Clinica Chimica Acta 471 (2017) 222–232

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/cca

Evaluation of whole exome sequencing by targeted gene sequencing and Sanger sequencing

Ya-Sian Chang[a,b,c], Hsien-Da Huang[d,e], Kun-Tu Yeh[f], Jan-Gowth Chang[a,b,g,h,*]

[a] Epigenome Research Center, China Medical University Hospital, Taichung, Taiwan
[b] Department of Laboratory Medicine, China Medical University Hospital, Taichung, Taiwan
[c] Department of Medical Laboratory Science and Biotechnology, China Medical University, Taichung, Taiwan
[d] Department of Biological Science and Technology, National Chiao Tung University, Hsinchu, Taiwan
[e] Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu, Taiwan
[f] Department of Pathology, Changhua Christian Hospital, Changhua, Taiwan
[g] School of Medicine, China Medical University, Taichung, Taiwan
[h] Department of Bioinformatics and Medical Engineering, Asia University, Taichung, Taiwan

ARTICLE INFO

Keywords:
Next-generation sequencing
Targeted gene sequencing
Whole exome sequencing

ABSTRACT

Background: Targeted gene sequencing (TGS) and whole exome sequencing (WES) are being used in clinical testing in laboratories. We compared the performances of TGS and WES using the same DNA samples.
Methods: DNA was extracted from 10 endometrial tumor tissue specimens. Sequencing were performed with an Illumina HiSeq 2000. We randomly selected variants to confirm through Sanger sequencing or mutant-enriched PCR with Sanger sequencing.
Results: We found that the variants identified in both TGS and WES were true positives (47/47), regardless of the sequencing depth. Most variants found in TGS only were true positives (34/40), and most of the variants found by WES only were false positives (8/18). From these results, we suggest that the sequencing depth may not play important role in the accuracy of NGS-based methods. After analysis, we found that WES had a sensitivity of 72.70%, specificity of 96.27%, precision of 99.44%, and accuracy of 75.03%.
Conclusions: The results of NGS-based methods must currently be validated, especially for important reported variants regardless of the methods used, and for the use of WES in cancers a higher false negative rate must be considered. More sensitive methods should be used to confirm the NGS results in uneven cancer tissues.

1. Introduction

Whole exome sequencing (WES) and targeted gene sequencing (TGS) are routinely used and are gradually being optimized for the detection of therapeutically targetable alterations. Recently, several studies have applied TGS for the personalized treatment of cancer. For example, ultra deep TGS has been used to develop personalized treatments for breast cancer [1]. Furthermore, next-generation sequencing (NGS) has also been applied for the detection of rare *RAS* mutations in metastatic colorectal cancer (mCRC), which has resulted in important clinical implications for patients with mCRC [2]. The National Cancer Institute Match Trial used an Oncomine Comprehensive Panel (OCP) to evaluate 3000 cancer samples for trial selection. One hundred and forty-three unique cancer genes were included in the OCP panel [3]. Comprehensive screening of genetic mutations in tumors has become an important part of the therapeutic decisions when treating cancer. Many commercial predesigned cancer-related gene panels are available, such as Roche/NimbleGen SeqCap EZ Comprehensive Cancer design (578 genes), Life/Ion AmpliSeq Comprehensive Cancer Panel (409 genes) and Illumina/TruSight Cancer Sequencing Panel (94 genes). Several laboratories have recently published their own approaches and experiences regarding validation and implementation of several NGS panels. The Ion Proton 409 gene panel assay has been demonstrated to be most suitable for use in a clinical molecular diagnostic laboratory [4].

The exome represents only ~2% of the human genome, but contains ~85% of known disease-related mutations, making WES an alternative to whole genome sequencing (WGS) [5,6]. WES has significant advantages over WGS: it is currently less expensive, has faster data analysis, and has easier data management. There are currently at least three commercial whole exome capture platforms available, including Agilent, NimbleGen, and Illumina, and new versions have also been released for each of these platforms. During the past few years, several

* Corresponding author at: Epigenome Research Center, China Medical University Hospital, 2 Yuh-Der Road, Taichung 404, Taiwan.
E-mail address: d6781@mail.cmuh.org.tw (J.-G. Chang).

http://dx.doi.org/10.1016/j.cca.2017.06.015
Received 24 March 2017; Received in revised form 16 May 2017; Accepted 14 June 2017
Available online 15 June 2017

performance comparison studies among these exome capture technologies have been published [7–9]. Investigators who are planning WES could select a suitable exome capture platform for their particular application.

Uneven distribution of cancer cells in the tumor tissues may influence the variant detection rate and accuracy of confirmation. In genomic analyses, sequencing depth and coverage are key considerations. The higher the sequencing depth and coverage, the higher the cost [10]. The number of samples included in one sequence run, the type of library constructed, and the type of NGS platform used all influence the coverage and length of the sequence reads. Targeted regions of GC-rich content and repetitive elements are generally harder to capture and/or sequence [11]. Adequate sequence coverage and depth and long sequence reads are sometimes needed for correct data analysis.

Performance comparison between TGS and WES has rarely been conducted and never in a comprehensive manner. In this study, we compared the two NGS techniques, performing TGS with the NimbleGen SeqCap EZ Comprehensive Cancer Design (578 cancer genes) and WES with the Agilent SureSelect Human All Exon kit 51 Mb (v4) on tissue samples from ten unrelated Taiwanese patients with endometrial cancer.

2. Materials and methods

2.1. Study subjects

DNA samples from ten unrelated subjects were selected for this study. Genomic DNA was extracted from frozen tumor resections using the QIAamp® DNA Micro kit (Qiagen, Heidelberg, Germany) according to the manufacturer's instructions. DNA amounts were quantified by Nanodrop 2000 (Thermal Fisher scientific, DE) and Qubit Fluorometer (Invitrogen, Carlsbad, CA, USA). Unamplified, high molecular weight and RNase treated genomic DNA was used for TGS and WES according to the platform's standard requirements, i.e., 1 μg for NimbleGen and 50 ng for Agilent. This study was approved by the Kaohsiung Medical University Hospital Institutional Review Board.

2.2. Next-generation sequencing

TGS and WES were performed with an Illumina HiSeq 2000. TGS was performed with NimbleGen SeqCap EZ Comprehensive Cancer Design (578 cancer genes). WES was performed with Agilent SureSelect Human All Exon kit 51 Mb (v4). Sequencing was carried out with 2 × 100 bp paired end reads.

2.3. Analysis of next-generation sequencing data

To filter the reads with bad sequencing quality, FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit) was used to process the raw read data files. There are two steps to sequence quality processing. The command was “fastq_quality_filter –Q33 –q 30 –p 70”. “-q 30” set the minimum quality score to keep to 30, and “-p 70” set the minimum percent of bases to have “-q” quality over or equaling 70%. Sequences were retained if both forward and reverse sequencing reads passed the first step.

An efficient sequence alignment tool, Bowtie2, was used to align the retained reads with the human genome (Grch38.p2) [12]. The sequence alignment results showed that the reads with only one chromosome location were retained for further analysis. The genome analysis toolkit (GATK), a widely used genetic variants discovery tool, was used to identify genetic variants according to the sequence alignment results [13].

2.4. PCR and Sanger sequencing

Randomly selected variants identified by both techniques and exclusively by WES or TGS were further confirmed by PCR and Sanger sequencing. The specific PCR primers were designed using Primer3 software (Supplementary Table 1). The products were directly sequenced with an ABI PRISM terminator cycle sequencing kit v3.1 on an ABI 3130 DNA sequencer (Applied Biosystems, Carlsbad, CA).

2.5. Mutant-enriched PCR assay to detect MDM4 codon 373 and JAK2 codon 830 mutations

This assay was performed as previously described [14]. Briefly, to detect the *MDM4* codon 373 mutation, the primer sequences for PCR amplification were as follows: 5′-ATTTCGGCTCCTGTCGTTAGG-3′ (forward primer which contains a mismatch base at the 3′ end marked by the underline) and 5′-CTGGCAATCCTCCATGTTTT-3′ (reverse primer). The forward primer harbors one mismatched base (A to G) to introduce a new GGCC sequence after PCR amplification of wild-type alleles. The restriction enzyme PhoI was used to digest the GGCC sequence in the amplicon of the wild-type. In contrast, codon 373 1st base mutant alleles were not digested because of the base substitution of a C to T at the third base of GGCC, resulting in the enrichment of mutant alleles.

To detect the *JAK2* codon 830 mutation, the primer sequences for PCR amplification were as follows: 5′-CGATTATTTTGGTCAACTTGAATG-3′ (forward primer) and 5′-CAAGCACTCCTTAAAATGTTGTAGA-3′ (reverse primer). The thermostable restriction enzyme PspGI was used to digest the CCWGG sequence in the wild-type amplicon. In contrast, codon 830 2nd base mutant alleles were not digested because of the base substitution of a T to C nucleotide at the third base of CCWGG, resulting in the enrichment of mutant alleles after PCR amplification with PspGI digestion. The unusually thermostable restriction enzyme PspGI, which resists deactivation during thermal cycling, is key for the mutant-enriched PCR assay. The nucleotide changes of *MDM4* codon 373 and *JAK2* codon 830 were examined by direct sequencing of the mutant-enriched PCR product.

3. Results

3.1. Overview

On average, 186 thousand reads per sample were mapped to TGS regions by WES, corresponding to a mean coverage of 11.5 × (Table 1). For comparison, we divided the results into four groups based on variant uniqueness and sequencing depth.

(i) Variants detected by both TGS and WES, and WES depth ≥ 20 ×

The results show that 1129 variants were identified by both TGS and WES, with a WES depth ≥ 20 ×. We performed Sanger sequencing on

Table 1
Reads and coverage statistics for each sample mapped to TGS by WES.

Sample	No. of WES reads aligned in TGS regions	WES mean coverage in TGS regions
03-3812T	188,687	11.62
F123	174,373	10.66
F132	188,789	11.54
F134	170,061	10.52
F146	216,672	13.35
F150T	191,121	11.59
F147T	189,889	11.69
F152T	187,542	11.83
F92T	173,132	10.55
F114T	182,051	11.28
Mean	186,232	11.5