Int J Clin Exp Med 2017;10(11):15294-15301
www.ijcem.com /ISSN:1940-5901/IJCEM0057067

Original Article

Novel and low allele frequency variants observed in the exomes of Southern Han Chinese

Minglin Ou[1,3], Weiguo Sui[1], Ming Yang[2], Zhen Xiang[1], Donge Tang[3], Hua Lin[1], Yue Zhang[1], Jiejing Chen[1], Peng Zhu[3], Wen Xue[1], Chengfu Dai[3], Yong Dai[3]

[1]Guangxi Key Laboratory of Metabolic Diseases Research, Guilin No. 181 Hospital, Guilin, China; [2]Reproductive Medicine Center of Guilin No. 181 Hospital, Guilin 541002, China; [3]Shenzhen People's Hospital, The Second Clinical Medical College, Jinan University, Shenzhen, China

Received March 21, 2017; Accepted October 12, 2017; Epub November 15, 2017; Published November 30, 2017

Abstract: Precision medicine is growing recognized by clinicians, and well understanding of personal genomics is important for precision medicine in clinical practice. However, human DNA genetic variants in different regions still need to be emulated today. Here, we performed variant analysis of ~30,000 genes that affect the protein coding portion in Southern Han Chinese by whole exome sequencing. The sequencing data was aligned to the human reference genome and filtered by NCBI dbSNP database, the novel variants that were absent in the HapMap and dpSNP database were further annotated by 1000 Genomes (1 KG) and Exome Aggregation Consortium (ExAC). We identified 11,081 novel variants that were absent in the HapMap and dpSNP database, including 7,748 uncommon variants that were novel or with a low frequency in 1 KG and ExAC in eight individuals from south of China. The uncommon variants were scattered across the exomes in the different chromosomes. There were 494 uncommon variants, including 389 SNPs and 105 InDels observed in 3 individuals at least, of which 20 SNPs and 22 InDels were observed in three-quarters individuals. Function analysis indicated that the majority of observed uncommon variants in this study appeared to be functionally neutral in the Southern Han Chinese. In summary, we discover a great number of novel and low allele frequency DNA variants recurring in the exomes of Southern Han Chinese, which may be valuable for variant filtering in whole exome sequencing, improving the current human reference genome and precision medicine in the future.

Keywords: Southern Han Chinese, DNA variant, exome, SNPs, InDels

Introduction

In the recent years, a great program has been made after the succession of Human Genome Project in sequencing the whole human genome. And associated experiments indicated that there are about 30,000-40,000 protein-coding genes in the human genome [1]. Some papers indicated that strategies focused on a subset of the human genome, such as protein-coding related genes known as the exome, have been proved to be cost-effective and powerful for human genetic evaluation and health research [2]. The exome accounts for a great part of mutations associating with Mendelian diseases and kinds of tumors, and the subset of sequence contained about 30 million base pairs [3]. As the application of the second DNA sequencing approaches, hundreds of genes have been characterized in human diseases by whole exome sequencing [4].

For the confirmation of disease candidate genes, cosegregation analysis of a genetic disease's family and mutation frequency detection in a control population are commonly included. However, cosegregation analyses may usually lack statistical power and lead to erroneous conclusion in the list with a great number of DNA variants; and some non-causative DNA variants will segregate with a disease by chance just because the linked region contains many genes [5]. In addition, there is a wide spectrum of mutations in kinds of human tumors. Although increasing efforts have been devoted in the identification of rare variants associated with human diseases, development of computation and bioinformatics are still

needed in the interpretation of sequencing data with a great amount of undefined DNA variants [4].

Fortunately, many researchers have been focus on the identification and characterization of DNA variants in human genome [6]. For example, a comprehensive research for the common genetic variants discovering in the individuals from multiple populations in the 1000 Genomes Project was performed by whole-genome sequencing [7]; The genome re-sequencing of an Asian individual revealed about 3,000,000 single-nucleotide polymorphisms (SNPs), of which 13.6% were new SNPs [8]. Furthermore, some researches have sequenced human exomes from Denmark with average depth of 12-fold, and identified an excess of low-frequency coding variants [9]. Importantly, most of the discovered SNPs of these researches are stored in the certain database, which may be valuable for medical genetics and the coming era of personalized medicine [10].

In our previous researches, some efforts have been made to try to discover pathogenic gene mutations by whole exome sequencing; and our studies indicated that whole exome sequencing was a powerful high-throughput tool for the investigation of rare or common coding variants in human genome [2, 11-13]. However, our current understanding of the DNA variants in human exomes limits our interrogation of sequencing data, because there are large amounts of variants that needed to be further deciphered, in spite of variants were filtered by HapMap and dbSNP database. Furthermore, the genetic make-up of Chinese people may quite different from different areas. Therefore, we try to reveal some specific DNA variants in the exomes in the individuals from south of China with high depth of more than 50-fold in this study, which may be valuable for the personalized genome medicine in the future.

Materials and methods

Exome capture and sequencing

Informed consents were received from each individual involved in this study, and the research was approved by Ethics Committee of Guangxi Key laboratory of Metabolic Diseases Research. Genomic DNA was extracted from eight Han ethnicity individuals from seven different large families in the south of China; there were no obviously common disease between families and all of the individuals were non-lineal relatives; they were male, had a mean age of 21 ± 17 years, ranged from 3 to 54 years; the genomic DNA samples were subjected to exome capture and sequencing as described in our previous researches [2, 11, 12]. Briefly, the DNA samples were extracted form peripheral venous blood by QIAamp DNA Blood Mini kit (Qiagen, Hilden, Germany) by standard procedures; for the exome library preparation, the extracted DNA was amplified, purified, and hybridization for sequence capture, non-hybridized fragments were then washed out using NimbleGen SeqCap EZ Human Exome Library v2.0 kit (Roche NimbleGen Inc, Wisconsin, USA). Both non-captured and captured products were subjected to quantitative PCR to estimate the magnitude of enrichment, and each qualified library was then loaded on Illumina Genome Analyzer. Raw image files were processed by Illumina basecalling Software 1.7 for base-calling with default parameters and the sequences of each individual were generated as 90 bp pair-end reads.

Sequencing data processing and mapping

The bioinformatics analysis was performed according to the pipeline of our previous researches [2, 11, 12]. The qualified sequencing data with a desired average depth of 50× was further analyzed. Briefly, the processing and mapping analysis begins with the raw data generated from the Illumina pipeline. Firstly, the raw data was clean, including removing the adapter and discarding low quality reads. And the filtering produce are as follows: (1) if some adapter bases were found in a read in the raw FASTQ data, the adapter reads would be eliminated; (2) if ≥50% bases in a read were defined as low-quality base with quality ≤5 in the raw FASTQ data, the low-quality reads would be eliminated; (3) the reads with unknown bases more than 10% would be eliminated. Secondly, Burrows-Wheeler Aligner (BWA) was used to align the clean reads to the human genome build37 (Hg19) with parameters set to bwa aln-o 1-e 50-m 100000-t 4-i 15-q 10 [14]. The produced BAM format files with alignment information were further processed with some

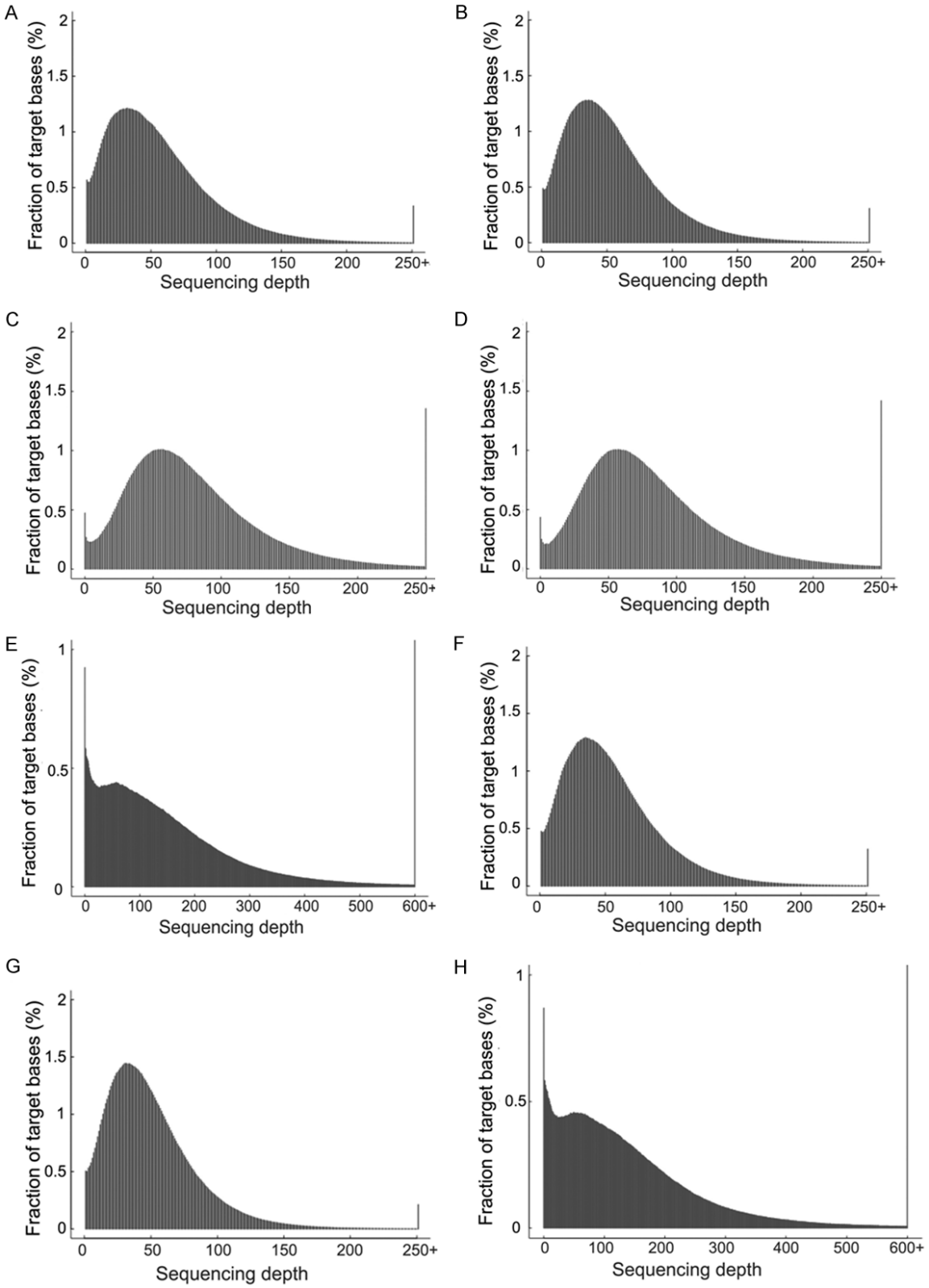

Figure 1. The distribution of per-base sequencing depth for the eight individuals. X-axis denoted sequencing depth, while Y-axis indicated the percentage of total target region under a given sequencing depth.

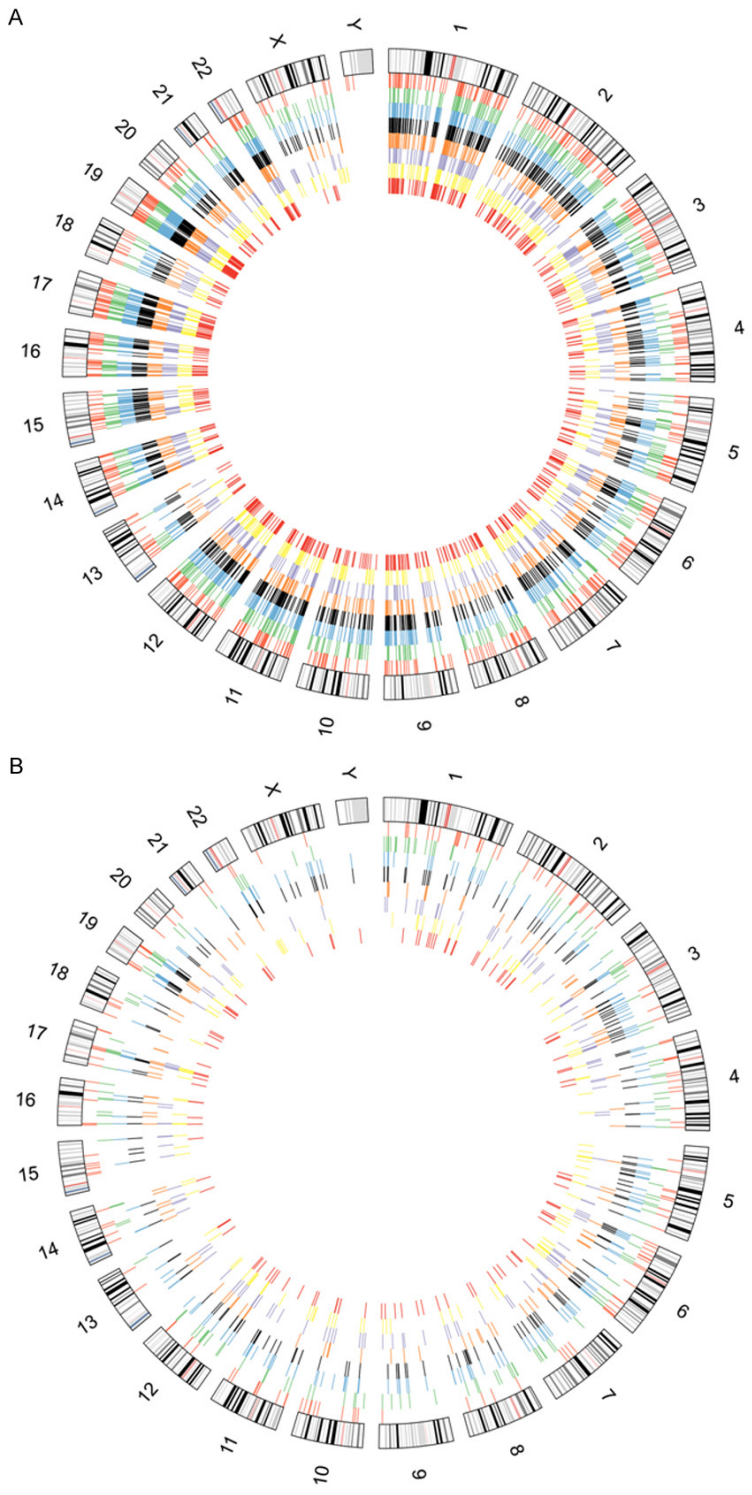

Figure 2. The variants that were absent in the HapMap and dpSNP database. A. The novel or low frequency SNPs variants that are annotated by 1000 Genomes and Exome Aggregation Consortium; B. The novel or low frequency InDels variants that are annotated by 1000 Genomes and Exome Aggregation Consortium.

steps, such as mate-pair information fixing, read group information adding, duplicate reads marking. After the previous processes, the data of final BAM files was performed variant calling. The SOAPsnp was used to detect SNPs with parameters set to soapsnp -t -u -z ! -r 0.00005 -e 0.0001 -Q L [15], and SAMtools/GATK was used to detect InDels [16]. Purity estimation was included in the previous pipeline also the, and filters were used to get variant results with higher confidence. Then ANNOVAR was used to annotate the variant data based on which the subsequent advanced analysis can be carried out [17]. Quality control was carried out in the each stage of the analysis pipeline, including data cleaning, the alignment, and the variant calling. The novel variants that were absent in the HapMap and dpSNP (v135) database were further annotated by 1000 Genomes (1 KG) and Exome Aggregation Consortium (ExAC). The uncommon variants that were novel or with a low frequency (≤0.05) in the 1 KG and ExAC were collected and analyzed. For statistical analysis, the different amount of novel SNPs and InDels distributing in the different chromosomes were using ANOVA and conducted by SPSS Statistics 17.0 (SPSS Inc., Chicago, USA).

Results

Screening of sequence variants in 30,000 coding genes

We captured and analyzed a collection of DNA sequence in the content of the Consensus CDS project, miRBase microRNA databases, and the protein coding portion of the NCBI RefSeq database for each sample, which including more than 30,000 coding genes and 36.5 Mb of the GRCh37/HG19 genome assembly. A total of 55,

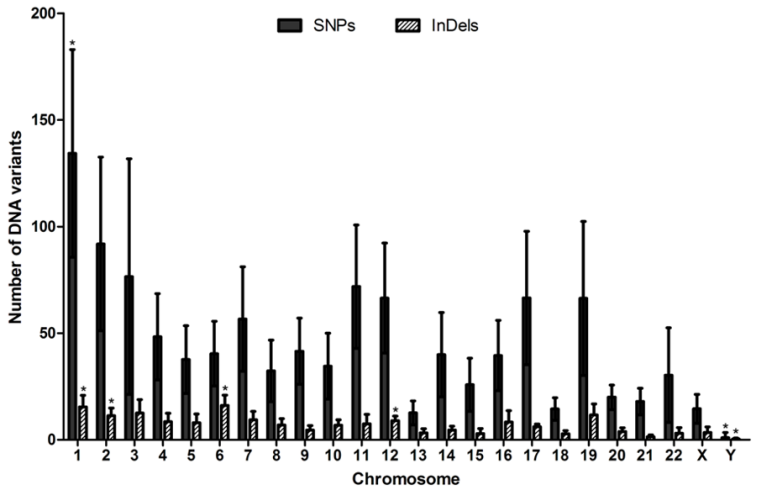

Figure 3. The novel SNPs and InDels distributing in the 46 chromosomes in the exomes of Southern Han Chinese; there were significantly more SNPs in the chromosome 1 comparing to chromosome 13, and significantly less SNPs in the chromosome Y comparing to chromosome 1, 6, 9, 12, 18 and 21 (*$P < 0.05$); there were significantly more InDels in chrome 1 comparing to chromosome 15 and 21, significantly more InDels in chrome 2 comparing to chromosome 18, 21 and 22, significantly more InDels in chrome 6 comparing to chromosome 9, 13, 14, 15, 18, 20, 21, 22 and X, significantly more InDels in chrome 17 comparing to chromosome 21, and ignificantly less InDels in the chromosome Y comparing to chromosome 1, 2, 6, 10, 12 and 17.

Table 1. The frequency of novel variants (SNPs) found in sites of missense, nonsense, read through and splicesite

Sample	Missense	Nonsense	Readthrough	SpliceSite
No.1	452	6	1	170
No.2	456	5	1	171
No.3	746	15	4	121
No.4	764	26	5	122
No.5	797	17	1	172
No.6	486	8	0	167
No.7	461	7	1	160
No.8	716	22	2	192

Table 2. The frequency of novel variants (InDels) found in sites of Frameshift, CDS-Indel, SpliceSite, 3-UTR and 5-UTR

Sample	Frameshift	CDS-Indel	SpliceSite	3-UTR	5-UTR
No.1	16	15	35	45	19
No.2	13	19	30	46	25
No.3	23	21	42	33	54
No.4	22	23	42	22	58
No.5	12	13	44	43	12
No.6	19	15	33	41	32
No.7	18	16	42	40	18
No.8	13	9	40	35	13

184 Mb raw data which containing 615,023,134 reads data was obtained from the eight samples. The average length of read was 89 bp in this research. The mean coverage depths of the 8 samples were 84.0 X, and the median coverage depths was 68.9 X, with more than 99% of targeted DNA sequences covered by at least one read and 97% covered by more than 4 reads. The distribution of per-base sequencing depth in target regions approximately followed a Poisson distribution, which showed the captured exome region was evenly sampled and the sequencing data was qualified enough to be further processed (**Figure 1**).

After mapping, we found 11,081 novel variants that were absent in the HapMap and dpSNP database. In the annotation of the novel variants, a total of 7,748 uncommon variants that were novel or with a low frequency (≤0.05) in the 1 KG and ExAC (**Figure 2**), which included 6,922 SNPs and 826 InDels distributing in the 46 chromosomes, and each chrome contain different amount of DNA variants (**Figure 3**). The uncommon variants including SNPs and InDels, the SNPs were mainly found in the sites of missense, nonsense, splicesite and read through (**Table 1**); the InDels were mainly found in the sites of frameshift, 3-UTR, cds-Indel, spliceSite and 5-UTR (**Table 2**).

Recurring variants in the Southern Han Chinese

In order to discover some specific variants in the Southern Han Chinese, the uncommon variants that were novel or with known allele frequencies ≤0.05 in 1 KG and ExAC recurring in multiple individuals were collected. We found 389 SNPs and 105 InDels recurring in 3 individuals at least, of which 20 SNPs and 22 InDels were recurring in three-quarters individuals (Supplementary Table S1 and S2). In the variants recurring in 3 individuals at least, there were 65

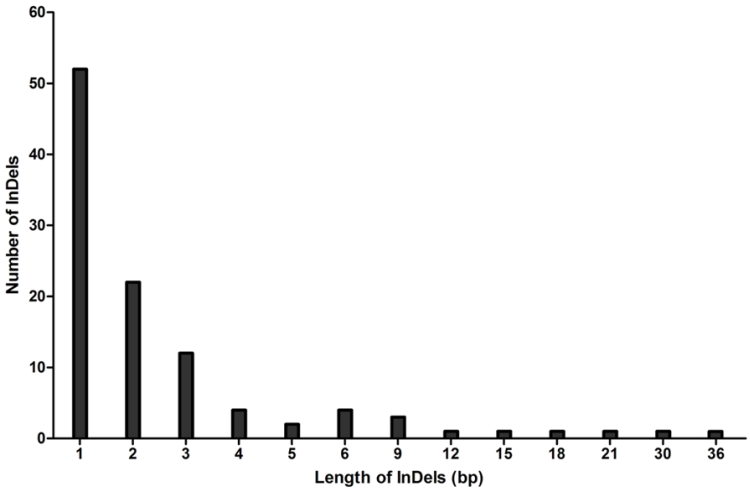

Figure 4. The size distribution of uncommon InDels recurring in the exomes of Southern Han Chinese.

SNPs that may affect protein function predicting by the algorithm of SIFT (Sorting Intolerant From Tolerant), and 105 InDels with a length of 1 bp to 36 bp (**Figure 4**).

Discussion

The main objective of the Human Genome Project is to set up a standard DNA sequence that can be used as a common reference for interpreting the genetic basis of human biology. In addition, the reference sequence of HapMap is playing important roles in genotyping, clinical research and so on. In our study, we try to find specific character of DNA polymorphism in the exomes of Southern Han Chinese. As a result, we found more than 11 thousands of novel DNA variants that were not existed in the HapMap and dpSNP database. It indicates that the current human reference DNA sequence built by limited number of individuals may be hard to represent the all range of human diversity [10]. Although a great number of DNA variants have been found and stored in some databases just like dpSNP database, there are still a great number of variants that are still needed to be further annotated in understanding deep sequencing data [2]. China is a country with the largest population in the world now. There are more than 1.4 billion people in China, which is accounting for one-fifth population of the world. Furthermore, China is a multi-ethnic country with many ethnic groups; the regional distribution of genetic background is poorly known. Our previous study also indicated that the exclusion of un-relative DNA polymorphism in the application of whole exome sequencing in the clinical was difficult. Therefore, as our deepening of understanding human genes, our data may be valuable to understand human DNA variants in different areas and different nations.

Single nucleotide polymorphisms (SNPs) are the most abundant type of genetic marker, which are high density and valuable for the inheritance of genomic regions studying [18]. As we know, the polymorphism of microsatellite DNA is playing important roles in the genomic linkage analysis. However, SNPs are not only distributed in the genomic non-coding region, but also in the gene coding sequence, and they are known as coding SNPs. Some studies have documented that there are 250,000 to 400,000 coding SNPs in the human genome and about 30% of them are non-synonymous SNPs, which may change the function of gene products and contribute to the differences between individuals [19]. About the distribution of coding SNPs in the exomes of Southern Han Chinese, most of them were missense and splice site variants, which may affect the function of gene product. And some studies have indicated that the effect of a missense mutation may be commonly multifaceted and involved in multiple perspectives including biochemical, medical, and evolutionary [20]. According to our study, a total of three variants locating in TMEM14B and SLC9B1 could be found in all of the individuals, and 389 variants observed in the exome of 3 Southern Han Chinese at least. It is well known that a mutation in a gene may involve in human disease; however, it is not easy to know a possible phenotypic effect of a variant, because different variant in the same gene may result in different phenotypic effects and some phenotypic effects of some variants could only be found under certain environments [21]. There are some amino acid substitution prediction algorithms, such as SIFT, which can predict the effect of a SNP [18]. About 16.7% recurring uncommon SNPs were predicted to impact pro-

tein function in our research, which indicated that the majority of the novel or low frequency SNPs recurring in the exomes of Southern Han Chinese might be functionally neutral. Although most of the SNPs cannot be well characterized by current literature and more individuals should be sequenced, we supposed that the SNPs recurring in our study might be some candidate population variants of Southern Han Chinese.

InDels are another kind of common DNA variant in human genome, and millions of InDels have been discovered in the recent years. However, there are much fewer studies have been focus on InDels polymorphisms in human populations and personal genomes comparing to SNPs. The quantity of InDels is second only to the quantity of SNPs in human genomes, and they may impact on human phenotypes and diseases just like other kinds of DNA variant including SNPs [22]. Some studies have indicated that InDels may be an important source of genetic markers that are easy and cheap for genotyping of natural populations [23]. In our study, we discovered more than one hundred uncommon InDels recurring in 3 or more individuals, and two novel InDels could be found in all of Southern Han Chinese individuals. The two novel InDels were located in RND3 and KRTAP4-3. The InDels with a length of 1 bp were the largest sets of InDels, they may cause insert or delete amino acids in coding regions, and result in a frameshift [21]. Although the InDel with a length from 1 bp to 36 bp, their lengths do not appear to be randomly distributed. Most kind of the InDels, their lengths are multiples of 3 base, they may cause amino acid insertion or deletion.

Our study is currently limited to eight individuals of Southern Han Chinese. However, it may bring some significant benefits for many individuals with known phenotypes in the genetic analysis by exome sequencing. For example, the patients with different phenotypes can be grouped, and the disease related mutation analysis could focus on the rare variants that are absent in the human exome, especially those common variants recurring in the Southern Han Chinese. Furthermore, these discovered DNA variants will be valuable for the interpreting a large proportion of Southern Han Chinese.

Summary, this study performed a variant analysis of ~30,000 genes that affect the protein-coding portion of human genome by whole exome sequencing with high depth in the Southern Han Chinese. Our data indicates that there are a great number of novel and low allele frequency DNA variants in the exomes of Southern Han Chinese, which may use to filter human exome sequencing data, improve the current human reference genome and valuable for the personalized genome medicine in the future.

Acknowledgements

This research was supported by Guangxi Natural Science Foundation under Grant No. 2015GXNSFBA139176 and Science and Technology Innovation Commission of Shenzhen Municipality under Grant No. JCYJ20150403-101146277.

Disclosure of conflict of interest

None.

Address correspondence to: Yong Dai and Chengfu Dai, The Second Clinical Medical College, Jinan University, No. 1017, North of Dongmen Road, Shenzhen 518020, China. Tel: +86-755-25533000-2780; E-mail: daiyong22@aliyun.com (YD); Tel: +86-755-22948520; E-mail: daicfu@126.com (CFD)

References

[1] Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M and FitzHugh W. Initial sequencing and analysis of the human genome. Nature 2001; 409: 860-921.

[2] Sui W, Ou M, Liang J, Ding M, Chen J, Liu W, Xiao R, Meng X, Wang L and Pan X. Rapid gene identification in a Chinese osteopetrosis family by whole exome sequencing. Gene 2013; 516: 311-315.

[3] Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A and Eichler EE. Targeted capture and massively parallel sequencing of 12 human exomes. Nature 2009; 461: 272-276.

[4] Wang Z, Liu X, Yang BZ and Gelernter J. The role and challenges of exome sequencing in studies of human diseases. Front Genet 2013; 4: 160.

[5] Piton A, Redin C and Mandel JL. XLID-causing mutations and associated genes challenged in light of data from large-scale human exome se-

quencing. Am J Hum Genet 2013; 93: 368-383.
[6] Chaisson MJ, Wilson RK and Eichler EE. Genetic variation and the de novo assembly of human genomes. Nat Rev Genet 2015; 16: 627-640.
[7] Moore CB, Wallace JR, Wolfe DJ, Frase AT, Pendergrass SA, Weiss KM and Ritchie MD. Low frequency variants, collapsed based on biological knowledge, uncover complexity of population stratification in 1000 genomes project data. PLoS Genet 2013; 9: e1003959.
[8] Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Guo Y, Feng B, Li H, Lu Y, Fang X, Liang H, Du Z, Li D, Zhao Y, Hu Y, Yang Z, Zheng H, Hellmann I, Inouye M, Pool J, Yi X, Zhao J, Duan J, Zhou Y, Qin J, Ma L, Li G, Zhang G, Yang B, Yu C, Liang F, Li W, Li S, Ni P, Ruan J, Li Q, Zhu H, Liu D, Lu Z, Li N, Guo G, Ye J, Fang L, Hao Q, Chen Q, Liang Y, Su Y, San A, Ping C, Yang S, Chen F, Li L, Zhou K, Ren Y, Yang L, Gao Y, Yang G, Li Z, Feng X, Kristiansen K, Wong GK, Nielsen R, Durbin R, Bolund L, Zhang X and Yang H. The diploid genome sequence of an Asian individual. Nature 2008; 456: 60-65.
[9] Li Y, Vinckenbosch N, Tian G, Huerta-Sanchez E, Jiang T, Jiang H, Albrechtsen A, Andersen G, Cao H, Korneliussen T, Grarup N, Guo Y, Hellman I, Jin X, Li Q, Liu J, Liu X, Sparsø T, Tang M, Wu H, Wu R, Yu C, Zheng H, Astrup A, Bolund L, Holmkvist J, Jørgensen T, Kristiansen K, Schmitz O, Schwartz TW, Zhang X, Li R, Yang H, Wang J, Hansen T, Pedersen O, Nielsen R and Wang J. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. Nat Genet 2010; 42: 969-972.
[10] Rosenfeld JA, Mason CE and Smith TM. Limitations of the human reference genome for personalized genomics. PLoS One 2012; 7: e40294.
[11] Liu F, Guo H, Ou M, Hou X, Sun G, Gong W, Jing H, Tan Q, Xue W and Dai Y. ARHGAP4 mutated in a Chinese intellectually challenged family. Gene 2016; 578: 205-209.
[12] Sui W, Hou X, Che W, Ou M, Sun G, Huang S, Liu F, Chen P, Wei X and Dai Y. CCDC40 mutation as a cause of primary ciliary dyskinesia: a case report and review of literature. Clin Respir J 2015; 10: 614-621.
[13] Huang S, Liang J, Sui W, Lin H, Xue W, Chen J, Zhang Y, Gong W, Dai Y and Ou M. EDA mutation as a cause of hypohidrotic ectodermal dysplasia: a case report and review of the literature. Genet Mol Res 2015; 14: 10344-10351.
[14] Li H and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009; 25: 1754-1760.
[15] Li R, Li Y, Fang X, Yang H, Wang J and Kristiansen K. SNP detection for massively parallel whole-genome resequencing. Genome Res 2009; 19: 1124-1132.
[16] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G and Durbin R. The sequence alignment/map format and SAMtools. Bioinformatics 2009; 25: 2078-2079.
[17] Wang K, Li M and Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010; 38: e164-e164.
[18] Sim NL, Kumar P, Hu J, Henikoff S, Schneider G and Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res 2012; 40: W452-W457.
[19] Chasman D and Adams RM. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. J Mol Biol 2001; 307: 683-706.
[20] Kryukov GV, Pennacchio LA and Sunyaev SR. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. Am J Hum Genet 2007; 80: 727-739.
[21] Ng PC, Levy S, Huang J, Stockwell TB, Walenz BP, Li K, Axelrod N, Busam DA, Strausberg RL and Venter JC. Genetic variation in an individual human exome. PLoS Genet 2008; 4: e1000160.
[22] Mills RE. An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res 2006; 16: 1182-1190.
[23] Väli U, Brandström M, Johansson M and Ellegren H. Insertion-deletion polymorphisms (indels) as genetic markers in natural populations. BMC Genet 2008; 9: 8.

Supplementary Table 1. The uncommon variants (SNPs) recurring in the samples

Gene	Position			Variant detail		Allele frequencies				Recurring number	SIFT score	Prediction from SIFT
	Chr	Start	End	Ref	Alt	Hapmap	dbSNP	1 KG	ExAC (162)			
SLC9B1	chr4	103832620	103832620	A	T	NA	NA	NA	Total (0.00140102), East Asian (0.00242014)	8	0.48	TOLERATED
SLC9B1	chr4	103832643	103832643	A	C	NA	NA	NA	Total (0.00001707), East Asian (0.00012042)	8	0	DAMAGING
TMEM14B	chr6	10756712	10756712	C	T	NA	NA	NA	Total (0.00420508), East Asian (0.00271290)	8	0	DAMAGING
NBPF10	chr1	145299925	145299925	A	G	NA	NA	NA	NA	7	0.36	TOLERATED
MLL3	chr7	151919751	151919751	A	G	NA	NA	NA	NA	7	0	DAMAGING
KRT6B	chr12	52843581	52843581	A	G	NA	NA	NA	Total (0.00338917), East Asian (0.00259067)	7	1	TOLERATED
CLEC18B	chr16	74443490	74443490	C	T	NA	NA	NA	Total (0.00006702), East Asian (0.00000000)	7	0.1	TOLERATED
METTL2A	chr17	60512662	60512662	A	G	NA	NA	NA	Total (0.00000825), East Asian (0.00000000)	7	-	-
NBPF15	chr1	148594489	148594489	A	T	NA	NA	NA	Total (0.00067392), East Asian (0.00000000)	6	1	TOLERATED
HRNR	chr1	152186960	152186960	C	T	NA	NA	NA	NA	6	0.61	TOLERATED
IGFN1	chr1	201178844	201178844	T	C	NA	NA	NA	NA	6	-	-
OR2T3	chr1	248637367	248637367	A	G	NA	NA	NA	Total (0.00890077), East Asian (0.00278358)	6	1	TOLERATED
BCLAF1	chr6	136582401	136582401	A	T	NA	NA	NA	Total (0.00103082), East Asian (0.00085616)	6	-	-
UBXN8	chr8	30620840	30620840	G	T	NA	NA	NA	NA	6	-	-
MUC6	chr11	1018473	1018473	A	T	NA	NA	NA	Total (0.00579240), East Asian (0.00162419)	6	0.05	DAMAGING
TAS2R30	chr12	11285909	11285909	G	C	NA	NA	NA	Total (0.00140619), East Asian (0.00453028)	6	0.13	TOLERATED
CDC27	chr17	45234367	45234367	A	T	NA	NA	NA	Total (0.00458591), East Asian (0.00308422)	6	0.45	TOLERATED
ANKRD20A5P	chr18	14184100	14184100	A	G	NA	NA	NA	Total (0.00025662), East Asian (0.00011563)	6	-	-
LILRB2	chr19	54778617	54778617	T	C	NA	NA	CHB (0.009709), CHS (0.004762), JPT (0.004808)	Total (0.00065294), East Asian (0.00081585)	6	0.94	TOLERATED
MUC16	chr19	8999421	8999421	G	C	NA	NA	NA	Total (0.00205404), East Asian (0.00069735)	6	0.27	TOLERATED

FLG	chr1	152284576	152284576	C	A	NA	NA	NA	Total (0.00294906), East Asian (0.00180159)	5	0.42	TOLERATED
IGFN1	chr1	201178667	201178667	C	G	NA	NA	NA	NA	5	-	-
OR2T33	chr1	248436638	248436638	A	G	NA	NA	NA	NA	5	1	TOLERATED
PCDP1	chr2	120366070	120366070	C	T	NA	NA	NA	NA	5	-	-
TUBA3D	chr2	132237750	132237750	G	A	NA	NA	NA	Total (0.00002472), East Asian (0.00000000)	5	0.01	DAMAGING
AP3S1	chr5	115238670	115238670	A	G	NA	NA	NA	Total (0.00064064), East Asian (0.00117730)	5	0.22	TOLERATED
SEC63	chr6	108197874	108197874	C	A	NA	NA	NA	NA	5	-	-
FIG4	chr6	110053838	110053838	A	T	NA	NA	NA	NA	5	-	-
FAM75C2	chr9	90744764	90744764	A	G	NA	NA	NA	NA	5	-	-
NSUN6	chr10	18937579	18937579	A	G	NA	NA	NA	NA	5	-	-
MUC6	chr11	1017498	1017498	C	G	NA	NA	NA	NA	5	0.91	TOLERATED
PRB2	chr12	11546309	11546309	C	G	NA	NA	NA	NA	5	0.74	TOLERATED
CCNT1	chr12	49091972	49091972	G	A	NA	NA	NA	NA	5	-	-
FAM186A	chr12	50745703	50745703	T	G	NA	NA	NA	Total (0.00158507), East Asian (0.00000000)	5	0.43	TOLERATED
FAM186A	chr12	50746243	50746243	T	G	NA	NA	NA	Total (0.00020812), East Asian (0.00000000)	5	0.59	TOLERATED
TPTE2	chr13	20066994	20066994	T	C	NA	NA	NA	Total (0.00058696), East Asian (0.00104505)	5	-	-
OR11H12	chr14	19378036	19378036	C	T	NA	NA	NA	Total (0.00046390), East Asian (0.00445487)	5	0	DAMAGING
POTEM	chr14	20007593	20007593	T	C	NA	NA	NA	Total (0.50000000), East Asian (NA)	5	-	-
EXD1	chr15	41483631	41483631	T	A	NA	NA	NA	NA	5	-	-
ACSF3	chr16	89178494	89178494	T	C	NA	NA	NA	Total (0.00776892), East Asian (0.00253566)	5	-	-
CDC27	chr17	45234327	45234327	C	T	NA	NA	NA	Total (0.00120694), East Asian (0.00081623)	5	0.24	TOLERATED
COL5A3	chr19	10097443	10097443	T	A	NA	NA	NA	NA	5	-	-
PLIN4	chr19	4511223	4511223	C	T	NA	NA	NA	Total (0.00025123), East Asian (0.00012080)	5	0.13	TOLERATED
DENND1B	chr1	197576309	197576309	A	T	NA	NA	CHB (0), CHS (0), JPT (0)	Total (0.00025948), East Asian (0.00000000)	4	-	-
IGSF3	chr1	117146423	117146423	C	T	NA	NA	NA	NA	4	0.4	TOLERATED
IGSF3	chr1	117146563	117146563	G	A	NA	NA	NA	Total (0.00001683), East Asian (0.00000000)	4	0.18	TOLERATED
IGSF3	chr1	117150617	117150617	A	G	NA	NA	NA	Total (0.00009087), East Asian (0.00023116)	4	0.39	TOLERATED

IGSF3	chr1	117150736	117150736	C	G	NA	NA	NA	NA	4	0.1	TOLERATED
IGSF3	chr1	117156600	117156600	T	C	NA	NA	NA	Total (0.00001650), East Asian (0.00000000)	4	1	TOLERATED
IGSF3	chr1	117158898	117158898	C	T	NA	NA	NA	NA	4	0.4	TOLERATED
IGSF3	chr1	117158909	117158909	C	T	NA	NA	NA	Total (0.00001053), East Asian (0.00000000)	4	0.8	TOLERATED
IGSF3	chr1	117158972	117158972	A	G	NA	NA	NA	Total (0.00000853), East Asian (0.00000000)	4	0.03	DAMAGING
PPIAL4G	chr1	143767778	143767778	T	A	NA	NA	NA	Total (0.00373706), East Asian (0.00348028)	4	1	TOLERATED
NBPF14	chr1	148004692	148004692	G	C	NA	NA	NA	Total (0.00181659), East Asian (0.00174581)	4	0.08	TOLERATED
NBPF20	chr1	148346650	148346650	C	A	NA	NA	NA	Total (0.00004186), East Asian (0.00023452)	4	0.06	TOLERATED
ANP32E	chr1	150199057	150199057	C	T	NA	NA	NA	NA	4	0.3	TOLERATED
HRNR	chr1	152189055	152189055	G	C	NA	NA	NA	Total (0.00103842), East Asian (NA)	4	0.99	TOLERATED
IGFN1	chr1	201178688	201178688	A	G	NA	NA	NA	NA	4	-	-
IGFN1	chr1	201178849	201178849	G	A	NA	NA	NA	Total (0.00096165), East Asian (0.00000000)	4	-	-
IGFN1	chr1	201180045	201180045	A	G	NA	NA	NA	Total (0.00011416), East Asian (0.00000000)	4	-	-
IGFN1	chr1	201180064	201180064	G	A	NA	NA	NA	NA	4	-	-
CR1	chr1	207753846	207753846	C	T	NA	NA	NA	Total (0.00715511), East Asian (0.00094922)	4	-	-
OR2L8	chr1	248112554	248112554	T	C	NA	NA	NA	Total (0.00004120), East Asian (0.00000000)	4	1	TOLERATED
NASP	chr1	46073299	46073299	C	T	NA	NA	NA	Total (0.00000831), East Asian (0.00000000)	4	-	-
AK5	chr1	78024352	78024352	C	T	NA	NA	NA	NA	4	1	TOLERATED
ABCA4	chr1	94544142	94544142	T	C	NA	NA	NA	NA	4	-	-
RGPD8	chr2	113147666	113147666	C	A	NA	NA	NA	NA	4	0.97	TOLERATED
POTEF	chr2	130872791	130872791	A	G	NA	NA	NA	Total (0.00001777), East Asian (0.00000000)	4	0.59	TOLERATED
HSPD1	chr2	198363406	198363406	C	T	NA	NA	NA	Total (0.00104669), East Asian (0.00147493)	4	0	DAMAGING
ANKRD36	chr2	97910889	97910889	C	T	NA	NA	NA	NA	4	-	-
ANKRD36	chr2	97910894	97910894	G	A	NA	NA	NA	NA	4	-	-
ANKRD36	chr2	97910960	97910960	A	T	NA	NA	NA	NA	4	0.02	DAMAGING
ANKRD36	chr2	97910988	97910988	A	G	NA	NA	NA	NA	4	0.06	TOLERATED
ANKRD36	chr2	97911007	97911007	G	T	NA	NA	NA	NA	4	0.1	TOLERATED
ANKRD36	chr2	97911022	97911022	C	A	NA	NA	NA	NA	4	0.14	TOLERATED

ANKRD36	chr2	97911045	97911045	A	G	NA	NA	NA	NA	4	0	DAMAGING
ANKRD36	chr2	97911096	97911096	C	T	NA	NA	NA	NA	4	0.01	DAMAGING
ANKRD36	chr2	97911395	97911395	A	G	NA	NA	NA	NA	4	1	TOLERATED
ANKRD36	chr2	97911416	97911416	G	A	NA	NA	NA	NA	4	1	TOLERATED
ANKRD36	chr2	97911498	97911498	T	C	NA	NA	NA	NA	4	0.07	TOLERATED
ANKRD36	chr2	97914920	97914920	T	A	NA	NA	NA	NA	4	0.32	TOLERATED
ANKRD36B	chr2	98124059	98124059	G	A	NA	NA	NA	NA	4	-	-
ANKRD36B	chr2	98124074	98124074	C	A	NA	NA	NA	NA	4	-	-
ANKRD36B	chr2	98128172	98128172	C	A	NA	NA	NA	NA	4	-	-
ANKRD36B	chr2	98128313	98128313	G	A	NA	NA	NA	NA	4	-	-
ANKRD36B	chr2	98128364	98128364	T	C	NA	NA	NA	NA	4	-	-
LNP1	chr3	100170589	100170589	A	G	NA	NA	NA	NA	4	0.05	DAMAGING
MUC13	chr3	124646709	124646709	G	A	NA	NA	NA	NA	4	0.75	TOLERATED
SI	chr3	164748665	164748665	C	A	NA	NA	NA	NA	4	-	-
MUC4	chr3	195510341	195510341	A	G	NA	NA	NA	Total (0.36538462), East Asian (0.00000000)	4	-	-
MUC4	chr3	195510718	195510718	G	T	NA	NA	NA	NA	4	-	-
CCDC66	chr3	56650057	56650057	A	T	NA	NA	NA	Total (0.00001658), East Asian (0.00000000)	4	0.74	TOLERATED
CCDC66	chr3	56650058	56650058	A	C	NA	NA	NA	Total (0.00001656), East Asian (0.00000000)	4	0.03	DAMAGING
DNAH12	chr3	57431752	57431752	T	A	NA	NA	NA	NA	4	-	-
PDZRN3	chr3	73437225	73437225	T	A	NA	NA	NA	NA	4	-	-
FRG2C	chr3	75713555	75713555	G	A	NA	NA	NA	NA	4	0.1	TOLERATED
FRG2C	chr3	75714771	75714771	A	G	NA	NA	NA	NA	4	0.31	TOLERATED
NHEDC1	chr4	103826757	103826757	T	C	NA	NA	NA	NA	4	0.35	TOLERATED
NHEDC1	chr4	103826769	103826769	G	A	NA	NA	NA	NA	4	1	TOLERATED
NHEDC1	chr4	103826813	103826813	A	G	NA	NA	NA	NA	4	-	-
NHEDC1	chr4	103832611	103832611	G	A	NA	NA	NA	NA	4	0.81	TOLERATED
SEC24B	chr4	110384047	110384047	G	T	NA	NA	NA	NA	4	-	-
FRG1	chr4	190874281	190874281	G	T	NA	NA	NA	Total (0.00163532), East Asian (0.00099094)	4	-	-
FRG1	chr4	190876196	190876196	G	A	NA	NA	NA	Total (0.00007433), East Asian (0.00000000)	4	0.29	TOLERATED
FRG1	chr4	190876257	190876257	G	A	NA	NA	NA	Total (0.00002475), East Asian(0.00000000)	4	0.04	DAMAGING
FRG1	chr4	190881933	190881933	A	G	NA	NA	NA	NA	4	1	TOLERATED
FRG1	chr4	190882973	190882973	A	G	NA	NA	NA	Total (0.00004297), East Asian (0.00000000)	4	-	-
MUC7	chr4	71347171	71347171	C	T	NA	NA	NA	NA	4	-	-

AP3S1	chr5	115238589	115238589	A	G	NA	NA	CHB (0), CHS (0), JPT (0)	Total (0.00012537), East Asian (0.00011740)	4	0.11	TOLERATED
DCP2	chr5	112346446	112346446	T	A	NA	NA	NA	Total (0.00006133), East Asian (0.00000000)	4	-	-
AP3S1	chr5	115238605	115238605	G	A	NA	NA	NA	Total (0.00169793), East Asian (0.00187091)	4	0.84	TOLERATED
PANK3	chr5	167988517	167988517	C	T	NA	NA	NA	Total (0.00116025), East Asian (0.00052576)	4	0	DAMAGING
SCAMP1	chr5	77745855	77745855	T	A	NA	NA	NA	NA	4	-	-
BCLAF1	chr6	136599000	136599000	T	A	NA	NA	NA	NA	4	-	-
FAM120B	chr6	170627869	170627869	A	G	NA	NA	NA	NA	4	0.2	TOLERATED
HLA-DRB5	chr6	32497960	32497960	C	A	NA	NA	NA	NA	4	0	DAMAGING
CFTR	chr7	117188750	117188750	C	T	NA	NA	NA	Total (0.00249148), East Asian (0.00132031)	4	0.03	DAMAGING
CFTR	chr7	117188797	117188797	A	G	NA	NA	NA	Total (0.00234815), East Asian (0.00222222)	4	0.63	TOLERATED
MLL3	chr7	151921099	151921099	C	T	NA	NA	NA	NA	4	-	-
MLL3	chr7	151921149	151921149	G	A	NA	NA	NA	NA	4	1	TOLERATED
MLL3	chr7	151927023	151927023	G	C	NA	NA	NA	NA	4	1	TOLERATED
MLL3	chr7	151927067	151927067	T	C	NA	NA	NA	NA	4	0.09	TOLERATED
IGF2BP3	chr7	23353160	23353160	A	G	NA	NA	NA	Total (0.00404366), East Asian (0.00127256)	4	0	DAMAGING
IGF2BP3	chr7	23353246	23353246	A	C	NA	NA	NA	Total (0.00000827), East Asian (0.00011650)	4	0.4	TOLERATED
PABPC1	chr8	101721442	101721442	G	A	NA	NA	NA	Total (0.00032383), East Asian (0.00046653)	4	0.03	DAMAGING
PABPC1	chr8	101721451	101721451	T	A	NA	NA	NA	Total (0.00012440), East Asian (0.00023337)	4	0.47	TOLERATED
ATAD2	chr8	124382164	124382164	A	T	NA	NA	NA	Total (0.00013868), East Asian (0.00092178)	4	1	TOLERATED
PCMTD1	chr8	52733066	52733066	T	C	NA	NA	NA	Total (0.00015656), East Asian (0.00011555)	4	0.9	TOLERATED
C8orf59	chr8	86126827	86126827	C	T	NA	NA	NA	Total (0.00004148), East Asian (0.00000000)	4	0.13	TOLERATED
OR1L4	chr9	125487011	125487011	T	C	NA	NA	NA	Total (0.00006622), East Asian (0.00011614)	4	0.03	DAMAGING
IFNA4	chr9	21187439	21187439	C	G	NA	NA	NA	Total (0.00005925), East Asian (0.00069946)	4	0.43	TOLERATED
PRSS3	chr9	33796647	33796647	T	C	NA	NA	NA	Total (0.00008244), East Asian (0.00000000)	4	0.55	TOLERATED
PRSS3	chr9	33796686	33796686	C	T	NA	NA	NA	NA	4	0.04	DAMAGING

PRSS3	chr9	33796703	33796703	C	G	NA	NA	NA	Total (0.00000824), East Asian (0.00000000)	4	0.94	TOLERATED
PRSS3	chr9	33798543	33798543	A	G	NA	NA	NA	Total (0.00054428), East Asian (0.00118850)	4	0.21	TOLERATED
FOXD4L6	chr9	69200516	69200516	C	T	NA	NA	NA	NA	4	1	TOLERATED
FRG2B	chr10	135438869	135438869	C	T	NA	NA	NA	Total (0.00096482), East Asian (0.00168878)	4	0.08	TOLERATED
PARG	chr10	51363249	51363249	A	G	NA	NA	NA	NA	4	-	-
PARG	chr10	51363348	51363348	G	C	NA	NA	NA	NA	4	-	-
LIPJ	chr10	90362324	90362324	A	C	NA	NA	NA	NA	4	-	-
MUC6	chr11	1017294	1017294	A	T	NA	NA	NA	Total (0.00501512), East Asian (0.00295250)	4	0.03	DAMAGING
MUC6	chr11	1017912	1017912	G	A	NA	NA	NA	Total (0.00595966), East Asian (0.00257069)	4	0.13	TOLERATED
MUC6	chr11	1018348	1018348	G	A	NA	NA	NA	Total (0.01812165), East Asian (0.00167866)	4	0.45	TOLERATED
MUC2	chr11	1093424	1093424	C	G	NA	NA	NA	Total (0.00017075), East Asian (0.00000000)	4	0.09	TOLERATED
MUC2	chr11	1093531	1093531	A	G	NA	NA	NA	NA	4	0.43	TOLERATED
MUC2	chr11	1093537	1093537	A	C	NA	NA	NA	Total (0.00005070), East Asian (0.00000000)	4	0.5	TOLERATED
MUC2	chr11	1093582	1093582	G	C	NA	NA	NA	NA	4	0.29	TOLERATED
LGR4	chr11	27414071	27414071	G	A	NA	NA	NA	NA	4	-	-
FOLH1	chr11	49204732	49204732	T	C	NA	NA	NA	Total (0.00358786), East Asian (0.00070110)	4	0.06	TOLERATED
KRTAP5-7	chr11	71238675	71238675	C	G	NA	NA	NA	NA	4	0.04	DAMAGING
RAB6A	chr11	73390774	73390774	A	G	NA	NA	NA	NA	4	-	-
CLEC9A	chr12	10217326	10217326	C	T	NA	NA	NA	NA	4	-	-
TAS2R30	chr12	11286117	11286117	G	C	NA	NA	NA	Total (0.00001651), East Asian (0.00000000)	4	0.09	TOLERATED
SETD8	chr12	123879666	123879666	A	G	NA	NA	NA	NA	4	0	DAMAGING
SETD8	chr12	123879668	123879668	G	C	NA	NA	NA	NA	4	0	DAMAGING
SETD8	chr12	123889492	123889492	A	C	NA	NA	NA	NA	4	0.03	DAMAGING
NCOR2	chr12	124824721	124824721	C	T	NA	NA	NA	NA	4	0.59	TOLERATED
LIMA1	chr12	50589674	50589674	C	A	NA	NA	NA	NA	4	-	-
LARP4	chr12	50829263	50829263	T	C	NA	NA	NA	NA	4	-	-
IGHV4-59	chr14	107083255	107083255	C	G	NA	NA	NA	Total (0.00176481), East Asian (0.00093023)	4	0.59	TOLERATED
POTEG	chr14	19553642	19553642	A	G	NA	NA	NA	Total (0.00022442), East Asian (0.00000000)	4	1	TOLERATED

HNRNPC	chr14	21679564	21679564	C	G	NA	NA	NA	Total (0.00014884), East Asian (0.00017259)	4	1	TOLERATED
GPHB5	chr14	63784409	63784409	C	G	NA	NA	NA	NA	4	-	-
ADAM21	chr14	70924335	70924335	C	T	NA	NA	NA	Total (0.00443192), East Asian (0.00034851)	4	0.25	TOLERATED
HYDIN	chr16	71012855	71012855	T	C	NA	NA	CHB (0.009709), CHS (0.004762), JPT (0.009615)	Total (0.00025011), East Asian (0.00035162)	4	0.31	TOLERATED
C16orf3	chr16	90095620	90095620	A	G	NA	NA	NA	Total (0.01293550), East Asian (0.00464331)	4	-	-
CDRT4	chr17	15343531	15343531	T	C	NA	NA	NA	NA	4	1	TOLERATED
NCOR1	chr17	16097792	16097792	T	C	NA	NA	NA	Total (0.00002401), East Asian (0.00000000)	4	0.14	TOLERATED
UBB	chr17	16285364	16285364	A	T	NA	NA	NA	NA	4	0	DAMAGING
UBB	chr17	16285381	16285381	C	T	NA	NA	NA	NA	4	0	DAMAGING
CDC27	chr17	45214564	45214564	A	T	NA	NA	NA	NA	4	0.34	TOLERATED
CDC27	chr17	45214606	45214606	G	T	NA	NA	NA	NA	4	0.03	DAMAGING
CDC27	chr17	45219332	45219332	T	G	NA	NA	NA	Total (0.00082573), East Asian (0.00093132)	4	0.05	DAMAGING
CDC27	chr17	45229257	45229257	T	C	NA	NA	NA	Total (0.00002480), East Asian (0.00034682)	4	0.68	TOLERATED
CDC27	chr17	45229262	45229262	G	A	NA	NA	NA	Total (0.00000825), East Asian (0.00000000)	4	0.34	TOLERATED
CDC27	chr17	45234337	45234337	G	C	NA	NA	NA	NA	4	0.69	TOLERATED
CDC27	chr17	45234753	45234753	G	A	NA	NA	NA	Total (0.00001737), East Asian (0.00000000)	4	-	-
CDC27	chr17	45235598	45235598	G	T	NA	NA	NA	NA	4	0.03	DAMAGING
CDC27	chr17	45249391	45249391	T	G	NA	NA	NA	NA	4	0.06	TOLERATED
TUBD1	chr17	57937792	57937792	G	A	NA	NA	NA	NA	4	-	-
SLC35G6	chr17	7386030	7386030	C	T	NA	NA	NA	Total (0.00009105), East Asian (0.00011566)	4	0.65	TOLERATED
POTEC	chr18	14513678	14513678	T	G	NA	NA	NA	NA	4	0.18	TOLERATED
POTEC	chr18	14542791	14542791	C	T	NA	NA	NA	NA	4	0.4	TOLERATED
WDR7	chr18	54354080	54354080	T	C	NA	NA	NA	NA	4	-	-
SERPINB3	chr18	61326628	61326628	A	T	NA	NA	NA	NA	4	-	-
CNDP2	chr18	72186175	72186175	T	C	NA	NA	NA	NA	4	-	-
OR7C2	chr19	15052988	15052988	G	T	NA	NA	NA	NA	4	0.53	TOLERATED
ZNF714	chr19	21299777	21299777	G	A	NA	NA	NA	NA	4	0.05	DAMAGING
ZNF285	chr19	44891202	44891202	C	G	NA	NA	NA	Total (0.00001655), East Asian (0.00000000)	4	0.03	DAMAGING

PLIN4	chr19	4511213	4511213	C	G	NA	NA	NA	Total (0.00003319), East Asian (0.00000000)	4	1	TOLERATED
MADCAM1	chr19	501762	501762	A	C	NA	NA	NA	NA	4	1	TOLERATED
LILRA6	chr19	54744722	54744722	T	C	NA	NA	NA	Total (0.03481647), East Asian (0.00452058)	4	1	TOLERATED
LILRA6	chr19	54745682	54745682	C	T	NA	NA	NA	Total (0.00167814), East Asian (0.00438483)	4	1	TOLERATED
MUC16	chr19	9025639	9025639	T	C	NA	NA	NA	Total (0.00084708), East Asian (0.00034932)	4	0	DAMAGING
MUC16	chr19	9025654	9025654	T	G	NA	NA	NA	Total (0.00184176), East Asian (0.00141543)	4	0	DAMAGING
ITCH	chr20	33033083	33033083	C	T	NA	NA	NA	NA	4	-	-
TPTE	chr21	10951267	10951267	T	C	NA	NA	NA	Total (0.00001660), East Asian (0.00000000)	4	1	TOLERATED
TPTE	chr21	10951311	10951311	G	C	NA	NA	NA	NA	4	0.26	TOLERATED
TPTE	chr21	10951387	10951387	T	G	NA	NA	NA	NA	4	0.68	TOLERATED
TPTE	chr21	10951426	10951426	G	T	NA	NA	NA	NA	4	1	TOLERATED
POTED	chr21	14982577	14982577	A	G	NA	NA	NA	NA	4	1	TOLERATED
POTED	chr21	14982586	14982586	A	G	NA	NA	NA	NA	4	0.92	TOLERATED
POTED	chr21	14982746	14982746	G	A	NA	NA	NA	NA	4	0.24	TOLERATED
GGT1	chr22	25023471	25023471	C	A	NA	NA	NA	NA	4	0.36	TOLERATED
GGT1	chr22	25023513	25023513	G	A	NA	NA	NA	Total (0.00012929), East Asian (0.00000000)	4	0.14	TOLERATED
GGT1	chr22	25024275	25024275	G	A	NA	NA	NA	Total (0.00003846), East Asian (0.00000000)	4	0.07	TOLERATED
ARSD	chrX	2832715	2832715	T	C	NA	NA	NA	Total (0.00103764), East Asian (0.00030152)	4	0.5	TOLERATED
GBP6	chr1	89845935	89845935	T	A	NA	NA	CHB (0), CHS (0), JPT (0)	Total (0.01264850), East Asian (0.00011579)	3	-	-
NBPF10	chr1	145299894	145299894	C	G	NA	NA	CHB (0), CHS (0), JPT (0)	Total (0.00028607), East Asian (0.00000000)	3	1	TOLERATED
ANP32E	chr1	150199027	150199027	C	T	NA	NA	CHB (0), CHS (0), JPT (0)	Total (0.00003303), East Asian (0.00000000)	3	1	TOLERATED
FLG	chr1	152276782	152276782	G	T	NA	NA	CHB (0.009709), CHS (0.004762), JPT (0.038462)	Total (0.00007421), East Asian (0.00081131)	3	1	TOLERATED
IGSF3	chr1	117142977	117142977	A	G	NA	NA	NA	Total (0.00042270), East Asian (0.00047393)	3	-	-
IGSF3	chr1	117158857	117158857	C	T	NA	NA	NA	NA	3	0.03	DAMAGING
IGSF3	chr1	117158983	117158983	T	G	NA	NA	NA	NA	3	0.01	DAMAGING

PRAMEF1	chr1	12854401	12854401	G	T	NA	NA	NA	Total (0.00004947), East Asian (0.00000000)	3	0.01	DAMAGING
NBPF9	chr1	144816621	144816621	A	G	NA	NA	NA	NA	3	1	TOLERATED
NBPF10	chr1	145299906	145299906	T	G	NA	NA	NA	Total (NA), East Asian (0.00000000)	3	1	TOLERATED
ANP32E	chr1	150199051	150199051	T	C	NA	NA	NA	Total (0.00001653), East Asian (0.00000000)	3	0.2	TOLERATED
HRNR	chr1	152186699	152186699	T	A	NA	NA	NA	Total (0.50000000), East Asian (NA)	3	0.76	TOLERATED
FLG	chr1	152276855	152276855	A	C	NA	NA	NA	Total (0.00003298), East Asian (0.00000000)	3	1	TOLERATED
FLG	chr1	152281635	152281635	C	A	NA	NA	NA	Total (0.00000824), East Asian (0.00000000)	3	0.93	TOLERATED
FLG	chr1	152285453	152285453	A	G	NA	NA	NA	NA	3	0.59	TOLERATED
FLG2	chr1	152327810	152327810	C	A	NA	NA	NA	Total (0.00002471), East Asian (0.00000000)	3	1	TOLERATED
GON4L	chr1	155733214	155733214	G	C	NA	NA	NA	Total (0.01010080), East Asian (0.00034317)	3	0.28	TOLERATED
FCGR3A	chr1	161518286	161518286	C	T	NA	NA	NA	Total (0.00136325), East Asian (0.00000000)	3	1	TOLERATED
TNN	chr1	175116062	175116062	G	T	NA	NA	NA	Total (0.00010588), East Asian (0.00019763)	3	-	-
IGFN1	chr1	201178637	201178637	C	G	NA	NA	NA	NA	3	-	-
IGFN1	chr1	201180409	201180409	A	G	NA	NA	NA	Total (0.00086997), East Asian (0.00000000)	3	-	-
IGFN1	chr1	201180496	201180496	G	A	NA	NA	NA	NA	3	-	-
CD55	chr1	207532882	207532882	A	T	NA	NA	NA	Total (0.00503094), East Asian (0.00197759)	3	-	-
OR2L8	chr1	248112794	248112794	G	C	NA	NA	NA	Total (0.00009101), East Asian (0.00000000)	3	0.11	TOLERATED
OR2M5	chr1	248309020	248309020	G	A	NA	NA	NA	Total (0.00095918), East Asian (0.00082508)	3	0.07	TOLERATED
WDR63	chr1	85598528	85598528	C	T	NA	NA	NA	Total (0.00284738), East Asian (0.00000000)	3	-	-
RBMXL1	chr1	89448868	89448868	A	T	NA	NA	NA	Total (0.00004947), East Asian (0.00000000)	3	1	TOLERATED
RBMXL1	chr1	89448896	89448896	A	T	NA	NA	NA	Total (NA), East Asian (0.00000000)	3	1	TOLERATED
NOL10	chr2	10808776	10808776	C	A	NA	NA	NA	Total (NA), East Asian (0.00000000)	3	-	-

RANBP2	chr2	109371399	109371399	T	G	NA	NA	NA	Total (0.00003325), East Asian (0.00000000)	3	0.41	TOLERATED
RGPD5	chr2	110613453	110613453	G	C	NA	NA	NA	NA	3	0.14	TOLERATED
RGPD6	chr2	111273208	111273208	C	G	NA	NA	NA	NA	3	0.14	TOLERATED
ZC3H6	chr2	113074790	113074790	G	T	NA	NA	NA	Total (0.00245724), East Asian (0.00000000)	3	-	-
TMEM37	chr2	120194651	120194651	A	C	NA	NA	NA	Total (0.00003302), East Asian (0.00000000)	3	0.02	DAMAGING
POTEF	chr2	130872761	130872761	T	C	NA	NA	NA	Total (0.00002624), East Asian (0.00024432)	3	0	DAMAGING
POTEF	chr2	130872772	130872772	T	G	NA	NA	NA	Total (0.00002625), East Asian (0.00024366)	3	0.05	DAMAGING
POTEE	chr2	132021629	132021629	G	T	NA	NA	NA	Total (0.00380829), East Asian (0.00023164)	3	0.1	TOLERATED
TUBA3D	chr2	132240243	132240243	A	T	NA	NA	NA	Total (0.00180608), East Asian (0.00105857)	3	0.01	DAMAGING
ORC4	chr2	148733548	148733548	T	A	NA	NA	NA	NA	3	-	-
PDE11A	chr2	178494180	178494180	T	G	NA	NA	NA	Total (0.00004122), East Asian (0.00000000)	3	1	TOLERATED
EEF1B2	chr2	207025358	207025358	A	G	NA	NA	NA	Total (0.00001648), East Asian (0.00000000)	3	0	DAMAGING
ANKRD36	chr2	97911057	97911057	A	T	NA	NA	NA	NA	3	0.07	TOLERATED
ANKRD36	chr2	97911071	97911071	C	T	NA	NA	NA	Total (0.00176567), East Asian (0.00000000)	3	0.01	DAMAGING
ANKRD36	chr2	97911266	97911266	A	G	NA	NA	NA	NA	3	0.17	TOLERATED
ANKRD36B	chr2	98124037	98124037	T	C	NA	NA	NA	NA	3	-	-
ANKRD36B	chr2	98127647	98127647	G	A	NA	NA	NA	NA	3	-	-
ANKRD36B	chr2	98128338	98128338	G	A	NA	NA	NA	NA	3	-	-
ANKRD36B	chr2	98128352	98128352	T	A	NA	NA	NA	NA	3	-	-
SLC9C1	chr3	111873902	111873902	T	A	NA	NA	NA	NA	3	-	-
ATG3	chr3	112253193	112253193	C	A	NA	NA	NA	NA	3	-	-
MFN1	chr3	179103352	179103352	G	T	NA	NA	NA	NA	3	-	-
MUC4	chr3	195507817	195507817	A	G	NA	NA	NA	Total (0.00422372), East Asian (0.00274725)	3	-	-
MUC4	chr3	195510310	195510310	T	G	NA	NA	NA	NA	3	-	-
MUC4	chr3	195510878	195510878	C	G	NA	NA	NA	Total (0.00016863), East Asian (0.00000000)	3	-	-
MUC4	chr3	195513502	195513502	T	G	NA	NA	NA	NA	3	-	-
MUC4	chr3	195515470	195515470	T	G	NA	NA	NA	Total (0.00880626), East Asian (0.00483871)	3	-	-

RBM6	chr3	50098993	50098993	C	T	NA	NA	NA	NA	3	-	-
UBA3	chr3	69112658	69112658	C	A	NA	NA	NA	NA	3	-	-
ADH6	chr4	100137316	100137316	A	T	NA	NA	NA	Total (0.00004019), East Asian (0.00017319)	3	-	-
SLC9B1	chr4	103822492	103822492	G	A	NA	NA	NA	Total (0.00285264), East Asian (0.00291852)	3	-	-
ALPK1	chr4	113347620	113347620	C	T	NA	NA	NA	NA	3	-	-
C4orf21	chr4	113506885	113506885	C	A	NA	NA	NA	NA	3	-	-
ARHGAP10	chr4	148778696	148778696	A	T	NA	NA	NA	Total (0.00004841), East Asian (0.00000000)	3	-	-
FBXW7	chr4	153268227	153268227	C	A	NA	NA	NA	NA	3	-	-
DDX60	chr4	169190129	169190129	C	A	NA	NA	NA	NA	3	-	-
SRA1	chr5	139931629	139931629	C	G	NA	NA	NA	Total (0.00098971), East Asian (0.00166044)	3	-	-
PCDHB9	chr5	140568036	140568036	T	A	NA	NA	NA	NA	3	-	-
PCDH12	chr5	141324975	141324975	C	T	NA	NA	NA	Total (0.00156119), East Asian (0.00065980)	3	0.62	TOLERATED
PCDH12	chr5	141324976	141324976	T	G	NA	NA	NA	Total (0.00157041), East Asian (0.00065274)	3	0.16	TOLERATED
RICTOR	chr5	38953665	38953665	G	A	NA	NA	NA	NA	3	-	-
SKIV2L2	chr5	54710072	54710072	G	T	NA	NA	NA	NA	3	-	-
RAD17	chr5	68692375	68692375	T	A	NA	NA	NA	NA	3	-	-
MED23	chr6	131948816	131948816	T	A	NA	NA	NA	NA	3	-	-
FAM120B	chr6	170627841	170627841	T	C	NA	NA	NA	Total (0.00156548), East Asian (0.00000000)	3	1	TOLERATED
FAM120B	chr6	170627883	170627883	T	G	NA	NA	NA	NA	3	0.81	TOLERATED
ZNF322A	chr6	26638444	26638444	G	A	NA	NA	NA	NA	3	0.39	TOLERATED
C4B	chr6	31996297	31996297	G	A	NA	NA	NA	Total (0.00038916), East Asian (0.00251331)	3	0.51	TOLERATED
COL12A1	chr6	75890933	75890933	T	A	NA	NA	NA	NA	3	-	-
C6orf165	chr6	88126380	88126380	A	T	NA	NA	NA	Total (0.00053441), East Asian (0.00026589)	3	-	-
TRIM73	chr7	75034174	75034174	G	A	NA	NA	CHB (0.019417), CHS (0.019048), JPT (0.004808)	Total (0.00101146), East Asian (0.00458716)	3	0.4	TOLERATED
PRSS1	chr7	142459797	142459797	A	G	NA	NA	NA	Total (0.00007436), East Asian (0.00000000)	3	0.14	TOLERATED
ARHGEF5	chr7	144059791	144059791	C	G	NA	NA	NA	NA	3	0	DAMAGING
MLL3	chr7	151933015	151933015	G	A	NA	NA	NA	NA	3	0	DAMAGING
DPY19L1	chr7	35009095	35009095	C	A	NA	NA	NA	Total (0.00026511), East Asian (0.00000000)	3	0.25	TOLERATED

STAG3L2	chr7	74300308	74300308	C	T	NA	NA	NA	NA	3	0.3	TOLERATED
ASNS	chr7	97493761	97493761	C	G	NA	NA	NA	NA	3	0.07	TOLERATED
SPDYE3	chr7	99917435	99917435	G	A	NA	NA	NA	NA	3	-	-
TEX15	chr8	30704247	30704247	C	T	NA	NA	CHB (0.009709), CHS (0.004762), JPT (0)	Total (0.00032995), East Asian (0.00427746)	3	0	DAMAGING
RP1L1	chr8	10467673	10467673	C	T	NA	NA	NA	Total (0.00535439), East Asian (0.00430223)	3	0.51	TOLERATED
AMAC1L2	chr8	11188806	11188806	G	C	NA	NA	NA	Total (0.00009065), East Asian (0.00000000)	3	0.81	TOLERATED
PCMTD1	chr8	52733171	52733171	C	G	NA	NA	NA	NA	3	0.6	TOLERATED
NSMAF	chr8	59514103	59514103	C	A	NA	NA	NA	Total (0.00057480), East Asian (0.00067705)	3	-	-
ASPH	chr8	62550924	62550924	T	A	NA	NA	NA	NA	3	-	-
IFNA4	chr9	21187434	21187434	C	G	NA	NA	CHB (0.009709), CHS (0), JPT (0.019231)	Total (0.00029555), East Asian (0.00092958)	3	0.45	TOLERATED
FSD1L	chr9	108234179	108234179	C	T	NA	NA	NA	NA	3	-	-
CBWD1	chr9	172167	172167	C	T	NA	NA	NA	Total (0.00000825), East Asian (0.00000000)	3	0.37	TOLERATED
PRSS3	chr9	33796734	33796734	C	A	NA	NA	NA	Total (0.00005769), East Asian (0.00000000)	3	1	TOLERATED
ANKRD20A1	chr9	67938626	67938626	A	G	NA	NA	NA	Total (0.00012051), East Asian (0.00000000)	3	0.11	TOLERATED
PCSK5	chr9	78790192	78790192	C	G	NA	NA	NA	NA	3	0	DAMAGING
FAM75C1	chr9	90535283	90535283	T	C	NA	NA	NA	Total (0.00005799), East Asian (0.00000000)	3	-	-
C10orf76	chr10	103717493	103717493	T	A	NA	NA	NA	NA	3	-	-
CELF2	chr10	11356093	11356093	T	G	NA	NA	NA	NA	3	-	-
SLC39A12	chr10	18280071	18280071	C	T	NA	NA	NA	Total (0.00037691), East Asian (0.00000000)	3	-	-
ENKUR	chr10	25273311	25273311	G	A	NA	NA	NA	NA	3	-	-
BMS1	chr10	43312088	43312088	G	A	NA	NA	NA	Total (0.00001670), East Asian (0.00000000)	3	0.11	TOLERATED
CSGALNACT2	chr10	43659339	43659339	T	G	NA	NA	NA	Total (0.00358806), East Asian (0.00244755)	3	0.71	TOLERATED
PARG	chr10	51363294	51363294	T	C	NA	NA	NA	NA	3	-	-
FAM208B	chr10	5804490	5804490	A	T	NA	NA	NA	Total (0.00183847), East Asian (0.00208986)	3	-	-
MUC5B	chr11	1271077	1271077	G	T	NA	NA	CHB (0), CHS (0), JPT (0.004808)	Total (0.02748383), East Asian (0.00150812)	3	0	DAMAGING
MUC5B	chr11	1271143	1271143	G	A	NA	NA	CHB (0.004854), CHS (0.009524), JPT (0)	Total (0.00027375), East Asian (0.00023196)	3	0	DAMAGING

OR10G8	chr11	123900411	123900411	G	A	NA	NA	CHB (0), CHS (0), JPT (0)	Total (0.00490334), East Asian (0.00186133)	3	0.69	TOLERATED
MUC6	chr11	1017945	1017945	A	G	NA	NA	NA	Total (0.00077882), East Asian (0.00023175)	3	0.88	TOLERATED
MUC6	chr11	1017974	1017974	C	G	NA	NA	NA	Total (0.00087472), East Asian (0.00000000)	3	0.4	TOLERATED
MUC6	chr11	1018390	1018390	C	A	NA	NA	NA	Total (0.00097601), East Asian (0.00131046)	3	1	TOLERATED
MUC6	chr11	1018462	1018462	C	T	NA	NA	NA	Total (0.00642747), East Asian (0.00109276)	3	0.53	TOLERATED
MUC2	chr11	1092595	1092595	C	T	NA	NA	NA	NA	3	0.78	TOLERATED
MUC2	chr11	1093195	1093195	T	A	NA	NA	NA	NA	3	0.45	TOLERATED
MUC5B	chr11	1265872	1265872	G	A	NA	NA	NA	Total (0.00125528), East Asian (0.00127581)	3	0	DAMAGING
MUC5B	chr11	1269602	1269602	C	T	NA	NA	NA	Total (0.00055677), East Asian (0.00034843)	3	0	DAMAGING
TRIM48	chr11	55032462	55032462	T	C	NA	NA	NA	Total (0.00016175), East Asian (0.00061989)	3	0.05	DAMAGING
TRIM48	chr11	55032500	55032500	C	T	NA	NA	NA	Total (0.00112047), East Asian (0.00049640)	3	0.09	TOLERATED
KRTAP5-10	chr11	71276725	71276725	A	G	NA	NA	NA	Total (0.00033193), East Asian (0.00011732)	3	0.1	TOLERATED
STK33	chr11	8486375	8486375	G	A	NA	NA	NA	NA	3	-	-
UBE3B	chr12	109924385	109924385	C	A	NA	NA	NA	NA	3	-	-
TAS2R31	chr12	11183142	11183142	T	G	NA	NA	NA	Total (0.00039761), East Asian (0.00069509)	3	1	TOLERATED
TAS2R30	chr12	11286309	11286309	G	C	NA	NA	NA	Total (0.00008256), East Asian (0.00000000)	3	1	TOLERATED
TAS2R30	chr12	11286336	11286336	T	G	NA	NA	NA	Total (0.00001652), East Asian (0.00000000)	3	0.1	TOLERATED
PRB4	chr12	11461633	11461633	T	C	NA	NA	NA	NA	3	0.81	TOLERATED
PRB2	chr12	11546677	11546677	C	T	NA	NA	NA	Total (0.00030781), East Asian (0.00011677)	3	0.46	TOLERATED
HCAR3	chr12	123200189	123200189	A	C	NA	NA	NA	Total (0.00004125), East Asian (0.00034682)	3	0.1	TOLERATED
SETD8	chr12	123875311	123875311	C	T	NA	NA	NA	Total (0.01022473), East Asian (0.00433967)	3	0	DAMAGING
SETD8	chr12	123879800	123879800	G	C	NA	NA	NA	NA	3	0.16	TOLERATED
SETD8	chr12	123889486	123889486	G	C	NA	NA	NA	NA	3	0	DAMAGING
DDX11	chr12	31244809	31244809	T	C	NA	NA	NA	NA	3	-	-
GXYLT1	chr12	42481688	42481688	T	A	NA	NA	NA	Total (0.00003340), East Asian (0.00000000)	3	0.13	TOLERATED

FAM186A	chr12	50746164	50746164	A	G	NA	NA	NA	NA	3	1	TOLERATED
VWF	chr12	6128067	6128067	G	A	NA	NA	NA	NA	3	0.01	DAMAGING
CLLU1OS	chr12	92821861	92821861	T	C	NA	NA	NA	Total (0.00002543), East Asian (0.00000000)	3	-	-
TPTE2	chr13	20048214	20048214	T	G	NA	NA	NA	Total (0.00307560), East Asian (0.00133085)	3	-	-
AHNAK2	chr14	105412163	105412163	C	G	NA	NA	NA	Total (0.00123627), East Asian (0.00058713)	3	0.4	TOLERATED
AHNAK2	chr14	105416323	105416323	T	C	NA	NA	NA	Total (0.00595338), East Asian (0.00128205)	3	0.13	TOLERATED
OR11H12	chr14	19378117	19378117	T	C	NA	NA	NA	Total (0.00172423), East Asian (0.00244641)	3	0.95	TOLERATED
HNRNPC	chr14	21679396	21679396	C	G	NA	NA	NA	NA	3	0.9	TOLERATED
PPL	chr16	4942126	4942126	C	G	NA	NA	CHB (0), CHS (0.004762), JPT (0)	Total (0.00035547), East Asian (0.00497570)	3	0.28	TOLERATED
AMDHD2	chr16	2579545	2579545	C	G	NA	NA	NA	Total (0.00372398), East Asian (0.00086555)	3	0.11	TOLERATED
FUS	chr16	31202716	31202716	G	T	NA	NA	NA	Total (0.00223874), East Asian (0.00202922)	3	-	-
PKD1L3	chr16	71981414	71981414	C	G	NA	NA	NA	NA	3	-	-
USP10	chr16	84797766	84797766	G	A	NA	NA	NA	Total (0.00068058), East Asian (0.00245327)	3	0.12	TOLERATED
CTU2	chr16	88780641	88780641	G	T	NA	NA	NA	NA	3	-	-
CTU2	chr16	88780645	88780645	T	G	NA	NA	NA	NA	3	-	-
FAM18B2-CDRT4	chr17	15343524	15343524	C	T	NA	NA	NA	NA	3	-	-
NCOR1	chr17	16097786	16097786	C	T	NA	NA	NA	Total (0.00001309), East Asian (0.00000000)	3	0	DAMAGING
CCDC144NL	chr17	20768757	20768757	T	C	NA	NA	NA	Total (0.00091142), East Asian (0.00060768)	3	0	DAMAGING
AMAC1	chr17	33520392	33520392	G	C	NA	NA	NA	NA	3	1	TOLERATED
RPL23	chr17	37006361	37006361	A	T	NA	NA	NA	Total (0.00002701), East Asian (0.00000000)	3	-	-
KRTAP4-3	chr17	39324258	39324258	A	T	NA	NA	NA	NA	3	0.33	TOLERATED
CDC27	chr17	45214623	45214623	G	A	NA	NA	NA	Total (0.00085983), East Asian (0.00000000)	3	0	DAMAGING
CDC27	chr17	45214636	45214636	T	C	NA	NA	NA	NA	3	0.71	TOLERATED
CDC27	chr17	45216160	45216160	A	C	NA	NA	NA	Total (0.00028450), East Asian (0.00011647)	3	0	DAMAGING
CDC27	chr17	45221298	45221298	A	G	NA	NA	NA	NA	3	0.28	TOLERATED

CDC27	chr17	45221318	45221318	A	C	NA	NA	NA	Total (0.00015973), East Asian (0.00000000)	3	0.12	TOLERATED
CDC27	chr17	45234321	45234321	C	G	NA	NA	NA	NA	3	0.4	TOLERATED
CDC27	chr17	45235636	45235636	A	G	NA	NA	NA	NA	3	0	DAMAGING
CDC27	chr17	45249316	45249316	T	C	NA	NA	NA	NA	3	0	DAMAGING
CDC27	chr17	45249365	45249365	C	T	NA	NA	NA	NA	3	0.05	DAMAGING
USP32	chr17	58288396	58288396	T	C	NA	NA	NA	Total (0.01442369), East Asian (0.00173611)	3	1	TOLERATED
RBFOX3	chr17	77111776	77111776	C	G	NA	NA	NA	NA	3	0.19	TOLERATED
POTEC	chr18	14542916	14542916	G	T	NA	NA	NA	NA	3	1	TOLERATED
CD97	chr19	14499537	14499537	A	G	NA	NA	NA	Total (0.00328211), East Asian (0.00127374)	3	0.2	TOLERATED
POTED	chr21	14982764	14982764	G	A	NA	NA	NA	NA	3	1	TOLERATED
POTED	chr21	14982904	14982904	G	A	NA	NA	NA	Total (0.00002650), East Asian (0.00000000)	3	0.54	TOLERATED
GART	chr21	34900917	34900917	T	A	NA	NA	NA	Total (0.00004750), East Asian (0.00000000)	3	-	-
GGTLC2	chr22	22989566	22989566	A	G	NA	NA	CHB (0), CHS (0), JPT (0)	Total (0.00121221), East Asian (0.00000000)	3	-	-
OR11H1	chr22	16449003	16449003	T	C	NA	NA	NA	Total (NA), East Asian (0.00000000)	3	1	TOLERATED
GGTLC2	chr22	22989571	22989571	T	C	NA	NA	NA	NA	3	-	-
GGT1	chr22	25023441	25023441	C	T	NA	NA	NA	Total (0.00001662), East Asian (0.00011582)	3	0.01	DAMAGING
GGT1	chr22	25023459	25023459	G	A	NA	NA	NA	Total (0.00000854), East Asian (0.00000000)	3	1	TOLERATED
TTC38	chr22	46679864	46679864	C	T	NA	NA	NA	NA	3	-	-
NXF2	chrX	101576806	101576806	G	A	NA	NA	NA	NA	3	-	-
MID1	chrX	10491233	10491233	C	A	NA	NA	NA	NA	3	-	-
FRMPD4	chrX	12725042	12725042	C	T	NA	NA	NA	NA	3	-	-
OPN1MW	chrX	153455671	153455671	G	T	NA	NA	NA	Total (NA), East Asian (NA)	3	0.4	TOLERATED

Chr: chromosome; 1 KG: 1000 Genomes; ExAC: Exome Aggregation Consortium; Recurring number, the variants observed in the number of the individuals; NA: Not applicable, the variant have not been included in the database.

Supplementary Table 2. The uncommon variants (InDels) recurring in the samples

Gene	Position			Variant detail	Allele frequencies				Recurring number
	Chr	Start (2)	end (3)		Hapmap	dbSNP	1 KG	ExAC	
RND3	chr2	151328285	151328285	-1A	NA	NA	NA	NA	8
KRTAP4-3	chr17	39324229	39324229	+15GCAGCAGGTGGTCAG	NA	NA	NA	NA	8
HTR3D	chr3	183756836	183756836	+1T	NA	NA	NA	NA	7
EVC2	chr4	5630471	5630471	-1A	NA	NA	NA	NA	7
MAP3K1	chr5	56177849	56177851	-3CAA	NA	NA	NA	NA	7
FGFR1OP	chr6	167453519	167453519	+1T	NA	NA	NA	NA	7
PRIM2	chr6	57512787	57512787	+9CACCAAGGC	NA	NA	NA	NA	7
PRIM2	chr6	57512788	57512788	+2TA	NA	NA	NA	NA	7
PRIM2	chr6	57512794	57512796	-3ATT	NA	NA	NA	NA	7
OSBPL8	chr12	76780051	76780051	+1A	NA	NA	NA	NA	7
COL11A1	chr1	103496806	103496806	-1A	NA	NA	NA	NA	6
AKNAD1	chr1	109377158	109377158	-1A	NA	NA	NA	NA	6
C1orf200	chr1	9714538	9714538	+2GA	NA	NA	NA	NA	6
TMIE	chr3	46751074	46751076	-3AAG	NA	NA	NA	NA	6
BRD8	chr5	137507105	137507105	-1A	NA	NA	NA	NA	6
RASGEF1C	chr5	179528352	179528352	-1G	NA	NA	NA	NA	6
SYNE1	chr6	152629773	152629773	-1A	NA	NA	NA	NA	6
MAP3K4	chr6	161519351	161519353	-3CTG	NA	NA	NA	NA	6
ADCY8	chr8	131792607	131792607	+1A	NA	NA	NA	NA	6
NSUN6	chr10	18940185	18940185	+1T	NA	NA	NA	NA	6
NDUFB1	chr14	92588156	92588159	-4TTTT	NA	NA	NA	NA	6
RPL23	chr17	37006350	37006351	-2TT	NA	NA	NA	NA	6
HTR3D	chr3	183756837	183756837	+12TCAGTCCAAATT	NA	NA	NA	NA	5
DSPP	chr4	88537205	88537213	-9GACAGCAGT	NA	NA	NA	NA	5
HGC6.3	chr6	168376880	168376880	+1G	NA	NA	NA	NA	5
HGC6.3	chr6	168376882	168376882	+1T	NA	NA	NA	Total (0.00041754), East_Asian (0.00000000)	5
HNF4G	chr8	76468310	76468311	-2TT	NA	NA	NA	NA	5
NRAP	chr10	115405703	115405703	+2AA	NA	NA	NA	NA	5
KIAA0174	chr16	71956592	71956592	+2AG	NA	NA	NA	NA	5
PLA2G4C	chr19	48613828	48613829	-2TA	NA	NA	NA	NA	5
HRNR	chr1	152195729	152195729	-1T	NA	NA	NA	NA	4
KCNH1	chr1	211280725	211280725	-1A	NA	NA	NA	NA	4
TTC13	chr1	231094051	231094051	-1A	NA	NA	NA	NA	4
SPC25	chr2	169728070	169728072	-3AAA	NA	NA	NA	NA	4
DNAH7	chr2	196756536	196756536	-1A	NA	NA	NA	NA	4

C1D	chr2	68269961	68269966	-6GGGGGG	NA	NA	NA	NA	4
ATG3	chr3	112253058	112253058	+1A	NA	NA	NA	NA	4
KIAA2018	chr3	113376111	113376113	-3TGC	NA	NA	NA	NA	4
LNX1	chr4	54327211	54327211	+1A	NA	NA	NA	NA	4
MACC1	chr7	20180507	20180507	+2AC	NA	NA	NA	NA	4
PMS2	chr7	6037058	6037058	-1A	NA	NA	NA	NA	4
C7orf62	chr7	88424366	88424367	-2AA	NA	NA	NA	NA	4
ATAD2	chr8	124382159	124382164	-6TCATCA	NA	NA	NA	NA	4
PSD3	chr8	18393159	18393159	+1T	NA	NA	NA	NA	4
TMEM2	chr9	74349918	74349919	-2AA	NA	NA	NA	NA	4
GPAM	chr10	113915574	113915574	+1A	NA	NA	NA	NA	4
ACADSB	chr10	124812686	124812686	-1T	NA	NA	NA	NA	4
AGAP4	chr10	46342668	46342688	-21GCTCCTGCCATCCTGTCCCCA	NA	NA	NA	Total (NA), East_Asian (NA)	4
PATE1	chr11	125618731	125618731	-1T	NA	NA	NA	NA	4
KRTAP5-1	chr11	1606121	1606150	-30CCACAGCCACCCTTGGATCCCCCACAAGAG	NA	NA	NA	NA	4
SLCO1B1	chr12	21327652	21327652	+1A	NA	NA	NA	NA	4
KDM5A	chr12	498268	498268	-1G	NA	NA	NA	NA	4
CLEC4E	chr12	8693419	8693419	+4TCTC	NA	NA	NA	NA	4
KPNA3	chr13	50366680	50366680	+3GCG	NA	NA	NA	NA	4
FLYWCH1	chr16	2987056	2987059	-4GGGG	NA	NA	NA	NA	4
FLYWCH1	chr16	2987064	2987065	-2GC	NA	NA	NA	NA	4
FLYWCH1	chr16	2987067	2987068	-2GA	NA	NA	NA	NA	4
FLYWCH1	chr16	2987072	2987073	-2GG	NA	NA	NA	NA	4
NETO1	chr18	70532186	70532186	-1A	NA	NA	NA	NA	4
PLA2G4C	chr19	48613839	48613840	-2TC	NA	NA	NA	NA	4
SIRPB1	chr20	1592043	1592044	-2CA	NA	NA	NA	Total (0.00001579), East_Asian (0.00000000)	4
SIRPB1	chr20	1592048	1592048	-1G	NA	NA	NA	Total (0.00001577), East_Asian (0.00000000)	4
TMEM164	chrX	109416689	109416689	-1T	NA	NA	NA	NA	4
CD34	chr1	208063100	208063100	+1A	NA	NA	NA	NA	3
AC012487.2	chr2	109123988	109123988	+1T	NA	NA	NA	NA	3
ORC4	chr2	148701100	148701100	-1A	NA	NA	NA	NA	3
LEPREL1	chr3	189690816	189690816	+1A	NA	NA	NA	NA	3
FAM157A	chr3	197880131	197880136	-6GCAGCA	NA	NA	NA	NA	3
PDZRN3	chr3	73651622	73651622	-1A	NA	NA	NA	NA	3
OTUD4	chr4	146076633	146076633	-1A	NA	NA	NA	NA	3
RWDD4	chr4	184570724	184570724	-1A	NA	NA	NA	NA	3

SEMA6A	chr5	115803579	115803581	-3GTG	NA	NA	NA	NA	3
SNX2	chr5	122165471	122165476	-6ACACAC	NA	NA	NA	NA	3
RNF145	chr5	158588634	158588638	-5AAAAA	NA	NA	NA	NA	3
AKD1	chr6	109954066	109954069	-4ATAG	NA	NA	NA	NA	3
ADTRP	chr6	11714639	11714639	+1A	NA	NA	NA	NA	3
EDN1	chr6	12294603	12294605	-3AGA	NA	NA	NA	Total (0.00051983), East_ Asian (0.00173531)	3
PDE7B	chr6	136172996	136172996	+1T	NA	NA	NA	NA	3
MAP3K5	chr6	136878752	136878752	-1T	NA	NA	NA	NA	3
IL20RA	chr6	137325902	137325903	-2AA	NA	NA	NA	NA	3
NUP43	chr6	150052698	150052698	+1A	NA	NA	NA	NA	3
ATXN1	chr6	16327910	16327918	-9TGCTGATGC	NA	NA	NA	NA	3
FAM120B	chr6	170627660	170627695	-36CCCTGAATCCAGGCGAGAAGTTCCCATGT-GTTCAGA	NA	NA	NA	Total (0.00005829), East_ Asian (0.00011979)	3
DUSP22	chr6	350940	350940	+1A	NA	NA	NA	NA	3
F13A1	chr6	6318921	6318921	+1A	NA	NA	NA	NA	3
COL12A1	chr6	75890933	75890933	+1A	NA	NA	NA	NA	3
POLR2J	chr7	102113952	102113953	-2GG	NA	NA	NA	NA	3
KB-1208A12.3	chr8	99054836	99054836	+1A	NA	NA	NA	NA	3
SETX	chr9	135152543	135152543	+1A	NA	NA	NA	NA	3
C10orf76	chr10	103717493	103717493	+1A	NA	NA	NA	NA	3
LOC729020	chr10	105005683	105005683	+1A	NA	NA	NA	NA	3
TMPRSS13	chr11	117772788	117772789	-2CA	NA	NA	NA	NA	3
PUS3	chr11	125763610	125763610	+1T	NA	NA	NA	NA	3
TSPAN8	chr12	71523200	71523201	-2AA	NA	NA	NA	NA	3
SKA3	chr13	21729952	21729952	+2AA	NA	NA	NA	NA	3
MEF2A	chr15	100252710	100252712	-3CAG	NA	NA	NA	NA	3
MEFV	chr16	3294155	3294155	-1T	NA	NA	NA	NA	3
MKNK2	chr19	2037647	2037647	-1G	NA	NA	NA	NA	3
CCDC9	chr19	47774984	47775001	-18GTGTGTGCGCGCGCGCGC	NA	NA	NA	NA	3
PLA2G4C	chr19	48613834	48613836	-3TGT	NA	NA	NA	NA	3
HRC	chr19	49657889	49657889	+3TCC	NA	NA	NA	NA	3
SPINT3	chr20	44141202	44141203	-2CA	NA	NA	NA	NA	3
UCKL1	chr20	62571418	62571418	-1G	NA	NA	NA	NA	3
ATP5L2	chr22	43035850	43035851	-2TT	NA	NA	NA	NA	3
UTY	chrY	15508700	15508704	-5CTTCT	NA	NA	NA	NA	3

Chr: chromosome; 1 KG: 1000 Genomes; ExAC: Exome Aggregation Consortium; Recurring number, the variants observed in the number of the individuals; NA: Not Applicable, the variant have not been included in the database.