TELL-Seq AppNote:

Highly Accurate Haplotype Phasing of PCR Products

UNIVERSAL SEQUENCING
innovation for all

For Research Use Only.
Not for use in diagnostic procedures.

February 2022
Document # 100030, version 1.0

Introduction

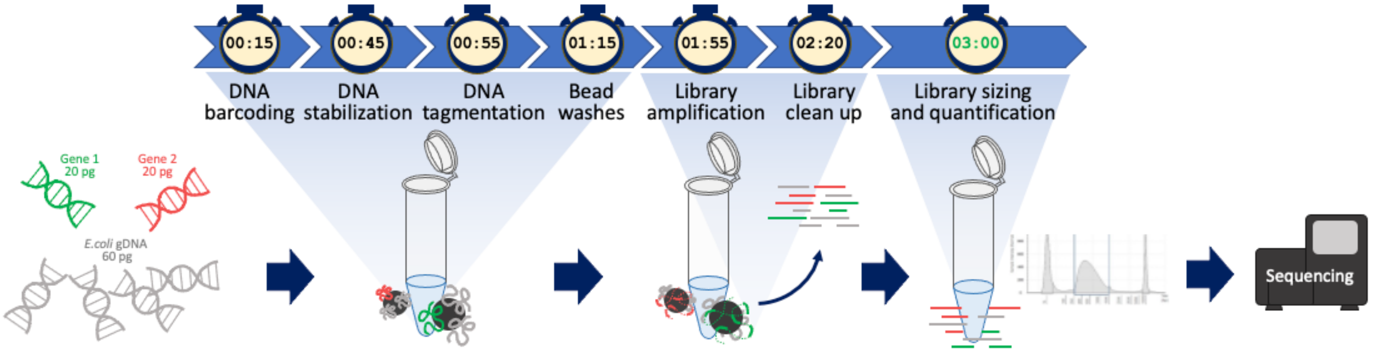
Every individual human genome differs from the reference human genome at 4.1 to 5.0 million sites across all chromosomes (The 1000 Genomes Project Consortium, 2015). However, most of the sequence differences (*alternate* alleles) are found in only one of the two copies for each chromosome—either the maternal or paternal copy (i.e., *heterozygous*). Determining whether multiple alternate alleles are in the same or in a different chromosomal copy—a process known as *haplotype phasing*—is critical for many research and clinical applications.

Widely used second-generation sequencing technology produces highly accurate but relatively short reads. Short reads rarely contain more than one heterozygous (het) site due to the relatively low frequency of these positions. Not having two het sites in one read makes the process of haplotype phasing unfeasible. Other third-generation continuous long-read platforms enable phasing on heterozygous sites located kilobases (kb) apart, but at the expense of sensitivity and fidelity, or at a higher cost.

To add long-read phasing capabilities to short-read next-generation sequencing (NGS) instruments, such as Illumina® systems, Universal Sequencing Technology (UST) has developed the TELL-Seq™ whole-genome sequencing (WGS) Library Prep kit. TELL-Seq (transposase enzyme-linked long-read sequencing) is a linked-read method that captures long-range information in short reads during NGS library preparation (Chen et al., 2020). TELL-Seq enables the rapid and efficient phasing of entire genomes using high-molecular-weight (HMW) DNA—50 to 200 kilobases (kb). The question has remained whether the TELL-Seq WGS Library Prep kit could also rapidly and efficiently phase relatively short DNA fragments, such as PCR products in the 1 kb to 15 kb range.

In this Application Note, we demonstrate that the TELL-Seq WGS Library Prep kit (a 3-hour protocol; workflow shown below) can phase PCR products of varying sizes. For our validation tests, we turned to clinically relevant examples. First, we sought an example of a *composite* mutation (i.e., when a gene has two or more mutations), which is relatively common in cancer (Gorelick et al., 2020). We phased a double *PIK3CA* mutation in 1.8 and 3.4 kb fragments amplified from breast cancer cells (Vasan et al., 2019). Second, we sought an example of genome-wide association study (GWAS) hits, which often emerge as clusters. We phased between four and twenty-three genomic sites in 2.7, 3.9, 4.1, and 13 kb fragments amplified from the *SCN10A* locus. Seven of these positions belong to a GWAS hit cluster in studies of sudden cardiac death (Pinsach-Albuin et al., 2020). *PIK3CA* and *SCN10A* libraries were sequenced on an Illumina MiSeq™ instrument and read outputs were processed using UST software (Tell-Read and Tell-Sort). Data visualization was based on the Integrative Genomics Viewer (IGV, Broad Institute). Taken together, our data show robust phasing of clinically relevant PCR products using the TELL-Seq WGS Library Prep kit with highly accurate and low-cost NGS data collection.

TELL-Seq WGS Library Prep workflow



Methods

DNA sources and preparation

Human epithelial breast cancer HCC202 cells were purchased from ATCC® (Cat. #CRL-2316) and used as cDNA source for the generation of *PIK3CA* PCR products. RNA was obtained using the *Quick*-RNA Miniprep Kit™ (Cat. #R1054, Zymo Research®), and 1.5ug of RNA was processed using the SuperScript™ III First-Strand Synthesis System and 1ul of random hexamers or oligodT primer, as indicated, to generate cDNA (Cat. #18080051; ThermoFisher Scientific®). We followed the manufacturer's instructions.

Human lymphoblastoid GM12878 cells were purchased from Coriell Cell Repositories® (Cat. #GM12878) and used as gDNA source for the generation of *SCN10A* PCR products. Genomic DNA (NA12878) was extracted from cultured cells using a salting out method that preserves DNA integrity. We note that alternative gDNA extraction methods could also be used. The protocol we followed can be downloaded from our website (TELL-Seq Demonstrated Protocol).

Bacterial and viral gDNAs were used to compete with human DNA in TELL-Seq reactions. The *Escherichia coli* DH10B (bacterial) strain was purchased from New England Biolabs® (Cat. #FEREC0113), and gDNA was extracted using the *Quick*-DNA Miniprep Kit (Zymo Research, Cat. #D3024). BstI P-digested lambda phage gDNA was purchased from TaKaRa® (Cat. #3402), and used directly.

PCR amplification

Primers were synthesized at Integrated DNA Technologies® and diluted to 10uM in 0.1x TE buffer. Primers (5'->3') for 3.4kb *PIK3CA*: Fw-TGGGACCCGATGCGGTTA and Rv-AATCGGTCTTTGCCTGCTGA (Vasan et al., 2019). Primers for 1.8kb *PIK3CA*: Fw-CAGACGCATTTCCACAGCTA and Rv-TGTGACGATCTCCAATTCCCA. Primers for 13kb *SCN10A* Fw-GCCATGACCATTGTTATTTGTCCAGA and Rv-CCTGAAGAAATGTCACGGCTTGTTAG (Pinsach-Albuin et al., 2020). Primers for 3.9kb *SCN10A*: Fw-CACTTTGCACGAAGTGCTTG and Rv-GCCCACACACCTCTCTTCAT. Primers for 4.1kb *SCN10A*: Fw-GTGTGGGCTCTTGCTCTCAT and Rv-GAGGTGGGAGGATGACTTGA. Primers for 2.7kb *SCN10A*: Fw-TGTAATTTCTGCAGCCACGA and Rv-CACTGGTTTCCCATTGCTCT.

We followed PCR conditions previously established to amplify 3.4 kb *PIK3CA* fragments (Vasan et al., 2019), and used the same conditions to amplify 1.8 kb *PIK3CA* fragments. PCR was conducted using Phusion Hot Start II High-Fidelity polymerase (Cat. #F-565L, Thermo-Fisher Scientific) and 2 ng of HCC202 cDNA. Cycling protocol: 98°C for 30 sec; 30 cycles of 98°C for 10 sec, 65°C for 20 sec, and 72°C for 1 min; final step at 72°C for 8 min. PCR products were cleaned up with ExoSap-IT (Cat. #78200.200.UL; Thermo-Fisher Scientific®).

We followed PCR conditions previously established to amplify 13kb *SCN10A* fragments (Pinsach-Albuin et al., 2020), and used the same conditions to amplify three non-overlapping internal regions: 3.9, 4.1, and 2.7 kb. PCR was conducted using NA12878 gDNA and Supreme NZYLong DNA Polymerase (Cat. #MB331, NZYTech®) according to the manufacturer's recommendations. Cycling protocol: 94°C for 5 min; 30 cycles of 94°C for 20 sec, 68°C for 30 sec, and 68°C for 14 min; final step at 68°C for 21 min. PCR products were processed with ExoSap-IT and two rounds of 0.41x HighPrep™ PCR Clean-up beads (Cat. #AC-60050; MagBio Genomics®).

PCR products were assessed by gel electrophoresis and quantified using the Qubit™ dsDNA HS Assay kit (Cat. #Q32854, Thermo-Fisher Scientific).

Library preparation and sequencing

TELL-Seq libraries were generated using the TELL-Seq WGS Library Prep kit developed by UST. We followed the instructions provided in a modified protocol for amplicons that can be downloaded from the UST website (TELL-Seq Amplicon-based Library Prep User Guide). TELL-Seq libraries were generated combining 20 pg of a single *PIK3CA* or *SCN10A* PCR product, 60 pg *E.coli* gDNA or 80 pg BstP I-digested lambda phage gDNA, and 3 million (M) TELL beads, as indicated. When testing a pool of *PIK3CA* and *SCN10A* amplicons, we combined the same amounts of fragments (in total, 80 pg) and 3M TELL beads (without adding non-human material as source of competitor DNA). In all tests, we used approximately 75,000 TELL beads for indexing (taking 1/40th of the processed bead solution; 18-21 PCR cycles). Libraries (2×145 bp illumine-compatible paired-end reads) were sequenced on a MiSeq instrument (Illumina) using the MiSeq 200 Micro kit v2 (Cat. #MS-103-1002, Illumina). Sequencing depths will be indicated for every experiment.

Data analysis and visualization

For sample demultiplexing and QC reporting, sequencing output files were processed with Tell-Read software developed by UST. Mapping, variant calling, and phasing were performed using Tell-Sort (UST), which incorporates BWA-MEM for read alignment, GATK-v4.0.3.0/HaplotypeCaller for variant calling (Broad Institute), and HapCUT2 (github/Vibansal) for phasing. The human genome (hg38) was used as reference for the analysis of *SCN10A* products. The targeted cDNA sequence from hg38 was used as reference for the analysis of *PIK3CA* products. Allele frequency threshold was set to 0.1. The Integrative Genomics Viewer, IGV-v2.11.1 (Broad Institute) was used for data visualization assisted with WhatsHap-v1.1 (github/whatshap) and UST software that can be downloaded from our website (Data Visualization on IGV).

Results

Phasing a double mutation in 1.8 and 3.4 kb fragments of amplified *PIK3CA* cDNA

To test the phasing capabilities of the TELL-Seq WGS Library Prep kit using PCR products, we first turned to the *PIK3CA* gene. *PIK3CA* is the second most frequently mutated gene across all human cancer types, mutated at least twice in the same tumor in 8-13% of clinical cases (Bailey et al., 2018; Vasan et al., 2019; Gorelick et al., 2020). The phasing of double *PIK3CA* mutations has prognostic value and is predictive of susceptibility to treatment (Vasan et al., 2019). A double *PIK3CA* mutation can be found in the human breast cancer HCC202 cell line: E545K (GAG->AAG) and L866F (TTG->TTC) separated by 964 bp but on a different chromosome copy (i.e., *in trans* configuration). HCC202 cells also carry an alternate allele (SNV) at the rs2230461 position, located 459 bp upstream of the E545K mutation, I391M (ATA->ATG; A is the reference allele and G is the alternate allele). E545K and I391M are carried by the same chromosomal copy (i.e., *in cis* configuration), as shown in **Figure 1** (Vasan et al., 2019). But I391M has not known clinical relevance.

Using HCC202 cDNA as DNA source, we amplified a 3.4 kb fragment containing all *PIK3CA* exons, including rs2230461 and the two mutations (**Figure 1**). We also generated a shorter 1.8 kb amplicon that contains fewer *PIK3CA* exons but still includes the three het sites (**Figure 1**). We generated six TELL-Seq libraries (in two rounds of three technical replicates, n = 2 x 3) using 20 pg of the 3.4 kb *PIK3CA* amplicon and 60 pg of competitor *E.coli* gDNA. Using Tell-Sort, sequencing outputs were aligned to the targeted region (*PIK3CA* exons) and followed variant calling and phasing after setting the allele frequency (AF) cutoff at 0.1. Finally, data visualization was based on IGV tracks (Broad Institute) generated with UST software that show phased sites as part of phase blocks (see Methods).

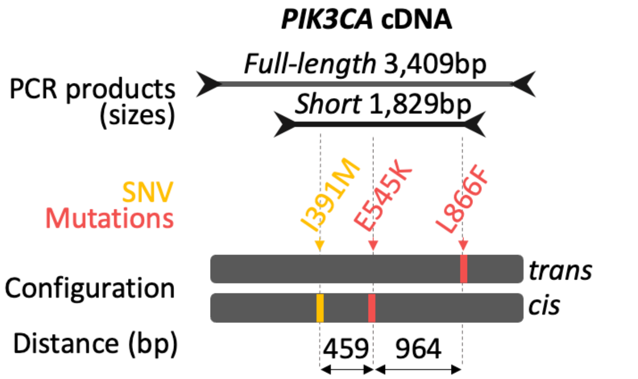


Figure 1. *Cis and trans configurations of three alternate alleles across the amplified PIK3CA coding region in breast cancer HCC202 cells. Distances indicated in bp. 1.8kb and 3.4kb amplicons used for phasing with TELL-Seq also indicated.*

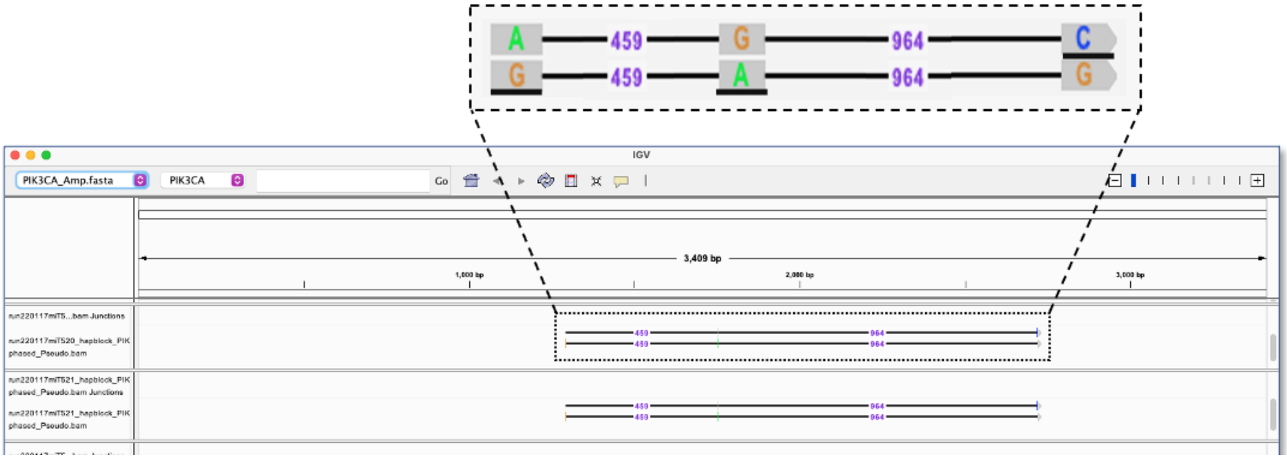

A
459
G
964
C
G
459
A
964
G
IGV
PIK3CA_Amp.fasta
PIK3CA
Go
3,409 bp
1,000 bp
2,000 bp
3,000 bp
run220117miT5...bam Junctions
run220117miT520_hapblock_PIK phased_Pseudo.bam
run220117miT521_hapblock_PIK phased_Pseudo.bam Junctions
run220117miT521_hapblock_PIK phased_Pseudo.bam
459
964

Figure 2. *TELL-Seq analysis of PIK3CA 3.4 kb PCR products amplified from HCC202 cDNA. Screenshot from IGV portal showing two phased.pseudo.bam tracks. Both tests show the same result, shown on top (magnification); distances between het sites indicated in bp). Targeted region (hg38) used as reference.*

Figure 2 (top) shows an example of correctly phased mutations with expected phasing GAG and AGC (alternate alleles highlighted in red). The three heterozygous sites were correctly recalled in the six tests in the absence of false positive calls (100% genotype re-call accuracy). Moreover, the three heterozygous sites were phased as expected in every case (AGC and GAG, according to Vasan et al., 2019), representing a 100% phasing recall accuracy (**Table 1**, *PIK3CA* amplicons section, C#1 column).

The six tests described above were based on cDNA primed with random hexamers. Alternatively, we sought to phase *PIK3CA* fragments using cDNA primed

	***PIK3CA* amplicons**							***SCN10A* amplicons**						
Sample characteristics	C#1	C#2	C#3	C#4	C#5	C#6	C#7	C#8	C#9	C#10	C#11	C#12	C#13	C#14
Priming method for cDNA (R6 vs dT)	R6	R6	R6	R6	dT	dT	dT	--	--	--	--	--	--	--
DNA source (genomic, gen.)	--	--	--	--	--	--	--	gDNA	gDNA	gDNA	gDNA	gDNA	gDNA	gDNA
PCR product size (kb)	3.4	3.4	1.8	1.8	3.4	3.4	3.4	13	2.7	2.7	3.9	3.9	4.1	4.1
Target input amount (pg)	20	20	20	5	20	5	0.4	20	20	20	20	20	20	20
Target fraction over total DNA	1/4	--	1/4	1/14	1/5	1/20	1/250	1/4	1/4	--	1/4	--	1/4	--
E.coli gDNA amount (pg)	60	--	60	65	--	--	--	60	60	--	60	--	60	--
Lambda phage gDNA amount (pg)	--	--	--	--	80	95	99.6	--	--	--	--	--	--	--
Other human fragments (pg)	--	60	--	--	--	--	--	--	--	60	--	60	--	60
Data analysis														
Technical replicates	**6**	**6**	**3**	**3**	**3**	**2**	**2**	**6**	**3**	**6**	**3**	**6**	**3**	**6**
Average on-target coverage (x)	213	257	3,574	834	423	73	58	480	307	728	301	397	58.8	124
Average on-target, dedup reads (number)	6.0k	7.2k	5.5k	12.8k	12.4k	2.1k	1.6k	57.8k	6.8k	15.9k	9.9k	12.6k	2.1k	4.2k
Expected hets per test	3	3	3	3	3	3	3	23	4	4	6	6	8	8
Tests w/all expected hets recalled correctly	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Tests w/all expected hets phased correctly	**100%**	**100%**	**100%**	**100%**	**66%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

Table 1. *Phasing of alternate/reference alleles in PIK3CA and SCN10A amplicons generated from breast cancer HCC202 cDNA (n = 25 tests) and NA12878 gDNA (n = 33 tests). Highlighted in red the number of technical replicates, and genotyping and phasing accuracies. R6 (random hexamers) and dT (oligo-dT) for cDNA priming, as indicated.*

with oligo-dT. In these new tests, we employed digested lambda gDNA as competitor DNA and titrated the amount of human amplicon to up to 1/250th of the total DNA input. After sequencing the (seven) new libraries, we obtained the expected phasing in all but one test (results shown in **Table 1**, *PIK3CA* amplicons section, C#5-7 columns). For the failed test, we note that two exact replicates showed the correct phasing.

Lastly, we processed six new TELL-Seq libraries using the shorter 1.8 kb fragment as input DNA. The three het sites were also called and phased as expected in all cases (**Table 1**, C#3-4 columns). Taken together, the TELL-Seq Library Prep kit was able to successfully phase the two *PKI3CA* mutations in 18 out of 19 tests.

Phasing of seven clinically relevant polymorphic sites in *SCN10A*-containing amplicons of different sizes (up to 13 kb)

To test the phasing capabilities of the TELL-Seq WGS Library Prep kit with larger PCR products, we amplified a 13 kb region from the *SCN10A* locus that contains seven clinically relevant, separated polymorphic sites associated with risk for heart arrythmia and sudden cardiac death—shown as red bars in **Figure 3** (Pinsach-Albuin et al., 2020). According to published phasing data (Zook et al., 2016; Chen et al., 2020), NA12878 carries the following two allelic combinations (haplotypes), TACCATT and CGGGGGC—known as Hap1 and Hap3, respectively (Pinsach-Albuin et al., 2020). Besides these seven heterozygous sites, however, NA12878 gDNA contains sixteen additional het positions along the 13 kb *SCN10A* region, all but one in *cis* configuration (**Figure 3**).

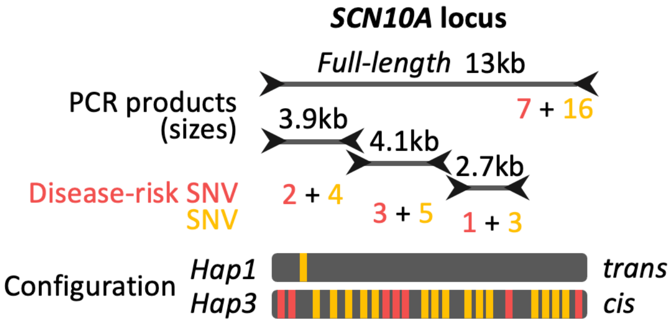


***Figure 3**. Cis and trans configurations for n = 23 alternate alleles (red/orange bars) in a 13 kb region derived from the SCN10A locus in NA12878 gDNA. Yellow bars indicate seven clinically relevant positions; their sequence and configuration correspond to Hap1 and Hap3 according to Pinsach-Albuin et al. (2020). Four PCR products are indicated on top.*

We amplified the 13 kb fragment from NA12878 gDNA and processed 20 pg with the TELL-Seq WGS Library Prep kit together with 60 pg of *E. coli* gDNA and 3M TELL beads. We also processed shorter *SCN10A* fragments containing six, eight, and four heterozygous sites (**Figure 3**). In total, we sequenced n = 15 libraries. After sequencing, using GRCh38 (hg38) as reference for mapping, we recalled the twenty-three positions expected heterozygous sites in every case (100% re-call accuracy). In addition, Tell-Sort was able to recall the expected phase for the twenty-three positions with 100% accuracy and all in a single *phase block*—a continuous set of phased positions (i.e., without intermediate gaps). **Figure 4** shows IGV tracks with the twenty-three phased alleles (**Table 1**, *SCN10A* section, C#8,9,11,13 columns). Taken together, the TELL-Seq Library Prep kit was able to successfully phase the twenty-three polymorphic sites along the *SCN10A* locus in 15 out of 15 tests.

Multiplexing TELL-Seq reactions with an amplicon panel

As described in the Methods section, we routinely include *E. coli* or lambda phage gDNA in TELL-Seq reactions to decrease the likelihood of two human DNA molecules sharing the same TELL bead. We next sought to determine whether using amplicons with different sequences could be an alternative competition method that reduces the phasing cost and labor. Thus, we pooled four PCR products (3.4 kb *PIK3CA* and 2.7, 3.9, and 4.1 kb *SCN10A*, 20 pg each) prior to TELL-Seq processing. Analysis of six technical replicates revealed 100% phasing recall accuracy for every fragment in every replicate (**Table 1**, C#2, 10,12,14 columns).

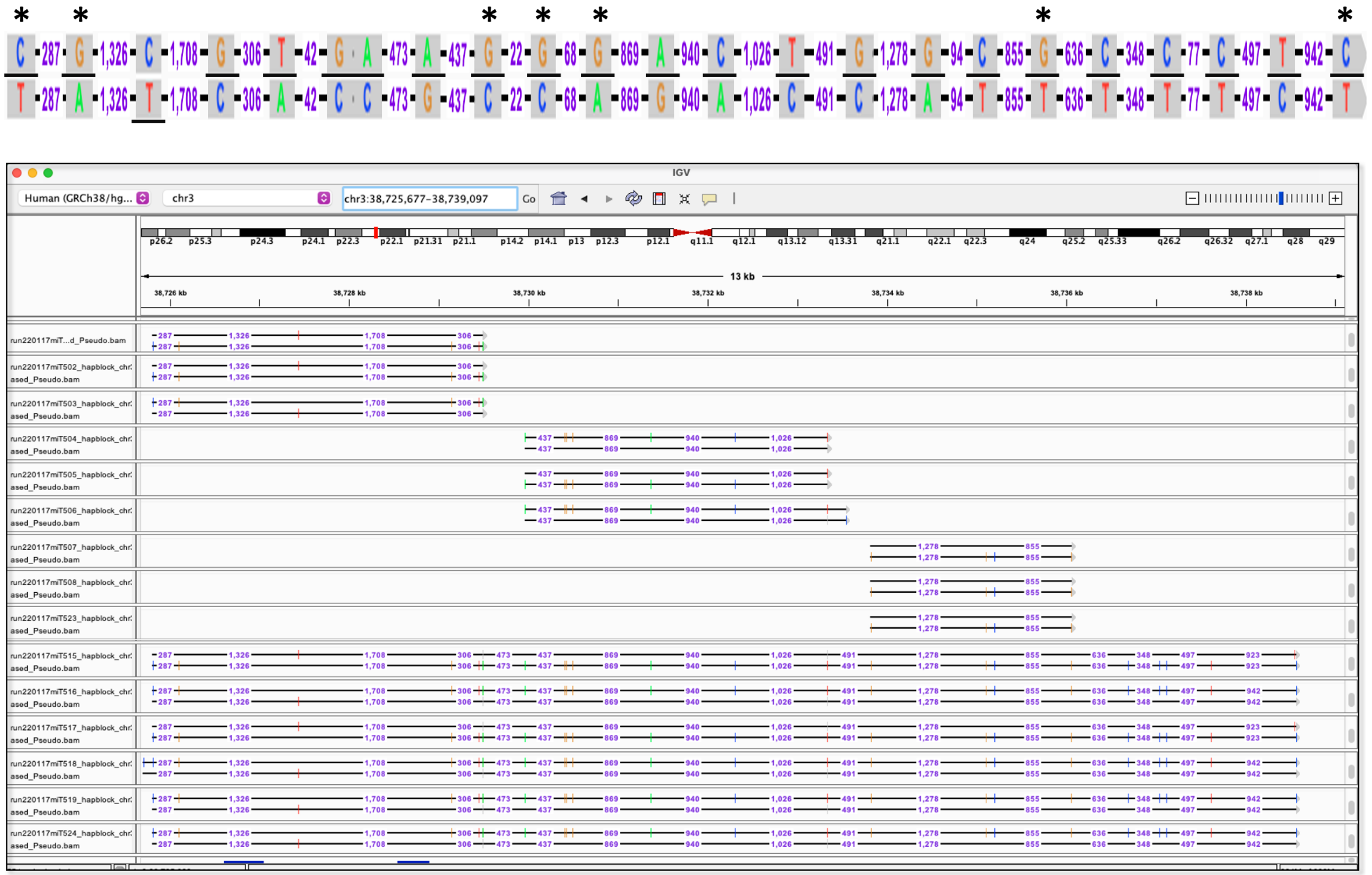

* * * * * * *
C 287 G 1,326 C 1,708 G 306 T 42 G A 473 A 437 G 22 G 68 G 869 A 940 C 1,026 T 491 G 1,278 G 94 C 855 G 636 C 348 C 77 C 497 T 942 C
T 287 A 1,326 T 1,708 C 306 A 42 C C 473 G 437 C 22 C 68 A 869 G 940 A 1,026 C 491 C 1,278 A 94 T 855 T 636 T 348 T 77 T 497 C 942 T
IGV
Human (GRCh38/hg...
chr3
chr3:38,725,677-38,739,097
Go
p26.2 p25.3 p24.3 p24.1 p22.3 p22.1 p21.31 p21.1 p14.2 p14.1 p13 p12.3 p12.1 q11.1 q12.1 q13.12 q13.31 q21.1 q22.1 q22.3 q24 q25.2 q25.33 q26.2 q26.32 q27.1 q28 q29
13 kb
38,726 kb
38,728 kb
38,730 kb
38,732 kb
38,734 kb
38,736 kb
38,738 kb
run220117miT...d_Pseudo.bam
run220117miT502_hapblock_chr...ased_Pseudo.bam
run220117miT503_hapblock_chr...ased_Pseudo.bam
run220117miT504_hapblock_chr...ased_Pseudo.bam
run220117miT505_hapblock_chr...ased_Pseudo.bam
run220117miT506_hapblock_chr...ased_Pseudo.bam
run220117miT507_hapblock_chr...ased_Pseudo.bam
run220117miT508_hapblock_chr...ased_Pseudo.bam
run220117miT523_hapblock_chr...ased_Pseudo.bam
run220117miT515_hapblock_chr...ased_Pseudo.bam
run220117miT516_hapblock_chr...ased_Pseudo.bam
run220117miT517_hapblock_chr...ased_Pseudo.bam
run220117miT518_hapblock_chr...ased_Pseudo.bam
run220117miT519_hapblock_chr...ased_Pseudo.bam
run220117miT524_hapblock_chr...ased_Pseudo.bam

Figure 4. *TELL-Seq analysis of SCN10A PCR products amplified from NA12878 gDNA. Fragment sizes: 3.9 kb (tracks 1-3), 4.1 kb (tracks 4-6), 2.7 kb (tracks 7-9) in triplicate (n = 3 replicates each), and 13 kb (tracks, 10-12; n = 6 replicates). On top, a representative example of n = 23 phased positions (100% genotype and phasing accuracy). Clinically relevant alleles (*) and alternate alleles are indicated (_). TACCATT corresponds to Hap1 (all-reference haplotype, bottom copy) and CGGGGGC corresponds to Hap3 (all-alternate haplotype, top copy). Screenshot shows phased.pseudo.bam tracks on IGV with the correct phasing for all the het positions regardless of fragment size.*

Summary

The TELL-Seq WGS Library Prep kit developed by UST enables haplotype phasing with the accuracy of Illumina sequencing (Chen et al., 2020). TELL-Seq is a rapid library preparation method that shows outstanding phasing performance with picogram amounts of input HMW DNA—required for *de novo* assembly of entire genomes (Chen et al., 2020). This Application Note shows similar performance on clinically relevant PCR products in the range of 2.7 to 13 kb. As a proof of principle, we have successfully phased a double mutation in the *PIK3CA* gene—one of the most frequently mutated genes in solid tumors—that has prognostic value and is used as a biomarker for therapy response (Vasan et al., 2019; Gorelick et al., 2020). We also show 100% phasing accuracy of haplotypes in the *SCN10A* locus, associated with risk for a life-threatening heart arrythmia (Pinsach-Abuin et al., 2020). Finally, the modified version of the TELL-Seq WGS Library Prep protocol used for these tests is available on our website (TELL-Seq Amplicon-based Library Prep User Guide).

Learn more

About the TELL-Seq technology:
www.universalsequencing.com/technology
www.youtube.com/watch?v=JkBRiGzttHM&t=3s

TELL-Seq software guides (Universal Sequencing)
Tell-Read, Tell-Sort, and IGV input files
www.universalsequencing.com/protocol-gate

IGV viewer (Broad Institute)
software.broadinstitute.org/software/igv/

Additional TELL-seq applications (Illumina)
www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/ilmn-ust-microbial-denovo-assembly-app-note-m-gl-00130/illmn-ust-microbial-denovo-assembly-app-note-m-gl-00130.pdf

Supreme NZTLong DNA polymerase (NZYTech)
https://www.nzytech.com/products-services/molecular-biology/end-point-pcr/dna-polymerases/mb331/

Data

Input and output files can be downloaded from PRJNA804227

Ordering information

https://www.universalsequencing.com

Reagent boxes

TELL-Seq WGS Library Reagent Box 1	#100001
TELL-Seq WGS Library Reagent Box 2	#100002

Primer boxes

TELL-Seq Library Index Primer Kit	#100003/9/10
TELL-Seq Illumina Sequencing Primer Kit	#100004

Safety data sheets:

https://www.universalsequencing.com/protocol-gate

Acknowledgements

Diogo Comprido, NZYTech, Lisbon (Portugal)

Adrian Perez Agustin, Mel.lina Pinsach Abuin, and Sara Pagans, Universitat de Girona, Catalonia (Spain)

References

Bailey et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell*, 371-385 (2018)

Chen et al. Ultra-low input single tube linked-read library method enables short-read second-generation sequencing systems to generate highly accurate and economical long-range sequencing information routinely. *Genome Res.* 30:898-909 (2020)

Gorelick et al., Phase and context shape the function of composite oncogenic mutations. *Nature*, 582, 100-103 (2020)

Klingstrom et al., A comprehensive model of DNA fragmentation for the preservation of High Molecular Weight DNA. *BioRxiv*, http://doi.org/10.1101/254276 (2018)

Pinsach-Abuin et al. Analysis of Brugada syndrome loci reveals that fine-mapping clustered GWAS hits enhances the annotation of disease-relevant variants. *Cell Reports Med.*, 2:100250 (2021)

The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature, 526: 68-74 (2015)*

Vasan et al. Double PIK3CA mutations in cis increase oncogenicity and sensitivity to PIKa inhibitors. *Science,* 366: 714-723 (2019)

Zook et al. Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Sci. Data*, 3, 160025 (2016)

For research use only. Not for use in diagnostic procedures