Nabsys HD-Mapping™

A novel system for human whole-genome structural variation analysis

Overview

- High-definition human whole-genome mapping for structural variation analysis
- The capture of balanced and unbalanced SVs at high resolution in a single run
- Lower cost per sample than long-read sequencing, OGM, or traditional cytogenetic techniques

INTRODUCTION

Structural variants (SV) are queried far less frequently than single nucleotide variants (SNV) in both clinical and research settings. This is despite the fact that SVs are widely recognized as drivers of constitutional disease and cancer as well as being responsible for 5 times more human genomic variability than SNVs alone[1]. The reason for this disparity is apparent in the genomics tools landscape in which there is a tradeoff between cost and the length-scales on which genomic variation can be queried.

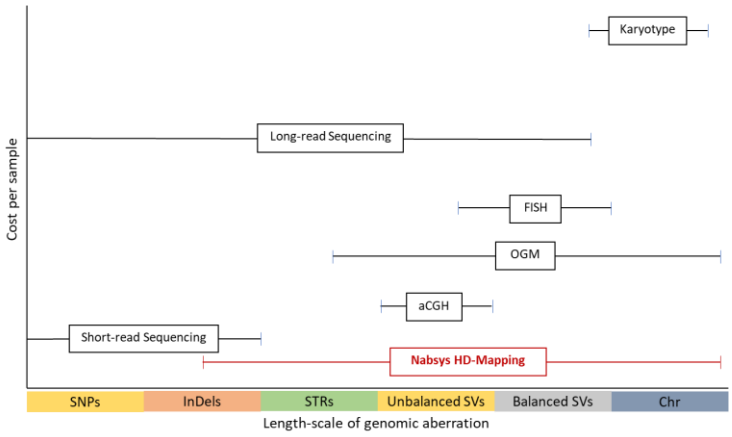

Figure 2: HD-Mapping closes the gap between short-read ensemble sequencing and OGM while expanding the length-scale of long read sequencing at a lower cost per sample.

As is well documented, short-read ensemble sequencing platforms offer low cost but read lengths are limited to ~500bp making the technologies adept at detecting SNVs and small indels, but not well suited for SVs. Hi-C and long-read single molecule sequencing enable the detection of SVs but at a higher cost. Optical genome mapping (OGM) improves price performance and length-scale but cannot consistently resolve SVs below ~1000bp. These tradeoffs result in continued reliance on legacy cytogenetic technologies, such as karyotyping, fluorescence in-situ hybridization (FISH), and chromosomal microarrays (aCGH) which are used reflexively or in parallel to capture the complexity of balanced and unbalanced SVs.

Figure 1: HD-Mapping's detector design and unique chemistry deliver cost-effective throughput allowing human whole-genome analysis using a single detector on an instrument the size of a desktop computer.

HD-Mapping™ is a novel system for high-definition human whole-genome structural variation analysis (Figure 1). By employing electronic detection as opposed to optical detection, HD-Mapping can capture balanced and unbalanced SVs at high resolution in a single run. The higher resolution captures repeat expansions and contractions at small enough intervals to be complimentary to short-read sequencing with reads long enough to capture large chromosomal rearrangements at a lower cost per sample than long-read sequencing, OGM, or traditional cytogenetic techniques (Figure 2).

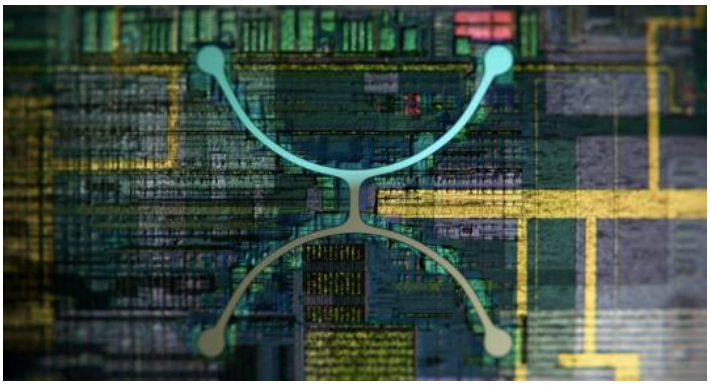

Figure 3: Two hundred and fifty-six nanochannels are tightly configured to capture a whole human genome in a single run.

© 2022 Nabsys All rights reserved. All trademarks are the property of their respective owners.

HD-MAPPING OVERVIEW

HD-Mapping detects labelled DNA through changes in electrical resistance as single molecules pass through solid-state nanochannels (Figure 3). Two hundred and fifty-six nanochannels are tightly configured into a single reusable detector to process millions of molecules across a whole human genome in a single run.

High molecular weight (HMW) DNA is isolated, and the molecules are labelled at known recognition sites. Once a detector is loaded on the instrument and the sample is injected, single DNA molecules are electrophoretically translocated through the nanochannel where the analytes are electronically detected by changes in resistance caused by the analyte. The results are single molecule maps which are *in silico* representations of each DNA molecule with the location of each label. The single molecule maps are assembled into map contigs to produce a genome-wide map of all labeled locations for the sample which are aligned against a reference genome map for SV analysis (Figure 4).

While HD-Mapping does not capture single nucleotides, the electronic detection method captures a broader length-scale of variation than OGM at a lower cost per sample than nanopore or nano-feature based sequencing or OGM.

Expanding the length-scale

OGM, as the name implies, relies on the optical detection of fluorescent labels. Light diffraction makes it difficult to detect labels in close proximity to each other, limiting resolution. Missing labels and the associated reduction in resolution increase error in sizing an SV and create uncertainty in the breakpoint locus. HD-Mapping does not use optics or florescence, eliminating diffraction and allowing for the detection of labels in closer proximity to each other. The increased label density in the HD-Mapping approach leads to greater resolution compared to OGM.

This difference in resolution becomes evident in the ability of HD-Mapping to close the gap in length-scale by consistently detecting SVs in the 300-1000bp range. On the other end of the length-scale, HD-Mapping's long read lengths allow for the detection of large SVs including chromosomal rearrangements.

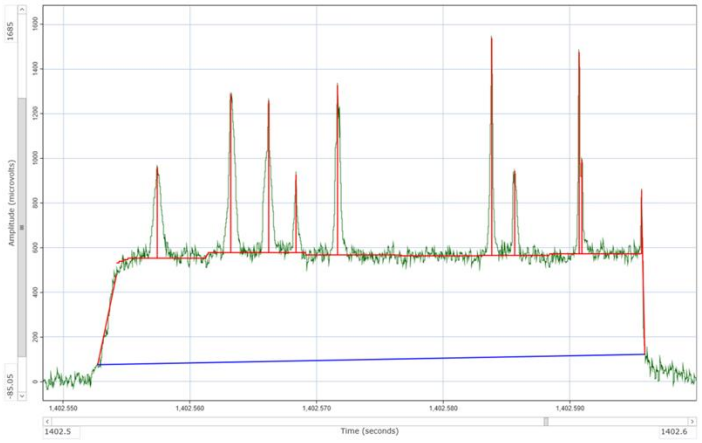

Figure 6: HD-Mapping achieves a high signal-to-noise ratio (in the above case SNR=30) without slowing or stretching the molecule which is traveling at 1 megabase / sec.

Cost-effective throughput

HD-Mapping's nanochannel approach reduces design complexity, consumable requirements, and associated costs. Nanopores, solid state or biological, are typically sized such that the majority of the pore is occupied by the DNA molecule in order to create a detectable signal. The large resistance change that occurs when DNA passes through a nanopore results in crosstalk in adjacent pores unless they are electrically isolated by a physical barrier. The requirement of a physical barrier between nanopores introduces a linear scaling effect in nanopore-based systems. As throughput scales, the platform footprint, consumable requirements, and associated costs scale linearly.

HD-Mapping's detectors (Figure 5) on the other hand use a solid-state nanochannel that only requires a <1% change in resistance for a label to be detected. The low change in electrical resistance as well as the inherent isolating qualities of the nanochannels eliminate crosstalk concerns. This, combined with HD-Mapping's unique RecA protein coating on the molecule, produces a high signal-to-noise ratio (SNR) (Figure 6). Because of the high SNR and low crosstalk, the nanochannels can be tightly configured. Further, the high SNR is achieved without having to slow or stretch the DNA molecule as it passes through the detection channel.

Figure 5: Detector

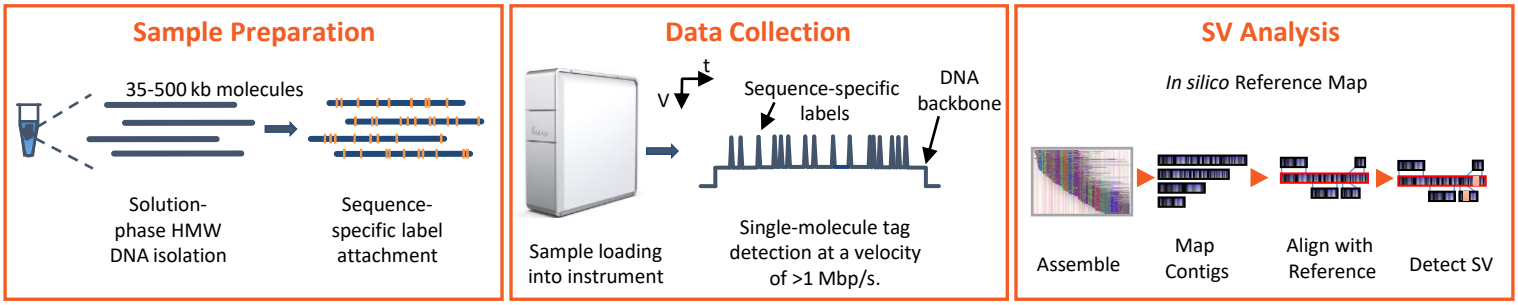

Figure 4: Human SV analysis workflow – The HD-Mapping platform enables whole-genome SV analysis using a straightforward workflow with minimal hands-on time.

© 2022 Nabsys All rights reserved. All trademarks are the property of their respective owners.

STRUCTURAL VARIATION ANALYSIS USING HD-MAPPING

Using HD-Mapping's onboard field-programmable gate array (FPGA), single molecule reads are processed in real-time into signal-processed files which are transferred to the cloud for SV analysis. The single molecule reads derived from the sample are assembled into map contigs. Before SV analysis can continue, a comparable reference is created. To accomplish this, a sequence-based reference genome is labelled *in silico* using the same sequence specific recognition sites as the sample to generate a reference genome map. Once a reference is created, the map contigs are aligned to the reference genome map. This alignment along with the assembly statistics are loaded into the SV analysis pipeline. SVs are then identified by comparing the observed label pattern of the sample map contigs with the expected pattern of the reference genome map. Changes to the loci of these labels represent structural changes in the genome (Figure 7). For example, two label sites measurably farther apart on a sample map contig compared to the corresponding locations on the reference map could indicate an insertion (Figure 8).

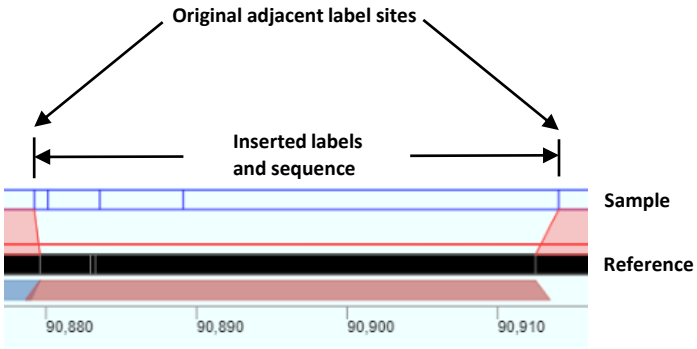

Figure 7: Screenshot of analysis software showing an insertion.

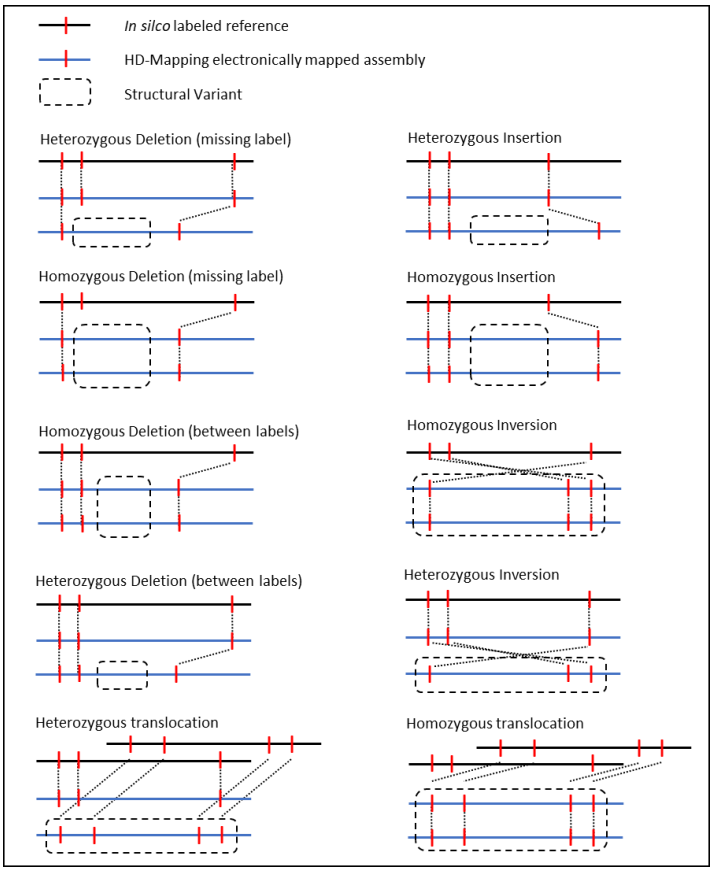

Figure 8: Comparing label patterns between sample map contigs and a reference genome map reveals structural variation in the sample.

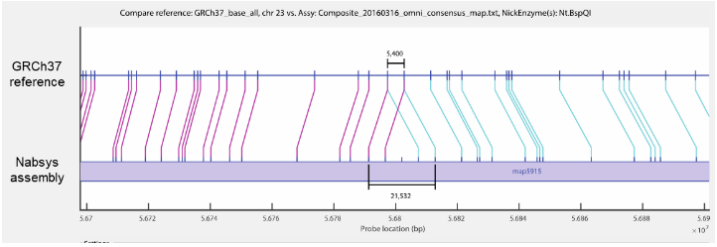

Figure 9: HD-Mapping assembly of NA24385 compared to GRCh37 reference showing a collapsed repeat region on the reference.

NABSYS AND GENOME IN A BOTTLE

Nabsys has demonstrated the utility of HD-Mapping for SV analysis in part through its work with the Genome in a Bottle (GIAB) Consortium. A study comparing NA24385 relative to GRCh37 examined a collapsed repeat region on the X chromosome (Figure 9). HD-Mapping was able to detect a 16,132bp insertion where no call had been made previously. The team was also able to pick up a 34kb balanced inversion on chromosome 10, a 1,531bp insertion in chromosome 11, and a 301bp deletion in chromosome 18 demonstrating the range of structural variation detectable by HD-Mapping. This range is further demonstrated when comparing the concordance in deletion sizing to a GIAB high confidence call-set (Figure 10).

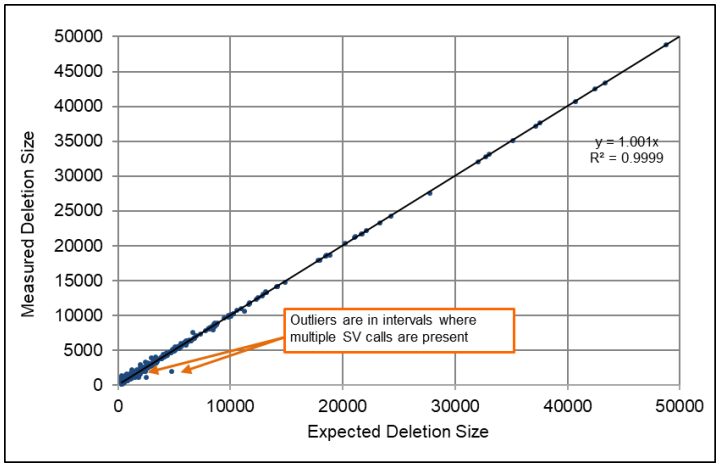

Figure 10: Concordance of HD-Mapping deletion size to GIAB high confidence call set. (Note: expected deletion size is an aggregate when multiple SVs are found between two label sites.)

© 2022 Nabsys All rights reserved. All trademarks are the property of their respective owners.

NABSYS AND HITACHI HIGH-TECH

In furthering its analytics, Nabsys has partnered with Hitachi High-Tech on a collaboration to develop Human Chromosome Explorer™ (HCE)(Figure 11). HCE exploits multiple threads, clusters, and compute nodes in Google Cloud's High Performance Computing cloud infrastructure to distribute the generation of long-range uniform map contigs into a haplotype aware assembly of a human whole-genome map for SV analysis through the web. For a detailed discussion of HCE & SV analysis, see *"Human Chromosome Explorer for HD-Mapping: Structural variation analysis using electronic whole-genome mapping data."*

CONCLUSION

The genomic instrumentation landscape is characterized by a proportional increase in cost as the length-scale of detecting genomic variation increases. Nabsys HD-Mapping's use of electronic detection and solid-state nanochannels combine to provide high signal-to-noise ratios, and nearly eliminates crosstalk capturing a broader length-scale of variation at a lower cost per sample than any other whole-genome analysis platform from an instrument the size of a desktop computer. When combined with HCE, the system captures and visualizes repeat expansions and contractions at small enough intervals to be complimentary to short-read sequencing with reads long enough to capture large structural rearrangements.

Figure 11: Screenshot of the Result Summary View in HCE

LEARN MORE

To learn more about Nabsys HD-Mapping visit https://nabsys.com

[1] *A New Human Genome Sequence Paves the Way for Individualized Genomics*, PLoS Biol. 2007 Oct; Liza Gross

© 2022 Nabsys All rights reserved. All trademarks are the property of their respective owners.