PacBio's 1000 Genomes RNA Sequencing Project – Automated on firefly®

Rebecca Martin[1], Katherine Munson[2], Kaitlyn Sun[2], Elizabeth Tseng[3], Melissa Netwal[4]

[1]Seattle Children's Research Institute, US [2]University of Washington School of Medicine, Department of Genome Sciences, US [3]Pacific Biosciences, US [4]SPT Labtech

PacBio's 1000 Genomes Project (1KGP) aims to use PacBio's Iso-Seq and Kinnex kits to generate a high-quality, comprehensive open-source whole transcriptome dataset to characterize transcript diversity across populations. This effort is being led by Danny Miller, MD, PhD, FACMG - Seattle Children's Research Institute / University of Washington, Rajiv McCoy, PhD – Johns Hopkins, Winston Timp, PhD – Johns Hopkins and Evan Eichler, PhD – University of Washington. More than a thousand RNA samples were isolated at Johns Hopkins. Samples were split and batched and are currently being processed across two firefly's at Seattle Children's Research Institute – Spatial and Genomics CoLab and University of Washington in the Eichler Lab.

Iso-Seq & Kinnex Workflow

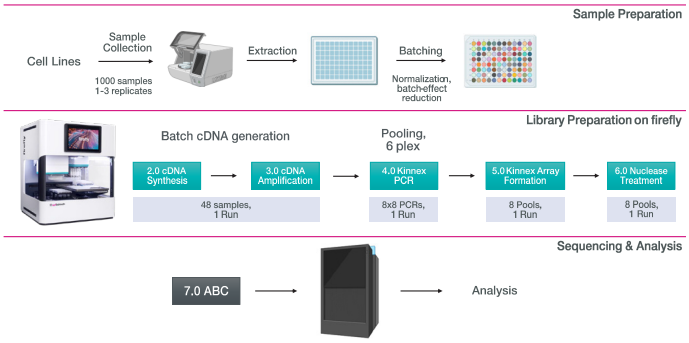

Figure 1. Overview of 1KGP workflow using SPT Labtech's firefly to automate library preparation steps. Processing began with sample collection, RNA extraction, and batching for normalization and batch-effect mitigation (Johns Hopkins). Iso-seq nd Kinnex libraries are being generated across two sites on firefly in 48 RNA sample batches (2.0 cDNA Synthesis, 3.0 cDNA Amplification), cDNA is pooled 6-plex for 8 Kinnex pools carried through library preparation on firefly (4.0 Kinnex PCR, 5.0 Kinnex Array Formation, 6.0 Nuclease Treatment). Final libraries are being sequenced on two PacBio Revio sequencing platforms at UW & SCRI. Analysis is being performed at Johns Hopkins (McCoy Lab, Miller Lab, Timp Lab).

cDNA Synthesis & Amplification

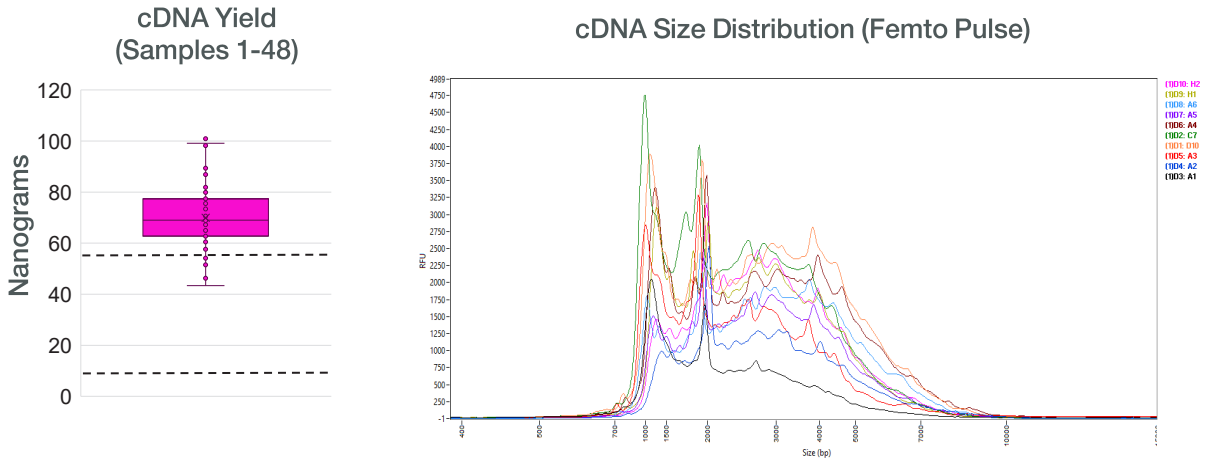

Figure 2. cDNA Synthesis & Amplification QC from a 48-sample batch automated on firefly Left: cDNA yield across 48 samples following synthesis and amplification on Firefly. Average yield ~70 ng, exceeding the typical input requirement for Kinnex.Right: Representative cDNA fragment size distributions measured using the Agilent Femto Pulse system, showing consistent amplification profiles across samples with dominant peaks near ~200 bp and broad fragment representation from 200–1000 bp, suitable for downstream Kinnex library prep.

sptlabtech

Sequencing Quality Control

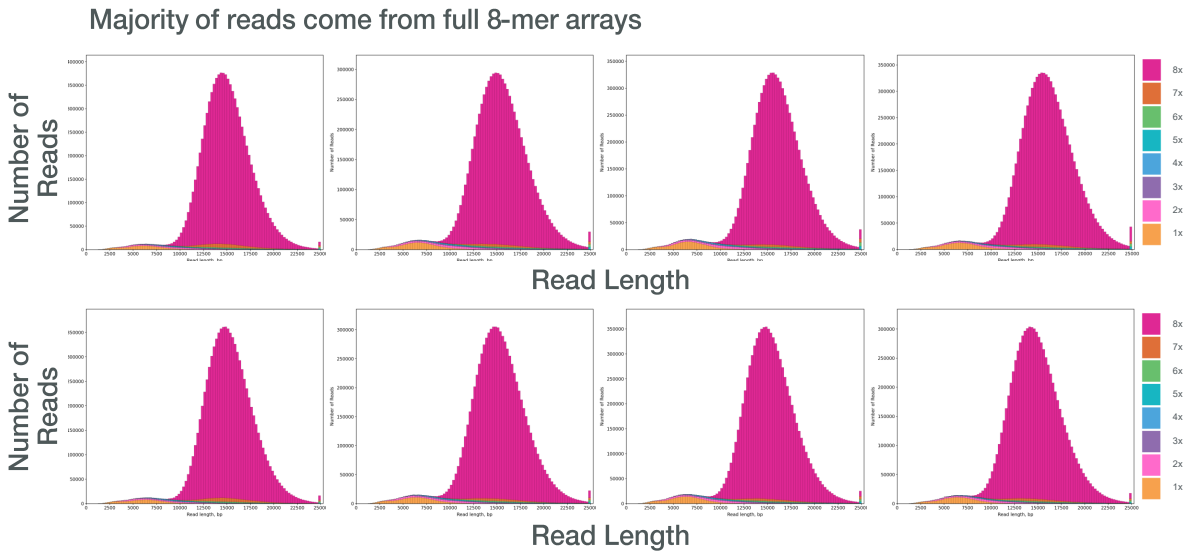

Figure 3. Kinnex Sequencing QC. Read length distribution across 8 samples shows a sharp peak at ~15,000–17,000 bp, indicating that the majority of sequencing reads originate from complete 8-mer arrays. Each panel represents a separate sequencing run with consistent results across chips. Stacked color traces distinguish different 8-mer repeat levels, with the magenta peak corresponding to full-length arrays (8x), comprising the dominant proportion of the reads.

Increased Reads Marginally Increase Gene Saturation

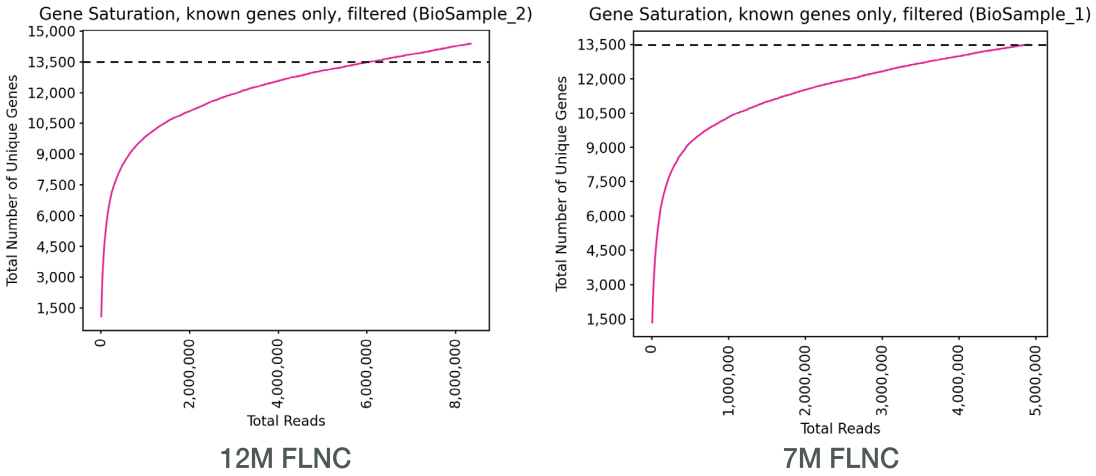

Figure 4. Gene saturation curves demonstrate the diminishing returns of increased sequencing depth on gene detection for two representative BioSamples (FLNC). While early increases in read count yield substantial gene discovery, the curves plateau around 12,000–13,500 unique genes, indicating saturation. These results suggest that ~7–12 million FLNC reads are sufficient to capture the transcriptome complexity.

Scan & Watch Our Webinar: Join Dr. Rebecca Martin, Scientific Lead at the Genomics and Spatial Biology CoLab, Seattle Children's Research Institute, and Dr. Liz Tseng, Product Manager at PacBio, as they discuss how automated workflows are revolutionizing reproducible, full-length transcript profiling at scale.

Scan & Explore Our firefly:
Discover our liquid handling plattform and see how it can transform your workflows!

sptlabtech.com