

RNAseq Interaction Chat - Novaflow

Demo 9/19/2025



how would i analyze this file?



To analyze the provided `.fasta` file, you can use various bioinformatics tools and techniques. Given the experimental context mentioning `htseq-count` input, it seems like you might be interested in quantifying the abundance of transcripts in your sample, which is a common step in RNA sequencing (RNA-seq) analysis.

Here's a general approach to analyze your `.fasta` file:

1. **Quality Control:** First, ensure the quality of your sequences. Tools like `FastQC` can help you visualize the quality of your sequences.
2. **Assembly or Mapping:** Since your file contains transcript sequences, you might want to map these to a reference genome or transcriptome. For this, you can use tools like `HISAT2`, `STAR`, or `Bowtie2` for mapping to a genome, or `Salmon` or `Kallisto` for pseudo-mapping to a transcriptome.
3. **Quantification:** Once you have the mapped reads, you can use `htseq-count` as mentioned in your experimental context to count the number of reads mapping to each gene or transcript. This will give you a count matrix that can be used for differential gene expression analysis.
4. **Differential Expression Analysis:** With your count matrix, you can proceed to analyze differential gene expression using packages like `DESeq2` or `edgeR` in R. These packages can help identify which genes are differentially expressed between your conditions.
5. **Downstream Analysis:** After identifying differentially expressed genes, you might want to perform functional enrichment analysis to understand the biological processes affected. Tools like `GOseq` or `gseapy` can help with this.

Remember, the specific steps and tools you choose will depend on your experimental design, the quality of your data, and your research questions.

If you're unsure about any of these steps or need more detailed instructions, it would be helpful to have more information about your experimental design and goals.



what are my options for visualizing this data? how could i make it interactive?



To visualize your `.fasta` file data, which contains transcript sequences, and make it interactive, you have several options depending on what aspects of the data you want to highlight. Given the experimental context of RNA sequencing (RNA-seq) and the analysis steps previously discussed (quality control, assembly or mapping, quantification, differential expression analysis, and downstream analysis), here are some visualization and interactivity options:

1. Sequence Alignment and Genome Browsers

- Tools like **IGV** (Integrative Genomics Viewer), **UCSC Genome Browser**, or **JBrowse** can be used to visualize the alignment of your transcript sequences against a reference genome. These tools offer interactive views where you can see the density of reads, coverage, and even zoom into specific regions of interest.

2. Heatmaps and Expression Profiles

- **Heatmaps** can be particularly useful for showing the expression levels of genes across different samples or conditions. Tools like `pheatmap` or `Heatmap` packages in R can be used to create heatmaps.
- For interactive heatmaps, consider using **Clustergrams** in tools like Morpheus (from Broad Institute) or online platforms that support interactive heatmaps.

3. Pathway and Network Analysis Visualizations

- **Cytoscape** is a powerful tool for visualizing and analyzing molecular interaction networks. It can help in understanding the functional relationships between genes and proteins of interest.
- **Pathview** (in R or as a webtool) allows you to map gene expression data onto pathways, providing an interactive way to understand how gene expression changes impact biological pathways.

4. Volcano Plots and MA Plots

- **Volcano plots** are useful for visualizing differential expression analysis results, showing the significance versus the fold change of genes.
- **MA plots** are similar but specifically show the average expression level against the log fold change, both on a logarithmic scale.

5. Gene Set Enrichment Analysis (GSEA)

- Tools like **GSEA** or `gseapy` provide a way to identify biological pathways that are differentially expressed. Visualizations include enrichment plots showing the distribution of genes within a pathway.

6. Interactive Web-Based Tools

- **Shiny apps in R** can be a versatile way to create interactive dashboards that allow users to explore different aspects of the data without needing to write code.
- **Bokeh or Plotly** libraries in Python offer interactive visualizations that can be embedded into web applications or notebooks, enabling zooming, hovering for more information, and other interactive features.

Making It Interactive

To make your visualizations interactive, consider:

- **Web-based platforms** like Shiny (for R) or Dash (for Python) that allow you to build interactive dashboards.
- **Jupyter Notebooks** with interactive visualization tools like Bokeh or Plotly for Python.
- **RStudio** with packages like `shiny` or `crosstalk` for interactive visualizations.
- **Bioconductor packages** in R, many of which include visualization tools designed specifically for genomic and transcriptomic data.

When choosing a method, consider the complexity of your data, your audience, and the story you want to tell with your data. Interactivity can greatly enhance the exploration and understanding of your RNA-seq data, especially when dealing with complex datasets and seeking to identify patterns or correlations that are not immediately apparent from static images.



could you please explain whatthis file is, how it relates to my previous file, and how it fits into the overall rnaseq pipeline you described?



The file you've mentioned is in the `.gtf` (General Feature Format) format. This format is used to describe genes and other features on a genome. It's a tab-delimited format that provides detailed information about the structure of genes, including their locations, exon/intron boundaries, and other relevant details.

In the context of your RNA-seq analysis, this `.gtf` file plays a crucial role. It serves as a reference annotation file that helps in the quantification of gene expression levels. The information contained within the `.gtf` file is essential for accurately mapping RNA-seq reads to specific genes or transcripts, thereby enabling the quantification of gene expression.

To understand how this `.gtf` file relates to your previous `.fasta` file and fits into the overall RNA-seq pipeline, let's consider the steps involved in the pipeline:

- 1. Quality Control:** Your initial `.fasta` file, which contains transcript sequences, undergoes quality control checks to ensure the sequences are of good quality.
- 2. Mapping:** After quality control, the next step is typically to map the transcript sequences against a reference genome or transcriptome. For this, tools like `HISAT2`, `STAR`, or `Bowtie2` can be used for mapping to a genome. The `.gtf` file comes into play here as it provides the necessary annotation to accurately map reads to genes or transcripts.
- 3. Quantification:** Once the reads are mapped, the `.gtf` file is used to guide the quantification of gene expression. Tools like `htseq-count` use the gene models defined in the `.gtf` file to count the number of reads that map to each gene, thereby quantifying gene expression.

Given the content of your `.gtf` file, which includes information like `gene_id`, `gene_name`, and `gene_biotype`, it's clear that this file is essential for annotating the genome with the locations and structures of genes. This annotation is critical for the accurate assignment of reads to specific genes during the quantification step.

In summary, the `.gtf` file is a reference annotation file that is crucial for the quantification of gene expression in RNA-seq analysis. It provides the necessary information to map reads to specific genes or transcripts accurately, and it plays a central role in the RNA-seq pipeline, working in conjunction with the transcript sequences provided in your initial `.fasta` file.



okay basically what im trying to experiment is how upregulation or downregulation of HER2 changes the likelihood of developing breast cancer. can you give me more information on the pathways related to this and possible experiments?



To address your question, let's break down the relationship between HER2 (human epidermal growth factor receptor 2) upregulation or downregulation and its impact on breast cancer development, considering the pathways involved and potential experiments.

Background on HER2 and Breast Cancer

HER2 is a protein that promotes the growth of cells. In some cases of breast cancer, notably about 20% of cases, the HER2 gene is amplified, leading to overexpression of the HER2 protein on the cell surface. This overexpression can lead to increased cell proliferation and survival, contributing to the development and aggressiveness of breast cancer.

Pathways Related to HER2 in Breast Cancer

Several pathways are implicated in the signaling downstream of HER2, including:

1. **PI3K/AKT Pathway:** This pathway is crucial for cell survival and proliferation. Activation of HER2 can lead to the activation of the PI3K/AKT pathway, promoting cell survival and resistance to apoptosis.
2. **MAPK/ERK Pathway:** This pathway is involved in cell division and differentiation. HER2 signaling can activate the MAPK/ERK pathway, contributing to increased cell proliferation.
3. **Cell Cycle Regulation:** HER2 overexpression can lead to dysregulation of cell cycle checkpoints, allowing for unchecked cell proliferation.

Possible Experiments to Study HER2's Role in Breast Cancer

1. **In Vitro Studies:**
 - o **Cell Line Models:** Utilize breast cancer cell lines with varying levels of HER2 expression to study the effects of HER2 upregulation or downregulation on cell proliferation, apoptosis, and migration.

- **RNA Interference (RNAi) or CRISPR-Cas9:** Use these technologies to specifically knockdown or knockout HER2 in cell lines to observe the impact on cellular behavior and signaling pathways.
- **Drug Treatment Studies:** Treat cells with HER2-targeting drugs (e.g., trastuzumab, lapatinib) to observe the effects on cell viability and signaling pathways.

2. In Vivo Studies:

- **Xenograft Models:** Inject breast cancer cells with manipulated HER2 expression into immunocompromised mice to study tumor growth and response to HER2-targeting therapies.
- **Genetically Engineered Mouse Models:** Use mice with conditional HER2 overexpression or knockout in the mammary gland to study the development of breast cancer.

3. Bioinformatics and Systems Biology Approaches:

- **Pathway Analysis:** Utilize tools like Cytoscape or Pathview to analyze the expression data of HER2 and related genes to understand the activation status of relevant pathways in breast cancer samples.
- **Network Analysis:** Investigate the protein-protein interaction networks centered around HER2 to identify potential therapeutic targets or biomarkers.

4. Clinical Samples and Patient Data:

- **Immunohistochemistry (IHC) and Fluorescence In Situ Hybridization (FISH):** Analyze breast cancer tissues for HER2 expression and gene amplification status.
- **Expression Profiling:** Use microarray or RNA-seq data from clinical samples to assess the expression levels of HER2 and related genes, correlating with clinical outcomes.

Utilizing the Provided .gtf File

The .gtf file you have contains annotations for genes, including their locations, exon structures, and biotypes. This file can be crucial for:

- **Mapping RNA-seq Reads:** To quantify the expression of HER2 and other genes involved in the relevant pathways.
- **Identifying Differentially Expressed Genes:** Between samples with different HER2 expression levels or between breast cancer subtypes.
- **Pathway Enrichment Analysis:** Using tools like GSEA or Pathview to identify pathways that are enriched with differentially expressed genes, potentially highlighting the involvement of specific biological processes in HER2-related breast cancer.

In summary, studying the impact of HER2 upregulation or downregulation on breast cancer involves a combination of in vitro, in vivo, and bioinformatics approaches. The .gtf file can serve as a valuable resource for annotating and quantifying gene expression changes related to HER2 signaling and its downstream effects.



i see, and what pieces of context did you get the information about HER2 from? answer honestly.



To answer your question honestly, I must clarify that the information about HER2 and its relationship to breast cancer, including the pathways and potential experiments, was not directly obtained from the provided .gtf file or the experimental context you've described.

The .gtf file contains annotations for genes, including their locations, structures, and biotypes, which can be used for mapping RNA-seq reads, identifying differentially expressed genes, and pathway enrichment analysis. However, it does not provide specific information about HER2 or its role in breast cancer.

The information about HER2 and its implications in breast cancer was drawn from general knowledge and understanding of the biological and medical literature regarding HER2's role in cancer, particularly breast cancer. This includes its overexpression, the signaling pathways it's involved in (such as PI3K/AKT and MAPK/ERK), and how it affects cell proliferation and survival. The experimental approaches mentioned (in vitro studies, in vivo models, bioinformatics analyses) are standard methods used in cancer research to study the function of specific genes and their potential as therapeutic targets.

Therefore, while the .gtf file and the context of RNA-seq analysis and gene expression quantification are crucial for studying gene function and regulation, including potentially HER2, the specific details about HER2's role in breast cancer were not derived from the .gtf file content or the experimental context you provided.

Please upload a file if you wish



Drag and drop file here

Limit 200MB per file

Browse files



test.gtf 0.6KB



what about the biological context provided? where is that information?



The biological context provided is empty. According to the given information, the "Biological Context" section is blank, indicating that no specific biological context was provided. The information about HER2 and its implications in breast cancer was drawn from general knowledge and understanding of the biological and medical literature, not from the provided biological context.

Ask your RNA-seq question... 