Detection of Ultra-rare Mutations *in vivo* Establish Biomarkers of Endogenous and Environmental Exposure

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

TwinStrand Biosciences, Seattle, WA[1] – Amgen, Thousand Oaks, CA[2] – MilliporeSigma/BioReliance, Rockville, MD[3]

DNA Sequencing of Ultra-Rare Mutations

Endogenous and environmental processes alter the genetic record through damage and mutation. Next-generation sequencing (NGS) technologies have been revolutionary in describing the genetic differences between clonal populations but are too error-prone to detect ultra-rare mutations.

We introduce the TwinStrand Duplex Sequencing™ assay that is sensitive enough to directly measure the faint signal of a mutagen within days of animal or cellular exposure using only bulk-extracted genomic DNA.

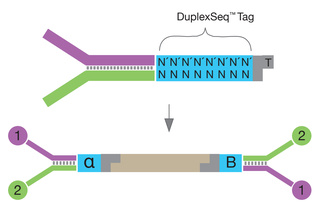
Sequencing Errors Obscure Truth

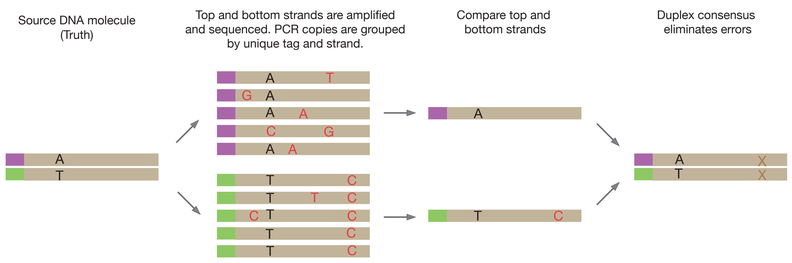
Next-Generation Sequencing (NGS) | Single Strand Error-Corrected NGS | Duplex Sequencing

TwinStrand Duplex Sequencing™ Technology

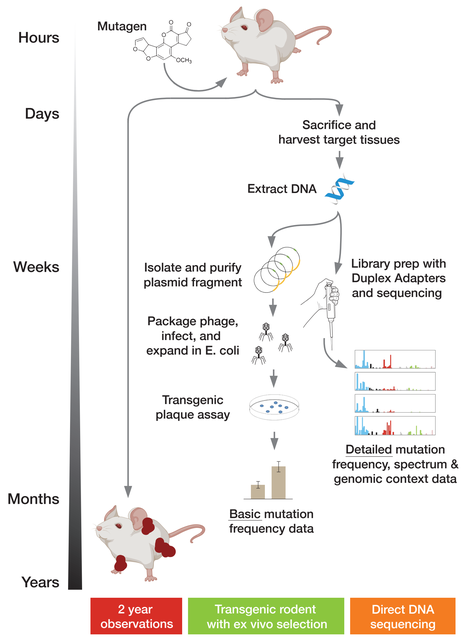
A DuplexSeq™ Adapter has:

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





Measuring Genotoxicity



Various methods for assessing genotoxicity exist although none, until now, could be performed in weeks time while providing a rich, mechanistically-insightful set of output data. TwinStrand 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.

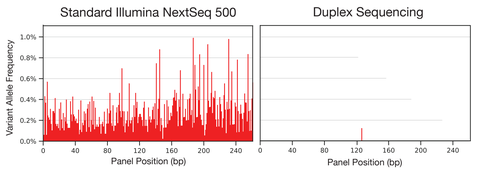
Common Sources of Error

- Sequencer Artifact
- 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.

Accuracy is Required

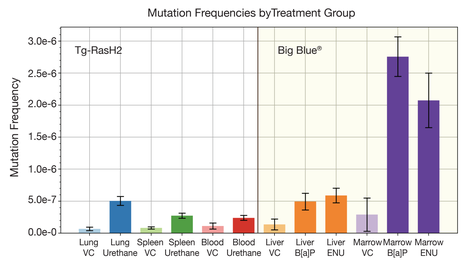


Raw bases:	183,976,776	Duplex bases:	10,220,932
Raw SNVs:	232,523	Duplex SNVs:	3
Mutation frequency:	1.26/1,000	Mutation frequency:	2.94/10,000,000

Experimental Design

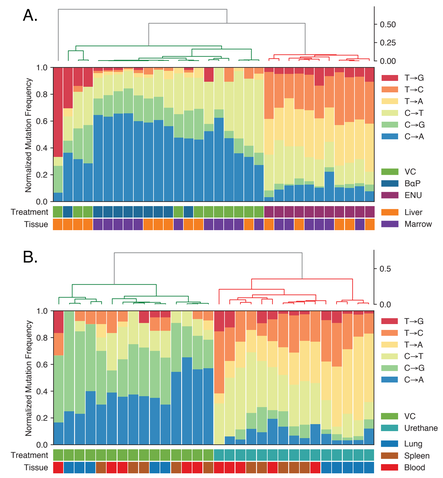
	Tg-RasH2 Mouse	BigBlue® Mouse
Tissues	Lung (10) Spleen (10) Blood (10)	Liver (15) Marrow (17)
Treatment	Urethane (15) VC (15)	B[α]P (10) ENU (11) VC (11)
Genomic Targets	*Polr1c, Rho, Ctnnb1 Hp, Hras, Nras, Kras*	*Polr1c, Rho, Ctnnb1, Hp*
Duplex BP	4,923,565,684	4,716,990,836

Mutation Frequencies



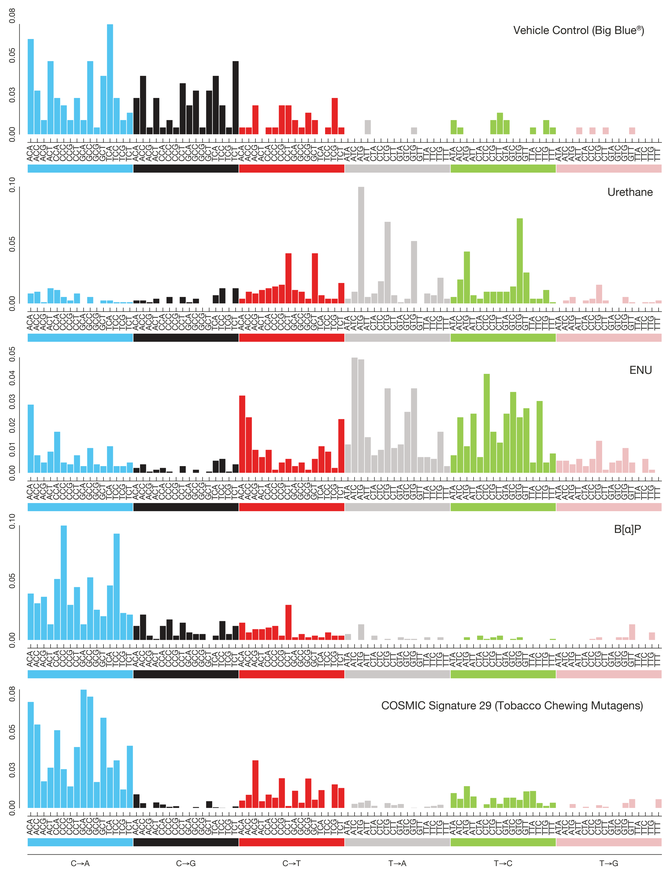
Mutation induction varies considerably between mutagens and tissue types. The fold-increase mutation frequency in this assay mirrors that found by using the BigBlue® Transgenic Rodent (TGR) assay.

Base Substitution Spectra



The simple base substitution spectrum of nucleotide variation can identify mutagenic tendency such as in distinguishing a urethane exposure from vehicle control. In other cases, the simple spectrum alone is not powerful enough to resolve distinct differences between treatments, such as in the lack of a clear separation for Benzo[a]pyrene and vehicle control in slow-dividing liver tissue.

Trinucleotide Spectra



The trinucleotide spectra of base substitutions reveals a unique fingerprint for each treatment group. These fingerprints are consistent with known signatures of clonal mutation observed in tumors caused by similar exposures. In untreated animals the spectra is abundant for C·G→A·T and C·G→G·C which is likely caused by oxidation, and C·G→T·A caused by deamination of cytosine and 5-methylcytosine. Urethane shows the characteristic exposure pattern of A·T→T·A in NTG contexts and B[a]P has a high similarity to the pattern of tobacco exposure in human cancer where B[a]P is a major carcinogen.

Conclusions

- Duplex Sequencing™ enables anyone to detect genotoxin induced variants below a frequency of one-in-a-million bases.
- We show a robust ability to detect, and precisely quantify, the effect of mutagen exposure on the genomic DNA of five tissue types from two mouse models against three mutagen treatments.
- Duplex Sequencing is a sensitive and data-rich assay for detecting mutagenesis of any genetic locus, in any tissue, in any organism.
- Duplex Sequencing-produced trinucleotide base substitution spectra enables the discovery of links between mutagenic exposure and human genetic disease. These spectra can also be used to infer the etiology of a mutagenic compound.

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.

MKT0001 v1.0