

# Deducing the mechanism of action of compounds identified in phenotypic screens by integrating their multiparametric profiles with a reference genetic screen

Varadharajan Sundaramurthy<sup>1</sup>, Rico Barsacchi<sup>1</sup>, Mikhail Chernykh<sup>1</sup>, Martin Stöter<sup>1</sup>, Nadine Tomschke<sup>1</sup>, Marc Bickle<sup>1</sup>, Yannis Kalaidzidis<sup>1,2</sup> & Marino Zerial<sup>1</sup>

<sup>1</sup>Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG), Dresden, Germany. <sup>2</sup>Faculty of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia. Correspondence should be addressed to M.Z. (zerial@mpi-cbg.de).

Published online 30 January 2014; doi:10.1038/nprot.2014.027

Cell-based high-content screens are increasingly used to discover bioactive small molecules. However, identifying the mechanism of action of the selected compounds is a major bottleneck. Here we describe a protocol consisting of experimental and computational steps to identify the cellular pathways modulated by chemicals, and their mechanism of action. The multiparametric profiles from a 'query' chemical screen are used as constraints to select genes with similar profiles from a 'reference' genetic screen. In our case, the query screen is the intracellular survival of mycobacteria and the reference is a genome-wide RNAi screen of endocytosis. The two disparate screens are bridged by an 'intermediate' chemical screen of endocytosis, so that the similarity in the multiparametric profiles between the chemical and the genetic perturbations can generate a testable hypothesis of the cellular pathways modulated by the chemicals. This approach is not assay specific, but it can be broadly applied to various quantitative, multiparametric data sets. Generation of the query system takes 3–6 weeks, and data analysis and integration with the reference data set takes an 3 additional weeks.

## INTRODUCTION

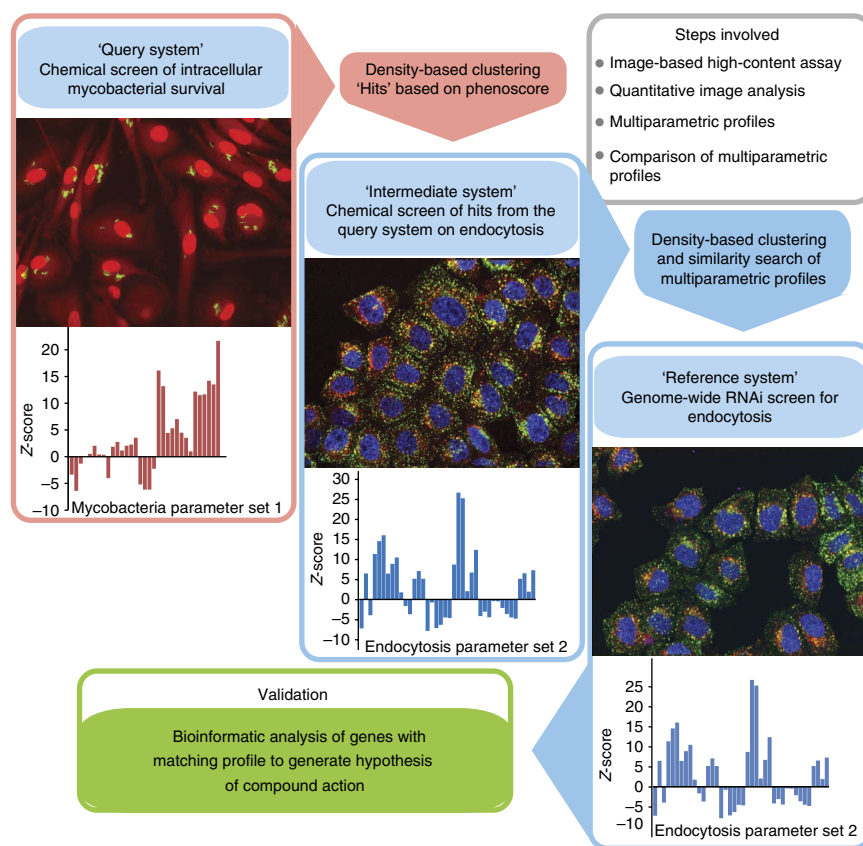
Biological systems are characterized by a high level of molecular complexity, which arises from the wiring between individual cellular components within a pathway and between different pathways. Chemical or genetic perturbations alter this wiring and thus affect phenotypes. These phenotypes can be captured comprehensively by high-content (HC), image-based cellular assays combined with computational analysis tools that can extract multiple parameters indicative of functional properties, in a quantitative and comprehensive manner<sup>1,2</sup>. Therefore, HC cell-based assays have been instrumental in functional genomics, but they have also emerged as promising tools in chemical genetics for exploring cellular mechanisms through small organic molecules<sup>3</sup>. Furthermore, HC screening (HCS) has held great promise for drug discovery<sup>4–6</sup>. HCS has substantial advantages over traditional single target-based high-throughput screens (HTS). First, whereas classical target-based screens involve single or a few experimental readouts, HCS captures the systems response with up to a few hundred parameters, some of which may be correlated, thus allowing profiling of a large part of the phenotypic space. This enables the detection of subtle phenotypes that may not be considered statistically significant on the basis of a single parameter, but that can be highly significant when multiple parameters are considered. Such phenotypes are typically missed in the case of monoparametric, target-based HTS. Second, target-based screens are typically performed *in vitro* by inspecting individual cellular components such as purified enzymes or ligand-receptor complexes. In such experiments, a small chemical molecule producing a strong phenotype *in vitro* may have side effects on the cell, owing to its binding of different proteins within the cell<sup>7</sup>. In addition, the premise upon which the target was selected is often based on incomplete knowledge at the systems level. This implies that the effect observed *in vivo* may be different from the expected phenotype, even in the case of validated targets<sup>7</sup>. These factors

have a key role in the high rate of attrition in drug discovery. In contrast, HCS allows for the identification of phenotypes at the cellular level even during the primary screen. For example, overtly toxic compounds can be eliminated at early stages.

Despite such advantages, phenotype screens have a major bottleneck in comparison with target-based HTS<sup>6,8</sup>. Besides the limitations in throughput owing to the complexity of the assay and microscopy measurements, the molecular target or the cellular pathways targeted by a 'hit compound' is not directly deducible from the primary screen. Indeed, laborious follow-up of these primary activities at the level of individual molecules is necessary to unravel the molecular processes underlying the compound action<sup>9,10</sup>. Thus, the lack of methods for gaining insights into the mechanism of action (MoA) of hit compounds remains a severe impediment to the use of HCS in chemical genetics and drug development.

To overcome this restriction, we have developed a combined computational and cell biological pipeline to predict the MoA of lead compounds emerging from primary screens. This method involves the extraction and integration of the quantitative multiparametric profiles between HC chemical screens and RNAi screens<sup>11</sup> (Fig. 1). The basic idea is that the similarity in the multiparametric profiles between a chemical and a genetic perturbation implies a mechanistic link between the compounds and the genes involved. Such a comparison is possible only under the condition of using a sufficient number of parameters to capture the phenotype, as the probability of a random match between the chemical and genetic profiles decreases rapidly with the increase in the number of parameters analyzed. However, one caveat to consider is that the phenotype of a compound and that of a siRNA may not be identical because they interact with the cells differently, and thus they could generate different multiparametric profiles. For this reason, we think that a precise matching between the

**Figure 1** | An overview of the distinct stages involved in the protocol. The protocol consists of the distinct stages of the generation of quantitative multiparametric profiles from the query and the reference data sets (pink and blue outlined boxes), followed by comparison of the multiparametric profiles to generate testable hypothesis (green box). The protocol involves image-based cellular assays. Representative images from the different assays involved, i.e., mycobacteria assay and endocytosis assays, are shown. Quantitative multiparametric profiles, as shown below the images, are extracted from the image analysis. Details of the assays, imaging and quantitative image analysis are provided in the protocol.



phenotypic profiles of a compound with that of a silenced gene is probably not feasible at the present stage. Nevertheless, as we explain below, matching the phenotypic profile of a group of compounds with that of a cellular pathway can be successful<sup>11</sup>.

Here we describe the protocol for the generation of the primary data sets and integration of multiparametric profiles of chemical and siRNA screen results to deduce the MoA of the selected compounds. We have successfully applied this procedure to the analysis of the MoA of chemicals that trigger the killing of intracellular mycobacterium in primary human macrophages<sup>11</sup>. In this protocol, we consider the chemical screen as the query system. In our specific case, it is the chemical screen for host modulators of intracellular mycobacterial survival<sup>11</sup>. We have used the genome-wide endocytosis RNAi screen<sup>1</sup> as a reference system. As the two screens are biologically different, we have performed an intermediate screen by using the chemicals from the mycobacteria assay in the endocytosis assay. The multiparametric profiles from this screen are compared with the reference system (Fig. 1).

There are four different stages in the protocol (Fig. 1): (i) generation of the reference data set, (ii) generation of the query data set, (iii) generation of the intermediate data set and (iv) comparison of the multiparametric profile of the intermediate with the reference data sets. Although the experimental procedure presented here for the first three stages (Steps 1–70) describes specific reference, query and intermediate systems, we would like to emphasize that the process for the comparison of multiparametric profiles between chemical and genetic screens to determine the MoA is not restricted to these systems. We have developed the tools to perform this process (the fourth stage of the protocol; Steps 71–83) with any HC reference data set.

### Experimental design

**The reference system: genome-wide, HC (multiparametric) screen of endocytosis.** We have previously studied the endocytic pathway in a HC image-based, genome-wide RNAi screen<sup>1</sup>. This screen has captured the vast phenotype space produced by systematically silencing individual human genes in an endocytosis assay. Notably, the screen has revealed the interaction of endocytosis

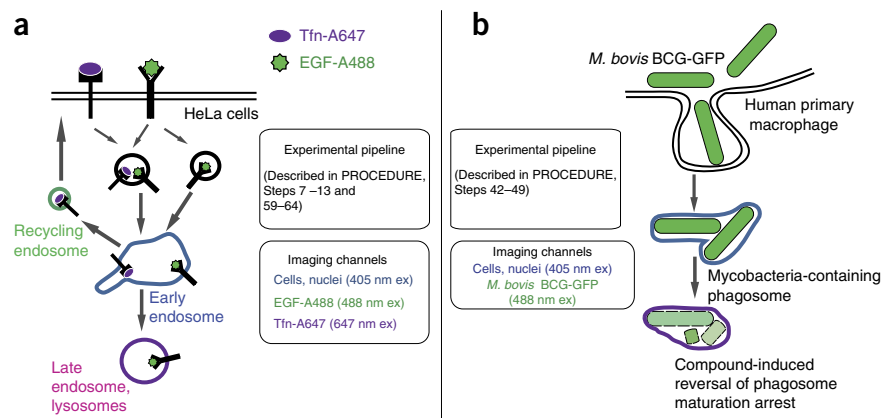
with other cellular pathways such as signaling and metabolic pathways<sup>1</sup>. The endocytosis assay involved the simultaneous uptake and trafficking of two different types of endocytic cargo: epidermal growth factor (EGF) and transferrin (Fig. 2a). The basic experimental steps involved in the generation of the reference data set are outlined in Steps 1–13. The resultant data set contains a comprehensive set of quantitative multiparametric profiles reporting various features of the different cargo-containing endosomes such as their number, morphology, content, position and distribution, as well as the association between the different endosomes<sup>1</sup>. This data set is publicly accessible at <http://endosomics.mpi-cbg.de>. Details on the organization of the database are provided in **Supplementary Figure 1**. This data set provides a wealth of information about the genes involved in endocytosis.

In our case, we considered the genome-wide endocytosis screen as our reference system. However, we would like to emphasize again that the procedure for comparing multiparametric profiles is not restricted to this particular reference system. In recent years, many other HC screens that yield multiparametric data sets have been published<sup>12–16</sup>. If the query system addresses a different biological process, an alternate reference data set could be considered. However, it must be stressed that the reference data set should be comprehensive, and it should capture the systems properties using multiple independent parameters.

**The query system: HC chemical screen of intracellular mycobacterial survival.** The query system we used is the HC chemical screen for intracellular modulators of mycobacterial infection of

## PROTOCOL

**Figure 2** | Schematics of the HC assays for endocytosis and intracellular mycobacterial survival. **(a)** The endocytosis assay is performed in HeLa cells by incubating cells with two different fluorescently labeled cargos: EGF conjugated to Alexa Fluor 488 and Transferrin (Tfn) conjugated to Alexa Fluor 647. The figure shows the schematic of the trafficking routes of the two different cargos after their cellular uptake. **(b)** The intracellular mycobacterial survival assay is performed by infecting human primary monocyte-derived macrophages with *M. bovis* BCG-GFP. Chemical perturbations are introduced after the infection is established. The figure shows the schematic of mycobacterial trafficking inside the infected cell. The experimental pipeline and the imaging conditions for the two assays are indicated.



human primary macrophages<sup>11</sup>. *Mycobacterium tuberculosis* is the causative agent of the dreaded disease tuberculosis (TB), which is the leading cause of death from bacterial infection<sup>17</sup>. Owing to the resurfacing of TB, as well as the emergence of multiple drug-resistant and extensively drug-resistant TB strains, there is a pressing need for the identification of new drugs against TB. Indeed, several new molecules have been identified and some are in various phases of clinical trials<sup>18</sup>. Most of these molecules target the bacteria directly. However, by focusing exclusively on bacteria as targets, the emerging therapies are likely to be of prolonged duration and vulnerable to the emergence of drug-resistant strains. These factors have imposed considerable burden on the management of the disease. Therefore, host-based therapies for targeting intracellular mycobacterial infections offer an attractive alternative. Such approaches are gaining interest in other therapeutic areas as well, such as cancer<sup>19</sup>.

*M. tuberculosis* is primarily an intracellular pathogen that infects host cells such as macrophages. Pathogenic mycobacteria establish and maintain their intracellular niche by influencing key host cell processes, such as autophagy<sup>20,21</sup>, trafficking<sup>22,23</sup>, signaling<sup>24,25</sup> and apoptosis<sup>26–28</sup>. In addition, systematic RNAi screens over the past years have uncovered a range of new host factors that affect intracellular mycobacterial survival<sup>21,29,30</sup>. Chemical modulation of these host cell targets is likely to affect intracellular mycobacterial survival.

We have recently published an image-based HC assay to systematically explore this target space (Fig. 2b)<sup>11</sup>. From these images, we extracted multiple parameters that capture diverse facets of the bacteria, such as intensity, number and morphology, as well as the host morphology and toxicity, as described in our recent publication<sup>11</sup>. By using this assay, we screened a library of 2,000 compounds that contain currently available drugs and other US Food and Drug Administration (FDA)-approved compounds. We used DMSO as a negative control. We then clustered the resulting data set by using a density-based clustering algorithm<sup>1</sup>. This resulted in the identification of a cluster of 115 compounds that showed a reduction in all parameters related to bacteria while having little or no effect on the host cells, indicating that these compounds were primarily detrimental to bacteria. Independent experiments revealed that the majority of these compounds were not directly toxic to bacteria, thereby indicating that they are likely

to modulate the host's response<sup>11</sup>. However, the primary assay did not provide information on the MoA of the compounds.

Experimentally, generation of the 'query data set' involves the preparation of human primary macrophages and infection with *Mycobacterium bovis* BCG-GFP (Steps 14–39). *M. bovis* BCG (Bacillus Calmette-Guérin) is an attenuated strain of *M. bovis* that is used as a vaccine against TB. In this study, we use a strain of *M. bovis* BCG that has been engineered to express GFP to enable easy detection by imaging methods<sup>31</sup>. After macrophage preparation and infection, HC screening and image analysis are performed to extract multiparametric profiles (Steps 40–56). We would like to emphasize here that the protocol quantifies intracellular survival of mycobacteria. Similar approaches could, in principle, be used for other intracellular pathogens, with appropriate modifications.

**The intermediate system.** As mycobacteria interfere with the host endolysosomal and phagolysosomal pathways<sup>22,23</sup>, compounds interfering with mycobacterial survival are likely to act in these pathways. This prompted us to search for mechanistic clues to the compound action in the endocytosis pathway.

The query and the reference systems constitute two different sets of multiparametric profiles that need to be compared in order to identify matching profiles. However, the two particular data sets used are disparate in terms of the perturbations, cell types and assays used, e.g., here the query system is a chemical screen in human primary macrophages for mycobacterial survival, whereas the reference system is an RNAi screen in HeLa cells for endocytosis. Hence, the two sets of multiparametric profiles cannot be directly compared. To circumvent this limitation, we generated an intermediate data set that encompasses crucial features of the query and the reference systems and bridges them (Fig. 1): we profiled the activities of the positive compounds from the mycobacterial screen in the HeLa cells for their effects on endocytosis<sup>11</sup>. In this chemical screen for endocytosis, we used two compounds that have known effects on the endocytic machinery, AG-1478 and nocodazole, as positive controls and DMSO as the negative control. AG-1478 is a specific inhibitor of EGF receptor tyrosine kinase and affects mostly EGF trafficking, whereas nocodazole interferes with the polymerization of microtubules and affects both EGF and transferrin trafficking. This intermediate screen of

the compounds that influence mycobacterial survival generated multiparametric profiles with respect to endocytosis in HeLa cells, thus providing the possibility of directly comparing these profiles with the corresponding ones from the genome-wide RNAi screen<sup>1</sup>. Experimental details for the generation of the intermediate system in the endocytosis assay in HeLa cells are provided in Steps 57–70.

In general, the intermediate system is not necessary if the query and reference data sets are directly comparable. However, it might not always be feasible to have the query and the reference data sets generated using the same assay at the same time because of technical and biological considerations. For example, the generation of a HC reference system is expensive and requires tedious effort. It is quite likely that a query system needs to be tailored to a pre-existing HC reference data set. Our results<sup>11</sup> suggest that this approach is possible with the use of an intermediate screen. Moreover, these data also suggest that a single HC reference data set can be used for multiple query systems through the use of suitable intermediate screens.

**Some important considerations for optimizing HC screens.**

Effective management of intraplate variability is a crucial aspect of any screening campaign. The spatial positioning of controls and library in the multiwell plates has a key role. There are two major issues that determine the choice of an appropriate layout: edge effect and the number of replicates for negative and positive controls. These issues are discussed in detail in the **Supplementary Discussion**, and some sample plate layouts are provided.

**Quantitative imaging and image analysis.** The protocol involves quantitative image analysis. The most crucial aspect of the protocol is the quality of the images. Hence, acquisition of images suitable for quantitative image analysis is of utmost importance for the generation of correct multiparametric profiles. Samples are prepared on plastic-bottomed surfaces suitable for imaging. In our case, we use CellCarrier imaging plates (384-well format).

The protocol can be modified to suit lower-throughput formats, including glass or plastic-bottom 96-well plates, cover glasses mounted on glass slides or other suitable imaging formats.

Most of our imaging is performed by using the Opera (Perkin-Elmer) automated spinning-disk confocal system. We acquire images in different channels with different cameras to avoid bleed-through between the channels. In each imaging session, along with our experimental images, we also acquire images of a fluorescent dye solution and fluorescent multicolor microbeads using the same settings. This allows us to correct the images for uneven illumination and chromatic shift. Procedures for the image corrections are detailed in the documentation provided with the software (see below). **Table 1** summarizes the details related to imaging for the two assays used here. Cameras 1–3 denote the different cameras that were used for the specific acquisition channels.

Images obtained by other automated or manual imaging systems can also be used for quantitative image analysis. We would like to stress that, irrespective of the imaging system used, careful consideration should be given to generate images that are suitable for quantitative image analysis. Care should be taken to avoid saturation of pixel intensities, to use similar microscopic settings between different imaging sessions, and to acquire images with sufficient bit depth (12 or 16 bits are better than an 8-bit image owing to the increased dynamic range).

**Software and tools required.** Quantitative analysis of a large number of images requires specialized software solutions. Several commercial and open-source tools are available. In our case, we use MotionTracking, a custom-made software platform (<http://motiontracking.mpi-cbg.de>) for the HC analysis of images and data processing. Equivalent analysis can be performed with other open-source software such as CellProfiler (<http://www.cellprofiler.org>).

MotionTracking is a software platform that is under continuous development in the Zerial laboratory at the MPI-CBG, Dresden. It is freely available at <http://motiontracking.mpi-cbg.de>. Detailed

© 2014 Nature America, Inc. All rights reserved.

**TABLE 1** | Clinical features of imaging conditions used in the assays.

Assay	Intracellular mycobacterial survival		Endocytosis		
Cell types used	Human primary macrophages		HeLa cells		
Number of images (per well)	30		12		
Z offset (µm)	1		–2.5		
Number of channels for detection	2		3		
Detection fluorophores	DAPI, CellMask blue	<i>M. bovis</i> BCG-GFP	DAPI, CellMask blue	EGF–Alexa Fluor 488	Transferrin– Alexa Fluor 647
Laser line (nm)	405	488	405	488	647
Laser power (µW)	1,130	14,300	1,130	14,300	2,510
Emission filter (nm)	450/50	520/35	450/50	540/75	700/90
Camera	Camera 3	Camera 1	Camera 3	Camera 1	Camera 2
Exposure time (ms)	600	400	400	400	400





instructions for installation and getting started are provided. Detailed documentation for the multiple steps involved in the image analysis pipeline can be found at <http://pluk.mpi-cbg.de/projects/motiontracking>. Additional interfaces are provided to visualize, manipulate and interpret data in graphical format. For data analysis, MotionTracking provides a statistical interface that can handle large data sets in multiple rows and columns and can perform complex data handling and statistical operations. Alternatively, CellProfiler is an HC analysis tool developed in the Carpenter laboratory at the Broad Institute, Cambridge<sup>32,33</sup>. It is freely available at <http://www.cellprofiler.org> and contains elaborate documentation of the various modules and options for image analysis. In addition, KNIME is an open-source software program that can be used for the analysis of HC data<sup>34</sup>. A free version is available at <http://www.knime.org>. Guidelines for combining image analysis from CellProfiler and data analysis with KNIME has been published recently<sup>35</sup>. Note that all the tools listed here provide documentation support, are intuitive and do not require coding skills.

**Description of the general pipeline used in image analysis.** The image analysis pipeline, irrespective of the software used, consists of the following distinct steps: (i) import and correction of images; (ii) automated identification of objects (segmentation); and (iii) extraction of multiple parameters.

First, images acquired from the microscopes are imported into the software to be used for image analysis according to the suggested procedures (as given in the documentation section of the respective software). The next step in the image analysis pipeline is the identification of individual objects (Fig. 3a). Objects in different channels are treated as distinct entities and identified separately. Nuclei and cells are identified in the DAPI channel based on the staining of DAPI and CellMask blue, respectively. Bacteria are identified in the 488-nm channel on the basis of the expression of GFP. For the endocytosis assay, EGF-containing endosomes are identified in the 488-nm channel and Transferrin-containing endosomes are identified in the 647-nm channel. Please note that correct object identification is crucial for subsequent steps.

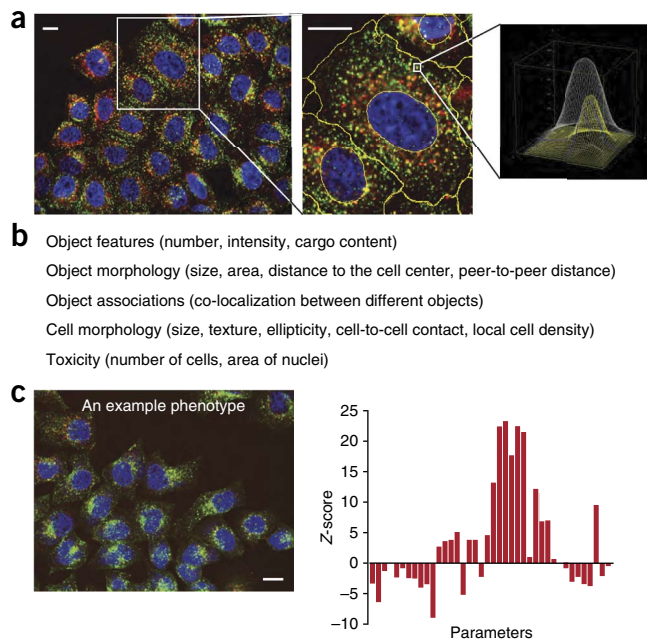
Once the objects are correctly identified, several features of the object can be extracted by choosing the appropriate parameters. These parameters include, for instance, the number of objects in the cells, their intensity, their position within the cell (distance to nucleus), their distribution with respect to each other and their association with other objects. Such multiparametric profiles are calculated for individual images. Images of single experimental conditions are combined to generate the multiparametric profile for each experimental condition (Fig. 3b,c). Statistics can be obtained at the level of either an individual image, an individual well or at the level of individual cell (this is more computationally demanding). For our experiments, we obtain statistics at the level of individual images and aggregate data from them.

**Image analysis with MotionTracking and/or CellProfiler.** We provide sample projects containing test images and the values we have used to segment the objects for analysis with both MotionTracking and CellProfiler (Supplementary Discussion, Supplementary Data 1 and 2). We emphasize that, in other imaging settings, the images will vary depending on the experimental and acquisition conditions. Hence, users have to choose the parameter settings for object analysis and the statistical filters to suit their particular images. The values provided in our pipeline are for guidance purpose only.

**Considerations for the comparison of chemical and RNAi multiparametric profiles.** Comparing chemical and RNAi multiparametric profiles to generate predictions about the intracellular MoA of compounds is an attractive concept. However, there are several significant caveats that must be taken into consideration for the approach to be successful. In this section, we discuss these concerns in detail.

• *Intrinsic differences in the nature of chemical and siRNA perturbation of cellular systems.* There are key differences in the way a chemical compound and a siRNA affect the cells. First, the effect of a compound on cells is usually rapid, whereas a siRNA acts slowly (within 48–72 h), thereby allowing the system to adjust to

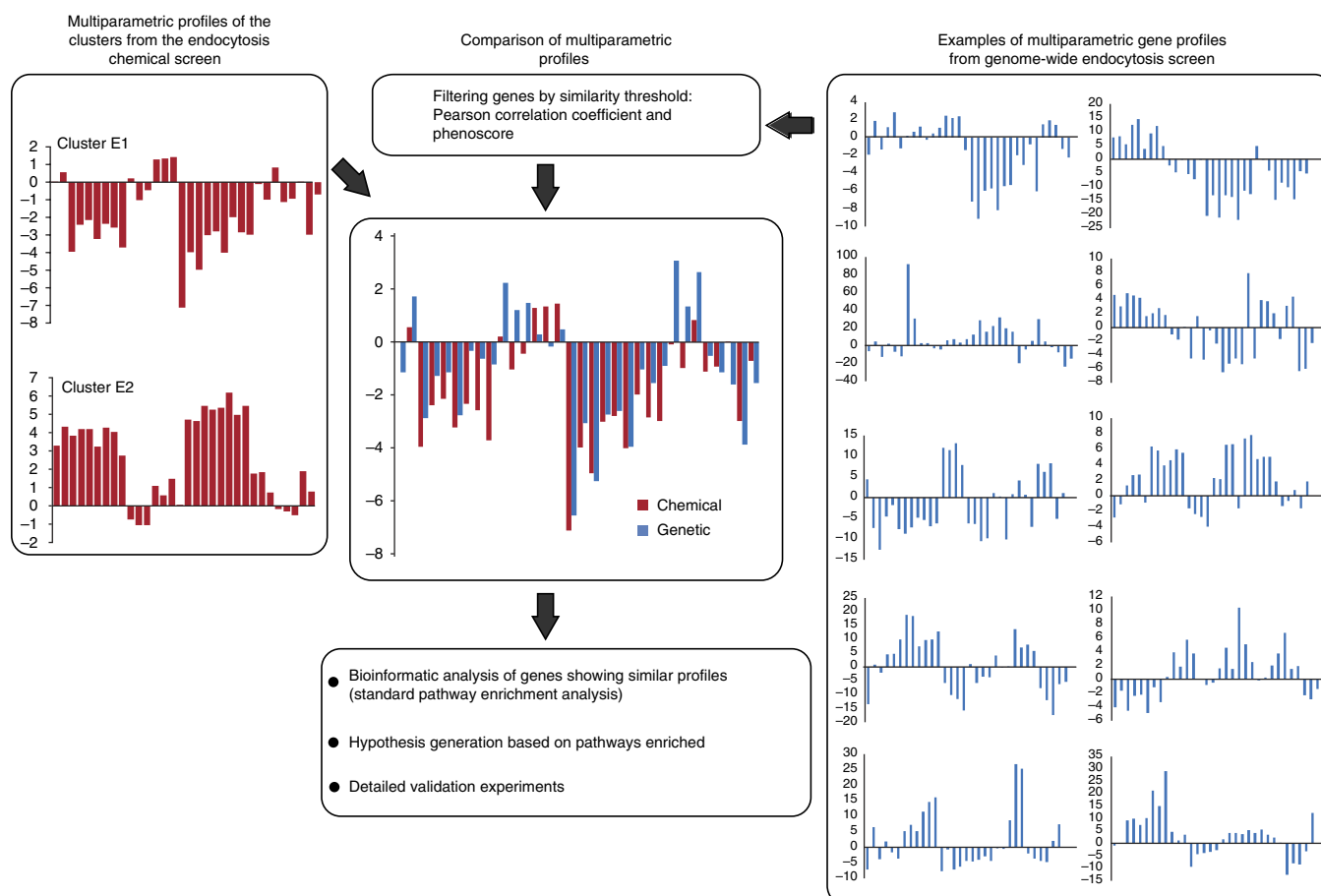
**Figure 3** | Image analysis using MotionTracking. (a) Representative image from the endocytosis assay of HeLa cells that have internalized EGF–Alexa Fluor 488 (green) and Transferrin–Alexa Fluor 647 and were stained with DAPI and SYTO42 to stain the nucleus and cytoplasm (blue), respectively (left). The central image is enlarged from the boxed region in the left and shows the contour of the cell (yellow line), as identified by automated image analysis. The right image shows the mathematical modeling of an individual endosome, marked by the box in the central image. This allows extraction of quantitative information from the images. In general, the cytoplasmic staining is very weak compared with the nuclear signal, although the intensity markedly higher than the background. It is masked here by the strong signals from the other channels. (b) A list of parameters that can be extracted from the images. Parameters reporting different features of the individual cargo, i.e., either EGF or transferrin, as well as parameters reporting the association of the cargos with each other such as colocalization, are extracted. In addition, parameters related to cell morphology and toxicity indicators are extracted. (c) A representative image from the endocytosis genome-wide RNAi screen (left panel) of a phenotype produced by a knockdown. Multiple parameters extracted from such images are represented as quantitative multiparametric profiles (right). In this illustration, each bar represents a single parameter, and all the parameters taken together represent the multiparametric profile. Scale bars, 10  $\mu$ m.



a new steady state. Second, an agonist or antagonist can bind to its molecular target within a complex without disrupting the entire complex. In contrast, silencing of a subunit of a complex can lead to disruption and even degradation of the entire complex. Third, both a compound and a siRNA can have multiple ‘off-targets’ in the cell<sup>7</sup>. Thus, even if the compound and the siRNA have the same molecular target, the off-target effects need not be conserved, resulting in differences in the multiparametric profiles produced. Fourth, the effect of a compound depends on the dose and time. In some instances, these could be nonlinear, which is mostly not the case for siRNA-mediated silencing<sup>1</sup>. Fifth, in the case of genetic screens, multiple siRNAs are typically used to target each gene. However, as responses to individual siRNAs targeting the same gene can vary significantly<sup>1</sup>, it can be difficult to estimate the multiparametric profile of the silenced gene. One possible solution to this problem was proposed in our previous genome-wide endocytosis screen<sup>1</sup>, in which the multiparametric gene profile is represented as a probabilistic estimate based on the phenotypes produced by the individual siRNAs. The gene profile estimations are stored in the database (<http://endosomics.mpi-cbg.de>) along with the individual siRNA profiles. This approach could potentially also be used with chemical compounds, where

a series of closely related chemical compounds could be used to distinguish on-target and off-target effects.

- *Use of different hardware and software in the query and reference data sets.* In our case, the hardware and software used for the generation of query and intermediate data sets<sup>11</sup> are the same as those used for the reference system<sup>1</sup>. However, generation of a robust HC reference system is a tedious and expensive endeavor. Hence, we anticipate that users might rely on existing reference data sets from the literature, and therefore it is quite likely that different hardware and software will be used to generate the query and reference data sets. Such differences might be unavoidable, and they can have a considerable effect on the multiparametric profiles generated. This effect can be minimized by using the same normalization steps for the query and intermediate data sets as are used for the reference data set. We recommend the use of the same software for normalization. In general, the closer the intermediate system is to the reference system in terms of software and normalization procedures, the better the chances of success.
- *Genotype-phenotype correlations.* Another important caveat to consider is the use of different cell types in the query versus reference screens. Our results on the use of an intermediate screen



**Figure 4** | Scheme for integration of multiparametric profiles from the mycobacteria HC chemical screen and the genome-wide reference endocytosis screen. The query profiles (E1 and E2, shown in red on left) are obtained by clustering the endocytosis multiparametric profiles of hit compounds from the mycobacteria intracellular assay. The reference genome-wide screen of endocytosis comprises gene knockdowns showing different phenotypes, as depicted by the different knockdown profiles (blue color, right). The chemical and genetic profiles are compared as shown in the middle images. Genes showing similar profiles in the query and reference system are subject to pathway enrichment analysis to generate hypothesis regarding the cellular MoA of selected compounds.

## PROTOCOL

to bridge between human primary macrophages and HeLa cells suggest that the interaction of small molecules with conserved cellular processes are likely to be similar across cell types<sup>11</sup>. The hit compounds we identified influence autophagy in both HeLa cells and human primary macrophages<sup>11</sup>. However, care should be taken while extending this to other cell types, as differences in drug effects across cell types could arise, for example, owing to differences in genotypes, metabolism and so on.

Owing to these caveats, a direct comparison of multiparametric profiles of a single compound with that of a silenced gene is not likely to yield a single direct target. To overcome this limitation, we first cluster all the compounds on the basis of their multiparametric profiles. Compounds acting on similar cellular pathways are likely to produce a similar phenotype, and hence they cluster together. We then take the average profile of the cluster of interest as our input query for comparison (Fig. 4). By doing so, we preserve the phenotype of the main targeted pathway and minimize the influence of the secondary targets of individual compounds. Next, we query the reference database for genes showing similarity to the average compound profile above a set threshold. We use a combination of two thresholds, namely Pearson correlation

and phenoscore<sup>11</sup>. Pearson correlation is a measure of similarity between two multiparametric profiles, whereas a phenoscore gives the probability of the contribution of the specific phenotypic trait (query profile) to the multiparametric profile of the reference<sup>1</sup>. We subject the list of genes thus obtained to standard bioinformatics enrichment analysis to classify the cellular pathways that are affected by the compounds in the cluster. By using this information, we formulate a hypothesis on the pathways to test for detailed validation experiments on the selected compounds from the cluster. We reiterate that every query and reference system is different and hence the factors mentioned above must be taken into consideration when extending the approach to other systems.

In general, the computational analysis involved in the generation of multiparametric profiles (Steps 55, 56 and 70) from HCS typically requires high-end computational capabilities. The timing depends on the scale of the screens and the number of images acquired and analyzed. HC analysis of 96- or 384-well imaging plates by using a standalone computer is time-consuming. We use a cluster of 1,024 Intel CPU cores for this purpose. By using this cluster, analysis of images from one 384-well plate with 15 images per well in three channels takes ~45 min with CellProfiler and ~60 min with MotionTracking.

## MATERIALS

### REAGENTS

- *M. bovis* BCG-GFP (can be obtained from different laboratories, for example, see Sundaramurthy *et al.*<sup>11</sup>, Jayachandran *et al.*<sup>24</sup> and Dhandayuthapani *et al.*<sup>31</sup>)
- Bottle-top filters, 0.22 µm (Corning, cat. no. 431118)
- Millex syringe filters, 0.22 µm (Millipore, cat. no. SLHV033RS)
- Centristar conical tubes, 15 ml (Corning, cat. no. 430791)
- CellCarrier imaging plates, 384 wells (PerkinElmer, cat. no. 600 7558)
- Deep-well plate, 384 wells, 120 µl, square wells (Nerbe, cat. no. 04.072.0120)
- Centristar conical tubes, 50 ml (Corning, cat. no. 430829)
- µClear imaging plates, 96 wells (Greiner, cat. no. 655090)
- AG-1478 (Sigma, cat. no. T4182)
- Biocoll (Biochrom, cat. no. L611)
- BSA (Roth, cat. no. 8076.2)
- Buffy coat from blood bank (e.g., Red Cross) **! CAUTION** Be sure to adhere to all relevant institutional ethics and safety guidelines when working with human blood or tissue products.
- Bacillol tissue disinfectant (Hartmann, cat. no. 975670)
- Cell culture flasks, T25 (Nunc, cat. no. 156367) and T75 (Nunc, cat. no. 156499)
- DAPI (Invitrogen, cat. no. D1306)
- DMEM (Invitrogen, cat. no. 31966-021)
- DMEM without phenol red (Invitrogen, cat. no. 31053-028)
- DMSO (Sigma, cat. no. 276855)
- DNase I (Sigma, cat. no. DN25)
- EDTA (Sigma, cat. no. E6758)
- EGF–Alexa Fluor 488 (Molecular Probes, cat. no. E13345)
- Transferrin–Alexa Fluor 647 (Molecular Probes, cat. no. T-23366)
- SYTO42, 5 mM solution in DMSO (Molecular Probes, cat. no. S-11353)
- FCS, heat inactivated (PAA Laboratories, cat. no. A15-102)
- HCS CellMask blue (Invitrogen, cat. no. H32720)
- Human monocyte colony-stimulating factor (hM-CSF; Peprotech, cat. no. 234376)
- L-Glutamine (Gibco, cat. no. 25030)
- LeucoSep tubes (Ficoll, cat. no. 227290)
- Middlebrook 7H11 agar base (Fluka, cat. no. M0428)
- Middlebrook 7H9 broth base (Fluka, cat. no. M0178)

- Middlebrook OADC growth supplement (Fluka, cat. no. M0678)
  - Kanamycin (Sigma, cat. no. K4000-5G)
  - Glycerol (VWR, cat. no. 24388.295)
  - MidiMACS LS column (Miltenyi Biotec, cat. no. 120-000-475)
  - Monocyte isolation kit (Miltenyi Biotec, cat. no. 130-091-153)
  - Nocodazole (Sigma, cat. no. M1404)
  - Nucleic acid staining solutions (see Reagent Setup)
  - Oligofectamine (Life Technologies, cat. no. 12252-11)
  - Opti-MEM (Life Technologies, cat. no. 11058-021)
  - Formaldehyde, 37% (wt/vol) solution (Merck, cat. no. 1040031000)
  - **! CAUTION** PFA is a toxic chemical. Wear suitable gloves and protective masks while making solutions and wear gloves while handling the liquid.
  - KCl (Merck Millipore, cat. no. 104936)
  - KH<sub>2</sub>PO<sub>4</sub> (Merck Millipore, cat. no. 104873)
  - Na<sub>2</sub>HPO<sub>4</sub> (Merck Millipore, cat. no. 106586)
  - NaCl (Merck Millipore, cat. no. 106404)
  - Penicillin-streptomycin (Gibco, cat. no. 10378-016)
  - RPMI 1640 (Gibco, cat. no. 61870-010)
  - Sodium azide (Sigma, cat. no. S8032) **! CAUTION** Sodium azide is a toxic chemical. Wear suitable gloves and protective masks while making solutions and wear gloves while handling the liquid.
  - Transferrin–Alexa Fluor 647 (Molecular Probes, cat. no. T-23366)
  - Trypsin–EDTA (PAA Laboratories, cat. no. L11-004)
  - RNAi libraries: we have used commercial RNAi libraries (Ambion, Qiagen) arrayed as single oligos, as well as endoribonuclease-prepared siRNA (esiRNA; generated in house)<sup>36</sup>
- ### EQUIPMENT
- Water bath (Isotemp 220, Fisher Scientific, cat. no. 15-462-21)
  - Incubator set at 37 °C, 5% CO<sub>2</sub> (HERACell 240, Thermo Scientific, cat. no. 51900293)
  - Table-top centrifuge Multifuge 3SR<sup>+</sup>, rotor no. 75006445, plate holder (Heraeus, cat. no. 75006449s)
  - Spectrometer (BioPhotomer, Eppendorf, cat. no. 6131 000.012)
  - QuadroMACS magnetic separator stand (Miltenyi Biotec, cat. no. 130-090-976) and MACS MultiStand (cat. no. 130-042-303)
  - Biomek FX (Beckman Coulter, cat. no. A31842)
  - Tecan Freedom Evo, 384-channel head (Tecan, cat. no. 612030)

- WellMate dispenser (Matrix Technology, cat. no. 5840150)
- Power washer 384 (Tecan)
- Petri dishes, 10 cm (Greiner, cat. no. 633180)
- 3M aluminum foil (Scotch Brand, Tape 431)
- Opera automatic spinning-disk confocal microscope (PerkinElmer)
- MotionTracking (<http://motiontracking.mpi-cbg.de>)
- CellProfiler (<http://www.cellprofiler.org>)
- KNIME (<http://www.knime.org>)

**REAGENT SETUP**

**PBS** Add 2 g of KCl, 2 g of KH<sub>2</sub>PO<sub>4</sub>, 11.5 g of Na<sub>2</sub>HPO<sub>4</sub> and 80 g of NaCl to 9 liters of water. Mix the solution well until all the contents have dissolved. Make up the volume to 10 liters and dispense the solution into 1-liter bottles. Autoclave PBS at 121 °C for 15 min and store it at room temperature (25 °C).

**PBS E/B** Add 0.74 g of EDTA and 0.5% (wt/vol) BSA (5 g) to 0.9 liter of PBS without Ca and Mg; mix for 10–30 min in a big beaker with a magnetic stirrer until the solution becomes clear, and then fill up to 1 liter with PBS without Ca and Mg. Filter the solution inside the laminar flow hood with a 0.2-µm filter to sterilize the solution. Dispense it into two 500-ml bottles; store one bottle at room temperature and another at 4 °C until use. These solutions should be prepared the day before the purification, and any remaining solution should not be recycled.

**BSA-coated 50-ml and 15-ml tubes** Dispense 4% (wt/vol) BSA (in PBS) in sterile 50-ml and 15-ml Centristar tubes and store them at 4 °C for at least 24 h for complete coating. When tubes are ready to use, transfer the 4% (wt/vol) BSA into a new 50-ml tube and invert the coated ones to dry before use. Coating solution can be reused for up to 3 months if it is kept at 4 °C.

**hM-CSF stock solution preparation** Add sterile distilled water to the hM-CSF powder inside the laminar flow chamber to obtain a 200 µg/ml solution; store it in 5-µl aliquots at –20 °C for 6 months.

**Monocyte differentiation medium** Add 50 ml of heat-inactivated FCS and 5 ml of penicillin-streptomycin to 500 ml of RPMI 1640. Add hM-CSF stock to the medium to obtain a final concentration of 50 ng/ml. The medium is stable for up to 1 month.

**DNase I dilution** Prepare a stock solution of 10 mg/ml in PBS. For working concentration, dilute the solution to 5 mg/ml in PBS. Filter-sterilize the solution through a 0.2-µm syringe filter. The stock solution is stored at –20 °C for up to 6 months. The working solution is freshly prepared each time.

**7H9 medium for *M. bovis* BCG-GFP culture** Suspend 4.7 g of dehydrated 7H9 powder in 900 ml of purified water. Autoclave it at 121 °C for 10 min. Wait for the bottle to cool down to 45 °C and

aseptically add 100 ml of Middlebrook OADC enrichment and appropriate antibiotic to allow selective growth of resistant bacteria (in this case, kanamycin, 25 µg/ml final concentration). Store the medium for up to 1 month at 4 °C.

**7H11 agar plates for *M. bovis* BCG-GFP culture** Suspend 10.25 g of Middlebrook 7H11 agar base in 450 ml of distilled water. Add 2.5 ml of glycerol. Boil the mixture to dissolve it completely, and then autoclave it at 121 °C for 15 min. Cool the medium to 50 °C, and then aseptically add 50 ml of OADC and 50 µl of kanamycin. Mix well and dispense the agar into sterile bacterial plates.

**HeLa cell culture medium** Add 50 ml of heat-inactivated FCS, 5 ml of L-glutamine and 5 ml of penicillin-streptomycin to 500 ml of DMEM. Store the medium at 4 °C for a maximum of 1 month.

**EGF-transferrin staining medium** Add 50 µl of EGF–Alexa Fluor 488 (200 µg/ml stock solution) and 100 µl of transferrin–Alexa Fluor 647 (5 mg/ml stock solution) to 100 ml of prewarmed DMEM (without serum, without phenol red, CO<sub>2</sub> equilibrated). This corresponds to a final concentration of 100 ng/ml for EGF–Alexa Fluor 488 and 5 µg/ml for transferrin–Alexa Fluor 647. Note that in our case, a large volume of the staining medium is needed because of the throughput used. Volumes should be scaled down to suit other throughput formats. The stock solutions are prepared at the indicated concentrations, aliquotted and stored at –20 °C, protected from light, for up to 1 year. The staining medium is freshly prepared for each experiment.

**DAPI** Resuspend DAPI in DMSO to obtain a 1 mg/ml stock solution. For working solutions, we use a final concentration of 0.4 µg/ml DAPI for HeLa cells and 1 µg/ml DAPI for human primary macrophages. The stock solution is stored at –20 °C for up to 6 months. The working solution is freshly prepared each time.

**SYTO42** SytoBlue 42 (SYTO42) is purchased as a 5 mM solution in DMSO. The final working concentration is 0.2 µM, diluted in PBS. The stock solution is stored at –20 °C for up to 6 months. The working solution is freshly prepared each time.

**HCS CellMask blue** Resuspend CellMask blue in DMSO to obtain a 10 mg/ml stock solution. The solutions are stable for up to 1 year at –20 °C. For mycobacterial assays, we use HCS CellMask Blue at a final concentration of 0.1 µg/ml.

**PFA solution, 3.7% (wt/vol)** We dilute a 37% (wt/vol) solution of formaldehyde solution in PBS to obtain a working solution of 3.7% (wt/vol) in PBS.

**Sodium azide, 0.04% (wt/vol)** Dissolve 400 mg of sodium azide in 1 liter of freshly prepared PBS. The solution can be stored at room temperature for up to 4 months.

**PROCEDURE**

**Generation of the reference data set: endocytosis assay ● TIMING 4 h**

1| Thaw HeLa cells 1 week before the experiment. Trypsinize and count the cells and prepare a cell suspension of 2.5 × 10<sup>4</sup> cells/ml for transfection.

▲ **CRITICAL STEP** We usually use cells in early passage for the experiments and never use cells above passage 10.

2| To prepare transfection complexes, for each siRNA or esiRNA mix 10 µl of diluted transfection reagent (40 µl Oligofectamine in 1 ml Opti-MEM) and 10 µl of diluted siRNA or esiRNA (200 nM siRNA or 5 ng/µl esiRNA) into an intermediate 384-well deep-well plate and mix.

3| Incubate the plate for 10 min at room temperature.

4| Transfer 10 µl of each transfection complex into a well in a 384-well imaging plate.

5| Seed 40 µl of the cell suspension into each well of the imaging plate. This will result in 1,000 cells per well transfected with 20 nM siRNA by using 0.02 µl of Oligofectamine per well of a 384-well plate.

▲ **CRITICAL STEP** Homogeneous distribution of cells in the wells is crucial for automated imaging.

6| Incubate the cells at 37 °C in 5% CO<sub>2</sub> for 72 h.





## PROTOCOL

- 7| On the day of the experiment, prewarm DMEM medium to 37 °C in 5% CO<sub>2</sub> for a minimum of 1 h, and freshly prepare EGF-transferrin staining medium.
  - 8| Take the cell culture plate from the incubator and remove the medium from the wells.
  - 9| Add 20 µl of EGF-transferrin staining medium to each well.
  - 10| Incubate the plate for 10 min at 37 °C.
  - 11| Wash the cells with PBS and fix them with 50 µl of 3.7% (wt/vol) PFA for 15 min at room temperature.
  - 12| Remove PBS and stain the nuclei and cytoplasm by incubating the fixed cells with 0.4 µg/ml DAPI and 0.2 µM SYTO42 for 1 h. Without removing the dyes, add 45 µl of PBS containing 0.04% (wt/vol) sodium azide to each well (to prevent fungal contamination), and then seal the plate with aluminum tapes. The plate is now ready for imaging, and it can be stored at this point at 4 °C until image acquisition.
    - ▲ **CRITICAL STEP** Leaving the dye in the wells increases the stability of staining.
    - **PAUSE POINT** We have stored stained plates for up to 6 months at 4 °C with no detectable loss of signal.
  - 13| Image analysis and extraction of multiparametric profiles: acquire images in four channels (**Table 1**) by using the automated microscope (Opera), and then analyze images as outlined in the Experimental design section. The HC data set from the genome-wide endocytosis screen is available publicly at <http://endosomics.mpi-cbg.de>. The organization of this data set is provided in **Supplementary Figure 1**.
    - ? **TROUBLESHOOTING**
- Generation of the query data set: preparation of human primary macrophages** ● **TIMING 5–7 d**
- ▲ **CRITICAL** The generation of the query data set with the examples used here includes the preparation of human primary macrophages (Steps 14–39), intracellular mycobacterial survival assay (Steps 40–54) and image analysis and extraction of query multiparametric profiles (Steps 55 and 56).
- 14| Collect buffy coat samples from a blood bank (e.g., the Red Cross). Buffy coat is derived from centrifugation of blood collected from volunteer donors and is enriched in leukocytes. For our mycobacteria chemical screens, we normally isolate PBMCs from four individual buffy coat samples and pool them.
    - ▲ **CRITICAL STEP** For our experiments, we isolate monocytes and derive them to macrophages in 384-well plates that can be directly infected and imaged. We prefer this approach, as cells do not have to be removed from the plates once they have adhered. The protocol can be suitably scaled to other lower-throughput formats such as 96- or 24-well plates.
  - 15| Sterilize the outside of the bag containing buffy coat with disinfectant spray, cut the tube with a sterile scalpel and dispense the contents into a 50-ml tube. Do not mix different buffy samples in single tube.
  - 16| Centrifuge the tubes at 400g for 40 min at 20 °C, with acceleration 0 and deceleration 0.
    - ▲ **CRITICAL STEP** It is important to accelerate and decelerate very slowly. If not, the quality of monocytes will be compromised. This step takes ~1.5 h.
  - 17| In the meantime, prepare the LeucoSep tubes for the next step. These tubes contain a retaining ring that stabilizes the position of the porous filter, which enables optimal separation of leukocytes by preventing recontamination with red blood cells (RBCs). Add 16 ml of Biocoll to each of the LeucoSep tubes and centrifuge them for 3 min at 1,000g at room temperature. Remove the Biocoll from the upper part of the tube while leaving ~250 µl in the tube on top of the filter.
  - 18| After the centrifugation at Step 16, leukocytes will form a layer on top of the filter, whereas the RBCs will have passed through. Take up all leukocytes with a 10-ml pipette, collecting ~10–15 ml, transfer them to a new 50-ml tube and make up the volume to 35 ml with PBS E/B at room temperature.
    - ▲ **CRITICAL STEP** A small amount of RBCs will be taken up along with leukocytes. This is not a problem, as they will be removed in the next step.
  - 19| Add 350 µl of DNase I and mix it carefully by inverting the tube.

**20|** Transfer the cells carefully into the prepared LeucoSep tube. Centrifuge the tube for 20 min at 1,000*g* at 20 °C, acceleration 0 and deceleration 0. Carefully inspect the tube after the centrifugation: the ring layer contains PBMCs and it should 'sit' on the filter in the tube while the RBCs will pass the filter and the bottom part.

**▲ CRITICAL STEP** Slow acceleration and deceleration is necessary at this step. The total centrifugation takes ~45 min.

**▲ CRITICAL STEP** After this step, cool the centrifuge to 4 °C.

**▲ CRITICAL STEP** From this point up to Step 35, work with the ice-cold PBS E/B that was stored at 4 °C (Reagent Setup).

**21|** Carefully transfer ~8–10 ml of the ring layer into a 50-ml BSA-coated tube. Use a 1-ml pipette and avoid taking up any RBC.

**22|** Make up to 35 ml with precooled PBS E/B, add 350 µl of DNase I and mix carefully by inversion. Centrifuge the tube for 10 min at 400*g* at 4 °C, with maximal acceleration and deceleration. In the meantime, prepare 400 ml of PBS E/B + 800 µl of DNase I and put it on ice.

**▲ CRITICAL STEP** Change the acceleration and deceleration settings of the centrifuge, as from here on all centrifugations are performed with maximal acceleration and deceleration.

**23|** Use a Pasteur pipette to remove the supernatant.

**24|** Resuspend the pellet carefully in 1 ml of the precooled PBS E/B + DNase I.

**25|** Make up to 35 ml with precooled PBS E/B + DNase I. Centrifuge the mixture for 8 min at 4 °C at 200*g*.

**26|** Repeat Steps 23–25 and take a 100-µl aliquot of the 35-ml cell suspension for cell counting. Count the cells while the centrifugation is on.

**27|** After washing, remove the supernatant and resuspend the cells in 1 ml of PBS E/B.

**28|** At this point, pool the cells from different donors such that equal cell numbers are taken from each donor. In the following steps, 200 million cells are used. Hence, take 50 million cells from each donor for the pool.

**▲ CRITICAL STEP** The cell numbers used depend on the throughput employed, and they have to be modified accordingly. In our case, we need ~5 million purified monocytes per 384-well plate.

**29|** Label the cells with the antibody cocktail from the monocyte isolation kit II, according to the manufacturer's protocol. This is detailed in Steps 30–39.

**30|** Take 200 million cells from the pool and resuspend them in 300 µl of PBS E/B. Add 100 µl of FcR blocking reagent (vial 1 from the monocyte isolation kit II), mix it slowly, and then add of 100 µl of biotin antibody cocktail (vial 2) and mix it slowly. Incubate the mixture for 10 min at 4–8 °C (on ice).

**▲ CRITICAL STEP** This cocktail contains antibodies for surface markers for cell types other than monocytes. Hence, this is a negative selection, which removes all cells but monocytes.

**31|** Add 300 µl of PBS E/B (including DNase I) and mix slowly, and then add 200 µl of anti-biotin microbeads (vial 3) and mix slowly. Incubate the mixture for 15 min at 4–8 °C (on ice).

**32|** Resuspend the cells in 10 ml of PBS E/B + DNase I and centrifuge them for 8 min at 4 °C at 200*g*.

**33|** Separation with Quatro MACS: remove the supernatant and resuspend the pellet in 500 µl of PBS E/B + DNase I. Assemble the magnetic column on the separator according to the manufacturer's protocol.

**34|** Equilibrate the column with 3 ml of PBS E/B + DNase I.

**35|** Load the cells and collect the flow-through into a 15-ml BSA-coated polypropylene tube. Wash the columns three times with 3 ml of PBS E/B + DNase I and collect the flow-through into the same BSA-coated polypropylene tube.

**▲ CRITICAL STEP** The flow-through contains the nonlabeled monocytes. The other cell populations are labeled with magnetic beads and are retained in the column.

## PROTOCOL

- 36| Centrifuge the collected flow-through for 10 min at 4 °C at 300g. Resuspend the cells in 1 ml of differentiation medium.
- 37| Count the cells and dilute them appropriately in the differentiation medium; seed them at a density of 15,000 cells per well in 384-well imaging plates or at 60,000 cells per well in 96-well imaging plates.
- ▲ **CRITICAL STEP** Homogeneous distribution of cells in the wells is crucial for automated imaging.
- 38| After 3 d, replace the medium with fresh differentiation medium.
- 39| Cells differentiate to macrophages 5–7 d after monocyte isolation. We typically use them on the seventh day after monocyte isolation. Cells can be kept in culture for up to 2 weeks, provided the medium is refreshed every 3 d.

### Intracellular mycobacterial survival assay ● **TIMING 2 d**

40| Thaw *M. bovis* BCG-GFP from stock ~10 d before infection, and maintain it in T25 flasks in liquid 7H9 medium containing 10% (vol/vol) OADC and 25 µg/ml kanamycin.

41| Monitor the growth of mycobacteria by measuring the OD<sub>600</sub>. Cells are maintained at OD<sub>600</sub> between 0.5 and 0.8.

▲ **CRITICAL STEP** Mycobacteria tend to clump in culture. For optimal infections, it is important to minimize the clumping. The culturing and infection conditions mentioned here ensure that the clumping is kept at a low level.

42| On the day of infection, freshly prepare the mycobacteria infection medium: transfer ~30 ml of bacteria to a 50-ml polypropylene tube and keep them at room temperature for 30 min to allow any bacterial clumps to sediment.

▲ **CRITICAL STEP** Steps 42–44 are designed to avoid bacterial clumps in the infection mix.

43| Collect ~90% of the suspension carefully from the top and transfer it to a new 50-ml polypropylene tube.

44| Centrifuge the tube at 30g for 10 min at room temperature. Collect the supernatant carefully to a new tube. Centrifuge this at 1,000g for 10 min at room temperature. Discard the supernatant and wash the pellet three times with RPMI containing 10% (vol/vol) FCS, without any antibiotics.

▲ **CRITICAL STEP** The RPMI medium used in Steps 44–49 should not contain any antibiotics.

45| Dilute the bacterial suspension to a final OD<sub>600</sub> of 0.06. This constitutes the infection mix. Put 5 µl of the mix on a glass slide and observe it under the microscope to ensure that there are no large clumps of bacteria present.

46| Before infection, remove the medium from the 384-well plate containing monocyte-derived macrophages (Steps 37–39) and wash it with fresh RPMI + 10% (vol/vol) FCS.

47| Infect the macrophages by adding 50 µl per well of the mycobacteria infection medium. This corresponds to a multiplicity of infection (MOI) of 10. Transfer the plate to the incubator at 37 °C and incubate it for 1 h.

▲ **CRITICAL STEP** The MOI must be tested for individual experiments. This can be done rapidly by counting the number of bacteria per unit volume of infection medium by using a cell culture fluorescence microscope. A more accurate way is to plate the infection mix on 7H11 agar plates and count the number of colonies after incubation at 37 °C for 3 weeks.

48| Wash the plate 3–4 times with fresh RPMI + 10% (vol/vol) FCS without any antibiotics.

▲ **CRITICAL STEP** This washing step is crucial to remove all extracellular bacteria. We wash each well extensively and confirm the absence of extracellular bacteria by observing the plates under the microscope before proceeding to the next step.

49| Add compounds from the compounds library to a final concentration of 10 µM and incubate the plate for 48 h at 37 °C.

▲ **CRITICAL STEP** The concentration of the compounds used in the screen is optimized to ensure uniformity in stock preparations and liquid handling. However, different compounds affect the system at different concentrations, and hence detailed dose-response studies must later be done on the selected hit compounds.

50| Fix the cells by adding 50 µl per well of 7.4% (vol/vol) PFA for 20 min. The final concentration of PFA is 3.7%.

51| Wash the plate with PBS to remove the PFA.

52| Remove the PBS and incubate the cells with 50  $\mu$ l of the staining solution (1  $\mu$ g/ml DAPI with 0.1  $\mu$ g/ml CellMask blue) for 30 min at room temperature. These dyes stain the nucleus and cytoplasm, respectively. This allows for automated detection of the cells and accurate enumeration of intracellular bacteria.

53| Add 50  $\mu$ l of PBS containing 0.04% (wt/vol) sodium azide (to prevent fungal contamination) on top and seal it with sealing foil.

54| The plate is now ready for imaging, and it can be stored at this point at 4  $^{\circ}$ C for up to 1 week until image acquisition.

#### Image analysis and extraction of query multiparametric profiles ● TIMING 2–5 weeks

55| Acquire images with the Opera automated microscope, by using the imaging conditions described in **Table 1**, and then analyze them with MotionTracking and/or CellProfiler software, as outlined in the Experimental design section.

#### ? TROUBLESHOOTING

56| After image analysis, select hit compounds by using a density-based clustering algorithm (ref. 1, and N. Samusik and Y.K., unpublished data). Users can adopt different methods, such as *k*-means clustering with the silhouette index, for *k* selection to identify clusters of hit compounds.

▲ **CRITICAL STEP** Individual compound hits must be validated with different experimental tools to confirm their effect. This protocol details the methods used to identify the cellular MoA of compounds that influence intracellular bacterial survival by modulating the host. Hence, the selected compounds must also be tested on the bacteria directly to check for direct antibacterial effects.

#### Generation of the intermediate data set endocytosis chemical screen ● TIMING 4 h

▲ **CRITICAL** The generation of the intermediate data set, with the examples used here, includes the endocytosis chemical screen (Steps 57–69) and imaging, analysis and extraction of intermediate multiparametric profiles (Step 70).

57| Thaw HeLa cells 1 week before the experiment.

▲ **CRITICAL STEP** We usually use cells in early passage for the experiments and never use cells above passage 10.

58| Three days before the experiment, split the cells and seed 500 cells per well in a volume of 50  $\mu$ l in 384-well plates. This cell density is optimal for automated imaging.

▲ **CRITICAL STEP** Homogeneous distribution of cells in the wells is crucial for automated imaging.

59| On the day of the experiment, prewarm DMEM medium at 37  $^{\circ}$ C in 5% CO<sub>2</sub> for a minimum of 1 h.

60| Add the compounds from the compounds library to a final concentration of 10  $\mu$ M to the cell culture plate. Incubate the plate for 2 h at 37  $^{\circ}$ C.

61| Mix 22.5  $\mu$ l of the EGF-transferrin staining medium with 2.5  $\mu$ l of 100  $\mu$ M compound (in DMEM). This creates an EGF-transferrin staining medium containing the compound at a 10  $\mu$ M final concentration.

62| Take the cell culture plate from the incubator, and remove the medium from the wells.

63| Add the EGF-transferrin staining medium containing the compounds to be tested.

64| Incubate the plate for 10 min at 37  $^{\circ}$ C.

65| Wash the plate with PBS and fix with 3.7% (vol/vol) PFA for 15 min at room temperature.

66| Remove PFA from the wells and wash the plate three times with PBS.

67| Stain the cells with 50  $\mu$ l of 1  $\mu$ g/ml DAPI and 0.1  $\mu$ g/ml CellMask blue for 30 min at room temperature.

68| Fill the wells with 50  $\mu$ l of PBS containing 0.04% (wt/vol) sodium azide (to prevent fungal contamination) and seal the plate with aluminum tape.



## PROTOCOL

69| The plate is now ready for imaging and can be stored at this point at 4 °C until image acquisition.

■ **PAUSE POINT** Plates can be stored 4 °C for up to 3 months without any apparent loss of signal intensities.

### Imaging, analysis and extraction of intermediate multiparametric profiles ● **TIMING 2–5 weeks**

70| Acquire images in four channels (**Table 1**) with the Opera automated microscope, and then analyze them as outlined in the Experimental design section.

#### ? **TROUBLESHOOTING**

### Comparison of chemical and genetic multiparametric profiles ● **TIMING 30 min–1 d**

▲ **CRITICAL** Perform Steps 71–75 if the genome-wide endocytosis data set<sup>1</sup> is used as your reference data set (timing: 30 min). Perform Steps 76–83 if you use a different HC reference data set (timing: 1 d).

71| The procedure so far described the generation of the reference, query and intermediate data sets by using the specific assays presented, i.e., endocytosis and intracellular mycobacterial survival assay. Users who want to compare multiparametric profiles generated with the intermediate endocytosis assay with the genome-wide endocytosis screen should proceed to Step 72. Alternatively, users can also import a completely different external reference data set. The requirement is that this data set should have sufficient parameter space (**Supplementary Discussion**). If you are using an external reference data set, go to Step 77.

72| We have generated a web interface (<http://endosomics.mpi-cbg.de/gws/search>) to compare multiparametric profiles. This interface is specific for the endocytosis genome-wide screen as a reference data set.

73| Enter the query data set into the search engine by selecting either the ‘graph’ or the ‘text’ option. In the ‘graph’ option, alter the red curve in the interactive graphical interface by dragging it along the y axis for each parameter to generate a semiquantitative multiparametric ‘query’ profile (**Supplementary Fig. 2a**).

74| If you are using the ‘text’ option, enter (or copy and paste) normalized values for the parameters into the space provided.

▲ **CRITICAL STEP** By default, all parameters listed will be included in the analysis. In case the query data set consists of only a subset of these parameters, it is very important to unselect the parameters that have not been computed and include only the relevant ones (**Supplementary Fig. 2b**).

▲ **CRITICAL STEP** When the ‘text’ option is used to input a query data set, the ordering of the parameters is important. The parameters list provided in the interface (**Supplementary Fig. 2b**) defines the order of the parameters. As the parameters of the genome-wide endocytosis reference data set are arranged in this order, it is imperative for the parameters of the query data set to have the same order.

▲ **CRITICAL STEP** For either the ‘graph’ or ‘text’ option, the input query data set must be normalized with respect to the controls of the assay and provided as z scores. Details on the normalization procedures are provided in the **Supplementary Discussion**. A sample data set from the genome-wide endocytosis screen containing normalized values for all the parameters is provided in **Supplementary Data 3**.

75| Provide the correlation and/or phenoscore thresholds required to compute the similarity of the multiparametric profiles (**Supplementary Fig. 2c**). These values range from 0 to 1, and they refer to the stringency criteria used to filter similar genes. In general, the greater the number of parameters, the lesser the probability that two nonsimilar profiles match by chance. The relationships between correlation and the number of parameters are explained in the **Supplementary Discussion**.

76| The search for similar genes might take up to 15 min. Once the search is completed, results are provided as a downloadable .csv file that contains the gene name, gene id and the values for phenoscore and correlation of the individual genes with the query data set.

77| If a different reference system not related to endocytosis is used, we provide tools in MotionTracking to allow users to compare it with the query data set. In order to do this, first import the reference data set into the local version of MotionTracking by following the steps: ‘Statistic’ → ‘Batch Statistic’ → ‘Open Batch Statistic’. This opens the ‘Image Set Analysis’ window. In this window, follow the path ‘File’ → ‘Import’ → ‘Import Table from CSV’ (**Supplementary Fig. 3a**).

▲ **CRITICAL STEP** The first row of the user-generated reference data set should contain the parameter names, and the first column should contain the gene names. As an example of the organization of the reference data set, we provide a subset of the multiparametric profiles from the genome-wide screen of endocytosis (**Supplementary Data 3**).

▲ **CRITICAL STEP** Users can upload a normalized .csv file or use the normalization features in the 'Image Set Analysis' window of MotionTracking. Details of the normalization procedures are provided in the **Supplementary Discussion**.

**78|** Import the normalized profiles of the query data set (or the intermediate data set) to compare it with the reference data set by following a two-step process. First, follow the steps 'Image Set → Analysis' → 'Profile 'Profile to User-Defined Profile' (**Supplementary Fig. 3b**). This opens a new pop-up window called 'condition list'.

**79|** Fill the query parameters in the pop-up window. Alternatively, a set of comma/space/tab-separated values could be copy-pasted in the first cell of the 'Condition List' table. After pressing enter, the set will be parsed and the values will be distributed to the selected parameters. By pressing 'OK', a new pop-up window called 'Profile Correlation' appears (**Supplementary Fig. 3c**).

**80|** Select 'Use Covariation Matrix'. Ensure the correct calculation of the covariation matrix on the basis of the selected controls in the reference data set (**Supplementary Discussion**). Enter the name of the covariation matrix and the appropriate control from the reference data set (such as 'Mock') and select 'OK'. This generates a new column in the reference data set in the 'Image Set Analysis' window containing the correlations for each element of the reference data set in relation to the query data set.

**81|** The second step is to calculate the phenoscores of the query (or intermediate) data sets with the reference data set. To do this, follow the steps 'Image Set Analysis' → 'Phenoscore Correlation' → 'Phenoscore Correlation to User-Defined Profile' (**Supplementary Fig. 3d**). This opens a new pop-up window called 'Condition list'.

**82|** Input the query (or intermediate) profile as above (Step 79) and press OK. This opens a new pop-up window called 'Phenoscore' (**Supplementary Fig. 3c**). Select the 'Use covariation matrix'. Provide the name of the control of the reference data set for calculation of the covariation matrix. This is similar to Step 80. This step generates a new column in the data set containing the correlation of the query (or intermediate) profile with each element of the reference data set.

▲ **CRITICAL STEP** By default, all the parameters provided in the reference data set are included in the pop-up windows. Users have to deselect the parameters that are not present in the query data set.

**83|** The resulting list contains the phenoscore and correlation values of all genes in the reference data set to the query (or intermediate profile). This table can be sorted and analyzed by MotionTracking Image Set Analysis tools (see the software's documentation). Alternatively, the table could be exported as a .csv file for analysis by third-party tools by selecting 'Menu' → 'File' → 'Export' → 'Export Data to CSV'. MotionTracking will offer to choose the naming of genes in the export file (e.g., Gene Name, Gene ID, siRNA and so on, depending on the input details provided by the user). Next, MotionTracking asks for the parameters to export. Choose phenoscore and correlation, in addition to any other parameters necessary. After this procedure, the user will have a .csv file of the list of genes from the reference set with the correlation and phenoscore for each gene relative to the query profile. This list can be analyzed in Excel to select for genes showing a specific correlation and phenoscore thresholds.

▲ **CRITICAL STEP** Genes showing similarity to the query (or intermediate) profile could be selected on the basis of their correlation and phenoscore values by sorting the table and selecting for genes above a defined similarity threshold value. The values range from 0 to 1: the closer the values are to 1, the higher the similarity of the query profile to the gene profile.

## ? TROUBLESHOOTING

### Bioinformatics analysis of similar genes ● TIMING 30 min–1 h

**84|** Feed the list of genes to any standard bioinformatic pathway enrichment program to get the cellular pathways enriched by the set of genes chosen. We have used the DAVID bioinformatic resource for this purpose (<http://david.abcc.ncifcrf.gov>)<sup>37</sup>.

## ? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 2**.

**TABLE 2** | Troubleshooting table.

Step	Problem	Possible reason	Possible solution
13, 55, 70	Many images have no cells, or cells are stacked on top of each other	Nonhomogeneous distribution of cells in the well	After cell seeding, keep the plate at room temperature without moving it for 15–30 min
	Systematic bias in measurements that are dependent on the position of the wells in the plate (for example, outer wells having fewer cells than the inner wells)	Edge effects in plates	Optimal plate layout is important. We have provided some generic guidelines for this in the <b>Supplementary Discussion</b>
	Images are fuzzy	Out-of-focus images	Optimize appropriate settings for the hardware autofocus controls Check for proper plate positioning in the microscope during acquisition. Uneven positioning of plates in the plate holder could lead to out-of-focus effects
	Imaging signals from one channel interfering with another	Bleed-through between channels	Take care to ensure correct microscopy settings and fluorophore choice. We use different cameras for the different acquisition channels to minimize bleed-through. <b>Table 1</b> provides the details of our imaging conditions
	Images are brighter in the central part of the image compared with the edges	Uneven illumination	Apply appropriate correction using homogeneous fluorescent dye solutions, as recommended by the software used
55, 70	Systematic <i>x</i> – <i>y</i> shift between signals from different channels	Chromatic shift between channels	Apply appropriate correction with the use of multicolor fluorescent beads, as recommended by the software used
	Interplate variations	Differences in cell culture between the experiments	Use a highly standardized cell culture procedure between different experiments. For example, use cells of the same passage number and the same cell viability, and follow a defined time table for culturing and splitting the cells
	Objects are not identified correctly	Incorrect segmentations	Optimize the correct parameters for segmentation for each experiment. These will vary depending on differences in hardware settings, fluorescence intensity, and so on
83	Poor correlation between chemical and genetic multiparametric profiles	Determinant of covariation matrix is close to zero	Recalculate the covariation matrix with control sets of sufficient size (more than the number of parameters). As a rule of thumb, the number of controls should be at least three times the number of parameters As an alternative, use identity covariation matrix See <b>Supplementary Discussion</b>
	Poor correlation between chemical and genetic multiparametric profiles	Incorrect normalization	Follow correct normalization procedures as described in the <b>Supplementary Discussion</b>

● **TIMING**

Steps 1–13, endocytosis assay: 4 h

Steps 14–39, preparation of human primary macrophages: 5–7 d

Steps 40–54, intracellular mycobacterial survival assay: 2 d

Steps 55 and 56, image analysis and extraction of query multiparametric profiles: 2–5 weeks

Steps 57–69, endocytosis chemical screen: 4 h

Step 70, imaging, analysis and extraction of intermediate multiparametric profiles: 2–5 weeks



Steps 71–83, comparison of chemical and genetic multiparametric profiles: 30 min–1 d (depending on the reference data set used)

Step 84, bioinformatics analysis of similar genes: 30 min–1 h

## ANTICIPATED RESULTS

Clustering of the HC chemical screens (query or intermediate) should produce different profiles that are likely to be indicative of different mechanisms of action of the compounds, as shown in **Figure 4** (clusters E1 and E2). In this example, clusters E1 and E2 are derived from the intermediate data set of the HC chemical endocytosis screen of the hit cluster from the intracellular mycobacterial survival screen<sup>11</sup>. The multiparametric profile of a particular cluster represents the average profile of all the compounds that are present in it. This is used for comparison with the reference genetic data set. After this comparison, highly correlated genes are selected. A typical example of such correlated profiles is provided in **Figure 4**, middle panel.

The end result of the procedure is the generation of a set of genes that show similar multiparametric profiles to the input chemical query. The output is in the form of a .csv file with four columns, including the name of the gene, the Entrez gene ID, the phenoscore and the correlation of the gene with respect to the chemical. Users are expected to select the genes on the basis of appropriate threshold values for phenoscore and correlation, and to use this as input into a bioinformatic resource (such as DAVID; Step 83) to identify the relevant cellular pathways.

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

**ACKNOWLEDGMENTS** We thank J. Pieters (Biozentrum, Basel) for *M. bovis* BCG-GFP strain and A. Roesen-Wolff (Technische Universität Dresden (TUD)) for the help with monocyte isolation. We are indebted to High-Throughput Technology Development Studio (HT-TDS), MPI-CBG for expert technical assistance, to the High-Performance Computing Center at TUD and the Computer Department for outstanding IT support and to H. Nonaka for critical comments on the manuscript. The work was supported by the EU-FP7-funded projects NATT, PHAGOSYS and APO-SYS, and by the Max Planck Society.

**AUTHOR CONTRIBUTIONS** V.S. and M.Z. conceived the project; V.S., R.B., M.S. and N.T. developed and performed the experimental steps; M.C. and Y.K. developed the computational tools and performed the analysis; and M.B. and M.Z. coordinated and supervised the project.

**COMPETING FINANCIAL INTERESTS** The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Collinet, C. *et al.* Systems survey of endocytosis by multiparametric image analysis. *Nature* **464**, 243–249 (2010).
- Danuser, G. Computer vision in cell biology. *Cell* **147**, 973–978 (2011).
- Stockwell, B.R. Chemical genetics: ligand-based discovery of gene function. *Nat. Rev. Genet.* **1**, 116–125 (2000).
- Bickle, M. The beautiful cell: high-content screening in drug discovery. *Anal. Bioanal. Chem.* **398**, 219–226 (2010).
- Young, D.W. *et al.* Integrating high-content screening and ligand-target prediction to identify mechanism of action. *Nat. Chem. Biol.* **4**, 59–68 (2008).
- Swinney, D.C. & Anthony, J. How were new medicines discovered? *Nat. Rev. Drug Discov.* **10**, 507–519 (2011).
- Merino, A., Bronowska, A.K., Jackson, D.B. & Cahill, D.J. Drug profiling: knowing where it hits. *Drug Discov. Today* **15**, 749–756 (2010).
- Cong, F., Cheung, A.K. & Huang, S.M. Chemical genetics-based target identification in drug discovery. *Annu. Rev. Pharmacol. Toxicol.* **52**, 57–78 (2012).
- Stockwell, B.R. Exploring biology with small organic molecules. *Nature* **432**, 846–854 (2004).
- Schenone, M., Dancik, V., Wagner, B.K. & Clemons, P.A. Target identification and mechanism of action in chemical biology and drug discovery. *Nat. Chem. Biol.* **9**, 232–240 (2013).
- Sundaramurthy, V. *et al.* Integration of chemical and RNAi multiparametric profiles identifies triggers of intracellular mycobacterial killing. *Cell Host Microbe* **13**, 129–142 (2013).

- Kittler, R. *et al.* Genome-scale RNAi profiling of cell division in human tissue culture cells. *Nat. Cell Biol.* **9**, 1401–1412 (2007).
- Neumann, B. *et al.* Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* **464**, 721–727 (2010).
- Pelkmans, L. *et al.* Genome-wide analysis of human kinases in clathrin- and caveolae/raft-mediated endocytosis. *Nature* **436**, 78–86 (2005).
- Paul, P. *et al.* A Genome-wide multidimensional RNAi screen reveals pathways controlling MHC class II antigen presentation. *Cell* **145**, 268–283 (2011).
- Yin, Z. *et al.* A screen for morphological complexity identifies regulators of switch-like transitions between discrete cell shapes. *Nat. Cell Biol.* **15**, 860–871 (2013).
- WHO. Global Tuberculosis Report. (World Health Organization, 2013).
- Zumla, A., Nahid, P. & Cole, S.T. Advances in the development of new tuberculosis drugs and treatment regimens. *Nat. Rev. Drug Discov.* **12**, 388–404 (2013).
- Goodwin, P.J., Meyerhardt, J.A. & Hursting, S.D. Host factors and cancer outcome. *J. Clin. Oncol.* **28**, 4019–4021 (2010).
- Gutierrez, M.G. *et al.* Autophagy is a defense mechanism inhibiting BCG and *Mycobacterium tuberculosis* survival in infected macrophages. *Cell* **119**, 753–766 (2004).
- Kumar, D. *et al.* Genome-wide analysis of the host intracellular network that regulates survival of *Mycobacterium tuberculosis*. *Cell* **140**, 731–743 (2010).
- Armstrong, J.A. & Hart, P.D. Response of cultured macrophages to *Mycobacterium tuberculosis*, with observations on fusion of lysosomes with phagosomes. *J. Exp. Med.* **134**, 713–740 (1971).
- Russell, D.G. *Mycobacterium tuberculosis*: here today, and here tomorrow. *Nat. Rev. Mol. Cell Biol.* **2**, 569–577 (2001).
- Jayachandran, R. *et al.* Survival of mycobacteria in macrophages is mediated by coronin 1-dependent activation of calcineurin. *Cell* **130**, 37–50 (2007).
- Koul, A., Herget, T., Klebl, B. & Ulrich, A. Interplay between mycobacteria and host signalling pathways. *Nat. Rev. Microbiol.* **2**, 189–202 (2004).
- Behar, S.M., Divangahi, M. & Remold, H.G. Evasion of innate immunity by *Mycobacterium tuberculosis*: is death an exit strategy? *Nat. Rev. Microbiol.* **8**, 668–674 (2010).
- Chen, M., Gan, H. & Remold, H.G. A mechanism of virulence: virulent *Mycobacterium tuberculosis* strain H37Rv, but not attenuated H37Ra, causes significant mitochondrial inner membrane disruption in macrophages leading to necrosis. *J. Immunol.* **176**, 3707–3716 (2006).
- Keane, J., Remold, H.G. & Kornfeld, H. Virulent *Mycobacterium tuberculosis* strains evade apoptosis of infected alveolar macrophages. *J. Immunol.* **164**, 2016–2020 (2000).
- Agaisse, H. *et al.* Genome-wide RNAi screen for host factors required for intracellular bacterial infection. *Science* **309**, 1248–1251 (2005).



30. Philips, J.A., Rubin, E.J. & Perrimon, N. *Drosophila* RNAi screen reveals CD36 family member required for mycobacterial infection. *Science* **309**, 1251–1253 (2005).
31. Dhandayuthapani, S. *et al.* Green fluorescent protein as a marker for gene expression and cell biology of mycobacterial interactions with macrophages. *Mol. Microbiol.* **17**, 901–912 (1995).
32. Carpenter, A.E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
33. Kametsky, L. *et al.* Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics* **27**, 1179–1180 (2011).
34. Berthold, MR *et al.* *KNIME: The Konstanz Information Miner* (Springer, 2007).
35. Stoter, M. *et al.* CellProfiler and KNIME: open source tools for high content screening. *Methods Mol. Biol.* **986**, 105–122 (2013).
36. Kittler, R. *et al.* Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies. *Nat. Methods* **4**, 337–344 (2007).
37. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).