PacBio HiFi 质控，比对和SVs变异检测方面的相关软件

- 1. 质控软件主要有：
  - FastQC：适用于快速检查碱基质量分布、GC含量、序列长度分布等。
  - NanoPlot：专门用于长读长数据，能够生成详细的读长分布、质量分布等可视化图表。最近更新2023.10.29
  - Nanofilt: 长读取测序数据的过滤和修整。最近更新 2022.12.21
  - chopper： Rust implementation of NanoFilt+NanoLyse, both originally written in Python. This tool, intended for long read sequencing such as PacBio or ONT, filters and trims a fastq file. 最近更新 2024.08.20
- 2. 比对软件主要有：
  - minimap2：最近更新2024.03.27
  - NGMLR: 文章于2018年发表于Nature Methods，最近更新2018.06.25
  - PBmm2：是PB官方基于minimap2进行优化的版本，最近更新2024.10.30
  - Winnowmap：一个基于minimap2改进的映射工具，它采用了独特的优化策略，尤其是在处理复杂重复区域时，如人类染色体中的长串联重复序列。最近更新2021.05.08
- 3. SVs变异检测软件主要有：
  - PBSV：PacBio官方开发的结构变异软件，最近更新2024.10.29
  - Sniffles2，版本一于2018年发表于Nature methods，最近更新2024.12.17
  - cuteSV2，文章于2020年发表于Genome Biology，最近更新2024.05.10
  - PAV，Phased Assembly Variant Caller，文章于2021年发表于Science，最近更新2024.12.18
  - DeBreak，文章于2023年发表在Nature Communication上，最近更新2022.12.01
  - SVIM，文章于2019年发表于Bioinformatics，最近更新2021.06.18
  - SVision, 文章于2022年发表于Nature Methods，最近更新2024.05.08

以下是两篇软件评估文章

- 评估文章：
- 第一篇文章于2024年发表在Genome Biology上, 文章强调Minimap2-cuteSV2、NGMLR-SVIM、PBMM2-pbsv、Winnowmap-Sniffles2 和 Winnowmap-SVision 等管道表现出相对较高的召回率和精度。
- 第二篇文章于2024年发表在Nature Communications上, 文章评论minimap2和winnowmap比对上优于ngmlr，但未对pbmm2进行评估，在pacbio hifi上，pbsv2整体表现较好

相关文章

2024 GB 中山大学李淼新团队 Comprehensive and deep evaluation of structural variation detection pipelines with third-generation sequencing data

- (https://genomebiology.biomedcentral.com/articles/10.1186/s13059-024-03324-5)
- 文章结果： 值得注意的是，Minimap2-cuteSV2、NGMLR-SVIM、PBMM2-pbsv、Winnowmap-Sniffles2 和 Winnowmap-SVision 等管道表现出相对较高的召回率和精度。

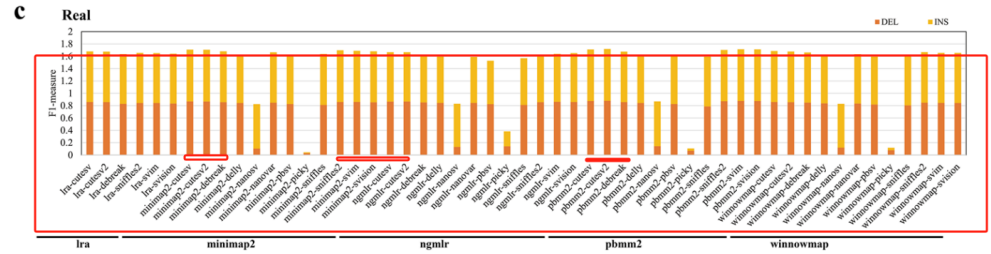


Fig. 1 Performance of SV detection pipelines in different SV types (CCS). Precision and recall of DEL, DUP, INS, INV, and BND were determined with the simulated (**a**, **b** (DUP_INS)) and the real data (**c**). F1 measures, which combine precision and recall statistics (see the "Methods" section for details), are depicted for pipelines distinguished by different colored bars. Pipelines are categorized according to the alignment tools (lra, minimap2, ngmlr, pbmm2, winnowmap)

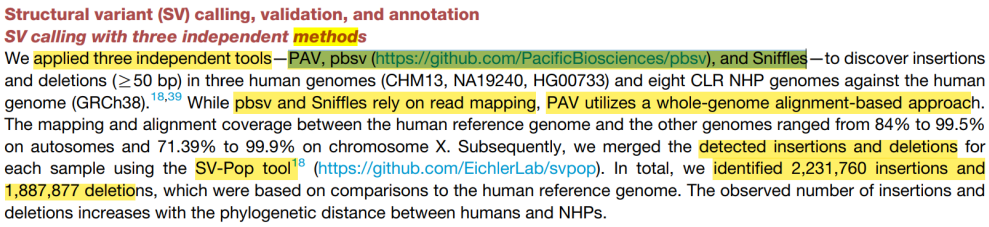
Results: This study comprehensively evaluates 53 SV detection pipelines using simulated and real data from PacBio (CLR: Continuous Long Read, CCS: Circular Consensus Sequencing) and Nanopore (ONT) platforms. We assess their performance in detecting various sizes and types of SVs, breakpoint biases, and genotyping accuracy with various sequencing depths. Notably, pipelines such as Minimap2-cuteSV2, NGMLR-SVIM, PBMM2-pbsv, Winnowmap-Sniffles2, and Winnowmap-SVision exhibit comparatively higher recall and precision. Our findings also show that combining multiple pipelines with the same aligner, like pbmm2 or winnowmap, can significantly enhance performance. The individual pipelines' detailed ranking and performance metrics can be viewed in a dynamic table: http://pmglab.top/SVPipelinesRanking.

2024 NC Tradeoffs in alignment and assembly-based methods for structural variant detection with long-read sequencing data

- https://doi.org/10.1038/s41467-024-46614-z
- 文章首先系统地分析和评估了 2014 年至 2022 年间在大量 PacBio 和 ONT 数据集上引入的 16 种最先进的长读长 SV 方法的性能。
- 不同比对工具：
  - a） 不同的对准器对基于对齐的工具的插入召回率有显着影响。一般来说，对于大多数基于比对的工具，minimap2、Winnowmap 和 LRA 在插入召回方面优于 NGMLR。
  - b） 插入精度、删除召回率和精度不受影响，或者仅受不同对准器的细微影响
- 不同的SV caller：
  - a） 对于易位，在模拟数据上，总体最佳工具是 cuteSV 和 pbsv（在所有合适的数据集上 F1 > 0.95），其次是 NanoSV 和 SVIM。
  - b） 对于倒置，没有工具达到高性能。NanoVar、Sniffles2 和 NanoVar 分别在 Hifi、CLR 和 ONT 上获得了最好的 F1 分数。
  - c） 对于重复，pbsv 是 Hifi 和 CLR 数据的最佳工具，其次是 DeBreak，而 NanoVar 是 ONT 数据的最佳工具。

对pacbio hifi测序数据进行SV检测的文章：

2024 Cell 上海交通大学毛亚飞 Structurally divergent and recurrently mutated regions of primate genomes

- 文章中通过minimap2比对，SVs检测是PAV, pbsv, and Sniffles进行检测，然后再通过SV-pop tool进行合并SVs
- https://doi.org/10.1016/j.cell.2024.01.052
- 

Structural variant (SV) calling, validation, and annotation

SV calling with three independent methods

We applied three independent tools—PAV, pbsv (https://github.com/PacificBiosciences/pbsv), and Sniffles—to discover insertions and deletions (≥50 bp) in three human genomes (CHM13, NA19240, HG00733) and eight CLR NHP genomes against the human genome (GRCh38).[18,39] While pbsv and Sniffles rely on read mapping, PAV utilizes a whole-genome alignment-based approach. The mapping and alignment coverage between the human reference genome and the other genomes ranged from 84% to 99.5% on autosomes and 71.39% to 99.9% on chromosome X. Subsequently, we merged the detected insertions and deletions for each sample using the SV-Pop tool[18] (https://github.com/EichlerLab/svpop). In total, we identified 2,231,760 insertions and 1,887,877 deletions, which were based on comparisons to the human reference genome. The observed number of insertions and deletions increases with the phylogenetic distance between humans and NHPs.

- 2024 NC 泛基因组图改进了对罕见遗传病结构变异的分析： 做的泛基因组，流程是minigraph + pbsv
- 2023 GB 利用HiFi长阅读测序进行全面的从头突变发现：分析流程是：pbmm2 + pbsv
- 2023 GM 利用HiFi长阅读测序进行全面的从头突变发现 分析流程是：pbmm2 + pbsv