Medizinische Fakultät
der
Universität Duisburg-Essen

Aus dem Institut für Pathologie

miRNA markers for subtyping of non-small cell lung cancer. A data-adaptive analysis of expression profiles

I n a u g u r a l - D i s s e r t a t i o n
zur
Erlangung des Doktorgrades der Medizin
durch die Medizinische Fakultät
der Universität Duisburg-Essen

Vorgelegt von
Felix Marcel Bertram
aus Bremerhaven
2020

DOI: 10.17185/duepublico/74281
URN: urn:nbn:de:hbz:464-20210611-083422-0

Dekan:	Herr Univ.-Prof. Dr. med. J. Buer
1. Gutachter:	Herr Univ.-Prof. Dr. K. W. Schmid
2. Gutachter:	Herr Priv.- Doz. Dr. med. W. E. E. Eberhardt
3. Gutachter:	Frau Prof. Dr. med. A. Tannapfel

Tag der mündlichen Prüfung: 18. März 2021

Contents

1. Introduction

1.1. Nucleotide markers in an evolving molecular pathology

Molecular markers, used for both diagnosing and monitoring diseases, are driving innovations across all medical fields. The discipline of pathology applies such markers for biological characterization of cancers. In this way, they facilitated a more personalized treatment of cancer, helping to improve prognosis in developed countries (Tsongalis and Silverman, 2006; Allemani *et al.*, 2018). This study intends to identify molecular markers for the differentiation of lung cancer subtypes.

An important criterion is stability of the marker of interest. This means, first, that the marker can be measured within a small range of statistical deviation in the diagnostic material. It also refers to biochemical stability, which allows reliable sampling and laboratory procession. Furthermore, technical measurement needs to be replicable (Chau *et al.*, 2010). As almost all detectable molecules in human materials are not universally specific, markers should only be used for narrowly defined questions (Holland, 2016).

Most molecular markers in routine use are proteins, examined through immuno-histochemistry. The reason for the preference of proteins is their relatively good conservation in formalin-fixed, paraffin-embedded (FFPE) samples, the main working material in pathology (True, 2014). While immunohistochemistry is a highly popular technique, it is only semi-quantitative, being affected by observer-dependence, compartmentalized expression and non-specificity of antigens (Walker, 2006).

Nucleotides, especially RNA, have a closer functional link to cancer biology than proteins, but their role in diagnostics is relatively small. In recent years, it has become possible to extract RNA from FFPE material (Patel *et al.*, 2017), where it may degrade after long time of storage. A molecular subgroup of RNA, the non-coding microRNA (miRNA) remain relatively stable even after multiple years of storage (Hall *et al.*, 2012). miRNA regulate gene expression, define cellular differentiation and their expression has a high tissue-specificity (Lu *et al.*, 2005), a useful characteristic for differential diagnosis of cancer subtypes.

1.2. Clinical importance and challenges of NSCLC subtyping

Lung cancer is the third most common cancer in Western Europe, where an individual has a cumulative life risk of 3.04% to die of this disease (Bray *et al.*, 2018). Almost all cases develop from epithelial lung cells and are divided into 40 histological subtypes (Travis *et al.*, 2015).

These subtypes are classified into small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) with respective ratios among all lung cancer cases of 13.3% and 83.3%, the remainder being sarcomas or non-specified tumors (Noone *et al.*, 2018). This classification is based on histology, as the highly aggressive SCLC show a very characteristic picture: Small, spindle-like cells with scant cytoplasm and granular chromatin in the nuclei (Nicholson *et al.*, 2002). Treatment is also different: SCLC require chemotherapy early, while the standard combination has basically consisted of cisplatin and etoposide for four decades, and overall prognosis is still poor and worse than in NSCLC (Pietanza *et al.*, 2015). Multiple chemotherapeutic regimens have been applied since the 1990s for NSCLC, mainly combining platins with paclitaxel, pemetrexed, targeted therapies and immune checkpoint inhibitors (Ettinger *et al.*, 2019). Several therapeutic approaches for NSCLC are subtype-specific and thus require an accurate differential diagnosis before the beginning of treatment.

The two most common NSCLC subtypes are adenocarcinoma (here termed “Adeno”, comprising 51.9% of NSCLCs) and squamous-cell carcinoma (SqCC) with 27.1% of NSCLCs. Another distinct group develops from neuroendocrine epithelium of the lung and accounts for ~7,5% of NSCLC, comprising large-cell neuroendocrine cancer (LCNEC) and pulmonary carcinoids. The remaining cases are mainly non-specified or poorly differentiated (Travis, 2010; Noone *et al.*, 2018).

Adeno and SqCC:

Multiple new chemotherapeutics have been identified for NSCLC in the past decade, mainly affecting Adeno (Neal, 2010; Zappa and Mousa, 2016). Pemetrexed and bevacizumab are not admitted for SqCC and thus are mainly used against Adeno (Johnson *et al.*, 2004; Scagliotti *et al.*, 2008). Tyrosine kinase inhibitors often depend on specific mutations, which are mainly present in Adeno (Thomas *et al.*,

2012). Due to these innovations, the prognosis of Adeno has improved most distinctively among all NSCLC subtypes (Neal, 2010).

Growing demands for precise subtyping have caused an increasing use of immunohistochemistry: The antigens TTF-1 and Napsin A are diagnostic for Adeno, p63, p40, CK5 and CK6 are diagnostic for SqCC (Travis *et al.*, 2011; Inamura, 2018). But these antigens have a limited specificity: TTF-1 is expressed in 3% of SqCCs and p63 in 32% of Adeno (Bishop *et al.*, 2011). Alien cells within tumor tissue may also express diagnostic antigens: TTF-1 and Napsin A are found on pneumocytes and Napsin A on infiltrating monocytes (Inamura, 2018). In conclusion, diagnostic accuracy in differentiating Adeno/SqCC is still not sufficient to give reliable guidance in clinical decision-making.

Neuroendocrine NSCLC subtypes:

The treatment of neuroendocrine NSCLC varies greatly with respect to the specific subtype: An early systemic treatment is recommended for LCNEC, the most aggressive subtype (Fernandez and Battafarrano, 2006). Pulmonary carcinoids are categorized into the relatively benign typical carcinoids (TC) and the more aggressive atypical carcinoids (AC) with a ratio of one to ten (Travis, 2010).

The two carcinoid subtypes are relatively rare. Their histomorphology is highly similar, but their distinction is nonetheless necessary for the choice of treatment: The standard treatment for TC is limited surgery, but AC require a more aggressive approach, often complemented with chemotherapy (Pusceddu *et al.*, 2016). The carcinoids can be distinguished relatively easy from the highly malignant LCNEC and SCLC due to their distinct histomorphology (Fisseler-Eckhoff and Demes, 2012). Two diagnostic criteria are available for differentiating TC and AC: TC have a mitosis count below 2 per $2mm^2$ high-power field, AC exhibit a count of 2-10 and AC show necrosis while TC do not (Travis, 2010).

Despite these criteria, diagnostic reproducibility is poor among experienced pathologists: One publication reported that 10% of TC were misdiagnosed as AC and that 15% of AC were misdiagnosed as TC (Travis *et al.*, 1998). This may be based on observer variances assessing the mitotic counts (Barry *et al.*, 2001) and because necrosis in AC is discrete and similar to artefacts (Rekhtman, 2010). Prognostic data indicate the potential fatality of false diagnoses: The survival rate

over five years is estimated as 87% for TC but only 56% for AC (Travis *et al.*, 1998). In order to enhance diagnostic accuracy in differentiating TC/AC, the marker KI-67, which is related to cell proliferation, has been proposed. However, the diagnostic accuracy of KI-67 is often confounded by an heterogeneous expression within tumor tissue (Pelosi *et al.*, 2017).

1.3. Diagnostic challenges through small biopsies and cytology

The previous section outlined needs and limitations of histological subtyping but referred to samples from surgical resections only. As such samples are relatively large, their morphological picture is mostly intact. However, surgery is today barely the primary source of diagnostic material. Instead, small biopsies (from bronchoscopy and CT-navigated puncture) and cytology (from pleural effusions and bronchioalveolar lavage) are becoming the main source of material for pathological investigations. The reasons for these changes are both technical availability and patient characteristics: Diagnosis is increasingly performed on patients with advanced age and multiple comorbidities, who are still offered individualized treatment which, in turn, requires a more precise characterization of disease (Yancik and Ries, 2004; Arruebo *et al.*, 2011).

Due to small size and the frequent loss of histological structures, diagnostic accuracy decreases with these samples (Travis *et al.*, 2015): Up to 30% of NSCLC in small biopsies and cytology cannot be differentiated (Travis *et al.*, 2010). The recommended approach to solve this problem is a routine use of immunohistochemistry (Travis *et al.*, 2015).

However, the accuracy of immunohistochemistry has the above-mentioned limitations with these materials: The differentiation of Adeno and SqCC is more difficult if a diagnostic sample shows a non-characteristic expression of antigens or is infiltrated by alien cells. In such a case, an interpretation cannot be obtained without a larger histomorphological picture (Gurda *et al.*, 2015). This issue is even more problematic in TC and AC, as no established profile of antigens is available for their differentiation. TC are also well vasculated and the likelihood of bleeding artefacts through puncture of vessels is high (Pusceddu *et al.*, 2016).

1.4. Relation of histological subtyping and survival

Facing diagnostic limitations, pathological registries report percentages of unspecified NSCLCs. Their diagnostic term is “not otherwise specified NSCLC” (NOS-NSCLC), which is a diagnosis of exclusion (Travis *et al.*, 2015).

The incidence of NOS-NSCLC depends on examiners and technical capabilities, which is the most likely reason for its regional variation: It ranges from 12.9% in Great Britain, where it decreases (McLean *et al.*, 2011), to 22.9% in California, where it increases (Ou and Zell, 2009). Regardless of incidence, all available data associate NOS-NSCLC with poor prognosis: The 5 year-survival rate lies between 5.8% in California (Ou and Zell, 2009) and 7.5% in Norway (Sagerup *et al.*, 2012). If compared to SCLC, which has a 5 year-survival rate of 6.5% (National Cancer Institute, SEER group, 2017), NOS-NSCLC has a very poor prognosis comparable to the most aggressive kind of lung cancer (SCLC).

Two aspects need to be explained in this context: First, the less differentiated a NSCLC is, the more likely is a non-specific diagnosis. Second, NOS-NSCLC is more common if diagnosis is based on cytology (Sagerup *et al.*, 2012). Systematic immunohistochemistry can reduce its incidence (McLean *et al.*, 2011), leaving a margin of cases where current diagnostic tools are insufficient.

1.5. Differentiating solid cancers through miRNA expression

miRNA define the characteristics of cells by regulating gene expression. In malignant transformation, their expression changes (Kong *et al.*, 2012). Most solid cancers show characteristic expression profiles, and one publication has demonstrated that miRNA profiles can distinguish 22 cancers of different primary sites (Rosenfeld *et al.*, 2008). Lung cancers develop from the same organ, but their biological heterogeneity may result in distinct expression profiles.

The miRNA mechanism of gene regulation is relatively predictable: Targeting the 3’UTR of messenger RNA (mRNA) in a sequence-specific manner, miRNA mark their targets for enzymatic degradation (Bartel, 2009). They may also interact with their targets in other ways, but this classical mechanism is the most important and allows for the prediction of targets within the transcriptome. Differences in a single

miRNA's expression can thus reveal a larger biological context (Riffo-Campos *et al.*, 2016).

In order to produce meaningful results, testing assays on miRNA expression need to minimize background noise. Most miRNA in scientific databases have not been experimentally validated. The registry "miRBase", for example, shows 1,917 entries (date: May 22th, 2020) which mainly represent miRNA-typical sequences, detected in large-scale sequencing data (Kozomara and Griffiths-Jones, 2014). It can be assumed that only a fraction of these entries has a stable expression in samples (Landgraf *et al.*, 2009; Gustafson *et al.*, 2016).

1.6. Clinical and statistical aspects within the study design

This study aims to identify differential miRNA markers for two comparisons: Adeno vs SqCC (Adeno/SqCC) and TC vs AC (TC/AC). There are other scenarios where NSCLC subtyping may be challenging as well, but these two comparisons were selected for clinical reasons: First, Adeno and SqCC are treated with distinct chemotherapeutical regimens and substances. Due to sampling techniques, the histological differentiation is often inaccurate. Immunohistochemistry helps to increase diagnostic accuracy but remains insufficient due to the mentioned reasons. Second, AC require a more aggressive treatment than TC, but the two subtypes have similar histomorphologies. Regarding TC and AC, immunohistochemistry does still not offer established tests for differentiating these two subtypes.

Adeno/SqCC and TC/AC represent, strictly speaking, two separate investigations, but this study combines them for two reasons:

- Logistics: The experimental process is identical for both.
- Statistical error reduction: Testing multiple variables (800 miRNA) in parallel induces a rate of false discoveries. The error rate will be minimal, if the analysis first examines all variables for differences between all subtypes, and then examines only the resulting significances in the comparisons of interest.

This study places high emphasis on the distinct distribution of each variable and error control, as expression count data often show high statistical dispersion. The optimal solution would be a larger number of samples, but this was impractical due

to logistical limitations. This study thus seeks to minimize these limitations through a relatively elaborate statistical approach.

It is common in biomedical research to apply statistical tests on data sets which do not meet their requirements, which results in high type I and type II error rates (Ghosh and Poisson, 2009). In this study, distributional characteristics determine for each specific miRNA which statistical test is performed. The implementation of error control consisted of hit selection as a method of quality control, and adjustment of the false discovery rate.

2. Material & Methods

2.1. Experimental process

2.1.1. Sample collection

For this study, tumor samples from the Institute of Pathology at the University Hospital Essen, University Duisburg-Essen were used. The samples were collected between 2006-2014 from patients who underwent surgery or diagnostic biopsies in the Ruhrlandklinik Essen, West German Lung Centre, University Hospital Essen, University Duisburg-Essen. All the experimental work was performed by Mr Robert Werner of the same institute. The use of these samples was approved by the Ethics Committee of the University Duisburg-Essen (ID: 17-7595-BO). A total of ten Adeno, seven SqCC, eight TC and eight AC were included in the analysis.

2.1.2. Sample deparaffinization and miRNA isolation

The samples were stored at ambient temperature as FFPE tissue. For deparaffinization, sections with a thickness of 5-10µm and a total weight of up to 10mg were washed with xylol and ethanol. The final miRNA isolation was performed using the miRNeasy FFPE Kit from Qiagen company (Hilden, Germany) following the manufacturer's instruction.

2.1.3. Digital expression analysis

The digital expression analysis of miRNA using the nCounter technique from NanoString (Seattle, United States of America) was carried out using the nCounter Human v2.1 miRNA Expression Assay from the same company. The testing panel investigated 800 miRNA per sample. Sample preparation and hybridization were performed using the automated nCounter Prep Station. Then, the cartridges

containing the miRNA hybrids were quantified using the Digital Analyzer (NanoString) measuring at 555 fields of view per sample.

2.2. Data preparation

2.2.1. Quality control through nCounter assay components

Potential sources of false measurement included sample preparation, binding density, digital imaging, and background noise.

Six positive controls formed a quality control (qc) value for each sample as depicted in the following formula:

Equation 1: Sample-specific quality control value

$$\mathrm{qc}_{\mathrm{sample}} = \frac{\text{mean of positive controls}_{\mathrm{sample}}}{\text{mean of positive controls}_{\text{all samples}}}$$

Following this equation, a $\mathrm{qc}_{\mathrm{sample}} < 0.33$ means a fold change > 3 for the respective sample relative to the others, and such a deviation would necessitate sample removal. The main reason for such deviations are irregularities in the preparation process. Also, the positive controls had to be correlated with factor ≥ 0.95 to expected concentrations, as provided by the supplier.

Six negative controls aimed to remove artefacts. Their mean value had to be below the lowest count value of any positive control, and the absolute difference had to be greater than two standard deviations of the negative controls. In order to reduce background noise, the mean value plus two standard deviations of the negative controls were subtracted from each raw count. To avoid negative values, the minimum count was set as 0.

All samples passed the quality control and were subjected to further analyses.

2.2.2. Reduction of miRNA testing panel and data normalization

After quality control, additional techniques were necessary to limit cofounding effects related to the samples and inaccuracy of the quantification.

Due to experience from previous experiments, several miRNA with no expression in any histological sample or only artefact counts were expected. To exclude these non-expressed miRNA, a minimum mean of 30 counts across all samples was

defined as a biological cut-off. All miRNA below this cut-off were deleted from further statistical analysis. This reduced the testing set from 800 to 543.

The data had to be normalized to minimize confounding effects from concentration differences in the experimental materials. First, the data were normalized through the technique „trimmed mean of M-values (TMM)". This technique not only minimizes confounding effects but also provides a uniform distribution of the expression counts in all miRNA across all samples. This is suitable for exploring the global expression of each subtype. It is, however, not appropriate for the identification of markers. For identifying the best miRNA markers, it was necessary to rule out confounding effects from the experiment but to retain the original distribution.

Trimmed mean of M-values (TMM):

This method, as designed by Robinson and Oshlack (2010), compared each individual count to a reference sample, calculating distance (termed "weight") and fold changes (termed "M"). The miRNA with the 30% highest M values represented outliers and were removed from the data set. Then, the individual count values were normalized through a statistic that comprised "weight" and "M" values of all miRNA. A normalized miRNA expression count was thus connected to the expression of all other miRNA, which transformed the individual distribution within each sample into a uniform distribution through:

Equation 2: Normalization of miRNA expression counts through TMM

$$\mathrm{miRNAcount}_{\mathrm{normalized}} \quad \frac{\mathrm{miRNAcount}}{\sqrt{\log_2 \sum_{\substack{\text{all miRNA} \\ \text{counts}}} \frac{\text{weight} + \mathrm{M}}{\text{weight}}}}$$

Normalization through NormFinder algorithm:

The algorithm NormFinder (Andersen *et al.*, 2004) assigned a stability value to each miRNA based on their expression across all 33 samples, identifying miR.140.5p, miR.28.5p and miR.361.3p as the three most stable miRNA. Their sample-specific geometric mean and the highest geometric mean among the 33 samples formed a normalization factor as shown in the following equation:

Equation 3: Normalization of miRNA expression counts through NormFinder algorithm

$$\mathrm{miRNAcount}_{\mathrm{normalized}} = \frac{\mathrm{miRNAcount}}{\mathrm{geoMean}_{\mathrm{sample}}} * \mathrm{geoMean}_{\mathrm{max}}$$

2.3. Data exploration

Cluster analysis provided a way to detect similarities in global miRNA expression between individual samples. The Euclidean distances between all the TMM-normalized counts in the 33 samples formed the basis of this analysis: The smallest distance formed the first cluster, connected two samples, and the other samples were sorted accordingly.

Each sample's global expression was graphically depicted in a heatmap, where a specific color represented the intensity of each expression count. The TMM-normalized counts were $\log_2$-scaled and transformed through:

Equation 4: Expression count transformation for cluster analysis

$$\mathrm{value}_{\mathrm{transformed}} = \frac{\mathrm{count} - \mathrm{mean\ count}_{\mathrm{column}}}{\mathrm{standard\ deviation}_{\mathrm{column}}}$$

2.4. miRNA marker discovery process

2.4.1. Hit selection by strictly standardized mean differences

As the first step of statistical analysis, the strictly standardized mean difference (SSMD) was calculated for each miRNA in both Adeno/SqCC and TC/AC. This measure refers to the ratio of the difference of the two group means divided through the square root of the sum of the two squared standard deviations. The SSMD is an effect size. While not appropriate for hypothesis testing, it is useful for error control of the subsequent analysis (Zhang *et al.*, 2007).

The set of selected hits was defined as the 5% of miRNA with the highest absolute SSMD values, containing 27 miRNA for each comparison with respect to Adeno/SqCC and TC/AC.

2.4.2. Sorting according to distributional characteristics

Two distributional criteria determined the appropriate statistical test for each miRNA: Normal vs non-normal distribution and equal vs unequal variances within the respective subtypes (homoscedasticity vs heteroscedasticity).

The Shapiro-Wilk test examined for normal distribution. It ranks all values of a group in ascending order and assigns a tabular value, thus generating a normal term $W_{critical,}$ corresponding to α=0.05. miRNA with W> $W_{critical}$ were considered to have a normal distribution. Otherwise, the distribution was considered non-normal.

Levene's test examined the expression counts for homoscedasticity. If Levene's statistic was L≤ F, where F is the critical upper quantile and corresponds to α= 0.05, then the respective miRNA's distribution was homoscedastic. If L> F, the distribution was considered heteroscedastic.

2.4.3. Distribution-adaptive hypothesis testing

Hypothesis testing for differential miRNA expression was performed in two steps: The first step examined for differential expression between all four subtypes. Significances from the first step (α= 0.05) were processed to the second step, which tested these miRNA for differences in Adeno/SqCC and TC/AC (α= 0.05). The first step intended to reduce the number of false discoveries in the later analysis, the second to finally identify the best miRNA markers.

<u>Statistical tests for all four subtypes:</u>

According to their distributional characteristics, the miRNA were assigned to statistical tests according to:

Normal+ homoscedastic	→	One-way analysis of variance (ANOVA)
Normal+ heteroscedastic	→	Welch's ANOVA
Non-normal+ homoscedastic	→	Kruskal-Wallis test
Non-normal+ heteroscedastic	→	Welch's ANOVA on ranks

Non-normal+heteroscedastic miRNA with a total 0 expression in at least one group were mathematically not compatible for Welch's ANOVA on ranks due to a group-specific variance= 0. Their variance was manually set as = 1.

Statistical tests for Adeno/SQCC and TC/AC:

A miRNA's distributional characteristics could change between the two steps: For example, a non-normal miRNA with a normal distribution in three subtypes and a non-normal distribution in one subtype was categorized as non-normal in the first step. If significant, it was processed to the second step, where its distribution became normal in one comparison and non-normal in another. Therefore, Shapiro-Wilk test and Levene's test examined the data before both the first and the second testing. Then, the miRNA were assigned according to:

Normal+ homoscedastic	→	Student's t-test
Normal+ heteroscedastic	→	Welch's t-test
Non-normal+ homoscedastic	→	Wilcoxon rank-sum test
Non-normal+ heteroscedastic	→	Welch's t-test on ranks

2.4.4. False discovery rate adjustment

As the analysis was performed on 543 miRNA in parallel, the high number of variables induced the "multiple comparisons problem". It means that a percentage of results will be significant by chance. For example, if all 543 miRNA (n= 543) had in fact the same expression in all subtypes and were tested with α= 0.05, this would result in 27 false discoveries as shown in

Equation 5: Expected false discoveries depending on the level of significance

$$\text{number of false discoveries} = \alpha \times n$$

The percentage of false positive results is termed "false discovery rate (FDR)". It is defined as the number of false discoveries divided by the number of all discoveries Several approaches are available for adjustment of the FDR. Unfortunately, the most popular approaches rely on a subjective estimation of the FDR.

In this study, the statistic "q-value" estimated the FDR through a calculation based on all p-values (Storey and Tibshirani, 2003). The more common technique of Hochberg & Benjamini (1995) then adjusted the p-values according to this estimation. This statistic also provided a graphical depiction of all results for the detection of systemic errors.

2.4.5. Calculating the markers' testing accuracy

Using "receiver operating curve (ROC)" statistics, the margin expression count was calculated for each identified miRNA marker. The margin expression count, also called "cut-off value", is the count value which differentiates the two subtypes of interest with the best sensitivity and specificity. These margin expression counts and their respective values of sensitivity and specificity were calculated with a previously determined confidence interval (CI) of 95%.

As multiple markers were identified for both Adeno/SqCC and TC/AC, a combined testing accuracy was calculated for the combination of these markers. In examining a sample, its histological subtype was determined if most of the markers were indicative of this subtype.

2.5. Computational methods

The program "R" (R development core team, 2008) was the primary environment for statistical calculations. Graphical images were produced by using the R-based program "plotly" (Plotly technologies Inc, Montréal, Canada). The "Shiny"-application DEBrowser provided the platform for data exploration through TMM-normalization and heatmapping (Kucukural *et. al*, 2018)**.** The statistic "qvalue", also based on R, estimated the FDR via a web-based application, as provided by the University of Princeton (http://qvalue. princeton.edu/, accessed January 6th, 2019). The proper FDR adjustment was performed by using an Excel spreadsheet (Microsoft corporation, Seattle, USA), provided from biostatshandbook.com (Sparky House publishing, Baltimore, USA, downloaded July 25th, 2017). The program for ROC statistics was the "Shiny" application "plotROC" (by M.C. Sachs: https://sachsmc.shinyapps. io/plotROC/, accessed November 1st, 2019).

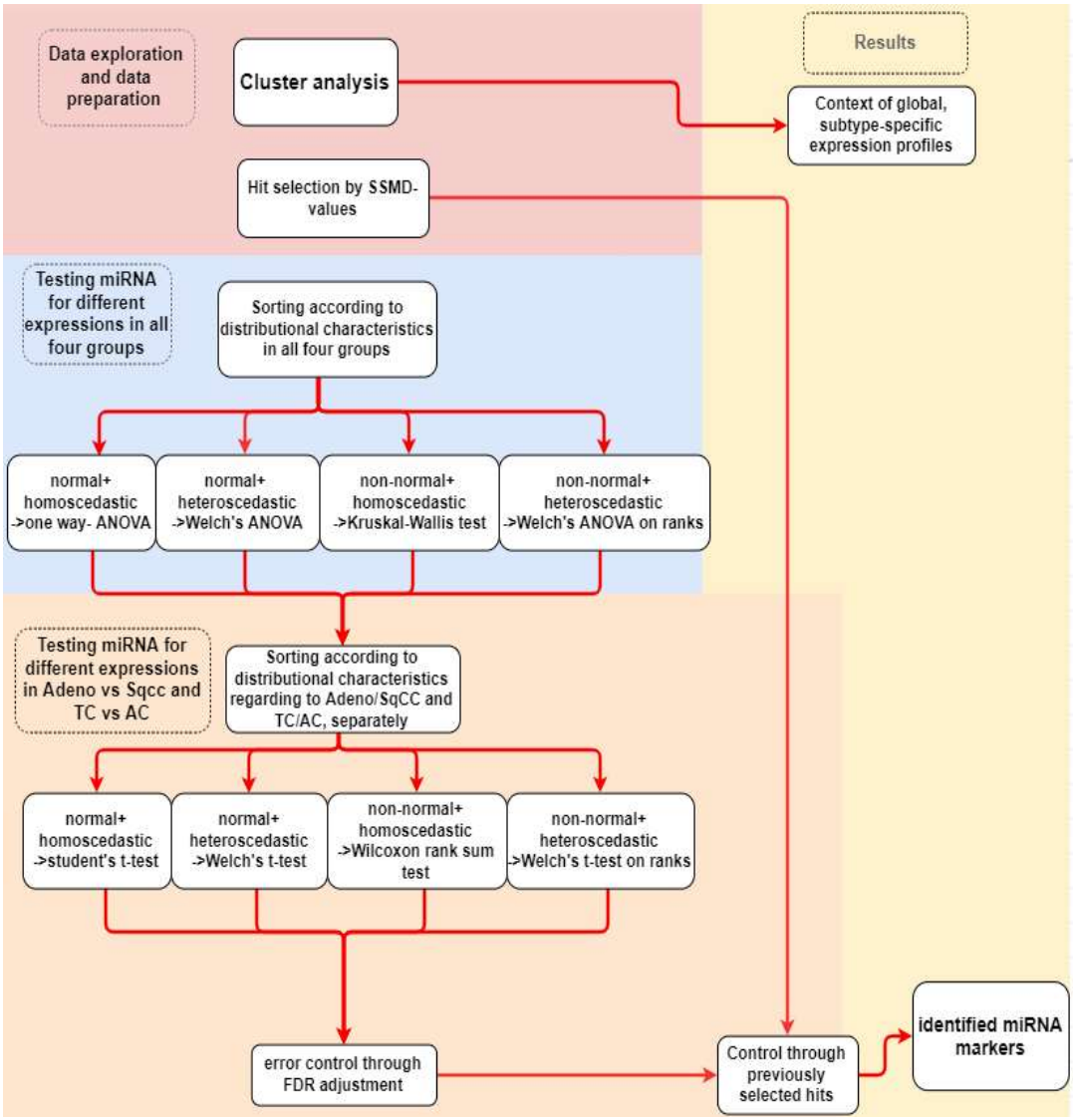

Figure 1: Workflow of the statistical analysis

- **The red field** depicts the exploration and preparation of the data. Cluster analysis (**section 2.3.)** provided an overview of the data set. Hit selection by SSMD values (**section 2.4.1.**) provided a mechanism of quality control for the later analysis
- **In the blue field**, the miRNA are tested for differences in all four subtypes according to their respective distributional characteristics (**sections 2.4.2./ 2.4.3**).
- **In the orange field**, the miRNA with significant differences between all four subtypes were again examined for their distributional characteristics and tested for differences in Adeno/SqCC and TC/AC, respectively (**section 2.4.3**). The procedures of FDR adjustment and quality control through SSMD are included.
- **The yellow field** depicts the results of the data exploration and the finally identified miRNA markers.

3. Results

3.1. Global miRNA expression profiles unable to distinguish Adeno/SqCC or TC/AC

Cluster analysis provided a global miRNA expression profile for all subtypes. These expression profiles clustered the 33 samples into two groups, the carcinoid samples, and the other samples **(Figure 2)**. Only one TC could not be separated from Adeno and SqCC. For the following identification of miRNA markers, these clusters emphasized a challenge. By means of the global miRNA expression, neither Adeno could be distinguished from SqCC nor TC from AC.

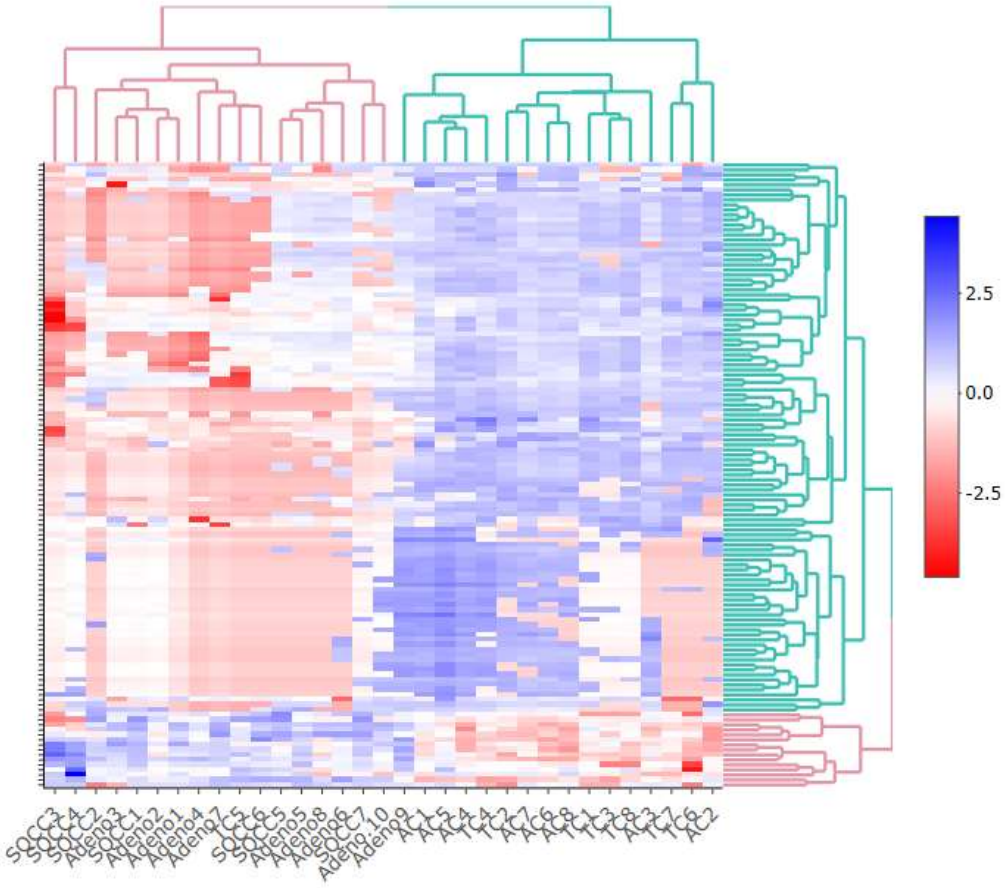

Figure 2: Cluster analysis of all four subtypes using a heatmap

The x-axis shows the 33 samples. The y axis represents the miRNA tested. Their singular annotations are omitted as they would be indistinguishable due to their large number. The colour bar to the right shows the technique of graphical depiction for the transformed count values, ranging from red for low values to blue for high values. The dendrogram at the top reflects the new arrangement of the 33 samples, the dendrogram to the right reflects the

clustering according to miRNA expression. Each dendrogram is depicted in two colors, which reflect the two principal clusters. The samples are clustered into two groups: The left cluster (coloured in red) contains SqCC and Adeno samples and one TC outlier (TC5). The right cluster (coloured in green) contains the remaining carcinoid samples. One Adeno sample (Adeno09) is within the right cluster but apart from the carcinoids.

3.2. Selected hits for both Adeno/SqCC and TC/AC

Hit selection provided an instrument to control the results of the following marker identification. According to SSMD values, 27 miRNAs with the strongest expression differences were calculated for Adeno/SqCC (**figure 3, table 1**) and TC/AC (**figure 4, table 2**). In Adeno/SqCC, the selected hits were relatively balanced between the two subtypes, with 13 overexpressed in SqCC and 14 in Adeno. In TC/AC, all but one of the selected hits were overexpressed in AC.

In the following text and tables, the depicted miRNA are named according to miRBase database. Technical annotations from the nCounter software are omitted, which accounts for decimal numbers after the second separator.

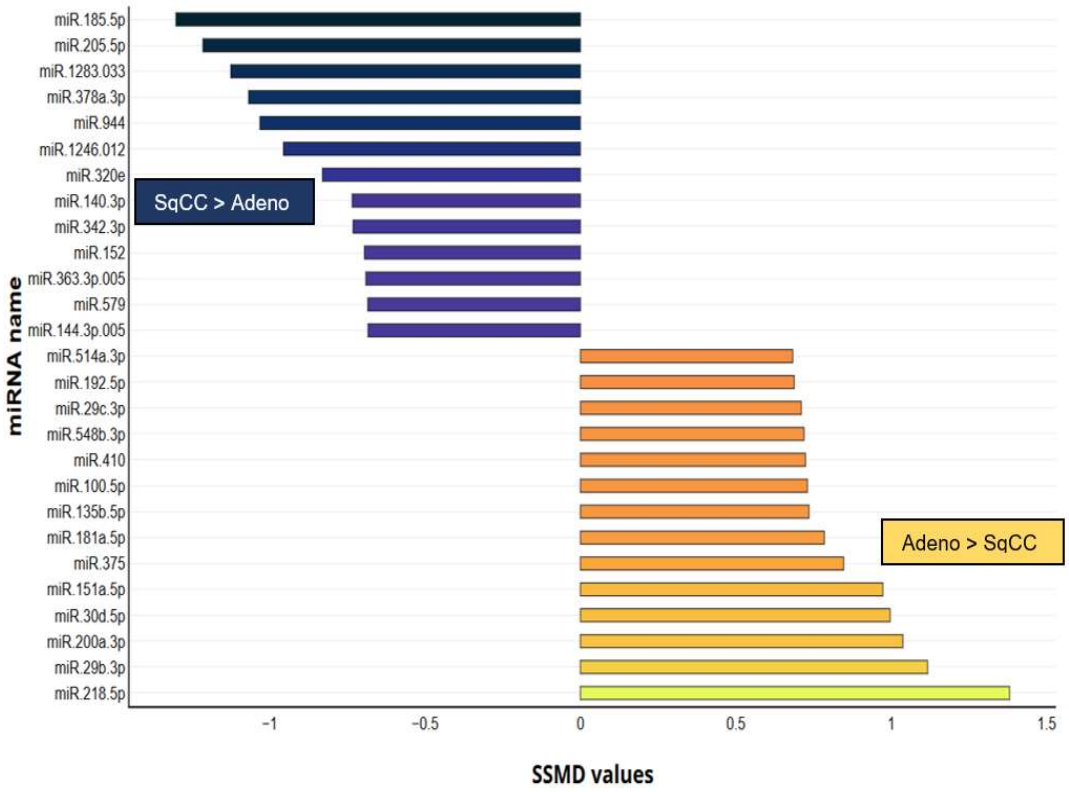

Figure 3: Selected hits for Adeno/SqCC

The x-axis depicts the SSMD values with SSMD= 0 in the center. The y-axis depicts the names of the selected 27 miRNA. SSMD stands for the difference of the mean expression counts in Adeno and SqCC, respectively, divided through the square root of the sum of the squared standard deviations in each subtype. The 27 miRNA with the highest absolute SSMD values thus represent the 5% of miRNA with the highest difference in expression between these subtypes. Positive SSMD values stand for a higher expression of miRNA in Adeno, negative values for a higher expression in SqCC.

Table 1: Selected hits for Adeno/SqCC in tabular presentation

The left column shows the miRNA, the two middle columns the respective mean expression counts of Adeno and SqCC and the right column depicts the SSMD values in ascending order.

miRNA	mean Adeno	mean SqCC	SSMD value
miR.185.5p	335.43	548.73	-1.3
miR.205.5p	1472.19	72817.34	-1.21
miR.1283	266.71	490.33	-1.12
miR.378a.3p	260.95	500.25	-1.07
miR.944	12.56	95.17	-1.03
miR.1246	549.15	9805.48	-0.96
miR.320e	396.79	758.61	-0.83
miR.342.3p	5121.21	8674.74	-0.73
miR.140.3p	57.31	123.73	-0.73
miR.152	344.17	611.54	-0.7
miR.363.3p	321.57	524.6	-0.69
miR.144.3p	1004.47	1697.79	-0.68
miR.579	161.02	221.65	-0.68
miR.514a.3p	99.8	12.24	0.68
miR.192.5p	366.2	164.27	0.69
miR.29c.3p	2634.02	1317.79	0.71
miR.410	101.48	14.93	0.72
miR.548b.3p	74.16	0	0.72
miR.100.5p	3739.94	2045.36	0.73
miR.135b.5p	4808.77	2153.92	0.74
miR.181a.5p	5219.22	2975.46	0.78
miR.375	1689.8	215.43	0.85
miR.151a.5p	328.56	183.11	0.97
miR.30d.5p	2965.72	1355.78	1
miR.200a.3p	4958.7	2105.11	1.04
miR.29b.3p	42696.07	19958.39	1.12
miR.218.5p	593.18	287.87	1.38

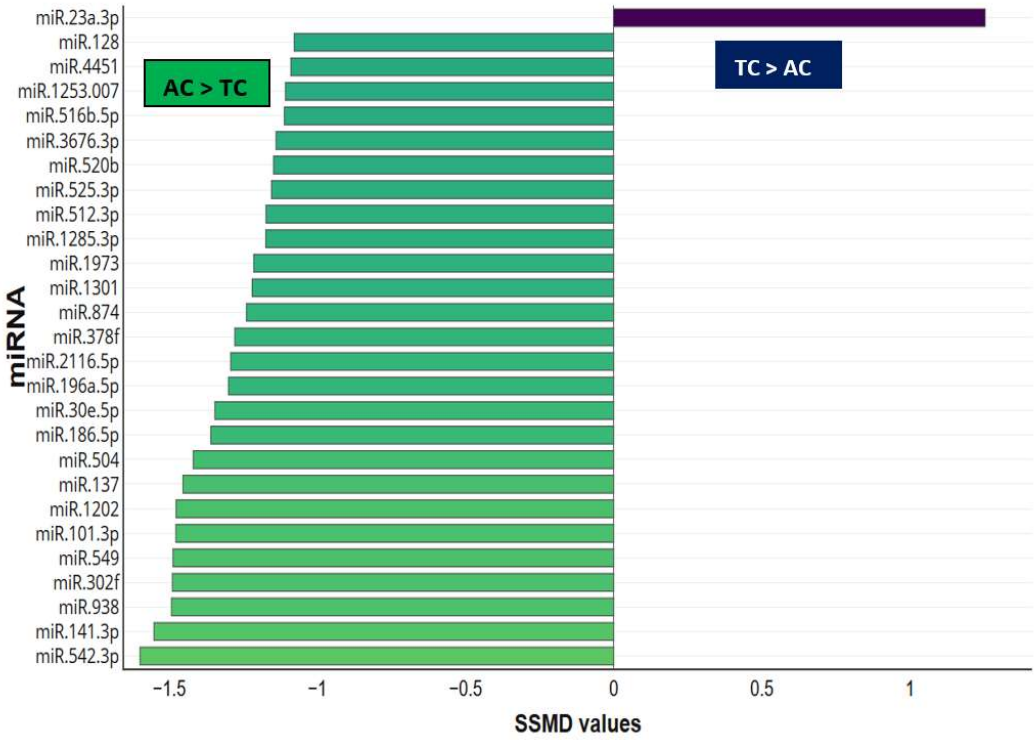

Figure 4: Selected hits for TC/AC

The x-axis depicts the SSMD values with SSMD= 0 in the center. The y-axis depicts the names of the selected 27 miRNA. SSMD stands for the difference of the mean expression counts in TC and AC, respectively, divided through the square root of the sum of the squared standard deviations in each subtype. The 27 miRNA with the highest absolute SSMD values thus represent the 5% of miRNA with the highest expression difference in these two subtypes. Positive SSMD values stand for a higher expression of miRNA in TC, negative values for a higher expression in AC.

Table 2: Selected hits for TC/AC in tabular presentation

The left column shows the miRNA, the two middle columns the respective mean expression counts of TC and AC, and the right column depicts the SSMD values in ascending order.

miRNA	mean TC	mean AC	SSMD value
miR.542.3p	10.73	132.65	-1.60
miR.141.3p	5597.85	12400.33	-1.55
miR.938	7.57	83.60	-1.49
miR.302f	28.37	140.53	-1.49
miR.549	0.00	126.55	-1.49
miR.101.3p	49.40	165.02	-1.48
miR.1202	6.94	103.39	-1.48
miR.137	3363.76	7321.56	-1.45
miR.504	9.47	172.09	-1.42
miR.186.5p	309.92	756.21	-1.36
miR.30e.5p	1789.57	4094.67	-1.35
miR.196a.5p	26.15	141.68	-1.30
miR.2116.5p	48.79	182.12	-1.29
miR.378f	6.94	92.39	-1.28
miR.874	6.94	125.82	-1.24
miR.1301	7.57	95.73	-1.22
miR.1973	31.06	144.70	-1.21
miR.1285.3p	7.57	89.34	-1.17
miR.512.3p	0.00	87.47	-1.17
miR.525.3p	32.71	105.11	-1.16
miR.520b	7.57	109.24	-1.15
miR.3676.3p	24.65	115.67	-1.14
miR.516b.5p	6.94	122.59	-1.11
miR.1253	165.02	438.48	-1.11
miR.4451	0.00	55.08	-1.09
miR.128	174.54	509.28	-1.08
miR.23a.3p	16655.40	9529.30	1.25

3.3. Differential expression with respect to all four subtypes

According to their distributional characteristics, the 543 miRNA were tested for differential expression in all four subtypes. At α= 0.05, a considerable number of 111 miRNA were tested as significant.

3.4. miRNA markers for Adeno/SqCC

Statistical testing in Adeno/SqCC reported three miRNA (miR.1246, miR.218.5p and miR.375) with significant p-values after FDR adjustment (α= 0.05, FDR= 0.12) and quality control. miR.1246 had a higher expression in SqCC, while miR.218.5p and miR.375 had a higher expression in Adeno (**Table 3, Figure 5**). For subtype differentiation with adequate measures of sensitivity and specificity, expression count cut-off values with CI= 95% were calculated (**Table 4**). In combination, these markers achieved high testing accuracy: If at least two of these markers were indicative for a specific subtype, all samples were identified correctly (**table 5**).

Table 3: Identified miRNA markers for Adeno/SqCC

The three markers are sorted according to their p-values, from lowest to highest in descending order. Due to the non-normal distribution of these markers, they are depicted through their median values, as appropriate for variables with these distributional characteristics. miR.1246 has a higher expression in SqCC while miR.218.5p and miR.375 have a higher expression in Adeno.

miRNA	p-value	median Adeno	median SqCC
miR.1246	<0.0001	371	4826
miR.218.5p	0.0001	559	244
miR.375	0.0006	724	198

Table 4: Calculated cut-off values discriminating Adeno/SqCC at CI= 95%

Through ROC statistics, appropriate cutt-off expression counts were calculated at CI= 95% for the differentiation of Adeno/SqCC. Each cut-off value infers distinct likelihoods to assign a given sample correctly to its subtype (sensitivity) and to rule out the other subtype (specificity).

miRNA	Cut-off with CI= 95%	sensitivity	specificity
miR.1246	1204	85%	100%
miR.218.5p	394	90%	100%
miR.375	496	90%	100%

Table 5: Combined testing accuracy of the three miRNA markers for Adeno/SqCC

This table shows the capacity of the three markers miR.1246, miR.218.5p and miR.375 to differentiate the Adeno and SqCC subtypes. If at least two markers were indicative of either Adeno or SqCC, according to the previously calculated cut-off values **(table 4)**, the respective sample was identified accordingly. Through this approach, all 17 samples (Adeno+ SqCC) were identified correctly, resulting in a discriminative accuracy of 100%.

	Adeno	SqCC	Total
≥2/3 markers indicative of Adeno	10	0	10
≥2/3 markers indicative of SqCC	0	7	7
Total	10	7	17

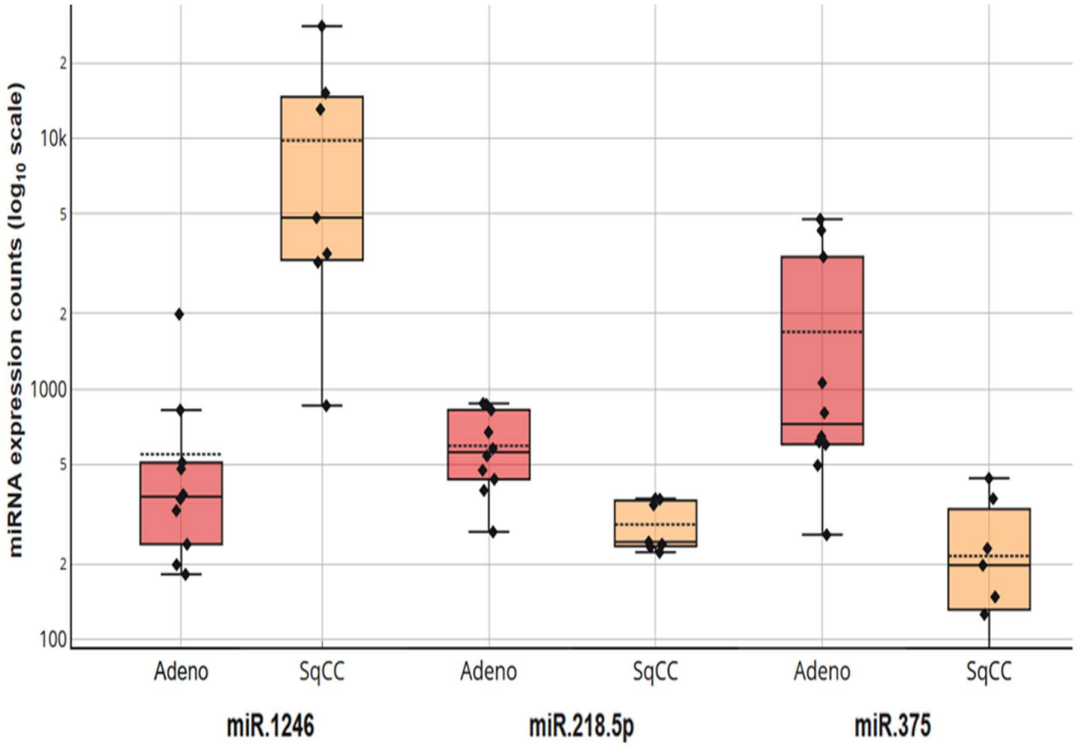

Figure 5: miRNA markers for Adeno/SqCC subtyping

These boxplots represent the three identified miRNA markers for Adeno/SqCC subtyping. Due to the large range of expression counts, they are shown on the y-axis in logarithmic scale (log_{10}-scaled). The red boxes represent Adeno, the orange boxes depict SqCC. The thick or bold lines in each box represent the median values, the dashed lines depict the mean values. The plots contain the values between the first and third quartile (termed "interquartile range" or "ICR"). The upper whiskers contain the first quartile minus 1.5× ICR, the lower whiskers the third quartile minus 1.5× ICR.

3.5. miRNA markers for TC/AC

The analysis of TC/AC resulted in six miRNA markers after application of appropriate statistics and error control. 110 miRNA were tested, 16 were significant at α= 0.05 and the FDR was estimated as FDR= 0.2497. Six miRNA remained after application of the above-mentioned statistical analysis and quality control (**Table 6**). All markers had a higher expression in AC compared to TC (**figures 6,7**). Using ROC statistics, expression count margins with appropriate sensitivity and specificity were calculated at CI=95% (**Table 7**). These six markers achieved a discriminative accuracy of 100% with a threshold of ≥ 4 markers being indicative of one subtype (**table 8**).

Table 6: Identified miRNA markers for TC/AC

The three markers are sorted according to their p-values, from lowest to highest value in descending order. The first two markers had a non-normal distribution, which resulted in the depiction of all six markers through their median values.

miRNA	p-value	median TC	median AC
miR.1202	0.0007	0	99
miR.549a	0.0007	0	115
miR.141.3p	0.0020	5544	14296
miR.137	0.0031	3563	8223
miR.1253	0.0073	153	405
miR.128	0.0117	177	527

Table 7: Calculated cut-off values discriminating TC/AC at CI= 95%

Through ROC statistics, appropriate cut-off expressions counts were calculated for the differentiation of TC/AC at CI= 95%. Each cut-off value infers a distinct likelihoods to assign a given sample correctly to its subtype (sensitivity) and to rule out the other subtype (specificity).

miRNA	Cut-off with CI= 95%	Sensitivity	specificity
miR.1202	60	87%	100%
miR.549a	80	87%	100%
miR.141.3p	7005	85%	87%
miR.137	4095	75%	87%
miR.1253	247	85%	87%
miR.128	377	75%	87%

Table 8: Combined testing accuracy of the three miRNA markers for TC/AC

The six miRNA markers miR.1202, miR.549a, miR.141.3p, miR.137, miR.1253 and miR.128 differentiated the TC and AC subtypes. If ≥ 4 markers were indicative of either TC or AC, the respective sample was determined as TC or AC. All samples were correctly identified, showing a discriminating accuracy of 100%.

	TC	AC	Total
≥2/3 markers indicative of TC	8	0	8
≥2/3 markers indicative of AC	0	8	8
Total	8	8	16

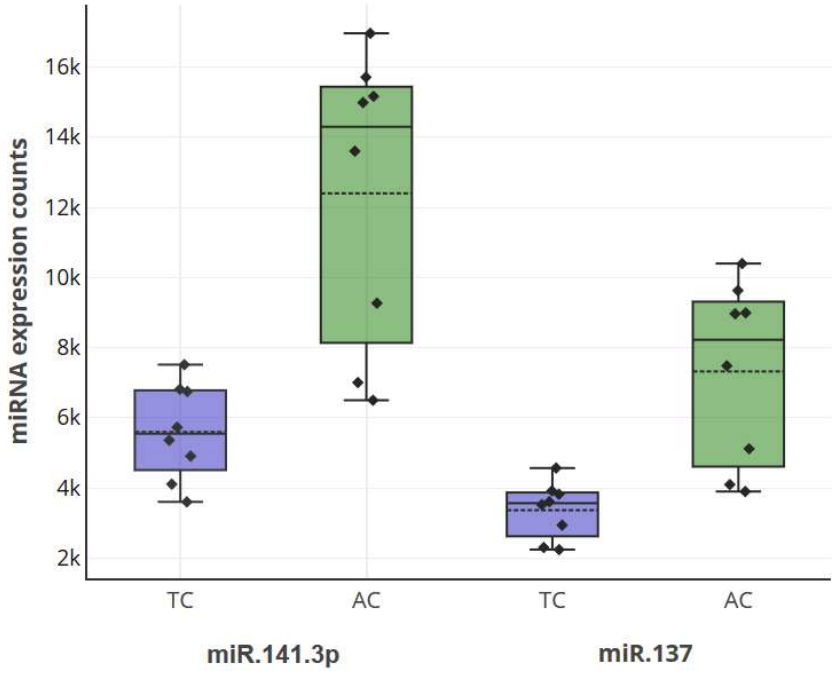

Figure 6: miRNA markers miR.141.3p and miR.137 for TC/AC subtyping

These boxplots represent two out of six identified miRNA markers for TC/SAC subtyping. The expression counts are shown on the y-axis. The blue boxes represent Adeno, the green boxes depict AC. The thick or bold lines in each box represent the median values, the dashed lines depict the mean values. The plots contain the values between the first and third quartile (termed “interquartile range” or “ICR”). The upper whiskers contain the first quartile minus 1.5× ICR, the lower whiskers the third quartile minus 1.5× ICR.

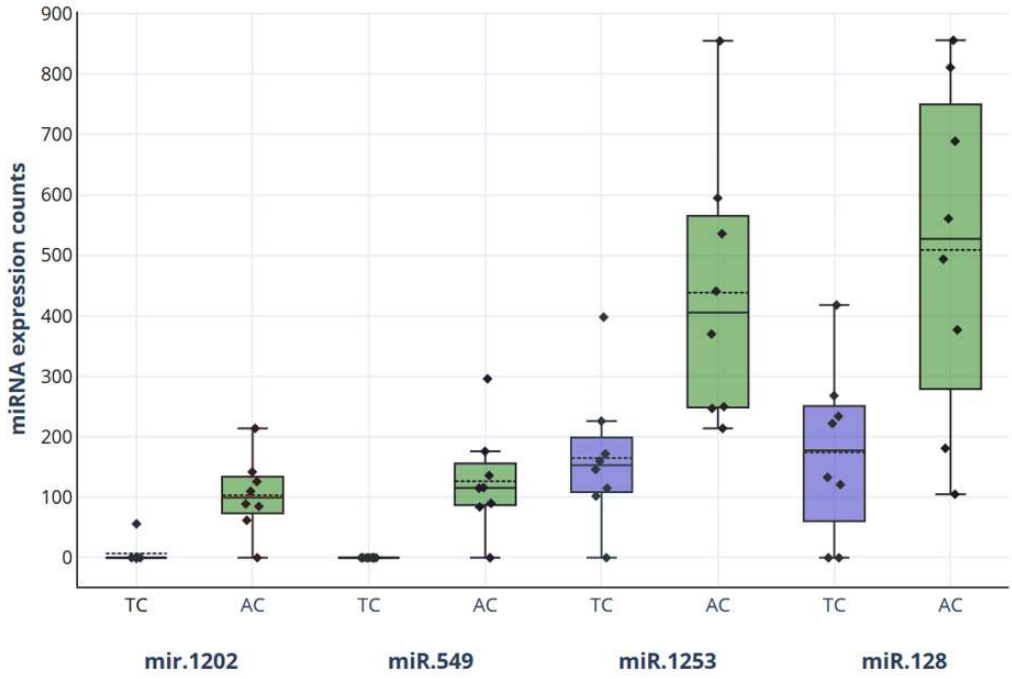

Figure 7: miRNA markers miR.1202, miR.549a, miR.1253 and miR.128 for TC/AC subtyping

These boxplots represent four out of six identified miRNA markers for TC/AC subtyping. The expression counts are shown on the y-axis. The blue boxes represent TC, the green boxes depict AC. The thick or bold lines in each box represent the median values, the dashed lines depict the mean values. The plots contain the values between the first and third quartile (termed “interquartile range” or “ICR”). The upper whiskers contain the first quartile minus 1.5× ICR, the lower whiskers the third quartile minus 1.5× ICR.

4. Discussion

4.1. Identified markers and global expression profiles

This study identified three miRNA markers for subtyping Adeno/SqCC (miR.1246, miR.218.5p, miR.375) and six markers for TC/AC (miR.1202, miR.549, miR.141.3p, miR.137, miR.1253, miR.128). With respect to Adeno/SqCC, miR.1246 had a higher expression in SqCC, and miR.218.5p and miR.375 a higher expression in Adeno. All six markers for TC/AC had a higher expression in AC.

The global miRNA expression profiles of all four subtypes could neither distinguish the more malign AC from the more benign TC nor Adeno from SqCC. It was unsurprising that the carcinoid subtypes and the non-carcinoid subtypes formed two separate groups, given their largely distinct biology. But, as Adeno and SqCC develop through distinct pathogenic patterns (Campbell *et al.*, 2016), the lack of separate clustering was contrary to expectations. Also, separate clustering of the more malignant AC and the more benign TC had been expected, as aggressive neoplasms often show a deranged expression (Lin and Gregory, 2015).

4.2. Remarks on the statistical analysis

This study combined several statistical methods in an effort to maximize testing accuracy. While this combination is relatively unconventional in this field of biomedical research, it addressed common issues: The number of variables in this data set vastly exceeded the number of samples, limiting testing power. In almost any measurement of biological expression values, data present strong statistical dispersion with non-normal and unequal distributions (Barton *et al.*, 2013). Instead of accounting for the distributional characteristics of each individual variable, it is common to apply indiscriminate statistical tests (Malo *et al.*, 2006). Several popular tests like student's t-test, Wilcoxon rank-sum test, one-way ANOVA and Kruskal-Wallis-test also require statistical homoscedasticity, but this requirement is frequently not taken into account (Barton *et al.*, 2013). Instead of choosing statistical tests according to characteristics of the data, researchers often transform them through, for example, log-scaling. But transformation effectively moves the data away from their native state and may render results non-reproducible (O'Hara and Kotze, 2010).

In this study's discovery process, a distribution-adaptive approach was performed instead of an indiscriminate statistic for all miRNA. A not very common feature is the qvalue statistic for error adjustment, which balances type I and type II errors better than a subjective estimation of the false discovery rate.

4.3. Markers for Adeno/SqCC subtyping

4.3.1. Roles in subtype-specific carcinogenesis

Experimental findings can be validated through their biological context. Indeed, the three miRNA markers for Adeno/SqCC showed links to key oncogenic mechanisms specific for both NSCLC subtypes (**Figure 8**).

A higher percentage of SqCC (59%) than Adeno (12%) contain activating mutations of the PI3K/AKT/mTOR pathway. Likewise, an inactivation of the tumor suppressive factor TP53 is found mainly in SqCC (80%), less in Adeno (50%). Inversely, activating mutations of tyrosine receptor kinases (TRKs) are more frequent in Adeno (50%) than in SqCC (27%) (Shtivelman *et al.*, 2014). Adeno also has high expressions of the molecule CADM1 relative to SqCC (Kitamura *et al.*, 2009). CADM1 is tumor suppressive in both subtypes, but apparently through different mechanisms (Vallath *et al.*, 2016).

The marker miR.1246, which is overexpressed in SqCC, targets CADM1 (Sun *et al.*, 2014). As CADM1 has a low expression in SqCC, their relation shows a subtype-specific oncogenic role of miR.1246. The remaining markers miR.218 and miR.375 have a higher expression in Adeno. In SqCC, they appear to suppress cancer growth: According to Kumamoto et al. (2016), artificial induction of miR.218 in SqCC effectively stalls proliferation of the tumor cells. miR.218 is encoded in a region of genomic losses typical for SqCC, resulting in depletion of this miRNA (Davidson *et al.*, 2010). It does apparently not affect the proliferation of Adeno, which can thus tolerate a stronger concentration of this marker. The second diagnostic marker for Adeno, miR.375, suppresses the PI3K/AKT/mTOR pathway (Yan *et al*, 2014), which is more active in SqCC than in Adeno, making the downregulation of miR.375 necessary for SqCC growth. No similar role is known in Adeno, which implies that this subtype can also tolerate a higher expression of miR.375.

4.3.2. Comparison to previous research

Due to the clinical importance of Adeno/SqCC subtyping, previous researchers have already investigated for differential miRNA marker expression. Two publications stand out through considerable resources, and this section discusses their findings:

Lebanony *et al.*, (2009) reported miR.205 as diagnostic for SqCC and, like this study, miR.375 as diagnostic for Adeno. They quantified 141 miRNA through a microarray applied on 62 cases of Adeno and 60 cases of SqCC, and afterwards, they validated their findings through real-time PCR with 20 NSCLC samples. For statistical analysis, they used student's t-test and applied error control through Bonferroni's correction (lowering the level of significance from $p<0.05$ to $p<0.00035$).

Hamamoto *et al.*, (2013) reported miR.205 and miR.196b as markers for SqCC and miR.375 as marker for Adeno. Their sample set contained 54 cases of Adeno and 25 cases of SqCC, and the expression of 377 miRNA in these samples was quantified through a microarray. Then, they validated their findings through real-time PCR with 44 samples each of Adeno and SqCC, performing statistical analysis through student's t-test without error control, combining it with a specific fold change for significance.

Both these publications reported miR.205 as diagnostic for SqCC and miR.375 as diagnostic for Adeno. This study also identified miR.375 but not miR.205, which also fits into the biological context through targeting PTEN, an inhibitor of the PI3K/AKT/mTor pathway in SqCC (Cai *et al.*, 2013). From a technical perspective, these previous publications had stronger sample sizes. From a statistical perspective, their use of student's t-test was less than optimal for expression data. miR.205 was non-significant in this study, possibly due to an aberrant expression in the smaller number of samples. Neither Lebanony et al. nor Hamamoto et al. reported miR.1246 or miR.218, perhaps missing them through inadequate statistics. **Figure 8** on the following page shows the markers miR.1246, miR.375, miR.218 and also miR.205 within the context of SqCC carcinogenesis.

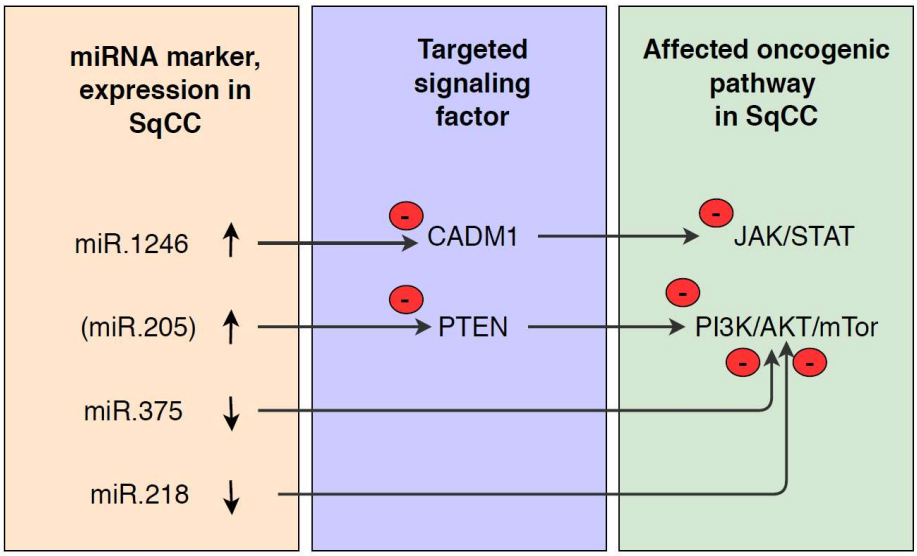

Figure 8: miRNA markers for Adeno/SqCC in relation to SqCC carcinogenesis

All three discovered miRNA markers for Adeno/SqCC (miR.1246, miR.375 and miR.218) have either direct or indirect relationships to molecular pathways driving cancer proliferation in SqCC. These relationships help to explain their relative expressions in these two subtypes.

In brackets, miR.205 is likewise depicted. This miRNA war tested as non-differential in this analysis, but featured prominently in the works of Lebanony *et al.* (2009) and Hamamoto *et al.*(2013). miR.1246 and, according to the mentioned previous reports, miR.205 are overexpressed in SqCC. They target tumor suppressive factors, thus contributing to proliferation of SqCC: miR.1246 targets CADM1 (Sun *et al.*, 2014), which would otherwise suppress the JAK/STAT pathway (Vallath *et al.*, 2016). miR.205 targets PTEN (Cai *et al.*, 2013), which would otherwise suppress the PI3K/AKT/mTor signaling pathway. miR.375 and miR.218 are downregulated in SqCC. They would otherwise downregulate the PI3K/AKT/mTor pathway (Yan *et al.*, 2014; Kumamoto *et al.*, 2016).

4.4. Markers for TC/AC subtyping

4.4.1. The identified markers as oncogenic factors

The global expression profiles depicted TC and AC as two closely related entities. The main biological difference is the more aggressive behavior of AC. A histological characteristic of AC is necrosis, which occurs in hypoxic tissue during malignant proliferation (Proskuryakov and Gabai, 2010). All identified markers had a higher

expression in AC, and they can at best be interpreted in the context of oncogenic dedifferentiation.

The markers miR.141.3p and miR.137 are both associated with lung cancer: Blood of NSCLC patients has a higher concentration of miR.141.3p compared to smokers without cancer (Nadal *et al.*, 2015). Both markers predict relapse and poor prognosis for NSCLC patients (Yu *et al.*, 2008; Zhang *et al.*, 2015). Considered alone, miR.141.3p has an ambiguous role in general carcinogenesis: It belongs to the tumor suppressive miR.200 molecular family (Gregory *et al.*, 2008) but also has an oncogenic role, targeting the tumor suppressor YAP1 (Imanaka *et al.*, 2011) which is depleted in high-grade neuroendocrine lung cancers (Ito *et al.*, 2016). Like miR.141.3p, the marker miR.137 was in several instances described as tumor suppressor, but it is also responsive to hypoxia and prevents autophagy in fast-proliferating tumors (Li *et al.*, 2014), thereby facilitating necrosis (Su *et al.*, 2015), a characteristic feature of AC.

Out of the remaining four markers, miR.1202 and miR.128 are in instances oncogenic: miR.1202 inhibits apoptosis in endometrial cancer (Chen *et al.*, 2017) and is elevated in both MALT lymphoma (Thorns *et al.*, 2012) and colon cancer (Hamfjord *et al.*, 2012). Similarly, miR.128 is also an inhibitor of apoptosis in T-cell leukemia (Yamada *et al.*, 2014). For the remaining two markers, miR.549 and miR.1253, the scientific reports are insufficient for interpretation.

4.4.2. Carcinoids in the context of neuroendocrine lung cancers

To the author's knowledge, this is currently the only study on differences in miRNA expression between the two carcinoid subtypes outside of the context of other neuroendocrine lung cancers. This is apparently due to the carcinoids' low incidence.

In the only available previous publication on miRNA expression differences between TC and AC, Rapa *et al.* (2015) quantified 752 miRNA in 6 samples of TC and 6 samples of AC. By using student's t-test and Wilcoxon rank-sum test, they reported four miRNA with a higher expression in AC and 20 miRNA with a higher expression in TC. These results were then transferred onto the highly aggressive LCNEC and SCLC by quantification of these miRNAs' expressions in a second sample set,

containing TC, AC, LCNEC and SCLC. There, the selected miRNA had distinct expression levels in the high-grade cancers but were statistically indistinguishable between the two carcinoid subtypes. In conclusion, it appears that this work was performed on two vastly different groups of neuroendocrine subtypes. The pathogenesis of the pulmonary carcinoids appears to have little in common with LCNEC and SCLC, with little association to smoking history in the carcinoids (Swarts *et al.*, 2012) and distinct patterns of genetic mutations (Armengol and Kaur, 2015). Further research into neuroendocrine lung cancers should thus treat the carcinoids separately. Otherwise, the highly aggressive biology of SCLC and LCNEC may overshadow the discrete differences between AC and TC.

4.5. Limitations and perspectives

4.5.1. Experimental validation

The reliability of experimental discoveries depends on study design, experimental techniques, and statistical methods. As the two previous publications concerning miRNA markers for Adeno/SqCC (Lebanony *et al.*, 2009; Hamamoto *et al.*, 2013) demonstrated, this study's sample sizes were below those of comparable experiments but also had different methodological features: Lebanony et al. used a microarray based on the MicroGrid II assay (Genomic Solutions Inc., Ann-Arbor, USA), Hamamoto et al. used the TaqMan human microRNA microarray V2.0 (Applied Biosystems Inc., Foster City, USA). This study quantified miRNA expression through NanoString technology. Different experimental techniques can cause different results: In a comparative analysis, the measurements of NanoString and TaqMan showed varying measurements, ranging from -0.75 and to +0.75 in $\log_2$-scale (Malkov *et al.*, 2009). Different normalization methods can also impair the comparison of studies, as Lebanony et al. normalized through median values, Hamamoto et al. through the z-score and this study through the three most stable miRNA.

Given the methodological differences, it will be necessary to validate the findings on Adeno/SqCC and TC/AC through a larger sample set and different experimental techniques. This study's data-adaptive statistical analysis should be retained, as it took several recommendations from theoretical biostatistics into account.

4.5.2. Practical issues for diagnostic application

Clinical use of miRNA panels already takes place in diagnosing thyroid cancer (Nishino and Nikiforova, 2018). Most diagnostic material for thyroid cancer is derived from fine needle biopsies, where histomorphological structures are mainly not intact. Currently, no routine miRNA expression application exists for the differentiation of different types of cancer from the same organ. While the case of thyroid cancer has proven that a diagnostic application is technically possible, the miRNA markers for Adeno/SqCC and TC/AC still face several practical issues:

1) Differential diagnosis of LCNEC in poorly differentiated NOS-NSCLC:

   Minimizing the rate of NOS-NSCLC is a primary intention of this study. Given the high percentages of Adeno and SqCC among all NSCLC, they are the two most likely differential diagnosis for NOS-NSCLC. But poorly differentiated NOS-NSCLC, which are especially difficult for subtyping, often exhibit large cells resembling LCNEC (Fasano *et al.*, 2015). It will thus be necessary to exclude LCNEC as a differential diagnosis either through immunohistochemistry or an additional miRNA marker panel.

2) Different benign pulmonary cells as confounders:

   Most diagnostic samples contain varying amounts of non-cancer cells that express their own distinct miRNA profiles. At best, miRNA content within a sample should be extracted from the cancer components only. This is relatively easy in surgical samples, but difficult in small biopsy-samples and cytological material. Only if the expressions of the diagnostic markers in healthy tissue are available, it will be possible to minimize confounding effects from benign cells.

   It can be assumed that each type of pulmonary tissue has to some degree a specific miRNA profile: Squamous epithelium in the upper airways may express the diagnostic markers for SqCC, and glandular cells the diagnostic markers for Adeno. To this day, the available expression data only refer to lung tissue in general (Ludwig *et al.*, 2016), but for diagnostic application of the markers in cytological material and small biopsy samples, these tissue-specific expressions are needed.

3) Effects from bleeding artefacts:

Bleeding artefacts frequently occur in well-vasculated tumors during biopsy, which mainly concerns TC (Fisseler-Eckhoff and Demes, 2012). Blood components, especially white blood cells, may alter a sample's miRNA expression and confound the concentrations of diagnostic markers (Pritchard *et al.*, 2011). Such confounding effects from bleeding artefacts are difficult to determine, as the markers' concentrations within these artefacts depend on its cellular composition. Thus, Samples with large bleeding artefacts do not qualify for diagnostic use of the miRNA markers.

4) Reference values for cytological material:

The samples in this study came from surgical resections and biopsies, but the greatest need for molecular markers lies in cytological fluids from broncho-alveolar lavages or pleural effusions. Such material will present different ranges of marker concentration (da Cunha Santos *et al.*, 2018). Thus, different reference values, taking total miRNA content into account, need to be established for each diagnostic material.

5) Different miRNA expression in never-smokers:

Nicotine consumption alters miRNA expression in lung tissue (Momi *et al.*, 2014), and NSCLC of never-smokers may exhibit different expressions of diagnostic markers. This issue concerns the miRNA marker panel for Adeno/SqCC, as nicotine consumption only has a small effect on the carcinoids. While the incidences of Adeno and SqCC are low in never-smokers, it will be difficult to quantify the marker expression in these patient groups. As long as these expressions are not available, samples from never-smokers should be treated with adequate caution.

4.5.3. Diagnostic application in other sample materials

Sputum is another type of cytological material, which can be obtained without invasive techniques. miRNA markers can be applied on this material and thereby help to diagnose patients which are not eligible for other sampling-techniques. Xing *et al.*, (2015) have already demonstrated the technical possibility to diagnose lung cancer from sputum but did not differentiate between histological subtypes.

Blood components are also a promising diagnostic material and can be gained with a minimum of invasiveness. The current interest in blood as diagnostic material is expressed in the term “liquid biopsy”, which mainly addresses the potential of circulating nucleotides, both RNA and DNA, for diagnosing and monitoring cancer (Perakis and Speicher, 2017). Circulating miRNA in blood are biochemically more stable than messenger RNA or DNA (Gustafson *et al.*, 2016). Cancer cells actively secrete miRNA in exosomes, perhaps creating a more proliferation-friendly environment in this way. Thus, Circulating miRNA in exosomes are not merely the product of dying cancer cells (Rabinowits *et al.*, 2009), but may reflect a tumor’s dynamic behavior and hence help to differentiate and monitor neoplastic diseases (Kosaka *et al*., 2010).

4.6. Conclusion

This study profiled the miRNA expression of four NSCLC subtypes, Adeno, SqCC, TC and AC, to discover markers differentiating Adeno from SqCC and TC from AC. The global expression profiles were considerably similar between Adeno and SqCC and also between TC and AC. Neither did the distinct histological background of Adeno and SqCC lead to subtype-specific expression clusters nor did the differences in aggressiveness between TC and AC.

The analysis discovered three markers for Adeno/SqCC. All three fit into the specific oncogenic mechanisms in these NSCLC subtypes, but only one corresponds to previous publications on the same issue. A direct comparison of this study’s findings to these reports is not possible due to different experimental and statistical techniques. Six markers were discovered for TC/AC subtyping, all with a higher expression in AC. An oncogenic role of these markers is possible, but as the knowledge of the carcinoids’ biology is relatively small, their functional connections to specific oncogenic pathways remain unclear. To the author’s knowledge, this is the first study to investigate for differential miRNA expression in the carcinoids outside of the context of other neuroendocrine lung cancers.

To transfer the markers into diagnostic application, it will be necessary to validate their expression values and to examine technical confounders in diagnostic materials. These practical issues do not require additional research into the

fundamental biology of lung cancer and can be undertaken through conventional methods. Failure to differentiate NSCLC subtypes is likely to lead to inadequate treatment, causing poor prognosis. These two miRNA marker panels may provide clinical benefit by supporting therapeutic decisions based on reliable NSCLC subtype identification.

5. Summary

Background: Within the group of non-small cell lung cancers (NSCLC), the number of clinically relevant subtypes is increasing. At the same time, diagnostic tissue is more often sampled through minimally invasive procedures, making accurate subtyping difficult. microRNA (miRNA) molecules may be able to satisfy the need for molecular markers in NSCLC subtyping.

Rationale: This study aims to identify miRNA markers for NSCLC subtyping. It is necessary to differentiate adenocarcinoma (Adeno) and squamous cell carcinoma (SqCC) because several therapeutics agents are only admitted for Adeno, not for SqCC. For his purpose, miRNA markers may provide diagnostic support. miRNA markers may also help to distinguish typical carcinoid (TC) and atypical carcinoid (AC), as the therapeutic approach is more aggressive for AC than for TC.

Methods: In 33 NSCLC samples (ten Adeno, seven SqCC, eight TC, eight AC), the expression of 800 miRNA was quantified through the nCounter technology from NanoString. The data were quality-controlled, normalized and the removal of artefacts reduced the panel size to 543 miRNA. Through cluster calculations, the data were explored. The subsequent analysis was designed as a data-adaptive approach: Each miRNA was examined for its distributional properties through the Shapiro-Wilk test (for normality) and Levene's test (for equal distribution) and accordingly categorized. Statistical testing for differences between all four groups was performed through one-way analysis of variance (ANOVA), Welch's ANOVA, the Kruskal-Wallis test, and Welch's ANOVA on ranks (all at $\alpha = 0.05$). The significances were examined for differences in the two comparisons of interest through student's t-test, Welch's t-test, the Wilcoxon rank-sum test, and Welch's t-test on ranks (all at $\alpha = 0.05$). The false discovery rate was controlled, and the results had to pass a final quality control.

Results and conclusion: Three miRNA markers were identified for Adeno/SqCC: miR.1246, miR.218.5p and miR.375, which fit into current knowledge of their biological role in lung cancers. For TC/AC subtyping, six markers were identified: miR.1202, miR.549a, miR.141.3p, miR.137, miR.1253 and miR.128. They are the first miRNA markers investigated specifically for differentiating the pulmonary carcinoids.

6. References

1. Allemani, C., Matsuda, T., Di Carlo, V., Harewood, R., Matz, M., Nikšić, M., Bonaventure, A., Valkov, M., Johnson, C. J., Estève, J., Ogunbiyi, O. J., Azevedo E Silva, G., Chen, W. Q., Eser, S., Engholm, G., Stiller, C. A., Monnereau, A., Woods, R. R., Visser, O., Lim, G. H., … CONCORD Working Group (2018): Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): Analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet. 391(20125), 1023–1075.
2. Andersen, C. L., Jensen, L. J. and Ørntoft, T. F. (2004): Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and Colon Cancer Data Sets. Cancer Res. 64(15), 5245-50.
3. Armengol, G. and Kaur, V. (2015): Driver Gene Mutations of Non-Small-Cell Lung Cancer are Rare in Primary Carcinoids of the Lung: NGS Study by Ion Torrent. Lung. 193(2), 303-8.
4. Arruebo, M., Vilaboa, N., Sáez-Gutierrez, B., Lambea, J., Tres, A., Valladares, M., González-Fernández, A. (2011): Assessment of the evolution of cancer treatment therapies. Cancers (Basel). 3(3), 3279-330.
5. Barry, M., Sinha, S.K., Leader, M.B., Kay E.W. (2001): Poor agreement in recognition of abnormal mitoses: requirement for standardized and robust definitions. Histopathology. 38(1), 68–72.
6. Bartel, D. P. (2009): MicroRNA Target Recognition and Regulatory Functions. Cell. 136(2), 215–233.
7. Barton, S.J., Crozier, S.R., Lillycrop, K.A., Inskip, H.M. (2013): Correction of unexpected distributions of P values from analysis of whole genome arrays by rectifying violation of statistical assumptions. BMC Genomics. 14(161).
8. Benjamini, Y. and Hochberg, Y. (1995): Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. Royal Stat. Soc. 57(1): 289–300.

9. Bishop, J.A., Teruya-Feldstein, J., Westra, W.H., Pelosi, G., Travis, W., Rekhtman, N. (2011): P40 (D Np63) is superior to p63 for the diagnosis of pulmonary squamous cell carcinoma. Mod Pathol. 25(3), 405–415.
10. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A (2018): Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 68(6), 394-424.
11. Cai, J., Fang, L., Huang, Y., Li, R., Yuan, J., Yang, Y., Zhu, X., Chen, B., Wu, J., Li, M. (2013): MiR-205 targets PTEN and PHLPP2 to augment AKT signaling and drive malignant phenotypes in non-small cell lung cancer. Cancer Res. 73(17), 5402–5415.
12. Campbell, J.D., Alexandrov, A., Kim, J., Wala, J., Berger, A.H., Pedamallu C.S., Shukla S.A., Guo, G., Brooks, A.N., Murray, B.A., Imielinski, M., Hu, X., Ling, S., Akbani, R., Rosenberg, M., Cibulskis, C., Ramachandran, A., Collisson, E.A., Kwiatkowski, D.J., Lawrence, M.S., Weinstein, J.N., Verhaak, R.G., Wu, C.J., Hammerman, P.S., Cherniack, A.D., Getz, G.; Cancer Genome Atlas Research Network, Artyomov MN, Schreiber R, Govindan R, Meyerson M. (2016): Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. Nat Genet. 48(6): 607-16.
13. Chau, H., Rixe, O., McLeod, H., Figg, W.D.(2010): Validation of Analytical Methods for Biomarkers Employed in Drug Development. Cancer. 14(19), 5967–5976.
14. Chen, H., Fan, Y., Xu, W., Chen, J., Meng, Y., Fang, D., & Wang, J. (2017): Exploration of miR-1202 and miR-196a in human endometrial cancer based on high throughout gene screening analysis. Oncol Rep. 37(6), 3493–3501.
15. da Cunha Santos, G., Saieg, M.A., Troncone, G., Zeppa, P. (2018): Cytological preparations for molecular analysis: A review of technical procedures, advantages and limitations for referring samples for testing. Cytopathology. 29(2), 125–132.
16. Davidson, M. R., Larsen, J. E., Yang, I. A., Hayward, N. K., Clarke, B. E., Duhig, E. E., Passmore, L. H., Bowman, R. V., & Fong, K. M. (2010): MicroRNA-218 is deleted and downregulated in lung squamous cell carcinoma. PloS One. 5(9), e12560.

17. Ettinger, D. S., Wood, D. E., Aggarwal, C., Aisner, D. L., Akerley, W., Bauman, J. R., Bharat, A., Bruno, D. S., Chang, J. Y., Chirieac, L. R., D'Amico, T. A., Dilling, T. J., Dobelbower, M., Gettinger, S., Govindan, R., Gubens, M. A., Hennon, M., Horn, L., Lackner, R. P., Lanuti, M., … Hughes, M. (2019): NCCN Guidelines Insights: Non-Small Cell Lung Cancer, Version 1.2020. J Natl Compr Canc. 17(12), 1464–1472.
18. Fasano, M., Della Corte, C. M., Papaccio, F., Ciardiello, F., & Morgillo, F. (2015): Pulmonary Large-Cell Neuroendocrine Carcinoma: From Epidemiology to Therapy. J Thorac Oncol. 10(8), 1133–1141.
19. Fernandez, F. G. and Battafarrano, R. J. (2006): Large-Cell Neuroendocrine Carcinoma of the Lung: An Aggressive Neuroendocrine Lung Cancer. Semin Thorac Cardiovasc Surg. 18(3), 206–210.
20. Fisseler-Eckhoff, A. and Demes, M. (2012): Neuroendocrine tumors of the lung. Cancers (Basel). 4(3), 777–798.
21. Ghosh, D. and Poisson, L. M. (2009): 'Omics' data and levels of evidence for biomarker discovery. Genomics. 93(1), 13–16.
22. Gregory, P.A., Bert, A.G., Paterson, E.L., Barry, S.C. (2008): The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nat Cell Biol. 10(5), 593–601.
23. Gurda, G. T., Zhang, L., Wang, Y., Chen, L., Geddes, S., Cho, W. C., Askin, F., Gabrielson, E., & Li, Q. K. (2015): Utility of five commonly used immunohistochemical markers TTF-1, Napsin A, CK7, CK5/6 and P63 in primary and metastatic adenocarcinoma and squamous cell carcinoma of the lung: a retrospective study of 246 fine needle aspiration cases. Clin Transl Med. 4, 16.
24. Gustafson, D., Tyryshkin, K. and Renwick, N. (2016): MicroRNA-guided diagnostics in clinical samples. Best Pract Res Clin Endocrinol Metab. 30(5), 563–575.
25. Hall, J. S., Taylor, J., Valentine, H. R., Irlam, J. J., Eustace, A., Hoskin, P. J., Miller, C. J., & West, C. M. (2012): Enhanced stability of microRNA expression facilitates classification of FFPE tumour samples exhibiting near total mRNA degradation. Br J Cancer. 107(4), 684–694.
26. Hamamoto, J., Soejima, K., Yoda, S., Naoki, K., Nakayama, S., Satomi, R., Terai, H., Ikemura, S., Sato, T., Yasuda, H., Hayashi, Y., Sakamoto, M.,

Takebayashi, T., & Betsuyaku, T. (2013): Identification of microRNAs differentially expressed between lung squamous cell carcinoma and lung adenocarcinoma. Mol Med Rep. 8(2), 456–462.

27. Hamfjord, J., Stangeland, A. M., Hughes, T., Skrede, M. L., Tveit, K. M., Ikdahl, T., & Kure, E. H. (2012): Differential expression of miRNAs in colorectal cancer: comparison of paired tumor tissue and adjacent normal mucosa using high-throughput sequencing. PloS One. 7(4), e34150.
28. Holland, R. L. (2016): What makes a good biomarker? Adv Precis Med. 1(1), 66.
29. Imanaka, Y., Tsuchiya, S., Sato, F. Yukako Imanaka, Soken Tsuchiya, Fumiaki Sato, Shimada, Y., Shimizu, K., Tsujimoto, G. (2011): MicroRNA-141 confers resistance to cisplatin-induced apoptosis by targeting YAP1 in human esophageal squamous cell carcinoma. J Hum Genet. 56, 270–276.
30. Inamura, K. (2018): Update on immunohistochemistry for the diagnosis of lung cancer. Cancers (Basel). 10(3), 1–15.
31. Ito, T., Matsubara, D., Tanaka, I., Makiya, K., Tanei, Z. I., Kumagai, Y., Shiu, S. J., Nakaoka, H. J., Ishikawa, S., Isagawa, T., Morikawa, T., Shinozaki-Ushiku, A., Goto, Y., Nakano, T., Tsuchiya, T., Tsubochi, H., Komura, D., Aburatani, H., Dobashi, Y., Nakajima, J., … Murakami, Y. (2016): Loss of YAP1 defines neuroendocrine differentiation of lung tumors. Cancer Sci. 107(10), 1527–1538.
32. Johnson, D., Fehrenbacher, L., Novotny, W., Herbst, R., Nemunaitis, J, Jablons, D., Langer, C., DeVore, R,, Gaudreault, J., Damico, L., Holmgren, E., Kabbinavar, F. (2004): Randomized Phase II Trial Comparing Bevacizumab Plus Carboplatin and Paclitaxel With Carboplatin and Paclitaxel Alone in Previously Untreated Locally Advanced or Metastatic Non-Small-Cell Lung Cancer. J Clin Oncol. 22(11). 2184-91.
33. Kitamura, Y., Kurosawa, G., Tanaka, M., Sumitomo, M., Muramatsu, C., Eguchi, K., Akahori, Y., Iba, Y., Tsuda, H., Sugiura, M., Hattori, Y., & Kurosawa, Y. (2009): Frequent overexpression of CADM1/IGSF4 in lung adenocarcinoma. Biochem Biophys Res Commun. 383(4), 480–484.
34. Kong, Y.W., Ferland-McCollough, D., Jackson, T.J., Bushell, M. (2012): MicroRNAs in cancer management. Lancet Oncol. 13(6), e249–e258.

35. Kosaka, N., Iguchi, H. and Ochiya, T. (2010): Circulating microRNA in body fluid: A new potential biomarker for cancer diagnosis and prognosis. Cancer Sci. 101(10), 2087–2092.
36. Kozomara, A. and Griffiths-Jones, S. (2014): MiRBase: Annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 42(D1), 68–73.
37. Kucukural, A., Yukselen, O., Ozata, D.M., Moore, M.J., Garber, M. (2018): DEBrowser: interactive differential expression analysis and visualization tool for count data. BMC Genomics. 20(1), 6.
38. Kumamoto, T., Seki, N., Mataki, H., Mizuno, K., Kamikawaji, K., Samukawa, T., Koshizuka, K., Goto, Y., & Inoue, H. (2016): Regulation of TPD52 by antitumor microRNA-218 suppresses cancer cell migration and invasion in lung squamous cell carcinoma. Int J Oncol. 49(5), 1870–1880.
39. Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A. O., Landthaler, M., Lin, C., Socci, N. D., Hermida, L., Fulci, V., Chiaretti, S., Foà, R., Schliwka, J., Fuchs, U., Novosel, A., Müller, R. U.,Tuschl, T. (2007): A mammalian microRNA expression atlas based on small RNA library sequencing. Cell. 129(7), 1401–1414.
40. Lebanony, D., Benjamin, H., Gilad, S., Ezagouri, M., Dov, A., Ashkenazi, K., Gefen, N., Izraeli, S., Rechavi, G., Pass, H., Nonaka, D., Li, J., Spector, Y., Rosenfeld, N., Chajut, A., Cohen, D., Aharonov, R., Mansukhani, M. (2009): Diagnostic assay based on hsa-miR-205 expression distinguishes squamous from nonsquamous non-small-cell lung carcinoma. Journal of Clin Oncol. 27(12), 2030–2037.
41. Li, W., Zhang, X., Zhuang, H., Chen, H. G., Chen, Y., Tian, W., Wu, W., Li, Y., Wang, S., Zhang, L., Chen, Y., Li, L., Zhao, B., Sui, S., Hu, Z., & Feng, D. (2014): MicroRNA-137 is a novel hypoxia-responsive microRNA that inhibits mitophagy via regulation of two mitophagy receptors FUNDC1 and NIX. J Biol Chem. 289(15), 10691–10701.
42. Lin, S. and Gregory, R. I. (2015): MicroRNA biogenesis pathways in cancer. Nat Rev Cancer. 15(6), 321–333.
43. Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., Downing, J.R.,

Jacks, T., Horvitz, H.R., Golub, T.R. (2005): MicroRNA expression profiles classify human cancers. Nature. 435(7043), 834–838.

44. Ludwig, N., Leidinger, P., Becker, K., Backes, C., Fehlmann, T., Pallasch, C., Rheinheimer, S., Meder, B., Stähler, C., Meese, E., & Keller, A. (2016): Distribution of miRNA expression across human tissues. Nucleic Acids Res. 44(8), 3865–3877.
45. Malkov, V. A., Serikawa, K. A., Balantac, N., Watters, J., Geiss, G., Mashadi-Hossein, A., & Fare, T. (2009): Multiplexed measurements of gene signatures in different analytes using the Nanostring nCounter Assay System. BMC Res Notes. 2, 80.
46. Malo, N., Hanley, J.A., Cerquozzi, S., Pelletier, J., Nadon, R. (2006): Statistical practice in high-throughput screening data analysis. Nat Biotechnol. 24(2), 167–175.
47. McLean, E.C., Monaghan, H., Salter, D.M., Wallace, W.A. (2011): Evaluation of adjunct immunohistochemistry on reporting patterns of non-small cell lung carcinoma diagnosed histologically in a regional pathology centre. J Clin Pathol. 64(12), 1136–1138.
48. Momi, N., Kaur, S., Rachagani, S., Ganti, A. K., & Batra, S. K. (2014): Smoking and microRNA dysregulation: a cancerous combination. Trends Mol. Med. 20(1), 36–47.
49. Nadal, E., Truini, A., Nakata, A., Lin, J., Reddy, R., Chang, A., Ramnath, N., Goto, N., Chen, G. (2015): A Novel Serum 4-microRNA Signature for Lung Cancer Detection. Sci Rep. 5(12464).
50. Neal, J. W. (2010): Histology matters: Individualizing treatment in non-small cell lung cancer. Oncologist. 15, 3–5.
51. Nicholson, S.A., Beasley, M.B., Brambilla, E., Hasleton, P.S., Colby, T.V., Sheppard, M.N., Falk, R., Travis, W.D. (2002): Small Cell Lung Carcinoma (SCLC): A Clinicopathologic Study of 100 Cases with Surgical Specimens. Am J Surg Pathol. 26(9), 1184–1197.
52. Nishino, M. and Nikiforova, M. (2018): Update on molecular testing for cytologically indeterminate thyroid nodules. Arch Pathol Lab Med. 142(4), 446–457.
53. Noone, A.M., Howlader, N., Krapcho, M., Miller, D., Brest, A. Yu, M. Ruhl, J., Tatalovich, Zn. Mariotto, A. Lewis D.R., Chen, H.S., Feuer, E.J., Cronin,

K.A. (2018): SEER Cancer Statistics Review, 1975-2015, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/csr/1975_2015/, based on November 2017 SEER data submission, posted to the SEER web site, April 2018.

54. O'Hara, R. B. and Kotze, D. J. (2010): Do not log-transform count data. Methods Ecol. Evol. 1(2), 118–122.
55. Ou, S.-H. I. and Zell, J. A. (2009): Carcinoma NOS is a common histologic diagnosis and is increasing in proportion among non-small cell lung cancer histologies. J Thorac Oncol. 4(10), 1202–11.
56. Patel, P. G., Selvarajah, S., Guérard, K. P., Bartlett, J., Lapointe, J., Berman, D. M., Okello, J., & Park, P. C. (2017): Reliability and performance of commercial RNA and DNA extraction kits for FFPE tissue cores. PloS One. 12(6), e0179732.
57. Pelosi, G., Sonzogni, A., Harari, S., Albini, A., Bresaola, E., Marchiò, C., Massa, F., Righi, L., Gatti, G., Papanikolaou, N., Vijayvergia, N., Calabrese, F., & Papotti, M. (2017): Classification of pulmonary neuroendocrine tumors: new insights. Transl Lung Cancer Res. 6(5), 513–529.
58. Perakis, S. and Speicher, M. R. (2017): Emerging concepts in liquid biopsies. BMC Med. 15(1), 1–12.
59. Pietanza, M. C., Byers, L. A., Minna, J. D., & Rudin, C. M. (2015): Small cell lung cancer: will recent progress lead to improved outcomes? Clin Cancer Res. 21(10), 2244-2255.
60. Pritchard, C. C., Kroh, E., Wood, B., Arroyo, J. D., Dougherty, K. J., Miyaji, M. M., Tait, J. F., & Tewari, M. (2012): Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. Cancer Prev Res (Phila). 5(3), 492–497.
61. Proskuryakov, S. and Gabai, V. (2010): Mechanisms of Tumor Cell Necrosis. Curr Pharm Des. 16(1), 56–68.
62. Pusceddu, S., Lo Russo, G., Macerelli, M., Proto, C., Vitali, M., Signorelli, D., Ganzinelli, M., Scanagatta, P., Duranti, L., Trama, A., Buzzoni, R., Pelosi, G., Pastorino, U., de Braud, F., Garassino, M.C. (2016): Diagnosis and management of typical and atypical lung carcinoids. Crit Rev Oncol Hematol. 100, 167–176.

63. R Core Team (2013): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
64. Rabinowits, G., Gerçel-Taylor, C., Day, J.M., Taylor, D.D., Kloecker, G.H. (2009): Exosomal microRNA: a diagnostic marker for lung cancer. Clin Lung Cancer. 10, 42–46
65. Rapa, I., Votta, A., Felice, B., Righi, L., Giorcelli, J., Scarpa, A., Speel, E.J., Scagliotti, G.V., Papotti, M., Volante, M. (2015): Identification of MicroRNAs Differentially Expressed in Lung Carcinoid Subtypes and Progression. Neuroendocrinology. 101(3), 246–255.
66. Rekhtman, N. (2010): Neuroendocrine Tumors of the Lung, an update. Achives of Pathology and Laboratory Medicine. 134, 1628–1638.
67. Riffo-Campos, Á. L., Riquelme, I. and Brebi-Mieville, P. (2016): Tools for sequence-based miRNA target prediction: What to choose? Int J Mol Sci. 17(12), 1987.
68. Robinson, M. D. and Oshlack, A. (2010): A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11(3), R25.
69. Rosenfeld, N., Aharonov, R., Meiri, E., Rosenwald, S., Spector, Y., Zepeniuk, M., Benjamin, H., Shabes, N., Tabak, S., Levy, A., Lebanony, D., Goren, Y., Silberschein, E., Targan, N., Ben-Ari, A., Gilad, S., Sion-Vardy, N., Tobar, A., Feinmesser, M., Kharenko, O., Nativ, O., Nass, D., Perelman, M, Yosepovich, A., Shalmon, B., Polak-Charcon, S., Fridman, E., Avniel, A., Bentwich, I.,, Bentwich, Z., Cohen, D., Chajut, A., Barshack, I. (2008): MicroRNAs accurately identify cancer tissue origin. Nat Biotechnol. 26(4), 462–469.
70. Sagerup, C.M., Småstuen, M., Johannesen, T.B., Helland, A., Brustugun O.T. (2012): Increasing age and carcinoma not otherwise specified: A 20-year population study of 40,118 lung cancer patients. J Thorac Oncol. 7(1), 57–63.
71. Scagliotti, G.V., Parikh, P., von Pawel, J., Biesma, B., Vansteenkiste, J., Manegold, C., Serwatowski, P., Gatzemeier, U., Digumarti, R., Zukin, M., Lee, J.S., Mellemgaard, A., Park, K., Patil, S., Rolski, J., Goksel, T., de Marinis, F., Simms, L., Sugarman, K.P., Gandara, D. (2008): Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed

in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. J Clin Oncol. 26(21), 3543–3551.

72. Shtivelman, E., Hensing, T., Simon, G. R., Dennis, P. A., Otterson, G. A., Bueno, R., & Salgia, R. (2014): Molecular pathways and therapeutic targets in lung cancer. Oncotarget. 5(6), 1392–1433.
73. Storey, J. D. and Tibshirani, R. (2003): Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 100(16), 9440–9445.
74. Su, Z., Yang, Z., Xu, Y., Chen, Y., & Yu, Q. (2015): MicroRNAs in apoptosis, autophagy and necroptosis. Oncotarget. 6(11), 8474–8490.
75. Sun, Z., Meng, C., Wang, S., Zhou, N., Guan, M., Bai, C., Lu, S., Han, Q., & Zhao, R. C. (2014): MicroRNA-1246 enhances migration and invasion through CADM1 in hepatocellular carcinoma. BMC cancer, 14, 616.
76. Swarts, D. R. a, Ramaekers, F. C. S. and Speel, E. J. M. (2012): Molecular and cellular biology of neuroendocrine lung tumors: Evidence for separate biological entities. Biochim Biophys Acta. 1826(2), 255–271.
77. Thomas, A., Rajan, A. and Giaccone, G. (2012): Tyrosine Kinase Inhibitors in Lung Cancer. Hematol Oncol Clin North Am. 26(3), 589–605.
78. Thorns, C., Kuba, J., Bernard, V., Senft, A., Szymczak, S., Feller, A. C., & Bernd, H. W. (2012): Deregulation of a distinct set of microRNAs is associated with transformation of gastritis into MALT lymphoma. Virchows Arch. 460. 371–377.
79. Travis W., Gal, A., Colby, T., Klimstra, D., Ralk, R. (1998): Reproducibility of Neuroendocrine Lung Tumor Classification. Hum Pathol. 23(3), 272–279.
80. Travis, W. D., Rekhtman, N., Riley, G. J., Geisinger, K. R., Asamura, H., Brambilla, E., Garg, K., Hirsch, F. R., Noguchi, M., Powell, C. A., Rusch, V. W., Scagliotti, G., & Yatabe, Y. (2010): Editorial: Pathologic diagnosis of advanced lung cancer based on small biopsies and cytology: A paradigm shift. J Thorac Oncol. 5(4), 411–414.
81. Travis, W. D., Rush, W., Flieder, D. B., Falk, R., Fleming, M. V., Gal, A. A., & Koss, M. N. (1998): Survival analysis of 200 pulmonary neuroendocrine tumors with clarification of criteria for atypical carcinoid and its separation from typical carcinoid. Am J Surg Pathol. 22(8), 934–44.
82. Travis, W. D. (2010): Advances in neuroendocrine lung tumors. Ann Oncol. 21(SUPPL. 7), 65–71.

83. Travis, W. D., Brambilla, E., Nicholson, A. G., Yatabe, Y., Austin, J., Beasley, M. B., Chirieac, L. R., Dacic, S., Duhig, E., Flieder, D. B., Geisinger, K., Hirsch, F. R., Ishikawa, Y., Kerr, K. M., Noguchi, M., Pelosi, G., Powell, C. A., Tsao, M. S., Wistuba, I., & WHO Panel (2015): The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. J Thorac Oncol. 10(9), 1243–1260.
84. Travis, W. D., Brambilla, E., Noguchi, M., Nicholson, A. G., Geisinger, K. R., Yatabe, Y., Beer, D. G., Powell, C. A., Riely, G. J., Van Schil, P. E., Garg, K., Austin, J. H., Asamura, H., Rusch, V. W., Hirsch, F. R., Scagliotti, G., Mitsudomi, T., Huber, D. (2011): International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma. J Thorac Oncol. 6(2), 244–285.
85. True, L. D. (2014): Methodological requirements for valid tissue-based biomarker studies that can be used in clinical practice. Virchows Arch. 464(3), 257–263.
86. Tsongalis, G. J. and Silverman, L. M. (2006): Molecular diagnostics: A historical perspective. Clin Chim Acta. 369(2), 188–192.
87. Vallath, S., Sage, E. K., Kolluri, K. K., Lourenco, S. N., Teixeira, V. S., Chimalapati, S., George, P. J., Janes, S. M., & Giangreco, A. (2016): CADM1 inhibits squamous cell carcinoma progression by reducing STAT3 activity. Sci Rep. 6, 24006.
88. Walker, R. A. (2006): Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment I. Histopathology. 49(4), 406–410.
89. Xing, L., Su, J., Guarnera, M. A., Zhang, H., Cai, L., Zhou, R., Stass, S. A., & Jiang, F. (2015): Sputum microRNA biomarkers for identifying lung cancer in indeterminate solitary pulmonary nodules. Clin Cancer Res. 21(2), 484–489.
90. Yamada, N., Noguchi, S., Kumazaki, M., Shinohara, H., Miki, K., Naoe, T., Akao, Y. (2014): Epigenetic regulation of microRNA-128a expression contributes to the apoptosis-resistance of human T-cell leukaemia Jurkat

cells by modulating expression of Fas-associated protein with death domain (FADD). Biochim Biophys Acta. 1843(3), 590–602.

91. Yan, J.-W., Lin, J.-S. and He, X.-X. (2014): The emerging role of miR-375 in cancer. Int J Cancer. 135(5), 1011–1018.
92. Yancik, R. and Ries, L. A. G. (2004): Cancer in older persons: an international issue in an aging world. Semin Oncol. 31(2), 128–136.
93. Yu, S. L., Chen, H. Y., Chang, G. C., Chen, C. Y., Chen, H. W., Singh, S., Cheng, C. L., Yu, C. J., Lee, Y. C., Chen, H. S., Su, T. J., Chiang, C. C., Li, H. N., Hong, Q. S., Su, H. Y., Chen, C. C., Chen, W. J., Liu, C. C., Chan, W. K., Chen, W. J., … Yang, P. C (2008): MicroRNA Signature Predicts Survival and Relapse in Lung Cancer. Cancer Cell. 13(1), 48-57.
94. Zappa, C. and Mousa, S. A. (2016): Non-small cell lung cancer: current treatment and future advances. Transl Lung Cancer Res. 5(3), 288–300.
95. Zhang, X., Li, P., Rong, M., He, R., Hou, X., Xie, Y., Chen, G. (2015): MicroRNA-141 Is a Biomarker for Progression of Squamous Cell Carcinoma and Adenocarcinoma of the Lung: Clinical Analysis of 125 Patients. Tohoku J Exp Med. 235(3), 161-9.
96. Zhang, X.D., Ferrer, M., Espeseth, A.S., Marine, S.D., Stec, E.M., Crackower, M.A., Holder, D.J,, Heyse, J.F., Strulovici, B. (2007): The use of strictly standardized mean difference for hit selection in primary RNA interference high-throughput screening experiments. J Biomol Screen. 12(4), 497–509.

7. Appendix

7.1. List of abbreviations

Abbreviation	Full term
AC	Atypical carcinoid of the lung
Adeno	Adenocarcinoma of the lung
AKT = PKB	Protein kinase B
ANOVA	Analysis of variance
CADM1	Cell adhesion molecule 1
CI	Confidence interval
CT	Computer tomography
FDR	False discovery rate
FFPE	Formalin-fixed, paraffin-embedded
JAK	Janus kinase
LCNEC	Large-cell neuroendocrine cancer of the lung
MALT	Mucosa- associated lymphoid tissue
miRNA	microRNA
mTOR	Mammalian target of rapamycin
NOS-NSCLC	Not otherwise specified NSCLC
NSCLC	Non-small-cell lung cancer
PI3K	Phosphatidylinositol 3-kinase
PTEN	Phosphatase and tensin homolog
qc_{sample}	Quality control value of a specific sample
ROC	Receiver operating curve
SCLC	Small cell lung cancer
SEER	Surveillance, Epidemiology, and End Results
SqCC	Squamous-cell carcinoma of the lung
SSMD	Strictly standardized mean difference
STAT	Signal transducer and activator of transcription
TC	Typical carcinoid of the lung
TMM	Trimmed mean of M-values
TP53	Tumor protein 53
TTF-1	Thyroid transcription factor 1
YAP1	Yes-associated protein 1
3'UTR	Three prime untranslated region (of messangerRNA)

7.2. List of figures, tables, and equations

Graphical figures:

Tables:

Equations:

8. Thesis acknowledgements

Several people helped me to carry through with this thesis. To them, I would like to express my gratitude.

First, I am thankful to Mr. Professor Kurt Werner Schmid for his patience and several instances of valuable professional advice.

My tutor Mr. Doctor Robert Walter helped me to find structure in my scientific topic, he corrected my work in methodological detail and greatly enhanced my understanding of scientific work. He not only demanded but also demonstrated impressive commitment.

All my analysis is based on experimental work performed by Mr. Robert Werner, who profiled the miRNA expression in an impressive amount of lung cancer samples at the Ruhrlandklinik, West German Lung Center, University Hospital Essen. He formulated the initial ideas of investigation. My thesis started from the basis which he had established.

Mr. Professor Heinz Jöckel (Institut für Medizinische Informatik, Biometrie und Epidemiologie) examined my statistical design twice, more often than he needed to, and approved it. He provided insight into statistics, which I hope to use again.

Several coworkers at the Institute of Pathology of the University Hospital Essen provided me with advice and help, if needed.

On a personal level, I would like to thank my parents, who never ceased to give me emotional support. Likewise, I am grateful to my girlfriend Larissa for her inspiration and encouragement.

9. Curriculum vitae

This section is not presented in the online version due to reasons of privacy.