Duplex Proximity Sequencing (Pro-Seq): A cost-effective, high accuracy duplex sequencing method

Joel Pel[1], Wendy Choi[1], Milenko Despotovic[1], Laura Gelinas[1], Amy Leung[1], Gosuke Shibahara[1], Lloyd Ung[1] and Andre Marziali[1,2]

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

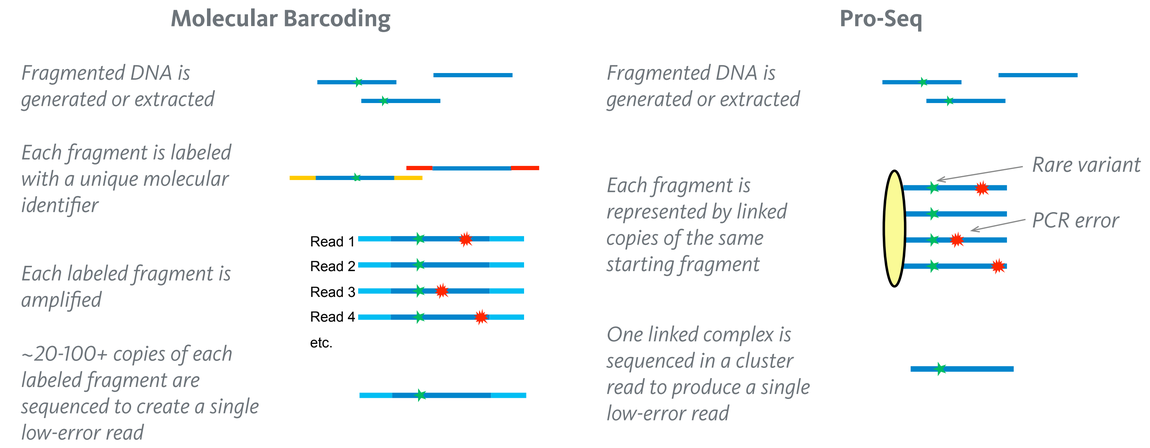
I. Proximity Sequencing (Pro-Seq)

We present a novel and scalable library construction process for NGS that increases sequencing accuracy and specificity of single reads >100-fold without additional reads and costs associated with consensus calling in barcoded sequencing. Pre-print publication in review - **doi:** https://doi.org/10.1101/163444

- See related poster for alternative Linked Molecule techniques as applied to whole genome sequencing, target capture and other workflows.
- Technology has been demonstrated on Illumina sequencing platforms, but is expected to be compatible with IonTorrent, GeneReader and other SBS platforms.

Concept

- Most NGS error reduction methods require reading many redundant clusters to create a consensus call in software after sequencing is complete.
- Pro-Seq duplicates the sequence of each sense of the original DNA strand ("Duplex" information is retained) prior to library construction to provide redundant, linked templates that form clusters together rather than separate clusters as required in barcoding methods.
- Effectively, Pro-Seq provides a "Q score" that describes the quality and accuracy of each base call as accumulated over the entire library construction process, thus allowing filtering of errors.

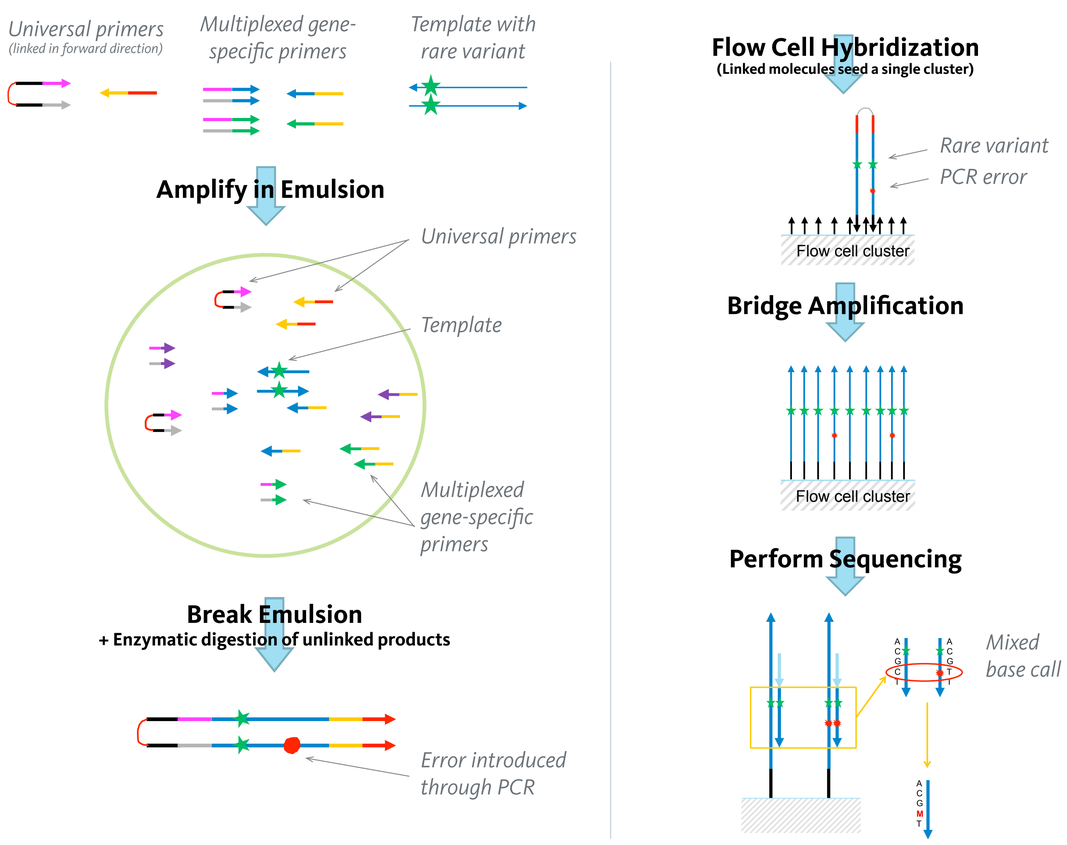


Consensus calling in Pro-Seq requires only a single read per template vs. many (6–100+) redundant reads with other error reduction methods

II. Droplet Pro-Seq Workflow

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

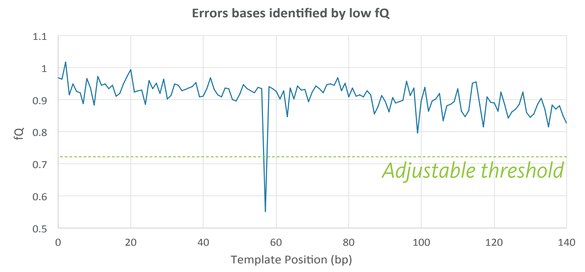
- Using a droplet generator, load one template of cell-free DNA per droplet with universal linked primers and multiplexed gene-specific primers
- Cycle reaction to saturation, so that linked complexes of two or more molecules are created (depending on type of linked primer), all from the same starting 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.



III. Results

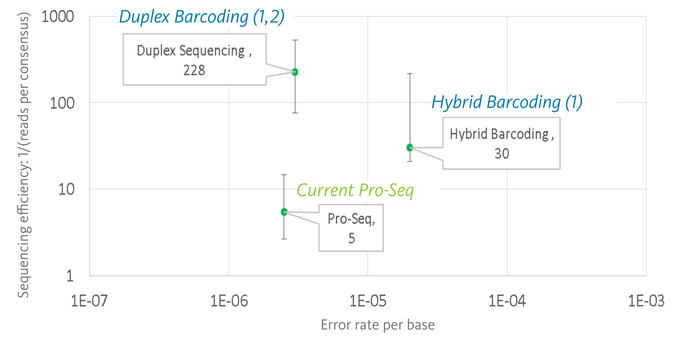
Error Identification

- Errors introduced through library construction and sequencing are identified in Pro-Seq clusters by measuring the fluorescent quality ratio (fQ)
- Bases with low fQ relative to surrounding bases are ignored as errors.



Current Sequencing Efficiency

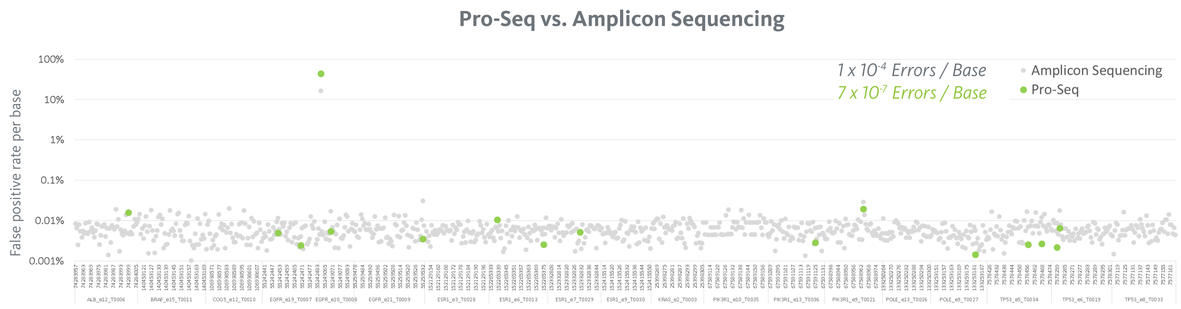
- Hybrid (1) and Duplex (2) barcoding require large redundancy to create a high fidelity consensus read (27 and >100 reads, respectively)
- Pro-Seq increases information density and achieves low error rate at much higher sequencing efficiency
- Pro-Seq efficiency still improving, other technologies limited by Poisson sampling



1. Newman, et al. Nat. Biotechnol 34, 547–555 (2016) 2. Krimmel, et al. Proc. Natl. Acad. Sci. (2016) – Supplementary Information

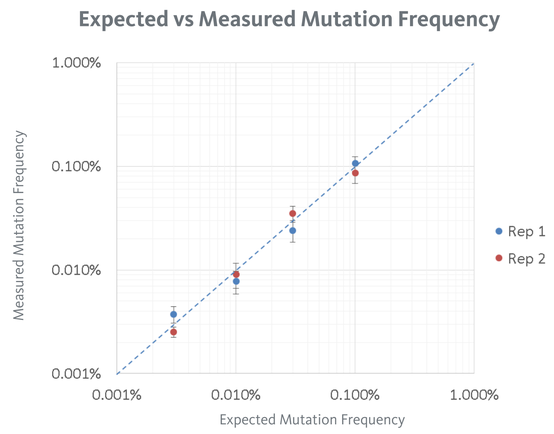
False Positive Rate Reduction

- Pro-Seq false positive rate measured to be 2.6 x 10^{-6} errors per base across 12 WT plasma samples, without in silico polishing or other computational techniques. The resulting analytical specificity is 99.9997%
- Pro-Seq false positive rate is 7 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%

Current Development Directions

- Custom disease or patient-specific panels with rapid turnaround, ranging in size from 10-50 amplicons for applications such as recurrence monitoring and therapy selection
- 10 kb panel in testing in conjunction with Boreal's Linked Target Capture Technology

With instrumentation and reagent support from RainDance Technologies Inc.