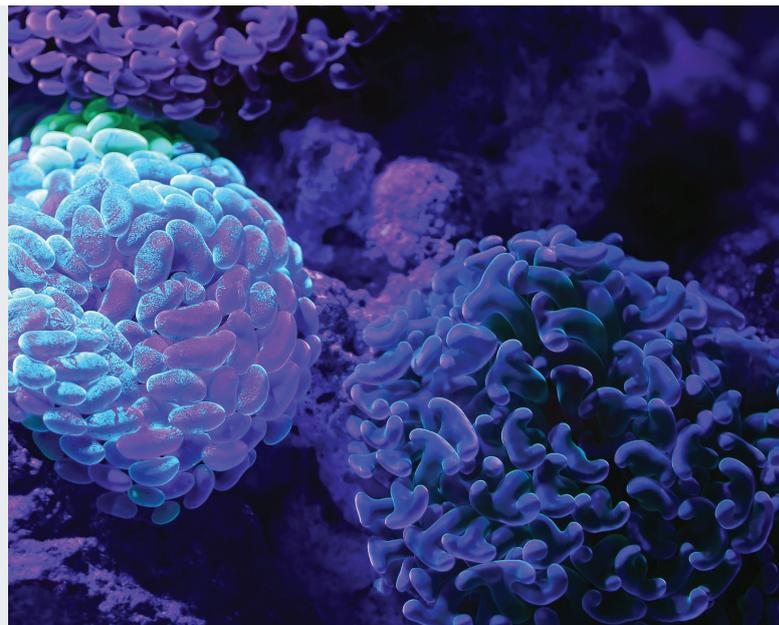


Gene Expression Analysis with KNIME Analytics Platform

Gene expression analysis is widely used in bioinformatics because it enables researchers to find gene products with increased or decreased synthesis in individuals with, for example, particular diseases. Typically, researchers find many genes that are differentially expressed in these analyses. To narrow that set of genes down to the ones of interest, scientists investigate the functional annotations of those genes and ones with similar expression patterns.



Analyzing gene expression data directly is a straightforward way to find connections between genes and diseases. The first step in gene expression is called transcription, during which DNA is transcribed to RNA. Advances in massively parallel sequencing enable the rapid sequencing of this RNA (RNA-Seq) in a genome-wide manner in order to quantify the amount of synthesized gene product. Here, RNA-Seq data from tumors and matched normal tissue from three patients with oral squamous cell carcinomas¹ are analyzed. Differentially expressed genes are discovered using the R integration in KNIME and then displayed in an interactive view. All statistically significant over/under expressed genes are investigated and interesting ones are selected by looking into their functional annotations. Using hierarchical clustering, a cluster of similarly expressed genes is selected. Their biological function is investigated through a shared component that allows researchers to perform a pathway enrichment analysis. Lastly, researchers can search for compounds that target the gene products by querying bioactivity data from Google BigQuery.

In this case, through the analysis it is possible to identify a gene that has been independently implicated as a therapeutic target for carcinomas². Moreover, it's been shown that the expression of that gene can predict disease progression as well as survival of oral squamous cell carcinoma patients³. Fig. 1 shows significantly over/under expressed genes in a scatter plot. Researchers can interactively filter and select interesting genes. Those genes are hierarchically clustered based on their expression pattern and the results are shown via a dendrogram alongside a heatmap. In addition, it's possible to perform a pathway enrichment analysis where the enriched pathways are displayed in a bar chart, making it possible to investigate compounds that target the selected gene products.

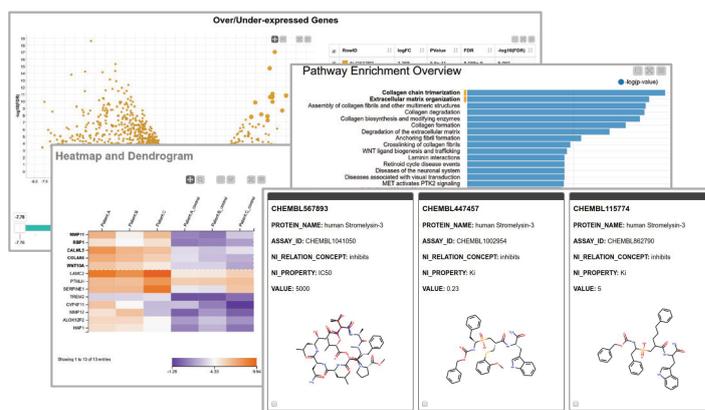


Fig. 1: Visualizations in KNIME Analytics Platform

Results:

- Detect significantly over/under expressed genes
- Find genes with similar expression patterns and common functions
- Mix and match different technologies: from your favourite R library to Google's BigQuery or using customized shared components

The open source KNIME Analytics Platform enables data scientists and researchers to mix and match tools, allowing them to create reproducible workflows in one platform. In this case, researchers can use their favorite R library, extract data from Google's BigQuery, and use shared components to customize the analysis in one reusable workflow. All steps of the analysis can also be performed on the KNIME WebPortal using the interactive views from the components.

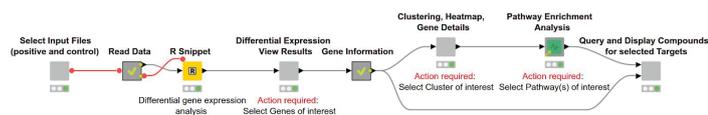


Fig. 2: Workflow to perform gene expression analysis. Differentially expressed genes are discovered using R, then genes are hierarchically clustered based on their expression pattern. Subsequently, a pathway enrichment analysis is performed, and lastly compounds targeting the gene product of interest are queried.

1. Tuch, B., Laborde, R., Xu, X., Gu, J., Chung, C., & Monighetti, C. et al. (2010). Tumor Transcriptome Sequencing Reveals Allelic Expression Imbalances Associated with Copy Number Alterations. *Plos ONE*, 5(2), e9317. doi: 10.1371/journal.pone.0009317
2. Yang, H., Jiang, P., Liu, D., Wang, H., Deng, Q., & Niu, X. et al. (2019). Matrix Metalloproteinase 11 Is a Potential Therapeutic Target in Lung Adenocarcinoma. *Molecular Therapy - Oncolytics*, 14, 82-93. doi: 10.1016/j.omto.2019.03.012
3. Hsin, C., Chou, Y., Yang, S., Su, S., Chuang, Y., Lin, S., & Lin, C. (2017). MMP-11 promoted the oral cancer migration and FAK/Src activation. *Oncotarget*, 8(20). doi: 10.18632/oncotarget.15824

