyanate-phenol-chloroform extraction (TriZol-LS and QIazol, Thermo Fisher Scientific) and spin column-based total RNA purification (MiRneasy Mini Kit, Qiagen). Exosomes pellet were resuspended in 300ul of 1X sterile PBS and were mixed with

Trizol-LS in volumetric ratios of 3:1 for lysis. After denaturation, one synthetic miRNA was added to the solution at fixed amount (3 x 10^9 copies of ath-159a) to monitor the extraction procedure. A volume of 240ul mL of chloroform is then added to the solution, which was centrifuged for 15 minutes at 11,800 x g at 4°C. A fixed amount (350ul; around 70%) of aqueous phase was recovered. This amount was chosen to limit contamination from the interphase and speed up the RNA isolation procedure. The following steps were performed using the miRNEASY Mini Kit (QIAGEN, ThermoFisher Scientific) standard protocol 525ul of 100% ethanol (1.5 volumes) were added and the mixture was then loaded onto a RNeasy Mini spin column that is subsequently washed with 700ul of RWT Buffer, and 500ul of RPE buffer, repeated two times. Total RNA, enriched in small RNAs, is finally eluted in 30ul of RNase-free water, transferred into 1.5 mL barcoded tubes and stored at -80°C, until further analysis.

For the intracellular RNA extraction, cells were pelleted with a low-speed centrifugation (300 x g, 5'), supernatant was discarded and 700ul of QIazol reagent was added on pelleted cells. The following steps are performed using the miRNEASY Mini Kit (QIAGEN, ThermoFisher Scientific) standard protocol, as previously described.

miRNA qRT-PCR in exosomes and cells

For miRNA expression analysis in exosomes, a total of 3ul of RNA was reverse-transcribed using the TaqMan Advanced miRNA cDNA Synthesis Kit (Thermo Fisher Scientific). Poly(A) tailing, adapter ligation, RT reaction and miR-Amp were performed following manufacturer's instructions. The resulting cDNA was then diluted 1:10 prior to perform PCR. TaqMan Fast Advance Master Mix (ThermoFisher Scientific) and Card Custom Advance (Thermo Fisher Scientific) were used for the analysis of subselected miRNAs. qRT-PCR reactions were run on the QuantStudio 12K Flex with an Array-block (ThermoFisher Scientific).

For miRNA expression analysis in cell lines, we used the same protocol, but we retrotranscribed a total of 10ng of intracellular RNA. Then, we diluted 1:10 the cDNA prior to proceed to qRT-PCR for miR-223-3p using TaqMan Fast Advance Master Mix and TaqMan hydrolysis probe in a QuantStudio 12K Flex with a 96-block (ThermoFisher Scientific). For primers list see Supplemental Table 2.

Exo-miRNA signature development

Unconditional logistic regression was applied in the cohort of 32 FFPE samples (N=17 nonC1-like; N=15 C1-like) to model the odds of C1-like as function of the 13 miRNAs expressed in at least 75% of samples. To reduce the number of predictors, the stepwise approach was used, obtaining a 7-miRNAs model. Probability of being C1-like was estimated, and models performance was assessed using the area under the receiver operating curve (AUC). Internal validation of model discrimination was conducted by calculating the bootstrap optimism-corrected AUC.

All statistical analyses were performed using SAS software, version 9.4 (SAS Institute, Inc., Cary, NC, USA), Graphpad 9 (Prism) and JMP 15 (SAS). p-values less than 0.05 were considered statistically significant.

miRPathDB Analysis

Has-miR-223-3p gene-pathways were analyzed with miRPathDB (https://mpd.bioinf.uni-sb.de/mirna.html?mirna=hsa-miR-223-3p&organism=hsa) [14]. Briefly, we filtered the gene pathways of Gene Ontology Biological Process based on their role in cancer (i.e. Proliferation, Migration, Immune, Stemness, EMT, Angiogenesis, Hypoxia). Then, we selected only these gene-pathways experimentally validated and significantly related to has-miR-223-3p, with at least 5 genes involved in the pathway (Hits). Finally, we reported the top 10 gene-pathways as the most representative of miR-223-3p function.

Migration and Invasion Assay

To evaluate migration of cell lines, cells were suspended in serum-free RPMI and plated at a density of 1x10^4 into the upper chamber of trans-well with 8.0-μm pore polycarbonate membrane (Falcon). RPMI with 20% of FBS was added to the lower well as chemoattractant. After 24h, trans-well was fixed using 4%PFA and nuclei of the cells in the trans-well chamber were stained using DAPI (Sigma Aldrich, working solution 1ng/ml). Non-migrated cells were removed from the upper chamber membrane with a cotton-tipped swabs and then the chamber was washed twice in PBS 1X. After DAPI staining ten images were acquired at 10X magnification for each cell line using the Axiocam 105 color (Zeiss) connected at the Eclipse TE300 Microscope (Nikon). Migrated cells were evaluated counting nuclei with Cell Count tool of Fiji software.

For invasion assay the protocol was the same of the migration, but we coated with Matrigel Growth Factor Reduced and Phenol red free (Corning, Lot 4132005) at a concentration of 150 μg/ml, the trans-well chambers with 8.0-μm pore polycarbonate membrane (Falcon). All experiments were performed in at least duplicate.

Proliferation Assay

To calculate doubling time, we performed the experiments s at 0h, 24h and 72h putting 10ul of cell suspension in Luna counter slide (Logos Byosistem) and analyzing the number of cells with Cell Luna II Counter (Logos Byosistem) with the followed parameters: Noise reduction: 5; Live cell sensitivity: 1; Roundness: 60%; Min cell size: 2um; Max cell size: 60um; Declustering level: medium. Then, it was obtained growth rate as follow:

$$gr = \frac{\ln\left(\frac{N(t)}{N(0)}\right)}{t}$$

- $N(t)$ = number of cells at time t
- $N(0)$ = number of cells at time 0
- gr = growth rate

- t = time in hours

Finally, it was calculated the doubling time as follow:

$$doubling\ time = \frac{\ln(2)}{growth\ rate}$$

All experiments were performed at least in triplicates.

Cell Viability assay

We evaluate vitality at 0h and 72h in Jurkat cell line adding 1:2 trypan blue 0.4% (Logos Byosistem) to cell solution and putting 10ul of cell suspension in Luna counter slide (Logos Byosistem). Then we automatically counted the live cells with the Luna II cell counter (Logos Byosistem). Percentage of live cells was calculated by the instrument based on the followed parameters: Noise reduction: 5; Live cell sensitivity: 1; Roundness: 90%; Min cell size: 10um; Max cell size: 60um; Declustering level: medium.

All experiments were performed at least in triplicates**.**

Statistical analysis

To perform hierarchical clustering, we used Cluster 3.0 and Java Tree View software, using the uncentered correlation and centroid linkage method. Heatmaps were generated with MORPHEUS or Java Tree View. Statistical analyses and graphs were generated using GraphPad 9 (Prism), SPSS version 15.0, SAS software, and R version 3.3.1. Normality of data distribution was checked using Shapiro-Wilk test and the other statistical test were performed based on data distribution. All the statistical tests used and the p-values were reported in figures, legends or in method details. A p-value less than 0.05 was considered as significant.

IV.6 References

[1] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, 'Cancer statistics, 2022', *CA Cancer J Clin*, vol. 72, no. 1, pp. 7–33, Jan. 2022, doi: 10.3322/CAAC.21708.

[2] M. Oudkerk, S. Y. Liu, M. A. Heuvelmans, J. E. Walter, and J. K. Field, 'Lung cancer LDCT screening and mortality reduction - evidence, pitfalls and future perspectives', *Nat Rev Clin Oncol*, vol. 18, no. 3, pp. 135–151, Mar. 2021, doi: 10.1038/S41571-020-00432-6.

[3] D. Bayik and J. D. Lathia, 'Cancer stem cell-immune cell crosstalk in tumour progression', *Nat Rev Cancer*, vol. 21, no. 8, pp. 526–536, Aug. 2021, doi: 10.1038/S41568-021-00366-W.

[4] V. Melocchi *et al.*, 'Aggressive early-stage lung adenocarcinoma is characterized by epithelial cell plasticity with acquirement of stem-like traits and immune evasion phenotype', *Oncogene*, 2021, doi: 10.1038/s41388-021-01909-z.

[5] E. Dama *et al.*, 'Non-coding RNAs as prognostic biomarkers: A mirna signature specific for aggressive early-stage lung adenocarcinomas', *Noncoding RNA*, vol. 6, no. 4, pp. 1–13, 2020, doi: 10.3390/ncrna6040048.

[6] M. Smolarz and P. Widlak, 'Serum exosomes and their miRNA load — A potential biomarker of lung cancer', *Cancers (Basel)*, vol. 13, no. 6, pp. 1–20, 2021, doi: 10.3390/cancers13061373.

[7] B. Zhou *et al.*, 'Application of exosomes as liquid biopsy in clinical diagnosis', *Signal Transduction and Targeted Therapy 2020 5:1*, vol. 5, no. 1, pp. 1–14, Aug. 2020, doi: 10.1038/s41392-020-00258-9.

[8] A. B, V. PF, and M. SA, 'Decoding the Biology of Exosomes in Metastasis', *Trends Cancer*, vol. 6, no. 1, pp. 20–30, Jan. 2020, doi: 10.1016/J.TRECAN.2019.11.007.

[9] J. Dai *et al.*, 'Exosomes: key players in cancer and potential therapeutic strategy', *Signal Transduct Target Ther*, vol. 5, no. 1, Dec. 2020, doi: 10.1038/S41392-020-00261-0.

[10] M. Y. Konoshenko, E. A. Lekchnov, A. V. Vlassov, and P. P. Laktionov, 'Isolation of Extracellular Vesicles: General Methodologies and Latest Trends', *Biomed Res Int*, vol. 2018, 2018, doi: 10.1155/2018/8545347.

[11] Y. T. Tang *et al.*, 'Comparison of isolation methods of exosomes and exosomal RNA from cell culture medium and serum', *Int J Mol Med*, vol. 40, no. 3, pp. 834–844, 2017, doi: 10.3892/ijmm.2017.3080.

[12] C. Théry *et al.*, 'Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines', *J Extracell Vesicles*, vol. 7, no. 1, Jan. 2018, doi: 10.1080/20013078.2018.1535750.

[13] T. Colangelo *et al.*, 'Extracellular vesicle microRNAs contribute to Notch signaling pathway in T-cell acute lymphoblastic leukemia', *Mol Cancer*, vol. 21, no. 1, pp. 1–8, Dec. 2022, doi: 10.1186/S12943-022-01698-3/FIGURES/2.

[14] T. Kehl *et al.*, 'miRPathDB 2.0: a novel release of the miRNA Pathway Dictionary Database', *Nucleic Acids Res*, vol. 48, no. D1, pp. D142–D147, Jan. 2020, doi: 10.1093/NAR/GKZ1022.

[15] R. Cuttano *et al.*, 'miRNome profiling of lung cancer metastases revealed a key role for miRNA-PD-L1 axis in the modulation of chemotherapy response', *J Hematol Oncol*, vol. 15, no. 1, Dec. 2022, doi: 10.1186/S13045-022-01394-1.

[16] A. M. Gocher, C. J. Workman, and D. A. A. Vignali, 'Interferon-γ: teammate or opponent in the tumour microenvironment?', *Nature Reviews Immunology 2021 22:3*, vol. 22, no. 3, pp. 158–172, Jun. 2021, doi: 10.1038/s41577-021-00566-3.

[17] F. Montani *et al.*, 'miR-Test: A Blood Test for Lung Cancer Early Detection', *JNCI: Journal of the National Cancer Institute*, vol. 107, no. 6, Jun. 2015, doi: 10.1093/JNCI/DJV063.

[18] M. J. Marzi *et al.*, 'Optimization and Standardization of Circulating MicroRNA Detection for Clinical Application : The miR-Test Case', vol. 754, pp. 743–754, 2016, doi: 10.1373/clinchem.2015.251942.

[19] T. Fehlmann *et al.*, 'Evaluating the Use of Circulating MicroRNA Profiles for Lung Cancer Detection in Symptomatic Patients', *JAMA Oncol*, vol. 6, no. 5, pp. 714–723, May 2020, doi: 10.1001/JAMAONCOL.2020.0001.

[20] T. L. Whiteside, 'Validation of plasma-derived small extracellular vesicles as cancer biomarkers', *Nature Reviews Clinical Oncology 2020 17:12*, vol. 17, no. 12, pp. 719–720, Sep. 2020, doi: 10.1038/s41571-020-00433-5.

[21] A. Hildebrandt, B. Kirchner, E. N. M. Nolte-'t Hoen, and M. W. Pfaffl, 'miREV: An Online Database and Tool to Uncover Potential Reference RNAs and Biomarkers in Small-RNA Sequencing Data Sets from Extracellular Vesicles Enriched Samples', *J Mol Biol*, vol. 433, no. 15, p. 167070, Jul. 2021, doi: 10.1016/J.JMB.2021.167070.

[22] M. Silva and S. A. Melo, 'Non-coding RNAs in Exosomes: New Players in Cancer Biology', *Curr Genomics*, vol. 16, pp. 295–303, 2015, doi: 10.2174/1389202916666150707154719.

[23] A. Mukherjee, B. Bisht, S. Dutta, and M. K. Paul, 'Current advances in the use of exosomes, liposomes, and bioengineered hybrid nanovesicles in cancer detection and therapy', *Acta Pharmacologica Sinica 2022 43:11*, vol. 43, no. 11, pp. 2759–2776, Apr. 2022, doi: 10.1038/s41401-022-00902-w.

[24] N. Tran, 'Cancer Exosomes as miRNA Factories', *Trends Cancer*, vol. 2, no. 7, pp. 329–331, 2016, doi: 10.1016/j.trecan.2016.05.008.

[25] M. Smolarz and P. Widlak, 'Serum Exosomes and Their miRNA Load-A Potential Biomarker of Lung Cancer', *Cancers (Basel)*, vol. 13, no. 6, pp. 1–20, Mar. 2021, doi: 10.3390/CANCERS13061373.

[26] C. Masaoutis, C. Mihailidou, G. Tsourouflis, and S. Theocharis, 'Exosomes in lung cancer diagnosis and treatment. From the translating research into future clinical practice', *Biochimie*, vol. 151, pp. 27–36, 2018, doi: 10.1016/j.biochi.2018.05.014.

[27] X. Zhang, X. Yuan, H. Shi, L. Wu, H. Qian, and W. Xu, 'Exosomes in cancer: small particle, big player.', *J Hematol Oncol*, vol. 8, p. 83, Jul. 2015, doi: 10.1186/s13045-015-0181-x.

[28] P. Kharaziha, S. Ceder, Q. Li, and T. Panaretakis, 'Tumor cell-derived exosomes: A message in a bottle', *Biochim Biophys Acta Rev Cancer*, vol. 1826, no. 1, pp. 103–111, 2012, doi: 10.1016/j.bbcan.2012.03.006.

[29] S. Duan, S. Yu, T. Yuan, S. Yao, and L. Zhang, 'Exogenous let-7a-5p induces A549 lung cancer cell death through BCL2L1-mediated PI3Kγ signaling pathway', *Front Oncol*, vol. 9, no. AUG, p. 808, Aug. 2019, doi: 10.3389/FONC.2019.00808/BIBTEX.

[30] D. Di Paolo *et al.*, 'Cotargeting of miR-126-3p and miR-221-3p inhibits PIK3R2 and PTEN, reducing lung cancer growth and metastasis by blocking AKT and CXCR4 signalling', *Mol Oncol*, vol. 15, no. 11, pp. 2969–2988, Nov. 2021, doi: 10.1002/1878-0261.13036.

[31] S. B. Lee, Y. S. Park, J. S. Sung, J. W. Lee, B. Kim, and Y. H. Kim, 'Tumor Suppressor miR-584-5p Inhibits Migration and Invasion in Smoking Related Non-Small Cell Lung Cancer Cells by Targeting YKT6', *Cancers 2021, Vol. 13, Page 1159*, vol. 13, no. 5, p. 1159, Mar. 2021, doi: 10.3390/CANCERS13051159.

[32] C. Zhu *et al.*, 'miR-31-5p modulates cell progression in lung adenocarcinoma through TNS1/p53 axis', *Strahlenther Onkol*, vol. 198, no. 3, pp. 304–314, Mar. 2022, doi: 10.1007/S00066-021-01895-X.

[33] Z. Lu *et al.*, 'miR-31-5p Is a Potential Circulating Biomarker and Therapeutic Target for Oral Cancer', *Molecular Therapy: Nucleic Acid*, vol. 16, pp. 471–480, 2019, doi: 10.1016/j.omtn.2019.03.012.

[34] Q. Zhao, X. Yuan, L. Zheng, and M. Xue, 'miR-30d-5p: A Non-Coding RNA With Potential Diagnostic, Prognostic and Therapeutic Applications', *Front Cell Dev Biol*, vol. 10, p. 125, Jan. 2022, doi: 10.3389/FCELL.2022.829435/BIBTEX.

[35] L. Gao *et al.*, 'Expression signature and role of miR-30d-5p in non-small cell lung cancer: A comprehensive study based on in silico analysis of public databases and in vitro experiments', *Cellular Physiology and Biochemistry*, vol. 50, no. 5, pp. 1964–1987, 2018, doi: 10.1159/000494875.

[36] M. Li, Q. Wang, X. Zhang, N. Yan, and X. Li, 'Exosomal miR-126 blocks the development of non-small cell lung cancer through the inhibition of ITGA6', *Cancer Cell Int*, vol. 20, no. 1, pp. 1–11, 2020, doi: 10.1186/s12935-020-01653-6.

[37] F. Pontis *et al.*, 'Circulating extracellular vesicles from individuals at high-risk of lung cancer induce pro-tumorigenic conversion of stromal cells through transfer of miR-126 and miR-320', *J Exp Clin Cancer Res*, vol. 40, no. 1, Dec. 2021, doi: 10.1186/S13046-021-02040-3.

[38] F. Yu, M. Liang, Y. Huang, W. Wu, B. Zheng, and C. Chen, 'Hypoxic tumor-derived exosomal miR-31-5p promotes lung adenocarcinoma metastasis by negatively regulating SATB2-reversed EMT and activating MEK/ERK signaling', *Journal of Experimental and Clinical Cancer Research*, vol. 40, no. 1, pp. 1–15, Dec. 2021, doi: 10.1186/S13046-021-01979-7/FIGURES/7.

[39] E. Dama *et al.*, 'An aggressive subtype of stage I lung adenocarcinoma with molecular and prognostic characteristics typical of advanced lung cancers', *Clinical Cancer Research*, vol. 23, no. 1, pp. 62–72, 2017, doi: 10.1158/1078-0432.CCR-15-3005.

[40] S. Ghafouri-Fard, T. Khoshbakht, B. M. Hussen, H. H. Jamal, M. Taheri, and M. Hajiesmaeili, 'A Comprehensive Review on Function of miR-15b-5p in Malignant and Non-Malignant Disorders', *Front Oncol*, vol. 12, May 2022, doi: 10.3389/FONC.2022.870996.

[41] Q. Ji *et al.*, 'miR-223-3p Inhibits Human Osteosarcoma Metastasis and Progression by Directly Targeting CDH6', *Molecular Therapy*, vol. 26, no. 5, pp. 1299–1312, 2018, doi: 10.1016/j.ymthe.2018.03.009.

[42] S. Zhu *et al.*, 'MiR-223-3p attenuates the migration and invasion of NSCLC cells by regulating NLRP3', *Front Oncol*, vol. 12, Oct. 2022, doi: 10.3389/FONC.2022.985962/FULL.

[43] L. Chen *et al.*, 'BMSCs-derived miR-223-containing exosomes contribute to liver protection in experimental autoimmune hepatitis', *Mol Immunol*, vol. 93, no. November 2017, pp. 38–46, 2018, doi: 10.1016/j.molimm.2017.11.008.

[44] L. Qin *et al.*, 'IncRNA F630028O10Rik Contributes to Suppress Lung Cancer in Mice Through Inhibiting miR-223-3p and VEGF Signaling Pathway', *Chest*, vol. 149, no. 4, p. A304, 2016, doi: 10.1016/j.chest.2016.02.317.

[45] P. Fuentes *et al.*, 'ITGB3-mediated uptake of small extracellular vesicles facilitates intercellular communication in breast cancer cells', *Nature Communications 2020 11:1*, vol. 11, no. 1, pp. 1–15, Aug. 2020, doi: 10.1038/s41467-020-18081-9.

[46] A. Möller and R. J. Lobb, 'The evolving translational potential of small extracellular vesicles in cancer', *Nature Reviews Cancer 2020 20:12*, vol. 20, no. 12, pp. 697–709, Sep. 2020, doi: 10.1038/s41568-020-00299-w.

[47] A. C. Dixson, T. R. Dawson, D. Di Vizio, and A. M. Weaver, 'Context-specific regulation of extracellular vesicle biogenesis and cargo selection', *Nature Reviews Molecular Cell Biology 2023*, pp. 1–23, Feb. 2023, doi: 10.1038/s41580-023-00576-0.

[48] H. Hasan *et al.*, 'Extracellular vesicles released by non-small cell lung cancer cells drive invasion and permeability in non-tumorigenic lung epithelial cells', *Scientific Reports 2022 12:1*, vol. 12, no. 1, pp. 1–18, Jan. 2022, doi: 10.1038/s41598-022-04940-6.

[49] J. Nigri *et al.*, 'CD9 mediates the uptake of extracellular vesicles from cancer-associated fibroblasts that promote pancreatic cancer cell aggressiveness', *Sci Signal*, vol. 15, no. 745, Aug. 2022, doi: 10.1126/SCISIGNAL.ABG8191/SUPPL_FILE/SCISIGNAL.ABG8191_MDAR_REPRODUCIBILITY_CHECKLIST.PDF.

[50] C. Jimenez Calvente, H. Del Pilar, M. Tameda, C. D. Johnson, and A. E. Feldstein, 'MicroRNA 223 3p Negatively Regulates the NLRP3 Inflammasome in Acute and Chronic Liver Injury', *Molecular Therapy*, vol. 28, no. 2, pp. 653–663, Feb. 2020, doi: 10.1016/j.ymthe.2019.09.013.

[51] S. Yuan *et al.*, 'miR-223: An Immune Regulator in Infectious Disorders', *Front Immunol*, vol. 12, p. 5346, Dec. 2021, doi: 10.3389/FIMMU.2021.781815/BIBTEX.

[52] X. Yuan *et al.*, 'MicroRNA miR-223 as Regulator of Innate Immunity', *J Leukoc Biol*, vol. 104, no. 3, p. 515, Sep. 2018, doi: 10.1002/JLB.3MR0218-079R.

[53] P. Jiao *et al.*, 'miR-223: An Effective Regulator of Immune Cell Differentiation and Inflammation', *Int J Biol Sci*, vol. 17, no. 9, p. 2308, 2021, doi: 10.7150/IJBS.59876.

[54] J. Jeffries, W. Zhou, A. Y. Hsu, and Q. Deng, 'miRNA-223 at the crossroads of inflammation and cancer', *Cancer Lett*, vol. 451, pp. 136–141, Jun. 2019, doi: 10.1016/J.CANLET.2019.02.051.

[55] J. Wu *et al.*, 'Impact of miR-223-3p and miR-2909 on inflammatory factors IL-6, IL-1ß, and TNF-α, and the TLR4/TLR2/NF-κB/STAT3 signaling pathway induced by lipopolysaccharide in human adipose stem cells', *PLoS One*, vol. 14, no. 2, p. e0212063, Feb. 2019, doi: 10.1371/JOURNAL.PONE.0212063.

[56] H. G. Zhang and W. E. Grizzle, 'Exosomes and cancer: A newly described pathway of immune suppression', *Clinical Cancer Research*, vol. 17, no. 5, pp. 959–964, 2011, doi: 10.1158/1078-0432.CCR-10-1489.

[57] R. CF, A. B, S. M, and M. SA, 'The Biology of Cancer Exosomes: Insights and New Perspectives', *Cancer Res*, vol. 77, no. 23, pp. 6480–6488, Dec. 2017, doi: 10.1158/0008-5472.CAN-17-0994.

[58] V. Appierto, M. Callari, E. Cavadini, D. Morelli, M. G. Daidone, and P. Tiberio, 'A lipemia-independent NanoDrop(®)-based score to identify hemolysis in plasma and serum samples.', *Bioanalysis*, vol. 6, no. 9, pp. 1215–26, May 2014, doi: 10.4155/bio.13.344.

IV.7 Supplemental Tables and Figures

Supplemental Table 1. Gene Primers List

Gene	Code	Company
E2F4	Hs00608100_g1	ThermoFIsher Scientific
SCGB3A1	Hs00369360_g1	ThermoFIsher Scientific
SF3B1	Hs00961636_m1	ThermoFIsher Scientific
HSPG2	Hs01078536_m1	ThermoFIsher Scientific
RRM2	Hs00357247_g1	ThermoFIsher Scientific
HPRT1	Hs02800695_m1	ThermoFIsher Scientific
GUSB	Hs99999908_m1	ThermoFIsher Scientific
ESD1	Hs00382667_m1	ThermoFIsher Scientific
HOXB7	NM_004502	ThermoFIsher Scientific
E2F1	NM_005225	ThermoFIsher Scientific
THY1	Hs00174816_m1	ThermoFIsher Scientific
NUDCD1	NM_032869	ThermoFIsher Scientific
RRM2	NM_001034	ThermoFIsher Scientific
SERPINB5	NM_002639	ThermoFIsher Scientific

Supplemental Table 2. miRNAs Primers List

miRNA	Code	Company
478197_mir	hsa-miR-708-5p	ThermoFisher Scientific
478003_mir	hsa-miR-29b-2-5p	ThermoFisher Scientific
478582_mir	hsa-miR-135b-5p	ThermoFisher Scientific
478262_mir	hsa-miR-192-5p	ThermoFisher Scientific
477941_mir	hsa-miR-187-3p	ThermoFisher Scientific
477970_mir	hsa-miR-210-3p	ThermoFisher Scientific
479605_mir	hsa-miR-3065-3p	ThermoFisher Scientific
478074_mir	hsa-miR-375-3p	ThermoFisher Scientific
478314_mir	hsa-miR-193b-3p	ThermoFisher Scientific
478214_mir	hsa-miR-9-5p	ThermoFisher Scientific
477860_mir	hsa-miR-16-5p	ThermoFisher Scientific
478606_mir	hsa-miR-30d-5p	ThermoFisher Scientific
479064_mir	hsa-miR-582-3p	ThermoFisher Scientific
478585_mir	hsa-miR-196b-5p	ThermoFisher Scientific
478167_mir	hsa-miR-584-5p	ThermoFisher Scientific
478575_mir	hsa-let-7a-5p	ThermoFisher Scientific
477887_mir	hsa-miR-126-3p	ThermoFisher Scientific
477992_mir	hsa-miR-24-3p	ThermoFisher Scientific
478313_mir	hsa-miR-15b-5p	ThermoFisher Scientific
478742_mir	hsa-miR-193b-5p	ThermoFisher Scientific
478012_mir	hsa-miR-31-3p	ThermoFisher Scientific
477852_mir	hsa-miR-550a-5p	ThermoFisher Scientific
478411_mir	ath-miR159a	ThermoFisher Scientific
478015_mir	hsa-miR-31-5p	ThermoFisher Scientific
477983_mir	hsa-miR-223-3p	ThermoFisher Scientific

Chapter V: General Conclusions and Perspectives

V.1 General Conclusion

Results of these studies allowed to characterize a newly identified aggressive molecular subtype of lung cancer, C1-LUAD, and to discover circulating biomarkers for early identification. In particular:

1)We used a multi-tiered approach, to analyze the transcriptional profile of three independent cohorts of LUAD patients, as well as we performed a phenotypic screening in 12 LUAD cell lines. We found that C1-LUAD exhibited a loss of lung cell lineage and an acquisition of stem cell-like features, as well as an increased mutator and immune evasion phenotype. These characteristics ultimately led to a more aggressive cancer with a worse prognosis, even at stage I.

2) To aid in a better identification of C1-LUAD, we performed miRNA expression profiling, ultimately identifying two signatures of 14- and 19-miRNA, which accurately stratified C1 and non-C1 cancer patients. We then explored the miRNA-mRNA interactome, identifying a set of seven miRNAs as "HUBs" which derived from both the 19-miRNA and 14-miRNA signatures. This seven-miRNA signature had a cross-validated AUC of 0.79 in C1/non-C1 patients' stratification, which was validated in an external cohort of 44 lung adenocarcinoma tissues.

3)Finally, we investigated exosomal miRNAs as potential circulating biomarkers for C1-LUAD identification, identifying a total of 13 miRNAs differentially expressed in C1 and non-C1 exosomes. Consistently, we demonstrated that a minimal set of 6-exo-miRNAs identified C1 patients with a bootstrap corrected AUC of 0.73. Moreover, we started to investigate the role of exo-miR-223-3p in C1-like aggressive cells.

Overall, this thesis provides valuable insights into the molecular characteristics of C1-LUAD and the development of potential biomarkers for early identification. These findings have important implications for the diagnosis, prognosis, and treatment of lung cancer patients, particularly those with the aggressive C1-LUAD subtype.

V.2 Future Perspectives

Despite the overall data here presented, there is still much work to be done to translate these findings into improved clinical outcomes for lung cancer patients.

One promising avenue for future research is the development of targeted therapies for C1-LUAD. The identification of stem cell-like features and immune evasion phenotype in C1-LUAD suggests that targeting these pathways may be effective in treating this subtype of lung cancer. Additionally, further exploration of the miRNA-mRNA interactome could reveal additional therapeutic targets for C1-LUAD.

Another important area of future research is the validation of the miRNA signatures and exosomal miRNA biomarkers in larger, independent cohorts. While the miRNA signatures

showed promising results in stratifying C1 and non-C1 patients, validation in larger patient populations will be necessary to establish their clinical utility. Additionally, further investigation into the functional roles of these miRNAs in C1-LUAD could provide important insights into the mechanisms underlying the aggressive behavior of this subtype. Lastly, the development of non-invasive diagnostic tests for the early detection of C1-LUAD could have a significant impact on the prognosis and survival of lung cancer patients. The identification of exosomal miRNA biomarkers suggests that liquid biopsy approaches may be a promising avenue for the development of such tests. However, further optimization and validation of these tests will be necessary before they can be implemented in a clinical setting.

In summary, the research presented in this thesis has provided important insights into the molecular characteristics of the aggressive C1-LUAD subtype and the development of potential biomarkers for early identification. Further research in the areas of targeted therapies, validation of biomarkers, and development of non-invasive diagnostic tests will be necessary to translate these findings into improved clinical outcomes for lung cancer patients.