miR-20b and miR-144* decrease in human pathologies

Andreas Keller[1], Petra Leidinger[2], Britta Vogel[3], Christina Backes[1], Abdou ElSharawy[4], Sabine Marquart[3], Michael G. Schrauder[5], Reiner Strick[5], Andrea Bauer[6], Hugo A. Katus[3], Jörg Hoheisel[6], Andre Franke[4], Benjamin Meder[3,+], Eckart Meese[2,+]

[1] Chair for Clinical Bioinformatics, Saarland University, Saarbrücken, Germany
[2] Institute of Human Genetics, Saarland University, Homburg, Germany
[3] Department of Internal Medicine III, University of Heidelberg, Heidelberg, Germany
[4] Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Kiel, Germany
[5] University Breast Center Franconia, Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany
[6] German Cancer Research Center, Heidelberg, Germany

* Equal contribution as first author, + Equal contribution as senior author

Abstract

Background: miRNA profiles are promising biomarker candidates for a manifold of human pathologies, opening new avenues for diagnosis and prognosis. Beyond studies that describe miRNAs frequently as markers for specific traits, we asked whether a general pattern for miRNAs across many diseases exists. We evaluated genome wide circulating profiles of 1,049 patients suffering from 19 different cancer and non-cancer diseases as well as unaffected controls. The results were validated on 319 individuals using qRT-PCR.

Results: We discovered 34 miRNAs with strong disease association. Among those, we found substantially decreased levels of hsa-miR-144* and hsa-miR-20b with AUC of 0.751 (95% CI: 0.703–0.799), respectively. We also discovered a set of miRNAs, including hsa-miR-155*, as rather stable markers, offering reasonable control miRNAs for future studies. The strong down-regulation of hsa-miR-144* and the less variable pattern of hsa-miR-155* has been validated in a cohort of 319 samples in three different centers. Here, breast cancer as additional disease phenotype not included in the screening phase has been included as 20th trait.

Conclusion: Our study on 1,368 patients including 1,049 genome wide miRNA profiles and 319 qRT-PCR validations further underscores the high potential of specific blood-borne miRNA patterns as molecular biomarkers. Importantly, we highlight 34 miRNAs that are generally dys-regulated in human pathologies and a valuable set of constant miRNAs as control markers.

Introduction

In the past decade, non-coding miRNAs have aroused the scientists` interest and their exploration has revolutionized biology. Since the first miRNA was discovered in *Caenorhabditis elegans* in 1993 [1], more and more miRNAs for various species have been reported. Currently, release 20 of the miRBase [2, 3] contains 24,521 entries representing hairpin precursor miRNAs, expressing 30,424 mature miRNA products in 206 species. For *Homo sapiens*, more than 2,500 different mature miRNAs are currently included in this database.

The small non-coding miRNAs are known to be involved in crucial biological processes such as proliferation, apoptosis, differentiation, or development [4-6]. More than 50% of all genes in the human genome are known to be miRNA targets and thus, miRNAs are involved in the regulation of a manifold of metabolic and regulatory pathways such that now the integrative network analysis of miRNAs and mRNAs becomes more and more possible [7-9]. Hence, abnormal miRNA profiles have been associated with many human pathogenic processes as shown by many studies that focused on tissue-derived miRNA profiles (e.g. from patients with lung cancer [10], breast cancer [11] or glioblastoma [12]), Since these small nucleic acids excel in their high stability, they have become even more attractive as biomarker candidates. This also underlines the potential of miRNA biomarkers derived from peripheral blood for diagnostic purposes. Many groups investigated circulating miRNA profiles from serum for various diseases (non-ischaemic systolic heart failure [13], pulmonary tuberculosis [14], non small-cell lung cancer [15, 16], breast cancer [17], prostate cancer [18], or ovarian cancer [19]) whereas we and others developed standardized operating procedures for measuring miRNA profiles from whole peripheral blood (myocardial infarction [20], lung cancer [21], multiple sclerosis [22, 23], melanoma [24], ovarian cancer [25], COPD [26], glioblastoma [27] and Alzheimer Disease [28]).

In the present meta-analysis we analyzed a total of 848 miRNAs in 1,049 samples (containing the 454 samples published in our previous study [29]) measured from whole blood collected in PAXgene blood tubes. The investigated cohort includes healthy controls as well as patients diagnosed with one of 19 diseases of different ICD10 classes (10 cancer entities and 9 non-cancer diseases; details on the different cohort sizes are presented in Table 1). Our results provide a comprehensive overview of the human disease miRNome. By using this rich data source, we aimed at identifying miRNA profiles representative for a general disease state, and to identify miRNA signatures that are suited to discriminate different diseases from controls and from each other.

Results

Identification of miRNAs generally indicating the presence of a disease

We asked if there is a general association between the expression of certain miRNAs in peripheral blood and the presence of a disease. To this end, we calculated 2-tailed unpaired t-test of all patients versus all controls and adjusted the significance values for multiple testing. Furthermore, we calculated for each miRNA separately the area under the receiver operator characteristics curve (AUC) together with the respective 95% confidence intervals. For the comparison diseases (cancer and non-cancer samples) versus healthy controls, we found 333 statistically significantly dys-regulated miRNAs (adjusted p-value <0.05), of which 254 have been up-regulated in diseases while only 79 have been down-regulated. The most significant miRNA hsa-miR-576-5p reached an adjusted significance value of $4.7*10^{-16}$ (raw p-value of $5.6*10^{-19}$). The miRNAs hsa-miR-144* and hsa-miR-20b were most down-regulated with AUC of 0.751 (95% CI: 0.703–0.799), followed by miR-17 and miR-20a. For the first 2 miRNAs, ROC curves are presented in Figure 1. In contrast, hsa-miR-720 was most up-regulated with AUC value of 0.68, followed by hsa-miR-302c. All AUC and p-values for this comparison are provided in Supplemental Table 2.

Focusing on onco-miRNAs by comparing cancer samples versus healthy controls we found 322 miRNAs with adjusted t-test p-values < 0.05, of which 101 have been down-regulated in cancer while 221 were up-regulated. The most significantly dys-regulated miRNA, hsa-miR-130b* reached an adjusted significance value of $1.9*10^{-14}$ (raw p-value of $2.2*10^{-17}$). In this analysis, again hsa-miR-144* and hsa-miR-20b showed the strongest down-regulation in diseases with AUC values of 0.771 (95% CI of 0.721–0.821) and 0.760 (95% CI of 0.71–0.811) while hsa-miR-194* was the most up-regulated miRNA with AUC value of 0.687. All AUC and p-value for this scenario are provided in Supplemental Table 3.

Notably, both comparisons described above showed a high concordance, demonstrated by a correlation of 0.95 of the AUC values and the significant overlap presented in the Venn diagram (Figure 2). This result indicates that most miRNAs are not specific for cancer but for diseases in general. Thus, it is not surprising that the maximal AUC between all cancer and non-cancer diseases computed for hsa-miR-574-5p was just 0.63 and thus substantially smaller than the AUCs for the comparison of diseases versus healthy control samples. Likewise, we found a decreased number of miRNAs significant for this comparison. Altogether, just 116 miRNAs reached a significance value of below 0.05 and remained significant after adjustment for

multiple testing. Of these, 61 were down-regulated in cancer and 55 were up-regulated. All AUC and p-values for this comparison are provided in Supplemental Table 4.

Disease-specificity of single miRNAs

As described above we found many miRNAs being dys-regulated in diseases in general. To further explore this we asked how specific miRNAs are with respect to a specific disease, e.g., whether they are up-regulated in one group and down-regulated in another group of diseases. First, we compared all diseases separately against controls. Of all miRNAs, seven (hsa-miR-380*, hsa-miR-106b, hsa-miR-17, hsa-miR-144*, hsa-miR-558, hsa-miR-548d-3p, and hsa-miR-222) were significantly down-regulated (adjusted 2-tailed t-test p-value < 0.05) in at least 13 of 19 disease conditions, representing most un-specific miRNAs. Further 6 miRNAs were significantly down-regulated in 12 pathologies, 7 miRNAs were down in 11 pathologies, and 6 miRNAs down in 10 diseases while not-being up-regulated in any other. Vice versa, three miRNAs (hsa-miR-130b*, hsa-miR-145, and hsa-miR-658) were up-regulated in 11 diseases while not being down-regulated in any other. Additional 9 miRNAs (hsa-miR-484, hsa-miR-499-5p, hsa-miR-126*, hsa-miR-491-5p, hsa-miR-1303, hsa-miR-539, hsa-miR-25*, hsa-let-7e* and hsa-miR-194*) were up-regulated in 10 diseases while not down-regulated in any other, as the balloon-plot (Figure 3) of all miRNAs significant in at least 8 of 19 diseases (>40%) shows. This result provides strong evidence that up- and down-regulation of miRNAs in diseases are anti-correlated, i.e. the dys-regulated miRNAs are either up- or down-regulated in diseases generally but very few miRNAs are up-regulated in several diseases while down-regulated in others. In our initial study [29], 62 miRNAs were found to be associated with over 40% of all tested disease conditions. Of these 62 miRNAs, 39 were found to be still dys-regulated in at least 40% of all diseases despite our substantial extension of the study.

Importantly, we found a substantial variance in miRNA expression related to human pathologies. Considering single diseases we found the highest number of 408 significantly dys-regulated miRNAs in the case of colon cancer and melanoma. The lowest number with 115 dys-regulated miRNAs was detected for pancreatitis. For each disease we were furthermore able to detect a unique signature, i.e. a combination of significant miRNAs that did not overlapped with any other signature, allowing for specific differentiating between normal controls and diseases.

Besides the comparison between controls and diseases we also asked for specific signatures between diseases at all. Altogether, our study includes 20 different classes, 19 diseases as well as controls. Thus, a total of 190 specific signatures, one for each possible pair of the 20 cohorts, can be calculated. We carried out all comparisons and computed the number of miRNAs

significant in each comparison as well as the number of comparisons where a certain miRNA was found to be significant. Thereby, we detected an average of 256 significant miRNAs per comparison. While some miRNAs were significant in many scenarios (including hsa-miR-106a (130 comparisons), hsa-miR-361-5p (130 comparisons), hsa-miR-17 (125 comparisons), hsa-miR-423-5p (125 comparisons), hsa-miR-320d (122 comparisons) and hsa-miR-20a (120 comparisons)), other were significantly dys-regulated just in few comparisons (including hsa-miR-506 (3 comparisons), hsa-miR-202* (5 comparisons), hsa-miR-361-3p (6 comparisons), hsa-miR-429 (7 comparisons), hsa-miR-548a-3p (9 comparisons) or hsa-miR-518e (9 comparisons)). All disease specific signatures are detailed in Supplemental Table 5. Especially the miRNAs that are significant in many different comparisons show a substantial data variance. To further evaluate this, we carried out an analysis of variance (ANOVA). Even after adjustment for multiple testing all but 19 miRNAs (2.2%) were significant in our ANOVA. Highest significance was reached for hsa-miR-151-3p (p-value of $4.03*10^{-89}$). Among the most significant miRNAs in the ANOVA was also hsa-miR-144*, being significant in 14 different diseases and representing the most generally dys-regulated miRNA with significance value of $1.88*10^{-33}$. The values for the 19 diseases and miR-144* are shown in Figure 6A, where it can be seen that none of the 19 diseases showed higher averaged expression values for this miRNA as compared to the controls. Among the miRNAs with higher significance values we found hsa-miR-155*, being significantly down-regulated in just 2 diseases, namely acute myocardial infarction and glioma.

qRT-PCR validation of microarray data

To validate our microarray results for two important disease miRNAs, hsa-miR-144* (non-specific) and hsa-miR-155* (specific), qRT-PCR was performed in two participating centers. Center 1 (Heidelberg University) analyzed a total of 172 samples from controls, and patients with acute myocardial infarction, non-ischaemic systolic heart failure, glioblastoma and pancreatic diseases. Center 2 (Saarland University) analyzed a total of 110 samples from controls, and patients with Wilms tumor, psoriasis, renal cancer, prostate cancer, lung cancer, multiple sclerosis, benign prostate hyperplasia, colon cancer and chronic obstructive pulmonary disease samples.

For miR-144*, we measured $\Delta\Delta$CT values of -1.93 in center 1. Thus, hsa-miR-144* was down-regulated 3.8 fold in diseases (p-value of $1.9*10^{-5}$). In center 2 we calculated $\Delta\Delta$CT values of -1.75 for, thus concordantly hsa-miR-144* was significantly lower expressed in diseases (p-value of 0.0096) with a fold change of 3.4.

As an independent set of patients and controls, we selected a third cohort of samples, containing blood samples from controls and from breast cancer patients. Notably, this validation was independent, in that the phenotype has not been included in the initial microarray screening and likewise, this center did not contributed any samples to the initial screening (sample details are provided in Supplemental Table 6). The qRT-PCR was performed from center 1. The $\Delta\Delta$CT value was -1.79. As for the first two validation approaches hsa-miR-144* was significantly lower (p-value of 0.04) expressed with a fold change of 3.5 in breast cancer samples compared to controls. In summary, we were able to successfully validate that hsa-miR-144* was significantly down-regulated in various diseases at a total of 319 samples in three approaches with consistent fold changes of 3.8, 3.4 and 3.5 respectively.

Analogously to hsa-miR-144* as general disease marker, we also validated the miRNA hsa-miR-155* as example of a rather specific miRNA. In our microarray experiments hsa-miR-155* was only significantly down-regulated in two diseases, namely acute myocardial infarction and glioma. The validation in center 1 reached a highly significant p-value of $3.66*10^{-6}$, showing a significant down-regulation of this miRNA in diseases (on average 2.8-fold). Remarkably, as mentioned above the sample cohort analyzed in this center contained acute myocardial infarction samples and glioma samples, as well as non-ischaemic systolic heart failure, and pancreatic diseases. In discordance with the screening results we found, however, also down-regulation of miR-155* for pancreatic diseases. In the second validation in center 2, analyzing besides controls the diseases Wilms tumor, psoriasis, renal cancer, prostate cancer, lung cancer, multiple sclerosis, benign prostate hyperplasia, colon cancer, melanoma, and chronic obstructive pulmonary disease we found a slightly up-regulation of has-miR-155* at a moderate fold change of 1.7 with a non-adjusted significance value of 0.008. After adjusting for multiple testing only one of the 10 tested diseases (prostate cancer) remained significant. For the breast cancer samples against controls we likewise did not detect any statistically significant difference (p-value of 0.42), providing evidence that hsa-miR-155* is in contrast to hsa-miR-144* no general disease marker but only significant in a restricted subset of diseases.

Improvement of AUC values by combining multiple miRNAs.

As demonstrated, miRNAs have the potential to differentiate between controls and patients in general with high AUC values up to 0.75. By combining the predictive power of different miRNAs it can be expected that the diagnostic power increases. To test this hypothesis we employed a machine learning procedure. We applied a stepwise forward subset selection

approach with radial basis function Support Vector Machines (SVM) and carried out 10 random repetition of 10-fold cross-validation.

For the classification in control and disease samples we reached maximal AUC values of 0.911, as the ROC curve in Figure 4A demonstrates. Our classifier outperformed the maximal AUC of the best single biomarker, i.e. hsa-miR-144* and hsa-miR-20a (AUC 0.751, respectively), by 16%. Altogether we reached classification accuracy, specificity and sensitivity of 78%, 81%, and 75%, as the box-plot in Figure 4B details. These results are significantly improved as compared to random permutation tests, presented as blue boxes in Figure 4B (p-value $<10^{-10}$). Figure 4C presents the classification example leading to the best AUC of 0.911, providing evidence that the majority of the samples have been classified correctly.

For the comparison of cancer versus controls the highest AUC was as high as 0.94, representing a 16.9% improvement over the best single miRNA for this comparison (hsa-miR-144*). Overall, classification accuracy of 82%, specificity of 81% and sensitivity of 83% were reached.

Discussion

For most diseases, early and specific markers are lacking. Hence, besides the continuous refinement of existing biomarkers, the search for novel, early disease predictors is belonging to the current challenges in biomarker research. miRNAs offer a new class of biologically active molecules that contribute to many disease processes and compensatory mechanisms. Accordingly, they might not only offer the ability to detect a disease early, but could also complement existing molecular and clinical markers by providing additional information, supporting a biomarker-guided differential diagnosis. Furthermore, miRNA signatures could support a differential diagnosis in clinically overlapping diseases, such as non-ischaemic systolic heart failure versus acute myocardial infarction. Consequently, miRNAs are increasingly recognized as valuable biomarkers for different pathologies. However, in most studies a case-control scenario has been applied and comprehensive comparisons between different diseases are largely missing.

The current meta-analysis aimed to compare the miRNA profiles from 1,049 samples belonging to 19 different disease as well as controls. Here, we not only identified disease specific miRNAs but also miRNAs associated with the presence of a disease in general. Moreover, we were able to show that miRNA patterns improve the diagnostic accuracy substantially and provide the required specificity for diagnostic purposes.

One key miRNA in our analysis is hsa-miR-144*, being down-regulated in almost all tested disease conditions. According to the Human MIRNA & Diseases Database (HMDD, [3, 33, 34]) several studies revealed hsa-miR-144* (in the current V20 miRBase: hsa-miR-144-5p) as disease-associated. Liu et al showed that hsa-miR-144* is over-expressed in PBMC of active TB patients [35] and Redova et al. showed that it is also up-regulated in serum of patients with renal cell carcinoma compared to healthy controls [36]. hsa-miR-144* was further identified as a new fecal-based marker for colon cancer [37], and as significantly up-regulated in primary Medulloblastoma samples compared to Neural Stem Cells [38], but as down-regulated in esophageal biopsy specimens of eosinophilic esophagitis patients [39]. These heterogeneous results demonstrate that it is essential to keep blood collection, miRNA extraction and measurement and bioinformatics as constant as possible.

The dataset used for our meta-analysis has been generated over three years and the samples have been collected at 9 different institutions. An obvious confounding variable that also may limit the applicability of miRNAs in clinical routine is the storage of samples over time. To minimize this we used PAXgene tubes containing RNA stabilizing agents, allowing for storage of samples between -20 to -70 degree for up to 50 months. We additionally checked the storage of RNA samples over a period of up to four months at -20 degree Celsius. After two months we still reached correlation of 0.89, which is well in the range of the platforms technical reproducibility for blood samples. Even after four months we still reached correlation of 0.865 (detail in Supplemental Figure 1). For serum samples we were even able to show the stability of miRNA expression for much longer periods of time (up to three decades) [16]. The highly consistent and significant results obtained in our meta-analysis thus confirm the robustness of the approach. Nevertheless, the identified signatures especially for the smaller cohorts need to be replicated in future studies in larger independent patient's cohorts. Additionally, in prospective studies one needs to investigate the outcome given clinical end-points associated with the different disease signatures. Another potential reason, which may delay or even hinder the translation in clinical routine is the measurement system. miRNAs are relatively stable molecules and their quantification can be achieved by different methodologies. As such, miRNA quantification by PCR based approaches shows a very high dynamic range and allows for absolute quantification, thus enabling testing in clinical routine. Furthermore, techniques for measuring sets of miRNAs as qRT-PCR are relatively inexpensive, fast and established in most clinical laboratories, enabling testing in clinical routine.

In summary, we present a substantial meta-analysis of high-throughput miRNA data from patients' blood samples. Our study presents miRNAs that are dys-regulated in almost all patients, such as miR-144*, which was also validated using qRT-PCR. Moreover we were able to present specific miRNA patterns for all diseases and for all inter-disease comparisons besides few cases such as the separation of pancreatistis from pancreatic cancer. Finally, we were able to report sets of miRNAs being dys-regulated in specific diseases, further promoting the investigation of miRNAs from peripheral blood as clinically relevant information carriers.

Methods

Blood samples: The blood samples were collected and processed from nine different institutions (see Table 1). Five centers provided samples from individuals with disease as well as controls from healthy individuals. Blood was collected in PAXgene Blood RNA tubes (BD). All blood donors participating in this study gave their informed consent. An overview of all patients is presented in Supplemental Table 1.

miRNA extraction and microarray screening: miRNA extraction and microarray measurement have been carried out as previously described [29].

Statistical analysis: All statistical computations were carried out using the publicly available statistical language R (http://www.r-project.org/). To assess the information content of single miRNAs and miRNA profiles the area under the receiver operator characteristics curve (AUC) was computed using the pROC package. The 95% confidence intervals for the ROC curves and AUC values were calculated using 2,000 bootstrap samples. To determine significance values for miRNAs, 2-tailed unpaired t-tests have been calculated and the significance values adjusted for multiple testing using the Benjamini-Hochberg approach.

Machine Learning Analysis: Supervised classification of samples was carried out using linear Support Vector Machines (SVM) as implemented in the R e1071 package. SVMs were evaluated by applying standard 10-fold cross-validation and a stepwise-forward filter subset selection technique. In order to account for variations in the random partitioning into sample subsets, cross-validation runs were repeated 10 times. Moreover, to test for potential overtraining, exactly the same procedure has been carried out using randomly permuted class

labels, such that 10 so-called permutation tests were applied for each subset size. All classifications have been carried out with equal cohort sizes, i.e., if one group was larger than the other, samples from the first group have been randomly selected in each repetition in order to simulate same cohort sizes.

qRT-PCR validation: qRT-PCR was performed in two participating centers (center 1: Heidelberg University, center 2: Saarland University) using the miScript PCR System (Qiagen) and the primer assays for hsa-miR-144* and hsa-miR-155*. We analyzed the expression of these two miRNAs in a total of 282 samples (center 1: 172 samples from controls, and patients with acute myocardial infarction, non-ischaemic systolic heart failure, glioblastoma, pancreatic diseases, and breast cancer; center 2: 110 samples from controls, and patients with Wilms tumor, psoriasis, renal cancer, prostate cancer, lung cancer, multiple sclerosis, benign prostate hyperplasia, colon cancer and chronic obstructive pulmonary disease). Additionally, a third cohort was included, providing 37 samples of a breast cancer study, a phenotype that was not included in the screening. As endogenous control, RNU6B was measured.

References

1. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 1993; **75**: 843-854.
2. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* 2011; **39**: D152-157.
3. Griffiths-Jones S, Grocock RJ, van Dongen S*, et al.* miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* 2006; **34**: D140-144.
4. Pogue AI, Cui JG, Li YY*, et al.* Micro RNA-125b (miRNA-125b) function in astrogliosis and glial cell proliferation. *Neurosci Lett* 2010; **476**: 18-22.
5. Cheng AM, Byrom MW, Shelton J*, et al.* Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. *Nucleic Acids Res* 2005; **33**: 1290-1297.
6. Yoshizawa M, Taguchi YH, Yasuda J. Inference of Gene Regulation via miRNAs During ES Cell Differentiation Using MiRaGE Method. *Int J Mol Sci* 2011; **12**: 9265-9276.
7. Laczny C, Leidinger P, Haas J*, et al.* miRTrail - a comprehensive webserver for analyzing gene and miRNA patterns to enhance the understanding of regulatory mechanisms in diseases. *BMC Bioinformatics* 2012; **13**: 36.
8. van Kouwenhove M, Kedde M, Agami R. MicroRNA regulation by RNA-binding proteins and its implications for cancer. *Nat Rev Cancer* 2011; **11**: 644-656.
9. Backes C, Meese E, Lenhof HP*, et al.* A dictionary on microRNAs and their putative target pathways. *Nucleic Acids Res* 2010; **38**: 4476-4486.
10. Yu SL, Chen HY, Chang GC*, et al.* MicroRNA signature predicts survival and relapse in lung cancer. *Cancer Cell* 2008; **13**: 48-57.
11. Ma L, Teruya-Feldstein J, Weinberg RA. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature* 2007; **449**: 682-688.

12. Chan JA, Krichevsky AM, Kosik KS. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer Res* 2005; **65**: 6029-6033.
13. Vogel B, Keller A, Frese KS*, et al.* Multivariate miRNA signatures as biomarkers for non-ischaemic systolic heart failure. *Eur Heart J* 2013.
14. Fu Y, Yi Z, Wu X*, et al.* Circulating microRNAs in patients with active pulmonary tuberculosis. *J Clin Microbiol* 2011; **49**: 4246-4251.
15. Wang ZX, Bian HB, Wang JR*, et al.* Prognostic significance of serum miRNA-21 expression in human non-small cell lung cancer. *J Surg Oncol* 2011; **104**: 847-851.
16. Keller A, Leidinger P, Gislefoss R*, et al.* Stable serum miRNA profiles as potential tool for non-invasive lung cancer diagnosis. *RNA Biol* 2011; **8**: 506-516.
17. Gotte M. MicroRNAs in breast cancer pathogenesis. *Minerva Ginecol* 2010; **62**: 559-571.
18. Yin Y, Li M, Li H*, et al.* [Expressions of 6 microRNAs in prostate cancer]. *Zhonghua Nan Ke Xue* 2010; **16**: 599-605.
19. Resnick KE, Alder H, Hagan JP*, et al.* The detection of differentially expressed microRNAs from the serum of ovarian cancer patients using a novel real-time PCR platform. *Gynecol Oncol* 2009; **112**: 55-59.
20. Meder B, Keller A, Vogel B*, et al.* MicroRNA signatures in total peripheral blood as novel biomarkers for acute myocardial infarction. *Basic Res Cardiol* 2011; **106**: 13-23.
21. Keller A, Leidinger P, Borries A*, et al.* miRNAs in lung cancer - studying complex fingerprints in patient's blood cells by microarray experiments. *BMC Cancer* 2009; **9**: 353.
22. Keller A, Leidinger P, Lange J*, et al.* Multiple sclerosis: microRNA expression profiles accurately differentiate patients with relapsing-remitting disease from healthy controls. *PLoS One* 2009; **4**: e7440.
23. Keller A, Leidinger P, Steinmeyer F*, et al.* Comprehensive analysis of microRNA profiles in multiple sclerosis including next-generation sequencing. *Mult Scler* 2013.
24. Leidinger P, Keller A, Borries A*, et al.* High-throughput miRNA profiling of human melanoma blood samples. *BMC Cancer* 2010; **10**: 262.
25. Hausler SF, Keller A, Chandran PA*, et al.* Whole blood-derived miRNA profiles as potential new tools for ovarian cancer screening. *Br J Cancer* 2010; **103**: 693-700.
26. Leidinger P, Keller A, Borries A*, et al.* Specific peripheral miRNA profiles for distinguishing lung cancer from COPD. *Lung Cancer* 2011; **74**: 41-47.
27. Roth P, Wischhusen J, Happold C*, et al.* A specific miRNA signature in the peripheral blood of glioblastoma patients. *J Neurochem* 2011; **118**: 449-457.
28. Leidinger P, Backes C, Deutscher S*, et al.* A blood based 12-miRNA signature of Alzheimer disease patients. *Genome Biol* 2013; **14**: R78.
29. Keller A, Leidinger P, Bauer A*, et al.* Toward the blood-borne miRNome of human diseases. *Nat Methods* 2011; **8**: 841-843.
30. Vorwerk S, Ganter K, Cheng Y*, et al.* Microfluidic-based enzymatic on-chip labeling of miRNAs. *N Biotechnol* 2008; **25**: 142-149.
31. Keller A, Ludwig N, Comtesse N*, et al.* A minimally invasive multiple marker approach allows highly efficient detection of meningioma tumors. *BMC Bioinformatics* 2006; **7**: 539.
32. Wang W, Peng B, Wang D*, et al.* Human tumor microRNA signatures derived from large-scale oligonucleotide microarray datasets. *Int J Cancer* 2011; **129**: 1624-1634.
33. Human MiRNA & Disease Database.
34. Griffiths-Jones S, Saini HK, van Dongen S*, et al.* miRBase: tools for microRNA genomics. *Nucleic Acids Res* 2008; **36**: D154-158.
35. Liu Y, Wang X, Jiang J*, et al.* Modulation of T cell cytokine production by miR-144* with elevated expression in patients with pulmonary tuberculosis. *Mol Immunol* 2011; **48**: 1084-1090.

36. Redova M, Poprach A, Nekvindova J*, et al.* Circulating miR-378 and miR-451 in serum are potential biomarkers for renal cell carcinoma. *J Transl Med* 2012; **10**: 55.
37. Kalimutho M, Del Vecchio Blanco G, Di Cecilia S*, et al.* Differential expression of miR-144* as a novel fecal-based diagnostic marker for colorectal cancer. *J Gastroenterol* 2011; **46**: 1391-1402.
38. Genovesi LA, Carter KW, Gottardo NG*, et al.* Integrated analysis of miRNA and mRNA expression in childhood medulloblastoma compared with neural stem cells. *PLoS One* 2011; **6**: e23935.
39. Lu TX, Sherrill JD, Wen T*, et al.* MicroRNA signature in patients with eosinophilic esophagitis, reversibility with glucocorticoids, and assessment as disease biomarkers. *J Allergy Clin Immunol* 2011; **129**: 1064-1075 e1069.

Tables

Table 1: Cohorts with ICD10 code and cohort sizes

Disease	ICD10	# Samples	Institution providing RNA
Normal	-	94	Saarland University DKFZ / Heidelberg University Heidelberg University Julius-Maximilians-University Wuerzburg Zürich University Christian-Albrechts-University Kiel
long-lived individuals	-	15	Christian-Albrechts-University Kiel
tumor of stomach	C16	13	DKFZ / Heidelberg University
colon cancer	C18	29	Saarland University
lung cancer	C24	73	Saarland University
pancreatic ductal adenocarcinoma	C25	45	DKFZ / Heidelberg University
Melanoma	C43	35	Saarland University
ovarian cancer	C56	24	Julius-Maximilians-University Wuerzburg
prostate cancer	C61	65	Saarland University
wilms tumor	C64	124	Saarland University
renal cancer	C65	20	Saarland University
Glioma	C71	20	Zürich University
Sarcoidosis	D86.0	45	Albrecht Ludwigs University, Freiburg
multiple sclerosis (MS)	G35	23	Saarland University
acute myocardial infarction	I21.3	62	Heidelberg University
non-ischaemic systolic heart failure	I42	33	Heidelberg University
chronic obstructive pulmonary disease (COPD)	J40-47	47	Saarland University
Peridontitis	K05.4	18	Christian-Albrechts-University Kiel
Pancreatitis	K85	37	DKFZ / Heidelberg University
Psoriasis	L40	43	Saarland University
benign prostate hyperplasia (BPH)	N40	35	Saarland University
Others	-	149	

Figures

Figure 1: ROC curves for disease specific miRNAs. A. The ROC curve for hsa-miR-144* is shown. B. The ROC curves for hsa-miR-20b is shown. The blue shaded area denotes the 95% confidence interval computed by 2,000 bootstrap samples.

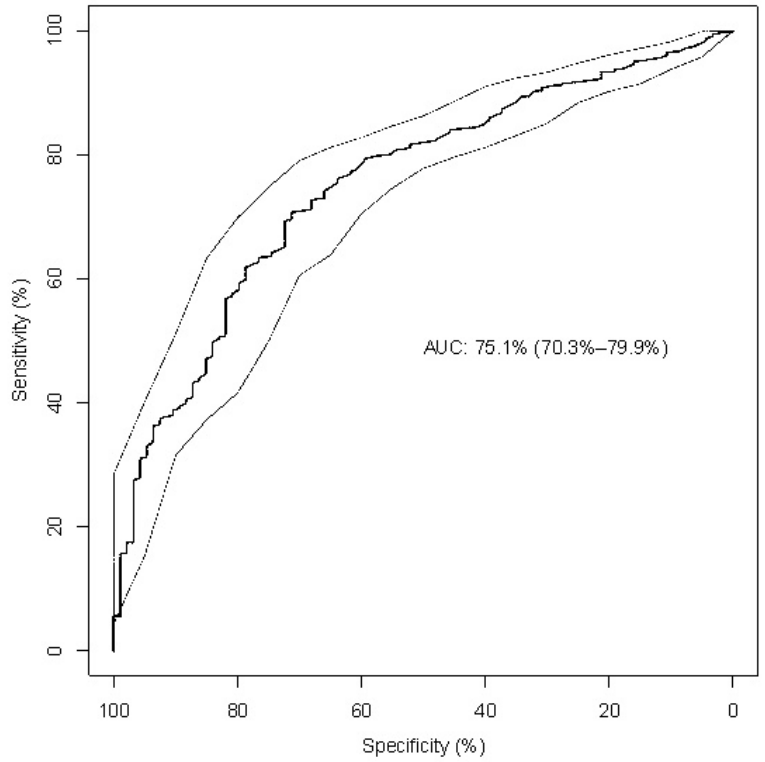
Figure 2: Area-proportional Venn diagram for miRNAs with highest AUC values in the comparisons disease versus healthy controls and cancer versus healthy controls. Green area shows up-regulated miRNAs while red area shows down-regulated miRNAs in cancer and diseases in general. Both comparisons show a high overlap of dys-regulated miRNAs, the respective miRNAs are presented on left and right of the Venn diagram.

Figure 3: Up- versus down-regulations. The balloon-plot shows for the different miRNAs in how many diseases the miRNAs are up- and respectively down-regulated. Orange bubbles belong to predominantly down-regulated while blue bubbles belong to predominantly up-regulated miRNAs. The two green bubbles represent 9 miRNAs that were equally up- and down regulated in disease.

Figure 4: Classification in patients (cancer and non-cancer) and controls. A. ROC curve for the best classification. B. Boxplots for accuracy, specificity and sensitivity for the 10 repeated cross validations in red and for 10 permutation tests in blue. C. The best classification. Samples above the horizontal black line are considered to be patients (denoted by 2) and below the black line as controls (denoted by 1).

Figure 1

A


Sensitivity (%)
100
80
60
40
20
0
AUC: 75.1% (70.3%–79.9%)
100
80
60
40
20
0
Specificity (%)

B

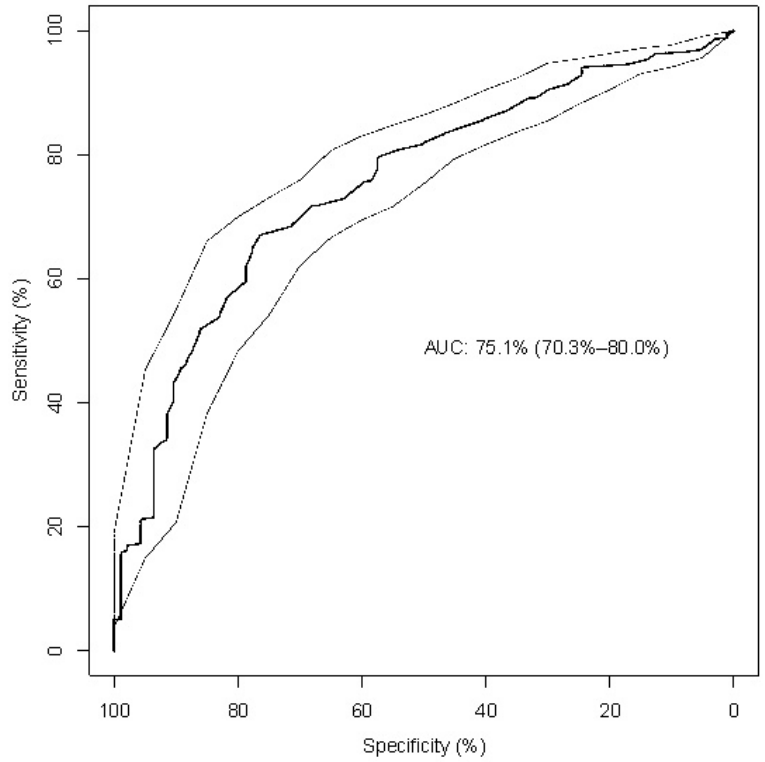

Sensitivity (%)
100
80
60
40
20
0
AUC: 75.1% (70.3%–80.0%)
100
80
60
40
20
0
Specificity (%)

Figure 2

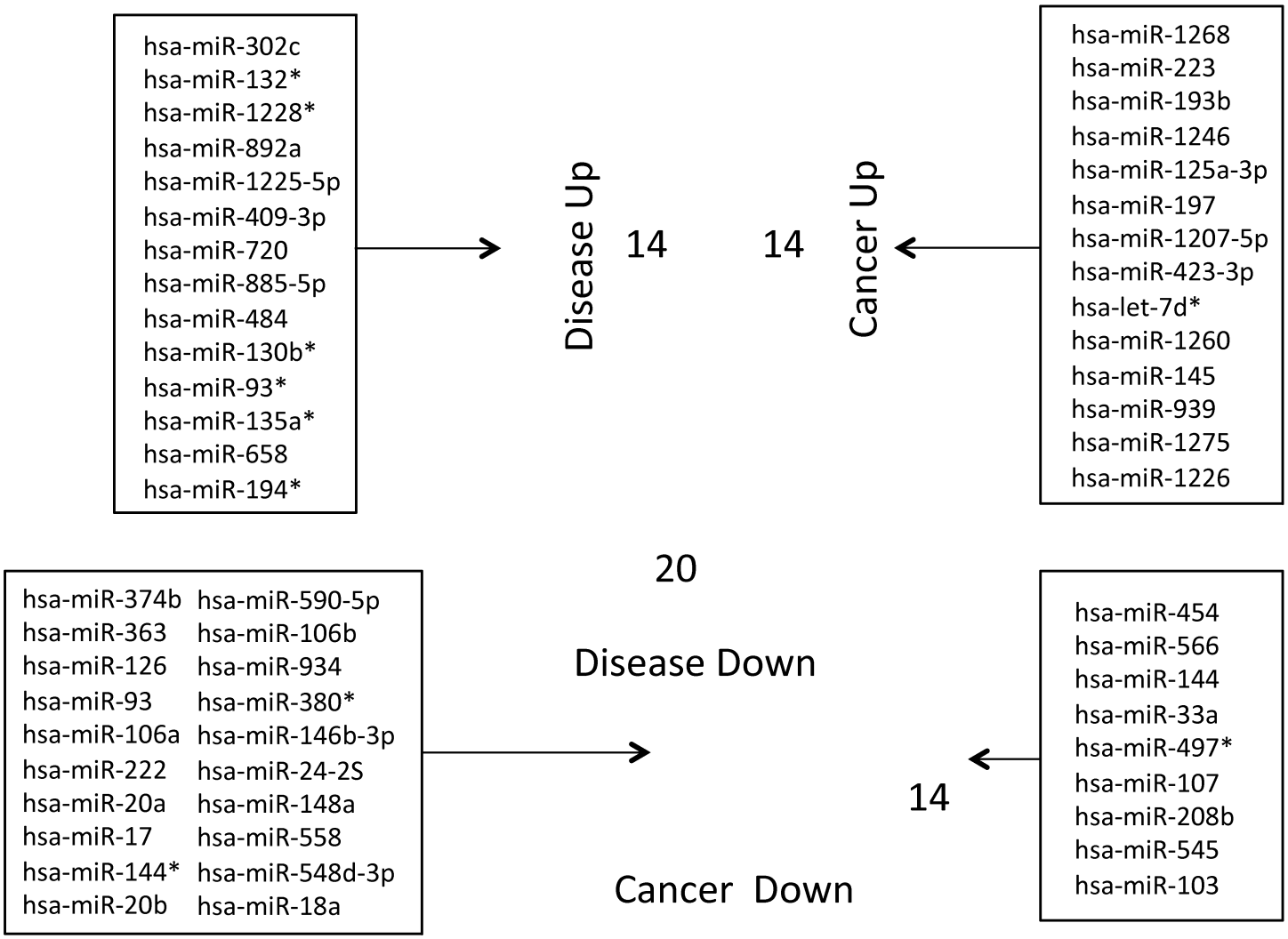

hsa-miR-302c
hsa-miR-132*
hsa-miR-1228*
hsa-miR-892a
hsa-miR-1225-5p
hsa-miR-409-3p
hsa-miR-720
hsa-miR-885-5p
hsa-miR-484
hsa-miR-130b*
hsa-miR-93*
hsa-miR-135a*
hsa-miR-658
hsa-miR-194*
Disease Up
14
14
Cancer Up
hsa-miR-1268
hsa-miR-223
hsa-miR-193b
hsa-miR-1246
hsa-miR-125a-3p
hsa-miR-197
hsa-miR-1207-5p
hsa-miR-423-3p
hsa-let-7d*
hsa-miR-1260
hsa-miR-145
hsa-miR-939
hsa-miR-1275
hsa-miR-1226
20
Disease Down
hsa-miR-374b
hsa-miR-363
hsa-miR-126
hsa-miR-93
hsa-miR-106a
hsa-miR-222
hsa-miR-20a
hsa-miR-17
hsa-miR-144*
hsa-miR-20b
hsa-miR-590-5p
hsa-miR-106b
hsa-miR-934
hsa-miR-380*
hsa-miR-146b-3p
hsa-miR-24-2S
hsa-miR-148a
hsa-miR-558
hsa-miR-548d-3p
hsa-miR-18a
14
Cancer Down
hsa-miR-454
hsa-miR-566
hsa-miR-144
hsa-miR-33a
hsa-miR-497*
hsa-miR-107
hsa-miR-208b
hsa-miR-545
hsa-miR-103

Figure 3

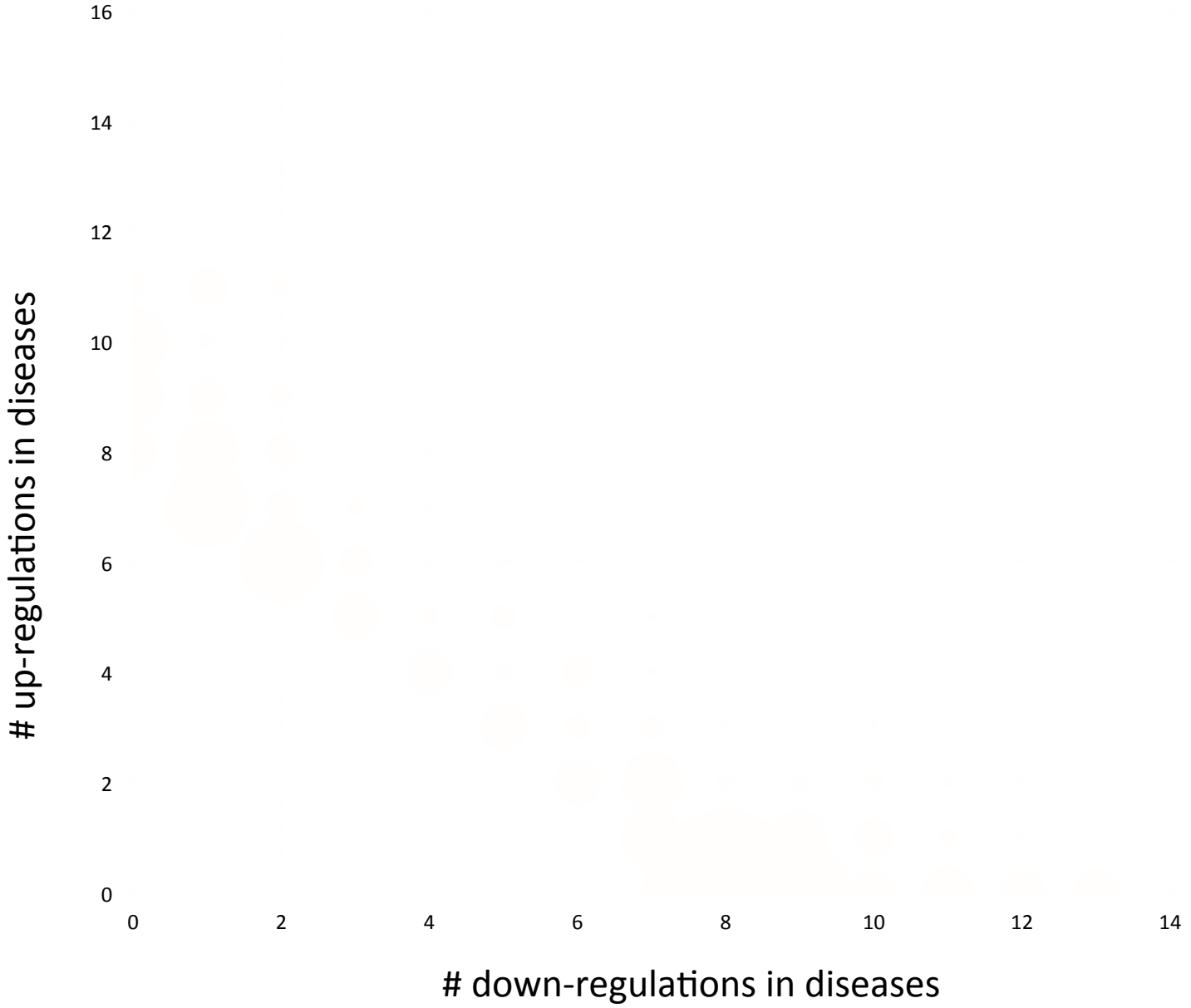
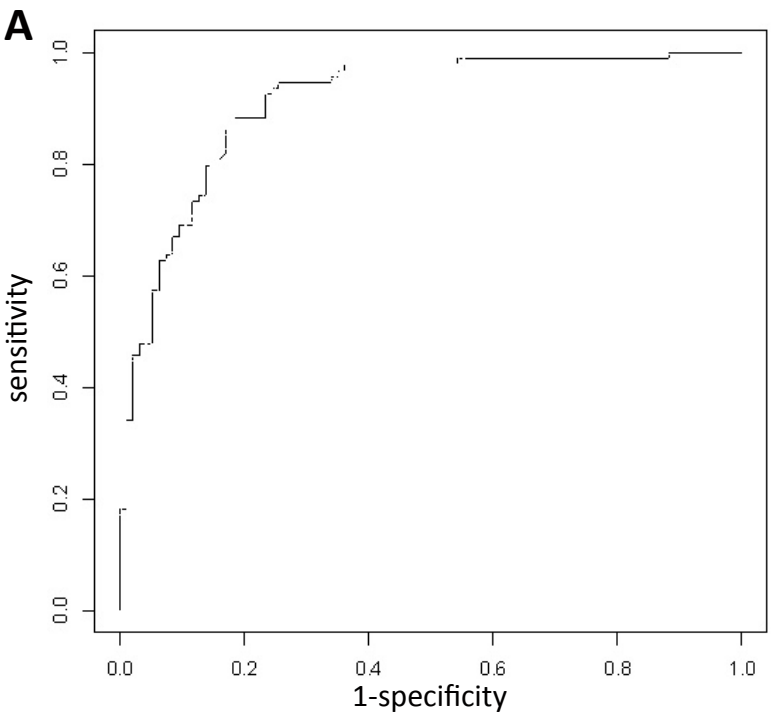
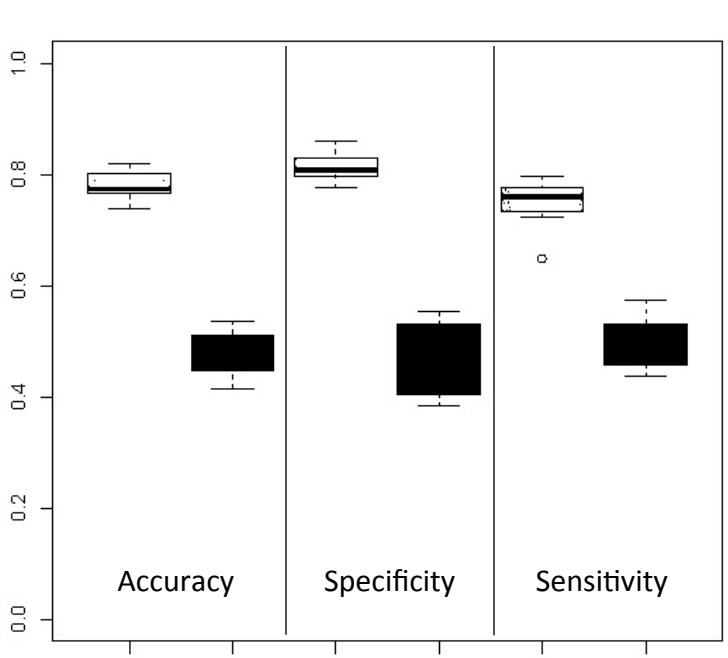
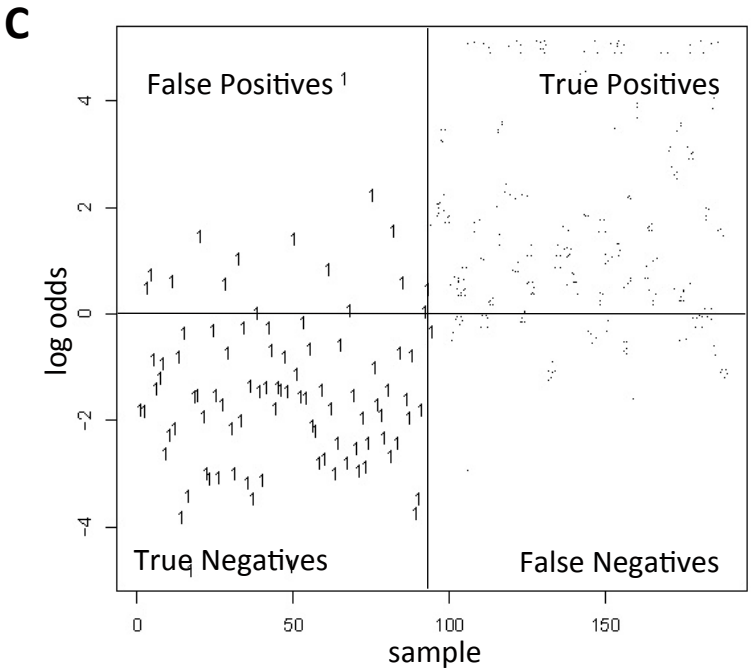

16
14
12
10
8
6
4
2
0
up-regulations in diseases
0
2
4
6
8
10
12
14
down-regulations in diseases

Figure 4


A
sensitivity
1-specificity
0.0
0.2
0.4
0.6
0.8
1.0


B
Accuracy
Specificity
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0


C
False Positives
True Positives
True Negatives
False Negatives
log odds
sample
-4
-2
0
2
4
0
50
100
150