TELL-Seq[™]: Ultra-low Input Single Tube Linked-Reads for Accurate Long-range Sequencing with Short Reads

Tom Chen¹, Long Pham¹, Donna Wu¹, Gloria Mo¹, Yu Xia¹, Peter Chang¹, Tan Phan², Huu Che², Nathan Houde³, Chris Boles³, Son Pham², Yong Wang¹, Ming Lei¹
1. Universal Sequencing Technology Corporation, Carlsbad, CA 92011
2. Bioturing Inc, San Diego, CA 92121
3. Sage Science, Inc., Beverly, MA 01915

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?

Transposase Enzyme Linked Long-Read Sequencing Mu Transpososc **Breaking** Barcoded TELL Bead Barcoding Amplificatio & Washing Non-barcoded $\bigvee \bigvee \bigvee \bigvee$ dsDNA Tell Bead 1 Imil Transposase (0.1ng-5ng)Transposon Barcode

TELL-Seq WGS Library Workflow is Simple



Whole Exome Phasing



^{cted}
5,700,704
7.00
122,890 bp

SNP Phased Phasing	85.4%
SNPs Phased	85.4%
Genes Phased	77.0%
Longest Phase Block	530,928 bp
N50 Phase Block	88,079 bp



TELL-Seq Library Structure



Sample	Escherichia coli	Escherichia coli	Arabidopsis	Homo sapiens
	MG1655	DH10B	thaliana	NA12878
gDNA input (ng)	0.5	0.1	3	5
Average gDNA length (Kb)	39	50	18	50
Assembler	TuringA	TuringAssembler		erNova
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

Targeted Full Gene Diplotyping

GEMs

N50 Li

Mean [

Target isolation by Sage HLS-CATCH



De novo assembly (200 Kb BRCA1 Locus)





Structural Variation



10X: called heterozygous deletion only (Zheng et al. 2016. Nat Biotech. 34: 303) stLFR: called homozygous deletion only (Wang et al. 2019. Genome Res. 29: 798)

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	



www.universalsequencing.com