Ultra-low input linked-read library method generates highly accurate and economical long-range information from short reads for broad applications

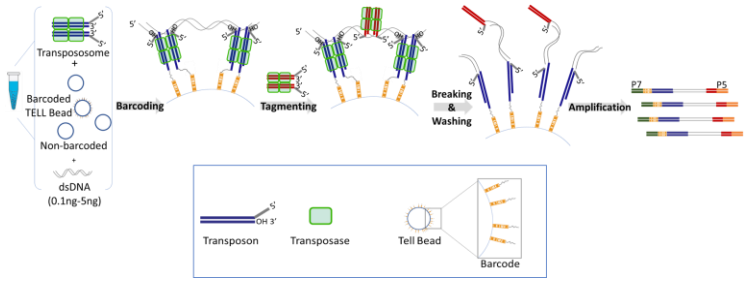
Tom Chen[1], Peter Chang[1], Long Pham[1], Donna Wu[1], Gloria Mo[1], Yu Xia[1], Chris Boles[2], Yong Wang[1]

1. Universal Sequencing Technology, Carlsbad, CA 92011 *2. Sage Science, Inc., Beverly, MA 01915* *email: tchen@universalsequencing.com*

Abstract

Transposase Enzyme Linked Long-read Sequencing (TELL-Seq™) is a simple and scalable NGS library technology to generate barcode linked-reads for genome scale haplotype phasing and de novo sequencing. In a PCR tube, under a standard NGS laboratory setting and without the need for any expensive protocol-specific instrument, TELL-Seq Whole Genome Sequencing (WGS) Library Prep Kit will generate an Illumina sequencing library in 3 hours from as low as 0.1ng DNA input. Multiple samples can be easily processed in parallel as needed. The library protocol can be adjusted and used for various sized genomes ranging from bacteria to human. The TELL-Seq library will replace both fragment library and mate-pair library altogether and become a new standard library method for WGS.

How does TELL-Seq Work?

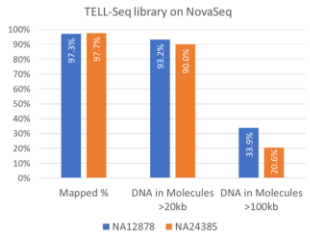


TELL-Seq WGS Library Workflow is Simple

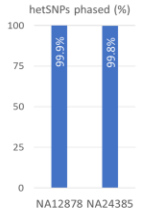
Genome Assembly

Sample	*Escherichia coli* MG1655	*Escherichia coli* DH10B	*Arabidopsis thaliana*	*Homo sapiens* NA12878
gDNA input (ng)	0.5	0.1	3	5
Average gDNA length (Kb)	**39**	**50**	**18**	**50**
Assembler	TuringAssembler		SuperNova	
Global/local k-mer sizes	105/69	105/31	N/A	N/A
Genome fraction (%)	99.9	99.3	97.4	93.7
Duplication ratio	1.009	1.012	1.061	1.075
Largest alignment	**4,637,716**	**3,459,910**	**5,613,314**	**23,573,913**
Total aligned length	4,678,679	4,705,787	119,712,720	2,771,334,486
NA50	**4,637,716**	**3,459,910**	**1,272,225**	**4,302,918**
# misassemblies	0	1	246	1,987
# mismatches per 100 kbp	5.37	12.31	14.17	114.07
# indels per 100 kbp	0.39	0.24	5.96	25.61
# N's per 100 kbp	0.0	0.0	4,499.1	6419.8
# contigs (>= 5000 bp)	3	2	382	13,584
# contigs (>= 10000 bp)	3	2	200	7,508
Largest contig	**4,640,256**	**4,657,354**	**10,292,118**	**109,183,970**
Total length (>= 1000 bp)	4,725,322	4,822,052	126,210,917	2,979,445,840
N50	**4,640,256**	**4,657,354**	**4,677,832**	**31,462,027**
GC (%)	50.75	50.67	36.29	40.91

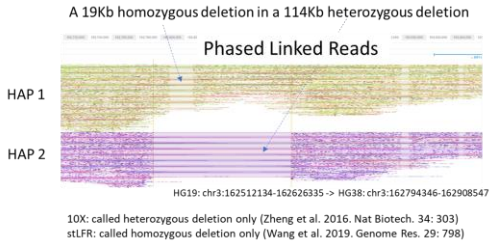
Whole Genome Phasing



TELL-seq Human Phasing	2×146 PE	
	NA12878	NA24385
Coverage Depth (Unique)	38×	46×
Longest Phased Block	67.5 Mb	59.2 Mb
N50 Phased Block	14.4 Mb	13.4 Mb
Switch Error Rate	0.04%	0.08%

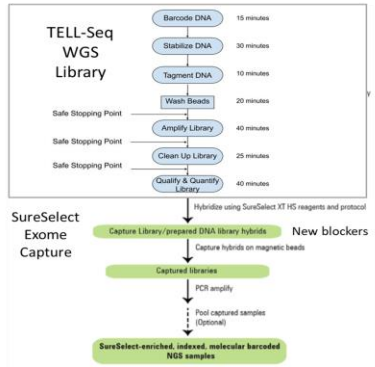


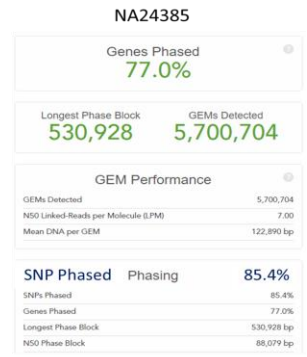
Structure Variation Detection



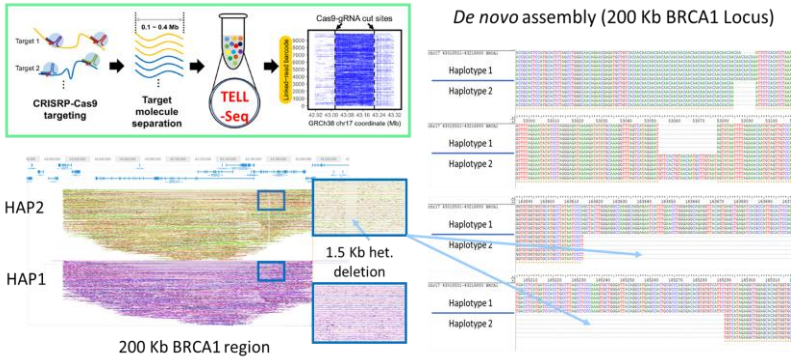
Technical Indices	TELL-Seq™
Instrument cost	Zero
Library prep cost	Low
Input DNA	0.1 ng to 5 ng
Processing time	3 hours
Barcode capacity	>2 billion
Genome size	≥ 0.1 Mb

Whole Exome Phasing





Targeted Full-Length Gene Diplotyping



References

Chen, Z et al. 2020. Ultralow-input single-tube linked-read library method enables short-read second-generation sequencing systems to routinely generate highly accurate and economical long-range sequencing information. ***Genome Research*** 30: 898-909.

UNIVERSAL SEQUENCING innovation for all