TITLE: RNA Sequencing

PRESENTER: Lauren Sanders

Slide 1:

Hello, my name is Lauren Sanders. I am a PhD candidate at the University of California Santa Cruz Genomics Institute. Welcome to this Pearl of Laboratory Medicine on "RNA Sequencing".

Slide 2:

RNA sequencing, or RNA-seq, is a method for high throughput sequencing of RNA molecules extracted from cells or tissue. This method measures RNA transcript nucleotide sequences, and quantifies abundance of each unique transcript. RNA-seq has the ability to detect multiple RNA types, including messenger RNA from coding genes, long non-coding RNA or lncRNA, microRNA and transfer RNA. RNA-seq has several clinical applications including detection of disease-relevant aberrant gene expression, gene fusions, and expressed sequence variants.

Slide 3:

An RNA-seq workflow starts by selecting for the RNA of interest from total RNA in the cell. The selected RNA is subjected to reverse transcription to produce complementary DNA or cDNA, which is then sequenced. One of the most common sequencing platforms in the clinical setting is Illumina sequencing. The resulting sequence reads are aligned to a genome assembly. After read alignment, downstream analyses quantify transcript abundance, detect fusions or unusual splice isoforms, and expressed sequence variants.

Slide 4:

There are several methods for selecting RNA of interest from total RNA in the cell. Poly-A selection captures mainly poly-adenylated transcripts. Ribo-deplete selection removes

ribosomal RNA which constitutes approximately 80% of RNA in the cell. This method leaves behind all non-ribosomal RNA which can contain both polyA and non-polyA transcripts. Targeted capture methods select specific RNA transcripts using hybridization to complementary probes. Finally, size selection methods can enrich for specific transcript lengths such as small RNA species. Because all of these methods sequence different RNA species from the cell, RNA-seq data from different methods cannot be directly compared without additional normalization steps.

Slide 5:

I will now discuss 4 main clinical applications for RNA sequencing: gene and exon expression quantification can identify aberrant expression of clinically relevant genes. Gene fusion detection can identify diagnostic fusions. Detection of unusual splice isoforms can identify novel or disease-relevant splicing events. Calling variants using RNA-seq can identify variants which are expressed in the RNA.

Slide 6:

There are two main RNA mapping strategies. Genome alignment involves aligning reads to an assembly of the entire genome. This method detects exonic and intronic reads. Exome alignment involves aligning reads to the exome only. With exome alignment, we can identify exon junction reads, but with genome alignment these reads are lost because they do not map to the genome assembly.

Slide 7:

RNA-seq can be used for fusion detection. Most RNA-seq in the clinic is paired-end sequencing, meaning that reads are measured from both ends of each RNA fragment. There are two ways of identifying RNA fusions: split reads and spanning reads. Split reads are reads which span exons from two different genes. Spanning reads are reads which map to exons from different reads but come from the same RNA fragment, indicating a fusion.

Slide 8:

RNA-seq can be used for detection of alternative splice isoforms by identifying reads which span exons which are normally not spliced together.

Slide 9:

RNA-seq detects sequence variants which are expressed in the RNA transcripts. These variants can arise in two different ways. Some variants occur through DNA mutation events in the genome DNA and can then be expressed through RNA transcription. However, other variants are not present in the genomic DNA because they occur through post-transcriptional RNA

editing events. RNA-seq is critical for detecting these variants which would be missed in whole genome or whole exome sequencing.

Slide 10:

All of the applications I have just discussed have applications in genomic medicine. Transcription quantification can be used in comparative gene expression analysis, for example comparing gene expression of tumor tissue to normal tissue. RNA-seq can detect diagnostic fusions, disease-specific alternative splicing, and disease-associated DNA or RNA variants.

Slide 11:

Expression quantification can be used for comparative gene expression analysis to identify over-expressed druggable genes in a diseased tissue. For example, in this 2018 case study, comparative gene expression analysis was used to identify *JAK1* as significantly over-expressed in RNA-seq data from a pediatric sarcoma lung metastasis, compared to other lung cancers, other sarcomas, or all other cancers in the background cohort. Administration of Ruxilitinib, a *JAK1* inhibitor, had clear benefits for the patient, demonstrating that comparative gene expression analysis can identify tumor driver genes.

Slide 12:

Several diseases harbor diagnostic gene fusions that can be detected through RNA-seq. For example, the *EWSR1-FLI1* fusion involves the fusion of the *FLI1* gene on chromosome 11 to the *EWSR1* gene on chromosome 22 and is one of several diagnostic *EWSR1* fusions for Ewing sarcoma.

Slide 13:

In addition to the applications of RNA-seq in genomic medicine that I have discussed, there are emerging applications for RNA-seq in diagnosis of rare germline disorders, including the detection of extreme gene expression, identification of mono-allelic expression, and aberrant splicing defects.

Slide 14:

Gene expression above physiological levels can cause rare germline disorders. Unusual gene expression can be caused by epigenetic dysregulation, or variants in promoter or enhancer regions. RNA-seq can be used to identify expression outliers by defining a cutoff for normal physiological gene expression using a Z-score or outlier threshold in a larger normal dataset.

Slide 15:

Mono-allelic expression can also cause rare germline disorders and be detected by RNA-seq. For example, thrombocytopenia with absent radii or TAR is a rare germline disorder characterized by a significant reduction of platelet number and the absence of radius bone in the forearm. Dr. Albers and colleagues found that TAR is caused by compound heterozygous

variants in the *RBM8A* gene, where one allele is deleted and the other allele displays reduced expression due to a non-coding variant. Allele-specific expression can be detected by RNA-seq while it would not be detectable using whole genome or whole exome sequencing.

Slide 16:

Alternative or aberrant splicing can also cause rare germline disorders and be detected by RNA-seq. For example, familial partial lipodystrophy type 2 or FPLD2 is a rare germline disorder characterized by abnormal distribution of adipose tissue in the body, and it is caused by mutations in the *LMNA* gene which codes for Lamin A. For example, Dr. Morel and colleagues found that a C to G mutation in the *LMNA* intron 8 consensus splice donor site causes aberrant splicing where intron 8 is retained. Because of a premature stop codon in intron 8, translation stops short of Exon 9 and results in a truncated Lamin A protein.

Slide 17:

In conclusion, RNA sequencing is a very promising technology with many applications in the clinic, including applications in genomic medicine and rare germline disorder identification and diagnosis. RNA sequencing and analysis is a developing field, and we look forward to new applications in clinical laboratory medicine and patient care as innovation progresses.

Slide 18: References

Slide 19: Disclosures

Slide 20: Thank You from www.TraineeCouncil.org

Thank you for joining me on this Pearl of Laboratory Medicine on "RNA Sequencing".