Cancer Treatment Reviews 40 (2014) 129–138

Contents lists available at SciVerse ScienceDirect

Cancer Treatment Reviews

journal homepage: www.elsevierhealth.com/journals/ctrv

Laboratory-Clinic Interface

Breast cancer classification by proteomic technologies: Current state of knowledge

S.W. Lam, C.R. Jimenez, E. Boven *

Department of Medical Oncology, VU University Medical Center, De Boelelaan 1117, 1081HV, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 26 March 2013
Received in revised form 20 June 2013
Accepted 25 June 2013

Keywords:
Breast cancer
Molecular subtypes
Proteomics
Biomarkers
Review

ABSTRACT

Breast cancer is traditionally considered as a heterogeneous disease. Molecular profiling of breast cancer by gene expression studies has provided us an important tool to discriminate a number of subtypes. These breast cancer subtypes have been shown to be associated with clinical outcome and treatment response. In order to elucidate the functional consequences of altered gene expressions related to each breast cancer subtype, proteomic technologies can provide further insight by identifying quantitative differences at the protein level. In recent years, proteomic technologies have matured to an extent that they can provide proteome-wide expressions in different clinical materials. This technology can be applied for the identification of proteins or protein profiles to further refine breast cancer subtypes or for discovery of novel protein biomarkers pointing towards metastatic potential or therapy resistance in a specific subtype. In this review, we summarize the current state of knowledge of proteomic research on molecular breast cancer classification and discuss important aspects of the potential usefulness of proteomics for discovery of breast cancer-associated protein biomarkers in the clinic.

Introduction

Breast cancer affects more than 1.3 million women worldwide each year and accounts for about 14% of cancer-related deaths [1]. The incidence of breast cancer has increased over the past decades and is expected to rise substantially in the coming years [2]. Hence, breast cancer will remain a considerable health burden.

Work on breast cancer has revealed substantial tumor heterogeneity consisting of different molecular subtypes, each with distinct biological and clinical characteristics [3,4]. In the pivotal study by Perou et al. [5], it has been shown that differential gene expression patterns account for heterogeneity among breast carcinomas. Based on the so-called intrinsic gene signatures, four major breast cancer subtypes were initially classified: luminal, HER2-enriched, basal-like and normal breast-like subtype. Subsequent studies by Perou et al. and others have expanded these initial findings by providing additional information for further refinements and adjustments of the breast cancer classification [3,4]. Within the luminal subtype characterized by the expression of luminal epithelial markers, three groups are currently recognized: luminal A, luminal B/HER2-negative and luminal B/HER2-positive. Basal-like breast cancer is a heterogeneous group of tumors comprising different histologies, which express basal epithelial markers. The normal breast-like subtype was located in a cluster containing normal breast and benign tumor samples and showed overexpression of genes related to adipose tissue and non-epithelial cell types in the original and subsequent validation studies [3–5]. This subtype may also be a technical artifact due to low tumor cellularity [6]. Hence, the normal breast-like subtype was often overlooked and was consequently poorly characterized.

The classification of breast cancer based on gene expression patterns has resulted into attempts to characterize clinically meaningful subgroups showing correlation with survival [7,8], disease relapse [8], site of preference of metastatic spread [9] and chemotherapy response [8,10]. Since microarray techniques are rather expensive and not readily available, immunohistochemistry (IHC) is an important method to define surrogate protein biomarkers for the classification of breast cancer [11]. The main advantages of IHC are its lower costs and easy implementation into standard pathology workflow. It has been shown that the molecular classification by microarray analysis corresponds reasonably well to IHC classification of different breast carcinomas [12,13]. Consequently, molecular and IHC classifications are concomitantly used to define the breast cancer subtypes (Table 1).

Ongoing research will identify new subtypes within the designated breast cancer classification [14]. Complementary to the genomic-based approach, proteomics might provide new insights into aberrant processes among breast cancer subtypes and may identify additional proteins or protein profiles to refine current breast cancer classifications. Moreover, proteomics might reveal

* Corresponding author. Tel.: +31 20 4444336; fax: +31 20 4444079.
E-mail addresses: sw.lam@vumc.nl (S.W. Lam), c.jimenez@vumc.nl (C.R. Jimenez), e.boven@vumc.nl (E. Boven).

0305-7372/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ctrv.2013.06.006

Table 1
Breast carcinoma subtypes: histopathological, molecular and clinical features.

Molecular subtype	Prevalence[a]	IHC definition[b]	Additional markers	Genes	Histological grade	TP53 mutation	Prognosis	Consensus recommendation for (Neo) adjuvant systemic treatment[b]
Luminal A	50–60%	ER+ and/or PgR+ HER2− Ki-67 low	CK8/18+ FOXA1+	*ESR1, GATA3, KRT8, KRT18, XBP1, FOXA1, TFF3, CCND1, LIV1*	Good differentiation	Low	Good	Endocrine therapy alone[d]
Luminal B	10–30%	–	FGFR1 and ZIC3 amplification	*ESR1, GATA3, KRT8, KRT18, XBP1, FOXA1, TFF3, SQLE, LAPTM4B*	Moderate differentiation	Intermediate high	Intermediate	–
Luminal B (HER2 negative)	15–20%	ER+ and/or PgR+ HER2− Ki-67 high	–	Not examined[c]	–	–	–	Endocrine therapy ± chemotherapy[e]
Luminal B (HER2 positive)	6%	ER+ and/or PgR+ Any Ki-67 HER2+	–	Not examined[c]	–	–	–	Endocrine + cytotoxic + anti-HER2 therapy
HER2-enriched	10–15%	HER2 + ER− and PgR−	CK5/6+ GRB7+	*ERBB2, GRB7*	Poor differentiation	High	Poor	Chemotherapy + anti-HER2 therapy[f]
Basal-like	10–20%	ER− and PgR− HER2−	EGFR+ CK5/6+ CK14+ CK17+ HER1+ Cyclin E+ CDKN2A+ RB1: low/− BRCA1: low/− FGFR2: amplification	*KRT5, CDH3, ID4, FABP7, KRT17, TRIM29, LAMC2, ITGB4*	Poor differentiation	High	Poor	Chemotherapy for triple negative breast cancer (ductal)

ER, estrogen receptor; IHC, immunohistochemistry; PgR, progesterone receptor.
[a] Prevalence data as reported in [7,9,12,25].
[b] According to St. Gallen International Expert Consensus 2011 [71].
[c] Not examined in the original discovery studies by Perou et al. [5] and Sørlie et al. [3].
[d] Some high-risk patients e.g. high nodal status require chemotherapy.
[e] Inclusion and type of chemotherapeutic agents may depend on level of endocrine receptor expression, perceived risk and patient preference.
[f] Patients at very low risk e.g. pT1aN0 may be observed without systemic adjuvant treatment.

biological insights and identify protein biomarkers defining differences in therapy resistance, prognosis and metastatic spread within a specific subtype. The purpose of this review is to discuss the current state of knowledge of proteomic studies conducted in relation with the molecular classification of breast cancer.

A brief overview of proteomic technologies

Definition of proteomics

Proteomics is a term which refers to a large-scale study of proteins encompassing several aspects, such as protein identification, protein ontology, protein–protein interaction, pathways involvement, quantification and functional analysis. In addition, proteomics involves the identification of protein subgroups, such as kinases (kinome), secreted proteins (secretome), phosphorylated proteins (phosphoproteomics), exosomal proteins (exosome) and proteases (degradome). A multitude of complex biological samples can be analyzed using proteomic technologies, such as tissue, serum, plasma, saliva, nipple fluid, urine, cerebrospinal fluid and so on, which makes proteomics an attractive strategy for biomarker discovery. This section aims to provide a brief overview of commonly used proteomic approaches based on the utilization of mass spectrometry (MS) as well as main methods for validation of protein candidates.

MS-based proteomics for biomarker discovery

Mass spectrometers have increasingly been employed as a platform for discovery proteomics or targeted follow-up of proteins in complex biological samples representing different disease conditions. In recent years, remarkable progress towards near complete proteome coverage and high sample throughput has been made by technological and methodological advancements, which is reviewed elsewhere [15]. Additionally, reproducible results can be achieved in an optimized workflow. It requires, however, a large initial investment to acquire a mass spectrometer, specialized staff and highly sophisticated bioinformatics to obtain reliable results.

The experimental design of earlier proteomic discovery studies are those of two-dimensional gel electrophoresis (2D-PAGE) coupled with MS. Proteins are separated in two dimensions based on mass and charge differences and can be visualized with e.g. Coomassie Brilliant Blue or silver staining. The resulting protein spots on a gel can be picked up for protein identification by MS instruments. Although 2D-PAGE has been widely used in proteomic studies, it suffers from inherent problems including the requirement of a large amount of material, gel-to-gel variation, limited dynamic range, low-throughput identification of proteins and bias towards abundant proteins. Two-dimensional difference gel electrophoresis (2D-DIGE) is designed for analysis of multiple samples on one gel after labeling with different fluorescent dyes, while it simplifies the analysis and reduces gel-to-gel variations.

A typical 'modern' MS-based proteomic workflow encompasses multidimensional protein separation by gel electrophoresis and nano-liquid chromatography prior to identification and quantification by MS. The selection of appropriate separation steps is crucial, because the number of identified proteins differs substantially [16]. Different MS instruments, including a variety of techniques, can be used for protein identification and quantification and have recently been reviewed [17]. In general, two established MS-based strategies have been widely adopted: stable isotope labels or label-free quantitative proteomics.

The isotope labeling technologies are based on the principle of isotope-induced shifts in mass, which can be detected in the same MS experiment. This allows simultaneous quantification of proteins among disease conditions. The use of isotope labels delivers significantly improved accuracy of protein quantification, but it comes at a cost in terms of expensive isotope labels, the requirement of specialized software and statistical challenges. In the label-free MS approach, all samples are processed and analyzed in parallel allowing the flexibility to conduct multiple comparisons [18]. The total number of identified peptides corresponding to a