Shayla Ghose

Thursday July 10th, 2025

**RNAseq:**

* Analyzing transcriptome, which is the complete set of RNA molecules in a cell.
* Technique used to analyze the quantity of RNA molecules in a biological sample.
* Provides insights into gene expression patterns and cellular processes.
* Allows researchers to see which genes are expressed, and when.
* The abundance of RNA sequences indicates the level of gene expression.

Process of RNAseq:

1. RNA Extraction: RNA extracted from biological samples (cells, tissues).
2. RNA Fragmentation: Extracted RNA is broken into smaller pieces to make it suitable for sequencing.
3. Synthesizing cDNA: Fragmented RNA converted to cDNA using reverse transcriptase, because sequencing technologies are used to analyze cDNA, not RNA directly.
4. Library Preparation: Sequencing adapters are attached to cDNA, creating a sequenced library, allowing cDNA fragments to be amplified and sequenced.
5. Sequencing: Prepared cDNA library sequenced using a high throughput sequencing platform, generating many short DNA sequences called “reads”.
6. Data Analysis: Reads are matched back with a transcriptome (reference sequence), determining what gene expression levels are present and how much.
7. Bioinformatics is used to identify DEGs and provide other insights.

**Microarray:**

* Technique used to study expression of thousands of genes simultaneously.
* Uses microscopic slides (microarrays) printed with known DNA sequences or genes.
* Sequences act as probes (fragments of DNA or RNA to detect the presence of a specific DNA fragment within a sample) to detect presence and abundance of specific RNA molecules (mRNA) in a sample.
* Used to gain insights into cellular processes, disease mechanisms, and potential drug targets.

Process of microarray analysis:

1. Sample Preparation: RNA extracted from cells, representing actively expressed genes, and is converted into cDNA using reverse transcriptase.
2. Fluorescent dyes: cDNA is labeled with fluorescent dyes with different colors to distinguish between control vs treated samples.
3. Hybridization: A microarray chip (glass slide) is spotted with probes (known DNA sequences), and the cDNA samples bind to their corresponding probes on the chip forming hybrids.
4. Microarray Scanner: Scanner is used to detect fluorescence intensity at each probe location, capturing images of the chip.
5. Image Analysis: Computer stores and analyzes images and data, measuring intensity of fluorescent signals.
6. Data Analysis: Fluorescence intensity data corrected for background noise, data is normalized to account for variations, and statistical methods applied to identify genes showing differences in expression.

Differences between RNAseq and Microarray Analysis:

* RNA-seq uses next generation sequencing to identify and quantify RNA molecules in a sample, while microarrays rely on hybridization of labeled cDNA to probes on a chip.
* RNA-seq is more comprehensive of the transcriptome, while microarrays rely on pre-defined probes to detect known sequences.
* RNA-seq has higher sensitivity and a wider dynamic range, but microarrays are a more established and cost-effective option for measuring expression of known transcripts.

**Differential Gene Expression Analysis:**

Differential gene expression analysis is a method used to identify differences in gene expression levels between two or more sample groups, such as healthy vs diseased tissue. This helps researchers find genes which are upregulated (expressed at higher levels) or downregulated (expressed at lower levels) depending on the conditions. Differential gene expression refers to differences in how much mRNA is produced by specific genes in different conditions (or between different cell types), and can be due to factors including gene regulation or responses to stimuli. Comparing gene expression profiles in healthy and diseased tissues allows researchers to identify genes that are involved with diseases. Biomarkers can be revealed from DGE analysis, leading to development of new drugs or therapies. This process is done through extracting data, preprocessing it by aligning it to a reference genome (entire set of DNA instructions found in a cell), and quantifying it to determine how much of each gene transcript there is. Tools like limma and DESeq2 can compare gene expression levels and identify DEGs, which are interpreted to understand their significance. Researchers use this process to gain valuable insights into biological processes and disease mechanisms.

**Biomarkers:**

* Measurable indicators of biological states or processes.
* Serve to understand what is happening in the body at a molecular level.
* Assess normal biological functions and disease processes, giving insights into a person’s health.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Type: | Diagnostic | Prognostic | Predictive | Pharmacodynamic | Monitoring | Safety |
| Definition: | Identifies presence or absence of a disease | Indicates likelihood of a disease worsening | Determine if a patient is likely to respond to a treatment | Measures effect of a drug on the body | Tracks progress of a disease or response to treatment | Assess potential side effects of a drug or treatment |
| Example: | Measuring levels of proteins in blood to diagnose a heart attack | Certain gene mutations predict aggressiveness of cancer | Testing for a specific genetic mutation that makes a tumor sensitive | Measuring level of a drug in blood or its effect on a biological pathway | Regularly measuring blood glucose levels in diabetic patients to monitor condition | Monitoring liver enzymes to detect drug induced damage |

|  |  |
| --- | --- |
| RNAseq | Microarray |
| -Less cost-effective | -More cost-effective |
| -Uses high-throughput platforms like Illumina and Pacific Biosciences | -Uses platforms that involve binding cDNA fragments to a chip, including Affymetrix microarrays or Illumina arrays |
| -Uses novel RNA and RNA variants as samples | -Uses RNA probes as samples |
| -Sequencing based | -Relies on hybridization |
| -More comprehensive and accurate view of transcriptome | -Less comprehensive and accurate view of transcriptome |

**R Packages**

**Limma:**

* R package for the analysis of gene expression data
* Can handle complex experimental designs and uses linear models
* Made by Professor Gordon Smyth along with contributions from a team of researchers
* First made available on December 2, 2002, latest contribution on April 14, 2025

**DESeq2:**

* Bioinformatics software package analyzing RNAseq data
* Identifies differentially expressed genes between different experimental conditions
* Implemented in R programming language
* Primarily developed by Michael Love, Wolfgang Huber, and Simon Anders along with other key contributors
* Initial release on March 22, 2013