Findings of a 1303 Korean whole-exome sequencing study

Soo Heon Kwak1,11, Jeesoo Chae2,3,11, Seongmin Choi4,11, Min Jung Kim2,3, Murim Choi3, Jong-Hee Chae5, Eun-hae Cho6, Tai ju Hwang7, Se Song Jang2,3, Jong-Il Kim2,3,8,11, Kyong Soo Park1,9,10,11 and Yung-Jue Bang1

Ethnically specific data on genetic variation are crucial for understanding human biology and for clinical interpretation of variant pathogenicity. We analyzed data obtained by deep sequencing 1303 Korean whole exomes; the data were generated by three independent whole-exome sequencing projects (named the KOEX study). The primary focus of this study was to comprehensively analyze the variant statistics, investigate secondary findings that may have clinical actionability, and identify loci that should be cautiously interpreted for pathogenicity. A total of 495,729 unique variants were identified at exonic regions, including 169,380 nonsynonymous variants and 4356 frameshift insertion/deletions. Among these, 76,607 were novel coding variants. On average, each individual had 7136 nonsynonymous single-nucleotide variants and 74 frameshift insertion/deletions. We classified 13 pathogenic and 13 likely pathogenic variants in 56 genes that may have clinical actionability according to the guidelines of the American College of Medical Genetics and Genomics, and the Association for Molecular Pathology. The carrier frequency of these 26 variants was 2.46% (95% confidence interval 1.73–3.46). To identify loci that require cautious interpretation in clinical sequencing, we identified 18 genes that are prone to sequencing errors, and 671 genes that are highly polymorphic and carry excess nonsynonymous variants. The catalog of identified variants, its annotation and frequency information are publicly available (http://koex.snu.ac.kr). These findings should be useful resources for investigating ethnically specific characteristics in human health and disease.

Experimental & Molecular Medicine (2017) 49, e356; doi:10.1038/emm.2017.142; published online 14 July 2017

INTRODUCTION

Technical advances in massive parallel sequencing have resulted in increased application of whole-exome sequencing (WES), not only for research purposes but also for clinical genetic diagnosis. As the United States Food and Drug Administration approved marketing authorization for the first next-generation sequencer in 2013, its clinical application is expected to expand more rapidly.1 It has been reported that WES can provide a potential molecular genetic diagnosis in ~25% of cases referred for suspected genetic disorders in clinical genetics laboratories.2 WES studies are also applied in genomics research on a large scale to identify causal coding variants of complex disorders.3 A comprehensive catalog of ethnically specific genetic variations and its frequency spectrum are crucial for determining variant pathogenicity as well as examining the quality of WES procedures. Currently, there are several genetic variation databases, such as those derived from the 1000 Genomes Project,4 NHLBI Exome Sequencing Project,5 and Exome Aggregation Consortium.6 However, most of these sequencing projects involved European populations, and more genetic information is required for other populations, including East Asians.

When WES is performed, a large number of variants are identified. Variants that are not directly related to the specific condition for which WES is performed are referred to as secondary or incidental findings.7 Some of these exonic variants

1Department of Internal Medicine, Seoul National University Hospital, Seoul, Korea; 2Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, Seoul, Korea; 3Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea; 4Biomedical Research Institute, Seoul National University Hospital, Seoul, Korea; 5Department of Pediatrics, Seoul National University College of Medicine, Seoul, Korea; 6Green Cross Genome, Yongin, Korea; 7Korean Hemophilia Foundation, Seoul, Korea; 8Cancer Research Institute, Seoul National University College of Medicine, Seoul, Korea; 9Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Korea and 10Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Korea

These authors contributed equally to this work.

Correspondence: Dr J-I Kim, Department of Biomedical Sciences, Seoul National University College of Medicine, 103 Daehak-ro, Jongno-gu, Seoul 03080, Korea.
E-mail: jongi@snu.ac.kr
or Dr KS Park, Department of Internal Medicine, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea.
E-mail: kspark@snu.ac.kr
Received 5 February 2017; revised 27 March 2017; accepted 5 April 2017
might be in genes that result in disorders that can be asymptomatic for a long period and can be prevented or treated. The American College of Medical Genetics and Genomics (ACMG) has recommended reporting pathogenic or likely pathogenic variants in a minimum list of 56 genes associated with 24 medical conditions. In addition, the ACMG and the Association for Molecular Pathology (AMP) have published standards and guidelines for interpreting the pathogenicity of sequence variation. These standards and guidelines consist of 28 attributes for evaluating evidence, including population frequency, functional experiments, familial segregation, computational prediction and rules to classify variants into five categories of pathogenicity. However, Amendola et al. have suggested that these criteria need to be clarified to reduce errors in the application of the ACMG-AMP standards and guidelines and discrepancies between laboratories. Although several studies have investigated the incidence of secondary findings in exome sequencing studies, secondary findings in East Asian populations have not been thoroughly investigated on a large scale using the strict criteria of the recent ACMG-AMP standards and guidelines.

On the other hand, there are loci that are prone to misinterpretation with regard to variant pathogenicity. These loci could be either (1) susceptible to sequencing errors or (2) highly polymorphic with excess nonsynonymous variants. There have been limited attempts to identify false-positive signals by filtering highly variable genes and to screen loci with excess heterozygosity in WES. Comprehensive investigations are required to identify genes that should be cautiously interpreted when WES is performed to identify disease-causing mutations. In addition, it is still unknown how these loci differ depending on the sequencing methods or the ethnicity of the study population. In this KORean whole EXome sequencing (KOEX) study, we investigated the following: (1) the characteristics of variants identified from 1303 Korean exomes, (2) the secondary findings of 56 ACMG-recommended genes and (3) genes that should be cautiously interpreted for pathogenicity.

**MATERIALS AND METHODS**

**Study design and participants**

We investigated exonic variants identified from 1303 participants of a KOEX study. The KOEX study consisted of three individual WES projects. Project 1 was the Seoul National University Hospital type 2 diabetes case–control WES study (SNUH project 1), which consisted of 910 participants. Project 2 consisted of 191 normal healthy parents from the Seoul National University rare disease WES study (SNUH project 2). Project 3 consisted of 202 subjects from the Green Cross hemophilia WES study (Green Cross project).

The Institutional Review Board (IRB) of the Seoul National University Hospital (IRB No. H-1205-130-411 for SNUH project 1, H-1406-081-588 for SNUH project 2) and the Green Cross Laboratories (IRB No. GCRL 2014-02) approved the projects, and written informed consent was obtained from each participant. Brief descriptions of each project are shown in Table 1.

**WES and variant calling**

Table 1 shows the whole-exome capture kit, sequencing system and median coverage information for each project. Paired-end sequence reads were aligned to the human reference genome (GRCh37). All data were processed using BWA, Picard software (http://broadinstitute.github.io/picard/), GATK pipelines of the Broad Institute to align the sequence reads. Variant calling was performed using GATK HaplotypeCaller in GVCF mode, and CombineGVCF was employed to merge the data into each cohort. GenotypeGVCF and VariantQualityScoreRecalibration (VQSR) were performed as recommended by the developers. ApplyRecalibration was performed with option -ts-filter-level 99.5 for single-nucleotide variants (SNVs) and 99.0 for insertion/deletions (INDELs).

**Quality control of WES and variant annotation**

To effectively remove sites with low depth (DP), genotype quality (GQ) and call rate, we imposed stringent genotype level filters: variants were called at sites where the DP was $\geq 7$, the GQ was $\geq 20$ and the call rate was $>0.90$. Bi-allelic variants with significant deviation

---

**Table 1** Brief description of studies and WES procedures

| Description of project | N  | Sex (M/F) | Sequencing system | Exome capture kit (targeted region size) | Median coverage (Min, Max) |
|------------------------|----|-----------|-------------------|----------------------------------------|--------------------------|
| SNUH project 1         | 910| 415/495   | Hiseq 2000        | SureSelect v4+UTR (71 Mb)               | 103.7 (66.9, 175.0)      |
| Type 2 diabetes mellitus whole-exome sequencing study | | | | | |
| SNUH project 2         | 191| 98/93     | Hiseq 2500        | NimbleGen SeqCapV2 (44 Mb)             | 65.2 (38.4, 118.1)       |
| Phenotypically normal parents of rare disease patients | | | | | |
| Green Cross project    | 202| 202/0     | Hiseq 2000        | SureSelect v5+UTR (75 Mb)              | 58.9 (32.3, 120.0)       |
| Hemophilia case study  | | | | | |

Abbreviations: F, female; M, male; Mb, mega base pair; N, sample size; UTR, untranslated regions. The present study is based on whole-exome sequence data of 1303 Koreans participating in three cohort studies. Collected data yield median on-target coverage between 58.9 × and 103.7 × by each cohort, producing high-quality sequencing data.
were adopted.9 Methods for applying each evidence criterion were described in the guidelines recommended by ACMG-AMP.8 Among all non-silent variants, we selected either low- or high-confidence disease-causing mutations using HGMD.22 We further applied a MAF filter of <0.5% and the evidence for pathogenicity was evaluated for 28 attributes. These data were combined using a scoring rule to classify each variant as either pathogenic (P), likely pathogenic (LP), benign (B), likely benign (LB) or variant of uncertain significance (VUS). Guidelines were strictly followed, and recently suggested modifications were adopted.9 Methods for applying each evidence criterion were similar to a recent report.12 Three investigators reviewed all the evidence attributes for each variant independently and made a consensus agreement for variants with a P or a LP classification. All lines of evidence were manually reviewed for P or LP variants, known pathogenic variants, and truncating variants. Detailed methods for variant pathogenicity classification are described in Supplementary Methods. We calculated the allele count and carrier frequency (95% confidence interval (CI)) of P or LP variants in our study participants. The CI of carrier frequency was calculated using a modified Wald method.27

Identification of loci prone for misinterpretation
We evaluated loci that are prone to misinterpretation and compiled a list of genes that should be provisionally excluded when investigating disease-causing mutations. First, we listed genes with low-quality protein coding DNA sequence (CDS) variants that were filtered during quality control procedures such as GATK-VQSR and HWE tests. We defined genes with (1) more than 100 variants filtered by VQSR or (2) more than five coding variants with significant deviation from HWE (P<0.001), as genes susceptible for sequencing errors. Second, we identified highly polymorphic loci by evaluating the per-gene metric for the nonsynonymous variant burden for each sample, considering the allele frequency and gene size in combination. Using the longest transcript of each gene, 10 714 RefSeq genes on autosomes with CDS lengths longer than 1 kb were evaluated. Genes containing more nonsynonymous variants than the third (upper) quartile +1.5 × (interquartile range) in terms of numbers or rates (number of variants divided by the length of the transcript) were considered to have excess nonsynonymous variants, according to the outlier detection method of Tukey.28 We then further defined highly polymorphic genes if the gene showed excess nonsynonymous variants in at least two of the following four categories: (1) excess absolute number of entire nonsynonymous variants, (2) excess absolute number of rare (MAF<0.5%) nonsynonymous variants, (3) excess rate of entire nonsynonymous variants and (4) excess rate of rare nonsynonymous variants. The same rule was also applied to the 1000 Genomes Projects to validate the misinterpretable genes suggested in this study.

RESULTS
Characteristics of variants identified by WES
A total of 1303 participants from three WES projects were included in the KOEX study. Brief descriptions of the three WES projects are shown in Table 1. After filtering low-quality samples and variants, we identified 495 729 unique variants at exonic regions (Table 2). The number of variants that were identified in each WES project and their overlap are displayed in Supplementary Figure 1. A total of 293 048 variants were located at CDS (169 380 nonsynonymous, 1665 splicing, 3642 stop gain/loss, 107 148 synonymous SNVs, and 4356 frameshift and 3221 in-frame INDELs). On average, each individual had 7136 nonsynonymous SNVs and 74 frameshift INDELs located in CDS. In addition, there was on average 177 variants that were predicted to result in protein truncation (splicing, stop gain/loss and frameshift). Further detailed information regarding the frequency distribution of INDELs with regard to its length, singleton and doubleton variant counts according to exonic variant annotation is shown in Supplementary Figure 1. There was a relatively large overlap between variants identified in this study and those of East Asian participants of the Exome Aggregation Consortium (59.1% of SNVs) and East Asian participants of the 1000 Genomes Project (34.7% of SNVs). We identified 76 607 novel coding variants (73 241 SNVs and 3366 INDELs) not cataloged in dbSNP build 147. Most of these variants were very rare (singleton: 87.0%, doubleton: 9.2%) or rare (MAF<0.5%: 99.8%). Population stratification and principal component analyses showed
that Korean participants clustered with East Asians of the 1000 Genomes Project and were separated from other populations (Figure 1).29,30

Secondary findings in the 56 ACMG recommended genes
A total of 1049 unique non-silent variants (1004 SNVs and 45 INDELs) were identified in 53 out of the 56 ACMG recommended genes (Supplementary Table 1). On average, there were 19.8 non-silent variants per gene for genes with at least one non-silent variant. For each individual, there was 19.8 non-silent variants per gene for genes with at least one non-silent variant. For each individual, there was at least one non-silent variant. For each individual, there was an average of 28.7 non-silent variants in these 53 genes. Among 1049 variants, 150 SNVs had a population MAF less than 0.5% and were reported as disease-causing mutations (with either high or low confidence); MAF, minor allele frequency; UTR, untranslated region.

Variant statistics according to functional annotation and frequency bin are shown. For functional CDS variants, the number of SNVs is shown with the number of INDELs given in parentheses. Calculations for UTR variants were performed with SNUH project 1 and the Green Cross project.

Table 2 Overall and per-sample variant statistics

| Genomic function | Total | AC = 1 | AC = 2 | MAF < 0.5% | MAF 0.5–5.0% | MAF ≥ 5% | Per sample |
|------------------|-------|--------|--------|------------|-------------|----------|------------|
| **SNV**          |       |        |        |            |             |          |            |
| CDS              | 284991| 145136 | 34162  | 229168     | 25722       | 30101    | 16070      |
| Nonsynonymous    | 169380| 91616  | 20917  | 142224     | 13680       | 13476    | 7136       |
| Synonymous       | 107148| 48484  | 12310  | 79703      | 11462       | 15983    | 8591       |
| Splice site      | 1665  | 1091   | 177    | 1498       | 96          | 71       | 38         |
| Stop gain/loss   | 3642  | 2454   | 405    | 3363       | 177         | 102      | 48         |
| Not in dbSNP 147 | 73241 | 63651  | 6807   | 73137      | 104         | 0        | 68         |
| UTR              | 181064| 83062  | 20659  | 1340722    | 19814       | 27178    | 13130      |
| **Indel**        |       |        |        |            |             |          |            |
| CDS              | 8057  | 4854   | 939    | 6979       | 634         | 444      | 224        |
| Frameshift       | 4356  | 2888   | 483    | 3920       | 276         | 160      | 74         |
| In-frame         | 3221  | 1701   | 402    | 2663       | 308         | 250      | 120        |
| Splice site      | 235   | 131    | 22     | 190        | 26          | 19       | 15         |
| Stop gain/loss   | 121   | 75     | 15     | 107        | 10          | 4        | 2          |
| Not in dbSNP147  | 3366  | 2959   | 260    | 3337       | 25          | 4        | 3          |
| UTR              | 21617 | 9449   | 2365   | 15606      | 2652        | 3359     | 1577       |

Functional CDS variants

| Genomic function | Total (AC = 1) | AC = 2 | MAF < 0.5% | MAF 0.5–5.0% | MAF ≥ 5% | Per sample |
|------------------|---------------|--------|------------|-------------|----------|------------|
| **HGMD-DM**      | 2897 (253)    | 1279 (146) | 351 (34)  | 2292 (222) | 431 (23) | 174 (8)    | 84.1 (3.0) |
| **ClinVar-P**    | 500 (36)      | 226 (21) | 57 (7)    | 389 (34)  | 74 (2)   | 37 (0)     | 19.2 (0.1) |

Abbreviations: AC, allele count; CDS, coding sequence; ClinVar-P, Pathogenic variant in the ClinVar database; HGMD-DM, Human Gene Mutation Database disease-causing variants (either low or high confidence); MAF, minor allele frequency; UTR, untranslated region.

For functional CDS variants, the number of SNVs is shown with the number of INDELs given in parentheses. Calculations for UTR variants were performed with SNUH project 1 and the Green Cross project.
Figure 1. Population stratification and principle component analyses. The merged data of 1303 Korean participants and the 1000 Genomes Project were investigated (a) for population structure analysis using the ADMIXTURE and (b) for principal component analysis. The Korean participants of our study clustered with East Asians and were separated from other populations. ACB, African Caribbeans in Barbados; AFR, African; AMR, American; ASW, Americans of African Ancestry in SW USA; BEB, Bengali from Bangladesh; CDX, Chinese Dai in Xishuangbanna, China; CEU, Utah Residents (CEPH) with Northern and Western Ancestry, USA; CHB, Han Chinese in Beijing, China; CHS, Southern Han Chinese; CLM, Colombians from Medellin, Colombia; EAS, East Asian; ESN, Esan in Nigeria; EUR, European; FIN, Finnish in Finland; GBR, British in England and Scotland, UK; GIH, Gujarati Indian from Houston, Texas, USA; GWD, Gambian in Western Divisions in the Gambia; IBS, Iberian Population in Spain; ITU, Indian Telugu from the UK; JPT, Japanese in Tokyo, Japan; KHV, Kinh in Ho Chi Minh City, Vietnam; KOR, Korean; LWK, Luhya in Webuye, Kenya; MSL, Mende in Sierra Leone; MXL, Mexican Ancestry from Los Angeles, USA; PEL, Peruvians from Lima, Peru; PJL, Punjabi from Lahore, Pakistan; PUR, Puerto Ricans from Puerto Rico; SAS, South Asian; STU, Sri Lankan Tamil from the UK; TSI, Toscani in Italy; YRI, Yoruba in Ibadan, Nigeria.
highly polymorphic genes were evaluated by assessing the burden of nonsynonymous variants. The excess of nonsynonymous variants was estimated using per-gene, per-sample statistics using a combination of burden (number versus rate) and allele frequency (entire versus rare variant). Genes with excess numbers of nonsynonymous variants that were relatively large were associated with extracellular matrix organization and the cytoskeleton (data not shown). By contrast, genes with excess rates of nonsynonymous variants that were mostly small were associated with olfactory transduction, keratin filament and the plasma membrane. We externally validated these findings using the 1000 Genomes Project data (Supplementary Table 5).

DISCUSSION

In this KOEX study, we comprehensively investigated genetic variations in a Korean population by analyzing the data obtained using high-quality whole-exome deep sequencing. Among the 1303 participants, we identified 495 729 unique

Table 3 Allele count and carrier frequency of P or LP variants in 56 clinically actionable genes

| Allele count | Carrier frequency | Allele count | Carrier frequency | Allele count | Carrier frequency | Allele count | Carrier frequency |
|--------------|------------------|--------------|------------------|--------------|------------------|--------------|------------------|
| P            | 12               | 1.32% (0.73–3.22%) | 2               | 1.05% (0.04–3.98%) | 1           | 0.50% (0.01–3.03%) | 15           | 1.15% (0.68–1.91%) |
| LP           | 13               | 1.43% (0.81–2.45%) | 2               | 1.05% (0.04–3.98%) | 2           | 0.99% (0.04–3.77%) | 17           | 1.30% (0.80–2.10%) |
| P+LP         | 25               | 2.75% (1.85–4.04%) | 4               | 2.09% (0.63–5.45%) | 3           | 1.49% (0.30–4.48%) | 32           | 2.46% (1.73–3.46%) |
| VUS          | 830              | 91.2% (89.2–92.9%) | 153             | 80.1% (73.8–85.2%) | 185         | 91.6% (86.9–94.8%) | 1168         | 89.6% (87.9–91.2%) |

Abbreviations: LP, likely pathogenic; P, pathogenic; VUS, variant of uncertain significance.
Allele count and carrier frequency of P or LP variants are shown for each WES project. Data are shown as the number, frequency (95% confidence interval).

The carrier frequency of VUS was calculated using the proportion of subjects with at least one VUS. A modified Wald test was used to calculate 95% confidence intervals.

![Figure 2](image-url)
Figure 3 Characterization of highly misinterpretable loci. (a) Example of a sequencing error prone gene, MUC6, showing excess coverage (gray) and imbalanced allelic fraction (blue) of VQSR filtered variants. (b) Distribution of variants with significant deviation from HWE (green), indicating sequencing error prone loci. (c–f) The frequency distribution of genes according to the burden of nonsynonymous variants and their cutoff values for highly polymorphic genes. The cutoff values for (c) excess number of entire nonsynonymous variants was 1.88, (d) excess rate of entire nonsynonymous variants was $0.78 	imes 10^{-3}$, (e) excess number of rare nonsynonymous variants was 0.064 and (f) excess rate of rare nonsynonymous variants was $0.027 	imes 10^{-3}$. The dashed line indicates the cut-off value for each category. The Venn diagrams show how the highly polymorphic genes in Koreans overlap with the 1000 Genomes Project (1KGP). VQSR, Variant Quality Score Recalibration.
We applied the guidelines we also tried to follow the specific details as published in recent studies. Whether carrier frequency varies depending on ethnicity and, if so, by how much, is unknown. There are suggestions that certain populations might have been underrepresented in the literature and in clinical genetics databases. As we filtered SNVs that were annotated in HGMD as disease-causing mutations for further clinical interpretation of pathogenicity, we might have underestimated the carrier frequency in our study population. Furthermore, a significant proportion of our participants were patients at SNUH, which is one of the largest tertiary hospitals in Korea. These participants might have complex medical presentations, and it is possible that our estimation could be different from that of the general population.

There could be several reasons why false positive genotypes are obtained when WES is performed. Even after these sequencing errors are excluded, there could still be loci that are highly polymorphic and have an excess of nonsynonymous variants. Here, we presented genes that should be cautiously interpreted as disease causing. Loci that were prone to sequencing error were evaluated by genotype quality (VQSR) and excess of heterozygosity (HWE). We further listed genes that were highly polymorphic by evaluating the number or rate of nonsynonymous variants (either entire or rare variants) per-gene for each sample. We provide a full list of genes that were filtered using these methods. The specific cut-off values used to define sequencing error prone genes and highly polymorphic genes were arbitrary. However, genes listed as prone to sequencing error had characteristics of segmental duplication, excess coverage, and excess allelic imbalance. Furthermore, genes listed as being highly polymorphic in our study were associated with categories that include olfactory receptors and keratin filaments known to contain multiple nonsynonymous variants. To evaluate whether the list of highly polymorphic genes was valid, we applied the same filter to the 1000 Genomes Project data and found a similar increased burden of nonsynonymous variants in the identified genes. It would be useful to further narrow down specific regions within the listed genes for cautious interpretation of pathogenicity.

There are certain limitations to this study. First, the sample size was modest. The lowest detectable MAF in our study was 0.038%. Nevertheless, this is one of the first major studies to investigate high-quality deep sequenced variants in East Asians. In addition, this is the largest WES study of Korean populations. Second, different whole-exome capture kits were used for each of the three projects. Although the target region covered for each project was different, most of the coding region in which we were primarily interested overlapped for the three capture kits. Third, large INDELs that might be important for certain Mendelian disorders were not included in this study. Methods for identifying large INDELs in WES are still incomplete. Further improvements and validations are required for these methods.

In conclusion, we have evaluated genetic variants in 1303 Korean exomes and provide ethnically specific variation information, including novel rare functional variants. The proportion of P or LP variants in our study population was estimated
to be 2.46%, which was comparable to the levels reported for other populations. Finally, we suggest a method for filtering genes that are prone to sequencing errors or that are highly polymorphic with excess nonsynonymous variants. We also provide a list of genes that require caution when interpreting disease causality. As WES is being increasingly used for both clinical and research purposes, our KOEX study results should serve as a valuable resource, especially regarding the exclusion of false positive findings and identifying true disease-causing pathogenic variants. Furthermore, large-scale sequencing studies are expected to broaden the catalog of rare variants and should accelerate the realization of precision medicine.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI15C1595, HI14C0006, HI15C3131, HI13C2148 and HI13C1468).

1 Collins FS, Hamburg MA. First FDA authorization for next-generation sequencer. N Engl J Med 2013; 369: 2369–2371.
2 Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y et al. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA 2014; 312: 1870–1879.
3 Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ et al. The genetic architecture of type 2 diabetes. Nature 2016; 536: 41–47.
4 The 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM et al. A global reference for human genetic variation. Nature 2015; 526: 68–74.
5 Fu W, O’Connor TD, Jun G, Kang HM, Abecasis G, Leal SM et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature 2013; 493: 216–220.
6 Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016; 536: 285–291.
7 Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med 2013; 15: 565–574.
8 Richards S, Aziz N, Bale S, Bick D, Brandt L et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015; 17: 405–423.
9 Amendola LM, Jarvik GP, Leo MC, McLaUnghlin HM, Akkar Y, Amaral MD et al. Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. Am J Hum Genet 2016; 98: 1067–1076.
10 Dorschner MO, Amendola LM, Turner EH, Robertson PD, Shirts BH, Gallego CJ et al. Actionable, pathogenic incidental findings in 1000 participants’ exomes. Am J Hum Genet 2013; 93: 631–640.
11 Amendola LM, Dorschner MO, Robertson PD, Salama JS, Hart R, Shirts BH et al. Actionable exonic incidental findings in 6503 participants: challenges of variant classification. Genome Res 2015; 25: 305–315.
12 Maxwell KN, Hart SN, Vijai J, Schrader KA, Slavin TP, Thomas T et al. Evaluation of ACMG-Guideline-Based Variant Classification of Cancer Susceptibility and Non-Cancer-Associated Genes in Families Affected by Breast Cancer. Am J Hum Genet 2016; 98: 801–817.