Dataset	Citation (if appropriate) / Name of resource	URL(s)	WES/ WGS/ both	GRCh37/ GRCh38/ both	Germline/ Somatic/ both	Benchmark/ Validation/ Call Sets	Technology (Short/long reads/both)	Description	Short reads/ long reads/ both
1	Database of Genomic Variants	http://dgv.tcag.ca/dgv/app/downloads?ref=GRCh37/hg19	both	both	Germline	Call Sets		Data from 72 studies (2006-2015) for healthy controls; main focus SVs but includes at least some data from CNVs	
2	Zarrei et al, 2015	https://www.nature.com/articles/nrg3871	WGS	both	Germline	Call Sets / Possible Validation		Paper describing final CNV Call set from 1KGP, date available in DGV (dataset 1 above)	
3	Zook et al, 2016	https://www.nature.com/articles/sdata201625	WGS	GRCh37	Germline	Validation		NIST/GIAB call data from 7 WGS samples being developed as reference materials (NA12878, AJ trio, Chinese trio), including CNV calls from Complete Genomics data.	
4	Zook et al, 2019	https://www.nature.com/articles/s41587-020-0538-8	WGS	both	Germline	Validation	both	benchmark for a broadly consented son in a Personal Genome Project trio with broadly available cells and DNA, the Genome in a Bottle (GIAB) Consortium integrated 19 sequence-resolved variant calling methods, both alignment- and de novo assembly-based, from short-, linked-, and long-read sequencing, as well as optical and electronic mapping.	
5	Kosugi et al, 2019	https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1720-5	WGS	GRCh37	Germline			Evaluate the performance of 69 existing SV detection algorithms using multiple simulated and real WGS datasets. datasets are : - A simulated short read dataset was generated using the VarSim simulator [37]: first, a simulated GRCh37 human diploid genome into which known SVs had been introduced at the known sites was generated, then this was used to generate simulated paired-end short reads (125 bp) with 500 bp insert size averaging 30× coverage of the simulated genome (Sim-A) - Four sets of the NA12878 Illumina short read data (data1, data2, data3, and data4) and three sets of PacBio long read data (PacBio-data1, PacBio-data2, and PacBio-data3) were used as real datasets and were acquired from different sources with different read lengths and/or insert sizes - A reference SV dataset for the real data was generated by merging the DGV dataset corresponding to NA12878 and the INS, DEL, and INV data detected from NA12878 long read assemblies	
6	Cameron et al 2019	https://www.nature.com/articles/s41467-019-11146-4	WGS	GRCh37	Germline			evaluation of 10 software with four cell line datasets with orthogonal validation data: NA12878, HG002, and CHM1 and CHM13 separately and merged as a synthetic diploid dataset	
7	Chaisson et al 2019	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6467913/#MOESM3	WGS / WES	GRCh37	Germline			As part of the Human Genome Structural Variation Consortium (HGSVC), we sought to comprehensively determine the complete spectrum of human genetic variation in three family trios. To overcome the barriers to SV detection from conventional algorithms, we integrate a suite of cutting-edge genomic technologies that, when used collectively, allow SVs to be comprehensively assessed in a haplotype-aware manner in diploid genomes. In addition, we also identify the optimal combination of technologies and algorithms that would maximize sensitivity and specificity for SV detection for future genomic studies.	
8	Dixon et al 2018	https://www.nature.com/articles/s41588-018-0195-8#MOESM3	WGS	GRCh37	Somatic / Germline			Here, we present a framework that integrates optical mapping, high-throughput chromosome conformation capture (Hi-C), and whole-genome sequencing to systematically detect SVs in a variety of normal or cancer samples and cell lines.	
9	Nam et al 2016	https://doi.org/10.1093/bib/bbv055	WGS/ WES/ SNP6.0	GRCh37	Somatic/ matched germline			We evaluate six WES somatic CNV detection tools. Using WES data from 50 kidney, 50 bladder, and 50 stomach cancer patients from The Cancer Genome Atlas, we compared the CNV calls from the six tools with a reference CNV set that was identified by both single nucleotide polymorphism array 6.0 and whole-genome sequencing data.	
10	Le Zhang et al 2019	https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007069	WGS		Germline			The sequencing data (94x) of the individual NA12878 were downloaded from the website of the 1000 Genomes Project[29] as evaluation data to compare the performance of CNV detection methods using real sequencing data. The DGV Gold Standard Variants for NA12878 were download from the Database of Genomic Variants (DGV)[28], and a previously published SV benchmark of NA12878[30] was also fetched from the FTP site (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase3/data/NA12878/) Comparison of results of 10 CNV detection systems	
11	Gross et al 2019	https://www.nature.com/articles/s41436-018-0295-y#Sec17	WGS	hg19	Germline	Benchmark/ Validation/ Call Sets		Coriell reference CNV call coordinates. We performed analytical validation of CNV calling on 17 reference samples, compared the sensitivity of GS-based variants with those from a clinical microarray, and set a bound on precision using orthogonal technologies. Twenty reference samples (Coriell, Camden, NJ) were chosen for validation (Table S1). Among these, 18 samples had known pathogenic CNVs representative of a large size range and inclusive of deletions and copy-number gains, and two samples were included as negative controls. Prior to sequencing and analysis, coordinates for this set of truth CNVs were compiled from descriptions on the Coriell website, reference publications, or previously conducted microarray-based CNV analyses27 (Table S1). Note that while all cell lines contain pathogenic CNVs, which established the baseline for our sensitivity analysis, we also examined all other CNVs detected in these samples by either microarray or cGS (see Results).	
12	Brett Trost et al 2018	https://www.sciencedirect.com/science/article/pii/S0002929717304962	WGS	GRCh37/ hg19 and GRCh38/ hg38	Germline	Benchmark/ Validation/ Call Sets		Benchmark using HuRef, NA12878, or AK1 benchmark datasets Deletion Breakpoint verification assays with PCR primers CNV coordinates for datasets are in supplemetary data	
13	gnomad-SV, Collins et al 2020	https://www.nature.com/articles/s41586-020-2287-8	WGS/ WES	hg19/ hg38(lifted)	Germline	Call set			
14	Abel et al 2020	https://www.nature.com/articles/s41586-020-2371-0	WGS	GRCh37/ hg19	Germline	Call set			
15	Peter Ebert et al 2021	https://science.sciencemag.org/content/372/6537/eabf7117	WGS					focused on comprehensive genetic variant discovery from a human diversity panel representing 25 human populations - (32 unrelated and three children from parent-child trios)	