

Title: Actionable secondary findings from whole-genome sequencing of 954 East Asians

Authors: Clara S Tang, PhD^{1,2*}, Saloni Dattani, BSc^{3,4}, Man-ting So, MPhil¹, Stacey S Cherny, PhD^{3,5}, Paul KH Tam, MD^{1,2}, Pak C Sham, MD, PhD^{3,5}, Maria-Mercè Garcia-Barcelo, PhD^{1*}

Affiliations:

1. Department of Surgery, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China
2. Dr. Li Dak-Sum Research Centre, The University of Hong Kong – Karolinska Institutet Collaboration in Regenerative Medicine, Hong Kong, China
3. Department of Psychiatry, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China
4. Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom
5. Centre for Genomic Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China

***Correspondence:**

Dr. Clara Sze-man TANG: claratang@hku.hk

Dr. Maria-Mercè GARCIA-BARCELO: mmgarcia@hku.hk

Division of Paediatric Surgery, Department of Surgery, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China

Tel : +852 28315073

Fax : +852 22554850

Abstract

Recently, the American College of Medical Genetics (ACMG) recommended the return of actionable secondary findings detected from clinical sequencing. The reported frequency of secondary findings in Asian populations were highly variable and it is unclear whether the uniformity in coverage offered by whole-genome sequencing (WGS) may impact the estimate. In this analysis, we aimed to refine the rate of secondary findings on East Asians through a large-scale WGS study. We classified 1,256 protein-altering or splicing variants of the 59 actionable genes detected from WGS of 954 East Asians in strict accordance with the ACMG and the Association for Molecular Pathology guidelines. A total of 21 pathogenic or likely pathogenic variants were detected in 24 of the 954 East Asian genomes with an estimate of 2.5% of East Asians carrying actionable variants. Although the overall estimate of secondary findings was consistent with those reported for non-East Asian ethnicities, genetic and allelic heterogeneity were observed. WGS offers a wider breadth of coverage over WES, which highlights the need to further investigate the variable sensitivity of WES and WGS in the detection of secondary findings. Identifying secondary findings in populations underrepresented in previous genetic literature might improve variant interpretation and has a profound impact on local decision-making with regard to the cost-effectiveness of returning the secondary findings from clinical sequencing.

Keywords: Secondary findings, whole-genome sequencing, East Asians, actionable genes

Introduction

With the decline in the cost of next generation sequencing (NGS), genome-scale sequencing, i.e. whole-exome (WES) and whole-genome (WGS) sequencing, has been increasingly adopted in clinical practice for molecular diagnosis of complex syndromes, particularly for rare congenital disorders. Complex issues, both ethnical and social, arise regarding the report of secondary findings (SF) unrelated to the primary medical condition in which the original sequencing is targeted for.

In 2013, the American College of Medical Genetics and Genomics (ACMG) released recommendations for reporting of secondary findings in clinical sequencing for a minimum list of 56 medically actionable genes on 24 rare Mendelian conditions (Green, Berg et al. 2013). These genes were selected on the basis of medical actionability, with substantial clinical evidence for the known or expected pathogenic variants resulting in a high penetrance of severe disease that is preventable prior to the occurrence of symptoms. The list was later revised to cover 59 actionable genes for 27 conditions in ACMG SF v2.0 (Kalia, Adelman et al. 2017). Evidence-based guidelines were also recommended by ACMG and Association for Molecular Pathology (AMP) in 2015 to standardize the clinical interpretation of sequence variants in general (Richards, Aziz et al. 2015).

Thus far, a number of studies have reported secondary findings found from genome-scale sequencing (Amendola, Dorschner et al. 2015, Jang, Lee et al. 2015, Dewey, Murray et al. 2016, Januar, Kuan et al. 2016, Natarajan, Gold et al. 2016). Most of these studies were conducted on WES data of Caucasians and not all of them adhered to the ACMG-AMP guidelines for variant classification. Among these studies, the reported incidence rate of pathogenic (P) or likely

pathogenic (LP) variants for the 56 actionable genes varied considerably, from 1% in Africans(Amendola, Dorschner et al. 2015) to 3.3% in Caucasians(Dewey, Murray et al. 2016). Similar figures were limited for Asians, with only two relatively small studies reporting highly variable frequencies of 1.6% in Singaporeans(Jamuar, Kuan et al. 2016) versus 6.1% in Koreans (after reclassification of a *RYRI* truncating variant to uncertain significance (VUS))(Jang, Lee et al. 2015). The variation in estimated secondary finding frequency could be due to the inconsistent use of variant classification scheme, the deficit of clinical database to catalog non-Caucasian pathogenic variants, or the true ancestral difference in disease prevalence. With the decline in the cost of WGS, it is expected that more WGS will be applied for clinical diagnosis in the future. While it is recognized that some exonic regions might not be well covered in WES due to the incomplete and uneven coverage, regardless of the high mean sequencing depth, the frequencies provided by these WES-based studies could be underestimated if WGS is to be applied(Meienberg, Zerjavic et al. 2015, Meienberg, Bruggmann et al. 2016).

Here, we report the frequency of secondary findings obtained from high-coverage WGS of 954 unrelated East Asians across the newly revised list of 59 actionable genes based on ACMG SF v2.0. By following the ACMG-AMP guidelines in variant classification, our study provides a refined estimate of secondary findings in East Asians with the use of WGS.

Materials and Methods

Subjects

WGS was performed on 447 patients suffering from Hirschsprung disease (HSCR), a congenital genetic disorder of colonic aganglionosis, and 507 non-HSCR subjects of East Asian ancestry from university, academic hospitals and general clinics from Hong Kong, China and Hanoi,

Vietnam, who were ascertained previously for a case-control association study on genetic factors predisposed to HSCR. Cases and controls were matched by sex and origin of sample collection. Except for reports on the association between HSCR and multiple endocrine neoplasia type 2, none of the remaining 26 conditions recommended for return from the ACMG SF v2.0 list are related to the genetic etiology of HSCR. Demographic information of these samples was provided in Table S1. Principal component analysis together with the 1000Genomes East Asian subpopulation confirmed the ancestry of these samples (Fig. S1). The informed consent was obtained from each participant. The study was approved by the institutional review board of The University of Hong Kong together with the Hospital Authority (IRB:UW 06–349 T/1374).

Sequencing and variant calling

Genomic DNA extracted from blood was assessed for quality by PicoGreen and gel electrophoresis. Library preparation was done in-house (The University of Hong Kong, Centre for Genomic Sciences) using TruSeq Nano DNA library preparation kit and then sequenced at 30X coverage using Illumina HiSeqX Ten (150-bp paired-end sequencing) by Macrogen. The paired-end sequence reads were processed according to the GATK best practice recommendations (DePristo, Banks et al. 2011). Sequencing output metrics are detailed in Table S2. Variants, including SNPs and insertions/deletions (indels), were called by GATK Haplotype Caller (v3.4) and variant quality recalibration was performed as variant-based quality control. We further performed genotype-based quality control by restricting the analysis to high quality calls with genotype quality (GQ)>20 and depth (DP)>8 using KGGseq (Li, Li et al. 2017).

Variant annotation and selection

Variant annotations, in regard to molecular function prediction, population allele frequencies in public databases and *in silico* prediction of deleterious effect, were carried out by KGGseq. We confined the analysis to the list of 59 actionable genes recommended by the ACMG SF v2.0(Kalia, Adelman et al. 2017) (Table S3). For each gene, we curated from literatures as well as ClinVar the most frequent isoform(s) documented for known pathogenic variants. A total of 2,133 coding or splicing variants were called for the 59 genes, of which 1,256 variants (1214 SNVs and 42 indels; Table S4) affect protein sequence or length (including nonsense, frameshift, splicing and missense variants, in-frame indels and start-loss variants) for the curated isoforms.

Known pathogenic (KP) variants

The 1,256 protein-altering or splicing variants were then cross-referenced with ClinVar for known pathogenicity. While variants in ClinVar are documented with varying levels of evidentiary support for clinical significance, we defined known pathogenic variants as those listed as “pathogenic” and have review status of 3 stars, i.e. reviewed by expert panels. Other variants reported as pathogenic but with review status of 2 stars, i.e. those with two or more submitters providing assertion criteria with the same interpretation, were subject to reclassification according to the ACMG-AMP guidelines. If there were sufficient evidence supporting these variants as pathogenic according to the guidelines, they were also interpreted as KP variants.

Variant classification

For variants not previously reported as pathogenic or insufficient supporting evidence for pathogenicity, we scored them with reference to the classification evidence suggested in the ACMG-AMP guidelines (Fig. S2). The evidence was later combined in accordance to the

suggested scoring rules to classify the variants into five tiers: P, LP, VUS, likely benign (LB) and benign (B)(Richards, Aziz et al. 2015). Detailed information about the criteria used for classification was provided in the Supplementary Materials and Methods.

Coverage of the 59 actionable genes in WGS and WES

For the WGS data, the depth of coverage was computed for the canonical transcripts of the curated isoforms using GATK DepthOfCoverage. For WES data, we downloaded the coverage metrics for exomes in the ExAC database, which record the per-base mean coverage as well as the proportion of samples with coverage above X where X is a range of coverage threshold (e.g. 10X, 20X, etc.). We considered 20X to be the minimum coverage to sensitively (>99%) detect heterozygous genotypes while most existing studies considered called variants with at least 10X(Dewey, Murray et al. 2016, Natarajan, Gold et al. 2016). As only protein-altering and splicing variants were included in assessing pathogenicity, we computed, per base of the coding exons (CDS), the proportion of samples with $\geq 10X$ and $\geq 20X$ coverage.

Results

Frequency of actionable secondary findings on 954 WGS samples

To provide a precise estimate of secondary findings in East Asians, we identified and categorized 1,256 protein-altering variants in the 59 actionable genes found in WGS of 954 East Asians. Following the ACMG-AMP variant classification recommendations, a total of 23 P/LP heterozygous variants were found in 26 individuals, including two *RET* missense mutations (p.Cys620Arg and p.Cys609Ser) identified in two patients with HSCR. While some pathogenic variants in the *RET* causing multiple endocrine neoplasia type 2 (MEN2) are associated with an

increased risk for HSCR, we excluded these two mutations found in our HSCR patients from our list of secondary findings. This resulted in a final set of 21 unique P/LP variants identified in 24 participants (12 cases and 12 controls), with one sample carrying one pathogenic variant and one likely pathogenic variant associated with two distinct medical conditions (Table 1). No actionable finding was discovered for the four genes newly added in the ACMG SF v2.0. We estimated that around 2.5% ($24/954=2.5\%$; 95% confidence interval (CI): 1.7-3.7%) of East Asians harbour deleterious variants that meet criteria for possible clinical action. Half of which (1.3%) carries KP variants (9 variants in 13 samples) with well-established evidence of pathogenicity. No difference in the overall rate of secondary findings was detected between cases and controls. The genes and the corresponding P/LP variants detected were summarized in Table 1. Criteria used to classify these variants were shown in Table S5.

Evidence of genetic heterogeneity by ethnicity

Among the 21 P/LP variants, two variants were detected in more than one sample, including a *APOB* missense change (p.Arg3527Trp; rs144467873) found in 4 individuals and a *BRCA2* nonsense mutation (p.Gln1037Ter; rs80358557) identified in 2 individuals each. These variants show large variation in allele frequency across ethnicities, with East Asian being the most common. The allele frequency of the *APOB* p.Arg3527Trp variant is 16 times higher in East Asians than Europeans (0.1% in ExAC East Asians compared to 0.006% in non-Finnish Europeans). This variant, like its counterpart p.Arg3527Gln in Europeans (0.04% in non-Finnish Europeans), was reported as the most frequent cause of familial hypercholesterolemia in Han Chinese (Chiou and Charng 2016). Similarly, *BRCA2* p.Gln1037Ter also accounted for a substantial proportion (~25%) of all *BRCA2* mutations found in breast and ovarian cancer

patients in Hong Kong(Kwong, Ng et al. 2012) and was predominantly found in patients of Asian ancestry. Interestingly, over 80% of our P/LP variants (17 out of 21; 95%CI: 60-92%) are indeed loss-of-function mutations. Such ratio is slightly higher than the reported proportion from similar studies with comparable or larger sample sizes, i.e. 48% (95%CI: 29-67%) in the African-American Jackson Heart Study(Natarajan, Gold et al. 2016) and 58% (95%CI: 43-72%) in the European DiscovEHR Study(Dewey, Murray et al. 2016).

Comparison of coverage between WGS and ExAC WES data for the 59 actionable genes

In fact, the existing reports of secondary findings on large cohorts are primarily conducted using WES under research settings, where variants were called with at least 10 sequencing reads (Dewey, Murray et al. 2016, Kwak, Chae et al. 2017). To specifically assess if WGS provides a wider breadth of coverage across the 59 actionable genes, we first compared the depth of our WGS against the ExAC database, which catalogues variants found in ~60,000 exomes sequenced using multiple exome capture kits and sequencing chemistries. The heterogeneity in capture allowed us to evaluate, in general, the sensitivity of detecting secondary findings using WGS against WES irrespective of a particular target capture. For both WGS and WES, most of the CDS regions (>97%) of the actionable genes were sequenced at $\geq 20X$ in at least half (50%) of the samples (Fig. 1). On the other hand, a larger proportion of the CDS regions could achieve $\geq 20X$ coverage (98% in WGS versus 85% in WES) as well as $\geq 10X$ coverage (99% in WGS versus 93% in WES) in the majority of samples (>90%) in WGS. Such uniformity offered a wider breadth of coverage and hence a high sensitivity across the entire exome in most of the samples. For example, the likely pathogenic *MYBPC3* p.Arg1073fs frameshift variant found in our study is well-covered ($>20X$) in the majority of the samples (~95%) in the gnomAD database;

however, it is of shallow coverage (<20X) in 40% of WES samples under the same database and can potentially be missed in WES study.

Discussion

In our study, we estimated that a nontrivial percentage of East Asians (24/954=2.5%; 95% CI: 1.7-3.7%) harbor P/LP variants across the 59 medically actionable genes. The rate is lower than the reported frequency in Koreans (12/196=6.1%; 95%CI: 3.5-10.3%) but is in agreement with those reported from large European cohorts (2-3.3%). The addition of four genes in the new recommendations did not substantially increase the rate of secondary finding. The disparity between our estimate and the Korean's is likely attributed to our large sample size, and hence a more precise estimate, as well as our strict adherence to the standard ACMG-AMP guidelines.

We also showed that WGS provides a wider breadth of coverage across the 59 actionable genes, such that some regions well covered by WGS could be missed in previous studies based on WES. Meanwhile, our WGS has a mean coverage of ~36X, which implies that the majority of the exonic region has lower absolute depth compared to WES. While sensitivity of variant calling is determined by both the depth and breadth of coverage, the extent to which they influence the sensitivity in WES compared to WGS remains uncertain. **Recent study (Meienberg, Bruggmann et al. 2016) using PCR-free WGS at 60X demonstrated unprecedented sensitivity over WES; however, the high cost of WGS at 60X renders its popularity and, more often, an average depth of 30-50X was chosen (Stavropoulos, Merico et al. 2016, Lionel, Costain et al. 2017).** A more in-depth analysis on the impact of varying sequencing depth and coverage on the sensitivity of detecting, particularly singleton variants in these 59 genes, from deep genome sequencing is warranted. In fact, our estimate of secondary finding frequency from WGS does not significantly differ from those obtained from WES, illustrating a rather consistent estimate

based on the ACMG guidelines across sequencing approaches. This might suggest a minimal impact of the wider breadth of coverage of WGS on the frequency estimate.

Nevertheless, genetic heterogeneity by ethnicity was observed compared to the published data. The difference was mostly driven by allelic heterogeneity in certain disorders but could also result from the lack of pathogenic variants reported on East Asians in public archives. The slightly higher proportion of loss-of-function secondary findings could be attributed to the fact that, for missense variant to be classified as P/LP, strong literature support on both functional and variant penetrance is needed. While existing publications are biased towards Caucasians, the deficit of evidence to support some truly pathogenic missense changes, particularly for non-Caucasian population-specific variants, could lead to a higher frequency of VUS in these populations. In the near future, as WES and WGS move into routine clinical practice, evidence regarding the penetrance and clinical impact of these variants will begin to accrue and some of these VUS can be reclassified into pathogenic or benign variants. By then, the frequency of secondary findings on East Asians will need to be revisited. Additional cost, policy and ethical questions regarding the re-contact of patients might also arise.

By linking with the electronic health records, recent studies on unselected populations demonstrated an increased aggregated risk of developing the relevant clinical features associated with the corresponding conditions for those carrying P/LP variants in any of the ACMG-actionable genes (Dewey, Murray et al. 2016, Natarajan, Gold et al. 2016). If the secondary findings were to be returned promptly, improved screening and early intervention could reduce their lifetime risk and improve the clinical outcome; however, it also poses a substantial economic burden to the public wealth fare, e.g. cost of counselling, diagnostic testing, and the lifetime cost of surveillance. With the rapid expansion of NGS into clinical diagnosis, it is now

timely to review the regional frequency of secondary finding that might impact local policy-making to maximize both the public and individual benefits.

In summary, our study provided a refined estimate of the frequency of actionable secondary findings based on a large group of East Asians while following the ACMG-AMP guidelines. Genetic heterogeneity by ethnicity was observed. In particular, the potential for a large number of VUS in populations with limited representation in genetic studies reporting variant pathogenicity suggests the need for periodic reclassification and re-contact in these groups, which might have a profound impact on local decision-making with regard to the cost-effectiveness of returning the secondary findings from clinical sequencing.

Conflict of interest: We declare no conflict of interest.

Acknowledgements

This work was supported by: Theme-based Research scheme T12C-714/14-R to P.T., Human Medical Research Fund (HMRF) 02131866 and 01121516 to M.M.G.B and 0451966 to C.S.T. and Seed Fund for Basic Research of The University of Hong Kong 201606159005 to C.S.T.

References

- Amendola, L. M., M. O. Dorschner, P. D. Robertson, J. S. Salama, R. Hart, B. H. Shirts, M. L. Murray, M. J. Tokita, C. J. Gallego, D. S. Kim, J. T. Bennett, D. R. Crosslin, J. Ranchalis, K. L. Jones, E. A. Rosenthal, E. R. Jarvik, A. Itsara, E. H. Turner, D. S. Herman, J. Schleit, A. Burt, S. M. Jamal, J. L. Abrudan, A. D. Johnson, L. K. Conlin, M. C. Dulik, A. Santani, D. R. Metterville, M. Kelly, A. K. Foreman, K. Lee, K. D. Taylor, X. Guo, K. Crooks, L. A. Kiedrowski, L. J. Raffel, O. Gordon, K. Machini, R. J. Desnick, L. G. Biesecker, S. A. Lubitz, S. Mulchandani, G. M. Cooper, S. Joffe, C. S. Richards, Y. Yang, J. I. Rotter, S. S. Rich, C. J. O'Donnell, J. S. Berg, N. B. Spinner, J. P. Evans, S. M. Fullerton, K. A. Leppig, R. L. Bennett, T. Bird, V. P. Sybert, W. M. Grady, H. K. Tabor, J. H. Kim, M. J. Bamshad, B. Wilfond, A. G. Motulsky, C. R. Scott, C. C. Pritchard, T. D. Walsh, W. Burke, W. H. Raskind, P. Byers, F. M. Hisama, H. Rehm, D. A. Nickerson and G. P. Jarvik (2015). "Actionable exomic incidental findings in 6503 participants: challenges of variant classification." Genome Res **25**(3): 305-315.
- Chiou, K. R. and M. J. Charng (2016). "Genetic diagnosis of familial hypercholesterolemia in Han Chinese." J Clin Lipidol **10**(3): 490-496.
- DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernysky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler and M. J. Daly (2011). "A framework for variation discovery and genotyping using next-generation DNA sequencing data." Nat Genet **43**(5): 491-498.
- Dewey, F. E., M. F. Murray, J. D. Overton, L. Habegger, J. B. Leader, S. N. Fetterolf, C. O'Dushlaine, C. V. Van Hout, J. Staples, C. Gonzaga-Jauregui, R. Metpally, S. A. Pendergrass, M. A. Giovanni, H. L. Kirchner, S. Balasubramanian, N. S. Abul-Husn, D. N. Hartzel, D. R. Lavage, K. A. Kost, J. S. Packer, A. E. Lopez, J. Penn, S. Mukherjee, N. Gosalia, M. Kanagaraj, A. H. Li, L. J. Mitnaul, L. J. Adams, T. N. Person, K. Praveen, A. Marcketta, M. S. Lebo, C. A. Austin-Tse, H. M. Mason-Suares, S. Bruse, S. Mellis, R. Phillips, N. Stahl, A. Murphy, A. Economides, K. A. Skelding, C. D. Still, J. R. Elmore, I. B. Borecki, G. D. Yancopoulos, F. D. Davis, W. A. Faucett, O. Gottesman, M. D. Ritchie, A. R. Shuldiner, J. G. Reid, D. H. Ledbetter, A. Baras and D. J. Carey (2016). "Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study." Science **354**(6319).
- Green, R. C., J. S. Berg, W. W. Grody, S. S. Kalia, B. R. Korf, C. L. Martin, A. L. McGuire, R. L. Nussbaum, J. M. O'Daniel, K. E. Ormond, H. L. Rehm, M. S. Watson, M. S. Williams, L. G. Biesecker, G. American College of Medical and Genomics (2013). "ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing." Genet Med **15**(7): 565-574.
- Jamuar, S. S., J. L. Kuan, M. Brett, Z. Tiang, W. L. Tan, J. Y. Lim, W. K. Liew, A. Javed, W. K. Liew, H. Y. Law, E. S. Tan, A. Lai, I. Ng, Y. Y. Teo, B. Venkatesh, B. Reversade, E. C. Tan and R. Foo (2016). "Incidentalome from Genomic Sequencing: A Barrier to Personalized Medicine?" EBioMedicine **5**: 211-216.

Jang, M. A., S. H. Lee, N. Kim and C. S. Ki (2015). "Frequency and spectrum of actionable pathogenic secondary findings in 196 Korean exomes." Genet Med **17**(12): 1007-1011.

Kalia, S. S., K. Adelman, S. J. Bale, W. K. Chung, C. Eng, J. P. Evans, G. E. Herman, S. B. Hufnagel, T. E. Klein, B. R. Korf, K. D. McKelvey, K. E. Ormond, C. S. Richards, C. N. Vlangos, M. Watson, C. L. Martin and D. T. Miller (2017). "Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics." Genet Med **19**(2): 249-255.

Kwak, S. H., J. Chae, S. Choi, M. J. Kim, M. Choi, J. H. Chae, E. H. Cho, T. J. Hwang, S. S. Jang, J. I. Kim, K. S. Park and Y. J. Bang (2017). "Findings of a 1303 Korean whole-exome sequencing study." Exp Mol Med **49**(7): e356.

Kwong, A., E. K. Ng, C. L. Wong, F. B. Law, T. Au, H. N. Wong, A. W. Kurian, D. W. West, J. M. Ford and E. S. Ma (2012). "Identification of BRCA1/2 founder mutations in Southern Chinese breast cancer patients using gene sequencing and high resolution DNA melting analysis." PLoS One **7**(9): e43994.

Li, M., J. Li, M. J. Li, Z. Pan, J. S. Hsu, D. J. Liu, X. Zhan, J. Wang, Y. Song and P. C. Sham (2017). "Robust and rapid algorithms facilitate large-scale whole genome sequencing downstream analysis in an integrative framework." Nucleic Acids Res **45**(9): e75.

Lionel, A. C., G. Costain, N. Monfared, S. Walker, M. S. Reuter, S. M. Hosseini, B. Thiruvahindrapuram, D. Merico, R. Jobling, T. Nalpathamkalam, G. Pellicchia, W. W. L. Sung, Z. Wang, P. Bikangaga, C. Boelman, M. T. Carter, D. Cordeiro, C. Cytrynbaum, S. D. Dell, P. Dhir, J. J. Dowling, E. Heon, S. Hewson, L. Hiraki, M. Inbar-Feigenberg, R. Klatt, J. Kronick, R. M. Laxer, C. Licht, H. MacDonald, S. Mercimek-Andrews, R. Mendoza-Londono, T. Piscione, R. Schneider, A. Schulze, E. Silverman, K. Siriwardena, O. C. Snead, N. Sondheimer, J. Sutherland, A. Vincent, J. D. Wasserman, R. Weksberg, C. Shuman, C. Carew, M. J. Szego, R. Z. Hayeems, R. Basran, D. J. Stavropoulos, P. N. Ray, S. Bowdin, M. S. Meyn, R. D. Cohn, S. W. Scherer and C. R. Marshall (2017). "Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test." Genet Med.

Meienberg, J., R. Bruggmann, K. Oexle and G. Matyas (2016). "Clinical sequencing: is WGS the better WES?" Hum Genet **135**(3): 359-362.

Meienberg, J., K. Zerjavic, I. Keller, M. Okoniewski, A. Patrignani, K. Ludin, Z. Xu, B. Steinmann, T. Carrel, B. Rothlisberger, R. Schlapbach, R. Bruggmann and G. Matyas (2015). "New insights into the performance of human whole-exome capture platforms." Nucleic Acids Res **43**(11): e76.

Natarajan, P., N. B. Gold, A. G. Bick, H. McLaughlin, P. Kraft, H. L. Rehm, G. M. Peloso, J. G. Wilson, A. Correa, J. G. Seidman, C. E. Seidman, S. Kathiresan and R. C. Green (2016). "Aggregate penetrance of genomic variants for actionable disorders in European and African Americans." Sci Transl Med **8**(364): 364ra151.

Richards, S., N. Aziz, S. Bale, D. Bick, S. Das, J. Gastier-Foster, W. W. Grody, M. Hegde, E. Lyon, E. Spector, K. Voelkerding, H. L. Rehm and A. L. Q. A. Committee (2015). "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 **17**(5): 405-424.

Stavropoulos, D. J., D. Merico, R. Jobling, S. Bowdin, N. Monfared, B. Thiruvahindrapuram, T. Nalpathamkalam, G. Pellicchia, R. K. C. Yuen, M. J. Szego, R. Z. Hayeems, R. Z. Shaul, M. Brudno, M. Girdea, B. Frey, B. Alipanahi, S. Ahmed, R. Babul-Hirji, R. B. Porras, M. T. Carter, L. Chad, A. Chaudhry, D. Chitayat, S. J. Doust, C. Cytrynbaum, L. Dupuis, R. Ejaz, L. Fishman, A. Guerin, B. Hashemi, M. Helal, S. Hewson, M. Inbar-Feigenberg, P. Kannu, N. Karp, R. Kim, J. Kronick, E. Liston, H. MacDonald, S. Mercimek-Mahmutoglu, R. Mendoza-Londono, E. Nasr, G. Nimmo, N. Parkinson, N. Quercia, J. Raiman, M. Roifman, A. Schulze, A. Shugar, C. Shuman, P. Sinajon, K. Siriwardena, R. Weksberg, G. Yoon, C. Carew, R. Erickson, R. A. Leach, R. Klein, P. N. Ray, M. S. Meyn, S. W. Scherer, R. D. Cohn and C. R. Marshall (2016). "Whole Genome Sequencing Expands Diagnostic Utility and Improves Clinical Management in Pediatric Medicine." NPJ Genom Med **1**.

FIGURE LEGENDS

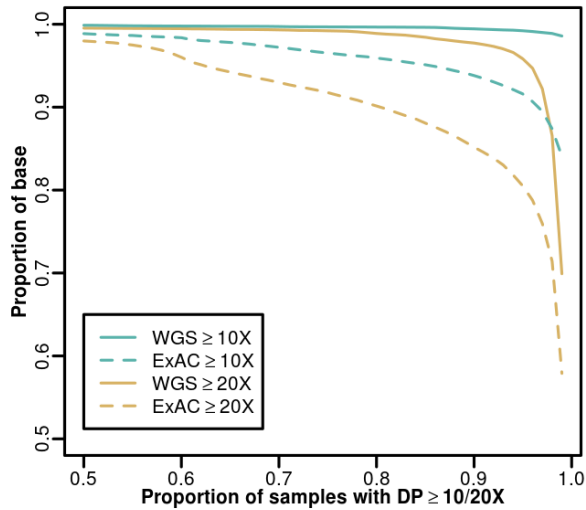


Figure 1: Uniformity in coverage of WGS. Proportion of coding region of the 59 actionable genes with coverage (DP) of at least 10X or 20X in a given proportion of samples is plotted. It shows that around 98% of the targeted region is covered by $\geq 20X$ in at least 90% of the WGS samples.

Table 1. Pathogenic and likely pathogenic variants in 59 ACMG SF v2.0 genes identified through whole-genome sequencing of 954 East Asians.

Associated condition(s)	Gene	Nucleotide change	Amino acid change	dbSNP ID	Heterozygous carriers	
					Chinese (n=793)	Vietnamese (n=161)
Known pathogenic variants						
Familial hypercholesterolemia	<i>APOB</i>	c.10579C>T	p.Arg3527Trp	rs144467873	4	0
Hereditary breast and ovarian cancer	<i>BRCA1</i>	c.5335delC	p.Gln1779fs	rs80357590	0	1 ^a
		<i>BRCA2</i>	c.2059_2063delGATTA	p.Asp687fs	rs587782780	1
		c.2808_2811delACAA	p.Ala938fs	rs80359351	1	0
		c.3109C>T	p.Gln1037Ter	rs80358557	2	0
		c.3599_3600delGT	p.Cys1200fs	rs80359391	1	0
		c.4471_4474delCTGA	p.Leu1491fs	rs80359451	1	0
Arrhythmogenic right-ventricular cardiomyopathy	<i>PKP2</i>	c.1613G>A	p.Trp538Ter	rs193922672	1	0
Hypertrophic cardiomyopathy	<i>MYH7</i>	c.4498C>T	p.Arg1500Trp	rs45544633	1	0
Likely pathogenic variants						
Familial hypercholesterolemia	<i>LDLR</i>	c.681C>A	p.Asp227Glu	rs121908028	1	0
Hereditary breast and ovarian cancer	<i>BRCA2</i>	c.2410G>T	p.Glu804Ter	NA	0	1
Arrhythmogenic right-ventricular cardiomyopathy	<i>PKP2</i>	c.1125_1132delTACTTTCA	p.Thr375fs	NA	0	1
		c.544delG	p.Glu182fs	NA	0	1
	<i>DSP</i>	c.5064delG	p.Ala1688fs	NA	1	0
	<i>DSC2</i>	c.2219C>A	p.Ser740Ter	NA	1	0
Hypertrophic cardiomyopathy	<i>MYBPC3</i>	c.3735delC	p.Pro1245fs	NA	1	0
		c.3217dupC	p.Arg1073fs	NA	1	0
		c.1504C>T	p.Arg502Trp	rs375882485	0	1
Brugada syndrome	<i>KCNH2</i>	c.2892delC	p.Pro964fs	NA	0	1 ^a
Lynch syndrome	<i>PMS2</i>	c.1206delA	p.Gln402fs	NA	1	0
Mafan syndrome, Loeys-Dietz syndromes, and familial thoracic aortic aneurysms and dissections	<i>MYH11</i>	c.4594_4595insGGTCCATGAGCTGG	p.Lys1532fs	NA	1	0

^a Same individual carry both *BRCA1* pathogenic and *KCNH2* likely pathogenic variants