Pro-Seq: A High Fidelity and Cost-Effective Duplex Sequencing Method for ctDNA Detection

Joel Pel[1], Matthew Wiggin[1], Wendy Choi[1], Milenko Despotovic[1], Laura Gelinas[1], Amy Leung[1], Laura Mai[1], Gosuke Shibahara[1], Tim Smith[1], Lloyd Ung[1] and **Andre Marziali**[1,2]

[1] Boreal Genomics Inc, Vancouver, BC Canada, [2] University of British Columbia, Vancouver, BC Canada

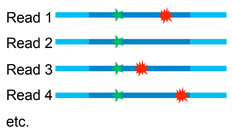
I. Proximity Sequencing (Pro-Seq)

We present a novel NGS library construction process that creates "self-checking" sequencing clusters for ultra-low sequencing error and *de novo* rare mutant detection with duplex information.

- See poster #712 for related Linked-molecule Target Capture.
- Both targeted (droplet) and whole genome (no droplet) versions of technology have been developed.
- Technology has been demonstrated on Illumina sequencing platforms, but is compatible with all SBS platforms.

Concept

- Barcoding error reduction methods require many (~20-100+) redundant clusters to create a consensus call in software after sequencing is complete.
- Pro-Seq duplicates each sense of the original DNA strand ("duplex" information is retained) with linked primers prior to library construction, eliminating the need for molecular barcodes and cost of redundant sequencing for error correction.



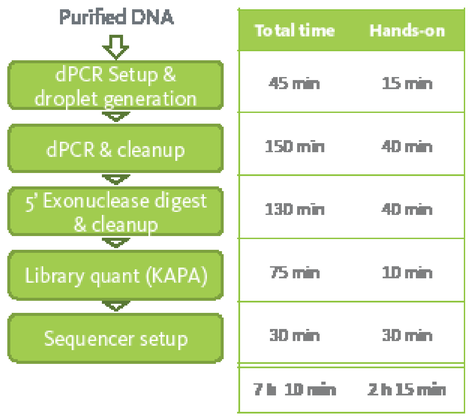
Molecular barcoding consensus calling requires many redundant reads for error correction

Pro-Seq performs error correction in a single sequencing cluster by linking copies of the template

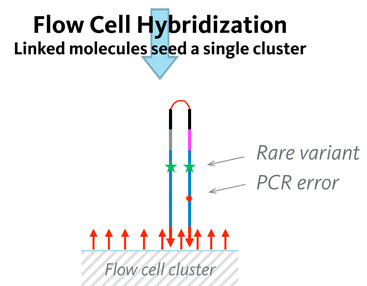
II. Targeted Pro-Seq Workflow

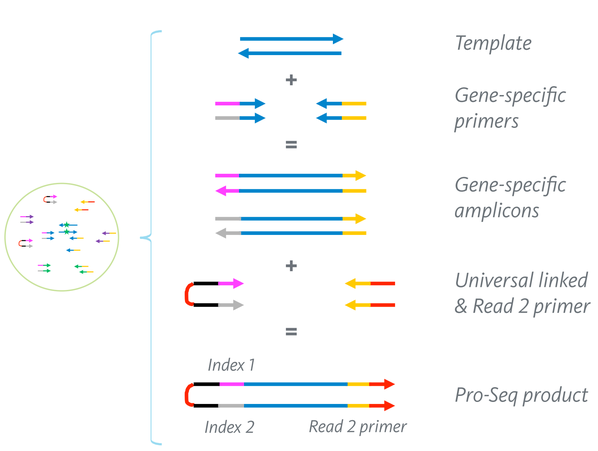
Pro-Seq offers a simple, single day sequencing workflow:

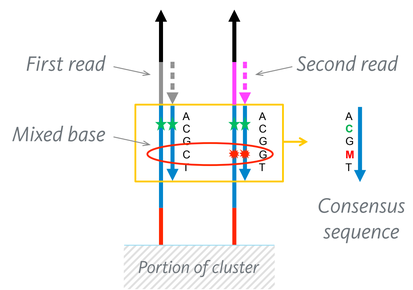
- Using a droplet generator, load one template of cell-free DNA per droplet with universal linked primers and multiplexed gene-specific primers
- Cycle reaction to saturation, so that linked complexes of two or more molecules are created (depending on type of linked primer), all from the same starting fragment
- Break emulsion and enzymatically digest un-linked molecules
- Sequencing clusters will be seeded by two or more linked strands originating from the same original strand. Library construction errors can be identified through the presence of mixed base signal from a single cluster, signifying that an error has been made at that position.



Total time	Hands-on
45 min	15 min
150 min	40 min
130 min	40 min
75 min	10 min
30 min	30 min
7 h 10 min	2 h 15 min



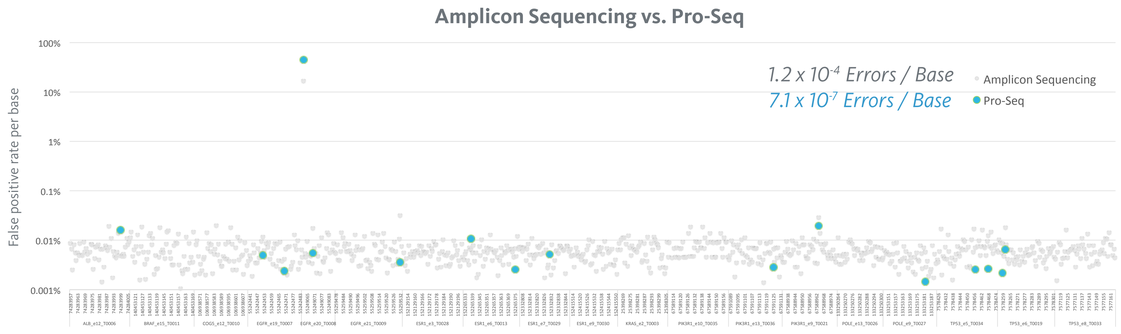




III. Results

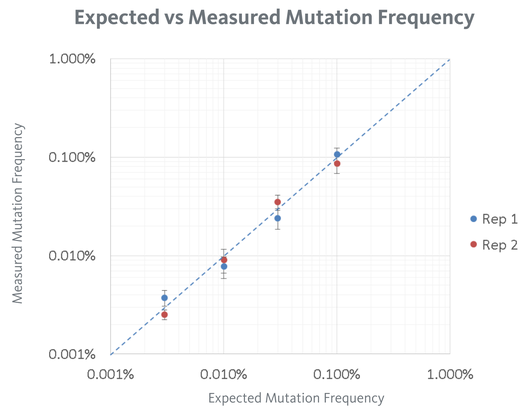
False Positive Rate Reduction

- Pro-Seq false positive rate is 7.1 x 10^{-7} (<0.0001%) on 19 amplicon panel and may be limited by true biological background mutations
- False positive rate improvement over amplicon sequencing is ~100–1,000-fold.



Mutation Detection

- 5 mutations titrated down from 0.1% to 0% (EGFR T790M, EGFR L858R, EGFR ΔE746-A750 KRAS G12D, PIK3CA E545K)
- To our knowledge, 0.003% is the lowest reported detection for an individual mutant
- Detection of mutations lower than this may be limited by biological background, but 0.003% is more than sufficient for a practical liquid biopsy



Molecular Sensitivity

- Molecular sensitivity of Pro-Seq is near single molecule and limited by sampling noise

Expected Number of Copies per Mutant	Expected Number of Mutants (both replicates)	Sampling-corrected Number of Mutants	Total Number of Mutants Detected	Fraction of Mutants Detected (sampling corrected)
45	10	10.0	10	100%
15	10	10.0	10	100%
4.5	10	9.9	7	71%
1.5	10	7.8	6	77%
0	0	0.0	0	0%

Sequencing Efficiency

- Pro-Seq has >2-fold higher sequencing efficiency than PCR or hybrid capture methods
- Unlike hybrid capture, Pro-Seq offers a simple one-day workflow that retains high on-target fraction with smaller panels

Input Mass (ng)	Reads (millions)	Sensitivity at AF %			Reads per target
		0.05%	0.1%	0.5%	
30	0.35	-	92%	100%	5
30	0.46	-	100%	100%	7
100	0.15	50%	-	100%	0.5
100	0.4	100%	-	100%	1

Colorectal Cancer Panel

- Hotspots of 5 genes: APC, BRAF, KRAS, PIK3CA, and TP53
- 20 amplicons covering ~1kb of total sequence with the ability to cover SNVs, indels and CNVs
- >83% on-target fraction, no PhiX required on sequencer (saving bandwidth)
- False positive rate measured at 8.6 x 10^{-6}
- High uniformity : Ratio of depth of highest and lowest amplicon is less than 2.5x