European Journal of Cancer 64 (2016) 149–158

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Original Research

RNASeq analysis reveals biological processes governing the clinical behaviour of endometrioid and serous endometrial cancers

Christophe Lemetre [a], Begoña Vieites [b], Charlotte K.Y. Ng [a], Salvatore Piscuoglio [a], Anne M. Schultheis [a], Caterina Marchiò [a,c], Rajmohan Murali [a,d], Maria A. Lopez-García [b], Jose C. Palacios [e], Achim A. Jungbluth [a], Luigi M. Terracciano [f], Jorge S. Reis-Filho [a,g], Britta Weigelt [a,*]

[a] *Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA*
[b] *Department of Pathology, University of Seville, University Hospital Virgen del Rocío, Seville, Spain*
[c] *Department of Medical Sciences, University of Turin, Turin, Italy*
[d] *Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, USA*
[e] *Department of Anatomic Pathology, University Hospital Ramón y Cajal, Madrid, Spain*
[f] *Molecular Pathology Division, Institute of Pathology, University Hospital Basel, Basel, Switzerland*
[g] *Human Oncology & Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA*

Received 15 January 2016; received in revised form 16 May 2016; accepted 25 May 2016

KEYWORDS
Endometrial cancer;
Gene expression;
Biological process;
Outcome

Abstract ***Background:*** Endometrial carcinoma comprises a group of tumours with distinct histologic and molecular features and clinical behaviour. Here, we sought to define the biological processes that govern the clinical behaviour of endometrial cancers.
Methods: Sixteen prototype genes representative of different biological processes that would likely play a role in endometrial and other hormone-driven cancers were defined. RNA-sequencing gene expression data from 323 endometrial cancers from The Cancer Genome Atlas (TCGA) were used to determine the transcription module of each prototype gene. The expression of prototype genes and modules and their association with outcome was assessed in univariate and multivariate survival analyses. The association of MSH6 expression with outcome was validated in an independent cohort of 243 primary endometrial cancers using immunohistochemistry.

* *Corresponding author*: Department of Pathology, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA. Tel.: +1 212 639 2332; fax: +1 212 639 2502.
E-mail address: weigeltb@mskcc.org (B. Weigelt).

http://dx.doi.org/10.1016/j.ejca.2016.05.028

Results: We observed that the clinical behaviour of endometrial cancers as a group was associated with hormone receptor signalling, PI3K pathway signalling and DNA mismatch repair processes. When analysed separately, in endometrioid carcinomas, hormone receptor, PI3K and DNA mismatch repair modules were significantly associated with outcome in univariate analysis, whereas the clinical behaviour of serous cancers was likely governed by apoptosis and Wnt signalling. Multivariate survival analysis revealed that *MSH6* gene expression was associated with outcome of endometrial cancer patients independently from traditional prognostic clinicopathologic parameters, which was confirmed in an independent cohort at the protein level.
Conclusion: Endometrioid and serous endometrial cancers are underpinned by distinct molecular pathways. MSH6 expression levels may be associated with outcome in endometrial cancers as a group.

1. Introduction

Endometrial cancer (EC), the most common gynaecologic malignancy in the USA, comprises a heterogeneous group of tumours with distinct histologic features, biological behaviour and treatment response. Endometrioid and serous carcinomas account for the majority of ECs. Treatment decisions for patients with EC are primarily determined by surgical stage at presentation, histologic type and grade [1].

There is evidence to suggest that some molecular alterations are preferentially found in endometrioid endometrial carcinomas (EECs), including mutations in *PTEN* and *CTNNB1*, whereas others are more prevalent in serous endometrial carcinomas (SECs), such as *TP53* mutations [2,3]. These observations have been corroborated and expanded by the transcriptomic and genomic analyses of a large set of EECs and SECs carried out by The Cancer Genome Atlas (TCGA) [4]. The analyses performed by TCGA have led to an integrated genomic classification of EECs and SECs and the identification of the POLE (ultramutated), microsatellite instability (MSI) (hypermutated), copy-number low (endometrioid) and copy-number high (serous-like) subtypes, which have been shown to be underpinned by distinct combinations of genomic and epigenetic alterations [4].

Based on the molecular heterogeneity observed in ECs, we posited that the biological processes and pathways associated with outcome may be distinct between ECs of different histologic types, grades or integrated genomic subtypes and that additional markers predictive of clinical behaviour may be present in different subsets of the disease.

The aims of this study were i) to determine genes or gene expression modules representative of biological processes known to play a role in EC or in other hormone-driven cancers, and ii) to define the association of these genes and/or gene expression modules with outcome in ECs as a group, and in subgroups stratified according to histologic type, grade or integrated genomic types using RNA-sequencing data from the TCGA study [4].

2. Methods

2.1. Transcription modules and univariate analysis

RNA-sequencing gene expression and clinicopathologic data including outcome from 323 ECs were retrieved from the TCGA data portal (https://tcga-data.nci.nih.gov/docs/publications/ucec_2013/, files 'RNASeq' and 'Key Clinical Data'; accessed December 2015) [4] (Supplementary Table 1). The RSEM normalised gene-level expression data were obtained. As described by TCGA [4], genes lacking HGNC annotation or with small expression values in at least one-fourth of the samples were removed. The expression values of the final set of 20,502 genes were $\log_2$ transformed. We selected 16 'prototype' genes representative of biological processes that have been shown to play a role in EC or other hormone-driven cancers. The details of the gene selection are described in the Supplementary Methods [3–5]. The transcription modules of each prototype gene, which comprise the genes specifically co-expressed with each prototype gene (Supplementary Tables 2 and 3), were defined essentially as described by Desmedt *et al.* [5] and are described in the Supplementary Methods. The R script for the univariate models, transcription module development and module score calculations is available in the Supplementary Methods.

2.2. Gene ontology enrichment analysis

Gene ontology enrichment analyses of the 16 transcription modules were performed using Cytoscape (v.2.8.3) with the BinGO plugin (v.2.44) [6], and a hypergeometric test, with the false discovery rate controlled using the Benjamini and Hochberg procedure. Biological processes with a corrected $P < 0.05$ were deemed significant.

2.3. Tissue microarrays (TMAs) and immunohistochemistry

Tissue microarrays from the University Hospital Basel, Basel, Switzerland, and University Hospital Virgen del