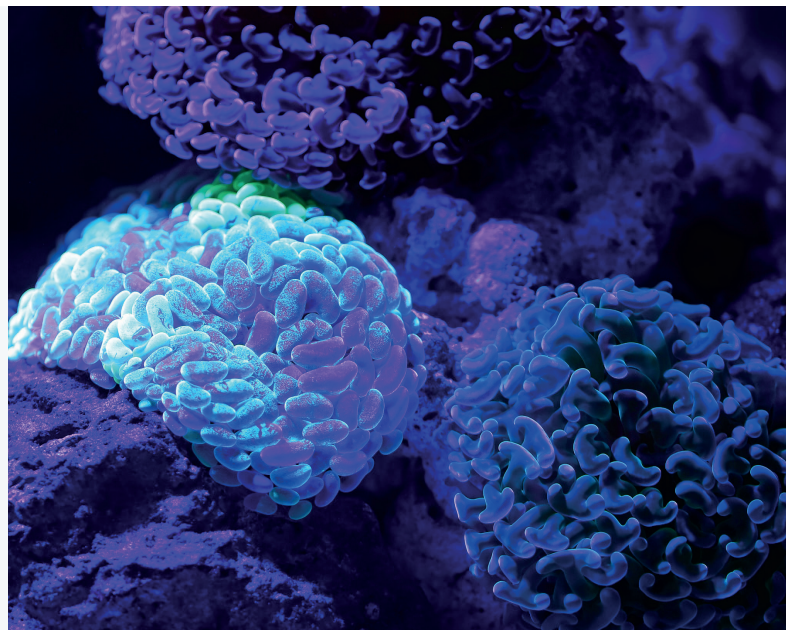


Gene expression analysis with KNIME Analytics Platform

Gene expression analysis is widely used in bioinformatics because it enables researchers to find gene products with increased or decreased synthesis in individuals with particular diseases, for example. In these analyses, researchers typically find many genes that are differentially expressed. 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.



One straightforward way to find connections between genes and diseases is analyzing gene expression data directly. The first step in gene expression is called transcription, during which DNA is transcribed to RNA. The 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, we see an analysis of RNA-Seq data from tumors and matched normal tissue from three patients with oral squamous cell carcinomas¹.

In KNIME, differentially expressed genes are discovered using the R integration, and then displayed in an interactive view. All statistically significant overexpressed and underexpressed genes are investigated; interesting ones are selected by looking into their functional annotations. A cluster of similarly expressed genes is selected using hierarchical clustering. A shared component allows researchers to perform a pathway enrichment analysis to investigate the cluster's biological function. Lastly, researchers can search for compounds that target the gene products by querying bioactivity data from Google's BigQuery.

In this case, the analysis makes it possible to identify a gene that's been independently implicated as a therapeutic target for carcinoma². 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 overexpressed and underexpressed genes in a scatter plot. Researchers can interactively filter and select interesting genes. Those genes are hierarchically clustered based on their expression pattern; 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.

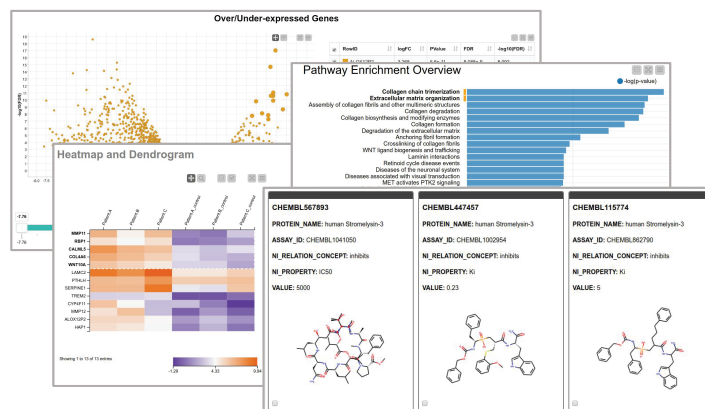


Fig.1: Visualizations in KNIME Analytics Platform

Results:

- Detect significantly overexpressed and underexpressed 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 Business Hub using the web-based application and interactive views from the components.

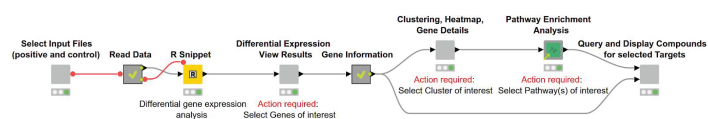


Fig.2: Workflow to perform gene expression analysis. Differentially expressed genes are discovered using R. Then, genes are hierarchically clustered based on their expression pattern. A pathway enrichment analysis is performed, and lastly, compounds targeting the gene product of interest are queried.

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 Metalloprotease 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

