P032

Error Rate Normalization to Establish Position-Specific LoB & LoD in Next Generation Sequencing Assays

Julian D'Angelo.[1], Stephanie Ferguson[2], Zhiyi Xie[1], Lisa Chamberlain[1], Andrew Carson[1], and Jeffrey E. Miller[1,2]

[1]Invivoscribe, Inc., San Diego, USA, [2] Lab PMM, LLC, San Diego, CA

Introduction

- Next-generation sequencing (NGS) continues to be the method of choice for high throughput genetic assays. Many assays employ a uniform assay-specific limit of blank (LoB) and limit of detection (LoD) that tends to be overly conservative due to "worst-case" error and artifact rates (generated by PCR, sequencing and alignment errors, etc.).
- The process of differentiating sequencing artifacts from real, low-frequency variants is extremely important for clinical utility of NGS-based assays, especially as clinicians push for tracking minimal residual disease (MRD).
- The motivation for this study was to identify a strategy for finding the inherent error-rate, at single base resolution, for our MyMRD® single site assay.
- This strategy dramatically improved the specificity and sensitivity of our MyMRD assay and can be implemented in any NGS-based assay.

Materials and Methods

- Extracted DNA from whole blood or cell lines for 209 samples
  - 36 control samples (GIAB: NA12878 replicates)
  - 173 clinical samples
- NGS performed using Illumina's MiSeq® instrument
- All samples analyzed using custom In-house MyMRD® analysis pipeline
- Interrogated every position across the panel for all samples
- Targets were split into exonic (17,538 targets) and intronic (20,039 targets) sets and same analysis was run for both sets
- Counted observations of each base substitution and insertion/deletion (indel) event
- Looked at the distributions of reference to alternate sequencing errors
- Investigated for trends or extreme biases which can influence error rate
  - i.e. repeat regions (using UCSC Genome Browser's RepeatMasker), high GC%, etc.
- Determined more accurate LoB/LoD on a per position basis
- Compared trends between exonic and intronic regions

Results:

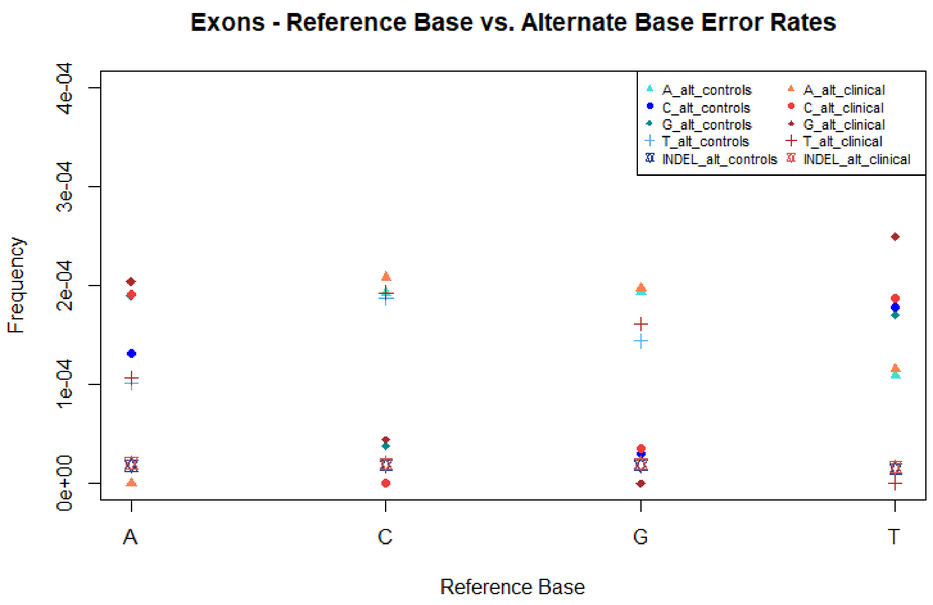


Figure 1: This plot shows the average error rate frequencies (across samples) for each substitution, for each reference base across targeted exonic positions. The control samples are labeled in various shades of blue whereas the clinical samples are labeled in various shades of red. The average error rates are all well below the generalized 0.1% error rate of the sequencer.

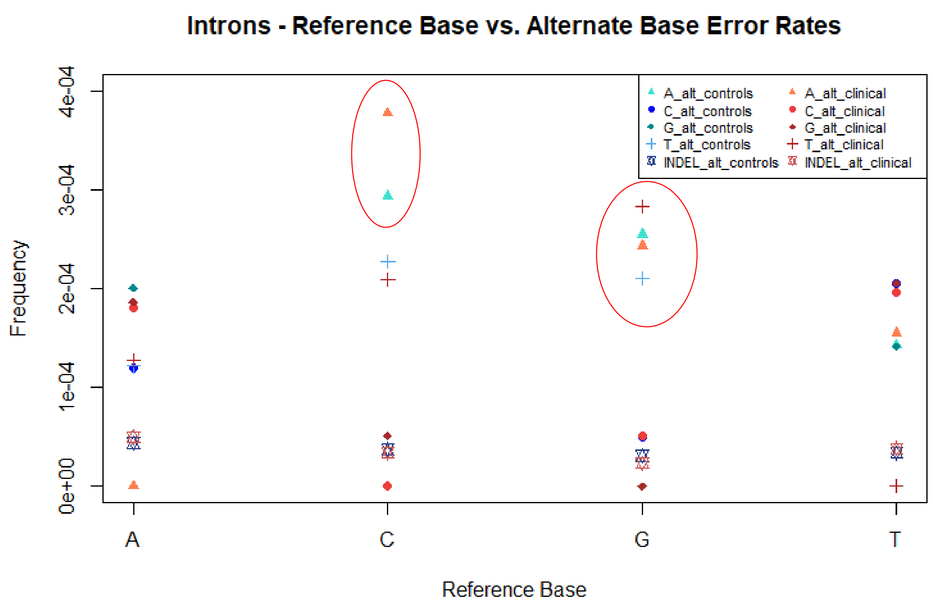


Figure 2: This plot shows the average error rate frequencies (across samples) for each substitution, for each reference base across targeted intronic positions. The control samples are labeled in various shades of blue whereas the clinical samples are labeled in various shades of red. The average error rates are all well below the generalized 0.1% error rate of the sequencer. Also note the higher frequency error rates in intronic positions for C→A and G→T.

Results:

Figure 3

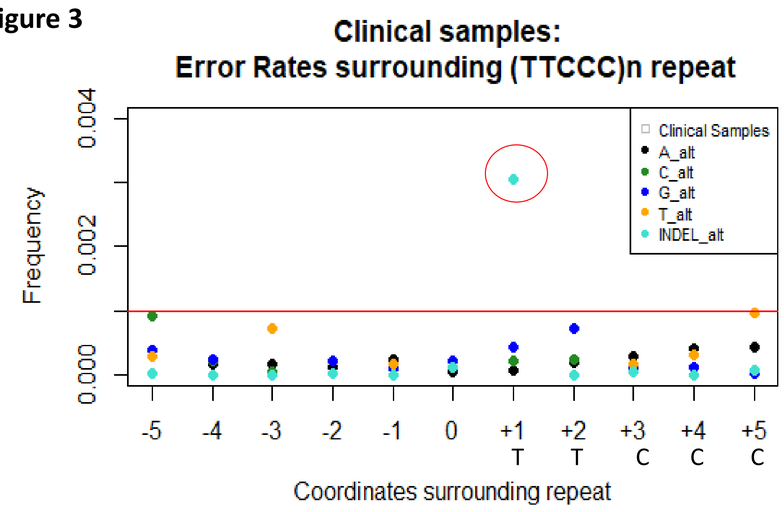
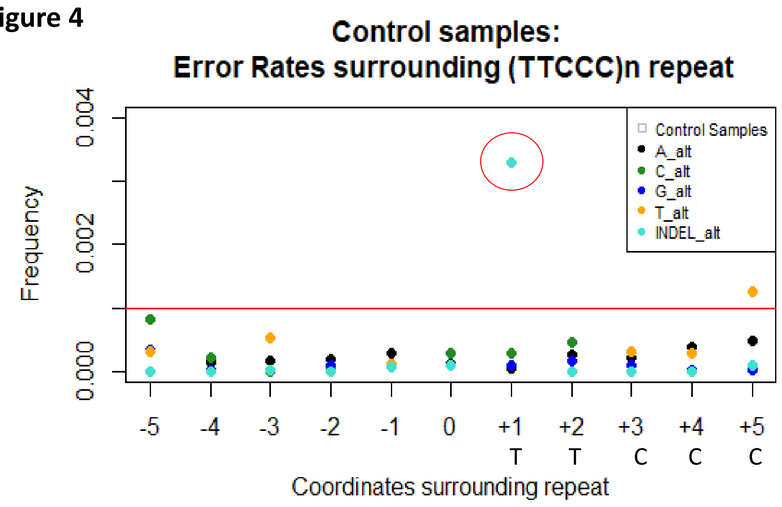


Figure 4



Figures 3 and 4: This plot shows the frequencies of each alternate call at +/- 5 bases from the start of the (TTCCC)n repeat. One of the repeats observed (TTCCCT)n had a clear affect on increased error rates in the surrounding area. The circled point highlights a drastically increased indel error rate one base upstream of the repeat start site. The observed indel was a deletion of the TTCCC repeat. This is interesting and not unexpected due to the potential of slippage around such regions. This observation held for both clinical (**Figure 3**) and control (**Figure 4**) samples.

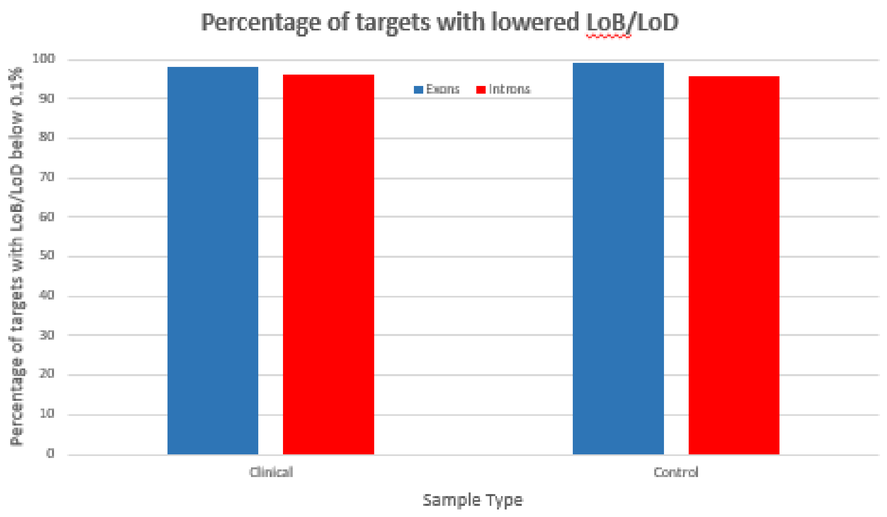


Figure 5: This bar plot shows the percentage of targets that have a background error rate below the 0.1% background error rate that many sequencers claim to have. It is separated by exonic (blue bars) and intronic (red bars) targets. As is expected, intronic targets had high background error rates in more positions compared with exonic targets.

Conclusions

- This study showed the importance of both assay specific and position specific error rates for LoB & LoD determinations due to high variability at each position
- Error rates and general trends aligned with expectations:
  - Intronic targets showed a higher overall incidence of error rate as compared to exonic targets (especially true for indels).
  - There was a simple repeat (TTCCC)n overlapping the MyMRD® targets that seemed to influence/increase error rates in the surrounding region.
- **>97% of all targets positions (both introns and exons) had a background error rate below the general 0.1% suggested error rate of the sequencing instrument.**
  - >98% of exonic targets
  - >95% of intronic targets

AMP Global, May 16-18 2019, Hong Kong