A Universal Data Model for longitudinal integration of Disparate Biomarker and in-life patient data augmented by machine learning

Dr. Rajalaxmi S, Arpitha Hanumanthiah (PointCross Inc.)

Abstract

Translational and precision medicine development in immune and gene therapies rely on biomarker data from assays including genomics, proteomics, IHC, Flow cytometry and cell phenotyping data. Biomarkers are not only generated from the patient biosamples, but also from the biomanufacturing sites such as in batch processing of the adoptive immune cell biomanufacturing for novel immunotherapies. Extracting valuable insights from these disparate data sources and integrating it to clinical data to relate to patient outcome and/or discover and validate relevant biomarkers are met with challenges of ingestion, harmonization and integration of disparate data with the clinical data. Key decisions and ideas that impact study design of future clinical trials depend on gaining insights rapidly from such integrated data on patients or stratified cohorts. Xbiom, built on Machine Learning and Universal Data Modelling effectively solve these challenges of disparate, big and varied data sources.

Systematic curation with self-validation for completeness and consistency is time-consuming and difficult without technology. Xbiom platform built on machine learning-enabled Smart Transformation is capable of making ingestion, curation and harmonization process automatic and are able to process both streamed data as well as in batch mode. Machine learning algorithms work on the principle of supervised learning and are initially trained within training sets. The Xbiom platform is not only agile by having the ability to retrain the algorithm for the newer/unseen type of data types and structures but also robust in predicting the column for a recommendation in manual approval in case of unseen data types. The curated and transformed output is quality checked and can be viewed in the QC dashboard for the process quality. Curated and harmonized data is stored in a Universal Data Model (UDM) that holds study and assay data of subjects or samples and their attributes in a simple, indexable form that facilitates instant search through query masks to yield highly stratified cohorts and data. Search and detection of signals is enabled by the integration of assay data into longitudinal patient data. UDM can be unique to the requirement of the organization and is extensible if the future requirement may arise in the drug development/discovery pipeline. Data in the UDM is ready for the statistical analysis and can also be used as a data set in machine learning designed for biomarker discovery and other insights generation as precision and translational medicine. Xbiom's data governance system allows for the role-based access to keep the data in the process for transformation or in the UDM secure when accessed in the browser or through the API.

Xbiom's Smart Transformation and UDM solutions, for automatic ingestion, harmonization and integrating of data in real-time, are compliant to quality standards and data governance enabling valuable insights for development in clinical trials and in biomarker discovery, making Xbiom a unique product to reduce the time and cost of drug development, immunotherapy as well as in the translational and precision medicine.

Authors

Dr Rajalaxmi S is a physician and data scientist with 15 years of experience in patient care and hospital administration. She is now focused on data science, analyzing patient data.

Arpitha Hanumanthaiah is a Subject Matter Expert with 8 years of experience in Data Standardization focused on Clinical and Biomarkers (Omics) data.