

Supplementary Figures and Tables

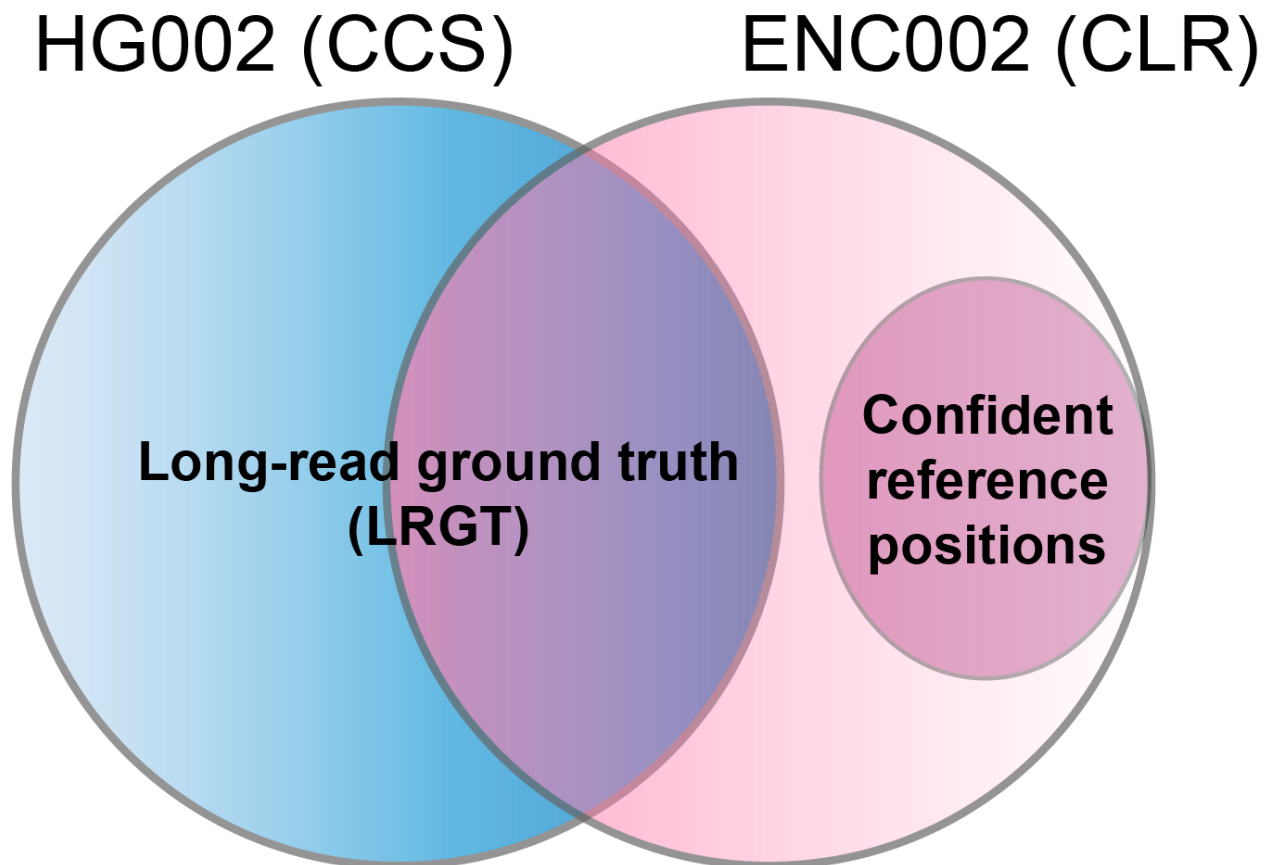


Figure S1. The scheme of building LRGT and confident reference positions. We define SVs called from CCS HG002 as LRGT. We define SVs that were only called in CLR ENC002 and have no same-type SVs called in the nearby region of CCS HG002 as confident reference positions. On short-read HG002, LRGT was used to calculate recall for genotypers and *de novo* callers' recall, while confident reference positions were used to calculate genotypers' precision. Genotypers' precision and F-score were estimated from LRGT and confident reference positions together.

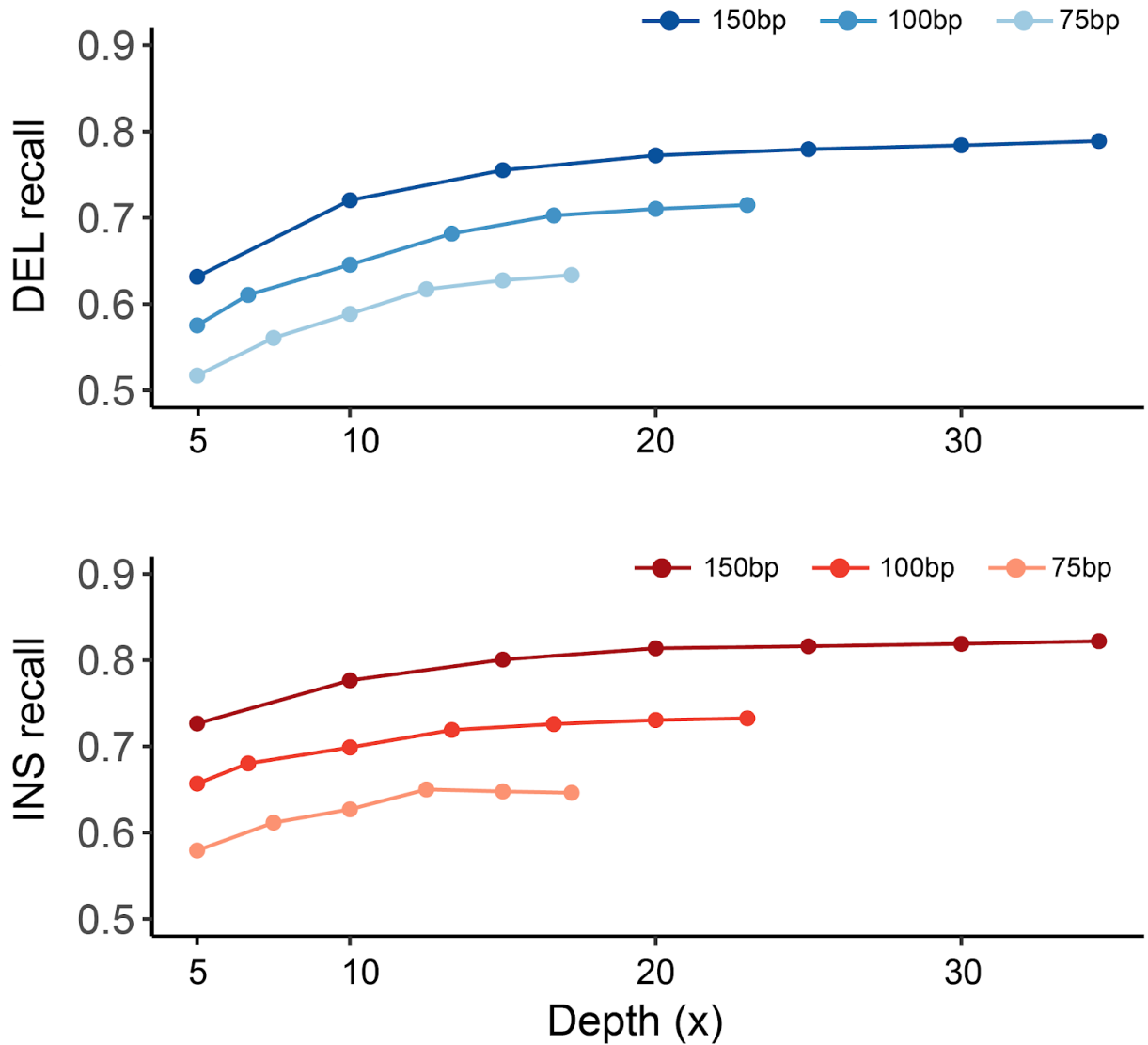


Figure S2. The recall from Paragraph under different read lengths and depths. The recall was calculated on the Illumina HiSeq X sequenced HG002 using LRGT. Each line represents recall under a fixed read length in different depths, generated from downsampling the 150 bp reads in the HiSeqX BAM. To generate data for 100bp and 75bp reads, besides downsampling for depths, the 150bp reads from HiSeqX HG002 were trimmed from their 3' end as well.

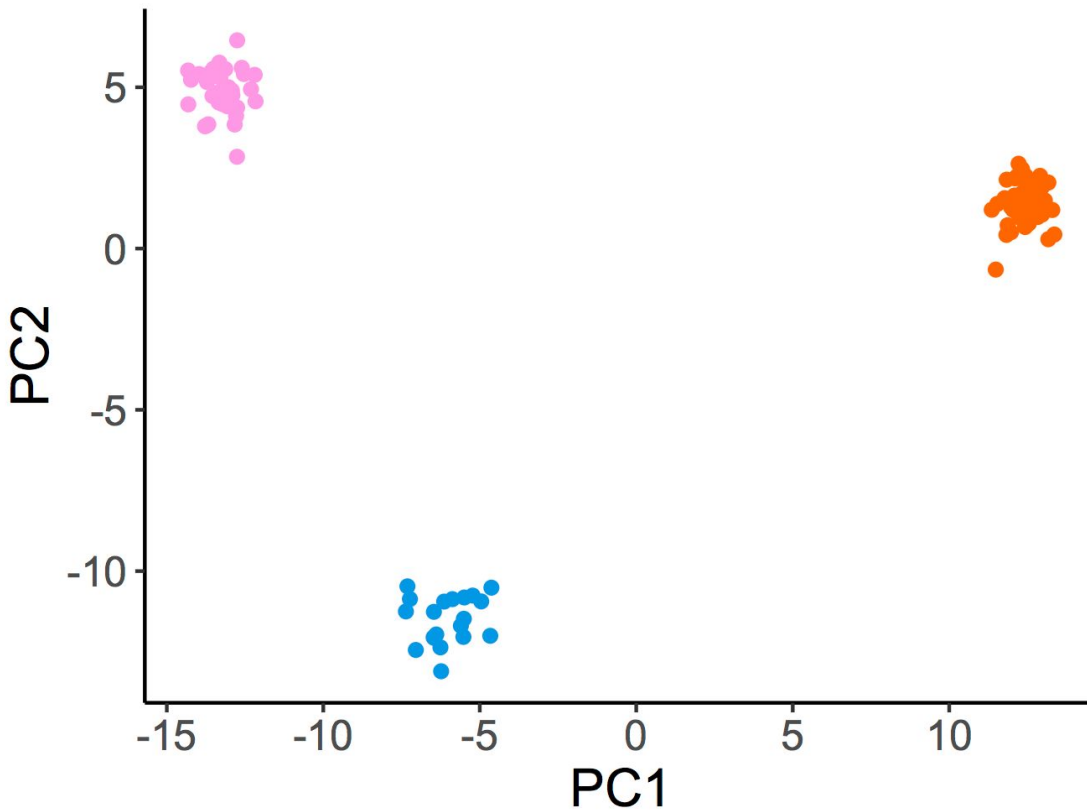


Figure S3. PCA plot of individuals in the population using just the SVs that failed our HWE test

Type	Deletion				Insertion
	Paragraph	Delly Genotyper	Paragraph (100+ bp)	SVTyper (100+ bp)	Paragraph
#True SVs	8,355		5,372		8,956
#Confident reference positions	2,366		1,001		2,855
Recall	0.82	0.68	0.89	0.35	0.82
Specificity	0.92	0.80	0.96	0.99	0.90
Precision	0.97	0.92	0.98	0.99	0.96
F-score	0.89	0.51	0.93	0.78	0.89

Table S1. Overall performance for different genotypers. The recall was evaluated on an Illumina HiSeqX sequenced HG002 data using LRGT (same as in **Table 1**). Specificity was evaluated on the same Illumina HiSeqX data using confident reference positions.

SV length (bp)		50~ 100	100~ 200	200~ 500	500~ 1k	1k~ 10k	>10kbp	All
Recall	#DEL	2,127	934	1,513	264	584	42	5,464
	Paragraph	0.85	0.68	0.83	0.65	0.79	0.83	0.80
	Manta	0.51	0.41	0.72	0.55	0.81	0.79	0.59
Recall	#INS	2,141	1,355	2,182	652	920	31	7,281
	Paragraph	0.94	0.88	0.83	0.84	0.81	0.81	0.87
	Manta	0.43	0.27	0.19	0.11	0.16	0	0.26

Table S2. Recall for Paragraph and Manta measured on the GIAB benchmarking dataset.

The recall was evaluated on the Illumina HiSeqX sequenced HG002. The NIST GIAB SV benchmarking dataset²³ version 0.6 is originally on GRCh37. We lifted to GRCh38 for calculating recall as the reads of HiSeqX HG002 were mapped on GRCh38.