Benchmarking Reports

```
SV_STAT command:
$ ./sv_stat -m 50000 -T
ASVCLR;SVDSS;DeBreak;Sniffles2;pbsv;cuteSV;SVIM -C
1;2;3;4;5;6;7;8;9;10;11;12;13;14;15;16;17;18;19;20;21;22;X;Y
./Data/output_ASVCLR.vcf ./Data/output_SVDSS.vcf
./Data/output_DeBreak.vcf ./Data/output_Sniffles2.vcf
./Data/output_pbsv.vcf ./Data/output_cuteSV.vcf
./Data/output_SVIM.vcf ./Data/HG002_SVs_Tier1_v0.6.vcf
./Data/hs37d5.fa -o HG002_CCS_30X
```

1. Benchmarking results

Variant type match mode: **loose (allow type match between DUPLICTION and INSERTION)**

The benchmarking metrics has two categories after filtering long SV regions: one category is used to highlight performance by metrics including Recall, Precision, F1 score, and sequence identity (Identity) and the other category presents benchmark results, which consists of TP_bench, TP_user, FP, FN. Visualizing these metrics through bar charts provides a more intuitive representation of the benchmarking results for the variation detection methods.

(1) The benchmarking results of the user-called set are as follows:

Table 1 Structural Variation Detection Method Performance Benchmarking

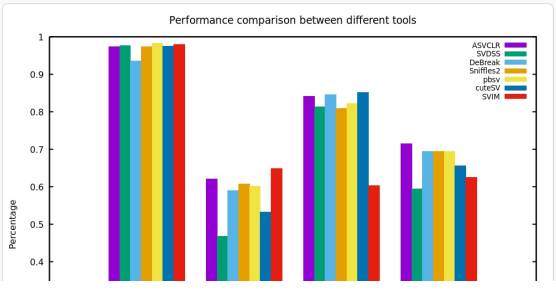
Tool	#SVs_bench	#SVs_user	#SVs_filtered_user	#TP_bench	#TP_user	#FP	#FN	Recall	Precision	F1 score	Identity
ASVCLR	74012	54423	54423	45986	45807	8616	28026	0.973949	0.621332	0.841685	0.714914
SVDSS	74012	45787	45787	34684	37221	8566	39328	0.977113	0.468627	0.812916	0.594524
DeBreak	74012	49868	49709	43644	41248	7565	30368	0.936125	0.589688	0.845021	0.694634
Sniffles2	74012	54545	54458	44973	43106	10168	29039	0.973331	0.607645	0.809138	0.694063
pbsv	74012	52807	52741	44492	42927	9253	29520	0.983104	0.601146	0.822672	0.694676
cuteSV	74012	44937	44928	39438	36952	6416	34574	0.975146	0.532860	0.852057	0.655674
SVIM	74012	116615	116427	48022	47230	30995	25990	0.979929	0.648841	0.603771	0.625495

The table 1 shows the benchmarking results of the variation identification result. Where #SVs_bench represents the number of identified structural variations (SVs) in the benchmark set, #SV_user represents the number of SVs in the called set, and #SV_filtered_user represents the number of SVs after filtering out large SVs. #TP stands for the number of True Positives, indicating correctly identified targets or events. #FP stands for the number of False Positives, representing falsely identified targets or events. #FN represents the number of False Negatives, referring to the targets or events that were missed or not identified correctly. Identity represents the measure of sequence identity, which is calculated for matched SV pairs that include sequences.

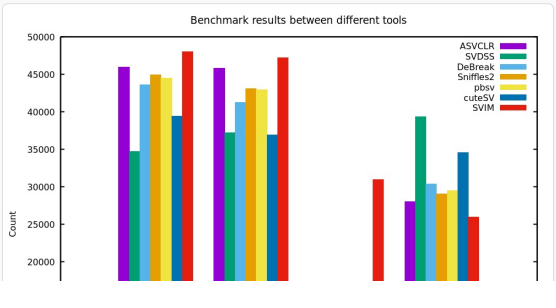
(2) The benchmarking results of two categorizes of metrics are shown in the figure:

Two categories of metrics are independently calculated: (a) one category includes Recall, Precision, F1 Score, and Identity; (b) the other category consists of #TP_bench, #TP_user, #FP, and #FN. The result statistics are as follows:

(a)



(b)



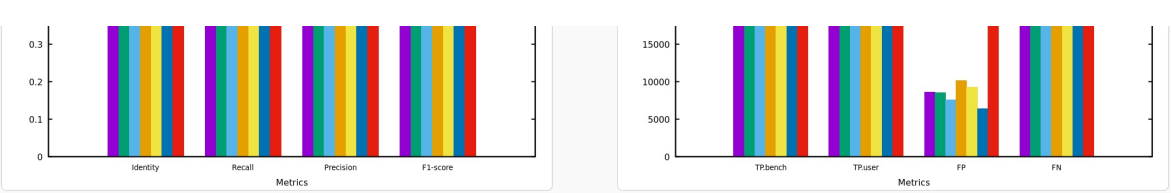


Figure 1 Benchmarking results of the user-call set

2. Statistical results of deviations for overlapping variants

For variations that overlap between the user-called set and the benchmark set, the deviations between them are quantified by calculating the breakpoint distance and the variant size ratio of the overlapping variations.

(1) Deviation of the center distance

As the center distance approaches 0, the deviation decreases, indicating a more precise identification result. Statistics results for eight size regions are presented in Table 2:

Table 2 Statistical results of breakpoint distance deviation

Tool	-200--151	-150--101	-100--51	-50--1	0-50	51-100	101-150	151-200
ASVCLR	241	358	756	6408	31364	4160	1924	1313
SVDSS	281	373	902	5463	27417	2188	1520	1122
DeBreak	278	451	849	3435	31649	3741	1546	963
Sniffles2	276	355	646	3154	32890	4559	2006	1394
pbsv	288	438	812	5027	32552	2314	1476	1071
cuteSV	182	276	604	4750	27134	3410	1541	1107
SVIM	553	686	1159	5988	35797	3855	2291	1737

(2) Deviation of the region size ratio

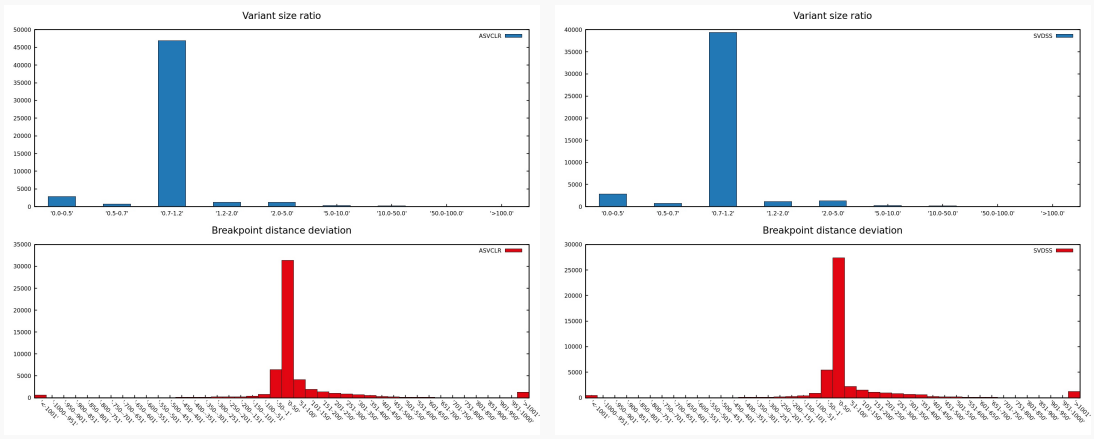
Calculating the variant size ratio for two overlapping variations based on the length of SVs, the closer the ratio is to 1, the smaller the deviation, indicating a more precise and accurate identification result. Statistics results for nine size regions are presented in Table 3:

Table 3 Statistical results of deviation of the variant region size ratio

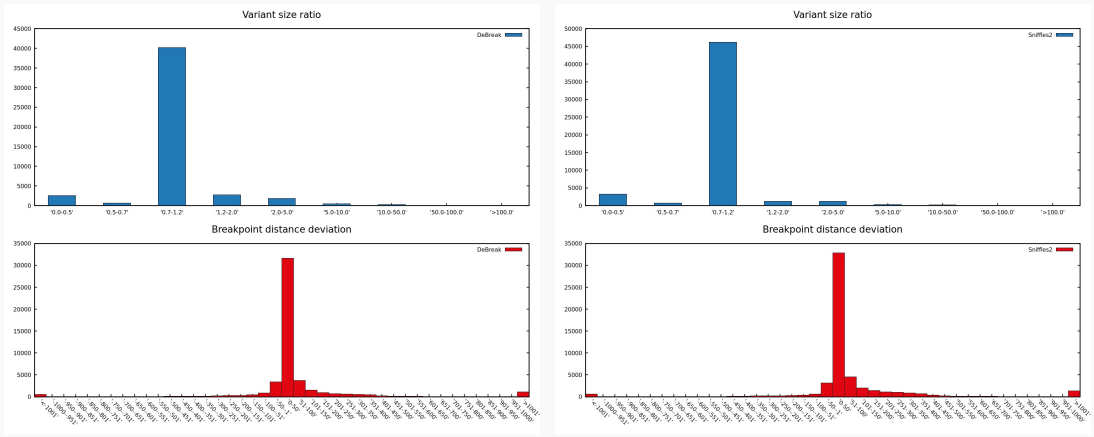
Tool	0.0-0.5	0.5-0.7	0.7-1.2	1.2-2.0	2.0-5.0	5.0-10.0	10.0-50.0	50.0-100.0	>100.0
ASVCLR	2885	773	46869	1287	1222	306	248	31	23
SVDSS	2874	737	39357	1186	1330	283	154	6	6
DeBreak	2559	657	40170	2723	1841	456	292	21	40
Sniffles2	3257	752	46170	1244	1236	342	231	31	43
pbsv	3500	840	43660	1724	1489	401	321	40	62
cuteSV	2916	615	39079	1039	934	213	175	16	24
SVIM	7184	1303	51061	1730	1752	607	524	85	173

The statistical results of user-called sets are as follows:

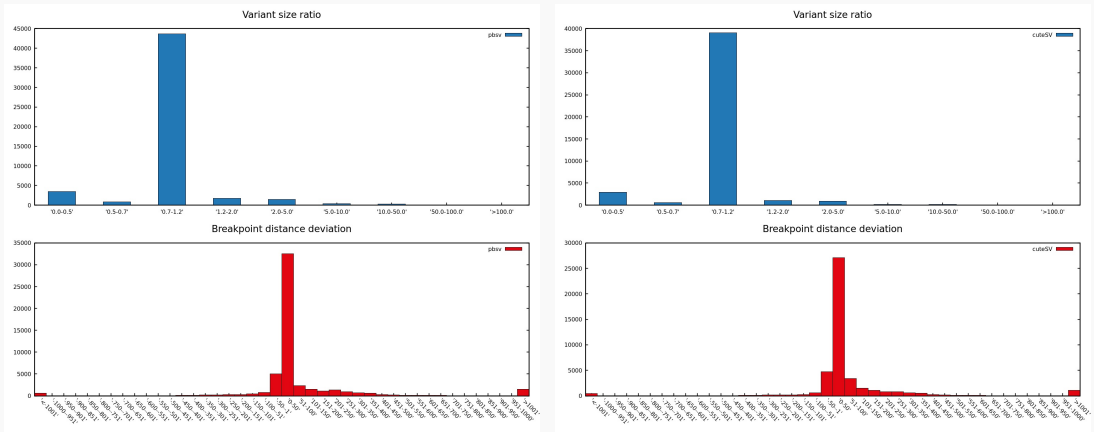
(a) ASVCLR: **(b) SVDSS:**



(c) DeBreak: **(d) Sniffles2:**



(e) pbsv: **(f) cuteSV:**



(g) SVIM:

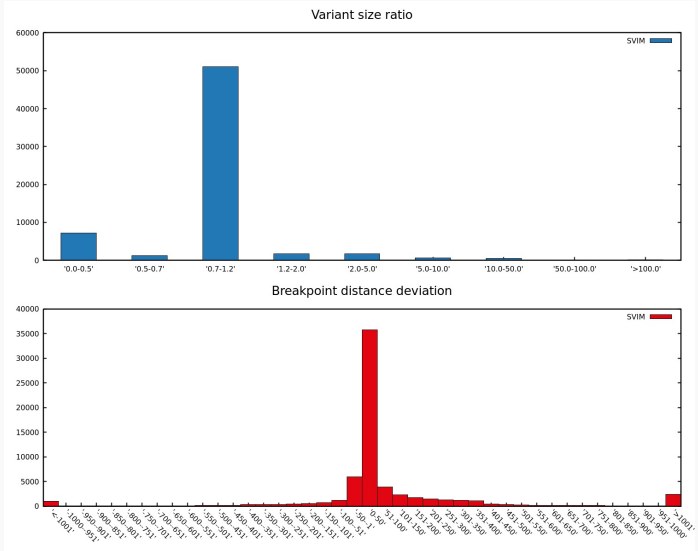


Figure 2 Deviation statistics with overlapping variations

3. Benchmarking results for metrics of different SV size regions

The SV identification results typically contain variations of various sizes, and categorize these variations into different size ranges could be used to explore the

identification results more detailed in a fine-grained manner, and could provide new insights into the sensitivity of SV callers to variations of different sizes. Detailed benchmarking results are presented in the table as followsï¼š

(1) Benchmarking results for metrics of different SV size regions with different methods

Variations are categorized into seven size regions and metrics are computed for comprehensive benchmarking for different detection methods within each region. The benchmarking results are as follows:

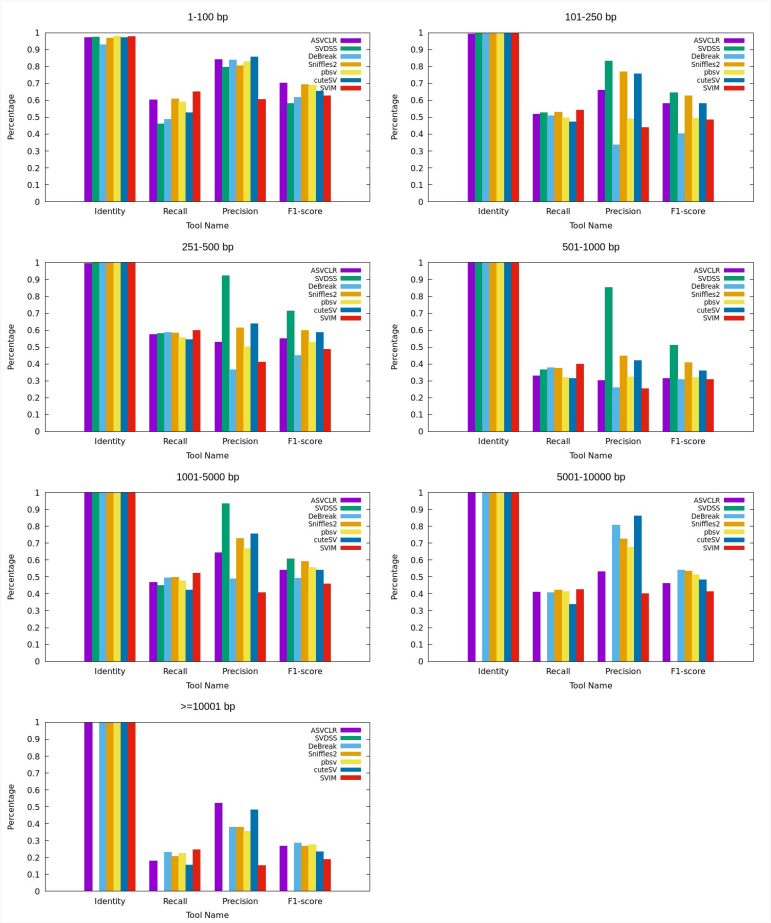


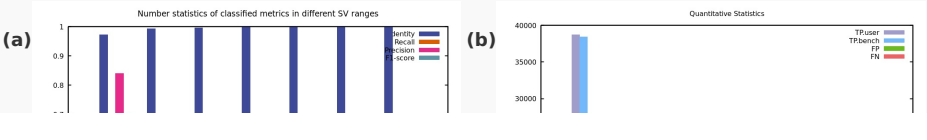
Figure 3 Statistics of metrics of different SV size region

(2) The user-called set (ASVCLR) of basic metrics results statistics

Table 4 The metric benchmarking results of ASVCLR in different SV regions

Region	#TP_bench	#TP_user	#FP	#FN	Recall	Precision	F1 score	Identity
1-100bp	38704	38396	7279	25440	0.972401	0.603392	0.840635	0.702525
101-250bp	2050	1889	968	1908	0.992270	0.517938	0.661183	0.580860
251-500bp	1707	1669	1481	1265	0.995352	0.574361	0.529841	0.551203
501-1000bp	365	366	849	747	0.997929	0.328237	0.301235	0.314157
1001-5000bp	579	754	420	659	0.999229	0.467690	0.642249	0.541243
5001-10000bp	138	140	124	198	1.000000	0.410714	0.530303	0.462910
>10000bp	45	46	42	207	1.000000	0.178571	0.522727	0.266204

Benchmarking results for metrics of different SV size regions show as following figures:



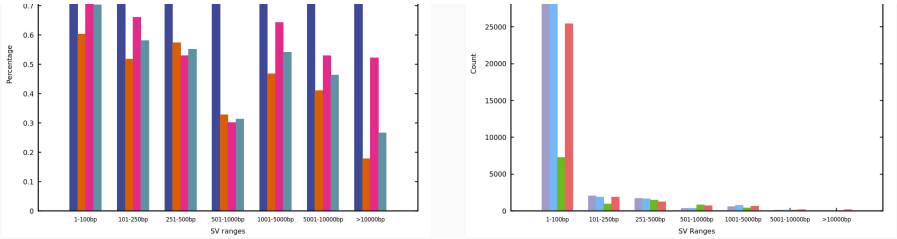


Figure 4 Result statistics of different SV size region

Figure (a) shows the statistical results of Recall, Precision, F1 score and Identity;
(b) shows the statistical results of #TP_benchmark, #TP_user,#FP and #FN.

(3) The user-called set (SVDSS) of basic metrics results statistics

Table 5 The metric benchmarking results of SVDSS in different SV regions

Region	#TP_bench	#TP_user	#FP	#FN	Recall	Precision	F1 score	Identity
1-100bp	29479	31916	8150	34665	0.973565	0.459575	0.796586	0.582873
101-250bp	2086	2126	428	1872	1.000000	0.527034	0.832420	0.645426
251-500bp	1728	1869	159	1244	1.000000	0.581427	0.921598	0.713018
501-1000bp	406	433	75	706	1.000000	0.365108	0.852362	0.511231
1001-5000bp	558	590	41	680	1.000000	0.450727	0.935024	0.608249
5001-10000bp	0	0	0	336	0.000000	0.000000	0.000000	0.000000
>10000bp	0	0	0	252	0.000000	0.000000	0.000000	0.000000

Benchmarking results for metrics of different SV size regions show as following figures:

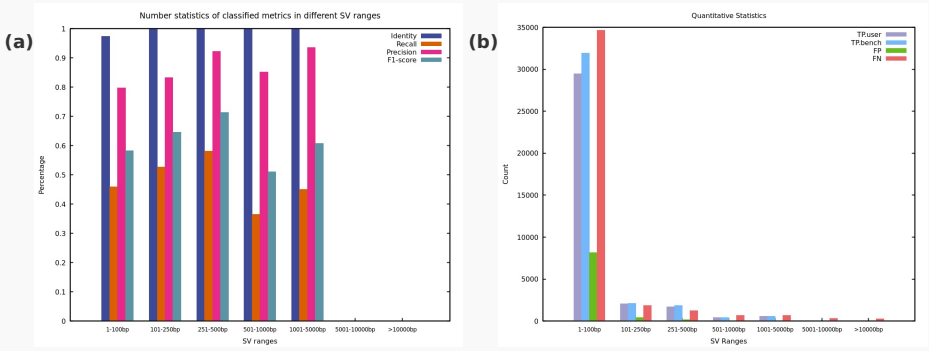


Figure 5 Result statistics of different SV size region

Figure (a) shows the statistical results of Recall, Precision, F1 score and Identity;
(b) shows the statistical results of #TP_benchmark, #TP_user,#FP and #FN.

(4) The user-called set (DeBreak) of basic metrics results statistics

Table 6 The metric benchmarking results of DeBreak in different SV regions

Region	#TP_bench	#TP_user	#FP	#FN	Recall	Precision	F1 score	Identity
1-100bp	31287	29770	5756	32857	0.930399	0.487762	0.837978	0.616612
101-250bp	2014	1843	3658	1944	0.999851	0.508843	0.335030	0.404036
251-500bp	1743	1699	2963	1229	0.999830	0.586474	0.364436	0.449532
501-1000bp	421	410	1165	691	0.998064	0.378597	0.260317	0.308509

1001-5000bp	613	598	625	625	1.000000	0.495153	0.488962	0.492038
5001-10000bp	137	138	33	199	1.000000	0.407738	0.807018	0.541758
>10000bp	58	59	96	194	1.000000	0.230159	0.380645	0.286864

Benchmarking results for metrics of different SV size regions show as following figures:

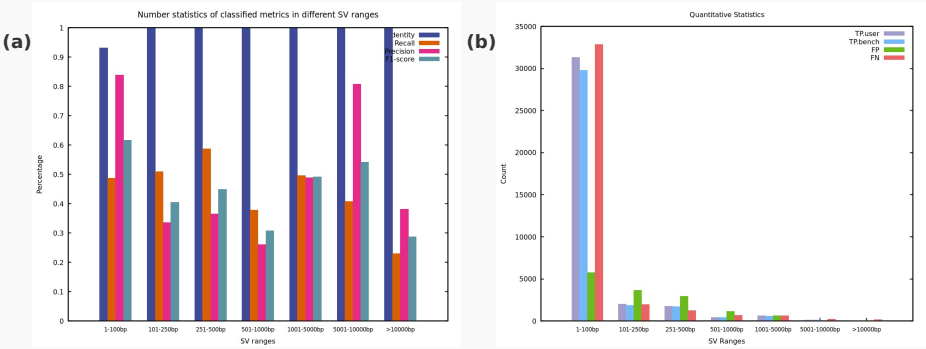


Figure 6 Result statistics of different SV size region

Figure (a) shows the statistical results of Recall, Precision, F1 score and Identity;
(b) shows the statistical results of #TP_benchmark, #TP_user,#FP and #FN.

(5) The user-called set (Sniffles2) of basic metrics results statistics

Table 7 The metric benchmarking results of Sniffles2 in different SV regions

Region	#TP_bench	#TP_user	#FP	#FN	Recall	Precision	F1 score	Identity
1-100bp	38987	36903	8913	25157	0.969980	0.607804	0.805461	0.692811
101-250bp	2098	1936	582	1860	1.000000	0.530066	0.768864	0.627514
251-500bp	1737	1705	1066	1235	1.000000	0.584455	0.615301	0.599482
501-1000bp	416	419	518	696	1.000000	0.374101	0.447172	0.407386
1001-5000bp	615	636	238	623	1.000000	0.496769	0.727689	0.590454
5001-10000bp	142	152	58	194	1.000000	0.422619	0.723810	0.533650
>10000bp	52	56	92	200	1.000000	0.206349	0.378378	0.267058

Benchmarking results for metrics of different SV size regions show as following figures:

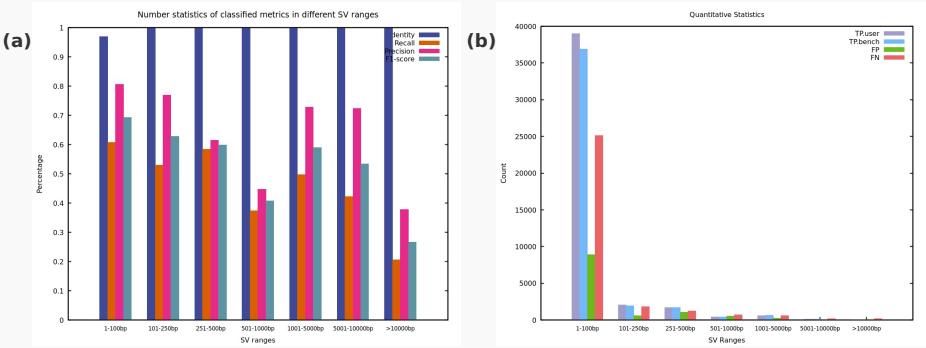


Figure 7 Result statistics of different SV size region

Figure (a) shows the statistical results of Recall, Precision, F1 score and Identity;
(b) shows the statistical results of #TP_benchmark, #TP_user,#FP and #FN.

(6) The user-called set (pbsv) of basic metrics results statistics

Table 8 The metric benchmarking results of pbsv in different SV regions

Region	#TP_bench	#TP_user	#FP	#FN	Recall	Precision	F1 score	Identity
1-100bp	37817	35709	7265	26327	0.980641	0.589564	0.830944	0.689746
101-250bp	1971	1794	1853	1987	1.000000	0.497979	0.491911	0.494926
251-500bp	1656	1620	1607	1316	1.000000	0.557201	0.502014	0.528170
501-1000bp	355	351	738	757	1.000000	0.319245	0.322314	0.320772
1001-5000bp	591	582	288	647	1.000000	0.477383	0.668966	0.557165
5001-10000bp	139	140	67	197	1.000000	0.413690	0.676328	0.513368
>10000bp	57	59	107	195	1.000000	0.226190	0.355422	0.276449

Benchmarking results for metrics of different SV size regions show as following figures:

(a) (b)

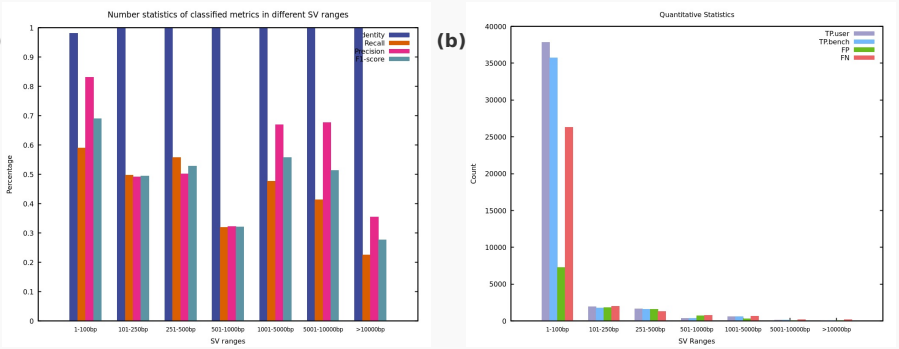


Figure 8 Result statistics of different SV size region

Figure (a) shows the statistical results of Recall, Precision, F1 score and Identity;
(b) shows the statistical results of #TP_benchmark, #TP_user,#FP and #FN.

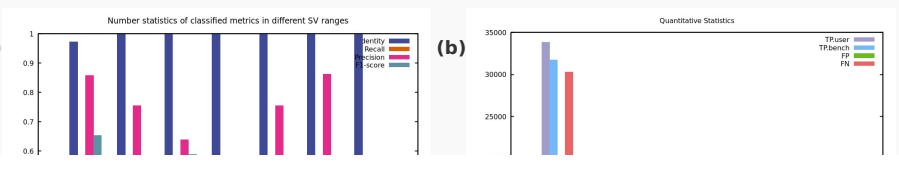
(7) The user-called set (cuteSV) of basic metrics results statistics

Table 9 The metric benchmarking results of cuteSV in different SV regions

Region	#TP_bench	#TP_user	#FP	#FN	Recall	Precision	F1 score	Identity
1-100bp	33846	31712	5251	30298	0.971827	0.527656	0.857939	0.653433
101-250bp	1866	1686	546	2092	1.000000	0.471450	0.755376	0.580559
251-500bp	1618	1577	891	1354	1.000000	0.544415	0.638979	0.587918
501-1000bp	349	341	472	763	1.000000	0.313849	0.419434	0.359040
1001-5000bp	522	512	166	716	1.000000	0.421648	0.755162	0.541145
5001-10000bp	113	113	18	223	1.000000	0.336310	0.862595	0.483940
>10000bp	39	40	43	213	1.000000	0.154762	0.481928	0.234287

Benchmarking results for metrics of different SV size regions show as following figures:

(a) (b)



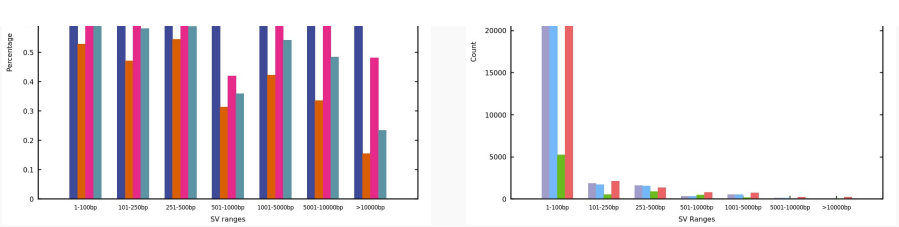


Figure 9 Result statistics of different SV size region

Figure (a) shows the statistical results of Recall, Precision, F1 score and Identity;
(b) shows the statistical results of #TP_benchmark, #TP_user,#FP and #FN.

(8) The user-called set (SVIM) of basic metrics results statistics

Table 10 The metric benchmarking results of SVIM in different SV regions

Region	#TP_bench	#TP_user	#FP	#FN	Recall	Precision	F1 score	Identity
1-100bp	41770	39594	25735	22374	0.977421	0.651191	0.606071	0.627821
101-250bp	2141	1960	2500	1817	1.000000	0.540930	0.439462	0.484945
251-500bp	1776	1747	2502	1196	1.000000	0.597577	0.411156	0.487140
501-1000bp	442	442	1308	670	1.000000	0.397482	0.252571	0.308875
1001-5000bp	645	656	953	593	1.000000	0.521002	0.407707	0.457444
5001-10000bp	143	146	218	193	1.000000	0.425595	0.401099	0.412984
>10000bp	62	71	393	190	1.000000	0.246032	0.153017	0.188684

Benchmarking results for metrics of different SV size regions show as following figures:

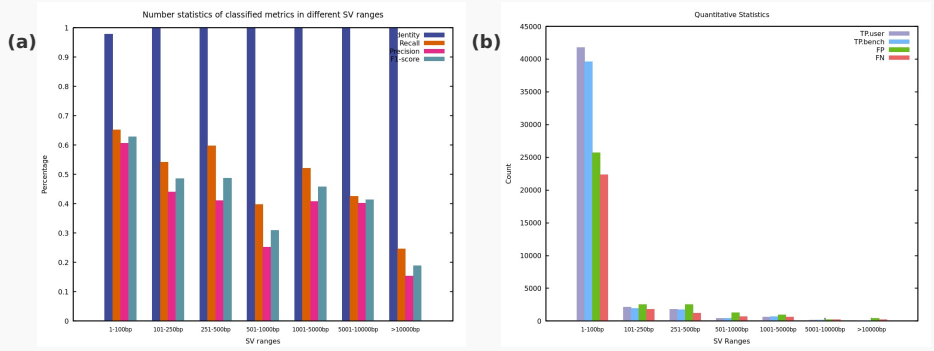


Figure 10 Result statistics of different SV size region

Figure (a) shows the statistical results of Recall, Precision, F1 score and Identity;
(b) shows the statistical results of #TP_benchmark, #TP_user,#FP and #FN.

4. Quantitative statistics

(1) Statistics of the count of different SV lengths in the benchmark set:

The SV reference region size statistics for benchmark set: Total SVs numberï¼š74012

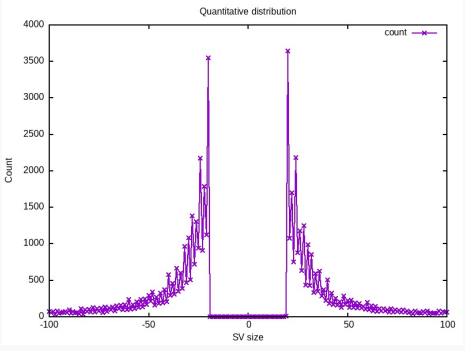


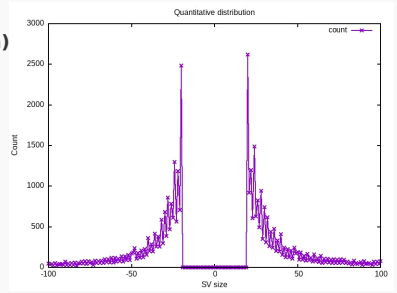
Figure 11 The quantity distribution of the benchmark set

The figure shows the distribution of SV counts of the benchmark set.

(2) Statistics of the count of different SV lengths in the user-called set (ASVCLR):

The SV reference region size statistics before filtering for user-called set (ASVCLR):Total SVs numberï¼š54423
The SV reference region size statistics after filtering for user-called set (ASVCLR):Total SVs numberï¼š54423

(a)



(b)

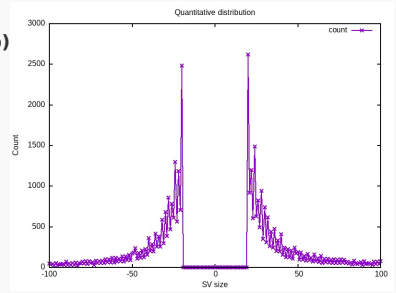


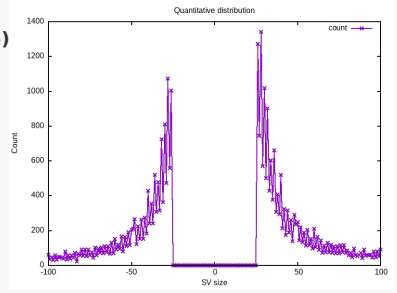
Figure 12 The quantity distribution of the user-called set

The figures show the distribution of SV counts, where (a) represents the result statistics before filtering large SVs, and (b) shows the result statistics after filtering large SVs.

(3) Statistics of the count of different SV lengths in the user-called set (SVDSS):

The SV reference region size statistics before filtering for user-called set (SVDSS):Total SVs numberï¼š45787
The SV reference region size statistics after filtering for user-called set (SVDSS):Total SVs numberï¼š45787

(a)



(b)

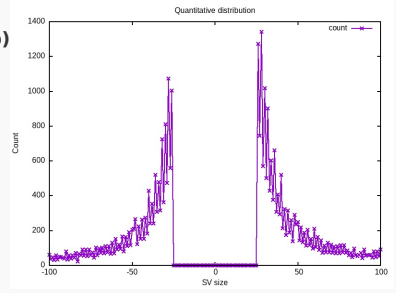


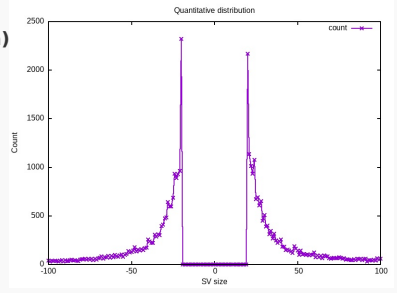
Figure 13 The quantity distribution of the user-called set

The figures show the distribution of SV counts, where (a) represents the result statistics before filtering large SVs, and (b) shows the result statistics after filtering large SVs.

(4) Statistics of the count of different SV lengths in the user-called set (DeBreak):

The SV reference region size statistics before filtering for user-called set (DeBreak):Total SVs numberï¼š49868
The SV reference region size statistics after filtering for user-called set (DeBreak):Total SVs numberï¼š49709

(a)



(b)

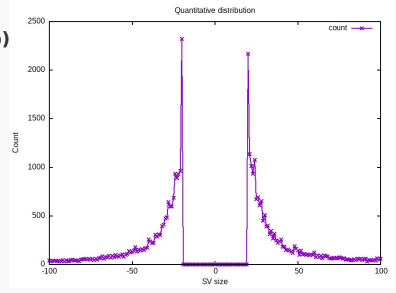


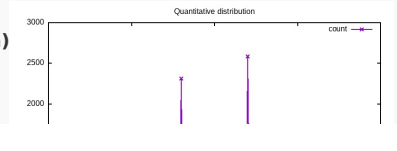
Figure 14 The quantity distribution of the user-called set

The figures show the distribution of SV counts, where (a) represents the result statistics before filtering large SVs, and (b) shows the result statistics after filtering large SVs.

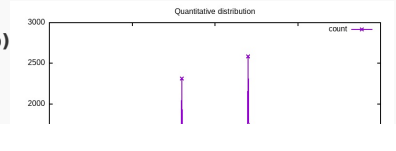
(5) Statistics of the count of different SV lengths in the user-called set (Sniffles2):

The SV reference region size statistics before filtering for user-called set (Sniffles2):Total SVs numberï¼š54545
The SV reference region size statistics after filtering for user-called set (Sniffles2):Total SVs numberï¼š54458

(a)



(b)



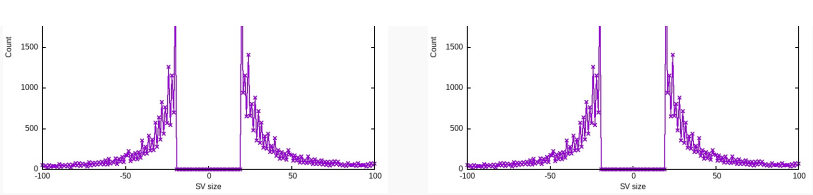


Figure 15 The quantity distribution of the user-called set

The figures show the distribution of SV counts, where (a) represents the result statistics before filtering large SVs, and (b) shows the result statistics after filtering large SVs.

(6) Statistics of the count of different SV lengths in the user-called set (pbsv):

The SV reference region size statistics before filtering for user-called set (pbsv):Total SVs numberï¼š52807

The SV reference region size statistics after filtering for user-called set (pbsv):Total SVs numberï¼š52741

(a) **(b)**

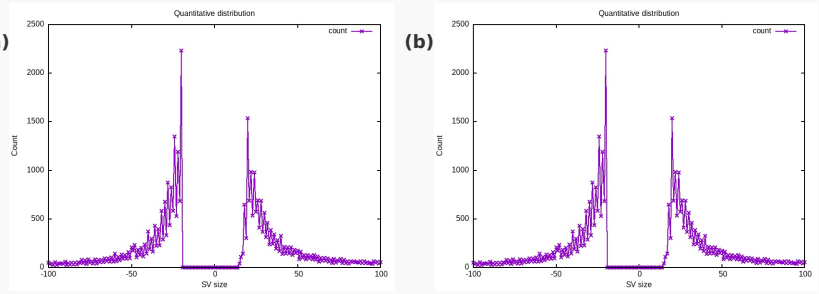


Figure 16 The quantity distribution of the user-called set

The figures show the distribution of SV counts, where (a) represents the result statistics before filtering large SVs, and (b) shows the result statistics after filtering large SVs.

(7) Statistics of the count of different SV lengths in the user-called set (cuteSV):

The SV reference region size statistics before filtering for user-called set (cuteSV):Total SVs numberï¼š44937

The SV reference region size statistics after filtering for user-called set (cuteSV):Total SVs numberï¼š44928

(a) **(b)**

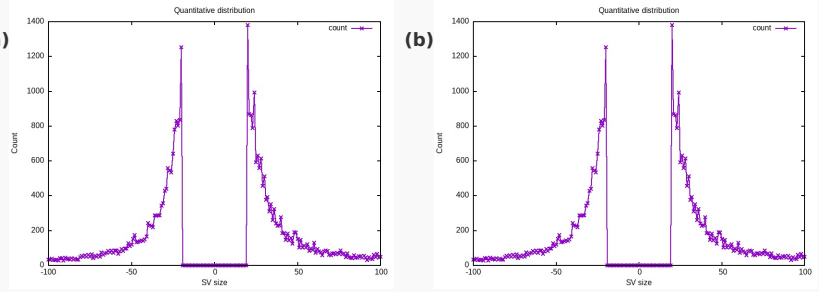


Figure 17 The quantity distribution of the user-called set

The figures show the distribution of SV counts, where (a) represents the result statistics before filtering large SVs, and (b) shows the result statistics after filtering large SVs.

(8) Statistics of the count of different SV lengths in the user-called set (SVIM):

The SV reference region size statistics before filtering for user-called set (SVIM):Total SVs numberï¼š116615

The SV reference region size statistics after filtering for user-called set (SVIM):Total SVs numberï¼š116427

(a) **(b)**

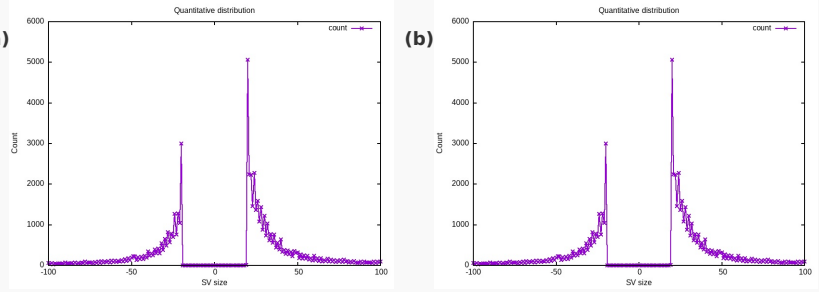


Figure 18 The quantity distribution of the user-called set

The figures show the distribution of SV counts, where (a) represents the result statistics before filtering large SVs, and (b) shows the result statistics after filtering large SVs.

More information

- For more detailed benchmarking results, please refer to the generated result information in the respective folders.
- For more detailed experiment information, please refer to the github repositories: sv_stat and sv_stat-experiments.
- If you have any problems, comments, or suggestions, please contact xzhu@ytu.edu.cn without hesitation. Thank you very much!

---------- This is the end of the Benchmarking Reports. ----------
-