

Structural Variation Detection of Adolescent Thyroid Cancer Using Optical Mapping

Lei yang

Department of Otolaryngology, Head and Neck Surgery, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health

Shaochang Wu

Department of Geriatrics, Lishui Second People's Hospital

Baosheng Liao

Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences

Shuai Guo

Pharmacy College, Chengdu University of Traditional Chinese Medicine

Shengcai Wang

Department of Otolaryngology, Head and Neck Surgery, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health

Yuwei Liu

Department of Otolaryngology, Head and Neck Surgery, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health

Jialin Fan

Department of Geriatrics, Lishui Second People's Hospital

Jun Tai

Department of Otolaryngology, Head and Neck Surgery, Children's Hospital Capital Institute of Pediatrics

Jun Chen (✉ bchebh@163.com)

Beijing Engineering Research Center of Pediatric Surgery, Engineering and Transformation Center, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health

Xin Ni

Department of Otolaryngology, Head and Neck Surgery, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health

Jiang Xu

Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences

Research Article

Keywords: Structural Variation (SV), adolescent thyroid cancer, Bionano optical mapping, Oxford Nanopore Technology (ONT), Next Generation Sequencing (NGS)

Posted Date: January 31st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1172959/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Structural variation (SV) is a fundamental genetic cause for cancer, with demonstrated correlation to disease progression and treatment response. Traditional sequencing method cannot provide full genomic landscape especially large-scale and complex structural variation. To overcome these limitations, we adopted a combined sequencing approaches, including optical mapping, single molecular sequencing and short reads shotgun sequencing, to evaluate the SV in thyroid cancer. Different numbers, length and types of structural variation, with genes affected by SV were scrutinized. Integrating these results could showed comprehensive scenario for thyroid cancer in a genomic view. We demonstrate that integrated approaches could provide a powerful tool for capturing a higher level of genomic SV, creating new interpretation of sequencing data of particular relevance to human cancer.

Introduction

Genomic variation, ranging from point mutation, deletion, duplication, insertion, inversion, translocation and complex structural variation (SV)[1], associated with poor prognosis in tumor [2]. Specific genomic SV conducted functional gene alteration, which affected essential signaling pathway of tumor biology of initial differential and progression [3, 4]. Traditional tumor genomic research preferred second/next generation sequence (NGS) technology, only detecting small sequence genomic variation including gene mutation, copy number alteration, methylation, and transcription using short reads. However, accompany with tumor progression, large genomic SV or rearrangements requires superior genomic detection methods to better comprehensively and accurately identify the specific changes in the tumor genomics, and help us better understand the characteristics of tumor, but there are still challenges.

We dedicated in researching the characteristic of thyroid cancer genome. Many studies have revealed the gene variation, biological signaling pathway, and relationship between molecular classification and clinical traits in papillary thyroid carcinoma (PTC). On the contract, large genomic SV and adolescent PTC studies were not clear, as the adolescent genome variation were more complicated than adults[5, 6]. Therefore, we need a thorough and comprehensive understanding of tumor genomic variation and biology. So, we integrated multi-sequence technology, NGS, Oxford Nanopore Technology (ONT) Sequencing and Bionano Optical Mapping, to inspect and complement the adolescent PTC genome characteristic.

Results

Genomic Structural Variation of thyroid cancer

One adolescent PTC sample (from a 15 years old female) was used in this test. General speaking, we conducted four sequencing technologies to reveal the characteristic genome structural variation, including Illumina NGS, ONT Sequencing, Bionano Genomics Nt.BspQI nickase endonuclease and Direct Labeling Enzyme (DLE) Optical Mapping. Sequencing blood was used as control. After initial tumor

sample sequencing and bioinformatic analysis, we detected five different kind of SV, including deletion, duplication, insertion, inversion and translocation, and the total numbers of SV within Bionano DLE, BspQI, ONT and NGS were 243, 127, 3120 and 2357 respectively (Table 1).

After scrutinizing the detecting SV numbers and length, huge differences among sequencing technologies were shown (Figure 1). ONT and NGS were obviously advanced in total SV detecting number, where the number of deletion and insertion were dominant within other SV types (Figure1-A). As for the length distribution of different SV types, Bionano sequencing DLE and BspQI enzyme were advanced in short-read SV detecting, but defect in total SV number (Figure1-B/C). ONT detecting SV length were concentrated in 0-1000bp, where deletion and insertion number were 1361/1631, however the other types of SV were relative less, which means the ONT was good at detecting short-read deletion and insertion (Figure1-D). NGS detecting processed a huge number of SV and distributed in all length, which could be a complement for long-read single molecular sequencing (Figure1-E).

Then, we integrated the SV of deletion, insertion and inversion of four sequencing method into one genomic circle plot (Figure 2). Moving inward from the outer the four ideograms are histograms of deletions, the median are insertions, and the inner are inversion in four detection method (NGS, ONT, BspQI, DLE for black, brown, purple, red). The SV disperse within all chromosome, and different sequencing method are inconsistent in detecting SV effects, which demonstrate that these sequencing methods are complementary in cancer genomic detection.

Genes affected by genomic SV

Large scale SV affected a wide region within genome, especially for the numerous essential genes. All genes located within the detecting tumor genomic SV region were screened. The total number of SV genes in Bionano DLE, BspQI, ONT and NGS were 1218, 454, 1219, and 1112 respectively (Table 2). Next, we proceeded preliminary screening and summarization for the different types of SV genes within different sequencing technologies, then took intersection of these SV genes lists (Figure 3). For the number of SV genes within deletion, duplication, and insertion affected, ONT and NGS sequencing were much than Bionano DLE and BspQI enzyme. For inversion, Bionano DLE and BspQI were much more than ONT and NGS. For translocation, Bionano DLE, BspQI and ONT technologies were advanced than NGS, indicating that the single molecular reads processed perceptible superiority in detection the affected genes of large-scale and complex genomic structural variations.

Gene enrichment functional analysis

The tumor occurrence and progression often accompanied with a large number of genomic SV, affected many key genes alteration [2, 3]. Each SV affected gene was manually filtered by fully reviewing known functions and literatures related with cancer, and 585 key tumor genes affected by SV were identified (supplementary file). Using these genes to proceed GO/KEGG analysis, the result showed that cancer related metabolism and pathway were significantly enriched, and some cell component, molecular function and biological process were elucidated which may be essential in tumor proliferation and

progression, such as gland development, DNA-binding transcription activator activity, DNA repair complex, P13K-Akt, Ras, Rap1 signaling pathway (Figure 4).

Discussion

At present, traditional tumor genomic variation research were mostly devoted in single nucleotide polymorphism, copy number variation, DNA methylation, mRNA, lncRNA, etc. However, large-scale variations and complex chromosomal rearrangement (like generated by double strands break) were common and essential in tumor differential and progression mechanisms [3, 4], and the short-read variation cannot describe genomic or chromosome alteration completely. Therefore, superior long-read or optical mapping genome detection methods were required for accurate identification of tumor-specific genetic variation with a better and comprehensive perspective. Here, we have integrated Bionano optical sequencing and ONT single-molecule sequencing, by extending the read length of sequenced DNA molecules, to comprehensively detect structural variations of tumor genomes, compared with traditional NGS technology.

Genomic Structural Variation in thyroid cancer

The onset and progression of cancer are often triggered by genome-wide accumulated structural abnormalities, especially by dysregulated oncogenes variation. Many genomic variations have been found in thyroid cancer, including point mutation, SV, gene fusion and other complex genomic variation [7-9], which were reported to be correlated with genomic instability and patient clinical traits [10-13]. BRAF^{V600E} and RAS gene mutation was the main mutation type in thyroid cancer, accompany with activity of MAPK signal pathway through different metabolism [7]. By summarizing the published research report so far, 95 papillary thyroid cancer mutant genes were identified (supplementary file). Common mutation genes in PTC include BRAF, RAS, EIF1AX, and rare mutations include PP1MD, CHEK2, TSHR, etc. Gene fusion was a diverse and common genomic variation in PTC, such as RET fusion which was related with genetic function abnormality and alteration in tumor progressive signaling pathway [7, 14]. Copy Number Alteration can be also detected in PTC, and approximately 10% occurring at 22q chromosome, 15% occurring at chromosome 1q, 2% owning a high frequency of focal CNV [7]. Adolescent thyroid cancer genomic structural variation was more complex than adults [5, 6], including several uncertain gene mutation or fusion and other complex chromosome rearrangement [5]. What's more, some specific family hereditary syndrome like Gardner, Cowden, DICER1, Werner and PPNAD syndrome, causing by certain gene mutation, could increase the risk of thyroid cancer incidence and progression [15-18]. The complex genomic structural variation and key oncogene could be better detected through the advanced sequencing technology [19], which can lead to improved diagnostics and therapies [20]. Through the Bionano optical sequencing, ONT single-molecule sequencing, and NGS technology, we got the adolescent thyroid cancer SV.

Discrepancy among Several Sequencing Technologies

Traditional Next Generation Sequencing processed advancement of mature technology, high quality and low cost. Nucleotide sequence were quantitatively analyzed through short-read DNA synthetic detection technology, which limited in detection of repeated sequence[21]. Nowadays, NGS technology can successfully identify different kind of variations through the bioinformatics technology update progress.

Bionano optical mapping was designed with optical fluorescent labeling technology which shearing DNA into a special nanochip for detecting. Scanning the optical signal, genome motives were directly measured through sheared high-quality molecular DNA. Bionano optical mapping was more precise than traditional fluorescent technologies like FISH, CNV array, karyotyping, in both simple and complex SV[22], which can be a novel clinical genetic diagnostic method in future [23, 24]. Bionano optical mapping reveals a wide spectrum of SVs through visualizing DNA structure, which can be a good complement to short-read sequencing[25]. Since points is limited to label density, more enzymes like DLE-1 or BspQI are needed to improve the performance of SVs characterization.

Recent advances in single-molecule sequencing are showing promising utility in detecting comprehensive structural variation analysis, particularly for duplication, deletion, insertion, inversion, and translocation[26]. Owned for long-read-based, the Pacific Biosciences and Oxford Nanopore Technologies are capable of reconstructing the entire human genome, which is impossible by short-read sequencing. The ONT single molecular sequencing technology was designed using nanopore protein motor, passing through exactly high-quality long read DNA and detecting electricity changes to analyzed nucleotide sequence. Due to the long-read DNA detection, the continuity and integrity was preferable to genomic de-novo assembly application. However, due to the sequencing technology itself, the error rates were 15% higher mismatch rate than NGS, and haplotype genome algorithm is weak in detection complex genomic structural variation [27], and it was either too expensive to obtain sufficient coverage, or suffered from poor accuracy[28]. However, combining the advantages of different sequencing technologies, the accuracy of SV detection will be greatly improved.

The mismatching for SV and affected genes

Our result showed that the number of genomic SV and affected genes were different between sequencing technologies. For Bionano DLE enzyme sequencing technology, the total SV number were 243, however, the total SV affected genes were 1218. Based on Bionano optical mapping, each SV segment contained more genes, which advanced in detecting extension SV affected genes. For ONT and NGS sequencing technology, the total SV number were more than 2000, however, the total SV affected genes were approximately 1000. Although large SV number were detected by ONT and NGS, the accuracy detecting genes were approximately equivalent within different technologies. The Bionano BspQI enzyme technology in detection SV genes were significantly less accuracy than other three method.

Limitation

There still some limitation in our study. There is only one thyroid cancer samples for genomic SV detection with several technologies, and validation still needed. For the sequencing thyroid cancer

sample, although tissue have been validated by pathology group, the cell heterogeneity still existed. Stromal cell and tumor cell DNA mixture could affect accuracy of tumor specific genomic SV. To comprehensive and precise detection genomic alteration in future, single cell sequencing technology can be combined with advanced sequencing technology in tumor genomic research.

Conclusion

We used four sequencing technologies to explore the genomic SV of this adolescent thyroid cancer. Different numbers, length and types of structural variation, with genes affected by SV were scrutinized. Integrating these results could showed comprehensive scenario for thyroid cancer in a genomic view. We demonstrated that integrated approaches could provide a powerful tool for capturing a higher level of genomic SV, creating new interpretation of sequencing data of particular relevance to human cancer.

Declarations

Ethical compliance

All methods were performed in accordance with the relevant guidelines and regulations based on declaration of Helsinki, and the written informed consent to attend this study was obtained from the parents of the subject prior to inclusion in the study, as approved by the Ethics Committee of Beijing Children's Hospital, Capital Medical University, National Center for Children's Health.

Consent for publication: Not Applicable.

Availability of data and materials: The datasets generated and analyzed during the current study are not publicly available due to privacy and ethical restrictions of the current teenager thyroid cancer patient, but are available from the corresponding author on reasonable request.

Competing interests: The authors declare no competing financial or non-financial interests.

Funding: Not Applicable

Authors' contributions: Lei yang wrote the main manuscript text. Shaochang Wu, Baosheng Liao, Shuai Guo made the initial experiment and validation. Shengcai Wang, Yuwei Liu, Jialin Fan, Jun Tai made investigation and quality check. Jun Chen, Xin Ni, Jiang Xu reviewed and supervised this research.

Acknowledgement: Thanks for all authors' contribution.

Methods

Briefly, we took four method, Bionano-DLE, Bionano-BspQI enzyme, Oxford Nanopore Technology, and Next Generation Sequence, to detect adolescent thyroid tumor genome SV using the paired blood sample as normal control.

Sample collection and Ethics

We collected thyroid tumor tissues and paired blood samples from one fifteen years old papillary thyroid cancer. Briefly, fresh tissue samples were collected at the time of resection of the abnormal thyroid mass. Samples were confirmed by H&E histology section content. The paired peripheral blood samples were drawn from the same patient for sequencing as control. Written informed consent to attend this study was obtained from the parents of the subject prior to inclusion in the study, as approved by the Ethics Committee of Beijing Children's Hospital, Capital Medical University, National Center for Children's Health.

Bionano optical mapping with DLE and BspQI enzyme:

The tumor sample and blood were sequenced with Bionano-DLE and Bionano-BspQI enzyme strategies. Briefly, High molecular weight (HMW) DNA was extracted using the Bionano Kits for the sample preparation. Then, the DNA was labeled by DLE and BspQI enzyme in two strategies. The labeled DNA was loaded on Bionano chips, and automatically running with Bionano system. Genomic mapping and assembling was performed with Bionano Access and Solve software. SV calling was performed with Bionano algorithm aligning hg38 genome and detecting genomic structural variation after filtering common SV and choosing the tumor specific SV. More specific procedures and parameters were in supplementary file method.

Oxford Nanopore Technology sequencing:

The tumor sample and blood were also sequenced using Oxford Nanopore Technology (ONT). The original sample treatment and HMW DNA were in accordance with Bionano process. Next, the library preparation was used ONT protocol suggestion, following MinION sequencing and software basecalling. After trimming, the reads were mapped with software NGMLR using reference Hg38 genome, and then genomic structural variations were called with software Sniffles and filtered with only precise SV which presented in tumor while not in blood sample. More specific procedures and parameters were in supplementary file method.

Next-Generation Sequencing:

The tumor sample and blood were also sequenced using Next-Generation Sequencing (NGS). The DNA was extracted from sample tissues and fragmented into small size for the library preparation followed the NGS DNA preparation protocols. After PCR amplification, the enriched DNA library was sequenced by HiSeq, and common bioinformatic procedures of mapping, alignment and SV calling was performed with Bcl2fastq, BWA and Delly softwares. After trimming the SVs lists excluding normal variation, the final specific thyroid tumor SVs were determined by those presented in tumor but not in blood sample. More specific procedures and parameters were in supplementary file method.

SV affected genes identification and classification:

All SV from four sequencing method were summarized and depicted different SV type number and length distribution with Origin software. Then, different types of SV (deletion, insertion, inversion) with different sequencing method were marked in one circle genome plot. Next, to explore the gene alteration affected with genomic structural variation, we collected the genes located within the genomic structural variation region. Different types SV affected genes were classified and took intersection to show the overlapped genes with Venn plot. Besides, each gene was manually scrutinized the relationship with tumor proliferation and progression and filtered the oncogenes affected by SV in thyroid cancer. The filtered gene list were performed enrichment analysis with Gene Ontology and KEGG pathway analysis, showed by four bubble plots.

Table 1: SV Number in four sequencing method						
	DEL	DUP	INS	INV	TRA	Total
DLE	64	4	109	58	8	243
BspQI	34	1	55	26	11	127
ONT	1361	115	1631	7	6	3120
NGS	588	112	1548	108	1	2357
DEL=Deletion; DUP=Duplication; INS=Insertion; INV=Inversion; TRA=Translocation						

Table 2: Different type SV genes number						
	DEL	DUP	INS	INV	TRA	Total
DLE	81	4	158	972	3	1218
BspQI	9	5	9	422	9	454
ONT	528	60	608	3	20	1219
NGS	256	79	606	171	0	1112
DEL=Deletion; DUP=Duplication; INS=Insertion; INV=Inversion; TRA=Translocation						

References

- [1] T. Lappalainen, A. J. Scott, M. Brandt, and I. M. Hall, "Genomic Analysis in the Age of Human Genome Sequencing," (in eng), *Cell*, vol. 177, no. 1, pp. 70-84, 2019.
- [2] N. Andor, C. C. Maley, and H. P. Ji, "Genomic Instability in Cancer: Teetering on the Limit of Tolerance," (in eng), *Cancer Res*, vol. 77, no. 9, pp. 2179-2185, May 1 2017.

- [3] J. W. Yun *et al.*, "Dysregulation of cancer genes by recurrent intergenic fusions," (in eng), *Genome biology*, vol. 21, no. 1, p. 166, 2020.
- [4] M. Dzamba *et al.*, "Identification of complex genomic rearrangements in cancers using CouGaR," (in eng), *Genome research*, vol. 27, no. 1, pp. 107-117, 2017.
- [5] M. I. C. V. Cordioli *et al.*, "Fusion Oncogenes Are the Main Genetic Events Found in Sporadic Papillary Thyroid Carcinomas from Children," (in eng), *Thyroid : official journal of the American Thyroid Association*, vol. 27, no. 2, pp. 182-188, 2017.
- [6] K. Iyama *et al.*, "Identification of Three Novel Fusion Oncogenes, SQSTM1/NTRK3, AFAP1L2/RET, and PPFIBP2/RET, in Thyroid Cancers of Young Patients in Fukushima," (in eng), *Thyroid : official journal of the American Thyroid Association*, vol. 27, no. 6, pp. 811-818, 2017.
- [7] N. Cancer Genome Atlas Research, "Integrated genomic characterization of papillary thyroid carcinoma," (in eng), *Cell*, vol. 159, no. 3, pp. 676-90, Oct 23 2014.
- [8] L. Chai, J. Li, and Z. Lv, "An integrated analysis of cancer genes in thyroid cancer," (in eng), *Oncology reports*, vol. 35, no. 2, pp. 962-970, 2016.
- [9] J. Gudmundsson *et al.*, "A genome-wide association study yields five novel thyroid cancer risk loci," (in eng), *Nature communications*, vol. 8, p. 14517, 2017.
- [10] J. C. Rubinstein *et al.*, "Shifting patterns of genomic variation in the somatic evolution of papillary thyroid carcinoma," (in eng), *BMC cancer*, vol. 16, p. 646, 2016.
- [11] H. He *et al.*, "Allelic variation in gene expression in thyroid tissue," (in eng), *Thyroid : official journal of the American Thyroid Association*, vol. 15, no. 7, pp. 660-667, 2005.
- [12] J. Gudmundsson *et al.*, "Discovery of common variants associated with low TSH levels and thyroid cancer risk," (in eng), *Nature genetics*, vol. 44, no. 3, pp. 319-322, 2012.
- [13] C. Campo *et al.*, "Inherited variants in genes somatically mutated in thyroid cancer," (in eng), *PloS one*, vol. 12, no. 4, p. e0174995, 2017.
- [14] M. Santoro, M. Moccia, G. Federico, and F. Carlomagno, "RET Gene Fusions in Malignancies of the Thyroid and Other Tissues," (in eng), *Genes*, vol. 11, no. 4, 2020.
- [15] S. Uchino *et al.*, "Age- and Gender-Specific Risk of Thyroid Cancer in Patients With Familial Adenomatous Polyposis," (in eng), *The Journal of clinical endocrinology and metabolism*, vol. 101, no. 12, pp. 4611-4617, 2016.
- [16] J. E. Hall, D. J. Abdollahian, and R. J. Sinard, "Thyroid disease associated with Cowden syndrome: A meta-analysis," (in eng), *Head & neck*, vol. 35, no. 8, pp. 1189-1194, 2013.

- [17] E. Bonora, G. Tallini, and G. Romeo, "Genetic Predisposition to Familial Nonmedullary Thyroid Cancer: An Update of Molecular Findings and State-of-the-Art Studies," (in eng), *Journal of oncology*, vol. 2010, p. 385206, 2010.
- [18] D. R. Stewart *et al.*, "Neoplasm Risk Among Individuals With a Pathogenic Germline Variant in DICER1," (in eng), *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 37, no. 8, pp. 668-676, 2019.
- [19] M. Chen *et al.*, "Identification of a likely pathogenic structural variation in the LAMA1 gene by Bionano optical mapping," *npj Genomic Medicine*, vol. 5, no. 1, p. 31, 2020/08/12 2020.
- [20] P. Vanden Borre *et al.*, "Pediatric, Adolescent, and Young Adult Thyroid Carcinoma Harbors Frequent and Diverse Targetable Genomic Alterations, Including Kinase Fusions," (in eng), *The oncologist*, vol. 22, no. 3, pp. 255-263, 2017.
- [21] M. J. P. Chaisson, R. K. Wilson, and E. E. Eichler, "Genetic variation and the de novo assembly of human genomes," (in eng), *Nature reviews. Genetics*, vol. 16, no. 11, pp. 627-640, 2015.
- [22] H. Wang *et al.*, "Analysis of balanced reciprocal translocations in patients with subfertility using single-molecule optical mapping," *Journal of Assisted Reproduction and Genetics*, vol. 37, no. 3, pp. 509-516, 2020/03/01 2020.
- [23] H. Barseghyan *et al.*, "Next-generation mapping: a novel approach for detection of pathogenic structural variants with a potential utility in clinical diagnosis," (in eng), *Genome medicine*, vol. 9, no. 1, p. 90, 2017.
- [24] K. Neveling *et al.*, "Next generation cytogenetics: comprehensive assessment of 48 leukemia genomes by genome imaging," *bioRxiv*, p. 2020.02.06.935742, 2020.
- [25] M. Pendleton *et al.*, "Assembly and diploid architecture of an individual human genome via single-molecule technologies," *Nat Methods*, vol. 12, no. 8, pp. 780-6, Aug 2015.
- [26] M. J. Chaisson *et al.*, "Resolving the complexity of the human genome using single-molecule sequencing," *Nature*, vol. 517, no. 7536, pp. 608-11, Jan 29 2015.
- [27] M. Pendleton *et al.*, "Assembly and diploid architecture of an individual human genome via single-molecule technologies," (in eng), *Nature methods*, vol. 12, no. 8, pp. 780-786, 2015.
- [28] F. J. Sedlazeck *et al.*, "Accurate detection of complex structural variations using single-molecule sequencing," *Nat Methods*, vol. 15, no. 6, pp. 461-468, Jun 2018.

Figures

Figure 1

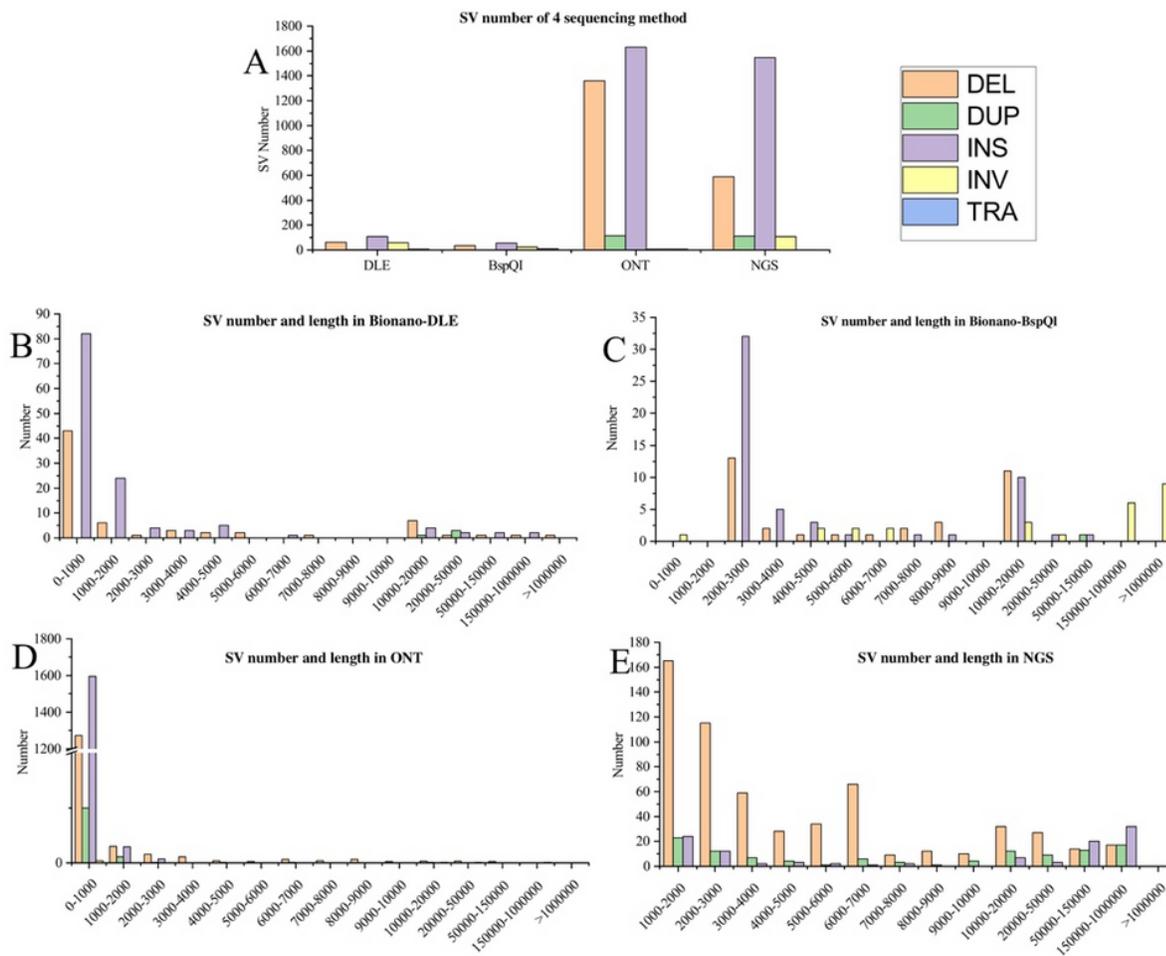


Figure 1

SV numbers and length among four sequencing technologies, Bionano-DLE, BspQI, ONT and NGS.

Figure 2

Circle plot of the SV of deletion, insertion and inversion within four method. Moving inward from the outer the four ideograms are histograms of deletions, the median are insertions, and the inner are inversion in four detection method (NGS, ONT, BspQI, DLE for black, brown, purple, red).

Figure 3

Venn plot of different types of SV genes lists within four sequencing technologies.

Figure 4

GO/KEGG analysis of manually filtered SV affected key genes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfilemethod.doc](#)
- [Supplementaryfile.xlsx](#)