

Supplementary Information

Supplementary Tables

Supplementary Table 1. Truvari benchmarking metrics for real genomic intervals (GRCh38)

Benchmarking results from Truvari for structural variant callers evaluated against the Genome in a Bottle (GIAB) truth set using real genomic intervals. Results are stratified by sequencing technology (Illumina WES, Illumina WGS, ONT, PacBio), SV caller (Manta, CuteSV, Sniffles, Pbsv), and target region (gene panel, high confidence, WES+UTR). TP.b = true positives (base), TP.c = true positives (comp), FP = false positives, FN = false negatives, Prec = precision, Rec = recall, Base = base count, Comp = comp count, GT = genotype concordance.

Tech	Caller	Region	TP.b	TP.c	FP	FN	Prec	Rec	F1	Base	Comp	GT
Illumina_WES	Manta	gene_panel	4	4	0	1297	1.000	0.003	0.006	1301	4	1.000
Illumina_WES	Manta	high_conf	14	14	19	9685	0.424	0.001	0.003	9699	33	1.000
Illumina_WES	Manta	wes_utr	7	7	18	254	0.280	0.027	0.049	261	25	1.000
Illumina_WGS	Manta	gene_panel	565	565	43	736	0.929	0.434	0.592	1301	608	0.961
Illumina_WGS	Manta	high_conf	4499	4499	326	5200	0.932	0.464	0.620	9699	4825	0.961
Illumina_WGS	Manta	wes_utr	117	117	24	144	0.830	0.448	0.582	261	141	0.991
ONT	CuteSV	gene_panel	1216	1216	205	85	0.856	0.935	0.893	1301	1421	0.000
ONT	CuteSV	high_conf	9130	9130	1417	569	0.866	0.941	0.902	9699	10547	0.000
ONT	CuteSV	wes_utr	225	225	73	36	0.755	0.862	0.805	261	298	0.000
ONT	Sniffles	gene_panel	1225	1225	96	76	0.927	0.942	0.934	1301	1321	0.970
ONT	Sniffles	high_conf	9140	9140	691	559	0.930	0.942	0.936	9699	9831	0.958
ONT	Sniffles	wes_utr	219	219	46	42	0.826	0.839	0.833	261	265	0.954
PacBio	CuteSV	gene_panel	1241	1241	112	60	0.917	0.954	0.935	1301	1353	0.000
PacBio	CuteSV	high_conf	9295	9295	819	404	0.919	0.958	0.938	9699	10114	0.000
PacBio	CuteSV	wes_utr	230	230	40	31	0.852	0.881	0.866	261	270	0.000
PacBio	Pbsv	gene_panel	1240	1240	90	61	0.932	0.953	0.943	1301	1330	0.957
PacBio	Pbsv	high_conf	9301	9301	747	398	0.926	0.959	0.942	9699	10048	0.965
PacBio	Pbsv	wes_utr	231	231	48	30	0.828	0.885	0.856	261	279	0.961

Supplementary Table 2. Truvari benchmarking metrics for real genomic intervals (GRCh37)

Benchmarking results from Truvari for structural variant callers evaluated against the Genome in a Bottle (GIAB) truth set using real genomic intervals on GRCh37 reference genome. Results are stratified by sequencing technology (Illumina WES, Illumina WGS, ONT, PacBio), SV caller (Manta, CuteSV, Sniffles, Pbsv), and target region (gene panel, high confidence, WES+UTR). Column abbreviations are as in Supplementary Table 1.

Tech	Caller	Region	TP.b	TP.c	FP	FN	Prec	Rec	F1	Base	Comp	GT
Illumina_WES	Manta	gene_panel	4	4	0	1297	1.000	0.003	0.006	1301	4	1.000
Illumina_WES	Manta	high_conf	14	14	19	9685	0.424	0.001	0.003	9699	33	1.000
Illumina_WES	Manta	wes_utr	7	7	18	254	0.280	0.027	0.049	261	25	1.000
Illumina_WGS	Manta	gene_panel	565	565	43	736	0.929	0.434	0.592	1301	608	0.961
Illumina_WGS	Manta	high_conf	4499	4499	326	5200	0.932	0.464	0.620	9699	4825	0.961
Illumina_WGS	Manta	wes_utr	117	117	24	144	0.830	0.448	0.582	261	141	0.991
ONT	CuteSV	gene_panel	1216	1216	205	85	0.856	0.935	0.893	1301	1421	0.000
ONT	CuteSV	high_conf	9130	9130	1417	569	0.866	0.941	0.902	9699	10547	0.000
ONT	CuteSV	wes_utr	225	225	73	36	0.755	0.862	0.805	261	298	0.000
ONT	Sniffles	gene_panel	1225	1225	96	76	0.927	0.942	0.934	1301	1321	0.970
ONT	Sniffles	high_conf	9140	9140	691	559	0.930	0.942	0.936	9699	9831	0.958
ONT	Sniffles	wes_utr	219	219	46	42	0.826	0.839	0.833	261	265	0.954
PacBio	CuteSV	gene_panel	1241	1241	112	60	0.917	0.954	0.935	1301	1353	0.000
PacBio	CuteSV	high_conf	9295	9295	819	404	0.919	0.958	0.938	9699	10114	0.000
PacBio	CuteSV	wes_utr	230	230	40	31	0.852	0.881	0.866	261	270	0.000
PacBio	Pbsv	gene_panel	1240	1240	90	61	0.932	0.953	0.943	1301	1330	0.957
PacBio	Pbsv	high_conf	9301	9301	747	398	0.926	0.959	0.942	9699	10048	0.965
PacBio	Pbsv	wes_utr	231	231	48	30	0.828	0.885	0.856	261	279	0.961

Supplementary Table 3a. Simulated intervals percentile ranks and KDE (GRCh38)

Percentile ranks and kernel density estimates (KDE) from simulation-based benchmarking on GRCh38. For each technology-caller-region combination, the table shows where real data metrics fall within the simulated distribution. Percentile rank indicates the proportion of simulated runs with lower metric values; KDE estimates the probability density at the real data value. Prec = precision, Rec = recall, %ile = percentile.

Combination	Prec %ile	Rec %ile	F1 %ile	Prec KDE	Rec KDE	F1 KDE
Illumina_WGS.Manta.gene_panel	25.64	0.00	0	0.247	1.000	1.000
ONT.CuteSV.gene_panel	98.29	99.15	100	0.874	0.976	0.986
PacBio.CuteSV.gene_panel	100.00	100.00	100	0.997	1.000	1.000
ONT.Sniffles.gene_panel	100.00	100.00	100	0.981	1.000	1.000
PacBio.Pbsv.gene_panel	100.00	100.00	100	1.000	0.985	1.000
Illumina_WGS.Manta.high_conf	36.75	0.00	0	0.125	1.000	1.000
ONT.CuteSV.high_conf	100.00	100.00	100	0.976	0.993	1.000
PacBio.CuteSV.high_conf	100.00	100.00	100	0.999	1.000	1.000
ONT.Sniffles.high_conf	100.00	100.00	100	0.994	1.000	1.000
PacBio.Pbsv.high_conf	100.00	100.00	100	1.000	0.999	1.000
Illumina_WGS.Manta.wes utr	0.00	0.00	0	1.000	1.000	1.000
ONT.CuteSV.wes utr	0.00	0.00	0	1.000	0.981	1.000
PacBio.CuteSV.wes utr	8.55	0.00	0	0.504	1.000	0.981
ONT.Sniffles.wes utr	0.00	0.00	0	1.000	1.000	1.000
PacBio.Pbsv.wes utr	0.00	0.00	0	0.989	1.000	1.000

Supplementary Table 3b. Simulated intervals medians, differences, and variability (GRCh38)

Median values, differences, and standard deviations from simulation-based benchmarking on GRCh38. For each technology-caller-region combination, the table shows median metric values across simulations, the difference between real data and simulated median (real minus simulated), and the standard deviation across simulation replicates. Prec = precision, Rec = recall, med = median, diff = difference, SD = standard deviation.

Combination	Prec med	Rec med	F1 med	Prec diff	Rec diff	F1 diff	Prec SD	Rec SD	F1 SD
Illumina_WGS.Manta.gene_panel	0.936	0.612	0.740	0.007	0.177	0.148	0.011	0.017	0.013
ONT.CuteSV.gene_panel	0.825	0.900	0.861	-0.030	-0.035	-0.033	0.014	0.011	0.010
PacBio.CuteSV.gene_panel	0.874	0.918	0.894	-0.044	-0.036	-0.041	0.015	0.009	0.010
ONT.Sniffles.gene_panel	0.889	0.888	0.888	-0.038	-0.053	-0.046	0.013	0.010	0.009
PacBio.Pbsv.gene_panel	0.866	0.926	0.895	-0.067	-0.028	-0.047	0.013	0.010	0.009
Illumina_WGS.Manta.high_conf	0.936	0.612	0.740	0.004	0.148	0.120	0.011	0.017	0.013
ONT.CuteSV.high_conf	0.825	0.900	0.861	-0.040	-0.042	-0.041	0.014	0.011	0.010
PacBio.CuteSV.high_conf	0.874	0.918	0.894	-0.045	-0.040	-0.044	0.015	0.009	0.010
ONT.Sniffles.high_conf	0.889	0.888	0.888	-0.041	-0.054	-0.048	0.013	0.010	0.009
PacBio.Pbsv.high_conf	0.866	0.926	0.895	-0.060	-0.033	-0.047	0.013	0.010	0.009
Illumina_WGS.Manta.wes utr	0.936	0.612	0.740	0.107	0.163	0.157	0.011	0.017	0.013
ONT.CuteSV.wes utr	0.825	0.900	0.861	0.070	0.038	0.056	0.014	0.011	0.010
PacBio.CuteSV.wes utr	0.874	0.918	0.894	0.022	0.037	0.028	0.015	0.009	0.010
ONT.Sniffles.wes utr	0.889	0.888	0.888	0.062	0.049	0.056	0.013	0.010	0.009
PacBio.Pbsv.wes utr	0.866	0.926	0.895	0.038	0.040	0.040	0.013	0.010	0.009

Supplementary Table 4a. Simulated intervals percentile ranks and KDE (GRCh37)

Percentile ranks and kernel density estimates (KDE) from simulation-based benchmarking on GRCh37. For each technology-caller-region combination, the table shows where real data metrics fall within the simulated distribution. Column abbreviations are as in Supplementary Table 3a.

Combination	Prec %ile	Rec %ile	F1 %ile	Prec KDE	Rec KDE	F1 KDE
Illumina_WGS-Manta-gene_panel	32.43	0.00	0	0.236	1.000	1.000
ONT-CuteSV-gene_panel	98.20	99.10	100	0.801	0.976	0.985
ONT-Sniffles-gene_panel	100.00	100.00	100	0.997	1.000	1.000
PacBio-CuteSV-gene_panel	100.00	100.00	100	0.999	1.000	1.000
PacBio-Pbsv-gene_panel	100.00	100.00	100	1.000	0.997	1.000
Illumina_WGS-Manta-high_conf	39.64	0.00	0	0.126	1.000	1.000
ONT-CuteSV-high_conf	100.00	100.00	100	0.949	0.993	1.000
ONT-Sniffles-high_conf	100.00	100.00	100	0.999	1.000	1.000
PacBio-CuteSV-high_conf	100.00	100.00	100	1.000	1.000	1.000
PacBio-Pbsv-high_conf	100.00	100.00	100	1.000	1.000	1.000
Illumina_WGS-Manta-wes_utr	0.00	0.00	0	1.000	1.000	1.000
ONT-CuteSV-wes_utr	0.00	0.90	0	1.000	0.961	1.000
ONT-Sniffles-wes_utr	0.00	0.00	0	1.000	1.000	1.000
PacBio-CuteSV-wes_utr	4.50	0.00	0	0.678	1.000	0.998
PacBio-Pbsv-wes_utr	0.90	0.00	0	0.971	1.000	0.999

Supplementary Table 4b. Simulated intervals medians, differences, and variability (GRCh37)

Median values, differences, and standard deviations from simulation-based benchmarking on GRCh37. For each technology-caller-region combination, the table shows median metric values across simulations, the difference between real data and simulated median, and the standard deviation. Column abbreviations are as in Supplementary Table 3b.

Combination	Prec med	Rec med	F1 med	Prec diff	Rec diff	F1 diff	Prec SD	Rec SD	F1 SD
Illumina_WGS-Manta-gene_panel	0.936	0.609	0.738	0.006	0.175	0.146	0.011	0.019	0.015
ONT-CuteSV-gene_panel	0.828	0.896	0.860	-0.028	-0.039	-0.034	0.014	0.013	0.011
ONT-Sniffles-gene_panel	0.889	0.885	0.887	-0.038	-0.057	-0.048	0.011	0.012	0.009
PacBio-CuteSV-gene_panel	0.873	0.919	0.897	-0.044	-0.035	-0.039	0.013	0.010	0.009
PacBio-Pbsv-gene_panel	0.867	0.923	0.894	-0.065	-0.031	-0.049	0.014	0.010	0.010
Illumina_WGS-Manta-high_conf	0.936	0.609	0.738	0.003	0.145	0.118	0.011	0.019	0.015
ONT-CuteSV-high_conf	0.828	0.896	0.860	-0.038	-0.046	-0.042	0.014	0.013	0.011
ONT-Sniffles-high_conf	0.889	0.885	0.887	-0.041	-0.058	-0.049	0.011	0.012	0.009
PacBio-CuteSV-high_conf	0.873	0.919	0.897	-0.046	-0.039	-0.042	0.013	0.010	0.009
PacBio-Pbsv-high_conf	0.867	0.923	0.894	-0.058	-0.036	-0.048	0.014	0.010	0.010
Illumina_WGS-Manta-wes_utr	0.936	0.609	0.738	0.106	0.161	0.156	0.011	0.019	0.015
ONT-CuteSV-wes_utr	0.828	0.896	0.860	0.073	0.033	0.055	0.014	0.013	0.011
ONT-Sniffles-wes_utr	0.889	0.885	0.887	0.063	0.046	0.054	0.011	0.012	0.009
PacBio-CuteSV-wes_utr	0.873	0.919	0.897	0.022	0.038	0.030	0.013	0.010	0.009
PacBio-Pbsv-wes_utr	0.867	0.923	0.894	0.039	0.038	0.038	0.014	0.010	0.010

Supplementary Data 1. Truvari benchmarking metrics for individual simulated intervals (GRCh38)

Title: Per-simulation Truvari benchmarking results across all technology-caller combinations.

Legend: Raw Truvari output metrics for each individual simulated target interval on GRCh38 reference genome. The table contains the same metrics as Supplementary Table 1 (TP.base, TP.comp, FP, FN, precision, recall, F1, base count, comp count, genotype concordance) but reported separately for each of the simulated intervals. Columns include technology, caller, and simulation identifier. Due to its size, this table is provided as a separate supplementary data file (TSV format).

File: *truvari_metrics_simulated_intervals_raw.tsv*

Supplementary Data 2. Truvari benchmarking metrics for individual simulated intervals (GRCh37)

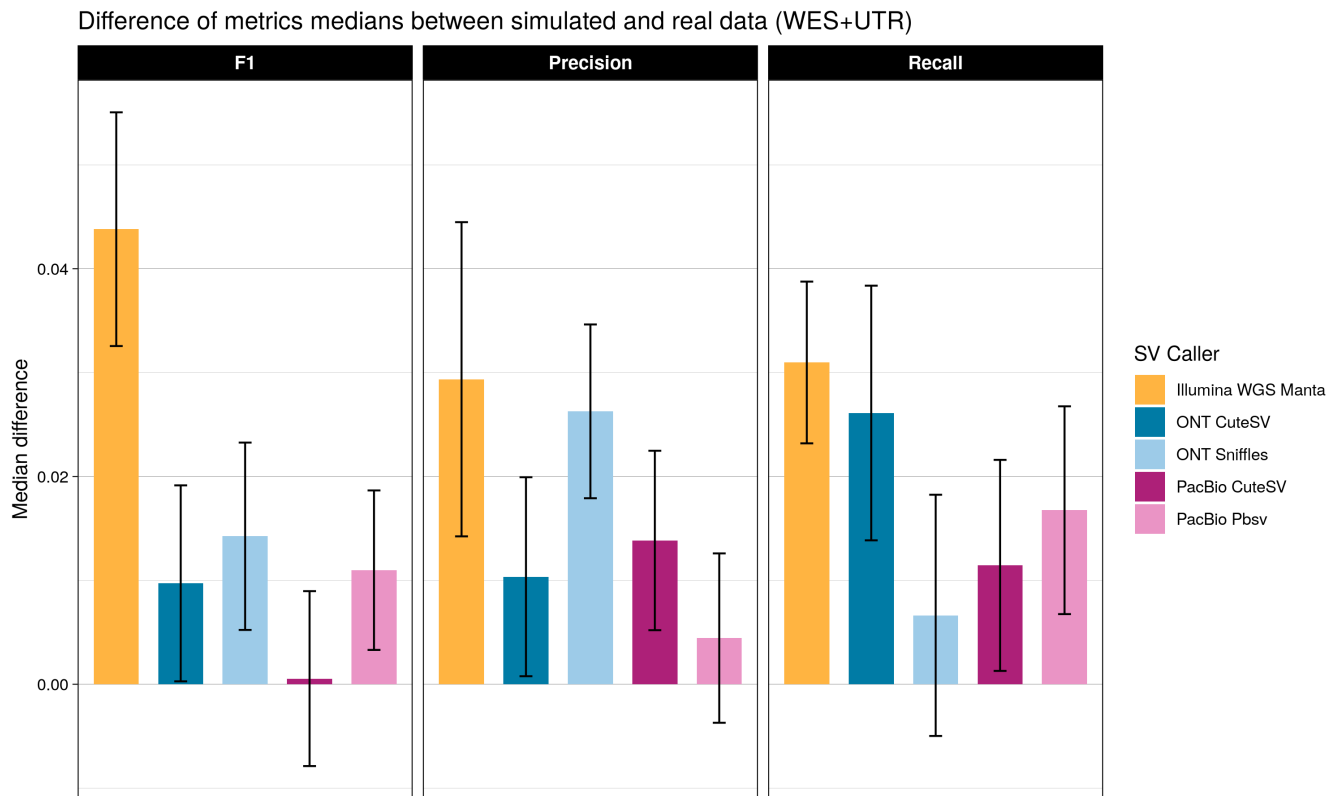
Title: Per-simulation Truvari benchmarking results across all technology-caller combinations (GRCh37).

Legend: Raw Truvari output metrics for each individual simulated target interval on GRCh37 reference genome. The table contains the same metrics as Supplementary Table 2 (TP.base, TP.comp, FP, FN, precision, recall, F1, base count, comp count, genotype concordance) but reported separately for each of the simulated intervals. Columns include technology, caller, and simulation identifier. Due to its size, this table is provided as a separate supplementary data file (TSV format).

File: *truvari_metrics_simulated_intervals_raw.tsv*

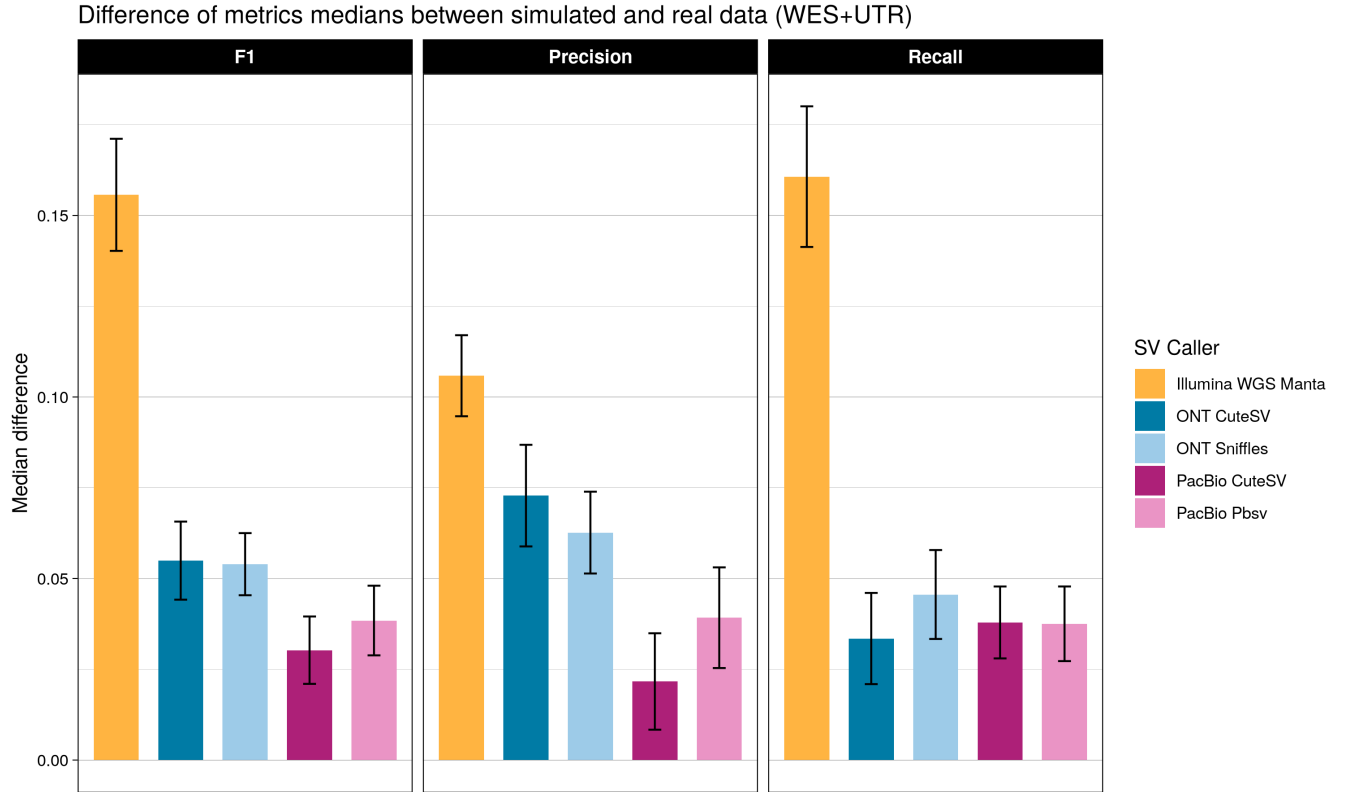
Supplementary Figures

Supplementary Figure 1. Difference of metrics medians between simulated and real data (GRCh38)



Bar plot showing the median difference in F1, precision, and recall between simulated and real benchmarking intervals (WES+UTR region) for each SV caller across sequencing technologies using the GRCh38 reference genome. Error bars represent the variability across simulation replicates. Illumina WGS Manta (orange) shows the largest median differences across all three metrics (F1, precision, and recall), indicating the greatest discrepancy between simulated and real interval performance. Long-read SV callers (ONT CuteSV, ONT Sniffles, PacBio CuteSV, PacBio Pbsv) demonstrate substantially smaller differences, suggesting that their benchmarking performance on simulated intervals more closely approximates performance on real genomic intervals.

Supplementary Figure 2. Difference of metrics medians between simulated and real data (GRCh37)



Bar plot showing the median difference in F1, precision, and recall between simulated and real benchmarking intervals (WES+UTR region) for each SV caller across sequencing technologies using the GRCh37 reference genome. Error bars represent the variability across simulation replicates. As with GRCh38, Illumina WGS Manta shows the largest median differences across all three metrics, while long-read SV callers demonstrate substantially smaller differences between simulated and real interval performance.