BIOINFORMATICS

Insights from the Industry

PROVENTA
INTERNATIONAL

Introduction

The last few years have seen the position of prominence occupied by bioinformatics across the pharmaceutical industry be further established. Advancements in computational techniques, Artificial Intelligence and even more cutting-edge technologies such as quantum computing, have touched upon every area in the life sciences. These include drug discovery, healthcare insights, clinical operations, diagnostics, virtual technologies and countless other fields. Proventa International's Strategy Meetings on Bioinformatics across Boston and London showcased these topics at the cutting edge of industrial discourse through our unique roundtable format - enabling key industry experts to facilitate in-depth discussions on the issues that matter. Common among the themes to further advance the industry was the need for greater collaboration and communication, with an urgent need to facilitate industry-wide standards and demolish silos. Alongside these, many other concerns were raised regarding the challenges the industry is facing as it seeks to adopt and develop the latest technologies in bioinformatics. These are challenges we have aimed to match with the unique capabilities of the solution-providers at our Strategy Meeting, and will further elaborate on in this report - we hope you find it informative.

The Editorial Team
Proventa International

TABLE OF Contents

TOP 10 Challenges 2022: What Peers are Focusing on

1 Data Organization and Infrastructure

As we mature from the Big Data revolution of the 2010s, the industry seeks better answers to questions such as how to organize their data and configure their data infrastructure to suit their needs and minimize siloing. The answer to such questions can often vary drastically between therapeutic areas and organizations.

2 Data Acquisition and Sustainable Data Handling

The acquisition of data and its sustainable handling is a crucial question, as we see gargantuan volumes of data being produced and stored. Many question the utility of such numbers of data and promote better protocols in gathering only the data that is necessary and clinically relevant, while others see no harm in an abundance of information. Divining the right balance to arrive at cost-effective solutions for data acquisition is a priority for the industry.

3 Profiling Across Large Populations

As we see a rise in targeted therapeutics, the immune and genomic profiling of large patient populations is an emergent need in predicting treatment response and resistance. Novel sampling methods, such as liquid biopsies, promise to improve the feasibility of such approaches, while bioinformatics models that can support them are also key.

4 Cybersecurity

With growing volumes of data, which is subject to an increasing degree of regulatory scrutiny and privacy concerns, delegates expressed the need to overcome cybersecurity challenges. This is particularly crucial with regards to data storage and access, and is likely to be an evergreen concern as the arms race between cybersecurity and those who wish to abuse it progresses, much like in other industries.

5 Data Silos

As we seek to adopt more complex Artificial Intelligence (AI) models, the downsides of compartmentalization and fragmentation across the global healthcare and pharma infrastructures are becoming more evident. Higher degrees of collaboration will be needed to train better machine learning models and improve harmonization in research, but delegates express that balancing this cooperation with competitive interests can be uniquely challenging.

TOP 10 Challenges 2022: What Peers are Focusing on

6 Data Heterogeneity

Data heterogeneity remains a key concern - limiting the cross-comparability between similar research, and resulting in wasted efforts to reproduce comparable reference data, or efforts to integrate disparate datasets. Improved communication and standardization of terminology and data used is an urgent concern as we seek to facilitate increased progress in the industry: COVID-19 and the efforts to respond to it showed the potent results of highly standardized, collaborative research pursuits.

7 Target-Disease Relationships

Elucidating the relationships between therapeutic targets and disease mechanisms remains a key priority - particularly for expanding the indications of targets or optimizing other targets for the same disease, as well as minimizing off-site activity. Delegates expressed a need for improved modeling methods to illuminate such relationships.

8 Data Scientist Recruitment

Data science often calls for cross-disciplinary involvement, which requires diverse competencies and varied talent pools. Delegates noted the challenge in bringing over talent which may not originate from the life sciences. Computer and data scientists will be increasingly crucial as firms seek to adapt to new technologies, and solutions to improve their acquisition need to be found.

9 Real World Data

Real World Data presents a treasure trove of insights - or a garbage landfill of gibberish. Ensuring the usability of data gathered through smart technologies remains a key question, with many delegates expressing skepticism on the true applications of real world data. The concept has shown promise, but standards require harmonization and better data acquisition protocols to ensure it works for everyone.

10 Agile Transformation

The COVID-19 pandemic and the increasing progress in technological advancement have shown the need for flexibility and nimbleness at an organizational, systemic level. Delegates expressed the need for their organizations to undergo their own agile transformations, though navigating the path to achieving this remains a challenge for many.

Solving the Big Data Conundrum – Insights from PharmaFEATURES

The life sciences industry may have felt the rush of the Big Data phenomenon later than other fields – but its significance was no less important. Advances in next-generation sequencing have made the data ambitions of yesteryear seem paltry to our current aspirations – we moved on from the Human Genome Project to now mapping out entire microbiomes and thousands of species. This, coupled with innovations in the other omics fields, as well as advances in smart technology that can accelerate the gathering of Real World Evidence – and the growing complexity of our own investigational products – have driven the generation of data to unprecedented levels. How we handle this data and maximize the value obtained from it remains a unique challenge going forward.

Multi-omics Approaches – A Growing Data Footprint

A good example of modern multi-omics data generation and management is the field of oncology. Cancers affect the body at all levels – from the genomic, to the epigenetic, transcriptomic and proteomic levels, and even epigenetics and metabolomics. Combined with novel high-throughput technologies to sequence and generate this omics data, the trend to approach our understanding of tumours through an approach that integrates multi-omics tools is on the rise. This can be illustrated through studies using proteomic understandings to identify critical genes in the development of colorectal cancers, or combining metabolomics and transcriptomics to identify oncogenic pathways for prostate cancer.

Oncology's need for more composite and reliable biomarkers, as outlined above, is only set to increase. Our aspirations of predicting treatment responses in immunotherapy require data-rich environments, as does the trial design of clinical studies in the field. Read more on these topics in our interviews with Jessicca Rege and Amit Agarwhal. But multi-omics approaches raise many questions – the different omics fields not only generate large data sets, but ones that are also heterogeneous and require specialized tools for their integration. Still, choosing the right methods for integrative approaches remains a challenge.

Most current approaches involve dataset manipulation prior to integration – known as mixed integration – to reduce heterogeneity and accommodate downstream integration.

Machine learning models are typically used at this stage, due to the complexity of designing models that can perform integrative analyses at earlier stages. Earlier integration tends to be more effective when datasets are smaller and more homogeneous to begin with – which often raises the question of whether larger data sets from more omics fields are truly necessary. This is due to the fact that the more omics layers are used in an integrative approach, the higher the likelihood of noisy data is. Integration of omics with non-omics – such as clinical – data remains yet more challenging. With the limits imposed by our current technologies, judiciousness and prudence in designing these approaches will be paramount to success. In the future, deep learning models show great promise in revolutionizing integrative strategies.

Real World Evidence (RWE) Ramps up Data Volumes

As the pandemic accelerated the appetite for innovation in clinical operations, we saw a phenomenal rise in the adoption of Decentralized Clinical Trials (DCTs) and nontraditional investigative methods and tools. This sparked further interest in the wider virtual trials space – particularly the concomitant use of smart technologies for real-time data gathering that they bring. DCTs, and their associated technologies, have proven that they offer advantages that make them solid contenders for the future. Real world evidence is here to stay, with both the FDA and the EMA indicating an interest in seeing greater adoption even before COVID-19.

RWE poses unique challenges however – even now there are no globally accepted standard definitions for Real World Data and Real World Evidence, and the difference between the two. Yet signals from regulatory agencies seem hopeful and greater cooperation can be expected in the future. However, the data footprint of RWE can be massive – which used to be, and sometimes still is, one of the deterrents for wider implementation. Sifting through every-day routine data can be daunting without sophisticated study and data handling designs.

The greater challenge in making proper use of RWE is the fragmented nature of our infrastructure – which has been noted as a problem in nations such as Japan and the United States. Effective database linkage and management will be crucial in extracting benefit from novel technologies. Defragmentation of data will require concerted efforts from across the industry and regulatory landscapes to produce deep data integration while remaining conscious of concerns such as privacy, security and trade interests.

Common Challenges

While the fields of multi-omics and RWE may present some of the biggest and most unique challenges in big data handling across the life sciences, they share many of the same obstacles towards future improvement – as does the entirety of the industry. Unlike other fields of science, such as physics where a lot of the big data enterprise is centered around a few colliders, biology has experienced a rapid transformation into big data science. This has brought about a decentralization of technologies such as sequencing and clinical data management. While the democratization of innovation in these ways is a noble affair, the need for improved standards, definitions and procedures for data handling and interpretation is much higher.

Concomitant to the more distributed nature of big data enterprise in the life sciences is the need for infrastructures that can support it. This includes more than just the technologies that can generate the needed data. Infrastructure such as improved software, cloud-based solutions and data storage facilities will need to be developed further.

Innovators across the industry, such as IQVIA and Quantum are already paving the way forwards towards supporting infrastructural needs. The rise of Artificial Intelligence and quantum computing also promise to not only generate more data, but also drastically expand our capabilities of data interpretation and handling. As the world approaches a point where more than one zettabase (a billion terabase) of sequence data may be generated per year, technological and computational advancements will be pivotal in ensuring we make the most of it.

Yet sometimes it is also a worthwhile endeavor to take an introspective approach and contemplate the quality of the data we use. Study design optimization has been the best approach to lowering the amount of data needed down to the most valuable parts of datasets. As biology navigates its own transformation to a big data enterprise, we must remember that data quality retains importance over data volume. Enriching the quality of data we gather will remain contingent on improving industry-wide, and regulatory, collaboration as well as technological advancements.

Read more on the latest trends in the life sciences industry on **PharmaFEATURES**

TOP Investment Areas 2022

Proventa asked delegates at its events to speak about their investments for the coming year. Artificial Intelligence remains a key topic, as are subjects surrounding data acquisition, organization, collaboration and concomitant concerns.

Artificial Intelligence and Machine Learning

Artificial Intelligence (AI) promises to unleash a bioinformatics revolution, much like we saw in the early 2000s, at an unprecedented scale. Peers are examining the key areas AI and related technologies are being applied to, with implementations across pharma and healthcare, while waiting to see which niches it is becoming a permanent fixture in.

Data Visualization and Integration

Data visualization and integration remain key areas for investment in bioinformatics, particularly as peers work with more complex datasets that require detailed examination. The integration of heterogeneous data remains a key challenge to be overcome, especially as multi-omics methods grow in popularity - with methods such as phenotypic drug discovery increasingly reliant on detailed chemoproteomics for target validation.

Data Management and High Performance Computing

Data generation is only one side of the big data coin - proper data management and robust computing power is necessary to use the ever-growing and increasing complex data generated in clinical research. Questions such as data storage and privacy concerns are growing as governments around the world seek to grapple with the tech revolution.

Biomarker Discovery

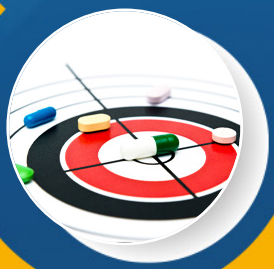
With the rise of precision medicine, we see a growing need for higher quality biomarkers, and even composite biomarkers such as Tumor Mutational Burden, or the microbiome. These will be key to predicting treatment responses and resistance, as well as patient selection. Delegates seek the best practice for guiding their research designs to best harness the potential of new biomarkers.

Cloud Computing

Cloud computing has become a permanent fixture across every industry, and the life sciences are no exception. The technology provides unparalleled flexibility, and as trials see an increase in decentralization and employees work from home, the importance of industry-tailored cloud computing solutions cannot be overstated.

TOP Investment Areas 2022

Proventa asked delegates at its events to speak about their investments for the coming year. Artificial Intelligence remains a key topic, as are subjects surrounding data acquisition, organization, collaboration and concomitant concerns.



Target Validation

Determining the functional relationship between genes, proteins and other targets and disease mechanisms remains key in pharmaceutical research and development. Bioinformatics can support target validation through powerful predictive modeling, and delegates expressed the need to invest in novel technologies such as text mining, which can efficiently process information that integrates knowledge about genes and proteins and linking them to biological function.

Integrated Drug Discovery

Integrated drug discovery has shown a lot of promise, enabling the provision of feedback back and forth throughout the drug development pipeline. This results in an iterative drug design process which can highly optimize putative molecule candidates - and delegates expressed interest in improving their own drug design approaches with integrated drug discovery.

Genomics

Delegates highlighted that genomics will remain of paramount importance to the pharmaceutical industry in 2022, and investment in bioinformatics-driven approaches to genomics are key to leveraging the full potential of the human genome. Genes offer attractive targets for many novel treatment modalities, while also illuminating data about protein products which may be relevant to disease mechanisms.

Laboratory Information Management Systems (LIMS)

As laboratories employ an increasing diversity of data sources and methodologies, they naturally grow in complexity. Delegates recognized this and the corollary need to invest in improved information management systems to better manage their own workplace while adhering to the latest regulations and good practice.

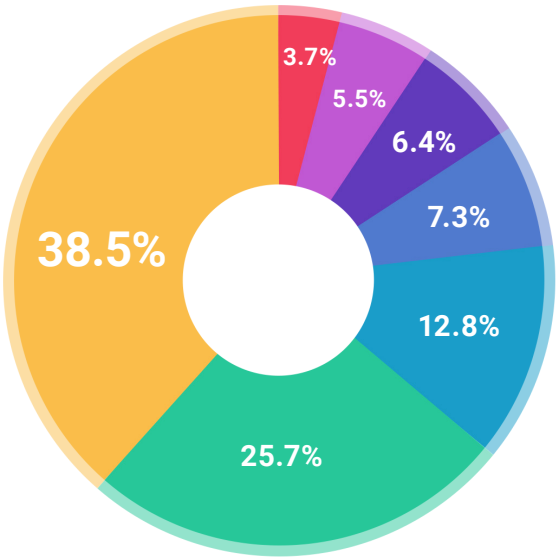
Consulting

COVID-19 has transformed many areas of the life sciences - and bioinformatics has certainly contributed greatly to that transformation. Firms seek external assistance to navigate the post-pandemic landscape, as well as grapple with the breathtaking pace of innovation and development of new technologies.

Delegate Breakdown:

Attendees at Proventa's 2022 Strategy Meetings

2022 Atendee Breakdown



38.5% Director level

25.7% Team Lead

12.8% Scientist

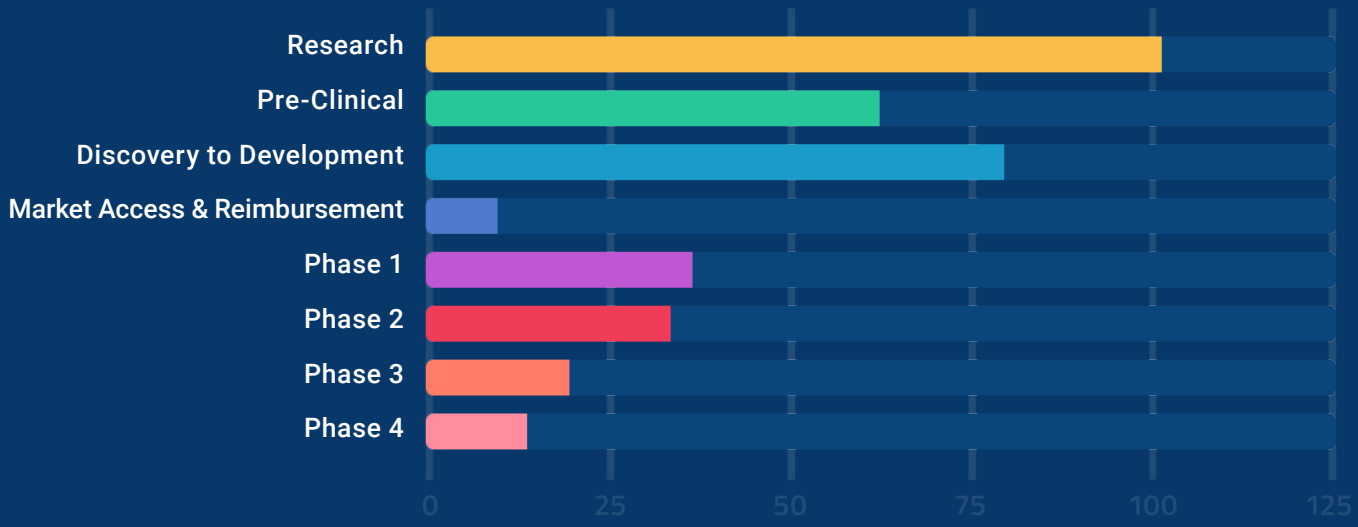
7.3% Vice President

6.4% Other

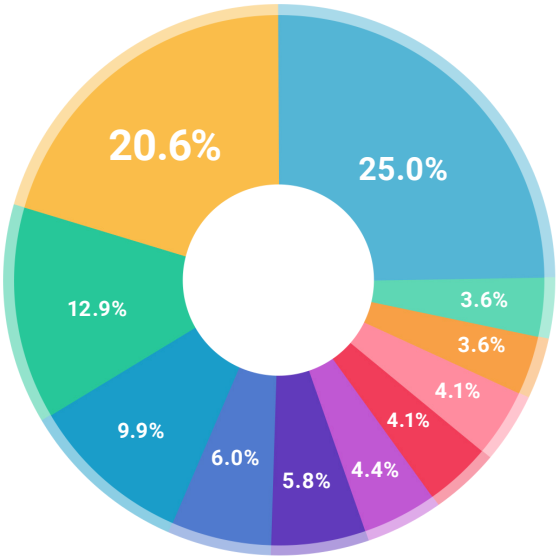
5.5% C-level

3.7% Academia

Drug Development Stages



Therapeutic Areas



- Oncology
- Immunology
- Rare Diseases
- CNS
- Neurology
- Inflammation
- Infectious diseases
- Vaccines
- Cardiovascular
- Metabolic diseases
- Other

2022 Event Sponsors

LEAD SPONSOR

BIOVIA

THOUGHT LEADER

CytoReason

CO-HOST SPONSORS

PRECISION quartzbio

Dotmatics

Quantum

excelra

Benchling

CrucialLogics
consulting with a conscience

SAPHETOR

bionamic

CodeWave

PerkinElmer
Informatics

TechnoLynx