excelra

Where data means more

CASE STUDY

Building large data assets allows MoA elucidation while keeping a view of the competitive landscape

Client's challenge and goal

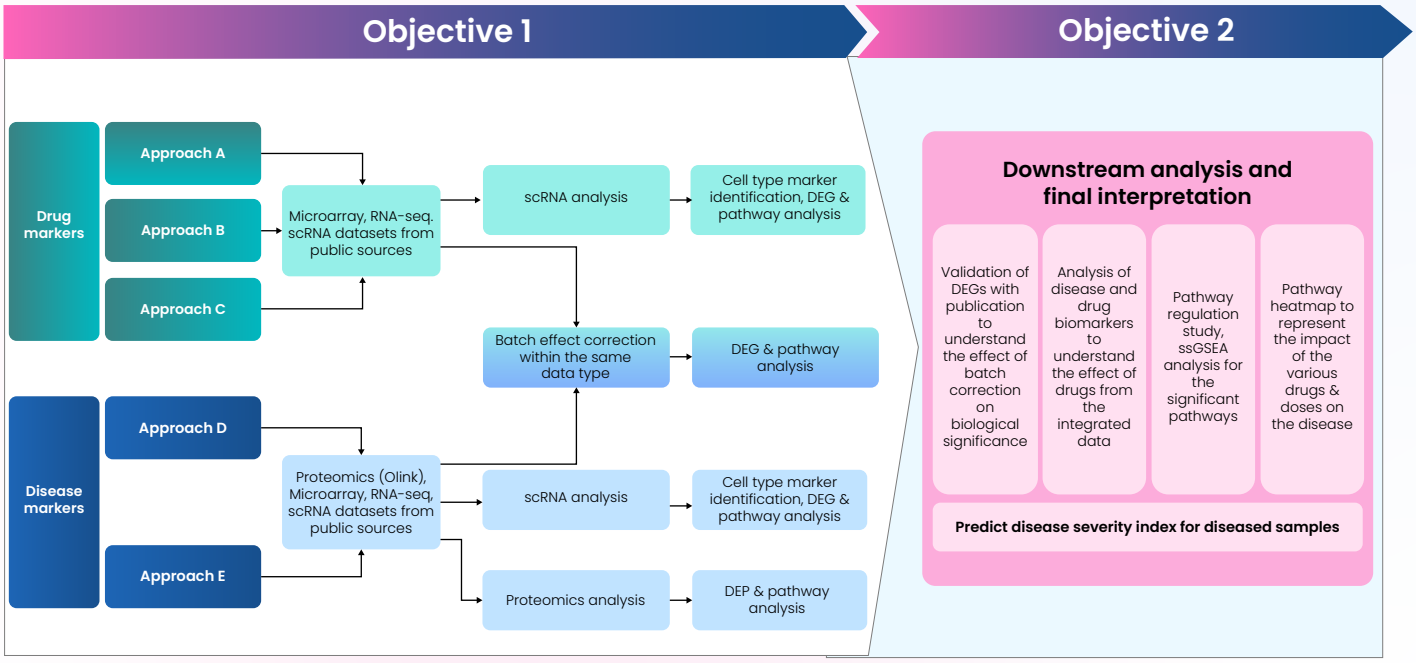
- The client was a global pharmaceutical company that majorly focusses on the RnD & identifies therapeutic solutions to diverse disease areas. The Client's asset (validated inhibitor of Target X) is successfully moving through the early stages of Clinical trials. However, their interest was to understand how their drug is going to perform for their disease of interest and whether it potentially holds best-in-class characteristics compared to already available marketed drugs.
- In doing so, the client wanted to create a large comprehensive data asset of publicly available datasets that can enable MoA elucidation while considering the competitive drugs for their disease of interest. Due to different data types, platforms & sources, data heterogeneity posed a major challenge. Moreover, different formats with various normalizations in the source added another layer of complexity, where batch effect correction and/or co-normalization became the essential step. And therefore, finding the best method for the purpose was a challenge because this might reduce the biological signal and relevancy of the dataset.
- The metadata was not standardized and required of much manual effort during structuring & data harmonization. Hence, extensive subject matter expertise and substantial efforts were required for data collection, cleaning, harmonization, and validation before the asset could be used to generate useful insights.
- The client was also interested in gauging the repurposing potential of their asset against other similar diseases. This required implementation of a scalable solution that could be applied to other diseases of interest.

Our client

A major Global Pharma Company

Our approach

- Excelra's approach was to create a large inventory of all public datasets, coming from different platforms & technologies but centered around the given disease and target as well as those downstream to their target of interest. In doing so, an important aspect of capturing reported evidence in associated publications was also considered. This helped in performing cross-validations when it comes to identifying drug & disease-associated biomarkers & how the signatures varied at different timepoints & doses for drugs studied in such datasets.
- The large project was divided into several approaches with the primary focus on dataset curation, followed by analysis & generating insights. The curation activity involved identifying & gathering different variables that can be captured from metadata associated with the reported dataset. The curation involved capturing details at the dataset as well as sample levels. In case of need, data from the publications were also extracted to make the data asset more self-contained enough to perform the analysis & co-relate with clinical captured parameters. Standard vocabulary was also taken into consideration while building the data asset.
- Due to the data coming from different platforms & technologies, it was challenging to retain the biological signal of the samples and genes and identify the significant pathway regulations after batch-corrections/co-normalizations. Multiple approaches had to be explored to come to the desired outcomes.
- Excelra leveraged AWS server instances and internal computational power for data collection, harmonization and analysis. Excelra's subject matter experts with experience in handling diverse data types were crucial in completing the project on time. The team used the knowledge gained from the first project to design plans to achieve homogenous results for the next projects in a shorter timeframe. The team handled the finer nuances of the objective to deliver quality output.



Our solution

- The team analyzed more than 4500 samples from different datasets for a given indication.
- The Excelra team divided the project into 5 different approaches to handle the client's expectations. Mainly, the team curated and shortlisted the transcriptomic, proteomic and scRNA-seq datasets relevant to drug, target, and disease of interest. The team also gathered datasets for other drugs targeting the same disease or those with knockout or inhibition studies for the target and the disease. The team planned and collected the raw data and metadata, classified, combined, and harmonized it, and analyzed the combined asset to report quantitative results regarding the drug and disease biomarker genes and pathways. The quality checks were performed at each step and the results were validated using reported literature evidence.
- As per the objective, various contrasts were prepared & studied that comprehensively provided a clear view of how the Client's asset would perform against the indication of their interest.
- The delivery team shared the results summary in the form of PowerPoint slides, while other files generated or needed during the analysis (such as raw count matrix, normalized matrix, metadata files, DEGs, DEPS) were shared in respective platform readable format.
- The client can use the data for MoA elucidation since the team has curated and analyzed data for all possible market-available drug datasets and analyzed the results. The impact of the various market-available drugs on the disease markers was analyzed which in turn will help the client to apply that knowledge for their own target molecule.
- The data asset created for the project will continue to help the client to identify the population of samples and their detailed metadata and perform additional analyses/ experiments of their own on the data in the future.
- The robust statistical framework developed by the team has enabled the client to estimate the drug-repurposing potential of their asset across other indications. This knowledge can help the client prioritize and design their upcoming projects. The framework can be generalized for any drug-indication combination.

Conclusion and CTA

- The client appreciated Excelra's capability to handle huge amounts of data and identify the best possible solutions within the given timeframe. The client was satisfied with the deliverables, approach, and quality of the results that has led to continuing engagement over other indications.
- Other than the bioinformatics and curation services used here, Excelra also offers services in the area of ontology management, database creation, and bioinformatics pipelining that are useful in data asset creation and management area. Interested parties can find additional information here

  https://www.excelra.com/applications/bioinformatics/

  https://www.excelra.com/databases/custom-biomarker-knowledgebase/

  https://www.excelra.com/blogs/fairification-a-path-to-scientific-data-connectivity/

  https://www.excelra.com/platforms/online-pipeline-platform/
- The reader can find out more about our offer or get in touch with us here

  https://www.excelra.com/contact-us/

Customer feedback

- The feedback on the deliverables was positive resulting in extended enagagements. During the entire course, the Excelra team played both roles of experienced consultants as well as strong collaborators.
- Excelra provided unique solutions that can be generalized for other problems within the same or other teams in the client company. The effort from the Team has successfully led to continuing engagement with the client over other indications of interest, having much larger objectives.

Where data means more

excelra

SAN FRANCISCO • BOSTON • LONDON • GHENT/GENT

SCHIPHOL • UTRECHT • BASEL • BIELEFELD • HYDERABAD

Connect with our experts: marketing@excelra.com

www.excelra.com