

# Parasites, proteomes and systems: has Descartes' clock run out of time?

J. M. WASTLING<sup>1\*</sup>, S. D. ARMSTRONG<sup>1</sup>, R. KRISHNA<sup>1,2</sup> and D. XIA<sup>1</sup>

<sup>1</sup> Department of Infection Biology, Institute of Infection and Global Health, Faculty of Health and Life Sciences, University of Liverpool, Liverpool L69 7ZJ, UK

<sup>1,2</sup> Department of Functional and Comparative Genomics, Institute of Integrative Biology, Faculty of Health and Life Sciences, University of Liverpool, Liverpool, L69 7ZB, UK

(Received 23 January 2012; revised 29 March 2012; accepted 5 April 2012; first published online 25 July 2012)

## SUMMARY

Systems biology aims to integrate multiple biological data types such as genomics, transcriptomics and proteomics across different levels of structure and scale; it represents an emerging paradigm in the scientific process which challenges the reductionism that has dominated biomedical research for hundreds of years. Systems biology will nevertheless only be successful if the technologies on which it is based are able to deliver the required type and quality of data. In this review we discuss how well positioned is proteomics to deliver the data necessary to support meaningful systems modelling in parasite biology. We summarise the current state of identification proteomics in parasites, but argue that a new generation of quantitative proteomics data is now needed to underpin effective systems modelling. We discuss the challenges faced to acquire more complete knowledge of protein post-translational modifications, protein turnover and protein-protein interactions in parasites. Finally we highlight the central role of proteome-informatics in ensuring that proteomics data is readily accessible to the user-community and can be translated and integrated with other relevant data types.

Key words: Systems-biology, proteomics, transcriptomics, genomics, parasites, host-parasite interactions, Descartes, reductionism.

## INTRODUCTION

*"I should like you to consider that these functions...follow from the mere arrangement of the machine's organs every bit as naturally as the movements of a clock or other automaton follow from the arrangement of its counter-weights and wheels."* Rene Descartes (1596–1650); *Treatise of Man* (Descartes and Hall, 1972)

Descartes' visionary approach to understanding the functioning of the human body almost 400 years ago brought him into conflict with the scientific establishment of the time, but marked the beginning of a new way of thinking in the natural sciences: one which saw the inexorable rise of a reductionist approach to resolving scientific problems. His appeal to *"Divide each difficulty into as many parts as is feasible and necessary to resolve it"* now underpins the way we think about science and has facilitated great advances in biomedicine, including an understanding of the mechanisms which control the replication, transfer and evolution of genetic information from one generation to another. Indeed, it is since the discovery of the structure of DNA and the

development in the later part of the 20th Century of tools to identify and manipulate individual genes and their products, that reductionist biology reached its high point through the paradigm of gene-centred, hypothesis-driven, experimentation. With gene identification becoming increasingly easy (including for most pathogens and hosts) and the availability of cheap commercial systems to manipulate gene expression both *in vitro* and *in vivo*, such experiments were irresistible. In infectious diseases research, this approach was successful in revealing many important biological functions that have transformed our understanding of the mechanisms which underpin host-pathogen interactions. Yet, within this vast amount of genomics data resides the challenge to reductionist thinking itself. The birth (or re-birth) of "Systems Biology" marks the beginning of a retreat from reductionism – a process encouraged by the fact that we are all being hopelessly outpaced by the scale of the biological data available. Some now argue that although reductionism has successfully identified most of the components and many of the interactions of biological systems, it offers no convincing concepts or methods to understand how system properties emerge (Sauer *et al.* 2007). By contrast, systems biology attempts to link high throughput molecular sciences such as genomics, proteomics and metabolomics by integrating data across different levels of structure and scale with the aim of understanding pathways, functional

\* Corresponding author: Department of Infection Biology, Institute of Infection and Global Health, Faculty of Health and Life Sciences, University of Liverpool, Liverpool L69 7ZJ, UK. Tel: 0151 794 4262. E-mail: j.wastling@liv.ac.uk

nodules and large-scale organisation (Oltvai and Barabasi, 2002). A systems approach is characterized by multi-disciplinary collaboration including the natural sciences, mathematics, computer science, engineering and medicine; and in its most tractable form it enables the linking of experiment and modelling processes. Above all, systems biology is about assembling, rather than disassembling structure; integration rather than reduction and “requires that we develop ways of thinking about integration that are as rigorous as our reductionist programmes, but different” (Noble, 2006).

Whether systems biology evolves to become the new paradigm for biological science experimentation in the 21st Century remains to be seen, but it is certainly here to stay. If it is to be a success then the technological tools required to implement its possibilities need to be fit for purpose. In the next few years the genomes from all known parasites either are, or will be available at the click of a mouse and the same will be true for most host species. The recording, storage and querying of genome sequence data is challenging because of its size, but is comparatively straightforward since genomes are basically stable entities which can be annotated and applied as a road map to the organism across multiple experimental scenarios. By contrast, the phenome, defined by the totality of all traits of an organism, or of one of its sub-systems (Mahner and Kary, 1997) is decidedly not stable and presents huge challenges in terms of analysis, representation and interpretation. Expression changes as measured by mRNA abundance using high-throughput mRNA sequencing provide great possibilities to provide signatures, for example, of how host cells or tissue respond to parasite challenge. These data are of course highly dynamic and more complex than sequence data, requiring careful regard of statistics during experimental design and need to consider the value of steady-state observations *versus* turnover. Although forming an integral part of the message signature in a biological system, mRNAs are of course not the functional molecules in a cell. Ultimately it is the dynamic of protein expression that is the most relevant functional measurement. Proteins, which more than any other component in the cell are the ‘*counter-weights and wheels.*’ of ‘the clock’, are in a constant state of flux and activation and as such provide an even greater analytical challenge than the measurement of mRNA.

In the past 12 years significant advances have been made in analysing the proteomes of parasites and their hosts. In 2002 only a handful of research publications existed on parasites and proteomics, whereas in 2011 over 70 articles were published in that year alone, with the field totalling over 2000 citations (ISI Web of Knowledge). The rapid growth of the field has been facilitated by major technical advances allowing access to a wide spectrum of

researchers. However, significant limitations still exist in the technologies constituting the science of proteomics, some of which present a real challenge to the emergence of systems biology as a tool to study host-parasite interactions. In this review we examine the current status of parasite proteomics and the scale of the tasks ahead. We review the current status of large-scale identification proteomics and discuss the need to apply more sophisticated quantitative proteomics approaches as we move from the era of descriptive proteomics to one in which we are concerned with understanding the dynamics of protein expression. We review recent developments in the availability of user-friendly publicly accessible interfaces for parasite proteomics data and their potential for integrating transcriptomics and other data into a wider systems biology analysis. Finally we discuss the relationship between the proteome and transcriptome and ask ourselves: *to what extent have we even begun to acquire the necessary baseline data required to realise a systems approach to host-pathogen interactions?*

#### HIGH-THROUGHPUT IDENTIFICATION OF PARASITE PROTEOMES

Before the availability of annotated parasite genomes, ‘top-down’ approaches in which intact proteins were analysed directly by mass spectrometry (MS) (along with non-MS based approaches such as amino acid sequencing by Edman degradation), were the only practical way to obtain protein sequence information. These approaches were challenging, laborious and had significant limitations, one of which was that they were suited to only very low-throughput experiments. Advances both in MS instrumentation and better annotated genomes meant that so-called ‘bottom-up’ proteomics was possible in which MS is used to analyse enzymatically-digested or chemically produced peptides from protein samples. The resulting MS fragmentation spectra (or fingerprint) can then be used to infer sequence by matching to a database derived from annotated genome sequence data. With genome sequence data now abundant, this approach is what most people today recognise as ‘proteomics’ and is especially suited to high-throughput experiments in which thousands of proteins can be identified from a sample in a single run on a suitably configured mass spectrometer. In high-throughput identification experiments, proteins samples can be analysed whole, or as sub-proteomes produced by a variety of approaches such as fractionation, organelle separation or affinity purification. Before MS takes place, separation at either the protein level or of peptides is almost always performed to reduce sample complexity. Protein separation is typically achieved by techniques such as gel electrophoresis or other forms of chromatography followed by in-gel digestion, while peptide

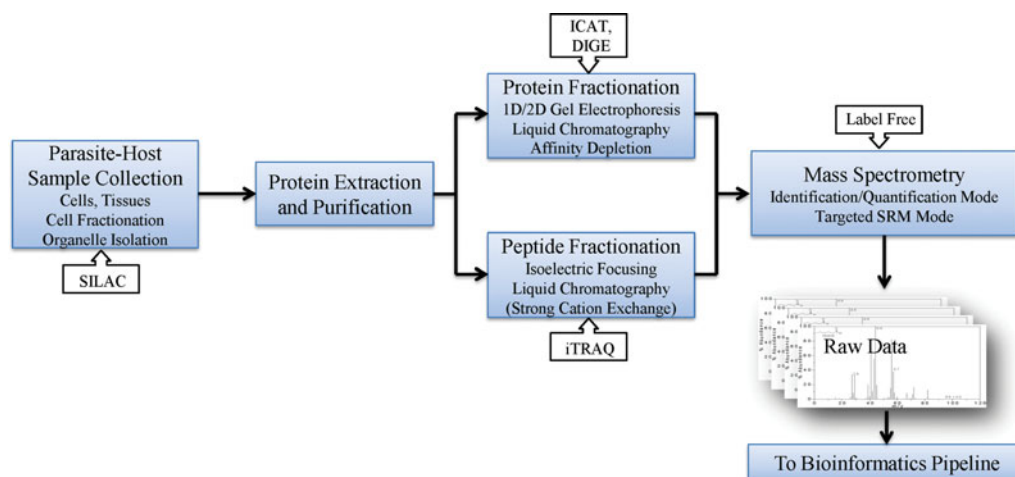


Fig. 1. A schematic diagram for a proteomics workflow. A typical workflow for a high-throughput identification and quantification proteomics experiment. The workflow consists of four major steps: sample collection, protein extraction and purification, sample fractionation (in either protein or peptide space) and mass spectrometry analysis. Various entry points for quantitative proteomics are shown by arrowed boxes. The collected raw mass spectra data are then analysed by various bioinformatics pipelines illustrated in Fig. 2.

separation is achieved by using liquid chromatography before analysis by tandem mass spectrometry (MS/MS). A typical workflow for high-throughput identification proteomics in parasite systems is shown in Fig. 1. Once the MS spectra have been produced identifications are made by matching the experimental peptide MS data, usually mass/charge ( $m/z$ ) data, to a theoretically calculated peptide  $m/z$  database using search engines such as Mascot (Perkins *et al.* 1999), SEQUEST (Yates *et al.* 1995) and X!Tandem (Craig and Beavis, 2004).

The field of parasitology has been quick to exploit emerging proteomics technologies. Early studies typically used in-gel digestion of protein spots separated by two-dimensional electrophoresis (2-DE) and identification made using peptide mass fingerprinting (PMF) on a MALDI-ToF instrument, initially on excretory–secretory (ES) products from parasitic worms (Jefferies *et al.* 2001; Yatsuda *et al.* 2003) then on more complex life stage proteomes (Cohen *et al.* 2002; Curwen *et al.* 2004). The more widespread use of CID (collisional induced dissociation) based LC-MS/MS analysis of peptides allowed the high-throughput identification of multiple proteins present in 1DE or 2DE spots or in whole digests—fractionated or un-fractionated. The advances in MS instrumentation enabled a more comprehensive analysis of parasite proteomes. Two pioneering examples took advantage of high-throughput proteomics for the analysis of the life cycle of *Plasmodium falciparum* (Florens *et al.* 2002; Lasonder *et al.* 2002). These global strategies were used to analyse the changes that the parasite undergoes as it traverses its life cycle in multiple hosts. Both studies utilised pre-fraction strategies, on-line multi-dimensional protein identification technology (MudPIT) which used the complementary separation

power of strong cation exchange and reverse phase chromatography or 1-D gel electrophoresis combined with MS/MS analysis to identify both parasite and human proteins. Evidence for the translation of more than 1000 predicted ‘hypothetical’ proteins were confirmed by use of high throughput proteomic techniques. Similar multiple separation technologies (MudPIT, 2-DE, and gel-LCMS) were used on the proteomes of other Apicomplexa including *Cryptosporidium parvum* (Sanderson *et al.* 2008) and *Toxoplasma gondii* (Xia *et al.* 2008). In the case of *T. gondii* in particular, peptide evidence was also used to help correctly assign exon-intron boundaries and make important refinements to the annotation of the genome (Xia *et al.* 2008). At present, much of the parasite high-throughput identification proteomics data is focused on protozoan parasites, probably largely due to the greater number of advanced annotated genomes compared to those available for helminths. There are now over 15 protozoan species with on-going proteomics projects, half of which have greater than 30% coverage and are headed by *Plasmodium* and *Toxoplasma*, each with around 70% coverage (Table 1). The next few years will see a similar advance in proteomics identification data for helminths as many of the genomes come on-line. However, some helminth parasites such as *Schistosoma mansoni* do have advanced proteomics projects (Curwen *et al.* 2004; van Balkom *et al.* 2005; Braschi and Wilson, 2006; Guillou *et al.* 2007). Some helminth studies have concentrated on sub-proteomes where the focus has been on the host-parasite interface and the possible roles played by excretory/secretory (ES) material on the host immune system. The ES of *Brugia malayi* has been extensively characterized using proteomic techniques. Several proteins with immune regulation

Table 1. Current proteomics projects in protozoan parasites. There are over 15 protozoan species with ongoing proteomics projects, half of which have greater than 30% coverage. EuPathDB ([www.eupathdb.org](http://www.eupathdb.org)) (Aurrecochea *et al.*, 2010) acts as one of the main portals for eukaryotic pathogen proteomics

Website	Species	Taxon ID	Predicted Proteome	Proteomics Data	Proteome Coverage	Enriched Subproteome Data	Multiple Life Cycle Stages Data
AmoebaDB	<i>E. histolytica</i>	294381	8306	2442	29.40%	Yes	No
CryptoDB	<i>C. parvum</i>	353152	3805	1320	34.69%	Yes	Yes
GiardiaDB	<i>G. intestinalis</i>	184922	5901	1744	29.55%	Yes	No
PlasmoDB	Assemblage A						
	<i>P. berghei</i>	5823	4904	1999	40.76%	Yes	No
	<i>P. falciparum</i>	36329	5524	3940	71.33%	Yes	Yes
	<i>P. vivax</i>	126793	5435	311	5.72%	No	No
	<i>P. yoelii</i>	352914	7724	674	8.73%	No	Yes
ToxoDB	<i>T. gondii</i>	508771	7993	5425	67.87%	Yes	Yes
TrichDB	<i>T. vaginalis</i>	412133	59672	1830	3.07%	Yes	No
TriTrypDB	<i>L. infantum</i>	5671	8241	538	6.53%	No	Yes
	<i>L. braziliensis</i>	420245	8357	40	0.47%	Yes	No
	<i>L. major</i>	347515	8412	502	5.97%	N.A.	N.A.
	<i>L. mexicana</i>	5665	8250	1682	20.39%	Yes	No
	<i>T. brucei</i>	185431	10533	3850	36.55%	Yes	No
	<i>T. cruzi</i>	353153	10342	1512	14.62%	Yes	Yes

properties were identified in the ES of individual life stages (Hewitson *et al.* 2008; Moreno and Geary, 2008) as well as in other nematodes such as *Teladorsagia circumcincta* (Craig *et al.* 2006). Further comprehensive analysis of the *B. malyai* secretome was achieved by (Bennuru *et al.* 2009), identifying over 800 proteins total in the analysis of several life stages of the parasite.

Advances in MS instrumentation and protein separation technology will continue to increase the number of identifications that can be obtained from a single sample. Increasing resolution and accuracy have improved the reliability of these identifications, as has the use of more sophisticated bioinformatic tools to improve processing of the MS/MS data and to ensure identifications are supported statistically. It is surely only a matter of time until relatively extensive proteomics coverage has been reported for most parasites of relevance to human and animal health. However, it is worth considering the extent to which a fully comprehensive proteome for some organisms is really achievable. Issues with sample preparation, incomplete tryptic digestion (Brownridge and Beynon, 2011), dynamic range, bias against certain classes of proteins, imperfect genome annotation, and concerns over peptide coverage, detectability and specificity ('proteotypic' peptides) (Beck *et al.* 2011) make full coverage of a proteome challenging. Extensive pre-fractionation of the protein or peptide samples can go some way to overcome some of these problems, but in doing so add a great deal of experimental redundancy as well as greatly increasing instrument analysis time and cost. Finally, parasites are characterized by often possessing complex life-cycles, sometimes in multiple hosts or survival in the external environment. A truly comprehensive proteome for any parasite is therefore

never expressed at any one moment; rather it is a dynamic and responsive facet of the host-parasite system for which a great range of temporal data is required before a full picture can be achieved.

#### ADVANCES IN QUANTITATIVE PROTEOMICS IN PARASITE AND HOSTS

Measurement of the changes in the abundance of proteins from one condition to the next, or determining changes in the protein composition of protein complexes and organelles under different conditions, are key data which can help us understand how the proteome responds in a dynamic host-parasite interaction. Quantification is now one of the foremost topics in proteomics and the most recent proteomic platforms are now geared not only to provide identification, but also some form of quantitative data. There are two main approaches used in quantitative proteomics: label-based methods and label-free methods.

Common labelling techniques involve either stable isotope labelling through *in vivo* metabolic labelling, chemical modification or labelling by fluorescent dyes. Popular chemical labelling includes proteins labelled with isotope-coded affinity tags (ICAT) (Gygi *et al.* 1999) and iTRAQ, which uses a multiplexed set of isobaric reagents that yield amine-derivatised peptides for relative and absolute quantitation (Ross *et al.* 2004). *In vivo* labelling metabolically labels proteins by incorporation of stable isotope labels with amino acids in cell culture (SILAC) (Ong *et al.* 2002). Fluorescent dye labelling is used in 2-DE fluorescence difference gel electrophoresis (2-DE DIGE) (Unlu *et al.* 1997). More recently, label-free quantification has gained popularity due to cheap and easy experimental



implementation. Label-free techniques directly use raw data from parallel MS runs to compare relative proteins abundance in different runs. Spectral counting and intensity-based methods are among the most commonly used approaches. Spectral counting infers protein abundance using the number of peptide-spectrum matches (PSMs) in a given run. However, due to different ionization efficiencies caused by biophysical properties of each peptide, the raw spectral counting has been proved to be less reliable as a quantification indication. Several software packages have been developed to normalize spectral counting, such as APEX (Lu *et al.* 2007) and emPAI (Ishihama *et al.* 2005). Intensity-based methods align precursor ion spectra of the same peptide from parallel runs according to their retention times (RT), and protein quantification is acquired by summing ion intensities that have been matched to peptides for a given protein. This approach has been implemented by several commercial software packages such as Progenesis LC-MS (NonLinear Dynamics) and Protein Lynx Global Server (Waters), as well as open-source packages such as MaxQuant (Cox and Mann, 2008), OpenMS (Sturm *et al.* 2008) and MSight (Palagi *et al.* 2005).

#### Label-based quantitative proteomics in host-parasite systems

Which quantitative approach to adopt depends on the nature of the host-parasite system under investigation. If the parasite can be cultured then *in vivo* metabolic labelling with a stable isotope can be achieved. SILAC (stable isotope labelling with amino acids in cell culture) has been extensively performed in cell culture, with virtually the entire proteomes of diploid and haploid yeast being compared using SILAC (de Godoy *et al.* 2008). There are only a few instances of *in vivo* metabolic labelling being used with parasites. Abundance changes in the proteome of the trophozoite stages of the malarial parasite *Plasmodium falciparum* following chloroquine and artemisinin treatment were examined using a stable isotope approach that used  $^{14}\text{N}$ -isoleucine and  $^{13}\text{C}_6$ ,  $^{15}\text{N}_4$ -isoleucine combined with a MudPIT peptide separation method (Prieto *et al.* 2008). The role of the antibiotic paromomycin on the global proteomes of susceptible and resistant strains of the protozoan parasite *Leishmania donovani* was examined with SILAC, using the conventional  $^{13}\text{C}_6$  L-lysine-HCl and  $^{13}\text{C}_6$ ,  $^{15}\text{N}_4$  L-arginine-HCl heavy isotopes (Chawla *et al.* 2011). Changes in the relative abundance and phosphorylation of protein components of the invasion motor complex during host cell invasion by the apicomplexan parasite *Toxoplasma* were also monitored by SILAC based quantitative proteomics (Nebl *et al.* 2011). While some progress has been made with single cell

parasites in cell culture, larger multicellular parasites are less amenable to SILAC labelling. Recent work by two groups has developed methodologies to perform SILAC based proteomics on the nematode *C. elegans*, using metabolically labelled *E. coli* as a food source in order to label the worms. Larance *et al.* (2011) characterized *C. elegans* protein abundance changes after heat shock treatment and Fredens *et al.* (2011) followed the *C. elegans* proteome response to the knockdown of the transcription factor nuclear hormone receptor 49 (NHR-49) with RNAi. Nearly 4700 proteins were identified (approx. 20% of predicted proteome) and 3470 of these quantified, with 330 significantly up- or down-regulated. SILAC based quantification benefits from high accuracy and the fact that the labelled 'heavy' proteome is essentially indistinguishable from the 'light' or normal proteome and can be combined early-on in the procedure, at the cell level or just after lysis, meaning that less variation or inaccuracies will be introduced during sample preparation and pre-fractionation before MS analysis. The success of these experiments opens up the possibility of using SILAC with other parasitic nematodes.

Many parasites cannot easily be labelled *in vivo* making SILAC an impractical approach. An alternative labelling technique is therefore to post-label by chemically modifying peptides or protein preparations from experiments using tagged stable isotope or fluorescent dyes. Difference gel electrophoresis (DIGE) can be used to label up to three different samples with a fluorescent dye. The samples are then mixed and analysed by 2D electrophoresis. Differences in protein abundance between the samples can be measured by excitation at different wavelengths and gel images are matched and analysed by image analysis software such as DeCyder™ (GE Healthcare). This technique helps reduce the variability between samples run on separate 2DE gels. DIGE has been used to monitor the changes in the host cell proteome to invasion by *Toxoplasma gondii*, highlighting significant changes in key metabolic pathways and in post-translational protein modification (Nelson *et al.* 2008), measuring the key protein changes during *Neospora* differentiation (Marugán-Hernández *et al.* 2010) and identifying changes in the abundance of proteins involved with energy metabolism in the head proteome of the *Anopheles* mosquito after infection with *Plasmodium* (Lefevre *et al.* 2007). The plasma proteomes of several individuals infected with *Leishmania donovani* were compared to control individuals using DIGE (Rukmangadachar *et al.* 2011) identifying several putative biomarkers. *Schistosoma japonicum* schistosomula from hosts with differing susceptibility to the parasite were also examined using DIGE (Hong *et al.* 2011) and several proteins were shown to be differentially expressed between schistosomula, highlighting the adaptation of *S. japonicum* to different host environments.

The most widespread of the stable isotope tag techniques is iTRAQ. Up to eight different samples can be labelled with isobaric tags that react with the primary amino groups of peptides. The samples can then be mixed and analysed in the same MS run. During MS the tags are fragmented into reporter groups with a different mass for each tag, the intensity of which can be used to derive the relative abundance of corresponding peptides in the starting sample. Protein abundance changes in the malaria parasite *P. falciparum* following doxycycline treatment were measured using iTRAQ (Briolant *et al.* 2010) as was the differential protein expression over the life stages of *Trypanosoma congolense* (Eyford *et al.* 2011). The protein abundance differential measured using iTRAQ was used to distinguish putative mitochondrial proteins from co-purified contamination in *Giardia intestinalis* extracts (Jedelský *et al.* 2011). Combining quantitative iTRAQ proteomic profiling with transcriptomics showed that the expression of merozoite proteins in *Plasmodium falciparum* were regulated post-translationally during invasion pathway switching as an adaptation to variations of the host cell (Kuss *et al.* 2011).

#### Label-free quantitative proteomics in parasite systems

Label-free methodologies are becoming an increasingly popular and widely used first approach to quantitative proteomics, although they do present statistical and bioinformatic challenges. The principle behind this technique is that two samples can be compared without the need to modify or label protein preparations using the mass spectra alone. Spectral counting is a straightforward way to obtain semi-quantitative data on protein abundances within a sample and is often automatically performed on peptide identification data sets (e.g. emPAI in the Mascot search engine). This way quantitative data can be obtained from intensive highly fractionated shotgun identification proteomics. Schrimpf *et al.* (2009) have identified more than half of the predicted *C. elegans* proteins and using a modified spectral counting algorithm estimate the abundances of over 1000 proteins. This information was used to validate gene models and to compare the abundance of orthologous proteins in another organism. Bennuru *et al.* (2011) identified approximately 60% of the predicted gene products from adults, microfilariae, L3 larvae and ES products of the lymphatic filarial worm *Brugia malayi*. Abundance was estimated using simple spectral counting. Several high-throughput proteomics studies have focused on the proteome of *Schistosoma* parasites. The proteomes of several developmental stages of *S. japonicum* as well as tissues at the host-parasite interface were characterized in tandem with transcriptomics (Liu *et al.* 2006) in parallel with a proteomic study of the host proteins that are associated with *S. japonicum*

(Liu *et al.* 2007) and the parasites excretory/secretory proteins (Liu *et al.* 2009). Label-free quantitative proteomics of the early gametocyte phase of *P. falciparum* identified that proteins involved in erythrocyte remodelling were enriched (Silvestrini *et al.* 2010).

An alternative to spectral counting is to align separate LC-MS/MS runs of peptide mixtures and to calculate the differences in intensities of the same peptides detected in each run. This approach tends to be more accurate than spectral counting but requires expensive instrumentation to ensure reproducibility. Software, both commercial and free, is available to perform the alignment and ion intensity comparison functions. To date no examples of its application to host-proteome interactions have been published, but the potential for increased accuracy with this approach means that its application is unlikely to be overlooked by the field.

#### Absolute protein quantification

The approaches described so far have been in terms of relative quantification. The precise determination of the concentration of specific protein is known as absolute quantification. This is performed by stable isotope dilution, where a reference standard to which a stable isotope has been incorporated is added in known amounts to the sample mixture. The reference peptide is in all respects the same as the analyte peptide apart from the mass difference due to the isotopic label. When analyzed by MS the ratio of the intensities of the analyte and standard ions will allow the calculation of the concentration of the analyte as the concentration of the standard is known. The reference standards can be synthesized chemically individually, e.g. AQUA peptides (Gerber *et al.* 2003) or expressed from synthetic genes in *E. coli* using stable isotopically enriched media e.g. QconCAT Absolute quantification (Beynon *et al.* 2005), can provide the copy number of proteins in a cell under a certain state, an important input for systems biology modelling. These absolute values can be compared across several studies including those from different groups and be more easily integrated into transcriptomic and metabolic system data. The use of absolute quantification in parasite studies is only just starting to emerge. The relative and absolute amounts of *Schistosoma* tegument proteins have been determined using a QconCAT methodology (Castro-Borges *et al.* 2011). Of course any isotope dilution experiment requires that the analyte be already characterized (in contrast to discovery based proteomics).

#### Targeted quantitative proteomics

So far we have discussed high-throughput proteomics approaches which have made no *a priori*

assumptions as to the proteins to be analysed. Whilst this is highly valuable as a tool to help generate assumption-free hypotheses, it has the disadvantage of pushing separation techniques and instrumentation to its limits because of the diversity of protein species and the dynamic range of the targets. Such an approach is unlikely ever to unravel a truly complete proteome and is even less able to make good quantitative measurements across such a wide dynamic range. Targeted proteomics approaches are now being developed that enable the MS instrument measurements to be focused only on specific peptides from predetermined sets of selected proteins. These techniques have the potential for greater accuracy since the instrumentation and bioinformatics can be tuned to a relatively small sub-set of protein targets. Selected reaction monitoring (SRM) is typically used to selectively record fragmentation events that are specific for the peptides of interest (Lange *et al.* 2008; Bertsch *et al.* 2010). Targeted proteomics allows a rapid and accurate quantitative profiling of a repeated set of proteins across samples from different conditions. A triple quadrupole mass spectrometer (QQQ) is used to achieve peptide targeting in SRM experiments. The first quadrupole is used to isolate precursor ions in a narrow mass range and the selected ions are then fragmented in the second quadrupole. The third quadrupole is used to specifically detect a set of fragment ions that is characteristic for the target peptides. This sequential isolation of targeted ions enables a great reduction in background noise and makes this approach the most sensitive MS strategy available. An application of this strategy has been used to obtain the absolute quantity of low abundance proteins in *P. falciparum* crude cell extracts (Southworth *et al.* 2011).

## PROTEIN MODIFICATIONS AND INTERACTIONS

### *Post-translational modifications*

Proteins can undergo a great range of chemical modifications after translation. These post-translational modifications (PTMs) can determine protein localisation, activity state, turnover, structure, as well as interactions with other proteins, cells or organisms. While >1000 PTMs have been assembled in UNIMOD ([www.unimod.org](http://www.unimod.org)) with more likely to be found, PTMs are generally not well characterized in parasites. Understanding the roles of these PTMs in parasite regulation, survival and pathogenesis as well their contribution to the adaptation and evolution of the host require both highly sensitive and precise detection and reliable high-throughput methodologies to quantify protein changes in a complex mixture.

The low dynamic range, stability and sometimes transient changes of protein modifications combined with attempting to relate these modifications to

biological events create a challenge to modern technologies. There are many proteomic approaches to studying PTMs, ranging from bottom-up and top-down mass spectrometry, gel and gel-free techniques and affinity based methodologies.

Classically, gel-based techniques paired with mass spectrometry (MS) have been used to highlight PTMs. Two-dimensional electrophoresis separates proteins by their charge and molecular weight. The resolving power of this technique can separate differentially expressed modified forms of a given protein. Further selectivity in the detection of specific PTMs by using certain stains, metabolic labelling, antibodies or specific probes can aid in detection and identification of PTMs and have been used to identify phosphorylated proteins in erythrocytes infected by the human malaria parasite *Plasmodium falciparum* (Wu *et al.* 2009). Fluorescent or colorimetric stains for gels or western blots (e.g. Pro-Q Diamond stain, Invitrogen) allow simple selective detection of phosphoproteins (Nunes *et al.* 2010). A more sensitive technique for phosphoprotein detection is the radiolabelling of proteins by  $^{32}\text{P}$  incorporation (Leykauf *et al.* 2010) or immunoblotting (Wu *et al.* 2009). Glycosylated protein can also be detected using specific stains such as Pro-Q-Emerald (Invitrogen) conjugated lectins or differential glycosidase digestion (Rebello *et al.* 2011).

The scarcity of many PTMs requires the enrichment of the sub-population of select modified proteins. Affinity based enrichment can be performed at the protein or peptide level and targets specific or groups of PTMs. Immobilised metal ion affinity chromatography (IMAC) utilises the affinity of chelated Fe(III) or Ga(III) ions to the phosphate group of phosphopeptides. Crude protein mixtures from *Leishmania donovani* extracts were enriched for phosphoproteins using IMAC then digested with trypsin and analysed for life stage specific phosphoprotein abundance (Hem *et al.* 2010). Oxides of metals such as titanium, zirconium and aluminium, can also be used to isolate phosphoproteins selectively. Anti-pSER/pThr/pTyr antibodies also facilitate the enrichment of phosphoproteins by immunoprecipitation followed by separation by 1-DE or 2-DE gels. Specific antibodies can also act to isolate other PTMs, for example the global analysis of acetylation, methylation and nitration of peptides. Carbohydrate-binding proteins (lectins) are used to enrich for glycoproteins and glycopeptides using affinity chromatography. Affinity resins that bind polyubiquitin protein conjugates are commercially available.

PTMs can also be specifically targeted by chemical derivatization. Affinity tags can be introduced by beta elimination of phosphoric acid from pSER or pThr followed by the addition of affinity groups such as biotin to allow enrichment of phosphoproteins by chromatography. Solid phase extraction of glycopeptides can be achieved by immobilisation of

carbohydrates to a hydrazide activated resin followed by release by PNGase F and analysis with LC-MS/MS. Hydrophilic interaction liquid chromatography (HILIC)-based methods can also be used to isolate glycopeptides. Recently hexapeptide libraries have been applied to large-scale glycomics analysis (Huhn *et al.* 2011) and arrays have been used to profile glycans (Lepenies and Seeberger, 2010; Lonardi *et al.* 2010; Ruhaak *et al.* 2010).

Covalent PTMs of cysteine are mediators of redox regulation and signalling. Cysteines are involved in many biochemical reactions, crucial in redox reactions and when involved in disulfide bonds, influence protein structure and stability. S-nitrosylation, s-glutathioylation, palmitoylation and prenylation are all PTMs of cysteine found in parasites (Jortzik *et al.*, 2011).

#### *Protein-protein interactions*

Most cellular processes are governed by protein-protein interactions. These can range from the interaction of two proteins to the formation of large macromolecular complexes consisting of many different proteins in differing ratios. Interactions can be strong or transient. There are several methods for experimentally determining protein-protein interaction. The two most widely used are the yeast two-hybrid system and affinity purification coupled with MS (Tandem affinity purification, TAP), where the c-terminus of a bait protein is fused to a TAP tag. The TAP tag consists of a calmodulin-binding peptide (CBP) and a IgG binding domain from protein A, separated by a TEV protease cleavage site. The TAP-tagged bait protein is isolated from a cell lysate using IgG-coated beads. After washing, the bait protein is released from the beads by incubating with TEV protease. A second round of purification uses calmodulin-coated beads to isolate the bait protein (and associated binding partners) via the CBP tag. Bound proteins are eluted and analysed with SDS-PAGE and mass spectrometry. The yeast two-hybrid system measures the interaction of a bait protein which is fused to the DNA binding domain of the yeast protein Gal4 and the prey protein, which is fused to the transactivation domain of Gal4. When the bait and prey interact, a downstream reporter gene is activated. A variation of this technique has been used to investigate interaction networks in *P. falciparum* (LaCount *et al.*, 2005), identifying 2846 unique interactions involving 1312 proteins and highlighting a group of interacting proteins involved with host cell invasion, including 19 uncharacterized proteins.

#### BIOINFORMATICS RESOURCES FOR PARASITE PROTEOMICS

There are two major limiting steps in any proteomics experiment. One is the limit imposed by the MS

instrumentation itself and second, but equally as crucial, is the bioinformatic processing of the large quantity of data generated by modern proteomics platforms. The involvement of various bioinformatics tasks in processing and interpreting proteomics data is summarised in Fig. 2. The scale and complexity of the data generated by such a workflow means that it is essential to develop integrated database pipelines if complex proteomics data are generated for multiple host-parasite systems. Proteomics databases for parasites are an essential component of ensuring that these data are stored and rendered accessible for easy use by the community. Subsequently, the focus is then on downstream interpretation in relation to protein function and localization prediction, pathway and network analysis, since these are the aspects of bioinformatics which have the potential to turn an elegant data gathering exercise into one which can reveal genuine insights into function. KEGG (Kanehisa and Goto, 2000), MetaCyc (Caspi *et al.* 2011) and Reactome (Croft *et al.* 2011) are such tools developed to facilitate pathway browsing and data analysis.

#### *Data repositories for parasite proteomics data*

Several public repositories host proteomics data for the research communities, such as the Proteomics identifications database (PRIDE) (Jones *et al.* 2006), the Global Proteome Machine databases (GPMDB) (Craig *et al.* 2004) and PeptideAtlas (Desiere *et al.* 2006). While these databases are useful for storage and re-querying of proteomics data generated, the integration of proteomics data with organism-specific genomic and proteomics resources provides an essential technical step to data interpretation (Xia *et al.* 2008). For parasites the most advanced example of this is the hosting of proteomics data in EuPathDB (Aurrecochea *et al.* 2010). Although these data deal with only protozoan parasites the easily accessible format has opened up proteomics data to the entire research community in a way that was difficult to envisage, even a few years ago. Proteomics data repositories for helminths are generally less unified and well resourced, possibly reflecting the fact that the respective genome sequencing projects lag behind those of the protozoa, although this is likely to change in the near future. WormBase (Yook *et al.* 2011), for example, now supports 15 helminth species with growing proteomics data resources.

#### *Proteomics resources at EuPathDB*

EuPathDB acts as a portal to eukaryotic pathogens (Aurrecochea *et al.* 2010). It is an integrated genome database composed of a family of dedicated pathogen databases including PlasmoDB, ToxoDB (also serves *Neospora caninum*), CryptoDB, GiardiaDB,



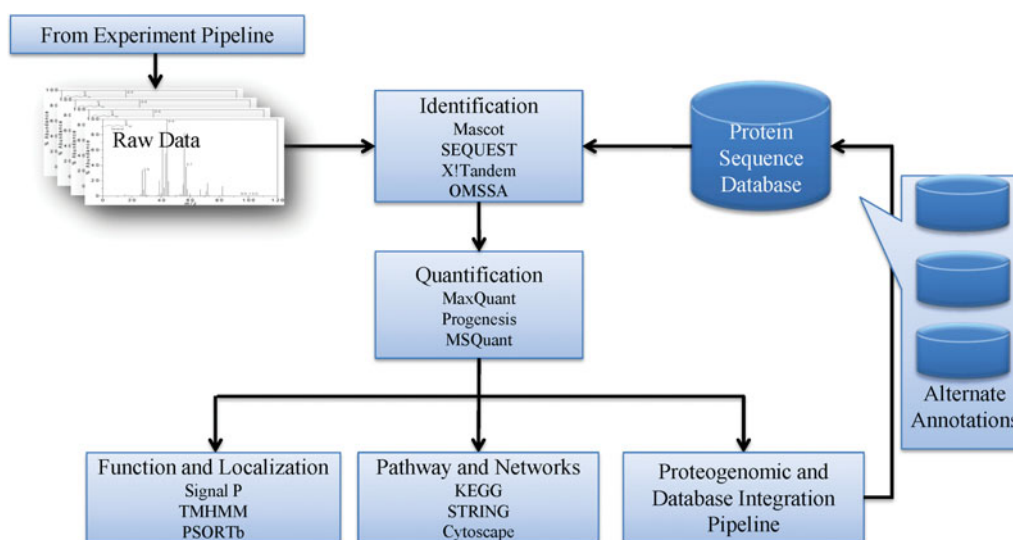


Fig. 2. A bioinformatics pipeline for large-scale proteomics data storage, querying and interpretation. The involvement of various bioinformatics tasks in processing and interpreting proteomics data is summarised. Raw mass spectra data collected from an experimental pipeline are subjected to proteomics identification packages, where protein identifications are acquired based on the comparison between raw data and protein sequence databases. The results are then analysed by appropriate quantification packages where relative or absolute quantifications of the analyte proteins are calculated. Protein function and localization prediction and pathway analysis tools have been developed to infer the biological meaning of the identification and quantification data. Proteogenomic and database integration pipelines are available to facilitate data integration with online databases and improve genome annotation using alternate annotations.

TrichDB, TriTrypDB, AmoebaDB and MicrosporidiaDB. More detailed introductions to these databases have been published through online tutorials, individual websites and journal publications (Heiges *et al.* 2006; Gajria *et al.* 2008; Aurrecochea *et al.* 2009a,b, 2010; Aslett *et al.* 2010). In this review, a selection of the important features that are relevant to parasite proteomics research will be highlighted.

Proteomics data of *T. gondii* and *C. parvum* pioneered the full integration of proteomics data into EuPathDB (Sanderson *et al.* 2008; Xia *et al.* 2008). The latest version of EuPathDB (v2.12) hosts proteomics data for 26,035 proteins from 16 species (<http://eupathdb.org/eupathdb/>). Data analysis tools have been developed to facilitate the browsing, functional prediction and comparison of proteomics data with other types of genomic data.

Fig. 3 shows how protein expression data for particular genes can be viewed on individual gene record pages using ToxoDB v7.2 as an example, where colour coded peptides are mapped to the gene sequence according to the experiments that identified them. Protein expression evidence can also be queried according to the experiments and samples using 'Identify Genes based on Mass Spec. Evidence.' tool (Fig. 4). Once a group of proteins of interest has been acquired based either on existing proteomics studies or user supplied lists, additional data analysis tools can be used to interact with genome information, functional predictions and other type of genome wide 'omics' data using 'Add Step.' tool in the results page. Fig. 4 also shows the comparison of proteomics data with mRNA expression data from

RNA sequencing experiments, where the relationship of the two datasets can be analysed side by side. This function has vastly improved the interaction of proteomics data with existing knowledge and other genomics data and is a first step to enabling a truly 'trans-omics' approach to studying parasite biology. In addition to these text based tools, the Generic Genome Browser (GBrowse) has also been incorporated in ToxoDB to improve visualization of data mining. GBrowse is a web-based application for displaying genomic annotations and other features developed by GMOD (Generic Model Organism Database project) (Stein *et al.* 2002). The implementation of ToxoDB allows expressed peptides to be visualized in relation to various gene models and the genomic region from which the sequence is predicted to have been produced. Fig. 4 shows the peptides identified from one of the proteomics datasets aligned with unified MS/MS peptides and one RNA-Seq datasets for a particular gene TGME49\_100100.

A more recent development involves the use of an automated proteogenomic pipeline for integration of mass spectrometry (MS) based proteomics evidence into genome databases (Krishna *et al.* 2011). The pipeline uses MS data for confirming official gene models on the database, but also examines whether there is supporting evidence for alternate annotations at particular loci and for identifying novel genes. The pipeline is currently being used to assist proteomics data integration and gene annotation for a number of EuPathDB supported species, including *T. gondii*, *N. caninum* and *C. parvum*. Similar to other generic tools developed in EuPathDB, the algorithms used in

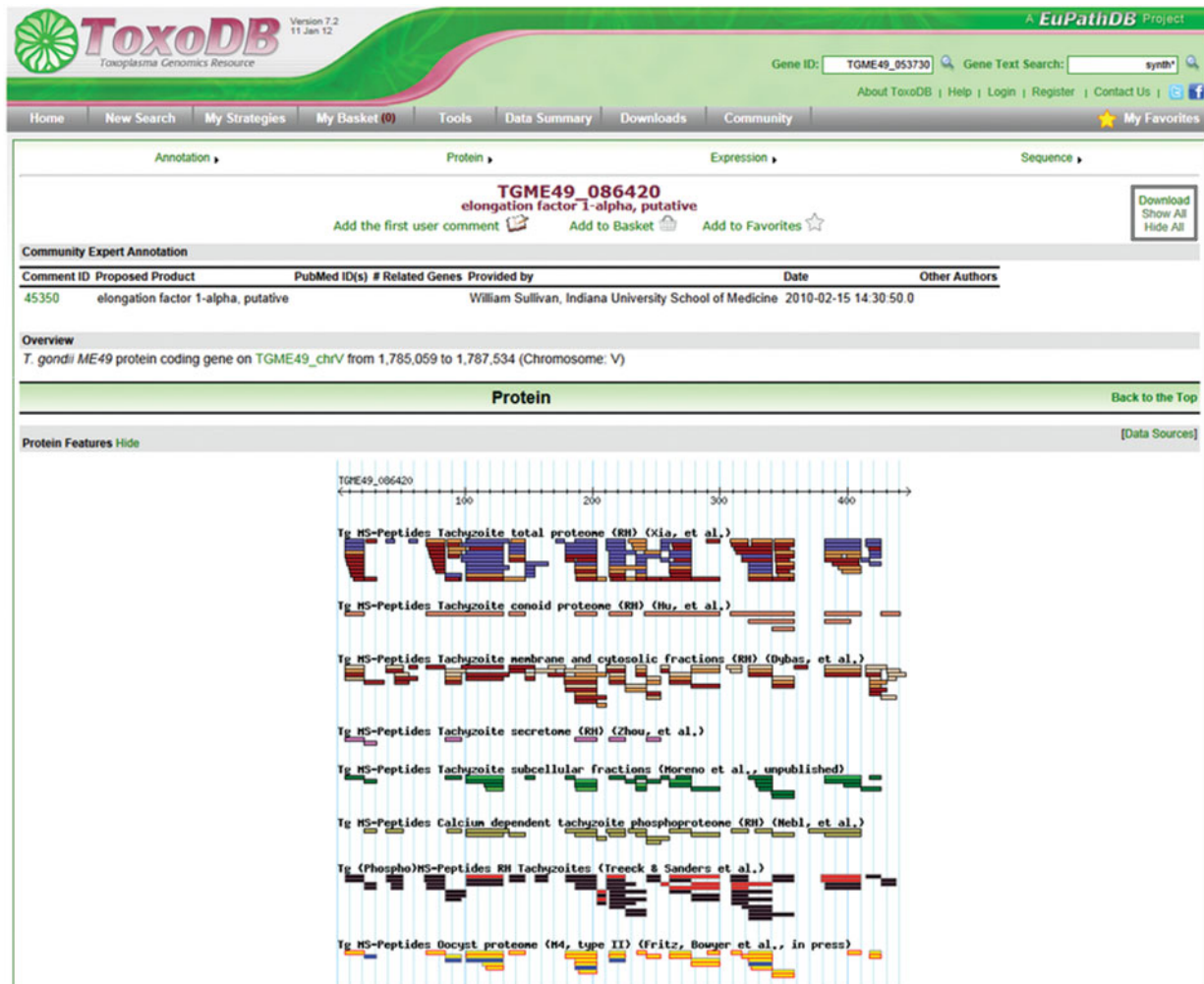


Fig. 3. Screenshot of protein expression data of individual genes displayed on ToxoDB. A screenshot illustrating protein expression data for gene a *Toxoplasma gondii* gene TGME49\_086420 viewed on an individual gene record page. Colour coded peptides are mapped to the gene sequence according to the various experiments and laboratories that identified them.

the pipeline are not specific to any organism, thus enabling the pipeline to be used in conjunction with any genome.

Around 590 *T. gondii* genes are annotated to ToxoDB metabolic pathways, which were automatically reconstructed from KEGG pathways maps. Many of the pathways are not organism specific and based on the presence of one or two enzymes present in other pathways or on the basis of less precise partial EC number annotations. A recent effort has been made to develop a manually curated metabolic pathways database for apicomplexan parasites using biochemical and physiological evidences available in the literature as well as proteomics evidence and gene annotations available on EuPathDB. The result is currently hosted at Liverpool Library of Apicomplexan Metabolic Pathways ([www.llamp.net](http://www.llamp.net)).

With the increasing availability of quantitative parasite proteomics data and that collected from host cell systems, new interfaces need to be developed to enable analysis of these data. While quantitative proteomic data can probably be visualised in a similar

way to transcriptomic data, a lack of data analysis tools for other protein features is holding back the wider integration and analysis of multiple ‘omics’ data. Not least is the complete lack of a platform on which host response data can be displayed and analysed alongside parasite data—something that is essential if we are to use these tools to answer questions about host-parasite interactions.

In a recent development outside EuPathDB, MOPED (Model Organism Protein Expression Database) provides for the rapid browsing of protein expression information in several model organisms including *Caenorhabditis elegans* (Kolker *et al.* 2012). It also offers data comparison tools to produce overlap plot and heatmaps between existing data and user-uploaded data with user-specified expression value thresholding (Kolker *et al.* 2012). A local data analysis tool, GProX, (Graphical Proteomics Data Explorer) is also available for comprehensive analysis, inspection and visualization of quantitative proteomics data (Rigbolt *et al.* 2011). Although the MOPED and GProX are currently

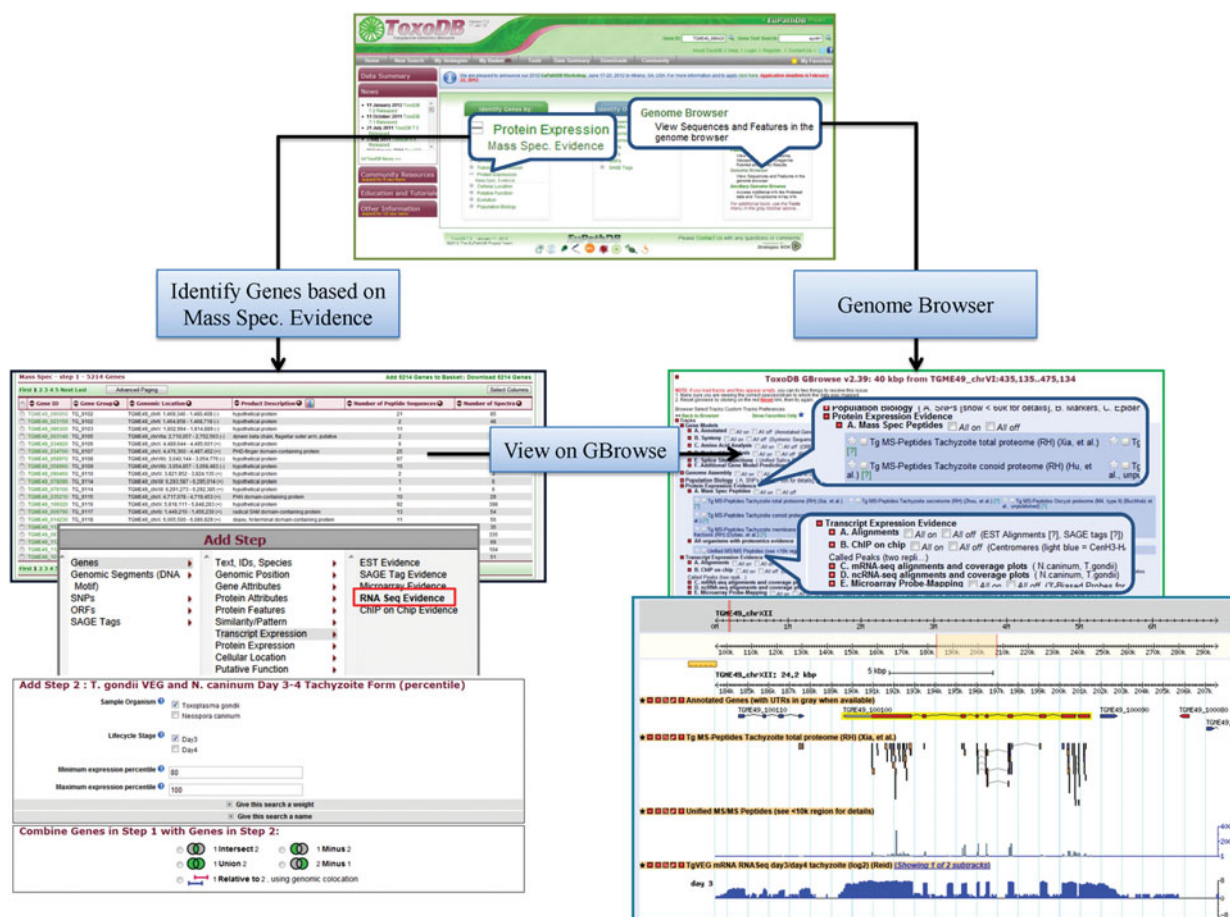


Fig. 4. Workflows for two important tools for proteomics informatics research on ToxoDB. Starting on ToxoDB the frontpage, protein expression evidence can be queried according to the experiments and samples using 'Identify Genes based on Mass Spec. Evidence.' tool. The results can be compared with other genome information using 'Add Step' tool in the result page. The example shown is the comparison of proteomics results with one of the RNA sequencing datasets (RNA Seq Evidence). 'Genome Browser' can also be accessed from ToxoDB frontpage, where various genome information (named tracks) can be selected to be aligned in a defined genomic region. Individual genes from proteomics results can also be viewed directly in GBrowse from the result page. The example shown is proteomics evidence of gene TGME49\_100100 aligned with RNA Seq evidence acquired.

limited to proteomics data only, a similar interface could be readily adapted on genome databases to facilitate large scale quantitative trans-omics studies.

#### Post-identification bioinformatic analysis

Once the protein identification and quantification has been determined, signature-based resources can be used to infer function and subcellular localizations where one or more protein signatures can be identified. Protein signatures are defined by either a regular expression method that shows patterns of conserved amino acid residues (Sigrist *et al.* 2002) or the Hidden Markov Model (HMM) method which provides a statistical profile based on probabilities of finding an amino acid at a given position in the sequence (Krogh *et al.* 1994). There are many publicly available signature databases of protein families and domains, including sequence-based PROSITE (Sigrist *et al.* 2002), Pfam (Finn *et al.* 2008), PRINTS (Attwood *et al.* 2003) PANTHER

(Mi *et al.* 2007) and structure-based SUPERFAMILY (Wilson *et al.*, 2007) and Gene3D (Yeats *et al.* 2006). Protein signatures can be used in combination to predict protein functions. For example, proteins with no significant sequence similarity but which have similar functions might be expected to share some common features like post-translational modifications, protein-sorting signals and similar subcellular localizations. In parasitology research, the identification of signal peptide and transmembrane domains are of special interests. The entry of virtually all proteins into the secretory pathway is controlled by signal peptides (Gierasch, 1989; von Heijne, 1990) and transmembrane proteins support essential biological functions acting as receptors, transporters or channels, which is essential in host-parasite interactions (Dowse and Soldati, 2005; O'Donnell *et al.* 2006; Baxt *et al.* 2008).

Universal software packages were developed to predict certain protein features based on a set of trained rules, such as Signal P for signal peptide

prediction (Bendtsen *et al.* 2004), TMHMM for transmembrane domain prediction (Krogh *et al.* 2001) and PSORTb for general sub-cellular localization prediction (Yu *et al.* 2010). Additional organism-specific prediction tools and databases were also developed to predict important features in the organism under study, with more targeted training data. For example, PSEApred (Verma *et al.* 2008), PlasMit (Bender *et al.* 2003) and ApiLoc (<http://apiloc.biochem.unimelb.edu.au/apiloc/apiloc>) for apicomplexan parasites.

#### TRANSCRIPTOMICS AND PROTEOMICS IN PARASITE SYSTEMS

Gene expression process can be simply summarized using the central dogma of Gene-Transcription-Translation. However, many levels of control and regulation events during this process introduce uncertainty to the system where a simple one-to-one expression is not achieved. Early comparisons between transcriptomics data and proteomics data have generally indicated a weak correlation (de Sousa Abreu *et al.* 2009; Maier *et al.* 2009). The same phenomenon has also been observed in some parasites and has been summarized in reviews of models systems such as *Apicomplexa* (Kooij *et al.* 2006; Wastling *et al.* 2009) and *Schistosoma* (Hokke *et al.* 2007). These studies highlighted the discrepancies of mRNA and protein expression and the important involvement of the regulation of expression, which is likely to involve biological explanations such as selective protein degradation and variations in protein turn-over rates (Yen *et al.* 2008; Doherty *et al.* 2009) as well as post-translational regulations such as mRNA decay and translational repression (Hakimi and Deitsch, 2007; Shock *et al.* 2007; Filipowicz *et al.* 2008).

However, despite the biological reasons, technical factors have also restricted the full insight of the correlation between transcriptome and proteome, namely the incomplete transcriptomics survey and rather basic quantitative proteomics techniques. The lack of simultaneously collected sample for both proteomics and transcriptomics analyses also contributes to the discrepancies observed. The recent introduction of high-throughput sequencing of mRNA and microRNA (Hall, 2007; Lister *et al.* 2009; Wang *et al.* 2009), and the development of quantitative proteomics techniques, in particular the absolute quantification methods, have significantly improved our ability to measure the correlation between transcriptome and proteome in biological systems. Studies carried out by Schwanhäusser *et al.* (2011) on mouse NIH3T3 cells and Nagaraj *et al.* (2011) on human Hela cells are among the first large-scale comparisons between RNA-Seq data and intensity-based absolute quantification (iBAQ) proteomics data from simultaneously collected samples

and report correlation coefficient of between 0.41 and 0.6 (Spearman) (Nagaraj *et al.* 2011; Schwanhäusser *et al.* 2011).

#### CONCLUSIONS

In the first decade of the 21st Century we have witnessed the beginnings of a new era in biomedical research in which we see scientific reductionism being challenged by the availability of vast collections of biological information. Much of this information, gathered in defined experimental contexts, constitutes genomic and genome-related expression data generated by technologies such as transcriptomics and proteomics. Together with other related data such as metabolomics, glycomics and lipidomics, attention has turned to ways in which individual streams of information can be processed as a whole, rather than remaining as an isolated descriptor of the action of individual components. The desire to combine these data into a unified system is in part driven by the pragmatic view that biological events are the result of the concerted action of individual system components. However, more than that, biological systems also exhibit *emergent properties*; and these properties cannot be fully predicted by the sum of the component parts. In host-parasite interactions, *emergent properties* are likely to be even more complex because the host-parasite relationship is a product of the interaction of two genetically distinct biological systems. Such systems interact in often unpredictable ways—after all, if they were completely predictable, science would have been far more successful in developing vaccines, drugs that were less susceptible to resistance and solved some of the paradoxes of parasite evolutionary biology. It is an interesting observation that the most successful vaccines are live-attenuated vaccines whose development has often by-passed a detailed understanding of the ‘black-box’ of the host-parasite system itself. When we try and break our knowledge down into its component parts to develop sub-unit vaccines for example, the outcomes are far less impressive. Moreover, efforts to understand the system as a whole must extend not just in the scope of the data, but in scale also; instead of considering the system only at an individual organism level we must be cognisant of population responses of both parasites and hosts.

So has Descartes’ reductionist vision really had its time—soon to be eclipsed by a return to a holistic view of the system? And if so, what are the implications for the facilitating technologies of systems biology such as proteomics? As with most paradigm shifts, any change will represent more of an evolution than a revolution. Few will seriously contend that we no-longer need to examine the fine detail of discrete biological components and their biochemical actions. On the contrary, experiments on



discrete gene function are arguably as important as ever since reliable information on protein function, localisation and modification is an essential element in generating accurate systems models. It is worth considering also the vast number of genes in parasite genome databases that are still annotated as 'unknown function', even though protein expression data generated by proteomics experiments such as those described here show clearly that they exist as functional molecules. Similarly, our understanding of protein expression, including information on post-translational modifications and absolute protein expression and turnover requires far more refinement before we can be confident that it will fulfil its role in meaningful systems modelling. As we have seen in this review, coverage of accurate quantitative proteomics is still relatively poor and focused on proteins with high to medium level expression. A trade-off still exists between our ability to identify proteins (which we can do in large numbers) and our ability to make accurate quantitative measurements. Finally, the non-linear relationship between transcription, protein expression and activity, still needs to be defined in host and parasite. These advances will require continued advancement of instrumentation and concomitant development in proteome bioinformatics as dramatic as any of those we have seen in the last decade.

## ACKNOWLEDGEMENTS

We thank the British Society for Parasitology for support in publishing this review.

## REFERENCES

- Aslett, M., Aurrecochea, C., Berriman, M., Brestelli, J., Brunk, B. P., Carrington, M., Depledge, D. P., Fischer, S., Gajria, B., Gao, X., Gardner, M. J., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Hertz-Fowler, C., Houston, R., Innamorato, F., Iodice, J., Kissinger, J. C., Kraemer, E., Li, W., Logan, F. J., Miller, J. A., Mitra, S., Myler, P. J., Nayak, V., Pennington, C., Phan, I., Pinney, D. F., Ramasamy, G., Rogers, M. B., Roos, D. S., Ross, C., Sivam, D., Smith, D. F., Srinivasamoorthy, G., Stoeckert, C. J., Jr., Subramanian, S., Thibodeau, R., Tivey, A., Treatman, C., Velarde, G. and Wang, H. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Research* **38** (Database issue), D457–462. doi: 10.1093/nar/gkp851.
- Attwood, T. K., Bradley, P., Flower, D. R., Gaulton, A., Maudling, N., Mitchell, A. L., Moulton, G., Nordle, A., Paine, K., Taylor, P., Uddin, A. and Zygouri, C. (2003). PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Research* **31**, 400–402. doi: 10.1093/nar/gkg030.
- Aurrecochea, C., Brestelli, J., Brunk, B. P., Carlton, J. M., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Innamorato, F., Iodice, J., Kissinger, J. C., Kraemer, E., Li, W., Miller, J. A., Morrison, H. G., Nayak, V., Pennington, C., Pinney, D. F., Roos, D. S., Ross, C., Stoeckert, C. J., Jr., Sullivan, S., Treatman, C. and Wang, H. (2009a). GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic protist pathogens *Giardia lamblia* and *Trichomonas vaginalis*. *Nucleic Acids Research* **37** (Database issue), D526–530. doi: 10.1093/nar/gkn631.
- Aurrecochea, C., Brestelli, J., Brunk, B. P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Innamorato, F., Iodice, J., Kissinger, J. C., Kraemer, E., Li, W., Miller, J. A., Nayak, V., Pennington, C., Pinney, D. F., Roos, D. S., Ross, C., Stoeckert, C. J., Jr., Treatman, C. and Wang, H. (2009b). PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Research* **37** (Database issue), D539–543. doi: 10.1093/nar/gkn814.
- Aurrecochea, C., Brestelli, J., Brunk, B. P., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Innamorato, F., Iodice, J., Kissinger, J. C., Kraemer, E. T., Li, W., Miller, J. A., Nayak, V., Pennington, C., Pinney, D. F., Roos, D. S., Ross, C., Srinivasamoorthy, G., Stoeckert, C. J., Jr., Thibodeau, R., Treatman, C. and Wang, H. (2010). EuPathDB: a portal to eukaryotic pathogen databases. *Nucleic Acids Research* **38** (Database issue), D415–419. doi: 10.1093/nar/gkp941.
- Baxt, L. A., Baker, R. P., Singh, U. and Urban, S. (2008). An *Entamoeba histolytica* rhomboid protease with atypical specificity cleaves a surface lectin involved in phagocytosis and immune evasion. *Genes and Development* **22**, 1636–1646. doi: 10.1101/gad.1667708.
- Beck, M., Claassen, M. and Aebersold, R. (2011). Comprehensive proteomics. *Current Opinion in Biotechnology* **22**, 3–8. doi: 10.1016/j.copbio.2010.09.002.
- Bender, A., van Dooren, G. G., Ralph, S. A., McFadden, G. I. and Schneider, G. (2003). Properties and prediction of mitochondrial transit peptides from *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* **132**, 59–66. doi: 10.1016/j.molbiopara.2003.07.001.
- Bendtsen, J. D., Nielsen, H., von Heijne, G. and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* **340**, 783–795. doi: 10.1016/j.jmb.2004.05.028.
- Bennuru, S., Meng, Z., Ribeiro, J. M. C., Semnani, R. T., Ghedin, E., Chan, K., Lucas, D. a., Veenstra, T. D. and Nutman, T. B. (2011). Stage-specific proteomic expression patterns of the human filarial parasite *Brugia malayi* and its endosymbiont *Wolbachia*. *Proceedings of the National Academy of Sciences, USA* **108**, 9649–9654. doi: 10.1073/pnas.1011481108.
- Bennuru, S., Semnani, R., Meng, Z., Ribeiro, J. M. C., Veenstra, T. D. and Nutman, T. B. (2009). *Brugia malayi* excreted/secreted proteins at the host/parasite interface: stage- and gender-specific proteomic profiling. *PLoS Neglected Tropical Diseases* **3**, e410. doi: 10.1371/journal.pntd.0000410.
- Bertsch, A., Jung, S., Zerck, A., Pfeifer, N., Nahnsen, S., Henneges, C., Nordheim, A. and Kohlbacher, O. (2010). Optimal de novo design of MRM experiments for rapid assay development in targeted proteomics. *Journal of Proteome Research* **9**, 2696–2704. doi: 10.1021/pr1001803.
- Beynon, R. J., Doherty, M. K., Pratt, J. M. and Gaskell, S. J. (2005). Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nature Methods* **2**, 587–589. doi: 10.1038/nmeth774.
- Braschi, S. and Wilson, R. A. (2006). Proteins exposed at the adult schistosome surface revealed by biotinylation. *Molecular and Cellular Proteomics* **5**, 347–356. doi: 10.1074/mcp.M500287-MCP200.
- Briolant, S., Almeras, L., Belghazi, M., Boucomont-Chapeaublanc, E., Wurtz, N., Fontaine, A., Granjeaud, S., Fusai, T., Rogier, C. and Pradines, B. (2010). *Plasmodium falciparum* proteome changes in response to doxycycline treatment. *Malaria journal* **9**, 141. doi: 10.1186/1475-2875-9-141.
- Brownridge, P. and Beynon, R. J. (2011). The importance of the digest: proteolysis and absolute quantification in proteomics. *Methods* **54**, 351–360. doi: 10.1016/j.ymeth.2011.05.005.
- Caspi, R., Altman, T., Dreher, K., Fulcher, C. A., Subhraveti, P., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Pujar, A., Shearer, A. G., Travers, M., Weerasinghe, D., Zhang, P. and Karp, P. D. (2011). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research* **2011**. doi: 10.1093/nar/gkr1014.
- Castro-Borges, W., Simpson, D. M., Dowle, A., Curwen, R. S., Thomas-Oates, J., Beynon, R. J. and Wilson, R. A. (2011). Abundance of tegument surface proteins in the human blood fluke *Schistosoma mansoni* determined by QconCAT proteomics. *Journal of Proteomics* **74**, 1519–1533. doi: 10.1016/j.jpro.2011.06.011.
- Chawla, B., Jhingran, A., Panigrahi, A., Stuart, K. D. and Madhubala, R. (2011). Paromomycin Affects Translation and Vesicle-Mediated Trafficking as Revealed by Proteomics of Paromomycin -Susceptible -Resistant *Leishmania donovani*. *PLoS One* **6**, e26660. doi: 10.1371/journal.pone.0026660.
- Cohen, A. M., Rumpel, K., Coombs, G. H. and Wastling, J. M. (2002). Characterisation of global protein expression by two-dimensional electrophoresis and mass spectrometry: proteomics of *Toxoplasma gondii*. *International Journal for Parasitology* **32**, 39–51. doi: 10.1016/S0020-7519(01)00308-3.
- Cox, J. and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and

- proteome-wide protein quantification. *Nature Biotechnology* **26**, 1367–1372. doi: 10.1038/nbt.1511.
- Craig, H., Wastling, J. M. and Knox, D. P. (2006). A preliminary proteomic survey of the *in vitro* excretory/secretory products of fourth-stage larval and adult *Teladorsagia circumcincta*. *Parasitology* **132**, 535–543. doi: 10.1017/S0031182005009510.
- Craig, R. and Beavis, R. C. (2004). TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**, 1466–1467. doi: 10.1093/bioinformatics/bth092.
- Craig, R., Cortens, J. P. and Beavis, R. C. (2004). Open source system for analyzing, validating, and storing protein identification data. *Journal of Proteome Research* **3**, 1234–1242. doi: 10.1021/pr049882h.
- Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., Jupe, S., Kalatskaya, I., Mahajan, S., May, B., Ndegwa, N., Schmidt, E., Shamovsky, V., Yung, C., Birney, E., Hermjakob, H., D'Eustachio, P. and Stein, L. (2011). Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Research* **39**(Database issue), D691–697. doi: 10.1093/nar/gkq1018.
- Curwen, R. S., Ashton, P. D., Johnston, D. A. and Wilson, R. A. (2004). The *Schistosoma mansoni* soluble proteome: a comparison across four life-cycle stages. *Molecular and Biochemical Parasitology* **138**, 57–66. doi: 10.1016/j.molbiopara.2004.06.016.
- de Godoy, L. M. F., Olsen, J. V., Cox, J., Nielsen, M. L., Hubner, N. C., Fröhlich, F., Walther, T. C. and Mann, M. (2008). Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**, 1251–1254. doi: 10.1038/nature07341.
- de Sousa Abreu, R., Penalva, L. O., Marcotte, E. M. and Vogel, C. (2009). Global signatures of protein and mRNA expression levels. *Molecular Biosystems* **5**, 1512–1526. doi: 10.1039/b908315d.
- Descartes, R. and Hall, T. S. (1972). *Treatise of Man. French text with translation and commentary by Thomas Steele Hall*, Cambridge, Harvard University Press.
- Desiere, F., Deutsch, E. W., King, N. L., Nesvizhskii, A. I., Mallick, P., Eng, J., Chen, S., Eddes, J., Loevenich, S. N. and Aebersold, R. (2006). The PeptideAtlas project. *Nucleic Acids Research* **34**(Database issue), D655–658. doi: 10.1093/nar/gkj040.
- Doherty, M. K., Hammond, D. E., Clague, M. J., Gaskell, S. J. and Beynon, R. J. (2009). Turnover of the human proteome: determination of protein intracellular stability by dynamic SILAC. *Journal of Proteome Research* **8**, 104–112. doi: 10.1021/pr800641v.
- Dowse, T. J. and Soldati, D. (2005). Rhomboid-like proteins in *Apicomplexa*: phylogeny and nomenclature. *Trends in Parasitology* **21**, 254–258. doi: 10.1016/j.pt.2005.04.009.
- Eyford, B. A., Sakurai, T., Smith, D., Loveless, B., Hertz-Fowler, C., Donelson, J. E., Inoue, N. and Pearson, T. W. (2011). Differential protein expression throughout the life cycle of *Trypanosoma congolense*, a major parasite of cattle in Africa. *Molecular and Biochemical Parasitology* **177**, 116–125. doi: 10.1016/j.molbiopara.2011.02.009.
- Filipowicz, W., Bhattacharyya, S. N. and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature Reviews Genetics* **9**, 102–114. doi: 10.1038/nrg2290.
- Finn, R. D., Tate, J., Mistry, J., Coghill, P. C., Sammut, S. J., Hotz, H. R., Ceric, G., Forslund, K., Eddy, S. R., Sonnhammer, E. L. and Bateman, A. (2008). The Pfam protein families database. *Nucleic Acids Research* **36**(Database issue), D281–288. doi: 10.1093/nar/gkm960.
- Florens, L., Washburn, M. P., Raine, J. D., Anthony, R. M., Grainger, M., Haynes, J. D., Moch, J. K., Muster, N., Sacci, J. B., Tabb, D. L., Witney, A. A., Wolters, D., Wu, Y., Gardner, M. J., Holder, A. A., Sinden, R. E., Yates, J. R. and Carucci, D. J. (2002). A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419**, 520–526. doi: 10.1038/nature01107.
- Fredens, J., Engholm-Keller, K., Giessing, A., Pultz, D., Larsen, M. R., Højrup, P., Møller-Jensen, J. and Førgeman, N. J. (2011). Quantitative proteomics by amino acid labeling in *C. elegans*. *Nature Methods* **8**, 845–847. doi: 10.1038/nmeth.1675.
- Gajria, B., Bahl, A., Brestelli, J., Dommer, J., Fischer, S., Gao, X., Heiges, M., Iodice, J., Kissinger, J. C., Mackey, A. J., Pinney, D. F., Roos, D. S., Stoekert, C. J., Jr., Wang, H. and Brunk, B. P. (2008). ToxoDB: an integrated *Toxoplasma gondii* database resource. *Nucleic Acids Research* **36**(Database issue), D553–556. doi: 10.1093/nar/gkm981.
- Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. and Gygi, S. P. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proceedings of the National Academy of Sciences, USA* **100**, 6940–6945. doi: 10.1073/pnas.0832254100.
- Gierasch, L. M. (1989). Signal sequences. *Biochemistry* **28**(3), 923–930. doi: 10.1021/bi00429a001.
- Guillou, F., Roger, E., Mone, Y., Rognon, A., Grunau, C., Theron, A., Mitta, G., Coustau, C. and Gourbal, B. E. (2007). Excretory-secretory proteome of larval *Schistosoma mansoni* and *Echinostoma caproni*, two parasites of *Biomphalaria glabrata*. *Molecular and Biochemical Parasitology* **155**, 45–56. doi: 10.1016/j.molbiopara.2007.05.009.
- Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H. and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* **17**, 994–999. doi: 10.1038/13690.
- Hakimi, M. A. and Deitsch, K. W. (2007). Epigenetics in *Apicomplexa*: control of gene expression during cell cycle progression, differentiation and antigenic variation. *Current Opinion in Microbiology* **10**, 357–362. doi: 10.1016/j.mib.2007.07.005.
- Hall, N. (2007). Advanced sequencing technologies and their wider impact in microbiology. *Journal of Experimental Biology* **210**, 1518–1525. doi: 10.1242/jeb.001370.
- Heiges, M., Wang, H., Robinson, E., Aurrecoechea, C., Gao, X., Kaluskar, N., Rhodes, P., Wang, S., He, C. Z., Su, Y., Miller, J., Kraemer, E. and Kissinger, J. C. (2006). CryptoDB: a *Cryptosporidium* bioinformatics resource update. *Nucleic Acids Research* **34**(Database issue), D419–422. doi: 10.1093/nar/gkj078.
- Hem, S., Gherardini, P. F., Osorio y Fortéa, J., Houldel, V., Morales, M. A., Watanabe, R., Pescher, P., Kuzyk, M. A., Smith, D., Borchers, C. H., Zilberstein, D., Helmer-Citterich, M., Namane, A. and Späth, G. F. (2010). Identification of *Leishmania*-specific protein phosphorylation sites by LC-ESI-MS/MS and comparative genomics analyses. *Proteomics* **10**, 3868–3883. doi: 10.1002/pmic.201000305.
- Hewitson, J. P., Harcus, Y. M., Curwen, R. S., Dowle, A. A., Atmadja, A. K., Ashton, P. D., Wilson, A. and Maizels, R. M. (2008). The secretome of the filarial parasite, *Brugia malayi*: proteomic profile of adult excretory-secretory products. *Molecular and Biochemical Parasitology* **160**, 8–21. doi: 10.1016/j.molbiopara.2008.02.007.
- Hokke, C. H., Fitzpatrick, J. M. and Hoffmann, K. F. (2007). Integrating transcriptome, proteome and glycome analyses of *Schistosoma* biology. *Trends in Parasitology* **23**, 165–174. doi: 10.1016/j.pt.2007.02.007.
- Hong, Y., Peng, J., Jiang, W., Fu, Z., Liu, J., Shi, Y., Li, X. and Lin, J. (2011). Proteomic analysis of *Schistosoma japonicum* schistosomulum proteins that are differentially expressed among hosts differing in their susceptibility to the infection. *Molecular and Cellular Proteomics: MCP* **10**(8), M110.006098. doi: 10.1074/mcp.M110.006098.
- Huhn, C., Ruhaak, L. R., Wührer, M. and Deelder, A. M. (2011). Hexapeptide library as a universal tool for sample preparation in protein glycosylation analysis. *Journal of Proteomics* **75**, 1515–1528. doi: 10.1016/j.jprot.2011.11.028.
- Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J. and Mann, M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Molecular and Cellular Proteomics* **4**, 1265–1272. doi: 10.1074/mcp.M500061-MCP200.
- Jedelský, P. L., Doležal, P., Rada, P., Pyrih, J., Smid, O., Hrdý, I., Sedinová, M., Marcinčíková, M., Voleman, L., Perry, A. J., Beltrán, N. C., Lithgow, T. and Tachezy, J. (2011). The minimal proteome in the reduced mitochondrion of the parasitic protist *Giardia intestinalis*. *PloS One* **6**, e17285. doi: 10.1371/journal.pone.0017285.
- Jefferies, J. R., Campbell, A. M., van Rossum, A. J., Barrett, J. and Brophy, P. M. (2001). Proteomic analysis of *Fasciola hepatica* excretory-secretory products. *Proteomics* **1**, 1128–1132. doi: 10.1002/1615-9861(200109)1:9<1128::AID-PROT1128>3.0.CO;2-0.
- Jones, P., Cote, R. G., Martens, L., Quinn, A. F., Taylor, C. F., Derache, W., Hermjakob, H. and Apweiler, R. (2006). PRIDE: a public repository of protein and peptide identifications for the proteomics community. *Nucleic Acids Research* **34**(Database issue), D659–663. doi: 10.1093/nar/gkj138.
- Jortzik, E., Wang, L. and Becker, K. (2011). Thiol-based posttranslational modifications in parasites. *Antioxidants and Redox Signaling*. doi: 10.1089/ars.2011.4266.
- Kanehisa, M. and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28**, 27–30. doi: 10.1093/nar/28.1.27.
- Kolker, E., Higdon, R., Haynes, W., Welch, D., Broomall, W., Lancet, D., Stanberry, L. and Kolker, N. (2012). MOPED: Model Organism Protein Expression Database. *Nucleic Acids Research* **40**(Database issue), D1093–1099. doi: 10.1093/nar/gkr1177.
- Kooij, T. W., Janse, C. J. and Waters, A. P. (2006). *Plasmodium* post-genomics: better the bug you know? *Nature Reviews Microbiology* **4**, 344–357. doi: 10.1038/nrmicro1392.

- Krishna, R., Wastling, J.M. and Jones, A.R. (2011). Automated integration of Mass Spectrometry based Proteomics evidence for improvement of gene annotations. In *8th BSPR-EBI Meeting—From the visible to the hidden proteome* Wellcome Trust Conference Centre, Cambridge, UK.
- Krogh, A., Brown, M., Mian, I. S., Sjolander, K. and Haussler, D. (1994). Hidden Markov models in computational biology. Applications to protein modeling. *Journal of Molecular Biology* **235**, 1501–1531. doi: 10.1006/jmbi.1994.1104.
- Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* **305**, 567–580. doi: 10.1006/jmbi.2000.4315.
- Kuss, C., Gan, C. S., Gunalan, K., Bozdech, Z., Sze, S. K. and Preiser, P. R. (2011). Quantitative proteomics reveals new insights into erythrocyte invasion by *Plasmodium falciparum*. *Molecular and Cellular Proteomics: MCP* **11**(2), M111.010645-. doi: 10.1074/mcp.M111.010645.
- LaCount, D. J., Vignali, M., Chettier, R., Phansalkar, A., Bell, R., Hesselberth, J. R., Schoenfeld, L. W., Ota, I., Sahasrabudhe, S., Kurschner, C., Fields, S. and Hughes, R. E. (2005). A protein interaction network of the malaria parasite *Plasmodium falciparum*. *Nature* **438**, 103–107. doi: 10.1038/nature04104.
- Lange, V., Picotti, P., Domon, B. and Aebersold, R. (2008). Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular Systems Biology* **4**, 222. doi: 10.1038/msb.2008.61.
- Larance, M., Bailly, A. P., Pourkarimi, E., Hay, R. T., Buchanan, G., Coulthurst, S., Xirodimas, D. P., Gartner, A. and Lamond, A. I. (2011). Stable-isotope labeling with amino acids in nematodes. *Nature methods* **8**, 849–851. doi: 10.1038/nmeth.1679.
- Lasonder, E., Ishihama, Y., Andersen, J. S., Vermunt, A. M., Pain, A., Sauerwein, R. W., Eling, W. M., Hall, N., Waters, A. P., Stunnenberg, H. G. and Mann, M. (2002). Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* **419**, 537–542. doi: 10.1038/nature01111.
- Lefevre, T., Thomas, F., Schwartz, A., Levashina, E., Blandin, S., Brizard, J.-P., Le Bourlignu, L., Demetree, E., Renaud, F. and Biron, D. G. (2007). Malaria *Plasmodium* agent induces alteration in the head proteome of their *Anopheles* mosquito host. *Proteomics* **7**, 1908–1915. doi: 10.1002/pmic.200601021.
- Lepeniev, B. and Seeberger, P. H. (2010). The promise of glycomics, glycan arrays and carbohydrate-based vaccines. *Immunopharmacology and Immunotoxicology* **32**, 196–207. doi: 10.3109/08923970903292663.
- Leykauf, K., Treeck, M., Gilson, P. R., Nebel, T., Brulke, T., Cowman, A. F., Gilberger, T. W. and Crabb, B. S. (2010). Protein kinase a dependent phosphorylation of apical membrane antigen 1 plays an important role in erythrocyte invasion by the malaria parasite. *PLoS Pathogens* **6**, e1000941. doi: 10.1371/journal.ppat.1000941.
- Lister, R., Gregory, B. D. and Ecker, J. R. (2009). Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Current Opinion in Plant Biology* **12**, 107–118. doi: 10.1016/j.pbi.2008.11.004.
- Liu, F., Cui, S.-J., Hu, W., Feng, Z., Wang, Z.-Q. and Han, Z.-G. (2009). Excretory/secretory proteome of the adult developmental stage of human blood fluke, *Schistosoma japonicum*. *Molecular and Cellular Proteomics MCP* **8**, 1236–1251. doi: 10.1074/mcp.M800538-MCP200.
- Liu, F., Hu, W., Cui, S.-J., Chi, M., Fang, C.-Y., Wang, Z.-Q., Yang, P.-Y. and Han, Z.-G. (2007). Insight into the host-parasite interplay by proteomic study of host proteins copurified with the human parasite, *Schistosoma japonicum*. *Proteomics* **7**, 450–462. doi: 10.1002/pmic.200600465.
- Liu, F., Lu, J., Hu, W., Wang, S.-Y., Cui, S.-J., Chi, M., Yan, Q., Wang, X.-R., Song, H.-D., Xu, X.-N., Wang, J.-J., Zhang, X.-L., Zhang, X., Wang, Z.-Q., Xue, C.-L., Brindley, P. J., McManus, D. P., Yang, P.-Y., Feng, Z., Chen, Z. and Han, Z.-G. (2006). New perspectives on host-parasite interplay by comparative transcriptomic and proteomic analyses of *Schistosoma japonicum*. *PLoS Pathogens* **2**, e29. doi: 10.1371/journal.ppat.0020029.
- Lonardi, E., Balog, C. I. a., Deelder, A. M. and Wührer, M. (2010). Natural glycan microarrays. *Expert Review of Proteomics* **7**, 761–774. doi: 10.1586/epr.10.41.
- Lu, P., Vogel, C., Wang, R., Yao, X. and Marcotte, E. M. (2007). Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nature Biotechnology* **25**, 117–124. doi: 10.1038/nbt1270.
- Mahner, M. and Kary, M. (1997). What exactly are genomes, genotypes and phenotypes? And what about phenomes? *Journal of Theoretical Biology* **186**, 55–63. doi: 10.1006/jtbi.1996.0335.
- Maier, T., Guell, M. and Serrano, L. (2009). Correlation of mRNA and protein in complex biological samples. *FEBS Letters* **583**, 3966–3973. doi: 10.1016/j.febslet.2009.10.036.
- Marugán-Hernández, V., Alvarez-García, G., Risco-Castillo, V., Regidor-Cerrillo, J. and Ortega-Mora, L. M. (2010). Identification of *Neospora caninum* proteins regulated during the differentiation process from tachyzoite to bradyzoite stage by DIGE. *Proteomics* **10**, 1740–1750. doi: 10.1002/pmic.200900664.
- Mi, H., Guo, N., Kejariwal, A. and Thomas, P. D. (2007). PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. *Nucleic Acids Research* **35**(Database issue), D247–252. doi: 10.1093/nar/gkl869.
- Moreno, Y. and Geary, T. G. (2008). Stage- and gender-specific proteomic analysis of *Brugia malayi* excretory-secretory products. *PLoS Neglected Tropical Diseases* **2**, e326. doi: 10.1371/journal.pntd.0000326.
- Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Paabo, S. and Mann, M. (2011). Deep proteome and transcriptome mapping of a human cancer cell line. *Molecular Systems Biology* **7**, 548. doi: 10.1038/msb.2011.81.
- Nebel, T., Prieto, J. H., Kapp, E., Smith, B. J., Williams, M. J., Yates, J. R., Cowman, A. F. and Tonkin, C. J. (2011). Quantitative *in vivo* Analyses Reveal Calcium-dependent Phosphorylation Sites and Identifies a Novel Component of the *Toxoplasma* Invasion Motor Complex. *PLoS Pathogens* **7**, e1002222. doi: 10.1371/journal.ppat.1002222.
- Nelson, M. M., Jones, A. R., Carmen, J. C., Sinai, A. P., Burchmore, R. and Wastling, J. M. (2008). Modulation of the host cell proteome by the intracellular apicomplexan parasite *Toxoplasma gondii*. *Infection and Immunity* **76**, 828–844. doi: 10.1128/IAI.01115-07.
- Noble, D. (2006). *The Music of Life: Biology Beyond the Genome*, Oxford University Press, Oxford.
- Nunes, M. C., Okada, M., Scheidig-Benatar, C., Cooke, B. M. and Scherf, A. (2010). *Plasmodium falciparum* FIKK kinase members target distinct components of the erythrocyte membrane. *PLoS One* **5**, e11747. doi: 10.1371/journal.pone.0011747.
- O'Donnell, R. A., Hackett, F., Howell, S. A., Treeck, M., Struck, N., Krnjanski, Z., Withers-Martinez, C., Gilberger, T. W. and Blackman, M. J. (2006). Intramembrane proteolysis mediates shedding of a key adhesin during erythrocyte invasion by the malaria parasite. *Journal of Cell Biology* **174**, 1023–1033. doi: 10.1083/jcb.200604136.
- Oltvai, Z. N. and Barabasi, A. L. (2002). Systems biology. Life's complexity pyramid. *Science* **298**, 763–764. doi: 10.1126/science.1078563.
- Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular and Cellular Proteomics* **1**, 376–386. doi: 10.1074/mcp.M200025-MCP200.
- Palagi, P. M., Walther, D., Quadroni, M., Catherinet, S., Burgess, J., Zimmermann-Ivol, C. G., Sanchez, J. C., Binz, P. A., Hochstrasser, D. F. and Appel, R. D. (2005). MSight: an image analysis software for liquid chromatography-mass spectrometry. *Proteomics* **5**(9), 2381–2384. doi: 10.1002/pmic.200401244.
- Perkins, D. N., Pappin, D. J., Creasy, D. M. and Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**(18), 3551–3567. doi: 10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.
- Prieto, J. H., Koncarevic, S., Park, S. K., Yates, J. and Becker, K. (2008). Large-scale differential proteome analysis in *Plasmodium falciparum* under drug treatment. *PLoS One* **3**, e4098. doi: 10.1371/journal.pone.0004098.
- Rebello, K. M., Barros, J. S. L., Mota, E. M., Carvalho, P. C., Perales, J., Lenzi, H. L. and Neves-Ferreira, A. G. C. (2011). Comprehensive proteomic profiling of adult *Angiostrongylus costaricensis*, a human parasitic nematode. *Journal of Proteomics* **74**, 1545–1559. doi: 10.1016/j.jprot.2011.04.031.
- Rigboltz, K. T., Vanselow, J. T. and Blagoev, B. (2011). GProX, a user-friendly platform for bioinformatics analysis and visualization of quantitative proteomics data. *Molecular and Cellular Proteomics* **10**, O110 007450. doi: 10.1074/mcp.O110.007450.
- Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Parkkayastha, S., Juhasz, P., Martin, S., Bartlett-Jones, M., He, F., Jacobson, A. and Pappin, D. J. (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Molecular and Cellular Proteomics* **3**, 1154–1169. doi: 10.1074/mcp.M400129-MCP200.

- Ruhaak, L. R., Zauner, G., Huhn, C., Bruggink, C., Deelder, a. M. and Wuhrer, M. (2010). Glycan labeling strategies and their use in identification and quantification. *Analytical and Bioanalytical Chemistry* **397**, 3457–3481. doi: 10.1007/s00216-010-3532-z.
- Rukmangadachar, L. A., Kataria, J., Hariprasad, G., Samantaray, J. C. and Srinivasan, A. (2011). Two-dimensional difference gel electrophoresis (DIGE) analysis of sera from visceral leishmaniasis patients. *Clinical Proteomics* **8**, 4. doi: 10.1186/1559-0275-8-4.
- Sanderson, S. J., Xia, D., Prieto, H., Yates, J., Heiges, M., Kissinger, J. C., Bromley, E., Lal, K., Sinden, R. E., Tomley, F. and Wastling, J. M. (2008). Determining the protein repertoire of *Cryptosporidium parvum* sporozoites. *Proteomics* **8**, 1398–1414. doi: 10.1002/pmic.200700804.
- Sauer, U., Heinemann, M. and Zamboni, N. (2007). Genetics. Getting closer to the whole picture. *Science* **316**, 550–551. doi: 10.1126/science.1142502.
- Schrimpf, S. P., Weiss, M., Reiter, L., Ahrens, C. H., Jovanovic, M., Malmström, J., Brunner, E., Mohanty, S., Lercher, M. J., Hunziker, P. E., Aebersold, R., von Mering, C. and Hengartner, M. O. (2009). Comparative functional analysis of the *Caenorhabditis elegans* and *Drosophila melanogaster* proteomes. *PLoS Biology* **7**, e48. doi: 10.1371/journal.pbio.1000048.
- Schwanhauser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature* **473**, 337–342. doi: 10.1038/nature10098.
- Shock, J. L., Fischer, K. F. and DeRisi, J. L. (2007). Whole-genome analysis of mRNA decay in *Plasmodium falciparum* reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle. *Genome Biology* **8**, R134. doi: 10.1186/gb-2007-8-7-r134.
- Sigrist, C. J., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., Bairoch, A. and Bucher, P. (2002). PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief Bioinformatics* **3**, 265–274. doi: 10.1093/bib/3.3.265.
- Silvestrini, F., Lasonder, E., Olivieri, A., Camarda, G., van Schaijk, B., Sanchez, M., Younis, S., Sauerwein, R. and Alano, P. (2010). Protein export marks the early phase of gametocytogenesis of the human malaria parasite *Plasmodium falciparum*. *Molecular and Cellular Proteomics: MCP* **9**, 1437–1448. doi: 10.1074/mcp.M900479-MCP200.
- Southworth, P. M., Hyde, J. E. and Sims, P. F. G. (2011). A mass spectrometric strategy for absolute quantification of *Plasmodium falciparum* proteins of low abundance. *Malaria Journal* **10**, 315. doi: 10.1186/1475-2875-10-315.
- Stein, L. D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J. E., Harris, T. W., Arva, A. and Lewis, S. (2002). The generic genome browser: a building block for a model organism system database. *Genome Research* **12**, 1599–1610. doi: 10.1101/gr.403602.
- Sturm, M., Bertsch, A., Gropl, C., Hildebrandt, A., Husson, R., Lange, E., Pfeifer, N., Schulz-Trieglaff, O., Zerck, A., Reinert, K. and Kohlbacher, O. (2008). OpenMS – an open-source software framework for mass spectrometry. *BMC Bioinformatics* **9**, 163. doi: 10.1186/1471-2105-9-163.
- Unlu, M., Morgan, M. E. and Minden, J. S. (1997). Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18**, 2071–2077. doi: 10.1002/elps.1150181133.
- van Balkom, B. W., van Gestel, R. A., Brouwers, J. F., Krijgsveld, J., Tielens, A. G., Heck, A. J. and van Hellemond, J. J. (2005). Mass spectrometric analysis of the *Schistosoma mansoni* tegumental sub-proteome. *Journal of Proteome Research* **4**, 958–966. doi: 10.1021/pr050036v.
- Verma, R., Tiwari, A., Kaur, S., Varshney, G. C. and Raghava, G. P. (2008). Identification of proteins secreted by malaria parasite into erythrocyte using SVM and PSSM profiles. *BMC Bioinformatics* **9**, 201. doi: 10.1186/1471-2105-9-201.
- von Heijne, G. (1990). The signal peptide. *Journal of Membrane Biology* **115**, 195–201.
- Wang, Z., Gerstein, M. and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57–63. doi: 10.1038/nrg2484.
- Wastling, J. M., Xia, D., Sohal, A., Chaussepied, M., Pain, A. and Langsley, G. (2009). Proteomes and transcriptomes of the *Apicomplexa* – where's the message? *International Journal for Parasitology* **39**, 135–143. doi: 10.1016/j.ijpara.2008.10.003.
- Wilson, D., Madera, M., Vogel, C., Chothia, C. and Gough, J. (2007). The SUPERFAMILY database in 2007: families and functions. *Nucleic Acids Research* **35**(Database issue), D308–313. doi: 10.1093/nar/gkl910.
- Wu, Y., Nelson, M. M., Quail, A., Xia, D., Wastling, J. M. and Craig, A. (2009). Identification of phosphorylated proteins in erythrocytes infected by the human malaria parasite *Plasmodium falciparum*. *Malaria Journal* **8**, 105. doi: 10.1186/1475-2875-8-105.
- Xia, D., Sanderson, S. J., Jones, A. R., Prieto, J. H., Yates, J. R., Bromley, E., Tomley, F. M., Lal, K., Sinden, R. E., Brunk, B. P., Roos, D. S. and Wastling, J. M. (2008). The proteome of *Toxoplasma gondii*: integration with the genome provides novel insights into gene expression and annotation. *Genome Biology* **9**, R116. doi: 10.1186/gb-2008-9-7-r116.
- Yates, J. R., 3rd, Eng, J. K., McCormack, A. L. and Schieltz, D. (1995). Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Analytical Chemistry* **67**, 1426–1436. doi: 10.1021/ac00104a020.
- Yatsuda, A. P., Krijgsveld, J., Cornelissen, A. W., Heck, A. J. and de Vries, E. (2003). Comprehensive analysis of the secreted proteins of the parasite *Haemonchus contortus* reveals extensive sequence variation and differential immune recognition. *Journal of Biological Chemistry* **278**, 16941–16951. doi: 10.1074/jbc.M212453200.
- Yeats, C., Maibaum, M., Marsden, R., Dibley, M., Lee, D., Addou, S. and Orengo, C. A. (2006). Gene3D: modelling protein structure, function and evolution. *Nucleic Acids Research* **34**(Database issue), D281–284. doi: 10.1093/nar/gkj057.
- Yen, H. C., Xu, Q., Chou, D. M., Zhao, Z. and Elledge, S. J. (2008). Global protein stability profiling in mammalian cells. *Science* **322**, 918–923. doi: 10.1126/science.1160489.
- Yook, K., Harris, T. W., Bieri, T., Cabunoc, A., Chan, J., Chen, W. J., Davis, P., de la Cruz, N., Duong, A., Fang, R., Ganesan, U., Grove, C., Howe, K., Kadam, S., Kishore, R., Lee, R., Li, Y., Muller, H. M., Nakamura, C., Nash, B., Ozersky, P., Paulini, M., Raciti, D., Rangarajan, A., Schindelman, G., Shi, X., Schwarz, E. M., Ann Tuli, M., Van Auken, K., Wang, D., Wang, X., Williams, G., Hodgkin, J., Berriman, M., Durbin, R., Kersey, P., Spieth, J., Stein, L. and Sternberg, P. W. (2011). WormBase 2012: more genomes, more data, new website. *Nucleic Acids Research* **40**(D1), D735–D741. doi: 10.1093/nar/gkr954.
- Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S. C., Ester, M., Foster, L. J. and Brinkman, F. S. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**, 1608–1615. doi: 10.1093/bioinformatics/btq249.