S5 Table. A detailed reference description of the WES method: DataBase information via snpeff software

No.	Name	Version	Reference (website)	Description
1	ClinVar	2015-05-07	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3965032/	Public archive of relationships among sequence variation and human phenotype
2	COSMIC	v71	http://cancer.sanger.ac.uk/cosmic	Catalogue of Somatic Mutations in Cancer
3	dbNSFP	2.9	https://www.ncbi.nlm.nih.gov/pubmed/21520341	A lightweight database of human nonsynonymous SNPs and their functional predictions
3-1	dnSNP	142	http://nar.oxfordjournals.org/content/29/1/308.full	The NCBI database of genetic variation
3-2	UniProt	2013-06-30	http://www.uniprot.org/	To provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information
3-3	InterPro	-	https://www.ebi.ac.uk/interpro/	Protein sequence analysis and classfication
3-4	SIFT	-	http://sift.bii.a-star.edu.sg/	Prodict effects of nonsynonmous and missense variants
3-5	PolyPhen-2	-	http://genetics.bwh.harvard.edu/pph2/	Prediction of fuctional effects of human nsSNPs (Polymorphism Phenotyping v2)
3-6	LRT	-	https://www.ncbi.nlm.nih.gov/pubmed/25117149	Likelihood ratio tests in rare variant detection for continuous phenotypes.
3-7	MutationTaster	-	http://www.nature.com/nmeth/journal/v7/n8/full/nmeth0810-575.html	MutationTastet evaluates disease-causing potential of sequence alterations
3-8	GERP	-	http://genome.cshlp.org/content/15/7/901.abstract	GERP identifies constrained elements in multiple alignments by quantifying substitution deficits
3-9	phastCons100way_vertebrate	3.4.0	https://bioconductor.org/packages/release/data/annotation/html/phastCons100way.UCSC.hg19.html	Store UCSC phastCons conservation scores for the human genome (hg19) calculated from multiple alignments with other 99 vertebrate species
3-10	1000g	phase3	http://www.nature.com/nature/journal/v526/n7571/full/nature15394.html	The 1000 Genome Project aims to provide a deep characterization of human genome sequence variantion as a foundation for investigating the relationship between genotype and phenotype
4	ESP6500	-	https://esp.gs.washington.edu/drupal/	The goal of the NHLBI GO Exome Sequencing Project is to discover novel genes and mechanisms contributing to heart, lung and blood disorders by pioneering the application of next-generation sequencing of the protein coding regions of the human genome across diverse, richly-phenotyped populations and to share these datasets and findings with the scientific community to extend and enrich the diagnosis, management and treatment of heart, lung and blood disorders.
5	ExAC	release0.3	http://exac.broadinstitute.org/	The Exome Aggregation Consortium is a coalition of investigators seeking to aggregate and harmonize exome sequencing data from a wide variety of large-scale sequencing projects, and to make summary data available for the wider scientfic community