Linked Target Capture: a rapid target enrichment method for diverse applications in oncology, gene editing and single cell genomics

Joel Pel[1], Milenko Despotovic[1], Gosuke Shibahara[1], Lloyd Ung[1], Matthew Wiggin[1], Rolf Turk[2], Matthew McNeill[2], and Gavin Kurgan[2], and Andre Marziali[1,3]

[1] Boreal Genomics Inc, Vancouver, BC, Canada, [2] Integrated DNA Technologies, Inc., Coralville, IA, USA, [3] University of British Columbia, Vancouver, BC, Canada

I. Linked Target Capture (LTC)

- **Rapid:** Total library prep and target capture time is ~6 hours, ~2 hours hands-on, and is easily automated
- **Scalable:** Single-tube panels have been demonstrated from <50bp to >1Mb, with exome-sized panels expected
- **Efficient:** High uniformity and on-target regardless of panel size, leading to reduced sequencing costs
- **Sensitive:** High performance with input mass as low as 0.1 ng (30 genomes)
- **Precise:** Detailed molecular information, including start/stop coordinates, unknown fusions, and UMIs, are captured compared to amplicon workflows
- **Flexible:** broad applications include oncology & liquid biopsy, gene editing characterization and single cell cDNA enrichment with a machine learning design pipeline that rapidly delivers high performance panels

II. Workflow

Ligation

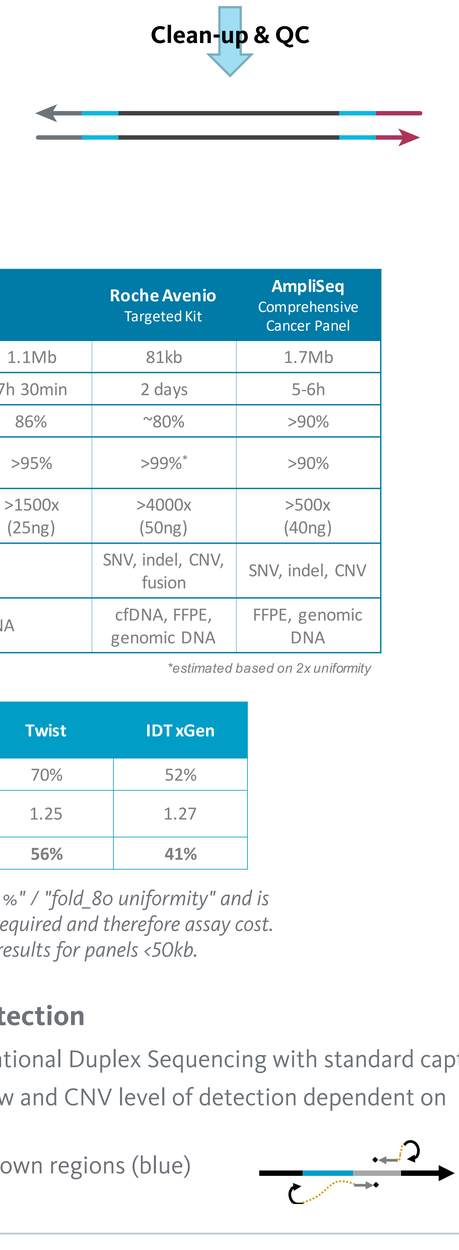
- Standard ligation with custom adapters and optional shearing, compatible with UMIs

Target Capture PCR

- Enrichment is achieved through proprietary Probe-Dependent Primers (PDPs) used in a PCR-like reaction
- Universal primer portion (light blue) of PDP only binds and extends if probe portion (black, blocked from extension) is bound to template
- Copies of each sense are collapsed into a single consensus read family at random to achieve low error rate (unlike Duplex Sequencing)
- Long captures and complicated pull downs are eliminated

Sequencing Clean-up & Quantification

- Library is cleaned up prior to quantification and then is ready for sequencing



Linked Target Capture

Purified DNA → Library prep (ligation) → Target Capture PCRs → Library quant (Qubit)

	Total time	Hands-on
Library prep (ligation)	120 min	45 min
Target Capture PCRs	241 min	80 min
Library quant (Qubit)	10 min	10 min
	6 h 11 min	2 h 15 min

IDT xGen

Purified DNA → Library prep (ligation) → Universal PCR → Target Capture → Capture Cleanup → Post-capture PCR → Library quant (Qubit)

	Total time	Hands-on
Library prep (ligation)	125 min	50 min
Universal PCR	80 min	50 min
Target Capture	285 min	50 min
Capture Cleanup	105 min	95 min
Post-capture PCR	150 min	40 min
Library quant (Qubit)	10 min	10 min
	12 h 35 min	4 h 55 min

III. Panel performance - Oncology applications

Performance Metrics	Boreal LTC Custom Oncology Panels			Roche Avenio Targeted Kit	AmpliSeq Comprehensive Cancer Panel
Panel size	11kb	166kb	1.1Mb	81kb	1.7Mb
Workflow time	6h 11min	6h 11min	7h 30min	2 days	5-6h
On-target %	92%	83%	86%	~80%	>90%
Coverage uniformity *% of bases with >0.2x mean coverage*	100%	>99%	>95%	>99%*	>90%
Mean de-duplicated coverage *input into ligation*	>5000x (50ng)	>4000x (50ng)	>1500x (25ng)	>4000x (50ng)	>500x (40ng)
Variant types	SNV, indel, CNV, fusion			SNV, indel, CNV, fusion	SNV, indel, CNV
Sample types	cfDNA, FFPE, genomic DNA			cfDNA, FFPE, genomic DNA	FFPE, genomic DNA

**estimated based on 2x uniformity*

Performance Metrics	Boreal LTC	Twist	IDT xGen
On-target %	90%	70%	52%
Fold_80 uniformity	1.4	1.25	1.27
Target enrichment efficiency	**64%**	**56%**	**41%**

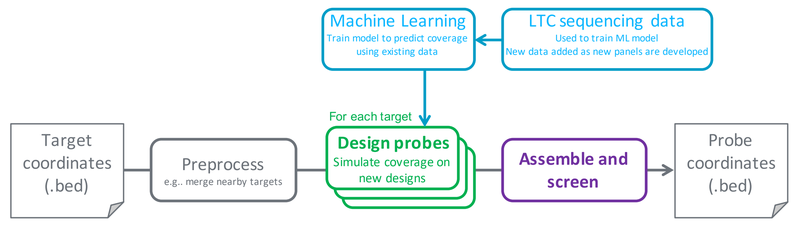
Target enrichment efficiency is defined as "on target %" / "fold_80 uniformity" and is inversely proportional to the amount of sequencing required and therefore assay cost. Data based on published and experimental results for panels <50kb.

Integration of UMIs enables high sensitivity variant detection

- Error rates below 1 x 10^{-6} are achievable, comparable to conventional Duplex Sequencing with standard capture
- SNV, fusion & indel detection is enabled down to 0.1% or below and CNV level of detection dependent on sequencing depth
- Novel fusion detection enabled by PDPs spanning across unknown regions (blue)

IV. Machine learning panel design pipeline

- Following the generation of oncology panel performance data presented here, Boreal has developed a new machine learning design pipeline that is expected to improve the on target and uniformity performance on all subsequent panels
- This pipeline enables rapid design of custom LTC panels for applications such as patient-specific monitoring of minimal residual disease (MRD)

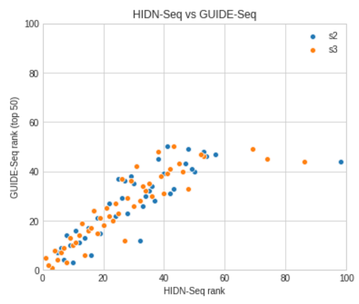


V. Gene editing characterization

CRISPR off-target detection assay

- Heterogeneously-Integrated DNA eNrichment (HIDN-Seq), an implementation of LTC, produces libraries with >95% of reads containing the tag sequence, while also providing insert sequence on both sides of the tag for improved identification of insert site
- Other methods, including GUIDE-Seq, typically produce ~20-40% tag-containing reads, while only providing sequence context on one side of the tag
- Using edited cell lines, HIDN-Seq was able to detect all top 50 insertion sites identified by GUIDE-Seq, while generating very few off-target reads (cell line DNA and Guide-Seq data provided by IDT – Coralville, IA).

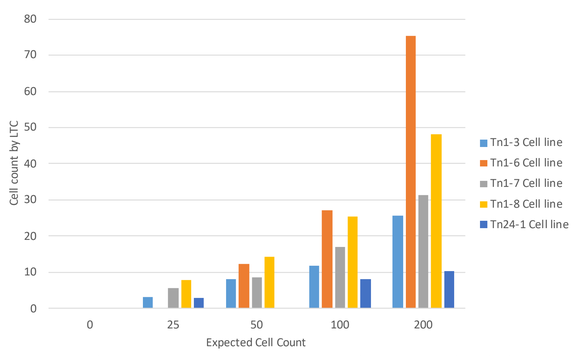
LTC can be used to efficiently determine on-target and off-target gene edits by enriching for inserted/edited sequences



LTC captures all top insertion sites identified by GUIDE-Seq with >95% of reads on target.

Synthetic transposon editing detection assay

- 5 human cell lines containing different known transposon integration sites were spiked into a background of mouse cells for a total of 7000 cells
- Each of 5 human cells types were spiked in at 0, 25, 50, 100 and 200 cells and subject to LTC enrichment for the transposon sequence.
- A simplistic analysis was conducted to discover integration sites, with detection demonstrated down to 25 cells (0.3%)
- The current analysis method is likely to miss targets since exact sequences matches were required, and strict thresholds were applied for UMI counting

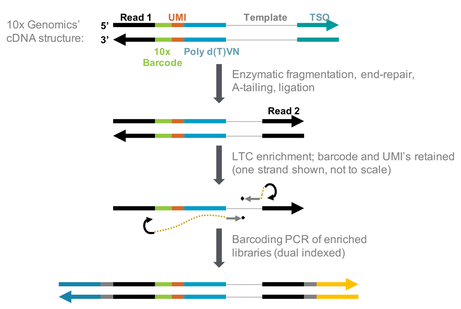


LTC can detect rare transposon insertion sites down to 25 cells spiked into a background of 7000 (0.3%) with no false positive calls (cell line DNA, expected insertion sites, and sequencing of LTC libraries provided by the Rossi Lab, UBC – Vancouver, BC)

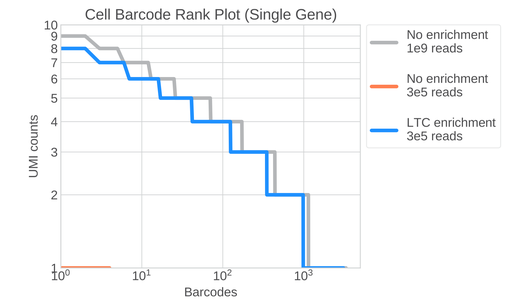
VI. Single cell rare transcript enrichment

Sequence enrichment from 10x Genomics cDNA libraries

- LTC enables the discovery of rare sequences that are missed by conventional transcriptome sequencing as their identification would take exceptional depth to observe
- LTC enrichment is performed off excess cDNA produced by the standard 10x Genomics Single Cell Gene Expression workflow, enabling enrichment of desired sequences before or after whole transcriptome sequencing has been performed
- LTC can be designed to enrich for few to many rare sequences while retaining cell barcode and UMI information



LTC enrichment and library generation using cDNA derived from 10x Genomics Single Cell Gene Expression workflow



LTC enriches for desired transcripts and reduces required sequencing >3000-fold compared to whole transcriptome sequencing (cDNA and whole transcriptome data provided by the Underhill lab, UBC – Vancouver, BC).

With reagent support from Integrated DNA Technologies Inc. (IDT)

Pel et al. PLOS ONE, Dec 2018. https://doi.org/10.1371/journal.pone.0208283