Constellation mapped read technology

SCAN ME

Fiona Kaper[1], Jacqui Weir[2], Mitch Bekritsky[3], Ali Crawford[3] Qing Zhang[2], Vicki Thomson[2], Ines Vitoriano[2], Niall Gormley[2], Arun Subramaniyan[1], Michael Ruehle[3], Drew Ellershaw[2], Shunhua Han[1], Daniel Andrews[2], Stephen Gaffney[2], Ritu Kundu[2], Gavin Parnaby[1], Pascal Grobecker[2], Maria Gardzielewska[2], Ivana Armogida[2], Maya Bajracharya[2], Ningxin Ouyang[1], Cande Rogert[1], Rami Mehio[1], Louise Fraser[1]

[1] Illumina Inc., San Diego, California, USA; [2] Illumina Inc., Cambridge, Cambridgeshire, UK; [3] Illumina Inc., Remote, USA

INTRODUCTION

Constellation mapped read technology is a novel approach that leverages on-flow cell library preparation and uses proximity information from neighboring nanowells to generate long-range genomic insights using standard SBS sequencing

HOW IT WORKS

Library prep is completely eliminated by the use of flow cell-bound transposomes which capture and tagment long molecules of DNA as they are flowed onto the flow cell surface. Clustering and SBS are as standard resulting in high-quality short-read sequencing quality. No modifications to the instrument hardware are required.

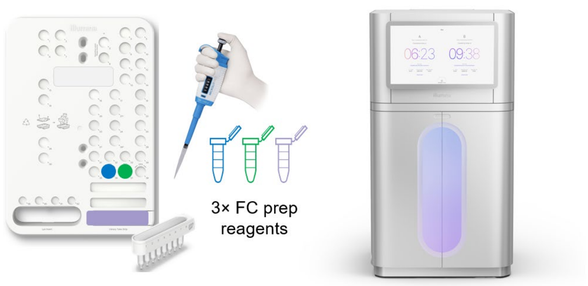

Figure 1. The Constellation Workflow has 3 Simple Steps:
1. Add 350ng extracted DNA to library strip tube
2. Add 2 reagents to cartridge
3. Load sequencer and start run

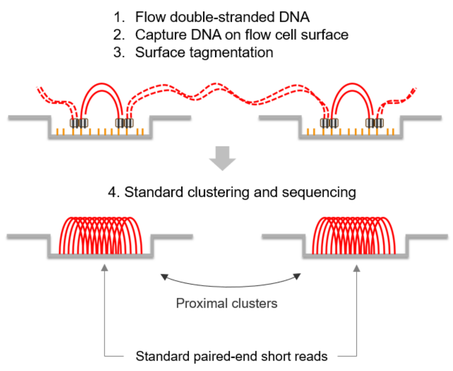

Figure 2. Double stranded DNA is flowed onto the flow cell. DNA is captured on the surface and tagmented. Clusters that originate from the same DNA template molecule are nearby on the flow cell surface.

Since DNA molecules are tagmented *in situ* on the flow cell surface, the resulting reads from neighboring clusters can be reconstructed into an interspersed version of the original DNA template molecule. This property is unique to Constellation and is not observed in any other NGS assay.

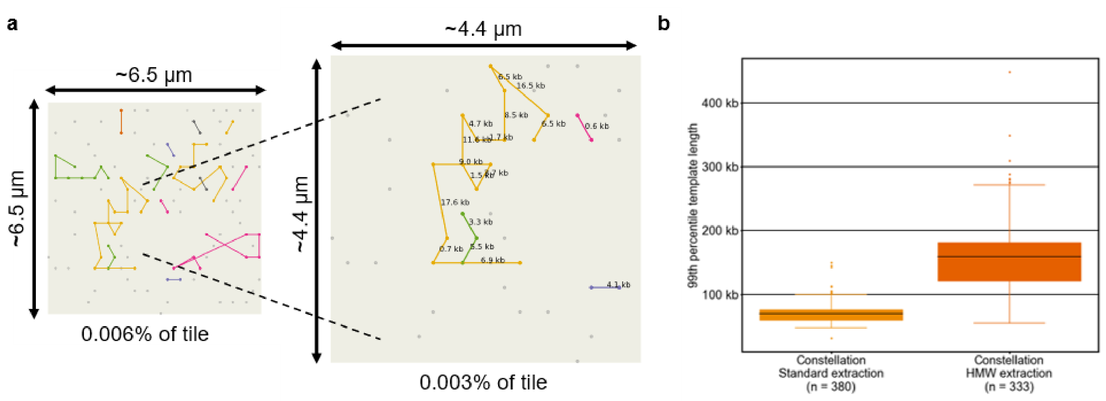

Figure 3. a. Clusters generated from the flow cell are organized into distinct templates, visually represented in different colors. Within each template, the cluster are sequentially arranged based on their genomic coordinates. The connections between paired reads are annotated with their corresponding genomic distances. **b.** Box plot shows the range of template sizes from std and HMW extractions.

IMPROVED PERFORMANCE IN DIFFICULT TO MAP REGIONS

In limited regions of the genome, uniquely mapping standard short reads is challenging due to high homology or other repetitive context which makes it difficult to distinguish among multiple candidate mapping positions. Constellation read mapping uses proximity information from neighboring clusters that *do* uniquely map to assign reads to the correct genomic location

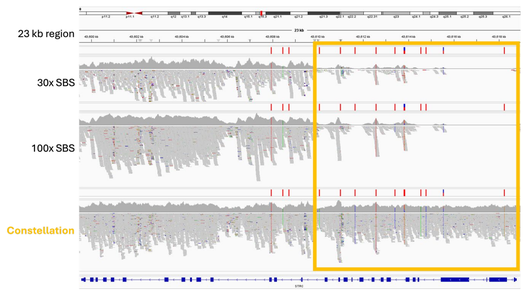

Figure 4. Recovery of coverage in the *STRC* gene.

STRC has a pseudogene, *STRCP1*, with >99% sequence identity, making mapping challenging with standard whole genome library prep. Some mutations in *STRC* are associated with pediatric nonsyndromic hearing loss.

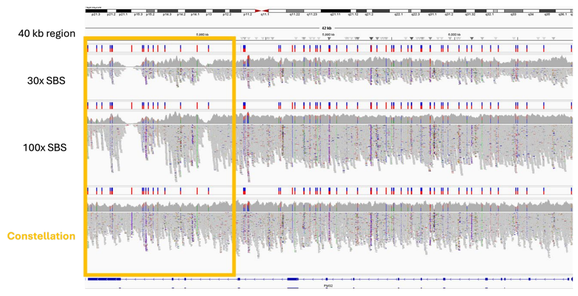

Figure 5. Recovery of coverage in the *PMS2* gene.

PMS2 has a pseudogene, *PMS2CL*, with homology >99% in some parts. Some mutations in *PMS2* are associated with Lynch syndrome, ovarian cancer, and other disorders.

The improved mapping resolution enabled by Constellation extends to improved small variant calling performance, particularly in difficult-to-map regions of the genome prone to low coverage.

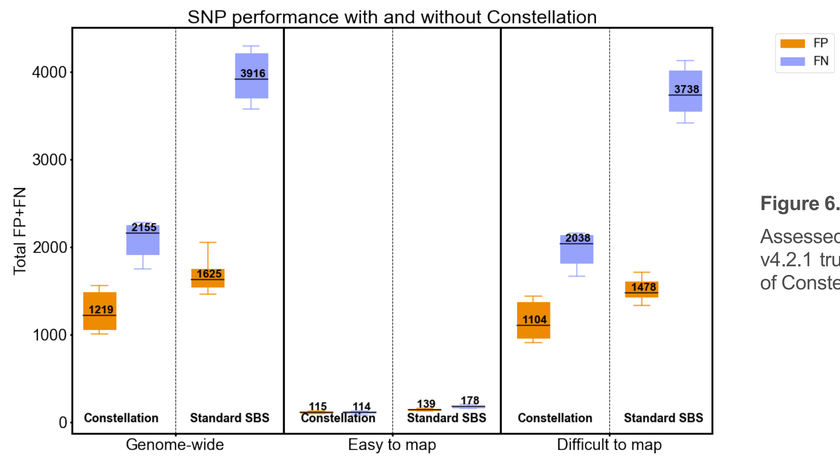

Figure 6. SNP variant calling performance

Assessed against NIST Genome in a Bottle v4.2.1 truthset using rtgeval for 11 HG002 runs of Constellation read mapping.

ULTRA LONG PHASING

Phasing with Constellation is especially powerful since its capabilities are defined solely by the native DNA template length captured on the flow cell, not read length, and currently extends from 100's of kilobases up to several megabases. High molecular weight (HMW) extraction methods that preserve larger templates are shown to contribute to larger phase blocks.

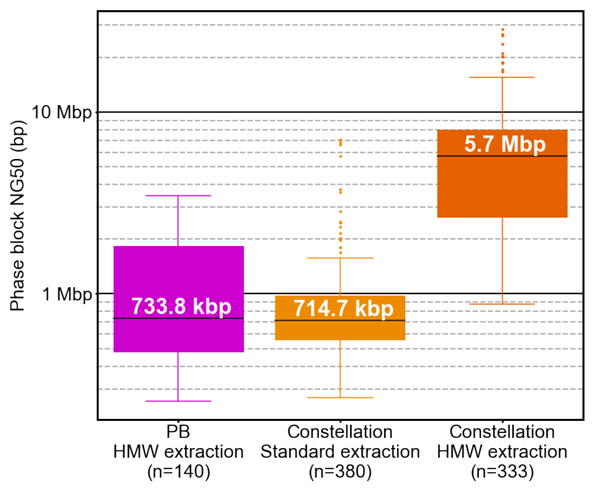

Figure 7. Phase block NG50 for constellation mapped reads with standard or HMW DNA extractions.

Phase block NG50 is measured over chromosomes 20–22 with WhatsHap stats. PacBio HiFi data (PB) data was obtained from the human pangenome reference consortium (HPRC) and was processed with pbmm2 v1.13, DeepVariant v1.6.0 and WhatsHap v2.2 on GRCh38.

IMPROVED STRUCTURAL VARIANT CALLING

Constellation technology has the added benefit of improved structural variant (> 50 bp) calling. Utilizing DRAGEN™ v4.3 secondary analysis, Constellation shows a dramatic improvement in SV recall from 51.5% with standard SBS to 87.8%. Constellation also enables improved detection of complex rearrangements, see poster 212.

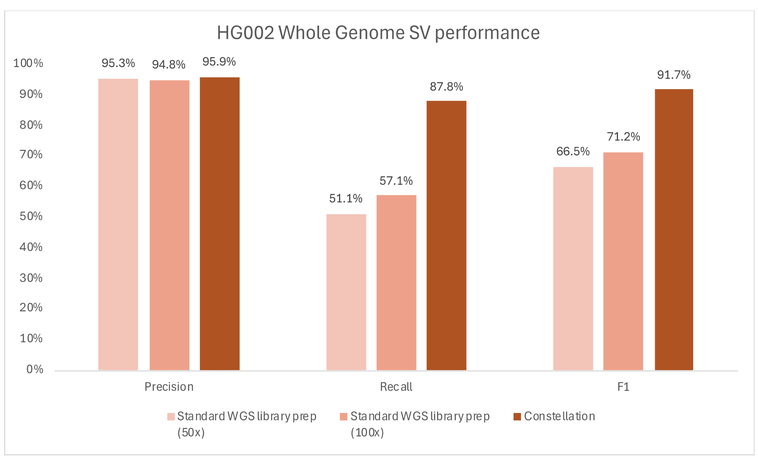

Figure 8. Constellation SV performance.

The analysis uses the Genome in a Bottle T2T-Q100 HG002 SV v1.1 truthset with the SV confident BED file

THIS IS JUST THE BEGINNING

Constellation Mapped Read technology is a new, powerful foundational technology with broad capabilities. Here we demonstrate some of the benefits for human genome sequencing, however multiple future applications are under evaluation. The first commercially available product based on Constellation technology is expected in 2026 and will leverage existing NovaSeq™ X platforms to create an accessible, cost-effective solution for comprehensive human WGS.

SP-00641

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

illumina®