JULY 2019

gW webinar

The 'Dark Matter' of Cancer Genomics:

Revealing Undetected Structural Variants in Leukemia

SPONSORED BY: bionano GENOMICS

A **GENOMEWEB** WHITE PAPER

JAMES R. BROACH,
Distinguished Professor and Chair, Department of Biochemistry and Molecular Biology
Director, Penn State Institute for Personalized Medicine, Penn State College of Medicine

This white paper is based on a webinar presentation by Dr. James Broach of the Penn State College of Medicine. He discussed methods for capturing a comprehensive snapshot of all variants – both point mutations and structural variants – present in a tumor sample in order to gain insights about the genetic and genomic basis of individual cancers.

Dr. Broach noted that cancer is a somatic genetic disease that often exhibits extreme instability – particularly with regard to rearrangements.

His team sought to identify methods that could "fairly exhaustively" identify structural rearrangements associated with cancer and decided to focus on acute myeloid leukemia (AML) for several reasons. The first reason is that leukemia patients provide a ready source of "relatively pure cancer material" because the blood sample drawn from the patient "is the tumor." In addition, he explained that the Cancer Genome Atlas project has performed "extensive" genomic analysis on AML that has identified a number of point mutations as well as a large number of structural variations – gene fusions, translocations, inversions, or amplifications – that are associated with the onset of cancer and could be potential drivers.

Additionally, the molecular subtypes of AML are associated with overall survival. For example, "if you're unlucky enough to have AML but are lucky enough to have that AML driven by an inversion of chromosome 16, the likelihood that you'll survive after 10 years is quite high. On the other hand, if your AML is driven by an inversion of chromosome 3 the likelihood of survival is quite low," Dr. Broach said. "So being able to identify those mutations and structural variations that are responsible for the AML in a patient is really important in being able to predict how aggressive one needs to be in treating that subtype."

In line with this goal, Dr. Broach explained that his team wanted to sequence individual cancer genomes "in the absence of a reference genome in order to be able to capture any large structural variation." This is challenging, however, because short-read sequencing methods are not well suited for *de novo* sequencing due to repetitive sequences in the human genome.

In order to address this, Dr. Broach and his colleagues used two complementary approaches: whole-genome sequencing with Illumina NovaSeq™ and optical mapping with Saphyr® system from Bionano Genomics.

The Bionano system relies on nanofluidic chips with nanochannels that are just the width of a molecule of DNA, so that barcoded DNA molecules that move through the nanochannels "have to assume a strictly linear form. This allows a precise measurements of distances along a particular individual DNA molecule," Dr. Broach explained.

Images of the labeled and linearized megabase-length DNA are then converted into digital molecules, which are pairwise aligned to create "a large contiguous map of the whole genome. Once that cancer genome is assembled *de novo*, it can be compared to a reference genome to identify structural variants.

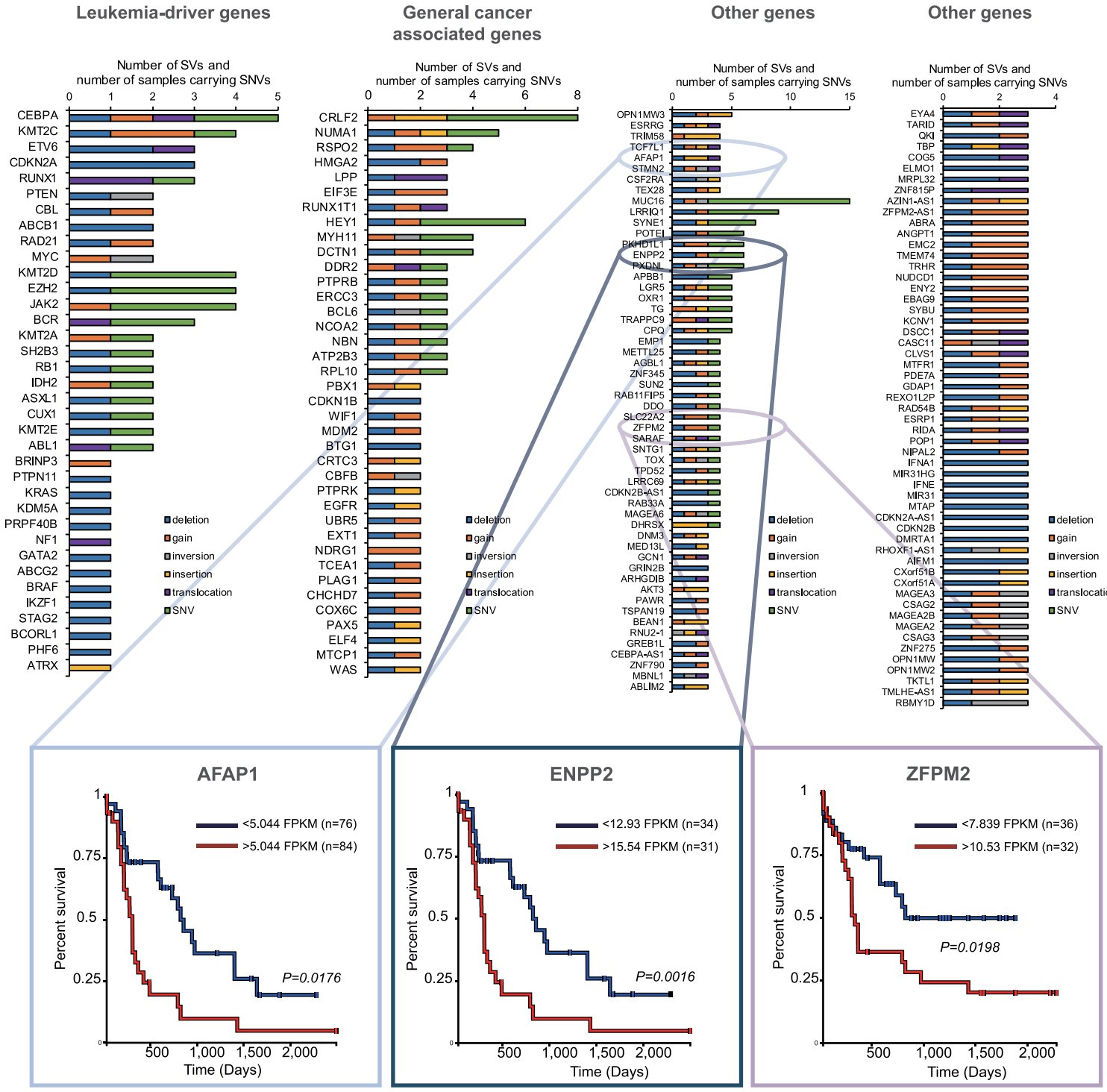
Dr. Broach described how his team applied this method in a study using blood samples from 12 leukemia patients. The approach was "to do optical mapping and whole-genome sequencing on each of these samples, take those data, merge them to generate a collection of structural variants and single nucleotide variants, and then use a bioinformatics pipeline to remove those polymorphisms that were likely present in the genome (i.e. germline variants) and not somatic mutations that were present only in the cancer."

The computational pipeline to remove germline polymorphisms is a "critical" component of the method, Dr. Broach said. The pipeline compares the variants identified in each of the patient samples to databases of known point mutations and structural variants. "We assume that if that polymorphism has been identified before, then it's likely a germline variant and not a somatic variant," he said.

The 'Dark Matter' of Cancer Genomics:
Revealing Undetected Structural Variants in Leukemia

Figure 1: New Cancer Genes Affect Patient Outcomes

The study identified structural variants that had previously been associated with leukemia as well as other cancers. In addition, a number of genes not previously implicated in cancer were associated with structural variants in multiple leukemia samples. These genes, which included AFAP1, ENPP2, and ZFPM2, were associated with patient outcomes.



The 'Dark Matter' of Cancer Genomics:
Revealing Undetected Structural Variants in Leukemia

The majority of structural variants that are detected in patient samples are germline, he noted, estimating that only about 5 percent of these variants are somatic. This approach addresses the fact that tumor/normal sequencing is not an option to identify somatic variants in leukemia because "the tumor is in fact the blood, so it's very difficult to get material that we can use as the reference for the germline for each individual."

In several cases, he said, "we were able to circumvent this by taking advantage of the fact that the patient's sample is not 100 percent tumor but has a small percentage of normal cells within that sample."

In these cases, Dr. Broach and colleagues stimulated growth of the resident normal T cells in the sample, performed cell sorting on this enriched population of T cells, and then analyzed them with the Bionano technology as well as whole-genome sequencing.

"Then, by comparing all of the structural variants that we see in the patient sample to what we see in the normal sample, we can determine how accurate our bioinformatics pipeline is for removing the germline polymorphism." Based on this analysis, he said, "we were able to calculate that our false discovery rate for structural variation was less than 6 percent (somatic versus germline)."

Dr. Broach said that somatic and germ line structural variants "differ" depending on the particular structural variation that we look at."

For example, "a reasonable fraction of the gains and duplications that we identify in our samples are actually somatic, whereas most of the deletions, insertions, and inversions that we identify are polymorphic and germline, and only a small fraction are actually somatic."

Ultimately, he said, whole-genome sequencing "is very good at picking up gains, whereas the optical mapping is much better at picking up losses and insertions and inversions and both technologies are very effective at picking up translocations."

Dr. Broach outlined a number of examples where this method identified structural variants with functional consequences that would not have been identified through cytogenetics or whole-genome sequencing alone. "Many of the structural variants are affecting cancer genes that have not been previously associated with leukemia," he said. Furthermore, they have also identified genes that had been associated with other cancers, "but we find recurrently affected by structural variants in our patient samples."

These genes, which include ENPP2, ZFPM2, and AFAP1, "are worth further study to understand the progression and etiology of leukemia," he said. (see Figure 1)

Dr. Broach's team also found a number of structural variants in intergenic regions that don't affect the sequence of any nearby coding regions but do reduce or increase the expression of nearby genes. "What we found is that for about 50 percent of these intragenic structural variants, we see an alteration in the expression of the associated gene," he said.

Dr. Broach and his colleagues are currently working with Bionano Genomics to apply these methods to solid tumors using as little as 10 mg of tissue. In one project, the team is studying HPV (human papilloma virus)-induced head and neck cancer in order to gain insight into the etiology of the disease.

Much of the current understanding of HPV-induced tumorogenesis comes from studies on cervical cancer, where integration of the HPV16 genome leads to the loss of the E2 gene, which then increases the expression of E6 and E7 and ultimately leads to increased cellular growth. Dr. Broach noted that "the data to date suggests that the model from cervical cancer doesn't really apply to head and neck cancer" but rather it appears that HPV is associated with amplification of specific regions of the genome.

"We've used our techniques to look at the structural alterations in various head and neck cancers and can identify a large number of structural variants," he said. "One of those affects the Rb gene, which is not quite consistent with the story of the cervical cancer paradigm."

This work is just beginning, Dr. Broach said, but "we're continuing these studies to be able to analyze further the nature of the genome rearrangements associated with head and neck cancer."

Importantly, he noted, "optical mapping provides otherwise inaccessible structural clarity of HPV-induced genome alterations."

For further details on the study Dr. Broach discussed in the webinar, please see: Xu J, Song F, Schleicher E, et al. 2019. ***An Integrated Framework for Genome Analysis Reveals Numerous Previously Unrecognizable Structural Variants in Leukemia Patients' Samples.***