Evaluation of long noncoding RNA MALAT1 as a candidate blood-based biomarker for the diagnosis of non-small cell lung cancer

Daniel Gilbert Weber1†, Georg Johnen1†, Swaantje Casjens1, Oleksandr Bryk1, Beate Pesch1, Karl-Heinz Jöckel2, Jens Kollmeier3 and Thomas Brüning1

Abstract

Background: The long noncoding RNA MALAT1 (metastasis-associated lung adenocarcinoma transcript 1) is described as a potential biomarker for NSCLC (non-small cell lung cancer). Diagnostic biomarkers need to be detectable in easily accessible body fluids, should be characterized by high specificity, sufficient sensitivity, and robustness against influencing factors. The aim of this study was to evaluate the performance of MALAT1 as a blood based biomarker for NSCLC.

Results: MALAT1 was shown to be detectable in the cellular fraction of peripheral human blood, showing different expression levels between cancer patients and cancer-free controls. For the discrimination of NSCLC patients from cancer-free controls a sensitivity of 56% was calculated conditional on a high specificity of 96%. No impact of tumor stage, age, gender, and smoking status on MALAT1 levels could be observed, but results based on small numbers.

Conclusions: The results of this study indicate that MALAT1 complies with key characteristics of diagnostic biomarkers, i.e., minimal invasiveness, high specificity, and robustness. Due to its relatively low sensitivity MALAT1 might not be feasible as a single biomarker for the diagnosis of NSCLC in the cellular fraction of blood. Alternatively, MALAT1 might be applicable as a complementary biomarker within a panel in order to improve the entire diagnostic performance.

Keywords: Lung, Cancer, NSCLC, Biomarkers, lncRNA, MALAT1, Minimally-invasive, Sensitivity, Specificity

Background

Lung cancer is the leading cause of cancer death worldwide [1] with NSCLC (non-small cell lung cancer) as the most prominent subgroup accounting for approximately 80% of all lung cancer cases. Commonly, the disease is detected in late stages resulting in short survival rates, whereas for patients with early-stage lung cancer longer survival rates could be observed [2]. Thus, the detection of lung cancer in early stages when clinical symptoms have not yet occurred appears to be a promising opportunity to decrease mortality, because in more cases a curative therapy might become possible.

In principal, biomarkers should be feasible for the detection of cancer in early stages. Thus, a major aim in cancer research is the identification of proper biomarkers. Key characteristics of diagnostic biomarkers among others are: (i) minimally-invasive to measure the biomarker in easily accessible body fluids, (ii) high specificity to avoid false-positive results in cancer-free individuals, (iii) sufficient sensitivity to detect the tumors, and (iv) robustness against potential influencing factors.

In recent years biomarker research focused on noncoding RNAs (ncRNAs), in particular microRNAs (miRNAs). MiRNAs are small RNA molecules with a length of ~ 22 nucleotides (nt), playing a central role in the regulation of gene expression [3] and acting as tumor suppressors or oncogenes in cancer [4]. Several studies show the feasibility of using miRNAs as biomarkers in body fluids for the diagnosis of lung cancer [5-8]. However, there is a lack of...
consistent results between studies focused on the identification of miRNAs as biomarkers [9]. Thus, the discovery of alternative or complementing biomarkers is essential.

In addition to miRNAs, long noncoding RNAs (lncRNAs) are a promising alternative within the group of ncRNAs. LncRNAs are commonly described as RNA molecules with a length > 200 nt, playing regulatory and structural roles in biological processes. As lncRNAs are implicated as tumor suppressors and oncogenes [10], they might be feasible as diagnostic biomarkers [11]. Currently, only few lncRNAs have been described as candidate biomarkers in human body fluids [10]. HULC (highly upregulated in liver cancer) is highly expressed in hepatocellular carcinoma patients and detectable in human blood [12]. PCA3 (prostate cancer gene 3) is detectable in urine of prostate cancer patients, showing high accuracy [13]. In addition, MALAT1 (metastasis-associated lung adenocarcinoma transcript 1) might be a candidate biomarker for NSCLC [14]. MALAT1 is a well-described lncRNA widely expressed in normal tissues [15]. In several human carcinomas MALAT1 was shown to be upregulated [16], particularly in early-stage metastasizing NSCLC.

The aim of this study was the evaluation of MALAT1 as a blood-based biomarker for NSCLC. The expression of MALAT1 was measured in the cellular fraction of peripheral human blood and the expression levels of NSCLC patients and cancer-free controls of the general population were compared.

**Methods**

**Study population**

The study was designed according to rules guarding patient privacy and with the approval from the ethics committee of the Ruhr-Universität Bochum (No. 3217–08). All participants provided written informed consent.

The cancer group of 45 NSCLC patients consisted of 21 patients with AdCa (adenocarcinoma) and 24 patients with SqCC (squamous cell carcinoma). Participants were recruited at the HELIOS Clinic Emil von Behring, Berlin, Germany. Tumor staging was performed according to the TNM classification of malignant tumors [17]. Cancer patients had not been treated by surgery, chemotherapy, or radiation therapy before blood collection. The control group of 25 cancer-free subjects was drawn from the Heinz Nixdorf Recall study, a population-based cohort of elderly subjects [18]. Characteristics of the study groups are summarized in Table 1. Detailed subject characteristics are listed in Additional file 1.

**RNA isolation**

Peripheral blood samples were collected from each participant in 9.0 ml S-Monovette EDTA gel tubes (Sarstedt, Nümbrecht, Germany) and centrifuged (2000 x g for 10 minutes) within 30 minutes after collection. The cellular fraction was separated from plasma and stored frozen until RNA isolation.

Samples were thawed at room temperature and RNA isolation including DNase I treatment was performed from 0.5 ml of the cellular fraction using the RiboPure Blood Kit according to the manufacturers’ instructions (Life Technologies, Darmstadt, Germany).

**Quantitative real-time PCR (qRT-PCR)**

TaQMan assays (Life Technologies) were used for quantitative expression analyses of MALAT1 (Hs00273907_s1) as potential biomarker and of GAPDH (Hs99999905_m1), HPRT1 (Hs02800696_m1), and RPLP0 (Hs99999902_m1) as potential reference genes for normalization. Quantitative real-time PCR (qRT-PCR) was performed using a 7900 HT Fast Real-Time PCR System (Life Technologies). For the reverse transcription reaction 12 μl RNA and for the PCR reaction 5 μl cDNA were used as templates. Samples were analyzed in duplicate and non-template controls were included. For cycle threshold (Ct) estimation a fixed threshold of 0.2 was used. Ct values > 35 were considered to be under the detection limit [19] and marked as 35 for analysis [20]. Raw Ct values are presented in Additional file 1.

The performance of potential references was analyzed utilizing RefFinder [21], a web-based comprehensive tool (www.leonxie.com/referencegene.php), including the four commonly used algorithms geNorm [22], NormFinder [23], BestKeeper [24], and comparative ΔCt method [25], to evaluate the most stable reference across study groups. As the geometric mean (GM) of several reference genes is more reasonable than a single reference gene [22], the GM of potential references was calculated. Normalized MALAT1 levels were expressed as ΔCt, with ΔCt = Ct(MALAT1) - Ct(Reference)-.
Statistical analysis
Median and inter-quartile range (IQR) were used to describe the distribution of MALAT1 levels. Groups were compared using the non-parametric Kruskal-Wallis test for continuous variables. Sensitivity and specificity of MALAT1 were determined from receiver operating characteristic (ROC) curves illustrating the performance of MALAT1 to discriminate the studied groups. In brief, NSCLC vs. controls, AdCa vs. controls, SqCC vs. controls, and AdCa vs. SqCC were analyzed. The bootstrap procedure (1000 runs) was used for internal validation of the estimates in the ROC analyses.

Potential factors influencing MALAT1 levels were evaluated using a linear regression model. Estimates were given as β with 95% confidence intervals (CI) and p values. Here, values of β > 0 indicate a negative association between the influencing factor and MALAT1 levels, values of β < 0 a positive association. Logistic regression modeling was performed to estimate the odds ratio (OR) with 95% CI for detecting NSCLC as a function of normalized MALAT1 levels.

Statistical analyses were performed using SAS/STAT and SAS/IML software, version 9.3 (SAS Institute Inc., Cary, NC).

Results
Expression stability of candidate references
The potential reference genes GAPDH, HPRT1, and RPLP0 were measured in all samples from NSCLC patients and cancer-free controls. Using raw Ct values no significant differences between NSCLC patients and controls could be observed for GAPDH and HPRT1 in contrast to RPLP0 (p = 0.0002), (Figure 1). Thus, RPLP0 was excluded from further evaluation as reference gene.

In order to identify the most stable reference across the study groups RefFinder was used to rank the analyzed references. Lowest rank represents the most stable reference and highest rank represents the least stable reference (Table 2). The GM of GAPDH and HPRT1 was identified as the most stable reference and used for normalization of MALAT1.

Distribution of MALAT1 in the study groups
Table 3 depicts the distribution of normalized MALAT1 levels in the study groups. The median normalized MALAT1 value for NSCLC was −0.35 (IQR −1.34; 1.00), for AdCa −0.59 (IQR −1.34; -0.25), and for SqCC −0.23 (IQR −1.31; 1.47), whereas for controls the median normalized MALAT1 level was −2.07 (IQR −2.53; -0.83).

Differences of normalized MALAT1 levels between cancer patients and cancer-free subjects were significant for NSCLC vs. controls (p <0.0001), AdCa vs. controls (p = 0.0043), and SqCC vs. controls (p = 0.0001), whereas the difference between AdCa and SqCC was not significant (Figure 2).

MALAT1 as biomarker of NSCLC
Using ROC analyses, for NSCLC patients and controls an area under the curve (AUC) of 0.79 (95% CI 0.68 – 0.89) (Figure 3A), for AdCa patients and controls an AUC of 0.75 (95% CI 0.59 – 0.90), (Figure 3B), for SqCC patients and controls an AUC of 0.82 (0.70 – 0.94) (Figure 3C), and for AdCa vs. SqCC an AUC of 0.58 (95% CI 0.42 – 0.76), (Figure 3D), were calculated.

Sensitivity and specificity of normalized MALAT1 are shown in Table 4, calculated due false-positive rates (FPR) of 0% (no false-positive test), 4% (one false-positive test), and 8% (two false-positive tests), and to the maximum Youden’s Index (YI = sensitivity + specificity −1), respectively. A FPR of 4%, representing 96% specificity, resulted in 56% sensitivity for the discrimination of NSCLC from controls. The sensitivity to discriminate SqCC from controls is higher (63%) than the sensitivity to discriminate AdCa from controls (48%). A FPR of 8% (92% specificity)

Table 2 Results of reference analysis using RefFinder [21] to evaluate the most stable reference across the study groups

| Algorithm            | Reference ranking | 1 (most stable) | 2 | 3 (least stable) |
|----------------------|-------------------|----------------|---|-----------------|
| RefFinder (comprehensive ranking) | GM*              | GAPDH          | HPRT1 |
| geNorm               | GM* & GAPDH       | HPRT1          |
| NormFinder           | GM*              | GAPDH          | HPRT1 |
| BestKeeper           | HPRT1            | GM*            | GAPDH |
| Comparative ΔCt method | GM*              | GAPDH          | HPRT1 |

*Geometric mean (GM) was calculated from GAPDH and HPRT1.
Table 3 Distribution of normalized MALAT1 levels by median and inter-quartile range (IQR) stratified by potential influencing factors

|                   | Total | Controls | NSCLC | AdCa | SqCC |
|-------------------|-------|----------|-------|------|------|
|                   | N    | Median   | IQR   | N    | Median | IQR   | N    | Median   | IQR   | N    | Median | IQR   |
| Total             | 70   | -0.86    | -2.10;| 25   | -2.07  | -2.53;| 45   | -0.35    | -1.34;| 21   | -0.59  | -1.34;| 24   | -0.23   | -1.31;|
| Gender            |       |          |       |      |        |       |      |          |       |      |        |       |      |        |
| Male              | 48   | -0.80    | -2.18;| 18   | -2.18  | -2.54;| 30   | -0.35    | -1.30;| 15   | -0.35  | -1.44;| 15   | -0.36   | -1.25;|
| Female            | 22   | -1.01    | -2.10;| 7    | -1.49  | -2.24;| 15   | -0.35    | -1.37;| 6    | -0.84  | -1.34;| 9    | 0.26    | -1.37;|
| Age               |       |          |       |      |        |       |      |          |       |      |        |       |      |        |
| < 60 years        | 11   | -1.93    | -2.69;| 4    | -2.77  | -3.58;| 7    | -1.10    | -2.01;| 2    | -1.14  | -2.54;| 5    | -1.10   | -1.93;|
| 60-69 years       | 26   | -0.78    | -1.69;| 7    | -1.69  | -2.24;| 19   | -0.41    | -1.35;| 9    | -0.89  | -1.34;| 10   | -0.19   | -0.74;|
| 70-79 years       | 27   | -0.78    | -2.29;| 11   | -2.07  | -2.53;| 16   | -0.29    | -1.37;| 8    | -0.57  | -1.37;| 8    | 0.54    | -1.39;|
| ≥ 80 years        | 6    | -0.78    | -1.25;| 3    | -0.92  | -2.64;| 3    | -0.35    | -1.25;| 2    | -0.06  | -0.35;| 1    | -1.25   | -1.25;|
| Smoking status    |       |          |       |      |        |       |      |          |       |      |        |       |      |        |
| Ever              | 64   | -0.86    | -2.12;| 20   | -2.15  | -2.53;| 44   | -0.36    | -1.35;| 20   | -0.69  | -1.39;| 24   | -0.23   | -1.31;|
| Never             | 6    | -1.01    | -2.10;| 5    | -1.38  | -2.10;| 1    | 1.54     |       | 1    | 1.54   |       |      |        |
| Tumor size        |       |          |       |      |        |       |      |          |       |      |        |       |      |        |
| I/II              | 3    | -1.10    | -2.17;| 21   | -0.59  | -1.34;| 3    | -1.10    | -2.17;| 21   | -0.15  | -1.25;| 21   | -0.15   | -1.25;|
| III/IV            | 42   | -0.29    | -1.34;| 21   | -0.59  | -1.34;| 3    | -1.10    | -2.17;| 21   | -0.15  | -1.25;| 21   | -0.15   | -1.25;|
| Metastasis status |       |          |       |      |        |       |      |          |       |      |        |       |      |        |
| M0                | 7    | -0.41    | -1.93;| 2    | -0.22  | -1.44;| 2    | -0.22    | -1.44;| 15   | -0.41  | -2.01;| 9    | -0.15   | -0.24;|
| M1                | 19   | -0.29    | -1.17;| 9    | -0.59  | -1.34;| 2    | -0.22    | -1.44;| 10   | -0.41  | -1.91;| 9    | -0.15   | -0.24;|
| Lymph node status |       |          |       |      |        |       |      |          |       |      |        |       |      |        |
| N0                | 3    | 0.24     | -2.17;| 1    | 0.24   |       | 2    | -0.42    | -2.17;| 22   | -0.23  | -1.25;| 22   | -0.23   | -1.25;|
| N1-N3             | 42   | -0.36    | -1.34;| 20   | -0.69  | 1.39; | 5    | 0.24     |       | 22   | -0.23  | -1.25;| 22   | -0.23   | -1.25;|

Figure 2 Scatter dot plots of MALAT1 levels in NSCLC (N = 45), AdCa (N = 21), SqCC (N = 24), and cancer-free controls (N = 25). Expression values were normalized by the geometric mean of GAPDH and HPTBI and expressed as ΔCt. Horizontal bars represent median and IQR. Groups were compared using the Kruskal-Wallis test. NSCLC: non-small cell lung cancer, AdCa: adenocarcinoma, SqCC: squamous cell carcinoma.

This did not lead to any increase in sensitivities, whereas a FPR of 0% (100% specificity) resulted in lowest sensitivities for the discrimination of controls from patients with NSCLC (47%), AdCa (38%), or SqCC (54%). Use of the maximum YI leads to an increase in sensitivity to 81% only for AdCa vs. controls, but specificity decreased to 64%. For the discrimination of AdCa from SqCC a FPR of 5% (95% specificity, one false-positive test) resulted in 8% sensitivity and a FPR of 10% (90% specificity, two false-positive tests) resulted in 21% sensitivity, whereas using the maximum YI resulted in 33% sensitivity and 86% specificity.

ROC analyses on 1000 bootstrap samples resulted in similar cutoffs, sensitivities, and specificities of MALAT1 in comparison to the original analyses. The calculated 95% CI regarding NSCLC vs. controls and SqCC vs. controls indicate a good precision of this assessment, whereas AdCa vs. controls the 95% CI shows a less precision (Additional file 2).

The application of logistic regression models revealed a two-fold increased risk of detecting NSCLC per normalized
The OR of MALAT1 was 1.88 (95% CI 1.26 - 2.83) without adjustment and 2.03 (95% CI 1.30-3.16) with adjustment for gender, age and smoking status.

Potential factors influencing MALAT1
The influence of tumor characteristics on MALAT1 levels is shown in Table 5. MALAT1 is not affected by tumor size, metastasis status or lymph node status.

The impact of potential influencing factors on the expression levels of MALAT1 are shown in Table 6. NSCLC showed a significant 1.63-fold (95% CI 0.75 – 2.51) decrease of MALAT1 (p = 0.0003), whereas the factors gender, age, and smoking status showed no impact on the MALAT1 levels in human blood.

Discussion
NSCLC is commonly detected in late stages of the disease. Biomarkers have the potential to detect cancer at early stages, facilitating an earlier and therefore more curative therapy that ideally results in decreased mortality. In NSCLC, Gutschner et al. showed that MALAT1 regulates the expression of several metastasis-associated genes, e.g. CDCP1 (CUB domain containing protein 1) and GPC6 (glypican 6), indicating a major role of MALAT1 in disease progression [26]. Additionally, it was suggested that MALAT1 might also regulate other important cellular processes in lung cancer [26]. Thus, MALAT1 is a candidate biomarker for NSCLC [14].

For quantitative expression analysis of messengerRNAs (mRNAs) and miRNAs qRT-PCR is considered to be the gold standard [27] and the same might be true for lncRNAs. However, to produce reliable data in qRT-PCR assays the use of appropriate reference genes for normalization is an important issue [28] and candidate reference genes need to be tested prior to application [29]. As no information regarding lncRNAs as references were accessible, mRNAs were selected as potential references. HPRT1 and RPLP0 are well-described reference genes for analyses in NSCLC tissues [30] and GAPDH

![Figure 3 Receiver operating characteristics (ROC) curves of normalized MALAT1.](image-url)
was already applied for normalization of MALAT1 [15]. However, in this study RPLP0 seems to be no feasible reference which is in agreement with Falkenberg et al., showing that RPLP0 is not appropriate as reference gene in human blood samples [31]. In this study, GAPDH and HPRT1 were suitable reference genes, particularly the GM of GAPDH and HPRT1 showed the best reference performance. This is in accordance with Ulivi et al., using GAPDH and HPRT1 for normalization of mRNAs in blood samples of NSCLC patients and controls [32].

One key characteristic of proper diagnostic biomarkers is the need to be detectable in easily accessible body fluids like peripheral blood. In this study MALAT1 was measured in the cellular fraction of human blood, showing that this matrix is in principle appropriate for the analysis of lncRNAs. Comparable results for the usability of the cellular blood fraction were shown for miRNAs [33,34]. Commonly, the cellular fraction obtained during plasma preparation is discarded, but it might be reasonable to collect this matrix in biobanks for subsequent biomarker discovery.

In this study, a significant downregulation of MALAT1 in NSCLC patients in comparison to cancer-free controls was shown. Comparable results were achieved by Zhang et al., showing a downregulation of MALAT1 in patients with hepatocellular carcinomas [35]. However, MALAT1 was implicated to play an oncogenic role [10] and upregulation of MALAT1 was observed in several

| Table 5 Analysis of tumor characteristics influencing MALAT1 levels in human blood |
|---------------------------------|-----------------|------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                                | β   | 95% CI     | p value         | β   | 95% CI     | p value         | β   | 95% CI     | p value         |
| Intercept                      | 1.23 | -3.49; 1.03 | 0.3670          | 1.52 | -0.82; 3.85 | 0.1911          | -1.23 | -3.41; 0.96 | 0.3670          |
| Tumor size (Reference: I/II)   | 1.06 | -1.28; 3.39 | 0.3670          | 1.52 | -0.82; 3.85 | 0.1911          | -1.23 | -3.41; 0.96 | 0.3670          |
| Intercept                      | -0.15 | -1.11; 0.80 | 0.22; 2.81      | -0.14 | -3.64; 2.72 | 0.7655          | 0.65  | -0.98; 2.26 | 0.4158          |
| Metastasis status (Reference: M0) | -0.14 | -1.36; 1.07 | 0.8143          | -0.46 | -3.64; 2.72 | 0.7655          | 0.65  | -0.98; 2.26 | 0.4158          |
| Intercept                      | -0.20 | -2.48; 2.08 | 0.24; 4.51      | -0.14 | -3.64; 2.72 | 0.7655          | 0.65  | -0.98; 2.26 | 0.4158          |
| Lymph node status (Reference: N0) | -0.04 | -2.40; 2.32 | 0.9714          | -0.02 | -5.29; 3.45 | 0.6654          | 0.57  | -2.32; 3.47 | 0.6860          |

Values of β > 0 indicate a negative association between the analyzed factor and MALAT1 levels, values < 0 a positive association.
such decay of is shorter than the median half-life of mRNA MALAT1 in various cell types and indicated that the half-life of MALAT1 or prostate cancer [37,38]. Thus, it would be reasonable to MALAT1 ever, further analyses are needed to evaluate the origin of might originate from leucocytes altered by the tumor. How-

lular fraction of blood remains unclear. Theoretically, it on the immune system. The source of MALAT1 in blood cells may be an indirect effect of the tumor, e.g., directly produced by the tumor tissue. Its downregulation was detectable in plasma of patients with gastric lncRNA-containing extracellular vesicles [39].

assumption needs to be verified in more detail. Theoretical meaning for lncRNA analyses in blood. However, this implication that the use of stabilization systems might be responding to larger quantities (data not shown). The results

Table 6 Analysis of potential factors influencing MALAT1 levels in human blood

|                      | β    | 95% CI | p value |
|----------------------|------|--------|---------|
| Intercept            | −4.64| −8.93; | 0.0004  |
|                      | −0.35|        |         |
| Group (Reference: controls) | NSCLC | 1.63 | 0.75;  | 0.0004  |
|                      | 2.51 |        |         |
| Gender (Reference: male) | Female | 0.22 | −0.66; | 0.6210  |
|                      | 1.10 |        |         |
| Age                  | 10 years | 0.45 | −0.10; | 0.1058  |
|                      | 1.00 |        |         |
| Smoking status (Reference: never) | Ever | −0.42 | −1.94; | 0.5848  |
|                      | 1.10 |        |         |

Values of β > 0 indicate a negative association between analyzed factor and MALAT1, values < 0 a positive association. Statistically significant changes are marked in bold.

other cancers, e.g. of the breast and prostate [16]. Such differences might be caused by the paradigm that in fact MALAT1 is expressed ubiquitously but fulfills tissue-specific functions depending on the cellular environment [36]. Commonly, MALAT1 is analyzed in tissues [14-16], whereas in this study and the study of Zhang et al. [35] MALAT1 was detected in blood. Because MALAT1 was detected in the cellular fraction of blood, it is unlikely to be directly produced by the tumor tissue. Its downregulation in blood cells may be an indirect effect of the tumor, e.g., on the immune system. The source of MALAT1 in the cellular fraction of blood remains unclear. Theoretically, it might originate from leucocytes altered by the tumor. However, further analyses are needed to evaluate the origin of MALAT1 in human blood. Very recently, it was shown that MALAT1 was detectable in plasma of patients with gastric or prostate cancer [37,38]. Thus, it would be reasonable to analyze MALAT1 in plasma of NSCLC patients instead of the cellular fraction, because the presence of MALAT1 in plasma might be a direct effect of the tumor, e.g., release of IncRNA-containing extracellular vesicles [39].

Tani et al. showed that the stability of MALAT1 varied in various cell types and indicated that the half-life of MALAT1 is shorter than the median half-life of mRNA [40]. Such decay of MALAT1 might also prevail in blood cells. It is well known that systems like PAXgene, Tempus, and RNAlater stabilize mRNAs and miRNAs in whole blood samples [41-43]. Thus, the performance of the assay was additionally tested in a few available blood samples stabilized by PAXgene or RNAlater. In the stabilized blood samples MALAT1 is detectable at lower Ct values corresponding to larger quantities (data not shown). The results implicate that the use of stabilization systems might be meaningful for IncRNA analyses in blood. However, this assumption needs to be verified in more detail.

Regarding the key characteristics of an obligatory high specificity and a sufficiently high sensitivity of diagnostic biomarkers, the candidate biomarker MALAT1 does not fulfill both criteria. Generally, in screening cohorts a high specificity is needed to avoid an unacceptably high number of false-positive tests that would result in psychological pressure and needless intervention for the patients. Thus, the sensitivity of candidate biomarkers should be calculated at a fixed high specificity level [44]. In regard to the relatively small study group the specificity of 96% is quite high, particularly as this corresponds to only one single false-positively tested control. On the other hand, the calculated sensitivity is too low (56%) for the use of MALAT1 as a single biomarker for the diagnosis of NSCLC, particularly for the subtype AdCa (48%). However, lower sensitivity could be balanced by the use of several biomarkers in a panel. Theoretically, in an optimal panel every single biomarker is characterized by sufficiently high sensitivity and the necessary high specificity, perfectly complement each other in order to obtain superior diagnostic performance [45]. Thus, it might be reasonable to verify MALAT1 in combination with other biomarkers in larger study groups to improve the entire diagnostic performance of the biomarker panel. However, for the discrimination of AdCa and SqCC, a sensitivity of only 8% precludes MALAT1 as a biomarker for the differential diagnosis of NSCLC subtypes.

Bootstrap analysis showed that the calculated cutoffs, sensitivities, and specificities remain stable, indicating that the calculated values are appropriate for the discrimination of patients and controls.

Regarding the fourth key characteristic of diagnostic biomarkers, the results indicate that MALAT1 values in blood are not correlated with tumor size, metastasis status, or lymph node status. However, more cases of early-stage metastasizing NSCLC need to be analyzed in subsequent studies because this study comprises only three cases with tumor stage T1 or T2. Additionally, MALAT1 seems to be relatively independent from common influencing factors like age, gender, and smoking status, indicating the robustness of the candidate biomarker. These observations are in agreement with MALAT1 expression in tissue [46]. However, it has to be clarified if other potential influencing factors from the multitude of biological, preanalytical, and analytical factors show an impact on MALAT1 levels in human blood.

Conclusions

MALAT1 could be detected in peripheral blood, showing different expression levels between NSCLC patients and cancer-free controls. It was shown that MALAT1 complies with key characteristics of diagnostic biomarkers, being minimally-invasive, exhibiting high specificity, and robustness. On the contrary, the observed sensitivity is too low for the use of MALAT1 as a single biomarker for the diagnosis of NSCLC using the cellular fraction of
peripheral blood. However, it might be reasonable to verify the performance of MALAT1 as a complementary biomarker within a panel in larger studies including more cases of early-stage metastasizing NSCLC.

Additional files

Additional file 1: Subject characteristics and raw data of MALAT1, GAPDH, and HPRT1 expression analysis.

Additional file 2: Marker cutoffs with 95% CI for NSCLC (non-small cell lung cancer) vs. controls, AdCa (adenocarcinoma) vs. controls, SqCC (squamous cell carcinoma) vs. controls, and AdCa vs. SqCC after bootstrap analysis with 1000 random samples, according to false positive rates (FPPR) corresponding to none, one, and two false-positive tests and maximum Youden's index (YI).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

DGW conceived of the study, participated in its design and coordination, and drafted the manuscript. GJ participated in study design and coordination and helped to draft the manuscript. SC performed the statistical analyses and helped to draft the manuscript. OB performed the coordination and helped to draft the manuscript. JK participated in statistical analyses and helped to draft the manuscript. Reis EM, Verjovski-Almeida S: Perspectives of long non-coding RNAs in cancer diagnostics. Front Genet 2012, 3:32.

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D: Global cancer statistics. CA Cancer J Clin 2011, 61:69–90.

2. Goldstraw P, Crowley J, Chansky K, Giroux D, Groome PA, Rami-Porta R, Postmus PE, Rusch V, Sobin L: International Association for the Study of Lung Cancer International Staging C, Participating I: the IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. J Thorac Oncol 2007, 2:706–714.

3. Ambros V: The functions of animal microRNAs. Nature 2004, 431:350–355.

4. Kent OA, Mendell JT: A small piece in the cancer puzzle: microRNAs as tumor suppressors and oncogenes. Oncogene 2006, 25:6188–6196.

5. Foss KM, Sima C, Ugolini D, Neri M, Allen KE, Weiss GJ: miR-1254 and miR-574-5p: Serum-based microRNA biomarkers for early-stage non-small cell lung cancer. J Thorac Oncol 2011, 6:482–488.

6. Zheng D, Haddadin S, Wang Y, Gu LQ, Perry MC, Feiter CE, Wang MX: Plasma microRNAs as novel biomarkers for early detection of lung cancer. Int J Exp Pathol 2011, 4:575–586.

7. Bianchi F, Nicassio F, Marzi M, Belloni E, Dall’olio V, Bernardi L, Pelosi G, Maisonneuve P, Veronese G, Di Fiore PP: A serum circulating miRNA diagnostic test to identify asymptomatic high-risk individuals with early stage lung cancer. Embo Mol Med 2011, 3:495–506.

8. Boeri M, Verri C, Conte D, Roz L, Modena P, Facchinetti F, Calabro E, Croce CM, Pastorino U, Sazgari G: MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. Proc Natl Acad Sci USA 2011, 108:373–377.

9. Pasi HJ, Beer DG, Joseph S, Masson P: Biomarkers and molecular testing for early detection, diagnosis, and therapeutic prediction of lung cancer. Thorac Surg Clin 2013, 23:211–224.

10. Reis EM, Verjovski-Almeida S: Perspectives of long non-coding RNAs in cancer diagnostics. Front Genet 2012, 3:32.

11. Tano K, Akimitsu N: Long non-coding RNAs in cancer progression. Front Genet 2013, 3:219.

12. Panatti K, Tschernatsch MM, Guellu C, Moustafa T, Strnad M, Strohmaier HM, Buck CR, Dunk H, Schroeder R, Trauner M, Zlatoukal K: Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA. Gastroenterology 2007, 132:330–342.

13. Fradet Y, Saaf D, Aprikian A, Dessureault J, Elhilali M, Trudel C, Masse B, Piche L, Chypre C: tPM3C, a new molecular urine test for the detection of prostate cancer. Urology 2004, 64:311–315. discussion 315–316.

14. Schmidt LH, Spiekert T, Kochschilder S, Schöllars S, Humberg J, Jungen D, Bulk E, Hascher A, Wittmer D, Marra A, et al: The long noncoding MALAT-1 RNA indicates a poor prognosis in non-small cell lung cancer and induces migration and tumor growth. J Thorac Oncol 2011, 6:1894–1992.

15. Ji P, Diederichs S, Wang W, Boing S, Metzger R, Schneider PM, Tidow N, Brandt B, Buerger H, Bulk E, et al: MALAT-1, a novel noncoding RNA, and thymosin betadetected metastasis and survival in early-stage non-small cell lung cancer. Oncogene 2001, 22:8301–8304.

16. Lin R, Maeda S, Liu C, Kariin M, Edgington TS: A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. Oncogene 2007, 26:851–858.

17. Sobin LH, Gospodarowicz MK, Wittekind C: International Union against Cancer: TNM classification of malignant tumours. 7th edn. Chichester, West Sussex, UK: Wiley-Blackwell; Hoboken, NJ, 2010.

18. Schmiedmann A, Möhlenkamp S, Stang A, Grönemeyer HU, Laedke F, Seibel R, Hirche H, Mann K, Siffert W, Lauterbuch K, Siegrist J, et al: Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL study. Risk factors, evaluation of coronary calcium and lifestyle. Am Heart J 2002, 144:212–218.

19. Guthrie JL, Seah C, Brown S, Tang P, Jamieson F, Drews SJ: Use of Bordetella pertussis BP3385 to establish a cutoff value for an IS481-targeted real-time PCR assay. J Clin Microbiol 2008, 46:3798–3799.

20. Ning B, Dial S, Yang J, Guo L: Gene expression studies in human reticulocytes using real-time PCR. J Mol Biol 2012, 420:8031–8041.

21. Xie F, Xiao P, Chen D, Xu L, Zhang B: miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. Plant Mol Biol 2012, 80:75–84.

22. Vandesompele J, De Preter K, Pattyn F, Poppe B, van Roy N, De Paepe A, Speleman F: Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002, 3:RESEARCH0034.1-0034.11.

23. Andersen CL, Jensen JL, Møller KJ, Orntoft TF: 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 2004, 64:2525–2526.

24. Pfaffi MW, Tichopad A, Pergomet C, Neuvians TP: Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper-Etch-based tool using pair-wise correlations. BioTechnol Lett 2004, 26:529–515.

25. Silver N, Best S, Jiang J, Thein SL: Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. BMC Mol Biol 2006, 7:3.

26. Gutschner T, Hämmerle MT, Eissmann M, Hsu J, Kim Y, Hung G, Revenko A, Brandt B, Buerger H, Bulk E, et al: MALAT-1, a novel noncoding RNA, and thymosin betadetected metastasis and survival in early-stage non-small cell lung cancer. Oncogene 2001, 22:8301–8304.

27. Becker C, Hämmerle-Fickinger A, Riedmaier I, Pfaffl MW: MicroRNA and mRNA quality control for RT-qPCR analysis. Methods 2010, 50:237–243.

28. Davoren PA, McNeill RE, Lowery AJ, Kerin MJ, Miller N: Identification of suitable endogenous control genes for microRNA gene expression analysis in human breast cancer. BMC Mol Biol 2008, 9:10.
30. Gresner P, Gromadzinska J, Wasowicz W: Reference genes for quantitative gene expression measurements in human blood RNA. Acta Biochim Pol 2009, 56:307–316.

31. Falkenberg VR, Whistler T, Murray JR, Unger ER, Rajeevan MS: Identification of Phosphoglycerate Kinase 1 (PGK1) as a reference gene for quantitative expression studies on non-small cell lung cancer. Acta Biochim Pol 2009, 56:307–316.

32. Ulivi P, Mercatali L, Casoni GL, Scarpi E, Bucchi L, Silvestrini R, Sanna S, Monteverde M, Amadori D, Polo L, Zoli W: Multiple marker detection in peripheral blood for NSCLC diagnosis. PLoS One 2013, 8:e74601.

33. Weber DG, Johnen G, Bryk O, Jöckel KH, Brüning T: Identification of miRNA-103 in the cellular fraction of human peripheral blood as a potential biomarker for malignant Mesothelioma - a pilot study. PLoS One 2012, 7:e30221.

34. Häusler SF, Keller A, Chandran PA, Ziegler K, Zipp K, Heuer S, Krockenberger M, Engel JB, Honig A, Scheffler M, et al: Whole blood-derived miRNA profiles as potential new tools for ovarian cancer screening. Br J Cancer 2010, 103:693–700.

35. Zhang PJ, Wei R, Wen YX, Ping L, Wang CB, Dong ZN, Deng XX, Bo W, Bin C, Tian YP: Genes expression profiling of peripheral blood cells of patients with hepatocellular carcinoma. Cell Biol Int 2012, 36:803–809.

36. Eissmann M, Gutschner T, Hämmerle M, Günther S, Caudron-Herger M, Gross M, Schirmacher P, Rippe K, Braun T, Zörnig M, Diederichs S: Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. RNA Biol 2012, 9:1076–1087.

37. Arita T, Ichikawa D, Konishi H, Komatsu S, Shiozaki A, Shoda K, Kawaguchi T, Kruhoffer M, Dyrskjot L, Voss T, Lindberg RL, Wyrich R, Thykjaer T, Orntoft TF: miRNA-103 in the cellular fraction of human peripheral blood as a potential biomarker for malignant Mesothelioma - a pilot study. PLoS One 2013, 8:e74601.

38. Eissmann M, Gutschner T, Hämmerle M, Günther S, Caudron-Herger M, Gross M, Schirmacher P, Rippe K, Braun T, Zörnig M, Diederichs S: Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. RNA Biol 2012, 9:1076–1087.

39. Tani H, Nakamura Y, Iiiri K, Akimitsu N: Stability of MALAT-1, a nuclear long non-coding RNA in mammalian cells, varies in various cancer cells. Drug Discov Today 2010, 4:235–239.

40. Weber DG, Casjens S, Zilch-Schöneweis S, Rozynek P, Lehnert M, Zöllner P, Bryk O, Jöckel KH, Brüning T: Multiple marker detection in peripheral blood for NSCLC diagnosis. PLoS One 2013, 8:e74601.

41. Weber DG, Johnen G, Bryk O, Jöckel KH, Brüning T: Identification of miRNA-103 in the cellular fraction of human peripheral blood as a potential biomarker for malignant Mesothelioma - a pilot study. PLoS One 2012, 7:e30221.

42. Matheson LA, Duong TT, Rosenberg AM, Yeung RS: Assessment of sample collection and storage methods for multicenter immunologic research in children. J Immunol Methods 2008, 339:82–89.

43. Taeger D, Gomolka M, Kreuzer M, Otten H, et al: Assessment of miRNA and microRNA stabilization in peripheral human blood for multicenter studies and biobanks. Biomark Insights 2010, 5:95–102.

44. Matheson LA, Duong TT, Rosenberg AM, Yeung RS: Assessment of sample collection and storage methods for multicenter immunologic research in children. J Immunol Methods 2008, 339:82–89.

45. Matheson LA, Duong TT, Rosenberg AM, Yeung RS: Assessment of sample collection and storage methods for multicenter immunologic research in children. J Immunol Methods 2008, 339:82–89.