TRG Clonality Interrogation by CE and NGS: Bridging the Gap Between Classical and Leading Edge Technologies

K. HUTT[1], A. ARIAS[1], E. FINNEGAN[1], J. PANGANIBAN[1], D. VISSERS[1], N. KHA[2], I. DANNER[2], D. MEDINA[2], D. CAGUIOA[2], J. LASSITER[2], Y. HUANG[1]

1 Invivoscribe, Inc., 10222 Barnes Canyon Road, Building 1, San Diego, CA 92121-2711 USA

2 LabPMM LLC, 10222 Barnes Canyon Road, Building 1, San Diego, CA 92121-2711 USA

EUROPEAN HEMATOLOGY ASSOCIATION

invivoscribe® *Improving Lives with Precision Diagnostics®*

INTRODUCTION

TRG clonality is a powerful tool to assist in diagnosing numerous lymphoproliferative disorders and is commonly assessed using PCR-based capillary electrophoresis (CE) assays. CE assays were popularized by their speed and low cost, but suffer from disadvantages such as subjective analysis and slow turnaround when using CE results for tracking. Next-generation sequencing (NGS) offers solutions to these disadvantages by producing less subjective results and almost immediate sequence information that is ready to use for tracking. While it is understandable for labs to be hesitant about switching to newer, unfamiliar technologies, we present the following data to show the concordance between CE and NGS in a format familiar to the current generation.

OBJECTIVES

First, to present *TRG* clonality results for 388 clinical samples using both CE and NGS.

Second, to transform the NGS data into an *in silico*, electropherogram format (called NGS-E), to enable more direct comparison to CE outputs.

Third, to present a simple heuristic to demonstrate a key analysis advantage of NGS data. Namely, elucidating highly complex peaks from what would be considered clonal peaks in CE, revealing false positives in CE.

METHODS

388 clinical samples representing a variety of suspect T-cell lymphoproliferative diseases routinely tested using *TRG* assays were tested using both the LymphoTrack® *TRG* Dx assay (in development) on the Illumina® MiSeqDx® platform, and the CE-based IdentiClone® assay for *TRG* clonality. For NGS results, the %Total Reads (*i.e.* the percentage of reads an amplicon takes up in the sequencing sample) was used as the metric for calling positives and negatives. For CE plots, a standard Relative-Peak-Ratio (RPR) calculation was used as a metric for calling positives and negatives.

The results from the NGS-based assay were converted, *in silico*, to NGS-E plots, by taking the sizes of the sequenced amplicons and plotting them on a histogram. On CE-based assays, electropherograms are the standard output. The two different groups were then compared (NGS-E vs CE). The NGS-E plots had an important difference: each size column often contained multiple different sequences. For any single column, an information-based metric for variability, Shannon entropy (*Figure 1*), is utilized to define columns as having many different sequences (high entropy, see Results: *Figure 2*) or few sequences (low entropy, see Results: *Figure 3*).

$$\mathrm{H}(X) = -\sum_{i=1}^{n} \mathrm{P}(x_i) \log_b \mathrm{P}(x_i)$$

***Figure 1**: Shannon Entropy equation*

To ensure fair comparisons between samples and size columns that had different numbers of participating amplicons, each size column was initialized with 65 participating amplicons (representing the highest number of participating amplicons across all samples/sizes), and each initialized amplicon was given a minimum %Total Reads value of 0.0003125. The probability value used for the entropy calculation was derived from the fraction of %Total reads each amplicon contributed to its size column. For a size column with only initialized amplicons (*i.e.* no actual data, only initialized data), each of the 65 participants contribute an equal probability to the entropy calculation, and the negative sum is maximized (high entropy). For a size column with a few, high-signal amplicons, these signal amplicons take up the vast majority of the entropy calculation, and the negative sum is minimized (low entropy).

RESULTS: NGS-E VERSUS CE PLOTS

388 clinical samples were run on both NGS and CE assays. Examples of concordant results between these two assays are shown in *Figures 2* and *3* for negative and positive *TRG* clonality, respectively. An example of a discordant result between NGS and CE is shown in *Figure 4*. As seen in this example, when multiple amplicons have the same size, the CE assay is not able to differentiate amplicons and the CE electropherograms look positive for *TRG* clonality. Alternatively, the NGS-E plot can delineate amplicons within size bins and shows that this sample should be negative for TRG clonality.

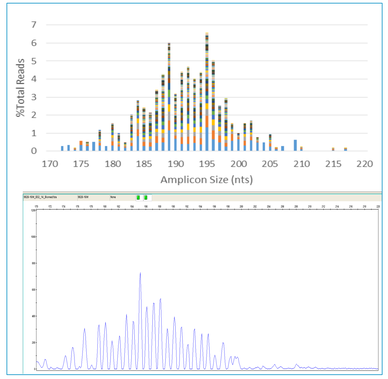

***Figure 2**: Example of a High Entropy NGS-E (top). Matching CE plot with amplicon size ranges aligned (bottom).*

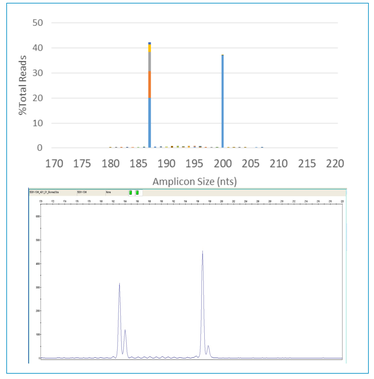

***Figure 3**: Example of a Low Entropy NGS-E (top). Each different color bar within a column represents a different rearrangement sequence, and are sorted from largest to smallest contributor, starting at the bottom.*

Matching CE plot with amplicon size ranges aligned (bottom). Note the shoulder peaks at +1 of the clonal peaks. These are common artefacts of CE. It should also be noted that peak sizes are generally 3-4nts smaller than detected by sequencing.

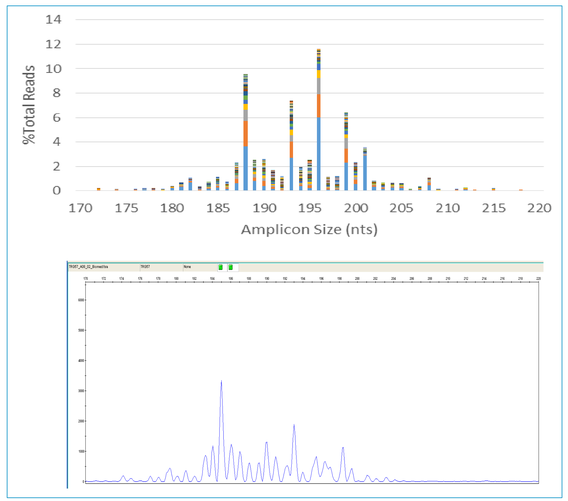

***Figure 4**: Example of a CE+, NGS- sample (NGS-E on top and matching CE bottom). While the electropherogram looks positive, the amplicon-specific delineation within size bins of the NGS-E plot shows the peaks are made of multiple, smaller signal amplicons. Therefore, this sample is an example of a false positive by CE.*

RESULTS: ENTROPY PLOT

For NGS results, the entropy calculation of the peak chosen for CE analysis was used for downstream comparison analysis. Of the 388 clinical samples, 256 (66.0%, average entropy 279.7, blue dots in *Figure 5*) were concordant negatives between the NGS and CE assays, and 90 (23.2%, average entropy 51.2, green dots in *Figure 5*) were concordant positives; the remaining 42 (10.8%) were discordant. Of the discordant samples, 16 were NGS+ (red dots, *Figure 5*), while 26 were CE+ (yellow dots in *Figure 5*). 16 of these 42 discordant samples were near the cut-off. However, an interesting pattern appeared for the discordant, CE+ samples. If the samples are ranked by the entropy value of the target peak from the NGS-E plot, 14 of the 26 were highly complex peaks (average entropy of 202.1, yellow dots with blue square, *Figure 5*) that were correctly distinguished in the NGS data as being polyclonal. These 14 represent missed clonality calls that were clear from the NGS data; comparison of the CE and NGS-E shows visually how this discordance could happen (example shown in *Figure 4*). Interestingly, none of the 16 NGS+ discordant samples could be distinguished by peak diversity. If we assume the 14 CE+ discordant samples are false positives, the NGS data was able to reduce the false positive rate by 54% (from 6.7% to 3.1%).

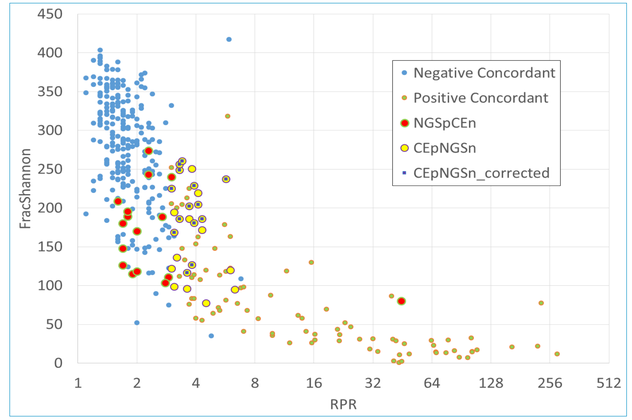

***Figure 5**: Entropy vs RPR plot for all 388 samples. RPR is shown in log2 scale to visually differentiate the datapoints.*

The 256 concordant negative samples had an average entropy of 279.7, while the 90 concordant positives had an average entropy of 51.2, demonstrating the ability of the entropy heuristic separate clear positives from negatives.

Of the 26 CE+ discordant samples, 14 demonstrated high entropy (average of 202.1), and are likely missed polyclonal calls in the CE assay due to the target peak containing multiple, different amplicons of the same size.

CONCLUSIONS

A significant hurdle in switching from CE to NGS for clonality testing is the assurance that the NGS assay gives more accurate results to offset the higher cost and longer time needed to generate data and results. However, it can be argued that the advantages of sequencing should be sufficient to offset the cost and time; the subjective nature of CE analysis and the slow turnaround time to obtaining sequence data from CE may nullify the NGS cost and time. In addition, NGS-E plots, which have the benefit of displaying the results in a manner familiar to those accustomed to CE results, in combination with a simple metric for complexity (which gives more accuracy and reduced false positives by 54% in our dataset), adds to the advantages of NGS over CE for clonality testing. These NGS-E plots can act as a bridge that the CE-centric labs can use to become more accustomed to NGS data, as they address potential concerns about converting to NGS-based testing.

CONTACT INFORMATION

Dr. Kasey Hutt
khutt@invivoscribe.com
1 858.224.6600