

miRNA Signatures in Sera of Patients with Active Pulmonary Tuberculosis

Paolo Miotto^{1,9}, Grace Mwangoka^{2,9}, Ilaria C. Valente¹, Luca Norbis¹, Giovanni Sotgiu³, Roberta Bosu³, Alessandro Ambrosi⁴, Luigi R. Codecasa⁵, Delia Goletti⁶, Alberto Matteelli⁷, Elias N. Ntinginya⁸, Francesco Aloï^{9,10}, Norbert Heinrich¹¹, Klaus Reither^{2,12}, Daniela M. Cirillo^{1*}

1 Emerging Bacterial Pathogens Unit, Division of Immunology, Transplantation and Infectious Diseases, San Raffaele Scientific Institute, Milan, Italy, **2** Ifakara Health Institute-Bagamoyo Research and Training Centre, Bagamoyo, United Republic of Tanzania, **3** Clinical Epidemiology and Medical Statistics Unit, Department of Biomedical Sciences, University of Sassari; Research, Medical Education and Professional Development Unit, AOU Sassari, Sassari, Italy, **4** Vita-Salute San Raffaele University, Milan, Italy, **5** Regional Reference Center for TB "Villa Marelli", Niguarda Ca' Granda Hospital, Milan, Italy, **6** Translational research Unit, National Institute for Infectious Diseases "L. Spallanzani", Rome, Italy, **7** Institute of Infectious and Tropical Diseases, World Health Organization Collaborating Centre for TB/HIV co-infection, Brescia University, Brescia, Italy, **8** National Institute of Medical Research-Mbeya Medical Research Centre, Mbeya, United Republic of Tanzania, **9** St. Francis Nsambya Hospital/AISPO, Kampala, Uganda, **10** Italian Association for Solidarity Among People (AISPO NGO), Kampala, Uganda, **11** Division for Infectious Diseases and Tropical Medicine, Ludwig Maximilian University, Munich, Germany, **12** Swiss Tropical and Public Health Institute and University of Basel, Basel, Switzerland

Abstract

Several studies showed that assessing levels of specific circulating microRNAs (miRNAs) is a non-invasive, rapid, and accurate method for diagnosing diseases or detecting alterations in physiological conditions. We aimed to identify a serum miRNA signature to be used for the diagnosis of tuberculosis (TB). To account for variations due to the genetic makeup, we enrolled adults from two study settings in Europe and Africa. The following categories of subjects were considered: healthy (H), active pulmonary TB (PTB), active pulmonary TB, HIV co-infected (PTB/HIV), latent TB infection (LTBI), other pulmonary infections (OPI), and active extra-pulmonary TB (EPTB). Sera from 10 subjects of the same category were pooled and, after total RNA extraction, screened for miRNA levels by TaqMan low-density arrays. After identification of "relevant miRNAs", we refined the serum miRNA signature discriminating between H and PTB on individual subjects. Signatures were analyzed for their diagnostic performances using a multivariate logistic model and a Relevance Vector Machine (RVM) model. A leave-one-out-cross-validation (LOOCV) approach was adopted for assessing how both models could perform in practice. The analysis on pooled specimens identified selected miRNAs as discriminatory for the categories analyzed. On individual serum samples, we showed that 15 miRNAs serve as signature for H and PTB categories with a diagnostic accuracy of 82% (CI 70.2–90.0), and 77% (CI 64.2–85.9) in a RVM and a logistic classification model, respectively. Considering the different ethnicity, by selecting the specific signature for the European group (10 miRNAs) the diagnostic accuracy increased up to 83% (CI 68.1–92.1), and 81% (65.0–90.3), respectively. The African-specific signature (12 miRNAs) increased the diagnostic accuracy up to 95% (CI 76.4–99.1), and 100% (83.9–100.0), respectively. Serum miRNA signatures represent an interesting source of biomarkers for TB disease with the potential to discriminate between PTB and LTBI, but also among the other categories.

Citation: Miotto P, Mwangoka G, Valente IC, Norbis L, Sotgiu G, et al. (2013) miRNA Signatures in Sera of Patients with Active Pulmonary Tuberculosis. PLoS ONE 8(11): e80149. doi:10.1371/journal.pone.0080149

Editor: Sunil K. Ahuja, South Texas Veterans Health Care System and University Health Science Center San Antonio, United States of America

Received: June 19, 2013; **Accepted:** September 30, 2013; **Published:** November 21, 2013

Copyright: © 2013 Miotto et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement FP7-223681 to DMC (<http://www.tbpannet.org/>) and by the EDCTP TB CHILD project (IP.2009.32040.007). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: cirillo.daniela@hsr.it

These authors contributed equally to this work.

Introduction

Tuberculosis (TB) remains one of the most relevant infectious diseases with nearly 9 million cases and 1.4 million deaths per year worldwide [1]. Due to the complexity of the clinical presentations of the infection caused by members of the *Mycobacterium tuberculosis* complex (latent asymptomatic infection, active pulmonary and/or extra-pulmonary disease), accurate classification of cases is essential to address the most appropriate clinical management [2–4].

Current standards for TB diagnosis, including the most sensitive molecular tests, rely on the detection of the pathogen, thus being dependent on the bacterial load in the specimen analyzed; indeed,

diagnosing TB in children is a difficult task because the mycobacterial load is often low [5–7]. Similarly, extra-pulmonary TB (EPTB) cases are often challenging to diagnose due to the difficulties in obtaining samples for microbiological investigations, and are affected by unpredictable distribution of bacteria in tissues. Therefore, EPTB and smear-negative pulmonary TB (PTB) are usually diagnosed *ex juvantibus*. Finally, despite the importance of utterly discriminating between latent TB infection (LTBI) and active TB, clear-cut biological markers separating the two conditions are not yet available [8–10]. Biomarkers and surrogate endpoints are therefore crucial tools for the development of innovative strategies for TB management [11].

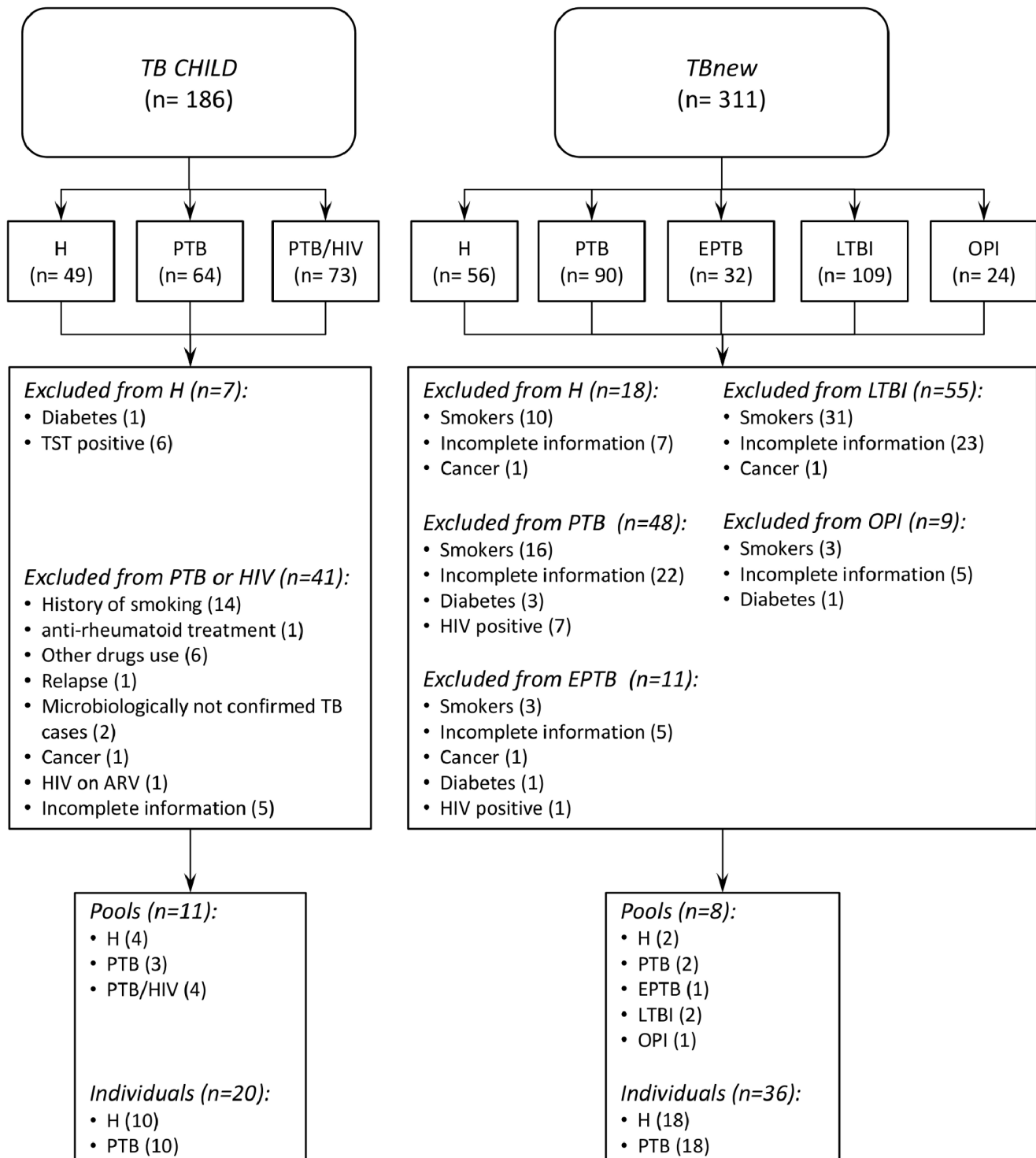


Figure 1. Criteria for the inclusion in the study population.
doi:10.1371/journal.pone.0080149.g001

The development of diagnostic tests based on host biomarkers is advocated to clinically categorize paucibacillary or not microbiologically-confirmed TB cases, and for proper identification of LTBI cases [11,12]. The need for biomarkers extends beyond the urgency for improved diagnostic tools: the assessment of the disease status and of the risk of progression to active disease or the assessment of treatment success early during therapy are critical

bottlenecks in the development of new vaccines and drugs, and could allow categorization of therapy based on individual risk of unfavorable outcome.

The ideal biomarker should be a stable molecule, and should be in sufficient amounts for easy detection in accessible body fluids [9,13]. The discovery that human microRNAs (miRNAs) expression is frequently altered in various diseases has uncovered a new

repertoire of molecular factors, which warrants investigation to further elucidate their role in physiology and disease [14]. In the cell, miRNAs post-transcriptionally regulate the expression level of target genes [15]. The miRNA highly-conserved regulatory system does not remain confined to the intracellular compartment: they can be transferred via body fluids (including plasma, serum, urine, and saliva), thus modulating translational responses by intercellular communication [16–19].

miRNAs could represent an ideal biomarker, owing to the sampling easiness, the inherent stability and the resilience [19,20]. However, their characterization during viral or bacterial infection has raised interest only recently, and data on TB remain limited to few studies [21–23].

This research aimed at identifying miRNA profiles associated with the different phases of *M. tuberculosis* infection (PTB, EPTB, LTBI), and non-tubercular lung infections as well as in healthy condition and showing their use as specific signatures for PTB diagnosis.

Materials and Methods

Ethical Statements

The protocol of the study was approved by the Ethical Committee of the San Raffaele Scientific Institute, Milano, Italy (GO/URC/ER/mm prot. N. 82/DG) and of the participating institutions in Uganda and Tanzania. The study was conducted in full compliance with the principles of the Declaration of Helsinki. All samples were collected from individuals who had signed an informed consent form for the purpose of the study and for cryopreservation of their biological samples.

Study Population

The following case definitions were adopted to categorize individuals enrolled in the study:

- Patients with active pulmonary TB (PTB): sputum smear microscopy positive for acid-fast bacilli, *M. tuberculosis* complex culture and/or Xpert MTB/RIF (Cepheid, Sunnyvale, CA)

positive patients with pulmonary disease, HIV negative individuals;

- Patients with active pulmonary TB and HIV co-infection (PTB/HIV): sputum smear microscopy positive for acid-fast bacilli, *M. tuberculosis* complex culture and/or Xpert MTB/RIF positive patients with pulmonary disease, HIV positive confirmed cases;
- Patients with active extra-pulmonary TB (EPTB): culture positive TB cases with any extra-pulmonary disease localization, HIV negative individuals;
- Latent TB infection cases (LTBI): subjects who resulted interferon- γ release assay (IGRA) or tuberculin skin test (TST) positive with no signs/symptoms of active disease;
- Subjects affected by pulmonary infectious diseases other than TB (OPI): clinical diagnosis (clinical exam plus imaging), with or without microbiological confirmation, IGRA and/or TST negative;
- Healthy subjects (H): IGRA and/or TST negative, without any known risk factors for LTBI, without any clinically relevant conditions.

The study populations of adult subjects were enrolled between September 2009 and December 2012 from two studies, namely *TBnew* and *TB CHILD*.

TBnew. Subjects enrolled at the San Raffaele Hospital (Milano, Italy), Spedali Civili of Brescia (Brescia, Italy), Regional Center for TB “Villa Marelli”, Niguarda Hospital (Milano, Italy), and at the National Institute for Infectious Diseases “L. Spallanzani” (Roma, Italy) belonged to the following categories: PTB, EPTB, LTBI, H, OPI.

TB CHILD. Subjects enrolled at Ifakara Health Institute - Bagamoyo Research and Training Centre (BRTC) (Pwani, Tanzania), NIMR-Mbeya Medical Research Programme (MMRC) (Mbeya, Tanzania), and at the St. Francis Nsambya Hospital (Kampala, Uganda) belonged to the following categories: PTB, PTB/HIV, H.

Population	Category	M	F	tot	Age
<i>TBnew</i>	EPTB	16	16	32	39 (19-78)
	PTB	58	32	90	40 (19-83)
	LTBI	40	69	109	40 (20-75)
	H	31	25	56	23 (19-51)
	OPI	14	10	24	61 (32-85)
	tot	159	152	311	39 (19-85)
<i>TB CHILD</i>	PTB	43	21	64	36 (18-73)
	H	27	22	49	22 (18-32)
	PTB/HIV	37	36	73	35 (19-61)
	tot	107	79	186	53 (18-73)

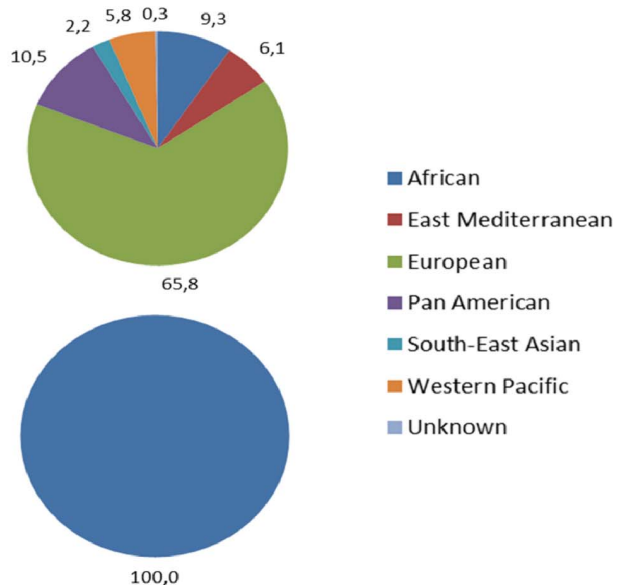


Figure 2. Demographic data of the *TBnew* and the *TB CHILD* populations. The Figure reports mean age for the categories of subjects enrolled and distribution of the country of birth in the different World Health Organization regions. doi:10.1371/journal.pone.0080149.g002

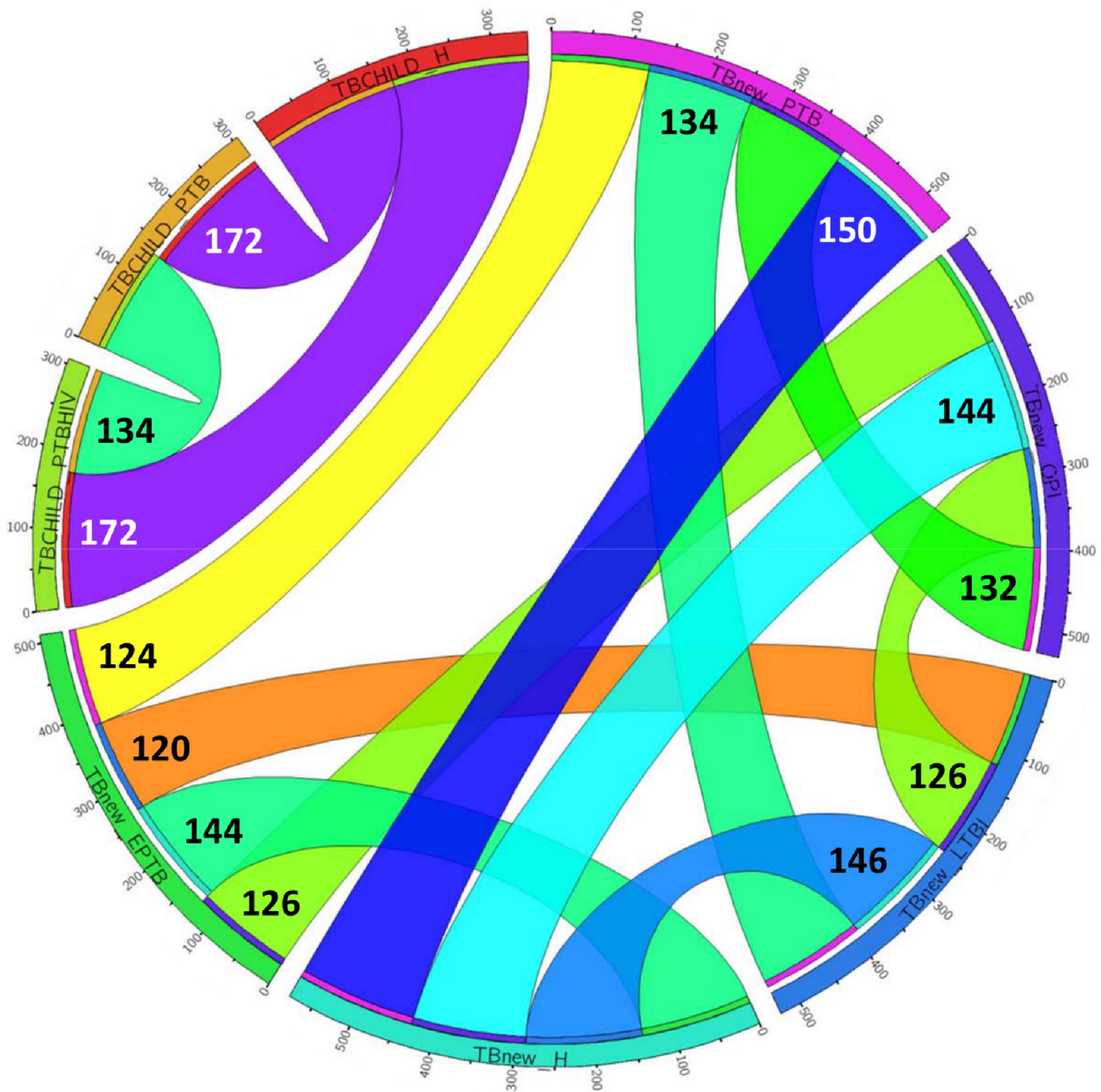


Figure 3. Summary of candidate serum miRNAs selected as relevant to discriminate among the categories according to the analysis of pooled specimens. Population and category of pooled specimens are reported on the circumference. The thickness of the ribbons connecting two categories is proportional to the number of miRNAs potentially interesting in the discrimination between the categories linked. doi:10.1371/journal.pone.0080149.g003

All subjects included in the study underwent the following procedures:

- phlebotomy through a 21G butterfly device to minimize hemolysis during specimen collection. Additive-free blood collection tubes were chosen to minimize unwanted modification of miRNA content in the serum;
- collection of clinically relevant data to determine TB status (with relevant confirmatory exams), and HIV status. A questionnaire to capture any additional information such as: pregnancy, smoking, current medical problems (diabetes,

transplant, silicosis, sarcoidosis, cancer), current therapies with particular focus on immunosuppressive, antiretroviral and anti-TB ones was compiled for all patients during enrolment.

All information was stored in an electronic data-protection system.

Enrolment and exclusion criteria of the study population are summarized in Figure 1.

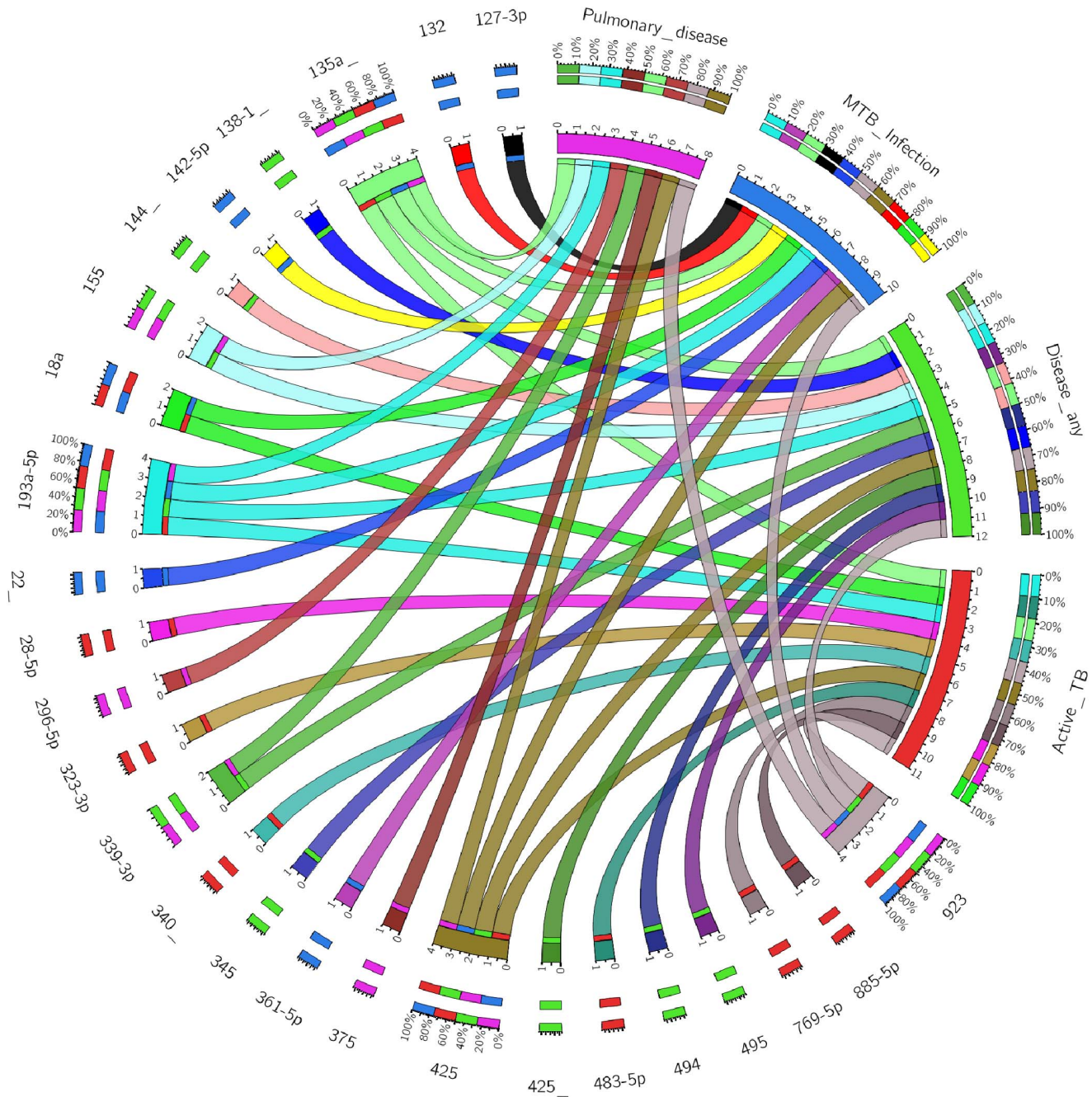


Figure 4. Summary of putative serum miRNA signatures discriminating among relevant clinical categories according to the analysis of pooled specimens. Clinical category and miRNAs are reported on the circumference. The ribbons represent the connection between miRNAs and clinical categories. The serum miRNAs have been defined by the following filters: Infection: H vs PTB, H vs LTBI, H vs EPTB, H vs PTB/HIV, OPI vs PTB, OPI vs LTBI, OPI vs EPTB; Active TB: H vs PTB, H vs EPTB, H vs PTB/HIV, OPI vs PTB, OPI vs EPTB, LTBI vs PTB, LTBI vs EPTB; Symptoms: H vs PTB, H vs OPI, H vs PTB/HIV, H vs EPTB, LTBI vs PTB, LTBI vs EPTB, LTBI vs OPI; Pulmonary disease (any): H vs PTB, H vs OPI, H vs PTB/HIV, LTBI vs PTB, LTBI vs OPI, EPTB vs PTB, EPTB vs OPI.
doi:10.1371/journal.pone.0080149.g004

Serum Preparation

Within 4 hours from time of phlebotomy, after coagulation, tubes were centrifuged at 2500 rpm for 10 minutes, with separation of serum from corpuscular fractions. Subsequently, serum fraction was transferred – in sterile environment – into 15 mL tubes and underwent a second centrifugation at 2800 rpm for 10 minutes, in order to achieve maximum removal of cellular component. Serum was hence transferred in aliquot of 1 mL into

cryogenic vials and stored at -80°C . For RNA extraction, sera were thawed on ice and the degree of hemolysis was determined through spectrometry analysis of free hemoglobin as previously described [24]. A cut-off value of hemoglobin concentration of >10 mg/dL was considered for hemolyzed samples.

Table 1. Serum miRNAs showing different levels in healthy (H) and pulmonary active tuberculosis (PTB) subjects from the two populations included in the study.

miRNA	Array	TBnew vs TB CHILD (p-val)	H vs PTB (p-val)
has-miR-155-4395459	A	0,01375	0,40055
hsa-miR-126-4395339	A	0,02580	0,76655
hsa-miR-129-5p-4373171	A	0,02600	0,84640
hsa-miR-139-3p-4395424	A	0,00975	0,91465
hsa-miR-142-5p-4395359	A	0,00025	0,33310
hsa-miR-145-4395389	A	<0,00001	0,88630
hsa-miR-146b-5p-4373178	A	0,02525	0,10360
hsa-miR-148b-4373129	A	0,00170	0,84365
hsa-miR-150-4373127	A	0,01065	0,37600
hsa-miR-152-4395170	A	<0,00001	0,43940
hsa-miR-17-4395419	A	0,00630	0,50405
hsa-miR-184-4373113	A	0,00510	0,32595
hsa-miR-195-4373105	A	<0,00001	0,16715
hsa-miR-19a-4373099	A	0,04110	0,34870
hsa-miR-20b-4373263	A	0,00315	0,52460
hsa-miR-220c-4395322	A	0,00100	0,29480
hsa-miR-29c-4395171	A	0,02855	0,29935
hsa-miR-302c-4378072	A	<0,00001	0,55550
hsa-miR-324-3p-4395272	A	0,00805	0,75735
hsa-miR-331-3p-4373046	A	0,00455	0,43190
hsa-miR-374b-4381045	A	0,00745	0,36145
hsa-miR-423-5p-4395451	A	0,04475	0,15395
hsa-miR-485-3p-4378095	A	<0,00001	0,54070
hsa-miR-574-3p-4395460	A	0,04470	0,05040
hsa-miR-597-4380960	A	0,00025	0,98915
hsa-miR-628-5p-4395544	A	0,01450	0,32615
hsa-miR-744-4395435	A	0,00100	0,40245
hsa-miR-872-4395375	A	<0,00001	0,06135
hsa-miR-9-4373285	A	0,04985	0,21925
MammU6-4395470	A	0,00140	0,51640
hsa-miR-135a*-4395343	B	0,00020	0,22165
hsa-miR-509-3p-4395347	B	0,01160	0,19885
hsa-miR-645-4381000	B	0,00005	0,27560
hsa-miR-801-4395183	B	<0,00001	0,87435
hsa-miR-923-4395264	B	0,00080	0,06465
MammU6-4395470	B	0,00045	0,58190
hsa-let-7e-4395517	A	0,80890	0,00395*
hsa-miR-10b-4395329	A	0,79615	0,0072*
hsa-miR-127-5p-4395340	A	0,96665	0,00495*
hsa-miR-130a-4373145	A	0,09690	0,02485
hsa-miR-146a-4373132	A	0,14045	0,0047*
hsa-miR-148a-4373130	A	0,91780	<0,00001*
hsa-miR-16-4373121	A	0,17145	0,00045*
hsa-miR-185-4395382	A	0,47350	0,00685*
hsa-miR-19b-4373098	A	0,42240	0,01720
hsa-miR-24-4373072	A	0,29355	0,04625
hsa-miR-25-4373071	A	0,18345	0,00065*
hsa-miR-27a-4373287	A	0,19845	0,01235

Table 1. Cont.

miRNA	Array	TBnew vs TB CHILD (p-val)	H vs PTB (p-val)
hsa-miR-27b-4373068	A	0,12255	0,03670
hsa-miR-342-3p-4395371	A	0,77485	0,01620
hsa-miR-365-4373194	A	0,48025	0,00345*
hsa-miR-374a-4373028	A	0,17845	0,03010
hsa-miR-376c-4395233	A	0,32505	0,04740
hsa-miR-451-4373360	A	0,50895	0,0082*
hsa-miR-532-5p-4380928	A	0,09210	0,00155*
hsa-miR-590-5p-4395176	A	0,46820	0,0039*
hsa-miR-660-4380925	A	0,48025	0,00095*
hsa-miR-885-5p-4395407	A	0,86025	0,003*
RNU48-4373383	A	0,90280	0,02350
hsa-miR-144*-4395259	B	0,63015	0,02595
hsa-miR-223*-4395209	B	0,79420	0,00025*
hsa-miR-30a-4373061	B	0,14410	0,00065*
hsa-miR-30a*-4373062	B	0,54630	0,03605
hsa-miR-30d-4373059	B	0,43060	0,00715
hsa-miR-30e-4395334	B	0,50865	0,0012*
hsa-miR-106a-4395280	A	0,03535	0,01145
hsa-miR-125a-5p-4395309	A	0,00380	0,03295
hsa-miR-192-4373108	A	0,04850	0,001*
hsa-miR-193a-5p-4395392	A	0,01675	0,00485*
hsa-miR-212-4373087	A	<0,00001	0,02215
hsa-miR-483-5p-4395449	A	0,00120	0,03800
hsa-miR-518d-3p-4373248	A	0,03350	0,00155*

*miRNAs showing a p-adj <0.05.
doi:10.1371/journal.pone.0080149.t001

Serum Pools

Each pool for qRT-PCR analysis was composed of 10 non-smoker subjects from the same category, equally distributed by gender. Selected subjects were free of co-morbidities (as for the pathologies investigated during the enrolment interview) to avoid confounding effects on miRNA profiles. A serum aliquot from each subject was thawed on ice, and 500 µL of serum from each

sample were mixed together in order to obtain a homogeneous pooled serum sample. One mL of pooled serum was used for RNA extraction and subsequent miRNA analysis in duplicate.

Serum from Single Individuals

To refine the results obtained from pooled sera, we performed individual serum analysis on 18 subjects belonging to the PTB and

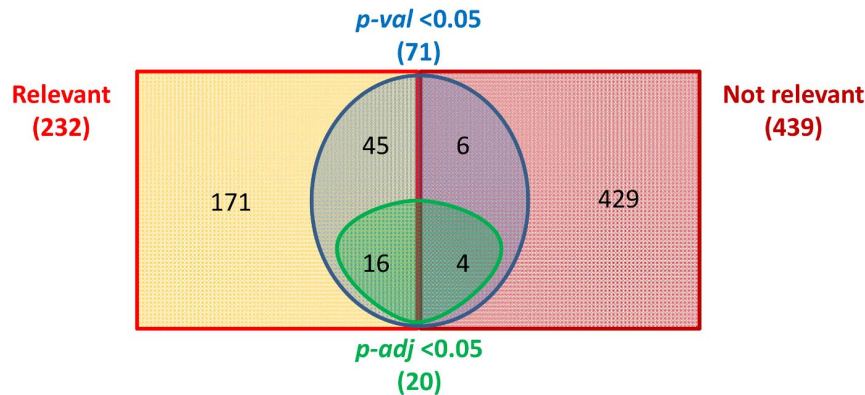


Figure 5. Analysis of the results of serum miRNAs in healthy controls (H) and pulmonary active tuberculosis (PTB) using pooled and individual specimens. Squares summarize results obtained by comparison of miRNAs in pooled specimens, whereas circles define results from individual sera.

doi:10.1371/journal.pone.0080149.g005

Table 2. Serum miRNA levels in pulmonary active tuberculosis (PTB) subjects as compared with healthy controls (H).

miRNAs p-adj <0.05	TBnew		TB CHILD		Combined		Relevance in pooled specimens	To be considered for		
	Individuals	Pools	Individuals	Pools	Individuals	Pools				
hsa-let-7e-4395517	↓	↓	↓	↓	↓	↓	TBnew, TB CHILD	TBnew, TB CHILD		
hsa-miR-146a-4373132	↓	↑	×	↓	↓	↑	TB CHILD only	TB CHILD only		
hsa-miR-148a-4373130	↑	↑	↑	↑	↑	↑	TBnew, TB CHILD	TBnew, TB CHILD		
hsa-miR-16-4373121	↑	↑	↑	↑	↑	↑	TBnew only	TBnew only		
hsa-miR-192-4373108	↑	↑	↑	↑	↑	↑	TBnew, TB CHILD	TBnew, TB CHILD		
hsa-miR-193a-5p-4395392	↑	↑	↑	↑	↑	↑	TBnew, TB CHILD	TBnew, TB CHILD		
hsa-miR-25-4373071	↑	↑	↑	↑	↑	↑	TBnew only	TBnew only		
hsa-miR-365-4373194	↑	↑	↑	↓	×	↑	TBnew only	TBnew only		
hsa-miR-451-4373360	↑	↑	↑	↑	↑	↑	TBnew, TB CHILD	TBnew, TB CHILD		
hsa-miR-532-5p-4380928	↑	↓	×	↑	↑	↑	TB CHILD only	TB CHILD only		
hsa-miR-590-5p-4395176	↑	↑	↑	↑	↑	↑	TBnew, TB CHILD	TBnew, TB CHILD		
hsa-miR-660-4380925	↑	↑	↑	↑	↑	↑	TB CHILD only	TB CHILD only		
hsa-miR-885-5p-4395407	↑	↑	↑	↑	↑	↑	TBnew, TB CHILD	TBnew, TB CHILD		
hsa-miR-223*-4395209	↑	↑	↑	↑	↑	↑	TB CHILD only	TB CHILD only		
hsa-miR-30a-4373061	↑	↓	×	↑	↓	×	↓	×	TBnew, TB CHILD	–
hsa-miR-30e-4395334	↑	↓	×	↑	↑	↓	×	TBnew, TB CHILD	TB CHILD only	

The table reports data from individual and pooled specimens. Discrepancies are marked by x.
doi:10.1371/journal.pone.0080149.t002

H categories from the *TBnew* group, and on 10 subjects from the *TB CHILD* group. As for pooled sera, individuals were selected free of co-morbidities among non smokers. A serum aliquot from each subject was thawed on ice, and 1 mL of serum from each individual was used for RNA extraction and subsequent miRNA analysis.

RNA Extraction

RNA extraction was performed using the mirVana miRNA isolation kit (Life Technologies) according to the manufacturer's instructions for isolating total RNA. RNA samples were stored at -80°C until use.

Serum miRNA Profiling

Serum miRNAs analysis was performed with TaqMan[®] Low Density Array (TLDA) Human MicroRNA Panels A and B, investigating an overall of 671 different miRNAs (Life Technologies). Retro-transcription was performed with TaqMan[®] microRNA reverse transcription kit and MegaPlex RT Primers Human Pool A v2.1 and B v2.0 components (Life Technologies), according to the manufacturer's instructions. For each RT reaction 15 ng of total RNA were used. After RT step, samples were pre-amplified using the MegaPlex PreAmp Primers A and B and the TaqMan[®] PreAmp Mater Mix according to the manufacturer's instructions (Life Technologies). TLDA were performed using a 7900HT Fast Real-Time PCR System or a ViiA[™] 7 Real-Time PCR System according to manufacturer's instructions (Life Technologies). On pooled sera, we performed analyses in biological and technical replicate for each array panel (A and B), whereas for each individual sera we performed single arrays (panels A+B).

Data Analysis and Normalization

Data from TaqMan[®] were collected with SDS v2.4.1, ViiA[™] 7 v1.1, and RQ Manager v1.2.2 software (Applied Biosystems),

(baseline: automatic; threshold: 0.20; maximum allowable C_t : 35.0). Data analysis was performed in R and Bioconductor environment [25]. We first preprocessed raw C_t values by means of quantile normalization, as described elsewhere [26–28]. This widely used approach is based on the assumption that only few miRNAs are differentially expressed. As a general result, this method provides homogenous data with the same distribution and the correlation coefficient between observations increases compared to raw data. Normalized data distribution was graphically inspected.

Statistical Analysis

Results from pools and individuals were analyzed separately.

Pools. We performed one-to-one category comparisons between mean C_t values fitting a constrained regression model with MM robust estimators [29,30]. These robust estimates have a high breakdown-point and are not affected by the presence of outliers or differently expressed miRNAs. We then computed the Empirical Distribution Function of residuals and filtered miRNAs associated with residuals outside the Inter Quartile range (*i.e.* outside the 1st quartile –3rd quartile interval) of the residuals distribution and we defined them as “interesting miRNAs” or “relevant miRNAs”. Circular visualization of data was made by Circos software [31].

Individuals. As our first step we filtered out miRNAs detected $C_t < 35$ in at least $\geq 80\%$ of subjects of at least one of the categories considered (H in *TBnew*, PTB in *TBnew*, H in *TB CHILD*, and PTB in *TB CHILD*). For these filtered out miRNAs we performed a two ways ANOVA for health status and genetic makeup (defined as for country of birth). P-values were computed non-parametrically by means of permutations [32]. We checked for False Discovery Rate (FDR) with the method described by Benjamini and Yekutieli [33]. miRNAs showing both (i) an adjusted p-value (*p-adj*) < 0.05 on individuals and (ii) relevant by

Table 3. Diagnostic performances of the serum miRNA signatures in the Relevance Vector Machine (RVM).

Serum miRNA signature (n)	All (15)	TBnew (10)	TB CHILD (12)
n of individuals (H/PTB)	56 (28/28)	36 (18/18)	20 (10/10)
Sensitivity % (95% CI)	85.71 (68.51–94.30)	77.78 (54.78–91.00)	100.00 (72.25–100.00)
Specificity % (95% CI)	78.57 (60.46–89.79)	88.89 (67.20–96.90)	90.00 (59.58–98.21)
PPV % (95% CI)	80.00 (62.69–90.50)	87.50 (63.98–96.50)	90.91 (62.26–98.38)
NPV % (95% CI)	84.62 (66.47–93.85)	80.00 (58.40–91.93)	100.00 (70.08–100.00)
Diagnostic accuracy % (95% CI)	82.14 (70.16–90.00)	83.33 (68.11–92.13)	95.00 (76.39–99.11)
Likelihood ratio of a positive test % (95% CI)	4 (2.846–5.621)	7 (2.524–19.41)	10 (1.409–70.99)
Likelihood ratio of a negative test % (95% CI)	0.1818 (0.1087–0.3041)	0.25 (0.1508–0.4144)	0

The Table reports diagnostic performances already corrected for the leave-one-out cross validation (LOOCV).
doi:10.1371/journal.pone.0080149.t003

pooled specimens analysis were considered for miRNA signature definition.

Performances of the signature. To assess the single miRNA performance in identifying health status a Receiver Operating Characteristic (ROC) curve based on kernel density distributions method fit as described in [34]. As overall measures of the performance in distinguish cases, the associated Area under the curve (AUC) was calculated and the p-values computed by means of permutations.

To assess and compare diagnostic performances of the miRNA signature identified, we fitted a multivariate logistic model selected by maximizing the Akaike Information Criteria (AIC) and a Relevance Vector Machine (RVM) model [35–37]. In contrast with Support Vector Machine, RVM follow a Bayesian approach giving *a posteriori* probability of the class. This makes the results from the two approaches more directly comparable. ROC curve and associated AUC were also computed for the logistic model.

A leave-one-out-cross-validation (LOOCV) approach was adopted for assessing how the results of both the RVM and AIC logistic regression predictive models would perform in practice. Performances were summarized in terms of sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and diagnostic accuracy.

Results

Study Population

A total of 311 subjects (159 males, 51.2%) were enrolled within the *TBnew* group. As summarized in Figure 2 there were 56

healthy subjects, 109 individuals with LTBI (75 recent contacts of active PTB cases; 34 non-recent contacts), 24 subjects affected by other pulmonary infections (OPI), 32 EPTB and 90 PTB patients. Subjects were recruited in Italy but were born in countries located in the different World Health Organization (WHO) regions (Figure 2). The *TB CHILD* group includes only individuals from Tanzania and Uganda. A total of 186 subjects were enrolled (107 males, 57.5%): 49 healthy subjects, 64 PTB patients, and 73 patients with PTB and HIV co-infection. In the latter category the median CD4⁺ cell count was 198.4 cells/mL (interquartile range: 277.1) compared with 707.7 cells/mL (interquartile range: 813.3) of PTB subjects without HIV infection. H and PTB populations were statistically comparable for mean age values of enrolled subjects (p-values for H and PTB: 0.29, and 0.80, respectively). Eight pools from the *TBnew* group, and 11 pools from the *TB CHILD* group were considered for the analysis. We used a subset of individuals to refine the miRNA signatures identified on pooled specimens. Briefly, we performed individual serum analysis for 36 and 20 subjects for the *TBnew*, and the *TB CHILD* group, respectively (details are reported in Table S1).

Normalization of qPCR Data

The TLDA is a 384-well microfluidic card containing dried TaqMan[®] primers and probes. Array A focuses on more highly characterized miRNAs while array B contains many of the more recently discovered miRNAs along with the miR* sequences. The use of two panels (array A and array B) enables quantitation of gene expression levels of up to 671 different miRNAs. This is

Table 4. Diagnostic performances of the serum miRNA signatures in the AIC logistic regression model.

Serum miRNA signature (n)	All (15)	TBnew (10)	TB CHILD (12)
n of individuals (H/PTB)	56 (28/28)	36 (18/18)	20 (10/10)
Sensitivity % (95% CI)	71.43 (52.94–84.75)	72.22 (49.13–87.50)	100.00 (72.25–100.00)
Specificity % (95% CI)	82.14 (64.41–92.12)	88.89 (67.20–96.90)	100.00 (72.25–100.00)
PPV % (95% CI)	80.00 (60.87–91.14)	86.67 (62.12–96.26)	100.00 (72.25–100.00)
NPV % (95% CI)	74.19 (56.75–86.30)	76.19 (54.91–89.37)	100.00 (72.25–100.00)
Diagnostic accuracy % (95% CI)	76.79 (64.23–85.90)	80.56 (64.97–90.25)	100.00 (83.89–100.00)
Likelihood ratio of a positive test % (95% CI)	4 (2.599–6.156)	6.5 (2.302–18.35)	undefined
Likelihood ratio of a negative test % (95% CI)	0.3478 (0.2672–0.4527)	0.3125 (0.2079–0.4696)	0

The Table reports diagnostic performances already corrected for the leave-one-out cross validation (LOOCV).
doi:10.1371/journal.pone.0080149.t004

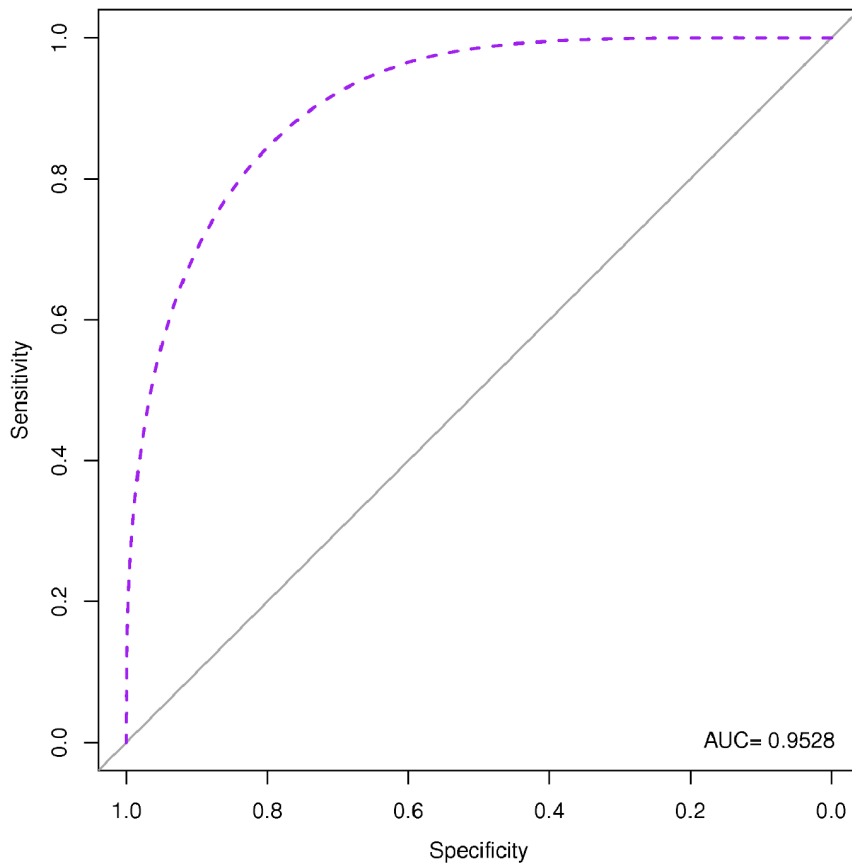


Figure 6. Receiver Operating Characteristic (ROC) based on Akaike information criterion (AIC) logistic regression for the 15-miRNA signature. AUC: area under the curve. The AIC model identified as the best performing miRNA signature the following: miR-let-7e, miR-146a, miR-16, miR-25, miR-365, miR-451, miR-885-5p, miR-223*. doi:10.1371/journal.pone.0080149.g006

accomplished by loading the cDNA product onto the array for PCR amplification and real-time analysis. MegaPlex Pools are designed to detect and quantitate up to 380 microRNAs (miRNAs) per pool in human species thanks to a set of stem-looped reverse transcription primers (MegaPlex RT Primers) that enable the simultaneous synthesis of cDNA and a set of miRNA-specific forward and reverse primers (MegaPlex PreAmp Primers) intended for use with very small quantities of starting material. The primers enable the unbiased preamplification of the miRNA cDNA target by PCR prior to loading the TaqMan[®] MicroRNA Array.

After the quantile normalization procedure, the C_t values of four miRNAs (ath-miR159a, MammU6, RNU44, and RNU48) detected by both array A and array B were compared. As showed in Table S2, C_t values were consistent between “pools – individuals”, “array A – array B”, and “*TBnew* – *TB CHILD*”. Normalized data from pooled and individual specimens are reported in Table S3.

Analysis of Serum miRNA Profiles in Pooled Samples

Normalized qPCR data from pools showed 277 miRNAs undetectable in the categories (H, PTB, LTBI, OPI, EPTB, and PTB/HIV) from both groups (*TBnew*, and *TB CHILD*).

The mean C_t value for each miRNA was calculated and a one-to-one comparison between different categories was carried out. Residual values are available in Table S4.

Figure 3 summarizes the number of miRNAs outside the 1st and 3rd quantile tails of the distribution of the residuals obtained by comparing two categories. The two percentiles considered, should most probably contain the miRNAs that are significantly different between the two compared categories. According to this qualitative analysis based on the distribution of the residuals, between 120 and 172 serum miRNAs could allow to discriminate among the categories considered in this study. For example, 134 miRNAs showed to be relevant in differentiating LTBI and PTB, whereas 132 miRNAs would allow discriminating between PTB and OPI. After filtering the pooled specimen results according to clinically relevant categories, we identified putative serum miRNA signatures defining MTB infection, active TB, pulmonary disease, or any of the disease statuses considered in this study (Figure 4).

To optimize our approach, for the PTB and H categories, we performed the analysis of individual sera.

Analysis of Serum miRNA Profiles in Individuals

Serum miRNA profiles from 18 H, and 18 PTB from *TBnew* group as well as 10 H and 10 PTB from the *TB CHILD* group were analyzed.

Out of the 671 miRNAs tested, 126 targets were detected in 80% of the subjects in at least one of the analyzed categories. As reported in Table 1, 71 miRNAs showing a p-value <0.05 were identified: we could detect differences in serum miRNA profiles between the *TBnew* and *TB CHILD* individuals (35 miRNAs, 49.3%), between H and PTB (29 miRNAs, 40.8%), and

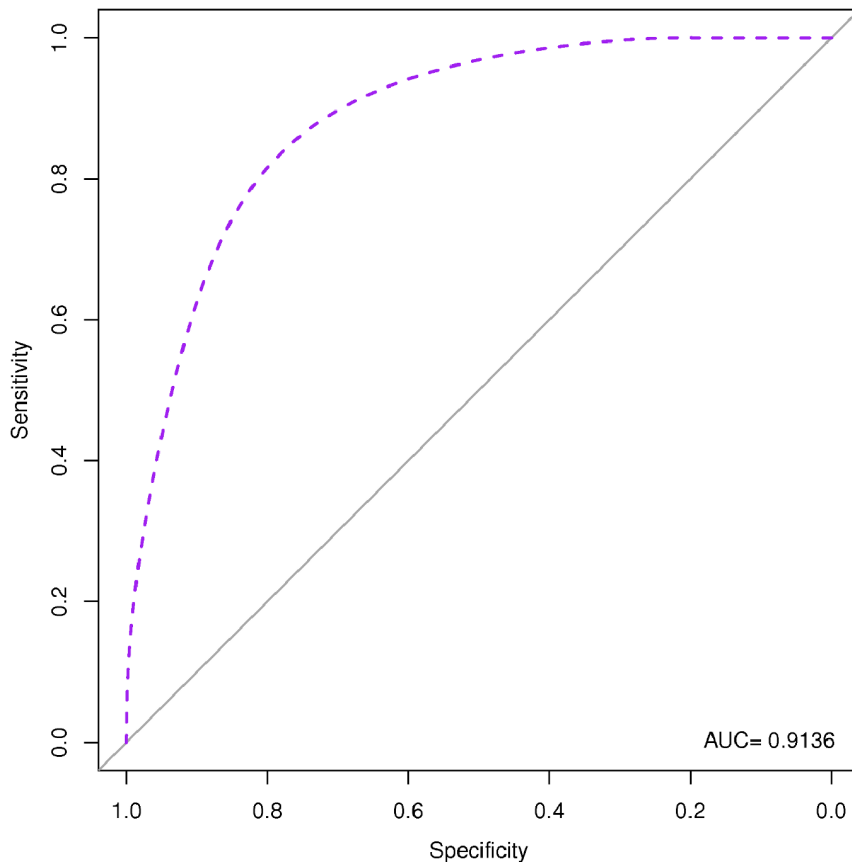


Figure 7. Receiver Operating Characteristic (ROC) based on Akaike information criterion (AIC) logistic regression for the 10-miRNA signature specific for *TBnew* population. AUC: area under the curve. The AIC model identified as the best performing miRNA signature the following: miR-let-7e, miR-192, miR-25, miR-451. doi:10.1371/journal.pone.0080149.g007

differences depending on both the population and the clinical status (7 miRNAs, 9.9%). After checking for FDR (p -adj), twenty miRNAs resulted to be significantly different between H and PTB (let-7e, miR-10b, miR-127-5p, miR-146a, miR-148a, miR-16, miR-185, miR-192, miR-193a-5p, miR-25, miR-365, miR-451, miR-518d-3p, miR-532-5p, miR-590-5p, miR-660, miR-885-5p, miR-223*, miR-30a, miR-30e). Complete results are available in Table S5.

Pools vs Individuals

Out of the 20 miRNAs showing a significant p -adj in the study on individuals (Table 1), 16 (80%) had already been identified as “relevant miRNA” in the analysis of pooled specimens (Figure 5). Among those miRNAs, nine showed differences between H and PTB pools in each group; four showed differences only in *TB CHILD* pooled specimens, whereas three only in the pools from *TBnew* group. Four miRNAs with a significant p -adj in the study on individuals had not been detected as relevant by the first screening on pooled specimens. Figure 5 shows that the analysis on pools correctly excluded 429 out of 439 “not relevant” miRNAs from further investigations.

Then, for the 16 miRNAs found to be relevant in the pooled specimens analysis and significant (p -adj < 0.05) in individual sera (let-7e, miR-146a, miR-148a, miR-16, miR-192, miR-193a-5p, miR-25, miR-365, miR-451, miR-532-5p, miR-590-5p, miR-660, miR-885-5p, miR-223*, miR-30a), we compared the direction of the variation (increase or decrease) in pooled and single PTB serum analyses. From

this comparison we found that some miRNAs showed inconsistent results between individual and pooled specimens (Table 2): eleven miRNAs showed the same type of variation in individual and pooled samples across both study groups considered, three miRNAs showed discordant variation only in *TBnew* or *TB CHILD* population, and one showed discordant variation in both *TBnew* and *TB CHILD* groups. Five and three targets were found to be associated only to the *TB CHILD* and to the *TBnew* group, respectively. In conclusion, from the combined analysis on pools and individuals, a total of 15 miRNAs were identified as a signature for discriminating between H and PTB (let-7e, miR-146a, miR-148a, miR-16, miR-192, miR-193a-5p, miR-25, miR-365, miR-451, miR-532-5p, miR-590-5p, miR-660, miR-885-5p, miR-223*, miR-30e).

Diagnostic Performances of the Identified Serum miRNA Signature

We used two approaches to evaluate the diagnostic performances of the identified serum miRNA signature: RVM model, and AIC logistic regression analysis. Both were cross-validated by the use of a LOOCV approach. All fifteen miRNAs (let-7e, miR-146a, miR-148a, miR-16, miR-192, miR-193a-5p, miR-25, miR-365, miR-451, miR-532-5p, miR-590-5p, miR-660, miR-885-5p, miR-223*, miR-30e) were used to re-classify the 56 subjects analyzed as single sera obtaining a diagnostic accuracy of 82% by RVM, and of 77% by logistic regression. The diagnostic accuracy increased up to 83% by RVM and 81% by logistic regression when miRNAs identified as “*TBnew*-specific” (let-7e, miR-148a,

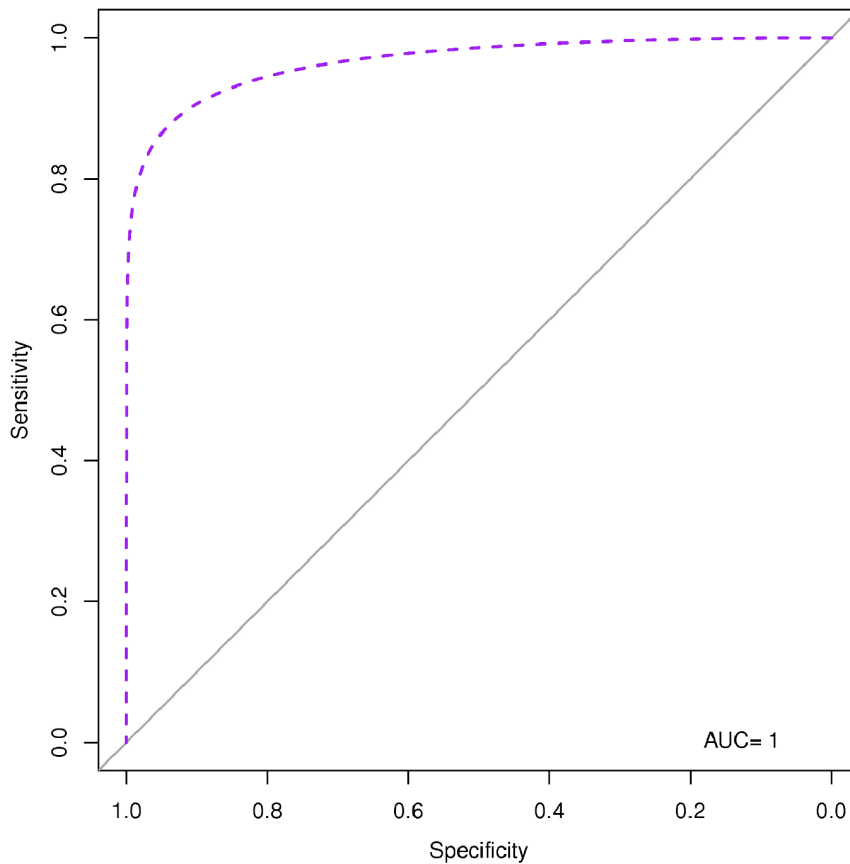


Figure 8. Receiver Operating Characteristic (ROC) based on Akaike information criterion (AIC) logistic regression for the 12-miRNA signature specific for *TB CHILD* population. AUC: area under the curve. doi:10.1371/journal.pone.0080149.g008

miR-16, miR-192, miR-193a-5p, miR-25, miR-365, miR-451, miR-590-5p, miR-885-5p) re-classify *TBnew* single serum samples. A similar increase was observed when the “*TB CHILD*-specific” signature (let-7e, miR-146a, miR-148a, miR-192, miR-193a-5p, miR-451, miR-532-5p, miR-590-5p, miR-660, miR-885-5p, miR-223*, miR-30e) was used on the 10 individual samples from *TB CHILD* group (95% by RVM, and 100% by logistic regression, respectively). Table 3 and Table 4 summarize the diagnostic performances of the miRNA signatures achieved by the different approaches used for the classification of subjects. AUCs for the regression logistic analyses are reported in Figure 6, Figure 7, and Figure 8.

Discussion

In TB biomarker research, published studies on highly multiplexed assays focus on proteomics, gene expression, transcriptomics, and miRNAs, as predictors of disease, disease recurrence or drug resistant infection [38–47]. Our study on serum miRNA signatures ascertains the value of these biomarkers for TB disease classification, as previously reported by others [21–23]. However, previous studies did not consider the impact of the genetic makeup on the miRNA signatures: ethnicity, together with age and gender, could influence the levels of circulating miRNAs [48]. By the comparison of two populations with different genetic makeup we showed that some population-specific miRNAs can increase the diagnostic accuracy for active TB.

While much research is still focused on assessing the quality of single biomarkers, there is an emerging interest in panels of biomarkers composed of multiple candidate targets which are neither specific nor sensitive when used as single tests, but which show a good performance when used in combination. In this study we used both RVM and logistic regression methods supported by a LOOCV approach to evaluate the diagnostic performances of serum miRNAs identified as a signature rather than as single biomarkers. The presence of miRNAs detectable in at least the 80% of one category indicates a higher diagnostic index; moreover, our filtering approach did not introduce biases in detecting potential on/off miRNAs between categories. Indeed, whenever a serum sample is tested for miRNA, its diagnostic relevance will be attested by its level rather than by their presence/absence in H and PTB categories.

To ascertain if a serum miRNA signature could discriminate between different categories of patients, we used a restrictive stratification screening approach to minimize the number of possible biases. Despite subjects within the H category showed a mean age lower than the other categories, we consider this to be only a minor drawback of the study. Particular attention should be reserved to specific subgroups (*e.g.* childhood and elderly) where age would expect to have a bigger impact. Another possible drawback could be the heterogeneity in terms of genetic background (estimated on the basis of the country of birth) in the *TBnew* group. This could partially explain lower performances of the miRNA signature on this population. The results on pooled specimens identified differences in serum miRNA profiles in the

Table 5. Comparison between the serum miRNA signature found in the present and the miRNAs reported in the previous studies.

miRNA	Fu et al, 2011	Qi et al, 2012
hsa-let-7e-4395517	let-7 family	let-7e, let-7 family
hsa-miR-146a-4373132	miR-146a	miR-146a
hsa-miR-148a-4373130	–	–
hsa-miR-16-4373121	–	–
hsa-miR-192-4373108	–	–
hsa-miR-193a-5p-4395392	–	miR-193 family
hsa-miR-25-4373071	–	miR-25
hsa-miR-365-4373194	miR-365 family	–
hsa-miR-451-4373360	–	–
hsa-miR-532-5p-4380928	–	miR-532 family
hsa-miR-590-5p-4395176	–	miR-590-5p
hsa-miR-660-4380925	–	miR-660
hsa-miR-885-5p-4395407	–	miR-885-5p
hsa-miR-223*-4395209	–	miR-223 family
hsa-miR-30e-4395334	miR-30 family	miR-30 family

doi:10.1371/journal.pone.0080149.t005

categories analyzed. Interestingly, according to our qualitative analysis based on the Empirical Distribution Function of residuals, serum miRNAs would allow not only to discriminate LTBI and PTB, but also PTB from OPI and EPTB. Indeed, the 134 miRNAs showing relevant differences in serum level of LTBI and PTB subjects include a smaller subset of miRNAs that could be used as specific signature to discriminate between these two categories. A similar approach could be applied for the 132 miRNAs showing relevant differences in serum levels of PTB and OPI subjects and for the 124 miRNAs differentiating PTB and EPTB subjects. Further studies on cohorts of LTBI, OPI and EPTB individuals will allow to identify specific miRNA patterns and to evaluate their diagnostic accuracy. The discriminatory power of serum miRNAs observed is further supported by the fact that the same findings have been confirmed in the two different groups (namely *TBnew* and *TB CHILD*) for the signature identified for the comparison H-PTB. Additionally, the use of pooled specimens allowed halving the number of targets to be analyzed by excluding miRNAs under the detection threshold or showing very little changes across categories.

To refine our findings, we performed serum miRNA analysis on individual sera from H and PTB subjects. Table 5 summarizes the comparison between our results and previously published studies. In the first study on circulating miRNAs as biomarkers for TB reported by Fu and colleagues [21], it was demonstrated that 92 miRNAs had significantly different levels in the sera of healthy controls *vs* PTB subjects: 59 miRNAs were down-regulated and 33 miRNAs were up-regulated in the serum of TB patients. One-by-one comparison is not possible due to different analytical platforms and normalization strategies, but some homologies between the study by Fu and our results can be observed. For example, miRNAs belonging the families let-7, miR-30, and miR-146 were found to be significantly different between H and PTB in both studies. miRNAs miR-590-5p, miR-185, miR-660, let-7e, miR-25, miR-146a, and miR-885-5p showed to be differentially expressed between healthy controls and PTB subject also in the study reported by Qi and colleagues [22]. miR-197 which was observed to be slightly

increased in our study was also reported to be increased in sera from pulmonary TB patients by Abd-El-Fattah and colleagues [23]. However, the previous studies did not consider the genetic background of the subjects enrolled. Differently, in the present study the inclusion of groups with different genetic background allowed us to better define serum miRNA signatures associated to a different (health) status. Indeed, subjects belonging to the same status (i.e. H or PTB) showed different serum miRNA levels between the *TBnew* and the *TB CHILD* groups in pooled specimens (Table S4). Comparing individual sera from subjects belonging to the two populations we found significant differences in the level of several miRNAs (Table S5). Despite larger population-based studies are still needed, our data support the hypothesis that the genetic background could influence the specific serum miRNA profiles. Interestingly, by matching miRNA signatures from pooled specimens, individual specimens, and direction of variation (increase/decrease) we identified 7 common discriminatory miRNAs (let-7e, miR-148a, miR-192, miR-193a-5p, miR-451, miR-590-5p, miR-885-5p) plus three miRNAs specific for the *TBnew* group (miR-16, miR-25, miR-365), and five miRNAs specific for the *TB CHILD* group (miR-146a, miR-532-5p, miR-660, miR-223*, miR-30e). The diagnostic accuracy for each single miRNA was found to be <75% (data not shown), while better results were achieved by using the approach of “signatures”: AUC values were above 0.90 and the use of the entire fifteen-miRNAs signature provided a diagnostic accuracy between 77% and 82% in a LOOCV approach (logistic regression and RVM, respectively). Population-specific signatures allowed to further improve classification accuracy (81–83% for *TBnew*, and 95–100% for *TB CHILD*, respectively). As mentioned before, serum miRNA signatures showed less efficiency in classifying subjects belonging to the *TBnew* group. Our hypothesis is that, despite the mild to moderate differences, the genetic background heterogeneity of this group is likely affecting the classification performances of the miRNA signature. The higher number of miRNAs with discrepancy variations (in terms of increase/decrease) between individual and pooled specimens provides some evidence on the heterogeneity of the *TBnew* population.

As for the function of extracellular miRNAs, current evidences suggest a regulatory role on the expression of target genes when taken up by recipient cells, with the peculiar capability to act on several targets at a time and to operate in a network with other extracellular/intracellular miRNAs [17,19]. A “hormonal” role to extracellular miRNAs was attributed [17,19]. If this interpretation proves to be true, we could then be facing a major advancement in modern biology not only for better understanding of biological complexity, but also in terms of diagnostic and even therapeutic possibilities [49]. Target cells for extracellular miRNAs and related targeting mechanism are poorly understood, thus careful interpretation of circulating miRNA origin and function should be considered.

Here we described a serum miRNA signature discriminating H and PTB subjects. Despite promising results, several challenges in pre-analytical and analytical phases remain in the analysis of circulating miRNAs. Accurate large-cohort studies are therefore required to validate PTB-specific miRNA signatures, and to identify miRNA signatures also for LTBI, EPTB and pathologies (like pneumonia, cancer, HIV infection and sarcoidosis) often in differential diagnosis with TB. The inclusion of different sets of biomarkers (*e.g.* cytokines, antibodies) could also help in achieving higher discrimination power amongst the closest categories.

Supporting Information

Table S1 Details on subjects included in pooled and individual specimens analyzed in the study. H: healthy

controls; LTBI: latent tuberculosis infection; PTB: pulmonary active tuberculosis; EPTB: active extra-pulmonary tuberculosis; OPI: other pulmonary infections; PTB/HIV: PTB in HIV-positive. SD: standard deviation. (XLSX)

Table S2 Comparison of the C_t values among the four miRNAs detected in both array A and array B types between different populations (*TBnew*, *TB CHILD*) and different specimens (*individual sera*, *pooled sera*). Mean C_t values are reported together with the standard deviation after quantile normalization. (XLSX)

Table S3 Normalized C_t values obtained from individual and pooled specimens after quantile normalization. (XLSX)

Table S4 Complete data analysis on pooled specimens. The Table reports residual values obtained by each comparison performed between the categories. Relevant differences are highlighted. (XLSX)

References

- World Health Organization (2012) Global tuberculosis report. Geneva, ISBN 978 92 4 156450 2.
- Zumla A, Atun R, Macurer M, Mwaba P, Ma Z, et al. (2011) Viewpoint: Scientific dogmas, paradoxes and mysteries of latent *Mycobacterium tuberculosis* infection. *Trop Med Int Health* 16(1): 79–83. doi: 10.1111/j.1365-3156.2010.02665.x.
- McNerney R, Macurer M, Abubakar I, Marais B, McHugh TD, et al. (2011) Tuberculosis diagnostics and biomarkers: needs, challenges, recent advances, and opportunities. *J Infect Dis* 205: S147–158. doi: 10.1093/infdis/jir860.
- Amanatidou V, Syridou G, Mavrikou M, Tsolia MN (2012) Latent tuberculosis infection in children: diagnostic approaches. *Eur J Clin Microbiol Infect Dis* 31(7): 1285–1294. doi: 10.1007/s10096-011-1524-3.
- Schaaf HS, Collins A, Bekker A, Davies PD (2010) Tuberculosis at extremes of age. *Respirology* 15(5): 747–763. doi: 10.1111/j.1440-1843.2010.01784.x.
- Perez-Velez CM, Marais BJ (2012) Tuberculosis in children. *N Engl J Med* 367(4): 348–361. doi: 10.1056/NEJMr1008049.
- Pearce EC, Woodward JF, Nyandiko WM, Vreeman RC, Ayaya SO (2012) A systematic review of clinical diagnostic systems used in the diagnosis of tuberculosis in children. *AIDS Res Treat* 2012: 401896. doi: 10.1155/2012/401896.
- Mack U, Migliori GB, Sester M, Rieder HL, Ehlers S, et al. (2009) LTBI: latent tuberculosis infection or lasting immune responses to *M. tuberculosis*? A TBNET consensus statement. *Eur Respir J* 33(5): 956–973. doi: 10.1183/09031936.00120908.
- Walzl G, Ronacher K, Hanekom W, Scriba TJ, Zumla A (2011) Immunological biomarkers of tuberculosis. *Nat Rev Immunol* 11(5): 343–354. doi: 10.1038/nri2960.
- Norbis L, Miotto P, Alagna R, Cirillo DM (2013) Tuberculosis: lights and shadows in the current diagnostic landscape. *New Microbiol* 36(2): 111–120.
- Wallis RS, Pai M, Menzies D, Doherty TM, Walzl G, et al. (2010) Biomarkers and diagnostics for tuberculosis: progress, needs, and translation into practice. *Lancet* 375(9729): 1920–1937. doi: 10.1016/S0140-6736(10)60359-5.
- Lawn SD, Zumla AI (2011) Tuberculosis. *Lancet* 378(9785): 57–72. doi: 10.1016/S0140-6736(10)62173-3.
- Wallis RS, Kim P, Cole S, Hanna D, Andrade BB, et al. (2013) Tuberculosis biomarkers discovery: developments, needs, and challenges. *Lancet Infect Dis* 13(4): 362–372. doi: 10.1016/S1473-3099(13)70034-3.
- Kim VN, Nam JW (2006) Genomics of microRNA. *Trends Genet* 22(3): 165–173.
- Flynt AS, Lai EC (2008) Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nat Rev Genet* 9(11): 831–842. doi: 10.1038/nrg2455.
- Vickers KC, Palmisano BT, Shoucri BM, Shamburek RD, Remaley AT (2011) MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nat Cell Biol* 13(4): 423–433. doi: 10.1038/ncb2210.
- Chen X, Liang H, Zhang J, Zen K, Zhang CY (2012) Secreted microRNAs: a new form of intercellular communication. *Trends Cell Biol* 22(3): 125–132. doi: 10.1016/j.tcb.2011.12.001.
- Weber JA, Baxter DH, Zhang S, Huang DY, Huang KH, et al. (2010) The MicroRNA spectrum in 12 body fluids. *Clin Chem* 56(11): 1733–1741. doi: 10.1373/clinchem.2010.147405.
- Cortez MA, Bueso-Ramos C, Ferdin J, Lopez-Berestein G, Sood AK, et al. (2011) MicroRNAs in body fluids – the mix of hormones and biomarkers. *Nat Rev Clin Oncol* 8(8): 467–477. doi: 10.1038/nrclinonc.2011.76.
- Gilad S, Meiri E, Yogev Y, Benjamin S, Lebanony D, et al. (2008) Serum microRNAs are promising novel biomarkers. *PLoS One* 3(9): e3148. doi: 10.1371/journal.pone.0003148.
- Fu Y, Yi Z, Wu X, Li J, Xu F (2011) Circulating microRNAs in patients with active pulmonary tuberculosis. *J Clin Microbiol* 49(12): 4246–4251. doi: 10.1128/JCM.05459-11.
- Qi Y, Cui L, Ge Y, Shi Z, Zhao K, et al. (2012) Altered serum microRNAs as biomarkers for the early diagnosis of pulmonary tuberculosis infection. *BMC Infect Dis* 12: 384. doi: 10.1186/1471-2334-12-384.
- Abd-El-Fattah AA, Sadik NA, Shaker OG, Aboultouh ML (2013) Differential microRNAs expression in serum of patients with lung cancer, pulmonary tuberculosis, and pneumonia. *Cell Biochem Biophys*. 2013 Apr 5. [Epub ahead of print].
- Kirschner MB, Kao SC, Edelman JJ, Armstrong NJ, Valley MP, et al. (2011) Haemolysis during sample preparation alters microRNA content of plasma. *PLoS One* 6(9): e24145. doi: 10.1371/journal.pone.0024145.
- R Development Core Team (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. Available: <http://www.R-project.org/>. Accessed 2013 June 14.
- Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19(2): 185–193.
- Rao YL, Lee Y, Jarjoura D, Ruppert AS, Liu CG, et al. (2008) A comparison of normalization techniques for microRNA microarray data. *Stat Appl Genet Mol Biol* 7(1): Article22. doi: 10.2202/1544-6115.1287.
- Deo A, Carlsson J, Lindlöf A (2011) How to choose a normalization strategy for miRNA quantitative real-time (qPCR) arrays. *J Bioinform Comput Biol* 9(6): 795–812.
- Yohai VJ (1987) High Breakdown-Point and High Efficiency Robust Estimates for Regression. *Ann Stat* 15: 642–656.
- Salibian-Barrera M, Yohai VJ (2008) High breakdown point robust regression with censored data. *Ann Stat* 36: 118–146.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, et al. (2009) Circos: An information aesthetic for comparative genomics. *Genome Res* 19: 1639–1645. doi: 10.1101/gr.092759.109.
- Good PI (2005) Permutation, parametric, and bootstrap tests of hypotheses. Springer. 336 p.
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29: 1165–1188.
- Zou KH, Hall WJ, Shapiro DE (1997) Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Stat Med* 16: 2143–2156.
- Tipping ME (2001) Sparse Bayesian learning and the relevance vector machine. *Journal Mach Learn Res* 1: 211–244.
- Lapedriza A, Segui S, Masip D, Vitria J (2008) A sparse Bayesian approach for joint feature selection and classifier learning. *Pattern Anal Appl* 11: 299–308.

Table S5 Complete data analysis on individual specimens. The Table reports the p-value for the following comparison on individual specimens: *TBnew vs TB CHILD*, and healthy control (H) vs pulmonary active tuberculosis (PTB). P-adjusted (*p-adj*) is reported for the comparison H vs PTB. Values <0.05 are highlighted. na: not applicable (miRNAs not detected in at least 80% of subjects in at least one of the category considered where not included in the analysis). (XLSX)

Acknowledgments

The authors thank the individuals and the study teams in each site for their support during implementation of this work. We are grateful to Valentina Vanini for the technical help and to Gilda Cuzzi for the enrollment help provided at INMI, and Guenter Froeschl as PhD coordinator at LMU.

Author Contributions

Conceived and designed the experiments: PM DMC. Performed the experiments: PM GM ICV. Analyzed the data: PM AA GS RB. Contributed reagents/materials/analysis tools: NH KR DMC. Wrote the paper: PM GM DMC. Enrolment of subjects and stratification: LN LRC DG AM ENN FA.

37. Braun AC, Weidner U, Hinz S (2012) Classification in High-Dimensional Feature Spaces-Assessment Using SVM, IVM and RVM With Focus on Simulated EnMAP Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5: 436–443.
38. Agranoff D, Fernandez-Reyes D, Papadopoulos MC, Rojas SA, Herbster M, et al. (2006) Identification of diagnostic markers for tuberculosis by proteomic fingerprinting of serum. *Lancet* 368(9540): 1012–1021.
39. Jacobsen M, Repsilber D, Gutschmidt A, Neher A, Feldmann K, et al. (2007) Candidate biomarkers for discrimination between infection and disease caused by *Mycobacterium tuberculosis*. *J Mol Med (Berl)* 85(6): 613–621.
40. Mistry R, Cliff JM, Clayton CL, Beyers N, Mohamed YS, et al. (2007) Gene-expression patterns in whole blood identify subjects at risk for recurrent tuberculosis. *J Infect Dis* 195(3): 357–365.
41. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, et al. (2010) An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 466(7309): 973–977. doi: 10.1038/nature09247.
42. Maertzdorf J, Repsilber D, Parida SK, Stanley K, Roberts T, et al. (2011) Human gene expression profiles of susceptibility and resistance in tuberculosis. *Genes Immun* 12(1): 15–22. doi: 10.1038/gene.2010.51.
43. Leshe E, Forestiero FJ, Hirata MH, Hirata RD, Cecon L, et al. (2011) Transcriptional responses of host peripheral blood cells to tuberculosis infection. *Tuberculosis (Edinb)* 91(5): 390–399. doi: 10.1016/j.tube.2011.07.002.
44. Wang C, Yang S, Sun G, Tang X, Lu S, et al. (2011) Comparative miRNA expression profiles in individuals with latent and active tuberculosis. *PLoS One* 6(10): e25832. doi: 10.1371/journal.pone.0025832.
45. Liu Y, Wang X, Jiang J, Cao Z, Yang B, et al. (2011) Modulation of T cell cytokine production by miR-144* with elevated expression in patients with pulmonary tuberculosis. *Mol Immunol* 48(9–10): 1084–1090. doi: 10.1016/j.molimm.2011.02.001.
46. Yi Z, Fu Y, Ji R, Li R, Guan Z (2012) Altered microRNA signatures in sputum of patients with active pulmonary tuberculosis. *PLoS One* 7(8): e43184. doi: 10.1371/journal.pone.0043184.
47. Olivier I, Loots du T (2012) A metabolomics approach to characterise and identify various *Mycobacterium* species. *J Microbiol Methods* 88(3): 419–426. doi: 10.1016/j.mimet.2012.01.012.
48. Ajit SK (2012) Circulating microRNAs as biomarkers, therapeutic targets, and signaling molecules. *Sensors (Basel)* 12(3): 3359–3369. doi: 10.3390/s120303359.
49. Schoof CR, Botelho EL, Izzotti A, Vasques Ldos R (2012) MicroRNAs in cancer treatment and prognosis. *Am J Cancer Res* 2(4): 414–433.