Analysis of Prognostic Marker Panel for High-Grade Serous Ovarian Cancer through Age-Dependent DNA Methylation

Cheng WS* and Chiang JH
Department of Computer Science and Information Engineering, National Cheng Kung University, No.1, University Road, Tainan City 70101, Taiwan

Abstract
Achieving early detection at the onset of cancer is a major goal of cancer research. The early presence of aberrant DNA methylation makes the use of DNA methylation biomarkers an attractive candidate for early detection. Altered DNA methylation is ubiquitous in human cancers and specific methylation changes are often correlated with clinical features. DNA methylation biomarkers provide a range of opportunities for early detection, diagnosis, prognosis, therapeutic stratification and post-therapeutic monitoring. Furthermore, aging is one of the primary risk factors associated with cancer development. We conducted computational biology analyses of published High-grade serous ovarian cancer (HGSOC) epigenetic profiles using gene lists bearing human embryonic stem cell (hESC) characteristics. Through aging correlated features, epigenetic age-dependent marker panel on HGSOC was conducted. It is to be noted, the genes in the refined marker panel were found all included in the age-dependent features. The further experiment results showed not only the refined marker panel is able to represent the age-dependent features but performed better performance than the reported marker panel without associated with age parameter. Furthermore, the refined prognostic marker panel, including HOXA9, HSPA1A, and CALCA-associated with ovarian cancer and tumor growth, is strongly connected with literature support the potential for considering into clinical assay for patients’ stratification and future personalized medicine interventions.

Keywords: High-grade serous ovarian cancer; Aging; Human embryonic stem cell; DNA methylation; Epigenetic clustering

Introduction
Ovarian cancer is the most lethal form of gynecological cancers, due in part to lack of a reliable early detection method, and thus majority of the patients are diagnosed at an advanced stage of the disease. In the absence of an early ovarian cancer detection test, improved therapies for advanced disease are paramount to improving the survival for women with ovarian cancer. A better understanding of the molecular pathogenesis of ovarian cancer is currently required to identify biomarkers to enable early detection and novel therapeutic targets. Cancer recurrence is thought to reflect the survival of a small percentage of ovarian cancer stem cells (CSCs), which are relatively resistant to chemotherapy, can repopulate the tumor, and can spread to distal sites [1-3]. Previous studies have identified certain phenotypic and functional characteristics of CSCs that may distinguish such cells from other neoplastic cells [3,4]. These CSCs are thought to be resistant to current chemotherapeutic strategies due to their intrinsic stem-like properties and thus may be responsible for recurrent tumor growth. Ovarian CSCs may have stem-cell-like gene expression signatures, possibly reflecting their relatively high capacity for self-renewal and capacity to regenerate the entire tumor population [2,5]. Early detection of cancer, mainly at the onset of the disease, is a major goal of cancer research. The changes present in precancerous lesions [6] and some cancers [7] revealed that DNA methylation alterations play a key role in the early steps of human carcinogenesis. It has been a few decades now since Riggs, Holliday and Pugh proposed that DNA methylation was part of a system for controlling gene expression, and play an essential role in genome regulation, development and disease in mammalian cells [8,9]. Profiling of specific sites or panels of sites aberrantly methylated within tumor cells are currently being investigated as biomarkers for early prediction and prognostication, where cancer-associated methylation can be detected in tumor biopsy samples, cell-free serum, urine and peritoneal fluid [10-12]. Mikelska et al. also concluded that altered DNA methylation is ubiquitous in human cancers and specific methylation changes are often correlated with clinical features. DNA methylation biomarkers with some specific methylation changes provide a range of opportunities for early detection, diagnosis, prognosis therapeutic stratification and post-therapeutic monitoring [13]. Aging is one of the most important risk factors for the development neoplasia [14]. Epigenetic modifications of DNA, such as altered DNA methylation patterns, have long been postulated to play a role in aging and the associated increased incidence of neoplasia. There are now a substantial amount of evidences that aging does affect DNA methylation in specific loci, including cancer-related genes [15-19] suggest that CpG island methylation may be one of the molecular mechanisms which is involved in altering gene expression during the aging process. Age-related methylation of CpG islands in critical genes may also prove importance for the pathophysiology of other human diseases. Based on these observations, a hypothesis was proposed in the current work which stated that age may induce DNA methylation of human embryonic stem cell-like genes, and thereby predispose to cancer. Hence, in order to identify epigenetic age-dependent genes that may be associated with HGSOC, the genes in stem cell gene sets that were closely correlated with age were retrieved as features for subsequent marker panel analysis. We first retrieved age-dependent genes as features from human embryonic stem cells, and validated the age-dependent signatures and the refined age-dependent marker panel in all samples. Furthermore, the statistical significance of the features and the refined marker panel were examined through the previously developed R package, CGPredictor [20]. It is to be noted that, three of the eleven genes in the refined marker panel were strongly connected.
with ovarian cancer and tumor growth, as reported in the literature. These results also indicated that our method reliably extracted age-dependent markers with the potential to consider further clinical testing.

Materials and Methods

Study population

DNA methylation and gene expression profiling data were obtained from the TCGA website. The training dataset contained 282 tumor samples and eight solid normal tissues as controls samples. DNA methylation data was generated on the HumanMethylation27 BeadChip (Illumina, Inc.) to include 27,578 CpG dinucleotides spanning ~14,000 genes. The probe information was available on the Illumina website, whereas the clinical information was downloaded from the TCGA Data Portal.

Generation of hESC related gene sets

According to previous reports [5,21] we compiled an hESC related gene panel, including ESC overexpressed genes [22], Nanog, Oct4 and Sox2 targets [23]. Polycomb targets in hESCs [24] and Myc targets [25,26] were also used for subsequent analyses. This hESC specific gene panel was found to be enriched with poorly differentiated tumors [5]. We limited our primary analysis to the common gene set between this hESC dataset and the Infinium platform -3,038 genes in total for subsequent analyses.

Statistical analysis

Kaplan-Meier analysis was used to generate survival curves and log-rank test was employed to determine univariate differences between phenotypes. Bootstrapping was used to evaluate the robustness: two phenotypes of patients, representing 47 and 25 patients the number of samples grouped in the O-CIMP negative and positive phenotype, respectively, were resampled from patients of training data with the original age-dependent features. F-score was used to measure the performance of each of the 1,000 resampling sets.

Results

Age-dependent genes was extracted as features

Because of the heterogeneity of HGSOC, there are at least two typical subgroups. Pearson Correlation Coefficient was performed in terms of methylation intensity to choose the two most-normal-like HGSOC patients as surrogate 1 and the 2 most-normal-unlike HGSOC samples as surrogate [2]. We performed genome-wide DNA methylation profiling on HGSOC specific to human ES cell identity gene sets. The enrichment pattern of the used gene sets have been used successfully in the past to analyze an enrichment pattern with 13 gene sets via gene expression and DNA methylation [5,21,27]. Accordingly, these gene sets were utilized for further feature extraction and unsupervised clustering. After implementing Pearson Correlation Coefficient, methylation intensity of the genes correlated with age was extracted (a minimum Pearson correlation threshold >0.1 was applied). The forty-four genes in features for subsequent clustering were extracted under genes in surrogate2 and control samples possess the great beta value difference with 0.45. The extracted age-dependent features were utilized for unsupervised clustering of k-means to assign tumors to subclasses, as shown in Figure 1.

The age-dependent marker panel is confident and is able to represent the feature genes

To compare O-CIMP-negative with O-CIMP-positive, hypomethylated and upregulated genes were extracted as age-dependent marker panel. Interestingly, the genes in the marker panel were all found in the age-dependent features. Kaplan-Meier (KM) plots for both the clustered phenotypes by features, and the retrieved marker panel showed a more favorable outcome for O-CIMP-negative phenotype than the O-CIMP-positive (Figure 2a and 2b). Moreover, the age-dependent marker panel could distinguish between the two phenotypes with high performance similarly as done by the feature. These results indicated that the retrieved age-dependent marker panel could represent the age-dependent features. However, the performance of the current panel was distinctly more favorable than the previously reported marker panel for HGSOC, which was without associated with age, indicating that the performance of age-dependent marker panel (as shown in Figure 2c [8] performed better than the marker panel without age as a parameter. Furthermore, some genes in the marker panel were strongly connected to the previous reports related to poor survival of ovarian cancer and tumor growth. Collectively, we utilized faithful method to identify two distinct DNA methylation subgroups of HGSOC (Figure 3), which had great potential to mine biomarkers provided clinical test consideration and provided novel insight regarding the role of O-CIMP phenotypes [3].
Figure 2: Similar performance as the features in KM plot.
Validations of age-dependent features and marker panel signatures

We next asked if this extracted features and marker panel reflected a consistent pattern of increased methylation with age in the large patients set. We found that the methylation signatures of the feature genes correlated significantly with age from the whole sample (Figure 4a and 4b). When scaled with respect to two extracted specific phenotypes with features, the age-dependent feature was strikingly different between age groups (Figure 4c). The results supported that the age-dependent feature is highly confident for clustering to distinct phenotypes. As for the age-dependent marker panel, the same experiment was implemented for validation. We also found that the marker panel correlated significantly with age with the whole sample (Figure 4d). When scaled with the intensity distribution of the two extreme phenotypes, the results showed the similar correlation within the same age group but strikingly different between different groups (Figure 4e). The results were consistent with the methylation intensity distribution of age-dependent features. Additionally, these results did show a consistent pattern of increased methylation with age in the HGSOC cancer patients. It was also indicated the phenotypes was distinctly by age-dependent marker panel, which were correlated with age. Together, these results did demonstrate the marker panel be able to perform the similar performance as the features in KM plot (shown in Figure 2) indicated the ability as well that age-dependent marker panel be representing the feature. Furthermore, the developed examining package, CGPredictor, was also utilized for further validation [2]. The designed bootstrap test was utilized to examine the associated significance between clustering genes and the phenotypes. Experimental result was obtained after 1000 iterations (p<0.001), which indicated the significant relationship between features and clustered results. Additionally, the other statistical examining mechanism, viz. the random selection test was utilized to examine the significance of the marker panel. The same number of genes was randomly selected as was originally extracted as biomarker candidates for HGSOC. The marker panel for HGSOC was also significant after 1000 iterations (p<0.001). The significance result showed that the gene sets in our age-dependent marker panel has not been selected randomly.

Discussions and Conclusions

Aging is the main risk factor associated with cancer development [30]. In this regard, early detection offers the opportunity for therapeutic intervention at an early stage of cancer development, and thus produces better outcome in the disease treatment. Obviously, there is a need for a preventive strategy that can utilize biomarkers in order to stratify patients into appropriate screening or surveillance programs for the early detection of cancer. Hence, reliable markers associated with a large proportion of tumors needs to be developed for their widespread use in the diagnosis and treatment of cancer. DNA methylation is an important regulator of gene transcription. Alterations in DNA methylation are common in a variety of tumors as well as in development. Then, the early presence of aberrant DNA methylation, even in precancerous lesions, makes the use of DNA methylation biomarkers an attractive possibility for early detection. An alternative DNA-based approach for early detection of ovarian cancer might be promising since DNA extracted from a patient's plasma, serum or other body fluids could be easily amplified by PCR technology, and is therefore potentially more sensitive than the conventional tests. The majority of the commercially available DNA methylation biomarker-based assays for colorectal, lung, and brain are thus directed toward early detection. For example, one of the promising methylation biomarkers, SHOX2, was employed to distinguish between malignant lung cancer and benign lung diseases [31,32]. In colorectal cancer, the DNA methylation markers, both SEPT9 and VIM, are potential candidates for the early detection [33-36]. Moreover, Hegi et al. confirmed that MGMT methylation was not only a predictive biomarker to achieve survival of glioblastoma patients treated with temozolomide, but also triggered an increased interest in DNA methylation biomarkers [37]. The development of DNA methylation biomarkers through several bioinformatics and statistical mechanism would provide a strong influence on a biomarker's accuracy and robustness. Promoter methylation can be seen as an alternative gene-inactivation mechanism to chromosomal loss and loss of functional mutations [38]. Both genetic (changes in DNA sequence, such as deletions/amplifications and mutations) and epigenetic changes can be defined as heritable changes in gene expression that occur without changes to the DNA sequence [39]. In contrast to DNA sequence changes, chromosomal loss and mutations are essentially irreversible. This change, though heritable, is reversible, making it a therapeutic target. The reversible nature of DNA methylation offers a therapeutic opportunity to globally restore the normal epigenome using demethylation agents [40]. Moreover, epigenetic alterations are age dependent, but genetic alterations are not [41]. In the present work, we conducted bioinformatics analysis integrating public epigenetic and transcriptomic data and gene lists bearing human Embryonic Stem Cell characteristics for HGSOC stratification. In this study, the DNA methylation profiles and associated clinical information were taken from a large population data set obtained from the TCGA website discarding those without clinical data and repeated samples were used for subsequent analyzing. O-CIMP-negative and O-CIMP-positive could be distinguished through the extracted age-dependent features.

Figure 4: Validations of Age-dependent Features and marker Panel Signatures.
and the identified age-dependent marker panel (Table 1). The validating results demonstrated the age-dependent marker panel be able to perform the similar performance as the features in KM plot (shown in Figure 2.) indicated the ability as well that age-dependent marker panel be representing the feature. Through the log-rank test, the refined age-dependent marker panel not only presented the similar performance as age-dependent feature, but also presented a more favorable performance than the reported marker panel without the age parameter. The statistical significance test through the developed R package [29], viz. CGPredictor, a statistical significance test was implement to examine the bootstrap resampling. The result showed the confidence of the extracted age-dependent features for the subsequent analysis. These results also indicated that the age-dependent features were reliable for the two phenotypes, and aging is an effective parameter for the distinction of HGSOC phenotypes as well. Moreover, the significance of the age-dependent marker panel in distinguishing the O-CIMP-negative and -positive phenotypes were also examined by evaluating their log-rank test after 1,000 iterations with randomly selected forty four genes from the original gene set. Furthermore, some genes in the age-dependent marker panel have been previously reported to have a strong connection with ovarian cancer and tumor growth. Among the age-dependent marker panel, Ko and co-workers [42] showed that the expression of HOXA9 was associated with poor survival of patients with ovarian cancer and in mouse xenograft models. Their results indicated that HOXA9 expression promotes a microenvironment that is permissive for tumor growth. Inducible heat shock protein (HSPA1A), is highly expressed on the cell surface of tumor cells, but not on normal cells, and promotes tumor cell growth and survival. It also interacts with effector cells of the innate immune system and affects their activity. As for the promoter of CALCA, it was also informative for differentiating methylation between the early stages of ovarian disease and the healthy maintenance of control. Our results may have a potential implication for the association of biomarker candidates with ovarian cancer especially when conducting the effects associated with age on DNA methylation profiles. Taken together, these results indicated that the extracted age-dependent marker panel deserves consideration for further clinical testing to identify their potential usefulness in clinical molecular diagnosis and targeted treatment of patients with HGSOC. Our results may have the potential implications for cancer prevention, risk prediction, detection, prognosis, and therapy.

Acknowledgments
This work was supported by Research Grants from the Ministry of Science and Technology, Taiwan (MOST 103-2221-E-006-254-MY2).

Disclosures
The authors declare that they have no competing interests.

References


