Cancer Letters 405 (2017) 22–28

Contents lists available at ScienceDirect

Cancer Letters

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

Original Article

Clonal composition of human ovarian cancer based on copy number analysis reveals a reciprocal relation with oncogenic mutation status

Kazuko Sakai [a], Masayo Ukita [b], Jeanette Schmidt [c], Longyang Wu [c], Marco A. De Velasco [a], Alan Roter [c], Luis Jevons [c], Kazuto Nishio [a,*], Masaki Mandai [b]

[a] Department of Genome Biology, Kindai University Faculty of Medicine, 377-2 Ohno-higashi, Osaka-Sayama, Osaka, 589-8511, Japan
[b] Department of Obstetrics and Gynecology, Kindai University Faculty of Medicine, 377-2 Ohno-higashi, Osaka-Sayama, Osaka, 589-8511, Japan
[c] Thermo Fisher Scientific, 3420 Central Expy, Santa Clara, USA

ARTICLE INFO

Article history:
Received 3 April 2017
Received in revised form
11 July 2017
Accepted 13 July 2017

Keywords:
Oncogenic mutation
Clonal composition
Ovarian cancer
Heterogeneity

ABSTRACT

Intratumoral heterogeneity of cancer cells remains largely unexplored. Here we investigated the composition of ovarian cancer and its biological relevance. A whole-genome single nucleotide polymorphism array was applied to detect the clonal composition of 24 formalin-fixed, paraffin-embedded samples of human ovarian cancer. Genome-wide segmentation data consisting of the log2 ratio (log2R) and B allele frequency (BAF) were used to calculate an estimate of the clonal composition number (CC number) for each tumor. Somatic mutation profiles of cancer-related genes were also determined for the same 24 samples by next-generation sequencing. The CC number was estimated successfully for 23 of the 24 cancer samples. The mean ± SD value for the CC number was 1.7 ± 1.1 (range of 0–4). A somatic mutation in at least one gene was identified in 22 of the 24 ovarian cancer samples, with the mutations including those in the oncogenes *KRAS* (29.2%), *PIK3CA* (12.5%), *BRAF* (8.3%), *FGFR2* (4.2%), and *JAK2* (4.2%) as well as those in the tumor suppressor genes *TP53* (54.2%), *FBXW7* (8.3%), *PTEN* (4.2%), and *RB1* (4.2%). Tumors with one or more oncogenic mutations had a significantly lower CC number than did those without such a mutation (1.0 ± 0.8 versus 2.3 ± 0.9, $P = 0.0027$), suggesting that cancers with driver oncogene mutations are less heterogeneous than those with other mutations. Our results thus reveal a reciprocal relation between oncogenic mutation status and clonal composition in ovarian cancer using the established method for the estimation of the CC number.

Introduction

Ovarian cancer is the most common cause of death from gynecologic malignancy [1]. Cytoreductive surgery followed by platinum-taxane chemotherapy is the standard treatment for ovarian cancer. Although the initial response rate is >80%, most patients experience relapse within 5 years as a result of the survival of chemoresistant clones [1,2]. A high degree of intratumoral heterogeneity for ovarian cancer also may give rise to clonal evolution, tumor progression, and resistance to chemotherapy [3,4]. Intratumoral heterogeneity of ovarian cancer cells has remained largely unexplored, however.

Next-generation sequencing (NGS) has been applied to the molecular characterization of tumors, to the identification of new druggable targets, and to the screening of patients for clinical trials [5]. NGS technology has uncovered novel mutations in a variety of cancer types including lung cancer [5,6] and ovarian cancer [7]. The mutation profile of cancer-related genes can now be determined by NGS from clinical formalin-fixed, paraffin-embedded (FFPE) tumor samples [8]. Mutation of the *TP53* tumor suppressor gene has been detected in >50% of ovarian cancers, predominantly in those at an advanced stage and with high-grade serous histology [9]. Mutations in *BRAF*, *KRAS*, *PTEN*, or *CTNNB1* have also been identified in ovarian tumors with endometrioid, mucinous, or low-grade histologies [10].

Copy number variations reflect genomic structural changes that give rise to gene amplification, deletion, or copy number gain and which result from selection pressures that favor cancer development. Chromosomal microarray and other array-based approaches

* Corresponding author. Department of Genome Biology, Kindai University Faculty of Medicine, 377-2 Ohno-higashi, Osaka-Sayama, Osaka, 589-8511, Japan.
E-mail address: knishio@med.kindai.ac.jp (K. Nishio).

http://dx.doi.org/10.1016/j.canlet.2017.07.013

have been widely adopted for detection of whole-genome copy number variation [11,12], but they can be difficult to perform with FFPE samples. The OncoScan® FFPE Assay Kit (a research use only product from Thermo Fisher Scientific) relies on molecular inversion probe (MIP) technology to detect genome-wide copy number alterations, loss of heterozygosity, and somatic mutations [13]. This assay provides the B allele frequency (BAF), log2 ratio (log2R), and copy number for each of ~220,000 analyzed polymorphic genomic locations. Copy number is derived from log2R and BAF. We hypothesized that clonal composition of a tumor can be analyzed on the basis of BAF and log2R, determined from whole-genome copy number profiles obtained with the OncoScan® FFPE Assay Kit. A similar approach had been described in Ref. [14], where clonal composition was derived from both somatic mutations as well as log2Rand BAF for loci with aberrant copy number. Others have previously described approaches mainly based on the variant allele frequency (VAF) of somatic mutations, [15–17]. We have now examined the feasibility of this approach with retrospectively collected FFPE samples of ovarian cancer FFPE samples of ovarian cancer. In addition, we examined the relation between mutation profile and clonal composition. Our results suggest that the presence of somatic mutations in ovarian cancer may be associated with a lower clonal composition (CC number).

Materials and methods

Ovarian cancer samples

Twenty-four previously untreated ovarian cancer samples were enrolled in this retrospective study. All cases were assigned a stage based on the 1988 International Federation of Gynecology and Obstetrics (FIGO) system.

Isolation of genomic DNA

Collected FFPE samples were subjected to a histological review, and only those containing sufficient tumor cells as determined by hematoxylin and eosin staining were subjected to DNA extraction. Genomic DNA was extracted from FFPE tumor samples with the use of a QIAamp DNA FFPE Kit (Qiagen, Valencia, CA). The quality and quantity of the DNA were determined with the use of a Thermo Scientific™ NanoDrop™ 2000/2000c Spectrophotometers (Thermo Fisher Scientific, Waltham, MA) and Invitrogen™ Quant-iT™ PicoGreen™ dsDNA Assay Kit (Thermo Fisher Scientific).

Sequencing analysis

Tumor DNA samples were analyzed with NGS panels for mutation detection. For library preparation, the DNA was subjected to multiplex amplification by the polymerase chain reaction (PCR) with the use of an Ion AmpliSeq™ Library Kit 2.0 and Ion AmpliSeq™ Cancer Hotspot Panel v2 (Thermo Fisher Scientific). The PCR products were ligated to Ion Xpress™ Barcode Adapters 1–16 Kit (Thermo Fisher Scientific) and purified with the use of Agencourt AMPure XP beads (Beckman Coulter, Brea, CA). The purified libraries were pooled and then sequenced with an Ion Proton™ System, Ion PI™ Hi-Q™ Sequencing 200 Kit and Ion PI™ Chip Kit v3 (all from Thermo Fisher Scientific). DNA sequencing data were accessed through the Torrent Suite™ Software v.4.4 program (Thermo Fisher Scientific). Reads were aligned with the hg19 human reference genome, and potential mutations were called with the use of Variant Call Format ver. 5.0. Raw variant calls were filtered with a quality score of <100 and were manually checked with the integrative genomics viewer (IGV; Broad Institute, Cambridge, MA). Germline mutations were excluded with the use of the Human Genetic Variation Database (http://www.genome.med.kyoto-u.ac.jp/SnpDB).

Array-based readout

DNA isolated from FFPE samples was analyzed with an OncoScan® FFPE Assay Kit (Thermo Fisher Scientific). Oncoscan FFPE Assay Kit contains ~220,000 molecular inversion probes (MIPs) that measure both the A (A/T) and B (G/C) allele at these (A/B) polymorphic loci.

In the Oncoscan FFPE Assay Kit, DNA (80 ng) is subjected to annealing with these MIPs for 16–18 h followed by enzyme digestion and two separate gap-fill reactions. The circular MIPs are then separately linearized for each gap fill with a cleavage enzyme and amplified by PCR. The PCR products are subjected to enzymatic cleavage and fragmentation followed by hybridization for 16–18 h with two OncoScan® arrays, (one for each gap fill). The arrays are then stained and washed with the use of a GeneChip Fluidics Station 450 and loaded into a GeneChip Scanner 3000 7G (Thermo Fisher Scientific). Array fluorescence intensity (CEL) files were generated with Affymetrix GeneChip Command Console (AGCC) software version 4.0, and the CEL files were converted to OSCHP files with OncoScan Console software 1.3.

Clonal composition analysis

The OncoScan® FFPE Assay Kit provides genome-wide segmentation data, with each segment having a log2R and BAF. Cancer samples are rarely pure and almost always have normal tissue contamination. In addition, cancer samples frequently have multiple clones. We assume, as was done in Ref. [14], that a given abnormal copy number segment is the result of a single underlying event attributable to a single clone. Indeed, rarely do different clones affect the same copy number segment. Therefore, as assumed in Ref. [14], a given copy number segment can be associated with a %AC – the percentage of aberrant tumor cells. Hence we assume that copy number segments associated with the same %AC belong to the same clone. OncoScan FFPE Assay Kit reports a single %AC if the vast majority of copy number changes are consistent with the same %AC, otherwise a value of NA is reported. Onco Clone Composition clusters segments with similar log2R and BAF and assigns a %AC to each such cluster. A second clustering is done to combine clusters assigned to the same %AC.

We now proceed to show how the %AC is derived from the log2R and BAF value of a segment. The log2R value is a function of the percentage of aberrant tumor cells (%AC) that contribute to the copy number alteration and of the copy number (C) in the cancer genome. In particular, for %AC = β, log2R has the following relationship to β = %AC.

$$\log 2R = \gamma \log 2\left(\frac{\beta C + 2(1-\beta)}{2}\right) \quad (1)$$

Where γ is a technology factor. For example, for β = 0.2 (%AC = 20%), 20% of cells have C copy number, while 80% cells have 2 copies.

BAF is defined by the equation.

$$BAF = \frac{B}{B+A} \quad (2)$$

Where A and B correspond to the number of A alleles and B alleles, respectively. The number of A alleles and B alleles depends again on the %AC, the copy number C in the cancer genome and the allelic imbalance in the cancer genome. The number of minor alleles at a heterozygous site (NOMA) [18], helps in defining the BAF of a segment. For normal segments (C = 2) NOMA is 1, given that both alleles are present at each heterozygous site. However, for a segment with 100% loss of heterozygosity and normal copy number (C = 2), NOMA is 0, given that only one of the alleles is present, and the BAF of such a segment is zero. In a segment with %AC = β, copy number C in the tumor portion, and NOMA = κ, the number of A alleles at a heterozygous site with less A alleles is βκ + (1−β), while the copy number of the segment is βC +2 (1−β). The BAF of such a segment can hence be written as:

$$BAFcs = \frac{\beta\kappa + (1-\beta)}{\beta C + 2(1-\beta)} \quad (3)$$

The key observation is that by combining Equations (1) and (3) for any given segment we get

$$^{1}/_{\gamma} * \log 2R + \log_2(BAFcs) = \log(\beta\kappa + (1-\beta)) - 1 \quad (4)$$

As a result when plotting log2R versus $\log_2(BAF_{cs})$ all copy number segments from a given clone with %AC = β and with same κ (NOMA) form a straight line with intercept 1−log (βκ + (1−β)), and constant slope 1/γ. Segments with κ = 0 and various copy numbers are particularly easy to analyze as the corresponding intercept is 1−log (1−β) and the relationship to β is obvious. In default mode Onco Clone composition only uses NOMA = 0 segments to estimate the CC number, which in practice gave the best results. Fortunately, such segments are very common in cancer research samples. They correspond to segments with loss of heterozygosity in the cancer, and residual heterozygosity from the normal tissue. Segments with κ = 1, do not provide information on the %AC (β), because β cancels out in the equation log (βκ + (1−β)). Such segments are hence not informative for their association with a given clone.

Analysis of the clonal composition number corresponds to the identification of the number of different β = %AC terms detected among the aberrant segments. This analysis determines a clone to be positive if the percentage of markers in the aberrant segments assigned to a given β is more than 1% of all ~220,000 markers. The 1% cut-off was validated by mixing studies of tumors with various amounts of matched normals and assures that noise in the data does not affect the results. The Onco Clone Composition program provides a graphical representation of the aberrated segments and their association to %AC, as well as analytical estimation of the clonal composition by using a clustering approach that clusters all segments with same logR and BAF and then combines clusters corresponding to the same % AC. Each solid circle, cross, diamond, and square in Fig. 2 corresponds to the theoretical position of a segment with a given copy number and given %AC (color coded by %AC from 20% to 90%, as per the legend of the figure). Each open red circle corresponds to one segment in the cancer sample. Red circles turn yellow upon