Experimental and Molecular Pathology 100 (2016) 8–12

Contents lists available at ScienceDirect

Experimental and Molecular Pathology

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

Genetic alterations in endometrial cancer by targeted next-generation sequencing

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

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

ARTICLE INFO

Article history:
Received 9 November 2015
Accepted 24 November 2015
Available online 25 November 2015

Keywords:
Endometrial cancer
Next-generation sequencing
PTEN mutation
IL-7 signaling pathway

ABSTRACT

Many genetic factors play important roles in the development of endometrial cancer. The aim of this study was to investigate genetic alterations in the Taiwanese population with endometrial cancer. DNA was extracted from 10 cases of fresh-frozen endometrial cancer tissue. The exomes of cancer-related genes were captured using the NimbleGen Comprehensive Cancer Panel (578 cancer-related genes) and sequenced using the Illumina Genomic Sequencing Platform. Our results revealed 120 variants in 99 genes, 21 of which were included in the Oncomine Cancer Research Panel used in the National Cancer Institute Match Trial. The 21 genes comprised 8 tumor suppressor candidates (*ATM*, *MSH2*, *PIK3R1*, *PTCH1*, *PTEN*, *TET2*, *TP53*, and *TSC1*) and 13 oncogene candidates (*ALK*, *BCL9*, *CTNNB1*, *ERBB2*, *FGFR2*, *FLT3*, *HNF1A*, *KIT*, *MTOR*, *PDGFRA*, *PPP2R1A*, *PTPN11*, and *SF3B1*). We identified a high frequency of mutations in *PTEN* (50%) and genes involved in the endometrial cancer-related molecular pathway, which involves the IL-7 signaling pathway (*PIK3R1*, n = 1; *AKT2*, n = 1; *FOXO1*, n = 1). We report the mutational landscape of endometrial cancer in the Taiwanese population. We believe that this study will shed new light on fundamental aspects for understanding the molecular pathogenesis of endometrial cancer and may aid in the development of new targeted therapies.

1. Introduction

Endometrial cancer is the fourth most common gynecological malignancy in Europe and the United States. More than 280,000 women are diagnosed each year worldwide and 74,000 women die from endometrial cancer annually. It is also the second most common gynecological cancer in Taiwan.

Endometrial cancer is broadly divided into two groups: endometrioid carcinoma and uterine serous carcinoma (Bokhman, 1983). Close to 80–90% of endometrial cancers are endometrioid carcinomas and 2–10% are uterine serous carcinomas (Dedes et al., 2011). Although uterine serous carcinoma is in the minority, it is much more aggressive than endometrioid carcinoma and has a poor outcome (del Carmen et al., 2012; Hamilton et al., 2006). Therefore, understanding the genomic alterations associated with this disease may provide opportunities for genome-guided clinical trials and drug development.

Next-generation sequencing (NGS) technology has emerged as a powerful tool to investigate the genetic etiology of diseases. NGS has been applied in a variety of ways, such as whole genome sequencing, targeted capture, high-throughput RNA sequencing, and chromatin immunoprecipitation followed by sequencing. The ability to generate a huge quantity of sequencing data also presents the challenge of deciding which variants to validate. Several organizations, such as the College of American Pathologists, Centers for Disease Control and Prevention, and U.S. Food and Drug Administration, have published guidelines for clinical NGS analysis.

Over the past two years (2012–2014), a number of studies have reported the genomic landscapes of endometrial cancer. The first study to characterize the genomic landscape of uterine serous carcinoma was described by Kuhn et al. (2012). Among the most frequently mutated genes in 10 uterine serous carcinomas were *TP53*, *PIK3CA*, *FBXW7*, and *PPP2R1A*. Subsequently, Gallo et al. found frequent somatic mutations not only in *FBXW7*, but also in ubiquitin ligase complex and chromatin remodeling genes (Le Gallo et al., 2012). Zhao et al. then decoded the exomes of a uterine serous carcinoma cohort five times larger than those reported earlier. They identified a significantly increased burden of mutation in 14 genes, including the previously reported and well-known cancer genes *TP53*, *PIK3CA*, *PPP2R1A*, *KRAS*, *PTEN*, *FBXW7*, and *CDKN1A* (Zhao et al., 2013). The Cancer Genome Atlas (TCGA) project revealed 14 pathogenic driver genes of uterine serous carcinoma

* 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.yexmp.2015.11.026
0014-4800/

(*TP53*, *PIK3CA*, *FBXW7*, *PPP2R1A*, *CHD4*, *CSMD3*, *SLC9A11*, *PTEN*, *COL11A1*, *PRPF18*, *SPOP*, *CDH19*, *HIST1H2AM*, and *CELP*) (Kandoth et al., 2013).

The first whole exome sequencing study of endometrioid carcinoma investigated 13 cases. Ten tumor suppressor genes (*ARID1A*, *INHBA*, *KMO*, *TTLL5*, *GRM8*, *IGFBP3*, *AKTIP*, *PHKA2*, *TRPS1*, and *WNT11*) and two oncogenes (*ERBB3* and *RPS6KC1*) were identified as potential candidate driver genes (Liang et al., 2012). In addition, the frequent occurrence of mutations in *PTEN* (64%), *PIK3CA* (59%), *ARID1A* (55%), *CTNNB1* (32%), *MLL2* (32%), *FBXW7* (27%), *RNF43* (27%), *APC* (23%), *FGFR2* (18%), and *EGFR* (14%) among endometrioid carcinomas was confirmed by Kinde et al. based on the exome sequencing results of 22 cases (Kinde et al., 2013).

The aim of the current study was to identify genetic alterations in endometrial cancer in the Taiwanese population. We performed deep sequencing (>500×) to detect the mutational status in 578 cancer-related genes (NimbleGen Comprehensive Cancer Panel) using fresh-frozen tissues from 10 Taiwanese patients with endometrial cancer.

2. Materials and methods

2.1. Sample preparation and DNA extraction

The specimens consisted of 10 fresh-frozen endometrial cancer tissues that were submitted for targeted sequencing. DNA was isolated using proteinase K and a QIAamp® DNA Micro Extraction Kit (QIAGEN) according to the manufacturer's protocol. This study was approved by the Institutional Review Board (KMUH-IRB-970488).

2.2. Library preparation and amplification, targeted capture, and Illumina-based sequencing

Genomic DNA (1 μg) was fragmented using a Covaris S2 Focused-ultrasonicator (Covaris, Woburn, MA), and quality control (QC) was performed using an Agilent Bioanalyzer 2100 (Agilent Technologies) to ensure a fragment size range from 200 to 400 bp. Fragmentation was followed by end repair, A-tailing and sequencing adapter ligation using an Illumina TruSeq DNA Sample Preparation Kit. The adapter-ligated DNA was amplified via selective, limited-cycle polymerase chain reaction (PCR) for a total of eight cycles. The prepared library (1 μg) was hybridized for 64 to 72 h to the NimbleGen SeqCap EZ Designs-Comprehensive Cancer library (Roche Diagnostics). The hybridized product was amplified for 18 PCR cycles using Roche post-capture primers. QC was performed on the amplified product using an Agilent Bioanalyzer DNA 1000 Kit to ensure that the final library fragment size ranged from 150 to 400 bp; the product was quantified using KAPA SYBR® Fast ABI Prism® 2× qPCR Master Mix (KAPA Biosystem) to ensure a sufficient yield for sequencing.

For paired-end 100 bp × 2 sequencing on the Illumina HiSeq 2000 instrument, captured libraries were denatured and loaded onto an Illumina cBot instrument at 9.5 pM for cluster generation according to the manufacturer's instructions. Ten libraries were sequenced per HiSeq lane. A 1% PhiX control DNA (Illumina) was added into libraries as the internal control. On each flow cell, one lane contained PhiX DNA as a control.

2.3. Bioinformatic analysis

To filter poor-quality reads, FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit) was employed to process the raw read data files. There are two steps for sequence quality processing. The command was "fastq_quality_filter –Q33 –q 30 –p 70"; "–q 30" indicates that the minimum quality score to keep is 30 and "–p 70" indicates that the minimum percent of bases must have "–q" quality greater than or equal to 70%. Sequences were retained if both forward and reverse sequencing reads passed the first step.

Bowtie2, an efficient sequence alignment tool (Langmead et al., 2009), was used to align the retained reads with the human genome (Grch38.p2). According to the sequence alignment, reads having only one chromosomal location were retained for further analysis. The Genome Analysis Toolkit (GATK) (McKenna et al., 2010), a widely used genetic variants discovery tool, was employed to identify genetic variants according to the sequence alignment results.

Several databases and tools that provide information on genetic variants were used to annotate identified genetic variants. dbSNP (b144) (Sherry et al., 2001) is a database that collects and provides information on genetic variants within different species. ClinVar (Landrum et al., 2014) is a database that collects significant clinical information on genetic variants from patients. Cosmic (v73) (Forbes et al., 2015) is a database that collects somatic mutation information in human cancers. Polyphen2 (Adzhubei et al., 2010) is a software that predicts the level of influence of non-synonymous substitution in the human proteins.

Five criteria were used to identify the genetic variants of cancer driver genes: 1) a read coverage greater than 500, 2) non-synonymous genetic variants, 3) the occurrence of genetic variants was not over 50%, 4) the genetic variants were considered or predicted as pathogenic or possibly/probably functionally impaired, and 5) the global minor allele frequency of the genetic variants was less than 0.02.

2.4. Sanger sequencing validation

Primers for Sanger sequencing validation were designed using Primer3 software. The PCR primers used are described in Supplemental 1. PCR amplifications were performed using Pro Taq Plus DNA Polymerase (Protech Technology Enterprise, Taiwan) following the manufacturer's instructions. Sanger sequencing was performed on an ABI Prism 3130 Genetic Analyzer.

3. Results

3.1. Pathway analysis of mutated genes in endometrial cancer

We sequenced each sample on the Illumina HiSeq Platform to an average sequencing depth of 535.79. We generated a mean of 16 M raw reads per sample, of which 98.45% to 99.71% were aligned to the human reference genome (Grch38, Table 1). Table 2 shows an overview of our approach to identify variants.

After data filtering, we identified 120 variants in 99 genes, including 107 missense variants, 7 nonsense variants and 6 frame shift variants (Supplemental 2). Functional annotation of the 99 mutated genes was performed by use of the Annotation, Visualization, and Integrated Discovery (DAVID) database. Eighteen of the 99 mutated genes were found in the Kyoto Encyclopedia of Genes and Genomics (KEGG) cancer pathways (hsa05200), including *CREBBP*, *CTNNB1*, *FGFR2*, *FLT3*, FOXO1, *MTOR*, *MSH2*, *MSH6*, *PTCH1*, *PTEN*, *PIK3R1*, *PDGFRA*, *PML*, *KIT*, *TPR*, *TP53*, *AKT2*, and *ERBB2* ($p = 2.6 \times 10^{-9}$, false discovery rate $= 2.7 \times 10^{-6}$,

Table 1
Alignment and coverage statistics for 10 endometrial cancer patients.

Patient ID	Total reads	Reads mapped to genome	Covered ≧ 500× (%)	Average target coverage
F114T	17,197,961	17,079,751	59.2%	569.37
F123	20,545,225	20,459,077	68.8%	691.13
F132	14,534,452	14,426,666	46.6%	479.77
F134	15,572,857	15,507,377	44.7%	523.69
F146	14,826,104	14,693,849	48.9%	491.33
F147T	11,099,160	10,925,387	23.4%	354.19
F150T	16,221,035	16,154,454	46.8%	531.05
F152T	16,256,250	16,180,893	46.6%	534.26
F92T	17,218,187	17,150,449	52.6%	585.51
03-3812T	17,730,300	17,676,534	51.1%	597.55
Average	16,120,153	16,027,138	48.9%	535.79