Ultra-Sensitive Duplex Sequencing for Quantifying Multi-Individual Cell Therapy Sub-Population Fractions

Zach Norgaard[1], Jake Higgins[1], Jeffry Yaplee[1], Joe Blake[2], Mary Prieve[2], Fang Yin Lo[1], Colleen Delaney[2], Charles C Valentine[1], Jesse J Salk[1]

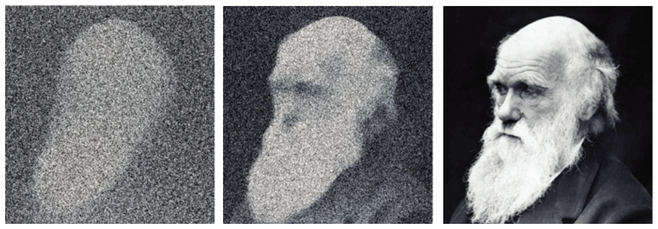
1. TwinStrand Biosciences, Seattle, WA
2. Deverra Therapeutics, Seattle, WA

Introduction

Umbilical cord blood (CB) contains CD34+ hematopoietic stem cells and is routinely used as a source of donor stem cells for allogeneic transplant. CB units with insufficient cell dose for transplant largely go unused but are increasingly being repurposed as source material to generate cell therapies and regenerative medicine products. If these small units are salvaged and pooled to generate a cell therapy product, current techniques are often insufficiently sensitive for quantitative deconvolution of these mixtures to assess the relative expansion of each subpopulation during the manufacturing process or after infusion into patients. A more sensitive and widely applicable assay would accelerate development of these new treatments. Duplex Sequencing (DS) compares both strands of each original DNA molecule to eliminate technical errors and achieve extreme accuracy and sensitivity, with an error rate below one-in-ten-million.

We designed a hybrid capture panel targeting 277 single nucleotide polymorphisms (SNPs) to distinguish each contributor genome in a complex mixture. From 16 individual CB samples, we prepared 10 synthetic mixtures of up to 8 contributing genomes at fractions ranging from 82% to 0.05%. To determine the estimated fraction of a specific CB genome in a mixture, we divided the total number of molecular counts of that individual's informative alleles (alleles unique to a specific CB sample) by the total effective duplex molecular depth at all relevant informative sites. Operators and analysts were blinded to the composition of 7/10 mixtures until after observed genomic fractions were calculated.

Sequencing errors obscure truth

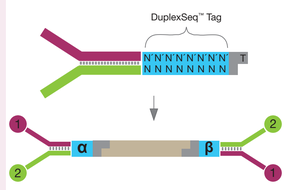


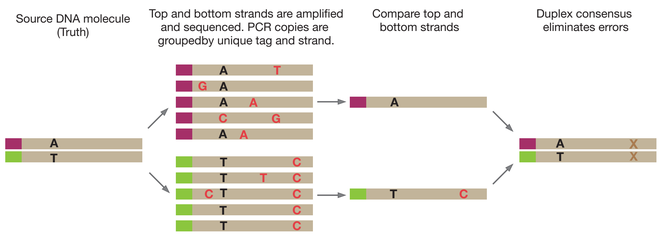
Next-Generation Sequencing (NGS) | Single Strand Error-Corrected NGS | Duplex Sequencing

TwinStrand Duplex Sequencing™ technology

A DuplexSeq™ adapter has:

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





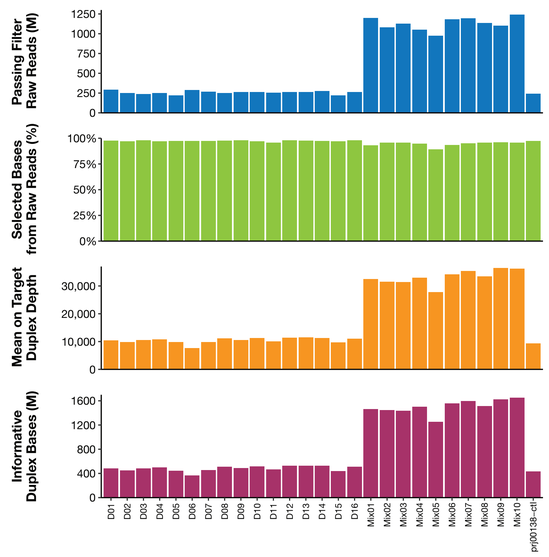
Study design

16 individual CB units were used to create 10 synthetic CB unit mixtures

We would like to acknowledge the Cleveland Cord Blood Center (CCBC) and CCBC's Volunteer Donating Communities in Cleveland, OH, Atlanta, GA, and San Francisco, CA, as a source of the material in this study to create 10 synthetic CB unit mixtures.

CB Unit	Mix01	Mix02	Mix03	Mix04	Mix05	Mix06	Mix07	Mix08	Mix09	Mix10
D01	12.5%	34.0%	1.0%		33.35%				12.8%	
D02	12.5%		1.0%	0.2%	65.00%					
D03	12.5%	31.0%	1.0%		0.40%		0.05%		20.0%	
D04	12.5%		14.0%	17.3%		25.0%		41.00%		
D05			1.0%	0.1%	0.30%	21.8%		57.00%		7.0%
D06	12.5%		80.0%							7.0%
D07	12.5%		1.0%	82.0%		10.0%	32.00%	0.10%		
D08	12.5%		1.0%			43.0%	47.80%			
D09				0.1%	0.25%			0.50%	10.0%	
D10				0.1%	0.40%		0.05%		11.0%	
D11				0.1%	0.20%				1.0%	
D12				0.1%			0.10%	0.05%		11.0%
D13	12.5%				0.10%	0.1%			17.0%	12.0%
D14						0.1%		1.00%	20.0%	14.0%
D15		35.0%					20.00%	0.30%		33.0%
D16								0.05%	8.2%	16.0%

Different sequencing goals require different duplex molecular depths



For pure CB Units, the objective of sequencing was to accurately characterize germline variants. For mixture samples, the objective was to accurately detect ultra low-frequency variants.

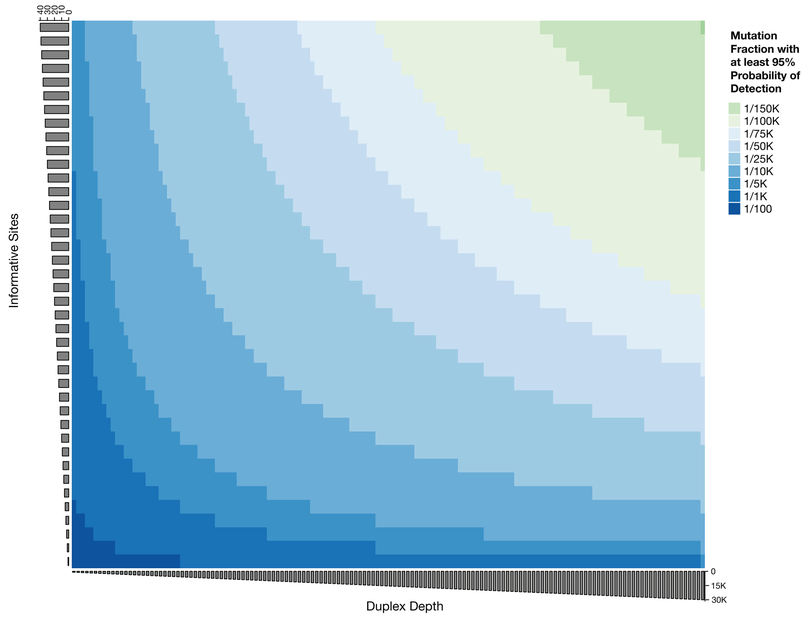
An average of 256 million raw reads were obtained for pure CB unit samples and 1.1 billion raw reads for mixture samples.

The average mean on target duplex depth for per CB unit samples was 10,352x and the average mean on target duplex depth for mixture samples was 33,191x.

At least one informative site was found for every sample in a sixteen-sample comparison

	D01	D02	D03	D04	D05	D06	D07	D08	D09	D10	D11	D12	D13	D14	D15	D16
Homozygous	2	0	1	0	1	0	1	3	0	0	0	0	0	2	1	0
Heterozygous	7	3	5	4	13	5	4	9	1	6	1	3	4	1	1	2
Total Sites	9	3	6	4	14	5	5	12	1	6	1	3	4	3	2	2

95% probability to detect contributing genotypes at all expected frequencies

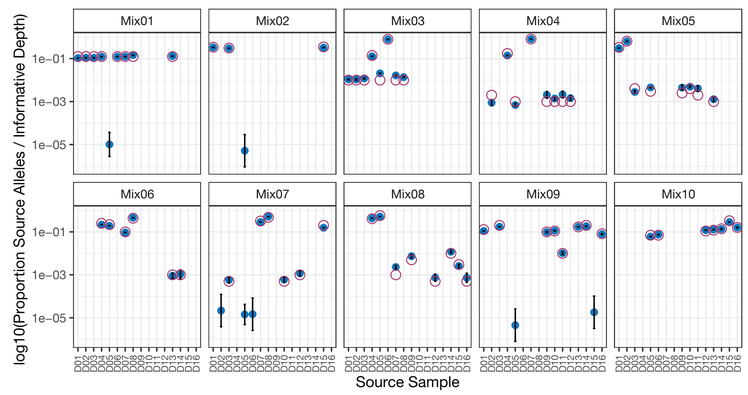


The probability of detecting a contributing genotype of a specific frequency in a mixture can be calculated as $1-(1-m*a)^{\wedge}(\frac{d}{2})$ where m is the contributing genotype frequency, a is the number of informative alleles, and d is the mean duplex depth.

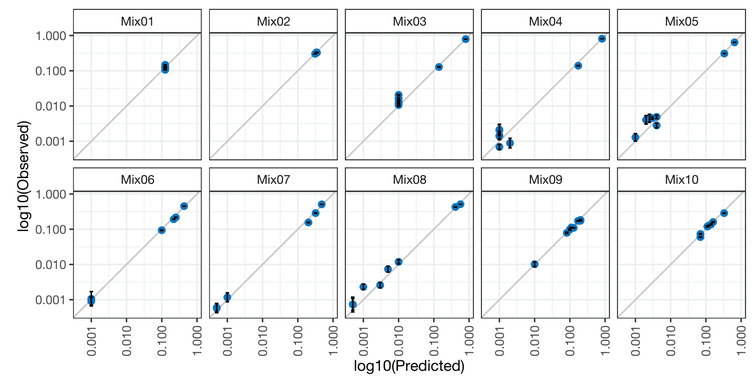
For Mixture 8, the lowest expected frequency is 0.0005 ($^{1}/_{2K}$) for D16, a sample with 2 informative alleles.

With a mean on target duplex depth of 33K for Mixture 8, we have a 95% probability to detect populations down to $^{1}/_{10K}$ with 2 informative alleles.

Duplex Sequencing accurately quantifies contributing genotypes down to proportions as low as 1 in 2,000 in complex mixtures



For each mixture, the frequency of each CB unit genotype was assessed. Observed frequency estimates are in blue with 95% Wilson intervals as black error bars. Expected frequencies are identified by plum circles. If a circle is not present, the CB unit genotype was not expected. There may be evidence of contamination at frequencies well below any expected values (1/100K) in four of the ten mixtures.



Predicted and observed values are plotted with 95% Wilson intervals for the observed values. The identity line ($x = y$) is plotted in gray.

Conclusions

- **Duplex Sequencing achieved excellent performance for individual and mixed CB unit samples**
  - Different analysis goals allows us to vary duplex sequencing depth. On average 256 million raw reads were sequenced for each pure CB unit sample and 1.1 billion for mixture samples.
  - Hybrid selection was 95% effective for pure CB unit samples and 89% for mixture samples.
  - The average mean on target DS depth for pure CB unit samples was 10,352× and 33,191× for mixture samples.
- **Duplex Sequencing allows for robust estimation of CB unit genotype frequencies in complex mixtures**
  - Using the hybrid capture panel targeting 277 single nucleotide polymorphisms (SNPs), we were able to identify at least one informative site for each of the sixteen cord blood samples and up to fourteen informative sites for a single sample.
  - Sufficient duplex depth and informative sites were obtained to achieve greater than a 95% probability of detection for all expected contributing genotype frequencies in the mixtures.
  - There may be evidence of contamination at frequencies well below any expected values (1/100K) in four of the ten mixtures.
  - Duplex sequencing allows for robust estimation of CB unit frequencies in complex mixtures with up to eight contribution genotypes at frequencies as low as 1/2K (0.05%).
  - This assay will allow for the tracking of CB unit genotype frequencies in culture to assess population dynamics through time and allow for tracking of individual CB unit frequencies in patients treated with a mixture of CB units.

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