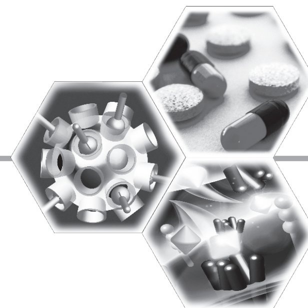

For reprint orders, please contact reprints@future-science.com

Application of genomics, proteomics and metabolomics in drug discovery, development and clinic

Genomics, proteomics and metabolomics are three areas that are routinely applied throughout the drug-development process as well as after a product enters the market. This review discusses all three 'omics, reporting on the key applications, techniques, recent advances and expectations of each. Genomics, mainly through the use of novel and next-generation sequencing techniques, has advanced areas of drug discovery and development through the comparative assessment of normal and diseased-state tissues, transcription and/or expression profiling, side-effect profiling, pharmacogenomics and the identification of biomarkers. Proteomics, through techniques including isotope coded affinity tags, stable isotopic labeling by amino acids in cell culture, isobaric tags for relative and absolute quantification, multidirectional protein identification technology, activity-based probes, protein/peptide arrays, phage displays and two-hybrid systems is utilized in multiple areas through the drug development pipeline including target and lead identification, compound optimization, throughout the clinical trials process and after market analysis. Metabolomics, although the most recent and least developed of the three 'omics considered in this review, provides a significant contribution to drug development through systems biology approaches. Already implemented to some degree in the drug-discovery industry and used in applications spanning target identification through to toxicological analysis, metabolic network understanding is essential in generating future discoveries.

Craig Russell, Ayesha Rahman & Afzal R Mohammed*

Aston School of Pharmacy, Aston University, Birmingham, UK
*Author for correspondence:
Tel.: +44 121 204 4183
E-mail: a.u.r.mohammed@aston.ac.uk

Genes are at the hub of so many biological events – focusing studies on human genetics allows for advances in drug discovery through a multitude of avenues, including the understanding of disease onset and progression, and providing a means for the identification of drug targets and even therapeutic agents by observing genes and their transcription products. The application of genomics to areas including, **biomarkers**, tissue expression profiling, side-effect profiling (SEP), **pharmacogenomics** (PGx) and genome-wide epigenetics are some of the main areas through which genomic techniques are contributing to the drug discovery and development processes. Similarly, proteomics applied to the field of drug discovery and development expand our knowledge relating to protein function, interaction and regulation. Proteomics finds application in most areas through the drug-development pipeline, including target and lead identification, compound optimization, throughout the clinical trials process and after market analysis [1]. Although genomic techniques are well established, proteomics holds some major advantages over genomics.

Where metabolomics are concerned, due to the almost overwhelming complexity of cells, research has tended to focus on isolated metabolic pathways and products and has increased the level of understanding on how cells function, revealing many significant mechanisms of disease. Unfortunately this 'one at a time' approach has drawbacks, for example, one of the main causes of drugs failing during clinical trials is the action of a drug molecule at a site other than the intended drug target, a potentially serious problem not normally picked up until later in the developmental process [2]. Possessing a holistic 'network' view of all of the cellular pathways and understanding how these pathways interact with each other in a **systems biology** approach through metabolomics allows for advances in drug discovery and development. The identification of a drug target and then the validation procedure that follows are initial stages in drug discovery, and the information provided by the metabolic network shows how the target in question behaves in a normal environment. By looking at the metabolic pathway network, the downstream action of the target can be determined. This allows the consequence of intervention to be predicted, whether this be beneficial or detrimental in terms of disease treatment. In addition, if the target proves to be involved in

Key Terms

Biomarkers: Quantifiable indicators of biological state that are measured to indicate normal biological processes, pathogenic processes or the pharmacologic response to therapeutic intervention.

Pharmacogenomics: Study of genetic effect on the performance of a drug to optimize drug therapy.

Systems biology: Relatively recent branch of science in which the whole picture is assessed, in that, instead of focusing on one event at a time, the whole complex network of events that come together to elicit the function or behavior of the system are considered.

multiple pathways then it may be advantageous to discard it and search for a target further down the specific pathway for which the intervention is desired, so as to focus the intervention effect. In the same way, side effects of drugs and also toxicology, both of which are a major cause of late-stage drug failure, can at times be predicted and avoided prior to the investment of significant time and money [3]. Already implemented to some degree in the drug-discovery industry and used in applications spanning target identification through to toxicological analysis, metabolic network understanding is essential in generating future discoveries.

In this review we provide an overview detailing some of the most common critical and rapidly developing areas of genomics, proteomics and metabolomics and discuss their application throughout the areas of drug-discovery, development and clinical practice.

Genomics

■ Genomic sequencing techniques

There are numerous techniques available for genetic sequencing that have come to fruition since the early methods of Maxam–Gilbert sequencing and chain-termination methods developed in the 1970s. The basic methods have been improved upon so as that large-scale sequencing and *de novo* sequencing became a reality using techniques such as shotgun sequencing, which allows for long DNA sequences to be derived. The most recent techniques finding application in a pharmaceutical setting fall into the class of next-generation sequencing and there are new techniques being developed continuously.

Next-generation sequencing

Massively parallel signature sequencing (MPSS) was the first next-generation sequencing technology developed during the 1990s and involved a complex bead-based process consisting of adapter ligation and adapter decoding followed by fluorescence-based signature sequencing [4]. MPSS had the disadvantage that the method was susceptible to the loss of some specific sequences and also sequence specific bias. Following the lead of MPSS, development of next-generation sequencing methods increased and now next-generation sequencing techniques come in a variety of forms with Illumina (Solexa®) sequencing, 454 – Roche, supported oligonucleotide ligation and detection (SOLiD) sequencing, ion semiconductor sequencing and single-molecule real-time (SMRT) sequencing being some of the most common.

Illumina (Solexa) sequencing

Illumina (Solexa) sequencing is a next-generation sequencing approach developed by Illumina (CA, USA), which can provide very high throughput at a relatively low cost. It involves the use of single-molecule arrays where DNA amplification is achieved using bridge amplification to produce a template for synthesis. This template is then used to generate complementary strands to eventually provide a flow cell containing somewhere in the region of 40 million DNA colonies, each made up of around 1000 clonal copies of the template. Sequencing in an MPSS fashion using fluorescent markers then follows using a different colored marker for each of the four nucleotide bases. Although it is a high-throughput, low-cost method there are limitations with the Illumina (Solexa) sequencing method, for example, it is unable to resolve short sequence repeats due to short sequence reads. Furthermore, sequencing mistakes stemming from substitution errors have been reported in published data using Illumina (Solexa) sequencing [5].

454 – Roche sequencing

454 – Roche sequencing can be used for whole genome sequencing, targeted resequencing, metagenomics and transcriptome sequencing making it one of the most efficient platforms on the market. The platform was designed to overcome one of the main limitations of the chain-termination methods developed in the 1970s. That limitation is the *in vivo* DNA amplification usually carried out using bacteria as hosts. This is overcome by implementing emulsion PCR [6]. DNA fragments are immobilized onto beads where amplification is then carried out in an emulsion. Amplification in this way is a reliable method that is well used and is very efficient, providing around 10^7 clonal copies per bead. The clonal copies are then sequenced using a technique known as pyrosequencing, which is a sequencing by synthesis technique running many samples in parallel and monitoring chemiluminescense in response to the release of inorganic pyrophosphate. This method of sequencing was the first platform commercially available on the market and to date is the most widely used next-generation sequencing technique [7].

SOLiD sequencing

SOLiD sequencing is MPSS by hybridization-ligation. It is able to generate up to 3 Gb of sequence in 35 bp reads from an 8 day run [7].

This method begins with emulsion PCR as used in the 454 – Roche sequencing approach. The amplified clonal copies are then immobilized onto a slide before hybridization and ligation cycles are used to perform the sequencing. Fluorescent labeling is again utilized as a means of differentiating the nucleotides from each other, however, SOLiD sequencing has an advantage over other techniques in that it is capable of identifying the difference between changes in the sequence due to sequencing error and changes in the sequence due to polymorphism. This is achieved by probing each sequence twice where sequencing error would only be observed on one occasion, whereas polymorphism would be observed on both [8].

Ion semiconductor sequencing

Ion Torrent™ from Life Technologies (Paisley, UK) is the ion semiconductor sequencing platform available on the market. It is unlike most other sequencing techniques in that it does not use optical output to sequence the genome, instead, it is able to sequence genomes directly by measuring ions that are produced by template-directed DNA polymerase synthesis. The two instruments available are the personal genome machine and the ion proton system. The clonal sequences are produced either on a massively parallel semiconductor-sensing device or an ion chip. The ion chip contains ion-sensitive sensors capable of testing an array of over a million wells allowing parallel, simultaneous detection of independent sequencing reactions. This method allows for low-cost, large-scale analysis [9].

SMRT sequencing

SMRT sequencing is based on the sequencing by synthesis approach and does not require the amplification of DNA, as is the case in other sequencing techniques [10]. This means problems that may stem from the amplification stage associated with introducing sequencing errors or altering abundance of DNA fragments are avoided. In addition, there is no need for the immobilization of cloned fragments onto a solid support, instead, polymerase molecules are fixed to a solid support. Samples are then passed over the immobilized polymerase and the DNA fragments interact when they encounter the polymerase active site. The nucleotide sequence is then imaged as this happens. In this way real-time sequencing is possible, which shortens the time for sequencing and also allows larger fragments to be sequenced [11].

■ Techniques currently in development

Nano-pore sequencing

Nano-pore sequencing is a method for the genetic sequencing of DNA without the need for an enzymatic replication step. The process for nano-pore sequencing is very simple with the method involving an electric current applied across a membrane containing nano-pores that has been submerged in salt solution. There are two variations of nano-pore methodologies, one which utilizes the application of biological membranes and a second that utilizes synthetic membranes [12].

Currently, it is biological membranes that are preferred as the pore size shows little variability across membrane sections and is also well defined. Examples of biological membranes used include an α-hemolysin pore [13] and a Mycobacterium smegmatis porin A pore [14]. When an electric current is applied to the system ions pass across the membrane through the nano-pore. DNA molecules are charged, and when they arrive at and pass through the pores in the membrane they cause a change in the passage of ions, this correlates to a measurable change in the electrical current. The detected change differs in magnitude depending upon the size of the molecule passing across the membrane and also the duration of the passage. The difficulty in nano-pore methods is providing a means by which the four DNA nucleotide bases can be told apart from one another, an essential feature if the DNA sequence is to be determined. This has recently been achieved through the genetic modification of the α-hemolysin pore to include a molecule that retards the passage of the DNA through the pores in order to improve detection [15].

The application of synthetic membranes to achieve DNA sequencing on a commercial level is still some way off due to complications involved in the manufacture of a membrane that provides a pore size with sufficiently low interpore variation, and currently grapheme is widely regarded as the most promising material [16]. In contrast, Oxford Nanopore Technologies (Oxford, UK) and Genia (CA, USA) are expected to deliver commercialized nano-pore sequencing platforms within the next 12 months [12].

Hybridization sequencing

DNA sequencing is a widely used genomic technique and there are a number of techniques that can be employed to carry out analysis. One of the most important additions to the world of genome profiling was the DNA array, which first

Key Terms

Microarray: Example of a lab-on-a-chip, they consist of probes immobilized onto a solid substrate, variations of microarray include examples such as DNA microarrays, protein/peptide microarrays, tissue microarrays and antibody microarrays to name a few.

Human Genome: Profile of the 25,000 or so genes that code for the human body. The Human Genome Project has so far accurately sequenced 99% of the human genome.

arrived on the scene in 1982 [17] before being miniaturized and producing **microarrays** in 1995 [18], allowing for the first mapping of a complete eukaryotic genome in 1997 [19]. DNA microarrays consist of a flat solid support onto which probes are immobilized in an organized fashion, with each probe able to unambiguously match and identify their corresponding molecule [20]. Both ssDNA and dsDNA oligomers can be used in the production of microarrays. There are three main types, consisting of short length (25 bp) oligonucleotide probes [21], longer and highly variable cDNA probes [22] and long (50–80 bp) oligonucleotide arrays [23,24]. The long oligonucleotide arrays are the most recent development and may leave the cDNA versions redundant [24]. Microarray techniques provide a powerful high-throughput tool in the assessment of mRNA abundance, producing expansive data readouts detailing the expression levels of thousands of genes in every sample. The detailed analysis of these data sheets provides insight into the biological processes being studied and also contributes to the production of genomic databases and genetic maps such as the **human genome** project [18]. In the context of drug discovery, microarray analysis has advanced areas of drug discovery and development through the comparative assessment of normal and diseased-state tissues, transcription and/or expression profiling, SEP, PGx and the identification of biomarkers.

Sequencing with mass spectrometry

Sequencing with matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) has been used as an alternative to gel electrophoresis in order to measure DNA fragments. The fragments are measured by mass, and as the mass of each nucleotide differs it is possible to detect this difference. In addition MALDI-TOF MS makes it easier when compared with gel electrophoresis to detect mutations that result from only one changed nucleotide in the sequence. The limitation of using MALDI-TOF MS and other MS techniques is that the largest manageable read length is around 100 bp. This means that this is not a particularly useful method in large-scale sequencing techniques. One of the best examples of MALDI-TOF MS in genetic analysis is by Sequenom Inc. (CA, USA), which has integrated the application of MALDI-TOF MS into a top-down approach for disease gene identification by genome-wide association studies [25].

Genomic applications

■ Biomarkers

Biomarkers are measurable entities that indicate a diseased state when compared with normal state and provide feedback on the behavior of a drug following drug administration. They have a range of uses including pharmokinetic, pharmodynamic and toxicological assessment. The focus of genomics in biomarker identification is to increase the number of biomarkers available and through this, increase the amount of data available earlier in the drug-development pipeline to base decisions upon regarding the progression of candidate compounds. Biomarker identification is often carried out in parallel with drug development and their use is applicable for many stages in the drug-development process, including follow up studies once the medication hits the market [26].

Biomarker discovery has been applied in the identification of biomarkers for osteoarthritis. A report on the second Osteoarthritis Biomarkers Global Initiative, discussed how the experiments in genetics, epigenetics and genomics are yielding valuable information into osteoarthritis, and combining the findings with conventional biomarkers and other techniques will provide scientists with increased understanding of the disease [27]. In order to ensure high quality standardized data generated from microarray investigations into biomarkers using genomic profiling, the US FDA along with various users of the microarray platforms established the microarray quality control consortium that settled upon methods and standards to maintain wide-reaching quality control. Through the application of this standardization and the continued investigation into biomarkers for disease it will one day be possible to apply a reliable, predictive tool to the development of new medicines selecting for improved efficacy and reduced toxicity at an early stage in the developmental programme, saving time and money in research and development, and also increasing the output of the pharmaceutical industry [28].

■ Tissue-expression profiling

During drug development, expression profiling of various tissues plays an instrumental role in many aspects of drug discovery, including target identification, target validation, identification of candidate compounds, PGx, biomarker development, the evaluation of clinical trials and the gathering of toxicological information [28].

Over the last 10 years, the pharmaceutical industry has been keen to develop large-scale, species-specific transcriptional profiling databases for a range of species. The profiling is normally performed through one of two routes, whole tissue profiling or regional tissue profiling using techniques including purified cells and physical or laser micro-dissection. The investigations are standardized by ensuring that the type of array selected and the control standards are used universally. In this way, the data generated will be comparable over the entire investigative span, and in addition, the standardization of the data means that multiple databases can be compared, and even cross-species comparisons can be performed with confidence, which is important when considering species for nonhuman drug testing.

When a pharmaceutical company wishes to establish profiling databases there are various ways in which to go about it. The first is to gather existing profiling data from previous genetic profiling investigations and from this, create the databases in their entirety. The second is to combine the existing data with data they generate through their own investigations. In the second instance it is important that the investigative procedures generate data sets that are comparable to the existing database. The third option is where the pharmaceutical company wishes to create a new transcriptional profiling database through the application of a previously unused array platform. In this case the initial step is extensive genomic profiling of the species of interest. This leads to the rapid expansion of the new database. This approach can be applied to various stages of the drug-development process.

During target identification it is important that the targets are isolated to one specific tissue. For this it is necessary to reveal transcripts that are observed only in one tissue type. A recent example of this used the Ilumina HumanHT-12_v3 Expression BeadChips® in expression profiling to examine the genetic control of gene expression in blood and brain [29]. It was found that although there were many similarities between the blood and brain transcript levels there were examples where differences were found in transcript expression either due to restricted gene-expression patterns in one tissue or due to differences in how genetic variants are associated with transcript levels. Alternatively in toxicology, toxicogenomics provides a means to predict the toxicity at an earlier stage than the previous methods of histopathology or clinical chemistry by looking at the response of the genome to substances toxic to the body [30]. Through gene silencing investigations the genomic characterization of delivery systems and identification of drug incompatibility issues is possible. This allows for a reduction in the investment of both time and money by means of earlier drug screening. Once the database is constructed containing the transcription profile data from the whole tissue samples, the next stage is to analyze individual regions of each tissue through regional tissue profiling approaches. This involves the application of microarray analysis and expands the transcriptome of the tissue in question dramatically.

The effect of expanding the transcriptome improves the scope for target identification by improving the potential to identify drug targets with localized expression. A recent example of this can be seen in the Affymetrix human genome U133 plus 2.0 microarray representing approximately the whole human genome [31]. Gene expression was examined in the epididymis and of the 15,000 transcripts they detected, 65% were present throughout the three tissue regions examined, 2.6% were present in only one region and the remaining 32.4% was found in two of the three regions. The transcripts present in a single region only could be used as a drug target, safe in the knowledge that a drug targeting one of these transcripts will have no site of action in the second and third regions [31].

Target identification is not the only area to benefit from the micro-dissection of tissue samples. The delineation of regulated and non-regulated genes reveals the genes (regulated) that are most likely to either be involved in the process being examined or alternatively become a biomarker if disease is to effect its expression. Furthermore, the fact that the gene location in the genome can be pinpointed often means that the protein for which that gene codes is often found in the same area, and thereby narrows the search if the aim is to isolate that particular protein. The application of tissue sample micro-dissection can expand the understanding of biological pathways and add to the disciplines of proteomics and metabolomics, especially when combined with *in silico* modeling techniques. Finally, having the most detailed databases possible with information on tissues from multiple species allows for the cross-species comparison of genes or groups of genes. If it transpires that a group of genes is present in multiple species

then it is normally the case that those genes are of high importance with regards to the biological processes involved in an organism and also that the genes are likely to able feature in the human genome.

■ SEP

SEP, also known as **toxicogenomics**, is the study of a chemicals toxicological behavior looking at effects on a transcriptional level. Although applicable throughout the drug-development process, SEP is at its most useful and is most widely used once a drug reaches the clinical trials phase. Where SEP really excels when compared with traditional toxicology is in its ability to provide data earlier in the drug development pipeline, and also the fact that SEP can predict patient segments that are likely to respond best to the drug and also identify any patient segments for which side effects may be more severe. SEP reports have already been included in the supporting data provided to regulatory authorities. There are currently a wide variety of tools available to carry out SEP including DNA microarrays, differential display analysis, subtractive hybridization, serial analysis of gene expression, MPSS and various proteomic techniques.

A recent example of such an application was published in 2012, in which biomarkers for an early time point gene expression pattern for a condition that leads to phospholipidosis as a side effect of many cationic amphiphilic drugs were identified [32]. To achieve this, Affymetrix GeneChip RAT Genome 230 2.0 arrays and identified 25 probe sets that were the biomarkers identified through genetic algorithm optimization. Following validation of their findings they were able to confirm that the biomarkers indicated statistically significant changes in the lipid metabolism which is associated with phospholipidosis. In this way they successfully developed a means of screening for phospholipidosis early in the drug development process which is a marked improvement over the previous method of histopathology which involves late time point analysis and repeated doses of the drug being tested. The early testing is a significance advantage to the pharmaceutical industry through the time and money it can save the companies concerned [32].

SEP is poised to play a vital role in the selection of lead compounds from the thousands of candidates which are available for selection and progression further down the line in drug development. Through the observation of normal state gene expression, it is possible to identify alterations in gene expression that are responsible for side effects stemming from the drug action. Through the logging of SEP data and the formation of toxicogenomic databases, it is possible to understand the mechanism by which toxicity results and allow for toxicological prediction of the behavior of new compounds entering the pipeline, through comparison with previously studied examples. It is important, however, that the databases generated are of sufficient depth and breadth, ideally containing an extensive variety of known reference drug toxicogenomic profiles to ensure their reliability. There are currently several databases such as Gene Logic [101] and Iconix Pharmaceuticals [102] that contain toxicogenomic profiles and they have been created using the guidelines of the standard of minimum information about microarray experiment (MIAME) and MIAME/Tox, which extends the MIAME guidelines to cover toxicogenomics [33]. The DrugMatrix database from Iconix was recently used in research into the application of black cohosh as a treatment for breast cancer [34]. DrugMatrix details organ specific gene expression data for over 650 reference compounds. Their findings indicated that black cohosh extract impacts the mitochondrial oxidative phosphorylation pathway and that the downregulation of mitochondrial genes may cause mitochondrial damage. The benefit of using expression data is that small changes in the liver can be used to predict toxicological and pathological effects in the liver and other tissues before these effects can actually be observed through the more traditional toxicological techniques [34].

■ PGx

It is common knowledge in the pharmaceutical industry that the number of candidate drug compounds identified in drug discovery is huge when compared with the number of compounds that actually reach the market. Compounds can fail at any point though the drug-development process, however the failures that are of the highest consequence are the failures that occur once a drug has reached or cleared clinical trials. Almost 50% of the drugs that reach the final phase of clinical trials progress no further due to issues associated with their efficacy or their toxicity. Despite the fact that some drugs pass the clinical trials, there are occasions when upon reaching the market and becoming available to a very large patient population efficacy and toxicity issues not detected previously may present

Key Terms

Toxicogenomics: Used in drug discovery and development to study adverse reactions by examining how the genome responds *in vitro* and/or *in vivo* following exposure to a pharmaceutical product.

Proteome: Comprehensive readout of protein concentrations from a given sample at any given time will change in response to physiological and environmental conditions in addition to drug intervention.

themselves. Failure of a drug in the final stages of clinical trials, or worse, once it has reached the market, carries a huge price tag for the company involved in its development. One of the main reasons behind the unpredicted failures is the heterogeneous way in which individuals respond to medications due to differences in metabolism, drug disposition and target protein polymorphism [35].

Pharmacogenetics investigates the correlation between an individual's genetic makeup and their response to medicines through genotype and phenotype differences in the pharmacodynamics and pharmacokinetics of drug metabolism. PGx, on the other hand, is a genome-scale application of these techniques used to elucidate correlations between gene expression or single nucleotide polymorphism and drug efficacy and toxicity [35]. Through the application of DNA microarrays and bioinformatic techniques the aim is to understand how particular gene sets bring about particular responses in drug effect.

Recently, two groups Xu and Wang [36] and Li and Lu [37] have both published work detailing methods to correlate gene expression and drug action via data mining of existing literature. Xu and Wang developed a conditional relationship extraction approach to extract PGx-specific drug–gene pairs from 20 million MEDLINE abstracts using known drug–gene pairs as prior knowledge. Li and Lu developed a systematic approach to automatically identify PGx relationships between genes, drugs and diseases from trial records in the clinical trials database [103]. In this way it is anticipated that medicines designed specifically for certain patient groups may be possible, providing less side effects and increased efficacy. The application will also be applicable to preventing adverse drug reactions that result from unexpected drug–drug interactions. Drug–drug interactions that lead to adverse drug reactions are a major problem in the pharmaceutical industry, especially in cases where patients are regularly receiving multiple medications – most often the elderly [38].

■ Genome-wide epigenetics

Sequencing-based approaches now allow high-resolution, genome-scale investigation of cellular epigenetic landscapes. For example, mapping of open chromatin regions, post-translational histone modifications and DNA methylation across a whole genome is now feasible, and new non-coding regulatory RNAs can be sensitively identified via RNA sequencing. The resulting large-scale data sets promise to contribute towards a more precise and complete understanding of gene regulation and to yield insights into the interplay between genomes and the environment [39]. Epigenetics has benefited greatly from increases in sequencing power. Next-generation sequencing technologies discussed earlier, as well as emerging platforms such as Pacific Biosciences (CA, USA), enable full genome mapping of many kinds of epigenetic characteristics. Genome-scale epigenetic research is still in its infancy and many issues and challenges in the bioinformatic aspects of sequence-based epigenetics still remain [40].

Proteomics

In the field of drug discovery and development the use of knowledge relating to protein function, interaction and regulation is known as proteomics and finds application in most areas including target and lead identification, compound optimization, throughout the clinical trials process and after market analysis [1]. Although genomic techniques are well established, proteomics hold some major advantages over genomics. First, gene expression is not necessarily an accurate reflection of protein regulation [41]. Second, genomics does not take into account cellular events that occur after gene transcription, such as histone modification with acetyl-, methyl- and phosphoryl-groups at distinct amino acid residues [42]. As such, current research has begun to focus more in the direction of proteomics (and metabolomics, discussed later). Proteomic techniques fall mainly into the categories of separation and identification and are applied to produce a comprehensive readout of protein concentrations from a given sample at any given time, this readout is known as a **proteome** and the proteome will change in response to physiological and environmental conditions in addition to drug intervention [43].

Proteomic techniques

Proteins are inherently complicated to gain information from. The main limitations are that the protein level in cells fluctuates in response to a multitude of factors and protein concentration may not correlate with protein activity. Protein interactions with other proteins and other molecules in the cell also complicate the analysis. In addition, protein concentrations tend to be very low, leading to problems in the detection and accurate quantification of measurements.

As such it is essential that techniques employed must be suitably sensitive, activity indicating and also rapid to cater for the high number of proteins to be analyzed, and also the fact that cellular protein changes alter over time [1].

In the past, the main method of proteomic analysis involved firstly the separation of proteins through a 2D gel electrophoresis followed by MS [1]. There were, however, limitations in the method sensitivity, limit of detection and separating power. Scientific advances in the analytical technology available including the advent of lab-on-a-chip systems and advances in MS techniques, coupled with huge advances in bioinformatic techniques and computing power have now provided new methods for proteomic assay including; multidirectional protein identification technology (MudPIT) [44], protein chips [45], yeast two-hybrid systems, isotope coded affinity tags (ICAT) [46], stable isotopic labeling by amino acids in cell culture (SILAC) [47], isobaric tags for relative and absolute quantification (iTRAQ) [48] and activity-based probes (ABPs) [49]. It must be noted, however, that protein arrays have not been found to be as widely applicable as their DNA counterparts owing to protein heterogeneity.

The information generated from proteomics provides at a minimum the identity and at times the abundance of the proteins being investigated. These minimum level findings are often generated by what is referred to as the classical approach to proteomics. Beyond the classical techniques, the abundance, the function and indications of interactions are resolved by more complex analysis, it is the functional proteomic approach that provides this more complete picture and in a drug-development context it is essential that information from both classical and functional proteomics be available [1].

■ Classical proteomic analysis

MudPIT

MudPIT is an example of 'shotgun proteomics', this name stemming from the analogy of the rapidly-expanding, quasi-random firing pattern of a shotgun that is applied to the shotgun sequencing of DNA. MudPIT falls into the classical branch of proteomics and utilizes micro-capillary HPLC–MS to characterize peptides following enzymatic digestion of protein samples. Following the initial stage of enzymatic digestion the digested peptides are separated using either a two or three stage LC process using either a biphasic column (two-phase MudPIT) consisting of reversed phase material flanked by strong cation exchange resin, or a triphasic column (three-phase MudPIT) consisting of reversed phase material flanked by strong cation exchange resin followed by additional reverse phase material. This means that the peptides are separated initially by their hydrophobicity and then by their charge. Once separated the peptides are then analyzed using MS methodologies. Three-phase MudPIT allows for superior analysis when compared with two-phase MudPIT, however, the volume of sample that can be analyzed is limited to approximately 100 µg to prevent damage to the columns. This can cause problems in the detection of low-abundance proteins and peptides. In addition the faster two-phase MudPIT or even simple liquid chromatography (LC)-MS/MS may be preferred if the sample to be analyzed is not of a sufficient complexity to merit three-phase MudPIT. Although MudPIT provides significant advances on the 2D gel electrophoresis methods of analysis there are still limitations in its application as protein activity and protein abundance are not quantified [44].

■ Isotopic labeling

ICAT

ICAT is a technique that provides a sensitive, relatively high-throughput method, which allows for the identification and quantification of proteins in a proteome. It involves the labeling of protein samples with known linkers that can be classified as being either heavy or light and different isotopes of labeling tags. These attach onto amino acids before samples are mixed and enzymatically digested, followed with the resulting peptides being separated and identified through HPLC–MS/MS or LC–MS/MS [1,50]. ICAT was used during the investigation of proteins that were associated with a high risk re-occurrence of stage IV colorectal cancer (CRC) to elucidate the underlying mechanisms of the disease, and also to glean information that could prove beneficial in the development of personalized medicines [50]. A number of proteins for the prognosis of CRC were identified and following statistical analysis on subsequent data produced to validate the findings, the authors confirmed the accuracy of their multi-marker biomarker panel. The results suggest that by analyzing expression levels of a select group of proteins that are considered to alter the drug sensitivity and proliferation of CRC cells, post-treatment prognosis of stage IV CRC patients may be possible [50].

SILAC

SILAC is a simple labeling technique to prepare cell samples for proteomic analysis by MS. SILAC is particularly useful as it removes false positives in protein-interaction studies, reveals large-scale kinetics of proteomes and, as a quantitative phosphoproteomics technology, directly uncovers important points in the signaling pathways that control cellular decisions [47]. Amino acids, often arginine and lysine that have substituted stable isotopic nuclei, are incorporated into newly synthesized proteins in the cell samples. The light isotopic form is incorporated into one cell sample and the heavy isotopic form is incorporated into another cell sample. These cell samples are then grown in parallel until, eventually, all of the cells in the samples have the labeled amino acid in their makeup. As the amino acid is almost identical to the natural form, the cells behave in exactly the same way as they would if they have remained unlabeled. Once the cells are ready for analysis they are digested with trypsin before undergoing analysis, often by LC–MS/MS. Despite the successful introduction of SILAC and the previously mentioned advantages to the method, there are however several major sources of quantification error with this technique. These include the incomplete incorporation of isotopic amino acids, the conversion of arginine to proline, and experimental errors that present themselves as a result of the final sample mixing [51].

iTRAQ

iTRAQ is a labeling technique that allows for the quantification of proteins in proteomic investigations. iTRAQ delivers the advantages that this labeling technique can be applied to cells and tissues after cell lysis. This means that iTRAQ is not limited to only the quantification of proteins in systems in which stable isotopes can be incorporated into cellular proteins during cell culture. In addition, it is possible to compare multiple samples in one MS run thanks to the availability of different iTRAQ labels [48].

■ Label-free proteomic analysis

Over recent years label-free quantitative proteomic techniques have developed rapidly, and as a result of this, there are now rapid, low-cost platforms for the analysis of protein expression levels. The most commonly used label-free quantification methods in use are peak intensity-based comparative LC–MS and spectral count-based LC–MS/MS.

The development of analytical methods to deliver nano-HPLC separation that is reliable, accurate and precise in addition to the availability of high resolution mass spectrometers and delicate computational tools has greatly improved the reliability and accuracy of label-free, comparative LC–MS. In addition to this, the computer software commercially available for the purpose of analyzing the data generated from such proteomic techniques is now automatically able to detect, match and analyze peptides from hundreds of different LC–MS experiments simultaneously, which provides a high-throughput technique for disease-related biomarker discovery [52].

ABPs

ABPs can fall into both the classical and functional branches of proteomics. The probes are capable of binding covalently to specific enzymes involved in particular cellular mechanisms with the probes, typically consisting of a reactive component and a tag, most often an affinity tag. All tags allow for the detection of the tagged protein however the preference for the affinity tag is that an affinity tag allows for enrichment during the MS stage of the analysis, where the other tags do not. One of the most important factors in the application of ABPs is the selection of the position of the tag through the choice of reactive group. Ideally the tag should be as close to the enzymes active site as possible, and the results produced following MS allow for the active proteins of a proteome to be identified. The main advantage of ABPs is that the post-translational modifications that are responsible for the precise regulation of protein activity are taken into account [53].

■ Functional proteomic analysis

Protein arrays

Protein arrays have advantages over other proteomic techniques in that they are both simple to work with and provide a high-throughput platform for proteomic analysis. That said, they are much more complicated to produce, calibrate and validate when compared with DNA microarray platforms. Protein arrays are capable of detecting specific proteins of interest and also at times have the ability to identify protein–protein and protein–small molecule interactions. There are three main types of protein array, an analytical protein array (APA), a functional protein array (FPA) and a reverse-phase protein array (RPPA) [54]. Within the three main subsets (APA, FPA and RPPA) there are numerous other forms of array

Key Term

Metabolome: Complete list of all the small-molecule metabolites present in an organism. In a similar fashion to the human genome project, since 2005 the human metabolome project has been in progress with the aim to identify, quantify, catalog and store all of the metabolites that can be found in human tissues and bio-fluids.

including 3D surface structure, nanowell and plain glass chip arrays, which are differentiated by the variations in chip design [55].

APAs most often consist of antibodies immobilized onto a glass chip and they are used in the analysis of samples that may contain a large and diverse mix of different proteins. APAs find application in clinical diagnostics through the measure of factors including protein expression, binding and specificity in differential expression profiling [54,56].

FPAs, in contrast, are constructed using full length proteins or their domains and they are used to measure protein–protein or protein–small molecule interactions on a whole proteome scale [54,57,58].

RPPAs are the final set of protein arrays and these involve the use of antibody probes that are applied to cell lysate from cells collected from the tissue that is under investigation. Assays to detect the antibody binding are then applied to achieve the quantification of the proteins in the lysate [54,59].

Phage displays

Phage display technology makes use of bacteriophages to study protein–protein, protein–peptide and protein–DNA interactions. This is achieved through the use of bacteriophages that have undergone batch cloning of DNA into their genome, with the idea that the proteins for which the inserted DNA codes are presented on the surface of the phenotype. The bacteriophages for which the batch-cloning process is successful proceed through an enrichment process, which eventually leads to bacteriophages that are ready for binding analysis. In the case of antibody-based phage displays, analysis often takes the form of ELISA methodologies and the data generated are used in the construction of large phage antibody libraries that are used in the identification of novel therapeutic targets. In addition to target identification, the libraries are also used to develop methods that allow for the selection of candidate and lead compounds based on the analysis of a particular ligands biological activity [60].

In addition to the biological approach to phage displays, there are nonbiological alternatives that use barcoded nanoparticles, these are single gold nanoparticles with each nanoparticle being covered by an antibody that is specific to a target protein plus thousands of specific DNA strands. The antibody is responsible for the protein binding while the DNA strands are what account for the barcode-like label for the bound protein. These barcode-like nanoparticles provide assays for identifying bound molecules though a similar principle to the biological alternatives [61].

Ribosomal display

Ribosome display technology presents an innovative cell-free *in vitro* technology that allows for a fast means of isolating and producing high-affinity peptides or proteins. The desired proteins are enriched once they are displayed by multiple selection rounds. An advantage of ribosomal display is that at this stage, no transformation stage is required meaning that a potential loss of library diversity is avoided [62]. A second advantage is that a complete cycle of display and selection can be carried out in 24 h, meaning that the existing gene repertoire can be rapidly scanned. Modification of the proteins produced can be achieved post-isolation via random or directed molecular evolution for affinity maturation, as well as selected for characteristics such as protein stability, folding and functional activity. Recently, the field of display technologies has become more prominent due to the generation of new scaffolds for ribosome display, isolation of high-affinity human antibodies by phage display, and their implementation in the discovery of novel protein–protein interactions. Applications for this technology extend into the broad field of antibody engineering, proteomics, and synthetic enzymes for diagnostics and therapeutics in cancer, autoimmune and infectious diseases, neurodegenerative diseases and inflammatory disorders [63].

Two-hybrid systems

Two-hybrid systems find application in drug discovery in the drug targeting of specific protein–protein interactions. They utilize a ‘lock and key’ type association between proteins to identify protein–protein interaction. The ‘lock’ protein or proteins are fused to the DNA binding domain of a transcription activator while the ‘key’ protein or proteins are fused to the activation domain of the transcription activator. If interaction between the ‘lock’ and ‘key’ proteins occurs this results in the reconstitution of a transcription factor, which in turn leads to the activation of a reporter gene and a measurable response, most often growth under selective conditions, or the development or change

of color through proteins such as β-galactosidase and green fluorescent protein [64].

Two hybrid systems used for the study of protein–protein interactions, were originally developed in yeast, and the use of yeast two-hybrid systems is widely reported. Despite their widespread use there are still some drawbacks of yeast two-hybrid systems including issues regarding the experimental reproducibility, false positive and negative results due to issues regarding the protein expression as fusion proteins and also the inability to analyze membrane-based protein–protein interactions. In order to overcome some of the limitations of yeast two-hybrid systems there have been many variations of the model with more recent developments including the membrane yeast two-hybrid system, which allows for the study of protein–protein interactions involving membrane proteins [65], something that was not possible through the use of the original yeast two-hybrid systems. Also the development of mammalian two-hybrid systems has been achieved. However, these are not well suited to the same degree of high-throughput analysis as yeast two-hybrid systems, and, in fact, are used more often for the confirmation of findings arising from the yeast two-hybrid systems [66].

Selected reaction monitoring quantification

Selected reaction monitoring is a quantitative proteomic technique that can be used for non-scanning targeted quantitative proteomics through the application of MS. Following ionization in an electrospray source, a peptide precursor is first isolated to obtain a substantial ion population that consists mostly of the intended species. This ion population is then separated into multiple fragments, each of which contain ions whose signal strength correlates to the abundance of the peptide in the sample. This experiment is performed on triple quadrupole mass spectrometers where fragmentation results in quantitative analyses with unmatched sensitivity [67].

Metabolomics

The main challenge in the mapping of the **metabolome** and its network of pathways is the diverse nature and sheer quantity of metabolites and metabolite concentrations present in the cellular environment. However, metabolomics does have one major benefit over other ‘omics in that the pathways must obey strict stoichiometric constraints, meaning metabolomics are more discriminating [68]. In metabolomics, small-scale changes in gene expression and protein concentration are amplified to produce significant large-scale changes in the metabolite concentration and in addition, metabolite concentration may be altered by both disease and drug intervention [69].

■ Generation of data for pathway databases

There are a number of methods by which metabolite concentrations can be measured. The three main categories include chromatographic methods, capillary electrophoresis (CE), and NMR techniques. The data gathered from these approaches are then used to generate pathway databases, which are used to understand the metabolome.

Chromatographic methods

With respect to chromatographic metabolic analyses, gas chromatography MS (GC-MS) and LC–MS are preformed most regularly. The major benefit of GC-MS is the high sensitivity that the technique offers. There are, however, drawbacks – GC-MS is unable to analyze metabolites that are non-volatile, temperature-sensitive or highly polar. Derivatization, a process by which compounds are chemically manipulated, can be implemented to broaden GC-MS application to metabolites for which analysis would previously be impossible. There are limitations and disadvantages to this, including the need to purify samples to remove reaction by-products, the production of highly unstable compounds for analysis and use of hazardous and odorous chemicals [70].

LC–MS has not been in use for metabolomic analysis for as long as GC-MS systems, however, they are useful tools and they are becoming increasingly popular as they are able to analyze the non-volatile, temperature-sensitive and polar metabolites without the need for a derivatization stage. In addition, the development of new stationary phase materials and advances in ion-pair techniques means that where analysis of ionic and highly polar compounds that are common place in bodily fluids was previously troublesome, it is now relatively simple [71].

CE

CE is a nonchromatic separation technique and brings with it advantages over GC-MS and LC–MS techniques, and is often used to complement analysis performed on these platforms. Analysis is quick as there is very little work involved in preparing samples for analysis, also the methods

use little or no organic solvents. In addition to this, the technique is inexpensive as it uses simple silica capillaries and this means there is no need for expensive stationary phases that are necessary for GC and LC techniques. CE separates compounds based on the ratio of their charge and mass. As such, CE is most useful for the separation of compounds that are polar or possess a charge. Operating as a standalone method, CE can encounter problems stemming from limited sensitivity, however, combining CE with MS it is possible to overcome this and in addition yield information on metabolite structural characteristics. Through this means, metabolic projects are using CE-MS with increasing frequency to perform their analysis [72].

NMR

NMR differs from the other techniques applied in metabolomics that apply first a separation stage followed by a second detection stage, in that for metabolite analysis by NMR, a stage for the separation of the metabolites is not required. This makes the method relatively simple and in addition, following analysis by NMR, the sample can be used for additional testing. NMR also carries with it the advantage that it is possible to analyze multiple metabolites present in a sample all at the same time. The main limitation of NMR application to metabolomics is that NMR cannot achieve the same level of sensitive detection that is achievable through the implementation of techniques that utilize MS in the detection stage of the analysis [73].

Pathway databases

One of the challenges in metabolomic analysis is handling the vast quantities of data that are generated, and using this to build a functional and coherent picture of the metabolome. To overcome this, there are initiatives integrating newly generated information with existing data sets and in this way developing large pathway databases such as: The Human Metabolome Database, Lipid maps, Mass Bank, National Institute of Standards and Technology, METLIN [74] and the Kyoto Encyclopaedia of Genes and Genomes (KEGG) [75]. It should be noted, however, that as it stands our understanding of pathway interaction is far from complete despite the vast quantities of research that have been undertaken.

METLIN

METLIN is an online resource in which metabolite data is stored in an open access database. Used in metabolomics for analyzing and archiving data allowing for predictive modeling, METLIN exists to link many biological resources and deliver:

- Physical properties of metabolites, drugs and drug metabolites;
- Metabolite, drug and drug metabolite structural information as well as identifying structurally similar metabolites;
- The CAS number of each metabolite and provide a direct link to the metabolites entry in KEGG;
- Highly accurate Fourier transform MS data from reference samples used in producing *m/z* measurements;
- Searchable MS/MS data for reference use gathered from known metabolites and their derivatives, used widely in metabolite identification.

Updated frequently, the METLIN database allows a user to search for a range of metabolic resources [76,104].

KEGG

KEGG is a widely used resource in the reconstruction of genome-scale metabolic models (GEMs) and in systems biology. It is a computer model for biological networks and includes data from three areas of interest including chemical-based, genome-based and biological network-based databases. Throughout KEGG, the information is interlinked and also connected to outside data resources that are most often open access. The usefulness of KEGG stems from its ability to integrate information gathered on molecular pathways, genome projects and compound databases that is complemented by the logging of information referring to cellular pathways, both metabolic and signaling, and also information on human disease and drug development [77]. Despite the best of efforts, there are still gaps in the pathway databases, although it is possible to loosely predict the missing data using interaction studies such as yeast two-hybrid [78] and tandem affinity purification studies [79], though this can induce a high degree of error. An initial limitation of pathway databases was software compatibility, and in an effort to minimize this problem, systems biology mark-up language was created to ensure interoperability. Systems biology mark-up language is the main way in which

metabolism, metabolomics and systems biology are integrated [68].

To summarize, metabolomics examines metabolite concentration that can be changed as a result of disease and/or drug intervention. Through techniques including chromatography, CE and NMR, data are collected and used to create pathway databases. These are designed to make sense of the huge amounts of data generated, and promote understanding of the metabolome and what changes in metabolite concentration result from or will result in. Thanks to the nature of metabolomic data producing large maps of the metabolome, metabolomics lends itself nicely to systems biology approaches.

The role of metabolomics in systems biology

Unlike the comparatively old fashioned reductionalist approach, which focuses on individual aspects of a system, systems biology links the individual aspects to provide a complete picture of a biological system. Systems biology is a research strategy by which *in silico* models are linked into multiple disciplines of experimental biology with the shared goal of unlocking the molecular mechanisms that network together to produce a complex biological system [80]. The area of metabolomics is an important contributor in a systems biology approach, linking pathways described within the field of metabolomics and also by combining with large-scale biochemical, signaling, protein, miRNA, and gene regulatory networks. It is possible to reconstruct and analyze the underlying mechanisms controlling human cellular processes and it is in this way that the recent developments in systems biology can be extended to medical applications [81]. Systems biology is rapidly expanding and developing, allowing for the analysis of complex systems by reconstructing biochemical networks through the combination of mathematical models and data generated by the 'omics, including genomics, transcriptomics, proteomics and metabolomics [82]. There are two approaches to systems biology, the bottom-up approach and the top-down approach [83].

The bottom-up approach is based around hypothesis and involves the production of a mathematical model based on extensive prior understanding of multiple sub-systems. This metabolic model is then used to describe the system as a whole.

In contrast, the top-down approach is based on data collection. The main advantage of the top-down approach is that existing knowledge and understanding of the system and its sub-systems is not necessary. In this method, data produced from high-throughput experiments are analyzed to determine the function and mechanism of many of the sub-systems. Combining the data from the high-throughput experiments with network reconstruction methodologies and also standard molecular biology, biochemistry and physiology allows for the revelation of patterns in the pathways that are biologically significant and also identify the individual constituent components that make up the network. This in turn allows for the determination of the molecular basis of disease as the widespread changes observed in the metabolome can be used to categorize the molecular phenotype of the samples in question. Moreover, identifying changes in individual metabolite concentrations within the metabolome can lead to the discovery of new biomarkers. Metabolomics have been applied in a systems biology approach to various disease states including cancer [84].

One such strategy involves the use of siRNA strands. A recent investigation sought to discover how ARNT2 functioned in breast cancer signaling and metabolic pathways regulated by HIF-1α [85]. To achieve this, the group used siRNA to knockdown *ARNT2* mRNA expression in human breast cancer cells (MCF7), and through metabolite profiling using NMR and metabolite databases, the effect of *ARNT2* downregulation was correlated to a change in metabolite concentration. Statistical applications were then implemented to reveal biomarkers. In this way it was observed that the levels of the metabolites; betaine, glucose, glycine, phosphocholine, pyruvate, lactate, ATP, ADP and AMP were significantly reduced in the cells that had been treated with the siRNA when compared with untreated control MCF7 cells.

In addition, gene expression analysis revealed that downregulation of *ARNT2* mRNA expression significantly reduced the expression of Glut-3, which is a predominant transporter in neurones and cancer cells involved in glucose uptake, indicating that the downregulation of *ARNT2* mRNA expression affects a shift in glucose metabolism. From their results, the authors were able to draw conclusions in agreement with other reports that the downregulation of *ARNT2* mRNA expression impairs the function of HIF-1 regulated glycine synthesis and glucose metabolism in MCF7 cells [86,87].

GEMs

Systems biology is innovative in its ambition to bring together information gained from various scientific disciplines and utilize the combined knowledge to improve the understanding of disease over a diverse span including immunological disorders, inflammatory-based disease, infectious disease and diseases affecting the respiratory or nervous systems. This includes diseases related to human metabolism and it is a common view that comprehending the underlying metabolic pathways will lead to the onset and progression of metabolic diseases being fully explained. Due to the almost insurmountable complexity of the human metabolic network, the implementation of computer modeling is essential. Through this, the development of GEMs may be possible, and through their creation and application in a systems biology approach, the window of opportunity is presented to analyze the specific state of any given human tissue. To date, GEMs have not been used for any practical application in identifying how errors in metabolism are related to disease states. It is clear however, that such an application providing detailed analysis of the metabolome of multiple subjects displaying various phenotypes could lead to answers regarding the molecular advance of disease and in this way, identify new biomarkers for that disease while in addition open the door to the possibility of systems medicine, providing tailor-made medicines dependent upon patient and condition.

GEMs are produced using a multitude of information sources with the majority of the data coming from public databases such as KEGG, METLIN, HumanCyc, Reactome, BioCarta, Rhea, the Human Protein Atlas and the Human Metabolome Database. Utilizing these sources, as well as including other data, it is possible to produce GEMs that are capable of running simulations that are reliable representations and are specific to a tissue of choice. Recon1 [88], the Edinburgh Human Metabolic Network [89] and HumanCyc [90] are three existing GEMs, however, these are all generic and nontissue specific, rendering them unsuitable for application to specific cell-type studies and diseases [91]. The importance of cell-type-specific GEMs is clear as intertissue and intercellular metabolism in humans varies. As reported recently [81], two groups independently published reports on GEMs that were developed manually and are specific for hepatocytes [92,93]. Being manually produced means that these models are a suitable standard in the evaluation of subsequent automated or semi-automated approaches. GEMs now exist for kidney [94], brain [95], erythrocytes [96] and alveolar macrophages [97]. A recent article described the implementation of the Integrative Network Interference for Tissues algorithm to generate genome-scale active metabolic models for 69 different cell types and 16 cancer types [91]. The result of which allowed for comparison between the metabolic profiles of healthy and cancer cells, which in turn lead to the identification of cancer-specific metabolic features that are generic potential drug targets for cancer treatment. The authors also stated that the models generated will form the first stage in establishing a Human metabolic atlas, a tool designed to better understand complex diseases by providing the tools for tissue- and organism-level simulations. With this in mind, it should be noted that even the most recent GEMs currently available are incomplete with many gaps in the data. Although computational approaches are employed to fill these unknowns, they may be prone to error if the automation is initiated with poor insight. At this point it can be said that there have been significant advances in the production of specific GEMs on both a tissue and a cellular level, however there is still some distance to go and the need to continue to develop and create GEMs is constant.

Metabolic disorders

There are three main diagnostic groups into which metabolic disorders stemming from congenital metabolic errors can be organized:

- Disorders that interfere with the production or breakdown of complex molecules causing symptoms that are permanent, increase in their intensity and are disassociated from other disease progression and food consumption;
- Disorders such as phenylketonurea, homocystinuria and maple syrup urine disease that stem from malfunctions in amino acid metabolism. Similarly, organic acidurias, congenital urea cycle defects and galactosaemia all result in a metabolic block, leading to the build up of toxic compounds;
- Disorders such as congenital lactic acidemias, fatty acid oxidation defects, gluconeogenesis and mitochondrial respiratory chain disorders, which result in symptoms due to insufficient energy production or utilization in key areas such as the brain, liver, myocardium or muscle.

Human metabolism is a highly complex process consisting of many reaction pathways, and

Executive summary

Genomics

- Illumina (Solexa) sequencing is a next-generation sequencing technique that provides a platform capable of delivering high throughput at relatively low cost using a bead-based method.
- 454 – Roche sequencing can be used for whole-genome sequencing, target resequencing, metagenomics and transcriptome sequencing making it one of the most efficient platforms on the market.
- Supported oligonucleotide ligation and detection sequencing is a variation of massively parallel signature sequencing that uses hybridization-ligation.
- Ion semiconductor sequencing is unlike most other sequencing techniques on the market, in that the analysis is not based on optical output. Instead the measurement of ions generated by template-directed DNA polymerase synthesis allows for genome sequencing.
- Single-molecule real-time sequencing is a sequencing based on synthesis approach that carries with it the benefit that DNA amplification is not necessary and the associated problems are avoided. Techniques currently in development that show promise include nano-pore sequencing, hybridization sequencing and sequencing with MS.
- Biomarkers can be identified through the application of genomic techniques and represent measurable entities, which indicate a diseased state when their level of expression differs from that of normal.
- Tissue expression profiling plays an instrumental role in many aspects of drug discovery including target identification, target validation, identification of candidate compounds, pharmacogenomics, biomarker development, the evaluation of clinical trials and the gathering of toxicological information.
- Pharmacogenomics investigates the correlation between an individual's genetic makeup and their response to medicines through genotype and phenotype differences in the pharmacodynamics and pharmacokinetics of drug metabolism.
- Genome-wide epigenetics involve sequencing based approaches and promise to contribute towards a more precise and complete understanding of gene regulation and to yield insights into the interplay between genomes and the environment.

Proteomics

- Classical proteomic analysis delivers the identify and abundance of proteins in the sample being analyzed. Multidirectional protein identification technology is an example of shotgun proteomics and utilizes HPLC–MS to characterize peptides following the enzymatic digestion of protein samples.
- Isotope coded affinity tags generates information on protein identification and quantification in a sensitive, high-throughput manor. Stable isotopic labeling by amino acids in cell culture is the labeling of cells in culture with stable isotopic markers to prepare cell samples for proteomic analysis by MS.
- Isobaric tags for relative and absolute quantification have the advantage that it can be applied to tissues and cells after cell lysis. This means that isobaric tags for relative and absolute quantification is not limited to use in cell cultures in the way stable isotope labeling by amino acids in cell culture techniques are.
- Label-free proteomic analysis has developed rapidly over recent years and as such, rapid, low-cost platforms for the analysis of protein expression are now available. The most common use is peak intensity based comparative LC–MS and spectral count-based LC–MS/MS.
- Activity-based probes can fall into classical or functional proteomics and utilize probes consisting of a reactive component and a tag. The main advantage of activity-based probes is that post-translational protein modifications are taken into account.
- Functional proteomic analysis differs from the classical proteomic analysis in that function proteomics generate a more complete picture of protein abundance, functions, indications and interactions.
- Two-hybrid systems find application in drug discovery in the drug targeting of specific protein–protein interactions. They utilize a 'lock and key' type association to identify protein–protein interaction.

Metabolomics

- The three main categories by which metabolite concentrations can be measured include chromatographic methods, capillary electrophoresis and NMR techniques. The data gathered from these approaches are then used to generate pathway databases that are used to understand the metabolome.
- Pathway databases are the result of large-scale metabolic analysis, and the only way to handle and make sense of the huge amounts of data that are being generated.
- Systems biology is a research strategy by which *in silico* models are linked into multiple disciplines of experimental biology. The area of metabolomics is an important contributor in a systems biology approach, linking pathways described within the field of metabolomics and also combining them with large-scale biochemical, signaling, protein, miRNA and gene regulatory networks.
- Genome-scale metabolic models, when applied in a systems biology approach, are expected to allow for the analysis and later the simulation of the specific state of any given human tissue. Metabolic disorders result from the malfunction of metabolic pathways. In mapping the metabolic network it is possible to derive the molecular basis of the metabolic disease. This, in turn leads, to the opportunity to develop a strategy to combat the health risk.

metabolic diseases result from pathway malfunction. In mapping the metabolic network it is possible to derive the molecular basis of the metabolic disease. From this follows the development of a strategy to combat this emerging health risk. Due to the complexity of the metabolic networks it is often challenging to narrow down the particular mechanism by which a metabolic disease manifests itself from initiation to progression. The complexity of the metabolic networks is not solely due to the number of metabolic pathways involved, there is also the involvement of various tissue types, and in addition, the role of numerous genes that are capable of producing an even greater number of different mechanistic permutations that deliver the same phenotype. Some of the most common metabolism-related diseases include obesity, diabetes, hypertension, hyperinsulinemia, dyslipidemia and cancer. Recent advances in computing power, metabolomics and systems biology are moving to manage the complexity of the task in hand and allow for the development of GEMs.

Future perspective

Genomics, proteomics and metabolomics continue to develop and expand, revealing new insights into the world of drug discovery and development. With the increasing focus, the power of the individual disciplines is further boosted and more rapid advances are anticipated. Systems biology approaches, coupled with improved computer systems to manage data handling and help make sense of the huge amounts of data generated hold great potential for the future of drug discovery. The combination of well established genomic, proteomic and metabolomic techniques and also the integration of novel and emerging technologies positions 'omics and systems biology approaches at the forefront of medical research. The scope in which systems biology is expected to provide significant insight and advance will grow to cover the whole of the discovery process, from drug development and clinical trials to personalized medicine. As this branch of the science matures, systems biology, advanced by the development of techniques discussed in this review, promises to become a dominant approach in drug discovery and development by overcoming the limitations and enhancing the prospects of individual omic technologies.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as:
▪ of interest

1 Burbaum J, Tobal GM. Proteomics in drug discovery. *Curr. Opin. Chem. Biol.* 6, 427–433 (2002).

▪ **Insight into the application of proteomics in drug discovery, detailing many of the major areas of interest.**

2 Butchdanger E, Zimmermann J, Mett H *et al.* Inhibition of the Abl protein-tyrosine kinase *in vitro* and *in vivo* by a 2-phenylaminopyrimidine derivative. *Cancer Res.* 56, 100–104 (1996).

3 Apic G, Ignjatovic T, Boyer S, Russell RB. Illuminating drug discovery with biological pathways. *FEBS Lett.* 579, 1872–1877 (2005).

4 Shah T, de Villiers E, Nene V *et al.* Using the transcriptome to annotage the genome revisited: application of massively parallel signature sequencing (MPSS). *Gene* 336, 104–108 (2006).

5 Bentley DR. Whole-genome re-sequencing. *Curr. Opin. Gen. Dev.* 16, 545–552 (2006).

6 Tawfik DS, Griffiths AD. Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.* 16, 652–656 (1997).

7 Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92, 255–264 (2008).

▪ **Useful in detailing the application of next-generation genetic sequencing techniques in functional genomics.**

8 Guo L, Liang T, Lu Z. A comprehensive study of multiple mapping and feature selection for correction strategy in the analysis of small RNAs from SOLiD sequencing. *Biosystems* 104, 87–93 (2011).

9 Rothberg JM, Hinz W, Rearick TM *et al.* An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475, 348–352 (2011).

10 Eid J, Fehr A, Gray J *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* 323, 133–138 (2009).

11 Milward EA, Daneshi N, Johnstone DM. Emerging real-time technologies in molecular medicine and the evolution of integrated 'pharmacomics' approaches to personalized medicine and drug discovery. *Pharmacol. Ther.* 136, 295–304 (2012).

12 McGinn S, Gut IG. DNA sequencing – spanning the generations. *N. Biotechnol.* S1871–S6784(12), 00869-2 (2012).

13 Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* 4(4), 265–270 (2009).

14 Derrington IM, Butler TZ, Collins MD *et al.* Nanopore DNA sequencing with MspA. *Proc. Natl Acad. Sci. USA* 107(37), 16060–16065 (2010).

15 Wu HC, Astier Y, Maglia G, Mikhailova E, Bayley H. Protein nanopores with covalently attached molecular adapters. *J. Am. Chem. Soc.* 129(51) 16142–16148 (2007).

16 Garaj S, Hubbard A, Reina A, Kong J, Branton D, Golovchenko JA. Graphene as a subnanometre trans-electrode membrane. *Nature* 467, 190–193 (2010).

17 Augenlicht LH, Kobrin D. Cloning and screening of sequences expressed in a mouse colon tumor. *Cancer Res.* 42(3), 1088–1093 (1982).

18 Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270(5235), 467–470 (1995).

19 Lashkari DA, DeRisi JL, McCusker JH *et al.* Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl Acad. Sci. USA* 94(24), 13057–13062 (1997).

20 Karakach TK, Flight RM, Douglas SE, Wentzell PD. An introduction to DNA microarrays for gene expression analysis. *Chemometr. Intell. Lab.* 104, 28–52 (2010).

- **Very good starting point for developing an understanding of the application of DNA microarrays in gene expression analysis.**

21 Li X, Qi X, Miao L *et al.* Detection and subtyping of influenza A virus based on a short oligonucleotide microarray. *Diagn. Microbiol. Infect. Dis.* 65, 261–270 (2012).

22 Grover V, Pierce ML, Hoyt P, Zhang F, Melcher U. Oligonucleotide-based microarray for detection of plant viruses employing sequence-independent amplification of targets. *J. Virol. Methods.* 163, 57–67 (2010).

23 Zhao S, Recknor J, Lunney JK *et al.* Validation of a first-generation long-oligonucleotide microarray for transcriptional profiling in the pig. *Genomics* 86, 618–625 (2005).

24 Barrett JC, Kawasaki ES. Microarrays: the use of oligonucleotides and cDNA for the analysis of gene expression. *Rev. Res. Focus* 8(3), 134–141 (2003).

25 van den Boom D, Beaulieu M, Oeth P *et al.* MALDI-TOF MS: a platform technology for genetic discovery. *Int. J. Mass Spectrom.* 238, 173–188 (2004).

26 Lewin DA, Weiner MP. Molecular biomarkers in drug development. *Drug Discov. Today* 9(22), 976–983 (2004).

27 Meulenbelt I, Kraus VB, Sandell LJ, Loughlin J. Summary of the OA biomarkers workshop 2010 – genetics and genomics: new targets in OA. *Osteoarthr. Cartil.* 19, 1091–1094 (2011).

28 Chengalvala MV, Chennathukuzhi VM, Johnston DS, Stevis PE, Kopf GS. Gene expression profiling and its practice in drug development. *Curr. Genomics* 8, 262–270 (2007).

29 Hernandez DG, Nalls MA, Moore M *et al.* Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain. *Neurobiol. Dis.* 47, 20–28 (2012).

30 Akhtar S, Benter I. Toxicogenomics of non-viral drug delivery systems for RNAi: potential impact on siRNA-mediated gene silencing activity and specificity. *Adv. Drug Deliv. Rev.* 59, 164–182 (2007).

31 Zhang J, Liu Q, Li Y, Hall SH, French FS, Zhang YL. Genome-wide profiling of segmental-regulated transcriptomes in human epididymis using oligo microarray. *Mol. Cell. Endocrinol.* 250, 169–177 (2006).

32 Yudate HT, Kai T, Aoki M *et al.* Identification of a novel set of biomarkers for evaluating phospholipidosis-inducing potential of compounds using rat liver microarray data measured 24 h after single dose administration. *Toxicology* 295, 1–7 (2012).

33 Knudson TB, Daston GP. MIAME guidelines. *Reprod. Toxicol.* 19, 263 (2005).

34 Einbond LS, Soffritti M, Esposti DD *et al.* Pharmacological mechanisms of clack cohosh in Sprague-Dawley rats. *Fitoterapia* 83, 461–468 (2012).

35 Hardiman G. Application of ultra-high throughput sequencing and microarray technologies in pharmacogenomics testing. In: *Therapeutic Drug Monitoring: Newer Drugs and Biomarkers* (Chapter 7). Dasgupta A (Ed.). Elsevier Science Publishing Co Inc., London, UK, 144–159 (2012).

36 Xu R, Wang Q. A knowledge-driven conditional approach to extract pharmacogenomics specific drug–gene relationships from free text. *J. Biomed. Inform.* 45(5), 827–834 (2012).

37 Li J, Lu Z. Systematic identification of pharmacogenomics information from clinical trials. *J. Biomed. Inform.* 45(5), 870–878 (2012).

38 Stone A, Bornhorst J. An introduction to personalized medicine. In: *Therapeutic Drug Monitoring: Newer Drugs and Biomarkers* (Chapter 6). Dasgupta A (Ed.). Elsevier Science Publishing Co Inc., London, UK, 122–142 (2012).

39 Sati S, Ghosh S, Jain V, Scaria V, Sengupta S. Genome-wide analysis reveals distinct patterns of epigenetic features in long non-coding RNA loci. *Nucleic Acids Res.* 40(20), 10018–31 (2012).

40 Huss M. Introduction into the analysis of high-throughput-sequencing based epigenome data. *Brief. Bioinform.* 11(5), 512–523 (2010).

41 Ohtsuki S, Schaefer O, Kawakami H *et al.* Simultaneous absolute protein quantification of transporters, cytochromes P450, and UDP-glucuronosyltransferases as a novel approach for the characterization of individual human liver: comparison with mRNA levels and activities. *Drug Metab. Dispos.* 40(1), 83–92 (2012).

42 Sidoli S, Cheng L, Jensen ON. Proteomics in chromatin biology and epigenetics: elucidation of post-translational modifications of histone proteins by mass spectrometry. *J. Proteomics* 75, 3419–3433 (2012).

43 Cruz-Monteagudo M, Munteanu CR, Borges F, Cordeiro MN, Uriarte E, Gonzalez-Diaz H. Quantitative proteome-property relationships (QPPRs). Part 1: finding biomarkers of organic drugs with mean Markov connectivity indices of spiral networks of blood mass spectra. *Bioorg. Med. Chem.* 16, 9684–9693 (2008).

44 McDonald WH, Ohi R, Miyamoto DT, Mitchison TJ, Yates JR. Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC–MS/MS, 2-phase MudPIT, and 3-phase MudPIT. *Int. J. Mass Spectrom.* 219, 245–251 (2002).

45 Yang Z, Chevolot Y, Ataman-Onal Y, Choquet-Kastylevsky G, Souteyrand E, Laurenceau E. Cancer biomarkers detection using 3D microstructured protein chip: implementation of customized multiplex immunoassay. *Sens Actuators B Chem.* 175, 22–28 (2011).

46 Tunon J, Martin-Ventura JL, Blanco-Colio LM, Lorenzo O, Lopez JA, Egido J. Proteomic strategies in the search of new biomarkers in atherothrombosis. *J. Am. Coll. Cardiol.* 55(19), 2009–2016 (2010).

47 Mann M. Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* 7(12), 952–958 (2006).

48 Ross PL, Huang YN, Marchese JN *et al.* Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* 3, 1154–1169 (2004).

49 Wiedl T, Arni S, Roschitzki B *et al.* Activity-based proteomics: identification of ABHD11 and ESD activities as potential biomarkers for human lung adenomacarconoma. *J. Proteomics* 74, 1884–1894 (2011).

50 Kim H, Kang U, Lee H *et al.* Profiling of differentially expressed proteins in stage IV colorectal cancers with good and poor outcomes. *J. Proteomics* 75, 2982–2997 (2011).

51 Park SS, Wu WW, Zhou Y *et al.* Effective correction of experimental errors in quantitative proteomics using stable isotope labeling by amino acids in cell culture (SILAC). *J. Proteomics* 75, 3720–3732, (2012).

52 Zhu W, Smith JW, Huang CM. Mass spectrometry-based label-free quantitative proteomics. *J. Biomed. Biotechnol.* 2010, 840518 (2010).

53 Kramer HB, Nicholson B, Kessler BM, Altun M. Detection of ubiquitin-proteasome enzymatic activities in cells: application of activity-based probes to inhibitor development. *Biochim. Biophys. Acta* 1823(11), 2029–2037 (2012).

54 Hall DA, Pyacek J, Snyder M. Protein microarray technology. *Mech. Ageing. Dev.* 128, 161–167 (2007).

55 Zhu H, Snyder M. Protein chip technology. *Curr. Opin. Chem. Biol.* 7, 55–63 (2003).

56 Sreekumar A, Nyati MK, Varambally S *et al.* Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. *Cancer Res.* 61, 7585–7593 (2001).

57 Hall DA, Zhu H, Zhu X *et al.* Regulation of gene expression by a metabolic enzyme. *Science* 306, 482–484 (2004).

58 Zhu H, Bilgin M, Bangham, R *et al.* Global analysis of protein activities using proteome chips. *Science* 293, 2101–2105 (2001).

59 Speer R, Wulfkuhle JD, Liotta LA, Petricoin EF 3rd. Reverse phase protein microarrays for tissue-based analysis. *Curr. Opin. Mol. Ther.* 7, 240–245 (2005).

60 Hoogenboom HR, de Bruine AP, Hufton SE, Hoet RM, Arends JW, Roovers RC. Antibody phage display and its applications. *Immunotechnology* 4, 1–20 (1998).

61 Nam J, Stoeva SI, Mirkin CA. Bio-bar-code-based DNA detection with PCR-like sensitivity. *J. Am. Chem. Soc.* 126, 5932–5933 (2004).

62 He M, Taussig MJ. Ribosome display: cell-free protein display technology. *Brief. Funct. Genomic. Proteomic.* 1(2), 204–212 (2002).

63 Rothe A, Hosse RJ, Power BE. Ribosomal display for improved biotherapeutic molecules. *Expert Opin. Biol. Ther.* 6(2), 177–187 (2006).

64 Suter B, Kittanakom S, Stagljar I. Two-hybrid technologies in proteomics research. *Curr. Opin. Biotechnol.* 19, 316–323 (2008).

65 Fetchko M, Stagljar I. Application of the split-ubiquitin membrane yeast two-hybrid system to investigate membrane protein interactions. *Methods* 32, 349–362 (2004).

66 Luo Y, Batalao A, Zhou H, Zhu L. Mammalian two-hybrid system: a complementary approach to the yeast two-hybrid system. *Biotechniques* 22, 350–352 (1997).

67 Lange V, Picotti P, Domon B *et al.* Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* 4(222), 1–14 (2008).

68 Kell DB, Systems biology, metabolic modeling and metabolomics in drug discovery and development. *Drug Discov. Today* 11(23–24), 1085–1092 (2006).

69 D'Alessandro A, Zolla L, Metabolomics and cancer drug discovery: let the cells do the talking. *Drug Discov. Today* 17(1–2), 3–9 (2012).

70 Schauer N, Steinhauser D, Strelkov S *et al.* GC-MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Lett.* 579(6), 1332–1337 (2005).

71 Gika HG, Theodoridis GA, Wingate JE, Wilson ID. Within-day reproducibility of an LC–MS-based method for metabonomic analysis: application to human urine. *J. Proteome Res.* 6(8), 3291–3303 (2007).

72 Ramautar R, Somsen GW, de Jong GJ. CE-MS in metabolomics. *Electrophoresis* 30, 276–291 (2009).

73 Griffin JL. Metabonomics: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterization of xenobiotic toxicity and disease diagnosis. *Curr. Opin. Chem. Biol.* 7(5), 648–654 (2003).

74 Smith C, O'Maille G, Want E *et al.* METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* 27(6), 747–751 (2005).

75 Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, D277–D280 (2004).

76 Smith CA, Maille GO, Want EJ *et al.* METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* 27(6), 747–751 (2005).

77 Kanehisa M, Goro S, Hattori M *et al.* From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* 34, D354–D357 (2006).

78 Aloy P, Russell RB. Potential artefacts in protein-interaction networks. *FEBS Lett.* 530, 253–254 (2002).

79 von Mering C, Krause R, Snel B *et al.* Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417, 399–403 (2002).

80 Sato T, Watanabe H, Tsuganezawa K *et al.* Identification of novel drug-resistant EGFR mutant inhibitors by *in silico* screening using comprehensive assessments of protein structures. *Bioorg. Med. Chem.* 20, 3756–3767 (2012).

81 Mardinoglu A, Nielsen J. Systems medicine and metabolic modeling. *J. Int. Med.* 271, 142–154 (2012).

82 Hawkins R, Hon G, Ren B. Next-generation genomics: an integrative approach. *Nat. Rev. Genet.* 11, 476–486 (2010).

83 Bruggeman J, Westerhoff HJ. The nature of systems biology. *Trends Microbiol.* 15(1), 45–50 (2006).

84 Vinayavekhin N, Homan E, Saghatelian A. Exploring disease through metabolomics. *ACS Chem. Biol.* 5(1), 91–103 (2010).

85 Qin XY, Wei F, Yoshinaga J, Yonemoto J, Tanokura M, Sone H. siRNA-mediated knockdown of aryl hydrocarbon receptor nuclear translocator 2 affects hypoxia-inducible factor-1 regulatory signaling and metabolism in human breast cancer cells. *FEBS Lett.* 585, 3310–3315 (2011).

86 Denko NC. Hypoxia, HIF1 and glucose metabolism in the solid tumor. *Nat. Rev. Cancer* 8, 705–713 (2008).

87 Griffiths JR, McSheehy PM, Robinson SP *et al.* Metabolic changes detected by *in vivo* magnetic resonance studies of HEPA-1 wild type tumors and tumors deficient in hypoxia-inducible factor-1 beta (HIF-1beta): evidence of an anabolic role for the HIF-1 pathway. *Cancer Res.* 62, 688–695 (2002).

88 Duarte NC, Becker SA, Jamshidi N *et al.* Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl Acad. Sci. USA* 104, 1777–1782 (2007).

89 Hao T, Ma HW, Zhao XM, Goryanin I. Compartmentalization of the Edinburgh Human Metabolic Network. *BMC Bioinformatics* 11(393), 1–12 (2010).

90 Romero P, Wagg J, Green ML *et al.* Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol.* 6(1), R2.1–R2.17 (2005).

91 Agren R, Bordel S, Mardinoglu A, Pornputtapong N, Nookaew I, Nielsen J. Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput. Biol.* 8(5), 1–9 (2012).

92 Gille C, Bolling C, Hoppe A *et al.* HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Mol. Syst. Biol.* 6(441), 1–13 (2010).

93 Jerby L, Shlomi T, Ruppin E. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol. Syst. Biol.* 6(401), 1–9 (2010).

94 Chang RL, Xie L, Bourne PE, Palsson BO. Drug off-target effects predicted using structural analysis in the context of a metabolic network model. *PLoS Comput Biol.* 6(9), e1000938 (2010).

95 Lewis NE, Schramm G, Bordbar A *et al.* Large-scale *in silico* modeling of metabolic

interactions between cell types in the human brain. *Nat Biotechnol*. 28, 1279–1285 (2010).

96 Bordbar A, Jamshidi N, Palsson BO. iAB-RBC-283: a proteomically derived knowledge-base of erythrocyte metabolism that can be used to simulate its physiological and patho-physiological states. *BMC Syst. Biol*. 5(110), 1–12 (2011).

97 Bordbar A, Lewis NE, Schellenberger J, Palsson BO, Jamshidi N. Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic reconstructions. *Mol. Syst. Biol*. 6(422), 1–14 (2010).

Websites

101 Gene Logic. www.genelogic.com

102 Iconix Pharmaceuticals. http://iconixpharm.com

103 Clinical Trials. www.ClinicalTrials.gov

104 METLIN. http://metlin.scripps.edu