Accelerating Biomarker Insights from Public Omics Data

AI-Powered Curation, Meta-Analysis, and Clinical Translation

Our fully automated platform combines expert-guided curation with advanced bioinformatics analysis to process large-scale public omics datasets and provide analysis-ready datasets optimized for downstream applications.

Analyzing Large-Scale Omics Datasets

Key Challenges

Manual curation of literature and datasets is resource-intensive and slow.

Omics datasets are heterogeneous, large, unstructured, and difficult to interpret and analyze.

Current workflows lack tools for seamless integration into data warehouses or analysis pipelines.

Integration of ML-based approaches are required for stratification and analysis of large-scale omics datasets.

Our Solution

Our automated platform integrates data curation, bioinformatics, and machine learning for biomarker discovery. The platform:

1. Automatically searches and retrieves relevant public datasets (e.g., GEO), based on disease, tissue type, and assay modality.
2. Generates structured outputs formatted for direct integration into client pipelines.
3. Has standardized pipelines that perform differential expression analysis and statistical validation.
4. Incorporates machine learning models to identify and rank predictive gene markers.
5. Enables reliable patient stratification and supports development of PCR-based companion diagnostics.
6. Allows the entire workflow to be potentially hyper-automated with the help of strand engineers and data scientists for any given problem set to ensure a rapid turnaround within weeks.

Recent Case Study

Goal: We aimed to demonstrate how our automated platform could stratify ulcerative colitis patient populations and help identify cohorts most likely to respond optimally to specific therapies or drug dosages.

Solution: Our three-step process included:

Curation & Dataset Wrangling

Identifying biomarker leads using public databases for targeted assay development

Validation with Existing Data

Evaluating selected markers for patient stratification using meta-analysis and modeling

Assay Translation

Developing PCR-based companion diagnostics suitable for use throughout clinical trials

- The curation criteria comprised parameters such as sample size, data quality, and tissue type to select high-value datasets for analysis.
- An interpretable scoring system was created using selected markers to stratify patients into response cohorts.
- ML models were applied to assign gene weights and classify patients by disease severity and likely therapeutic benefit.

Our standardized summary for an Ulcerative Colitis dataset included:

Total Number of Samples

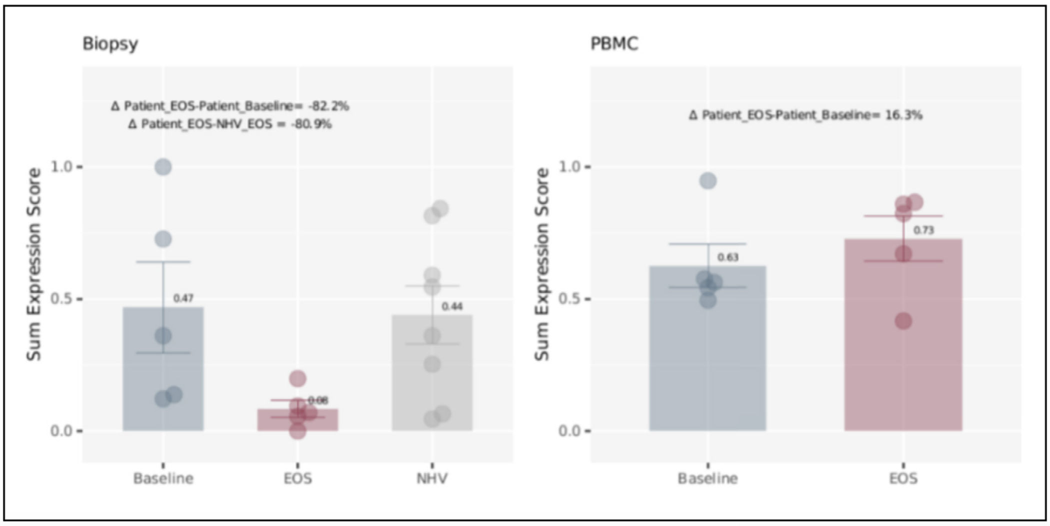
Biomarker Identification

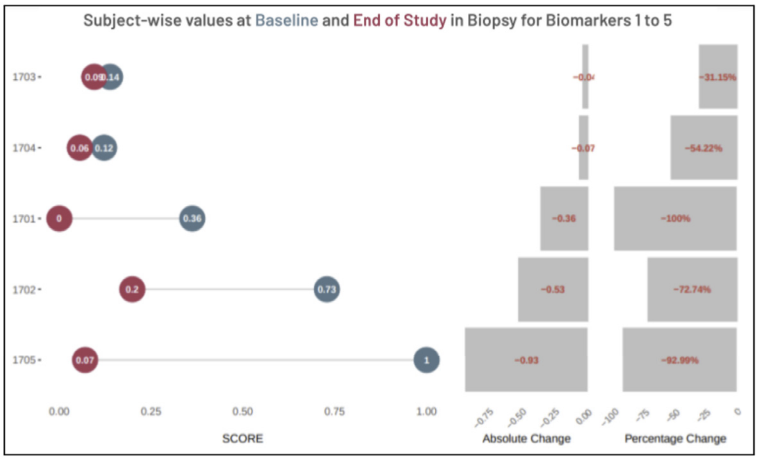
Reference Publication

Dataset Identifier

Results

The graphs in the figures below describe the performance of selected 5 markers on ulcerative colitis patient biopsies. The prioritised subset of markers identified by ML from the set of markers analyzed and curated from public sources by the team at Strand showed a collective reduction in expression at the end of the study (8 days post treatment) compared to baseline in ulcerative colitis patient biopsies.




Subject-wise values at Baseline and End of Study in Biopsy for Biomarkers 1 to 5

Impact

Data Analysis

15+ datasets and 3,000 samples

Biomarker Selection

Selected the top 5 markers from 130

Sensitivity Improvement

Sensitivity at 90% specificity improved from 38% to 63%

Sensitivity Achievement

100% sensitivity at 70% specificity using only 5 markers

Assay Format

The assay format now supports PCR implementation.

7th Floor, MSR North Tower, #144, Outer Ring Road, Nagavara, Bengaluru - 560045
+91 9980448044
hello@strandls.com
strandls.com