NEBNext Direct® HS

High Sensitivity Target Enrichment

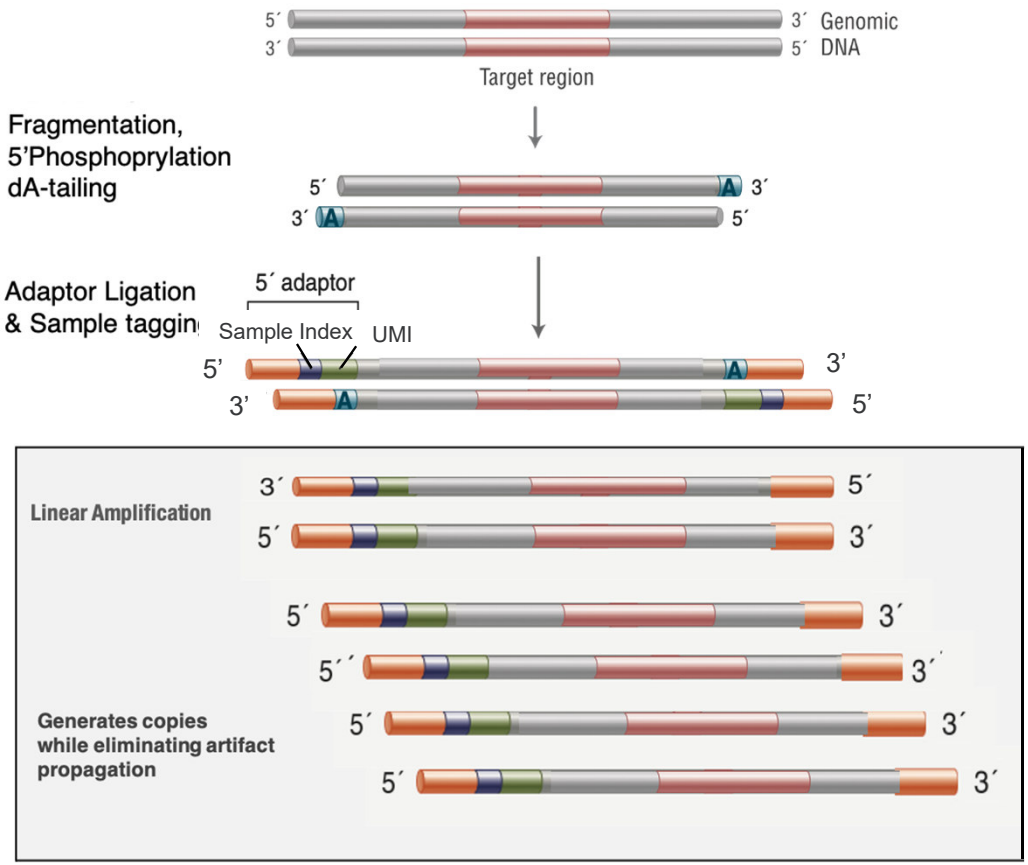
May 18, 2021

NEBNext Direct

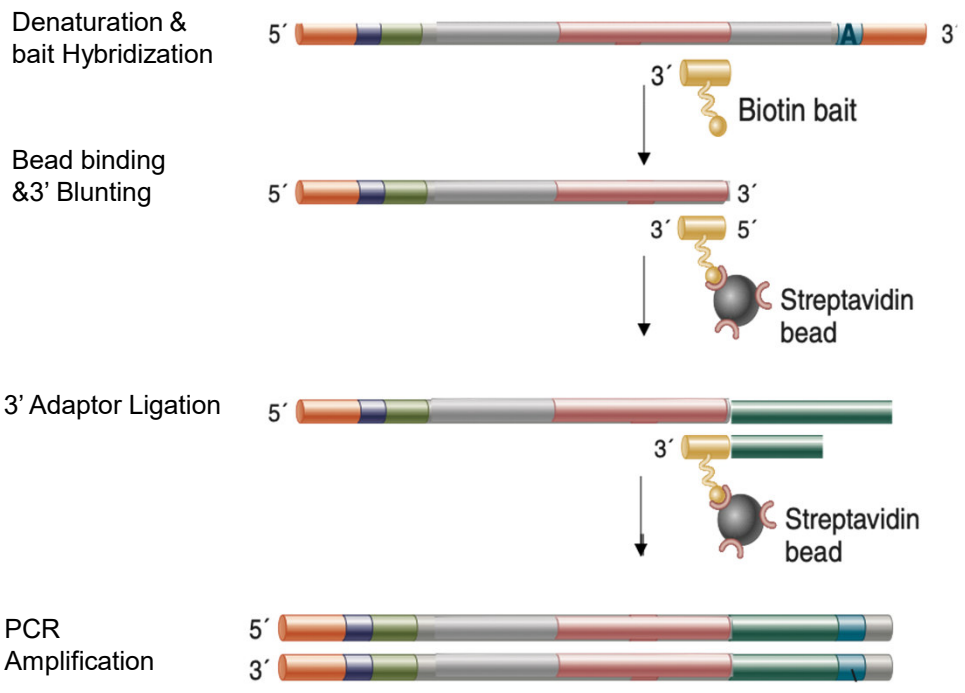
- Initial NEBNext Direct (v1.0) chemistry launched in 2016
- Needed higher sensitivity assay for clinical sample types (FFPE, cfDNA, ≤ 50ng)
- Goals:
    - Maintain High Specificity (% Selected)
    - Leverage Unique Chemistry for best-in-class coverage uniformity
    - Integrated, streamlined workflow (1-2 days)
- Currently in final stages of development, seeking strategic partnerships

NEBNext Direct HS Chemistry

Sample barcoding & Linear Amp



Target Capture

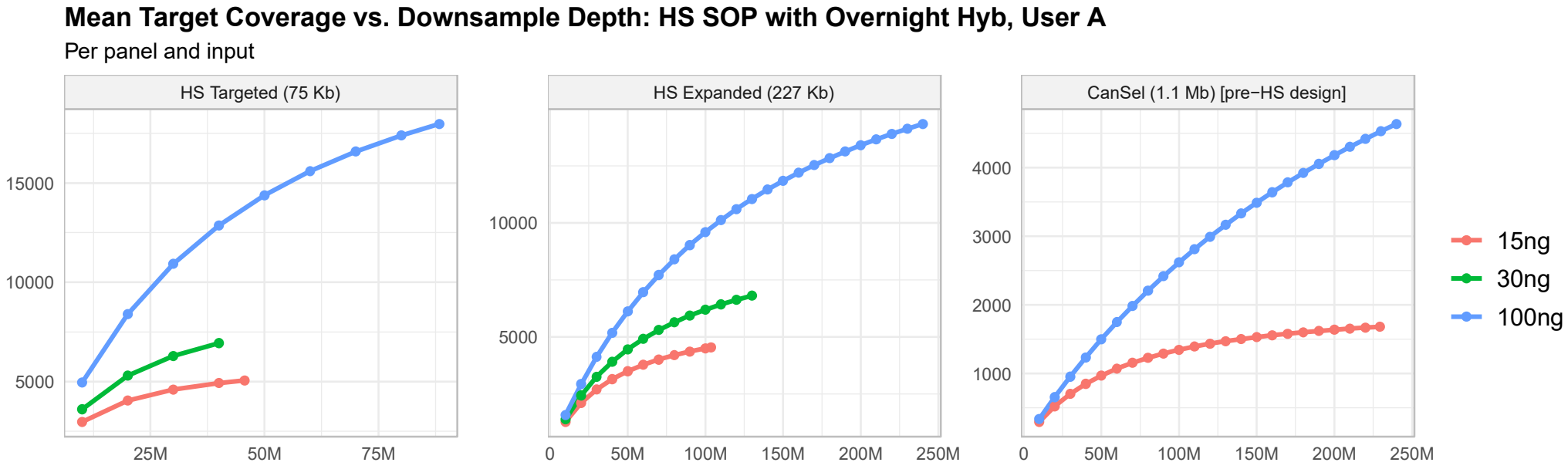


NEW ENGLAND BioLabs Inc.

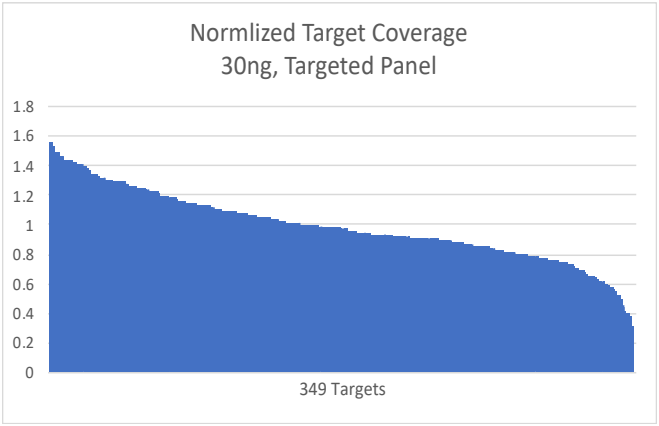
Testing across Panels, Input DNA

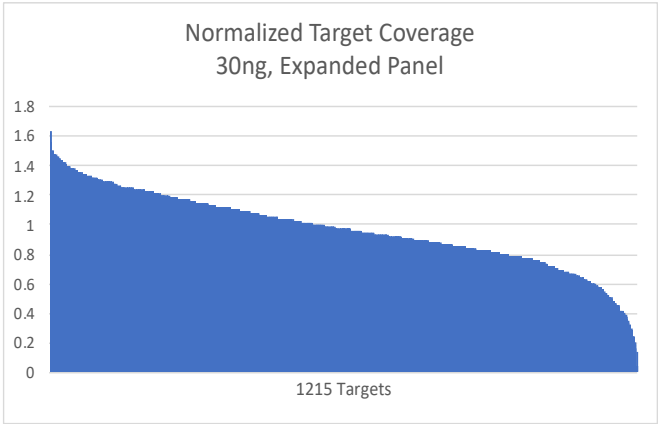
- 15, 30, 100ng input DNA
- 75kb Targeted Panel, 227kb Expanded Panel, 1.1 Mb Cancer Select Panel
- Raw sequencing data (NextSeq) 2 samples run in triplicate
- Demultiplexed and aligned with internal target enrichment pipeline

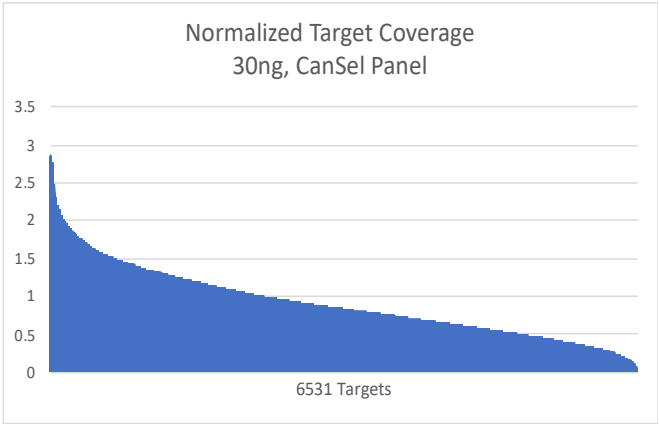
Unique Coverage Obtained



Unique Coverage as a function of read depth



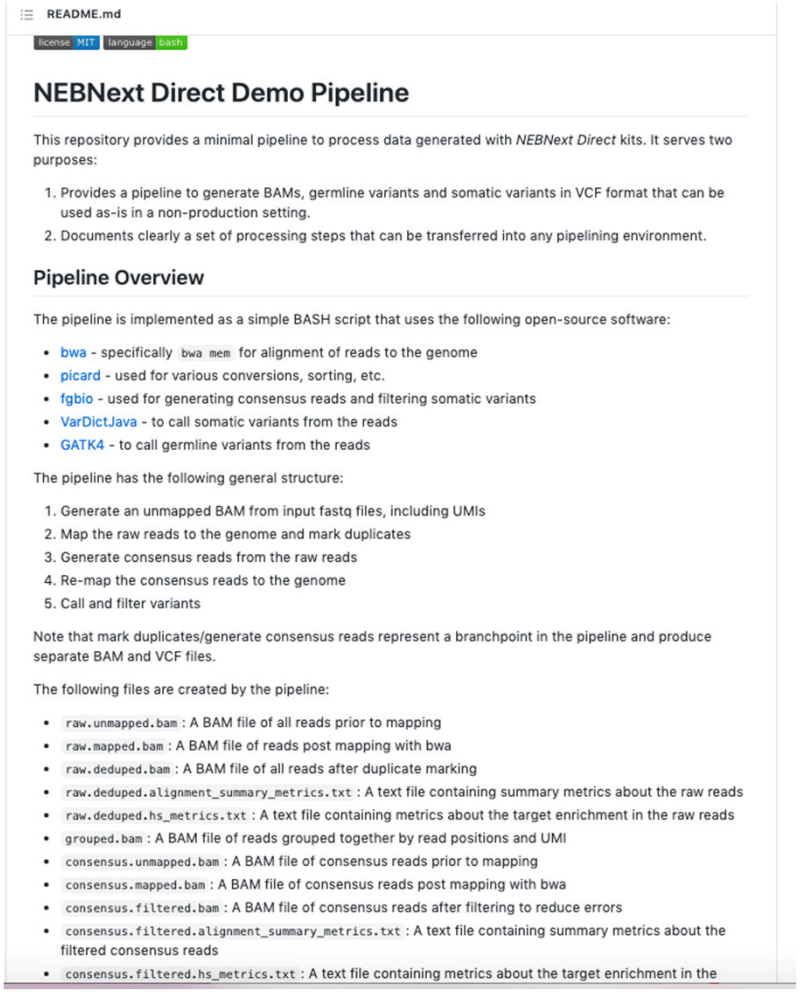




Normalized Coverage Across Targets

NEW ENGLAND BioLabs Inc.

Bioinformatic analysis



README.md

license MIT language bash

NEBNext Direct Demo Pipeline

This repository provides a minimal pipeline to process data generated with *NEBNext Direct* kits. It serves two purposes:

1. Provides a pipeline to generate BAMs, germline variants and somatic variants in VCF format that can be used as-is in a non-production setting.
2. Documents clearly a set of processing steps that can be transferred into any pipelining environment.

Pipeline Overview

The pipeline is implemented as a simple BASH script that uses the following open-source software:

- bwa - specifically `bwa mem` for alignment of reads to the genome
- picard - used for various conversions, sorting, etc.
- fgbio - used for generating consensus reads and filtering somatic variants
- VarDictJava - to call somatic variants from the reads
- GATK4 - to call germline variants from the reads

The pipeline has the following general structure:

1. Generate an unmapped BAM from input fastq files, including UMIs
2. Map the raw reads to the genome and mark duplicates
3. Generate consensus reads from the raw reads
4. Re-map the consensus reads to the genome
5. Call and filter variants

Note that mark duplicates/generate consensus reads represent a branchpoint in the pipeline and produce separate BAM and VCF files.

The following files are created by the pipeline:

- `raw.unmapped.bam` : A BAM file of all reads prior to mapping
- `raw.mapped.bam` : A BAM file of reads post mapping with bwa
- `raw.deduped.bam` : A BAM file of all reads after duplicate marking
- `raw.deduped.alignment_summary_metrics.txt` : A text file containing summary metrics about the raw reads
- `raw.deduped.hs_metrics.txt` : A text file containing metrics about the target enrichment in the raw reads
- `grouped.bam` : A BAM file of reads grouped together by read positions and UMI
- `consensus.unmapped.bam` : A BAM file of consensus reads prior to mapping
- `consensus.mapped.bam` : A BAM file of consensus reads post mapping with bwa
- `consensus.filtered.bam` : A BAM file of consensus reads after filtering to reduce errors
- `consensus.filtered.alignment_summary_metrics.txt` : A text file containing summary metrics about the filtered consensus reads
- `consensus.filtered.hs_metrics.txt` : A text file containing metrics about the target enrichment in the

- Developed open-source workflow for data analysis
- Leverages existing tools (picard, VarDict, FgBio)
- Includes tools for consensus variant calling
- Published on GitHub repository

Seeking partnership for Lot 1:

- Nucleic Acid Extraction:
  - FFPE & ctDNA
  - QC analytics

- Bioinformatics:
  - Implement variant calling pipeline

Please contact Andrew Barry (barrya@neb.com)