Cancer Letters 396 (2017) 110–116

Contents lists available at ScienceDirect

Cancer Letters

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

Original Article

Identifying the clonal relationship model of multifocal papillary thyroid carcinoma by whole genome sequencing

Mao Xia [a, 1], Hengyu Li [b, 1], Qian Ma [c, 1], Dong Yu [a], Jing Li [a], Yi Zhang [d], Yuan Sheng [b, **], Yingjun Guo [a, *]

[a] Institute of Translational Medicine, Second Military Medical University, Shanghai 200433, PR China
[b] Department of Breast and Thyroid Surgery, Changhai Hospital, Second Military Medical University, Shanghai 200433, PR China
[c] International Joint Cancer Institute, Second Military Medical University, Shanghai 200433, PR China
[d] Department of Medical Genetics, College of Basic Medical Science, Second Military Medical University, Shanghai 200433, PR China

ARTICLE INFO

Article history:
Received 23 November 2016
Received in revised form
7 March 2017
Accepted 8 March 2017

Keywords:
Clonal relationship
Whole genome sequencing
Multifocality
Multifocal papillary thyroid carcinoma

ABSTRACT

Purpose: To evaluate the application of whole genome sequencing (WGS) in determining the inter-foci clonal relationship of multifocal papillary thyroid carcinoma (mPTC).
Methods: After reviewing PTC patient profiles, 8 cancer foci and germline control samples from 3 mPTC patients were analyzed by WGS. Single nucleotide variations (SNVs), copy number variation (CNV), structural variation and mutational signature were examined.
Results: The multifocality rate of PTC was 35.1% and mPTC were shown to have larger primary tumor diameter, higher rate of lymph node metastasis and less number of accompanying non-cancerous lesions than single PTC in one or both gender groups. Out of the 8 cancer foci, 5 foci were identified as clonal-independent model with the rest 3 foci as clonal-derived model according to exonic SNVs spectrum. Non-exonic mutations provided compelling evidence at the genome-wide level for the classification. Specific CNV and 12 mutational signatures were also identified.
Conclusions: WGS could be an impressive tool in clonal relationship determination of PTC by providing a panoramic view of genome-wide somatic mutations. The substantial sequencing data provided additional information that could help studying the mechanism of carcinogenesis.

Introduction

Thyroid cancer is the most common endocrine malignant tumor and its incidence has continuously increased within the past three decades all over the world [1]. Thyroid cancer consists of four main types according to pathological features: papillary, follicular, medullary and anaplastic thyroid carcinoma. The most common form, known as papillary thyroid carcinoma (PTC), accounts for 80–85% of all malignant thyroid tumors [2]. PTC commonly appears as multifocal pattern at a high rate ranging from 24% to 46% [3–6]. Recently, an international multicenter study showed that up to 36% (288/794) of PTC larger than 1 cm had a contralateral carcinoma [7].

The clonal relationship behind the common clinical phenomenon of multifocality in PTC is an old-aged question. There have been enormous efforts in this origin-tracing deduction within different PTC foci in the past two decades [8–13]. Yet, whether multifocal tumors share the same origin or not remains controversial due largely to the technical limitation and incomplete estimation of the detection methods so far. One of the frequently used strategies was measuring the X-chromosome inactivation state of each foci. As a concrete implementation, the human androgen-receptor gene assay (HUMARA) is only applicable to only female patients. And it requires the heterogeneous allele of AR gene to be informative. Besides, this method sets a high threshold for tumor purity analyzed so that most related studies has to collect tumor cells using laser capture microdissection. Another tactic of inferring from *BRAF* V600E or other known driver mutations has the premise that the tumor evaluated has the specific mutation. Last but not least, the main defect of utilizing HUMARA or driver mutations for

* Corresponding author. Institute of Translational Medicine, Second Military Medical University, 800 Xiangyin Road, Shanghai 200433, PR China.
** Corresponding author. Department of Breast and Thyroid Surgery, Changhai Hospital, Second Military Medical University, 168 Changhai Road, Shanghai, PR China.
E-mail addresses: sheng528yuan@163.com (Y. Sheng), guoyingjun2017@163.com (Y. Guo).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.canlet.2017.03.012
0304-3835/

judging the clonal relationship is the restriction that definite conclusion could be achieved only based on inconsistent profiles from different foci, and no explicit statement could be reached in the case of identical profiles considering patch size [14] or high coincidence of driver mutation validated in different types of cancer including PTC [15].

Next generation sequencing emerged as a powerful tool for deciphering cancer from the genetic level [16], which could be used to decrypt the unique "fingerprint" of each cancer clone. Different DNA sequencing strategies including whole genome sequencing (WGS), whole exon sequencing (WES) and targeted sequencing differ in the scale and depth of the genome region sequenced. Among them, WGS grants the access to a detailed landscape of the entire genome. The high rate of technological advance of sequencing has also transformed WGS into an affordable, efficient and widely-used application [17].

Here, medical records in our department within the last two years were reviewed to define the clinicopathological characteristics of patients with multifocal PTC (mPTC). Then we tried to answer the old question of cancer clonal relationship in a new way by sequencing the whole genome of 8 cancer foci and their germline control from 3 mPTC patients. We demonstrate that a clear inter-foci clonal relationship is revealed using the somatic genetic alterations. To the best of our knowledge, this approach is the most comprehensive and versatile way for clonal relationship determination in mPTC to date.

Materials and methods

Clinicopathological data collection and patient selection

The electronic medical records of the newly diagnosed PTC patients from the Department of Breast and Thyroid Surgery at Changhai hospital ranging from January 2013 to May 2015 were retrieved for further review. Gender, age, number of cancer foci, primary tumor size, foci location, coexistent non-cancerous lesion, coexistent Hashimoto thyroiditis, lymph node metastasis, distant metastasis and extrathyroidal extension, were collected for each patient.

The patients who received a curative thyroidectomy and had no preoperative chemotherapy or radiotherapy were included. Whole blood was collected preoperatively into 2 ml EDTA containing vacationer tubes. Fresh tissue samples were acquired in the operating room and handled immediately within 20 min after resection. Samples of the tumor and adjacent non-tumor tissues were snap frozen in liquid nitrogen and transferred to the lab for storage at −80 °C. The remaining resected specimen was sent for routine pathological examination. Only those with diagnosis of PTC were accepted for further analysis. Before nucleic acid extraction, each foci was subject to cryosection and hematoxylin-eosin staining to ensure the existence of cancer cells. The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Second Military Medical University (approval number was 2015-002). Informed consent was obtained from all patients.

Determination of BRAF V600E by Sanger sequencing

The *BRAF* exon 15 was amplified by PCR using the same primers in [18]. Approximately 100 ng genomic DNA was used as a template in a total 25 μl PCR mixture containing 12.5 μl Premix Taq™ Hot Start Version (Takara Bio Inc., Shiga, Japan), 200 nM each primer. Cycling conditions were initial denaturation at 95 °C for 10 min, then 40 cycles with denaturation at 94 °C for 30 s, annealing at 58 °C for 30 s, and extension at 72 °C for 30 s, ended with a final extension at 72 °C for 10 min. The PCR product was loaded onto 2% agarose gel for electrophoresis. The 224 bp amplified fragment was purified using Purelink Quick Gel Extraction Kit (Invitrogen, Carlsbad, USA) and then underwent direct DNA bidirectional sequencing with an ABI 3730xl Genetic Analyzer and a BigDye Terminator cycle sequencing kit (Applied Biosystems, Foster City, USA). The obtained sequences were compared with the reference *BRAF* gene in the GenBank database and the *BRAF* T1799A was identified using chromas software (chromas v2.1.1, http://www.technelysium.com.au).

Validating BRAF V600E by immunohistochemistry

For each cancer foci, 5 μm-thick FFPE sections were prepared in the Department of Pathology. The primary antibody selected was mouse anti-human *BRAF* V600E monoclonal antibody clone VE1 (Spring Bioscience, Pleasanton, USA). The slides were subjected to a sequential procedure of deparaffinization, cell-conditioning and incubated with the primary antibody (1:100 dilution) at 37 °C for 30 min. After completing the immunostaining using OptiView detection kit (Ventana Medical Systems, Arizona, USA), the slides were conterstained with Mayer's hematoxylin.

The software ImageJ (https://imagej.nih.gov/ij/index.html) was used to quantify the intensity of the stained slides. Each slide was inspected by two investigators (MX and QM) independently for obtaining six random images. By installing the plugin IHC-Toolbox, we deconvoluted the images using the built-in H-DAB color model. After adjusting the threshold of all images to the same level, the integrated density was measured.

Library preparation and whole genome sequencing

TruSeq Nano DNA HT Library Prep Kit (Illunima, San Diego, USA) was used for sequencing library construction. All the steps were accomplished according to the standard protocol of the manufacturer. Briefly, fragmentation of genomic DNA was completed using Covaris (Covaris, Woburn, USA). End repair and size selection were performed before adenylate 3′end. Then, adapters were ligated to the end of the DNA fragments, followed by PCR-based DNA fragments enrichment. Qubit 2.0 (Life technologies, Carlsbad, USA) and Agilent 2100 bioanalyzer (Agilent Technologies, Palo Alto, USA) were used for concentration determination and size verification respectively. Finally, Cluster amplification was done on cBot (Illunima, San Diego, USA) and paired end sequencing (150bp × 2) was performed on a Hiseq X10 platform (Illumina, San Diego, USA).

Reads mapping and somatic alteration discovery

After removing adapters and low-quality reads using Trimmomatic v0.33 [19], all short reads were aligned to the NCBI human reference genome hg19 (GRCh37/hg19) using Burrows-Wheeler Aligner (BWA) v0.7.12 with default parameters [20]. Duplicates marking was completed using Picard v1.130 (http://broadinstitute.github.io/picard/) and then processing with GATK v3.2.2 (https://software.broadinstitute.org/gatk/) following the best practices.

For somatic SNVs analysis, Varsan2 v2.3.9 [21], Mutect v1.1.4 [22] and Strelka v1.0.15 [23] were used at default settings and the intersection among three callers were accepted for further inspection. Those who failed the two criteria were discarded: 1) The tumor sample contained a total number of reads exceeding 20 and the mutation was supported by at least 6 alteration reads; 2) There were at least 12 reads coverage in the normal sample at the same position. Finally, Mutations located within the exonic region, frequency higher than 40% or other selected mutations were confirmed by IGV [24]. It was those mutations having at least 6 alteration reads with mapping quality equaling 60 and phred quality surpassing 20 that were retained. Furthermore, there should be no obvious strand bias either. The vcf files were annotated by ANNOVAR [25].

Copy number variations were detected in the matched tumor-normal way using Control-FREEC v8.7 [26]. BreakDancer v1.3.6 [27], Manta v1.0.1 [28] and FACTERA v1.4.4 [29] were used for somatic structural variation analysis.

Statistical analysis

Quantitative data were presented as the mean ± SD. Comparisons of the categorical variables between two groups were carried out using Pearson Chi square test. Mann–Whitney U test was applied for comparison of continuous variables. All *P* values were two sided and a *P* value of less than 0.05 was regarded as statistically significant. The above statistics were performed using SPSS v17.0 (SPSS Inc., Chicago, USA).

Results

Multifocal papillary thyroid carcinomas harbor unique clinicopathological characteristics

Overall, a number of 920 patients received total or partial thyroidectomy and were diagnosed with PTC in our department between January 2013 and May 2015. The sex ratio was 2.24, contributing by 636 female and 284 male patients. The PTC patients were diagnosed at a mean age of 45.69 ± 12.68 and there was no significant difference in diagnostic age between the two genders ($P = .426$). The top three major spatial distribution pattern of cancer foci location for both genders in all patients was the right lobe, left lobe and bilateral (Supplementary Fig. 1A and B). A multifocal pattern was observed in 323 (35.1%) patients, which elevated the mean number of cancer foci for all PTC patients to 1.49 ± 0.83. Of the mPTC patients, 144 in female (61.8%) and 50 in male (55.6%) had bilateral distribution, making it the most common distribution pattern (Supplementary Fig. 1C and D).

Clinicopathological factors were compared between single PTC (sPTC) and multifocal PTC by patient gender (Table 1). It turned out