

Enhanced Detection of Structural Variants, VNTRs, and Haplotype Phasing with SBX Simplex Longer Sequencing (SBX-SL)

Chen Zhao,² Aaron Jacobs,¹ Alberto Gatto,² Alex Lehmann,¹ Alex Mitchell,¹ Chad Rosevear,¹ Cynthia Cech,¹ Daniel Zinder,² Dilmi Perera,² Donald Pacaba,¹ Erfan Sayyari,² Grant Kingsley,¹ Jagadeeswaran Chandrasekar,¹ Janie Johnson,¹ John Mannion,² Jordan Eizenga,² Kendall Berg,¹ Kevin Smith,¹ Lacey McGee,¹ Mahdi Golkaram,² Majid Babazadeh,¹ Marc Prindle,¹ Melanie Kirsche,² Melud Nabavi,¹ Mitchell Wolfin,¹ Robert McRuer,¹ Salka Barrett,¹ Sam Salari,² Scott Miller,¹ Svetlana Kritzer,¹ Thomas Reid,¹ Upneet Bala,¹ Yanli Hou,² Mark Kokoris¹

¹Roche Sequencing Solutions, Inc, Seattle, WA, USA; ²Roche Sequencing Solutions, Inc, Santa Clara, CA, USA; ³Roche Diagnostics GmbH, Penzberg, Germany



Copies of this poster obtained through QR (Quick Response), and/or text key codes are for personal use only and may not be reproduced without written permission of the authors.

Poster no. 4081T

Introduction

Sequencing By Expansion (SBX) is a high-throughput, flexible, and rapid-turnaround sequencing technology.¹ While SBX duplex sequencing (SBX-D) offers high accuracy for detecting small germline variants such as SNVs and INDELs, its ability to resolve larger genomic features—such as structural variants (SVs), tandem repeats (TRs), and long-range haplotypes—is limited. Here, we present SBX Simplex Longer sequencing (SBX-SL), an extension of the SBX platform that significantly enhances the detection of SVs and improves haplotype phasing in whole-genome sequencing (WGS) workflows. SBX data were processed through SBX-optimized open source (XOOS) variant callers.²

Introduction to SBX-SL

SBX Simplex Longer (SBX-SL) is a library prep and sequencing protocol that is able to achieve longer read lengths than the standard SBX duplex (SBX-D) workflow. It includes adapter ligation, linear amplification, Xpandomer synthesis, and sequencing

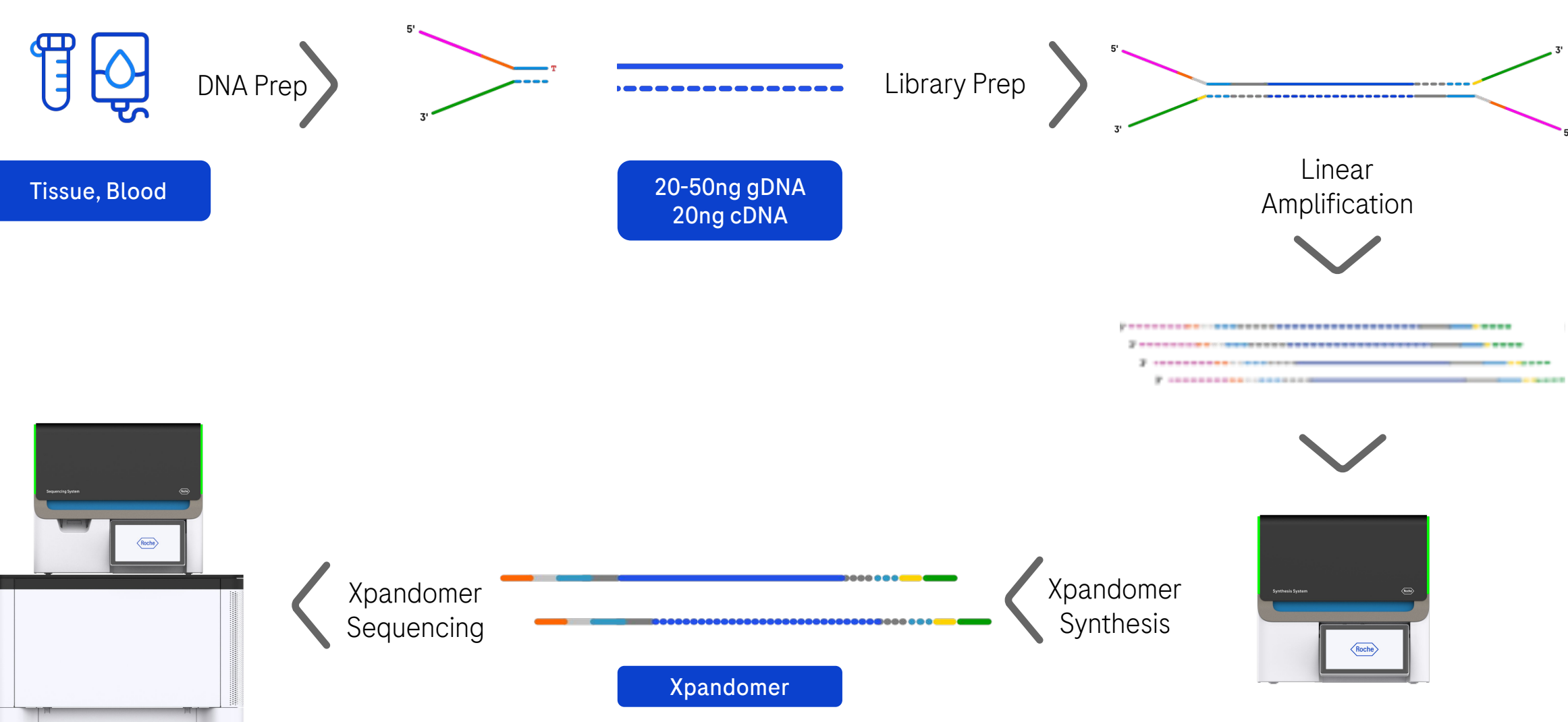


Fig 1. SBX-SL Workflow

This workflow is compatible with extracted DNA as well as already amplified cDNA libraries for RNA sequencing.

Results

The extended read length offered by SBX-SL sequencing provides a range of advantages for genomic discovery, particularly in regions where traditional short-read sequencing is limited. SBX-SL can improve the detection of structural variants and enables more accurate long-range phasing of variants on the same haplotype. To demonstrate these capabilities, we benchmarked SBX-SL using well-characterized cell lines, including somatic tumor/normal pairs such as HCC1395 and H2009. As shown in Fig. 2, longer reads show more evidence for expected SVs in DNA and RNA sequencing.

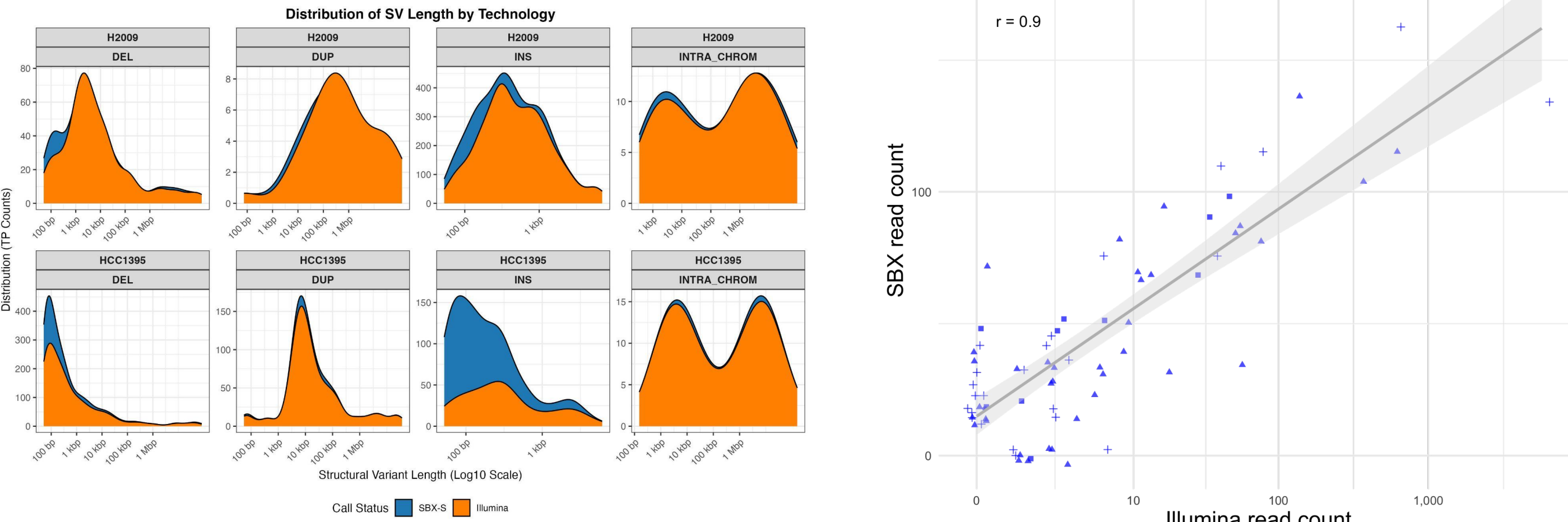


Figure 2. (left) break down of SV detection performance of SBX SL DNA on 2 well-characterized cancer cell lines.^{3,4} (right) performance of SBX-SLR in measurements of SV expression from RNAseq (HCC1395).⁴

Results

Variable Number Tandem Repeats (VNTRs) are short DNA sequences repeated a variable number of times in tandem at specific genomic locations. Their high variability between individuals makes them valuable for genetic studies, including disease association, forensics, and population genetics. However, due to their repetitive nature, short-read sequencing struggles to accurately genotype VNTRs in whole-genome sequencing (WGS) data. In contrast, the longer read lengths of SBX-SL sequencing can effectively span most VNTR regions, enabling accurate detection of VNTR expansions in the human genome.⁵⁻⁷ Benchmarking on the HG002 reference cell line shows that SBX-SL provides better performance compared to SBX-D in detecting these events—particularly those larger than 100 base pairs (Fig. 3).

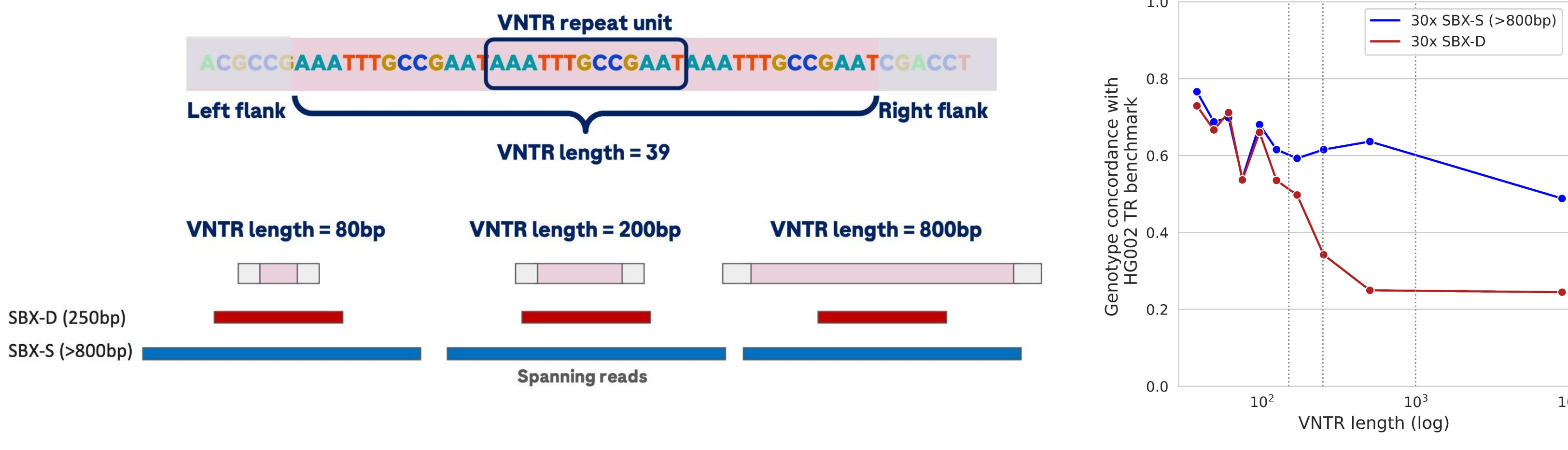


Figure 3. Accurate detection of VNTR variants requires long reads that can span VNTR sequence. SBX-SL can improve VNTR genotyping compared to SBX-D.

Variant phasing is the process of determining which genetic variants are inherited together on the same chromosome (haplotype). Long-read sequencing helps with phasing by spanning multiple variants within a single read, allowing accurate reconstruction of haplotypes over longer genomic distances. Benchmarking on HG002, SBX-SL is able to show better performance compared to short read sequencing specially in variants further apart than 500 bp (Fig. 4). Moreover, long range phasing can also improve detection of compound heterozygous variants.

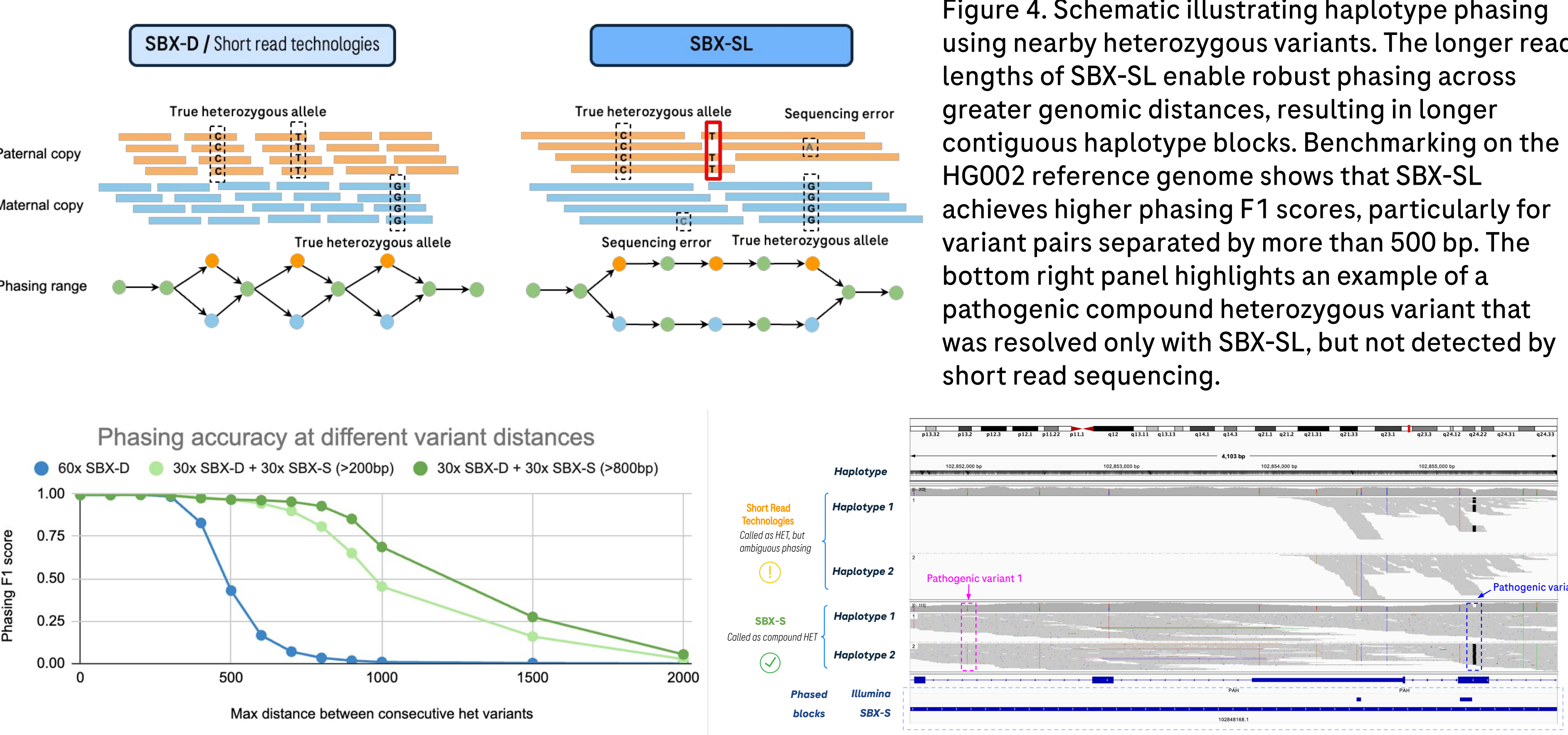


Figure 4. Schematic illustrating haplotype phasing using nearby heterozygous variants. The longer read lengths of SBX-SL enable robust phasing across greater genomic distances, resulting in longer contiguous haplotype blocks. Benchmarking on the HG002 reference genome shows that SBX-SL achieves higher phasing F1 scores, particularly for variant pairs separated by more than 500 bp. The bottom right panel highlights an example of a pathogenic compound heterozygous variant that was resolved only with SBX-SL, but not detected by short read sequencing.

Results

Finally, SBX-SL is designed to enable multi-omics discovery by integrating longer sequencing from both DNA and RNA. As illustrated in Fig. 5, allele-specific RNA expression can be resolved through phasing of SBX-SL RNA reads. SBX-SLR accurately detected exonic variants seen in SBX-SLD; however, one allele is significantly over expressed.

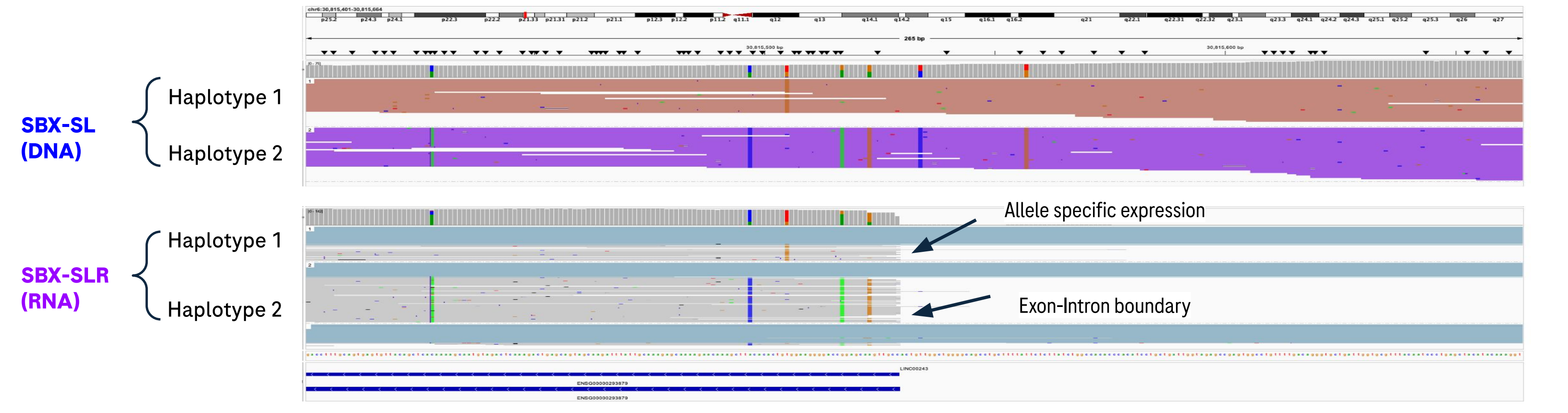


Figure 5. IGV snapshot illustrating multi-omics phasing using SBX-SL technology in the HG002 cell line. Reads were first phased using SBX-SL DNA data. RNA reads were phased and accurately assigned to paternal and maternal alleles. While SBX-SLR and SBX-SLD exonic variants show high concordance, SBX-SLR indicates strong allelic imbalance in expression of the second allele.

Conclusion

In summary, SBX-SL showed improvement in the detection of large-scale germline and somatic events—including structural variants, tandem repeats, and long-range phasing—from both DNA and RNA sequencing. It serves as a powerful complement to SBX-D across whole-genome sequencing (WGS) applications. Combined, SBX-SL and SBX-D can leverage SBX-D accuracy and SBX-SL longer reads to enable broader genomics applications.

Disclosures

This study was funded by Roche Diagnostics International Ltd (Rotkreuz, Switzerland)
All authors are employees of Roche Sequencing Solutions, Inc. or Roche Diagnostics GMBH and may hold non-voting equity securities in F. Hoffmann-La Roche Ltd.

References

- https://doi.org/10.1101/2025.02.19.639056
- https://roche-axelios.gitbook.io/xaos
- Keskus, Ayse G., et al. Nature Biotechnology (2025): 1-11.
- Talsania, Keyur, et al. Genome biology 23.1 (2022): 255.
- English, Adam C., et al. Nature Biotechnology (2024): 1-12.
- Javadzadeh, Sara, et al. PLOS Computational Biology 21.4 (2025): e1012885.
- Method: adVNTR used for genotyping of pVNTR + gVNTR catalog in Javadzadeh, et al.