Next-Generation Phenotypic Screening in Early Drug Discovery for Infectious Diseases

Nathalie Aulner, Anne Danckaert, Jongeun Ihm, David Shum, Spencer L Shorte

Trends in Parasitology

CellPress
REVIEWS

Review

Next-Generation Phenotypic Screening in Early Drug Discovery for Infectious Diseases

Nathalie Aulner,[1] Anne Danckaert,[1] JongEun Ihm,[1] David Shum,[2] and Spencer L. Shorte [1,2,*]

Cell-based phenotypic screening has proven to be valuable, notably in recapitulating relevant biological conditions, for example, the host cell/pathogen niche. However, the corresponding methodological complexity is not readily compatible with high-throughput pipelines, and fails to inform either molecular target or mechanism of action, which frustrates conventional drug-discovery roadmaps. We review the state-of-the-art and emerging technologies that suggest new strategies for harnessing value from the complexity of phenotypic screening and augmenting powerful utility for translational drug discovery. Advances in cellular, molecular, and bioinformatics technologies are converging at a cutting edge where the complexity of phenotypic screening may no longer be considered a hinderance but rather a catalyst to chemotherapeutic discovery for infectious diseases.

Highlights

The conventional wisdom of the drug-development roadmap strategy that seeks to establish therapeutic efficacy based upon knowing the molecular target and mechanism of action is being challenged.

Cell-based phenotypic methods are allowing the high-fidelity host-cell/pathogen niche to be used for high-throughput image-based cell analysis for drug screening.

Phenotypic screening recapitulates biological relevance at the heart of drug-screening campaigns and is becoming a primary determinate to improved therapeutic outcome.

Emerging methods using transcript quantification, public databases (chem/bioactivity profiles, ontologies, image-based screening results), combined with machine-learning tools, are providing ground-breaking new and alternative screening strategies by augmenting phenotypic screening results.

Empirical drug discovery refers to the emergent paradigm that harnesses the inherent complementarity of phenotypic and target-based screening strategies and augments predictivity of chemotherapeutic efficacy.

Phenotypic Screening for Infectious Disease Drug Discovery

The term 'phenotype' (Figure 1) was coined originally in 1903 by the Danish botanist Wilhelm Johannsen [1–3] and emerged as foundational in experimental, theoretical, and fundamental biology juxtaposed with 'genotype'. However, unlike genotype, for which a definition arises from tangible minimal information and molecular typing, the term phenotype lacks clear definition [4], relying on semantic description, for example, 'morphology', 'behavior', 'appearance', 'structure' etc. Indeed, there is an entire literature on the meaning of 'phenotype' but, more than 100 years after its conceptualization, it was only in recent years that the discipline and tools enabling **ontology** (see Glossary) for high-throughput and high-dimensional phenotyping began to emerge [4–7]. Specifically, for microbiology (MicroO [6]), there are efforts to align with other community-based efforts establishing standards for ontology in genetics (GO[i]), phenotype (PATO[ii]), small-molecule chemical entities of biological interest (ChEBI[iii]), and PubChem[iv]. Indeed, such efforts are invigorated, in part, by the realization that the performance of powerful natural language processing with neural-networks, machine-learning, and deep-learning methods is wholly dependent on the underlying means for data extraction that are linked to such ontologies, and ultimately the ability to relate them [7]. For the purposes of the current article we use the term 'phenotype' in the restrictive context of phenotypic screening [8,9].

In terms of drug discovery, phenotypic screening describes the original nascent methodology that allowed evaluation of the biological effects of chemical entities, revealing desirable or potentially therapeutic effects linked to a disease. During three decades, drug discovery has shifted away from *in vivo/in situ* phenotypic screening toward molecular target-based strategies [10]. Generally, target-based methodologies use simple *in vitro* biochemical readouts (e.g., at the level of a well in a plate), whereas phenotypic-based methods use more elaborate cell-based readouts, for example, imaging single cells [10], or even whole organisms [11]. Target-based screening, using rational **mechanism of action (MoA)**, and/or hypothesis-driven approaches, dominated the pharma industry mainly because of its efficiency, cost economy, and massive throughput. By contrast, phenotypic methods, using mostly imaging and cell-based detection,

[1]Institut Pasteur Paris, UTechS-PBI/Imagopole, 25-28 rue du Docteur Roux, 75015, France
[2]Institut Pasteur Korea, 16 Daewangpangyo-ro 712 beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do, 13488, Republic of Korea

*Correspondence:
sshorte@pasteur.fr (S.L. Shorte).

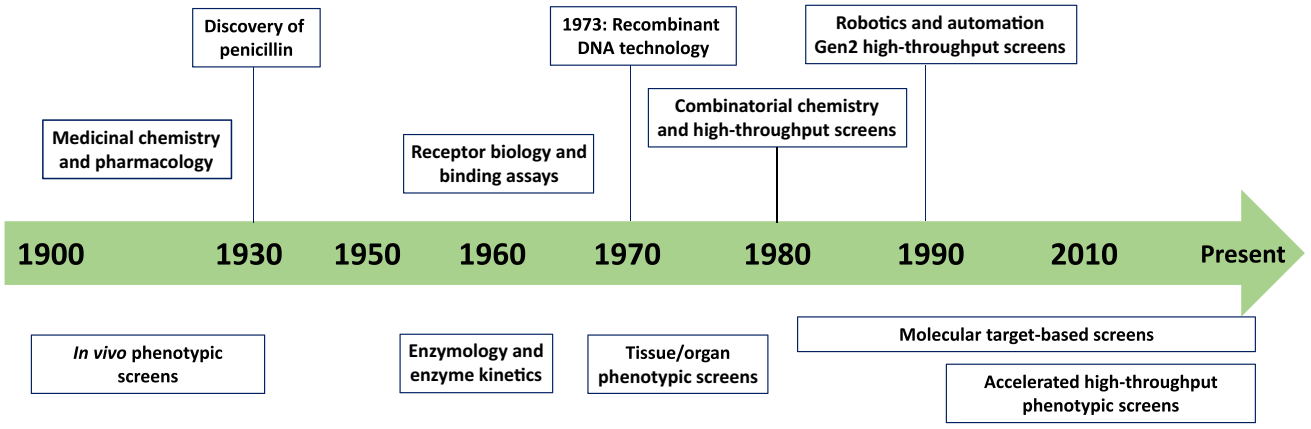

Figure 1. Phenotypic Screening Predates Target-based Screening by Decades. Phenotypic screening (also referred to as forward chemical screening [10]) is, in its simplest form, the basic activity of a cell biologist or microbiologist in observing, visually, the characteristics of cell biology either isolated *in vitro*, or in tissue, or in the whole organism context. Systematic phenotypic analysis began back in the early 1900s and is emerging hand in hand with microscopy and imaging. On the other hand, target-based screening was made possible in the wake of advances in our knowledge of biochemistry and molecular genetics that were pioneered in the postwar years (1950–1970s). Consequently, target-driven screening became a defining methodology during the 1980s and 1990s, and was adopted by the pharmaceutical industry to displace phenotypic methods. Adapted, with permission, from [10].

seemed comparatively inefficient and expensive (per sample), yielding a much lower throughput. Consequently, faced with industrial chemical libraries numbering hundreds of thousands, if not millions, of compounds, phenotypic screening lost traction as compared with target-based screening strategies. However, during the last 10 years, rapidly evolving image-detection, automation, computation and biotechnologies have started to change the *status quo*. Driven partly by what, at times, has been characterized as a deficit in discovery of new **first-in-class** chemical scaffolds using traditional target-based approaches, there has been a resurgence of interest in phenotypic screening for a variety of reasons [12–16]. Not least, the emergence of promising new improved phenotypic methods, which are based on enhanced biological relevance [17,18], has led some to claim that phenotypic screens surpass target-based approaches in identification of first-in-class scaffolds [14,16]. Today, it is becoming clear that the convergence of phenotypic-based and target-based methods is more important than their 'either/or' comparison. The trend toward 'empirical drug discovery' [15,19] recognizes the complementarity with target-driven campaigns whereby the biological relevance as revealed using phenotypic screening mitigates the risk of downstream failure in the translation **pipeline** from the preclinical to the clinical phase [18].

Phenotypic Screening Reveals Chemotherapeutic Potential

In infection biology the aim of phenotypic screening is to discover lead compounds that lend themselves to chemotherapeutic development in order to generate lead compounds capable of halting, and ideally reversing, infection and disease progression. The current conventional approach fits in the drug-screening framework that anticipates a molecular target which can be analyzed for sensitivity to a chemical compound, halting a disease **etiology** and lending itself to drug development by a precisely known MoA. However, because of our generally incomplete understanding of disease, combined with evidence suggesting increased promise for first-in-class hits [15], phenotypic screening approaches are appealing, precisely because they liberate screening campaigns from the burden of having *a priori* knowledge of either target, or MoA. In the search for chemotherapies targeting infectious diseases, successful drug-discovery campaigns using hypothesis-independent 'black-box' phenotypic screening strategies are abundant. For example, those having reached clinical use for the treatment of infectious disease include (see

[16]): the antifungal Caspofungin (Merck, 2001); antibiotics Daptomycin (Cubist, 2003), Linezolid (Pfizer, 2000), Retapamulin (GSK, 2007); and antivirals Docosanol (Avanir, 2000) and Sinecatechins (Medigene, 2006). More recently, while targeting neglected diseases, acoziborole (2009, Anacor Pharmaceuticals) was discovered using phenotypic screening with *Trypanosoma brucei* growth-inhibition assays [20], and a singular overexpression screening approach identified the molecular target of the benzoxaboroles to be the cleavage and polyadenylation specificity factor 3 (CPSF3) [21] currently in Phase II/III trials (sponsored by **DNDi**). Benzoxaboroles have broad-spectrum effects, which is a desirable property for chemotherapeutics. Similarly, fexinidazole, first discovered in 1983, shows a broad-spectrum antimicrobial activity against *Trypanosoma cruzi*, *Tritrichomonas foetus*, *Trichomonas vaginalis*, and *Entamoeba histolytica* [22,23]. Most recently, fexinidazole has been shown to be effective for clinical treatment of *T. brucei* infection [24] and is approved for global therapeutic use [25]. Broad-spectrum activity is desirable in efforts to develop chemotherapeutics, and can actually be used by design to enhance the drug-discovery process. For example, three million compounds were screened using three distinct proliferation assays on *Leishmania donovani*, *T. cruzi*, and *T. brucei* [26], identifying azabenzoxazole, GNF5343, as a hit in the *L. donovani* and *T. brucei* screens. Interestingly, GNF5343 was not active on *T. cruzi* in the primary screen. However, lead optimization yielded GNF6702 that displayed unprecedented *in vivo* efficacy, entirely clearing parasites from mice in all three models of infection and raising hopes of the possibility of developing a single class of drugs for these neglected diseases. Similarly, a weak hit from a phenotypic screen, performed across related flagellated protozoan parasites, was subjected to lead optimization, yielding GSK3186899/DDD853651 that proved a preclinical development candidate for the treatment of visceral leishmaniasis [27].

Broad-spectrum therapeutic effects can be the result of a drug interacting with a molecular target preserved commonly among different related pathogens. However, a more complex manifestation of broad-spectrum activity can be due to **polypharmacology**. In the 1970s, searching for new drugs for the treatment of malaria and intestinal parasites, Raymond Cavier and Jean-Francois Rossignol (of the Institut Pasteur Paris) discovered nitazoxanide, the first of the thiazolides [28]. Nitazoxanide (NTZ) was tested and found to be effective against both intestinal protozoa and helminths *in vitro* and in laboratory animal models. Several decades later, while the mechanisms still have yet to be elucidated, NTZ has lived up to more than its initial promise as an important antiparasitic drug [29–31]. Indeed, NTZ displays strikingly wider anti-infective efficacy than was expected, or even sought, upon its discovery [32]. Today, NTZ is revealed to be active against a range of extracellular and intracellular protozoans [33,34], helminths [35], anaerobic and microaerophilic bacteria [36,37], and viruses [38–45], including a broad range of respiratory viruses [46,47], and even neoplastic cells [48]. At least one recent report claims successful clinical treatment of cutaneous leishmaniasis with oral nitazoxanide [49]. Such is the diversity of broad-spectrum effects of NTZ that a common molecular target hypothesis seems less probable. An emerging idea is that polypharmacology may be at play whereby a compound (or its derivatives) is able to exert manifold effects. At this time, polypharmacology is not well understood, and it often eludes experimental design, making it difficult to even rule out (e.g., see Box 1). In the case of NTZ there is evidence that polypharmacology arises from multiple mechanisms, including metabolic, immunomodulatory host-cell/pathogen, tissue and cellular responses [50], oxidative stress [51], and in a manner additive with other known antileishmanials [51,52].

While broad-spectrum and polypharmacological effects are not exclusive to phenotypic screening, it is becoming clear that disease models of infection searching singular target-driven MoA microbicidal effects may be overly simplistic and not necessarily correlated with desirable clinical outcome. From the point of view of studies in infectious disease, it is most recently argued that

Glossary

Amastigote: a protist cell that does not have visible external flagella or cilia.
DNDi: Drugs for Neglected Diseases initiative, a nongovernmental international organization operating a collaborative, patients' needs-driven, nonprofit drug research and development pipeline for new treatments to help globally those suffering neglected diseases.
Etiology: the cause, set of causes, or manner of causation of a disease or condition.
First-in-class: drugs that use a new and unique mechanism of action for treating a medical condition. The first drug product to be marketed that contains a compound which acts on a specific (and new) target. The drug molecule and the target are both novel.
Fuzzy logic: a term derived from an area of mathematical modeling that was popularized to indicate the notion of 'gray area' in the certainty, or not of any decision-making process.
Mechanism of action (MoA): in pharmacology, it refers to the specific biochemical interaction through which a drug substance produces its pharmacological effect.
Ontology: a logic-based organizational structure for knowledge comprised from shared semantic descriptions; in effect, providing a formal representation for organizing the abstraction of knowledge into accessible (shared) semantic terminology.
Pipeline (drug): the set of drug candidates that a pharmaceutical research organization has under development at any given point in time, and the fixed stages through which they must proceed.
Polypharmacology: the design or use of pharmaceutical agents that act on multiple targets or disease pathways; it remains one of the major challenges in drug development, potentially opening novel avenues to rational next-generation drug design for more effective but less toxic therapeutic agents.

Box 1. A Case Study for Cutting-Edge Drug-Target Identification in the Parasite *Leishmania*

While phenotypic screening can help to replenish the pipeline with new lead candidate compounds, it does not change the fundamental need to identify the molecular target(s) and develop a rational strategy therein to produce a clinically efficacious drug product. Toward these ends a recent study reported pyrazolopyrimidine compounds that act principally by inhibiting *Leishmania* cdc-2-related kinase 12 (CRK12), thus defining a new druggable target for visceral leishmaniasis [88]. Pyrazolopyrimidines were optimized by assessing their effect on *in vitro* infection of macrophages by *Leishmania* and a mouse model of visceral leishmaniasis. Selecting one candidate based on good safety profile, high potency, and suitable properties for development as an orally administered drug, they next used a chemical proteomics strategy to identify the target(s), three enzymes: CRK3, CRK6, and CRK12. Next, whole-genome DNA sequencing revealed that drug-resistant parasites express CRK12 at higher than usual levels, and identified a mutation in the CRK12 gene that, when introduced back into wild-type parasites, conferred drug resistance. From the collective chemical proteomics studies of Wylie *et al.* [88], the authors recognized that, while CRK12 is undoubtedly the principal target of their compound series, they could not rule out 'polypharmacology' resulting from inhibition of secondary kinase targets responsible for some of the observed phenotypic effects in drug-treated parasites, such as cell cycle arrest.

the search for drugs should be better guided by the mantra '…improve host health, and outcome…' [11]. Albeit a rather **fuzzy logic** [53], phenotypic screening in relevant biological models may be better in helping to find molecules with the likelihood of improved therapeutic outcome [11]. Indeed, the discovery of clinically efficacious antibiotics has been attributed to assays using '…whole-cell screening…', and their therapeutic utility is, in part, attributed to their polypharmacology diminishing the risk of drug resistance in bacteria, viruses, and parasites ([54]: review article, and see references therein). While designing drugs with multitarget profiles has proved both complex and difficult, new computational probabilistic activity modeling approaches and synthetic medicinal chemistry has opened promising avenues toward automated design of ligands with polypharmacological profiles [55]. Having been mainly established in drug discovery aimed at cancer, and neuropsychiatric disorders, such polypharmacological profiling strategies have potentially high value for drug discovery for infectious disease. The key to unlocking such potential lays with the emergence of more complex phenotypic assays, better mimicking the biology underlying clinically relevant host-cell/pathogen interactions.

State-of-the-Art Phenotypic Assays Mimicking the Host-Cell/Pathogen Niche

During the last decade of drug screening, efforts in *Leishmania* have focused on extrapolating druggable targets in the parasite with some success (Box 1), reasoning that this could leverage rational development of chemotherapeutics suitable to address the public health challenges therein. However, while successful in identifying targets that allow large screening campaigns to be run, some groups have shifted their focus away from deconvolution of druggable targets in the parasite [56] and are turning their attention toward screening strategies targeting host-cell/pathogen interactions [57,58] (Figure 2) and phenotypic screening is well suited to this. For example, a novel strategy, based on *ex vivo* biology, used phenotypic assays combining primary murine macrophages and lesion-derived, virulent *L. donovani* and *Leishmania amazonensis* **amastigotes** to validate antileishmanial hit compounds ('GSK Leish-Box'). Strikingly, the *ex vivo* approach validated antileishmanial activity on intramacrophagic *L. donovani* for only 23 out of the 188 GSK Leish-Box hits previously identified using immortalized THP1 macrophage cell lines [58]. Presumably the greater physiological relevance of *ex vivo* assays, compared with immortalized cells, provides for a more discriminating assay sensitivity. Such assertion supports the rationale to use even more complex *ex vivo* assays – for example, splenic explant cultures from hamsters infected with luciferase-transfected *L. donovani* – to screen chemical compounds for antileishmanial activity [59]. A specific advantage of *ex vivo* explant cultures is that the host–parasite interaction may be better preserved, including aspects of the immune response. So, for drug screening, explant tissue assays can identify compounds that have both direct and indirect antiparasitic activity, and these studies have allowed identification of new compounds active against *L. donovani* within the pathophysiologic environment of the infected spleen [59].

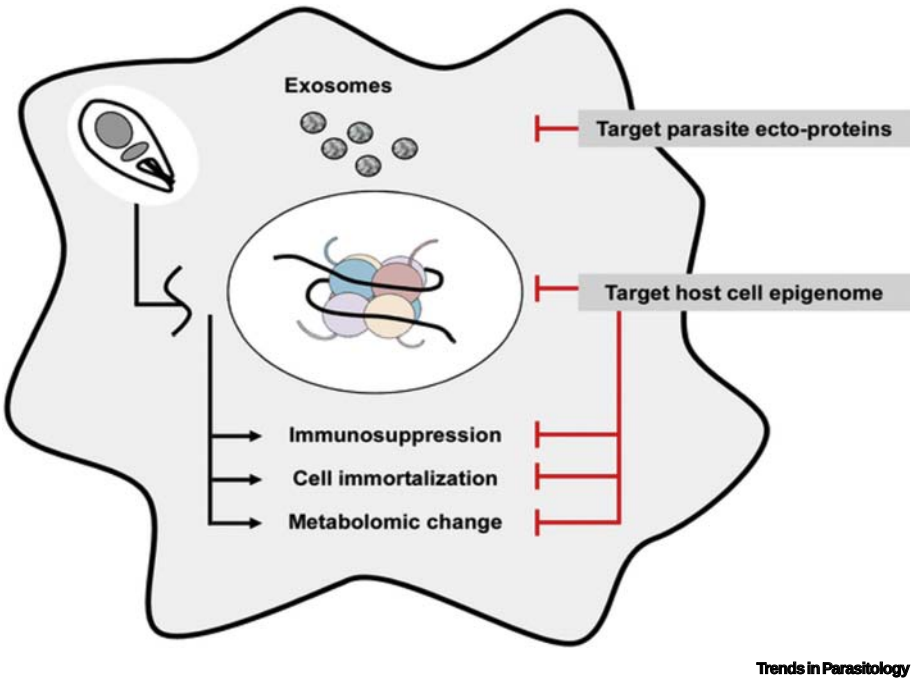

Figure 2. Targeting Host–Parasite Interaction as a New Venue for Antileishmanial Drug Discovery. Exosomal or secreted parasite factors released into the host cell likely modulate the macrophage epigenome, causing phenotypic changes that favor parasite survival, including suppression of immune functions, prolongation of host cell survival, and metabolic changes necessary for parasite proliferation. Interfering with parasite factors that act in *trans* on the host cell, or restoration of the normal host cell epigenome, will likely interfere with intracellular parasite survival and may thus be exploited for antileishmanial drug discovery. Reproduced, with permission, from [57].

Phenotypic assays using experimental models aiming to reconstitute the complexity of the host-cell/pathogen niche with greater fidelity can provide important and therapeutically relevant insights. For example, *Toxoplasma gondii* infection was used in a phenotypic screen examining the effects of compounds on cells stimulated by the powerful cytokine interferon gamma (IFN-γ; [60]). The rationale in this study was based on the observation that IFN-γ activates a variety of antimicrobial mechanisms in host cells, which are then able to control intracellular parasites such as *T. gondii*; however, despite the effectiveness of these pathways in controlling acute infection, the immune system is unable to eradicate chronic infections that can persist for life. The screen therefore used detection of parasite infection and autophagy in cells moderately stimulated with IFN-γ to screen for compounds whose effects could therefore be distinguished as dependent or not on IFN-γ activation. They reported a number of compounds that inhibited parasite growth *in vitro*, with enhanced potency in the presence of a low level of IFN-γ stimulation. Further, they demonstrated that a subset of these compounds acted by enhancing the recruitment of light chain 3 (LC3) to the parasite-containing vacuole, suggesting that they work by a noncannonical autophagy-related process, while others were independent of the autophagy pathway. Such studies indicate that synergistic interactions with immune responses are of high significance when considering the likely potential of therapeutic value and the predicted clinical outcome for any candidate molecule.

During infection there are many examples in which pathogens, sensing their environment, hide themselves or otherwise avoid host defense mechanisms; this can derail drug-discovery efforts. Cell-based assays accommodating such biologically relevant considerations into the experimental model can allow phenotypic screening to nonetheless proceed and identify compounds with

therapeutic potential, despite not necessarily knowing the underlying mechanisms. For example, the phenomenon of *T. cruzi* resisting extended exposure to trypanocidal compounds was shown to be most likely due to its ability to establish dormancy *in vivo* [61]. Most significantly, dormant *T. cruzi* amastigotes were uniquely resistant to extended drug treatment *in vivo* and *in vitro* and could re-establish infection after weeks of drug exposure. Combined with suitably adapted cell-based models, these results encourage the use of novel phenotypic screening approaches to seek alternative chemotherapeutic strategies. Similarly, in another study using a panel of *T. cruzi* strains, evidence for heterogeneity among parasites in the same population *in vitro* suggested the presence of quiescent parasites underlying differential drug sensitivity, and provided a possible mechanism explaining clinical failure of drug regimens to entirely clear parasite infection [62].

Some promising phenotypic cell-based assay methods attempt to enhance physiological/therapeutic relevance by synthetically reconstituting sophisticated tissue/organ-like microenvironment properties using a combination of microfluidics, micropatterning, and organotypic technologies. For example, Morada *et al*. [63] adapted a hollow fiber technology as a 3D substrate to culture human ileocecal colorectal adenocarcinoma cells (HCT-8), providing a long-term (>1 month) culture environment mimicking the gut. Importantly, the culture system allowed the delivery of nutrients and oxygen from the basal layer upwards concurrently with separate (anaerobic) redox and nutrient control of the lumen supporting infection-competent, fully virulent *Cryptosporidium* spp. In another example, micropatterned primary human hepatocyte cocultures (MPCCs) provided a microscale human liver platform, with stromal cells in a multiwell micropatterned coculture format supporting stable hepatocyte-specific function and metabolism during 4–6 weeks [64–66]. In addition to providing a permissive host, hepatocytes cultured in MPCC environments exhibited human-specific drug metabolism and long-term stability, which is ideal for drug screening and studies of long-term dormancy and reactivation. The multiwell *in vitro* platform has been demonstrated to be useful for transcriptional characterization using a customized capture method prior to RNA sequencing, demonstrating its potential as a drug-screening platform for studying *Plasmodium vivax* in a high content screening (HCS)-compatible 384-well format, leading the way for fully automated high-throughput drug screening. The further significance of this platform for supporting screening is supported by its compatibility with infection by hepatitis C and B viruses, *P. falciparum*, and *P. vivax* [64].

Next-Generation Phenotypic Screening in Infection

Offering great promise, data-driven computer-vision methods using machine-learning, neural networks, and deep-learning are capable of augmenting and facilitating throughput and analyses on ever more complex phenotypic screening pipelines – for example, 'artificial-intelligence (AI) workflow' adopting computational approaches to enhance phenotypic (high-content) imaging [67]. Fisch *et al.* [67] released a complete self-contained software platform (HRMAn: Host Response to Microbe Analysis) for image-based infection biology, providing fully automated, accurate unbiased quantification of host–pathogen interactions. Until now such essential analyses for infection biology were most often performed manually, or by using limited enumeration employing image analysis algorithms based on image segmentation. By contrast, HRMAn provides intuitive intelligent image analysis software capable of actually learning from the fluorescently labeled host-cell/pathogen images exactly how to distinguish and quantitatively assess host protein recruitment during infection. The open-source image analysis platform is based on machine-learning algorithms and deep learning, and is highly flexible, as evidenced by its capacity to learn phenotypes from the data without relying on researcher-based assumptions. Indeed, the system was shown to perform equally well on both parasite (*T. gondii*) and bacterial (*Salmonella enterica* Typhimurium) experimental models. HRMAn's capacity to recognize, classify, and quantify

pathogen killing, replication, and cellular defense responses is likely just the first example of many to come using AI-driven approaches to enhance early drug discovery for infection.

The utility of AI to enhance or augment high-content imaging is already advanced beyond replacing the classical image-processing and/or segmentation workflows. Indeed, an exciting avenue has been triggered by the work of Carpenter *et al*. who asserted a disruptive new experimental paradigm termed 'cell painting' [68–71]. Cell painting (also known as image-based morphological cell profiling) uses standardized methods for cell labeling and high-content imaging in order to establish a morphological profile based upon quantitative data extracted from microscopy images of cells. Using automated image analysis software, measuring approximately 1500 morphological features/cell, the data-rich cell profile is proposed to be suitable for detecting subtle phenotypes (Figure 3A, Key Figure). Profiles of cell populations treated with different experimental perturbations can be compared in order to suit many goals, such as identifying the phenotypic impact of chemical or genetic perturbations, grouping (clustering) compounds and/or genes into functional pathways, and identifying signatures of disease [72]. More recently, this idea has been extended further: Ceulemans *et al*. [73] demonstrated that 'image-based fingerprints' of compounds derived from a given image-based cellular assay can be repurposed to predict the biological activity of those same compounds in other seemingly unrelated assays, even those targeting alternate pathways or biological processes. The approach has the potential to greatly reduce unnecessarily voluminous scaling of screens by predicting the likelihood of activity on a new target. An 842-dimensional quantitative vector readout calculated for each single cell was extracted from a three-channel microscopy-based screen performed for 524 371 chemical compounds measuring glucocorticoid receptor translocation. A bioactivity matrix was used to document the available experimental activities of the 524 371 imaged compounds in some 1200 assays that could be attributed to a specific protein target, and from a subset of 535 assays revealed by machine-learning, several dozen returned robust criteria for predictivity performance. These data were used to predict assay-specific biological activity in two ongoing drug-discovery projects. In these projects, repurposing increased hit rates by 50- to 250-fold over that of the initial project assays while increasing the chemical structure diversity of the hits.

Underlying the tendency toward convergence between imaging and omics for empirical screening, there are exciting new technologies – for example, the development of a new method using digital amplicon-based RNA quantification by sequencing [74]. Armed with this technology it has been proposed that the 'phenotype' of a living cell may be usefully described by its pattern of active signaling networks, giving rise to a so-called 'molecular phenotype' (Figure 3B). Reasoning that the activity of signaling networks can be assessed based on a set of established key regulators and expression targets rather than the entire transcriptome, these proof-of-concept studies compiled a panel of 917 human genes, representing 154 human signaling and metabolic networks that they termed 'pathway reporter genes' [75]. In effect, allowing differential gene expression quantification, this method was used to characterize the molecular phenotype in developing human induced pluripotent stem cell (iPSC)-derived cardiomyocytes, and the drug/toxicity response in primary human hepatocytes [75]. The reporter genes were significantly enriched for regulators and effectors covering a wide range of biological processes, and were shown to report gene-level and pathway-level changes associated with differentiation and drug action. The pathway reporter genes delivered an accurate pathway-centric view of the biological system under study, and revealed known and novel modulation of signaling networks consistent with the literature and with experimental data. Focusing on the cardiomyocyte model, the same teams most recently went on to report the use of molecular phenotyping as a means to augment high-content image-based drug screening by helping to assure and stratify lead compound selection and characterize MoA classes [76].

CellPress
REVIEWS

Key Figure

Molecular Phenotyping and Image-based Fingerprinting

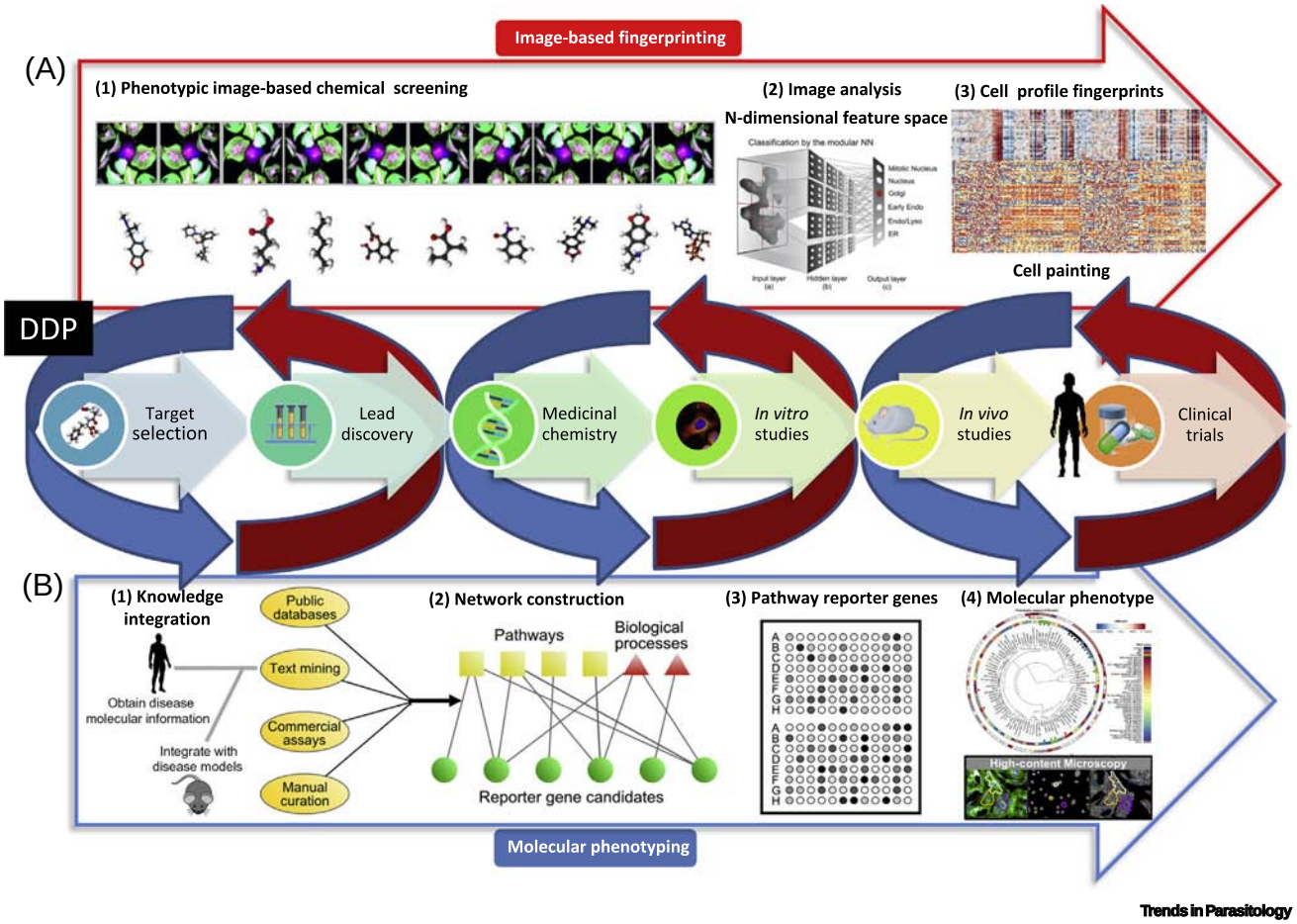

Figure 3. (A) Image-based fingerprinting (adapted from [68,73]). Set inside the red arrow, left to right, is a schematic representation workflow showing steps: (1) chemical phenotypic screening that used a three to five channel readout, >500 000 compounds, against several distinct cell lines; (2) image analysis, N-dimensional feature space extrapolating >800 independent feature parameters per single cell; and (3) cell profile fingerprint yielding a large array of single-cell morphological feature data. In the context of 1200 previously performed bioactivity assays applicable to the compound library used by Simm *et al*. [73], this shows that machine-learning methods can be applied to yield useful enhanced predictivity for these same compounds in new assays [73]. (B) Molecular phenotyping (adapted from [75,76]), set inside the blue arrow, showing a schematic illustration of the workflow for molecular phenotyping. Left to right: (1) knowledge integration using a variety of data sources, including public databases, text publication, shared assay data, clinical and preclinical disease model results, using both automated and manual data curation; (2) network construction links biological knowledge of signaling pathways and biological processes, with genes involved therein – this step links genes in pathways and transcriptional regulatory networks to biological processes of interest, judged as relevant to a particular pathology; (3) pathway reporter genes assembled into an experimental panel to be subject to quantification using highly sensitive amplicon-based mRNA transcript quantification (see [74]); and (4) the molecular phenotype allows reporter pathway genes and signaling networks therein to be effectively correlated with cell-based (high-content microscopy) assay readouts, allowing drug effects to be better assessed, screened, and modeled according to their likely therapeutic outcome [76]. Centre. DDP (drug discovery pipeline) shows, left to right, an illustration of the steps comprising the drug-discovery process from target selection through compound lead discovery to medicinal chemistry (lead optimization) through *in vitro* and *in vivo* preclinical screening, up to clinical translation. This pipeline lays in proximity to the underlying modeling premise and data exchange of molecular phenotyping and image-based fingerprinting; the cycling red/blue arrows indicate how DDP data exchange thereby powers, and is powered by, the integration of these methodological frameworks together, and underlies their inherent complementarity.

Interestingly, for the combined molecular phenotyping approach, one important limitation was attributed to the inability to recapitulate *in vitro* a single cellular phenotype model [76]. The authors used a classical segmentation-based image analysis method to score the cardiomyocyte phenotype, identifying sarcomere striation area, nuclei, and alpha-actinin (corresponding to cytoplasm). One wonders at the alternative strategy where molecular phenotyping might better be combined with the alternative image-based fingerprinting (cell painting [68,73]) described above (Figure 3). There is complementarity between these emergent technologies that both aim to power early drug discovery by drawing on data that inform on the biological relevance of the input (chemical libraries and cell-based assay) in order to enhance the quality of the data output (chemotherapeutic leads, likely MoA, beneficial therapeutic outcome). Importantly these strategies also aim to reduce the requirement for unnecessary throughput while simultaneously augmenting the richness and content of the data output. This is analogous to efforts toward improving predictions on safety and efficacy where chemical biology, cheminformatics, target-based, and phenotypic readouts help to inform hypotheses and predictive modeling in so-called 'phenotypic chemical biology' [77–79].

Concluding Remarks

New computational methods and concepts allow the handling of phenotypic models [13,15,67,68,73,76,80,81] such that the added complexity arising from considering biological relevance can actually help to improve productivity of early drug discovery. However, there are challenges ahead (see Outstanding Questions). Among the technical challenges, image-based morphological profiling requires a certain minimum information in terms of the resolution of the optical system used to collect the data and the dimensions of the target. For microorganisms, cell-morphological information may be obscured, or even entirely lost, due to insufficient optical resolution. Emerging super-resolution microscopy methods are only just beginning to show potential for high-throughput phenotypic screening, for example, in the form of deep-learning-based acceleration of single-molecule super-resolution microscopy: ANNA-PALM (artificial neural network accelerated photo-activated light microscopy; [82]). Such computational advances are certainly a significant step forwards, but more fundamental is the limitation whereby optical resolution and field size are inversely related; in lay terms, this means that higher optical resolution comes at the cost of diminishing optical field size. In the extreme, only a few cells can be imaged at a time using high-resolution (oil immersion) optics, thus, making impractical any possibility of automated imaging matching the throughput needs of drug screening. Interestingly, among a variety of optical methods aiming to overcome this major constraint, multibeam interferometric illumination [83] has yielded a commercialized cell-imaging system (Optical Biosystems, Santa Clara, CA, USA) capable of yielding optimized high-resolution equivalency in thousands of cells simultaneously using low-magnification optics. It will be interesting to see how such systems perform on highly multiplexed labeling of protein networks [84], and the transcriptome [85] visualized at mesoscopic scales. Such approaches could allow high-fidelity interactome analysis of host-cell/pathogen specific responses obviating background signal from noninfected cells. In combination with deep-learning and machine-learning automation, frameworks like ANNA and HRMAn might help to assure unprecedented genotype–phenotype classification, annotation, and clustering that is currently incomplete for many infectious disease and microbial paradigms.

Realizing opportunities from image-based morphological cell profiling highlights the imperative need for community coordination and collaborative efforts at every level from open-source software to open-access data and protocol sharing [71]. For infectious biology, combining phenotypic imaging and molecular readouts represents a doorway to the next generation of drug-discovery technologies. New bioinformatics tools promise a future in drug discovery that will leverage the biological relevance of phenotypic screening as a powerful means to better inform

Outstanding Questions

What types of optical imaging modality will emerge to address the concurrent need for high-resolution and high-throughput capacity?

Genotype–phenotype characterization requires imaging protein–protein interactions and transcriptome detection at mesoscopic scales, including the single cell. Will disruptive optical/computer-vision technologies overcome this challenge?

Will artificial-intelligence frameworks (machine-learning, neural networks, and deep-learning) leverage fully automated unsupervised genotype–phenotype annotation and classification?

How should the parasitology community engage to establish standards facilitating useful sharing of phenotypic screening data amenable as a sustained community resource?

cheminformatics, target-based MoA rationalization, or even multitarget drug profiling. Specifically, in parasitology, it is the clinical need in the field that is driving the shift toward screening campaign paradigms using phenotypic approaches [86]. For Chagas' disease, *T. cruzi* drug-discovery campaigns have, for several years, employed computational approaches using data from several public whole-cell, phenotypic high-throughput screens made available by the Broad Institute (including a single screen of over 300 000 molecules) [87]. Based on compiling and curating relevant biological and chemical compound screening data 584 compounds with activity data against *T. cruzi* were identified and made publicly available in the CDD database (Collaborative Drug Discovery Inc. Burlingame, CA, USA). Nonetheless, while providing a powerful proof-of-concept and tool in hand for future screening efforts, there is still a need to advance the field and facilitate better data annotation, curation, and public sharing therein. To achieve this we must consider how high-content screening data, software tools, and protocols will be validated and shared (e.g., [67]) and how, within this framework, the parasitology community shall contribute and benefit.

Acknowledgments

We are grateful for support from the Seventh Framework Program of the European Commission (LEISHDRUG project, 223414), the French Government [L'Agence nationale de la recherche (ANR)] programs: Investissements d'Avenir programme ('Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases'; grant ANR-10-LABX-62-IBEID and 'Laboratoire Revive'; grant ANR 10-LBX-73-REVIVE); France BioImaging (FBI; grant ANR-10-INSB-04-01)), the Région Ile de France (Domaine d'intêret majeur Innovative technologies for life sciences, DIM ELICIT; and DIM1Health); the GIS IBiSA (Infrastructures en biologie santé et agronomie); the National Research Foundation (NRF) of Korea (NRF-2017M3A9G6068254 Global Network and Research Infrastructure to Infectious Diseases); the Gyeonggi Provincial Government; the Institut Pasteur Korea, and Institut Pasteur Paris. We also thank the reviewers for their helpful and constructive suggestions.

Resources

[i]http://geneontology.org/
[ii]www.ebi.ac.uk/ols/ontologies/pato
[iii]www.ebi.ac.uk/chebi/init.do
[iv]https://pubchem.ncbi.nlm.nih.gov/

References

1. Roll-Hansen, N. (1979) The genotype theory of Wilhelm Johannsen and its relation to plant breeding and the study of evolution. *Centaurus* 22, 201–235
2. Johannsen, W. (1911) The genotype conception of heredity. *Am. Nat.* 45, 129–159
3. Johannsen, W. (1903) Om arvelighed i samfund og i rene linier. *Oversigt over det Kongelige Danske Videnskabernes Selskabs Forhandlinger* (3), p. 23
4. Mahner, M. and Kary, M. (1997) What exactly are genomes, genotypes and phenotypes? And what about phenomes? *J. Theoret. Biol.* 186, 55–63
5. Houle, D. *et al.* (2010) Phenomics: the next challenge. *Nat. Rev. Genet.* 11, 855
6. Blank, C.E. *et al.* (2016) MicrO: an ontology of phenotypic and metabolic characters, assays, and culture media found in prokaryotic taxonomic descriptions. *J. Biomed. Semantics* 7, 18
7. Smith, B. *et al.* (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* 25, 1251–1255
8. Brodin, P. and Christophe, T. (2011) High-content screening in infectious diseases. *Curr. Opin. Chem. Biol.* 15, 534–539
9. Mattiazzi Usaj, M. *et al.* (2016) High-content screening for quantitative cell biology. *Trends Cell Biol.* 26, 598–611
10. Zheng, W. *et al.* (2013) Phenotypic screens as a renewed approach for drug discovery. *Drug Discov. Today* 18, 1067–1073
11. Clatworthy, A.E. *et al.* (2018) Whole-organism phenotypic screening for anti-infectives promoting host health. *Nat. Chem. Biol.* 14, 331–341
12. Reaume, A.G. (2011) Drug repurposing through nonhypothesis driven phenotypic screening. *Drug Discov. Today* 8, 85–88
13. Moffat, J.G. (2017) Turning the light on in the phenotypic drug discovery black box. *Cell Chem. Biol.* 24, 545–547
14. Moffat, J.G. *et al.* (2014) Phenotypic screening in cancer drug discovery – past, present and future. *Nat. Rev. Drug Discov.* 13, 588–602
15. Moffat, J.G. *et al.* (2017) Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nat. Rev. Drug Discov.* 16, 531–543
16. Swinney, D.C. and Anthony, J. (2011) How were new medicines discovered? *Nat. Rev. Drug Discov.* 10, 507–519
17. Horvath, P. *et al.* (2016) Screening out irrelevant cell-based models of disease. *Nat. Rev. Drug Discov.* 15, 751–769
18. Vincent, F. *et al.* (2015) Developing predictive assays: The phenotypic screening 'rule of 3'. *Sci. Translat. Med.* 7 293ps15-293ps15
19. Lee, J.A. *et al.* (2017) Empirical drug discovery: a view from the proteome. *Drug Discov. Today Technol.* 23, 1–5
20. Jacobs, R.T. *et al.* (2011) SCYX-7158, an orally-active benzoxaborole for the treatment of stage 2 human African trypanosomiasis. *PLoS Negl. Trop. Dis.* 5, e1151
21. Wall, R.J. *et al.* (2018) Clinical and veterinary trypanocidal benzoxaboroles target CPSF3. *Proc. Natl. Acad. Sci. U. S. A.* 115, 9616–9621
22. Jennings, F.W. and Urquhart, G.M. (1983) The use of the 2 substituted 5-nitroimidazole, Fexinidazole (Hoe 239) in the treatment of chronic T. brucei infections in mice. *Z. Parasitenkd.* 69, 577–581

23. Raether, W. and Seidenath, H. (1983) The activity of fexinidazole (HOE 239) against experimental infections with *Trypanosoma cruzi*, trichomonads and *Entamoeba histolytica*. *Ann. Trop. Med. Parasitol.* 77, 13–26
24. Watson, J.A. *et al.* (2019) Pharmacokinetic-pharmacodynamic assessment of the hepatic and bone marrow toxicities of the new trypanoside fexinidazole. *Antimicrob. Agents Chemother.* 63, e02515-18
25. Deeks, E.D. (2019) Fexinidazole: first global approval. *Drugs* 79, 215–220
26. Khare, S. *et al.* (2016) Proteasome inhibition for treatment of leishmaniasis, Chagas disease and sleeping sickness. *Nature* 537, 229–233
27. Thomas, M.G. *et al.* (2019) Identification of GSK3186899/DDD853651 as a preclinical development candidate for the treatment of visceral leishmaniasis. *J. Med. Chem.* 62, 1180–1202
28. Rossignol, J.F. and Cavier, R. New derivatives of 2-benzamido-5-nitro thiazoles, in: Patent, U.S. (ed.) Phavic Sprl, USA, 1974 US Patent#US3950351A
29. White Jr., A.C. (2003) Nitazoxanide: an important advance in anti-parasitic therapy. *Am. J. Trop. Med. Hyg.* 68, 382–383
30. Lobovska, A. and Nohynkova, E. (2003) New drugs for treatment of parasitic infections. *Cas. Lek. Cesk.* 142, 177–181
31. Diaz, E. *et al.* (2003) Epidemiology and control of intestinal parasites with nitazoxanide in children in Mexico. *Am. J. Trop. Med. Hyg.* 68, 384–385
32. Raether, W. and Hanel, H. (2003) Nitroheterocyclic drugs with broad spectrum activity. *Parasitol. Res.* 90, S19–S39
33. Chavez, M.A. and White Jr., A.C. (2018) Novel treatment strategies and drugs in development for cryptosporidiosis. *Expert Rev. Anti-Infect. Ther.* 16, 655–661
34. Dunne, R.L. *et al.* (2003) Drug resistance in the sexually transmitted protozoan *Trichomonas vaginalis*. *Cell Res.* 13, 239–249
35. Beshay, E.V.N. (2018) Therapeutic efficacy of *Artemisia absinthium* against *Hymenolepis nana*: *in vitro* and *in vivo* studies in comparison with the anthelmintic praziquantel. *J. Helminthol.* 92, 298–308
36. Dowzicky, M.J. and Chmelarova, E. (2018) Antimicrobial susceptibility of Gram-negative and Gram-positive bacteria collected from Eastern Europe: results from the Tigecycline Evaluation and Surveillance Trial (T.E.S.T.) 2011–2016. *J. Glob. Antimicrob. Resist.* 17, 44–52
37. Guttner, Y. *et al.* (2003) Nitazoxanide in treatment of *Helicobacter pylori*: a clinical and in vitro study. *Antimicrob. Agents Chemother.* 47, 3780–3783
38. Piacentini, S. *et al.* (2018) Nitazoxanide inhibits paramyxovirus replication by targeting the Fusion protein folding: role of glycoprotein-specific thiol oxidoreductase ERp57. *Sci. Rep.* 8, 10425
39. Hickson, S.E. *et al.* (2018) Inhibition of vaccinia virus replication by nitazoxanide. *Virology* 518, 398–405
40. Haubrich, K. *et al.* (2018) Successful treatment of chronic norovirus gastroenteritis with nitazoxanide in a pediatric kidney transplant recipient. *Pediatr. Transplant.* 22, e13186
41. Ghusson, N. and Vasquez, G. (2018) Successfully treated Norovirus- and Sapovirus-associated diarrhea in three renal transplant patients. *Case Rep. Infect. Dis.* 2018, 6846873
42. Fumian, T.M. *et al.* (2018) Potential therapeutic agents for feline Calicivirus infection. *Viruses* 10, 433
43. Flerlage, T. *et al.* (2018) Rotavirus infection in pediatric allogeneic hematopoietic cell transplant recipients: clinical course and experience using nitazoxanide and enterally administered immunoglobulins. *Pediatr. Infect. Dis. J.* 37, 176–181
44. Esquer Garrigos, Z. *et al.* (2018) Nitazoxanide is a therapeutic option for adenovirus-related enteritis in immunocompromised adults. *Antimicrob. Agents Chemother.* 62, e01937-18
45. Dang, W. *et al.* (2018) Nitazoxanide inhibits human norovirus replication and synergizes with ribavirin by activation of cellular antiviral response. *Antimicrob. Agents Chemother.* 62 AAC.00707-18
46. La Frazia, S. *et al.* (2018) The second-generation thiazolide haloxanide is a potent inhibitor of avian influenza virus replication. *Antivir. Res.* 157, 159–168
47. Rossignol, J.-F. (2014) Nitazoxanide: A first-in-class broad-spectrum antiviral agent. *Antivir. Res.* 110, 94–103
48. Wang, X. *et al.* (2018) Nitazoxanide, an antiprotozoal drug, inhibits late-stage autophagy and promotes ING1-induced cell cycle arrest in glioblastoma. *Cell Death Dis.* 9, 1032
49. Dhawan, A.K. *et al.* (2015) Successful treatment of cutaneous leishmaniasis with nitazoxanide. *Indian J. Dermatol. Venereol. Leprol.* 81, 644–646
50. Shakya, A. *et al.* (2018) Update on nitazoxanide: a multifunctional chemotherapeutic agent. *Curr. Drug Discov. Technol.* 15, 201–213
51. Mesquita, J.T. *et al.* (2013) Lethal action of the nitrothiazolyl-salicylamide derivative nitazoxanide via induction of oxidative stress in *Leishmania* (L.) *infantum*. *Acta Trop.* 128, 666–673
52. Mesquita, J.T. *et al.* (2014) Combination therapy with nitazoxanide and amphotericin B, Glucantime(R), miltefosine and sitamaquine against *Leishmania* (Leishmania) *infantum* intracellular amastigotes. *Acta Trop.* 130, 112–116
53. Novák, V. *et al.* (1999) *Mathematical Principles of Fuzzy Logic*. Kluwer Academic Publishers, Boston/Dordrecht/London. ISBN 0-7923-8595-0
54. Peters, J.U. (2013) Polypharmacology – foe or friend? *J. Med. Chem.* 56, 8955–8971
55. Besnard, J. *et al.* (2012) Automated design of ligands to polypharmacological profiles. *Nature* 492, 215–220
56. Durieu, E. *et al.* (2016) From drug screening to target deconvolution: a target-based drug discovery pipeline using *Leishmania* casein kinase 1 isoform 2 to identify compounds with antileishmanial activity. *Antimicrob. Agents Chemother.* 60, 2822–2833
57. Lamotte, S. *et al.* (2017) The enemy within: Targeting host–parasite interaction for antileishmanial drug discovery. *PLoS Negl. Trop. Dis.* 11, e0005480
58. Lamotte, S. *et al.* (2019) Discovery of novel hit compounds with broad activity against visceral and cutaneous Leishmania species by comparative phenotypic screening. *Sci. Rep.* 9, 438
59. Osorio, Y. *et al.* (2011) Identification of small molecule lead compounds for visceral leishmaniasis using a novel *ex vivo* splenic explant model system. *PLoS Negl. Trop. Dis.* 5, e962
60. Radke, J.B. *et al.* (2018) High throughput screen identifies interferon γ-dependent inhibitors of *Toxoplasma gondii* growth. *ACS Infect. Dis.* 4, 1499–1507
61. Sanchez-Valdez, F.J. *et al.* (2018) Spontaneous dormancy protects *Trypanosoma cruzi* during extended drug exposure. *eLife* 7, e34039
62. MacLean, L.M. *et al.* (2018) Development of *Trypanosoma cruzi in vitro* assays to identify compounds suitable for progression in Chagas' disease drug discovery. *PLoS Negl. Trop. Dis.* 12, e0006612
63. Morada, M. *et al.* (2016) Continuous culture of *Cryptosporidium parvum* using hollow fiber technology. *Int. J. Parasitol.* 46, 21–29
64. Gural, N. *et al.* (2018) Engineered livers for infectious diseases. *Cell. Mol. Gastroenterol. Hepatol.* 5, 131–144
65. Gural, N. *et al.* (2018) *In vitro* culture, drug sensitivity, and transcriptome of *Plasmodium vivax* hypnozoites. *Cell Host Microbe* 23, 395–406.e4
66. Khetani, S.R. and Bhatia, S.N. (2008) Microscale culture of human liver cells for drug development. *Nat. Biotechnol.* 26, 120–126
67. Fisch, D. *et al.* (2019) Defining host–pathogen interactions employing an artificial intelligence workflow. *eLife* 8, e40560
68. Bray, M.A. *et al.* (2016) Cell painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.* 11, 1757–1774
69. Caicedo, J.C. *et al.* (2016) Applications in image-based profiling of perturbations. *Curr. Opin. Biotechnol.* 39, 134–142
70. Gustafsdottir, S.M. *et al.* (2013) Multiplex cytological profiling assay to measure diverse cellular states. *PLoS One* 8, e80999
71. Caicedo, J.C. *et al.* (2017) Data-analysis strategies for image-based cell profiling. *Nat. Methods* 14, 849–863
72. Scheeder, C. *et al.* (2018) Machine learning and image-based profiling in drug discovery. *Curr. Opin. Syst. Biol.* 10, 43–52
73. Simm, J. *et al.* (2018) Repurposing high-throughput image assays enables biological activity prediction for drug discovery. *Cell Chem. Biol.* 25, 611–618 e3

74. Zhang, J.D. *et al.* (2014) Highly sensitive amplicon-based transcript quantification by semiconductor sequencing. *BMC Genom.* 15, 565
75. Zhang, J.D. *et al.* (2015) Pathway reporter genes define molecular phenotypes of human cells. *BMC Genom.* 16, 342
76. Drawnel, F.M. *et al.* (2017) Molecular phenotyping combines molecular information, biological relevance, and patient data to improve productivity of early drug discovery. *Cell Chem. Biol.* 24, 624–634.e3
77. Berg, E.L. (2017) Phenotypic chemical biology for predicting safety and efficacy. *Drug Discov. Today* 23, 53–60
78. Liu, X. *et al.* (2016) Systematic identification of pharmacological targets from small-molecule phenotypic screens. *Cell Chem. Biol.* 23, 1302–1313
79. Prinz, J. *et al.* (2016) A novel drug-mouse phenotypic similarity method detects molecular determinants of drug effects. *PLoS Comput. Biol.* 12, e1005111
80. Bray, M.A. and Carpenter, A.E. (2018) Quality control for high-throughput imaging experiments using machine learning in cellprofiler. *Methods Mol. Biol.* 1683, 89–112
81. Bray, M.A. *et al.* (2017) A dataset of images and morphological profiles of 30 000 small-molecule treatments using the cell painting assay. *Gigascience* 6, 1–5
82. Ouyang, W. *et al.* (2018) Deep learning massively accelerates super-resolution localization microscopy. *Nat. Biotechnol.* 36, 460–468
83. Ryu, J. *et al.* (2006) Multibeam interferometric illumination as the primary source of resolution in optical microscopy. *Appl. Physics Lett.* 88, 171112
84. Gut, G. *et al.* (2018) Multiplexed protein maps link subcellular organization to cellular states. *Science* 361, eaar7042
85. Stoeger, T. *et al.* (2015) Computer vision for image-based transcriptomics. *Methods* 85, 44–53
86. Chatelain, E. and Loset, J.-R. (2018) Phenotypic screening approaches for Chagas disease drug discovery. *Expert Opin. Drug Discov.* 13, 141–153
87. Ekins, S. *et al.* (2015) Machine learning models and pathway genome data base for *Trypanosoma cruzi* drug discovery. *PLoS Negl. Trop. Dis.* 9, e0003878
88. Wyllie, S. *et al.* (2018) Cyclin-dependent kinase 12 is a drug target for visceral leishmaniasis. *Nature* 560, 192–197