
Figures and figure supplements

Benchmarking reveals superiority of deep learning variant callers on bacterial nanopore sequence data

Michael B Hall *et al.*

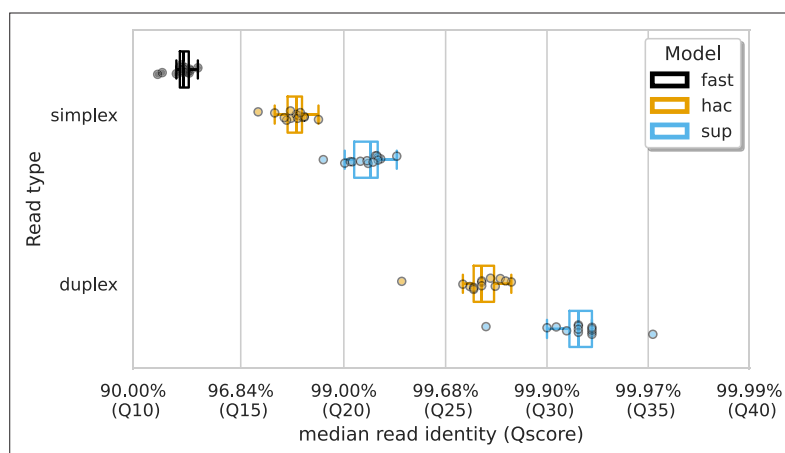


Figure 1. Median alignment-based read identity (x-axis) for each sample (points) stratified by basecalling model (colours) and read type (y-axis). The Qscore is the logarithmic transformation of the read identity, $Q = -10 \log_{10} P$, where P is the read identity.

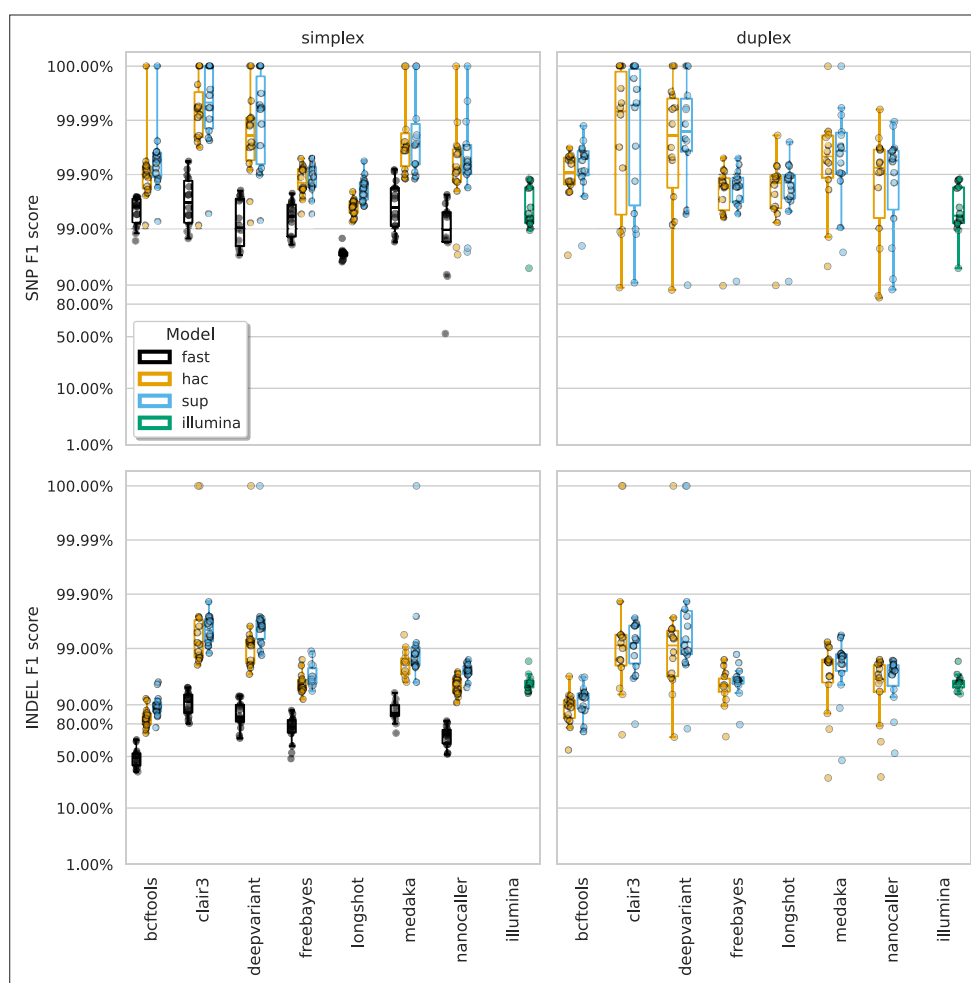


Figure 2. The highest F1 score for each sample (points), stratified by basecalling model (colours), variant type (rows), and read type (columns). Illumina results (green) are included as a reference and do not have different basecalling models or read types. Note, Longshot does not provide indel calls.

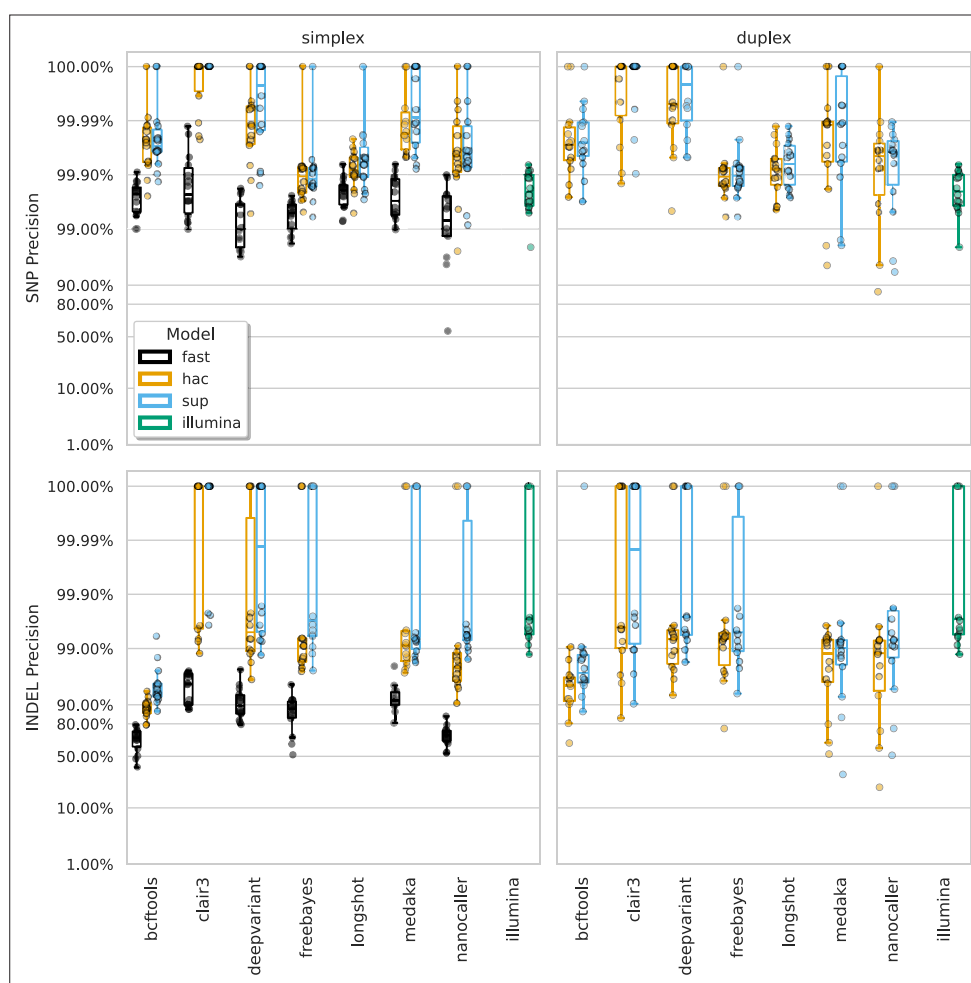


Figure 2—figure supplement 1. Precision at the highest F1 score for each sample (points), stratified by basecalling model (colours), variant type (rows), and read type (columns). Illumina results (green) are included as a reference and do not have different basecalling models or read types. Note, Longshot does not provide indel calls.

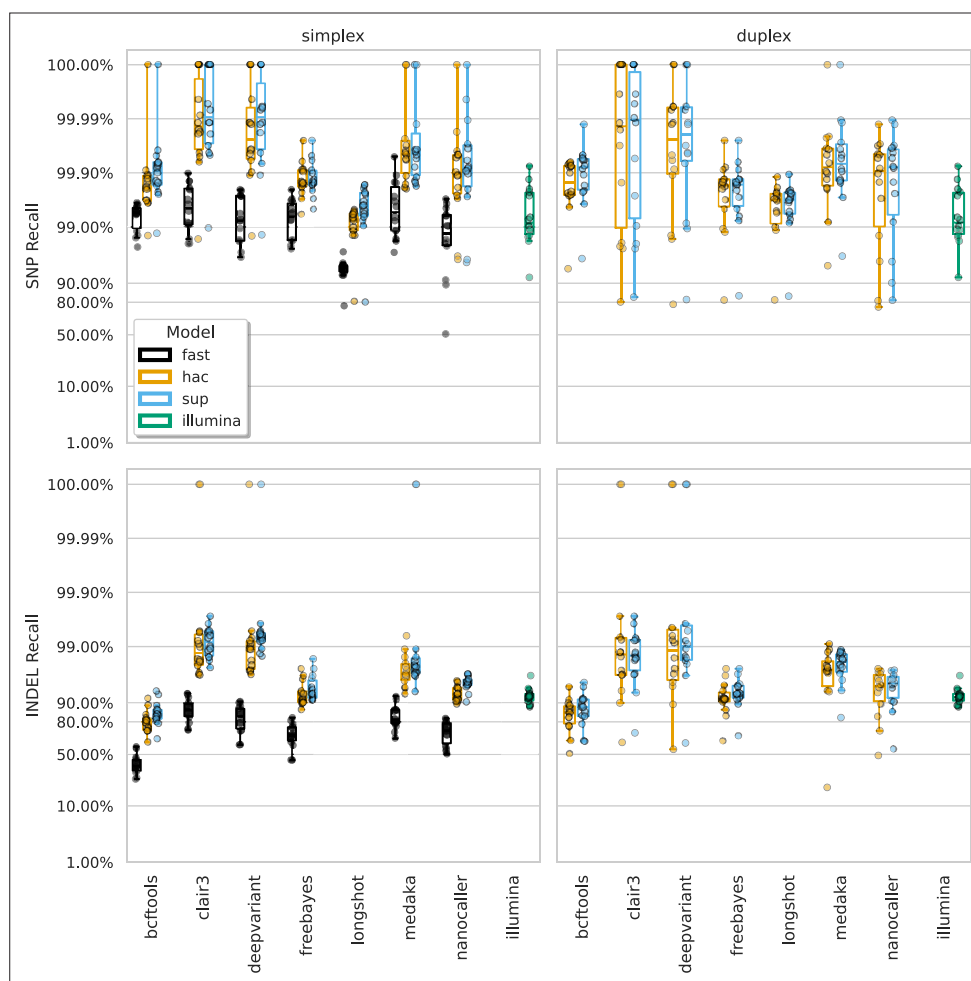


Figure 2—figure supplement 2. Recall at the highest F1 score for each sample (points), stratified by basecalling model (colours), variant type (rows), and read type (columns). Illumina results (green) are included as a reference and do not have different basecalling models or read types. Note, Longshot does not provide indel calls.

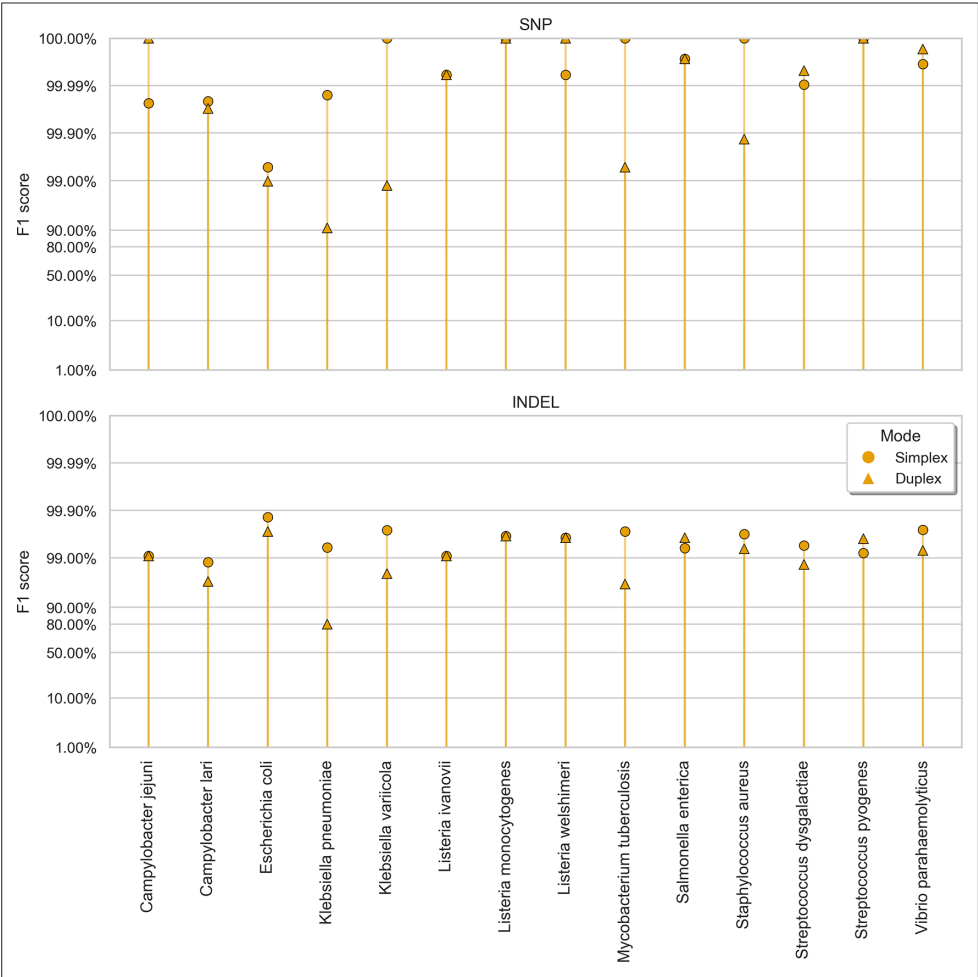


Figure 2—figure supplement 3. Clair3 sup model F1 score (y-axis) at the highest F1 score for each sample (x-axis), stratified by variant type (rows), and read type (shapes).

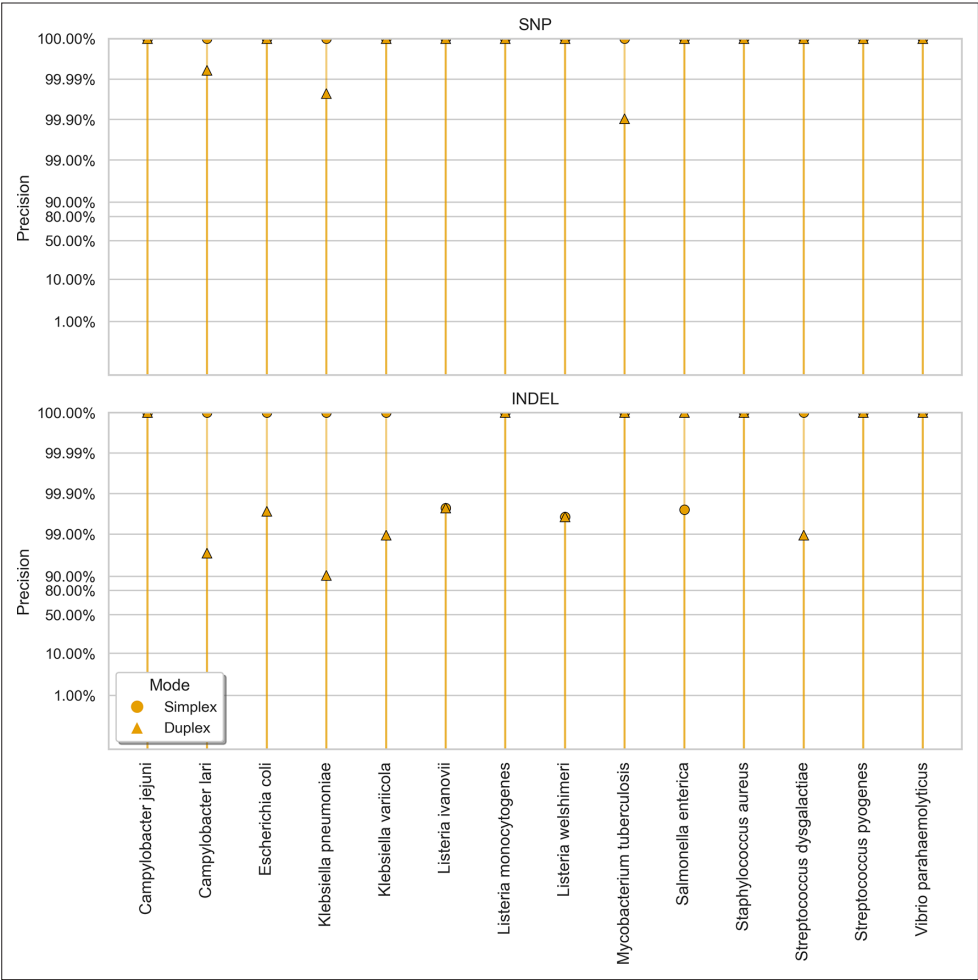


Figure 2—figure supplement 4. Clair3 sup model precision (y-axis) at the highest F1 score for each sample (x-axis), stratified by variant type (rows), and read type (shapes).

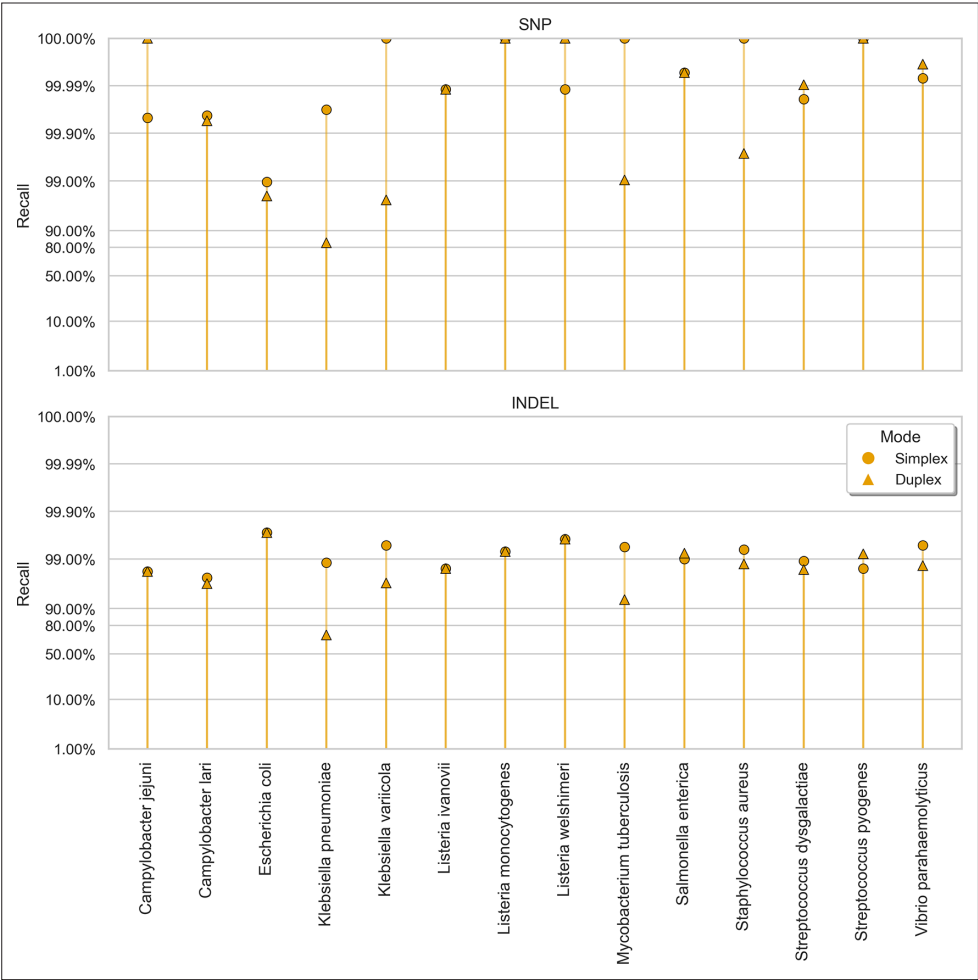


Figure 2—figure supplement 5. Clair3 sup model recall (y-axis) at the highest F1 score for each sample (x-axis), stratified by variant type (rows), and read type (shapes).

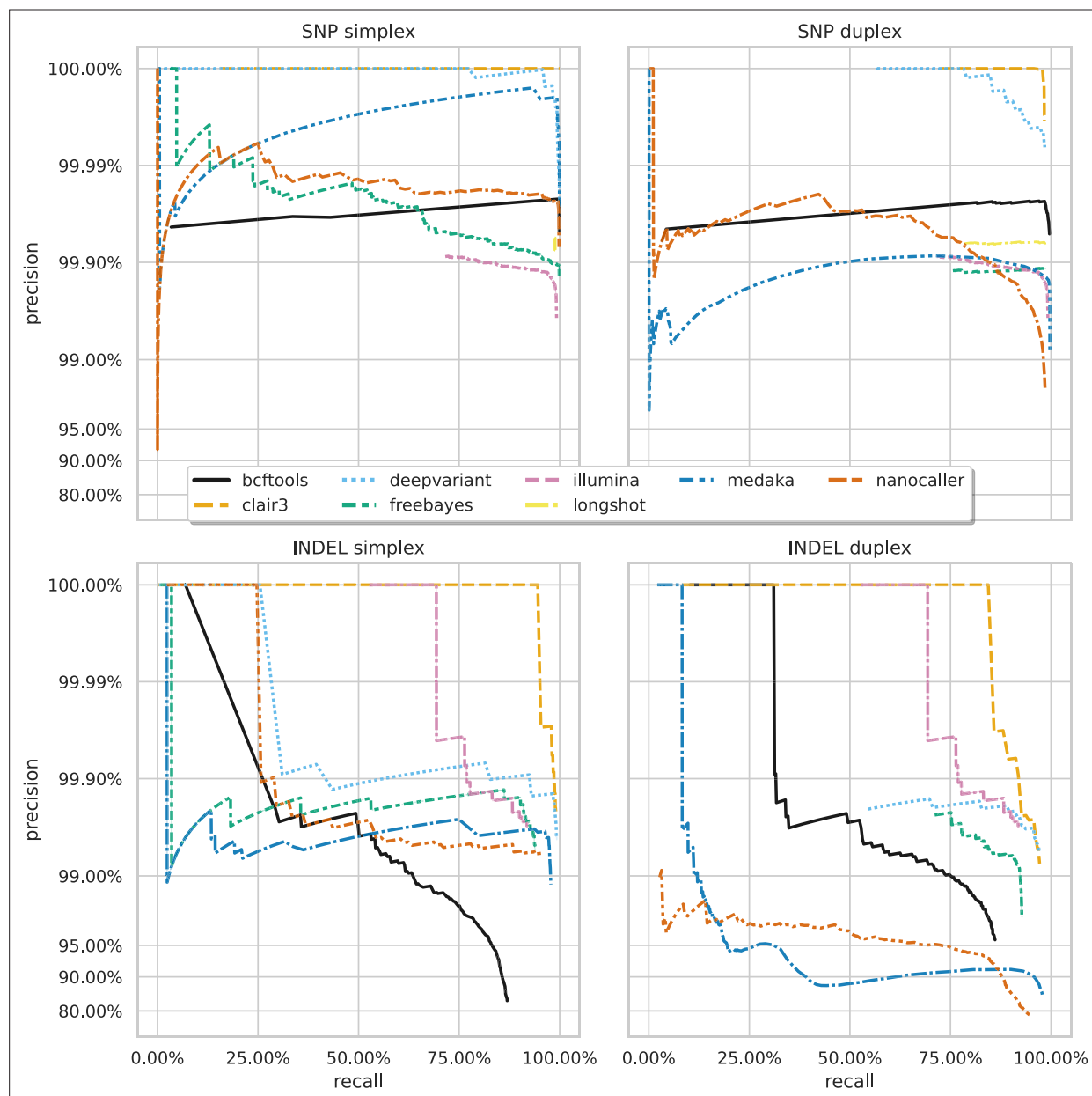


Figure 3. Precision and recall curves for each variant caller (colours and line styles) on sequencing data basecalled with the sup model, stratified by variant type (rows) and read type (columns) and aggregated across samples. The curves are generated by using increasing variant quality score thresholds to filter variants and calculating precision and recall at each threshold. The lowest threshold is the lower right part of the curve, moving to the highest at the top left. Note, Longshot does not provide indel calls.

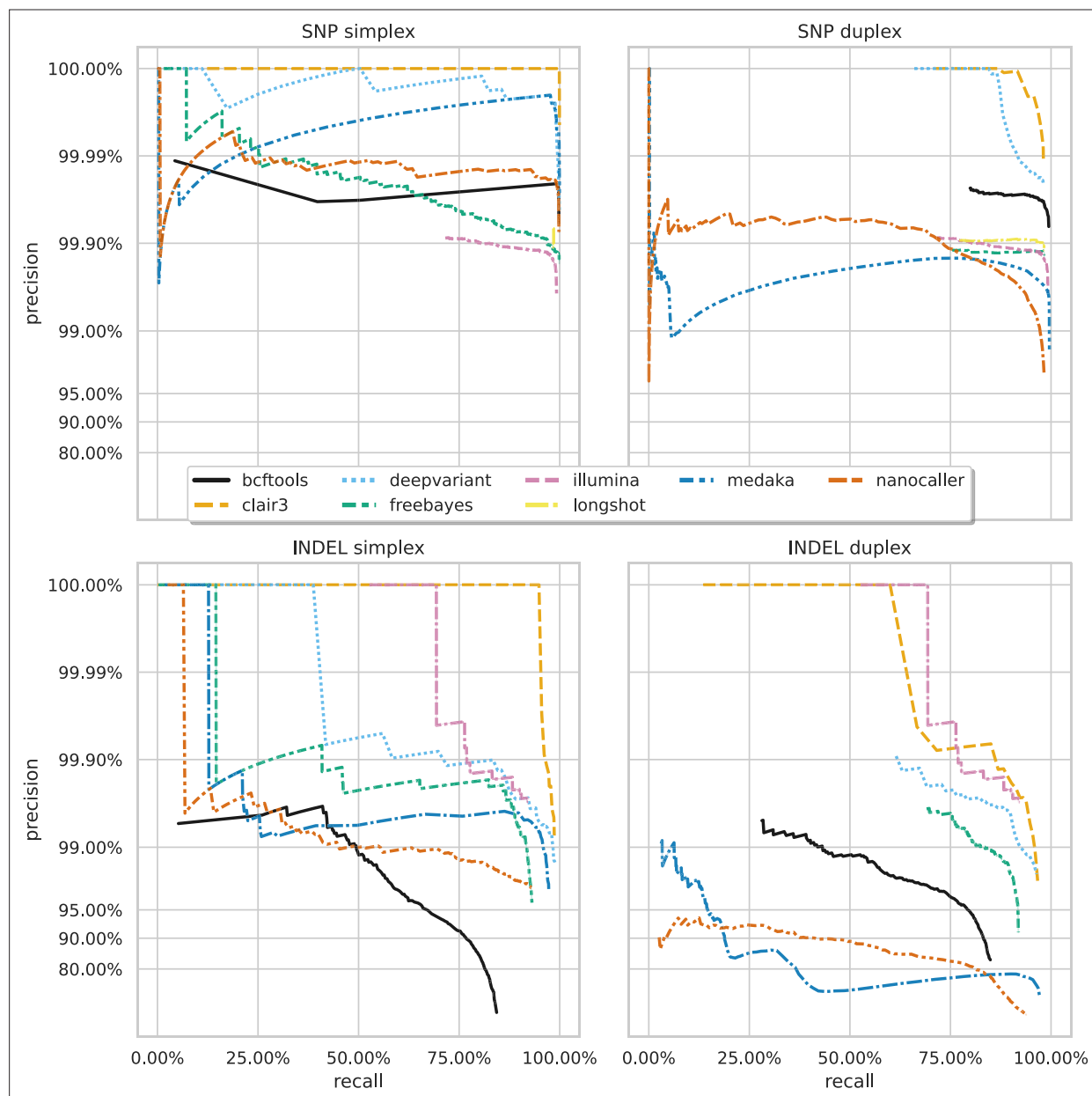


Figure 3—figure supplement 1. Precision and recall curves for each variant caller (colours and line styles) on sequencing data basecalled with the hac model, stratified by variant type (rows) and ready type (columns). The curves are generated by using increasing variant quality score thresholds to filter variants and calculating precision and recall at each threshold. The lowest threshold is the lower right part of the curve, moving to the highest at the top left. Note, Longshot does not provide indel calls.

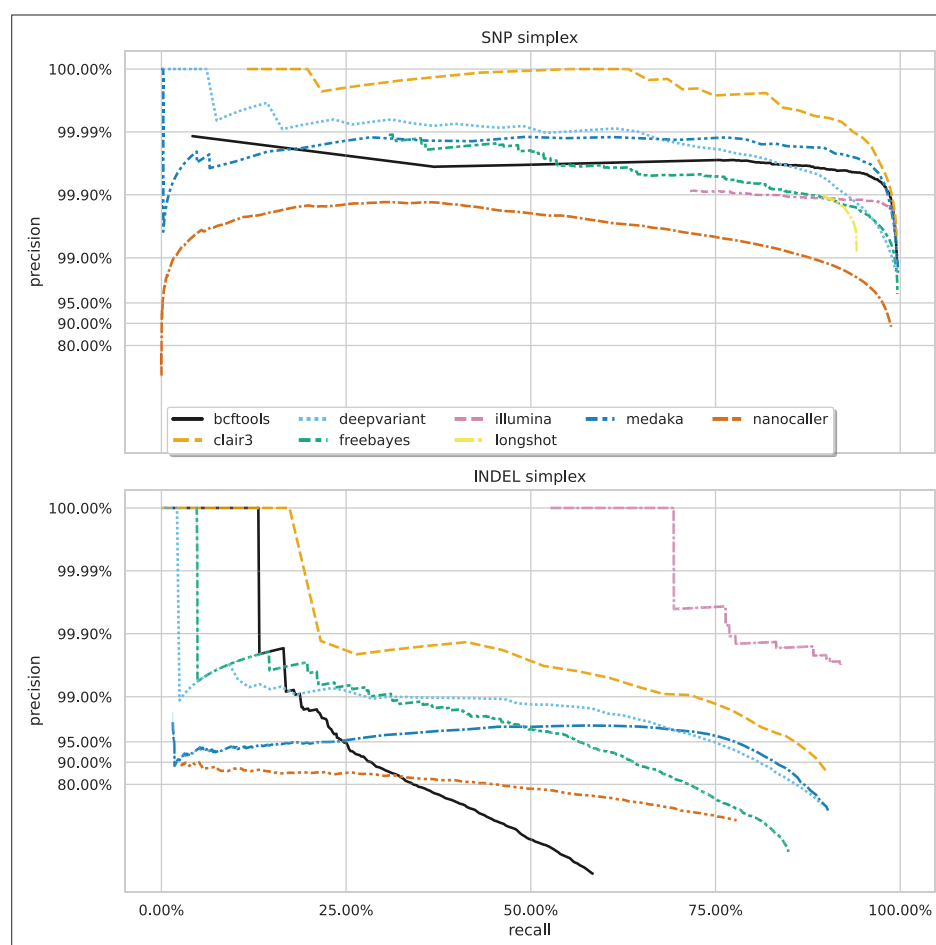


Figure 3—figure supplement 2. Precision and recall curves for each variant caller (colours and line styles) on sequencing data basecalled with the fast model, stratified by variant type (rows). The curves are generated by using increasing variant quality score thresholds to filter variants and calculating precision and recall at each threshold. The lowest threshold is the lower right part of the curve, moving to the highest at the top left. Note, Longshot does not provide indel calls.

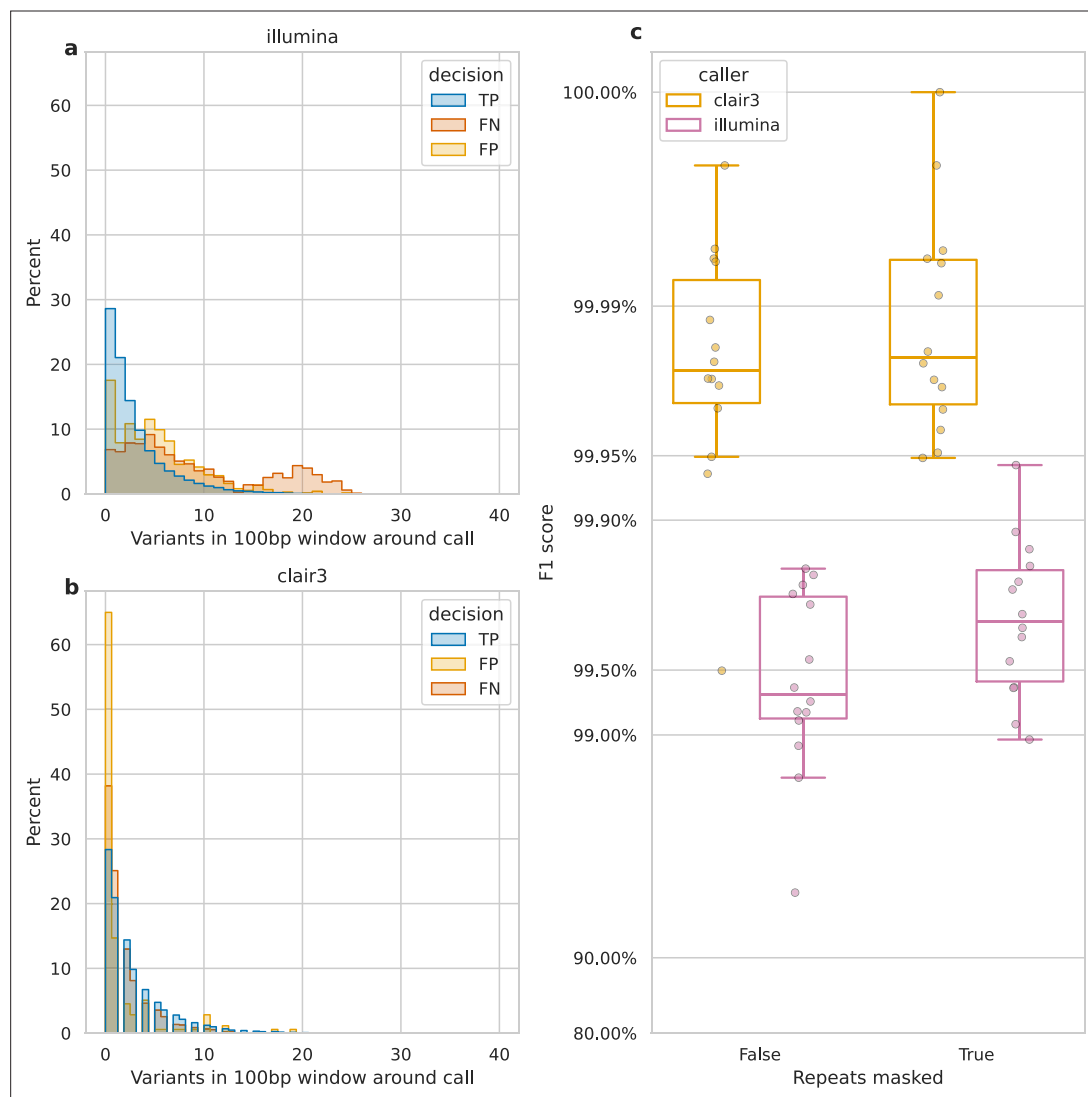


Figure 4. Impact of variant density and repetitive regions on Illumina variant calling. Variant density is the number of (true or false) variants in a 100 bp window centred on a call. **(a and b)** The distribution of variant densities for true positive (TP), false positive (FP), and false negative (FN) calls. The y-axis, percent, indicates the percent of all calls of that decision that fall within the density bin on the x-axis. Illumina calls, aggregated across all samples, are shown in a, while b shows Clair3 calls from simplex sup-basecalled reads at 100× depth. **(c)** Impact of repetitive regions on the F1 score (y-axis) for Clair3 (100× simplex sup) and Illumina. The x-axis indicates whether variants that fall within repetitive regions are excluded from the calculation of the F1 score. Points indicate the F1 score for a single sample.

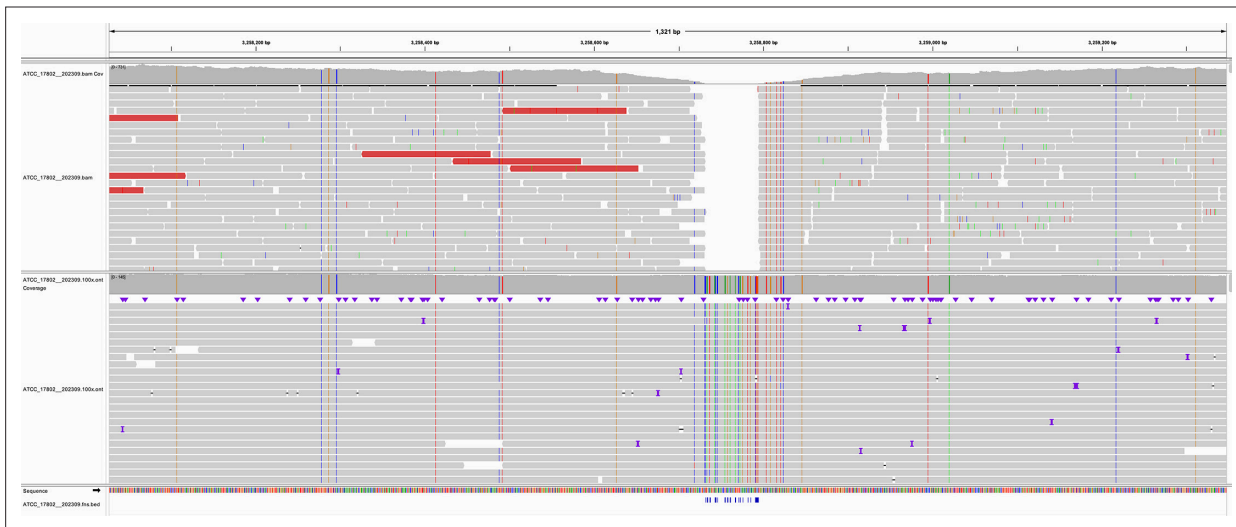


Figure 4—figure supplement 1. Read pileup in a variant-dense region in the genome of sample ATCC_17802_202309. The top track shows the alignment of the sample's Illumina reads, while the lower track is the Oxford Nanopore Technologies (ONT) reads. Missed variant calls (false negatives) are shown by small blue notches at the bottom of the figure, but are also identifiable by vertical coloured lines in the ONT reads. Pileup is visualised in IGV (*Robinson et al., 2011*).

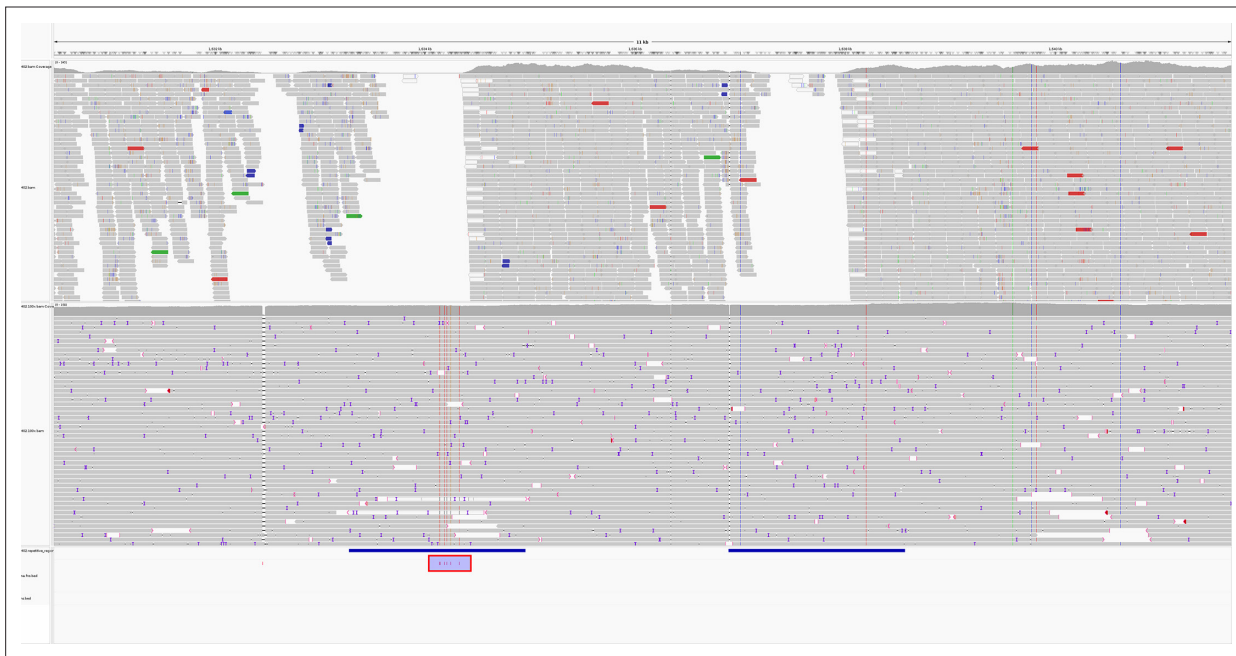


Figure 4—figure supplement 2. Read pileup around two repetitive regions (horizontal blue bars) in the genome of sample AMtb_1__202401. The top track shows the alignment of the sample's Illumina reads, while the lower track is the Oxford Nanopore Technologies (ONT) reads. Missed variant calls (false negatives) are shown by small red notches in a purple box with a red border at the bottom of the figure, but are also identifiable by vertical coloured lines in the ONT reads. Pileup is visualised in IGV ([Robinson et al., 2011](#)).

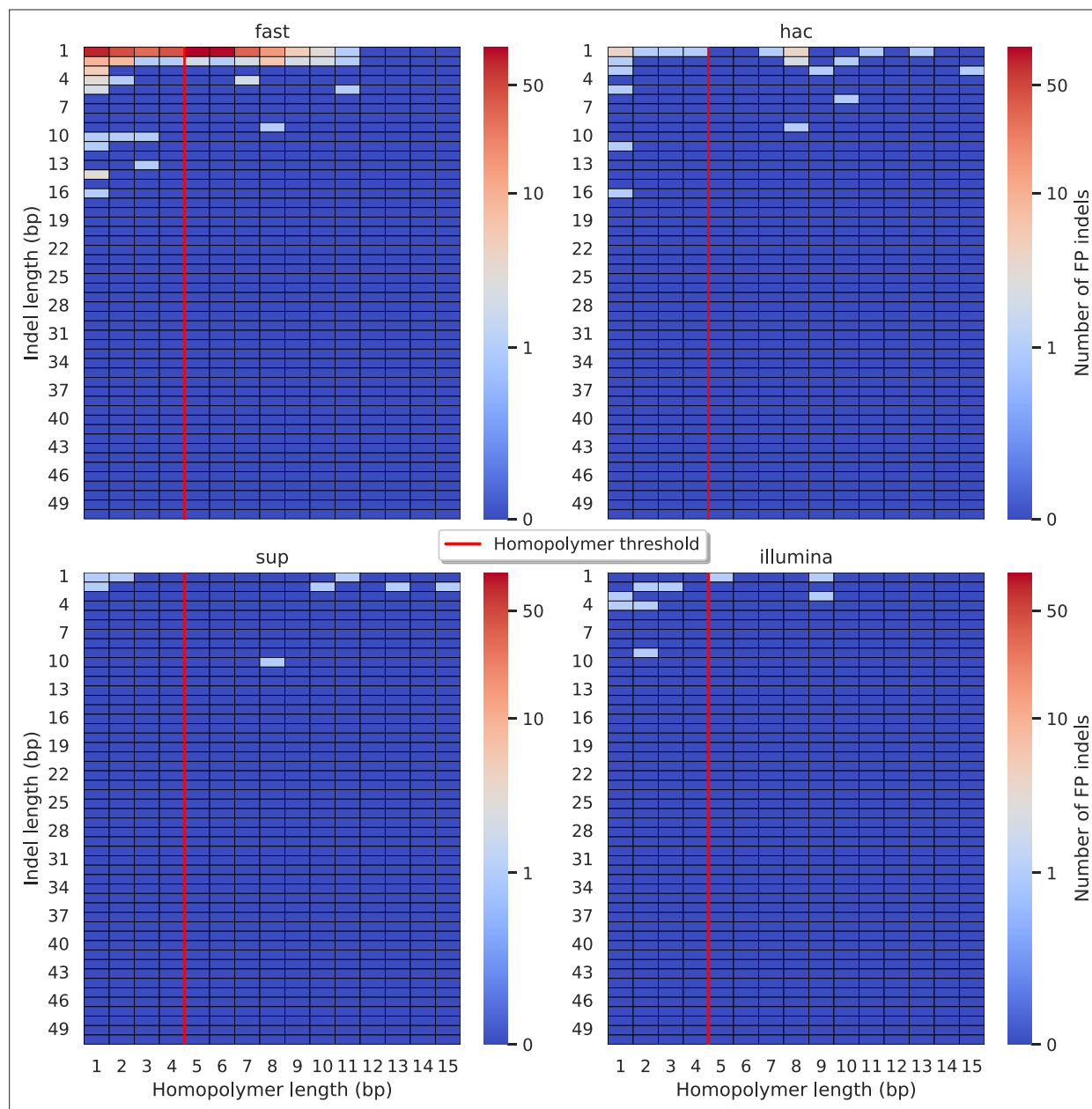


Figure 5. Relationship between indel length (y-axis) and homopolymer length (x-axis) for false positive (FP) indel calls for Clair3 100× simplex fast (top left), hac (top right), and sup (lower left) calls. Illumina is shown in the lower right for reference. The vertical red line indicates the threshold above which we deem a run of the same nucleotide to be a 'true' homopolymer. Indel length is the number of bases inserted/deleted for an indel, whereas the homopolymer length indicates how long the tract of the same nucleotide is after the indel. The colour of a cell indicates number of FP indels of that indel-homopolymer length combination.

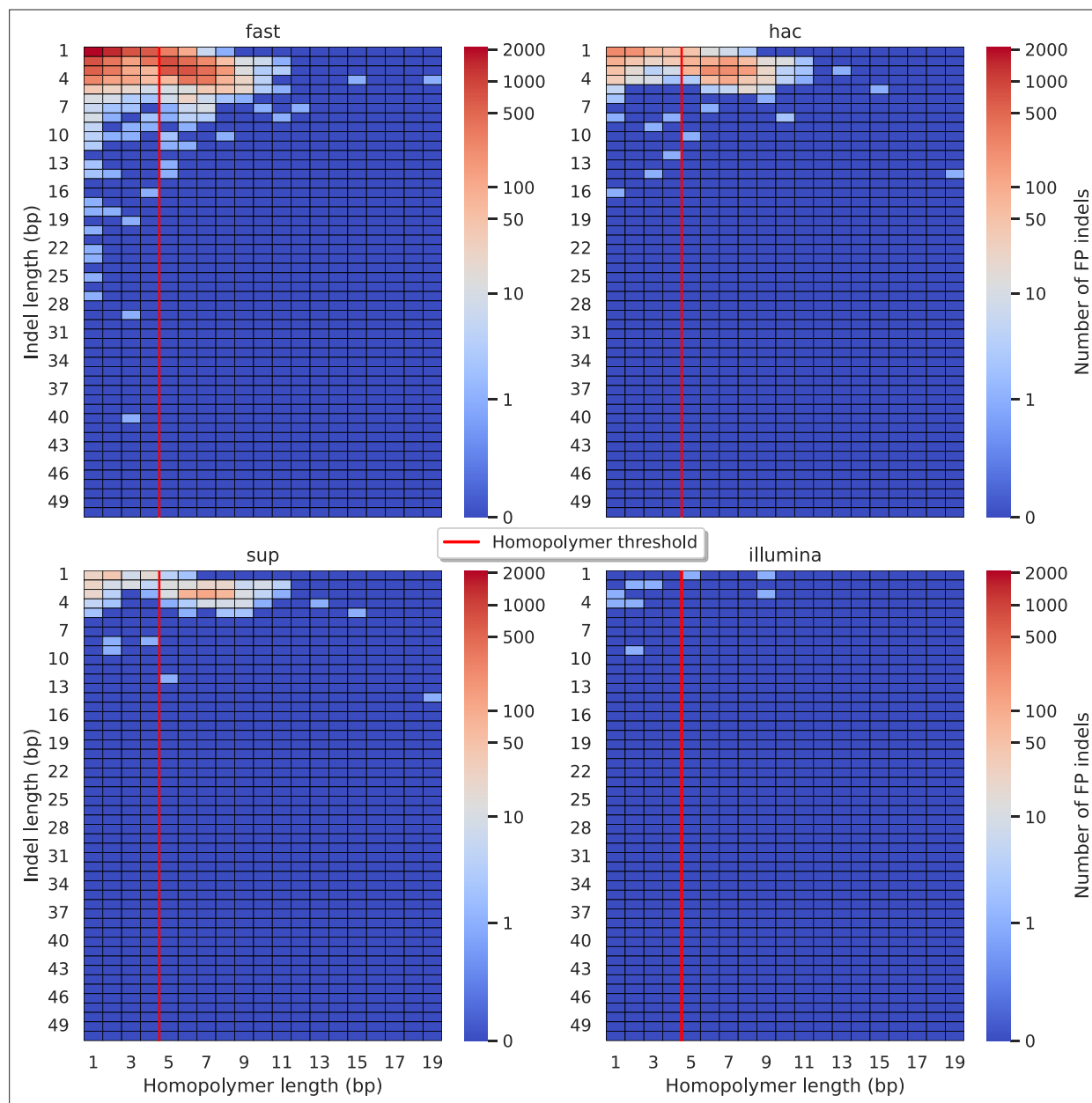


Figure 5—figure supplement 1. Relationship between indel length (x-axis) and homopolymer length (y-axis) for false positive (FP) indel calls for BCFtools 100× simplex fast (top left), hac (top right), and sup (lower left) calls. Illumina is shown in the lower right for reference. The vertical red line indicates the threshold above which we deem a run of the same nucleotide to be a 'true' homopolymer. Indel length is the number of bases inserted/deleted for an indel, whereas the homopolymer length indicates how long the tract of the same nucleotide is after the indel. The colour of a cell indicates number of FP indels of that indel-homopolymer length combination.

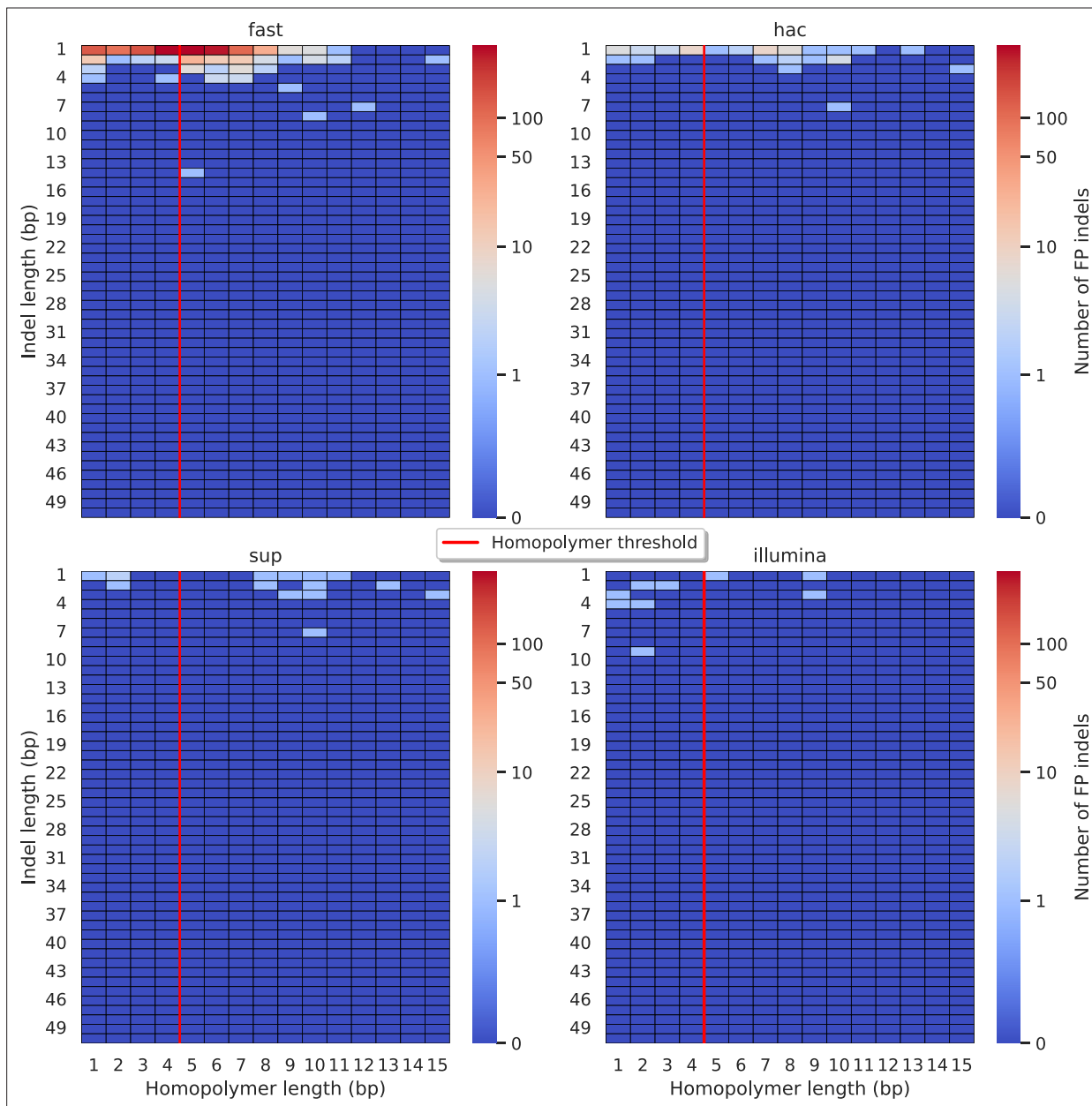


Figure 5—figure supplement 2. Relationship between indel length (x-axis) and homopolymer length (y-axis) for false positive (FP) indel calls for DeepVariant 100× simplex fast (top left), hac (top right), and sup (lower left) calls. Illumina is shown in the lower right for reference. The vertical red line indicates the threshold above which we deem a run of the same nucleotide to be a 'true' homopolymer. Indel length is the number of bases inserted/deleted for an indel, whereas the homopolymer length indicates how long the tract of the same nucleotide is after the indel. The colour of a cell indicates number of FP indels of that indel-homopolymer length combination.

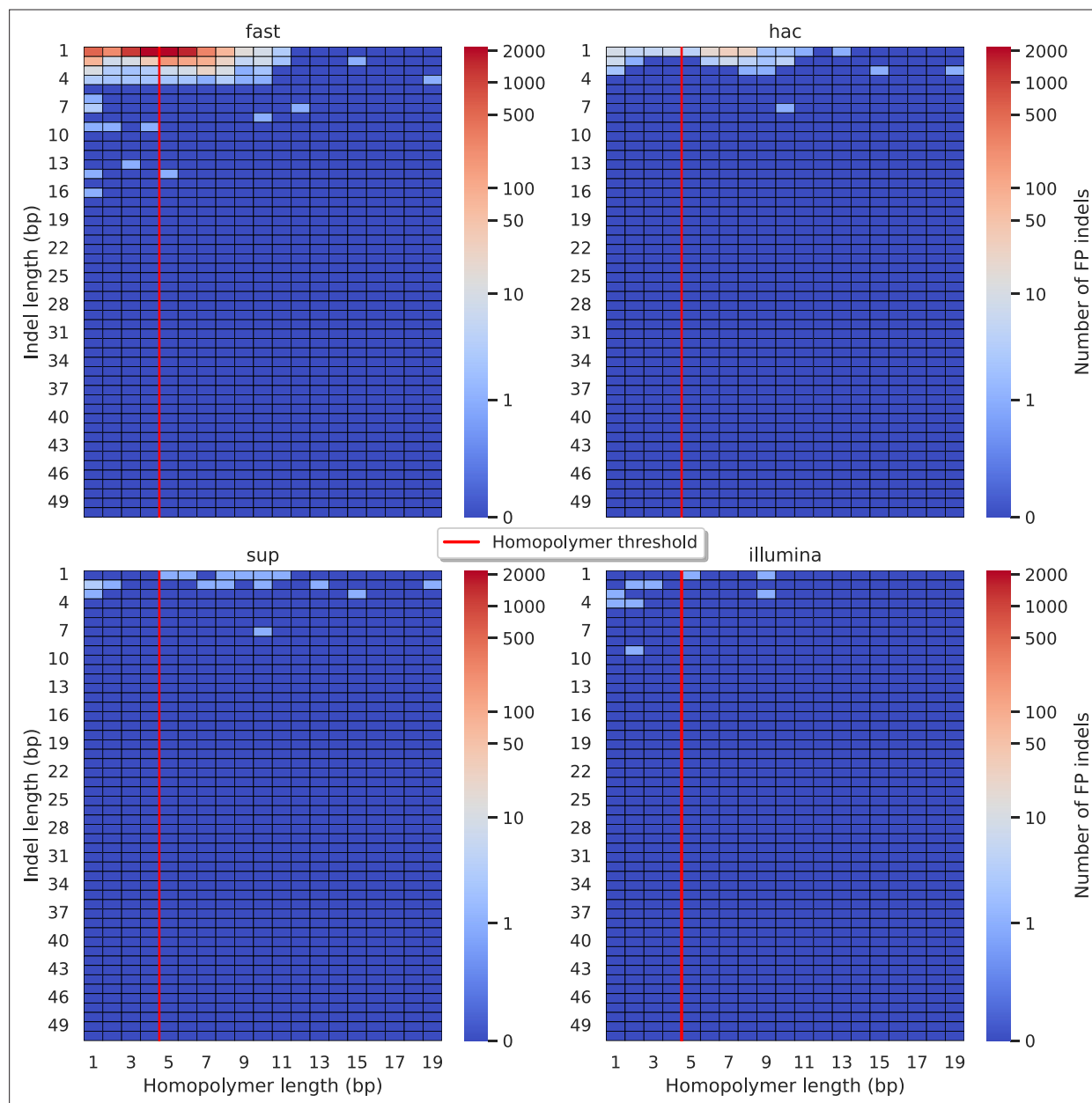


Figure 5—figure supplement 3. Relationship between indel length (x-axis) and homopolymer length (y-axis) for false positive (FP) indel calls for FreeBayes 100× simplex fast (top left), hac (top right), and sup (lower left) calls. Illumina is shown in the lower right for reference. The vertical red line indicates the threshold above which we deem a run of the same nucleotide to be a 'true' homopolymer. Indel length is the number of bases inserted/deleted for an indel, whereas the homopolymer length indicates how long the tract of the same nucleotide is after the indel. The colour of a cell indicates number of FP indels of that indel-homopolymer length combination.

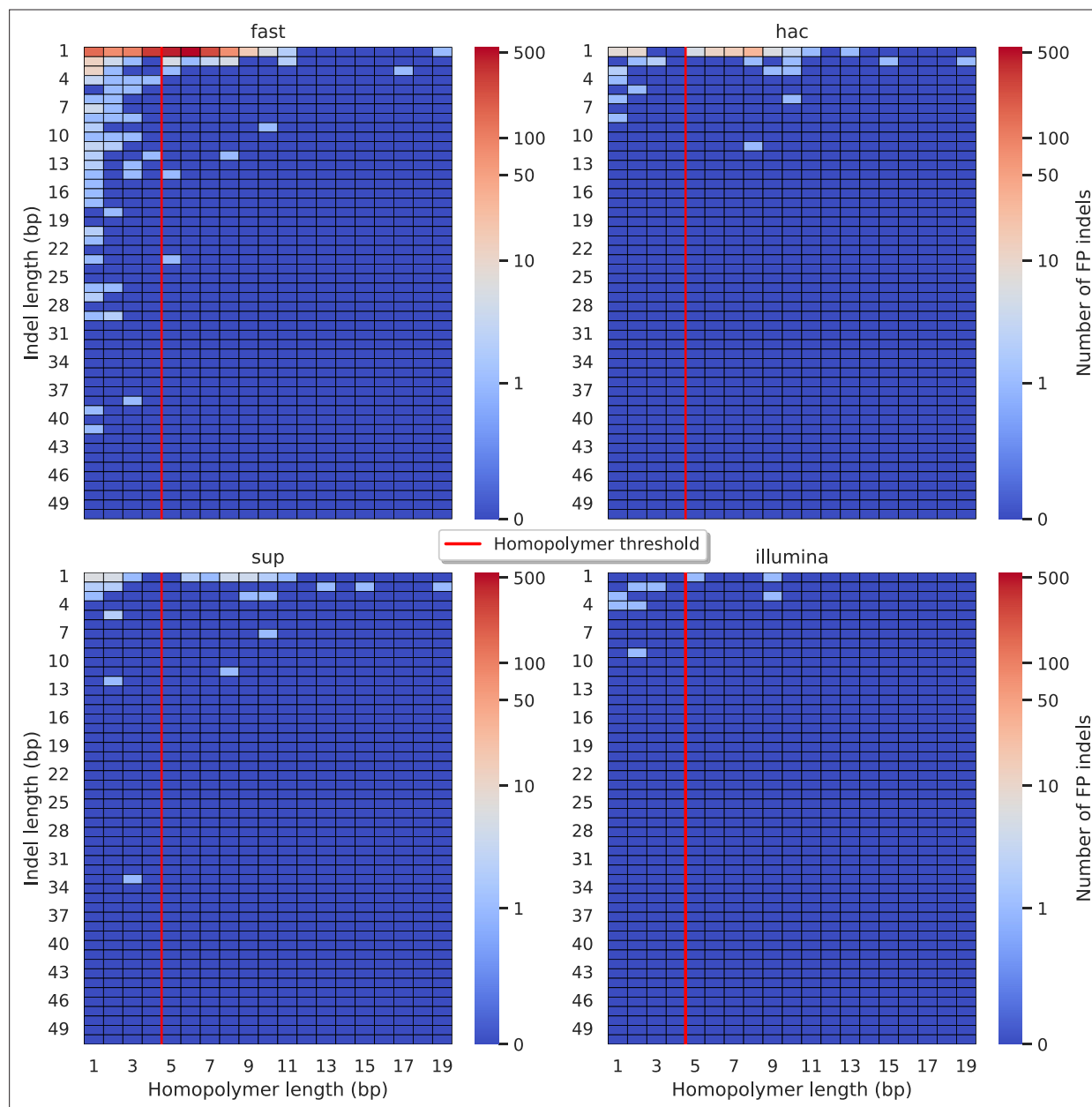


Figure 5—figure supplement 4. Relationship between indel length (x-axis) and homopolymer length (y-axis) for false positive (FP) indel calls for Medaka 100× simplex fast (top left), hac (top right), and sup (lower left) calls. Illumina is shown in the lower right for reference. The vertical red line indicates the threshold above which we deem a run of the same nucleotide to be a 'true' homopolymer. Indel length is the number of bases inserted/deleted for an indel, whereas the homopolymer length indicates how long the tract of the same nucleotide is after the indel. The colour of a cell indicates number of FP indels of that indel-homopolymer length combination.

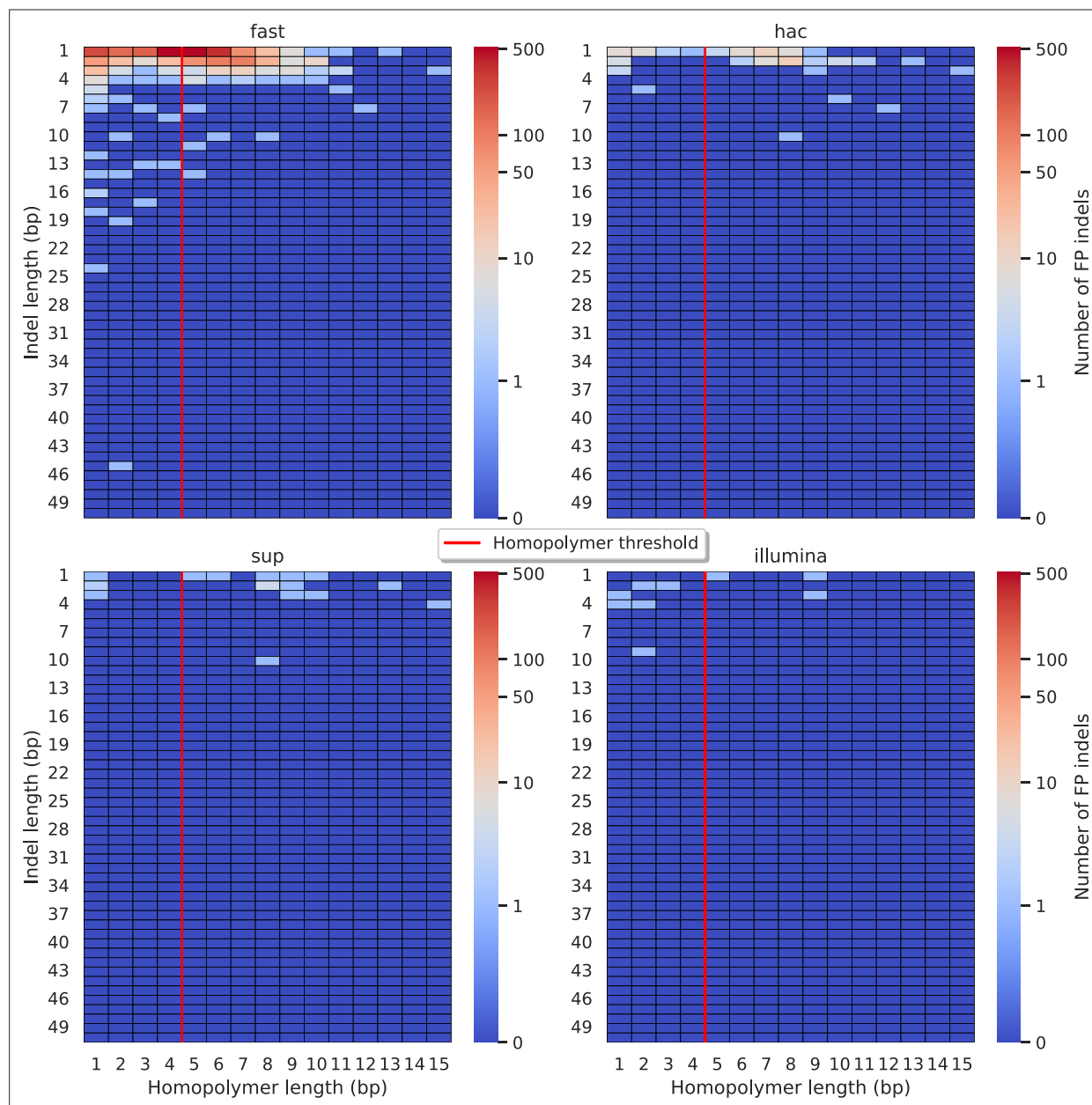


Figure 5—figure supplement 5. Relationship between indel length (x-axis) and homopolymer length (y-axis) for false positive (FP) indel calls for NanoCatcher 100× simplex fast (top left), hac (top right), and sup (lower left) calls. Illumina is shown in the lower right for reference. The vertical red line indicates the threshold above which we deem a run of the same nucleotide to be a 'true' homopolymer. Indel length is the number of bases inserted/deleted for an indel, whereas the homopolymer length indicates how long the tract of the same nucleotide is after the indel. The colour of a cell indicates number of FP indels of that indel-homopolymer length combination.

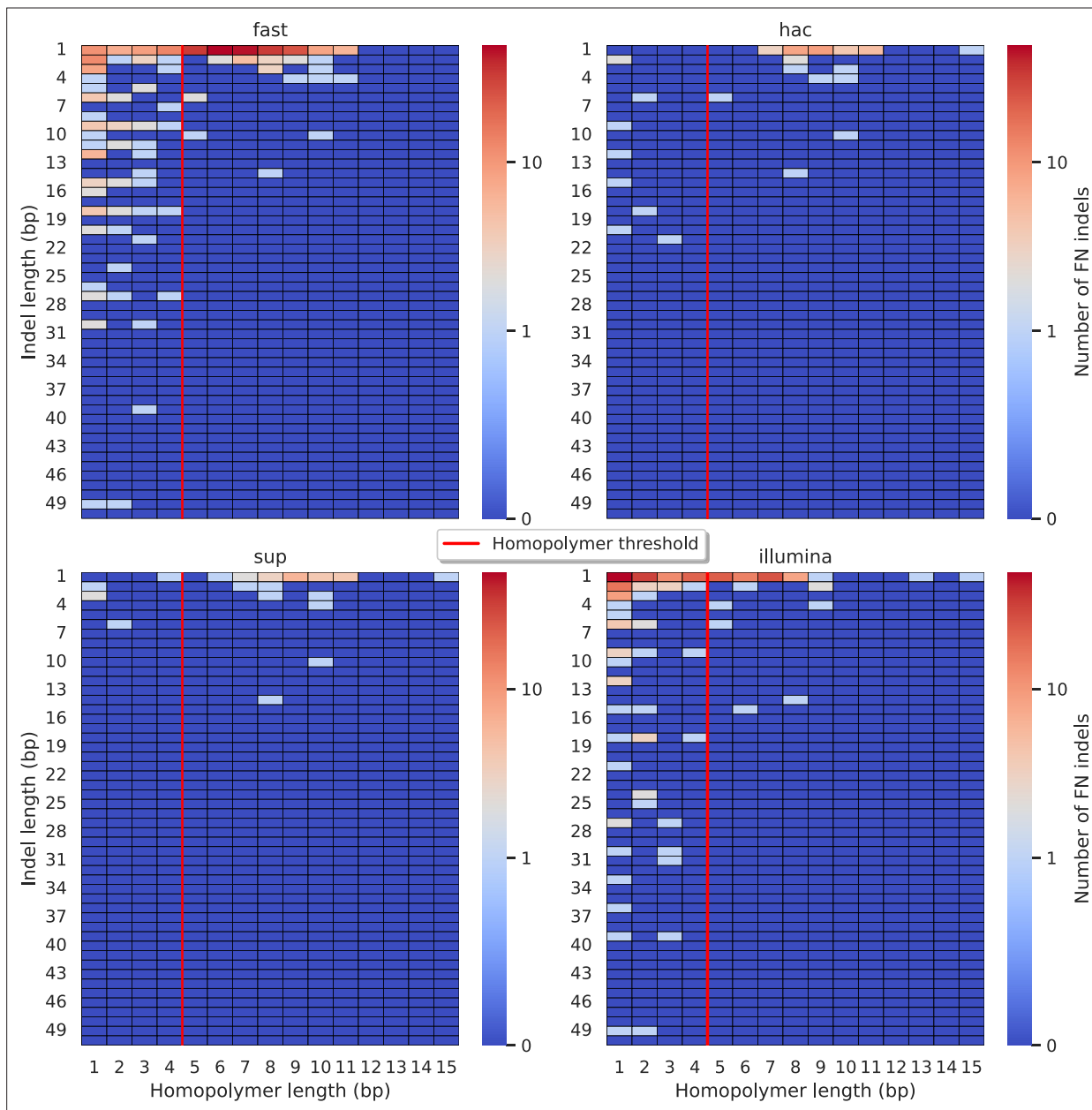


Figure 5—figure supplement 6. Relationship between indel length (x-axis) and homopolymer length (y-axis) for false negative (FN) indel calls for Clair3 100× simplex fast (top left), hac (top right), and sup (lower left) calls. Illumina is shown in the lower right for reference. The vertical red line indicates the threshold above which we deem a run of the same nucleotide to be a ‘true’ homopolymer. Indel length is the number of bases inserted/deleted for an indel, whereas the homopolymer length indicates how long the tract of the same nucleotide is after the indel. The colour of a cell indicates number of FN indels of that indel-homopolymer length combination.

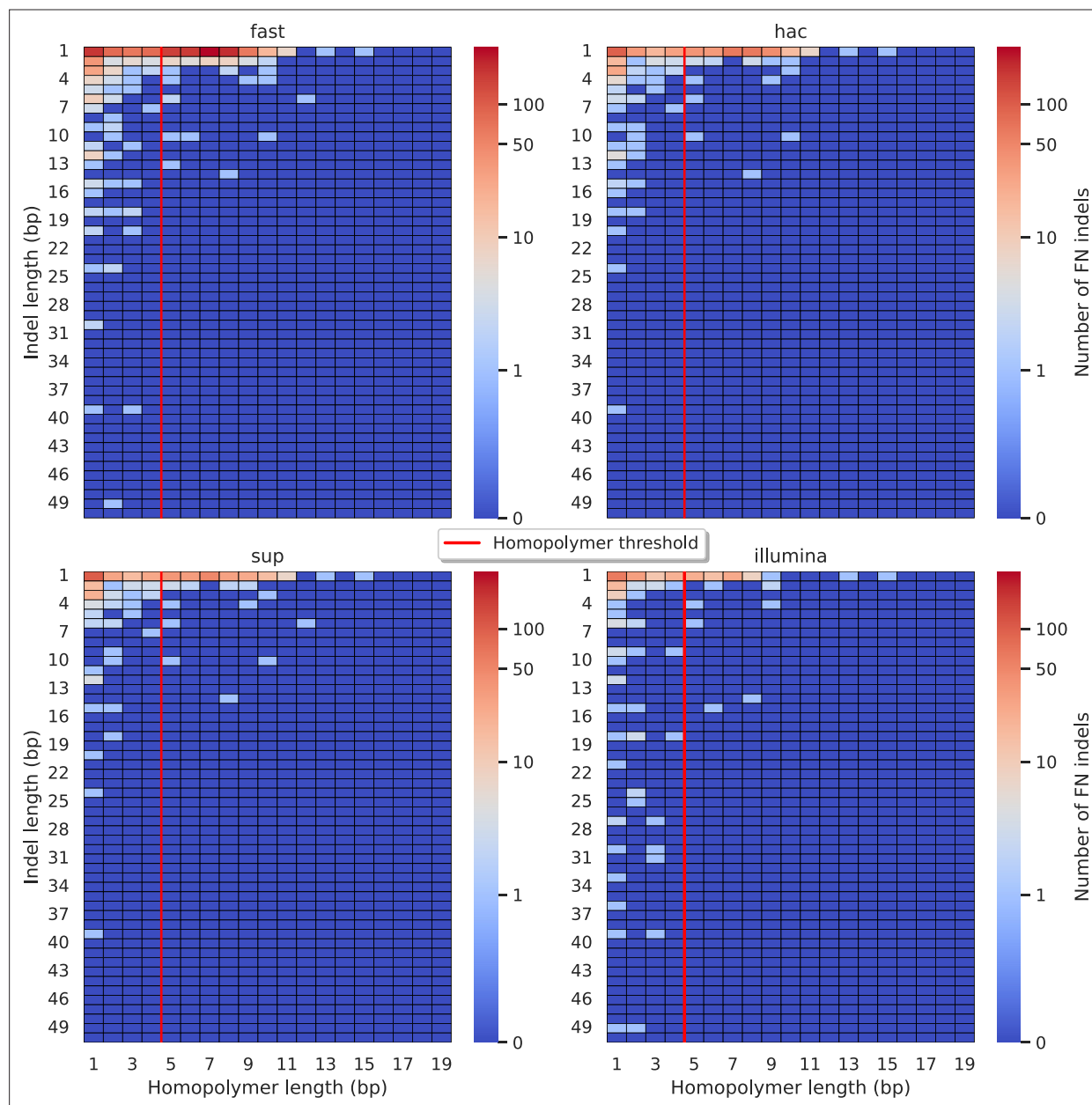


Figure 5—figure supplement 7. Relationship between indel length (x-axis) and homopolymer length (y-axis) for false negative (FN) indel calls for BCFtools 100× simplex fast (top left), hac (top right), and sup (lower left) calls. Illumina is shown in the lower right for reference. The vertical red line indicates the threshold above which we deem a run of the same nucleotide to be a 'true' homopolymer. Indel length is the number of bases inserted/deleted for an indel, whereas the homopolymer length indicates how long the tract of the same nucleotide is after the indel. The colour of a cell indicates number of FN indels of that indel-homopolymer length combination.

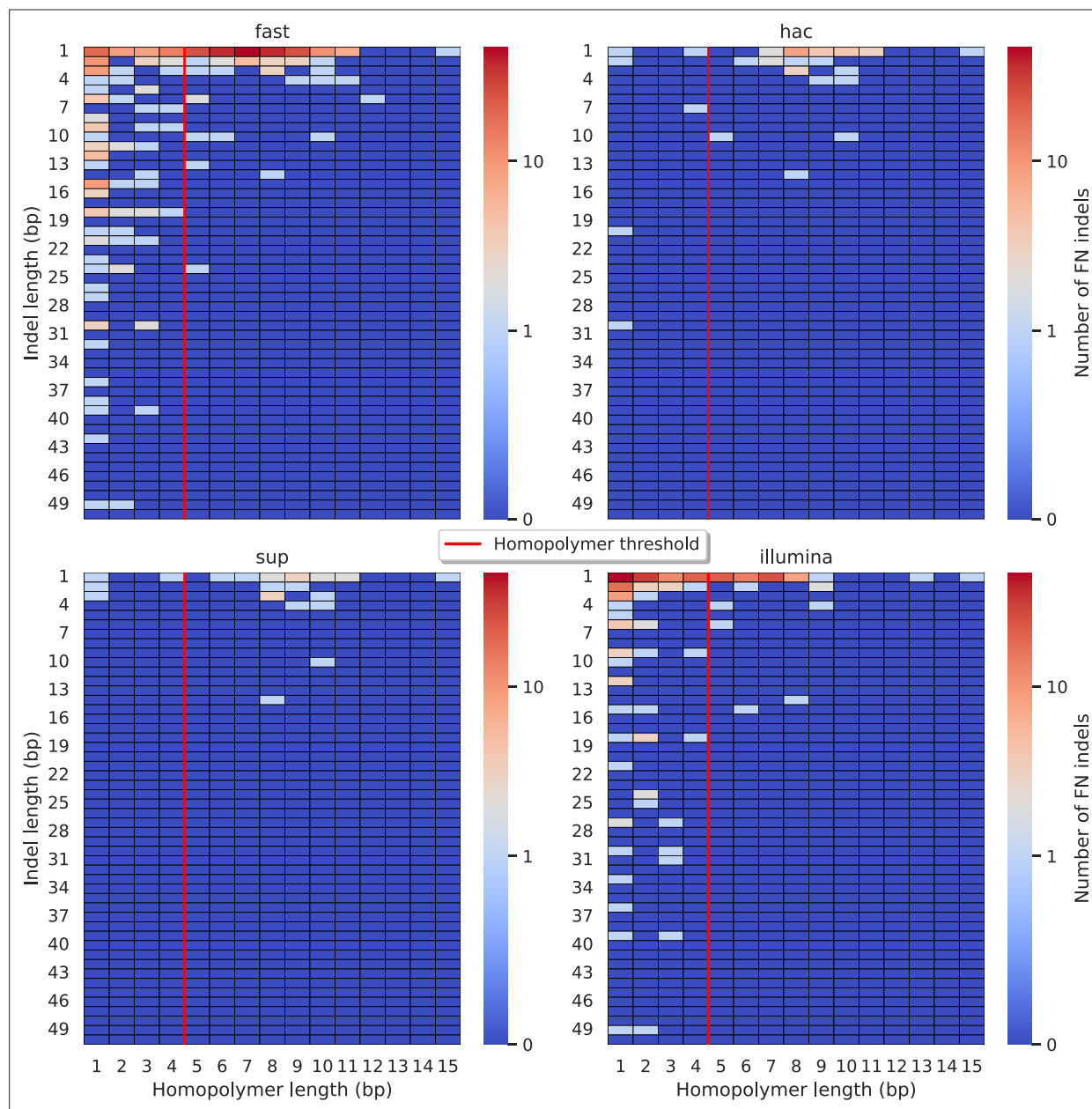


Figure 5—figure supplement 8. Relationship between indel length (x-axis) and homopolymer length (y-axis) for false negative (FN) indel calls for DeepVariant 100× simplex fast (top left), hac (top right), and sup (lower left) calls. Illumina is shown in the lower right for reference. The vertical red line indicates the threshold above which we deem a run of the same nucleotide to be a 'true' homopolymer. Indel length is the number of bases inserted/deleted for an indel, whereas the homopolymer length indicates how long the tract of the same nucleotide is after the indel. The colour of a cell indicates number of FN indels of that indel-homopolymer length combination.

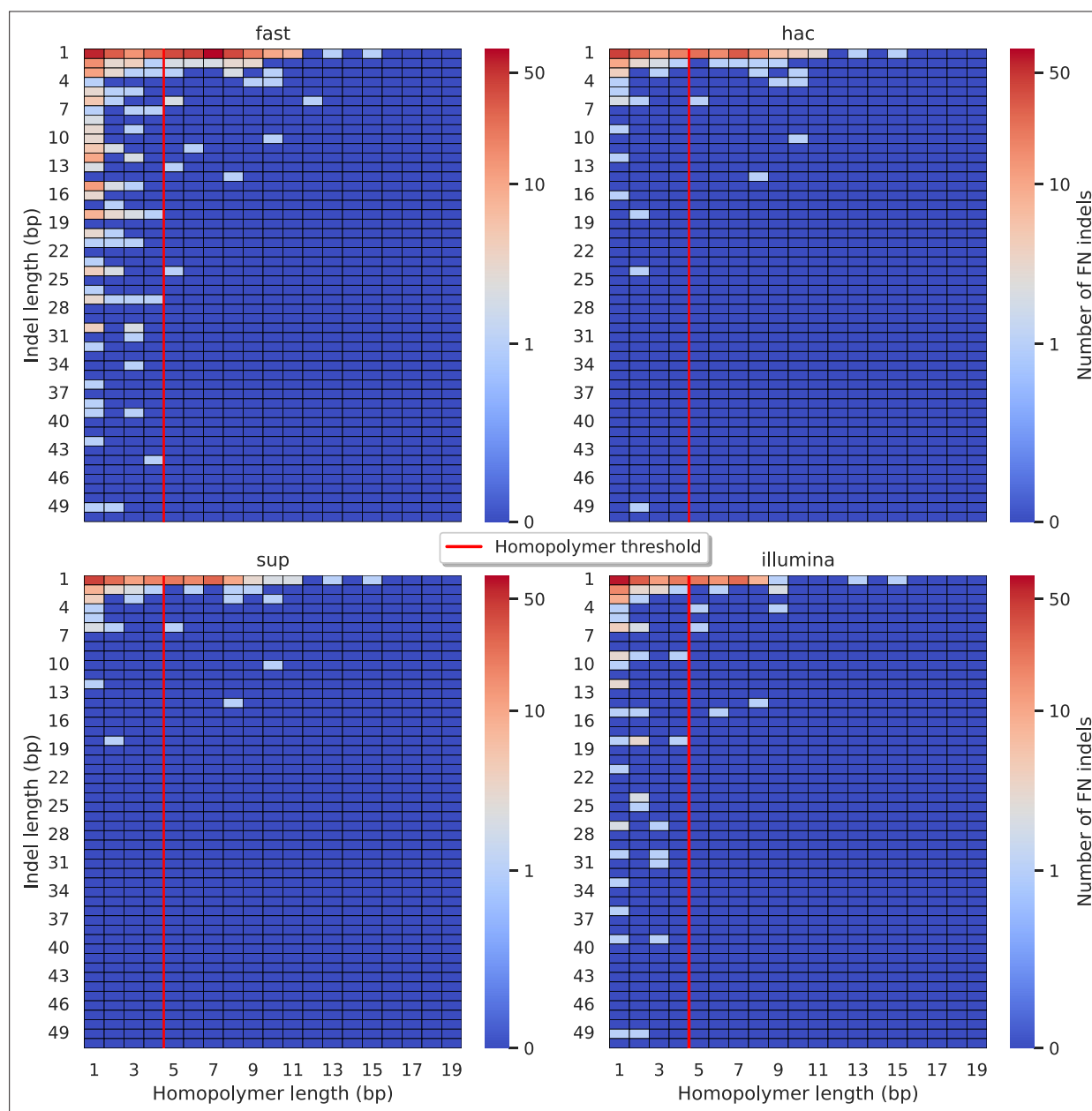


Figure 5—figure supplement 9. Relationship between indel length (x-axis) and homopolymer length (y-axis) for false negative (FN) indel calls for FreeBayes 100× simplex fast (top left), hac (top right), and sup (lower left) calls. Illumina is shown in the lower right for reference. The vertical red line indicates the threshold above which we deem a run of the same nucleotide to be a 'true' homopolymer. Indel length is the number of bases inserted/deleted for an indel, whereas the homopolymer length indicates how long the tract of the same nucleotide is after the indel. The colour of a cell indicates number of FN indels of that indel-homopolymer length combination.

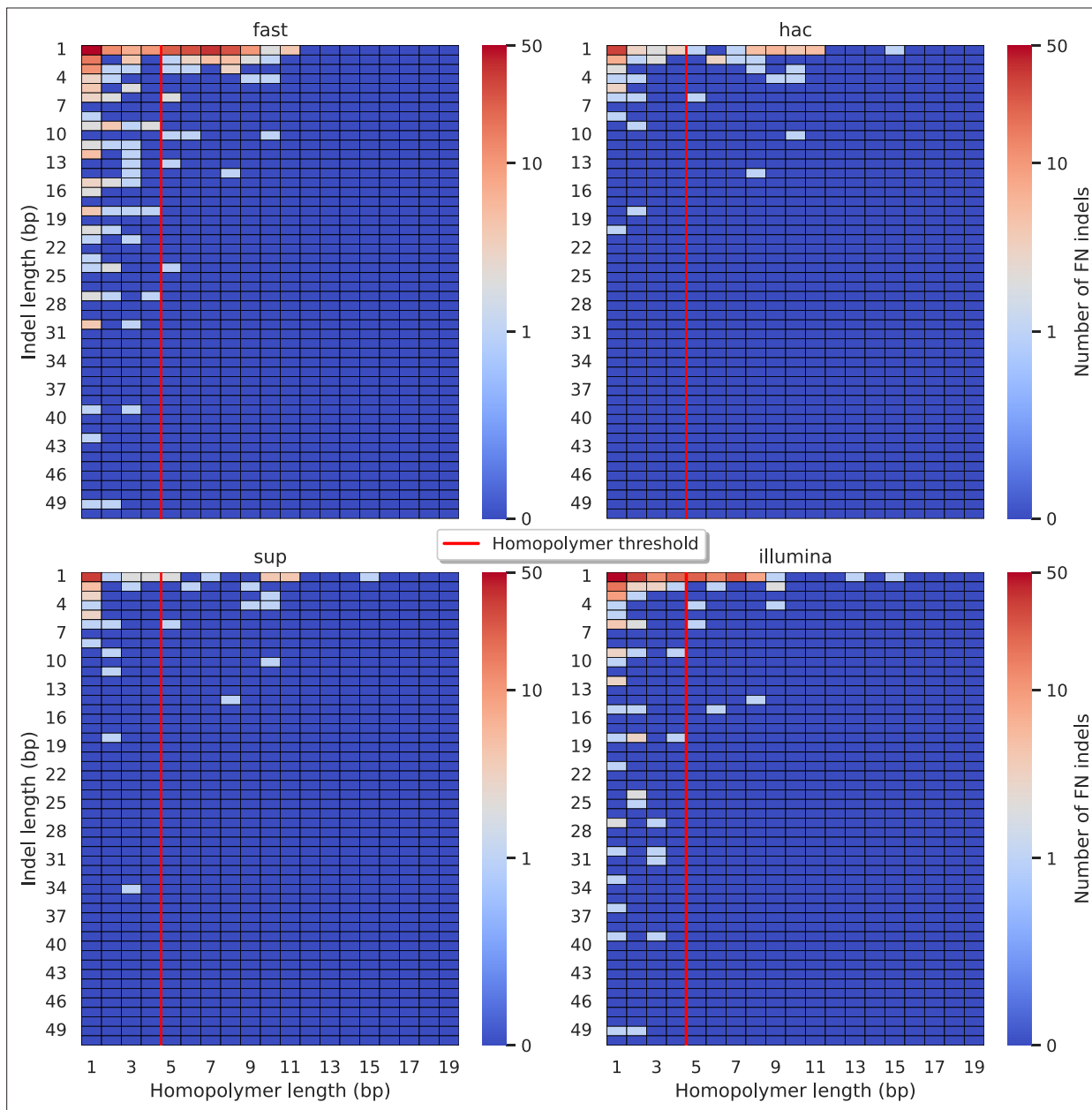


Figure 5—figure supplement 10. Relationship between indel length (x-axis) and homopolymer length (y-axis) for false negative (FN) indel calls for Medaka 100× simplex fast (top left), hac (top right), and sup (lower left) calls. Illumina is shown in the lower right for reference. The vertical red line indicates the threshold above which we deem a run of the same nucleotide to be a 'true' homopolymer. Indel length is the number of bases inserted/deleted for an indel, whereas the homopolymer length indicates how long the tract of the same nucleotide is after the indel. The colour of a cell indicates number of FN indels of that indel-homopolymer length combination.

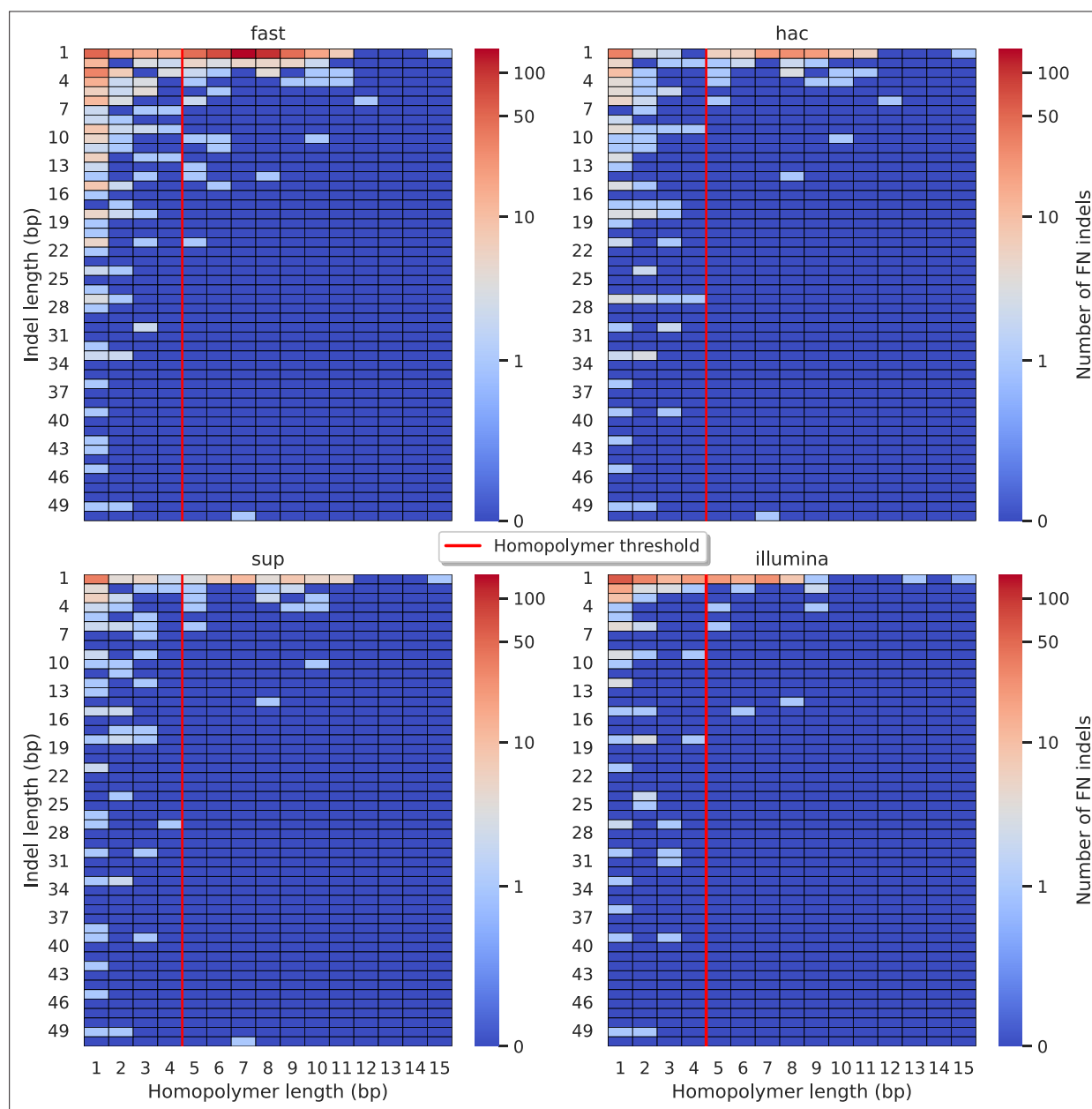


Figure 5—figure supplement 11. Relationship between indel length (x-axis) and homopolymer length (y-axis) for false negative (FN) indel calls for NanoCaller 100× simplex fast (top left), hac (top right), and sup (lower left) calls. Illumina is shown in the lower right for reference. The vertical red line indicates the threshold above which we deem a run of the same nucleotide to be a 'true' homopolymer. Indel length is the number of bases inserted/deleted for an indel, whereas the homopolymer length indicates how long the tract of the same nucleotide is after the indel. The colour of a cell indicates number of FN indels of that indel-homopolymer length combination.

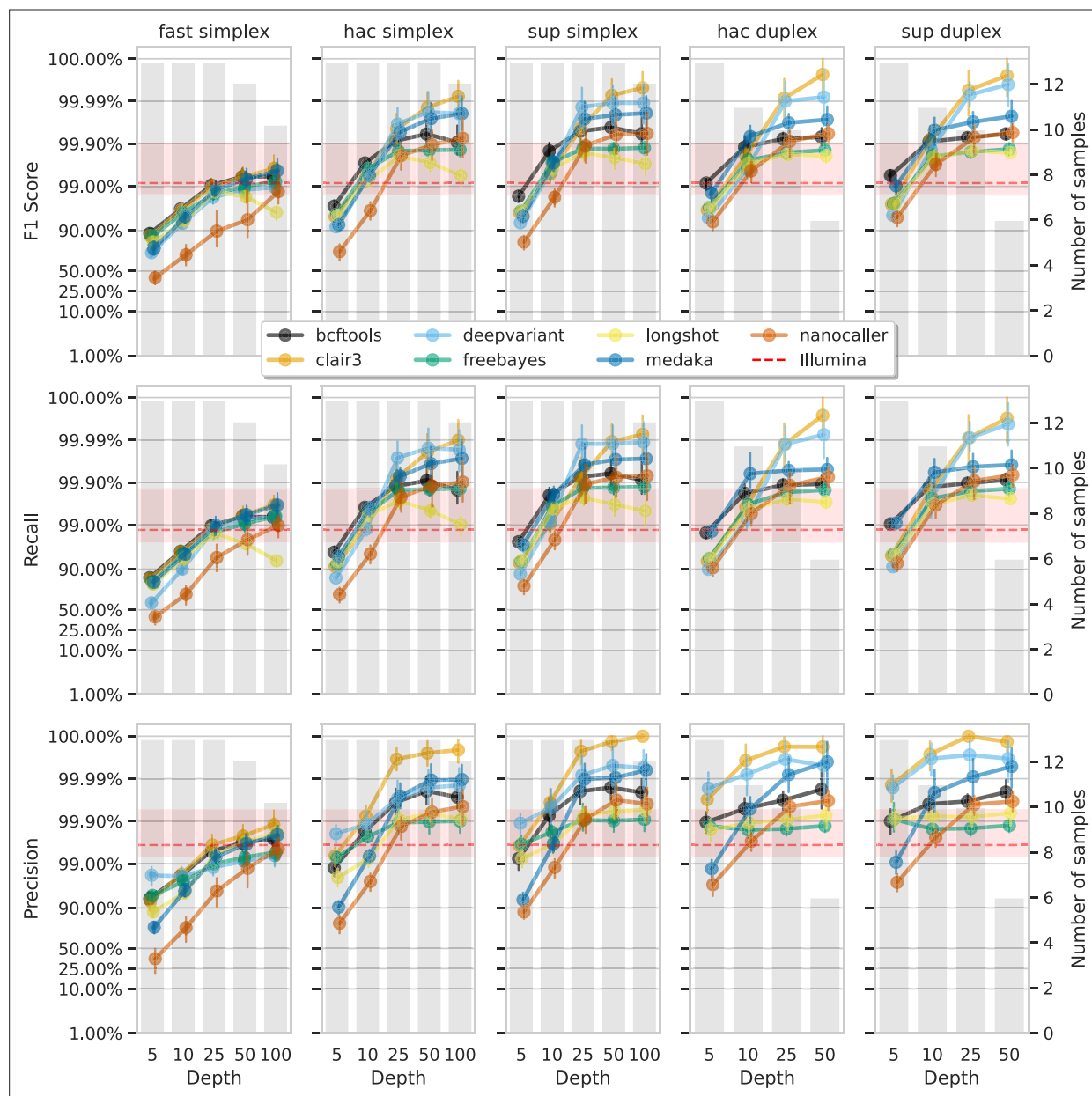


Figure 6. Effect of read depth (x-axis) on the highest SNP F1 score, and precision and recall at that F1 score (y-axis), for each variant caller (colours). Each column is a basecall model and read type combination. The grey bars indicate the number of samples with at least that much read depth in the full read set. Samples with less than that depth were not used to calculate that depth's metrics. Bars on each point at each depth depict the 95% confidence interval. The horizontal red dashed line is the full-depth Illumina value for that metric, with the red bands indicating the 95% confidence interval.

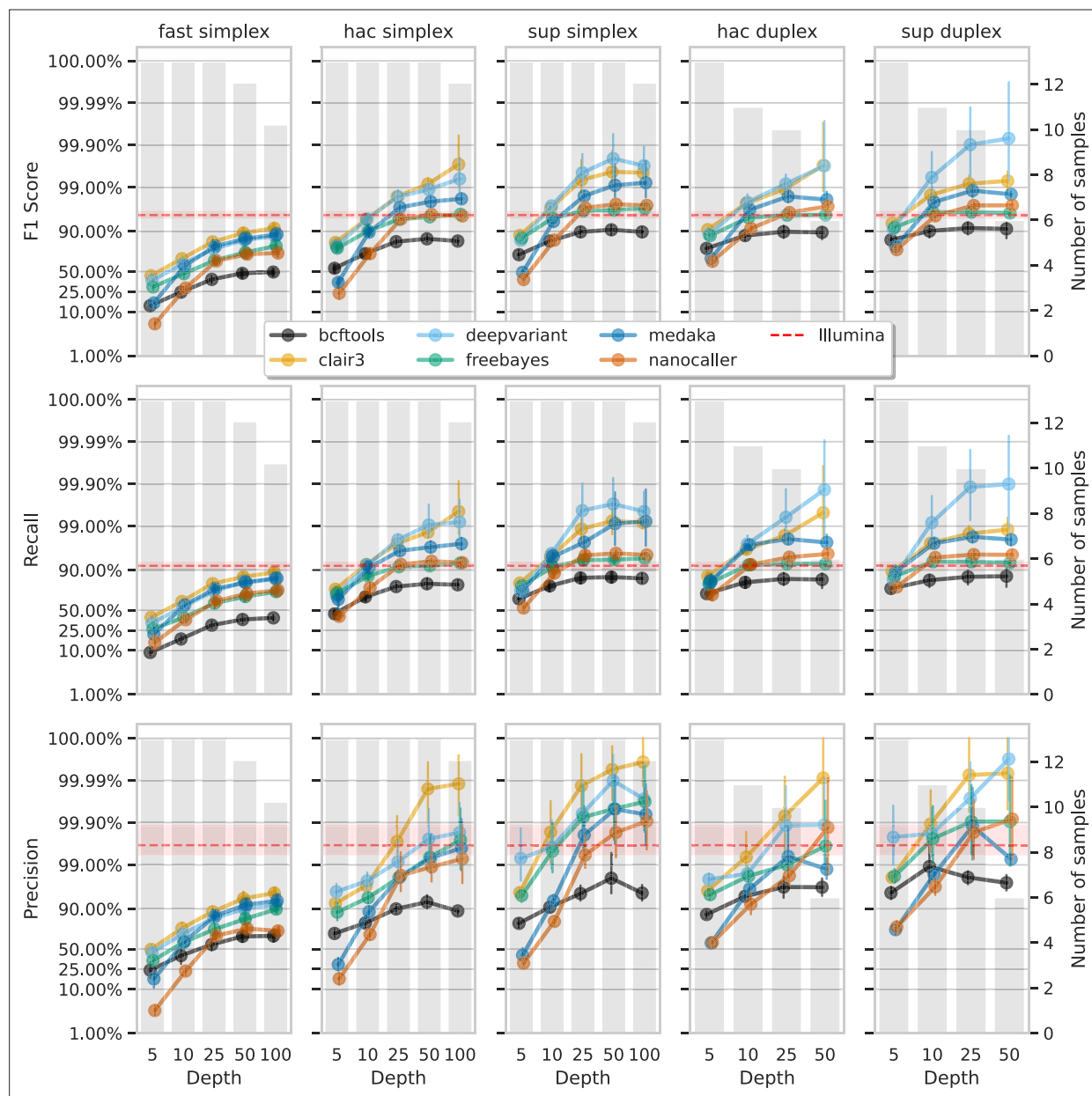


Figure 7. Effect of read depth (x-axis) on the highest indel F1 score, and precision and recall at that F1 score (y-axis), for each variant caller (colours). Each column is a basecall model and read type combination. The grey bars indicate the number of samples with at least that much read depth in the full read set. Samples with less than that depth were not used to calculate that depth's metrics. Bars on each point at each depth depict the 95% confidence interval. The horizontal red dashed line is the full-depth Illumina value for that metric, with the red bands indicating the 95% confidence interval.

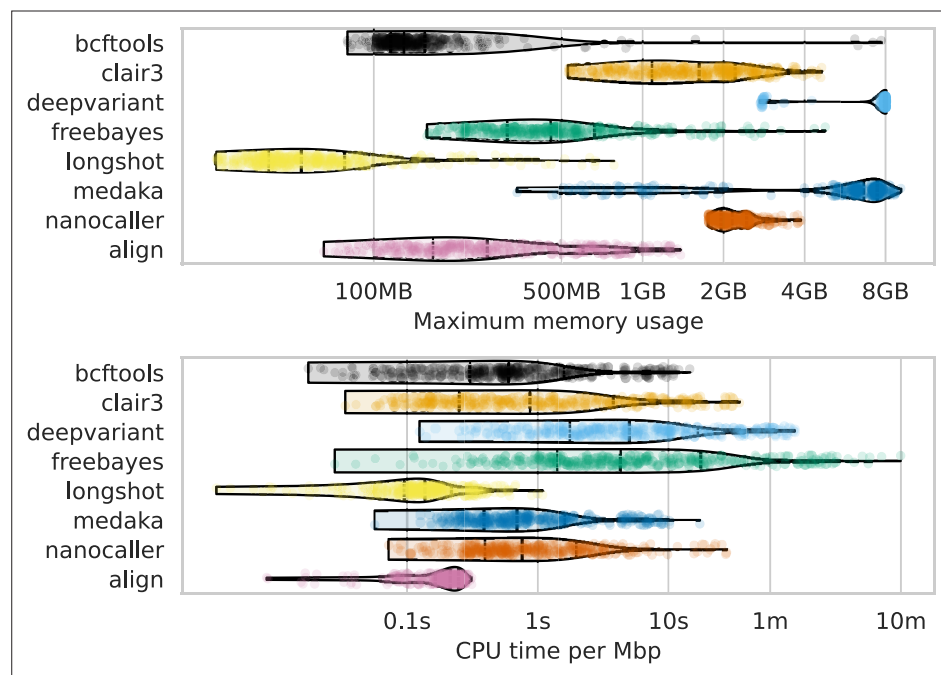


Figure 8. Computational resource usage of alignment and each variant caller (y-axis and colours). The top panel shows the maximum memory usage (x-axis) and the lower panel shows the runtime as a function of the CPU time (seconds) divided by the number of basepairs in the readset (seconds per megabasepairs; x-axis). Each point represents a single run across read depths, basecalling models, read types, and samples for that variant caller (or alignment). s=seconds; m=minutes; MB = megabytes; GB = gigabytes; Mbp = megabasepairs.

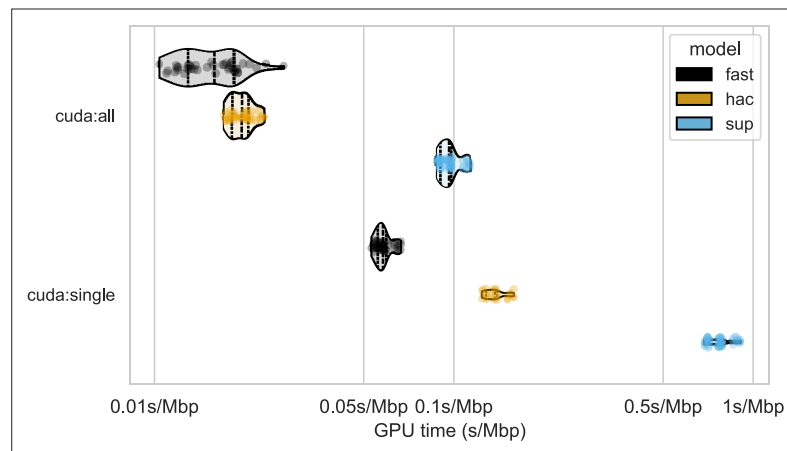
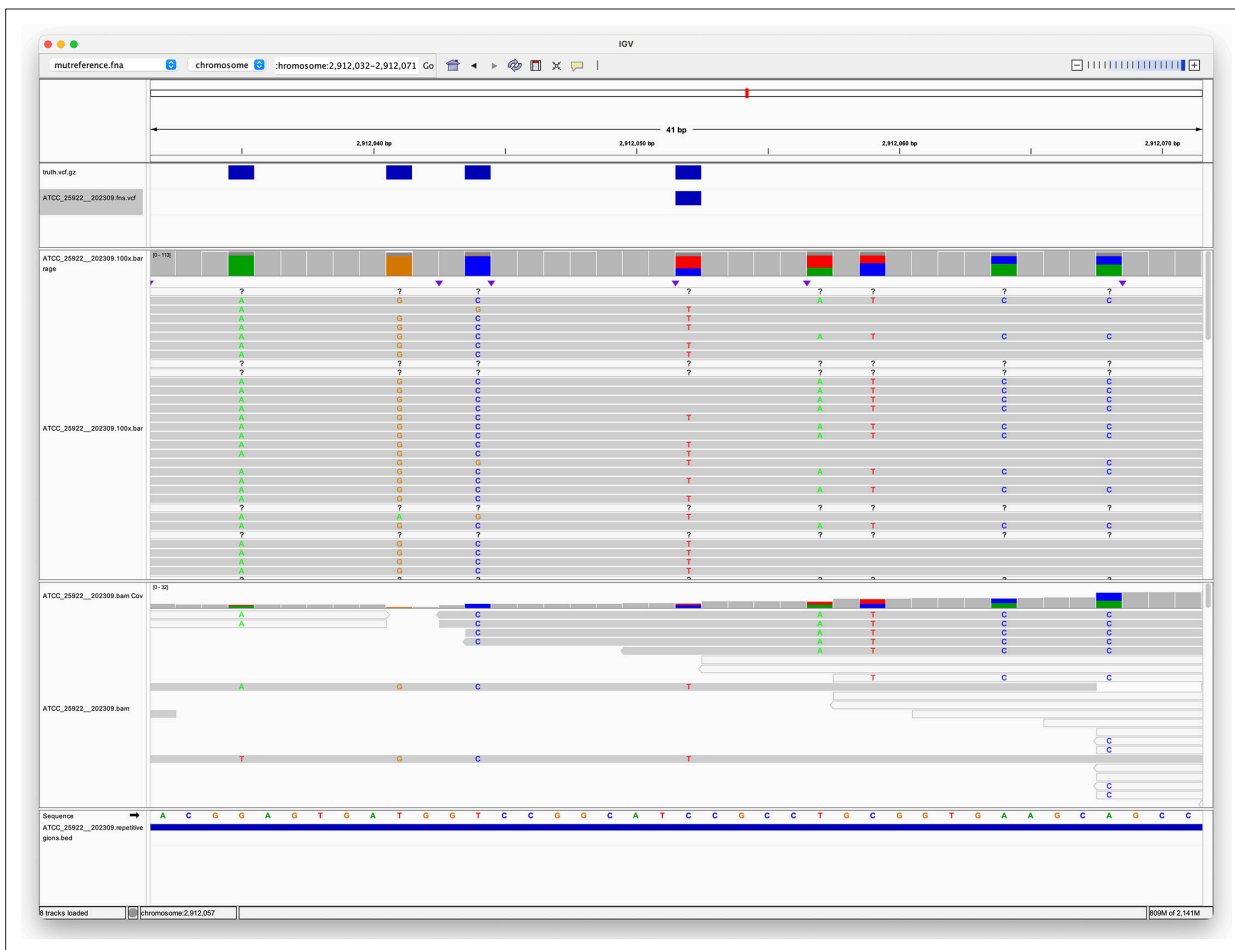


Figure 8—figure supplement 1. Runtime of basecalling Oxford Nanopore Technologies (ONT) data with different models on GPUs. Runtime is presented as seconds per megabasepairs, keeping consistent with **Figure 8** from the main text. The y-axis represents basecalling with either 8 Nvidia A100 GPUs (cuda:all) or 1 A100 (cuda:single), with each stratified by basecalling model (colours). Points represent running a single sample-model combination, with each combination being run three times.



Appendix 2—figure 1. Pileup surrounding a Clair3 false negative variant. The top track (four blue rectangles) shows true variants – i.e., variants we expect to find. The second track shows a single false negative variant. The third track shows the alignment of Oxford Nanopore Technologies (ONT) reads to this region, with coloured letters indicating where the aligned sequence disagrees with the reference. The fourth track shows the alignment of Illumina reads. The fifth track is the reference sequence. The bottom track shows repetitive regions (the whole region is repetitive as the blue rectangle spans the whole region in view).

	1	2	3
S1	-	{T}GG{T}CCGGCAT{C}CGCC	
S2	-	{G}GG{C}CCGGCAT{T}CGCC	
	1	2	3

Appendix 2—figure 2. Textual example of two small repetitive regions in sample ATCC_25922 that lead to a false negative at position 3 in sequence S1. Positions 1 and 2 are two other variant positions where a true positive was obtained for S1 (the expected variants at positions 1 and 2 match the sequence in S2).