Cancer Epidemiology 38 (2014) 321–327

Contents lists available at ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.cancerepidemiology.net

Systems approach to identify environmental exposures contributing to organ-specific carcinogenesis

J. Christopher States [a,*], Ming Ouyang [b,1], C. William Helm [c,2]

[a] Department of Pharmacology & Toxicology, University of Louisville, Louisville, KY, USA
[b] Department of Computer Engineering & Computer Science, University of Louisville, Louisville, KY, USA
[c] Department of Obstetrics, Gynecology & Women's Health, Saint Louis University, St. Louis, MO, USA

ARTICLE INFO

Article history:
Received 12 September 2013
Received in revised form 27 January 2014
Accepted 10 March 2014
Available online 12 April 2014

Keywords:
miRNA
mRNA
Chemicals
Carcinogenesis

ABSTRACT

Background: The most effective way to reduce cancer burden is Q2 prevention which is dependent on identifying individuals at risk for a particular cancer and counseling them to avoid exposure to causative agents. Other than a few well characterized environmental agents linked to specific cancers, linkage between any particular environmental exposure and a specific type of cancer is mostly unknown. Thus, we propose a systems approach to analyze publicly available large datasets to identify candidate agents that play a role in organ-specific carcinogenesis.
Methods: Publicly available datasets for mRNA and miRNA expression in ovarian cancer were queried to define the differentially expressed genes that are also targets of differentially expressed miRNAs. These target genes were then used to query the Comparative Toxicogenomics Database to identify interacting chemicals and also were analyzed by Ingenuity Pathway Analysis to identify pathways.
Results: The interacting chemicals interact with genes in known pathways in ovarian carcinogenesis and support the hypothesis that these chemicals are likely etiologic agents in ovarian carcinogenesis.
Conclusion: A systems approach may prove useful to identify specific etiologic agents to better develop personalized preventive medicine strategies for those most at risk.

1. Introduction

It is clear that carcinogenesis is driven by mutations and epigenetic changes in critical genes regulating cell replication and maintenance of genome integrity [1]. What is less clear are the specific environmental agents inducing oncogenic changes driving most cancers. Sometimes the oncogenic changes are epigenetic with changes in DNA methylation patterns playing a role in gene inactivation and activation [2]. More recently, changes in microRNA expression have been identified as playing a role in carcinogenesis [2–5]. Environmental agents that act as epimutagens have been identified [6–9], but not yet causally linked to specific cancers. The agents driving mutations in some cancers are well understood. For instance, ultraviolet light clearly is the environmental agent causing DNA damage resulting in mutations driving most skin cancers [10]. Likewise, carcinogens in tobacco smoke are known drivers of the mutations in lung cancers [11]. However, for many other tumors, the etiologic agents are unknown.

Identification of specific environmental exposures linked to organ-specific carcinogenesis can inform efforts to prevent those cancers from developing, especially in genetically susceptible individuals. Thus, development of ways to identify candidate exposures for further study and confirmation of linkage is vital for individualized preventive medicine [12].

We propose here that available large data sets provide the means to identify candidate agents for study of causality. Identifying environmental exposures that may contribute to organ-specific carcinogenesis would provide the basis for epidemiological studies to determine the causal relationships.

Environmental exposures are capable of inducing epigenetic changes, including changes in microRNA expression, that result in differential gene expression [13–15]. Most importantly, changes in microRNA expression may cause permanent changes in regulation of gene expression patterns leading to altered cellular phenotypes.

* Corresponding author. Tel.: +1 502 852 5347; fax: +1 502 852 2123.
E-mail addresses: jcstates@louisville.edu, jcstates@earthlink.net (J.C. States).
[1] Current address: Department of Computer Science, University of Massachusetts Boston, Boston, MA, USA.
[2] Current address: Northern Gynaecological Oncology Centre, Queen Elizabeth Hospital, Gateshead, Tyne and Wear NE9 5XT, United Kingdom.

http://dx.doi.org/10.1016/j.canep.2014.03.005
1877-7821/

These alterations in microRNAs represent dysregulation of gene expression in the cancer cells and may be a manifestation of environmental exposures [13]. Thus, we hypothesize that differential expression of microRNAs in a cancer are a consequence of carcinogenic environmental exposure and drive gene expression changes resulting from these environmental exposures. Hence, the cancer stem cell phenotype may arise as a consequence of environmental exposure driven changes in gene expression whether these changes are the consequence of mutation or epigenetic modification. We propose that a systems approach utilizing publicly available data for differential microRNA expression associated with a cancer and for differential mRNA expression in that cancer in conjunction with known associations between chemical exposures and gene expression will identify chemicals that play a role in carcinogenesis. We demonstrate the potential for this approach to identifying environmental contributions to development of a specific class of cancers using ovarian cancer (OvCa) as an example.

Other than an association of a small fraction with inherited mutations in BRCA1 and BRCA2 [12], most cases of OvCa are sporadic with cause unknown [16]. Using publicly available datasets, we identified both mRNAs and microRNAs differentially expressed in ovarian cancers. Intersection of genes with differential expression in ovarian cancer and putative gene targets of differentially expressed microRNAs in ovarian carcinomas gave a list of candidate genes for chemical interactions. Ingenuity Pathway analyses of these genes show that they participate in pathways involved in cancer and other chronic diseases. This list of genes was submitted to the Comparative Toxicogenomics Database (CTD) and a list of chemicals with known interactions with these genes was collected. The list of chemicals interacting with the OvCa specific genes included known carcinogens (arsenic, asbestos, benzo[a]pyrene), co-carcinogens (tetrachlorodibenzodioxin, tetradecanoylphorbol acetate), peroxides, heavy metals (Cd, Cu, Zn), epigenotoxicants (hydralazine, valproic acid), inflammation inducers (zymosan, lipopolysaccharide), steroid hormones (estradiol, progesterone, dexamethasone), dietary chemopreventives (indole-3-carbinol, curcumin) and chemotherapeutics (doxorubicin, tamoxifen, tretinoin). While some of these chemicals and pathways have been implicated previously in ovarian carcinogenesis (for instance, asbestos [17], tetrachlorodibenzodioxin [18]), many have not, and their identification will enable the development of targeted epidemiological studies to determine the role of exposure to these chemicals in ovarian carcinogenesis. This systems approach will likely be useful to identify environmental exposures contributing to cancer of other organs with unknown etiologies providing information essential to personalized preventive approaches to reducing cancer incidence.

2. Materials and methods

Genes differentially expressed in ovarian cancer compared to normal ovary tissue were determined using publicly available data sets. The two data sets of gene expression used in this study were as follows. Normal ovary gene expression was extracted from NCBI Gene Expression Omnibus (GEO) GSE1133. This data set contains baseline mRNA abundance of various human and mouse tissues. The human ovary data were assayed by two Affymetrix chip platforms: U95 and U133. Ovarian tumor gene expression was extracted from NCBI GEO GSE6822. This data set contains human ovarian cancer gene expression data assayed by the Affymetrix Hu6800 platform from a large heterogeneous set of 74 ovarian tumors. The Affymetrix U95 platform assays 8075 genes. The company provides a list of good matches between U95 and U133 platforms. From this list we extracted 5804 genes as being present in both platforms.

We used Perl scripts to match the three platforms, and to extract genes that were present on all three platforms. The *p*-values of differential expression between healthy ovaries and ovarian cancer were calculated with the regularized *t*-test [19], and the Benjamini and Hochberg correction [20] was applied to control the false discovery rate. After analysis of significance, 723 genes (Set A) were designated as differentially expressed between ovary and ovarian cancer.

The microRNAs differentially expressed in ovarian cancer cells identified by Wyman et al. [21] were used to query the Sanger miRBase for target mRNAs. Wyman et al. identified 56 differentially expressed microRNAs. The Sanger miRBase lists 3916 mRNAs (Set B) as targets of these 56 microRNAs. The intersection of Sets A and B has 380 genes (Set C).

Next we took the intersection of Set A and Set B, and obtained 380 genes. These 380 differentially expressed mRNA targets of differentially expressed microRNAs were designated gene Set C. Set C was submitted to the Comparative Toxicogenomics Database (CTD, http://ctdbase.org/ [22]) to find the known interactions of these genes with chemicals. The CTD is a curated database compiling associations between gene expression, chemical exposure and disease data reported in the literature. The interacting chemicals were ranked by their numbers of gene interactions. Lastly, Set C was analyzed for gene interaction networks and pathways with Ingenuity Pathway Analysis software.

3. Results

The 380 genes identified as differentially expressed in ovarian cancer and also as targets of microRNAs differentially expressed in ovarian cancer are listed in Fig. 1A. The majority of the 380 genes had decreased expression (Fig. 1A, green) in ovarian cancer cells with only 99 showing increased expression (Fig. 1A, red).

Query of the CTD with the 380 differentially expressed genes that also were targets of differentially expressed microRNAs in ovarian cancer cells (Fig. 1A) identified 889 chemicals with known interactions with these genes. Some genes had interactions with many chemicals, some with as few as one. Likewise, some chemicals had interactions with a large number of genes, and some with as few as one. The number of genes interacting with each chemical ranged from 1 to 162 as shown in Fig. 1B.

We arbitrarily reduced the set to the 110 chemicals interacting with at least nine genes to show in Table 1. Among these chemicals are known carcinogens, heavy metals, steroid hormones, chemopreventives, and cancer chemotherapeutics, to name just a few of the classes. It should be noted that some of the listings are different forms of the same active agent, *e.g.* arsenic is listed also as sodium arsenite and arsenic trioxide.

We next performed network analysis of the genes with differential expression in ovarian cancer cells using IPA. The top three networks are shown in Figs. 2–4. Included in these networks are chemical interactions and functional pathways in the IPA knowledge base. Most of the chemicals identified as interacting with genes in these networks by IPA are found in the list of chemicals identified by the CTD analysis and include PAH, arsenicals, asbestos, cadmium, dioxin and PCBs. These networks include ones involved with tissue morphology, cancer and nucleic acid metabolism. The top network (Hematological System Development and Function, Humoral Immune Response, Tissue Morphology; Fig. 2) includes genes in three ovarian cancer subnetworks: "Response to therapy" (platelet derived growth factor complex [PDGF], chemokine (c-c motif) ligand 2 [CCL2]), "Prognosis" (CCL2, FBJ murine osteosarcoma viral oncogene homolog B [FOSB], nuclear factor kappa B [NFKB]), and "Diagnosis"