Genedata

The Opportunities and Challenges of Data Science in Cell Therapy Development

PROFILER

Based on a Genedata Profiler Podcast

Author:
Justyna Lisowska, Ph.D.,
Scientific Marketing Manager,
Genedata AG

Featuring:
Daniel Nunez, Ph.D.,
Director of Computational Biology, Cabaletta Bio

Ohad Manor, Ph.D.,
Director of Computational Biology & Data Science, Century Therapeutics, Inc

Industry: Biotechnology
Therapeutic Modalities: Cell Therapy

Listen to the Podcast

Cell therapy is an exciting, rapidly advancing area of research and development transforming the paradigm of treatment. It is also a very complex landscape encompassing various therapeutic modalities and technologies. Currently approved genetically modified immune cell therapies (autologous CAR-T cells) have found a niche in treating haematological malignancies, but continuous efforts are being made to broaden their application to solid tumors and other indications beyond cancer and scale their production by developing standardized off-the-shelf, allogeneic products.

Recently, we had a chance to speak with Dr. Daniel Nunez from Cabaletta Bio and Dr. Ohad Manor from Century Therapeutics, two experts in the field leading digital strategies in pioneering cell therapy companies, about the importance of data and data-centric decisions in maximizing the potential of these transformative treatments. During our discussion, they shed light on the data challenges faced as they empower their companies to innovate and develop disruptive cell therapy modalities for cancer and autoimmune diseases.

Why is Data Important?

In today's biopharmaceutical industry, data drives innovation and advances the progress of clinical development. With complex modalities like cell therapy, the importance of data is even more critical to enable smarter clinical decisions and deliver high-quality, efficacious products. Unlike standard therapeutics, cell therapies are evolving living entities, whose activity and quality attributes depend not only on their inherent physiological characteristics but also on their interaction with their surrounding environment. Therefore, when it comes to making decisions regarding product safety, potency, manufacturability, or scalability, it is crucial to thoroughly investigate the whole biological system at different molecular levels in conjunction with the patient-specific features and manufacturing parameters. All this requires various types of assays and analytical approaches and results in large amounts of diverse, complex data.

The collection and analysis of varied data types throughout the product and patient journey enable scientists to evaluate product safety, efficacy, and manufacturing characteristics. Before infusion, cell therapy products are extensively characterized for their cell phenotypic and genetic features using flow cytometry, next-generation sequencing, or PCR to determine their purity and key proprietary attributes. In patients, the evaluation of antigen expression levels on target cells serves to predict the outcome of treatment, while the measurement of cytokines post-infusion is key to assessing product safety and mitigating potential toxicity. Unlike other drugs, manufacturing data plays a critical role in cell therapy product development; for instance in autologous therapies, the number of generated cells from the patient, their size and biological activity are indicative of the patient's response to treatment. These data points provide areas of focus for companies to improve their treatment's probability of success.

For allogeneic therapies where immunogenicity issues can arise, it's important to measure any markers of immune reaction: cytokines or antibodies generated following patient treatment. Finally, engineering iPSC-derived allogeneic off-the-shelf products requires consistent and thorough cell phenotypic and genomic profiling along the differentiation process to generate a high-quality, potent, and safe master cell bank. Full characterization of this source material is needed to ensure new desired features are incorporated but no important signalling pathways have been disrupted in the final immune effectors following genetic engineering. As mentioned by Dr. Manor, clones with unwanted, off-target, modifications are dropped from the pool to have only genetically clean cells for further product development.

Where Does This Data Come From and How Is It Handled?

Most of these cellular product characterization and manufacturing data are produced by companies in-house, although data from specific technologies such as Next Generation Sequencing (NGS) may be generated by external service providers or even collaborators from Academia. On the other hand, clinical trial data from patient samples are usually handled by clinical research organizations (CROs) contracted by the company. The data provided by CROs are not always standardized and need to be pre-processed and analyzed. "There are no agreed formats across companies for different analysis types," says Dr Ohad Manor. Bringing such data from disparate sources under one roof does not allow researchers to seamlessly integrate suitable data sets for further analysis and insight generation.

This is why data scientists at Cabaletta Bio and Century Therapeutics need to organize and structure such data first. They harmonize diverse data formats and naming conventions to ensure data is meaningful, reliable, and interoperable before they can productively use it to address critical scientific questions. These processes are time-consuming, and some are manual and error-prone. Thus, companies would benefit from a tool that could automate these steps to accelerate data-driven decision-making.

How Can We Extract Knowledge from Data?

Several bioinformatics tools can be subsequently used for exploratory data analysis and visualization to generate insights about product safety and potency. Computational biologists at Cabaletta Bio and Century Therapeutics use a programmatic environment with Python and R to build applications or data-specific analytical pipelines to support their scientists in making scientific decisions. Both companies also see great potential in AI-powered solutions. Some cited examples include the use of the Google Alpha Fold AI-based tool to predict the structure of functional chimeric

antigen receptors (CARs) and machine learning to forecast the cell differentiation trajectory of iPSCs based on differentially expressed genes while generating the master cell bank. Dr. Daniel Nunez also sees the benefit of using generative AI-based tools such as ChatGPT to support data scientists in developing analytical tools or analyzing data programmatically. Such solutions can also help to quickly address pending scientific questions and generate new hypotheses on the fly based on accumulated data. Finally, knowledge graphs can be utilized for identifying unexpected correlations between data and building predictive models for instance to identify biomarkers of treatment response.

However, according to Dr. Manor, what prevents data scientists from leveraging such innovative, advanced analytical approaches is the hindered access to consolidated data generated across teams or external collaborators. While attractive, the idea of having a centralized storage location for all the data may not be viable due to the ever-increasing amounts of research data requiring a scalable digital data hub for its aggregation. As an alternative, he would like to see a digital solution that can virtually federate relevant datasets from different data sources (S3 buckets, PC, SharePoint, ELN systems) in real-time, creating a single point of access to easily connect and integrate suitable datasets when needed.

Both Dr. Nunez and Manor also recognize the need to improve the data-sharing experience and enhance the quality of reporting. Instead of using static PowerPoint-type presentations, which is a gold standard in most companies, they would like to automatically link out the most recent data from various sources to different analytical and business intelligence tools to generate interactive self-service dashboards. “It would be really nice to have a centralized dashboarding area where different people could see the reports in real life as the data is coming in”, says Dr. Nunez. Such user-friendly real-time visualizations would empower scientists to derive knowledge efficiently. They would also provide insight into data provenance and lineage for a better understanding of the analytical approaches behind them.

Obstacles to Digitalization and Data Culture

While there is an emphasis on collecting diverse data types in the cell therapy space to enhance product development, our speakers acknowledge there is still a lot to learn. Currently, companies are trying to understand what kind of data to use and how to best leverage them to support their research questions and enhance clinical decisions. “There is a lot of interest in making things more digital and making the data science work better for the company,” said Dr. Manor. Nevertheless, to digitalize cell therapy R&D workflows and establish a data culture in biopharmaceutical companies, the prerequisite is for people to be ready to work differently and embrace data democratization and sharing principles. Users need to be familiar with the benefits such a transition can bring to their daily work and be incentivized from early on. Once onboard, the company can invest in an easy-to-use, intuitive platform to deliver on its promise and empower its teams to innovate.

To find out more, have a listen to our podcast.

Listen to the Podcast

Visit our website to learn how Genedata Profiler supports biopharmaceutical companies in maximizing the value of data.

Genedata PROFILER

Genedata Profiler® is part of the Genedata portfolio of advanced software solutions, which digitalize and automate data-rich and complex biopharma R&D processes. From early discovery all the way to the clinic, Genedata solutions maximize the ROI in R&D expenditure.

#PathToPrecision