JOURNAL OF PROTEOMICS 89 (2013) 165–178

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Protein isoform-specific validation defines multiple chloride intracellular channel and tropomyosin isoforms as serological biomarkers of ovarian cancer

Hsin-Yao Tang[a], Lynn A. Beer[a], Janos L. Tanyi[b], Rugang Zhang[c], Qin Liu[a], David W. Speicher[a,*]

[a]Center for Systems and Computational Biology and Molecular and Cellular Oncogenesis Program, The Wistar Institute, Philadelphia, PA, USA
[b]Ovarian Cancer Research Center, University of Pennsylvania, Philadelphia, PA, USA
[c]Gene Expression and Regulation Program, The Wistar Institute, Philadelphia, PA, USA

ARTICLE INFO

Article history:
Received 22 January 2013
Accepted 10 June 2013
Available online 21 June 2013

Keywords:
Biomarker validation
Biomarker isoforms
Protein families
Ovarian cancer biomarkers
Multiple reaction monitoring
Proteomics

ABSTRACT

New serological biomarkers for early detection and clinical management of ovarian cancer are urgently needed, and many candidates have been reported. A major challenge frequently encountered when validating candidates in patients is establishing quantitative assays that distinguish between highly homologous proteins. The current study tested whether multiple members of two recently discovered ovarian cancer biomarker protein families, chloride intracellular channel (CLIC) proteins and tropomyosins (TPM), were detectable in ovarian cancer patient sera. A multiplexed, label-free multiple reaction monitoring (MRM) assay was established to target peptides specific to all detected CLIC and TPM family members, and their serum levels were quantitated for ovarian cancer patients and non-cancer controls. In addition to CLIC1 and TPM1, which were the proteins initially discovered in a xenograft mouse model, CLIC4, TPM2, TPM3, and TPM4 were present in ovarian cancer patient sera at significantly elevated levels compared with controls. Some of the additional biomarkers identified in this homolog-centric verification and validation approach may be superior to the previously identified biomarkers at discriminating between ovarian cancer and non-cancer patients. This demonstrates the importance of considering all potential protein homologs and using quantitative assays for cancer biomarker validation with well-defined isoform specificity.

Biological significance
This manuscript addresses the importance of distinguishing between protein homologs and isoforms when identifying and validating cancer biomarkers in plasma or serum. Specifically, it describes the use of targeted in-depth LC–MS/MS analysis to determine the members of two protein families, chloride intracellular channel (CLIC) and tropomyosin (TPM) proteins that are detectable in sera of ovarian cancer patients. It then establishes a multiplexed isoform- and homology-specific MRM assay to quantify all observed gene products in these two protein families as well as many of the closely related tropomyosin isoforms. Using this assay, levels of all detected CLICs and TPMs were quantified in ovarian cancer patient and control subject sera. These results demonstrate that in addition to the previously known CLIC1, multiple tropomyosins and CLIC4 are promising new ovarian

* *Corresponding author at:* The Wistar Institute, 3601 Spruce St., Room 272A, Philadelphia, PA 19104, USA. Tel.: +1 215 898 3972; fax: +1 215 495 6915.
E-mail address: speicher@wistar.org (D.W. Speicher).

1874-3919/$ – see front matter
http://dx.doi.org/10.1016/j.jprot.2013.06.016

cancer biomarkers. Based on these initial validation studies, these new ovarian cancer biomarkers appear to be superior to most previously known ovarian cancer biomarkers.

1. Introduction

Epithelial ovarian cancer (EOC) is the most lethal gynecological cancer in the United States, with an estimated 22,280 new cases detected and 15,500 deaths in 2012 [1]. When diagnosed early (Stages I/II), treatment is generally successful, with a five-year survival rate of up to 90%; but unfortunately, most cases are not detected until after the cancer has spread, resulting in a dismal five-year survival rate of 30% or less [2,3]. There are currently no effective screening tests for EOC early detection, and existing clinical tests using protein biomarkers, such as cancer antigen 125 (CA-125), human epididymis protein-4 (HE4), or multivariate OVA1, are only approved for monitoring disease recurrence, therapeutic response, or for use in managing women with an ovarian adnexal mass [4–7]. The most commonly used EOC biomarker, CA125, is recognized as a poor biomarker for early detection due to its high false-positive rate and poor sensitivity and specificity [8,9]. Better biomarkers or, more likely, panels of markers are urgently needed to diagnose early-stage EOC with high sensitivity and specificity, and for clinical management of the disease after initial diagnosis.

We and others have leveraged proteomics to discover new EOC biomarkers. Diverse experimental systems, including cancer cell cultures, tissue specimens, ascites fluid, secretomes, and mouse models, have been investigated using numerous proteomics strategies in attempts to identify better EOC biomarkers [10–21]. Using an in-depth 4D analysis of serum from severe combined immunodeficiency (SCID) mice containing a human endometrial ovarian cancer tumor, we recently identified 106 candidate human proteins with at least two peptides [21]. In that study, we performed a pilot validation on candidate biomarkers in the 20–55 kDa region of 1D SDS gels and found that nearly half the proteins discovered in the xenograft mouse model could be detected in human serum using multiple reaction monitoring analysis. Two of the tested candidates, chloride intracellular channel 1 (CLIC1) and cathepsin D 30 kDa fragment (CTSD-30 kDa), showed significantly elevated serum levels in cancer patients compared with non-cancer controls [21].

A major advantage of xenograft mouse models is that proteins shed by human tumors into mouse blood can be unambiguously distinguished by exploiting species differences in peptide sequences identified by liquid chromatography–tandem mass spectrometry (LC–MS/MS). However, the capacity to distinguish species differences diminishes as the sequence homology between the two species for specific proteins increases, especially with lower-abundance proteins where sequence coverage is typically low. Similarly, the capacity to distinguish between homologous human members of protein families during the discovery phase is often limited by low sequence coverage of candidate biomarkers.

The high number of candidates identified using current proteomics methods, coupled with the lack of well-characterized immunoassays for most of the new candidates, necessitates the use of alternative quantitative techniques capable of screening candidates in patient serum or plasma. MRM has recently emerged as the most effective targeted quantitative technique for biomarker verification due to its high selectivity and multiplexing capability [22,23]. Despite the recent wide use of MRM as a quantitative tool in proteomics, concerns have been raised regarding its specificity and selectivity [24,25]. One of the major concerns is the specificity of peptide signals chosen for MRM quantitation, especially in a complex proteome background. In addition, the use of proteolytic fragmentation in discovery experiments creates a fundamental problem of protein inference [26]. In complex proteomes such as serum or plasma, it is often not trivial to reconnect peptides to a precise protein of origin due to the presence of shared tryptic peptides from protein isoforms (alternatively splicoforms from a single gene) and members of a protein family (proteins produced by homologous genes). Although the term "protein isoform" has been variably defined, in this study, protein isoform will be used to indicate related protein sequences from a single gene. These isoforms may be produced by alternative splicing, proteolytic processing, or other posttranslational modifications. Homologs or homologous proteins will be used to refer to members of a protein family produced by different genes.

The protein inference problem is further aggravated by substantial redundancy in the database caused by completely or nearly completely redundant entries, partial sequences, polymorphisms, sequencing errors, etc. In order to prevent inflation of protein counts in discovery experiments, a common practice is to use the parsimony principle or the Occam's razor constraint to report a minimal list of proteins that can account for all observed peptides [26]. This often results in multiple protein entries being assigned to a set of peptides due to homologous proteins and redundancy in the database, and the actual protein(s) present in the sample can be ambiguous.

Importantly, the failure to detect peptides unique to an isoform does not prove that the isoform is not present in the sample. This is particularly critical in biomarker studies, as the correct or best biomarker may be a protein family member or isoform that was not explicitly identified in the discovery phase. It is important to differentiate isoforms and homologous family members because these related proteins are often associated with distinct structural or functional roles [27–29]. This is particularly important in cancer biomarker validation studies, because only the specific members of a protein family that are shed by the tumor into the blood should ideally be quantified. In this regard, any assay, including sandwich ELISA assays, could give misleading results if the isoform and family member specificity is unknown or if multiple related proteins are quantitated as a group. MRM has been used successfully to quantitate specific protein isoforms or mutant proteins especially from cell extracts [30–32]. However, targeted serological quantitation of serum biomarkers using MRM methods is a relatively new approach, and little is known about how the accuracy of protein quantitation is affected by the isoform and family member specificity of the peptides chosen.