SPECIAL INTEREST GROUP MEETING JUNE 2018

DEVELOPING INNOVATIVE APPROACHES TO USE HUMAN DATA TO DISCOVER AND VALIDATE NEW TARGETS

Richard Lumb
FRONT LINE GENOMICS

Front Line Genomics

Foreword

Despite huge advances in technology, drug discovery and development remains a time-consuming, high-risk and incredibly difficult process, with a worryingly small success rate. Many of the answers for overcoming the challenges in the drug development pipeline remain locked inside unstructured data and in the silos created at each stage of the process.

To understand how we might liberate this data and overcome the challenges being faced, experts in the field were brought together for a meeting under Chatham House Rules, in Boston on June 20th. In order to maintain a balanced perspective, the participants of the meeting were selected from a range of backgrounds, with many different specialties and focuses.

The purpose of this meeting was to address the efficiency and innovation gap in drug development through peer-level discussion and the sharing of experiences of innovative approaches that have been used to achieve common goals. With priorities set in mind, continued conversation and collaborative opportunities will set a path to a dramatically increased success rate of lead generation for drug discovery and a more robust drug development pipeline.

This report presents an accurate reflection of the discussions that took place at this meeting and does not constitute an official statement from any of the individual participants, their organizations, or Front Line Genomics.

Attendees:

Aaron Day-Williams, Director of Quantitative Target Sciences, **Merck**

Diane Joseph-McCarthy, Vice President, Translational Science, **Enbiotix**

Bin Li, Director of Computational Biology, **Takeda**

Vibhor Gupta, Director, **Pangaea Group**

Gautham Sridharan, Research Scientist, **Alnylam Pharmaceuticals**

Leila Pirhaji, Founder and CEO, **ReviveMed**

David Orlando, Principal Scientist, Computational Biology, **Syros Pharmaceuticals**

Ping Zhu, Head of Target Discovery and Genomics, **H3 Biomedicine**

Ken Rhodes, CSO, **Yumanity Therapeutics**

Enrique Neumann, Product and Application Manager, Genomics, **Tecan**

Agustin Chicas, Head of Genomic Target Discovery and Validation, **Forma Therapeutics**

Melanie Adams, Business Development Director, **TTP Labtech**

Derek Stevens, Director, Business Development and Marketing, **Childrens Hospital of Philadelphia**

Helena Deus, Director of Disruptive Technologies, **Elsevier**

Jennifer Wortman, Senior Director of Bioinformatics, **Seres Therapeutics**

Matthew Eaton, Principal Scientist, Computational Biology, **Syros Pharmaceuticals**

Frances Shaw, Producer, **Front Line Genomics**

Richard Lumb, CEO and Founder, **Front Line Genomics**

Report Author: Frances Addison, Staff Writer, Front Line Genomics

Points of Discussion:

Target Validation

- **The definition of target validation is flexible** and will change depending on which stage the drug development has reached. This can make communication between different groups difficult, even when working internally within a single company, as they can be interpreting the word in different ways. One solution may be to introduce a new language to discuss the same principle that is more concise, but this would then necessitate scientists having to relearn terminology.

- **Some drugs reach the market without ever identifying a validated target**. While the majority of drugs being brought to market have validated targets, it is not a necessity if the compound is proven to treat its target condition in accordance with drug regulations.

Artificial Intelligence and Data

- **AI is a black box** and for researchers to be able to trust the data, the technology has to become more transparent. When scientists cannot see how their data is being generated or handled, it is difficult to trust in the reliability of their results. Until AI algorithms are more transparent and comprehensible for non-specialists, a trust issue will remain when using the technology.

- **AIs are highly complex and are rarely user friendly.** The lack of accessibility of AI for researchers who are not familiar with the technology has meant that it is not being used in instances where it may be highly beneficial. There needs to be a push to make the technology more approachable for people who have not received extensive training in computing.

- **The amount of structured biological data is very limited** and thus researchers are restricted when training new AIs. Collecting more data from patients introduces greater privacy risks and can risk skewed datasets with biases towards certain populations. However, by only gathering a specific level of data from patients who consent to share it, the available data pool can be very limited.

- **Many researchers do not know the value of their data**. In organizations such as hospitals and academic institutions, there is little understanding of how pharma can process and utilize data. As a result, the data stewards cannot accurately assess the value of their own resources.

- Genetic data suffers from an 'exceptionalism', when, in practice, it may be no more valuable and personal than other data that many people share freely through outlets such as social media. The pri mary differences are that genomic data can provide information on family members and it may influence the patient's insurance prospects in future.

- AI can remove or minimize researcher biases. Some drug developers will have natural biases as a result of work done in the past and previous examples they've experienced. Because of this, they may overlook an effective drug target. Artificial intelligence programs will not carry this bias (although they can become biased during their training process) and thus may be able to identify targets that would be otherwise overlooked.

Regulations and Funding

- Research funding is largely reserved for data generation, not processing. As very little funding is reserved for data processing and refinement, which can be an expensive and time-consuming process, laboratories are not encouraged to improve their pre-

existing data. Instead, it is more rewarding to spend the time generating more data. As a result, researchers are left with large quantities of unstructured, low quality data that cannot be used.

- Regulations enforce data sharing but do not specify how to do so. Researchers are being asked and, in some cases, required by law, to share aspects of their data, but they have not been provided with guidelines for how that data should be shared. Further, there is no reward structure i n place to encourage groups to spend time developing data sharing protocols.

- New regulations for data sharing are poorly understood, and so researchers are keen to avoid any data that they may not have permission to handle. For example, the introduction of GDPR in Europe has meant that it is easier for researchers to avoid using EU data altogether instead of adapting to become GDPR compliant.

- Pharma companies don't want to fund long-term research that might not benefit them. Inves tigating single genes to provide effective drug targets may allow a development company to bring a new drug to market a decade later, but there is also a strong possibility that the research will be wasted. For pharma companies to sink money into this research, they need to be convinced that there will be value for them and under the present system, this is not the case.

Communication

- There is little communication between researchers in different stages of the analysis pipeline. This means that the researchers generating the data are often unaware which information will be most useful for the scientists working on the downstream analysis. With better communication and collaboration between different stages of the pipeline, researchers are more likely to be getting the data they need without having to complete their own processing.

- It is difficult to identify what data is available and where it is. Even at an internal level, different groups may not know what data is available to them when they are trying to build a dataset. Because the availability of data is not well communicated both within and between companies, researchers can spend much longer than necessary bringing together the data they need for their work.

- The context by which data was collected is often separate from the data. When data is published alongside a paper, the explanation of how the data was collected and considerations that need to be made when using it (potential shortcomings, identified biases, etc.) is typically separate from the data itself. When the data is shared

between multiple groups, this context may be lost.

- Large scale genomic healthcare projects need to involve pharma companies. Projects like All of Us in the US need to bring pharma companies into the conversation so that the research can help to direct drug development and vice versa. If the ultimate goal of the project is to improve healthcare, then drug developers need to be a part of the process.

- The stewards of data generated during collaborations can be difficult to determine. If data is generated or processed during a collaboration between multiple organizations or companies, it can unclear as to which entity has the rights to that data. In many cases, a lack of clear communication regarding the matter leads to assumptions being made that can cause issues at a later date. To establish models that enable transparent collaborations, better communication needs to be encouraged between different partners.