A Comprehensive Multi-Platform Approach for Structural Variant Analysis in a Clinical Cohort of Cancer Patients

Camir Ricketts[1], Dmitrii Meleshko[1], Rui Yang[1], Alicia Alonso[1], Can Alkan[2], Olivier Elemento[1], Iman Hajirasouliha[1]

[1] Institute of Computational Biomedicine, Weill Cornell Medicine of Cornell University
[2] Department of Computer Engineering, Bilkent University

TRI-INSTITUTIONAL PHD PROGRAM Computational Biology & Medicine

Summary

Question: What patterns of structural variations can a multi-platform approach uncover within a cohort of cancer patients?

Task: Use orthogonal sequencing data from Bionano, Linked-Reads, and short-read WGS to show how each can complement each other in identifying variations.

Results: We identified multiple classes of SVs with each platform. 115 average sample specific calls with Bionano, 80 variants overall (Valor2) and 693 variants overall (longranger).

Discussion: The length distributions of the SV calls from the platforms show that these technologies can complement each other in identifying small to large SVs. We also identify high confidence SVs that potentially interrupt known cancer genes.

Abstract

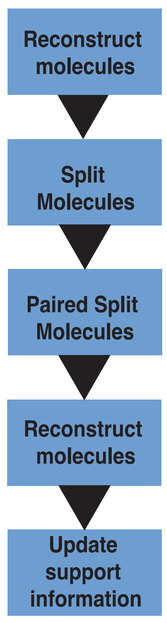
While there have been notable advancements in structural variant analysis in tumor genomes, there is still a lack of comprehensive analysis of tumors integrating multiple technologies to fully characterize events not previously discovered using current tools. In this study, we have applied a non-sequencing-based genome mapping technology, Bionano, together with 10x Genomics linked-reads and standard whole genome sequencing to three gastric cancer, two brain cancer and one gastrointestinal stromal tumor (GIST) samples. Leveraging the Bionano technology, we captured structural variants including those at low frequency by combining assembly and rare bionano SV pipelines. We identified 24 sample-specific variants in brain cancer samples, 90 sample-specific variants in GIST and average of 231 sample-specific variants in gastric cancer. Additionally, we present results of Longranger SV calling as well as our newly developed tool VALOR2, a mapping based algorithm for characterizing complex large-scale SVs from 10x linked-reads on this cohort. We show that an integrated approach is able to identify more variants than one single approach (an average increase of 400 inversions over any one method, for example), and characterize the composition of structural variants for each sample. Consensus calls allows for the identification of false positives or resolving more accurate breakpoint locations. Valor2 is able to identify an average of 80 variants across all samples, a majority of which is not identified by longranger. We also identified events which affect candidate cancer genes such as PSD3, IFNA6, MIR4767, among others. We present a comprehensive set of variations in gastric and brain cancer ranging from simple events to large and complex events such as reciprocal and inverted-reciprocal translocations in order to identify clinically relevant events.

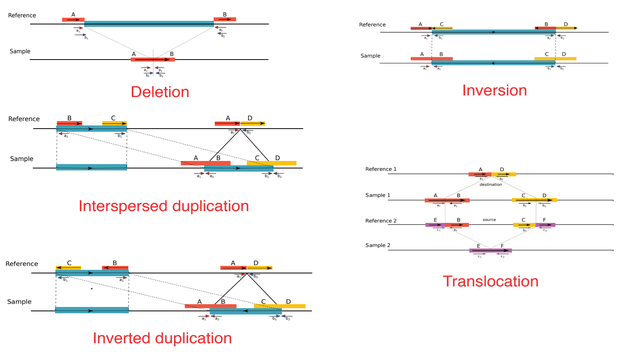
Methods

Patient	Tumor Type	Fusions	Bionano	10X Linked-Read	WGS
001	GIST	46	✓	✓	✓
002	Gastric Cancer	31	✓	✓	✓
003	Brain Cancer	27	✓	✓	✓
004	Brain Cancer	27		✓	
005	Gastric Cancer	23		✓	✓
006	Gastric Cancer	24	✓	✓	✓

Table 1: Sequencing plan and details of eacg patient in pilot cohort

VALOR2





Results

Our early analysis includes 6 pilot genomes from the Englander Institute for Precision Medicine.

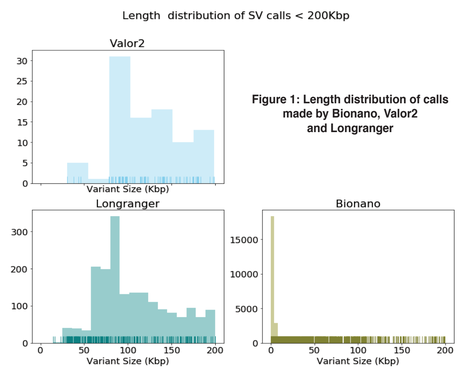


Figure 1: Length distribution of calls made by Bionano, Valor2 and Longranger

SVs not Bionano database	Patient 3	Patient 2	Patient 1	Patient 6
Insertions	12	31	25	31
Deletions	21	42	38	67
Inversion	1	19	21	23
Duplications	1	77	10	5
Interchr-translocations	0	31	16	78
Intrachr-translocations	0	13	14	89

Table 2: Sample-specific SVs identified by Bionano pipeline.

Figure 1: Number of structural variants identified for all samples across population.

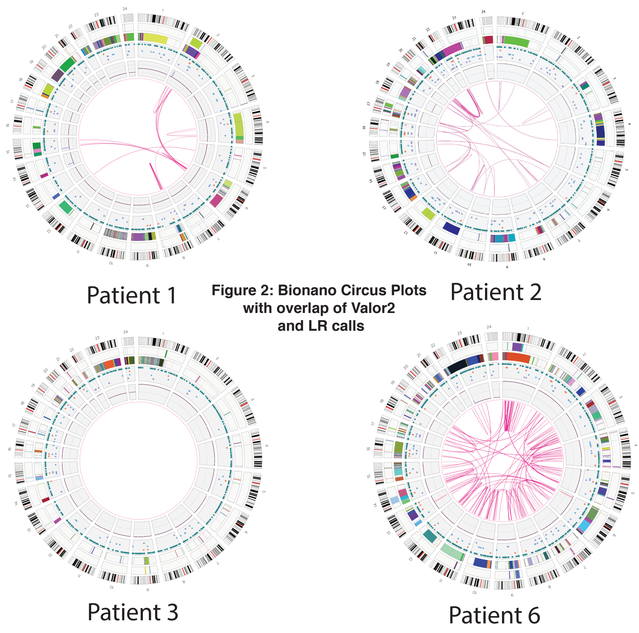


Figure 2: Bionano Circus Plots with overlap of Valor2 and LR calls

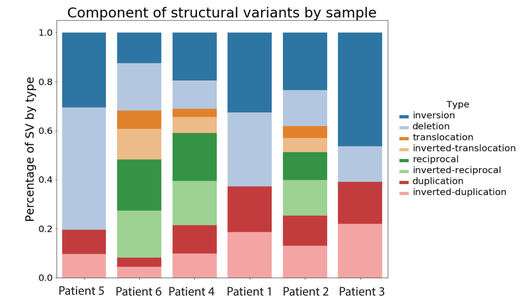


Figure 3: Structural variants within pilot cohort identified by Valor2

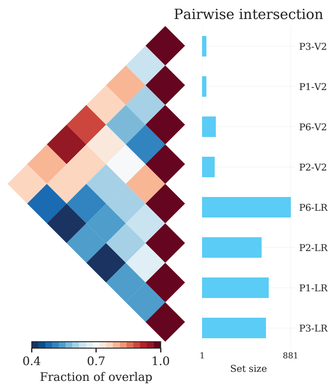


Figure 4: Overlap comparison of Valor2 and Longranger calls

Discussion

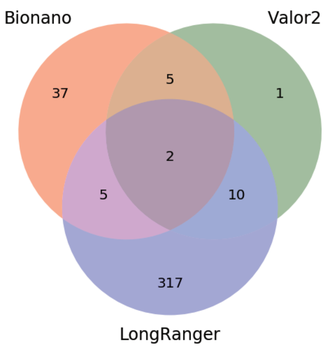


Figure 5: Overlap of large deletion calls in Patient 2

Figure 6: 333 Kbp homozygous inversion in Patient 1. Potentially interrupts FRMD5

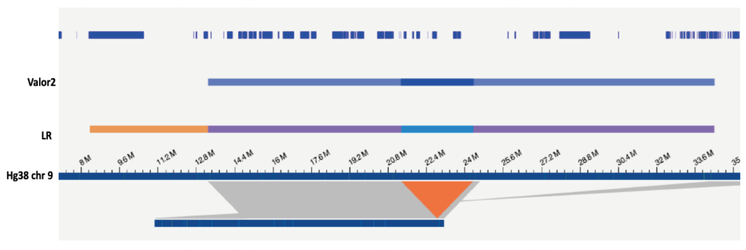


Figure 7: 21 Kbp deletion in Patient 6. Interrupts IFNA6

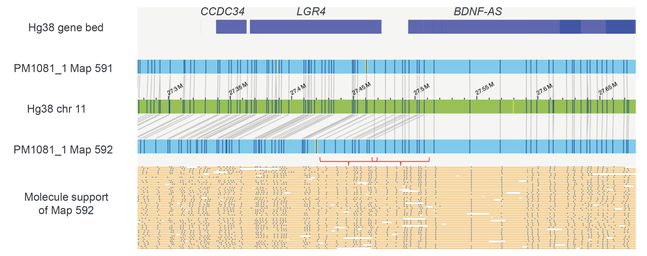


Figure 8: 40 Kbp duplication identified by Bionano in Patient 2 disrupting LGR4 and BDNF-AS

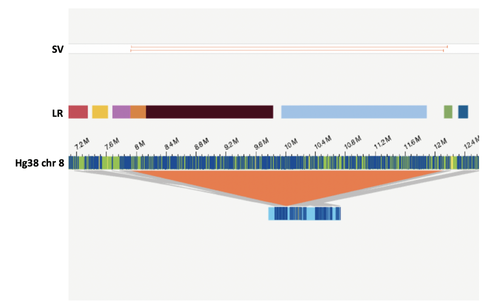


Figure 9: 4.2 Mbp heterozygous deletion idenitified in Patient 6. Bionano reveals a single large deletion event which is covered by multiple predicted events by Linked-Reads.

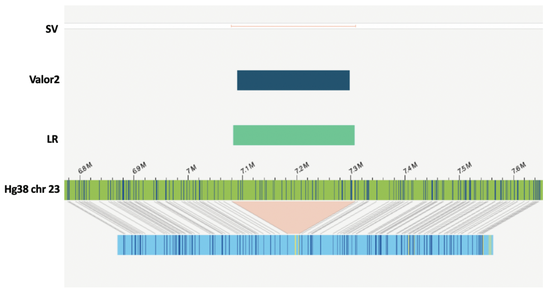


Figure 10: 200 Kbp heterozygous deletion in Patient 2 showing concordance among platforms. Interrupts micro RNAs, including MIR-4767

References

- Chaisson, M. J.et al.(2019). Multi-platform discovery of haplotype-resolved structural variation in human genomes.Nature communications,10.
- Marks, P.et al.(2019). Resolving the full spectrum of human genome variation using linked-reads.Genome Research.
- Karaoglanoglu, Fatih, et al.(2018). “Characterization of Segmental Duplications and Large Inversions Using Linked-Reads.” Biorxiv, 394528. https://github.com/BilkentCompGen/valor
- Software Downloads - Bionano Genomics. Available at: https://bionanogenomics.com/support/software-downloads/.