strand

Double the Data, Double the Discovery:

Hunting Publicly Available and Generated Transcriptomes

Preface

The wide landscape of biological investigation is always changing, as are the technologies used to uncover the mysteries hidden within cells. Generating transcriptome data is a time-consuming and resource intensive operation. However, researchers also have access to a variety of transcriptome data stored in public archives. Deep diving into this publicly available data/meta-data, utilizing state-of-the-art mining and analytics paired with curation helps us uncover unique biological insights. This in turn has the potential to transform our understanding of disease processes and drug target response.

Our team is skilled in RNA-seq analysis and bioinformatics pipeline building. We employ reproducible bioinformatics pipelines with extensive QC for primary, secondary and tertiary analysis to ensure data accuracy and integrity. Furthermore, we specialize in designing and implementing customized bioinformatics pipelines based on individual study goals. These pipelines streamline the analysis process, and help extract from raw RNA-seq data, disease/pathway associated marker lists. De-novo ML approaches or weighted expression sums of select markers lead to a scoring system with valuable biological insights towards disease prognosis, therapy and response.

Fig. 1: Mining public and generated transcriptome data for maker discovery

Example: **Workflow to explore Disease/target associated Markers in Ulcerative Colitis**

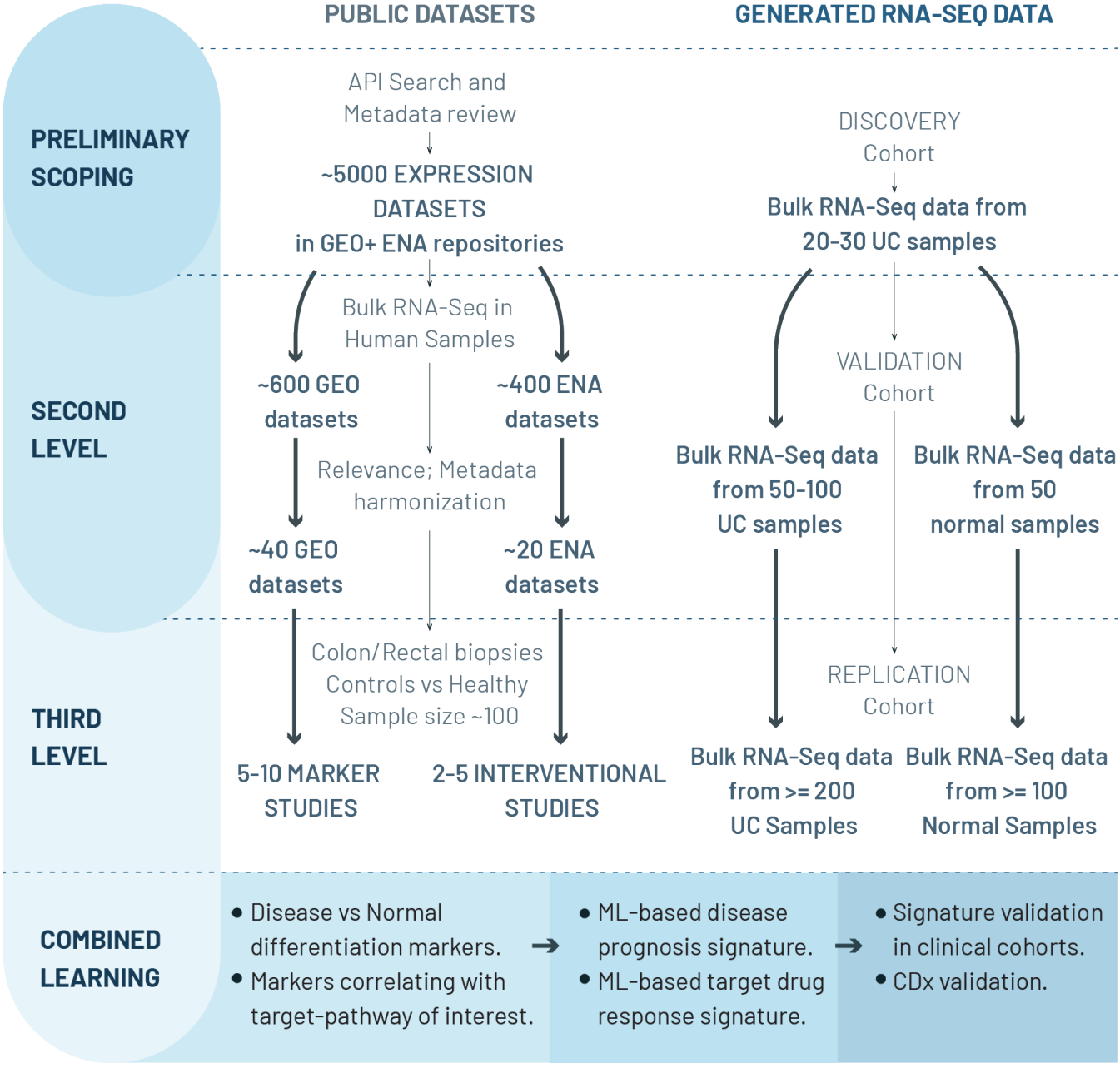
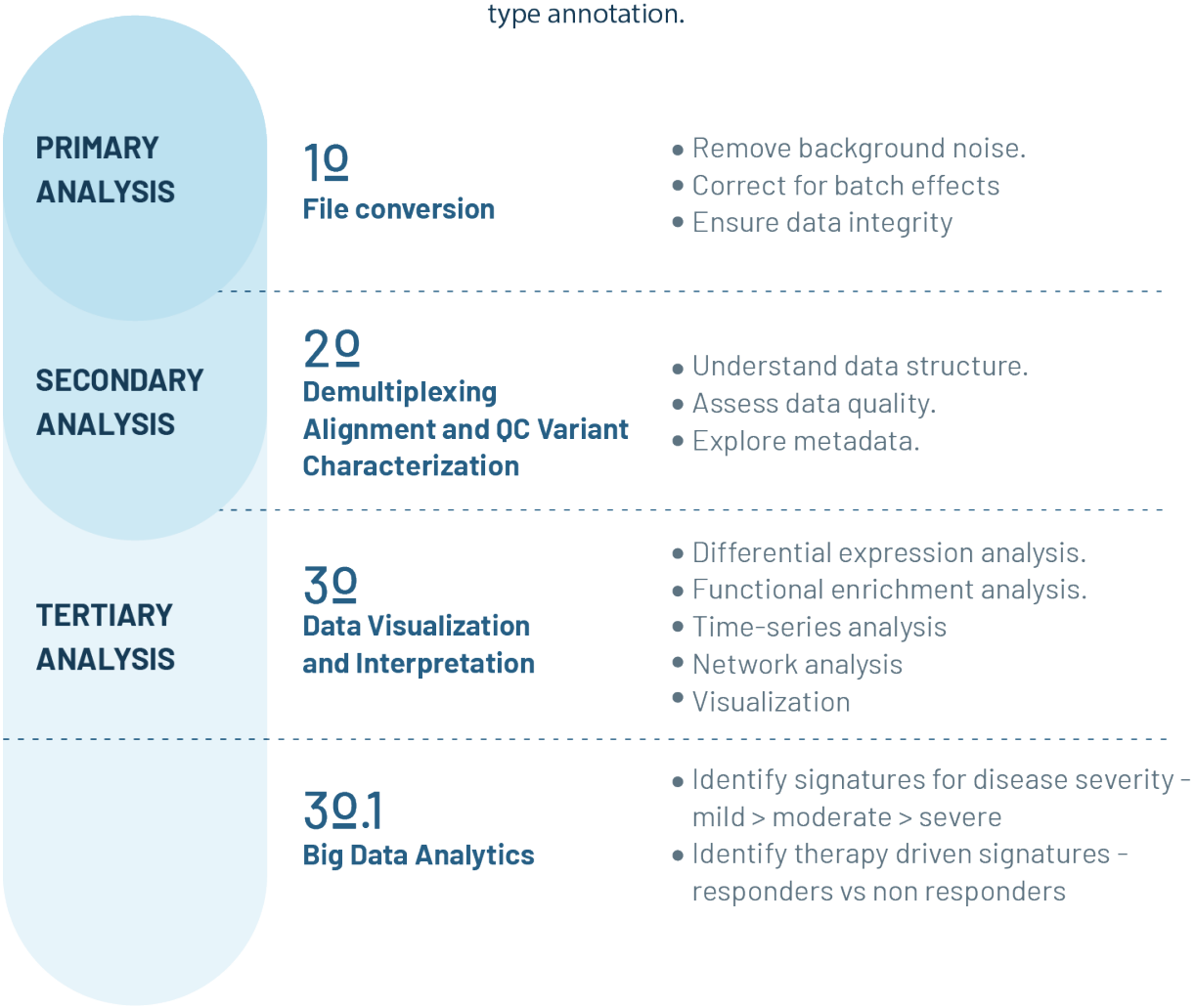


Fig. 2: Raw reads to insights: key steps in decoding transcriptome data

Strand has developed robust and custom bioinformatics pipelines tailored for a variety of RNA-Seq technologies such as bulk RNA-Seq, Slide-Seq, TCR-Seq, scRNA-Seq and stRNA-Seq including cell type annotation.



Example tools/software; In built and custom built scripts in StrandNGS, RnaSeqSampleSize (power analysis), STAR, Salmon, nf-core/rna seq, RnaSeqSampleSize, DESeq2, Topconfects, metaVolcanoR, DecoipleR, PROGENy, clusterProfiler, various R packages such as R caret, Cytoscape, etc.

Our Process

1. Our science team works with you to outline a statement of work
2. The statement of work
   - Describes the overall scope of work with an outline of the approach to use bioinformatics tools and pipelines to analyze sequencing data and assess oligonucleotide specificity and gapmer affinity, including data preprocessing, alignment, etc followed by the application of specific algorithms for identifying potential off-target effects and assessing the binding affinity of oligonucleotides.
   - Estimates costs and timelines for a go/no-go decision
   - Is at zero-cost to you
   - Takes ≈1-2 weeks to complete
3. If you choose to execute the SOW, then you receive a report with a summary of all publicly archived omics results in the space, key findings, as well as evidence specifically towards a go/no-go decisions within the timeline listed.

Recent stories

1. We have meticulously engineered software and cloud infrastructure to seamlessly combine WSI pathology images, spatial transcriptomics, cancer genomics data, and single-cell data meticulously curated from public sources, which are all housed in dedicated data lakes. This approach enables cell-type imputation powered by curated data, integration of WSIs and transcriptomics data for cell type assessment and evaluation across modalities, and a seamless assembly of patient cohorts based on criteria of interest.

   Value: Dr. Molly Heyer of Mount Sinai's Icahn School of Medicine used Curio Seeker, a spatial transcriptomics mapping kit based on Slide-seqV2, to investigate the effects of a microRNA in the cerebellum linked to schizophrenia. Strand's bioinformatics team utilized the integrated workflow in analyzing the data. The approach involved eliminating outlier cells while retaining high-quality data, which was then paired with similar data to perform differential gene expression analysis. After miRNA KO, the cerebellum's Purkinje neurons showed upregulation of 3-4 genes that influence calcium signaling and may be linked to hyperactivity and schizophrenia-like characteristics. Further information is available from Dr. Heyer's recent webinar presentation on the Curio Biosciences website.

2. For a mid-sized biotech we used the classical marker analysis method to identify significant markers in a specific inflammatory condition, by examining their transcriptome/genome changes across multiple published studies; with confidence weighted by the consistency of effects across independent experiments. Subsequently, an additional layer of annotation was added to selected genes known for their pro or anti-inflammatory characteristics and as downstream effectors of the target protein of interest. This annotated gene list, along with data on their differential expression across various pathways in control and disease samples, facilitated the refinement of marker selection. These markers were then utilized to derive a Disease Activity Score (DAS), that could inform further on patient subgroups, in terms of disease severity and responder/non-responders.
   Value: The client has collaborated with Strand to leverage our bioinformatics and curation capabilities in further investigations including their patient cohort datasets to derive predictive markers.

3. Ulcerative Colitis (UC) is a chronic inflammatory disease with a significant unmet clinical need. To enable identification of new therapeutic targets for UC, our team has harnessed, curated and performed data harmonization on scRNA-seq UC datasets using internationally recognized ontologies for defined metadata fields. The showcase currently hosts ~900k cells and 46 different cell types associated with Ulcerative Colitis. Equipped with 100 metadata fields and 27 filters, our showcase aims to enable UC researchers in navigating the complexities of publicly available datasets in a matter of seconds.
   Value: Customers get free access to the power of our scRNA data curation service, which includes comprehensive data curation tailored to your individual interests, whether they are centered on Ulcerative Colitis or any other topic.

GENOMICS BIOTECH & LARGE PHARMA

Eg- Custom pipelines

- Develop, test and implement pipelines to analyze CRISPR assays. Implementation of pipelines in
- Nextflow & Snakemake.

STRAND INDIA R&D

Omics data pipelines

- Fragmentomics analysis
- Spatial Transcriptomics data analysis
- Single cell transcriptomics data analysis
- Big genomics data ingestion, warehousing, visualization platform development

TOP-20 PHARMA

Eg- Data Analysis

- RNA-Seq data analysis to infer mechanisms causing difference in cytokine production observed in samples of interest.

STRAND INDIA CLIN.DX

NGS data pipelines

- 7 different bioinformatics pipelines supporting secondary and tertiary germline/somatic data interpretation for panels such as: RNAseq/ WXS/ WGS and targeted panels (including TSO500)

Strand Life Sciences works with marquee genetics diagnostics, sequencing instrument, pharma and biotech companies to accelerate bioinformatics and software development.

- See our website for recent Case Studies and to get in touch!

Pushing the Boundaries of Data Science: Strand's Advances in Bioinformatics

Strand's bioinformatics capabilities are supported by our team of 66 specialists who are experts in NGS-related analytics and the development of variant annotation systems. Our portfolio includes participation in over 20 projects, as well as the implementation of pipelines and secondary analytic tools that have enabled the clinical reporting of over 30,000 patient samples. These efforts have resulted in approximately ten published articles, showing our dedication to advancing our bioinformatics expertise. Additionally, our in-house StrandNGS tool, an integrated platform for the analysis, management and visualization of next-generation sequencing multi-omics data, has been cited in over 700 articles. Our diverse offerings include:

- NGS and multi-omics analytics
- Variant annotation systems
- Bioinformatics pipeline building and benchmarking
- Data curation and harmonization
- Analytical validation
- Supervised/unsupervised learning on genomic/proteomic signatures

The Strand bioinformatics team has a proven track record of success in completing several diverse projects for major US healthcare companies and in-house projects:

24+ YEARS OF EXPERIENCE

80,000+ Genetic Tests Reported

500+ Projects Executed for Genomics Majors Globally

Presence in **20+** Countries

strand

7th Floor, MSR North Tower, #144, Outer Ring Road, Nagavara, Bengaluru - 560045

+91 9980448044 hello@strandls.com strandls.com