Benchmarking NGS Integration Site Analysis Methods in Support of Long-Term Safety Monitoring of Cell and Gene Therapy Products

Jeffrey Otto, Yongjun Fan, Guanhui Bao, David Corney, Ben Niu, Qiu Yu, Baursha Khan, Laure Turner, Chris Mozdzierz, Haythem Latif, Ginger Zhou
Azenta, South Plainfield, NJ 07080

Abstract

The FDA Guidance to Industry on Long Term Follow-Up (LTFU) After Administration of Human Gene Therapy Products states the importance of longitudinal testing of gene products introduced into human subjects. Depending on the delivery mechanism, the therapeutic gene product may or may not integrate into the genome. Of particular interest are gene-product integrations near proto-oncogenes which might lead to malignancies. The FDA LTFU guidance states that recipients of an integrating gene therapy modality should be tracked for 15 years, while those receiving a non-integrating therapy modality should be tracked for 5 years. Therefore, advanced analytical methods are needed to identify, quantify, and track integration events across the genome. Here, we provide a comprehensive evaluation of methods leveraging next-generation sequencing approaches for genome-wide analysis of lentiviral integration events. Our analysis employed well-characterized standards consisting of varying copy number and known integration sites.

The approaches we characterized can be bucketed into two major groups: PCR amplification approaches and target capture-based approaches. All methods detected true positives with strong correlation to theoretical integration site dosage levels down to 1% allele frequency. Comparatively, PCR amplification-based approaches have lower data requirement per sample suggesting higher sensitivity, greater molecular capture, and lower limit of detection compared with target enrichment-based approaches. Target enrichment-based approaches can afford the flexibility to capture the integrated vector, which is of interest for characterizing partial integration events. While all methodologies performed well in our study, the choice of assay (or assays) for testing will depend on numerous factors including but not limited to the viral vector system and construct and starting material availability.

AZENTA
LIFE SCIENCES

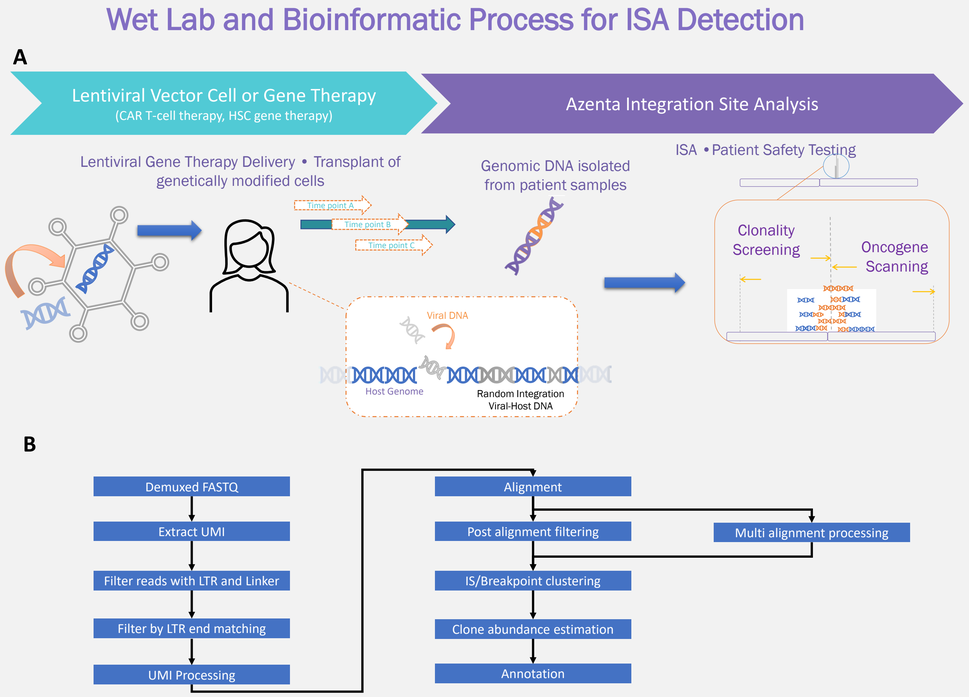

Figure 1. Schematic of sample processing logistics and wet lab/bioinformatics workflow for integration site analysis (ISA). A: Following initial lentiviral vector transduction and LTFU samples (typically whole blood and/or enriched populations) are collected and transported to Azenta for ISA. In this study, several workflows were developed and benchmarked including two PCR amplification based-approaches and target capture/enrichment sequencing (TES) based approach using baits targeting the LTR region. To aid in benchmarking these approaches, well-characterized reference samples were obtained. Each reference sample consisted of a known number and location of integration sites, as well as contrived mixtures of these reference samples. In each case, data was processed through a pipeline as shown in (B) to identify potential clonal expansions, as well as annotate the breakpoint/integration site based on proximity to known oncogenes.

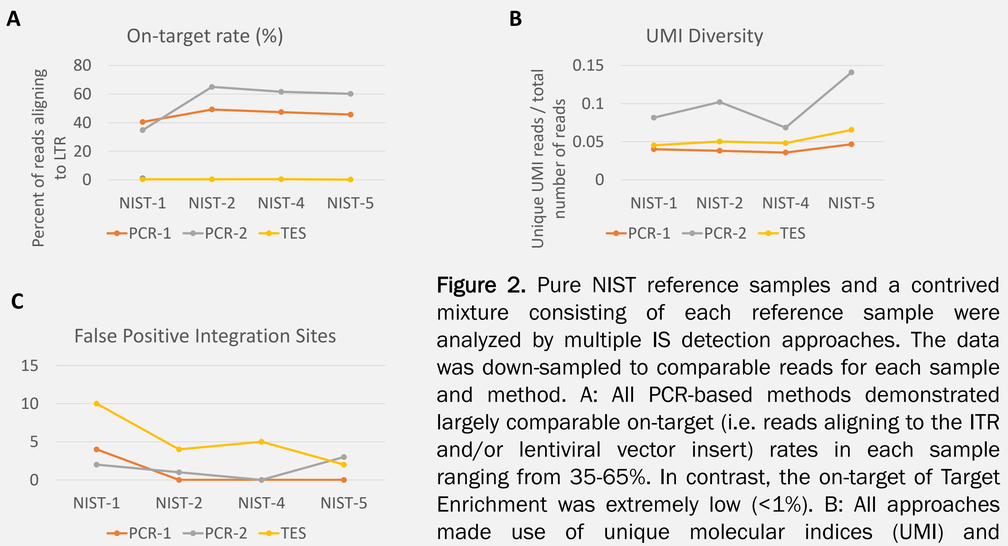

Figure 2. Pure NIST reference samples and a contrived mixture consisting of each reference sample were analyzed by multiple IS detection approaches. The data was down-sampled to comparable reads for each sample and method. A: All PCR-based methods demonstrated largely comparable on-target (i.e. reads aligning to the ITR and/or lentiviral vector insert) rates in each sample ranging from 35-65%. In contrast, the on-target of Target Enrichment was extremely low (<1%). B: All approaches made use of unique molecular indices (UMI) and comparable UMI diversity was observed. C: All methods detected the known IS, yet the number of false positive IS detected with > 5 supporting reads was elevated in TES.

A

Sample	Number of expected IS	Expected IS frequency in Mixture
NIST-1	3	22.9%
NIST-2	2	0.8%
NIST-3	0	0.0%
NIST-4	3	8.3%
NIST-5	1	5.0%

B

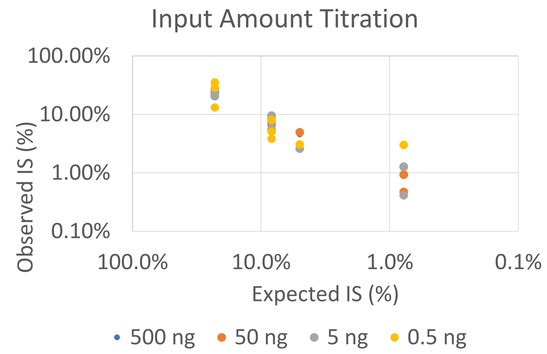

Figure 3. Input titration correlation analysis. A: A contrived mix of four NIST reference samples with known numbers of IS was prepared and 0.5 – 500 ng template used as input for method PCR-1. B: The observed level of each IS closely correlated with the expected level at the higher input amounts, but quantitation at reduced input amount, as well as IS at levels < 1%, was less consistent and precise illustrated by a wider spread of IS frequencies and correlation coefficient (500 ng: 0.993; 50 ng: 0.996; 5 ng: 0.957; 5 ng: 0.777)

Conclusions

- Lentiviral gene therapy has extraordinary potential as a therapy for a wide range of diseases and disorders but has potential for off-target insertional mutagenesis. For this reason, regulatory agencies require long-term surveillance.
- Two PCR-based approaches and one target capture approach was used to identify integration sites. All were able to successfully detect integration sites.
- PCR-based approaches confers high sensitivity and specificity, as well as having low input requirements.
- Enrichment-based approaches enable detection of truncated insertions which may occur with AAV-based therapies but requires higher input amounts and has reduced specificity.
- All tested approaches were robust and successfully identified known IS within reference samples. It is recommended to consider their relative advantages and disadvantages during study design.

Acknowledgements

We gratefully acknowledge Hua-Jun He and Zhiyong He (NIST Biomarker and Genomics Sciences Group) for providing well-characterized IS reference samples and helpful discussions.