Duplex Sequencing™ Detects Rare Subclonal Variants that Mark Early Carcinogenesis and Preneoplastic Clonal Evolution

Charles Valentine[1] – Mark Fielden[2] – Robert Young[3] – Jake Higgins[1] – Lindsey Williams[1] – Tan Li[1] – Rohan Kulkarni[3]
Sheroy Minocherhomji[2] – Rosana Risques[4] – Jesse Salk[1,5]

TwinStrand Biosciences, Seattle, WA[1] – Amgen, Thousand Oaks, CA[2] – MilliporeSigma/BioReliance, Rockville, MD[3]
Univ. of Washington Dept. of Pathology, Seattle, WA[4] – Univ. of Washington Division of Medical Oncology, Seattle, WA[5]

TWINSTRAND™ BIOSCIENCES

DNA Sequencing of Ultra-Rare Variants

Cancer is a disease of somatic evolution characterized by the natural selection of genomic mutations that facilitate enhanced cell survival and proliferation. Historically, our ability to identify early genetic patterns of clonal selection in both human and model organisms has been hampered by inadequately sensitive methods for identifying variants during the long period between their occurance and the final outgrowth of a clinically apparent tumor.

We present the use of Duplex Sequencing of both human and mouse tissues to demonstrate the detection of subclones at allelic fractions below 1×10^{-4} showing that this technology can be used as both a new preclinical and clinical tool for assessing life-integrated carcinogenic processes and ultimate cancer risk.

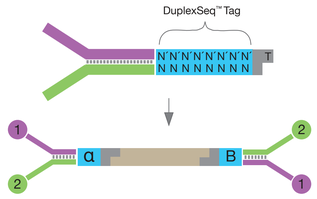
Sequencing Errors Obscure Truth

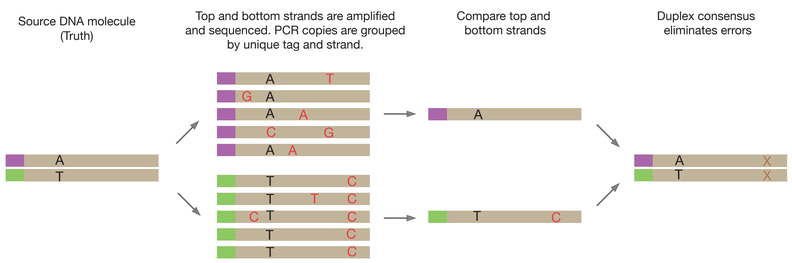
Next-Generation Sequencing (NGS) | Single Strand Error-Corrected NGS | Duplex Sequencing

TwinStrand Duplex Sequencing™ Technology

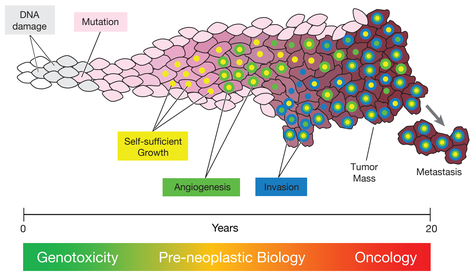
A DuplexSeq™ Adapter has:

1. Identical (or relatable) degenerate tags in each strand.
2. An asymmetry allowing independent strand identification.



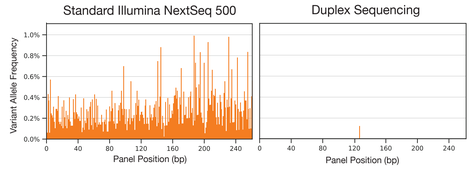


Neoplastic Evolution



The progression of cancer recapitulates Darwinian evolution. Through time, cells and tissues accumulate mutations as a result of unrepaired DNA damage or as errors that are introduced through cellular replication. Many of these mutations have either neutral or deterimental effects on the cell, however, few mutations will benefit the cell causing the cell and its lineage to preferentially amplify. As additional mutations are gained, selection and expansion continue as the field progresses into a genetic disease.

Sequencing Accuracy Required



Bases	183,976,776	10,220,932
SNVs	232,523	3
Mutation Frequency	1.26/1,000	2.94/10,000,000

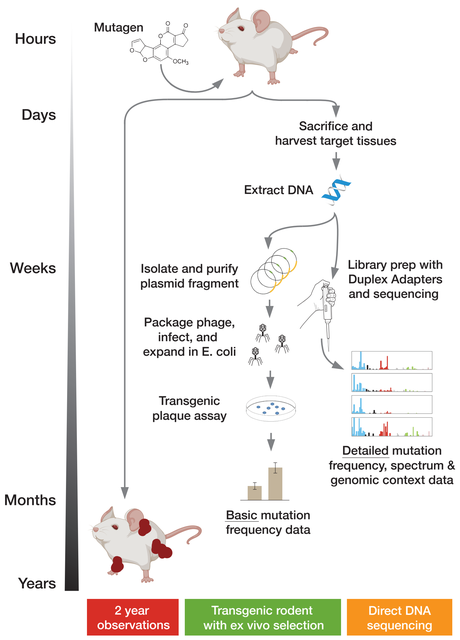
Common Sources of DNA Sequencing Error

- Sequencer Miscall
- PCR Misincorporation
- DNA Damage

→ 8-oxoguanine, Deaminated cytosine, Abasic sites, Many others...

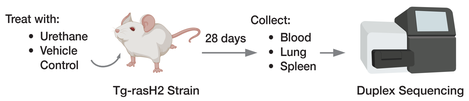
The error rate of NGS is ~0.1% which creates a background that obscures rare variants. Duplex Sequencing overcomes these errors by forming consensus among PCR duplicates from the same source molecule and increases the overall accuracy of sequence data by more than 10,000x.

Measuring Pre-neoplastic Clones in Murine Species



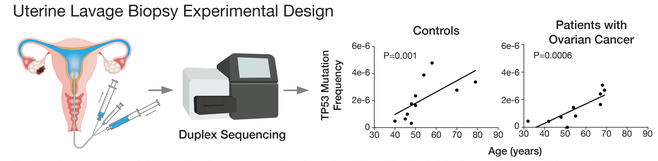
TwinStrand Duplex Sequencing™ is the only genotoxicity assay that delivers rich, mechanistically-informative readouts in the short timescale of weeks instead of months. Duplex Sequencing not only reports mutation frequency but also yields valuable information about mutation type and nucleotide context from anywhere in the genome of any organism.

Murine Treatment Experimental Design

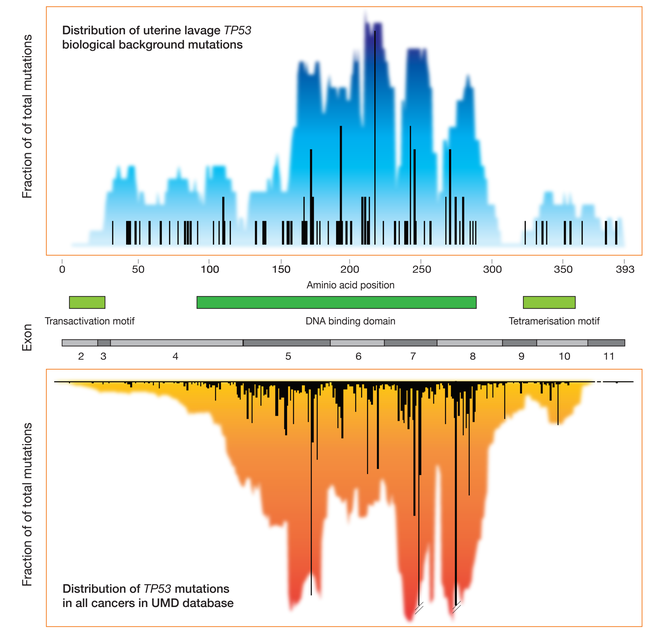


TP53 Mutations Correlate with Age and Those Seen in Human Cancers

Uterine Lavage Biopsy Experimental Design

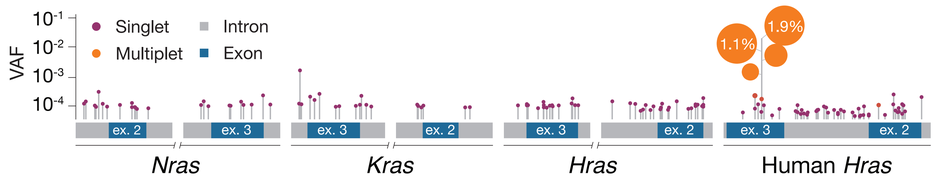


Duplex Sequencing of 21 uterine lavage samples from both healthy individuals and those with ovarian cancer show that the frequency of ultra-rare background *TP53* mutations correlate with age.



TP53 biological background mutations detected using Duplex Sequencing from uterine lavage samples (upper) mirror the prevalence of clonal mutations seen in all human cancers from the UMD Seshat database (lower). Mutations are distributed in the residues and domains of *TP53* that cause a functional disruption to the translated protein.

Oncogenic Variants Observed Soon after Mutagen Treatment



Single nucleotide variants (SNV) plotted over the genomic intervals for the exons captured from the Ras family of genes, including the transgenic loci, in the Tg-rasH2 mouse model treated with Urethane or Vehicle Control in the blood, lung, and spleen tissues using OECD guidelines. Singlets are mutations found in a single source molecule. Multiplets are identical mutations identified within multiple source molecules within the same sample and may represent clonal expansion events. A notable observation is the cluster of mutations in codon 61 of the Hras transgene which is a known cancer hotspot in human tumors. The mutations in codon 61 are A·T → T·A transversions which are indicative of Urethane mutagenesis. This work shows an early detection opportunity in discovering subclones that bear cancer-driving mutations in normal mouse tissues mere weeks after mutagen exposure.

Conclusions

- Duplex Sequencing™ enables detection of variants below a frequency of one-in-a-million bases, including those that have been induced by carcinogens.
- Quantification of subclones with frequencies below the error rate of traditional NGS technologies enables shorter timescale assays for detecting early cancer initiating events.
- Duplex Sequencing illustrates how similar patterns of clonal selection can be seen in multiple otherwise healthy tissues of humans as a part of normal aging.
- Duplex Sequencing is a sensitive and data-rich assay for detecting both mutagenesis and carcingoenesis of any genetic locus, in any tissue, in any organism.

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

PUBLICATIONS

Salk JJ, Schmitt MW, Loeb LA. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations **Nature Reviews Genetics,** 2018, 19(5):269-285. PMID: 29576615.

Chawanthayatham S, Valentine CC, Fedeles BI, Fox EJ, Loeb LA, Levine SS, Slocum SL, Wogan GN, Croy RG, Essigmann JM. **Proc. Nat. Acad. Sci. USA,** 2017, 114(15):E3101-9. PMID: 28351974.

Schmitt MW, Fox EJ, Prindle MJ, Reid-Bayliss KS, True LD, Radich JP, Loeb LA. Sequencing small genomic targets with high efficiency and extreme accuracy. **Nature Methods,** 2015, 12(5):423-5. PMID: 25849638.

Kennedy SR, Schmitt MW, Fox EJ, Kohrn BF, Salk JJ, Ahn EH, Prindle MJ, Kawai JK, Risques RA, Loeb LA. Detecting ultra-low frequency mutations using Duplex Sequencing. **Nature Protocols,** 2014, 9(11) 2586-606. PMID: 25411958.

Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA. Detection of ultra-rare mutations by next-generation sequencing. **Proc. Nat. Acad. Sci USA,** 2012, 109(36):14508-13. PMID: 24086148.

MKT0002 v1.0