Whole blood transcriptome analysis in onchocerciasis

Ole Lagatie a,*, Linda Batsa Debrab, d, Alex Y. Debrac, c, d, Lieven J. Stuyver a

a Johnson & Johnson Global Public Health, Janssen R&D, Turnhoutseweg 30, 2340 Beerse, Belgium
b Faculty of Allied Health Sciences, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
c Department of Clinical Microbiology, School of Medicine and Dentistry, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
d Kumasi Centre for Collaborative Research in Tropical Medicine, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

A R T I C L E   I N F O

Keywords:
Onchocerca volvulus
Onchocerciasis
Gene expression
Immune evasion
Th17
RICTOR

A B S T R A C T

Identifying the molecular mechanisms controlling the host’s response to infection with Onchocerca volvulus is important to understand how the human host controls such parasitic infection. Little is known of the cellular immune response upon infection with O. volvulus. We performed a transcriptomic study using PAXgene-preserved whole blood from 30 nodule-positive individuals and 21 non-endemic controls. It was found that of the 45,042 transcripts that were mapped to the human genome, 544 were found to be upregulated and 447 to be downregulated in nodule-positive individuals (adjusted P-value < 0.05). Pathway analysis was performed on this set of differentially expressed genes, which demonstrated an impact on oxidative phosphorylation and protein translation. Upstream regulator analysis showed that the mTOR associated protein RICTOR appears to play an important role in inducing the transcriptional changes in infected individuals. Functional analysis of the genes affected by infection indicated a suppression of antibody response, Th17 immune response and proliferation of activated T lymphocytes. Multiple regression models were used to select 22 genes that could contribute significantly in the generation of a classifier to predict infection with O. volvulus. For these 22 genes, as well as for 8 reference target genes, validated RT-qPCR assays were developed and used to re-analyze the discovery sample set. These data were used to perform elastic net regularized logistic regression and a panel of 7 genes was found to be the most performing classifier. The resulting algorithm returns a value between 0 and 1, reflecting the predicted probability of being infected. A validation panel of 69 nodule-positive individuals and 5 non-endemic controls was used to validate the performance of this classifier. Based on this validation set only, a sensitivity of 94.2% and a specificity of 60.0% was obtained. When combining the discovery test set and validation set, a sensitivity of 96.0% and a specificity of 92.3% was obtained. Large-scale validation approaches will be necessary to define the intended use for this classifier. Besides the use as marker for infection in MDA efficacy surveys and epidemiological transmission studies, this classifier might also hold potential as pharmacodynamic marker in macrofilaricide clinical trials.

1. Introduction

Onchocerciasis, also known as river blindness, is a neglected tropical disease caused by infection with the filarial nematode Onchocerca volvulus (Enk, 2006). It is estimated that currently 218 million people are living in areas at risk for transmission. Most infected individuals live in Africa, but some smaller foci still exist in Latin America and the Eastern Mediterranean Region (WHO, 2020). The parasite is transmitted to humans through the bite of a blackfly (Simulium spp.). These flies breed in fast-flowing streams and rivers, increasing the risk of blindness to individuals living nearby, hence the commonly known name of “river blindness”. Life-cycle stages of O. volvulus in the human host consist of macrofilariae (adult worms) and microfilariae (first-stage larvae, L1). While the macrofilariae accumulate in subcutaneous or deep nodules called onchocercomas, microfilariae migrate to the skin, eyes and other organs. Treatment is based on microfilaricidal agents such as ivermectin (Mectizan, Merck, Rahway, NJ, USA) as no approved macrofilaricide drugs or vaccines are available. Only doxycycline treatment has been found to result in an enhanced killing of female O. volvulus worms, besides its clear sterilizing effect (Debrab et al., 2015). Since microfilaricides only affect the larval stage of O. volvulus with little or no impact on the adult worm, annual or bi-annual treatments for several
years are required in so-called mass drug administration (MDA) programmes (Diawara et al., 2009; Traore et al., 2012). Evaluation of these MDA programmes, and ultimately also guidance to stop them, is mainly based on monitoring of infection levels in human populations. Besides clinical examination by palpation of nodules formed by adult worms, diagnostic tools for detection of *O. volvulus* infection also involve the detection of microfilariae (mf) in small, superficial skin biopsy samples (skin snips) using microscopy or nucleic acid-based detection of *O. volvulus* DNA (Taylor et al., 1989; Toe et al., 1998; Boatin et al., 2002; Fink et al., 2011; Lloyd et al., 2015; Lagatie et al., 2016a). More recently, non-invasive tools have been developed that are based on Ov16 IgG4 antibodies in the blood (Lavebratt et al., 1994; Chandrashekar et al., 1996; Golden et al., 2013; Steel et al., 2015; Lont et al., 2017; Coffeng et al., 2019). Also parasite-derived metabolites in urine of infected individuals have been discovered and proposed as diagnostic tool but clinical accuracy was found to be limited (Globisch et al., 2013, 2015; Lagatie et al., 2016b; Globisch et al., 2017; Shirey et al., 2018; Wewer et al., 2021). Based on the diagnostic target product profile put forward by the WHO it is expected that these metabolites will play a minor role in decision-making to eliminate transmission of onchocerciasis (WHO, 2021).

Although extensive research has been done to understand the antibody response upon infection, less is known about the contribution of cellular immune responses to the parasite-host interaction (Greene et al., 1985; Lagatie et al., 2017b). This host-parasite interaction is particularly important in onchocerciasis as pathological manifestations of *O. volvulus* infection does not only result from the parasite, but also from the magnitude and quality of the host immune response (Mackenzie et al., 1987; Ali et al., 2003). Parasite-derived immunosuppression is one of the most central mechanisms to the persistence of the infection. CD4⁺CD25⁺Foxp3⁺ regulatory T (Treg) cells and Th17 cells, CD4⁺ T helper cells that secrete IL-17, play a central role in controlling the immune balance required to maintain self-tolerance, preventing allergies, and confining inflammatory responses to infectious diseases (Langrish et al., 2005; Sakaguchi, 2005; Sakaguchi et al., 2008; Tesmer et al., 2008). Increasing evidence demonstrate that a Treg/Th17 imbalance plays an important role in parasitic pathogenesis (Zhang et al., 2012; Chen et al., 2013). In onchocerciasis, an increase in Th17 cells and concomitant reduction in Treg cells were associated to hyperreactive inflammatory responses during chronic infection. Also, a potential role for RICTOR, a key protein in TOR2 signaling, in modulating the host response upon infection is demonstrated. Furthermore, based on the expression profiling, a 7-gene classifier was explored as a diagnostic tool for detection of onchocerciasis.

### 2. Methods

#### 2.1. Study population and sample collection

Whole blood was collected in PAXgene Blood RNA tubes (PreAnalytiX, Franklin Lakes, NJ, USA) as part of a field study in Ghana. This study was undertaken in an onchocerciasis-endemic community located in Adansi South District along the Pra River basins in the Ashanti region of Ghana. Physical examinations were performed to identify those subjects having palpable nodules. Skin snips (biopsies) were then taken from those with nodules in order to determine the microfilarial (mf) load in the skin (Debrah et al., 2015). Most subjects were participating in mass drug administration programmes. A total of 99 nodule-positive subjects that donated whole blood samples were included, as well as 51 endemic controls that had no visible signs of onchocerciasis. Additionally, whole blood samples from 29 non-endemic controls (from Kumasi, Ashanti Region) were available for testing. In both sampling locations, the ethnicity of the population is primarily Akan. The study population was split in a discovery cohort consisting of 30 nodule-positive individuals and 21 non-endemic controls and a validation cohort consisting of 69 nodule-positive individuals, 51 endemic controls and 8 non-endemic controls. An overview of the patient demographics is provided in Table 1.

#### 2.2. Onchocerciasis IgG4 rapid test

The presence of IgG4 antibodies against the *O. volvulus* antigen Ov16 was determined using the SD BIOLINE Onchocerciasis IgG4 test (Standard Diagnostics, Gyeonggi-do, Republic of Korea), according to the

### Table 1

| Characteristic          | Discovery cohort | Validation cohort |
|-------------------------|------------------|-------------------|
|                         | Nodule-positive  | Non-endemic controls | Nodule-positive  | Non-endemic controls | Endemic controls |
| No. of subjects         | 30               | 21                 | 69               | 8                    | 51               |
| Age, median (min-max)   | 48.5 (22–78)     | 22 (18–56)         | 45 (21–85)       | 22 (18–26)           | 35 (18–81)       |
| Gender, n (%)           |                  |                    |                  |                      |                  |
| Male                    | 14 (46.7)        | 15 (71.4)          | 41 (59.4)        | 5 (62.5)             | 26 (51.0)        |
| Female                  | 16 (53.3)        | 6 (28.6)           | 28 (40.6)        | 3 (37.5)             | 25 (49.0)        |
| Nodules, median (min-max) | 1.5 (1–4)       | 0 (0–0)            | 1 (1–5)          | 0 (0–0)              | 0 (0–0)          |
| mf status, n (%)        |                  |                    |                  |                      |                  |
| 0–0.5 mf/mg             |                  |                    |                  |                      |                  |
| 0–5 mf/mg               |                  |                    | 69 (100)         | na                   | 51 (100)         |
| 5–10 mf/mg              |                  |                    | 0 (0)            | na                   | 0 (0)            |
| IVM rounds, median (min-max) | 2 (0–10) | 3 (0–10) | na | 0 (0–1) |
| Time since last ivermectin treatment (n) |                  |                    |                  |                      |                  |
| Not treated             | 5 (16.7)         | 11 (15.9)          | 34 (66.7)        |                      |                  |
| <20 months              | 16 (53.3)        | 18 (26.1)          | 0 (0)            |                      |                  |
| >20 months              | 9 (30.0)         | 40 (58.0)          | 17 (33.3)        |                      |                  |
| Ov16 status, n (%)      |                  |                    |                  |                      |                  |
| Positive                | 25 (83.3)        | 43 (62.3)          | 7 (87.5)         | 26 (51.0)            |                  |
| Negative                | 5 (16.7)         | 26 (37.7)          | 1 (12.5)         | 25 (49.0)            |                  |

Abbreviations: na, not applicable; mf, microfilariae; min, minimum; max, maximum.
manufacturer’s instructions. Briefly, 10 µl of plasma was added to the round sample well on the lateral flow strip, immediately followed by the addition of 4 drops of assay diluent into the square assay diluent well. After 1 h, tests were scored. Tests were considered positive only when both the test and control line were visible. Faint lines were considered positive, as recommended by the manufacturer.

2.3. Isolation of RNA from whole blood

RNA was isolated from whole blood by Biogazelle (Ghent, Belgium) using the PAXgene Blood miRNA Kit (Qiagen, Hilden, Germany) following the manufacturer’s instructions. On-column DNase digestion was performed during the RNA extraction. The RNA concentration was determined using the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). RNA quality control was performed using the Fragment Analyzer microfluidic gel electrophoresis system (Sage Science, Beverly, MA, USA). Samples with low RNA quality (i.e. RIN value < 7) were excluded for PolyA+ RNA sequencing.

2.4. PolyA+ RNA sequencing and data processing

PolyA+ RNA sequencing was performed by Biogazelle (Ghent, Belgium). Libraries for mRNA sequencing were prepared using the TruSeq stranded mRNA sample prep kit (Illumina, San Diego, CA, USA). The starting material (100 ng) of total RNA was mRNA-enriched using the oligoT bead system (Illumina). The isolated mRNA was subsequently fragmented enzymatically. Then, first-strand synthesis and second-strand synthesis were performed, and the double-stranded cDNA was purified (Agencourt AMPure XP, Beckman Coulter, Passadena, CA, USA). The cDNA was end-repaired, 3’ adenylated and Illumina sequencing adapters ligated onto the fragments ends, and the library was purified (Agencourt AMPure XP). The polyA+ RNA stranded libraries were pre-amplified with PCR and purified (Agencourt AMPure XP). The libraries size distribution was validated, and quality inspected on the 2100 Bioanalyzer (high sensitivity DNA chip, Agilent, Santa Clara, CA, USA). High quality libraries were quantified using the Qubit Fluorometer (Life Technologies, Carlsbad, CA, USA), the concentration normalized, and the samples pooled. Single-end sequencing was performed on NextSeq500 instrument according to the manufacturer instructions (Illumina). The Illumina sequencing sequences were trimmed to remove adapter content using Trimmomatic (v0.35). Reads with very low quality were also trimmed before mapping using Trimmomatic. TopHat (v2.1.0) was used to align the sequences to the reference transcriptome which was based on Ensembl gene annotation (release 84) and enriched for long non-coding genes present in LIGNpedia (www.lincpedia.org, release 4.0).

2.5. Gene expression analysis

Raw reads were normalized using the geometric mean-based method implemented in DESeq2 R package. Differential gene expression was performed using the same package. A false discovery rate (FDR) of 5% was applied on P-values adjusted according to Benjamini-Hochberg method. Principal components analysis was performed using prcomp function in the stats package in R. Heatmaps and PCA plots were calculated based on FPKM-values.

2.6. RT-qPCR confirmation

Before analysis, gDNA removal was performed on all samples with the HeatRUn gDNA removal kit (ArcticZymes, Tromsø, Norway) using 254.6 ng of total RNA. gDNA contamination was assessed on all samples by means of two DNA-specific qPCR assays using RNA as input. cDNA synthesis was performed using the iScript Advanced cDNA Synthesis Kit (Bio-Rad, Hercules, CA, USA) with 200 ng of total RNA as input. The cDNA quality of each sample was assessed by means of two universally expressed genes. For gene expression profiling we used validated approaches for qPCR amplification. All measurements were performed in 384-well plates (CFX384 – Bio-Rad) in a reaction volume of 5 µl, using SYBR Green I. Each PCR reaction was done in duplicate. cDNA equivalents of RNA (5 ng) were used as input in each PCR reaction. Upon transfer of the RT-qPCR assays to another laboratory, the same procedures were implemented, except that measurements were performed on a Roche LightCycler 480 in a reaction volume of 10 µl.

2.7. Elastic net regularisation for selection of the targets in the classifier

To select the optimum set of targets for constructing a RT-qPCR-based classifier for detection of onchocerciasis, we used the R package glmnet (v2.0-16) for the elasticNet algorithm. The model was built using α = 0.8, being close to lasso regression. For each gene, a penalty value was fed into the elastic net regularizer. This means that the regularization algorithm used the penalty values during each cycle of finding the next most appropriate target to expand the panel with. These penalty values were based on Cq-dependent precision and the spread in Cq-values for each target.

2.8. Logistic regression modeling for determination of the final qPCR-based classifier

Based on the ΔCq-values of the 7 selected targets in the discovery sample set, a logistic regression model was built. For this, we used the R package glmnet (v2.0-16) to fit a logistic regression, where onchocerciasis status, nodule-positive (1) or non-endemic control (0), was the binary response variable (Friedman et al., 2010). Glmnet fits generalized linear models with penalized maximum likelihood using ridge or lasso regression which provides more precise model fitting than other methods, such as least-squares, when the predictors are intercorrelated. In ridge/lasso regression, a penalty is imposed when the sum of the correlation coefficients of the predictors is large. Therefore, when predictors are intercorrelated, the values of their combined correlation coefficients are reduced either by shrinking one of the coefficients towards zero (lasso regression) or by shrinking both toward some medium value (ridge regression). The strength of the penalty against large correlation coefficients is determined by the regularization parameter (λ), which therefore determines the complexity of the model (the number predictors with non-zero correlation coefficients). The value of λ was chosen using the in-built cross validation function (we selected the default option of lambda.1se which gives the most regularized model such that error is within one standard error of the minimum). The elastic net mixing parameter (α) determines whether lasso (α = 1) or ridge regression (α = 0) is used. The model we built was based on α = 1 (lasso), where for correlated predictors one of the coefficients are shrunk towards zero. Using the ‘predict’ function of glmnet we determined for every individual its probability of being infected with O. volvulus based on the expression levels of the 7 genes.

3. Results

3.1. Generation of RNA-seq data

Thirty onchocerciasis patients who were nodule-positive and that were Ov16 IgG4-positive and/or microfilaria-positive were included for RNA-seq analysis. Additionally, 25 non-endemic controls who were Ov16 IgG4-negative were included as control population. After RNA extraction, RNA quality control was performed. On average, a RIN value of 8.1 was obtained, with low RIN values (< 6) for 4 non-endemic control samples. These RNA samples were excluded for subsequent RNA-seq analysis. For the remaining 51 samples PolyA+ RNA sequencing was performed (Table 1). The average number of mapped reads per sample was 34.7 million with a minimal read count of 21.4 million reads (Supplementary Table S1). For each sample, gene-level FPKM-values were generated based on HTSeq (Anders et al., 2015) raw counts (Supplementary Table S2). FPKM-values (fragments per kilobase of exon per million reads) represent
read counts normalized for sequencing depth and transcript length allowing comparison of genes across and within samples.

3.2. Differential gene expression upon O. volvulus infection

The present analysis aimed to evaluate changes in expression between non-endemic controls and nodule-positive individuals. A total of 991 targets were differentially expressed between both groups (multiple testing adjusted \( P \)-values at FDR < 0.05) with 544 of them being upregulated and 447 being downregulated (Supplementary Table S1). Of those targets, 800 represent coding genes. The results are displayed as a Volcano plot including all genes (Fig. 1A), as an MA plot indicating the relationship between expression level and fold change (Fig. 1B) and as a heatmap including the top 250 differentially expressed genes (Fig. 1C). Principal components analysis (PCA) was performed on the entire data set which showed no separation of both groups (Supplementary Fig. S1).

Re-analysis based on the 991 differentially expressed targets only resulted in a clear separation of both groups and this separation was mainly based on the second principal component (Fig. 1D). Since the two groups of individuals investigated in this study were not age- and gender-matched, PCA on the 991 differentially expressed targets was also performed with grouping based on gender or age (Supplementary Fig. S1). This analysis showed no effect of gender on the gene expression levels of the 991 differentially expressed targets. For age there appeared to be some separation of the group of 18–25 years, but all other age groups overlapped, indicating that age is unlikely to play an important role in the expression levels of these 991 targets.

3.3. Biological pathways affected during O. volvulus infection

To identify the biological pathways associated with O. volvulus infection, we performed Ingenuity canonical pathway analysis based on...

**Fig. 1.** Whole-blood transcriptomic analysis of non-endemic controls and O. volvulus-infected individuals. Characteristics of subjects used in this analysis are provided in Table 1. **A** Volcano plot of differential gene expression analysis for infected vs uninfected showing log₂FC (x-axis) plotted against − \( \log_{10} \) adjusted \( P \)-value (y-axis). DEGs with adjusted \( P \)-value < 0.05 are visualized as red points. **B** MA plot of differential gene expression analysis for infected vs uninfected showing mean of normalized (for sequencing depth) counts of all samples (x-axis) plotted against log₂FC (y-axis). Genes are colored according to their \( P \)-value. **C** An unsupervised clustering heat map (Spearman’s correlation with Ward’s linkage) of the top 250 most variably expressed genes. Red intensity indicates increased expression with infection, and blue intensity indicates decreased expression with infection. Individuals are denoted as blue (nodule-positive) and red (non-endemic control). **D** Principal components analysis of the RNA-seq samples (infected – uninfected) using the 991 differentially expressed genes across all samples based on FPKM data. Individuals are denoted as blue (nodule-positive) and red (non-endemic control).
the 991 differentially expressed targets. Using a cut-off of $P < 0.05$ (Fisher’s exact test) and an absolute z-score of $> 2$, onchocerciasis significantly affected 11 pathways, with the most significantly affected pathways being eIF2 signaling and oxidative phosphorylation (Fig. 2A). Both pathways appeared to be inhibited upon Onchocerca infection. Of the pathways that were found to be activated, most of them were related to mTOR signaling and its upstream regulatory pathways (PI3K, AKT) and calcium signaling (including phospholipase C and protein kinase D).

### 3.4. Predicted upstream regulators induced during *O. volvulus* infection

To determine which molecules might be important in the response to *O. volvulus* infection, we applied DEGs with a FDR $< 0.05$ (no fold change filter) to Ingenuity’s upstream regulator analysis to determine genes that are predicted to be important regulators during *O. volvulus* infection. Using an absolute z-score of $> 2$ and $P$-value $< 0.05$ as the cut-off for significance, RICTOR was found to have strong activation score and MYCN was found to be significantly inhibited (Fig. 2B). There appeared to be a substantial overlap between genes regulated by RICTOR and MYCN, with a number of genes specifically for MYCN that are inconsistent with the state of the downstream molecule, thereby suggesting that mainly the activation of RICTOR can be considered as relevant (Fig. 2C).

### 3.5. Predicted diseases or functions related to *O. volvulus* infection

Ingenuity’s diseases and functions analysis was performed to determine cellular or physiological functions that are predicted to be affected during *O. volvulus* infection. Using an absolute z-score of $> 2$ and $P$-value $< 0.05$ as...
the cut-off for significance, it was found that 6 functions were decreased, and 6 were increased (Fig. 2D). Besides a clear impact on cell adhesion and cell death/survival related functions, three of the downregulated functions were related to suppression of the humoral and adaptive immune systems: antibody response, Th17 immune response, and proliferation of activated T lymphocytes. Detailed investigation of the differentially expressed genes involved in these cellular functions shows that some of them encode for proteins involved in immune checkpoint regulation. Among them are Programmed Cell Death 1 (PDCD1), Inducible T cell Costimulator Ligand (ICOSLG), High-Mobility Group Box 1 (HMGB1) and Cytotoxic T-Lymphocyte Associated protein 4 (CTLA4) (Fig. 2E).

3.6. Selection of a set of 22 genes for development of a classifier and validation of the RNA-seq data

We performed both elastic net regularization, linear regression, and multivariate modeling on the RNA-seq dataset. Based on these modeling approaches, a set of 22 genes was selected that had the largest contribution to a predictive model (Fig. 3). The expression of these 22 genes and 9 reference genes (PPIA, HPRT1, B2M, GAPDH, HMBS, RPL13A, RPS18, TUBB and YWHAZ) was assessed using RT-qPCR in the same RNA samples as used for RNA-seq, using a Bio-Rad CFX384 platform (Supplementary Table S3). Comparison of the RNA-seq data with qPCR data...
showed that good correlation was achieved for all targets, except for KDM5D (Fig. 4A). This target codes for Lysine Demethylase 5D and is located on the Y-chromosome (and thus expressed solely in male subjects). Hence, it was decided to remove it from further analysis. A geneNorm pilot experiment was performed to select proper reference genes for data normalization using the set of 9 candidate reference genes on all samples. Based on this analysis, it was found that the optimal normalization factor can be calculated as the geometric mean of reference targets PPIA and HPRT1.

3.7. Selection of a 7-gene classifier

Before selecting the best out of the 21 remaining (non-reference) targets, we have weighed and ranked them according to Cq-dependent precision and the spread in Cq-values. Based on these two characteristics, penalties were assigned to each target and the overall penalty was fed into the elastic net regularizer. This means that the regularization algorithm used the penalty values during each cycle of finding the next most appropriate target to expand the panel with. The source data for the regression was the log10-transformed CNRQ-values of the qPCR experiment. Upon cross-validation of the elastic net regularizer, the ideal panel size could be estimated (Fig. 4B). With increasing panel size (top row of numbers, read from right to left) the performance first improves (deviance diminishes), but after a panel size of about 12 the performance starts to deteriorate (due to overfitting). Based on the elastic net regularizer, it was decided that a panel size of 7 would result in sufficient performance and avoid overfitting. The panel of the top 7 contributing genes, listed in Table 2, were transferred to another laboratory and another qPCR platform (Roche LightCycler 480) and the expression of these 7 genes and the 2 reference genes (PPIA, HPRT1) was assessed in the same RNA samples (Supplementary Table S4). A logistic regression model was developed based on ΔCq-values and for each subject a probability of infection was calculated (Fig. 4C). Based on this training dataset, the algorithm had an AUC of 100% (95% CI: 100–100%) in ROC analysis. A cut-off of 0.497 was defined based on this ROC analysis and used to calculate sensitivity and specificity (100% and 100%, respectively).

3.8. Validation of the 7-gene classifier

The RT-qPCR-based 7-gene expression classifier as described above was validated in an independent sample set (called test set). This set consisted of 69 onchocerciasis patients, 8 non-endemic controls, and contained 51 endemic controls as well (Supplementary Table S5). Since for 3 non-endemic control samples, the RIN value was low (< 6), these were excluded from analysis, resulting in a total of 5 non-endemic control samples. The endemic control group had no palpable nodules but lived and worked in the same area as the onchocerciasis population. These samples were tested for the 7 selected genes and 2 reference genes (Supplementary Table S4). In this setting, the classifier developed in the training set was used to classify the independent samples in the entire test.

Fig. 4. Development of a 7-gene expression classifier. A Correlation between RNA-seq and qPCR data for the 22 selected genes. Plot shows ln-transformed FPKM-values of all samples for all targets (x-axis) plotted against their corresponding log10-transformed CNRQ-values (y-axis). Data points are colored according to the target of interest and linear trendline is added for each target in the same color. B Cross-validation of the elastic net regularizer based on log10-transformed CNRQ-values of the training data set. Graph must be read from right to left. With increasing panel size (top row of numbers) the performance first improves (deviance diminishes), but after a panel size of about 12 the performance starts to deteriorate (due to overfitting). Based on this plot, a panel size of 7 was chosen as the most appropriate. First vertical dotted line gives minimum mean cross-validated error ($\lambda_{min}$) and second vertical dotted line ($\lambda_{1se}$) gives a model such that error is within one standard error of the minimum. C Classifier based on a 7-gene logistic regression model. For all samples used in the training set, predicted probability of infection was calculated based on the 7-gene logistic regression model. D Validation of the 7-gene expression classifier. Classifier-based predicted probability of infection was determined in both training and test sample set of nodule-positive and non-endemic controls, complemented with a sample set of endemic controls.
set (Fig. 4D). Classification based on the algorithm had a sensitivity of 94.2% and a specificity of 60.0%. In the endemic control group, also 68.6% of the individuals was found to have a classifier outcome above the threshold. Upon combining the discovery and validation set, the classifier was found to have a total sensitivity of 96.0% and total specificity of 92.3%.

4. Discussion

In the work presented here, we used RNA-seq to evaluate changes in whole-blood transcription profiles in O. volvulus-infected individuals compared to non-infected controls from the city Kumasi, which is located close to the endemic community (non-endemic controls). All infected individuals had palpable nodules detected and were Ov16-seropositive and/or microfilaria-positive. These individuals had no or very low levels of microfilariae (< 3 mf/mg, except for one who had 9.9 mf/mg) in skin snips. This can be attributed to the fact that the village where samples were collected has been involved in multiple rounds of MDA, but some of the nodules might also contain non-fertile or monosexual worms. The primary objective of this study was therefore to investigate the effect of adult O. volvulus worms on the host’s blood transcriptome. We acknowledge that the presence of microfilariae in the skin might have a different effect on the host, but this effect could not be studied in our study population due to the very low microfilarialdermia observed in this population.

Using Ingenuity pathway analysis, we showed that the blood transcriptome in nodule-positive individuals differs from the blood transcriptome in non-infected individuals and that it might play an important role in the host-pathogen relationship. Infected individuals showed reduced oxidative phosphorylation and elf2a signaling, both indicating reduced cellular activity and proliferation. The effect of infection on oxidative phosphorylation has previously been reported in the context of e.g. viral infections where this also induces increased apoptosis (El-Bacha & Da Poian, 2013). Suppression of elf2a signaling leads to protein translation inhibition and correlates with the increased autophagy signature documented in our samples (Harding et al., 2003; Holcik & Sonenberg, 2005; Hotamisligil, 2010). Downregulation of elf2a signaling in infected individuals may be a host response of circulating hematopoietic cells to cope with a stressful environment imposed by the localized helmithinfection in the skin. Regulation of oxidative phosphorylation and elf2a signaling is driven by mTOR and AKT (Ramanathan & Schreiber, 2009; Altomare & Khaled, 2012; Tenkerian et al., 2015).

In line with this, we also found that RICTOR, an essential subunit of the mTORC2 complex, might be an upstream regulator activated in infected individuals (Sarbassov et al., 2004). It is not the first time that mTOR signaling was found to play an important role in the host-parasite interaction upon filarial infection. A proteomics study performed on dendritic cells from Brugia malayi-infected individuals showed that the parasite modulates these cells to influence the regulation of the host immune response by downregulating mTOR signaling, resulting in autophagy (Narasimhan et al., 2016). In the present study, we cannot define which cell types drive the observed differences in gene expression as whole blood was used, but our results confirm that TOR signaling is impacted by O. volvulus infection leading towards a switch in the metabolic state of circulating hematopoietic cells. Thus, our data suggest that the host’s immune system responds to the persistent presence of the parasite by modulating its metabolism towards a more glycolytic metabolism and the induction of cell-death mechanisms. This adaptation might be specifically induced by the parasite itself (through the production of so far unidentified soluble factors that act on immune cells), or it may represent an indirect consequence of the altered availability of nutrients in the blood upon parasite infection given that mTOR signaling is regulated by availability of nutrients such as glucose and amino acids, which are known to be changed in plasma of infected individuals (Jewell & Guan, 2013; Bennuru et al., 2017).

Next to the cell death and growth-related effects, we also observed a negative impact on specific immune functions induced by O. volvulus infection: proliferation of activated T lymphocytes, antibody response and Th17 immune response. Detailed investigation showed that several of the differentially expressed genes involved in these functions encode proteins that play a role in immune checkpoint mechanisms (Pardoll, 2012). These transcriptional differences will likely have an impact on immune modulation thereby dampening the immune response towards the parasite infection. Previous work has shown that Th17 differentiation and T-cell activation is regulated by the mTOR pathway and as such it is well possible that this connection also plays a role in the observations here (Kurebayashi et al., 2012; O’Brien & Zhong, 2012). Similar findings were also observed for Wuchereria bancrofti, where co-infection with this filarial parasite diminished dramatically the frequencies of malaria-specific Th1 and Th17 T-cells (Metenou et al., 2011). It has been suggested that helmithins are ‘old friends’ that have long coevolved with their hosts and act as inducers of immunoregulatory circuits (Rook, 2012; Rook et al., 2013). Prevalence of certain helmithins has been inversely associated to autoimmune diseases which are characterized by dysregulation of the specific T cell subsets Th1 and Th17 (Smallwood et al., 2017). In this respect, our findings might hint towards a transcriptional mechanism driving the downregulation of the Th17 immune response by O. volvulus, whereby this persistent immune tempering is essential for the helmithin to remain undetected in human tissues for many years but potentially also leads to tempering of otherwise overresponsive activity towards self-antigens. The strong upregulation of EB3, which is an IL-27 subunit, we observed in onchocerciasis patients might be a key event leading to the suppression of the Th17 immune response, as was already shown before for Trypanosoma cruzi and Leishmania (Anderson et al., 2009; Böhme et al., 2016).

Besides the mechanistic analysis of the gene expression data, we also sought to use the observed differences in gene expression as a biomarker for ongoing O. volvulus infection. An initial panel of 22 targets (including coding genes and lncRNAs) was selected for confirmation using RT-qPCR in the same RNA extracts that were used for RNA-seq. For all targets a statistically significant positive correlation was found between RNA-seq data and RT-qPCR data, except for KDM5D, which is a Y-chromosome linked gene and therefore only expressed in male subjects, which was also confirmed in RT-qPCR. Using the results of the 21 targets, elastic net regularization was used to identify how many, and which targets are needed to build a classifier that enables separation of infected individuals.
from uninfected controls. It was found that a classifier based on logistic regression of 7 targets was ideal and this classifier had an AUC in ROC analysis of 100% in the training data set. To validate this classifier, the same set of 7 targets was tested in a second sample set of both nodules-positive and non-endemic controls. Sensitivity was found to be 94.2% in the test set only and 96.0% in the combined training and test set. Specificity was found to be 60.0% in the test set only and 92.3% in the combined training and test set. It is important to mention that the two non-endemic control samples that were found to be positive based on this gene expression classifier were Ov16 IgG4-positive and as such might be truly infected with *O. volvulus*. These data demonstrate the potential utility of this gene expression classifier as a diagnostic tool for detection of *O. volvulus* infection. The analysis of the endemic control population indicated that 68.6% of those individuals were infected, an observation that is supported by the fact that also half of this population was positive for Ov16 IgG4 and this emphasizes the need for better diagnostic tools for detection of *O. volvulus* infection. The combination of Ov16 IgG4 and this new gene expression classifier might overcome some of the shortcomings of the Ov16 IgG4 test to meet the diagnostic specificity criteria put forward by the WHO to decide on stopping MDA (WHO, 2021). It will however require more studies in other populations with e.g. untreated microfilaridermic individuals and subjects with more recent and even pre-patent infection, but also with other filarial infections or from non-endemic countries to clearly define the intended use of this classifier. It will also be of interest to investigate the utility of this classifier as pharmacodynamic marker in studies evaluating novel macrofilaricide treatments as today no marker is available that can be used for that purpose besides the indirect measurement of the effect of treatment on microfilariae in the skin.

We need to emphasize that all the observations and conclusions made in this study should be confirmed in additional studies as described above. Some of the observations made here might be related to unknown environmental factors or other pathogen exposures. In this regard, we have recently shown that other putative onchocerciasis biomarkers could not be confirmed in an independent non-endemic population, and it cannot be excluded that this is also the case for the gene expression markers found here (Lagatie et al., 2019). Furthermore, the different groups used in this study were not age-matched and results might therefore be confounded by general metabolic health status of the study participants. Lastly, the size of this study was limited which might have skewed the analysis.

### 5. Conclusions

In conclusion, this RNA-seq study on whole-blood of *O. volvulus*-infected individuals provides insights in the mechanism of immune modulation by this helminth parasite. Transcriptional downregulation of Th17 and metabolic modulation by RICTOR appear to be the most important findings in this study. A 7-gene classifier was built, based on the RNA-seq data, confirmed in RT-qPCR and validated in an unseen data set. With a total sensitivity and specificity of 96.0% and 92.3%, respectively, this classifier has the potential to become an important addition to the onchocerciasis diagnostic toolbox.

### Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Ethical approval

Field study performed in Ghana was approved by the Committee on Human Research, Publications and Ethics of the School of Medical Sciences of the Kwame Nkrumah University of Science and Technology, Kumasi, Ghana and Permission was sought from Ashanti Regional Health Directorate and District Health Directorate of Adansi South District. All study subjects signed an informed consent form. All samples used in this study were anonymized and were collected from adults (18 years or above) only.

### CRediT author statement

Ole Lagatie: Conceptualization, methodology, formal analysis, investigation, writing - original draft, visualisation. Linda Batsa Debrah: Resources, investigation, writing - review & editing. Alex Y. Debrah: Resources, investigation, writing - review & editing. Lieven J. Stuyver: Conceptualization, supervision, investigation, writing - review & editing.

### Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

### Declaration of competing interests

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Ole Lagatie and Lieven J. Stuyver are current employees of Janssen Pharmaceutica NV, a Johnson and Johnson Company. Ole Lagatie and Lieven J. Stuyver own stock or stock options in that company.

### Acknowledgements

We thank Janssen Biobank for logistic support, Mohamed Lamkanfi and Cheryl Sweeney for scientific discussions and critically reading this manuscript and Benny Baeten and Marc Engelen from J&J Global Public Health for programmatic support.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jcrpbvd.2022.100100.

### References

Ali, M.M., Baraka, O.Z., Abdelrahman, S.I., Sulaiman, S.M., Williams, J.F., Homeida, M.M., Mackenzie, C.D., 2003. Immune responses directed against microfilariae correlate with severity of clinical onchodermatitis and treatment history. J. Infect. Dis. 187, 714–717. https://doi.org/10.1086/367709.
Altomare, D.A., Khaleed, A.R., 2012. Homeostasis and the importance for a balance between AKT/mTOR activity and intracellular signaling. Curr. Med. Chem. 19, 3748–3762.
Anders, S., Pyl, P.T., Huber, W., 2015. Htseq - a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–166. https://doi.org/10.1093/bioinformatics/btu638.
Anderson, C.F., Stumhofer, J.S., Hunter, C.A., Sacks, D., 2009. IL-27 regulates IL-10 and IL-17 from CD4+ cells in nonhealing leishmania major infection. J. Immunol. 183, 4619–4627. https://doi.org/10.4049/jimmunol.0804024.
Bennett, S., Lustigman, S., Abraham, D., Nutman, T.B., 2017. Metabolite profiling of infection-associated metabolic markers of onchocerciasis. Mol. Biochem. Parasitol. 215, 58–69. https://doi.org/10.1016/j.molbiopara.2017.01.008.
Boatín, B.A., To, L., Alley, E.S., Nagelkerke, N.J., Boshuizen, G., Habibena, J.D.F., 2002. Detection of Onchocerca volvulus infection in low-prevalence areas: A comparison of three diagnostic methods. Parasitology 125, 545–552.
Böhme, J., Rossnegel, C., Jacobs, T., Behrends, J., Hölscher, C., Erdmann, H., 2016. Epstein-Barr virus-induced gene 3 suppresses T helper type 1, type 17 and type 2 immune responses after Trypanosoma cruzi infection and inhibits parasite replication by interfering with alternative macrophage activation. Immunology 147, 338–348. https://doi.org/10.1111/imm.12565.
Chandrashankar, R., Ogunrinade, A.F., Weil, G.J., 1996. Use of recombinant Onchocerca volvulus antigens for diagnosis and surveillance of human onchocerciasis. Trop. Med. Int. Health 1, 575–580.
Chen, D., Luo, X., Xie, H., Gao, Z., Fang, H., Huang, J., 2013. Characteristics of IL-17 induction by Schistosoma japonicum infection in C57BL/6 mouse liver. Immunology 139, 523–532. https://doi.org/10.1111/jmim.12105.
Coffeng, L.E., Stolk, W.A., Golden, A., de Los Santos, T., Domingo, G.J., de Vlas, S.J., 2013. Characteristics of IL-17 from CD4– cells in nonhealing leishmania major infection. J. Immunol. 183, 4619–4627. https://doi.org/10.4049/jimmunol.0804024.
Bennett, S., Lustigman, S., Abraham, D., Nutman, T.B., 2017. Metabolite profiling of infection-associated metabolic markers of onchocerciasis. Mol. Biochem. Parasitol. 215, 58–69. https://doi.org/10.1016/j.molbiopara.2017.01.008.
Boatín, B.A., To, L., Alley, E.S., Nagelkerke, N.J., Boshuizen, G., Habibena, J.D.F., 2002. Detection of Onchocerca volvulus infection in low-prevalence areas: A comparison of three diagnostic methods. Parasitology 125, 545–552.
Böhme, J., Rossnegel, C., Jacobs, T., Behrends, J., Hölscher, C., Erdmann, H., 2016. Epstein-Barr virus-induced gene 3 suppresses T helper type 1, type 17 and type 2 immune responses after Trypanosoma cruzi infection and inhibits parasite replication by interfering with alternative macrophage activation. Immunology 147, 338–348. https://doi.org/10.1111/imm.12565.
Chandrashankar, R., Ogunrinade, A.F., Weil, G.J., 1996. Use of recombinant Onchocerca volvulus antigens for diagnosis and surveillance of human onchocerciasis. Trop. Med. Int. Health 1, 575–580.
Chen, D., Luo, X., Xie, H., Gao, Z., Fang, H., Huang, J., 2013. Characteristics of IL-17 induction by Schistosoma japonicum infection in C57BL/6 mouse liver. Immunology 139, 523–532. https://doi.org/10.1111/jmim.12105.
Coffeng, L.E., Stolk, W.A., Golden, A., de Los Santos, T., Domingo, G.J., de Vlas, S.J., 2013. Predictive value of Ov16 antibody prevalence in different subpopulations for elimination of African onchocerciasis. Am. J. Epidemiol. 188, 1723–1732. https://doi.org/10.1093/aje/kwz109.
