



Why interest in TSLR:

Long and accurate reads for denovo genome assembly
that we can sequence on our existing HiSeq instruments

Alvaro G. Hernandez, Ph.D.



Overview of TSLR library workflow:

Fractionate HMW genomic DNA to 8-10kb

Dispense ~ 300 adapted fragments per well on 384-well plate

Make a barcoded Nextera XT library in each well

Pool all mini-libraries and sequence on one PE lane

Expected results from one library:

~ 100k to 150k reads

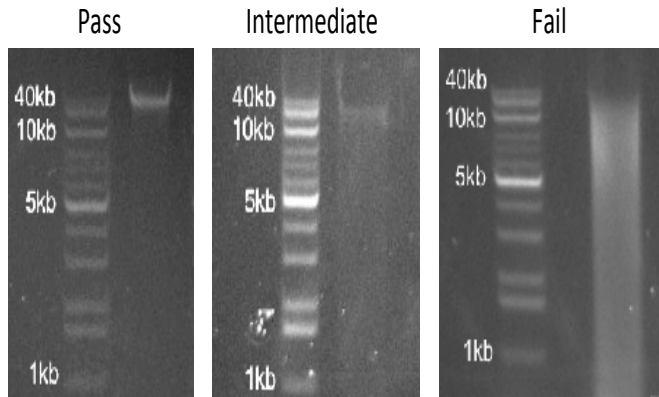
1.5kb to 10kb in length (N50 ~ 7kb)

> 500 MB

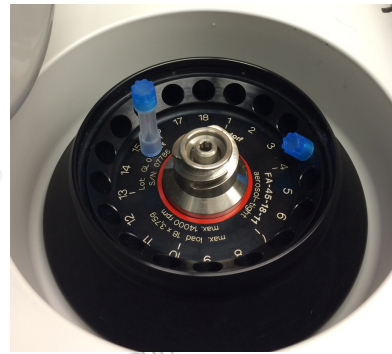


Day 1

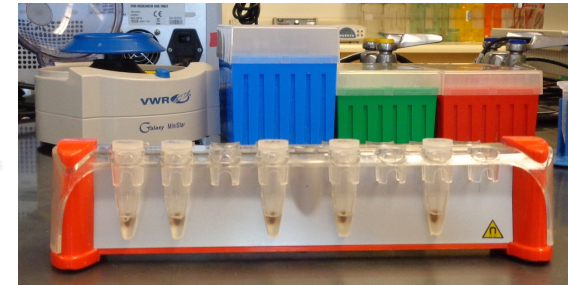
HMW gDNA



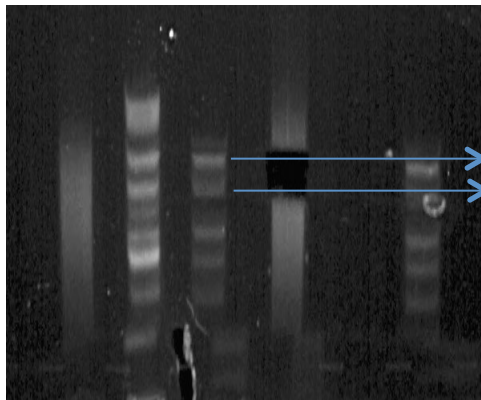
Fractionation of 0.5ug of DNA in gTube



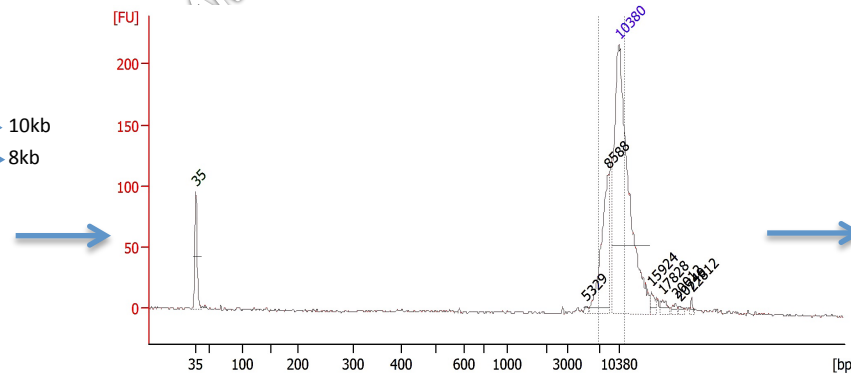
End Repair 3'-end A overhang End Adapter Ligation



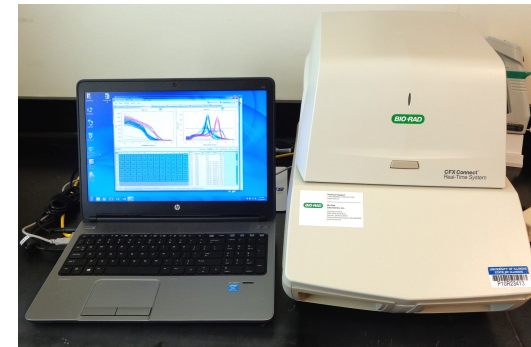
0.8% E-gel



QC of size selection



qPCR quantitation (8 hrs)





Day 2

qPCR results:

200 fg/ul

400 fg/ul

1,100 fg/ul

1,800 fg/ul



Dispense ~1,500 fg to a 384-well plate
with Long-Range PCR reagents (5ul reactions)



Dr. G. Hernandez, Ph.D.



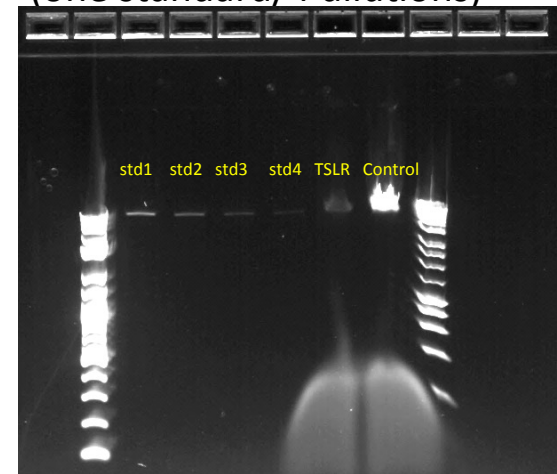
21 cycles for 384-well with 5ul reactions

Control 96-well plate: 50ul reactions, 26 cycles

***** 4.5hrs



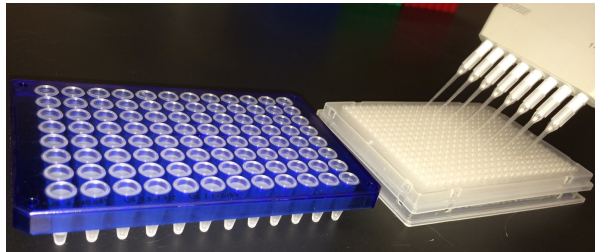
QC of long-Range PCR
(one standard/4 dilutions)





Day 3

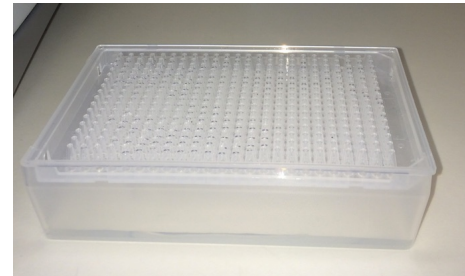
Tagmentation reagents (25min)



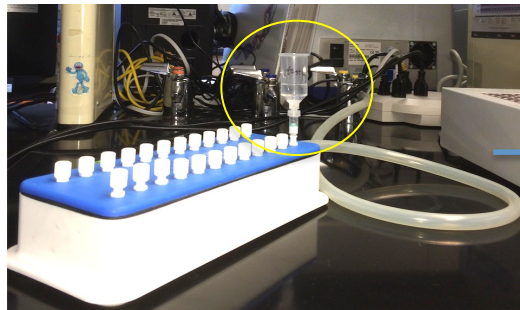
P5, P7 and Index primers
(10 cycles)



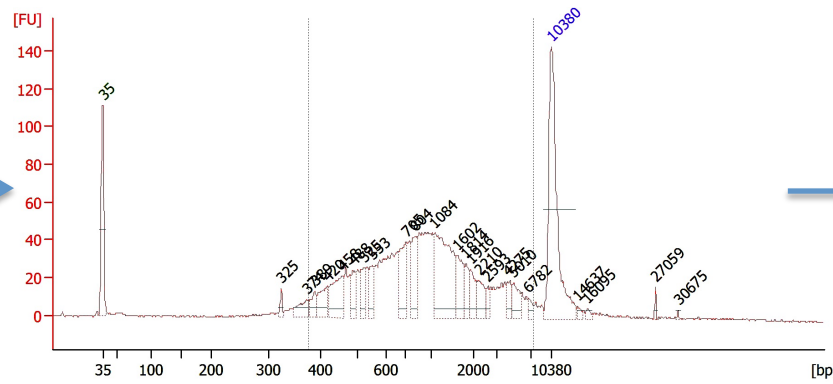
Pooling (24ml)



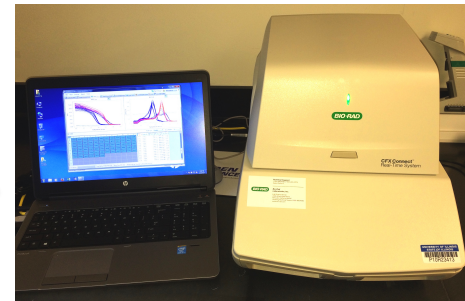
Clean up on
large Zymo columns



Bead size selection and QC



qPCR of final library





Day 4

151 x 8 x 151 Rapid Run 9pM
one lane per plate



Must stream data
to BaseSpace



Day 6

 SPAdes Genome Assembler 3.0 ALGORITHMIC BIOLOGY LAB Launch	 SWATHAtlas Ion Library Generator INSTITUTE FOR SYSTEMS BIOLOGY Launch	 The Broad's IGV BROAD INSTITUTE Launch
 TopHat Alignment ILLUMINA, INC. Launch	 TruSeq Amplicon v1.1 ILLUMINA, INC. Launch	 TruSeq Long-Read Assembly ILLUMINA, INC. Launch
 TruSeq Phasing Analysis ILLUMINA, INC. Launch	 Tumor Normal v1.0 ILLUMINA, INC. Launch	 TUTE GENOMICS TUTE GENOMICS Launch

Alvaro G. Hernandez, Ph.D.

TruSeq Long-Read Assembly
ILLUMINA, INC.

Analysis Name:

Save Results To:
TSLR_beta_10kb_ReadyLibrary

Sample:
Plate1

Reference Genome Alignment for Quality Assessment: Human (UCSC hg19)

This app is free.

[Continue](#)

2 days to final assembly

Expected results=
>500MB of reads 1.5kb to 10kb in length



App Output

PDF Summary Report

Long Reads FASTQ (>= 1500bp)

Long Reads FASTQ (500 - 1499bp)

Sample Information

Total PF Reads	Read 1 Length	Read 2 Length	Genome
364,178,661	101	101	-

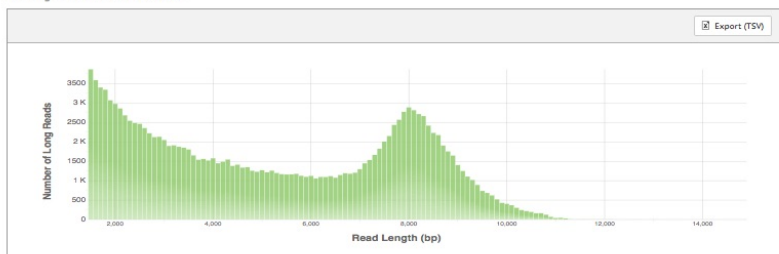
Well Metrics ⁱ

	n	Mean	Median	Span90
Reads PF	384	946,029	962,041	2.48
Percent End Marker	370	1.94 %	1.86 %	1.89
Number Unique Fragments per Well (End Markers)	384	235	242	1.37
Total Unique DNA Sequences Assembled (bp)	364	2,502,185	2,492,100	1.30
Assembly NSO Length (pre-QC)	364	7,566	7,619	1.08

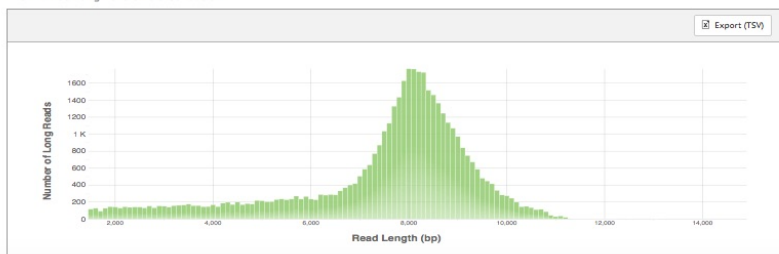
Assembly Metrics ⁱ

	500 - 1499bp	>= 1500bp
Number of Long Reads Assembled	65,638	151,117
Number of Bases in Assembled Long Reads	58,561,175	807,025,570
Long Reads Assembly NSO Length	970	7,447

All Long Read Size Distribution ⁱ



End-Marked Long Read Size Distribution ⁱ



Report Date: 05/16/2014

Library: gDNAcontrol

Analysis Software Version: v2.0.11

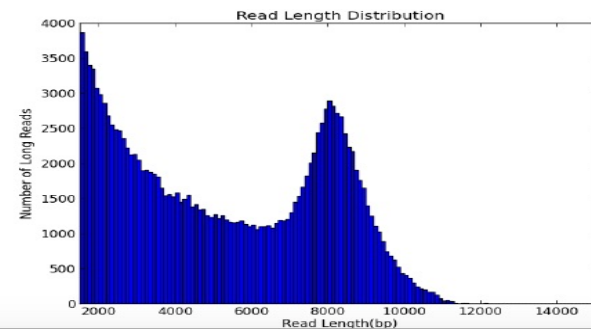
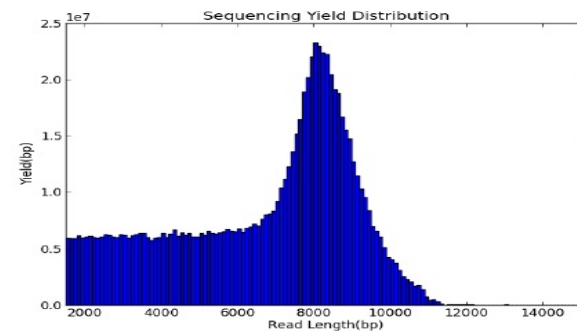
App Version: v1.0

Sequencing Metrics

Number of Long Reads Assembled >=1500bp: 151,117

Number of Bases in Assembled Long Reads >=1500bp: 807,025,570 bp

Number of Bases in Assembled Long Reads >=500bp: 865,586,745 bp





TSLR library from human gDNA (2 rapid lanes 2x100 cycles)

Sample Information

Total PF Reads	Read 1 Length	Read 2 Length	Genome
364,178,661	101	101	-

Well Metrics ⁱ

	n	Mean	Median	Span90
Reads PF	384	946,029	962,041	2.48
Percent End Marker	370	1.94 %	1.86 %	1.89
Number Unique Fragments per Well (End Markers)	384	235	242	1.37
Total Unique DNA Sequences Assembled (bp)	364	2,502,185	2,492,100	1.30
Assembly N50 Length (pre-QC)	364	7,566	7,619	1.08

Assembly Metrics ⁱ

	500 - 1499bp	≥ 1500bp
Number of Long Reads Assembled	65,638	151,117
Number of Bases in Assembled Long Reads	58,561,175	807,025,570
Long Reads Assembly N50 Length	970	7,447



TSLR library from human gDNA (2 rapid lanes 2x100 cycles)

Sample Information

Total PF Reads	Read 1 Length	Read 2 Length	Genome
364,178,661	101	101	-

Well Metrics ⁱ

	n	Mean	Median	Span90
Reads PF	384	946,029	962,041	2.48
Percent End Marker	370	1.94 %	1.86 %	1.89
Number Unique Fragments per Well (End Markers)	384	235	242	1.37
Total Unique DNA Sequences Assembled (bp)	364	2,502,185	2,492,100	1.30
Assembly N50 Length (pre-QC)	364	7,566	7,619	1.08

Assembly Metrics ⁱ

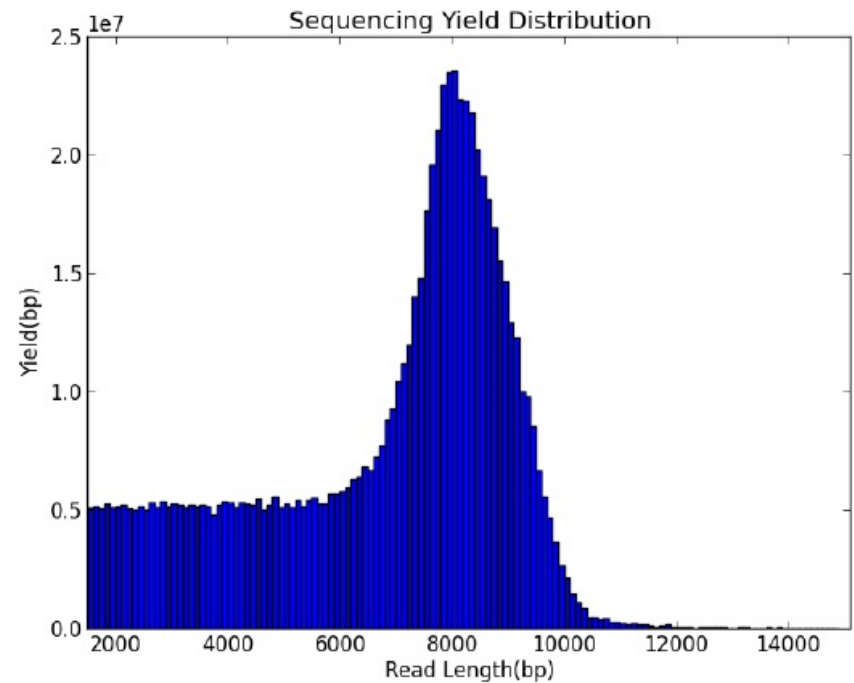
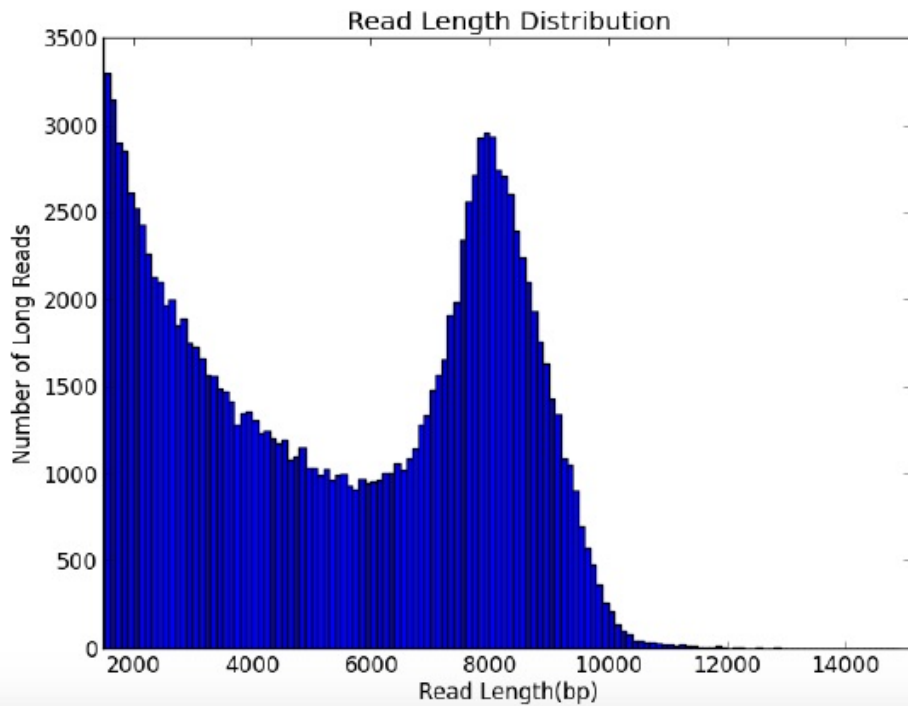
	500 - 1499bp	≥ 1500bp
Number of Long Reads Assembled	65,638	151,117
Number of Bases in Assembled Long Reads	58,561,175	807,025,570
Long Reads Assembly N50 Length	970	7,447



TSLR library from human gDNA (2 rapid lanes 2x100 cycles)

Sample Information

Total PF Reads	Read 1 Length	Read 2 Length	Genome
77,470,224	101	101	-



Number of Long Reads Assembled		
Number of Bases in Assembled Long Reads	58,561,175	807,025,570
Long Reads Assembly N50 Length	970	7,447



Other TSLR libraries

	Plant (500 MB)	Insect (2.3 GB)	Insect (300 MB)	Metagenome (600MB)	Plant (600 MB)
Sequencing	Two lanes 2x100nt	One lane 2 x 150	One lane 2 x 150	One lane 2 x 150	One lane 2 x 150
# of reads (millions)	706 (70 Gbases)	330 (50 Gbases)	358 (53 Gbases)	288 (43 Gbases)	291 (43 Gbases)
Fragments per well	327	210	206	164	163
MB of reads >1.5kb	824	642	722	541	508
N50 Read Length (kb)	6.6	6.8	6.8	7.8	7.4



Other TSLR libraries

	Plant (500 MB)	Insect (2.3 GB)	Insect (300 MB)	Metagenome (600MB)	Plant (600 MB)
Sequencing	Two lanes 2x100nt	One lane 2 x 150	One lane 2 x 150	One lane 2 x 150	One lane 2 x 150
# of reads (millions)	706 (70 Gbases)	330 (50 Gbases)	358 (53 Gbases)	288 (43 Gbases)	291 (43 Gbases)
Fragments per well	327	210	206	164	163
MB of reads >1.5kb	824	642	722	541	508
N50 Read Length (kb)	6.6	6.8	6.8	7.8	7.4



Processing of TSLR reads:

* Initial assembly of the reads is done in BaseSpace, no need for in-house processing before data delivery

* Two of the software we tested can be used with TSLR data:

PBJelly2 (BLASR) → adds TSLR reads to existing assembly to fill gaps and join scaffolds (genome finishing)
Use stringent alignment parameters for high confidence of gap filling
Fast (hours)

Celera Assembler 8.1 and minimus2 = successfully used for denovo assembly using shotgun, mate-pairs and TSLR

Recommendation for denovo assembly:

80x coverage of shotgun libraries 400bp and 700bp in length

50x coverage of mate-pair libraries 3kb, 8kb and 15kb

RNAseq library

3-5x coverage of TSLR

500 MB genome:

* 2 HiSeq lanes 2x150nt for shotgun, mate-pair and RNAseq libraries (>90 Gbases)

* 3 to 4 TSLR plates (1 library, 3-4 plates)