EBioMedicine 5 (2016) 211–216

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Research Paper

Incidentalome from Genomic Sequencing: A Barrier to Personalized Medicine?

Saumya Shekhar Jamuar [a,b], Jyn Ling Kuan [c,h], Maggie Brett [d], Zenia Tiang [c,h], Wilson Lek Wen Tan [c,h], Jiin Ying Lim [a], Wendy Kein Meng Liew [a,b], Asif Javed [c], Woei Kang Liew [a], Hai Yang Law [a,b], Ee Shien Tan [a,b], Angeline Lai [a,b], Ivy Ng [a,b], Yik Ying Teo [e], Byrappa Venkatesh [f], Bruno Reversade [g], Ene Choo Tan [d], Roger Foo [c,h,*]

[a] Department of Paediatrics, KK Women's and Children's Hospital, Singapore
[b] Paediatric Academic Clinical Programme, Singhealth Duke-NUS Graduate Medical School, Singapore
[c] Genome Institute of Singapore, A*STAR, Singapore
[d] KK Research Center, KK Women's and Children's Hospital, Singapore
[e] Saw Swee Hock School of Public Health, National University of Singapore, Singapore
[f] Institute of Molecular and Cell Biology, A*STAR, Singapore
[g] Institute of Medical Biology, A*STAR, Singapore
[h] Cardiovascular Research Institute, National University of Singapore, National University Health System, Singapore

ARTICLE INFO

Article history:
Received 26 November 2015
Received in revised form 16 January 2016
Accepted 25 January 2016
Available online 4 February 2016

Keywords:
Incidental findings
Genomic sequencing
Personalized medicine

ABSTRACT

Background: In Western cohorts, the prevalence of incidental findings (IFs) or incidentalome, referring to variants in genes that are unrelated to the patient's primary condition, is between 0.86% and 8.8%. However, data on prevalence and type of IFs in Asian population is lacking.
Methods: In 2 cohorts of individuals with genomic sequencing performed in Singapore (total n = 377), we extracted and annotated variants in the 56 ACMG-recommended genes and filtered these variants based on the level of pathogenicity. We then analyzed the precise distribution of IFs, class of genes, related medical conditions, and potential clinical impact.
Results: We found a total of 41,607 variants in the 56 genes in our cohort of 377 individuals. After filtering for rare and coding variants, we identified 14 potential variants. After reviewing primary literature, only 4 out of the 14 variants were classified to be pathogenic, while an additional two variants were classified as likely pathogenic. Overall, the cumulative prevalence of IFs (pathogenic and likely pathogenic variants) in our cohort was 1.6%.
Conclusion: The cumulative prevalence of IFs through genomic sequencing is low and the incidentalome may not be a significant barrier to implementation of genomics for personalized medicine.

1. Introduction

Incorporation of whole genome or exome sequencing (WGS/WES), hereafter referred to as genomic sequencing, in medical practice raises the disquieting issue of incidental findings (IFs), which has important and potentially far-reaching implications (Green et al., 2012; Knoppers et al., 2013; Wolf et al., 2012; Roche and Berg, 2015; Hegde et al., 2015; Ayuso et al., 2015; Krier and Green, 2015). IF, also called secondary findings and occasionally referred to as incidentalomes, are mutations in genes unrelated to the primary condition (phenotype) of the patient (Krier and Green, 2013). As genomic sequencing is a phenotype-agnostic test, it is not surprising that detection of IFs is of major concern and requires the decision of whether and how return of these results to the individual should be practiced (Krier and Green, 2013). Another concern revolves around the additional burden this creates on the healthcare system. Individuals with medically actionable IFs will require long-term surveillance and anticipatory care, which is acceptable when it is appropriate, but may be hard to justify if there is only uncorroborated evidence for the pathogenicity of the mutation in question: e.g. even if the gene is a causal gene for the condition, the mutation could be a novel one and never reported before for the condition.

The American College of Medical Genetics and Genomics (ACMG) has recommended return of IFs for a minimum set of 56 actionable genes, where prevention and surveillance may significantly reduce mortality and morbidity (Green et al., 2013). While these 56 genes represent rare Mendelian disorders, backed by substantial years of prior research and clinical experience, mutations in these genes are indeed highly medically actionable and include well-publicized ones like *BRCA1* and *BRCA2*. By being classed as "medically actionable", these

* Corresponding author at: Cardiovascular Research Institute, Centre for Translational Medicine MD6, 14 Medical Drive, National University of Singapore, 117599, Singapore.
E-mail address: mdcrfsy@nus.edu.sg (R. Foo).

http://dx.doi.org/10.1016/j.ebiom.2016.01.030

genes were prioritized to include disorders where preventative measures and/or treatments are available. For example, patients with mutations in cardiomyopathy-causing genes such as *MYH7* may have annual electrocardiogram (ECG) and echocardiography. Individuals with pathogenic mutations in these genes might be asymptomatic for long periods of time and therefore amenable to early intervention and prevention to reduce mortality and long-term morbidity. Patients with pathogenic *BRCA1* and *BRCA2* mutations have an 80% and 45% risk of developing breast cancer, respectively (Ford et al., 1998). Identification of a pathogenic variant in *BRCA1/BRCA2* significantly reduces the risk of developing breast cancer as close surveillance by an oncologist with clinical assessment, self-examination, mammogram and/or breast MRI allows for early detection, which, in turn leads to reduced morbidity and improved survival (Krier and Green, 2015; Krier and Green, 2013; Green et al., 2013; Ford et al., 1998; Warner et al., 2004). However about 50% of women harboring *BRCA1/BRCA2* mutation do not have a family history (Loman et al., 2001) and, hence, screening for breast cancer may not be recommended in this group of women. Indeed the absolute prevalence of breast cancer in this group of women is unclear. The incidental identification of individuals with these disease causing variants therefore allows the recommendation of follow-up screening offers a net benefit to individuals and society. Overall, the selection of these 56 genes may be conservative because many other genes are becoming medically important and actionable by the month.

The European Society of Human Genetics (ESHG), on the other hand, has recommended against using genomic sequencing in the clinic, and instead recommends the use of targeted genomic tests; clearly in an effort to avoid the scenario of unexpected IF (van El et al., 2013). However, as the cost of genomic sequencing continues to drop, genomic sequencing in practice is inevitable and IF will be clearly an issue that we cannot avoid.

Although with good intent, the first issue that arises from such recommendations includes lack of data on the frequency of IFs to determine the burden on the testing laboratory as well as the referring physician and even the healthcare system. A recent review of exome sequencing data from 1000 individuals (500 European and 500 African descents) recruited in the National Heart, Blood and Lung Initiative (NHBLI) Exome Sequencing Project (ESP) estimated the prevalence of IFs at 3.4% for European descent and 1.2% for African descent (Dorschner et al., 2013). A follow-up study expanded to include 6503 individuals (4300 European and 2203 African ancestry) estimated the frequency of IFs at 1.7% for individuals of European ancestry and 1.0% for African ancestry (Amendola et al., 2015). In unrelated cohorts, IFs were detected in 8.8% of the participants recruited through National Institute of Health Undiagnosed Disease Program ($n = 543$) (Lawrence et al., 2014), 0.86% in the Baylor-Hopkins Center for Mendelian Genomics ($n = 232$) (Jurgens et al., 2015), and 1.9% in the UK WGS500 cohort ($n = 500$) (Taylor et al., 2015). Within Asia, a review of 196 Korean exomes detected IFs in 7% of control subjects ($n = 100$) and 6% of patients with disease ($n = 96$) (Jang et al., 2015). Although Singapore is a nation gearing up for genomics (Manolio et al., 2015), such data is lacking for an Asian population and hence, there are no policies and recommendations regarding IFs in Singapore. In this study, we set out to estimate the prevalence as well as define the types of IFs found in genomic sequencing that was performed in 2 cohorts of the Asian population in Singapore. Singapore, as an island country in South East Asia, is uniquely dominated by immigrant ethnic groups, comprising Chinese, Malays, and Indians. The IF analysis from our study should hence be representative of the profile in South East Asia.

2. Methods

2.1. Patient Recruitment

Individuals were recruited through institutional ethics review board approved genomics projects. Informed consent was obtained from the eligible individual (or parent/legal guardian, when the individual is a minor). Sequencing data of these individuals was anonymized, de-identified and analyzed in a cumulative manner.

2.2. Genomic Sequencing

Blood samples were obtained from the consented individuals and DNA was extracted by established methods. Samples were sequenced on HiSeq 2000/HiSeq 2500 or Ion Proton using established protocols. Data generated from genomic sequencing were aligned to the human reference genome using established bioinformatic algorithms and software (e.g. BWA-MEM followed by SAMtools to generate SAM/BAM files) (Li and Durbin, 2009). BAM files were processed using GATK (DePristo et al., 2011) to generate variant calling format (VCF) files.

2.3. Gene List Development

The list of genes we chose to analyze was confined to the 56 actionable genes recommended by ACMG (Supplementary Table 1) (Green et al., 2013). These genes were selected on the basis that deleterious variants would lead to specific conditions of high disease penetrance, for which evidence-based medical recommendations are available, implementation of which would arguably help towards preventing significant morbidity and mortality.

2.4. Bioinformatic Filtering

Variants were quality filtered to exclude false positives according to standard thresholds (Quality scores >30, coverage >10×, and absence of clustered variants within a window size of 10 variants). From variants that passed this threshold, we extracted variants in each of the 56 genes in our gene list (Appendix 1) (Green et al., 2013). We then annotated the variants using our in-house bioinformatic pipeline to include information regarding the gene, chromosomal coordinate(s), genetic change, protein change, type of mutation (frameshift, nonsense, nonsynonymous, splicing, and synonymous); prediction of the variant from multiple algorithms (Polyphen-2, Adzhubei et al., 2013, SIFT, Ng and Henikoff, 2003, likelihood ratio test and MutationTaster2, Schwarz et al., 2014), allele frequencies in different databases (Exome Sequencing Project, dbSNP, 1000 Genomes, Complete Genomics, Exome Aggregation Consortium, and our in-house database of common variants (present in >5% of the population)), and annotation of variants in clinical mutation databases like Clinvar (http://www.ncbi.nlm.nih.gov/clinvar/), and Human Genetic Mutation Database (http://www.hgmd.cf.ac.uk/ac/index.php) (Fig. 1). We then further analyzed the variant as per our filtering strategy illustrated in Fig. 1.

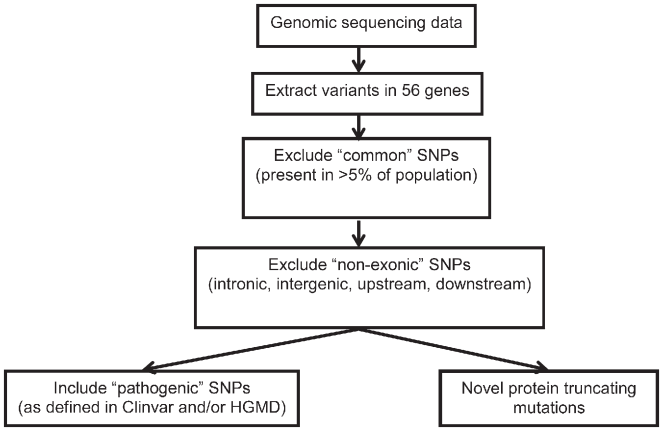

Fig. 1. Filtering strategy.