EBioMedicine 2 (2015) 74–81

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Original Article

Use of Whole Genome Sequencing for Diagnosis and Discovery in the Cancer Genetics Clinic

Samantha B. Foley [a,1], Jonathan J. Rios [b,c,1], Victoria E. Mgbemena [a], Linda S. Robinson [d], Heather L. Hampel [e], Amanda E. Toland [e], Leslie Durham [a,b], Theodora S. Ross [a,d,*]

[a] Department of Internal Medicine, UT Southwestern Medical Center, Dallas TX, 75390, USA
[b] Sarah M. and Charles E. Seay Center for Musculoskeletal Research, Texas Scottish Rite Hospital for Children, Dallas, Texas 75219, USA
[c] Department of Pediatrics, McDermott Center for Human Growth and Development, and Department of Orthopaedic Surgery, UT Southwestern Medical Center, Dallas, Texas 75390, USA
[d] Department of Cancer Genetics, UT Southwestern Medical Center, Dallas, TX, 75390, USA
[e] Department of Human Genetics, Ohio State University, Columbus, OH, 43210, USA

ARTICLE INFO

Article history:
Received 4 October 2014
Received in revised form 1 December 2014
Accepted 2 December 2014
Available online 3 December 2014

Keywords:
Cancer genetics
BRCA1/2
Whole-genome sequence
ClinVar
Pathogenic variants
Single nucleotide variant

ABSTRACT

Despite the potential of whole-genome sequencing (WGS) to improve patient diagnosis and care, the empirical value of WGS in the cancer genetics clinic is unknown. We performed WGS on members of two cohorts of cancer genetics patients: those with *BRCA1/2* mutations (n = 176) and those without (n = 82). Initial analysis of potentially pathogenic variants (PPVs, defined as nonsynonymous variants with allele frequency < 1% in ESP6500) in 163 clinically-relevant genes suggested that WGS will provide useful clinical results. This is despite the fact that a majority of PPVs were novel missense variants likely to be classified as variants of unknown significance (VUS). Furthermore, previously reported pathogenic missense variants did not always associate with their predicted diseases in our patients. This suggests that the clinical use of WGS will require large-scale efforts to consolidate WGS and patient data to improve accuracy of interpretation of rare variants. While loss-of-function (LoF) variants represented only a small fraction of PPVs, WGS identified additional cancer risk LoF PPVs in patients with known *BRCA1/2* mutations and led to cancer risk diagnoses in 21% of non-BRCA cancer genetics patients after expanding our analysis to 3209 ClinVar genes. These data illustrate how WGS can be used to improve our ability to discover patients' cancer genetic risks.

1. Introduction

With the rapid development and decreasing cost of whole-genome sequencing (WGS) technologies, this genetic testing tool will soon be readily available for common use in the laboratory and clinic (Collins and Hamburg, 2013). Though the research community is rejoicing in the newfound ability to use WGS to investigate patients' genomes in better detail, clinicians are more cautious. The unknown, but potentially significant, burden of delivering genetic results of uncertain value to patients weighs heavily on practitioners, such as oncologists, cardiologists and neurologists.

WGS, and whole-exome sequencing, have already been used to provide genetic diagnoses that inform clinical care (Worthey et al., 2011; Bainbridge et al., 2011; Rios et al., 2010). These early successes in individual patients prompted expanded studies to investigate a more general use of WGS in clinical settings (Saunders et al., 2012). Indeed the growing adoption of WGS in the clinic and the potential to positively impact patient care contributed, at least in part, to the UK100K Project, an effort by the Department of Health (United Kingdom) to provide high-coverage WGS for clinical interpretation in 100,000 participants focusing initially on rare diseases, cancer and infectious disease (www.genomicsengland.co.uk). Therefore, we sought to investigate the value of WGS in two cohorts of cancer genetics patients to begin to address the challenges associated with the identification and clinical interpretation of WGS and potentially pathogenic variants (PPVs) in the genetics clinic.

Although patients with a family history of cancer are currently evaluated with single-gene or gene panel tests, it is not clear whether WGS will replicate these findings or potentially increase the rate of identification of genetic risk factors. In this study, we modeled a scenario where WGS replaced gene-specific testing in 176 *BRCA1/2*-carriers and 82 non-*BRCA* patients from our cancer genetics clinics. We sought to examine whether WGS could be easily and quickly mined to identify PPVs

* Corresponding author at: 5383 Harry Hines Blvd Dallas, TX 75390, USA.
E-mail address: theo.ross@utsouthwestern.edu (T.S. Ross).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ebiom.2014.12.003

highly likely to increase cancer risk as well as potentially expand WGS to assess genetic risk for non-cancer conditions.

2. Methods

2.1. Participants

Individuals from the cancer genetics clinics of the University of Texas Southwestern Medical Center (UTSW) and the Ohio State University (OSU) cancer genetics programs were recruited to the study following informed consent approved by the Institutional Review Boards of both institutions. Only unrelated individuals were included in this study. Blood samples were obtained and de-identified. Subsequent genetic results were not returned to participants.

2.2. Whole-Genome Sequencing and Variant Analysis

WGS of DNA was performed by Complete Genomics Inc. (Mountain View, CA, USA) as previously described (Drmanac et al., 2010). Sequence analysis and variant detection were performed by Complete Genomics Inc., as well. Variant analysis was performed as previously described (Soyombo et al., 2013); however, variant quality measures were investigated to determine appropriate quality control parameters for identifying high-quality PPVs from WGS. To determine the variant quality score threshold, we measured genotype concordance between individual-matched WGS from lymphoblast and fibroblast samples from seven individuals, which are expected to have minimal discrepancies. Genotype concordance was measured using quality score thresholds ranging from 50 to 100 (Supplementary Fig. 1). Single nucleotide variants (SNVs) with quality scores less than 100 for both alleles were excluded, resulting in an average SNV genotype concordance rate of 98.88%. Because of the systematically lower concordance and higher error rate for detecting insertions and deletion (indels) compared to SNVs, indel concordance was measured using quality scores ranging from 0 to 300 (Supplementary Fig. 2). Indels with quality scores greater than 150 were included in the study, resulting in an average genotype concordance rate of 97.14%, although an average 63.07% of indels originally identified by Complete Genomics were excluded.

While numerous aspects of WGS performance are considered and represented in the variant quality score provided by Complete Genomics, individual quality parameters may improve WGS specificity and sensitivity, as are commonly used by other sequencing technologies. These include, but are not limited to, overall sequence depth, sequence coverage at variant positions, variant allele fraction, individual read quality and mapping quality, and read directionality for paired-end reads. Though these measures were not individually included in the analysis of WGS from Complete Genomics, the quality score procedure used here was previously shown to improve concordance to orthogonal validation by Sanger sequencing (Soyombo et al., 2013). All sequenced genomes were mapped to the human reference sequence (b37) and analyzed using Complete Genomic's software (version 2.4). A summary of sequencing statistics (coverage, amount of sequence, total variants and genome-wide QC measures) is reported per sample in Supplementary Table 1.

Potentially pathogenic variants (PPVs) were determined primarily by population frequency using the Exome Variant Server (ESP6500) and a population of HapMap individuals sequenced by Complete Genomics. To simplify detection of PPVs, nonsynonymous variants with frequency less than 1% in both the ESP6500 and HapMap datasets were considered potentially pathogenic. Loss-of-function (LoF) nonsynonymous variants were defined as SNVs predicted to create a premature truncation (nonsense), alter canonical splicing (disrupt), alter the initiating methionine of the protein (misstart), or alter the final stop codon of the transcript (nonstop). LoF indels included only those resulting in a frameshift and did not include in-frame deletions and insertions.

2.3. Statistical Analyses

All statistical analyses were performed using the R statistical framework. 95% confidence intervals were determined from the binomial probability. Fisher's exact test was used to test for significant differences between WGS and published gene-panel sequencing methods (Kurian et al., 2014).

3. Results

3.1. WGS Confirms Clinically Diagnosed BRCA1/2 Mutations

WGS was performed on a series of patients from the cancer genetics clinic that included those found to have *BRCA1* (n = 88; Supplementary Table 2) or *BRCA2* (n = 88; Supplementary Table 3) mutations, as well as those that were not carriers of a *BRCA1/2* mutation (n = 82; Supplementary Table 4). The genomes of the 176 unrelated *BRCA*-carriers at high risk for breast and ovarian cancer were first investigated to determine if WGS confirmed the clinically-diagnosed mutations (Myriad Genetics).

Similar to previous reports (Worthey, 2013), PPVs were defined as rare (less than 1% frequency in the ESP6500 and HapMap samples sequenced using the same technology) nonsynonymous variants. Because of the variability provided by different computational methods, *in silico* predictions of variant pathogenicity were not used to evaluate missense PPVs. Finally, due to technical limitations in detection of copy number variants (CNVs) and functional annotation of intronic intervening sequence (IVS) variants, these variants were not considered in our WGS analysis of PPVs. After applying quality control measures (see Methods), WGS identified the majority of clinically-diagnosed *BRCA1* and *BRCA2* mutations. Of the 75 patients with *BRCA1* mutations for which our method was expected to detect the clinically-diagnosed PPV, WGS detected 89.3% of the *BRCA1* mutations; the remaining eight mutations were identified in the WGS data but at lower quality (Supplementary Table 5). Among 88 patients with *BRCA2* mutations, WGS confirmed 88.6% of the *BRCA2* mutations; the remaining ten mutations were identified by WGS but again at lower quality (Supplementary Table 5). In sum, WGS in this cohort detected all *BRCA1/2* mutations expected to be identified by our WGS approach, although limitations in sequence quality prevented confident reporting of ~12% of *BRCA1/2* PPVs. We expect this result to be a lower-bound, as sequencing technologies and computational methods continue to improve. Indeed, 16 of 18 (89%) low quality *BRCA1/2* variants were indels, and indels were previously reported to be poorly sequenced using this WGS method (Drmanac et al., 2010). WGS results are summarized for all *BRCA1/2* mutations in Supplementary Tables 6 and 7.

3.2. WGS Detects Cancer-risk PPVs in BRCA1/2 Patients

To model clinical WGS in our patients, we evaluated 163 clinically-relevant genes (Table 1) to identify PPVs in the BRCA1/2 cohort. These genes included cancer genes evaluated on commercial cancer-susceptibility gene panels, genes selected by Dorschner et al. (2013) as genes that might impact reproductive decision making (e.g. carrier-status reporting) as well as genes initially recommended for reporting by ACMG (Green et al., 2013). In total, the initial gene panel represented 135 dominant, 24 recessive and 4 X-linked disease-associated genes.

WGS identified 1207 PPVs in the 176 patient genomes with *BRCA* mutations, representing 695 unique variants. Of these, most were missense variants and 46.33% of all PPVs were novel and not present in the ESP6500 database (Supplementary Table 8). On average, WGS identified 6.8 and 6.9 PPVs per patient in the *BRCA1*- and *BRCA2*-carrier cohorts, respectively (Fig. 1A). Among the 163 genes, the number of PPVs per gene varied greatly (Supplementary Tables 9 and 10). Similar gene variance results were obtained for *BRCA1*- or *BRCA2*-carrier cohorts evaluated separately (Supplementary Tables 11 and 12). As