PROVENTA
INTERNATIONAL

Bioinformatics

Insights from the Industry

BIOINFORMATICS

www.proventainternational.com

Introduction

2021 has so far seen an increased focus on innovations within bioinformatics across the pharmaceutical industry. Streamlining drug discovery through computational techniques is becoming a priority more than ever, to improve the management and analysis of vast data sets and accelerate target identification.

At the same time, it is a period of exciting innovation in several areas. In particular, the increased prevalence and investment in AI to accelerate drug discovery and support data analysis. The integration of -omics technologies is becoming increasingly popular across the field, although the challenge of data management in this area continues to receive attention.

Proventa's 2021 event on Bioinformatics examined these innovations in depth. From the integration of modern technologies like cloud-based software and text mining, to hit-to-lead optimisation through innovative techniques like fragment screening, to addressing the challenges within data visualisation like driving FAIRification, to the development of robust infrastructure for data management was discussed by experts in the field.

This report looks to provide greater information on the near future of both medicinal chemistry and biology fields: using data from our expert facilitators, it will discuss the top strategic challenges facing those who attended Proventa's event, as well as the major investments they will be making over the next 12 months. It will also show the quality of attendees at Proventa's event, and feature an article on an important issue in the field.

We hope this report proves both engaging and useful,

Charlotte Di Salvo, *Writer*
Proventa International

TABLE OF Contents

TOP 10 Challenges 2021:
What Peers are Focusing on

1 Artificial intelligence
Despite the success of AI across drug development, a number of challenges were noted by delegates across the pharmaceutical industry. Enabling better decision-making with AI-based platforms continues to be a priority and was highlighted as a challenge for those working within drug discovery and development. In addition, the development of AI for commercial usage and utilising ML solutions, which has become a desirable target for a number of delegates, was also expressed as a difficult area to develop.

2 Data management
With the industry shifting towards a more virtual platform and the analysis of medical records for research - data security has become a priority. The governance of data was raised as a challenge by several delegates, especially so with regards to general data protection regulation. Developing protocols for data minimisation and accountability whilst ensuring it is used in the correct manner for research is a frequently mentioned issue in the field.

3 Data analysis
As pharmaceutical companies invest in technological innovations, whether it be preclinical *in silico* modelling or machine learning approaches for data interpretation, data analysis is changing. The recruitment of data scientists in particular was an area of concern for delegates, as the complexity of artificial intelligence especially requires field experts to run simulations and software. Longitudinal data analysis is vital for therapeutic areas like neurology, however there are a number of emerging issues including missing data and highly complicated data trajectories.

4 -Omics
-Omics technologies have proven to be a popular tool across many therapeutic areas for qualitative and quantitative tissue analysis. The integration of omics analysis into existing workflows and multi-omics research, however, is proving to be a difficult task expressed by many. Utilising genomics was also highlighted as a challenging area, of which a number of delegates focused on technological demands like the analysis of extensive genomic datasets.

5 Digitalisation
Adaptation to emerging technologies was highlighted as a particular issue for a number of delegates. Some delegates inferred the difficulty of developing tools and resources for biological data integration was also an issue. Creating the value out of digital solutions and justifying their integration continues to be a focus for a number of companies. While digital innovations are revolutionising research and drug development, the management of different technologies is an issue for many.

TOP 10 Challenges 2021:
What Peers are Focusing on

6 Costs/funding
The consensus across the board highlighted a number of potential challenges in the future regarding costs and funding. Delegates currently focusing on research inferred that budget cuts were a particular challenge that needed to be addressed. Cost justification, stakeholder management and investments were other areas of focus for improvement in clinical research.

7 Drug discovery
The incorporation of novel modalities into discovery and development appears to be a challenging area across the pharmaceutical industry. In addition to target discovery, assay development was also highlighted as a challenging area within this stage of drug development. Furthermore, the implementation of AI in drug discovery which is becoming increasingly popular is proving difficult with the integration of new technologies.

8 Logistics
On the back of the COVID-19 pandemic, a number of logistical issues remain prominent for clinical and pharmaceutical organisations. The delivery of goods and chemicals continues to be affected by travel restrictions, halting drug production and putting pressure on drug manufacturers to continue to meet targets. Maintaining business continuity remains difficult for many delegates, especially so for those managing clinical research. Patient recruitment was an area significantly impacted by the pandemic and has been slow to recover and meet targets.

9 Regulatory processes
Disruptions in the regulatory environment and pathways continue to pose a challenge for many organisations, and was a point emphasised by delegates. Global market access for a drug product is often hindered by differing regulations between the FDA and EMA. Achieving drug approval continues to pose a challenge for many organisations on the basis of various regulatory processes.

10 Bioinformatics
The integration of computer science and mathematical methods to interpret biological data continues to innovate research and drug development. Unfortunately the integration of bioinformatics, and medical informatics, was noted as a particular challenge for delegates. The transition from traditional to modern software and systems has resulted in many organisations struggling to integrate new technologies. Biostatistics was another area of concern raised, potentially due to issues like a limited number of experts who can use the statistical software and the increasing amounts of data available.

The Integration of Omic Technologies in Drug Development:

An Interview with

Pawel Ciborowski

-Omics technologies have proven to be useful molecular tools across many therapeutic areas in facilitating qualitative and quantitative tissue analysis. Spatially Resolved Transcriptomics (SRT) is an example of -omics technology used in precision oncology. -Omics technologies like SRT continue to develop, with further integration into computational biology. We spoke to **Pawel Ciborowski, Professor at the University of Nebraska Medical Center**, about his thoughts and experience with -omics technologies supporting drug development.

P PharmaFeatures: Would you like to explain your background and experience with -omics technologies?

PC Pawel Ciborowski: I am an analytical biochemist and currently using -omics technologies - primarily proteomics, metabolomics and mass spectrometry - to address questions of biological importance. In my systems biology studies I use human primary macrophages infected with HIV-1. In recent years I have looked at the effect of drug abuse on the macrophage as part of the innate immune system. There are two perspectives on how we look at -omics technologies: technologically, or from the perspective of biological systems.

> **"There are two perspectives on how we look at -omics technologies: technologically, or from the perspective of biological systems."**

P **In comparison with traditional biochemical methods, how can an -omics approach improve drug discovery? Does it support target validation?**

PC The first and foremost way to improve output of -omics studies is to join teams of researchers from the clinics who are facilitating the treatment, and find out the specific question they want to answer. The more specific the question is, the better outcome of the study in question. If I want to answer a question, I have to ask 'how will I design an experiment' and then you fill in everything in the middle i.e. cell model, animal models, human clinical material, sample preparation, analytics, bioinformatics, etc. This is how -omics high-throughput technologies can be used. We have to accept that high-throughput platforms may generate relatively high rates of borderline or false positives results.

In other words, we are also coming to the question of how much analytical change is important for biological effect, and whether the choice to use high-throughput -omics technology will give us such an answer. Recently, I found that a lot of information is hidden by high-throughput -omics, so it acts more as guidance rather than a straight answer.

P **Spatially resolved transcriptomics is an example of -omics application in precision oncology. Are there other therapeutic areas which have benefits?**

PC We have to accept the fact that every -omics approach, whether genomics or proteomics etc, are equally important. This is because they are a part of the function of a biological system as a whole. Every segment has its own group of expertise which can analyse the data properly from experiments. I think this is where we have our weakness at the moment - integration of closely collaborating teams.

P **What are some of the limitations of -omics technologies like SRT? For example poor resolution with single cell experiments?**

PC Single cell technology is great. However, we have to understand that in the human body, we have various cells at different levels of development, differentiation and interaction. If you think about a cancer cell, they respond differently to treatment and treatment may lead to selection of drug resistant clones. So, I think that single cell analysis is a great technology, but the data interpretation and integration with other parts of the whole body are highly challenging.

P I'm aware -omics technology is becoming integrated further with AI. How will AI integration support -omics technologies? How can -omics further improve the drug development process in the future?

PC -Omics high-throughput should be the very first step and be used as a guidance in discovery and preclinical studies. After the -omics phase is completed, it will require several rounds of thorough bioinformatic analysis to allow its use as guidance. Preclinical studies should be structured more like a pyramid, where at the base we have all this high-throughput analysis and at the top it narrows to the specific drug targets/treatment/biomarker, not another way around.

For high-throughput studies, you use cell lines, animals, rodents, monkeys etc... Of course, with cell lines you can get very nice statistics, but how much of the transformed cell lines reflect what is going on in the human body is another question. We need to ask 'how does this information apply to the physiology of humans?' In the human body there are a lot of other mechanisms like secretion, metabolism etc.

It's not the question of whether -omics technology will improve it or not. -Omics high-throughput technologies have evolved and have improved a lot compared to those 15 years ago or so.

P How do you see -omics technology developing in the next few years?

PC We need higher sensitivity and precision from the high-throughput technology, focusing on downstream data analysis and the conclusions that we draw from that. The other problem with this is, yes, we have all these beautiful technologies, but on the other side we have human beings that are so diverse and respond so differently to stimuli, drug treatments etc. Let me go back to the point where we need to ask, 'how much analytical change discovered by high-throughput studies will translate into biological effect? The integration of -omics analytics with data interpretation and biological effect is the direction where -omics studies need to go - I don't think there is any way around that at this point.

"-Omics high-throughput should be the very first step and be used as a guidance in discovery and preclinical studies. After the -omics phase is completed, it will require several rounds of thorough bioinformatic analysis to allow its use as guidance. "

TOP 10 Delegate Investments for the Next 12 Months

Proventa asked delegates at its events to speak about their investments for the coming year. Those surveyed attested to increasing moves towards technology and innovative developments such as AI, as well as the need to further develop target validation, data analysis and resource management.

1 Artificial Intelligence

Significant investment in artificial intelligence has seen a number of new innovations to support drug development. Most notably, delegates have focused on using AI/ML to accelerate drug discovery via AI-based drug screening technologies. The mining of genotypic and phenotypic data was also highlighted. Other companies appear to be focusing on using these platforms for data capacities for example, maximising the value of existing data and supporting knowledge extraction.

2 Data Visualisation

Investment in data visualisation appears to be a priority across the pharmaceutical industry. Delegates expressed the value of investing in data visualisation across all platforms within a business. Examples include driving FAIRification, optimising data analysis as well as improving the access to data in general. Other delegates believe that investment in this area could improve the integration of Integrating data silos for data management and visualisation across R&D functions.

3 Cloud Computing

The increasing number of virtual trials has seen some organisations struggle to integrate emerging software with older systems in addition to the increasing amounts of data. A number of delegates emphasised switching from on premise to cloud solution as a key part of integrating new technologies. Cloud-based software is becoming a key investment area for a number of reasons, including the development of a universally accessible platform for resource management and optimising data analytics.

4 Genomics

Genomic technologies are becoming increasingly popular in providing solutions across drug development, especially so in drug discovery. Several delegates expressed investment in ongoing genomic projects including next generation sequencing and drug response signature detection. Those already utilising these technologies hope to expand and standardise genomic pipelines, as well as ensuring a return on their investment.

5 Data Management

As a result of the integration of modern technologies and progressive digitalisation, managing the increasing amounts of incoming data has become imperative. This supports the decision for a number of delegates who are investing in 'refreshing' their data management, whether this is computer applications and/or database systems, in order to support the collection and management of data. The overall consensus emphasised the importance of developing the infrastructure of data management which is critical for continued progress.

TOP 10 Delegate Investments for the Next 12 Months

Proventa asked delegates at its events to speak about their investments for the coming year. Those surveyed attested to increasing moves towards technology and innovative developments such as AI, as well as the need to further develop target validation, data analysis and resource management.

Text Mining

6

Investment in text mining for drug discovery was expressed as a particular area of focus to which in-text mining of electronic health records is a prime example. The computational analysis of text-locked data in these records has shown to be an important tool for identifying disease associations and healthcare disparities across patient populations. Some delegates also expressed how text mining could enable better use of available real world data, as well as enabling insights from the large corpus of documents.

Knowledge Management

7

Knowledge management promotes an integrated approach to identifying, capturing and evaluating collective knowledge within a system. Investment in this acknowledgement management has been evidence for pharmaceutical companies and CROS. A number of delegates emphasise how knowledge management facilitates internal discussions and specifically could enable consultants to keep up to date with optimal protocols.

RNA/mRNA

8

Ribonucleic acid (RNA) has proven in the last decade to be a valid therapeutic target for a multitude of diseases. More recently, the development and success of the mRNA COVID-19 vaccines has encouraged a number of companies to invest in RNA and mRNA research, investigating potential therapeutic applications across a number of therapeutic areas including cancer vaccines. Some delegates expressed RNA research as a company focus, while others focused on the return of investment in this area.

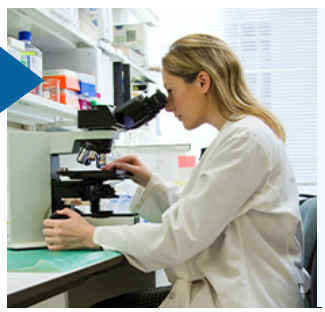
Proteomics

9

Across many therapeutic areas, proteomics has seen significant investment. One of the main challenges with proteomics is the vast amount of data produced, hence many delegates intend to invest in optimising data extraction techniques and analysis. Increasing data mining capabilities to optimise data capture was one particular area of investment for proteomic applications.

Hit-To-Lead Optimisation

10

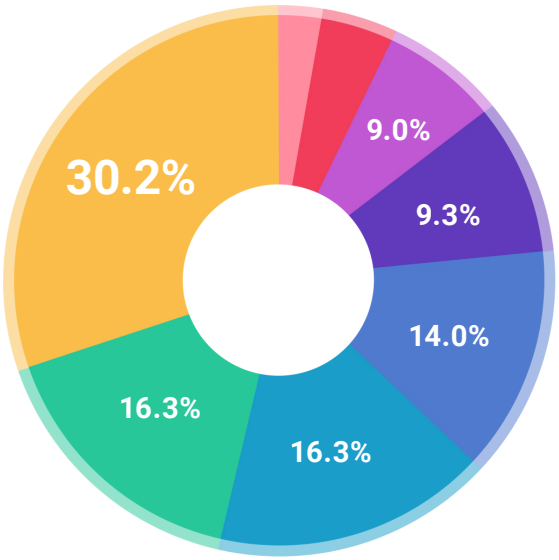


Innovations in hit-to-lead optimisation has resulted in many organisations seeking to invest in some of the emerging techniques. Fragment and virtual screening are prime examples of the latest advancements in demand for companies focusing on optimising drug discovery pipeline. Better characterisation of hits was a particular target for optimising hit-to-lead.

Delegate Breakdown:

Attendees at Proventa's 2021 Strategy Meetings

2021 Atendee Breakdown



30.2% Director level

16.3% Team Lead

16.3% Scientist

14.0% Academia

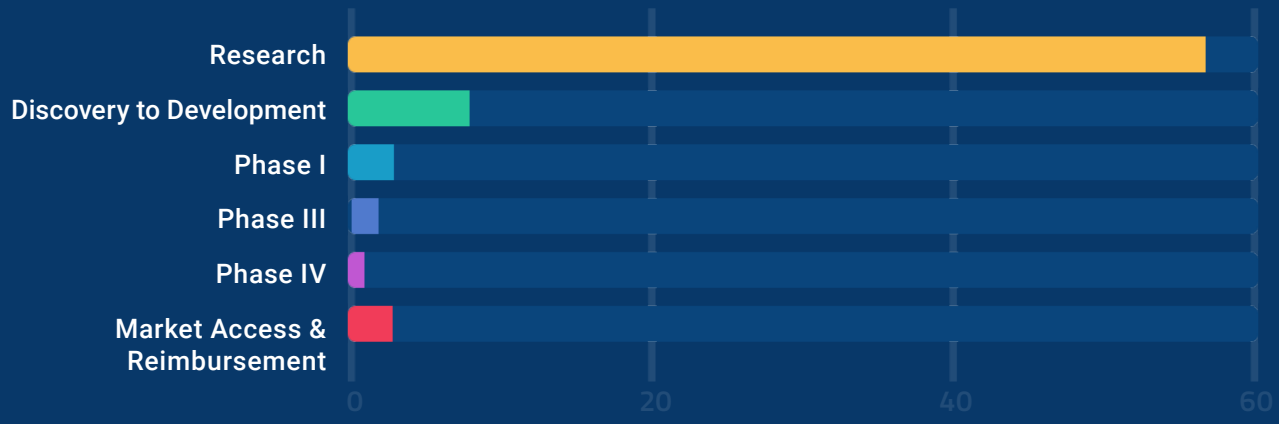
9.3% Manager

9.0% Other

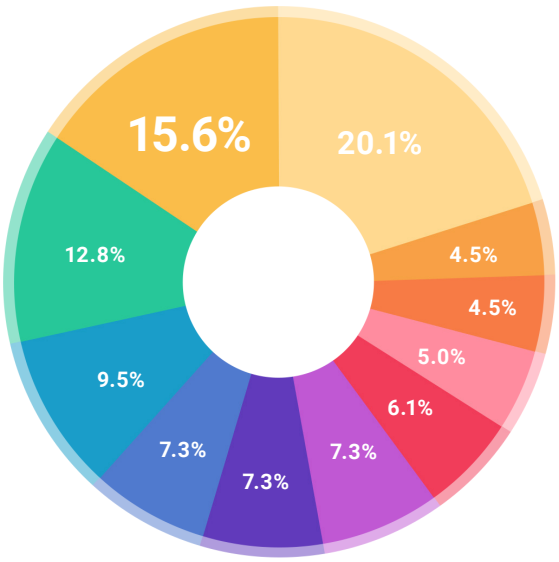
C-class

President / VP

Drug Development Stages



Therapeutic Areas



- Oncology
- Immunology
- Bone/Osteoporosis
- Inflammation
- Metabolic diseases
- Cardiovascular
- Rare diseases
- Infectious diseases
- CNS
- Neurology
- Other

2021 Event Sponsors

LEAD SPONSOR

CLOUDERA

CO-HOST SPONSORS

ITTM
Information Technology for Translational Medicine

Genestack

Mestrelab Research

QUANTORI

Biomax INFORMATICS

10x GENOMICS