Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Biotechnology

Networks in proteomics analysis of cancer

Wilson Wen Bin Goh[1,2] and Limsoon Wong[1]

Proteomics provides direct biological information on proteins but is still a limited platform. Borrowing from genomics, its cancer-specific applications can be broadly categorized as (1) pure diagnostics, (2) biomarkers, (3) identification of root causes and (4) identification of cancer-specific network rewirings. Biological networks capture complex relationships between proteins and provide an appropriate means of contextualization. While playing significantly larger roles, especially in 1 and 3, progress in proteomics-specific network-based methods is lagging as compared to genomics. Rapid hardware advances and improvements in proteomic identification and quantification have given rise to much better quality data alongside advent of new network-based analysis methods. However, a tighter integration between analytics and hardware is still essential for network analysis to play more significant roles in proteomics analysis.

Addresses
[1] Department of Computer Science, National University of Singapore, COM1 Building, 13 Computing Drive, Singapore 117417, Singapore
[2] Department of Computing, Imperial College London, Exhibition Road, London SW7 2AZ, United Kingdom

Corresponding author: Wong, Limsoon (WongLS@Comp.NUS.EDU.SG)

Current Opinion in Biotechnology 2013, **24**:1122–1128

This review comes from a themed issue on **Pharmaceutical biotechnology**

Edited by **Ajikumar Parayil** and **Federico Gago**

For a complete overview see the Issue and the Editorial

Available online 5th March 2013

0958-1669/$ – see front matter, © 2013 Elsevier Ltd. All rights reserved.

http://dx.doi.org/10.1016/j.copbio.2013.02.011

Introduction

Cancer is a heterogeneous disease driven by mutation. A cancerous growth or tumor consists of various subcolonies while individuals presenting similar cancers possess vastly different etiologies. Thus even with the sophistication of current genomics platforms, resolving the complexity is difficult. Furthermore, genomics provide only indirect information. To gain deeper perspective, the proteome must be assayed. In normal cells, the proteome consists of at least 500k moieties, not including PTMs [1]. Given that cancers are highly mutated, the cancer proteome is expected to be vastly larger [2].

Current advances in proteomics advances boast heightened reliability and deeper profiling. Yet, considerable improvement (in coverage, consistency, quantitation and scalability) is still needed for robust analysis [3].

General principles for cancer systems biology have been discussed elaborately elsewhere [4,5] but these derive mostly from genomics. The integration of proteomics with systems biology is relatively lagging. Drawing parallels from genomics, four primary utilizations can be envisaged in proteomics. These are (1) pure diagnostics, (2) biomarkers, (3) identification of root causes and (4) reconstruction of regulatory circuits/networks (the major publications and bugbears for each utilization are summarized in Tables 1 and 2).

To elaborate, pure diagnostics is concerned with obtaining consistent measurements and obtaining good coverage of the proteome. Biomarkers allow for condition-specific identification — for example, confirming disease (state) or establishing prognosis — by detecting a set of protein(s) consistently differential between normal and cancer state from non-tissue fluids. Identification of root causes is concerned with functional analysis, and establishing the set of critical and causal alterations. Finally, reconstruction of regulatory circuits and networks allows for inference of cancer-specific rewirings/alterations.

For each of these utilizations, biological networks can play important roles. These are abstract representations of relationships between biological entities. As a whole, they provide emergent insight of the system. Networks typically used in contemporary research include protein–protein interaction networks (PPINs) and biochemical pathways (e.g. signaling and metabolic). As models, they can be used to overcome fundamental issues such as lack of coverage or reproducibility [6]. Using recent works (2010–2012), this survey examines improvements in each utilization: where network-based techniques have played a significant role in advancement, we highlight and evaluate the contribution. If not, and where applicable, we indicate how appropriate network-based methods could play future roles.

Diagnostics

Diagnostics is concerned with obtaining comprehensive and inclusive readings with high levels of reproducibility. Typically, the lack of coverage and consistency stems from extensive sample complexity and technical limitations — for example, high instrument sensitivity, limited detection range and limited peptide reference library.

Hence, in general, proteomics generally reports few proteins while reproducibility and inter-sample agreement

Table 1

Key network-related references on each proteomic utilization and whether networks have played a significant role in its recent advancement

Utilization	Key references	Have networks played a significant role in advancement?
Pure diagnostics	[7,8,12,45–47]	Yes
Biomarkers	[17••,19–22,49,50,52,53]	No. There is room for greater implementation
Cause identification	[26,27,34,35•,48,54]	Yes
Cancer-specific circuits/rewirings	[37,38••,39,43,44••,51]	This is driven more by technological advances in proteomics; genomics far more developed in this area

can be potentially abysmal [7]. Although high resolution mass spectrometry (HRMS) can report many proteins — for example, Nagaraj *et al.* [8] and Beck *et al.* [9] identified ~10k proteins in HeLa and U2OS cells respectively — these methods are laborious and non-scalable. Instead, HRMS is better suited to producing better reference networks by identifying the relevant proteome. These reference networks provide more suitable analytical contexts.

On protein identifications and proteome coverage, most peptide matching approaches depend on reference databases (e.g. NCBI or UniProt). Since cancers are highly mutative, many relevant matches may be missed. *De novo* identification algorithms can help, particularly as they become more sophisticated and rapid — a subset of unmatched spectra could be reanalyzed selectively. Networks can play an integral role here: combining networks with untargeted proteomics allows coverage expansion via associations. The most straightforward way is to enumerate network links to identified critical proteins [10,11]. More sophisticated generalizations involve identification of critical clusters and then reconfirming via searching the mass spectra using both peptide reference library matches and *de novo* identifications particularly where mutations are suspected (e.g. only one or few high confidence peptides identified for a seed-associated protein; Figure 1) [12]. These can be further verified downstream using fast and robust targeted approaches such as SWATH MS [13]. In turn, the network clusters derived therein can be used for resolving agreement issues, and for class discovery among other applications (Figure 1). These cluster-based signatures can also prove to be more powerful biomarkers (see next section) [14••].

To elaborate, when samples are highly variable with little overlaps, analysis becomes difficult. This inconsistency issue can be resolved by first identifying the critical subnets in which detected proteins localize — this greatly increases the confidence that the identified proteins are bona fide via their associations (Figure 1). Applying this principle, contextualizing the detected proteins to their respective relevant subnets is sufficient to recover the underlying patient subclasses. This is not normally possible from the data itself due to high variability in terms of reported proteins per patient, and between reported protein expressions [6].

Table 2

Major bugbears with each category and how networks can alleviate the issues

Utilization	Bugbears	Relevant networks usage
Pure diagnostics	• Coverage — incomplete information limits identification of robust molecular signatures • Consistency — patients with same disease present grossly different molecular profiles	• Coverage can be improved by identifying closely associated network proteins • Consistency can be vastly improved at the subnet level
Biomarkers	• High noise levels — highly complex proteome; far removed from cancer site	• Chance associations can be removed by identifying key network subnets • Use proteomic network studies to first isolate candidates followed by a directed search on the sample
Cause identification	• Incomplete and inconsistent assays • Difficult to translate protein lists into functional understanding	• Coverage and consistency can be greatly improved using networks to mine for associations • Pathways and networks can be used for identifying high-level proteins
Cancer-specific circuits/rewirings	• Incomplete pathway sets • Establishing novel pathway links is non-trivial	• Integrated databases provide a means of higher-resolution analysis • Can be elucidated experimentally but networks can be used to provide higher level explanations

Download English Version:

https://daneshyari.com/en/article/10232135

Download Persian Version:

https://daneshyari.com/article/10232135

Daneshyari.com