R&D SERIES EU

Drug Discovery | Discovery Chemistry & Drug Design
Bispecifics In Discovery & Development | Neuroscience In Discovery & Development

Pre-Event Newsletter May 2019

Thalidomide Revisited

A CETSA MS mode of action study of the Thalidomide analogue Pomalidomide

Applying AI to Drug Discovery

Key indicators that "AI"/ML can be a driver of large-scale disruption

National Phenotypic Screening Centre

How advancing phenotypic screening can tackle unmet medical needs

MARITIM PROARTE HOTEL
11 - 12 JUNE 2019 | BERLIN, GERMANY

OXFORD GLOBAL CONFERENCES

Contents

This is an Interactive Newsletter.
You can click on elements such as website links or the contents below.

Meet the Team

Cerlin Roberts
Director

Chris Davies
General Manager and Portfolio Director

Tom Cashman
Operations & Events Executive

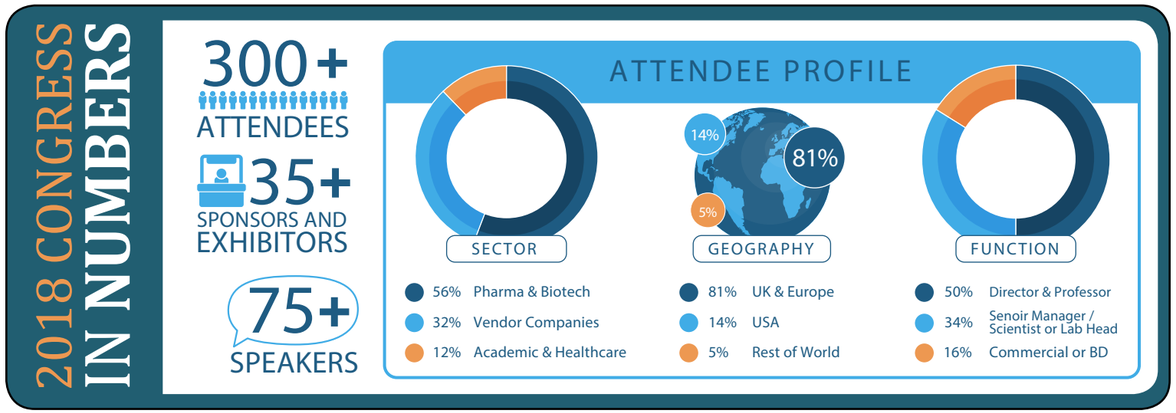
Guillaume Alonso
Marketing & CRM Manager

Alex Broad
Sponsorship Sales Executive

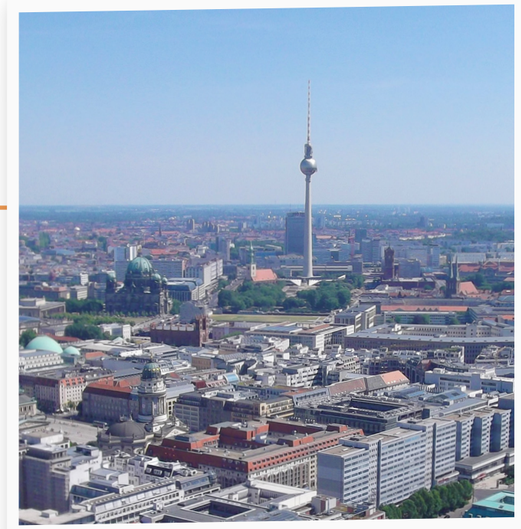
Henry Whitehouse
Delegate Sales Manager

Matei Marin - Assistant Conference Producer
Emma Richardson - Production and Marketing Assistant

We look forward to meeting you in June!



WELCOME TO OXFORD GLOBAL'S R&D SERIES MAY 2019 NEWSLETTER!



The R&D Series Europe event returns to Berlin this June in something of a milestone year, with the Drug Discovery Summit programme celebrating its 20th Anniversary!

The past 20 years have seen huge changes in the Pharma & Biotech sector as a whole, and in approaches to drug discovery. This has been reflected in the programme itself, where we're proud to have worked with leading representatives and companies each year since 1999 to explore the topical issues and developments facing the industry.

In 2019, the Drug Discovery programme will again be joined by a dedicated Discovery Chemistry & Drug Design programme (now in its 7th year) and two inaugural programmes; Neuroscience in Discovery & Development and Bispecifics in Discovery & Development. We'll be welcoming over 350 attendees to cover the latest developments in;

20th Annual Drug Discovery Summit: Protein degradation, AI/ML in drug discovery and updates in screening and modelling including CRISPR and organ-on-a-chip techniques.

7th Annual Discovery Chemistry & Drug Design Congress: DNA encoded libraries, protein-protein interactions, AI/ML in drug design and medicinal chemistry for oncology.

Bispecifics in Discovery & Development Congress: Developability, immunogenicity, safety and CMC considerations to bring bispecifics into the clinic successfully.

Neuroscience in Discovery & Development Congress: Tools & technologies enabling novel treatment developments, critical biomarker & translational work in preclinical development.

The Series continues to offer an increasingly interactive environment with more panel discussions, roundtables and workshops taking place in 2019 than any previous year. We'll also again include ample opportunity for informal interactions, with several networking sessions throughout the event. All attendees are also invited to join us at our Drinks & Canapes reception, which closes the first day of the event.

The central ethos behind the event is to provide an environment where senior representatives from Global Pharma, Biotech and Academia can share knowledge and build networks that will help to sustain and develop the industry, and we look forward to seeing this development continue over the next 20 years!

Please read on for some pre-event insight from leading speakers and sponsors of this year's event and we look forward to welcoming you to Berlin in June!

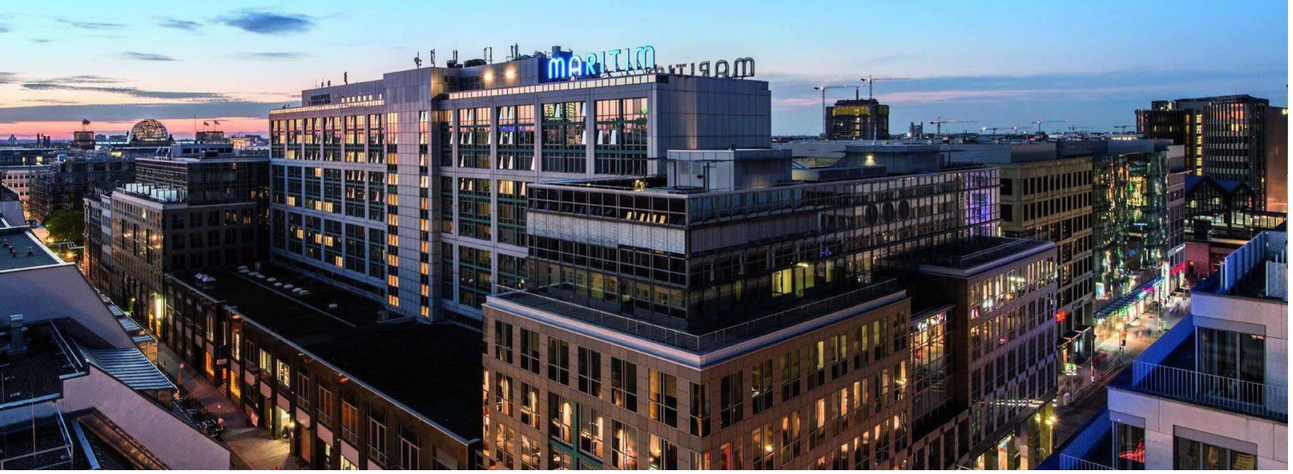
- Chris Davies, Portfolio Director

R&D SERIES EU

20TH ANNUAL **DRUG DISCOVERY** SUMMIT
7TH ANNUAL **DISCOVERY CHEMISTRY & DRUG DESIGN** CONGRESS
BISPECIFIC DRUG DEVELOPMENT CONGRESS
NEUROSCIENCE DRUG DEVELOPMENT CONGRESS

MARITIM PROARTE HOTEL
BERLIN, GERMANY

11 - 12 JUNE 2019



WHO IS ATTENDING?

For the full attendee list please contact
marketing@oxfordglobal.co.uk

- **350+ senior level delegates** representing leading biotech companies, global pharmaceutical organisations and internationally renowned academic institutions.
- **Directors, VPs, CEOs and Heads** of Drug Discovery, Drug Development, Drug Design and Medicinal Chemistry.
- **Highly esteemed members** of academic and government institutions.

These companies and many more:

AstraZeneca, UCB, Lilly, Bayer, Novartis, Merck, Sanofi Genzyme, Servier, LEO, BERG, Takeda, AbbVie, Amgen

Sponsors 2019

GOLD

evotec
PELAGO BIOSCIENCE

SILVER

dotmatics knowledge solutions
Malvern Panalytical a spectris company
MITRA BIOTECH
Syngene

WORKSHOP

Chemical Computing Group

BRONZE

CELLECTA Discovery is *Yours*.
CYCLICA
Onco design BIOTECHNOLOGY

NETWORK AND PROGRAMME

admescope
AnalytiCon discovery The Natural Product Company
Biomax INFORMATICS
ChemAxon
CDD VAULT Complexity Simplified
CONCEPT LIFE SCIENCES DELIVERING SCIENCE
creoptix sensors
GalChimia
GLYCOTOPE
LCC Liverpool ChiroChem
Luminex complexity simplified.
mcule MORE THAN MOLECULES
MercachemSyncom
MIMETAS the organ-on-a-chip company
SCHRÖDINGER
Selvita
SYGNATURE DISCOVERY
药明康德 WuXi AppTec
ZeClinics Powering discovery with Zebrafish

Thalidomide revisited

A CETSA MS mode of action study of the Thalidomide analogue Pomalidomide

Thalidomide was first introduced in Germany under the trade name Contergan in the late 1950s as a "completely safe" and "non-toxic" sedative and a remedy for morning sickness suffered by pregnant women. Despite the lack of knowledge about the mechanism of action of the drug, it was aggressively marketed and soon sold around the world. A few years later it became apparent that Thalidomide had caused limb malformations and other severe birth defects in thousands of children when the drug had been taken by their pregnant mothers during a sensitive time of embryonic development. Although the use of Thalidomide was subsequently banned in most countries at that time, the drug proved to be a useful treatment for leprosy and later multiple myeloma.

It was, however, not until 2010 that the primary molecular target of Thalidomide was discovered by high-performance affinity bead purification of Thalidomide-binding proteins from crude human HeLa cell extracts. Two proteins, CRBN and DDB1, which form part of an E3 ubiquitin ligase complex, were isolated, suggesting CRBN as a direct interaction partner of Thalidomide (1). Research in the field took-off after this discovery and different hypotheses arose around the mechanism of action of Thalidomide and its more potent analogues Lenalidomide and Pomalidomide, drugs commonly referred to as IMiDs.

In 2018 Donovan et al. and shortly thereafter Matyskiela et al. discovered that IMiD-mediated ubiquitination and subsequent degradation of the transcriptional factor SALL4 was responsible for the developmental defects and limb malformations (2,3). Along with SALL4, Donovan et al. published a range of additional substrates of the IMiD-bound E3 ubiquitin ligase complex.

> "One of the best target ID technologies available, MS CETSA"
> Derek Lowe, KOL in Drug Discovery

Experiment

Human iPS cells were treated with Pomalidomide or control and harvested at different timepoints between 15 minutes and 8 hours. In the herein applied compressed 2D CETSA MS format, individual melt curves are pooled over all temperatures which effectively compares treatment and control at multiple concentrations or timepoints. The changes over time then resolves both a time-dependence of drug-protein binding and provides a quantitative concentration response curve

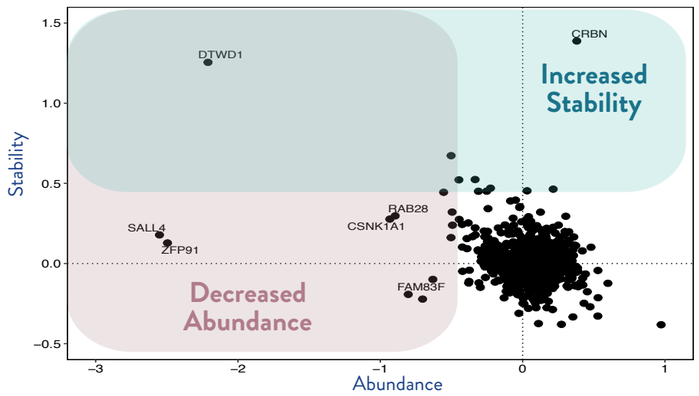


Figure: Using the CETSA MS methodology, CRBN is identified as a target of Pomalidomide, accompanied by a time-dependent degradation or decrease in protein abundance of several previously published and novel protein targets of the E3 ubiquitin ligase complex. Highlighted in the figure are SALL4, ZFP91, RAB28, CSNK1A1, FAM83F and DTWD1 which had been discovered previously as substrates of the IMiD-activated E3 ubiquitin ligase.

PELAGO BIOSCIENCE BANVAKTSVÄGEN 20 SE-171 48 SOLNA INFO@PELAGOBIO.COM WWW.PELAGOBIO.COM

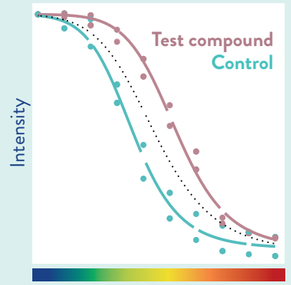
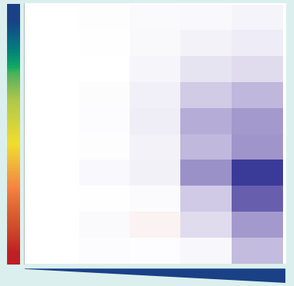
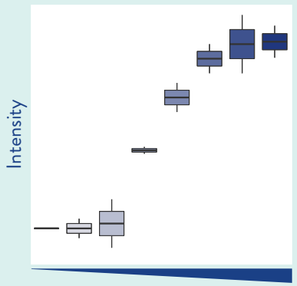
Melt curve CETSA MS	2D CETSA MS	Compressed CETSA MS
Ten-temperature melt curves measured for thousands of proteins in the presence or absence of the test compound. This format is the best choice if treatment with only a single concentration is possible or sufficient (e.g. *in vivo* studies of dosed animals or patient material).	**Five-point isothermal concentration- response curves accessed at 10-12 different temperatures** (usually presented as heat maps). This heat map clustering allows grouping of protein hits. Most detailed characterization of the compound effect in one or two sample matrices.	**Eight-point concentration-response curves** where each concentration is comprised of a range of pooled temperatures. Reflects compound-induced changes in thermal stability of the proteins. This is the best choice for large scale projects, e.g. multiple compounds, samples matrices, time points, etc.
		

Table: introducing the different CETSA MS formats

Technology

The combination of the patented Cellular Thermal Shift Assay (CETSA®) with quantitative proteomics (CETSA MS) offers unbiased proteome-wide thermal profiling in label-free and disease-relevant systems. The entire proteome is monitored for changes in protein thermal stability during drug treatment. The CETSA MS read-out is then the classical melt curves, concentration and temperature heat maps or a compressed format to quantify affinity and abundance (see table). In this case study we applied our compressed 2D CETSA MS format to find cellular targets of Pomalidomide and explore the downstream pathway effects of the drug treatment in living cells.

Pelago Bioscience is a specialist contract research organization (CRO) founded in 2013 with the aim to develop the patented Cellular Thermal Shift Assay (CETSA®) for maximum utility and customer value in drug discovery and diagnostics. The company delivers *in situ* target engagement studies to accelerate preclinical and clinical drug discovery and diagnostics development. Using CETSA data and applications, drug discovery R&D companies are able to make better and more informed decisions at earlier stages in their projects. Our scientific team are the experts in both targeted (Classics & HT) and unbiased proteomics (Mass Spectrometry) CETSA applications. For more information please visit our webpage www.pelagobio.com

Conclusion

Conventional proteomic approaches simply monitor protein abundance levels, either between treatment groups or at a phenotypic endpoint. CETSA MS measures the direct functional consequences of a drug on both its intended target and the immediate signaling cascade perturbed by the addition of your molecule. This enables us to discover markers of activity that that would otherwise not be identified using traditional methods. The study of Pomalidomide shows the power of the CETSA technology applied on a real case, identifying the detailed mode of action that took more than 50 years to resolve - all in a single experiment.

References:

1. Ito et al. Science 2010
2. Donovan et al. Elife 2018
3. Matyskiela et al. Nature chem biol 2018

GENERATING INNOVATION AND IMPACT FROM PHENOTYPIC SCREENING

DEN BARRAULT

Executive Director, National Phenotypic Screening Centre

Features of phenotypic screening

A phenotype is one or more observable features that report changes in the genotype, epigenotype or environmental response of a single cell, a group of cells, organ or of a whole organism. Phenotypic changes are the causes of disease, whether this is a cancer cell undergoing uncontrolled cell division, a motor neuron which fails to connect to muscle tissue in motor neuron disease, or the complex defects seen in the brain of a patient with schizophrenia or bipolar disease. Screening for agents that alter a phenotype is an effective route to finding new therapies, and has been found to be more efficient at finding "first-in-class" drugs that the more traditional target-based approaches.

Phenotypic screening takes a more holistic and empirical approach to discovery than the mechanistic approaches used in target-based screening. We develop and exploit validated biological assays that better reflect the pathophysiology of human disease to find actives that could be starting points for new drugs.

NPSC offers an ideal opportunity for biologists to validate their phenotypic assays, translate their biological innovations, and create impact from their research.

Our phenotypic screening facilities are state-of-the-art, and on-par with what can be found in the pharmaceutical industry. We benefit from access to a range of different high content readers, which allow us to tackle most assay configurations and readouts, as well as a wide range of robots and liquid handlers that fully automate complex workflows for high throughput screening performances.

University of Dundee
THE UNIVERSITY of EDINBURGH
UNIVERSITY OF OXFORD
NATIONAL PHENOTYPIC SCREENING CENTRE

NPSC is a world-class facility for phenotypic screening run by an interdisciplinary team of experts from the Universities of Dundee, Edinburgh and Oxford.

NPSC screens predictive phenotypic assays, translating world-class biological and clinical research to tackle grand challenges in the life sciences.

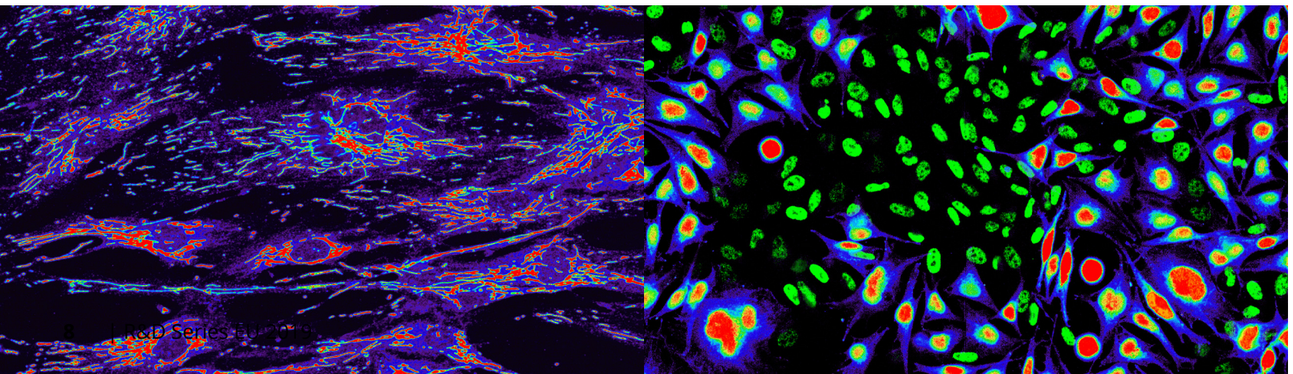
We bridge the gap between academics, clinicians and scientists in Pharma and SMEs, and catalyze collaborations that advance the use of complex biology to drive innovation in drug discovery and the life sciences

Our team exploits our facilities' advanced equipment and technologies to establish the most effective screening cascade for your cellular assay, facilitating the generation of efficient, high quality data that can be validated and exploited in downstream applications. Our approach will deliver robust data, allowing evidence-based results interpretation that meet with the programme objectives. By combining tried and tested hit identification protocols with available orthogonal and selectivity assays, we can ensure that hits are validated and moved on into the lead generation process.

Why use phenotypic screening?

Phenotypic screening searches for bioactive agents, such as small molecules or antibodies that reverse a disease phenotype and can be a shorter route to clinical efficacy. There is a new interest from industry in phenotypic screening since the current pharma research model is under stress:

- Productivity is low, so cost to market is increasing
- Clinical attrition rates are high due to lack of efficacy and poor safety.



By employing state-of-the-art screening platforms we have the capacity to quantitatively interrogate biology at multiple levels, and develop bioassays at an industrial standard and scale.

Our core philosophy is to foster scientific creativity whilst delivering high quality data and validated hit matter to collaborators and partners. We are passionate about building an interactive global network of leading scientists from multiple disciplines in medicine, academic bioscience and industry to advance the science of phenotypic screening and tackle unmet medical needs.

NPSC carries out assay development and high throughput multi-parametric phenotypic screening to industry standards. Significant capacity allows us to tackle a range of screen sizes and types.

- High content imaging platforms generate rich data – covering everything from rapid acquisition kinetic imaging, 2D deconvolution, 3D confocal imaging, and high throughput flow cytometry. These allow the capture of multiple parameters in parallel, delivering fast, information rich, quantitative results.
- Acoustic liquid handlers deliver fast and accurate low volume dispensing into 96 to 1536-well assay plates for primary screening and dose-response.
- Three High Res Biosolutions robotic systems are driven by sophisticated scheduling software allowing multiple interlaced assays. Key readers are mounted on plug-and-play carts allowing workflow flexibility.
- NPSC has a 500,000 capacity compound store that is environmentally-controlled, and benefits from immediate access and on the fly cherry-picking.
- NPSC holds several high-quality screening sets that can be accessed for client screening campaigns. These include a large diversity set, as well as several smaller annotated sets, which are useful for determining mode of action and for informing target deconvolution strategies.
- NPSC can accommodate large private collections on site for screening campaigns.
- Two systems are housed in Biological Safety Level II enclosures, live cells are protected and user safety for Cat II organisms ensured.

IT infrastructure is enterprise-level and ensures data integrity and confidentiality.

NPSC provides the perfect environment to exploit phenomics:

- Excellent academic and clinical networks
- High throughput industry platforms
- Fast sophisticated imaging capabilities
- Multiparametric phenotypic profiling
- Rapid advances in biological model systems
- hiPSCs, CRISPR/Cas9, 3D and organoids
- Enterprise image informatics and machine learning
- Smart screening and new annotated chemical libraries
- Powerful chemoinformatics tools
- Advanced modelling and statistical analysis
- Multiple routes to pathway and target identification

Hear Den's presentation, "*High Content Phenotypic Screening Increases The Translational Impact Of Academic Biological Research*", on Day One in the stream *Phenotypic Screening In Drug Discovery And High Content Screening Tools.*

HIGH THROUGHPUT SELECTION OF PEPTIDES WITH BIOLOGICAL FUNCTION

ALEX BATCHELOR

Alex Batchelor,
Chief Executive Officer,
Orbit Discovery

Alex has more than 20 years' experience of sales, marketing and business development in the life science research and drug discovery markets in Europe, US and Asia.

He is an experienced bridge between technical and commercial groups, having started his career in research for the UK Ministry of Defence before moving into industry where he has managed staged R&D processes, implemented IP strategies and launched many successful products and services at companies including Lonza and Genalyte.

Alex holds a BA in Pure and Applied Biology from Brasenose College, Oxford.

ORBIT DISCOVERY

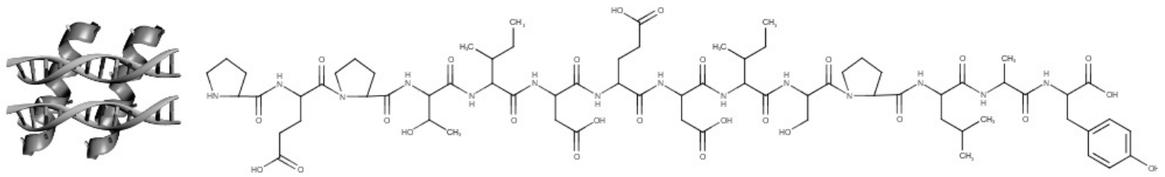
The need for identification of functional hits to cell surface targets

There is a large unmet need for novel therapeutics to address an array of diseases. The costs of drug discovery are large, and the process is inefficient at best. Initial drug discovery efforts fall into two categories: rational design, where there is a good understanding of target structure and knowledge of a ligand, and screening, which involves binding-based approaches with libraries of molecules to identify binders to the target of interest. These approaches are inadequate as they both require massive optimisation after hit identification, and while they may bind to the target of interest, the hits may be non-functional.

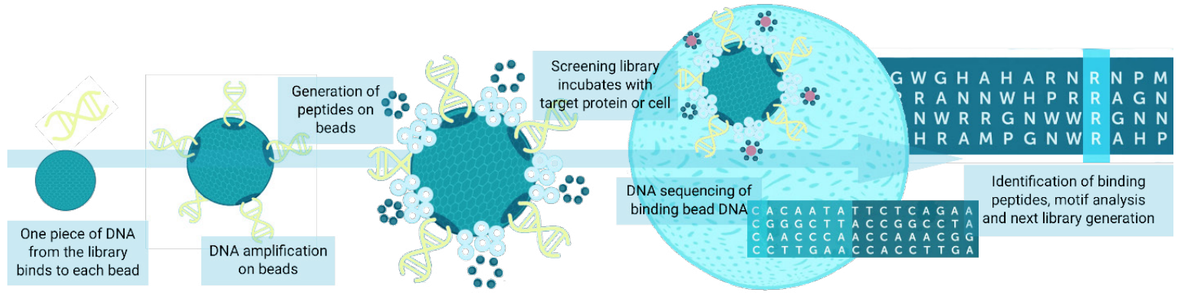
Who are Orbit Discovery?

Orbit Discovery, based in Oxford UK, was formed in 2015 by Alex Batchelor, and Professors Graham Ogg and Terrance Rabbitts, both University of Oxford. Orbit has built a refined discovery platform and pipeline for the identification of functional hits directly from the initial screening campaign. The platform is best utilised for the identification of ligands that interact directly with cellular targets in their natural context with a specific focus on the identification of molecules that agonize GPCRs, T-cell receptors, and other cell surface targets. This is achieved by the construction of cell lines expressing a target of interest coupled to a reporter system whereby stimulation of the target induces a measurable response within the cell, and these cells are subsequently sorted away from non-activated cells to identify the ligand.

Our discovery process begins with a bead-based, DNA encoded, peptide display platform. A single DNA construct is tethered to a bead, *in vitro* transcription and translation synthesises the peptide it encodes which once produced is also tethered to the bead thereby linking genotype to phenotype with each bead displaying a unique peptide. The valency of the display is tuneable with the ability to display hundreds to tens of thousands of copies of the same peptide on each bead. Our libraries can capture the full sequence diversity of a given size of peptide (i.e. $>10^{11}$ for a 9-mer).



THE ORBIT DISCOVERY PROCESS

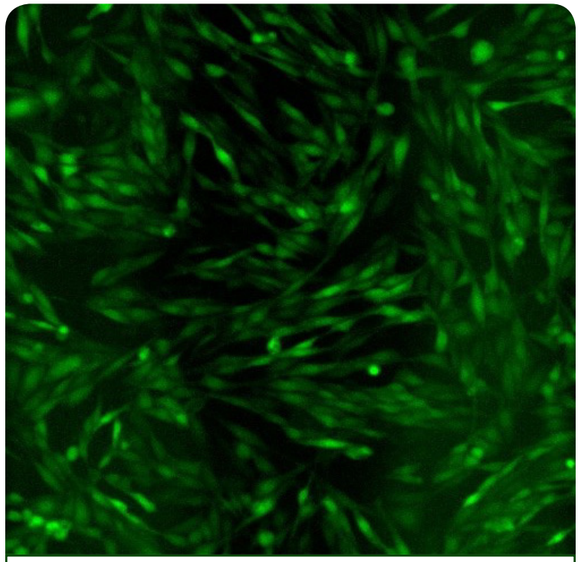


T-cell epitope identification

The ORBIT platform was originally developed for the determination of peptide fragments which our immune system recognises via their interaction with T-Cell receptors. By presenting a diverse library of peptide fragments on the ORBIT beads via an MHC like display scaffold we can explore and identify neo-antigens for vaccine discovery and additionally identify novel T-cell epitopes to allow for the purification and propagation of endogenous T-cells without genetic engineering.

Business Model

The ORBIT platform facilitates novel approaches to drug discovery and development to address unmet need. Combining our platform with other pharmaceutical companies' discovery programs allows us to develop unique therapeutics with unprecedented efficacy. Our business model is simple – collaborative programs with industry and academia and establishment of internal discovery programs which we will out-license in the future. We aim to balance the two efforts through the efficiency that our technology enables by identifying well characterised, selective hits for existing and novel drug targets ■



Pictured: Cells expressing a cell surface receptor coupled to a reporter system were challenged with ORBIT beads displaying a control peptide. The peptide activated the receptor and in turn the reporter system resulting in cellular fluorescence.

Alex will be expanding on this topic at our 20th Drug Discovery Summit. Hear his presentation on Day One in the stream '*Drug Discovery Innovation And Strategies*'.

APPLYING "ARTIFICIAL INTELLIGENCE" TO DRUG DISCOVERY

YOLANDA CHONG

SVP of Disease Biology, Recursion

RECURSION pharmaceuticals

Recursion is a clinical-stage biotechnology company combining experimental biology and automation with artificial intelligence in a massively parallel system to efficiently discover potential drugs for diverse indications, including genetic disease, inflammation, immunology, and infectious disease. Recursion applies causative perturbations to human cells to generate disease models and associated biological image data. Recursion's rich, relatable database of more than a petabyte of biological images generated in-house on the company's robotics platform enables advanced machine learning approaches to reveal drug candidates, mechanisms of action, and potential toxicity, with the eventual goal of decoding biology and advancing new therapeutics to radically improve lives.

Applying "artificial intelligence" to drug discovery

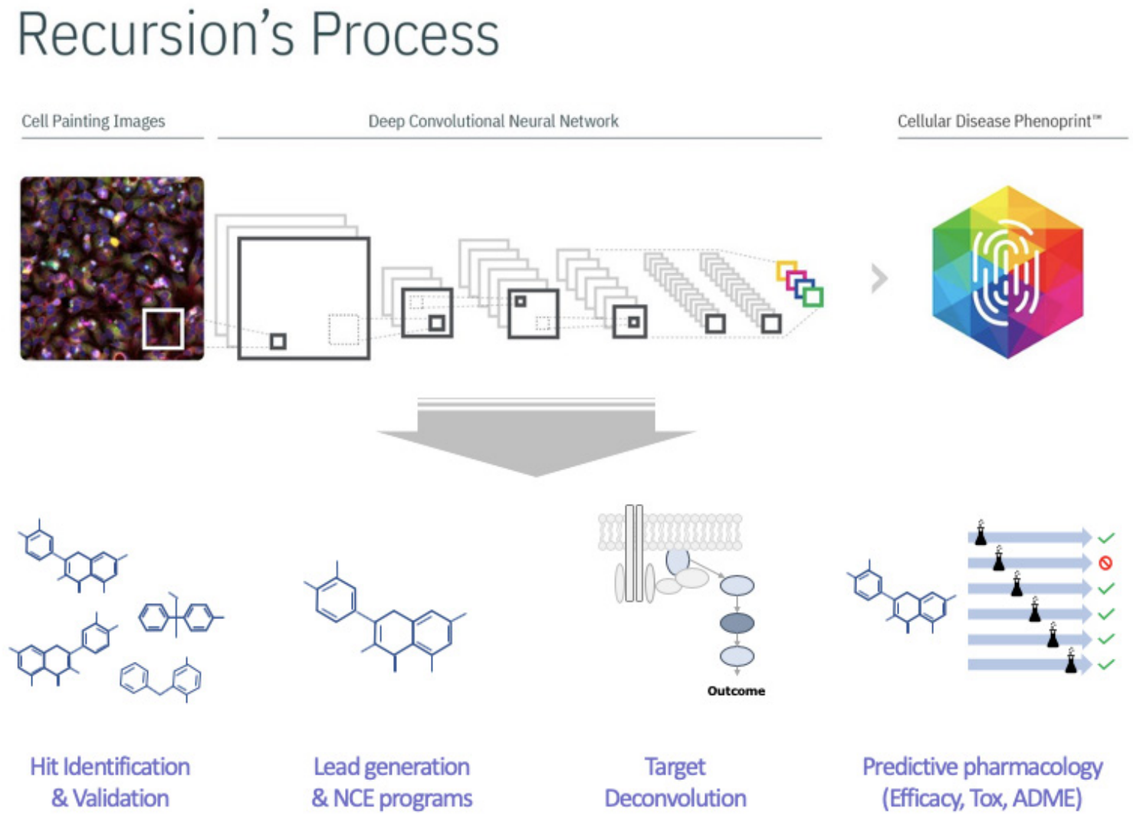
Every decade or so, new trends emerge in pharma that promise to shorten preclinical discovery timelines, minimize failure rates, and increase productivity. In the 80s and 90s, with the maturation of molecular biology techniques, the hope was that protein crystal structures and rational design would find the ideal molecules. In the same time period, combinatorial chemistry and high throughput screening was hailed as the answer to improve our ability to identify more clinical candidates. In the early 21st century, with the large-scale generation of sequencing data, we thought having the biological source code would provide us with the key for unlocking secrets associated with identifying the perfect targets. Phenotypic discovery, while the primary method in which drugs were discovered in the 60s and 70s made a comeback a decade ago, promising first in class medicines through the use of more relevant disease models, including organoids from patient derived material. More recently, life sciences companies across the board are injecting artificial intelligence (AI) into their R&D processes, soaking up over $1.14 billion in venture capital investment in 2018.

What does applying AI to the pharma process mean? Colloquially, the term AI is used to describe algorithms and machines that mimic cognitive functions, such as learning or problem solving, that we typically associate with humans. Unlike the images of Skynet and sentient machines taking over our lives, when people refer to utilizing "AI" in pharma, it is actually much more mundane. Those working in this space are applying modern machine learning (ML) approaches such as deep neural nets, while still relying on workhorse methods such as random forest, which are improved through new architecture enablement. These ML methods are being used to build predictive models across the entire R&D pipeline from identifying novel targets to predicting patient enrollment strategies. With that said, machine learning methods like phenotypic screening have been around for decades, so why do we believe they will make a greater impact today?

With advances in ML, many anticipate that the life sciences and health care industry is on the brink of a large-scale disruption. So, what are the key indicators that "AI"/ML can be a driver of this disruption?

1) More compute power. These days, individuals working at small startups can commercially access compute power that was previously exclusive to institutions (academic or industrial) that could invest and build internal infrastructure. Without pre-planning, you can now instantaneously scale up to thousands for GPUs or ramp down to zero, similar to how you place an order online for groceries. Specialized hardware has also been created to enhance usage of ML, such as Google's TPUs and constant improvements to GPUs are still being made.

2) New algorithms and platforms. Algorithmic improvements in the form of BatchNorm and ResNet, have allowed for big gains to be made by significantly decreasing misclassification rates on certain tasks and are highly efficient at handling imaging data. Modern platforms like Tensorflow and PyTorch, have made it feasible for broader access to deep learning and easy adoption of multi-task and transfer learning, which are critical for regularization and generalization, rather than traditional de novo builds for every new question.



3) Purpose-generated big data. The generation of massive datasets critical for drug discovery applications of ML has been enabled by advances in robotics, miniaturization of experimental biology, and cost reductions in technologies, such as sequencing. Today, high quality and well-controlled biological datasets can be generated specifically for ML applications.

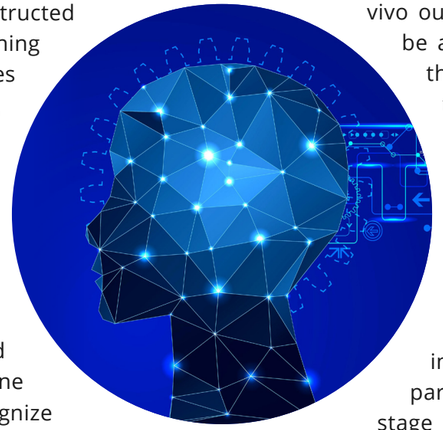
Not all data are (made) equal

Most companies in pharma applying ML to drug discovery believe a) they are sitting on a gold mine of untapped high quality data or b) there is a wealth of information out there in the public domain, for building predictive models. Gaining new insights from public data, however, is a challenging task. While human beings are not the fastest data crunchers, there are a lot of scientists who have been pouring over the publicly available data and are highly skilled at extrapolating and processing complex information. Those applying "AI" to scour public domain knowledge and build predictive models, tend to find themselves rediscovering, rather than observing novel findings.

Large pharmaceutical companies have an enormous repository of historical data, from both research and clinical development efforts. A few dozen companies have a very large quantity of research data, generated over decades, since the introduction of high-throughput screening, but efforts to create centralized information repositories, requiring standardized data generation and curation are not trivial. Although a company may be sitting on a mountain of data, the excavation of that data for mining with ML can be an insurmountable task. Companies that have spent the effort to appropriately curate data for ML use, typically have very low dimensional, single endpoint assay data, disparately screened across biased chemical libraries, creating large, but very sparse matrices of information.

Large quantities of low-dimensional data, while having been useful to address problems we are familiar with in our daily lives such as shopping preferences, may not be sufficient for building robust models that capture the complexity of biology and its interaction with chemistry. The data utilized, whether it be transcriptomics,

proteomic or morphology-based, needs to encompass as much biological signal as possible in order for the models to be constructed effectively. Representation learning or transfer learning approaches will also be critical to combining different data types and building models that can be used to predict highly complex problems in a generalizable fashion. The scientists at Recursion, recently released a large, longitudinally collected dataset of biological experiments specifically designed for the development of machine learning applications to recognize fundamental biology from experimental variability.



(See **www.rxrx.ai** - "CellSignal: Disentangling biological signal from experimental noise in cellular images.")

Defining ground truth

As we build and train our models, we need ground truth to help us establish their predictive capacity. Establishing positive controls are necessary to vet models, but defining the appropriate ground truth will depend on the question(s) being asked. The holy grail would be the ability to predict if a molecule will be efficacious in the clinic and if it will have intolerable toxicities. Unlike tech industries, where the rapid cycles of predicting and testing can be borne out in a matter of days, proving out these clinical predictions could be a decade long endeavour. Defining ground truth in the context of drug discovery is intrinsically multi-layered, but at each stage, a set of well defined pharmacological probes are key to helping optimize the performance of the predictive models.

At the earliest stages of any given program, where hypotheses are driven by knowledge of the underlying mechanisms of disease, ground truth is needed to confirm the molecular and cellular signals, such as engaging protein X or impinging on pathway Y. The primary challenge at this layer is identifying selective pharmacological probes that can help map subcellular target engagement in a predictive model. A significant milestone for a given drug discovery program is typically the identification and establishment of an in vivo signal of efficacy. Again, it will be necessary to use a set of pharmacological probes to help establish the computational models ability to predict in vivo outcomes, but the question begs to be asked, "How useful is it to predict the outcome of a questionably translatable model?"

Molecules fail in the clinic for a complexity of reasons including lack of efficacy, poor absorption, distribution, metabolism, excretion, and unanticipated toxicity. Preclinical toxicity, as it is currently applied in more traditional drug discovery paradigms, occurs at a relatively late stage in the discovery process, where these data are generally viewed as "compound terminators" and moving past a negative result is almost impossible. Predictive models could be transformative in delivering this type of information much earlier in a program, not necessarily to terminate molecules, but rather to inform and monitor molecule progression and design.

While a host of advancements have taken place allowing for broader application of machine learning, significant care and attention still needs to be taken in order to develop models that are reliable for predicting meaningful insights. ML methods can be thought of as young children with enormous potential and, like raising a child, significant effort is required to develop robust models. During childhood, if the wrong behaviours are reinforced, those behaviours will be learned and carried forward in life. So it is with machine learning algorithms. When biased or sparse data sets are used for training, inappropriate signal recognition will lead to a model recognizing the "wrong" type of information to achieve a given task and will fail or mislead the user. Applying ML to drug discovery has great potential, but like raising a child, it will require hard work, persistence and patience to yield useful tools to accelerate the drug discovery process and ultimately bring better therapies to patients ■

Yolanda will be expanding on this topic on Day One of our 20th Annual Drug Discovery Summit with her presentation '*Can AI Help Us In Drug Discovery?*'.