Direct proteomic detection and prioritization of 19 onchocerciasis biomarker candidates in humans

Bruce A. Rosa, Kurt Curtis, Petra Erdmann Gilmore, John Martin, Qiang Zhang, Robert Sprung, Gary J. Weil, R. Reid Townsend, Peter U. Fischer, Makedonka Mitreva

PII: S1535-9476(22)00262-6
DOI: https://doi.org/10.1016/j.mcpro.2022.100454
Reference: MCPRO 100454

To appear in: Molecular & Cellular Proteomics

Received Date: 19 July 2022
Revised Date: 30 October 2022
Accepted Date: 21 November 2022


This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology.
Direct proteomic detection and prioritization of 19 onchocerciasis biomarker candidates in humans

Authors: Bruce A. Rosa¹, Kurt Curtis¹, Petra Erdmann Gilmore², John Martin¹, Qiang Zhang², Robert Sprung², Gary J. Weil¹, R. Reid Townsend², Peter U. Fischer¹, Makedonka Mitreva¹,³,⁴*

Affiliations:
¹Division of Infectious Diseases, Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA
²Division of Endocrinology, Metabolism and Lipid Research, Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA.
³Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110, USA
⁴McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO 63108, USA

*Correspondence and request for material should be addressed to M.M. (email: mmitreva@wustl.edu). Makedonka Mitreva ORCID: 0000-0001-9572-3436
Running title: Direct proteomic detection of 19 onchocerciasis biomarkers

Abbreviations:

OV-pos: Samples from individuals infected with *Onchocerca volvulus*

OV-neg: Samples from uninfected individuals

PSM: Peptide-spectral matches

MDA: Mass drug administration

OPF: orthologous protein family

FPKM: Fragments per kilobase per million of reads mapped (relative gene expression)

DTT: Dithiothreitol

IAM: Iodoacetamide

FASP: Filter aided sample preparation

DB: Digestion buffer

TFA: Trifluoroacetic acid

PASEF: Parallel accumulation-serial fragmentation
Abstract

*Onchocerca volvulus*, the causative agent of onchocerciasis, infects over 20 million people and can cause severe dermatitis and ocular conditions including blindness. Current treatments employed in mass drug administration programs do not kill adult female worms, and common diagnostic tests cannot reliably assess viability of adult worms. There is an urgent need for better diagnostic tests to facilitate monitoring the efficacy of new treatments and disease elimination efforts. Here, eight plasma samples collected from individuals infected with *O. volvulus* and seven from uninfected individuals were analyzed by MS/MS spectrometry to directly identify *O. volvulus* proteins present in infected but absent in uninfected control samples. This direct proteomic approach for biomarker discovery had not been previously employed for onchocerciasis. Among all detected proteins, 19 biomarker candidates were supported by two or more unique peptides, identified in the plasma of at least three *O. volvulus*-infected human samples and absent in all control samples. Comprehensive analysis and ranking of these candidates included detailed functional annotation and a review of RNA-seq gene expression profiles. Isotope-labeled standard peptides were ran in parallel and validated MS/MS peptide identifications for 15 peptides from 11 of the 19 proteins, and two infected urine and one uninfected urine sample was used for additional validation. A major antigen / OVOC11613 was identified as the most promising candidate with eight unique peptides across five plasma samples and one urine sample. Additional strong candidates included OVOC1523 / ATP synthase, OVOC247 / laminin and OVOC11626 / PLK5, and along with OVOC11613, and were also detected in urine samples from onchocerciasis patients. This study has identified a promising novel set of proteins that will be carried forward to develop assays that can be used for diagnosis of *O. volvulus* infections and for monitoring treatment efficacy.

**Keywords:** Onchocerciasis; biomarker; proteomics; bioinformatics
Introduction

*Onchocerca volvulus* is a parasitic nematode transmitted by *Simulium* blackflies and is the causative agent of onchocerciasis, which is also known as “river blindness”. This debilitating neglected tropical disease can cause severe dermatitis and a variety of ocular conditions that can lead to visual impairment and permanent blindness (1). As of 2017, 20.9 million people were estimated to be infected, with 14.6 million afflicted with skin disease and 1.2 million with vision loss (2). Onchocerciasis is currently targeted for elimination, and in 2020 112 million people in Africa received ivermectin by mass drug administration (MDA) of a total of 239 million people who require MDA (1, 3).

While ivermectin clears first stage larval parasites (microfilariae, Mf) from the skin and eyes, adult *O. volvulus* female worms are not killed by this treatment. Mf repopulate the skin a few months after ivermectin treatment, so repeated rounds of MDA are required to reduce Mf in the skin and reduce transmission of infective larvae (4). There is an urgent need for treatments that effectively kill or permanently sterilize adult female worms. The efficacy of triple drug combination therapy with ivermectin, diethylcarbamazine and albendazole (IDA), which kills or sterilizes other filarial parasites for onchocerciasis is unknown (5). Diagnostic tests commonly used to detect the Mf such as skin snips, diethylcarbamazine patch tests and PCR cannot reliably assess the viability of adult worms. Adult worm viability can be assessed by microscopic examination of surgically removed onchocercal nodules, but this is invasive and expensive, and it requires special expertise. Antibody tests for onchocerciasis such as those based on recombinant antigen Ov16 (6, 7) are not reliable markers for the presence of viable adult worms, because antibodies persist long after the death of adult worms (8, 9). This makes it difficult to use this test to inform decisions about whether MDA should be continued (10). Other studies have identified additional antigen candidates, including OVOC3261 (11) and other *O. volvulus* proteins, that are specifically bound by IgG from infected individuals that are *O. volvulus*-specific and highly expressed by adult female worms (12). While these antigens may improve sensitivity, it is not
clear that any antibody assay will prove to be specific for current infection. Thus, there is an urgent need for better diagnostic tests for onchocerciasis to detect active infections and to monitor the efficacy of new treatments (13).

Problems with antibody serology have led to attempts to detect worm products or metabolites in bodily fluids. Early claims of success (for example with N-acetyltyramine-O,β-glucuronide (14)) were not independently confirmed for sensitivity or specificity (15, 16). Another metabolomics study has identified two potential plasma biomarkers for *O. volvulus* infection (inosine and hypoxanthine) which had generally low detection levels in a separate non-endemic cohort, but have yet to be validated in a separate cohort of infected individuals (17). Phospholipid profiling of *O. volvulus*, *Onchocerca ochengi* and *Litomosoides sigmodontis* identified species-specific lipids, but while the lipid biomarker phosphatidylethanolamine was detected at high levels in *O. ochengi* nodules in cows, the potential plasma or serum lipid biomarkers were below detection limits in infected animals and humans (18). Additionally, miRNA profiling study identified four *O. volvulus*-derived mature miRNAs in the serum and plasma of infected individuals and not in uninfected individuals, with two (miR-71 and lin-4) being detected in individuals from both Ghana and Cameroon (19). At present no commercially available test is sensitive or specific for detection of active infections by adult worms.

In this study we have generated new mass-spectrometry data from *O. volvulus* infected and uninfected subjects and performed a multi-platform analysis of proteomic, genomic, and transcriptomic *O. volvulus* datasets that resulted in identification, characterization and prioritization of *O. volvulus* candidate biomarkers that could be used to develop assays for diagnosis of adult females. These candidate biomarkers have been identified by proteomic detection in the blood and urine of several infected individuals and not in uninfected individuals, and additional proteomic validation is provided by the manual evaluation of spiked-in heavy peptides. The long-term goal of this initiative is to further evaluate these promising biomarker candidates for their use to detect active infections and monitor treatment efficacy.
Experimental Procedures

Experimental Design and Statistical Rationale

A total of 18 independent biological replicates were analyzed in the MS/MS proteomics analysis: Plasma and urine samples collected from individuals infected with *O. volvulus* (OV-pos; 8 plasma replicates, 2 urine replicates), and from uninfected individuals used as negative controls for detection (OV-neg; 7 plasma replicates, 1 urine replicate). Some samples were collected from individuals and some represented pooled samples as indicated in supplemental Table S1. Because data collection for this project has spanned many years, several different LC:MS technologies were used for data collection, including hybrid quadrupole Orbitrap™ (Q-Exactive) and trapped ion mobility time of flight (timsTOF) mass spectrometers. Technical replicates were collected from some samples from two different MS/MS proteomics machines (Q Exactive or timsTOF), as indicated per sample in supplemental Table S1. Fractionation of timsTOF samples was also performed, with separate runs per sample. For all technical replicates (including multiple fractions of the same sample), peptide-spectral matches (PSMs) were summed for each peptide in each sample, and they were not considered as replicates for any statistical analyses or presented as replicates in any figures or tables. A stepwise filtering process based on presence / absence across OV-pos and OV-neg samples is described in detail in the results / discussion, and did not rely on any statistical test validation. The overlap between some identified gene sets of interest with previous results including functional annotations and RNA-seq data were tested using a Fischer exact test or a two-tailed T-test (unequal variance). Informed consent was provided by the patients in accordance with the Declaration of Helsinki.

Plasma and urine sample collection

Origins, collection dates and sample processing data for all samples analyzed by proteomics are summarized in supplemental Table S1. Each plasma sample contained either an individual plasma sample or a pooled sample with up to 15 plasma samples, containing equal
volumes of each patient sample (usually 100 to 250 µL plasma each). Samples from subjects infected with *O. volvulus* were collected in Uganda (20) or Cameroon (21), and Mf density was determined by duplicate skin snips as described previously (20). Plasma samples from Uganda were from an area hyperendemic for onchocerciasis before any intervention. All donors had palpable nodules and high skin Mf densities over 50 Mf/mg (see supplemental Table S1). Samples from Cameroon were from subjects with low to moderate Mf densities. Additional samples from uninfected individuals were collected from India (22) and from the USA. No samples were collected from lymphatic filariasis (LF)-endemic areas. Ugandan samples were collected from an area co-endemic for *Mansonella perstans*, but all tested samples were negative for *Mansonella* Mf by a modified Knott technique. *Wuchereria bancrofti* is not endemic most parts of Cameroon, and blood samples collected during the day were negative for *L. loa* and *M. perstans* Mf. Onchocerciasis is not endemic to India and blood samples were negative for circulating *W. bancrofti* antigen. The uninfected control samples from the USA were collected from uninfected volunteers from Barnes Jewish Hospital, St Louis, MO, and represent serum samples, but for the sake of consistency in labelling, are labelled here as plasma samples. For these samples, non-endemic control sera samples were obtained de-identified from the Barnes-Jewish-Christian Hospital Clinical laboratory in St Louis. The Washington University in St. Louis Human Research Protection Office (an institutional review board) determined that work with these de-identified samples did not constitute human subjects research. In addition, the use of human samples for the development of new diagnostics was approved by the IRB at the Washington University School of Medicine under the ID #201102546.

Plasma samples were collected in EDTA-coated tubes, and stored in 1 mL aliquots in cryovials (Nunc, Thermo Fisher Scientific). Plasma samples collected in filarial parasite endemic countries underwent at least one freeze-thaw cycle, but not more than three freeze-thaw cycles. Plasma samples were stored long-term without specific protease inhibitors at -80°C. Urine samples were treated similarly, but aliquots were stored in 5 mL tubes.
Samples were assessed by LC-MS following (i) depletion to enrich for potential target proteins by removal of common human plasma proteins, (ii) immunoprecipitation (IP) using anti-parasite antibodies, or (iii) by precipitation of immune complexes (IC). All of these approaches are detailed below. Urine samples were exclusively individual samples, and were concentrated but were not depleted for host proteins.

IgG and Immunoprecipitation preparation

Rabbits were immunized with *O. volvulus* adult worm extract, and IgG was prepared from serum by ammonium sulfate precipitation as previously described (23). Total IgG was isolated from plasma samples by ammonium sulfate (Sigma-Aldrich, St. Louis, USA) precipitation. Ammonium sulfate was added drop-wise, with constant sample mixing for a final concentration of 50%. The mixture was incubated overnight at 4°C causing IgG precipitation, and then centrifuged at 3000 x g for 30 minutes at 4°C. The IgG pellet was suspended in PBS and dialyzed extensively against PBS (Thermo Fisher Scientific, Waltham, MA, USA) using 20kDa MWCO membrane cassette (Thermo Fisher Scientific) to remove ammonium sulfate. IgG samples were coupled to Pierce NHS-Active Agarose (Thermo Fisher Scientific). For 1 mL of IgG bound agarose, 2 mL of 50% NHS-Active Agarose was added to a 2 mL Pierce centrifuge column (Thermo Fisher Scientific), and the column was centrifuged at 1000 X g for 1 minute in a swinging bucket rotor to remove the column buffer. The agarose was washed by adding 2 mL of ultra-pure water to the column and centrifuging at 1000 X g for 1 minute. Two mL IgG in PBS (20 mg IgG) was added to the agarose and incubated for 2 hours at room temperature with constant mixing, then it was centrifuged at 1000 X g for 1 minute. Two mL PBS was then added to the column as a wash, and centrifuged at 1000 X g for 1 minute. The column was then blocked using 2 mL 1.0 M ethanolamine (pH 7.4; Sigma-Aldrich), incubated for 20 minutes at room temperature, and centrifuged at 1000 X g for 1 minute. Six mL PBS was added to the column, centrifuged at 1000 X g for 1 minute as a final
wash. The final 1 mL of beads were then diluted in PBS to a final volume of 10 mL (10% working dilution for immunoprecipitation).

Rabbit anti-\textit{O. volvulus} IgG beads were used to capture \textit{O. volvulus} antigens in human sera. Human plasma (250 µL) was added to an equal volume of 10% IgG beads in a 1.5 mL micro centrifuge tube (Sigma Aldrich) and incubated overnight at 4°C. The sample was centrifuged 12,400 x g for 2 minutes, the supernatant was discarded, and the pelleted beads were washed three times in PBS with Tween (Thermo Fisher Scientific).

**Immune complex preparation**

Human plasma (400 µL) was added to an equal volume of 0.2 M EDTA pH 7.5 and 12% PEG 8000 (Thermo Fisher Scientific) in 0.1 M borate buffer pH 8.3. The mixture was incubated overnight at 4°C and then centrifuged at 16000 X g for 30 minutes. The supernatant was discarded, and pellet was washed three times with ice cold 2.4% PEG 8000 in 0.1 M borate buffer pH 8.3. The final pellet was then submitted for MS analysis.

**Multi-affinity fractionation of plasma**

Fourteen of the highest abundant plasma proteins (albumin, IgG, α1-antitrypsin, IgA, IgM, transferrin, haptoglobin, α2-macroglobulin, fibrinogen, complement C3, α1-acid glycoprotein [orosomucoid], HDL [apolipoproteins A-I and A-II], LDL [mainly apolipoprotein B]) were removed (>95%) using Seppro LC5IgY14 columns (Sigma Aldrich, St. Louis MO USA) according to vendor protocol.

Processed plasma samples as described above were thawed and centrifuged at 15,000 x g for 10 min. The supernatants were transferred to a 1.5 mL tube and an equal volume of 2X Tris buffered saline (300 mM NaCl, 20 mM Tris-HCl, pH 7.4) was added. The diluted supernatant was filtered with a 0.45 µm spin filter and centrifuged at 10,000 x g for 10 min. A 25 µL aliquot of the concentrate was transferred to a 0.5 mL tube and 85 µL TBS (150 mM NaCl, 10 mM Tris-HCl, pH
7.4) was added. Then, 110 µL of sample was transferred to an autosampler vial and kept at 4°C. Multi-affinity removal of high-abundance proteins was performed using an Agilent 1260 Infinity liquid chromatography system with an Agilent Hu-14 column, 4.6 mm x 50 mm. The samples (0.1 mL) were loaded onto the column at a flow rate of 0.5 mL/min. The column effluent was monitored at A280. The flowthrough fractions (2.25 mL) were collected in TBS. The bound fractions (4.25 mL) were eluted with 0.1 M glycine, pH 2.5. The flow through and bound fractions were concentrated using a 3000 MWCO spin filter. For deep-scale proteomic analysis runs, multiple aliquots were prepared using multi-affinity chromatography as described above, and transferred into separate 3K filter units. Protein concentration was determined by a BCA assay before freezing samples.

Protein was prepared from urine (15 mL) by concentration using a 5000 MWCO K filter to ~ 200 µL. The concentrated urine was precipitated with cold acetone (1:5 parts of sample to cold acetone (vol/vol)). The samples in cold acetone were kept at -20°C overnight followed by centrifugation for 10 min at 20,000 x g. The protein pellets were solubilized in SDS buffer with sonication and protein concentration was determined by BCA assay. Immune complex pellets were solubilized in SDS buffer with sonication and protein concentration was determined by BCA assay. Immunoprecipitated proteins were eluted from antibody beads with 0.1M glycine (pH 2) followed by solubilization with SDS buffer. The eluates were lyophilized and digested as described below using filter-aided sample preparation methods (FASP).

Preparation of peptides
A summary of peptide preparation methods for the each of the samples is provided in supplemental Table S1. Peptides were prepared from multi-affinity fractionated sera by endoprotease digestion of ~ 250 µg of the “depleted” sera. The buffer (100 mM Tris-HCl, pH 8.0) containing 10 M urea was added to the “depleted” samples to a final concentration of 8 M urea. Disulfide bonds were reduced with 6.5 mM dithiothreitol (Pierce Chemical) for 45 min at room temperature and the Cys residues were alkylated with 13 mM iodoacetamide (IAM; Pierce
Chemical) for 45 min at room temperature in the dark. The urea concentration was diluted to 2 M by addition of 100 mM Tris-HCl buffer (pH 8.0) prior to addition of 5 µg of LysC (Wako Chemicals). Samples were incubated at room temperature for 2 h, followed by addition of trypsin (1:100 enzyme/substrate) and digestion overnight at room temperature. Peptides were desalted on a 200 mg Sep-PAK column. Peptides were eluted with 1.5 ml 50% (vol/vol) MeCN, 0.1% (vol/vol) FA. An aliquot (10%) of the eluate was assayed for peptide quantity and the remainder was lyophilized and stored at -80°C.

The immune complex pellets were solubilized with 30 µl of SDT buffer (4% [wt/vol], 100 mM Tris-HCl pH 8.0). Protein disulfide bonds were reduced using 100 mM DTT with heating to 95°C for 10 min. Peptides were prepared as previously described using a modification (24) of the filter-aided sample preparation method (eFASP) (25). The samples were mixed with 200 µl of 100 mM Tris-HCL buffer, pH 8.5 containing 8 M urea (UA buffer). The samples were transferred to the top chamber of a 30,000 MWCO cutoff filtration unit (Millipore) and processed to peptides as previously described (25). The peptides were dried in a Speedvac concentrator (Thermo Fisher Scientific, Savant DNA 120 Speedvac Concentrator) for 15 min. The dried peptides were dissolved in 1% (vol/vol) TFA and desalted using two micro-tips (porous graphite carbon, BIOMEKNT3CAR) (Glygen) on a Beckman robot (Biomek NX), as previously described (Chen Mol. Cell Proteomics PMID 22338125). The peptides were eluted with 60 µl of 60% (vol/vol) MeCN in 0.1% (vol/vol) TFA and dried in a Speed-Vac (Thermo Fisher Scientific, Model No. Savant DNA 120 concentrator) after adding TFA to 5% (vol/vol). The peptides were dissolved in 20 µl of 1% (vol/vol) MeCN in water. An aliquot (10%) was removed for quantification using the Pierce Quantitative Fluorometric Peptide Assay kit (Thermo Fisher Scientific, Cat. No. 23290). The remaining peptides were transferred to autosampler vials (Sun-Sri, Cat. No. 200046), dried and stored at -80°C. For samples treated with PNGaseF before digestion, samples were exchanged into ammonium bicarbonate buffer (100 mM, pH 8.0) after alkylation, then incubated
with 20 U PNGaseF in 100 µl of the buffer for 1 hour in a thermomixer at 600 rpm. Protein digestion and peptide preparation were then performed using the FASP method.

**Off-line fractionation of peptides**

The peptides prepared from depleted sera, immune complexes and urine were fractionated using basic pH reverse-phase chromatography prior to LC-MS (26). The chromatograph was an Agilent HPLC system 1260 equipped with an Agilent Zorbax 300Extend-C18, 4.6 x 250 mm column. The peptides were separated using a binary gradient of solvent A (4.5 mM ammonium formate (pH 10) in 2% (vol/vol) MeCN) and Solvent B (4.5 mM ammonium formate pH 10 in 90% (vol/vol) MeCN) at a flow rate of 1 mL/min. The gradient for equilibrating the column was as follows (time in min): %B: 0, 0; 4, 0; 19, 100; 21, 100; 22, 0; 24, 0; 39, 100; 41, 100; 42, 0; 70, 0.

The gradient method for peptide chromatography was as follows (time in min.): %B: 0, 0; 9, 0; 13, 6; 63, 28.5; 68.5, 34; 81.5, 60; 98.1, 60; 100, 0; 120, 0. The lyophilized peptides were dissolved in 465 µL of 4.5 mM ammonium formate (pH 10) in 2% (vol/vol) MeCN by vortexing. Each sample was centrifuged at 18,000 x g at RT for 5 min and transferred to the autosampler vial. The sample (450 µL) was loaded into the sample loop and injected onto the column for 1.2 min at a flow rate of 1 mL/min, and the gradient program was initiated. Column effluent was monitored at 214 nm. The column flow through and 96 fractions were collected. The 96 fractions were concatenated to 24 as previously described (26). The fractions were lyophilized to complete dryness (~ 16 h). Gradient optimization and system monitoring was performed using standard of bovine serum albumin tryptic peptides (500 µg / analysis). System performance was evaluated prior to each sample analysis by comparison with legacy peak resolution and retention times from previously optimized chromatographies using standard peptides.

**Mass spectrometry**
The purified peptides from depleted plasma, immune complexes and urine were analyzed using ultraperformance HPLC (UPLC) mass spectrometry. Initial studies were performed with a hybrid quadrupole Orbitrap instrument and more recently with a trapped ion mobility time-of-flight mass spectrometer (timsTOF) (27).

For ultraperformance HPLC (UPLC) hybrid quadrupole Orbitrap mass spectrometry (28) the peptides in FA (1% (vol/vol)) were loaded in 2.5 µl onto a 75 µm i.d. × 50 cm Acclaim® PepMap 100 C18 RSLC column (Thermo Fisher Scientific) on an EASY nanoLC (Thermo Fisher Scientific) at a constant pressure of 700 bar at 100% A (0.1%FA). Prior to sample loading the column was equilibrated to 100%A for a total of 11 µl at 700 bar pressure. Peptide chromatography was initiated with mobile phase A (1% FA) containing 2%B (100%ACN, 1%FA) for 5 min, then increased to 20% B over 100 min, to 32% B over 20 min, to 95% B over 1 min and held at 95% B for 19 min, with a flow rate of 250 nL /min. The data was acquired in data-dependent acquisition (DDA) mode. The full-scan mass spectra were acquired with the Orbitrap mass analyzer with a scan range of \( m/z = 325 \) to 1500 and a mass resolving power set to 70,000. Ten data-dependent high-energy collisional dissociations were performed with a mass resolving power set to 17,500, a fixed lower value of \( m/z \) 100, an isolation width of 2 Da, and a normalized collision energy setting of 27. The maximum injection time was 60 ms for parent-ion analysis and product-ion analysis. The target ions that were selected for MS/MS were dynamically excluded for 20 sec. The automatic gain control (AGC) was set at a target value of 1e6 ions for full MS scans and 1e5 ions for MS2. Peptide ions with charge states of one or > 8 were excluded for MS2 acquisition. QE parameter setting details are provided in **supplemental Table S2A**, and EASYnLC parameters are provided in **supplemental Table S2B**.

For timsTOF LC-MS, the peptides were separated using a *nano-ELUTE*® chromatograph (Bruker Daltonics, Bremen Germany) interfaced to a timsTOF Pro mass spectrometer (Bruker Daltonics) with a modified nano-electrospray source (CaptiveSpray, Bruker Daltonics). The mass
spectrometer was operated in Parallel accumulation-serial fragmentation (PASEF) mode (27). The samples in 2 µL of 1% (vol/vol) FA were injected at a flow rate of 400 nL / min onto the column (75 µm i.d. × 25 cm Aurora Series) with a CSI emitter (Ionopticks). The column temperature was set at 50 °C. The column was equilibrated using constant pressure (800 bar) with 8 column volumes of solvent A (0.1% (vol/vol) FA). Sample loading was performed at constant pressure (800 bar) in 2 µL of Solvent A.

The peptides were eluted using the following gradient of Solvent B (0.1% (vol/vol) FA/MeCN): solvent A containing 2% B was increased to 17% B over 60 min, to 25% B over 30 min, to 37% B over 10 min, to 80% B over 10 min and constant 80% B for 10 min. The MS1 and MS2 spectra were recorded from m/z 100 to 1700. Suitable precursor ions for PASEF-MS/MS were selected from TIMS-MS survey scans by a PASEF scheduling algorithm (27). A polygon filter was applied to the m/z and ion mobility plane to select features most likely to be multiply charged peptide precursor ions. Quadrupole isolation width was set to 2 Th for m/z < 700 and 3 Th for m/z > 700. The collision energy was ramped stepwise as a function of increasing ion mobility: 52 eV for 0–19% of the ramp time; 47 eV from 19–38%; 42 eV from 38–57%; 37 eV from 57–76%; and 32 eV for the remainder. The TIMS elution voltage was calibrated linearly using the Agilent ESI-L Tuning Mix (m/z 622, 922, 1222). Detailed timsTOF parameters are provided in supplemental Table S2C and Nanoelute parameters are provided in supplemental Table S2D.

Peptide detection from MS/MS spectra
Detected MS/MS proteomics spectra were assigned to *O. volvulus* peptides using MaxQuant (29) (version 1.6.17), by searching against a database of proteins based on the current *O. volvulus* genome (30) annotation (PRJEB513, WBPS15, WS276; 12,224 proteins) and the “common Repository of Adventitious Proteins” database (cRAP, v2012.01.01, https://www.thegpm.org/crap/; 115 proteins) to screen for common contaminants. The searches were performed with a fragment ion mass tolerance of 40 ppm and a parent ion tolerance of 20
ppm. The enzyme search specificity was selected as tryptic/P, allowing for a maximum of 4 missed cleavages. Carbamidomethylation of cysteine residues was specified as a fixed modification. The searches were conducted with the following variable modifications: deamidation of asparagine, deamidation of glutamine, formation of pyro-glutamic acid from N-terminal glutamine, acetylation of protein N-terminus, oxidation of methionine, and pyro-carbamidomethylation of N-terminal cysteine residues. Peptides and protein results were filtered at 1% false-discovery rate (FDR) by searching against a reversed protein sequence database.

MaxQuant settings used for spectra-peptide assignments were defined in the xml configuration file, which is provided in supplemental Table S2E, and all MaxQuant output files are provided as part of the raw data upload to the iProX database (IPX0004317000, ProteomeXchange Consortium identifier PXD033659; accession numbers provided in supplemental Table S1).

Resulting identified peptides were screened against the *Homo sapiens* proteome (GRCh38 release 105 (32) and including all isoforms; 117,909 isoform sequences from 20,465 coding proteins), treating leucine and isoleucine amino acids as matches since they are indistinguishable in the spectral matching. Any peptides exactly matching human sequences were discarded for all downstream analysis. Peptides were quantified per sample by the number of peptide-spectral matches (PSMs; supplemental Table S3A), but PSM counts were not utilized for any peptide filtering or statistical analysis. All 22,648 peptides assigned to *O. volvulus* sequences were also screened against *Wolbachia* (*Wolbachia* endosymbiont of *O. volvulus* str. Cameroon, NCBI tax id:1410384, GenBank release 250), and only four peptides had an exact match to *Wolbachia*, all of which were only detected in the whole-worm lysate sample. Those peptides are indicated with an asterisk in supplemental Table S3A. All data for *O. volvulus*-assigned spectral-peptide matches are provided in supplemental Table S3B, including charges, masses, modifications and scores.

The number of peptides per protein per sample are provided as part of supplemental Table S4. Although all peptides matching all proteins are reported in the supplemental tables,
only proteins supported by two or more unique peptides (as indicated in supplemental Table S4) were considered to be reliably detected in the absence of the detailed annotation of spectra, which can be performed using the available raw data (iProX database (31) accession IPX0004317000, ProteomeXchange Consortium identifier PXD03365).

A step-wise filtering and ranking process for detected proteins was performed to identify those with the highest potential to serve as biomarkers for the presence of viable adult *O. volvulus* females. This analysis compared results obtained with onchocerciasis samples to those obtained with negative control samples as described in detail in the Results/Discussion section and in Fig. 1. Proteins identified at each step in the filtering and ranking process are indicated in supplemental Table S4.

**Transcriptome sample collection and analysis**

Unprocessed RNA-seq reads were retrieved from analyses described in a previous study (from eight adult female samples (12)). RNA-seq reads were mapped to the *O. volvulus* genome (30) (PRJEB513, WBPS15, WS276) using STAR (33), quantified with featureCounts (34), and normalized using DESeq2 (35). Relative expression was calculated using fragments per kilobase per million reads (FPKM) normalization per gene per sample. All normalized expression values are provided in supplemental Table S4.

***O. volvulus* deduced protein annotation database**

The *O. volvulus* genome / proteome annotation was downloaded from WormBase Parasite (36) (PRJEB513, version WBPS15 WS276). The retrieved functional annotations were complemented with additional assignments using InterProScan v5.42 (37) to identify gene ontology (38) classifications and InterPro functional domains (39), and GhostKOALA v2.2 (40) to assign KEGG (41) annotations. Potentially secreted proteins were identified using both SignalP v5.0 (42) for signal peptides and transmembrane domains, and SecretomeP v2.0 (43) to identify
proteins with non-classical secretion sequences (where any proteins with 2 or more transmembrane domains were not classified as secreted).

Protein conservation data across nematodes and hosts was quantified using BLAST (44) and OrthoFinder (v2.4.1) (45), used to compare protein sequences across all species and define orthologous protein families (OPFs), using default values. Protein sequence data was downloaded from WormBase Parasite (36) (WBPS15 WS276) for *Ancylostoma ceylanicum* (PRJNA72583), *Loa loa* (PRJNA246086), *Necator americanus* (PRJNA72135), *Ascaris lumbricoides* (PRJEB4950), *Caenorhabditis elegans* (PRJNA13758), *Brugia malayi* (PRJNA10729), *Trichuris trichiura* (PRJEB535), *Strongyloides stercoralis* (PRJEB528), *O. ochengi* (PRJEB1465) and *W. bancrofti* (PRJNA275548; improved annotation). In addition, host outgroups were downloaded from Ensembl (32) for *Oryctolagus cuniculus* (Rabbit; OryCun2.0), *Homo sapiens* (Human; GRCh38.p13) and *Drosophila melanogaster* (BDGP6.32). The top BLAST (44) hit for each *O. volvulus* protein was identified, including the E value, alignment length, % identity, and whether the top hit was reciprocal (NCBI blastp v2.7.1+, default settings. All protein functional annotations and orthology information including OrthoFinder groups and BLAST results for all *O. volvulus* proteins are provided in supplemental Table S4.

**Heavy peptide spectral matching for detected peptide validation**

Isotope-labeled standard peptides (“heavy peptides”) were synthesized (Vivitide, Gardner, MA) for peptides shorter than 18 amino acids that were detected from the top 19 proteins from the proteomics analysis. These isotope-labeled heavy peptides were spiked into an aliquot of the sample where the candidate peptide was initially observed. Supplemental Table S5 shows the observed m/z values, intensities, and mass error for the fragmentation ions from candidate biomarker peptides and complimentary values for the corresponding stable isotope labeled peptides which were used to corroborate peptide sequence assignment. The fragment ion mass tolerance for matching used was agreement within +/- 20 ppm and +/- 40 ppm for the parent and
fragment ions respectively. Given the expected low abundance of *O. volvulus* proteins in blood and the potential for heterogeneous antibody responses across individuals, we anticipated that spectra for candidate biomarkers would be of low intensity and contain ions below the limit of detection, requiring manual interpretation. The minimum criteria for spectral matching / verification were agreement of the parent ion m/z value with the calculated value within mass error specification and observation of a series of fragment ions that defined a sequential amino acid sequence of four residues. Where only spectra from the modified peptide were observed, both the calculated values and the spectra from the synthetic, unmodified peptide were used for interpretation.
Results and Discussion

To identify candidate biomarkers with potential to detect adult *O. volvulus* females, we performed mass spectrometry (MS) analyses with human plasma and urine samples from infected and uninfected individuals. False positive identification of biomarker candidates (a common problem in MS studies) was minimized by comparing MS results from eight individual or pooled plasma samples from infected people with results obtained from seven individual or pooled samples from uninfected people. Peptides or proteins were excluded as infection-detection candidates if they were also detected in negative samples. Analysis of three types of samples derived from plasma (immune complexes or IC, proteins immunoprecipitated from human samples with antibodies to *O. volvulus* or IP, and samples depleted of common plasma proteins; “depleted plasma”) increased chances for detecting signals. Detection in multiple positive samples prepared by different methods further increased confidence in the results. We also analyzed concentrated urine samples from infected and uninfected individuals to assess the potential of urine (with much less host protein than plasma) as an alternative to plasma for biomarker detection, and for additional confirmation for candidate proteins. Samples for the study were from Uganda and Cameroon (infected and uninfected samples), India, Indonesia, and the USA (uninfected samples; Table 1). This approach of using MS proteomics to directly screen both infected and uninfected human plasma and urine samples has not been previously reported for onchocerciasis.

Production of a comprehensive dataset to support biomarker identification and characterization

Plasma and urine samples collected from individuals infected with *O. volvulus* (OV-pos; 8 plasma, 2 urine) and from uninfected individuals (OV-neg; 7 plasma, 1 urine) were analyzed by MS/MS proteomics. Several proteomic workflows were used to prepare peptides for LC/MS analysis, including from plasma with 14 high abundance proteins removed, immunoprecipitates and
immune complexes. Details of sample processing and MS configuration are provided in the materials and methods, and sample metadata, raw data accessions (from the iProX database (31) accession IPX0004317000, ProteomeXchange Consortium identifier PXD033659), and processing information are provided in supplemental Table S1. Mass spectrometric configuration data for all machines are provided in supplemental Table S2A-D.

For all datasets, MaxQuant (29) was used to match MS/MS spectra to O. volvulus peptides, using the same settings and approach for all samples (configuration file provided in supplemental Table S2E). After screening out identified peptides that had exact amino acid sequence matches in the human proteome (treating leucine and isoleucine as identical matches), an average of 339.4 peptide-spectral matches (PSMs) matching an average of 169.8 peptides in 159.4 proteins were identified in the samples from O. volvulus-positive individuals, compared to an average of 114.7 PSMs, 72.8 peptides and 70.4 proteins in the samples from uninfected individuals (false positives; Table 1). In the whole-worm lysate sample from adult female O. volvulus, 70,226 PSMs identified 21,109 peptides in 2,405 proteins. Data from this sample was analyzed in order to identify proteins, peptides and relative abundance (quantified by PSMs) in adult female worms. This analysis provided additional confidence in results from human samples when the same peptides and PSMs were detected by MS in O. volvulus-positive plasma and urine samples. The PSM counts for every detected peptide in every sample are provided in supplemental Table S3.

We used genomic and extensive transcriptomic (including RNAseq profiles of individual adult females) databases to provide additional insights into putative protein functions and specificity of candidate peptides for O. volvulus (compared to human and to other parasitic nematodes co-endemic to O. volvulus) to assess each identified protein’s potential for use as a biomarker for adult O. volvulus infection. Our comprehensive functional annotation data ensures that the maximum number of proteins have useful descriptions for interpretation of results. We also utilized gene expression data from eight adult female O. volvulus samples (12) to identify
genes that are consistently and highly expressed in adult females. We also evaluated the orthology of each protein by examining orthologous protein families (OPFs) spanning the phylogeny of nematodes and hosts, and by quantifying protein similarity by BLAST searches to species of interest. All of this detailed annotation data is provided for all *O. volvulus* proteins in an MS Excel database saved as supplemental Table S4; we used this information to evaluate and prioritize a list of top-ranked proteins as described below.

Evidence based detection of promising biomarker candidates

High-priority *O. volvulus* biomarkers were those that were present in multiple samples from infected individuals (OV-pos), but not in samples from uninfected individuals (OV-neg). The overall workflow for the filtering and ranking of candidate protein biomarkers is shown in Fig. 1. Out of 12,224 total annotated *O. volvulus* proteins, 989 were detected across any of the OV-pos plasma samples, and 803 of those were not also detected in any OV-neg plasma samples (allowing up to one false-positive peptide match from a candidate biomarker protein from a single OV-neg sample as long as that peptide was not also detected in an OV-pos sample). Of these 803 proteins, 90 had two or more peptides detected across OV-pos samples. This requirement increased our confidence in protein identifications, and these 90 proteins were considered to represent the “true positive” protein set. Compared to the “false positive” set of 207 proteins with detection in OV-neg samples (not part of the 803 as defined in Fig. 1), the true positives (i) were more likely to be functionally annotated (36) (56.7% of true positives vs. 41.1% of false positives, Fischer exact test $P = 0.016$); (ii) had much higher gene expression levels across the adult female RNA-seq dataset (average FPKM of 420.3 among true positives vs. 49.0 in false positives, two tailed-T-test $P = 0.0098$); and (iii) were more likely to be phylogenetically conserved across nematodes (55.5% of true positives vs. 41.1% of false positives, Fisher exact test $P = 0.023$). Of these 90 “true positive” proteins, 19 were detected in at least three of the OV-pos plasma samples.
Direct proteomic detection of 19 onchocerciasis biomarkers (Fig. 2). These 19 proteins were considered to be “promising biomarker candidates” that merit further study.

Confirmation of detected peptide sequences from the 19 promising biomarker candidates was performed by spectral matching to isotope-labeled synthetic peptides (“heavy peptides”). The MS/MS spectra from the heavy peptides were compared to the observed spectra of candidate biomarker peptides with regard to agreement to m/z values, intensities, and mass errors (supplemental Table S5). This analysis was used to confirm peptides for 15 peptides from 11 of the top-ranked 19 proteins (described below). This analysis was intended to provide an additional level of confidence for detected proteins. However, failure of this verification does not exclude proteins as potential biomarkers, because some peptides could not be successfully synthesized and because low concentrations of native biomarker in plasma samples can result in an insufficient number of fragment ions to align confidently confirm a match between synthetic heavy and naturally present peptides.

Favorable properties-based ranking of the 19 high priority candidates

The 19 most promising biomarker candidates were subsequently ranked based on several properties (Fig. 1). First, the proteins identified in the highest number of OV-pos plasma samples were ranked at the top of the list, with 4 proteins being identified in four or more OV-pos plasma samples. Proteins present in the same number of samples were then ranked by the number of unique peptides that supported their detection and then based on the number of PSMs that supported their detection in different plasma samples (Fig. 2). Additional functional annotation data for each of the proteins are provided in Table 2, and phylogenetic conservation data are shown in Fig. 3. Functional annotation data includes missense + nonsense single nucleotide polymorphism (SNP) rates for each protein across sequenced strains of O. volvulus, as identified in a previous study (46). Across all O. volvulus proteins, this average SNP rate was 5.06 SNPs/kb,
so our top 3 candidates were all below average in terms of strain-to-strain sequence variation, and none of our top 19 candidates showed significant variability between forest and savannah strains, or between West Africa and Ecuador strains of *O. volvulus* (46).

We identified four *O. volvulus* proteins in the OV-pos urine samples that were absent in the OV-neg urine sample. Urine results were used to corroborate plasma results, and urine testing was not part of the initial identification of the 19 high-priority candidates. Two of the proteins detected in urine were present in four or more OV-pos plasma samples (OVOC11613 and OVOC1523), and two were detected in three OV-pos plasma samples (OVOC247 and OVOC11626).

The top 6 proteins among the top 19 prioritized protein candidates (see bottom of Fig. 1) were the four supported by identification in the urine plus two additional proteins (OVOC9912 and OVOC4989), that were detected in four or more OV-pos plasma. These six proteins are described in greater detail below. It should be noted that one of the 19 prioritized biomarker candidates (OVOC9445) was also detected in an OV-pos urine sample. However, because one OV-neg urine sample contained one peptide from that protein (TFTDFNCLPVAAILDEK), the protein was ranked at the bottom of the list of 19 prioritized proteins. Below, additional details are provided for the top prioritized candidates, as well as other proteins of interest from the consistently detected and ranked list.

**Major Antigen / LFI1 (OVOC11613) is the biomarker candidate with the most supporting evidence**

OVOC11613 was identified as the top prioritized candidate because (i) peptides from it were detected in four OV-pos plasma samples with a total of six unique peptides including as many as four peptides in a single OV-pos sample (Fig. 2); (ii) two peptides supported its identification in an OV-pos urine sample: IEMMILEENKR (verified by heavy peptide spectral matching) and NADMKEDNDR (also detected in OV-pos plasma samples and both of which were not identically matched in other filarial nematode species; Fig. 4); (iii) it is nematode-specific,
sharing no significant homology with human protein, either by orthologous protein grouping or by BLAST (Fig. 3); (iv) It had moderate but consistent gene expression in individual adult *O. volvulus* females; Table 2); (v) four of the detected peptides were confirmed by the synthetic heavy peptide spectral matching and (vi) it has a very close ortholog in *Onchocerca lupi* (Ol-MJA, “Major antigen”, as annotated by UniProt (47), P21249) that was identified as a promising target for canine antibody serology in in a prior study (48). Additionally, the *O. ochengi* OVOC11613 ortholog was the third-most abundant protein detected in adult female worms in a previous mass spectrometry study (49). While one peptide (out of six) from this protein was detected in one OV-neg plasma sample, the peptide sequence was only supported by a single PSM. Since that peptide was not detected in any of the OV-pos samples, it did not eliminate the protein in the filtering process (since this single detection could be an artifact/false positive). In addition, OVOC11613 was detected by 219 unique peptides in the *O. volvulus* lysate sample, suggesting that it is a highly abundant protein in adult worms, and further supporting its adult *O. volvulus* origin in the plasma and urine.

OVOC11613 was called a "major antigen" based on its reactivity with antibodies in sera from humans with onchocerciasis (50). A subsequent study confirmed that the protein was expressed in adult female *O. volvulus* worms (51). In the recent *O. lupi* study (48), an immunoblotting / nLC-MS/MS proteomics approach was used to identify *O. lupi* proteins that were immunoreactive with antibodies in plasma from infected but not uninfected dogs. One peptide for OVOC11613 from that study (QQLSLALNERNTLQNDLR) was also identified here in three different OV-pos depleted plasma samples, and one OV-pos immune complex plasma sample. This *O. volvulus* peptide has at least 3 amino acid sequence differences relative to related peptides in other filarial parasite species that infect humans. These features highlight the potential for this peptide as a biomarker candidate and show that it warrants further evaluation.

The biological function of OVOC11613 / major antigen has not been described, and no primary sequence similarity was detected to any protein domain(s) or proteins in InterPro (39) or
KEGG (41). Its top BLAST match in *C. elegans* was to “LFI1, Lin-5 interacting protein” (ZC8.4), which is involved in spindle formation (52), but this had very low overall sequence similarity to OVOC11613 (62.5% dissimilarity over 80% of the protein length).

**OVOC1523 / ATP synthase alpha was the second-ranked overall candidate**

OVOC1523 (ATP synthase subunit alpha, mitochondrial homolog) was prioritized because (i) it was identified with two unique peptides in four OV-pos plasma samples (two depleted plasma samples and two immune complex preparations) (**Fig. 2**); (ii) it was identified in an OV-pos urine sample by the peptide GMALNLEPDNVVVFGNDK, which was also detected in all four plasma samples that contained this protein (**Fig. 5**). The presence of that peptide was confirmed by the synthetic heavy peptide spectral matching and (iii) it was more highly expressed than 99.67% of all genes in adult female *O. volvulus*. Although the *C. elegans* ortholog of OVOC1523 (ATP-1, H28O16.1a) functions as an ATP synthase in mitochondria, it also has an independent function responsible for iron acquisition in nematodes (53). An additional 24 unique peptides supported the identification of this protein in *O. volvulus* adult worm lysate. Although OVOC1523 is broadly conserved in nematodes and mammals (**Fig. 3**), it has 15% sequence dissimilarity to the closest human protein ortholog (ATP5F1A). Because some of the dissimilar sequences include portions of the molecule detected by MS in human samples, the protein has some potential as a biomarker for *O. volvulus* infection.

**OVOC247 and OVOC11626 were detected in both, the OV-pos plasma and urine samples**

OVOC247 / laminin alpha chain was ranked third overall because (i) it was detected by three peptides in three OV-pos plasma samples (one depleted plasma, two immune complex); (ii) it was identified in an OV-pos urine sample by the peptide QEYVEKNER (**Fig. 2; Fig. 6**); (iii) it has consistent and moderately high expression in adult female worms, (iv) it contains a signal peptide for secretion which increases the likelihood that viable adult worms might release the
protein into the host; (v) it was highly abundant in whole-worm lysate (with 90 detected peptides) and (vi) one peptide from OVOC247 was confirmed by synthetic heavy peptide spectral matching. Although it has an ortholog in human hosts, its human ortholog (LAMA1) only has 27.7% sequence similarity. Nematode laminins are glycoproteins that form a major polymer within basement membranes, and they are crucial for organizing extracellular matrix (54), which may explain why they would be identified in the plasma and urine.

OVOC11626 / PLK5 (inactive serine/threonine-protein kinase) was highly ranked because (i) it was identified by three unique peptides that were detected in four different OV-pos plasma samples (three depleted, 1 immune complex; Fig. 2); (ii) it was identified in an OV-pos urine sample by the peptide KENNIFQLSK, which was also detected in two of the depleted plasma samples and was supported by the synthetic heavy peptide spectral matching (Fig. 7); (iii) it was nematode-specific by the OrthoFinder classification, with only 31% similarity to any human protein (Fig. 3); (iv) it was predicted to be non-classically secreted and (v) it had consistent expression across all adult females in the RNA-seq dataset (Table 2). PLK5 has weak homology with C. elegans PLK2 (28% similarity) which is involved in crossover during meiosis (55), but the function of PLK5 in O. volvulus is unknown.

OVOC9912 and OVOC4989 were identified in more than three OV-pos plasma samples

OVOC9912 (PAN/apple domain) was the only protein other than the top prioritized candidate (OVOC11613) to be identified in five different OV-pos plasma samples (two depleted and three immune complex samples), and one of the detected peptides was confirmed by synthetic heavy peptide spectral matching. It is nematode-specific, with no orthology to any human protein, and has the lowest orthology of the top six proteins with other parasitic nematodes that infect humans (Fig. 3), making it an attractive biomarker candidate due to its O. volvulus-specificity. It contains a predicted signal peptide, suggesting that it is secreted by living worms, which might explain its relatively consistent detection in human samples despite its low gene
expression in RNA-seq datasets. OVOC9912 lacks detailed functional annotation, with the only functional evidence being a “PAN/Apple domain” (InterPro domain annotation, IPR003609), a domain which is commonly found in plasminogen proteins as well as in nematode proteins (56).

OVOC4989 (unc-22, twitchin homolog) is the last of the 19 candidates to be detected in more than three OV-pos plasma samples (3 depleted, 1 immune complex). Four unique peptides were detected in plasma samples. This protein was highly abundant in the worm lysate sample (180 peptides), and it has low conservation with human proteins. Nematode unc-22 is involved in myosin functional regulation in muscle tissue (57). It is an extremely large protein (6,785 amino acids in the current OVOC4989 protein model), so its presence in the blood may be due to its high relative abundance in the *O. volvulus* muscle tissue.

**Additional candidates of interest**

OVOC10067 (Basement membrane proteoglycan) was identified here by eight unique peptides across three OV-pos plasma samples (one depleted, two immune complex), and it was identified as one of 33 serodiagnostic candidate proteins in our prior proteomics study of *O. volvulus* proteins that were bound by antibodies present in plasma from onchocerciasis patients (12). One of its supporting peptides in human plasma was confirmed by the synthetic heavy peptide spectral matching, and this protein also contains a signal peptide sequence. Additionally, the *O. ochengi* OVOC10067 ortholog was the second-most abundant protein detected in adult female worms, and was highly abundant in *O. ochengi* nodule fluid in a previous mass spectrometry study (49). Thus, this candidate, although not detected in our pilot studies with urine, is an attractive target for an adult female-specific biomarker.

OVOC2745 (Actin-4) was supported by four peptides across 3 OV-pos plasma samples (one confirmed by synthetic heavy peptide spectral matching), and it was also identified as an antibody target in the canine *O. lupi* study (48). It had high and consistent gene expression in adult female *O. volvulus*. However, this protein has high homology (97.3%) with the human protein
ACTG1 and its function as a structural protein reduces its potential as biomarker candidate for detection by immunoassay.

Overall, all of the proteins shown in Fig. 2 are supported by the proteomics evidence, but some candidates such as OVOC7004, OVOC3364 and OVOC8003 have lower RNA-seq expression and are inconsistently expressed across adult female worm RNA-seq samples (Table 2). All of the protein properties presented in Figs. 2 and 3, Table 2 and supplemental Table S4 should be considered in future evaluations of these biomarker candidates.

Conclusions
The direct proteomics discovery approach utilized here facilitated the discovery of a short list of prioritized potential novel biomarkers with a relatively high rate of true positivity, with some candidates detected in multiple samples from persons with onchocerciasis. This is the first report of regular and deep scale MS proteomics being used to directly screen human plasma and urine samples for parasite proteins that might serve as biomarkers for active O. volvulus infection. Other studies have focused on identifying O. volvulus proteins for use as targets for antibody assays (e.g. (6, 7)) or on the prioritization of identified potential biomarkers based on an analysis of gene expression profiles or other bioinformatic data (e.g. (12)). However, this type of evidence only indicates detection of potential circulating antigens and does not quantify direct detection in the blood or urine, as performed here. Direct detection has been identified for O. volvulus-originated metabolites (e.g. (14) (17)) and phospholipids (18), but the consistent detection of small amounts of metabolites and phospholipids among all the metabolites found in human blood and urine is very difficult to achieve, and they may also be conserved among other endemic helminths. Additional validation of the potential biomarkers presented here is provided by their absence of detection in uninfected samples, and the manual validation of spiked-in heavy peptides to confirm the proteomic detection of key peptides.
There have been *Onchocerca* biomarkers proposed in the past which have later failed validation, and likewise, additional work will be needed to assess the sensitivity and specificity of the biomarker candidates identified in this study. This will require the development of sensitive assays and the testing of large panels of positive and negative samples. The lack of available genome data for the co-endemic filarial nematode *Mansonella* spp. necessitates future screening to ensure that biomarker peptide sequences are not conserved with this species, ideally by testing individuals infected only with *Mansonella*. Likewise, the testing of *O. volvulus*-infected individuals treated with ivermectin who are negative for the skin snip test would support the validation of the biomarkers as being derived from the adult females and not being from microfilariae. It is possible that multiplex assays for multiple biomarkers will be required to achieve high sensitivity, and the top 19 candidates are presented here in order to provide the maximum potential for a functional test, with the assumption that many will not work in a final biomarker test. Overall, this study has identified a promising novel set of proteins that will be carried forward to develop assays that can be used for diagnosis of *O. volvulus* infections and for assessing the efficacy of new treatments with samples from clinical trials. Such a test would be invaluable for screening individuals who currently test negative in traditional skin snip tests, but who still harbor live adult *O. volvulus* worms, necessitating further treatments in the area which currently may not be administered.

**Data availability**

The MS proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the iProX partner repository with the data set identifier PXD033659. Proteomics sample metadata and accession IDs per sample are provided in *supplemental Table S1*. Machine setting parameters for proteomics runs including QE, EASYnLC, timsTOF and Nanoelute, as well as MaxQuant configuration parameters are provided in *supplemental Table S2*. Peptide-spectral match (PSM) counts for each peptide across each
sample, including precursor charge, mass/charge, modifications observed and scores are provided in supplemental Table S3. Detected peptide counts per sample, complete functional annotation data, orthology information, and gene expression data, for all O. volvulus proteins are provided in supplemental Table S4. Output from the isotope-labeled standard peptide (heavy peptide) fragment ion spectral matching analysis for all confirmed proteins is provided in supplemental Table S5.

Author contributions
Bruce A. Rosa: Data curation, Investigation, Formal analysis, Methodology, Visualization, Writing – original draft preparation. Kurt Curtis: Investigation, Methodology. Petra Erdmann Gilmore: Data curation, Investigation. John Martin: Data curation, Methodology. Qiang Zhang: Methodology. Robert Sprung: Data curation, Methodology. Gary J. Weil: Conceptualization, Writing - review & editing, Project administration. R. Reid Townsend: Investigation, Methodology, Resources, Writing - review & editing, Supervision. Peter U. Fischer: Conceptualization, Writing - review & editing, Funding acquisition, Project administration, Supervision. Makedonka Mitreva: Conceptualization, Writing - review & editing, Funding acquisition, Project administration, Supervision.

Acknowledgements
This research was funded by the Bill and Melinda Gates foundation and the Drugs for Neglected Diseases Initiative, Grant No. INV-021433 awarded to Peter U. Fischer and Makedonka Mitreva (co-PIs) and OPP1083853 awarded to Gary Weil. The expert technical assistance of Dr. Yiling Mi, Alan Davis and Rose Connors is gratefully acknowledged. The proteomic experiments were performed at the Washington University Proteomics Shared Resource (WU-PSR), R Reid Townsend MD.PhD., Director. The WU-PSR is supported in part by the WU Institute of Clinical
and Translational Sciences (NCATS UL1 TR000448), the Mass Spectrometry Research Resource (NIGMS P41 GM103422; R24GM136766) and the Siteman Comprehensive Cancer Center Support Grant (NCI P30 CA091842).

Funding sources had no involvement in the collection, analysis and interpretation of data, in the writing of the report, or in the decision to submit the article for publication.

**Conflicts of interest:** none
References


Direct proteomic detection of 19 onchocerciasis biomarkers


Figure captions

**Fig. 1**: Overview of the step-wise filtering and ranking process for the identification of potential *O. volvulus* biomarkers among human plasma proteomics samples.

**Fig. 2**: A summary of results for the 19 high-priority protein targets identified according to the filtering criteria shown in Fig. 1. Green numbers represent the number of detected peptides in each sample, blue numbers represent the number of unique peptides from the protein that were identified in *O. volvulus*-positive plasma samples, and orange numbers represent the sum of peptide:spectral matches (PSMs) across all peptides in all *O. volvulus*-positive plasma samples. Proteins are annotated by assigned protein IDs, KEGG enzyme annotations, and InterPro functional domains. I = plasma / urine from individual person, P = Pooled plasma from multiple individuals. Dep = Depleted plasma, IP = immunoprecipitation from plasma samples, IC = immune complex from plasma samples.

**Fig. 3**: Conservation of the 19 prioritized proteins based on Orthofinder results and BLAST amino acid sequence similarity of *O. volvulus* proteins to orthologous proteins in other relevant species.

**Fig. 4**: Multiple protein alignment (M-Coffee) for Major Antigen (OVOC11613) and its most significant BLAST hits in *L. loa* (LLOA), *W. bancrofti* (WBAN), *B. malayi* (BMAL), *N. americanus* (NAME) and *A. ceylanicum* (ACEY). Shades of blue indicate level of amino acid conservation. Peptides highlighted in red were detected in plasma samples from *O. volvulus* infected individuals, peptides highlighted in black were detected in plasma samples and verified by the heavy peptide spectral matching, peptides highlighted in green were detected in urine samples from *O. volvulus* infected individuals, and the peptide highlighted in orange was detected in urine and verified by the heavy peptide spectral matching. The sequence marked with a black bracket was a peptide
also identified for Major Antigen protein in *Onchocerca lupi* (Latrofa et al, 2021). *Numbers shown indicate that a peptide was detected in multiple samples, red = plasma, green = urine.

**Fig. 5:** Multiple protein alignment (M-Coffee) for ATP synthase subunit alpha (OVOC1523) and its most significant BLAST hits in *L. loa* (LLOA), *W. bancrofti* (WBAN), *B. malayi* (BMAL), *N. americanus* (NAME), *A. ceylanicum* (ACEY) and human. Shades of blue indicate level of amino acid conservation. Peptides highlighted in red were detected in plasma samples from *O. volvulus* infected individuals, peptides highlighted in black were detected in plasma samples and verified by the heavy peptide spectral matching, and peptides highlighted in green were detected in urine samples from *O. volvulus* infected individuals. *Numbers shown indicate that a peptide was detected in multiple samples, red = plasma, green = urine.

**Fig. 6:** Multiple protein alignment (M-Coffee) for laminin alpha chain (OVOC247) and its most significant BLAST hits in *L. loa* (LLOA), *W. bancrofti* (WBAN), *B. malayi* (BMAL), *N. americanus* (NAME), *A. ceylanicum* (ACEY) and human. Shades of blue indicate level of amino acid conservation. Peptides highlighted in red were detected in plasma samples from *O. volvulus* infected individuals, peptides highlighted in black were detected in plasma samples and verified by the heavy peptide spectral matching, and peptides highlighted in green were detected in urine samples from *O. volvulus* infected individuals. *Numbers shown indicate that a peptide was detected in multiple samples, red = plasma, green = urine.

**Fig. 7:** Multiple protein alignment (M-Coffee) for PLK-5 (OVOC11626) and its most significant BLAST hits in *L. loa* (LLOA), *W. bancrofti* (WBAN), *B. malayi* (BMAL), *N. americanus* (NAME), *A. ceylanicum* (ACEY) and human. Shades of blue indicate level of amino acid conservation. Peptides highlighted in red were detected in plasma samples from *O. volvulus* infected individuals,
and peptides highlighted in green were detected in urine samples from *O. volvulus* infected individuals. *Numbers shown indicate that a peptide was detected in multiple samples.

**Supplemental Information**

This article contains supplemental data.

**Supplemental Table S1**: Proteomics sample metadata and accessions.

**Supplemental Table S2**: Machine setting parameters for proteomics runs including (A) QE, (B) EASYnLC, (C) timsTOF and (D) Nanoelute, and MaxQuant configuration parameters (E).

**Supplemental Table S3**: Complete peptide and spectral data. (A) Peptide counts per protein per sample. (B) Spectral data per detected peptide, including precursor charge and mass/charge, modifications observed and scores.

**Supplemental Table S4**: *O. volvulus* protein data summary, including peptide counts per sample, complete functional annotation data, orthology information, and gene expression data, for all *O. volvulus* proteins.

**Supplemental Table S5**: Isotope-labeled standard peptide (heavy peptide) fragment ion spectral matching data for all confirmed proteins.
Table 1: Overview of mass spectrometry proteomics sample sets used as part of the candidate biomarker filtering and ranking process. “OV-pos” = sample from individual(s) infected with *O. volvulus*; “OV-neg” = sample from individual(s) not infected with *O. volvulus*. “Ind.” = sample from single individual, “Pool” = sample from a pool of individuals.

<table>
<thead>
<tr>
<th>Sample group</th>
<th>Plasma treatment</th>
<th>Pooled / Individual</th>
<th>Country</th>
<th>Number of PSMs</th>
<th>Number of peptides</th>
<th>Number of proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>OV-pos plasma</td>
<td>Depleted (Dep)</td>
<td>Pool</td>
<td>Uganda</td>
<td>220</td>
<td>121</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pool</td>
<td>Uganda</td>
<td>439</td>
<td>190</td>
<td>181</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pool</td>
<td>Uganda</td>
<td>565</td>
<td>252</td>
<td>241</td>
</tr>
<tr>
<td></td>
<td>Immunoprecipitated (IP)</td>
<td>Pool</td>
<td>Uganda</td>
<td>370</td>
<td>211</td>
<td>203</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ind.</td>
<td>Uganda</td>
<td>495</td>
<td>283</td>
<td>235</td>
</tr>
<tr>
<td>Immune complex (IC)</td>
<td></td>
<td>Pool</td>
<td>Uganda</td>
<td>337</td>
<td>97</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pool</td>
<td>Uganda</td>
<td>295</td>
<td>134</td>
<td>133</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ind.</td>
<td>Cameroon</td>
<td>199</td>
<td>130</td>
<td>129</td>
</tr>
<tr>
<td>OV-pos urine</td>
<td>-</td>
<td>Ind.</td>
<td>Cameroon</td>
<td>266</td>
<td>163</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>Ind.</td>
<td>Cameroon</td>
<td>104</td>
<td>75</td>
<td>73</td>
</tr>
<tr>
<td>OV-neg plasma</td>
<td>Depleted (Dep)</td>
<td>Pool</td>
<td>India</td>
<td>181</td>
<td>103</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>Immunoprecipitated (IP)</td>
<td>Pool</td>
<td>Uganda</td>
<td>186</td>
<td>115</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pool</td>
<td>India</td>
<td>172</td>
<td>98</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ind.</td>
<td>India</td>
<td>91</td>
<td>72</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Immune complex (IC)</td>
<td>Ind.</td>
<td>India</td>
<td>65</td>
<td>55</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ind.</td>
<td>India</td>
<td>36</td>
<td>31</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pool</td>
<td>Indonesia</td>
<td>242</td>
<td>148</td>
<td>144</td>
</tr>
<tr>
<td>OV-neg Urine</td>
<td>-</td>
<td>Ind.</td>
<td>Cameroon</td>
<td>90</td>
<td>62</td>
<td>59</td>
</tr>
<tr>
<td>Worm lysate</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>70,226</td>
<td>21,109</td>
<td>2,405</td>
</tr>
</tbody>
</table>
Table 2: Additional annotation and gene expression data for 19 prioritized biomarker candidate proteins. Included are the protein lengths (amino acids), predicted secretion, gene expression in adult females (RNA-seq), gene expression consistency among adult females, and the rate of missense + nonsense single nucleotide polymorphisms (SNPs) per kilobase of protein sequence, based on results from Choi et al, 2016 (46).

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Protein functional annotation</th>
<th>Protein length (AA)</th>
<th>Predicted secretion</th>
<th>Gene exp. level (Avg. adult female FPKM)</th>
<th>Gene expression consistency (% of adult females)</th>
<th>Missense + nonsense SNP rate (Choi et al, 2016)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OVOC11613</td>
<td>Major antigen / Ili-1 - Lin-5 (Five) Interacting</td>
<td>2021</td>
<td>-</td>
<td>17.0</td>
<td>100%</td>
<td>3.6</td>
</tr>
<tr>
<td>OVOC1523</td>
<td>ATP synthase subunit alpha, mitochondrial</td>
<td>535</td>
<td>-</td>
<td>1963.9</td>
<td>100%</td>
<td>1.2</td>
</tr>
<tr>
<td>OVOC247</td>
<td>Ovo-lam-3 - Laminin alpha chain</td>
<td>3358</td>
<td>Signal peptide</td>
<td>24.9</td>
<td>100%</td>
<td>2.7</td>
</tr>
<tr>
<td>OVOC11626</td>
<td>Inactive serine/threonine-protein kinase PLK5</td>
<td>268</td>
<td>Nonclassical</td>
<td>2.0</td>
<td>100%</td>
<td>11.2</td>
</tr>
<tr>
<td>OVOC9912</td>
<td>IPR003609:PAN/Apple domain</td>
<td>628</td>
<td>Signal peptide</td>
<td>0.3</td>
<td>87.5%</td>
<td>15.4</td>
</tr>
<tr>
<td>OVOC4989</td>
<td>Ovo-unc-22 - Twitchin homolog</td>
<td>6785</td>
<td>-</td>
<td>28.9</td>
<td>100%</td>
<td>2.3</td>
</tr>
<tr>
<td>OVOC10067</td>
<td>Basement membrane proteoglycan</td>
<td>3415</td>
<td>Signal peptide</td>
<td>0.9</td>
<td>100%</td>
<td>7.1</td>
</tr>
<tr>
<td>OVOC2745</td>
<td>Actin-4</td>
<td>376</td>
<td>-</td>
<td>359.1</td>
<td>100%</td>
<td>0.9</td>
</tr>
<tr>
<td>OVOC7004</td>
<td>Putative cubilin</td>
<td>4112</td>
<td>Signal peptide</td>
<td>0.0</td>
<td>12.5%</td>
<td>22.0</td>
</tr>
<tr>
<td>OVOC7893</td>
<td>H2B: histone H2B</td>
<td>380</td>
<td>-</td>
<td>72.2</td>
<td>100%</td>
<td>2.6</td>
</tr>
<tr>
<td>OVOC10971</td>
<td>Ovo-phb-2 - Mitochondrial prohibitin complex protein 2</td>
<td>568</td>
<td>-</td>
<td>275.7</td>
<td>100%</td>
<td>4.1</td>
</tr>
<tr>
<td>OVOC1599</td>
<td>Ovo-adpr-1 - WDTC1, WD and tetratricopeptide repeats</td>
<td>597</td>
<td>-</td>
<td>21.8</td>
<td>100%</td>
<td>0.6</td>
</tr>
<tr>
<td>OVOC7471</td>
<td>Ovo-magu-2 - MAGUK p55 subfamily member 5</td>
<td>861</td>
<td>-</td>
<td>0.6</td>
<td>100%</td>
<td>8.1</td>
</tr>
<tr>
<td>OVOC216</td>
<td>still life protein homolog</td>
<td>987</td>
<td>-</td>
<td>12.8</td>
<td>100%</td>
<td>1.7</td>
</tr>
<tr>
<td>OVOC1112</td>
<td>Ovo-smg-1 - Serine/threonine-protein kinase smg-1</td>
<td>2527</td>
<td>-</td>
<td>13.8</td>
<td>100%</td>
<td>3.8</td>
</tr>
<tr>
<td>OVOC3364</td>
<td>(Filarial-specific)</td>
<td>166</td>
<td>Signal peptide</td>
<td>0.0</td>
<td>38%</td>
<td>12.5</td>
</tr>
<tr>
<td>OVOC874</td>
<td>(Nematode-specific)</td>
<td>239</td>
<td>-</td>
<td>106.1</td>
<td>100.0%</td>
<td>0.0</td>
</tr>
<tr>
<td>OVOC8003</td>
<td>Ovo-lgc-50 - Ligand-gated ion channel 50 homolog</td>
<td>221</td>
<td>-</td>
<td>0.1</td>
<td>63%</td>
<td>8.7</td>
</tr>
<tr>
<td>OVOC9445</td>
<td>PPP1C: serine/threonine phosphatase PP1 catalytic</td>
<td>671</td>
<td>-</td>
<td>9.8</td>
<td>100.0%</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Number of *O. volvulus* proteins

- 12,224: All annotated *O. volvulus* proteins
- 989: Detected in any OV-pos sample
- 803: Not detected in OV-neg samples*
- 90: ≥ 2 peptides across OV-pos samples
- 19: Detected in ≥ 3 OV-pos samples

*Ranking criteria:
1. # positive samples
2. # unique peptides
3. # PSMs

Detected in OV-pos urine:
- OVOC247
- OVOC11626
- OVOC11613
- OVOC1523
- OVOC9912
- OVOC4989

Detected in ≥ 4 OV-pos plasma samples

*Allowing 1 peptide in 1 negative sample, as long as it is not also in any positive samples
<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Protein functional annotation</th>
<th>Number of peptides detected</th>
<th>O. vulvulus-infected individuals (OV-pos)</th>
<th>Uninfected individuals (OV-neg)</th>
<th>Whole-body host</th>
<th>Number of unique peptides across OV-pos plasma samples</th>
<th>Number of PSVs across OV-pos plasma samples</th>
<th>Number of verified heavy peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td>OVOC11613</td>
<td>Major antigen / IF-1 - Lh-5 (Five) Interacting</td>
<td>1 1 2 1</td>
<td>1 1</td>
<td>0 0 0 2</td>
<td>0 0 0 1</td>
<td>0 0</td>
<td>219 6 9 4</td>
<td></td>
</tr>
<tr>
<td>OVO21523</td>
<td>ATP synthase subunit alpha, mitochondrial</td>
<td>0 1 2 0</td>
<td>0 0</td>
<td>0 1</td>
<td>0 0</td>
<td>0 0 0 1</td>
<td>24 2 17 1</td>
<td></td>
</tr>
<tr>
<td>OVO2477</td>
<td>Ovo-lam-3 - Lamin alpha chain</td>
<td>1 1 1 1</td>
<td>1 1</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>20 3 4 1</td>
<td></td>
</tr>
<tr>
<td>OVO21626</td>
<td>Inactive serine/threonine-protein kinase PLK5</td>
<td>1 0 0 0</td>
<td>1 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>2 6 1</td>
<td></td>
</tr>
<tr>
<td>OVO25912</td>
<td>IPR003609 PAN/Apil domain</td>
<td>1 2 2 1</td>
<td>1 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>2 12 1</td>
<td></td>
</tr>
<tr>
<td>OVO24689</td>
<td>Ovo-unc-22 - Twinkle homolog</td>
<td>0 1 1 1</td>
<td>1 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>180 0 6 2</td>
<td></td>
</tr>
<tr>
<td>OVO19097</td>
<td>Basement membrane proteoglycan</td>
<td>2 1 3 1</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>3 6 1</td>
<td></td>
</tr>
<tr>
<td>OVO27455</td>
<td>Acat-4</td>
<td>0 1 3 1</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>2 6 1</td>
<td></td>
</tr>
<tr>
<td>OVO27004</td>
<td>Putative cubulin</td>
<td>1 1 1 1</td>
<td>1 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>16 2 5 1</td>
<td></td>
</tr>
<tr>
<td>OVO27892</td>
<td>H2B: histone H2B</td>
<td>0 0 0 0</td>
<td>1 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>17 3 3 2</td>
<td></td>
</tr>
<tr>
<td>OVO10971</td>
<td>Ovo-phb-2 - Mitochondrial probolin complex protein 2</td>
<td>0 0 0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>17 3 3 2</td>
<td></td>
</tr>
<tr>
<td>OVO15999</td>
<td>Ovo-adip-1 - WDTC1, WD and tetra-tripeptide repeats</td>
<td>0 0 0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>17 3 3 2</td>
<td></td>
</tr>
<tr>
<td>OVO17471</td>
<td>Ovo-mp5-2 - MAGUK, p55 suinlarity membcr 5</td>
<td>0 1 1 1</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>17 3 3 2</td>
<td></td>
</tr>
<tr>
<td>OVO2216</td>
<td>still life protein homolog</td>
<td>0 0 0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>17 3 3 2</td>
<td></td>
</tr>
<tr>
<td>OVO11102</td>
<td>Ovo-smg-1 - Serine/threonine-protein kinase smg-1</td>
<td>0 0 0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>17 3 3 2</td>
<td></td>
</tr>
<tr>
<td>OVO23664</td>
<td>[Inter-related-specific]</td>
<td>0 0 0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>17 3 3 2</td>
<td></td>
</tr>
<tr>
<td>OVO28874</td>
<td>[Nematode-specific]</td>
<td>0 0 0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>17 3 3 2</td>
<td></td>
</tr>
<tr>
<td>OVO26503</td>
<td>Ovo-lgc-50 - Ligand-gated ion channel 50 homolog</td>
<td>0 0 0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>17 3 3 2</td>
<td></td>
</tr>
<tr>
<td>OVO20445</td>
<td>PRPT1C: serine/threonine phosphatase PP1 catalytic</td>
<td>0 0 0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>17 3 3 2</td>
<td></td>
</tr>
<tr>
<td>Protein ID</td>
<td>Protein functional annotation</td>
<td>Overall conservation of protein across species</td>
<td>I</td>
<td>Counts of orthology in OPF per species</td>
<td>Nematode clade</td>
<td>Outgroup</td>
<td>BLAST % similarity to top hit (only if E &lt; 10^-2)</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-------------------------------</td>
<td>-----------------------------------------------</td>
<td>---</td>
<td>--------------------------------------</td>
<td>----------------</td>
<td>----------</td>
<td>-----------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>OVC11613</td>
<td>Major antigen / Il-1 - Lin-5 (Five) Interacting</td>
<td>Nematode-specific</td>
<td>0</td>
<td>2 3 3 2 2 33 0 1 1 1 0 0 0</td>
<td>71 64 70 40 37 45</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVC1523</td>
<td>ATP synthase subunit alpha, mitochondrial</td>
<td>Conserved in hosts</td>
<td>3</td>
<td>1 1 1 1 1 1 1 1 1 1 1 2 1</td>
<td>97 96 98 91 90 94 85</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVC247</td>
<td>Ovo-lam-3 - Lamin alpha chain</td>
<td>Conserved in hosts</td>
<td>1</td>
<td>1 1 1 1 1 1 2 1 1 1 1 2 2 1</td>
<td>84 81 84 61 41 65 28</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVC11626</td>
<td>Inactive serine/threonine-protein kinase PLK5</td>
<td>Nematode-specific</td>
<td>0</td>
<td>2 3 1 1 1 2 0 1 1 0 0 0</td>
<td>80 77 75 38 28 27 31</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVC5992</td>
<td>iPR003065.PAN/Apple domain</td>
<td>Conserved in hosts</td>
<td>1</td>
<td>1 1 1 1 1 1 4 1 4 1 2 1 3 2 1</td>
<td>84 83 82 68 49 61 28</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVC4989</td>
<td>Ovo-unc-22 - Twitchin homolog</td>
<td>Conserved in hosts</td>
<td>1</td>
<td>1 1 1 1 1 1 4 1 4 1 2 1 3 2 1</td>
<td>84 76 80 44 70 45 27</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVC10067</td>
<td>Basement membrane proteoglycan</td>
<td>Conserved in hosts</td>
<td>3</td>
<td>1 1 5 2 1 1 1 4 1 4 4 1 2 3 2 2</td>
<td>84 76 80 44 70 45 27</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVC2745</td>
<td>Actin-4</td>
<td>Conserved in hosts</td>
<td>3</td>
<td>4 3 4 3 6 19 4 0 7 6 17 14 8 96 7 97 97 69 99 97</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVC7004</td>
<td>Putative cubulin</td>
<td>Conserved in hosts</td>
<td>1</td>
<td>4 10 2 1 1 6 1 0 6 1 1 1 2</td>
<td>84 82 84 61 44 37 23</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVC7893</td>
<td>H2B histone H2B</td>
<td>O. volvulus-specific</td>
<td>0</td>
<td>1 0 0 0 0 0 0 0 0 0 0 0</td>
<td>98 86 90 93 86 90 79</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVC10671</td>
<td>Ovo-phb-2 - Mitochondrial prohibitin complex protein 2</td>
<td>Conserved in hosts</td>
<td>1</td>
<td>1 1 1 1 1 1 1 1 1 1 1 1 1 1</td>
<td>84 96 97 72 62 81 66</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVC15999</td>
<td>Ovo-adpr-1 - WDTC1, WD and tetratricopeptide repeats</td>
<td>Conserved in hosts</td>
<td>1</td>
<td>2 1 1 1 1 1 0 1 1 1 1 1 1 1 2</td>
<td>92 93 91 72 39 32 32</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVC7471</td>
<td>Ovo-mag-2 - MAGUK p55 subfamily member 5</td>
<td>Conserved in hosts</td>
<td>1</td>
<td>2 3 3 2 2 2 2 2 2 2 2 6 6 2</td>
<td>75 74 72 52 50 36 47</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVC216</td>
<td>still life protein homolog</td>
<td>Conserved in hosts</td>
<td>0</td>
<td>1 4 1 1 2 2 1 2 1 1 2 1</td>
<td>82 82 62 64 41 37 27</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVC1112</td>
<td>Ovo-smg-1 - Serine/threonine-protein kinase smg-1</td>
<td>Conserved in hosts</td>
<td>1</td>
<td>1 1 1 1 1 1 5 1 4 1 2 1 2 1 1</td>
<td>84 84 84 57 48 35 27</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVC3364</td>
<td>(Filarial-specific)</td>
<td>Filarial-specific</td>
<td>0</td>
<td>1 0 0 1 1 0 0 0 0 0</td>
<td>35 41</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>OVC674</td>
<td>(Nematode-specific)</td>
<td>Nematode-specific</td>
<td>1</td>
<td>1 1 1 1 1 1 1 1 2 0 0 0 0 0</td>
<td>78 71 73 32</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>OVC8003</td>
<td>Ovo-lgc-50 - Ligand-gated ion channel 50 homolog</td>
<td>Nematode-specific</td>
<td>0</td>
<td>1 1 1 1 1 1 0 1 1 0 0 0</td>
<td>78 82 72 69 49 59 38</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVC9445</td>
<td>PPP1C - serine/threonine phosphatase PPI catalytic</td>
<td>Conserved in hosts</td>
<td>9</td>
<td>11 12 9 11 13 14 12 11 9 18 3 3 10 99 99 99 99 99 99 99</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Journal Pre-proof</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
• MS proteomics was performed directly on plasma and urine samples
• Matched negative controls were used to reduce false positives
• 19 novel *O. volvulus* biomarker candidates were identified and prioritized
• Candidates were validated by isotope-labeled heavy peptides
*Onchocerca volvulus* infects over 20 million people, causing severe disability. Diagnostic tests cannot reliably detect live adult worms, which is critical for disease elimination efforts. Plasma and urine samples from *O. volvulus*-infected individuals (and uninfected controls) were analyzed by MS proteomics, to directly identify *O. volvulus* proteins in hosts. A total of 19 proteins were prioritized. This set of biomarker candidates is a valuable resource to further explore for the diagnosis of active *O. volvulus* infections.
Author contributions

Bruce A. Rosa: Data curation, Investigation, Formal analysis, Methodology, Visualization, Writing – original draft preparation. Kurt Curtis: Investigation, Methodology. Petra Erdmann Gilmore: Data curation, Investigation. John Martin: Data curation, Methodology. Qiang Zhang: Methodology. Robert Sprung: Data curation, Methodology. Gary J. Weil: Conceptualization, Writing - review & editing, Project administration. R. Reid Townsend: Investigation, Methodology, Resources, Writing - review & editing, Supervision. Peter U. Fischer: Conceptualization, Writing - review & editing, Funding acquisition, Project administration, Supervision. Makedonka Mitreva: Conceptualization, Writing - review & editing, Funding acquisition, Project administration, Supervision.