MNG LABORATORIES
A LabCorp Company

Clinical Indications for Transcriptome Testing

All cells in the human body have the same exact gene content, with the exception of a few specialized cell types such as germ cells, T-cells, B-cells and red blood cells. Yet, cells are varied in their functions and their appearance. The reason for this variability is that different sets of genes are expressed or transcribed in each cell type based on its specific structural and functional requirements. The comprehensive cell-type specific set of transcripts is referred to as the transcriptome. The transcriptome is a set of RNA molecules that carry the genetic information from the DNA in the cell nucleus to the cytoplasm. In the cytoplasm, the RNA transcripts are used as the blueprint for the synthesis of proteins, the building blocks of the cells and tissues. The transcriptome is the summation of all genetic and epigenetic (not resulting in sequence alteration) changes that can occur in the DNA of the cell. For this reason, the analysis of the transcriptome allows assessment of the significance of genetic and epigenetic changes revealed by DNA sequencing or implied from phenotype. This is true not just for single nucleotide variants, SNV, but also copy number variants, CNVs, repeat expansions or imprinting defects due to genetic causes or epimutations. Until we acquire the ability to predict the functional outcomes of genomic and epigenetic changes based on DNA sequencing data alone, the transcriptome is a powerful tool to establish the clinical significance of the ever-growing list of variants of uncertain significance (VUS). In the following paragraphs, we describe the most important clinical indications for the use of transcriptome studies.

VUS in Regulatory Regions

Enhancer and promoter VUS: These will change the level and type of transcripts present in the transcriptome.

Genomic translocations, inversions, duplications point mutations and repeat expansions can place regulatory regions in a new context changing levels of transcripts for specific genes and changing the level of corresponding proteins in the cell. These changes can be demonstrated clearly and easily using transcriptome testing, while testing protein levels is technically more challenging or requires a targeted approach.

Clinical indications: Traditional cytogenetic studies or microarray (CMA) testing can reveal changes in the structure of the patient's genome that juxtapose regions in the genome not normally adjacent to each other. These changes can alter the regulatory context of genes resulting in over or under expression of genes or expressions of genes that are normally not expressed in a given tissue. The transcriptome is a great tool to assess the consequences of novel rearrangements detected by cytogenetic studies or CMA. The best examples for this kind of mutations are the juxtaposition of immunoglobulin and T-cell receptor promoters to regulatory factors directing cell differentiation in various hematological malignancies. Alternatively, repeat expansions and point mutations in the regulatory regions can alter the transcriptional accessibility of genes, as is seen in Fragile X syndrome. Short of methylation analysis or targeted repeat expansion studies these mutations that would be missed with sequencing alone but can be readily demonstrated using transcriptome studies. Transcriptome can also demonstrate X-chromosome silencing bias in carrier females of such mutations.

5' and 3' UTR (untranslated region) variants: Variants in the regions of genes that are copied into transcripts, but fulfill regulatory role for the processing and stability of the transcript made, rather than being directly used to build proteins.

Transcripts contain regions that are not translated, meaning they are not used for protein synthesis directly. These regions are referred to as untranslated regions, UTRs. Depending on whether they are at the beginning or the end of the transcript, we call them 5'UTRs or 3'UTRs. These regions are responsible for the ability of the transcripts to engage the protein synthesis machinery in the cytoplasm and also to regulate the stability and turnover of the transcripts. Analyzing the transcriptome allows detection of the transcript level changes of genes carrying variants in these regions.

Clinical indications: Genomic sequencing can identify SNVs and CNV in the 5' and 3' UTRs that can potentially affect the structure and stability of transcripts and thus their ability to direct protein synthesis over time. Although there are known examples where such changes severely alter the transcript level of genes, as for example in the case of 3'UTR mutations in the prothrombin gene, there are no good in silico methods to predict the effects of novel UTR variants without performing a transcriptome study.

Mutations affecting splicing: SNVs and CNV in exon/intron boundaries and deep within introns or exons might affect the level and stability of transcripts, as well as the ability of the transcript to direct the synthesis of functional proteins.

Human genes have large segments (introns), that are copied into RNA initially, but are removed from the transcripts during the process of splicing. A gene can contain dozens or even hundreds of introns. Splicing is a complex process controlled by large multiprotein complexes as well as variably conserved RNA sequences adjacent to the positions in the transcript where the cutting and pasting of the transcript takes place (splice site variants) and also deep in the interior of the regions removed (deep intronic variants). Variants in these regions can result in cutting the transcript at the wrong position, or splicing the wrong fragments together resulting in a poorly functioning or nonfunctional protein. Mis-spliced messages are often completely degraded resulting in loss of protein synthesis from the affected gene. Transcriptome analysis can detect the loss of splicing event that have been described previously for a given gene, and can also detect

Clinical Indications for Transcriptome Testing

unusual, novel splicing events causing productions of proteins with compromised function. Changes in the level of transcripts can detect degradation of transcripts that are not properly spliced.

Clinical indications: Genomic sequencing can identify SNVs and CNV in the regions of genes that can potentially affect the splicing and thus the functionality and stability of transcripts. Although variants immediately adjacent to the splice sites are generally known to have deleterious effects, changes that are within the proximity of such sites or even variants deep in exons or introns can affect splicing in ways that can only be ascertained using transcriptome studies. The examples for this type of variants are too numerous to list, but the case of SpinoMuscular Atrophy (SMA) is a fine example. Transcriptome can easily show the level and ratio of correctly spliced *SMN* gene product versus the incorrectly spliced nonfunctional product irrespective of whether it originated from the *SMN1* or *SMN2* gene.
Implication of a second deep intronic or exonic variant affecting splicing becomes especially important when in a recessive gene matching the patient's phenotype only one disruptive mutation is found. Transcriptome in such cases can demonstrate the presence of not one, but two affected alleles resulting in two mis-spliced transcripts, or the absence of a normally spliced transcript if the second mis-spliced message is unstable and is destroyed.

VUS in Coding Regions

Large genomic deletions or duplications: These changes can lead to complete loss or duplication of large segments or sometimes entire chromosomes encompassing an entire gene or many genes.

These can result in decrease or increase in transcript levels leading to functionally unacceptable drop in protein levels, potentially even to complete loss of transcripts in the case of homozygous deletions or loss of expressed alleles in imprinting disorders.

Clinical indications: Traditional cytogenetic studies, CMA or sequencing studies can identify large deletions or duplications affecting multiple genes. Many genes have fixed expression levels and thus changes in copy number are associated with changes in the level of transcript present. This is especially true for genes that are "imprinted", which means they are expressed only on chromosomes derived from one or the other parent. Genomic information does not reveal whether it is a paternal or a maternal chromosome segment is lost, but transcriptome studies can clearly identify the loss of expression from an imprinted locus. A good example for this is the ability to diagnose Prader-Willi Syndrome using transcriptome analysis. However, transcriptome is not only useful for imprinted genes, but also for other genes that when expressed only from a single chromosome cannot produce sufficient transcript and thus sufficient protein levels for the functioning of the cell.

Intragenic deletions or duplications: These are deletions or duplications within the confines of a single gene.

CNVs detected within coding regions of a gene, besides the loss of information for the synthesis of a given segment of a protein, can result in changes affecting the entire gene/transcript. This is due to the fact that the genetic code consists of three letters and thus insertions or deletions that are not exactly multiples of three can alter the reading frame of the message. This can lead to early termination of the protein synthesis or nonfunctional protein sequences or a complete removal of the resulting transcript through a process called nonsense mediated decay (NMD). The cell might attempt to overcome these effects using alternative transcription initiation and termination sites or altered splicing patterns including exon skipping which can result in reestablishment of the reading frame of the transcript and thus restore at least in part protein production. These complex outcomes cannot be predicted from the DNA sequence changes alone but can be precisely detected and measured using the transcriptome.

Clinical indications: DNA sequencing and CMA analysis can detect loss of the genetic information extending over multiple coding as well as noncoding exons, but fails to inform about the functional consequences of such changes. The best examples for this kind of situation is deletions in the *DMD* gene causing Duchenne and Becker Muscular Dystrophy. However, this type of mutation is common and thus transcriptome analysis can shed light on the exact functional consequences of such mutations any protein with a precision and accuracy currently not available from the genomic data alone.

Summary & Future Considerations

Transcriptome studies are relatively inexpensive and provide a wealth of information about the consequences of a wide array of genetic and epigenetic changes including insertions, deletions, CNVs, SNVs, repeat expansions as well as imprinting mistakes. With the development of splicing modifiers, transcriptome studies can also have importance in identification of patients who would benefit from such interventions. To accommodate the many types of diagnostic situations, MNG provides targeted as well as global transcriptome analysis. The lability of RNA and tissue specific nature of the transcriptome makes it imperative that before a plan for a transcriptome study is formulated, the ordering healthcare provider contacts MNG to select the best diagnostic approach to follow.

Peter L. Nagy, MD, PhD
Chief Medical Officer
MNG Laboratories