CRISPR/Cas9 customization of electronic genome mapping for targeted analysis

Lindsay Opsahl,[1] Jay Sage,[1] Mike Kaiser,[1] John F Thompson[1]
[1]Nabsys 2.0 LLC, Providence, RI

Nabsys

Abstract

Genome mapping of ultra-long, single-molecule DNA is an important addition to sequencing for genome assembly and structural variant (SV) detection. Electronic genome mapping (EGM) enables high-resolution, whole-genome SV analysis. DNA is nicked at sequence-specific sites, followed by adding tags to each nick site. The tagged DNA is analyzed as it passes through nanochannels on the OhmX™ Detector. The standard EGM nickases cover > 90% of the human genome, but there are some regions where the nick density is too low and others where the density is too high for optimal performance.

The CRISPR/Cas9 system can be used to customize the tagging pattern by adding or removing nick sites. Sequence-specific blocking of nick sites is achieved with the relevant sgRNAs and dead Cas9 (dCas9), and nick sites can be created using Cas9 D10A nickase. Customizing nick sites requires minimal changes to the standard protocol.

Blocking sgRNAs allows directed improvements in contig length and coverage, which is advantageous across many genes of interest where too many nearby nick sites can systematically decrease the coverage in that region. By reducing the number of nick sites around important genes, such as *AKT1*, we have demonstrated contig alignment to this region of the genome that was previously inaccessible with the standard EGM genome prep chemistry.

Adding nicks using Cas9 D10A nickase is also helpful for introducing coverage across genomic regions with few nick sites in the standard EGM workflow. This is important for detecting large-scale genetic abnormalities involving centromeres, such as Robertsonian translocations. Nicks can also be added to optimize interval size, especially near genes of interest where the interval between nicks can impact how well an SV is detected.

These CRISPR/Cas9 approaches can also be used simultaneously to add and block nick sites in the same sample. This was utilized for the analysis of the repeat expansion causing Fragile X syndrome where multiple interval sizes and samples were successfully tested with this combined wet lab procedure. Samples were accurately analyzed with our RepX™ Repeat Expansion Analysis (RepX) pipeline, which quantified the number of repeats in the Fragile X syndrome samples tested.

These results demonstrate how CRISPR functions enhance the power and flexibility of EGM. By tuning nick site density, improved accuracy and coverage are obtained, and SVs can be detected in previously inaccessible genomic regions. This advancement expands the utility of EGM for comprehensive genome mapping and SV detection.

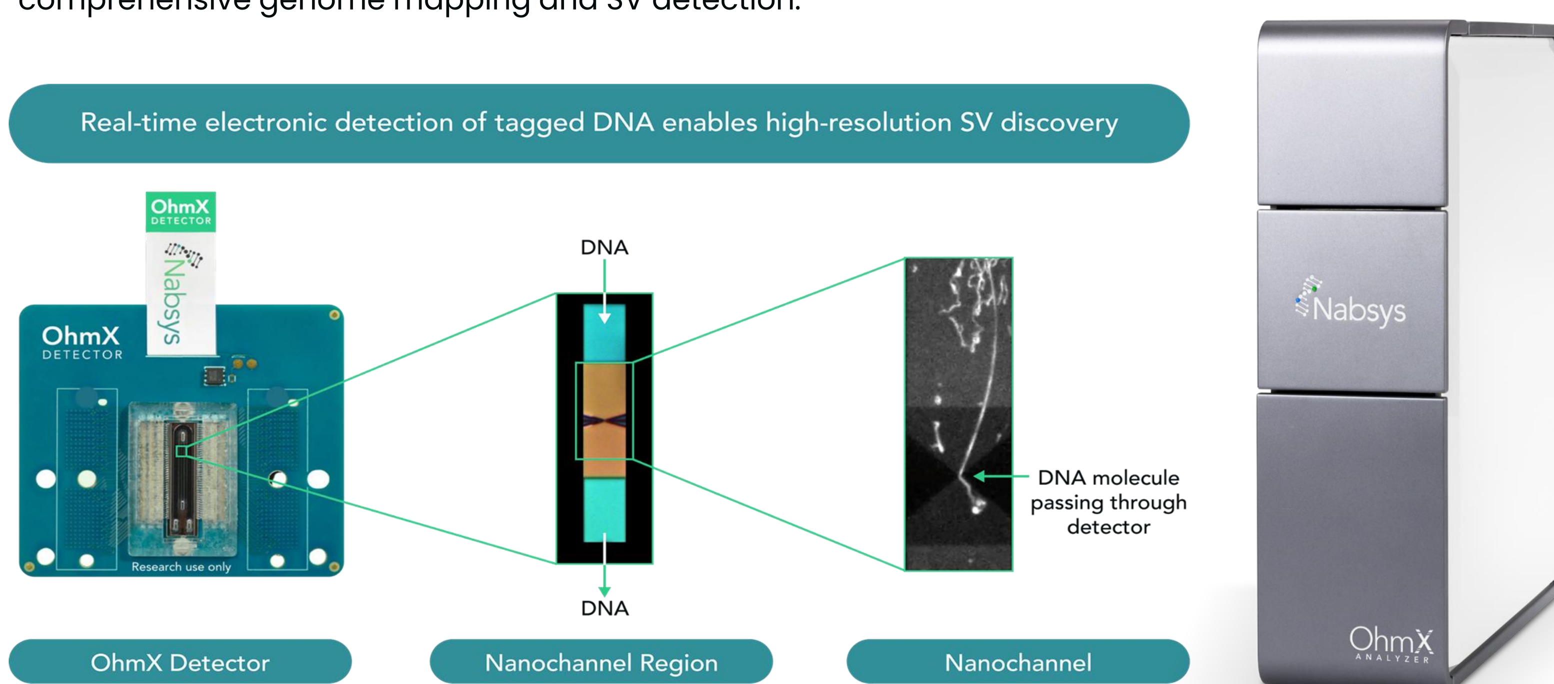

Figure 1. EGM detector and analyzer. EGM is enabled by the novel, solid-state nanodetectors developed by Nabsys. The OhmX Detector houses 256 parallel nanochannels, each with its own electronic sensor. Nabsys has developed DNA tagging chemistry that provides a high signal-to-noise. The high signal and spatial resolution allows tags to be closely spaced on the high molecular weight DNA molecules. This yields optimal tag density such that structural variations as small as 300 bp can be routinely identified and mapped genome-wide.

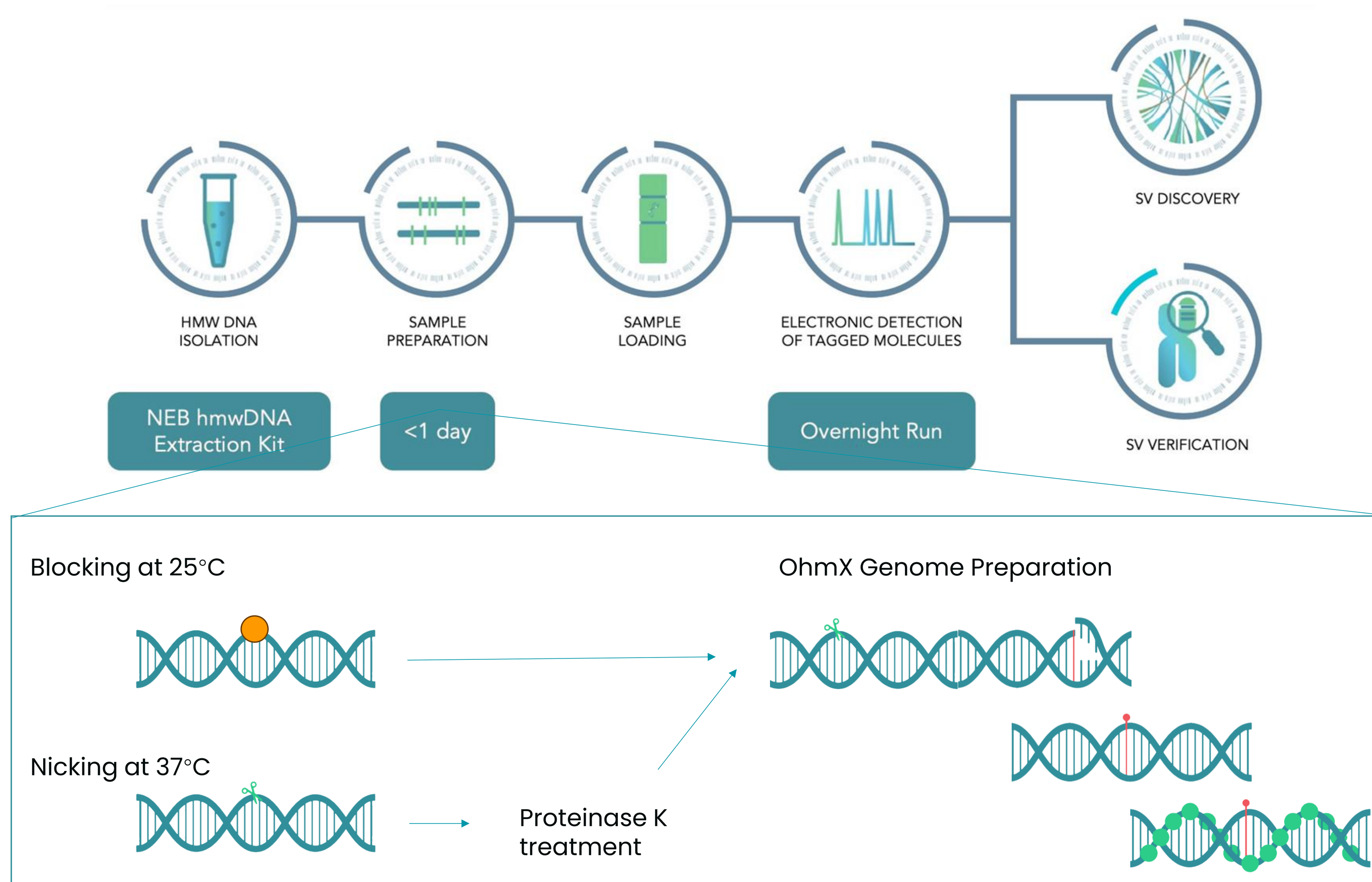

Figure 2. Overview of EGM and OhmX Genome Preparation workflow plus CRISPR. Cas9 ribonucleoprotein (RNP) complexes are made using sgRNA and dCas9 to block nick sites, or Cas9 D10A Nickase to add nick sites to target DNA. RNPs are mixed with DNA prior to the OhmX Genome Preparation workflow. Then, the DNA is loaded onto the OhmX Detector on the OhmX Analyzer for sample analysis.

Targeted Blocking of Nick Sites in Human DNA

Multiple classes of blocking sgRNA have been developed and tested for human sample applications:

- Repeat sgRNA that target many sites, not all of which are proximity breaks—some non-problematic sites will also be targeted (**Figures 3** and **4**)
- Unique sgRNA targeted to the following:
  - Proximity breaks in genes of high interest—one site per genome
  - Single-nucleotide polymorphism (SNP) detection

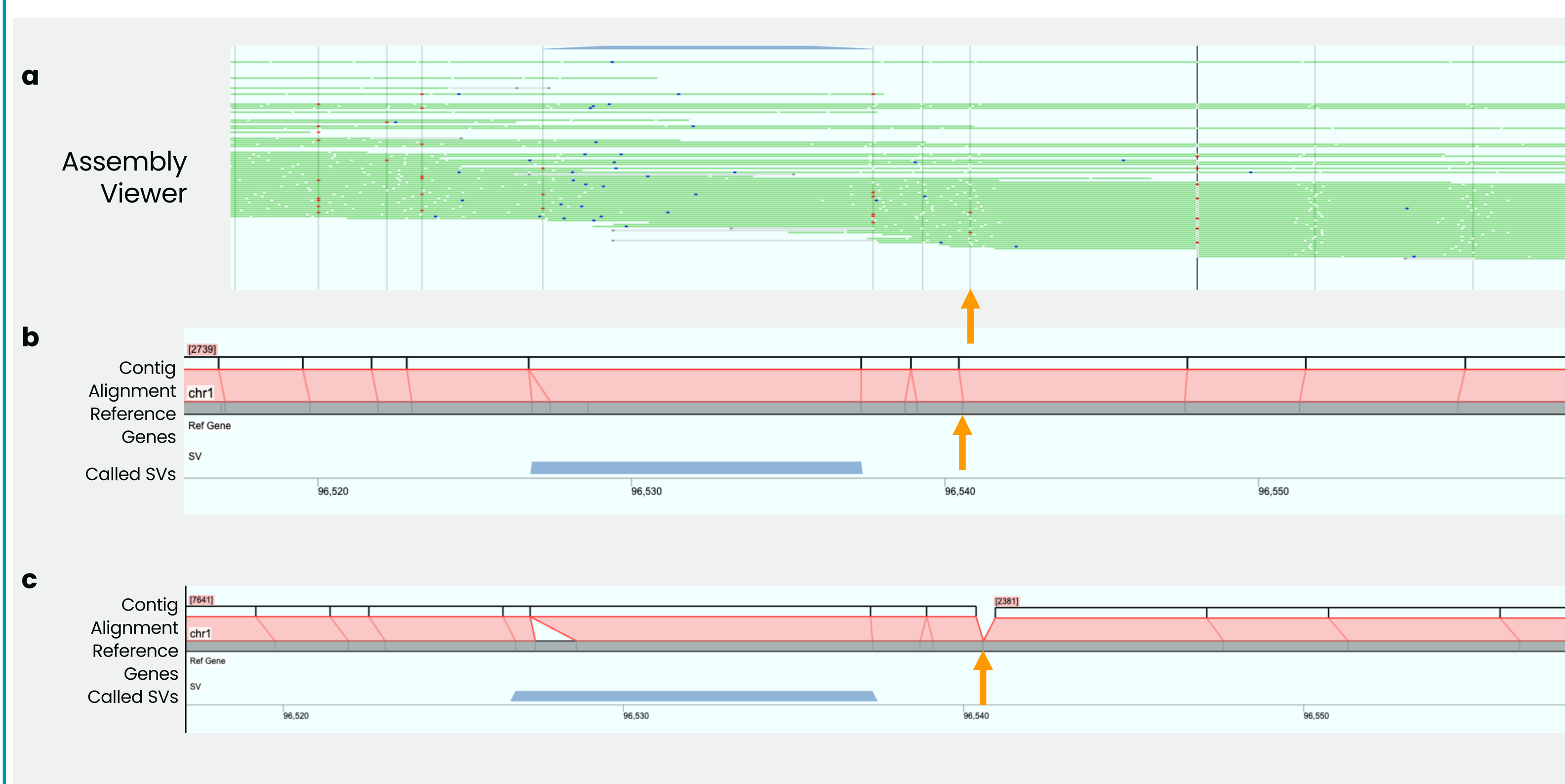

Figure 3. sgRNA blocking a proximity break to improve contig alignment on Human Chromosome Explorer® (HCE) assembly analysis. (a) Assembly view of molecules spanning across a proximity break that has been blocked using the sgRNA/dCas9. **(b)** Alignment view of a contig across a proximity break in a genome prepared with sgRNA/dCas9. **(c)** Alignment view of contig across proximity break using the standard OhmX Genome Preparation workflow.

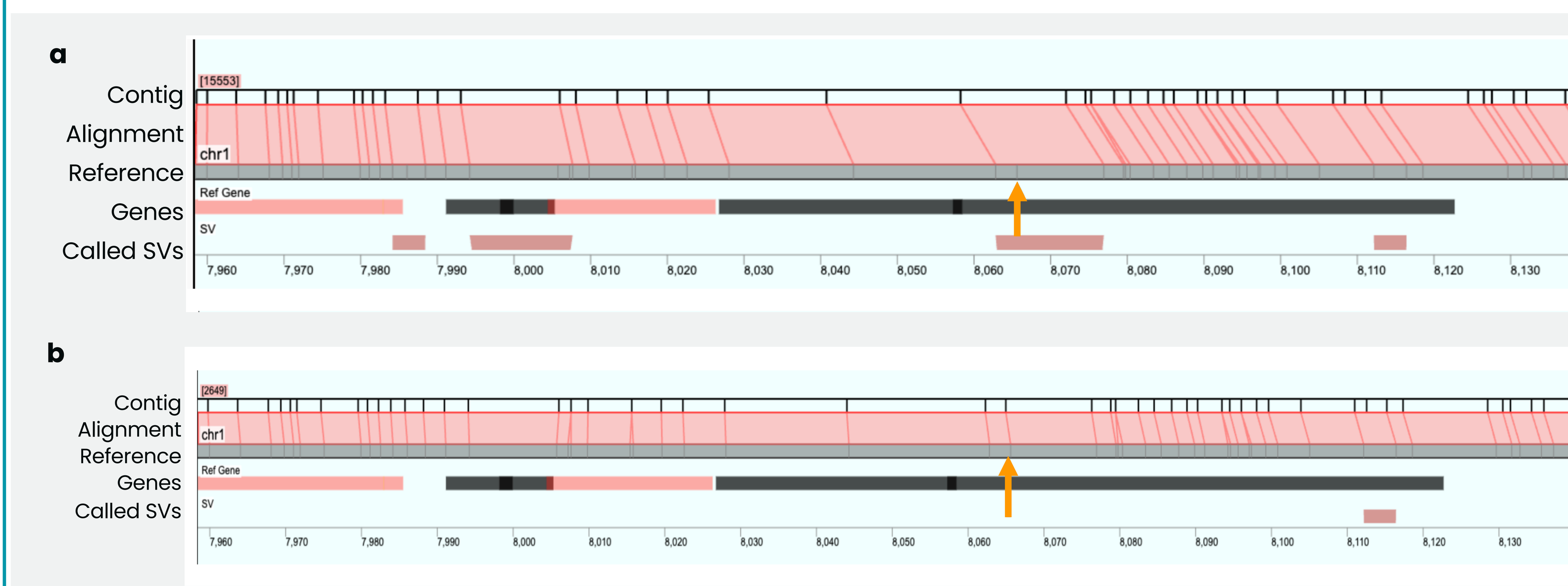

Figure 4. sgRNA blocking a non-problematic nick site on HCE assembly analysis. (a) Alignment view of a contig across the region with the blocked nick site in a genome prepared with sgRNA/dCas9. **(b)** Alignment view of a contig across the region with a blocked nick site in a genome prepared with the standard OhmX Genome Preparation workflow.

AKT1 is a gene of interest surrounded by multiple proximity breaks. This gene is important for regulating cell survival, proliferation, metabolism, and angiogenesis. Dysregulation is heavily associated with cancers, such as breast cancer and ovarian cancer.

Figure 5 outlines the impact of utilizing a unique sgRNA to target and block proximity breaks around *AKT1*. Without this sgRNA/dCas9 blocking, researchers would be unable to evaluate structural variants in this gene.

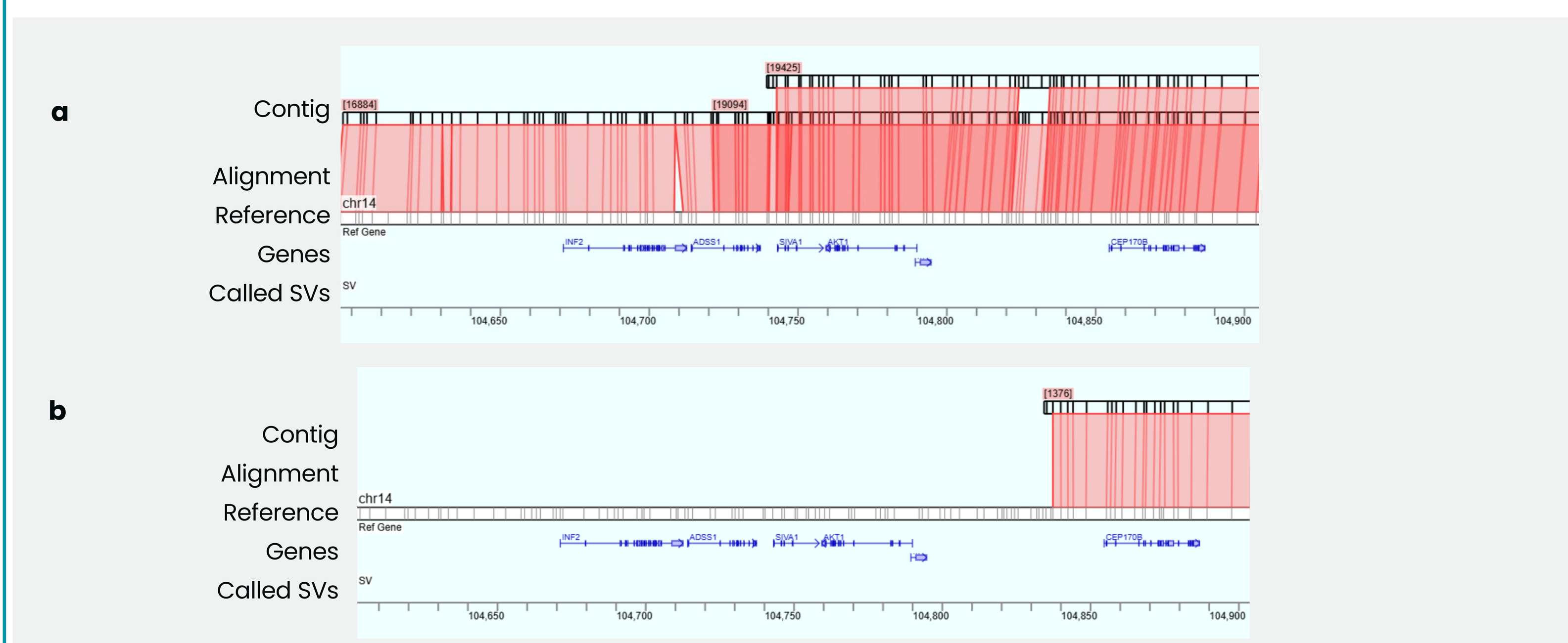

Figure 5. Targeted blocking of a problematic nick site near a gene of interest (*AKT1*) on HCE assembly analysis. (a) Alignment view of a contig across the region with blocked nick sites in a genome prepared with sgRNA/dCas9. **(b)** Alignment view of a contig across the region without blocked nick sites in a genome prepared with the standard OhmX Genome Preparation workflow.

Targeted Nicking of Custom Sites in Human DNA

Multiple classes of nicking sgRNA have been developed and tested for human sample applications:

- Repeat sgRNA that target many sites across the genome where the nicking density is not optimal for molecule alignment, including targeted nicking of the centromere for Robertsonian translocation analysis
- Unique sgRNA targeted to the following:
  - Repeat Expansions—such as Fragile X
  - Genes of interest—such as genes involved in evaluating FSHD

The RepX pipeline maps ultra-long DNA molecules to quantify the number of repeats associated with a repeat expansion. When a repeat expansion of interest falls within an interval that is not optimal for the RepX pipeline, using CRISPR/Cas9 allows for optimization of the target interval size and surrounding nick sites. Fragile X was studied as a proof of concept.

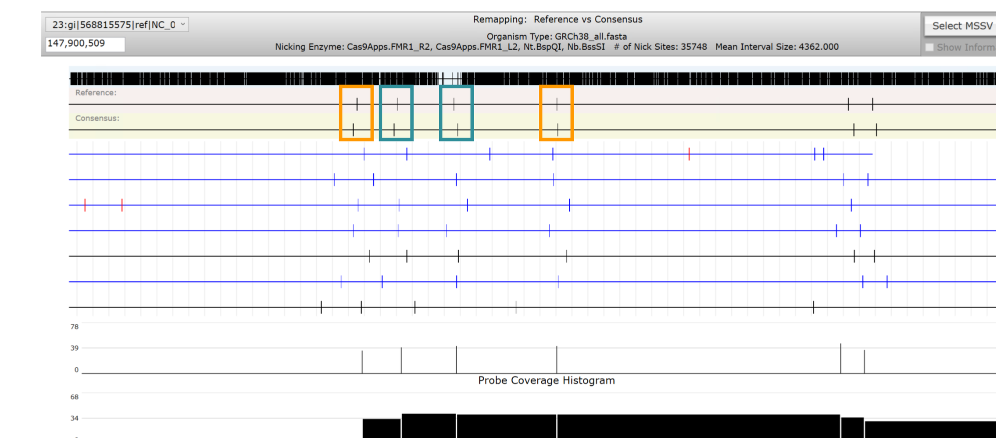

Figure 6. sgRNA nicking changes the interval size for repeat expansion sizing of *the FMR1* locus. The interval size is decreased from 6,099 bp (orange boxes) to 1,728 bp (blue boxes).

Table 1. Repeat sizing of Fragile X samples with and without CRISPR nicking. The CRISPR assay is efficient at introducing nicks into the sample, allowing accurate repeat expansion calling.

Sample	CRISPR Nicking	Expanded Allele				
		Repeat Annotation	P (Elevated)	P (Full Mutation)	Repeat Count	95% Confidence Interval
GM20239	No	183-193	100%	43.5%	192	117-271
GM20239	Yes	183-193	100%	48.8%	198	151-247
GM20239	Yes	183-193	100%	21.9%	184	148-221
GM20239	Yes	183-193	100%	65.7%	207	170-247

Using EGM for sgRNA/Cas9 Characterization

Further analysis of samples using Cas9 D10A nickase and dCas9 shows that EGM can be used for base-pair-level analysis of sgRNA and for assessing Cas9 proteins.

In addition to targeted genetic analysis, proof of concept testing shows that EGM can be used for the following:

- Test the distance an sgRNA can be from a nick site and still block using dCas9 (Table 2)
- Characterize the relative efficiency of Cas9 enzymes to nick and block target sites

Distance from sgRNA	Number of examples	Median Percent Tagged (Assembly 1)	Median Percent Tagged (Assembly 2)
5	18	53.4%	45.7%
9	40	96.9%	97.1%
10	4	94.9%	91.8%
12	1	94.4%	97.1%
14	4	92.7%	96.6%
15	1	98.7%	97.4%
24	1	96.4%	95.0%
9-24	51	96.4%	96.9%

Table 2. Distance of sgRNA from nick site for blocking evaluated across 2 assemblies using HCE.

Conclusions

Proof-of-concept testing across multiple applications shows that CRISPR/Cas9 can be used to introduce nick sites into regions of the genome where they are not prevalent or to remove nick sites that might result in proximity breaks. The CRISPR workflow adds minimal time to the existing OhmX Genome Preparation workflow and can be customized to the needs of the application.

Further work involves the following:

- Continuing to develop protocols for adding nicks to repeat sequences across the genome, including centromeres
- Designing sgRNA for more targeted gene applications to improve coverage across regions of interest that may not be well covered using the standard OhmX Genome Preparation workflow:
  - Through adding nick sites in larger intervals near genes of interest
  - Through blocking nick sites that impact contig alignment near genes of interest

Learn more at Nabsys.com