Chapter for 'Methods in *Toxoplasma* Research*'*

'Image-based Quantitation of Host Cell-*Toxoplasma gondii* Interplay using HRMAn; a Host Response to Microbe Analysis Pipeline'

Daniel Fisch[1], Artur Yakimovich[2], Barbara Clough[1], Jason Mercer[2], Eva-Maria Frickel[1]

[1]Host-*Toxoplasma* Interaction Laboratory, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK

[2]MRC-Laboratory for Molecular Cell Biology, University College London, Gower Street, London, WC1E 6BT, UK

To whom correspondence should be addressed:

Eva-Maria Frickel

Host-*Toxoplasma* Interactions Laboratory

The Francis Crick Institute

1 Midland Road, London NW1 1AT, UK

Tel: **+44 203 796 1438**

eva.frickel@crick.ac.uk

Abstract

Research on *Toxoplasma gondii* and its interplay with the host is often performed using fluorescence microscopy-based imaging experiments combined with manual quantification of acquired images. We present here an accurate and unbiased quantification method for host-pathogen interactions. We describe how to plan experiments and prepare, stain and image infected specimens and analyse them with the program HRMAn (Host Response to Microbe Analysis). HRMAn is a high-content image analysis method based on KNIME Analytics Platform. Users of this guide will be able to perform infection studies in high-throughput volume and to a greater level of detail. Relying on cutting edge machine learning algorithms, HRMAn can be trained and tailored to many experimental settings and questions.

Keywords

1. Introduction

Toxoplasma gondii (*Tg*) is an obligate intracellular, eukaryotic parasite that can infect any nucleated cell of a warm-blooded animal [1]. Belonging to the family of apicomplexan parasites, it has adapted to an intricate life cycle centred on its feline definitive host [2]. *Tg* commonly infects humans through food contamination or contact with cats, albeit infection does not endanger healthy individuals, and will persist latently for the lifetime of the infected human host (Global seroprevalence estimated around 30%) [3]. However, infection in immunocompromised humans or primary infection during pregnancy, can cause severe disease called *toxoplasmosis* or cause congenital birth defects [4, 5].

Tg parasites can take different forms during the lifecycle, with tachyzoites being the highly replicative and infectious form present during primary infection. Tachyzoites can be propagated easily and indefinitely in tissue culture, making them perfect for studying infection processes with *Tg in vitro*. To do so, infection experiments are performed by infecting host cells with freshly prepared tachyzoites. During infection, *Tg* rapidly invades host cells and establishes its own replicative niche within the cytosol, called the parasitophorous vacuole (PV) [6, 7]. *Tg* injects a plethora of effector proteins upon invasion and exports further proteins once residing within the PV in its new host cell, modulating the cell's response by, for example, suppressing cell-intrinsic immune responses or by blocking certain types of cell death [8, 9]. Within the PV *Tg* continues to grow and replicate and duplicates every 6 to 10 hours [10]. Once *Tg* has replicated to a sufficient number or if the host cell cannot sustain more parasites, *Tg* egresses from the host cell, thereby rupturing the PV and the cell, and spreads to infect neighbouring cells.

The PV is a highly modified compartment that was thought to be nonfusogenic with lysosomes and comprises a safe haven in which the parasite can grow [11–13]. However, more recently, it has become clear that PVs can be recognized by host factors within immune-stimulated murine [14–16] and human cells [17–19]. Consequently, this triggers cell-intrinsic immune responses, either slowing the growth of or killing the parasites.

To study infection processes, parasites or host cells are often manipulated, and infection is simply a readout for the effect of the manipulation. These can include, genetic modification (deletion of certain genes or overexpression), silencing of genes with RNAi or blocking specific enzymes with inhibitors. Whether the focus of the experiment is on the host cell or on the parasite itself will dictate the readout of the experiment.

As infection is a highly dynamic process, so is the readout chosen for quantification. Researchers often tend to score certain parameters by manual enumeration of immunofluorescence images. These parameters include the number of cells and parasites, replication of the parasite or deposition of host/parasite protein on the PV. However, manual counting limits the number of quantified parameters and valuable information may be lost, or interesting phenotypes missed. Furthermore, the process of manual counting is flawed, as human error is likely to be introduced. Thus, we identified the necessity to automate the process of image-based infection quantification.

We recently published an automated image analysis pipeline called HRMAn for quantification of host response to *Tg* infection [20]. As described above, it is imperative for infection biology to be able to quantify infection dynamics reliably and consistently. Here, we describe how to prepare samples for microscopy that give reproducible

results and provide details on how to image them. We also briefly describe how to use the HRMAn analysis pipeline, but we want to note that for a more detailed instruction, we recommend visiting the homepage (https://hrman.org/) and use the provided tutorials. To learn more about the information that is extracted from the immunofluorescence images of infected cells, we provide an explanation of their meaning and how they can be interpreted from the host side "Host response" or the parasite side "Parasite Biology" (Table 1).

To compute the readouts, HRMAn uses well-established image analysis and segmentation algorithms to detect cells and pathogens and combines this information with classical machine learning and cutting-edge artificial intelligence algorithms in the form of deep convolutional neural networks. This is wrapped into a user-friendly environment based on KNIME Analytics platform [21]. KNIME provides an intuitive graphical user interface and a modular architecture that will allow more experienced users to adapt and modify the analysis pipeline and tailor it to their needs. HRMAn can be used without any coding experience, but we recommend watching the tutorial videos and gaining some hands-on experience by using the example datasets provided on the homepage and on Dryad (See 2.6. Online resources). Once comfortable with the analysis pipeline, users are enabled to perform infection experiment on a different scale, with virtually no limitation to the number of samples or replicates within each experiment.

2. Materials

1. General

- Micropipettes

- Pipette controller

- 15 mL and 50 mL conical tubes

- Microcentrifuge tubes

- Serological pipettes

- ddH$_2$O

- Phosphate buffered saline (PBS), sterile filtered, suitable for cell culture

2. Cell/ parasite culture and infection

- Tissue culture plates: 24-well or black-wall 96-well plates (Optional with coverslips glass-bottom e.g. MACS Miltenyi, Imaging Plate CG 1.0 (96 well), # 130-098-264)

- Cell culture bench-top centrifuge with adaptors for multi-well plates.

- Appropriate cell culture medium with supplements for the used cell line

- Optional: Trypsin/EDTA, for dislodging adherent cells

- Human foreskin fibroblasts (HFFs), grown confluent in T25 flasks for propagation of *Tg* tachyzoites

- DMEM with added 10% FCS, for propagation of HFFs and *Tg*-infected HFFs

- Cell scrapers, 25 G needles and 10 mL syringes, for syringe-lysing *Tg* cultures

- Trypan blue and haemocytometer (e.g. Immune Systems, Fastread disposable counting slides #BVS100), for counting cells and parasite numbers

- Gelatin; porcine skin gelatin (Sigma #G1890) made to 1% in ddH$_2$O.

- Ethanol, for disinfection

- Coverslips; #1.5 thickness, 9-12mm diameter

- Optional: Phorbol 12-myristate 13-acetate (PMA) for differentiation of THP-1 cells into macrophages

3. Specimens preparation and staining

- Formaldehyde; methanol-free formaldehyde 16% (e.g. Pierce #28908) diluted to give a final concentration of 4%.

- Optional: Ammonium chloride (NH_4Cl) solution, 50 mM dissolved in ddH_2O, sterile filtered

- Permeabilization buffer; Prepare 10x stock in 50 mL volume. 2%(w/v) Bovine Serum Albumin Fraction V, 0.2%(w/v) saponin (e.g. Merck #47036) make to 50mL in 10x PBS. Dispense 5 mL aliquots in 50 mL Falcon tubes, store -20°C and make up to 50 mL with ddH_2O when required.

- Nuclear stain e.g. Hoechst 33342 or DAPI

- CellMask, plasma membrane stains for high-content screening (Thermo Fisher Scientific; HCS cell mask red #H32712, HCS cell mask green #H32714)

- Antibodies for staining and appropriate fluorescently-labelled secondary antibodies for visualisation (e.g. Alexa Fluor secondary antibodies by Molecular Probes, ThermoFisher)

- Mounting medium, a hardening mounting medium is required, for example Mowiol 4-88 (Sigma #81381) or ProLong Gold (ThermoFisher #P10144), when working with cells on coverslips

- Microscope slides

- Non-transparent plate seal, for sealing multi-well plates

4. Imaging

- Fluorescence microscope or high-content imager equipped with CCD high-resolution, digital camera with ideally >12-bit dynamic range.

5. Image analysis

- Computer capable of running KNIME Analytics platform version 3.4.2 or newer

6. Online resources

- The HRMAn analysis pipeline can be downloaded from:

  https://github.com/HRMAn-Org/HRMAn

- More information and tutorial videos can be found here: https://hrman.org

- Four different sample datasets for testing HRMAn can be downloaded from

  Dryad: 10.5061/dryad.6vq2mp0

3. Methods

Quantification of the host cell response to *Tg* infection by imaging can be performed on different scales: small pilot experiments can be performed using manual image acquisition and specimens prepared on coverslips (**A**), whereas experiments that require testing of many different conditions can be performed in a high-throughput manner and are usually performed using multi-well plates (96-well or more, **B**). In steps where the sample preparation and analysis differ, we will refer to them as **A** or **B**, as indicated above. Furthermore, two types of analyses are possible using HRMAn, an 'infection analysis' which quantifies infection dynamics as well as host cell responses and a 'recruitment analysis' which additionally quantifies protein recruitment parameters at the PV (see Section 5).

1. Preparation of host cells for infection

In general, cells need to be seeded in defined numbers at the start of an experiment. It is important to first determine the optimal seeding conditions and cell number for the experiment. If the cellular response is the target of the analysis, cells should be seeded as densely as feasible (to get as many of them imaged as possible) but they should not touch each other or grow fully confluent, otherwise image segmentation of the cells will be difficult and might result in inaccurate results (some example seeding densities and time lines of commonly used cell lines are provided in Table 2). For some cells it is advisable to coat plates in order for the cells to adhere better. We can recommend using gelatin as a substrate (See information box: Gelatin coating), but other substrates may be used if they work well for a particular cell type.

> **Gelatin coating**
>
> We find that gelatin-coating of plates significantly improves the attachment of cells during infection processes, which maintains equal numbers per well for imaging experiments. Perform the following steps to make up gelatin and coat plates:
>
> - **Prepare gelatin:**
>   Add 1%(w/v) gelatin powder (e.g. Sigma #G1890 porcine skin gelatin) to ddH$_2$O and dissolve at 60°C with stirring for 30 mins. While liquid, filter through a bottle top filtration system (0.22 μm; e.g. StarLab CytoOne C6032-8233). Aliquot (10-50 mL) and store at 4°C.
> - **Coating plates:**
>   Before using, melt gelatin at 37°C for 20-30 mins until fully liquid. Add enough gelatin to cover bottom of the well(s) and incubate at 37°C for 20 mins or longer.
> - **Seed cells:**
>   Aspirate excess gelatin from the well(s) and plate cells on top immediately.

In general, cells on coverslips are seeded in 24-well plates (**A, Note 1**) whereas for high-throughput experiments they are seeded in 96-well plates (or other multi-well plates, **B, Note 1-3**). It is advisable to omit wells on the outer edge of a plate because they show more evaporation which can skew results. Instead fill them with sterile PBS or culture medium (whatever is available) to prevent evaporation from the cell-containing wells in the centre of the plate. Another consideration during the planning of experiments is the position of different treatment conditions on a plate. Make sure to change them during repeats of experiments to exclude any positional effects that may distort the data. When planning an experiment, it should also be considered to have some wells as uninfected controls, which can be used to track the effect of infection on the number of cells. We also recommend creating a plate map before starting the experiment to keep track of experiment conditions (For an example see **Figure 1A**).

In some experimental settings the cells can be transfected with DNA or RNA before the infection. Additionally, cells can be pre-treated with cytokines the evening

before infection to induce responses or treated with inhibitors to manipulate their response. This depends on the type of experiment that is being performed (**Note 4**).

2. Infection procedure and fixation of samples

1. The day before infection, split *Tg* tachyzoites from freshly syringe-lysed HFFs onto new HFFs cells and let them grow to sufficient numbers overnight. Usually 1:5 splitting from a T25 flask into a new T25 flask with confluent HFFs gives sufficient high viability parasites (HFFs and *Tg*-infected HFFs are maintained in DMEM + 10% FCS).

2. For infection, harvest tachyzoites from HFFs by scraping the cells and syringe lysing through a 25 G needle into a 15 mL conical tube.

3. Clear lysate of cell debris by low speed centrifugation at 50 x g for 5 minutes. Maximum centrifuge breaking is suitable for this and all subsequent steps.

4. Transfer supernatant into a new 15 mL conical tube and pellet tachyzoites by centrifugation at 600 x g for 7 minutes.

5. Remove supernatant and resuspend *Tg* tachyzoites in complete medium, that is appropriate for the host cell line used for infection.

6. Take 10 µL tachyzoite suspension and mix 1:1 with trypan blue and fill 10 µL into a haemocytometer. Count viable parasites (unstained) and calculate their concentration.

7. Prepare dilutions of tachyzoites to have the appropriate number for the designated multiplicity of infection (MOI) in the appropriate volume:
Infection in 24-well plates is performed using 450 µL *Tg* suspension per well (**A**), infection in 96-well plates is performed using 150 µL *Tg* suspension per well (**B**) (see information box: Example: Infection of THP-1 cells in 60 wells of

a 96-well plate). During infection, fresh cytokines and/or inhibitors can be added

(if the inhibitors do not affect *Tg* itself, Note 4).

Example: Infection of THP-1 cells in 60 wells of a 96-well plate
THP-1s were seeded as described above with 30,000 cells per well in the centre 60 wells of a 96-well plate. They all will be infected with *Tg* using a MOI = 3 in 150 μL per well.
Calculation:
- **Total *Tg* suspension needed:**
60 wells x 150 μL/well = 9 mL medium, make 10 mL to have some additional volume.

- **Number of tachyzoites needed:**
30,000 host cells/well x 3 (MOI) = 90,000 *Tg*/well

- **Dilution:**
(90,000 *Tg*/well) / (150 μL/well) = 600,000 *Tg*/mL

- **Result:**
10 mL x 600,000 *Tg*/mL = 6,000,000 *Tg* = 6×10^6 *Tg*
→ Prepare 10 mL medium with 6×10^6 *Tg* tachyzoites and add 150 μL per well.

---

8. Once the *Tg* dilution for infection is prepared, add the appropriate volume per well and spin the plates at 500 x g for 5 mins (this helps to synchronise the infection). Place back in the tissue culture incubator at 37°C, 5% CO_2.

9. 2 hours post infection, any uninvaded/non-viable parasites are removed from the wells by aspirating the medium and carefully washing the infected cells three times with pre-warmed PBS (Note 5-6). This prevents uninvaded parasites being included in the final analysis. The wells are then replenished with pre-warmed complete medium. At this stage, cytokines and/or inhibitors can be added to the wells as appropriate (Note 4).

10. After the required time of infection (2-24 hours or more, depending on the experiment), wash cells twice with pre-warmed PBS (**Note 5**) and afterwards add 450 µL PBS per well of a 24-well plate (**A**) or 150 µL per well of a 96-well plate (**B**).

11. To fix the cells add methanol-free 16% PFA to the samples,150 µL per well of a 24-well plate (**A**) or 50 µL per well of a 96-well plate (**B**). This achieves 4% PFA final concentration.

12. Move samples to a light protected place (or cover with foil) and incubate for 15 minutes at room temperature.

13. Following incubation with 4% PFA, aspirate and discard the PFA into the appropriate waste and wash the cells three times with PBS (**Note 5+6**). Then fill all wells with PBS to cover the cells and leave at 4°C overnight to quench residual PFA.

14. OPTIONAL: If specimen needs to be processed immediately, wash with 50 mM NH_4Cl to quench residual PFA and then wash again twice with PBS.

3. Staining of fixed specimens for immunofluorescence microscopy

1. Permeabilize the cells: Remove PBS from the samples and add 300 µL of permeabilization buffer per well of a 24-well plate (**A**) or 100 µL per well of a 96-well plate (**B**) and incubate 30 minutes at room temperature in the dark.

2. (**A**) Aspirate permeabilization buffer from the coverslips and proceed with staining. For coverslips, it is efficient to carefully pick them from the plate using fine point forceps and incubate inverted on a drop of staining solution on parafilm (10-30 µL solution per 9-12mm coverslip respectively, is sufficient). In this instance, it is advisable to have the parafilm overlaying a sheet of pre-wet

filter paper in a square petri-dish to maintain a humid environment during incubation. (**B**) Remove permeabilization buffer from the 96-well plate (**Note 7**) and proceed with adding staining solution (see below). For 96-well plates the minimum volume to add to cover all cells is 70 µL per well!

Depending on the analysis you want to perform later on, decide on the staining procedure. Continue with step 3 if you simply want to analyse the infection process or continue with step 4 if you additionally want to analyse protein recruitment to the *Tg* vacuoles.

3. Staining for infection analysis:

    For simple infection analysis it is sufficient to have stained nuclei, cytosol and parasites. If non-fluorescent *Tg* lines were used for infection, perform antibody (Ab) staining as described below for host/parasite proteins (**Note 8**).

    a. Prepare staining solution: Permeabilization buffer + 1:2500 CellMask far red (for green parasites, use different colour if the parasites express a different fluorophore) and 1:5000 Hoechst 33342 (or other nuclear stain).

    b. Replace permeabilization buffer with 70 µL staining solution per well and incubate for 1 hour at room temperature in the dark.

    c. Continue with final wash steps.

4. Antibody-staining for protein recruitment analysis:

    For protein recruitment to *Tg* vacuoles analysis the cells have to be stained with an Ab against the protein of interest, if the protein is not tagged with a fluorescent protein. Furthermore, if non-fluorescent *Tg* lines were used for infection, perform Ab staining of them as well (**Note 8**) (see information box: Example: staining of HeLa cells infected with colourless *Tg*).

a. Prepare first staining solution: Permeabilization buffer with added dilution of primary antibodies, add to the samples and incubate for 1 hour at room temperature in the dark (**Note 9**).

b. Wash three times with PBS.

c. Prepare second staining solution: Permeabilization buffer with added dilution of appropriate secondary antibodies (coupled to fluorophores of choice) and 1:5000 Hoechst 33342, add to the samples and incubate for 1 hour at room temperature in the dark (**Note 9**).

d. Continue with final wash steps.

5. Following staining wash all specimens at least five times with PBS to remove as much secondary antibody or unused stain as possible. This will result in less background during image acquisition. Coverslips need to be washed in ddH_2O and then mounted using Mowiol (or other mounting medium) onto glass microscope slides (**A**) or fill wells of the 96-well plates with 200 µL PBS per well and seal the plate using non-transparent plate seal (**B**).

At this point samples can be stored at 4°C for more than a month, albeit some signal loss has to be expected.

> **Example: staining of HeLa cells infected with colourless *Tg***
>
> HeLa cells were infected with colourless *Tg* at a MOI = 3 for 6 hours and we wanted to visualise ubiquitin and p62 decoration of vacuoles. To do this, we needed to stain the parasites and host cell proteins with antibodies:
>
> - **1ˢᵗ stain:**
>
> Prepare permeabilization buffer with added rabbit anti-ubiquitin, goat anti-p62 and mouse anti-*Tg*SAG1*, add to samples and incubate for 1 hour. Then perform wash steps.
>
> - **2ⁿᵈ stain:**
>
> Prepare permeabilization buffer with added Hoechst 33342, anti-mouse-Alexa488, anti-rabbit-Alexa568 and anti-goat-Alexa647*, add to samples and incubate for 1 hour. Then perform final wash steps.
>
> This will result in specimen that have blue nuclei, green *Tg*, red ubiquitin, and far-red p62. These samples can then be used to analyse (co-)recruitment of the two proteins to the PV using HRMAn (see below)
>
> *Example Abs! Use as appropriate and optimize for your own experiment.

4. Image acquisition and data preparation

For image acquisition make sure to use the same standards as are applicable in general when using fluorescence microscopes:

1. Images should not only look 'pretty' but represent quantifiable data.

2. Settings should be adjusted using a real sample and kept the same for an entire image acquisition dataset (i.e. several 96-well plates or coverslips) in order to be able to compare the different conditions later on.

3. Because images are obtained for quantitative analysis ensure that gain and offset are set to use the entire dynamic range of the detector, this prevents clipping of the dataset at either end.

4. Use of higher bit cameras (e.g.12-bit) is advisable as they offer a higher dynamic range.

5. Data should be saved in a raw, uncompressed format such as .tiff

In general, the use of high-content imaging systems, which were developed for automated image acquisition will solve many of these problems.

When imaging coverslips ensure that you acquire enough non-overlapping and randomly chosen fields of view (even if this means you image an empty spot on your coverslip)! Only like this can you ensure to get real insight into what was happening on your coverslip/ in your well during infection (Loss of many cells or clumping of them in certain spots might actually represent an interesting phenotype!). To obtain an appropriately sized dataset, you should aim to image at least 1000 cells per condition which at a MOI of 1 should approximately result in a similar number of *Tg* vacuoles (this number may be smaller if cells were treated to become parasiticidal). The number of images required to reach these amount of cells depends on the magnification and camera detector size used for imaging. For example, collecting a grid of 5-by-5 adjacent fields of view containing approximately 50 cells each (at a magnification of 20x and when cells have been seeded according to the recommendations in Table 2) would give a suitable sample size for analysis.

As *Tg* is a reasonably large object, 20x magnification is sufficient for infection analysis and 40x magnification for protein recruitment analysis. Use of confocal imaging systems is not necessary at this point. If confocal images were acquired use the appropriate option in the image pre-processing step to create z-projections (see below).

1. Once the images are acquired and saved as single channel files they need to be prepared for analysis using HRMAn. Use the **"Microscope image pre-processing"** KNIME pipeline (Figure 1B) from the HRMAn GitHub repository to prepare your dataset.

2. Load the images into the program by specifying the location in the **"File upload"** node.

3. Load your plate map for the experiment (see below) into the program by specifying the .csv file in the **"Table reader"** node (**Table 3**).

4. Select the **"Image writer"** node on the right-hand side and start the pipeline by executing the node (Right click and click execute in the menu that opens).

5. The program will ask for:

   a. the number of samples

   b. the number of fields per sample

   c. the number of channels

   d. and the number of planes per field

   Enter the values and confirm by clicking ok. Once all parameters are entered the program will display a summary and ask if they are correct.

6. The program will then automatically rename all images and write them into the same folder as your input images. If several planes were acquired, a z-projection is performed to reduce the amount of data (subject to future 3D-analysis update).

7. The files are now named according to the following scheme:

A01f00d00*, with A01 specifying the first well of the plate or representing the first sample/coverslip, f00 being the first field and d00 being the first channel.

*It is important that all input images are named exactly according to this format!

As HRMAn was developed to work with images acquired in a plate format, this step ensures that data that did not originate from automated imaging of 96-well plates, or was named differently, will be analysed and clustered correctly using similar plate maps (see below).

5. High-content image analysis using HRMAn

Once images are acquired and have been pre-processed (re-named), they are ready to be analysed using HRMAn. The analysis runs in two stages: Stage 1 simply analyses common infection parameters and uses a decision tree machine learning algorithm to classify replication of the *Tg* parasites. This stage is based on well-established image analysis methods. Stage 2 uses a deep convolutional neural network (CNN) to classify recruitment of proteins to the PV. Before they can be used, the machine learning algorithms need to be trained:

1. *Creating a dataset for the decision tree machine learner:*

For this step you can use the **"Create learning datasets"** KNIME pipeline (Figure 1C) that can be found in the HRMAn GitHub repository. It will help you to annotate individual *Tg* vacuoles based on how many parasites they contain. To run the pipeline, open it in KNIME and perform the following steps (to learn how to set up KNIME and the individual analysis pipelines, you can find video tutorials on https://hrman.org/):

a. Use the **"File Upload"** node to load your images into the pipeline. It is important to have images acquired at a magnification of 20x to have the right scaling! Make sure you have re-named your images according to the format described above (File name A01f00d01, A01f00d02, etc).

b. Select the **"Image Viewer"-Input images** node and execute it. This will start the analysis.

c. The program will ask you for some information:

    i.   The channel number and their order.

    ii.   The magnification you used to image.

    iii. The pathogen type (enter a 1 for *Tg*).

d. The program will then perform the first steps of the image analysis up to the segmentation. Once finished, the little traffic light underneath the node **"Image Viewer"-Input images** turns green. You can use this node to view the images that you have provided.

e. Next use the **"Image Viewer"-Nuclei labels, -Pathogen labels** or **-Cell labels** to see the segmentation the program has performed and make sure they are good (**Note 10**).

f. Then select the **"Image Viewer"-View Vacuoles** node and the **"CSV Writer"-Learning dataset** node execute them. The csv writer node will write a table with the attributes of the *Tg* vacuoles to the defined output location. The last column in this document "Parasites/vacuole" will be empty.

g. Then use the **"Image Viewer"-View Vacuoles** node to look at the individual cropped vacuoles. On the bottom you will see the vacuole ID, which will match the rows in the csv file. Determine the number of parasites (1, 2, 4, 4 or more) and type the value into the empty column of the csv file. As it becomes very difficult to determine the exact number of parasites if more than 4 parasites are contained within the same vacuole, we decided to pool them. If you encounter vacuoles like that, simply type in an 8 and the machine learner will know what to do. Once you have finished annotating the vacuoles and every row in the csv

file is filled in with the number of contained parasites, you can use this file for all future experiments using HRMAn analysis. Simply read it into the pipeline as described below.

2. Creating a dataset for training the CNN and training:

For creating datasets and training the deep neural network implemented in HRMAn ("HRMAlexNet"), we recommend watching the tutorial videos on hrman.org and following the detailed instructions provided there. Already trained models can be obtained from the GitHub repository (https://github.com/HRMAn-Org/HRMAn) or the homepage (https://hrman.org/).

Now the HRMAn analysis itself can be run. Please follow the instructions provided there to set up the program properly; Figure 1D):

1. Use the **"File upload"** node to specify where HRMAn can find your images

2. Using the **"File Reader"** node provide the plate map:

   The template for a plate map can be found in the HRMAn repository (see Table 3 for the layout of the spreadsheet).

   It is important that the file is saved as .csv and that the structure of the table (i.e. the index and the name of the columns) is not changed, as this could lead to errors in the program. In the plate map define which wells represent repeats of the same conditions by using continuous numbering of the different conditions on your plate. The program also needs to know if the cells in the sample were infected or if they represent uninfected, cells-only conditions. If they were infected this has to be represented with a 1, otherwise fill the cell with a 0.

3. Using the **"File Reader"** node to provide a learning model for the decision tree (see above for instructions on how to create this file).

4. Using the **"DL4J Model Reader"** node read in the file for the trained CNN (even if you are not performing recruitment analysis, provide a model here, as otherwise the program might crash. You can use the trained HRMAlexNet-Ubiquitin that can be downloaded from the GitHub repository).

5. Select the **"Image Viewer"-Input images** node and execute it. This will start the analysis.

6. HRMAn will ask you for some information:

    a. The magnification that was used to image the cells: Simply enter the magnification as a number.

    b. The pathogen you want to analyse: For *Tg* type a 1 and confirm by clicking ok.

    c. The type of analysis you want to perform:

    For infection analysis (stage 1 only) type a 1,

    For recruitment analysis (stage 1 + 2) type a 2

    and confirm by clicking ok (see information box: Example: Recruitment analysis using HRMAn).

    d. The number of channels you used to image your cells (the default for infection analysis is 3 channels).

    e. The order of your channels.

7. HRMAn will then perform the first steps of the analysis up to the segmentation. Once this is done, the traffic light indicator underneath the node **"Image**

Viewer"-Input images turns green. You can use this node to view the images that you have provided.

8. Next use the **"Image Viewer"-Nuclei labels, -Pathogen labels** or **-Cell labels** to see the segmentation HRMAn has performed and make sure they are good (**Note 10**).

9. If you are happy with the segmentation you can select and execute the **"XLS Writer" -Properties of the cells** and **-Properties Pathogens** nodes. This will start the rest of the analysis and write an .xlsx into your selected working directory. HRMAn indicates when the analysis is complete and following this the results file will open.

10. Save the file in a different location, for later analysis. (IMPORTANT: File will be overwritten once the analysis pipeline is executed again!).

Example: Recruitment analysis using HRMAn

HeLa cells were infected with *Tg* for 6 hours and stained for ubiquitin. The images were re-named and are ready for analysis.

- **Load the images, the plate map, learning models and start the analysis.**
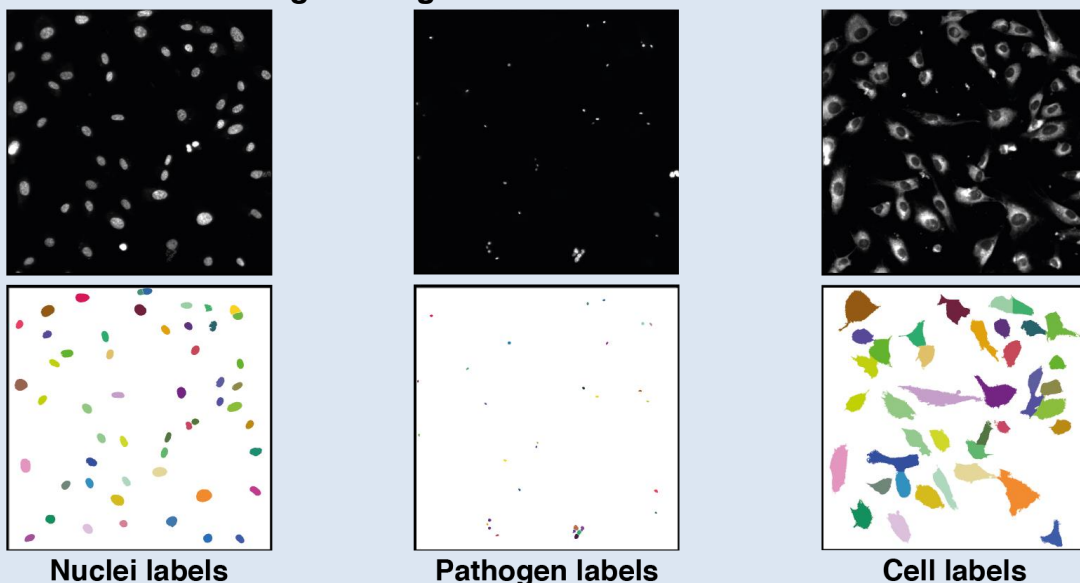- **Type in the information:**

Cells were imaged at a magnification of 40x (type in 40), the pathogen was *Tg* (type in 1), HRMAn shall perform recruitment analysis (type in 2). Then double check the summary and confirm if correct (type in 1). Next, HRMAn wants to know the channel number, you have 3 channels (type in 3) and their order (type in as they come up). Then HRMAn wants to make sure that you have provided the correct plate map (type in 1). Then the analysis and segmentation will be performed.

- **Check the input images:**

Infected cells   **Nuclei**   ***Toxoplasma***   **Ubiquitin**
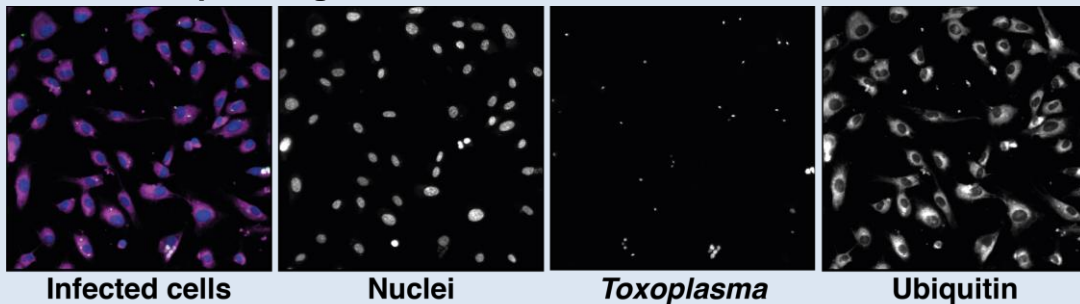
All images loaded? Yes!

- **Check the nuclei/*Tg*/cell segmentation:**

Nuclei labels   **Pathogen labels**   **Cell labels**

Segmentation ok? Yes!

- **Finish the analysis:**

If all segmentations are good, select the output nodes and execute them. This will start the rest of the analysis and write the output file to your computer.

When the HRMAn analysis of the images is finished, the user is provided with a spreadsheet file that contains all computed results. This data can then be visualised using graph plotting programs. Here users need to ensure that all segmentation steps of the analysis have performed well and that the data is of the expected quality. Further, it is the user's responsibility to ensure data integrity and perform the appropriate statistical analysis.

Once familiar with HRMAn and KNIME analytics platform, it is easy to adapt the analysis pipeline exactly to the needs of the researcher using it, or extend it to measure more parameters. This allows study of the parasite biology, the host response to infection and the interplay of both sides. Importantly, HRMAn is open-source and employs neural network precision for the analysis of host protein recruitment to the pathogen. HRMAn has the potential to be tailored for the analysis of other intracellular pathogens, e.g. *Salmonella, Plasmodium*.

4. Notes

Note 1: Use tissue culture treated plates.

Note 2: Use black-wall plates to reduce scattered light. If higher magnification imaging is required (>20x) it is advisable to use plates with a coverslip-thickness glass bottom. Handle glass-bottom plates with extra care, they are extremely fragile! We find that taping cardboard to the bottom helps prevent the glass from breaking. To reduce costs, higher magnification imaging can also be performed in plastic-bottom plates, depending on the microscope used for image acquisition, though this might increase background noise and negatively affect image quality.

Note 3: Omit wells on the edge of the plate and fill them with sterile PBS to prevent evaporation and edge effects.

Note 4: When using cell treatments such as cytokines or inhibitors, it is important to include appropriate no treatment and vehicle-only controls. Some inhibitors will need to be washed off the cells prior to infection as they may affect the invasion or growth of the parasite directly. We suggest washing twice with warmed medium or PBS.

Note 5: When performing wash steps of cells, be very gentle to not wash away the cells. For cells that do not adhere well, we recommend using a wire manifold (e.g. Drummond Scientific Co., Straight, 8 places for 96 well plate, # 3-000-093) which is gentler and set to leave a residual amount of medium/PBS behind on washing. This not only minimises damage to the cell layer by tips, but also prevents poorly adherent cells from being drawn up with the medium/PBS on aspiration.

Note 6: All solutions used during cell culture and staining should be filter sterilised through a 0.22 µm filter to prevent microbe and dust contamination of samples.

Note 7: We find that following fixation, inverting the plates and drying them by banging on a stack of paper is efficient and does not result in cell loss.

Note 8: We find that anti-SAG1 (*Tg* surface antigen 1) is a well-performing stain for *Tg*. This antibody is available commercially e.g. GeneTex mouse monoclonal, clone B754M #GTX38936.

Note 9: Staining with antibodies needs to be optimised first! As antibodies are highly variable, we recommend using antibodies that have been used for immunofluorescence experiments before. Some companies validate their primary antibodies using siRNA negative controls and provide staining images for endogenous levels of protein, which can be helpful when selecting reliable and specific antibodies. These steps can also be performed by the user. Initially it will be important to include primary only and secondary only antibody controls and where co-staining is required to incorporate single stain controls. Make sure to use enough antibody to get a good signal without overstaining the sample which will create non-specific background.

Note 10: The HRMAn analysis pipeline has been designed to perform segmentation of nuclei, cells and *Tg* automatically using different automated normalisation, thresholding and connected component analysis strategies combined with filtering. If the segmentation does not perform to the satisfaction of the user, it is possible to manually adjust the threshold used for segmentation. This can be achieved by opening the "Segmentation" node contained within the "Analysis" metanode and altering the threshold within the "Global Thresholder" used for the respective segmentation. However, the most common reasons for a failed segmentation is poor image acquisition/quality or a too high cell density. In these cases, we recommend preparing new samples.

Figure legend

Figure 1: Example plate layout and overview of the KNIME analysis pipelines used for HRMAn.

(A) Example plate map of a 96-well plate, wells on the edge of the plate were omitted (grey) and individual sample groups are performed in triplicate wells (= technical replicates). We recommend making plate maps like the one depicted to keep track of the layout.

(B) Overview of the "Microscope image pre-processing" pipeline (depicts user interface), used to rename and format images from any imaging platform before analysing them with HRMAn.

(C) Overview of the "Create learning datasets" pipeline (depicts user interface), used to create an annotated dataset for training of a decision-tree machine learning algorithm implemented in HRMAn to classify replication of *Toxoplasma gondii* (*Tg*) inside of its vacuole.

(D) Overview of the "HRMAn" analysis pipeline (depicts user interface), used to perform high-content/ -throughput image analysis to study host-pathogen interaction of *Tg*.

References

1. Sabin AB, Olitsky PK. TOXOPLASMA AND OBLIGATE INTRACELLULAR PARASITISM. Science (80-). 1937;85(2205):336–8.

2. Dubey JP, Miller NL, Frenkel JK. The Toxoplasma gondii oocyst from cat feces. J Exp Med. 1970 Oct 1;132(4):636–62.

3. Pappas G, Roussos N, Falagas ME. Toxoplasmosis snapshots: Global status of Toxoplasma gondii seroprevalence and implications for pregnancy and congenital toxoplasmosis. Int J Parasitol. 2009 Oct 1;39(12):1385–94.

4. Beverley JK. Toxoplasmosis. Br Med J. 1973 May 26;2(5864):475–8.

5. Luft BJ, Remington JS. Toxoplasmic encephalitis in AIDS. Clin Infect Dis. 1992 Aug;15(2):211–22.

6. Morisaki JH, Heuser JE, Sibley LD. Invasion of Toxoplasma gondii occurs by active penetration of the host cell. J Cell Sci. 1995 Jun 1;108(6):2457 LP – 2464.

7. Suss-Toby E, Zimmerberg J, Ward GE. Toxoplasma invasion: the parasitophorous vacuole is formed from host cell plasma membrane and pinches off via a fission pore. Proc Natl Acad Sci U S A. 1996 Aug 6;93(16):8413–8.

8. Besteiro S. Toxoplasma control of host apoptosis: the art of not biting too hard the hand that feeds you. Microb cell. 2015 May 30;2(6):178–81.

9. Krishnamurthy S, Konstantinou EK, Young LH, Gold DA, Saeij JPJ, Roth M. The human immune response to Toxoplasma: Autophagy versus cell death. Knoll LJ, editor. PLOS Pathog. 2017 Mar 9;13(3):e1006176.

10. Goldman M, Carver RK, Sulzer AJ. Reproduction of Toxoplasma gondii by internal budding. J Parasitol. 1958 Apr;44(2):161–71.

11. Jones TC, Hirsch JG. The interaction between Toxoplasma gondii and mammalian cells. II. The absence of lysosomal fusion with phagocytic vacuoles containing living parasites. J Exp Med. 1972 Nov 1;136(5):1173–94.

12. Mordue DG, Desai N, Dustin M, Sibley LD. Invasion by Toxoplasma gondii establishes a moving junction that selectively excludes host cell plasma membrane proteins on the basis of their membrane anchoring. J Exp Med. 1999 Dec 20;190(12):1783–92.

13. Charron AJ, Sibley LD. Molecular Partitioning during Host Cell Penetration by Toxoplasma gondii. Traffic. 2004 Sep 1;5(11):855–67.

14. Virreira Winter S, Niedelman W, Jensen KD, Rosowski EE, Julien L, Spooner E, et al. Determinants of GBP Recruitment to Toxoplasma gondii Vacuoles and the Parasitic Factors That Control It. Moreno SN, editor. PLoS One. 2011 Sep 8;6(9):e24434.

15. Degrandi D, Kravets E, Konermann C, Beuter-Gunia C, Klümpers V, Lahme S, et al. Murine guanylate binding protein 2 (mGBP2) controls Toxoplasma gondii replication. Proc Natl Acad Sci U S A. 2013 Jan 2;110(1):294–9.

16. Foltz C, Napolitano A, Khan R, Clough B, Hirst EM, Frickel E-M. TRIM21 is critical for survival of Toxoplasma gondii infection and localises to GBP-positive parasite vacuoles. Sci Rep. 2017 Jul 12;7(1):5209.

17. Selleck EM, Orchard RC, Lassen KG, Beatty WL, Xavier RJ, Levine B, et al. A Noncanonical Autophagy Pathway Restricts Toxoplasma gondii Growth in a Strain-Specific Manner in IFN-γ-Activated Human Cells. MBio. 2015 Sep 8;6(5):e01157-15.

18. Clough B, Wright JD, Pereira PM, Hirst EM, Johnston AC, Henriques R, et al. K63-Linked Ubiquitination Targets Toxoplasma gondii for Endo-lysosomal

Destruction in IFNγ-Stimulated Human Cells. PLOS Pathog. 2016;12(11):e1006027.

19. Clough B, Frickel E-M. The Toxoplasma Parasitophorous Vacuole: An Evolving Host-Parasite Frontier. Trends Parasitol. 2017 Jun 1;33(6):473–88.

20. Fisch D, Yakimovich A, Clough B, Wright J, Bunyan M, Howell M, et al. Defining host–pathogen interactions employing an artificial intelligence workflow. Elife. 2019 Feb 12;8.

21. Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, et al. KNIME: The Konstanz Information Miner. In: Data Analysis, Machine Learning and Applications Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg; 2008. p. 319–26.

Acknowledgements

Author Contribution

DF, BC and EMF wrote the manuscript. All authors contributed to revising the manuscript.

Competing financial interests

The authors declare no competing financial interest.