High throughput workflow for human whole genome sequencing using PacBio HiFi

Julian Rocha[1], Jackson Mingle[1], Aurelie Souppe[1], Kaitlyn Scott[1], Renee Fedak[1], Duncan Kilburn[1], Enrique Bayo Iglesias[2], Dominik Laubscher[2], Birgit Ottenwälder[2], Suzanne Dee[1], Heather Ferrao[1], Jeffrey Burke[1], Kelvin J Liu[1]
[1]PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025, [2]Hamilton Bonaduz AG, Via Crusch 8, 7402 Bonaduz, Switzerland

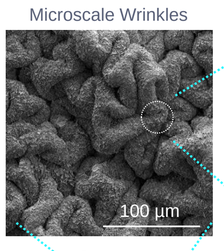
High-throughput HiFi sequencing workflow

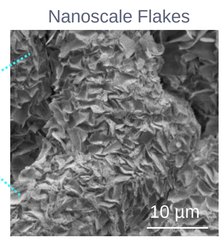
Extraction → Shearing → Size selection → Library prep → Sequencing → Analysis

Sample preparation workflows for long-read sequencing often contain bottlenecks in extraction, size selection, shearing, and library preparation that limit sample throughput, add cost, and create variability in data yields and quality. We present a high-throughput, fully automated 96 sample workflow that can be used to sequence and analyze a variety of human sample types and thus support growing numbers of large-scale population genomics and clinical research studies in a manner that rivals standard NGS workflows in simplicity.

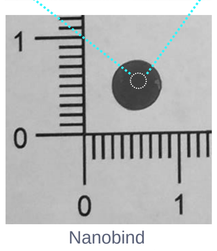
Nanobind HT (high-throughput) kits for HMW DNA extraction

The kits contain Nanobind magnetic disks which can extract HMW DNA using a simple bind, wash, and elute process. Scripts will be released for a wide variety of sample types including cells, blood, bacteria, and tissues.



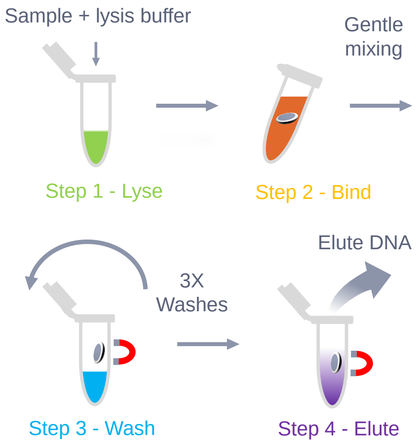


Nanobind magnetic disks
- 3 – 5 mm diameter
- Covered by high density of micro- and nanostructured silica
- High binding capacity
- High purity



Nanobind magnetic disk

Rapid magnetic purification
- Rapid bind, wash, and elute protocol
- HMW protocols generate 50 kb – 300+ kb sized DNA
- Manual processing using magnetic rack
- Automation compatible for high-throughput applications



Flexible instrument compatibility

SEMI-AUTOMATED
Manual plate filling + requires limited user interaction

FULLY AUTOMATED
Automated plate filling + fully walk-away

ThermoFisher KingFisher™ DUO
12 samples

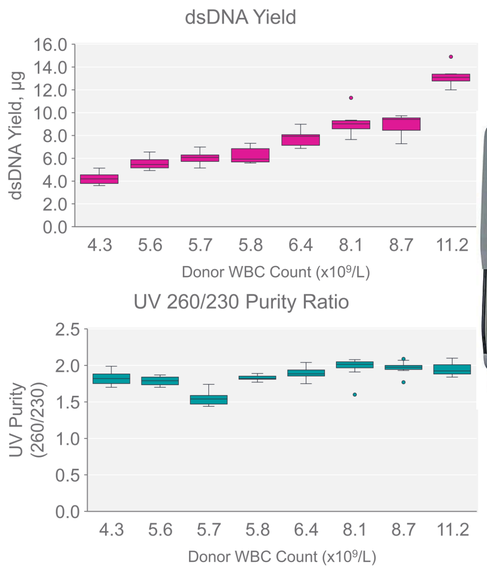
ThermoFisher KingFisher™ FLEX / APEX
96 samples

Hamilton NIMBUS Presto
96 samples

		DUO 12 samples	FLEX / APEX 96 samples	NIMBUS Presto 96 samples
200 µL blood/cultured cells/bacteria	Hands-on time	15 min	45 min	20 min
	Total time	90 min	120 min	200 min
1 mL blood	Hands-on time	10 min	40 min	20 min
	Total time	115 min	145 min	155 min

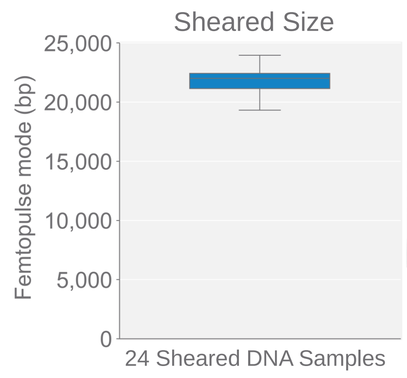
- Scripts are available for 4 different instruments to accommodate an array of sample throughput needs
- Semi-automated solutions require manual plate filling and limited user interaction
- Fully automated solutions will fill plates and are fully walk-away
- Total time is the hands-on time + automation runtime

Fully automated extraction of 96 blood samples on Hamilton NIMBUS Presto



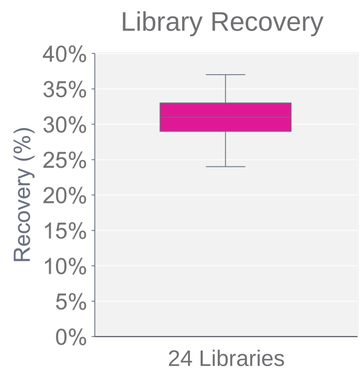
96 200 µL blood samples (8 donors x 12 replicates) were extracted using a single plate on Hamilton NIMBUS Presto
- Donors were selected to represent normal physiological range of white blood cell (WBC) counts
- Consistent DNA recovery and purity across all donors
- High reproducibility across replicates

High-throughput DNA shearing, size selection, and library prep on Hamilton NGS STAR



	Samples per run	Time per run	Time per plate	Sheared DNA size	Cost per sample	Cost per plate
HT shear	96	8 min	8 min	15 – 100 kb	$0.08	$7.68

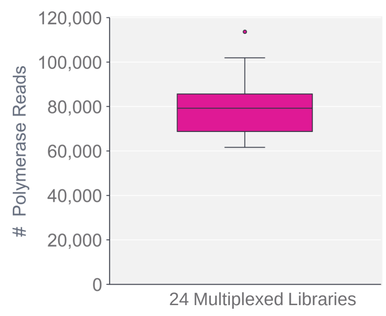
High-throughput shearing was performed directly on the Hamilton NGS STAR by repeated pipetting of the HMW blood DNA
- Consistent shearing performance observed across 96 samples
- This method can shear a plate of 96 samples in <10 min at <$0.10 per sample (cost of a pipette tip)
- It further allows transitioning directly from shearing to library prep on the instrument deck

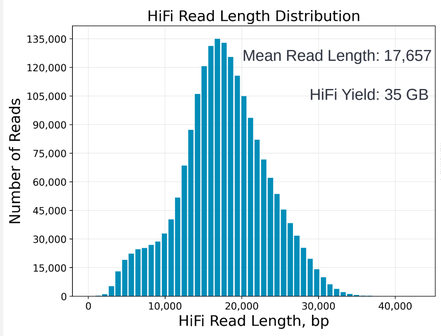


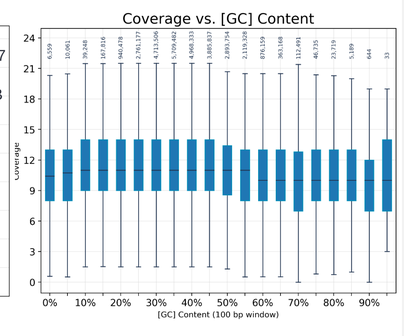
	Hands-on time	Automation runtime	Total time	Library Prep + ABC
24-library prep run	55 min	4 hours	5 hours	7 hours
24-ABC run	30 min	1.5 hour	2 hours	
96-library prep run	1.17 hours	4.75 hours	6 hours	8.5 hours
96-ABC run	40 min	1.75 hours	2.5 hours	

Library prep, size selection, and ABC (anneal, bind, cleanup) were then performed on 24 sheared blood DNA samples (3 µg ea)
- Consistent library recovery across 24 samples run on Hamilton NGS STAR with SMRTbell prep kit 3.0/Sequel II binding kit 3.2 to prepare for sequencing
- The method can fully prep, size select, and ABC a plate of 24-96 samples in 1 day

HiFi sequencing of high-throughput samples

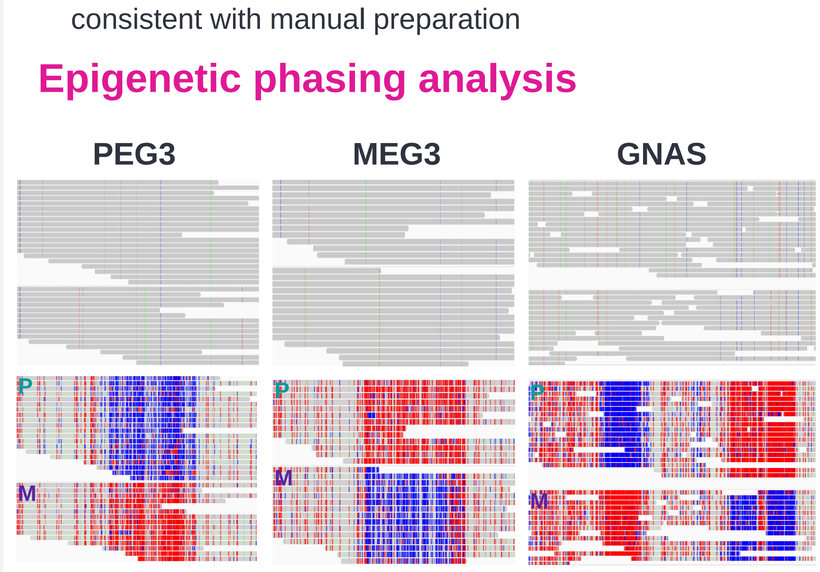






The 24 previously prepared samples were sequenced in multiplex on a Sequel IIe system
- Consistent sequencing performance was seen across all 24 samples
- A single SMRT Cell (30 hr movie) produced 11-fold coverage evenly across regions of varying GC content
- Overall HiFi yield and mean HiFi read length are consistent with manual preparation

Epigenetic phasing analysis



PEG3 = paternally expressed gene 3
MEG3 = maternally expressed gene 3
GNAS = guanine nucleotide binding protein, alpha stimulating activity polypeptide

Hypermethylation
Hypomethylation

The Sequel IIe system performs, as standard, 5 base sequencing (A, C, G, T + 5mC) with on-instrument basecalling
- One blood sample was sequenced using 3 SMRT Cells to obtain ~33-fold coverage
- Haplotype phasing and assembly were then performed
- Methylation analysis allows clear identification of paternal (P) and maternal (M) alleles with methylation pattens of genes that are characteristic of genomic imprinting

	Polished Contigs	Maximum Contig Length	Mean Contig Length	N50 Contig Length	Sum of Contig Lengths
Primary Contigs	2,503	104,819,458	1,214,921	39,295,530	3,040,947,438
Haplotigs	12,568	8,465,803	221,369	1,198,218	2,782,167,443

Contig length metrics for genomic assembly of a human genome.

Conclusions

We've demonstrated a high-throughput automated workflow for processing Human blood samples from extraction through HiFi sequencing. Nanobind HT extraction kits will be available in Q1 of 2023. Contact us for early access.