Redefining “Gold Standard”: Ultra-Sensitive Characterization of Commercial DNA Standards with Duplex Sequencing

Jacob Higgins, PhD¹ – Gabriel Pratt, PhD¹ – Charles C. Valentine, MS¹ – Lindsey N Williams, PhD¹ – and Jesse J. Salk, MD, PhD²

TwinStrand Biosciences, Seattle, WA¹ – Division of Medical Oncology, University of Washington, Seattle, WA²

Abstract

DNA standards are an essential resource in diagnostic and research labs. However, as genomic technologies become more sensitive, the metrics defining what a “gold standard” entails have not been carefully re-evaluated. Cell culture exposes DNA to oxidative damage and clonal selection, and synthetic DNA has a mutation frequency inherently higher than biologically-derived genomic DNA. Here we apply ultra-accurate Duplex Sequencing to characterize the background mutation frequency in two commercial mutation standards, in control genomic DNA from a healthy young donor and in custom mutation mixes using that young donor DNA as a base. We observe that synthetic (highest) and cell line (intermediate) DNA-based mixes exhibit significantly elevated background mutation frequency relative to carefully selected genomic DNA or mutation mixes where genomic DNA is the diluent into which patient or cell-line derived mutant molecules are spiked. When employing high-sensitivity genomic technologies with the ability to resolve mutations at levels below one-in-one-million, rigorous attention to the techniques and substrates used for manufacturing mutation standards is critical.

Sequencing Errors Obscure Truth

TwinStrand Duplex Sequencing™ Technology

A DuplexSeq™ Adapter has:
1. Identical (or relatable) degenerate tags in each strand.

Duplex Sequencing of Commericially Available DNA Mutation Standards

Duplex Sequencing (DS) libraries from commercial standard A (CS-A) and commercial standard B (CS-B) were captured with probes targeting 29 genes recurrently mutated in acute myeloid leukemia (AML). CS-A is a mix of cell line DNA with known mutations, and CS-B is a mix of synthetic mutant DNA fragments spiked into DNA from a single cell line. Fifteen mutations at predicted variant allele frequencies (VAF) from 5-40% were captured in CS-A, and 10 mutations at predicted VAF of 5-15% in CS-B. All mutations were identified by DS.

Background Mutations in Commercial Standards vs. Negative Control

CS-A and CS-B were sequenced to mean Duplex depth of 7.07x and 8.17x, respectively. Negative control DNA from a healthy 18-year-old never-smoker was sequenced to mean Duplex depth of 24.75x. Background mutations were compared in CS-A, CS-B and the control. Despite being sequenced to approximately 3-fold greater Duplex depth, the control DNA had fewer low-frequency clonal mutations (mutations with < 1% VAF and ≥2 alternate allele counts) than CS-A or CS-B, in terms of mutant sites and proportion of total background mutations. The control also had lower overall background frequency (4.7x10⁻⁴) than either CS-A or CS-B.

Duplex Sequencing Correlation with ddPCR and Standard NGS

In CS-A, DS VAF was highly correlated to vendor ddPCR measurements, with r² = 0.98. In CS-B, DS VAF correlated more weakly to predicted values, with r² = 0.74 vs. vendor ddPCR quantification. The DS successfully identified FLT3 internal tandem duplications (ITD) from 33-302 bp, as well as other indels up to 23 bp.

Conclusions

• Duplex Sequencing of cancer driver genes in spherosoma DNA from a healthy 18-year-old control reveals a background mutation frequency below 1-in-2-million, which is biological, not technical.
• Cell lines have a higher background mutation frequency and a greater proportion of clonal (multi-count) background mutations than DNA from a healthy young donor, even when sequencing the donor to 3x greater depth.
• For mutation mixes, use of DNA from a young donor as the base results in a lower background mutation than with cell line DNA as the base diluent.
• Duplex Sequencing readily detects single-nucleotide variants (SNVs) as well as large insertions and deletions at ultra-low frequencies.
• Given sufficient sequencing depth, Duplex Sequencing detects mutations at frequencies from 1/30,000 to 1/100,000 with high sensitivity and specificity.

Contact: https://twinstrandbio.com/contact/