THE POST GENOMICS ERA
what does it all mean and how can we translate it?

Information may no longer be a bottleneck to understanding and tackling complex genetic diseases such as cancer. For those who cut their research teeth not so very long ago, before PCR was thought of by Kary Mullis, this is truly an amazing position to be in. To think that effective management of cancer is potentially just a time and money issue equation, where new disease causation data needs to be converted into novel 'targeted' or 'personalised' therapies, how can we eliminate drug discovery economics as an overriding limiting factor?

In contrast to the process of drug discovery, which has remained much the same for the past 10 years, advances in genome sequencing and research have been exponential during this time. Moreover, we are probably not seeing a plateau in these capabilities yet and even now we have literally hundreds of potential new targets in cancer alone to pursue for new therapies. So what now must be done to speed up the translation of these targets into drugs?

One of the first issues is that many candidates are not directly 'drugable' and thus require additional searches downstream for drugable proteins. The second issue is that many putative new cancer targets are low in frequency, making them challenging to pick up by biotech and pharma until new information on their role in normal and disease biology is defined; especially if they can be placed in higher frequency cancer pathways. This is one area, therefore, that academia can have a major impact on by continuing to de-orphan these targets and perhaps even perform end-to-end drug discovery. Finally, new technologies are sorely needed to reduce the large amounts of time, money and attrition associated with all stages of drug discovery and development, in particular those associated with performing large and unselected clinical trials.

It is the thesis of this editorial that academia and industry must heavily invest in a post-genomics world; firstly to understand 'what it all means', ie decipher which genetic variations are consequential and which are merely random noise; secondly, design early and accurate diagnostic tests to enable potentially remedial therapies to be given before cancers become incurable; and thirdly, to have accurate and predictive models of human cancer, so that novel treatments can be developed quicker and directed to the patients most likely to respond. Such focused drug development will not only be faster and more likely to succeed, but also be more ethical to the patient, who has a better chance of entering into a worthwhile trial. To usher in this new era of high-throughput functional genomics,

By Dr Chris Torrance

Genomics

dsDNA and rAAV Homologous Recombination Vectors

ITR | LH arm | loxP | Marker | loxP | RH arm | ITR

Mut

WT

Conventional HR vectors: dsDNA
Contemporary AAV HR vectors: ssDNA virus

LH : Left Homology arm
RH : Right Homology arm
Marker : Selection Marker e.g. Neo
Mut : Point mutation being knocked-in
ITR : AAV Inverted Terminal Repeat
loxP : Cre-recombinase sites to remove selection markers after gene editing

Homologous Recombination Efficiency
Measured as neo-resistance correction rate

1.00E-02
1.00E-03
1.00E-04
1.00E-05
1.00E-06

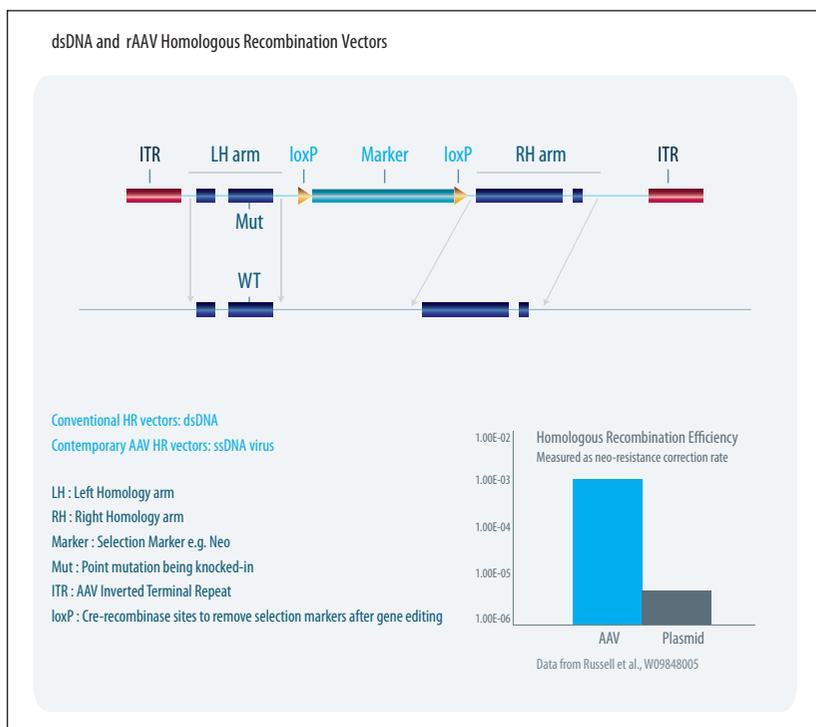AAV   Plasmid

Data from Russell et al., W09848005

Figure 1

predictive disease modelling and ultimately the design of rational clinical trials, scientists will need to be able to alter the DNA-sequence of a human cellular genome in a manner that is now routine and facile in mice and other lower organisms.

Why are we moving towards 'personalised' medicine?

Very few diseases are simple. They are either highly multi-factorial like cancer, and so require many different treatments tailored to the right patient; or they can be caused by a single agent, such as HIV or a bacterial infection, but it mutates over time; and thus requires doctors to keep rapid pace with a moving target. In the case of cancer it is both these things, which makes it such a challenge to manage. In the future, however, we will have the ability to rationally prescribe and adapt the right drug, drug combination, or drug dose, to each patient based on having a detailed understanding of their disease genetics, to far more effectively manage their disease.

This is, in essence, the concept of personalised medicine; a phenomenon that is already happening, but has a long way to go to realise its full potential. The principle issue is that there are currently nowhere near enough drugs in the personalised medicine toolbox to tailor to the right patients. The reality is we have only scratched the surface of 'drugging' the cancer genome and will

fail to do so in the present generation unless some hard decisions are made. Another issue is that an entirely new industry and service needs to be developed to provide early, routine and accurate diagnostic tests to support the development and tailoring of any future novel therapies, which has its own harsh economic models to deal with if performed outside of the established pharma industry. Finally, regulatory and healthcare agencies will need to foster these endeavours and ultimately be convinced of why this isn't going to cost them a lot more money. Wisdom would predict this will be true, given that it will enable us to move away from blanket or over-prescription of expensive new drugs, where we are again only at the tip of the iceberg right now.

The stark facts are that for every drug developed, approximately nine fail and the cost of these failures are ultimately passed on to the consumer. Combined with the approximate 10 years and $1 billion spent per drug to reach Phase III and then fail, this is clearly an unsustainable situation moving forward into a more personalised, or segmented, therapy world. There will be many reasons why drugs fail, but one that in principle can be fixed is to better understand which patients are more or less likely to respond. With the advent of clinical diagnostics and accurate models of human disease, new drugs in development or even already approved treatments, can increasingly be targeted to the 'right' patient populations who possess unambiguous 'biomarker' signatures of response. A clear example of the impact of such predictive models and biomarkers on directing patient therapy will be given in a later section, as will their benefits to accelerating earlier stages of drug discovery. First, a review of the state-of-the-art in genome editing, which will underpin both the generation of genetically-defined human disease models and the growing needs of performing precision functional genomics.

Human genome editing

The routine and accurate editing of human cellular genomes to permit disease modelling, functional genomics and possibly even corrective gene therapy, will represent the next important breakthrough in science. Given the 'hardware' in this case is a live human cell, one will always be restricted to working with cells' natural biology. Historically, altering the sequence of endogenous genes within differentiated human cell types has proven to be orders of magnitude less efficient compared to lower organisms. Only recently has it become technically feasible to perform the

accurate and stable engineering of human genomes without the introduction of potentially confounding or dangerous off-target errors or mutations; and improving the speed of gene-editing still comes at a cost of precision at this time.

Modern gene-editing technologies currently fall into two broad categories, those that rely exclusively on homologous recombination, a natural DNA-repair mechanism, to perform endogenous DNA alterations; and those that stimulate locus specific events as a consequence of introducing double strand DNA breaks. Each approach has its advantages: the latter allows the rapid and efficient 'knock-out' of specific genes, but also introduces unwanted off-target cuts in the genome and is comparatively inefficient at performing subtle 'knock-ins' of disease causing point mutations; and the former, which is currently less efficient at performing gene knock-outs, but does not introduce any off-target events and routinely can perform any genomic alteration, large or small at any target locus. The principle techniques are discussed further:

Linear double-stranded DNA homology vectors: This technique relies solely on homologous recombination (HR) and has been around for more than 10 years to create precision transgenic 'knock-in' and 'knock-out' mice. Vectors are simple stretches of homology to the target locus with a selectable marker in the middle (**Figure 1**). While this approach is inefficient compared to other more recent techniques, because mouse ES cells have very high natural rates of HR, this technique is perfectly adequate and still used commonly today. However, in human or other mammalian somatic cell-types, which have a far lower rate of homologous recombination, this technique is too unwieldy and has been abandoned in favour of more contemporary approaches that stimulate HR in some way.

Zinc-finger nucleases: These are relatively bespoke hybrid vectors that combine an adaptable, sequence specific Zinc-finger DNA-recognition domain, fused to a dimerisation-dependent nuclease, usually Fok1. When two zinc-finger nucleases (ZFNs) co-locate at a bipartite recognition sequence they create a dsDNA break, typically in both alleles, and thus elicit the rapid, permanent and relatively specific (compared to siRNA) deletion of a target gene. Absolute specificity is hard to achieve, and typically many off-target dsDNA breaks are introduced as well as the intended one. For human disease modelling and the creation of highly characterised bioproducer cell-lines, where complete precision is required, this is a significant

Genomics

issue. Subtle gene 'knock-in' or correction events in human cells, while possible when ZFNs are co-delivered with a second donor-homology vector, are less efficient due to the fact that dsDNA breaks are predominantly repaired by error-prone non-homologous end joining pathways.

Meganucleases: These are analogous to ZFNs, but possess greater specificity due to having a much larger DNA-recognition footprint. However, they are not as flexible in their design as ZFNs and the high cost of creating Meganucleases to a target locus of interest means they have been limited to performing high-value projects such as transgenesis in plants.

TALEN-nucleases: This is a new nuclease player in the field. Their advantage is that they are almost completely modular and deterministic in their assembly, allowing the simple and cost-effective design to almost any genome location in theory. Moreover, since modules are extendible into very large DNA-recognition footprints they are possibly much more specific. For these reasons and public domain construction algorithms, they look set to compete with established nuclease methods. Their ability to perform subtle knock-ins has not yet been tested, but they will logically be subject to the same limitations as other dsDNA break-inducing methodologies. Consequently, their prime advantage will lie in performing efficient gene knock-outs and transgene insertions in 'safe harbour' loci in a broad range of animal and plant systems.

rAAV: Recombinant adeno-associated viruses are non-pathogenic single stranded DNA-viruses and they have a unique and powerful capability to convert direct 'HR-only' vectors into a system that is ~1,000-fold greater efficiency at performing all forms of gene-editing compared to older-style dsDNA homology vectors (**Figure 1**). It is not entirely known why they are so efficient, but it appears that a distinct form of DNA-repair operates to faithfully recombine ss-DNA species into target genomic loci, which is independent of many of the factors typically seen to be important for dsDNA-mediated HR, eg Rad51 and Rad54b. rAAV was first used in human gene-therapy due to its efficient mode of delivery and its ability to perform precise targeted gene corrections, and then more widely in the field of *in vitro* genome editing and disease modelling. While not as efficient as nuclease methods at performing bi-allelic gene knock-outs currently, this may improve with further research into ssDNA-mediated HR mechanism, and in the meantime allows users complete

confidence that when you have successfully targeted a gene, it does not also come with other confounding off-target events. For this reason, it is becoming the method of choice for the definitive dissection of gene-function, as well as performing precision disease modelling. It is also likely to be preferred for creating enhanced bioproduction cell-lines, especially now the CHO genome has been sequenced, wherein dsDNA break methods can cause long-term genome stability issues and are even prone to integrating trace levels of non-human DNA present in cell-culture media, via highly efficient non-homologous recombination events.

Disease models in early stage drug discovery

Coming now to applications of gene targeting and genetically-defined disease models, the first thing any drug developer has to do is to choose a specific target, preferably a good one given how long and expensive it is. However, prior to recent large-scale, consortium-based, cancer genome profiling efforts, choosing a 'good' cancer target was very hard to qualify or quantitate in some way. All too often a 'validated' target was simply one that another company was working on, but not too many as this would be overly competitive. True disease validation was effectively minimal, with elevated expression or perceived pathway relevance typically being the best marker of cancer relevance, which is often misleading.

DNA-alterations (mutations and/or copy number gains or losses) in contrast, are unambiguous events and, if present in high enough frequency in a cancer type, are, more often than not, key drivers of the disease. Now we have a plethora of such information, several new issues actually arise: Firstly, most 'cancer genes' are tumour suppressors, which are either inactivated or completely lost in tumours, and thus are unrealistic targets for small-molecules that are typically easier to design as inhibitors of protein function. Secondly, many gain-of-function 'oncogenes' are also hard to drug, such as non-enzymatic transcription factors. Thirdly and of practical importance, most newly identified candidate cancer genes, including the drugable ones, have very low tumour mutation frequencies (often <5%), which could simply represent passenger 'noise' in genetically unstable tumours. Due to all these factors, there is currently a heavy operational bias towards drugging signalling pathway kinases, which if directly implicated in disease progression can be highly effective, but are also subject to rapid onset of resistance via compensatory signalling pathways or events. This may also be exacerbated

by the typical cytostatic nature seen for single agent pathway-targeted drugs.

All this represents a major challenge for the next wave of targeted drug discovery. In the conventional arena, we need to find more functionally characterised targets, ie which of the many mutant genes are drivers vs passengers, and then determine which of these stack-up into more frequent mutated pathways so that becomes viable for drug developers. Here the ability to alter gene function positively and negatively will enable the dissection of their normal vs disease biology. Moreover, the pointed search for key downstream effectors of undrugable genes will be significantly aided by simple 'isogenic' model systems, which will enable high-throughput expression profiling and siRNA screens to be performed. Such isogenic 'X-MAN' (gene-X; Mutant And Normal) cell-lines are being created by Horizon Discovery using rAAV, which now comprises more than 300 different disease models and will grow to thousands in the next two years based on internal production and the establishment of 50 academic centres of excellence.

As well as feeding the conventional drug discovery process, X-MAN disease models can also be used in 'chemical genetic' screens to identify new drugable targets that impact tumour-specific defects, especially those that are undrugable tumour suppressors. Moreover, if the compound libraries are chosen wisely, ie *in vivo* validated compounds, perhaps even isolate drug candidates directly. Many examples of such 'synthetically lethality' screens are now being described, the most notable and advanced of which is the toxic interaction of inhibiting PARP activity in BRCA2-null cancers. This observation first gleaned from an isogenic BRCA-2 null mES cell system, is now being used to successfully treat BRCA-null breast cancer patients.

Disease models in later stages of drug discovery

In conventional drug discovery, once sufficient 'on-target' chemistry has been obtained, sufficient on-target biology is next addressed. The question then arises, however, what is sufficient on-target biology? One could argue that this should be the selective death of tumour cells carrying a specific cancer causing mutations given the direction we like to take as a field. At this stage, X-MAN disease models can be assayed for target patient-specific activity *in vitro* or in animal models, which often reveal phenotypes and drug effects that were unexpected (**Figure 2**). Moreover, if a target patient population is unknown, a wide range of patient-genotypes can be rapidly profiled prospectively *in vitro* for those
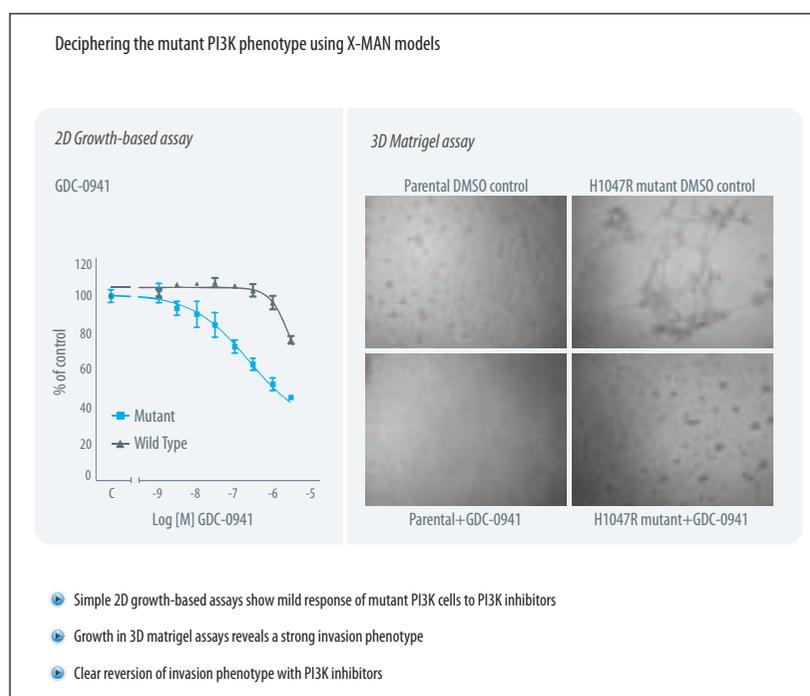
that are likely to respond the best. Together, these profiling tools will allow the design of smaller clinical trials centred on the patients most likely to respond, and if a drug fails here, it fails quickly rather than continuing for many years in larger trials, probably to the same result. The massive amounts of money saved can then be used to bring a wider set of next-generation targets and drugs into the same efficient process, allowing the best chance for single agents to show anticancer activity, and thus build a diverse enough drug portfolio to be mixed and matched in the right ways. This is ultimately where we will need to be to significantly impact cancer. Moreover, this strategy will form a sustainable biotechnology model moving forward.

Supporting this concept is AstraZeneca's Iressa (gefitinib), the first clear example of how knowing ahead of time which patients would respond (in this case mutant EGFR lung cancer patients) could have saved many development years and dollars. We also know now that in addition to primary 'sensitivity biomarkers', one also needs to define other pre-existing, or treatment acquired alterations, that cause drug resistance. Here genome editing techniques can be used to create isogenic models that harbour defined combinations of disease causing and/or candidate drug resistance genes and then used to prospectively profile for potential resistance mechanisms.

As a landmark example of this approach, Horizon's co-founder Alberto Bardelli and research

Deciphering the mutant PI3K phenotype using X-MAN models

2D Growth-based assay
GDC-0941

% of control
120
100
80
60
40
20
0
C -9 -8 -7 -6 -5
Log [M] GDC-0941

Mutant
Wild Type

3D Matrigel assay

Parental DMSO control   H1047R mutant DMSO control

Parental+GDC-0941   H1047R mutant+GDC-0941

● Simple 2D growth-based assays show mild response of mutant PI3K cells to PI3K inhibitors

● Growth in 3D matrigel assays reveals a strong invasion phenotype

● Clear reversion of invasion phenotype with PI3K inhibitors

Figure 2

Genomics

X-MAN in vivo

G12V

G13D

Tumor Volume (mm³)

Days

Tumor Volume (mm³)

Days

■ Control ■ Cetuximab

X-MAN predictions confirmed in patients

p.G13D mutation

Other K-Ras mutation

% Alive

% Alive

Months since randomisation

Months since randomisation

■ Cetuximab monotherapy ■ No cetuximab therapy

Adapted from De Roock W et al, JAMA Oct. 2010

◉ Currently all K-Ras mutant patients excluded from EGFR-targeted therapy

◉ X-MAN models harbouring specific K-Ras variants showed not all K-Ras variants impart resistance

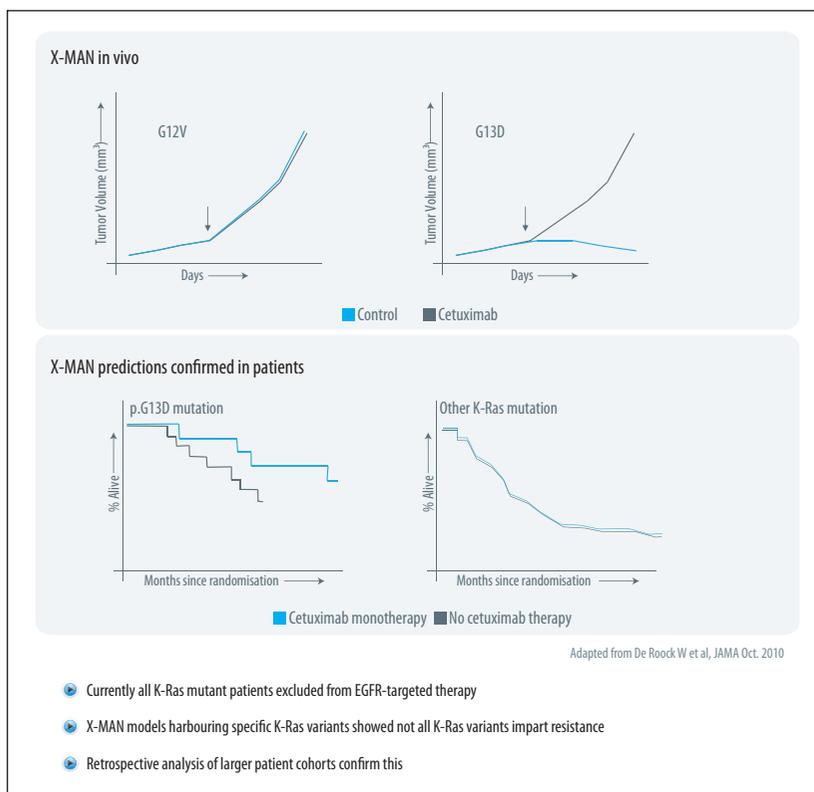◉ Retrospective analysis of larger patient cohorts confirm this

Figure 3

colleagues used X-MAN disease models harbouring different K-Ras variants (G12V and G13D) to test whether they are both equally resistant to Cetuximab therapy. *In vitro* proliferation assays and xenografted tumours, both determined unambiguously that G13D and WT K-Ras containing cells were highly responsive to Cetuximab, whereas G12V containing cells were not (**Figure 3**). Subsequent sequence analysis of tumour samples taken from actual patients treated with Cetuximab confirmed this picture; and thus with follow-up prospective clinical trials, these data may lead to changes in the rules for prescribing EGFR targeted therapies in colon cancer, where currently patients carrying any K-Ras mutation are excluded from therapy. Isogenic models will also form the ideal tool to rationally find rational drug combinations to reverse resistance.

One final area where genetically-defined disease models will help the both later stages of drug development and the prescription of approved drugs is in the development of reliable diagnostic kits and platforms. Armed with isogenic gDNA, which can be mixed in fixed proportions to mimic the heterogenous nature of tumour samples, these will form the perfect precision standards to determine the performance envelope of emerging diagnostics and as patient-relevant controls for CLIA labs running them.

Concluding remarks

Personalised therapies and diagnostics represent the logical direction for providing effective cancer therapy in the future and for the most part, the pharma industry has embraced these ideals; especially since healthcare reimbursers are moving to a system where they pay handsomely for effective medicines and not at all for marginal ones. There is still a limit to how infrequent a target industry will currently tackle, so it is imperative for new functional genomics technologies and genetically-defined disease models to connect the dots. Given the complexity of future clinical trials, pharmaceutical companies will probably continue the trend of focusing on late stage development and divesting early-stage research. Here, predictive and patient-relevant disease models will enable the triage of patients into more focused trials with greater certainty of positive outcome and drug approval; and at earlier stages of drug discovery, will enable academia and biotechnology companies to increasingly feed validated targets and drug candidates into pharma in a sustainable way. Finally, as a society, it would pay for us to explore ways to incentivise academia and industry to perform research on the wide diversity of rare cancer targets now presented to us, such as targeted translational funding for academia, early pre-approval from successful focused clinical trials and extended patent lifetimes for industry on any new 'first-in-class' drugs. These measures would stimulate in an entrepreneurial way the breadth of research and drugs we need; and once available, will likely have larger patient populations than anticipated once they are studied and combined in the right ways.   DDW

Dr Chris Torrance has a bachelor's degree in Biomedical Technology from Sheffield Polytechnic; a PhD in Biochemistry from East Carolina University (U.S.A) and completed Post-Doctoral training with Professor Bert Vogelstein at the Johns Hopkins University (U.S.A), where he pioneered the use of X-MAN cancer models in high-throughput screening and drug discovery. Prior to founding Horizon Discovery, Dr Torrance was Head of Oncology and Biology at the UK Biotechnology company Vernalis PLC (LSE: VER), where he was also responsible for progressing several novel kinase oncology programmes.