

## 2020 ACCOMPLISHMENTS

---

- **Project Survival (Corporate goals – Generate corporate value, F.A.I.R. data structures)**
  - Built data ingestion pipeline to obtain clinical data from StudyTrax every week via StudyTrax API.
  - Upgrade omics processing pipeline to support newer batches of omics sample from BIDMC.
  - Improved data ingestion speeds from LIMS system (68% faster transfer) by applying filter on SQL server, cutting down on size of data transmitted.
  - Performed survival analysis to find effectiveness of various chemotherapy regimens for treatment naïve PDAC patients.
  - Applied statistical analysis and hypothesis testing methods on omics data to find predictive biomarker panels for survival.
  - Generated bAlcis networks for visit one omics data and all participants.
  - Generated bAlcis network for clinical only data at visit one for all participants.
  - Created plots and tables to summarize results for conference posters.
- **LIMS and FreezerWorks (Corporate goals – F.A.I.R. data structures)**
  - Ported existing dashboard to run with R 3.6 (previous version R 3.4).
  - Added dashboard modules to provide an aggregate view of patients and samples on LIMS and FreezerWorks systems.
  - Added ability to point-and-click on distribution plot (bar plots) to drill deeper into a category and explore subsets of the data.
  - Added ability to export subsets of data from the dashboard as CSV or Excel file.
  - Fixed issue with FreezerWorks (FW) server not sending across all requested data via API by updating configurations on FW terminal.
- **BETSY (Corporate goals – F.A.I.R. data structures)**
  - Deployed updated internal databases - Patent DB, DrugBank, and PubMed – and set up automated pipelines to perform scheduled updates for the same.
  - Implement functions to use BETSY web API to obtain Patent DB, DrugBank, and PubMed information about any analytes detected at BERG.
  - Implement function to add color to XGMML node description based on sample type and omic type to add support for multiple samples and omic type in a project on BETSY.
  - Built data pipeline to push list of all analytes – proteomics, lipidomics, metabolomics – processed at BERG to BETSY on a weekly schedule.
  - Process bAlcis networks from legacy projects and push the same to BETSY under network exploration projects.
  - Fix authentication for BETSY dashboard by configuring password tables on BETSY table using Postgres SQL.
  - Improve and update project documentation with updated commands and methods to deploy networks/analytes.

- Smart bAlcis (Corporate goals - Increase usability, stability, and capabilities of analytics algorithms)
  - Implement model pipeline to process data and train models for prediction of various metrics and running time of bAlcis algorithm for a given dataset and parameters.
  - Trained over 2000 neural network models to find optimal hyper parameters for prediction of each metric and obtained best performing models.
  - Implemented prediction pipeline that takes a dataset and generates optimal bAlcis parameters that satisfy Precision and running time constraints provided by user.
- Advent Health COVID project (Corporate goals – Generate corporate value, F.A.I.R. data structures)
  - Built data processing pipeline to generate wide format data (clinical events as columns, patients as rows, event value as cell value) for bAlcis processing.
  - Implement functions to split events into multiple time bins for pre and post COVID datasets, with added checks to set events to “Not Applicable” for patients once they are discharged.
- Boehringer Ingelheim: Proteomics Data Processing and QC (Corporate goals - Increase usability, stability, and capabilities of analytics algorithms)
  - Repurposed proteomics processing scripts to add support for glycolipid processing.
  - Built QC and plotting pipeline for processing BI data and to automate QC checks and report generation.
  - Repurposed outlier detection scripts for BI data and ran PCA outlier detection to flag erroneous samples for further QC.
  - Implement function to process Tout data objects to generate tables in excel and export PNG files with cosmetic updates and fixes for external review.
- Metabolomics name standardization (Corporate goals – F.A.I.R. data structures)
  - Assisted in creating a standard list of metabolites processed at BERG by fixing naming inconsistencies in analytes names using IDOME dataset.
- Misc. analytics support (Corporate goals - Increase usability, stability, and capabilities of analytics algorithms)
  - Implement function to split merged analyte nodes in bn bAlcis object for compatibility of analytes with BETSY standards.
  - Implement function to generate count and distribution plots quickly with visualization best practices and BERG color pallet.
  - Find fix for magic number issue when loading R data objects between different R version (R 3.6 -> R 3.4).