The Versatility of Next-Generation Mapping: An Encyclopedia of Genomes

S. Chan, A. Hastie, M. Saghbini, T. Anantharaman, H. Cao

BioNano Genomics, Inc., San Diego, CA, USA

Abstract

Genomics is an important field which has found many applications in medical diagnostics, generating and finishing reference genome assemblies, understanding trait inheritance, and more. Despite the obvious benefits genomics offers, its routine implementation remains impractical due to high costs and/or long turnaround time to generate results. Traditional and next-generation sequencing (NGS) techniques have been the standard in genomics, but sometimes sequencing alone is overkill and/or doesn't provide the desired information. Next-generation mapping (NGM) with the Irys® System from BioNano Genomics® provides a quick, cost-effective method to fill these gaps.

NGM complements and in some cases can even replace the current methods. It can be used in conjunction with sequencing by scaffolding contigs, defining the structure between contigs, and correcting errors in sequence assemblies – especially important with plant samples, which not only have long tandem repeats but are also often polyploid, making NGS assembly difficult. For some applications, NGM can potentially replace sequencing; for example, quick screening of patient samples for diseases such as cancer or congenital diseases can be done using NGM to detect disease-causing structural variation events for a fraction of the cost and time of NGS.

In addition to having countless applications, the Irys System is compatible with many sample types. Here, we demonstrate its versatility, as it was used to map hundreds of genomes from a variety of sample sources, including blood, plant tissue, and whole animals. We present a catalog of genomes assembled using the Irys System to date, showing how NGM was used to scaffold and finish reference genomes from birds, mammals, invertebrates, and plants.

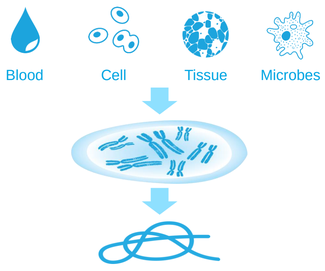
Background

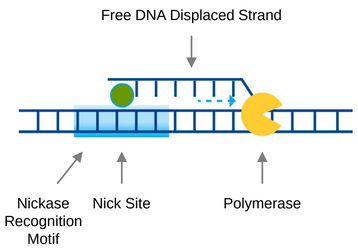
The Irys System utilizes a next-generation mapping technique in which labeled DNA is linearized and imaged in massively parallel NanoChannel arrays. Each imaged DNA molecule is a map of a portion of the sample genome, and these individual reads are *de novo* assembled into highly accurate consensus genome maps that cover the entire genome.

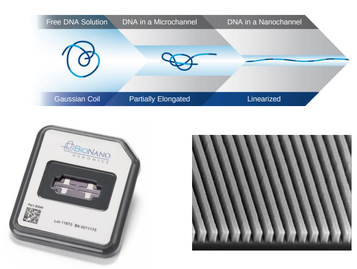
A major advantage of the Irys System is that it provides direct visualization of intact DNA molecules up to 2 Mbp in their native state, reliably preserving long range genomic information that is usually lost in other, less direct approaches. Since there are no amplification steps, errors and artifacts are extremely rare. For these reasons, the Irys System is an ideal candidate for projects involving structural variation analysis, and diploid (or polyploid) assembly with proper phasing. Genome maps can also be used to scaffold and validate NGS contigs, resulting in highly contiguous and correct genome assemblies.

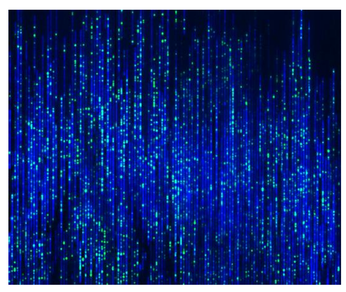
Methods

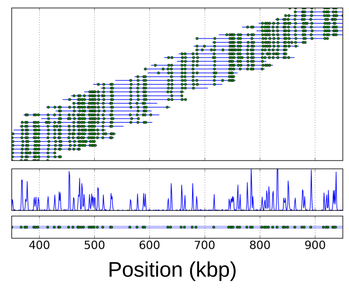
1. Extraction of long DNA molecules using IrysPrep kit
2. Fluorescent labeling of DNA at specific sequence motifs, with IrysPrep reagents
3. Linearization of DNA in NanoChannel arrays on IrysChip
4. Automated imaging of single molecules in NanoChannel arrays
5. Automated detection of molecules and labels in images, by instrument software
6. Assembly of data into optical consensus maps by IrysView software



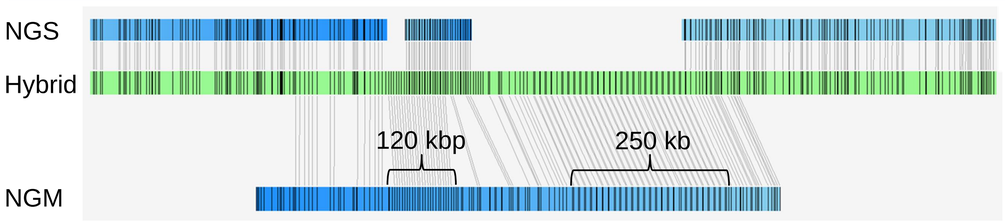








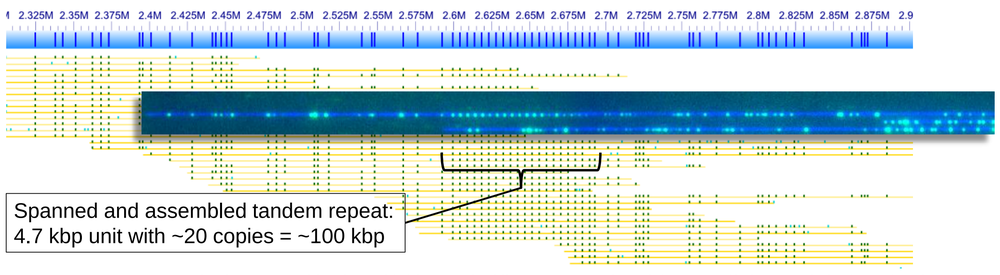
Hybrid Scaffolding: Spanning Long Tandem Repeats in the Hummingbird Genome



Blue bars represent PacBio (NGS) contigs which were digested *in silico* at nickase motif sites, and BioNano (NGM) genome maps assembled from Irys data. Motif sites are represented by vertical black lines. Alignments of sites is shown as gray lines.

NGS and NGM assemblies can be merged to form a hybrid scaffold. In this example two long tandem repeat regions in a hummingbird genome were spanned by a single Irys genome map. This map was used to anchor PacBio contigs on both sides. The final hybrid assembly had chromosome arm-length scaffolds.

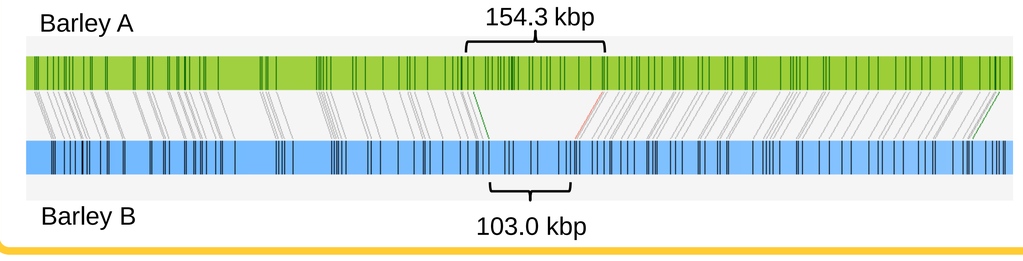
Single Molecule Validation of Assembled Repeat in Asparagus Genome



Consensus genome map is shown as a blue bar with blue vertical lines representing nickase motif sites, and single molecule reads are shown as yellow lines with green dots representing labeled motif sites.

Unlike many sequencing methods, the IrysPrep does not have any amplification steps, and intact DNA molecules up to 2 Mbp long are captured entirely in a single image, so artifacts are rare. Here, a single molecule read was used to validate an assembled repeat in an asparagus genome. The same method can be used to validate (or invalidate) a hybrid scaffold or NGS assembly.

Structural Variation Detection Between Strains of Barley



Irys NGM was used to identify structural variation between two strains of barley. The genome maps (blue and green bars) were *de* novo assembled independently and then aligned to each other. NGS contigs may be aligned to these maps to identify which gene(s) are affected.

Examples of Genomes Assembled from Various Sample Sources

Over 200 different organisms have been mapped using the Irys System to date. The IrysPrep is compatible with a wide range of input material. The table below highlights some interesting representative examples. The genome maps were used for a variety of purposes, including SV analysis, reference genome finishing, and disease diagnosis.

common rabbit
Oryctolagus cuniculus

Anna's hummingbird
Calypte anna

Malayan pit viper
Calloselasma rhodostoma

yellow fever mosquito
Aedes aegypti

common wheat
Triticum aestivum

soybean (soya bean)
Glycine max

Organism	Sample Source
Rabbit	Frozen spleen
White-tailed deer	Frozen muscle
Woodchuck	Frozen blood
Hummingbird	Frozen muscle
Blackcap	Frozen blood in EtOH
Human	Fresh/frozen blood, liver, kidney, spleen, brain, cell line (primary or transformed), buccal wash, brain
Goat	Frozen liver
Pit viper	Frozen kidney
Lamprey	Frozen/fresh blood, frozen sperm
Yellowtail kingfish	Fresh blood
Salmon	Frozen brain, frozen liver
Mosquito	Whole animals (pupae)
Spidermite	Whole animals (adult)
Drosophila	Whole animals (larvae)
Honeybee	Partial animals (heads)
Sea squirt	Partial animals
Trypanosoma cruzi	Frozen cell culture
Plasmodium falciparum	Frozen cell culture
Rice	Fresh leaves
Maize	Fresh/frozen leaves
Wheat	Fresh leaves, flow sorted chromosomes
Tomato	Fresh/frozen leaves, flowers
Barley	Fresh leaves, flow sorted chromosomes
Soybean	Fresh/frozen leaves
Chickpea	Fresh leaves
Papaya	Fresh leaves
Tobacco	Fresh/frozen leaves

white-tailed deer
Odocoileus virginianus

blackcap
Sylvia atricapilla

kingfish (yellowtail amberjack)
Seriola lalandi

sea squirt
Ciona intestinalis

maize (corn)
Zea mays

asparagus
Asparagus officinalis

Conclusions

BioNano next-generation mapping (NGM) is a powerful tool with a variety of applications, including – but not limited to: genome finishing, structural variation analysis, medical diagnostics and disease discovery, structural variation detection and analysis, and mass cataloguing of multiple samples. NGM and NGS are highly complementary, orthogonal methods, which can be combined to provide highly contiguous assemblies. In addition to having many applications, NGM has high utility potential, as it is compatible with a wide range of starting material. It is a quick and cost-effective method that will soon become the gold standard in genomics.

Related posters: P0033, P0712, P0957, P0961

References

- Lam, E.T., et al. Genome mapping on NanoChannel arrays for structural variation analysis and sequence assembly. Nature Biotechnology (2012); 10: 2303
- Hastie, A.R., et al. Rapid Genome Mapping in Nanochannel Arrays for Highly Complete and Accurate De Novo Sequence Assembly of the Complex Aegilops tauschii Genome. PLoS ONE (2013); 8(2): e55864

Acknowledgments

- Nils Stein & Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)
- Erich Jarvis & Duke University
- James Leebens-Mack & University of Georgia