Col	ID	annotation
1	***Chromosome***	Chromosome containing the variant in "chr#" format where # is 1 - 22 or X or Y.
2	***Begin***	Physical start position of the variant. 0-based coordinates. Coordinates correspond to hg19.
3	***End***	Physical end position of the variant. 0-based coordinates. Coordinates correspond to hg19.
4	***VarType***	Variant type ('loss' and 'gain').
5	***CNV***	Copy number (0:two copy loss, 1: one copy loss, 3:one copy gain, etc).
6	***Gene***	The transcript(s) nearest to the variant by physical distance. Gene models are derived from the UCSC genome browser known genes track (Meyer et al. 2012). HUGO gene symbol is provided for each gene followed by the UCSC transcript ID in parenthesis. Format: Gene_Symbol(UCSC_transcript_ID)
7	***Gene_Type***	The transcript type. Possible values are "Protein_Coding" or "Noncoding_RNA."
8	***Location***	Location of the variant relative to the nearest transcript(s). Exons and introns are numbered in the direction of the reading frame. Multiple nucleotide subsitutions may span multiple locations (e.g. Exon_6-Intron_6). Potential values are "Upstream", "Downstream", "5UTR" (5' untranslated region), "3UTR" (3' untranslated region), Exon_# (where # is the coding exon number), Intron_# (includes introns flanking coding and non-coding exons), and noncoding_rna for variants landing in non-intronic noncoding RNA sites.
9	***Coding_Impact***	The impact of a variant on a protein coding transcript(s). Multiple transcripts are delimited by "///". It is possible for transcripts with the same gene symbol to receive different values. Potential values include: • Frameshift: applies to deletions completely within a gene body • Partial_Deletion: deletion overlapping either end of a gene • Amplified: complete coding region amplification Partial_Amplification: amplification overlapping either end of a gene • Deleted: complete coding region deletion Out_of_Frame: applies to insertions completely within a gene body.
10	***Known_Gain_Syndrome***	Overlap of CNV with region known to result in pathogenic syndrome due to chromosomal gain. Format syndrome_name ~ %_of_reference_region_contained_in_overlap ~ %_of_CNV_contained_in_overlap ~ allele_frequency.
11	***Known_Loss_Syndrome***	Overlap of CNV with region known to result in pathogenic syndrome due to chromosomal loss. Format syndrome_name

		~ %_of_reference_region_contained_in_overlap ~ %_of_CNV_contained_in_overlap ~ allele_frequency.
12	***ClinVar_Gain***	Overlap of CNV with gain region annotated in Clinvar. Format Disease_Name ~ %_of_reference_region_contained_in_overlap ~ %_of_CNV_contained_in_overlap ~ Pathogenicity ~ Evidence ~ Accession.
13	***ClinVar_Loss***	Overlap of CNV with loss region annotated in Clinvar. Format Disease_Name ~ %_of_reference_region_contained_in_overlap ~ %_of_CNV_contained_in_overlap ~ Pathogenicity ~ Evidence ~ Accession.
14	***1000genomes_Gain***	Overlap of CNV with gain region annotated in 1000 genomes. Format ID ~ percOverlapIn1000genomes ~ percOverlapInAnnotatedCNV ~ Aggregate_AF ~ AFR_AF ~ AMR_AF ~ ASN_AF ~ EUR_AF.
15	***1000genomes_Loss***	Overlap of CNV with loss region annotated in 1000 genomes. Format ID ~ percOverlapIn1000genomes ~ percOverlapInAnnotatedCNV ~ Aggregate_AF ~ AFR_AF ~ AMR_AF ~ ASN_AF ~ EUR_AF.
16	***miRNA_genomic***	List of microRNAs whose non-coding pre-miRNA reading frame within the genome houses the variant. Multiple microRNAs are separated by '///'. Note the different microRNAs listed here have no assumed relationship with the nearest gene nor does the order of presentation have any bearing on the order of presentation of the nearest transcript.
17	***omimGene_ID~omimGene_association***	OMIM gene ID and associated phenotype if any (McKusick 1998). Presented as OMIM_ID~Phenotype. Transcripts delimited by "///".
18	***HGMD_Gene~disease_association***	List of phenotype associated in HGMD with the gene(s) nearest to the variant. Different phenotypes for the same transcript delimited by "$", transcripts delimited by "///".
19	***COSMIC_Gene~NumSamples***	Number of times the gene(s) impacted by the variant have been observed mutated in cancer samples. Transcript specific. Format: cancer_type~number_of_observations. Multiple tumor types separated by "$". Transcripts separated by "///".
20	***MSKCC_CancerGenes***	Determination as to whether the impacted gene(s) are considered cancer genes as catalogued by the Memorial Sloan Kettering Cancer Center (Higgins et al 2007). Potential values: Tumor Suppressor or Oncogene.
21	***Atlas_Oncology***	Determination as to whether the impacted gene(s) are considered cancer genes as catalogued by Atlas Oncology .
22	***Sanger_network-informed_C***	Significant cancer genes imputed by network connectivity to known cancer genes. Manuscript under preparation. Format: gene_name~p-value.

	ancerGenes~Pval	
23	***Mitelman_Database***	Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer. Format: gene_name~Pubmed ID(s).
24	***DrugBank***	DrugBank ID (Wishart et al. 2006) of compounds known to target the impacted gene(s).
25	***ADVISER_Score~Disease_Entry~Explanation***	Modified American College of Medical Genetics summary categorization for variants residing in genes previously causally associated with disease. See ADVISER Scoring for details.

ADVISER Scoring Table

Category 1: sequence variation is previously reported and is a recognized cause of the disorder

- Sequence variant is reported in the literature as causal to disease
- Variant is rare

Category 2: sequence variation is previously unreported and is of the type that is expected to cause the disorder

- Variant is rare
- Variant is predicted to damage protein function with high confidence

Category 3: sequence variation is previously unreported and is of the type which may or may not be causative of the disorder

- Variant is rare
- Variant is predicted to damage protein function with moderate confidence

OR

- Variant is uncommon
- Variant is predicted to damage protein function with high confidence

Category 4: sequence variation is previously unreported and is probably not causative of disease

- Variant is rare
- Variant is predicted to damage protein function with low confidence
- Variant is predicted to damage regulatory function with high confidence

OR

- Variant is uncommon
- Variant is predicted to damage protein function with moderate confidence

Category 5: sequence variation is previously reported and is a recognized neutral variant

- Variant is uncommon
- Variant is not predicted to damage protein function

OR

- Variant is common

If none of the above are met, "-~-~-" is reported in the ADVISER score column.