Supplemental Tables

Supplemental Table 1. Statistics over SVs of different types (as determined by the breakpoint orientation signatures) inferred by individual short-read callers and their concordance with SVs supported by long-reads. *specific* determines the number of SVs >=50bp output by a given method. *dd* determines the number of SVs in a deduplicated subset of *specific*. *lr-supp* determines the number of SVs in a set *dd* that have a matching SV in a long-read dataset. External File: "Supplemental_Table_1.xlsx"

Supplemental Table 2. 500bp regions shared across samples 51T, 51N, SKBR3, NA12878, and HG002 with abnormally high 10x Genomics alignment coverage in every sample. External file: "Supplemental_Table_2.bed.gz".

SV set	*total*	*promoter*	*enhancer*	*enhancer (DE)*	*genic*	*exonic*
51T	23,466 [3,258]	100 [18]	3,031 [399]	1,256 [139]	12,046 [1,643]	855 [130]
51T somatic	3,368 [140]	28 [3]	477 [23]	206 [8]	1,721 [62]	161 [10]
51T somatic (-A)	811 [75]	17 [1]	144 [12]	57 [5]	395 [29]	61 [5]
48T	21,333 [NA]	81 [NA]	2,873 [NA]	1,194 [NA]	10,773 [NA]	785 [NA]
SKBR3	19,316 [3,332]	84 [26]	2,488 [455]	1,022 [172]	9,650 [1,563]	691 [161]

Supplemental Table 3. Number of long-read supported SVs overlapping regulatory and coding genomic regions. The subset of long-read supported variants which are also supported by short-read SV calls (at least two callers) is reported in brackets. Overlaps were computed with BEDTools intersect commands with SVs longer than 100kbp were not considered for overlap analysis. Promoters were determined using the UCSC EPDnew (Dreos et al. 2013) track. Enhancers were determined by UCSC GeneHancer (Fishilevich et al. 2017) regular and the higher confidence double elite (DE) tracks. Gene and exons coordinates were downloaded from the Genecode portal (The ENCODE Project Consortium 2012; Davis et al. 2018). Somatic SV set with a "-A" suffix corresponds to somatic SVs in 51T thata re also absent in 15 healthy genomes (Audano et al. 2019).

Sample	*<5% 1KGP freq*	*< 5% 1KGP freq and ≥ 20% Audano freq*	*Variants from previous column overlapping a TR region*
51T	186 [185 | 25]	114	108
48T	188 [188 | NA]	93	85
SKBR3	216 [213 |31]	68	64

Supplemental Table 4. Discordance in population frequencies of COSMIC variants, as determined by short-read genotyping of 1KGP data vs comparison with SVs called from 15 long-read sequenced samples from Audano *et al.* Tandem repeats (TR) were determined using the UCSC TRF track.

Supplemental Table 5. SVs overlapping COSMIC SVs in SKBR3 that are genotyped in <= 5% of the 1KGP samples. The columns are as follows: A) SVid is a unique identifier for the variant; B) chr is the chromosome number for the variant; C) start is the starting position for the SV; D) end is the ending position for the SV; E) type is the type of the SV; F) len is the length of the SV; G) COSMIC is the name of gene that is overlapped by the SV; H) is the fraction of the sample from the 1000 Genomes Project containing this SV as determined by Paragraph; I) Audano is the number of samples described in (Audano et al, Cell, 2019) that contain this variant. External File: "Supplemental_Table_5.xlsx"