Laboratory automation advances high throughput bio-inspired research into personalised medicine

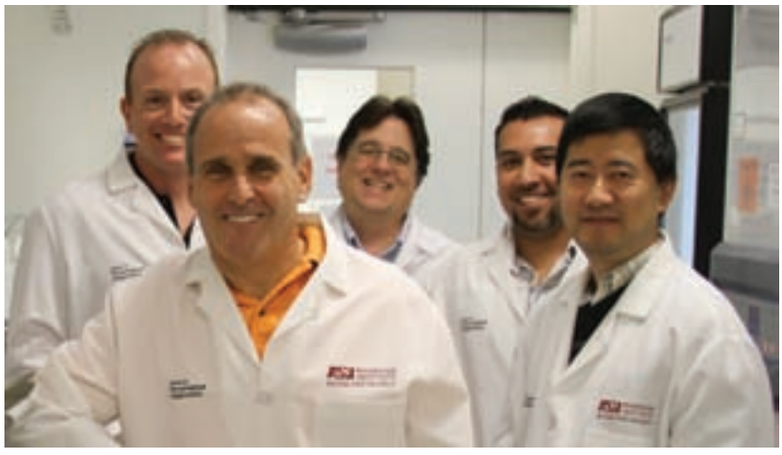
The Biodesign Institute at Arizona State University plays an important role in addressing global healthcare challenges by developing bio-inspired solutions that can be translated into commercially viable products and clinical practices. With 11 different research centres, ranging from environmental biotechnology to infectious diseases and vaccinology, and evolutionary medicine and informatics to personalised diagnostics, the Institute is ideally positioned to capitalise on an extensive range of skills and capabilities to resolve complex issues affecting human health and the environment.

By Dr Joshua LaBaer

With the completion of the human genome project came a phenomenal increase in the pace of biological research, with large-scale and high throughput approaches to biology heralding a new era of technology development and information collection. Advances in molecular medicine have demonstrated that medical conditions previously thought of as single diseases may, in fact, comprise a multitude of disparate molecular variants, each of which responds in a unique manner to any given therapy and has a different prognosis. A perfect example of this is breast cancer, which is now known to exist in a number of forms. Each form of the disease responds to the various chemotherapies and drugs available in a slightly different manner and diagnostic tools are required to establish which therapies are most suitable for a particular patient, an approach known as personalised medicine.

Using transdisciplinary science to tackle personalised medicine

Personalised medicine is changing the way that medical diagnostics, as well as therapeutics, is viewed. For this reason, the Biodesign Institute at Arizona State University has established the Virginia G. Piper Center for Personalized Diagnostics to work specifically in this new field. The Center's mission is to target translational research right at the early development stages of new diagnostic tools, so that they are able to identify molecular manifestations of disease based on



Some key members of the Center's staff. Front row (left to right): Dr Josh LaBaer and Dr Ji Qiu; Back row (left to right): Mike Gaskin, Dr Mitch Magee and Alex Mendoza

individual patient profiles. With a primary focus on the discovery and validation of blood-based biomarkers, the Center is involved in a number of projects across the entire healthcare pathway, investigating strategies for everything from earlier diagnosis to better management of disease, with the ultimate goal of improving patient care and reducing treatment costs through early detection of major illnesses.

The Institute as a whole has long-since adopted the philosophy of transdisciplinary science, bringing researchers from a broad range of backgrounds together to collaborate on a vast array of projects. This ethos is shared by the Center for Personalized Diagnostics, where a multidisciplinary approach has been taken to explore first line diagnostics, drawing on the expertise of engineers, physicists, software engineers, informaticists, molecular biologists and clinicians to study a range of different diseases.

Dr Fernanda Festa loading slides on to the HS 4800 Pro

Targeting the proteome

While the genome is the template leading to the production of proteins, ultimately it is the proteins – which can be thought of as providing the verbs to biology – that exert control over everything from catalysis to digestion to signalling. Most human disease is the result of protein dysfunction and nearly all drugs either act through proteins or are themselves proteins, and a key part of the Center for Personalized Diagnostics' research is the study of functional proteomics; the evaluation of human proteins according to their specific roles in living systems. However, creating technologies that enable the proteome to be studied on the same scale as the genome presents a considerable challenge. Sequencing techniques, such as the traditional Sanger sequencing and next generation sequencing, enable large-scale genomics studies to be performed in a comparatively uncomplicated manner, whereas the unique nature of the protein chemistries means that their measurement is not as straightforward as that of DNA.

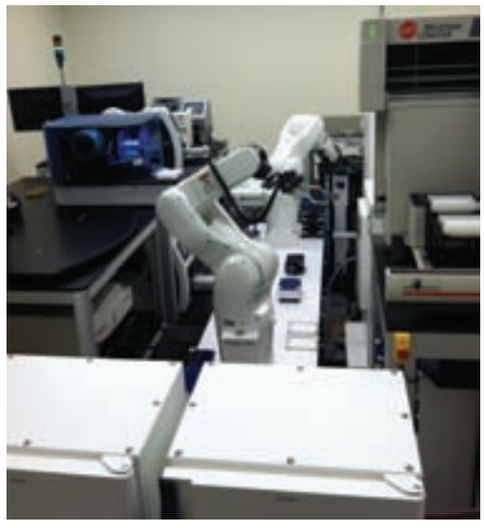
Initially, the Center for Personalized Diagnostics requires cloned copies of the genes that encode the proteins and, through the work of a laboratory consortium – the ORFeome Collaboration – a library of around 15,000 cloned copies of full length human genes has been established to meet the needs of scientists worldwide. These clones are available to the global scientific community via the nonprofit DNASU plasmid repository; a library of cloned genes stored at -80°C in individual, barcoded tubes, in an 880,000 tube capacity, fully automated freezer system (Brooks).

High throughput technology

The proteomics studies carried out at the Center for Personalized Diagnostics involve everything from the production of DNA through to cutting, sequencing and verification, and require high throughput methods for performing molecular biology. Most groups making protein arrays do so by developing high throughput platforms to purify proteins, and then spot the proteins on to the slides. However, this method does not work well for protein purification, particularly with regard to

The Center's PowerScanners provide reproducible, high throughput scanning

yield. The Center for Personalized Diagnostics has developed a novel protein microarray, NAPPA (Nucleic Acid-Programmable Protein Array), where plasmid DNA is printed on the slide and a cell-free extract is added to synthesise the proteins *in situ*, minimising the need for direct manipulation and avoiding the issues associated with protein purification and stability. Thousands of individual spots, each representing a different protein, are printed on slides in a perfect array, then stored dry at room temperature; the DNA on the slides is quite stable under these conditions. The proteins themselves are synthesised on the day of the experiment, enabling researchers to investigate which proteins are the targets of a particular enzyme, partners in an intermolecular interaction, or antigens recognised in an immune response. For personalised diagnostic applications, the arrays are probed with patients' serum to look for responses to specific proteins indicative of disease.



The Center's DNA Factory offers fully automated production of purified DNA, with a capacity of up to 4,600 genes per run

Minimising variation

Preparing DNA for thousands of different proteins is a laborious task when performed manually and, as much of its work requires purified DNA, the Center for Personalized Diagnostics has recently installed a HighRes Biosolutions robotic 'DNA Factory'. This unique, fully automated system for the production of purified DNA has a capacity of around 4,600 genes per run, and is ideal for high throughput applications. The system comprises three fully-articulated Denso robotic arms, three LiCONic 37°C shaking incubators, a Beckman Biomek FX dual arm liquid handling platform, two Hettich automated centrifuges, two KBio Wasp automated plate sealers, three Thermo bulk reagent dispensers, a Brooks XPeel® automated plate de-sealer, a LiCONic -20°C incubator, two HighRes ambient plate storage hotels, a BioTek six-mode plate washer, and a Molecular Devices plate reader. Scheduling software controls all the devices and data is tracked to the Center's Laboratory Information Management System (LIMS) in real time. Bacterial cultures are grown to a suitable density in 96-well plates and then centrifuged. The supernatant is removed and the residual pellets are frozen. Once frozen, the pellets are stable and can be stored indefinitely.

The system's scheduling software has been programmed to recognise which stages are time sensitive and which are not, enabling it to automatically calculate which actions to carry out and when; time-critical steps are performed precisely, while the less time-sensitive steps act as a buffer. Based on its pre-calculated schedule, a microplate of frozen pellets is removed from the freezer and a buffer is added. After shaking to resuspend the bacteria, a lysis buffer is added; this time-critical stage must be performed very precisely. The proteins are precipitated by the addition of a further buffer, followed by centrifugation to remove unwanted sediment and allow the supernatant to be collected. At this stage, the supernatant is typically purified by passing it over an extraction resin that binds to, and retains, the DNA, allowing unwanted material to be washed through and removed. The purified DNA is subsequently eluted, and may be precipitated with ethanol and redissolved in a smaller volume of solvent prior to measuring the optical density of each well. The final yield is calculated from the optical density, and an appropriate volume of buffer is added to ensure that each well contains an equal concentration of purified DNA. On completion of

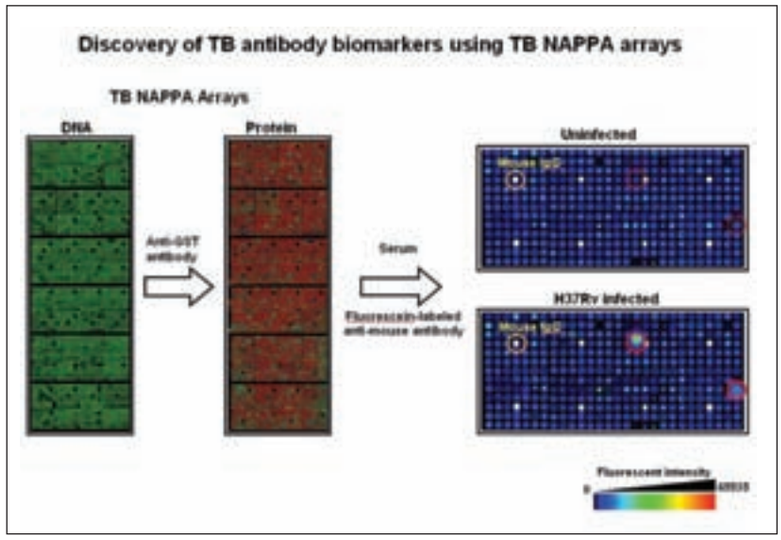
the purification process, a full set of array slides is printed (QArray, Genetix). QC procedures ensure that the array has been printed effectively and that the total protein levels across the slide are consistent, prior to proceeding with the next stage of the study.

Robust analysis

Automation also plays an important role in processing the microarrays, where reproducibility is vital. Historically, the arrays have been processed manually, which required great care as opportunities for variation are inherent. Automated processes, however, offer a precision and reproducibility unmatched by man, particularly in a high throughput setting, and enable studies to be performed on a large scale.

Preparing arrays for between 10,000 and 12,000 proteins involves significant labour, and the Center for Personalized Diagnostics has chosen to automate this procedure on a HS 4800™ Pro automated hybridisation station (Tecan). This system allows the proteins to be synthesised *in situ*, followed by washing and incubation to generate analysis-ready microarrays. Since no manual intervention is required, the arrays can be processed overnight; conversion of the DNA into proteins, washing and incubation steps are accomplished in a single run, allowing sample throughput to be increased considerably. Slide-to-slide reproducibility has also improved compared to manual processing; the same array can be analysed on two different days and still generate the same answer. This set-up also ensures a full audit trail detailing exactly how each sample was treated, virtually eliminating the risk of processing and transcription errors.

A huge consideration for the Center for Personalized Diagnostics' proteomics studies is the need to have statistically valid sample sizes and, as far as possible, to minimise operator-to-operator variation; statistics are far more powerful when the variation is minimised. Each study must have sufficient statistical power to generate the answers needed, taking into account that there is a huge amount of biological variation from person to person. It is also essential to ensure that study sizes are adequate and, consequently, many slides for many patients must be run over and over again. With such a vast number of slides to analyse, reliable automation is a prerequisite; manually loading 80 slides at a time is extremely tedious and very time-consuming. Analysis of the slides is also automated using a PowerScanner™ (Tecan), which allows the array slides to be placed in a 48-slide magazine from where they are automatically loaded and scanned to measure the fluorescence signal and determine which of the thousands of spots have responded to patient serum.

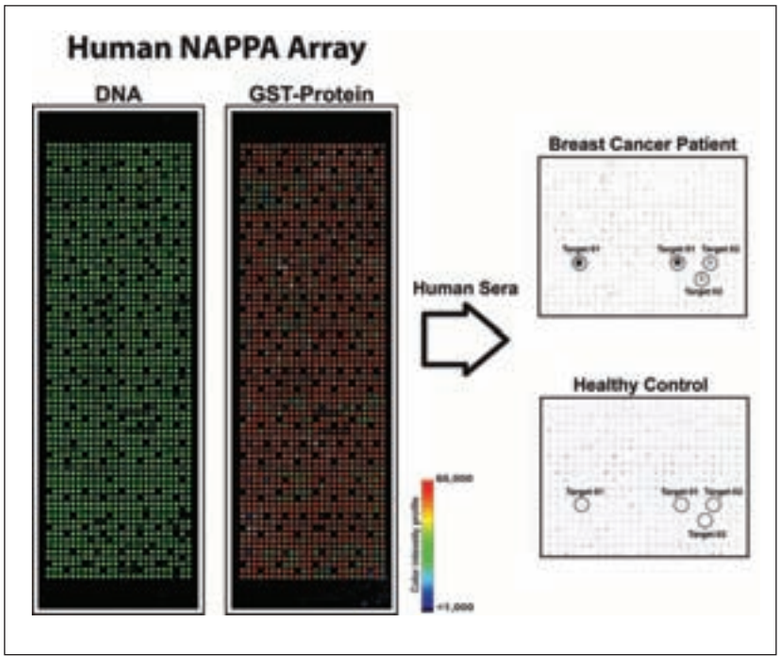


Discovery of antibody biomarkers indicating tuberculosis (TB) infection using TB NAPPA arrays. 2,103 expression clones encoding the TB target proteins fused to C-terminal GST tags were printed along with a polyclonal anti-GST antibody in single spot on the array surface. DNA capture was confirmed by PicoGreen staining (DNA). The quantity of protein displayed *in situ* was assessed by a monoclonal anti-GST antibody (GST-protein). After probing the NAPPA protein arrays with mouse sera (without and with H37Rv infection), the antibody bound to the target was detected using fluorescein conjugated anti-mouse IgG antibody. Mouse IgG was included as a registration control (yellow circle) and candidate targets are indicated (red circle)

Informatics also has a role to play. The fluorescence readings obtained are adjusted for background signal and, as some individuals have a stronger response than others, each array must be normalised before statistical analysis can take place. This enables elevated patient responses compared to the control arrays to be reliably calculated, forming the basis of the investigation into potential biomarkers.

Future clinical applications

One key project currently being undertaken by the Center for Personalized Diagnostics is an investigation into the early diagnosis of active tuberculosis (TB), which currently affects a third of the world's population. The first stage of this study requires DNA to be prepared and a TB protein array to be printed. The clones responsible for encoding the genes for TB are requested from the DNASU repository, then used to grow bacterial cultures from which the DNA is prepared. To establish whether a patient is producing antibodies that are specific to proteins on the TB array, patient samples and controls are loaded on to the HS 4800 Pro. Any unbound antibody is then washed away and a fluorescently-labelled secondary antibody, which recognises human antibodies, is added. The fluorescence signal is measured using the PowerScanner to determine which proteins have responded to the patient serum. Once the antibodies made very frequently by patients with active TB have been identified, the most promising antigens



Discovery of Basal-like breast cancer serum autoantibody biomarkers using human NAPPA arrays. 768 expression clones encoding human proteins fused to C-terminal GST tags were printed along with a polyclonal anti-GST antibody in duplicate spots on the array surface. DNA capture was confirmed by PicoGreen staining (left). The protein displayed *in situ* was assessed by a monoclonal anti-GST antibody (middle). After probing the NAPPA protein arrays with sera from either healthy individual (bottom right), or basal-like breast cancer patients (top right), auto-antibody bound to the target was detected using fluorescein conjugated anti-human IgG antibody. Circled spots indicate positive auto-antibody responses in patient's serum

are selected for further studies designed to reveal their effectiveness as a predictor for the disease. These antigens are then re-evaluated in a different patient set to discover whether they continue to act as efficient markers for TB.

Another ongoing project is looking into potential biomarkers for triple negative breast cancer using human protein arrays. 'Triple negative' refers to women whose tumours stain negative by immunohistochemistry for estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2). This particular form of breast cancer is a very challenging one; tumors tend to be more aggressive than other forms of the disease, and many of the most efficacious medications (eg tamoxifen and trastuzumab), which target ER and HER2, are not effective. In addition, these tumours are often discovered as interval cancers, and are not often detected by mammography. As a result, the overall prognosis is very poor for these patients and finding blood-based biomarkers for early detection would be of significant benefit.

Conclusions

At the Virginia G. Piper Center for Personalized Diagnostics, automation is the key to establishing reliable, high throughput protocols. Manual processes, no matter how experienced and well trained the scientist, are very time-consuming and will always be susceptible to human error. It is crucial to minimise the potential for variation during clinical studies, as any variation makes it much more difficult to differentiate between those patients with a disease and those without. If the technology is not reproducible, then the results may not be accurate, and incorrect conclusions may be drawn. The power of automated systems is that, once optimised, they are extremely reproducible and provide consistent results, day in, day out. DDW

Dr Joshua LaBaer *is currently Director of the Virginia G. Piper Centre for Personalised Diagnostics and Center Director, Chair of Personalised Medicine, Professor of Chemistry and Biochemistry at the Biodesign Institute, Arizona State University. He is one of America's foremost investigators in personalised medicine involving the discovery and validation of biomarkers which can provide early warning for those at risk of major illnesses, including cancer and diabetes. He earned his medical degree and a doctorate in biochemistry and biophysics from the University of California, San Francisco, completed both his medical internship and residency at the Brigham and Women's Hospital in Boston, and undertook a clinical fellowship in oncology at the Dana-Farber Cancer Institute, also in Boston.*