Translocator: local realignment and global remapping enabling accurate translocation detection using single-molecule sequencing long reads

Ye Wu
Department of Computer Science
The University of Hong Kong
Hong Kong, China
ywu@cs.hku.hk

Ruibang Luo
Department of Computer Science
The University of Hong Kong
Hong Kong, China
rbluo@cs.hku.hk

Tak-Wah Lam
Department of Computer Science
The University of Hong Kong
Hong Kong, China
twlam@cs.hku.hk

Hing-Fung Ting
Department of Computer Science
The University of Hong Kong
Hong Kong, China
hfting@cs.hku.hk

Junwen Wang
Department of Health Sciences Research
Mayo Clinic Arizona
Scottsdale, United States
wang.junwen@mayo.edu

ABSTRACT

Translocation is an important class of structural variants known to be associated with cancer formation and treatment. The recent development in single-molecule sequencing technologies that produce long reads has promised an advance in detecting translocations accurately. However, existing tools struggled with the high base error-rate of the long reads. Figuring out the correct translocation breakpoints is especially challenging due to suboptimally aligned reads. To address the problem, we developed Translocator, a robust and accurate translocation detection method that implements an effective realignment algorithm to recover the correct alignments. For benchmarking, we analyzed using NA12878 long reads against a modified GRCh38 reference genome embedded with translocations at known locations. Our results show that Translocator significantly outperformed other state-of-the-art methods, including Sniffles and PBSV. On Oxford Nanopore data, the recall improved from 48.2% to 87.5% and the precision from 88.7% to 92.7%. Translocator is available open-source at https://github.com/HKU-BAL/Translocator.

KEYWORDS

Structural Variant, Translocation, Local realignment, Global Remapping

BCB '20, September 21–24, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-7964-9/20/09…$15.00
https://doi.org/10.1145/3388440.3412457

ACM Reference format:

Ye Wu, Ruibang Luo, Tak-Wah Lam, Hing-Fung Ting and Junwen Wang. 2020. Translocator: local realignment and global remapping enabling accurate translocation detection using single-molecule sequencing long reads. In *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB'20). ACM-BCB,* September 21–24, 2020, Virtual Event, USA, 7 pages. https://doi.org/10.1145/3388440.3412457

1 Introduction

Translocation, often referred to as its abbreviation "TRA", is a type of structural variants (SV) defined as reciprocal exchange of parts at least 50 base-pairs (bp) in size between non-homologous chromosomes [1]. Translocations were found to play an important role in the early steps of tumorigenesis [2-4]. At the molecular level, the consequence of translocation is manifold. For example, the regulatory elements of a normal gene might be replaced, resulting in abnormal expression of a normal gene product [2, 5]; or two genes might be placed together and form a chimeric fusion gene [6, 7]. A previous study has shown that translocations that lead to gene fusion account for 20% of human cancer morbidity [2]. Thus, identifying translocation breakpoints sensitively and accurately is essential to pinpoint the genes being affected and lead to valid cancer diagnosis and treatment.

Traditionally, translocations are detected using the Next-Generation Sequencing (NGS) short-reads with methods devised for distinguishing genuine SV signals from the background noise [8-10]. These methods are based on one or more information on "read-depth", "pair-end", "split-read", and "assembly" [9]. However, the short read length of NGS (ranging from 100bp to 250bp) induces poor read alignments in the low-complexity regions [11], leading to an excessive amount of spurious translocations being detected among these regions. Insertions and deletions are

easily misreported as translocations. It is estimated that more than 80% of the translocations detected from Illumina short-reads are false positives [12].

Single molecular sequencing that produces long read promises more reliable and accurate detection of SVs [13]. Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) are two major technology providers and can produce reads averaging several thousand base-pairs or even up to 2Mbp for ONT [14]. Longer read length results in more accurate alignments in the low-complexity regions that should, in turn, leads to more high-confidence alignments spanning SV breakpoints to be produced. However, challenges remained, with the top one being to accommodate the very high per-base error rate in the long reads (13–15% in PacBio and 5-20% in ONT) [15].

Two key steps are crucial to the quality of SV detection using long reads. The first step is to align the long reads to a reference genome rapidly and correctly. A few aligners were developed for this purpose, including LAST [16], BLASR [17], GraphMap [18], NGMLR [12] and minimap2 [19]. Specifically designed for aligning the PacBio reads, NGMLR uses an SV-aware seeding strategy and a convex gap cost model to compute precise alignments. It achieved the best alignment accuracy on PacBio reads [12]. In contrast, aligning ONT reads is much more computational demanding due to longer average read length and higher per-base error rate at homopolymers. Minimap2 effectively sped up aligning ONT reads by doing the split-read alignment and employing concave gap cost for long insertions and deletions. With comparable accuracies, minimap2 is over 30 times faster than the other long-read aligners [19]. Later in our results, NGMLR and minimap2 were used for aligning PacBio and ONT reads, respectively.

The second step is to detect multiple classes of SV including translocation using the alignment results. Multiple methods for detecting SVs using long reads were released recently, with one or more limitations applied. Most of the methods apply to either PacBio or ONT reads, and require a specific aligner [14]. For example, PBHoney [20] works for PacBio with BLASR alignments only; Picky [21] and NanoSV [22] were designed for ONT reads and require using the LAST aligner. The limitation of these methods hindered them from leveraging the power of the new aligners such as NGMLR and minimap2 and being applied to new types of sequencing reads. Sniffles removed these limitations as it implements a parameter estimation step to fit its error model to different aligners and sequencing technologies [12].

However, compared to detecting unbalanced SV classes such as insertion and deletion, translocations were usually detected with much lower sensitivity and accuracy [12]. Existing methods commonly detect translocations base on split reads. Whether a translocation could be identified confidently depends on the number of read alignments consistently supporting a set of potential breakpoints. However, due to the complexities in real sequencing data and the limitation of alignment algorithms, reads are often misaligned around the translocation breakpoints. As a result, translocations without enough alignment supports are being ignored or mistaken for deletions or inversions.

In this study, we present Translocator, a robust and accurate translocation detecting method for both PacBio and ONT reads. Translocator detects misaligned reads around candidate translocation breakpoints and tries to realign them to increase potential supports at the breakpoints to rescue translocations with weak supports in low-complexity regions or with low sequencing depth coverage. Translocator was implemented in Sniffles' framework to leverage its high efficiency and compatibility with various aligners.

To evaluate the performance of Translocator, we benchmarked it using real NA12878 PacBio, and ONT reads on simulated translocations with length ranging from 100 to 3,000 base-pairs embedded into the GRCh38 reference genome. Compared with state-of-the-art methods including Sniffles and PBSV [23], Translocator performed outstandingly. On ONT data, Translocator improved the recall from 48.2% to 87.5% and the precision from 88.7% to 92.7%. We evaluated Translocator at multiple subsampled depth coverages (from 5-fold to 30-fold) and it outperformed existing methods consistently. At 10-fold,

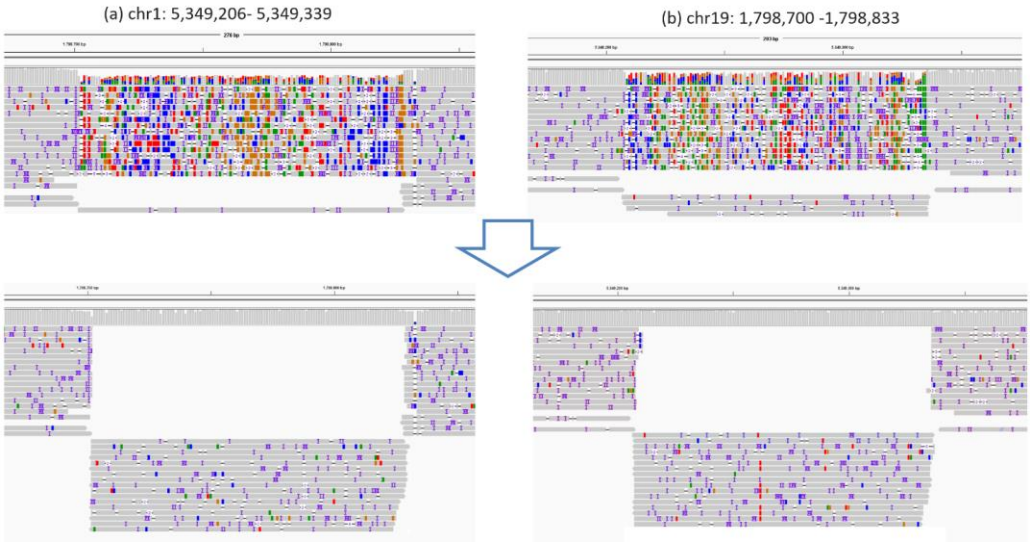

Figure 1: IGV screen capture of four random examples of suboptimal alignments and their realignment results.

Translocator was able to detect more than 70% of the embedded translocations with over 90% precision, enabling flexible and cost-efficient translocation detection using long reads. Finally, we benchmarked Translocator on real cancer cell-line datasets with PCR validated translocations and also observed outstanding performance than other existing methods.

2 Results

2.1 Performance on simulated translocations embedded into the reference genome

Translocator improves upon existing translocation detection methods by identifying suboptimally aligned reads and realigning them to achieve better translocation detection performance. We have shown in Figure 1 two possible scenarios of suboptimally aligned reads at the breakpoints of a translocation, including 1) undivided reads that are supposed to be divided, causing excessive amount of erroneous mismatch and small indel signals (noisy regions) at the translocation (Figure 1a, b), and 2) reads clipped at the breakpoints that the clipped parts should be aligned but did not, causing a sharp decrease in-depth coverage at the translocation (Figure 1c, d). Both scenarios lead to insufficient support against noise for confidently detecting a translocation. To solve the problem, Translocator scans all read alignment and identifies questionable regions with potential translocation signals. Then it retrieves the read sequences from these regions and tries to find optimal alignment for them via both local realignment and global remapping (more details are available in the Methods section).

We benchmarked Translocator against two state-of-the-art methods including Sniffles [12] and PBSV [23]. Similar to the method used for benchmarking Sniffles against other methods in Sedlazeck et al. [12], we randomly simulated 2,800 translocations (~1 translocation per 1Mbp in the human genome excluding the 'N' gaps) with length ranging from 100 to 3,000 base-pairs and embedded them at random positions of the GRCh38 reference genome (more details are available in the Methods section). Then we applied Translocator, Sniffles, and PBSV on the real PacBio or ONT sequencing reads of NA12878 [24] against the modified

		Speed (min)	TP	FN	FP	Precision	Recall	F1-score
ONT	Sniffles	108	1,349	1,451	186	87.88%	48.18%	62.24%
	PBSV	130	668	2,132	85	88.71%	23.86%	37.60%
	Translocator	226	2,450	350	192	**92.73%**	**87.50%**	**90.04%**
PacBio	Sniffles	111	2,618	182	56	**97.91%**	93.50%	95.65%
	PBSV	262	1,970	830	1,044	65.36%	70.36%	67.77%
	Translocator	169	2,752	48	76	97.31%	**98.29%**	**97.80%**

Table 1: Benchmarking results of 2,800 simulated homozygous translocations on two datasets (PacBio and ONT) and three methods (Sniffles, PBSV, and Translocator). Minimum supporting reads for calling a translocation was set as 10 as recommended by Sniffles. TP: True Positives. FN: False Negatives. FP: False Positives.

GRCh38 reference genome. The NA12878's innate translocations were distinguished and removed from our subsequent analyses (See Methods). The advantage of embedding simulated translocations into the reference genome instead of the sequencing reads is that it retains the full sequencing error profile by using real sequencing reads. The disadvantage is that only homozygous translocations can be simulated. So, in this section, we focused on establishing the baseline performance of different methods on homozygous translocations. Later in the "real cancer cell-line datasets" section, PCR validated heterozygous translocations were used for benchmarking.

We used two real datasets of NA12878, including a 43.0-fold ONT dataset [25] (rel6, Jain et al.), and a 44.2-fold PacBio dataset [26] (Mt. Sinai, Zook, et al.). For read alignment, we used NGMLR for PacBio and minimap2 for ONT, respectively. The results are shown in Table 1. Translocator outperformed Sniffles and PBSV on both the ONT and PacBio datasets. On the ONT dataset, Translocator achieved 90.04% F1-score, which is 27.80% higher than Sniffles and 52.44% higher than PBSV. While the precision of Translocator is just a few percent higher (92.73% vs. 87.88% and 88.71%), the recall has been tremendously improved (87.50% vs. 48.18% and 23.86%), confirming the power of local realignment and global remapping. On the PacBio dataset, the conclusion is similar. Translocator achieved the best F1-score (97.80% vs. 95.65% and 67.77%) and improved the recall significantly (98.29% vs. 93.50% and 70.36%).

2.2 Performance at lower depth coverages

The sequencing cost of long reads is continuously decreasing. Using ONT as an example, its massive-parallel sequencer PromethION can yield over 30-fold of a human genome on a flowcell, and the cost of a flowcell can be as low as 625 U.S. dollars [27]. However, the portable and most prevalent sequencer up to date is MinION. It costs (as low as) 475 U.S. dollars per flow cell and yields up to 30Gb (~10-fold of a human genome). In order to evaluate Translocator's robustness in different settings, we benchmarked Translocator against Sniffles and PBSV at lower depth coverages. We subsampled both the NA12878 PacBio and ONT datasets to 5-, 10-, 15-, 20- and 30-fold. We reused the 2,800 simulated translocations generated in the last section for benchmarking. For all methods, the minimum supporting reads for calling a translocation was set to one-third of the depth coverage.

The results are shown in Figure 2. Translocator consistently outperformed Sniffles and PBSV on both datasets at different depth

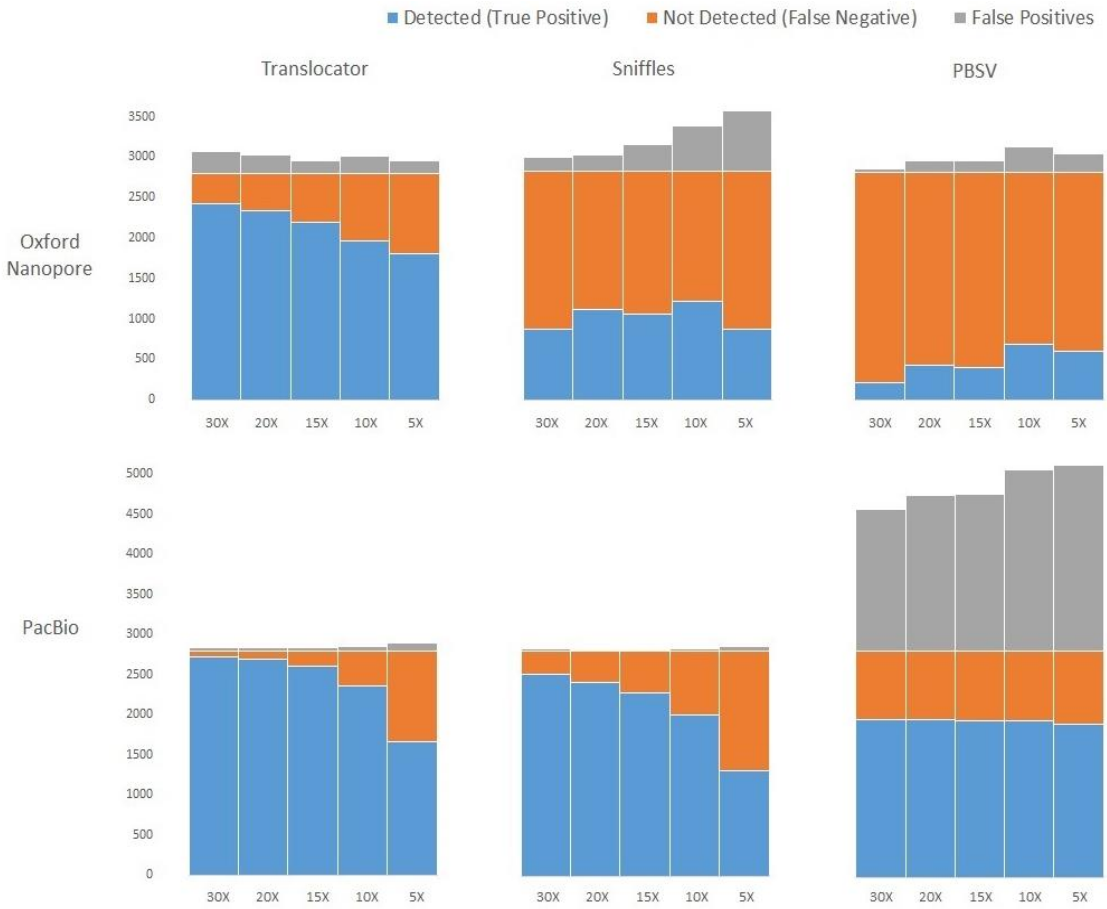

Figure 2: Benchmarking results of 2,800 simulated homozygous translocations on two datasets (PacBio and ONT) and three methods (Sniffles, PBSV, and Translocator) at five subsampled depth coverages (5-, 10-, 15-, 20- and 30-fold).

coverages. On the ONT dataset, Translocator detected over one time more translocations than the other two methods, while having lower false positive numbers than Sniffles especially at lower coverages. At 10-fold coverage, Translocator was still able to detect 70.1% of the translocations with a high precision at 90.4%. The results are in line with our expectations because due to a higher per-base error rate in ONT reads, we observed over half of the reads suboptimally aligned to the breakpoints of a translocation. At lower depth coverages, the absolute number of correctly aligned reads that can define the breakpoints of a translocation drops. Thus, the performance of Sniffles and PBSV suffered. With local realignment and global remapping, Translocator fixed the suboptimally aligned reads as much as possible, led to a significant increase in the recall rate at lower depth coverages.

On the PacBio dataset, Translocator also performed the best. Compared to PBSV, both Translocator and Sniffles controlled the number of false-positives well. Translocator has an edge over Sniffles on sensitivity, especially at lower depth coverages. At 10-fold coverage, Translocator detected 84.1% of the translocations with 97.6% precision, while Sniffle detected 71.6% of the translocations with 98.7% precision. We conclude that by using Translocator for both the ONT and PacBio datasets, 10-fold depth coverage is enough for detecting over 70% of the translocations over 90% precision.

2.3 Performance on PCR validated translocations in real cancer cell-line datasets

To further demonstrate Translocator's performance on real datasets with heterozygous translocations, we benchmarked using PCR-validated translocations in two real cancer cell-line datasets. One is a 4.5-fold ONT dataset of the HCC1187 cell-line with 17 PCR validated translocations from Gong et al. [21]. Another is a 60-fold PacBio dataset of the SK-BR-3 cell-line with 26 PCR validated translocations from Nattestad et al. [28]. The results are shown in Table 2.

While the PCR validated translocations are just a small subset of translocations in the two cancer cell lines, we have shown both the "total # of translocations detected" and "# of PCR validated translocations detected". On the 4.5-fold HCC1187 ONT dataset, although the depth coverage is low to call heterozygous translocations, Translocator consistently outperformed Sniffles and PBSV (9 vs. 8 and 5). On the 60-fold SK-BR-3 PacBio dataset, Translocator detected 23 out of 26 PCR validated translocators, while Sniffles and PBSV detected only 21 and 15 translocations, respectively. We conclude that Translocator is capable of detecting heterozygous translocations in real datasets and consistently outperformed other state-of-the-art methods.

3 Conclusion and discussions

In this study, we present Translocator, a method to detect translocations sensitively and accurately using single-molecule sequencing long reads. Translocator improves upon the existing methods by identifying suboptimally aligned reads at the breakpoints of candidate translocations, then use local realignment and global remapping to find the optimal alignment of the reads to improve the signal of supports against noise at the breakpoints. To benchmark Translocator and other methods, we analyzed real NA12878 PacBio and ONT long reads against a modified GRCh38 reference genome with 2,800 translocations of various lengths inserted at random locations. Translocator significantly outperformed other state-of-the-art methods including Sniffles and PBSV, especially on the recall rate. Benchmarks at subsampled depth coverages have shown Translocator outperformed existing methods consistently even with depth as low as 5-fold. At last, we applied Translocator to two real cancer cell-line datasets with PCR-validated heterozygous translocations, and it again outperformed other methods significantly.

We focused on translocation detection in this study, but we believe that both the local realignment and global remapping techniques are also applicable to improving the performance of detecting other classes of balanced SV, including inversion and nested SVs. Although we benchmarked heterozygous translocations using real cell-line datasets in this study, the simulated translocations remained homozygous constrained by our current method for generating a dataset with a large number of translocations with known positions and meanwhile retaining a full error profile of the sequencing reads. We look forward to devising a simulation strategy to address heterozygous translocations while fulfilling other requirements.

		Translocator	Sniffles	PBSV
HCC1187 ONT (4.5-fold)	total # of translocations detected	538	479	206
	# of PCR validated translocations detected (out of 17)	9	8	5
SKBR3 PacBio (60-fold)	total # of translocations detected	724	599	1016
	# of PCR validated translocations detected (out of 26)	23	21	15

Table 2: Benchmarking results of PCR validated translocations in two real cancer cell-line datasets.

4 Methods

4.1 An overview of Translocator

Translocator takes read alignments as input and outputs translocations together with other SVs in the VCF format. The workflow of Translocator is depicted in Figure 3. While we use Sniffles to provide basic functions for processing input and output, we also use it for calling SV classes other than translocation. Translocator finds both noisy reads and clipped reads in the alignments. The noisy reads are more likely to be suboptimally aligned. Then Translocator clusters the noisy reads into regions to be further worked at. Those noisy or clipped reads around the candidate breakpoints are realigned using local realignment. The remaining reads are globally remapped to recover their correct alignments. Finally, the translocations detected by Translocator and other classes of SV from Sniffle are combined, with the SVs overlapping with a translocation being removed.

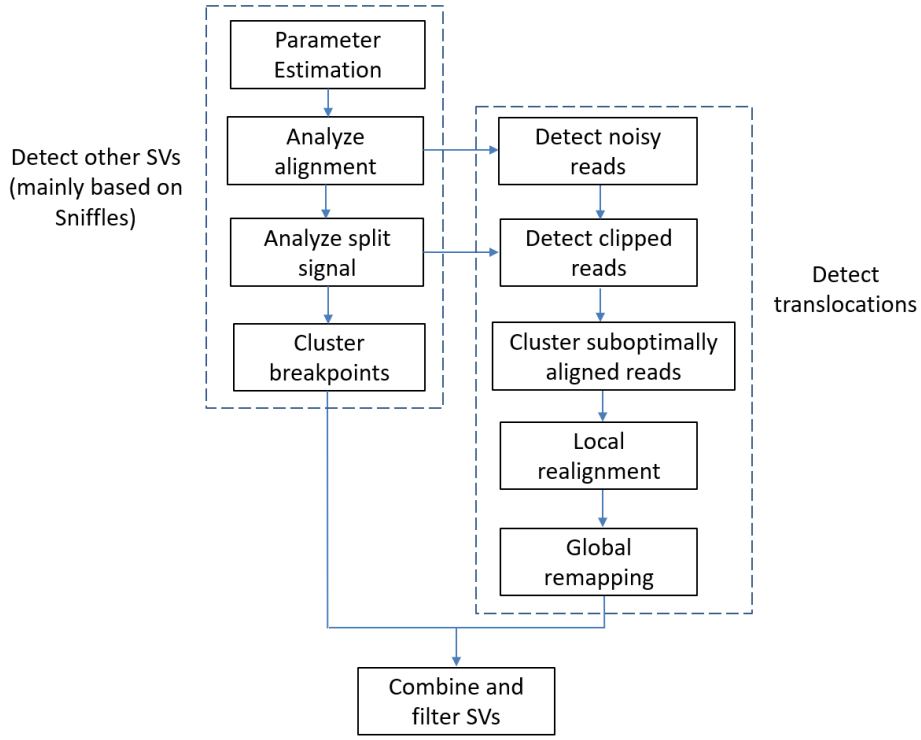

Figure 3: The workflow of Translocator.

4.2 Detecting noisy reads

To find noisy reads with blocks of excessive errors, Translocator scans through the alignment mismatches and indels extractable from the CIGAR and MD strings. A sliding window algorithm was implemented to find the blocks with excessive errors. We keep track of cumulated mismatches and small indels 100bp trailing each position. If the cumulated mismatches and small indels go beyond 50 (50% error rate), Translocator marks the block and extend it both forward and backward until the error rate goes below 30% to determine the farthest starting and ending position.

4.3 Local realignment

If the starting or ending position of a block with excess errors is close to candidate breakpoints (within 20bp), Translocator will perform local realignment of the reads in the block. These breakpoints are usually supported by one or more reads correctly split and aligned, but too few if compared against the noises. Translocator first determines the consensus of a breakpoint for realignment. Then Translocator extracts the segments of the suboptimally aligned reads in the block and aligns them to the destination using "overlap alignment" in SeqAn [29] to allow free end gaps. If the mismatch and small indel bases in the block go below 20% after realignment, Translocator considers the realignment successful and updates the alignments accordingly.

4.4 Global remapping

If no candidate breakpoint is found near a block, we remap the reads in the block globally to find their correct alignments. We found this method most helpful to the ONT datasets because almost all the reads covering translocations shorter than 500bp were found misaligned. Translocator first filters the blocks without enough read supports (default to 10). Then it maps the reads in the blocks to the reference genome again using minimap2 [19]. A successful remapping is considered having mapping quality ≥3, read coverage in a single alignment ≥80%, and clipped bases <20bp in at least an

end. After global remapping, the number of supporting reads for each candidate breakpoint is updated. Some breakpoints without enough read supports will be rescued in this process.

4.5 Combine and filter SVs

Finally, Translocator combines the translocations it called with other classes of SV called by Sniffles. The deletions and inversions called by Sniffles that overlaps with the translocations called by Translocator are regarded as negative and removed.

4.6 Generating simulated translocations for benchmarking

To best retaining the error profile of the real sequencing data, we analyzed real NA12878 reads against a modified GRCh38 human reference genome. We introduced 2,800 non-overlapping translocations into the referenced. Meanwhile, we also introduced 1,400 insertions, 1,400 deletions and 1,400 inversions as noises. We used SURVIVOR [30] for SV simulation. To remove NA12878's innate translocations from benchmarking, we identified them by analyzing the real NA12878 reads against the original GRCh38 reference. We considered an identified breakpoint within 10bp from the known position as a match.

4.7 Benchmarking PCR validated heterozygous translocations in real cancer cell-line datasets

We retrieved the SK-BR-3 PacBio reads from SRA accession number SRX4220390, and HCC1187 ONT reads (12 runs) from accession number SRP115881. For the PCR validated heterozygous translocations in the two cancer cell lines, we considered an identified breakpoint within 1,000bp from the known position as a match.

4.8 Code Availability

The source code and documentation are available at https://github.com/HKU-BAL/Translocator.

ACKNOWLEDGMENTS

This work is partially supported by the ECS (Grant No. 27204518), GRF (Grant No. 17208019) of the HKSAR government, and the ITF (Grant No. ITF/331/17FP) from the Innovation and Technology Commission, HKSAR government.

REFERENCES

[1] Weischenfeldt, J., Symmons, O., Spitz, F. and Korbel, J. O. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nature Reviews Genetics*, 14, 2 (2013/02/01 2013), 125-138.
[2] Mitelman, F., Johansson, B. and Mertens, F. The impact of translocations and gene fusions on cancer causation. *Nature Reviews Cancer*, 7, 4 (2007), 233-245.
[3] Nambiar, M. and Raghavan, S. C. How does DNA break during chromosomal translocations? *Nucleic Acids Research*, 39, 14 (2011), 5813-5825.
[4] Rabbitts, T. H. Chromosomal translocations in human cancer. *Nature*, 372, 6502 (1994/11/01 1994), 143-149.
[5] Nambiar, M., Kari, V. and Raghavan, S. C. Chromosomal translocations in cancer. *Biochimica et Biophysica Acta -Reviews on Cancer*, 1786, 2 (2008), 139-152.
[6] Aplan, P. D. Causes of oncogenic chromosomal translocation. *Trends in Genetics*, 22, 1 (2006), 46-55.
[7] Rowley, J. D. A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukaemia identified by Quinacrine Fluorescence and Giemsa Staining. *Nature*, 243, 5405 (1973/06/01 1973), 290-293.
[8] Kadalayil, L., Rafiq, S., Rose-Zerilli, M. J. J., Pengelly, R. J., Parker, H., Oscier, D., Strefford, J. C., Tapper, W. J., Gibson, J., Ennis, S. and Collins, A. Exome sequence read depth methods for identifying copy number changes. *Briefings in Bioinformatics*, 16, 3 (2014), 380-392.
[9] Alkan, C., Coe, B. P. and Eichler, E. E. Genome structural variation discovery and genotyping. *Nat Rev Genet*, 12, 5 (2011), 363-376.
[10] Layer, R. M., Chiang, C., Quinlan, A. R. and Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biology*, 15, 6 (2014/06/26 2014), R84.
[11] Dozmorov, M. G., Adrianto, I., Giles, C. B., Glass, E., Glenn, S. B., Montgomery, C., Sivils, K. L., Olson, L. E., Iwayama, T., Freeman, W. M., Lessard, C. J. and Wren, J. D. Detrimental effects of duplicate reads and low complexity regions on RNA- and ChIP-seq data. *BMC bioinformatics*, 16 Suppl 13, Suppl 13 (2015), S10-S10.
[12] Sedlazeck, F. J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A. and Schatz, M. C. Accurate detection of complex structural variations using single-molecule sequencing. *Nature methods*, 15, 6 (2018), 461–468.
[13] The long view on sequencing. *Nature Biotechnology*, 36, 4 (2018/04/01 2018), 287-287.
[14] Mahmoud, M., Gobet, N., Cruz-Dávalos, D. I., Mounier, N., Dessimoz, C. and Sedlazeck, F. J. Structural variant calling: the long and the short of it. *Genome Biology*, 20, 1 (2019/11/20 2019), 246.
[15] Weirather, J. L., de Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X.-J., Buck, D. and Au, K. F. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research*, 6 (2017), 100.
[16] Kiełbasa, S. M., Wan, R., Sato, K., Horton, P. and Frith, M. C. Adaptive seeds tame genomic sequence comparison. *Genome Res*, 21, 3 (2011), 487-493.
[17] Chaisson, M. J. and Tesler, G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics*, 13, 1 (2012/09/19 2012), 238.
[18] Sović, I., Šikić, M., Wilm, A., Fenlon, S. N., Chen, S. and Nagarajan, N. Fast and sensitive mapping of nanopore sequencing reads with GraphMap. *Nature Communications*, 7, 1 (2016/04/15 2016), 11307.
[19] Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34, 18 (2018), 3094-3100.
[20] English, A. C., Salerno, W. J. and Reid, J. G. PBHoney: identifying genomic variants via long-read discordance and interrupted mapping. *BMC Bioinformatics*, 15, 1 (2014/06/10 2014), 180.
[21] Gong, L., Wong, C.-H., Cheng, W.-C., Tjong, H., Menghi, F., Ngan, C. Y., Liu, E. T. and Wei, C.-L. Picky comprehensively detects high-resolution structural variants in nanopore long reads. *Nature methods*, 15, 6 (2018), 455–460.
[22] Cretu Stancu, M., van Roosmalen, M. J., Renkens, I., Nieboer, M. M., Middelkamp, S., de Ligt, J., Pregno, G., Giachino, D., Mandrile, G., Espejo Valle-Inclan, J., Korzelius, J., de Bruijn, E., Cuppen, E., Talkowski, M. E., Marschall, T., de Ridder, J. and Kloosterman, W. P. Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nature Communications*, 8, 1 (2017/11/06 2017), 1326.
[23] Biosciences, P. *pbsv*. City, 2019.
[24] Zook, J. M., Catoe, D., McDaniel, J., Vang, L., Spies, N., Sidow, A., Weng, Z., Liu, Y., Mason, C. E., Alexander, N., Henaff, E., McIntyre, A. B. R., Chandramohan, D., Chen, F., Jaeger, E., Moshrefi, A., Pham, K., Stedman, W., Liang, T., Saghbini, M., Dzakula, Z., Hastie, A., Cao, H., Deikus, G., Schadt, E., Sebra, R., Bashir, A., Truty, R. M., Chang, C. C., Gulbahce, N., Zhao, K., Ghosh, S., Hyland, F., Fu, Y., Chaisson, M., Xiao, C., Trow, J., Sherry, S. T., Zaranek, A. W., Ball, M., Bobe, J., Estep, P., Church, G. M., Marks, P., Kyriazopoulou-Panagiotopoulou, S., Zheng, G. X. Y., Schnall-Levin, M., Ordonez, H. S., Mudivarti, P. A., Giorda, K., Sheng, Y., Rypdal, K. B. and Salit, M. Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Scientific Data*, 3, 1 (2016/06/07 2016), 160025.
[25] Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., Tyson, J. R., Beggs, A. D., Dilthey, A. T., Fiddes, I. T., Malla, S., Marriott, H., Nieto, T., O'Grady, J., Olsen, H. E., Pedersen, B. S., Rhie, A., Richardson, H., Quinlan, A. R., Snutch, T. P., Tee, L., Paten, B., Phillippy, A. M., Simpson, J. T., Loman, N. J. and Loose, M. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*, 36, 4 (2018/04/01 2018), 338-345.
[26] Zook, J. M., Chapman, B., Wang, J., Mittelman, D., Hofmann, O., Hide, W. and Salit, M. Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nature Biotechnology*, 32, 3 (2014/03/01 2014), 246-251.
[27] Technologies, O. N. *Products Overview*. City, 2019.
[28] Nattestad, M., Goodwin, S., Ng, K., Baslan, T., Sedlazeck, F. J., Rescheneder, P., Garvin, T., Fang, H., Gurtowski, J. and Hutton, E. J. G. r. Complex rearrangements and oncogene amplifications revealed by long-read DNA and RNA sequencing of a breast cancer cell line, 28, 8 (2018), 1126-1135.
[29] Döring, A., Weese, D., Rausch, T. and Reinert, K. SeqAn An efficient, generic C++ library for sequence analysis. *BMC Bioinformatics*, 9, 1 (2008/01/09 2008), 11.
[30] Jeffares, D. C., Jolly, C., Hoti, M., Speed, D., Shaw, L., Rallis, C., Balloux, F., Dessimoz, C., Bähler, J. and Sedlazeck, F. J. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nature Communications*, 8, 1 (2017/01/24 2017), 14061.