Proteomic investigation of intra-tumor heterogeneity using network-based contextualization - a case study on prostate cancer

Goh, Wilson Wen Bin; Zhao, Yaxing; Sue, Andrew Chi-Hau; Guo, Tiannan; Wong, Limsoon

2019

Goh, W. W. B., Zhao, Y., Sue, A. C., Guo, T. & Wong, L. (2019). Proteomic investigation of intra-tumor heterogeneity using network-based contextualization - a case study on prostate cancer. *Journal of Proteomics*, 206, 103446-.
<https://dx.doi.org/10.1016/j.jprot.2019.103446>

<https://hdl.handle.net/10356/150776>

<https://doi.org/10.1016/j.jprot.2019.103446>

© 2019 Elsevier B.V. All rights reserved. This paper was published in *Journal of Proteomics* and is made available with permission of Elsevier B.V.

Downloaded on 26 Apr 2023 19:36:30 SGT

Proteomic investigation of intra-tumor heterogeneity using network-based contextualization --- A case study on prostate cancer

Wilson Wen Bin Goh^{1*#}, Yaxing Zhao^{2#}, Andrew Chi-Hau Sue², Tiannan Guo³, Limsoon Wong^{4,5*}

1. School of Biological Sciences, Nanyang Technological University, Singapore
2. School of Pharmaceutical Science and Technology, Tianjin University, China
3. School of Life Sciences, Westlake University, China.
4. Department of Computer Science, National University of Singapore, Singapore
5. Department of Pathology, National University of Singapore, Singapore

*Corresponding Author(s): Dr. Wilson Wen Bin Goh

School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore

637551; Tel: +65 63162800, Fax: +65 6791 3856;

Email: wilsongoh@ntu.edu.sg, goh.informatics@gmail.com

[Limsoon Wong](#)

[Department of Computer Science, National University of Singapore, 13 Computing Drive, Singapore](#)

[117417; Tel: +65-6516-2902; Fax: +65-6779-7465;](#)

Email: wongls@comp.nus.edu.sg

[#First Author\(s\)](#)

Acknowledgements

This research was supported by a NRF-NSFC (Grant No. NRF2018NRF-NSFC003SB-006) to WWBG, the Westlake Startup Grant to TG, Zhejiang Provincial Natural Science Foundation of China (Grant No. LR19C050001) to TG, and a Kwan Im Thong Hood Cho Temple Chair Professorship to LW.

Author contribution

WWBG and LW designed, implemented the bioinformatics method and pipeline, performed analysis, and wrote the manuscript. YZ performed analysis and co-wrote the manuscript. TG developed the PCT-SWATH technology, seeded the direction on investigating ITH, and generated the prostate cancer datasets. WWBG and ASCH both acted as co-supervisors.

Conflict of Interest

The research group of TG is supported by SCIEX, which provided access to prototype instrumentation, and Pressure Biosciences Inc, which provided access to advanced sample preparation instrumentation.

1 Abstract (113 words)

2 Cancer is a heterogeneous disease, confounding the identification of relevant
3 markers and drug targets. Network-based analysis is robust against noise, potentially
4 offering a promising approach towards biomarker identification. We describe here the
5 application of two network-based methods, qPSP (Quantitative Proteomics Signature
6 Profiling) and PFSNet (Paired Fuzzy SubNetworks), in an intra-tissue proteome data
7 set of prostate tissue samples. Despite high basal variation, we find that traditional
8 statistical analysis may exaggerate the extent of heterogeneity. We also report that
9 network-based analysis outperforms protein-based feature selection with
10 concomitantly higher cross-validation accuracy. Overall, network-based analysis
11 provides emergent signal that boosts sensitivity while retaining good precision. It is a
12 potential means of circumventing heterogeneity for stable biomarker discovery.

13 **Keywords:** Proteomics; Networks; Bioinformatics; Systems Biology; Biomarker.

14 1. Introduction

15 Although cancers can be characterized by a set of phenotypic hallmarks --- e.g.
16 unlimited replicative potential, immunity to death signalling, angiogenesis amongst
17 other traits [1, 2] --- at its heart, cancers are a collection of highly complex diseases.
18 The main driver that initiates cancer progression is mutations alongside aggressive
19 proliferation [3-6]. And it is through acquired mutations that cancers acquire their
20 characteristic hallmarks [1].

21 Different cancers may be characterized via a unique set of mutations and/or
22 expressional aberrations. For example, the BCR-ABL fusion protein arises due to a
23 translocation event unique in CML [7] and drives uncontrolled cellular proliferation
24 which allows it to serve as an (albeit imperfect) diagnostic marker. Understanding and
25 characterizing the unique set of mutations and/or expressional aberrations that
26 constitute a particular cancer improves our understanding and may lead to crucial
27 therapies and diagnostic markers [8].

28 The current paradigm for achieving the above takes advantage of modern
29 high-throughput analysis tools such as DNA arrays, RNA sequencing and proteomics
30 to assay the unique expressional profile of each sample. However, these studies usually
31 consider each patient sample only once as a “homogeneous representation” of that
32 individual’s tumor. The true complexity that underlies each patient’s tumor profile
33 becomes averaged, and some information is lost --- i.e., we do not consider the
34 intra-tumor heterogeneity (ITH) for each patient. Given the background complexity for
35 any given patient’s cancer, it is possible that different cancer data when merged
36 together would yield significant proteins or genes that would on the whole be
37 meaningful but may not lead to specific therapeutic options for the individual patient.

38 On the whole, the different cancer cells within a cancer tissue from a single
39 patient, or between patients, would still have a lot in common due to their clonal origins
40 [9]. Significantly altered proteins/genes could still be derived and be meaningful (in the
41 sense that they lie on the critical path leading towards the development of the cancer

phenotype), even though these proteins/genes may not explain the heterogeneity. The significant common genes/proteins may well be a common starting point for turning on uncontrolled growth for all the cancer cells in the tissue and is thus the primary treatment target.

However, the differences of the cancer cells in the sample could be responsible for differences in resistance to treatment or could be different root causes that converge to a common point [10]. Of course, it could also be that different cancer cells in the same tissue do have completely different paths (from start to finish) of becoming cancerous, but this is rather unlikely since all cancer cells tend to exhibit common hallmarks [1, 2]. Properly understanding root causes and drug-resistance potential is crucial to developing better treatments [11]. So, assuming the heterogeneity of the cancer cells in a tissue could be linked to these two important considerations, it is therefore important to also investigate ITH, besides simply identifying significant proteins/genes common to a cancer.

High-throughput technologies such as RNA-Seq, and its counterpart, exome sequencing, can be applied towards ITH investigation. In a study by Zhang *et al*, multi-region whole-exome sequencing was performed on 11 localized lung adenocarcinomas [12]. They reported extensive ITH; and that the presence of a particular subclonal population may be correlated with relapse. In another study by de Bruin *et al*, the authors hypothesized that a lack of specific understanding of ITH in non-small-cell lung carcinoma (NSCLC) may account for its dismal prognosis [13].

Investigation of 25 spatially-distinct regions from seven operable NSCLCs revealed subclonal diversification, confirming that the etiologies for different samples were profoundly and significantly different. These tumors cannot be naively considered or studied as a single homogeneous class [13].

Many studies on ITH, including the aforementioned, are based on genomic information. And in accordance with the considerations we put forth, they lend justification for an in-depth ITH investigation: Firstly, that ITH is linked with drug resistance; and secondly, diversification of the branches in cancer evolution converges towards the common cancer hallmarks.

While these genomics-based studies do contribute significantly to our understanding of ITH, it is imperative that we also consider the corresponding protein-expression profiles as a necessary complement. Unfortunately, the high-throughput study of proteins (proteomics) already suffers from coverage, consistency and reproducibility issues [14]. Recent advances in next-generation proteomic technologies make it possible to assay large sample sizes with higher consistency, reproducibility and coverage, given limited clinical samples. In particular, Guo *et al* [15] recently developed a novel pipeline, PCT-SWATH, towards achieving this objective. Briefly, PCT is a means of using pressure cycling to maximize protein extraction and minimize technical variation given a limited quantity of clinical sample. SWATH is a Data-Independent Acquisition (DIA) method that collects mass

spectra information by repeatedly scanning through a given m/z range in defined windows (termed a swath).

As a current technique, SWATH is purportedly able to overcome scalability issues with regards to sample size, with improved reproducibility and measurement reliability. It also produces good coverage, but this is currently dependent on the reference transition library (identifications are made based on matches between the SWATH profiles and known spectral patterns --- a pattern of a known peptide from a given protein). SWATH does not give full information on the proteome as yet.

Since there are analytical limitations, a network-based method, qPSP (quantitative Proteomics Signature Profiling) was developed for studying SWATH proteomic profiles [16]. Networks are useful for summarizing and contextualizing biological data [14]. The intuition is that if a group of proteins within a network cluster are co-ordinately over-expressed, it is far more likely to be true positive than a single protein. Similarly, if a group of proteins in a protein complex are co-ordinately over-expressed with the exception of one, it is far more likely that the exception is a false negative or a significant biological outlier [17]. qPSP uses this principle and has been shown to be robust against noise (and missing data), as well as outperforming conventional ways of studying sub-networks and complexes [16]. It may be applied towards studying noisy and complex scenarios, such as ITH.

In recent years, network-based methods have advanced and now cover five paradigms [18]. Of these, a similar but more advanced approach to qPSP, PFSNet

(Paired Fuzzy Scoring SubNetworks) was developed, utilizing a reciprocal class-weighting procedure prior to statistical filtering [19, 20].

As a standard clinical/analytical practice, the pooling of cancer cells in a tissue altogether and the focus (in the past) on significant genes common across the whole tissue (or a whole pool of patients) have certainly caused information loss or, at least, caused the underlying heterogeneity of cancer cells to be ignored. Yet as we know, ITH is implicated in tumor resistance and in differential pathway convergence towards a given set of cancer hallmarks. Moreover, in a clinical setting, ITH has true practical applications --- given the high cost of obtaining sufficient samplings or even the act of obtaining the clinical samples itself, these costs may become justifiable if practical therapy-guiding information can be obtained. The problem however, is that ITH is neither straightforward nor easy to study systematically. Since mutations are stochastic, it is expected that many of the differences between individual samplings from the same tumor are simply noise.

This paper contributes towards ITH proteomics studies in two ways: First, we take a network-based approach in a pilot proteomics study of heterogeneity in a prostate cancer data set (based on multiple tissue punches from normal and cancer tissues, in a small cohort of 3 patients). The second goal looks at stability issues arising from noise and confounding signal contributions from different tissue regions/punches, and how network-based techniques are able to produce more robust insights relative to protein-based analyses.

2. Materials and Methods

2.1 Prostate Cancer Data

There are a total of 60 samples derived from 3 patients: 12 (6 normal, 6 acinar) from patients 1 and 2, and 36 (12 normal, 12 acinar, 12 ductal) from patient 3. Patients 1 and 2 are designated P1 and P2 respectively. For patient 3, analysis based on acinar or ductal tumor tissues are designated as P3A and P3D respectively (see [Supplementary Table 1](#)) [21]. A disclaimer: To be more clinically precise, the ‘normal’ tissue taken from patients with existing cancers should be more appropriately termed ‘benign’. While these tissues do not display the outward characteristics of a malignancy, it does not mean they are truly normal tissue. We use the term ‘normal’ to describe these tissues, mostly as a semantic contrast against outwardly obvious ‘cancer’ tissue. The data is available for download at PRIDE database (Project accession: PXD003497). [The conditions reproduced from Guo et al. are described in supplementary methods \[21\].](#)

2.2 CORUM

Protein complexes may be considered as a generalization of network clusters. Protein-complex prediction from biological networks is an active area of research [22, 23]. As functional units, protein complexes are enriched for biological coherence, outperforming most other forms of biological meta-data [24]. In earlier work, we demonstrated that using real complexes outperforms predicted complexes for the

purpose of proteomic-profile analysis, and so for comprehensiveness, we use only real complexes in this analysis [25].

The set of protein complexes are derived from CORUM, a database containing manually-annotated protein complexes from mammalian organisms [26]. Of approximately 1200 human complexes, 380 (minimum size 5) are retained. The minimum size requirement is introduced to minimize large fluctuations on effect size calculations [16], cf. the following sections on qPSP and PFSNet.

2.3 GFS transformation procedure

Gene fuzzy scoring (GFS) is implemented in the NetProt package [16, 27] (Supplementary Figure 1).

A GFS transformation can be applied to the protein expression matrix, where the columns are individual samples, and the rows are identified proteins. Each cell in the matrix contains the protein expression data (see Supplementary Figure 1A) [28].

In GFS, each sample (column) is sorted in inverse order based on its quantitation. A cut-off level (α) is then applied for each sample.

To improve sensitivity, we deployed a strategy using two α levels [20] --- The top 10% proteins are defined as α_1 . While a single α cut-off is quite precise, it is also lacking in sensitivity. Introducing a second α level (α_2 – top 10%-20%) mitigates this but runs the risk of incorporating noise due to less informative rank shifts amongst the proteins. To reduce noise due to incorporation of α_2 , weighting is performed. In Lim *et al* [20], linear interpolation is used to produce a

series of continuous values from 1 to 0 for all genes between α_1 and 2. This generates a list of many possible continuous values. To simplify matters, the continuous range from α_1 to 2 was discretized into four bins, 10-12.5% (weight 0.8), 12.5% to 15% (weight 0.6), 15-17.5% (0.4), 17.5 to 20% (0.2). All other proteins below α_2 have a weight factor of 0, which therefore means they are excluded. α_1 proteins have weight factor of 1. (Later, in the description of PFSNet, this weight factor of a protein p in a sample s is denoted $fs(p,s)$, and is also called the fuzzy value of p in s .) The proteins that pass the α cut-off are collectively referred to as α proteins.

2.4 qPSP procedure

qPSP is implemented in the NetProt package [16, 27]. The procedure is diagrammatized in **Supplementary Figure 1B**:

Following obtaining the α proteins from samples, qPSP is performed as follows: For each sample, a vector of hit-rates was calculated by examining the overlap of α proteins in that sample against a vector of sub-networks (e.g. real complexes, predicted complexes and pathway subnets). The vector of hit-rates is also referred to as the *quantitative Proteomics Signature Profile* (a qPSP) for the sample itself.

For the purpose of clustering, the distance between the qPSP profiles of two samples (u, v) is defined as their Euclidean distance, $E_{u,v}$, across a given set of complexes C_i , where $i \in \{1..x\}$. Specifically, let $C_{i,u}$ and $C_{i,v}$ correspond to the hit-rates for samples u and v respectively for complex C_i , then

$$E_{u,v} = \sqrt{(C_{1,u} - C_{1,v})^2 + (C_{2,u} - C_{2,v})^2 \dots + (C_{x,u} - C_{x,v})^2}$$

The qPSPs are then clustered using Hierarchical Clustering (Ward's). Bootstrap using PVClust was performed to evaluate the stability of the clustering [29]. For each cluster induced by the hierarchical clustering, scoring values (between 0 and 100) are calculated via multiscale bootstrap resampling. PVClust provides two types of score values: AU (Approximately Unbiased) and BP (Bootstrap Probability). AU, which is computed by multiscale bootstrap resampling, is reported as a better approximation to unbiased p-value than BP value computed by normal bootstrap resampling but PVClust reports both anyway [29]. Clusters with significant AU scores (i.e., above 95) are encased in a red perimeter to highlight strong grouping.

Identification of significant complexes was performed using the standard t-test between the control and test state in terms of hit-rates [30]. The t-score is then evaluated for its corresponding p-value using the theoretical t-distribution given n-1 (where n is sample size) degrees of freedom. The t-distribution is similar to the normal or z-distribution, and approximates it when n is high. It is suitable for evaluating smaller-sample-size situations, when $n \ll 30$, and the conditions of the central limit theorem are not met. In the smaller n situation, the t-distribution has **fatter** tail areas, and smaller central area compared to the z-distribution, making it harder to reject the null hypothesis that the complexes are not significantly differentially expressed between normal and tumor.

2.5 PFSNet procedure

PFSNet is implemented in the NetProt package [19, 27]. The procedure is diagrammatized in Supplementary Figure 1C and is as follows:

As with qPSP, GFS transformation is first applied to the protein matrix (Supplementary Figure 1A).

Given a sample class X , we let the average fuzzy value of a protein p_i over each sample s_j in X be a class-proportion factor of the protein, $\beta(p_i, X)$.

$$\beta(p_i, X) = \sum_{s_j \in X} \frac{fs(p_i, s_j)}{|X|}$$

A protein complex C is scored for a sample s_j in class X by summing the fuzzy value of each constitutional protein in C for sample s_j , weighted by the proteins' class-proportion factors in X . For class Y , similarly, we can get a complex score based on Y .

$$\text{Score}(C, s_j, X) = \sum_{p_i \in C} fs(p_i, s_j) \times \beta(p_i, X)$$

$$\text{Score}(C, s_j, Y) = \sum_{p_i \in C} fs(p_i, s_j) \times \beta(p_i, Y)$$

PFSNet defines a score $\text{delta}(C, s_j, X, Y)$ for a complex and sample with respect to classes X and Y . Specifically,

$$\text{Delta}(C, s_j, X, Y) = \text{Score}(C, s_j, X) - \text{Score}(C, s_j, Y)$$

Statistical feature selection is achieved via one-sample one-sided t-tests with the null hypothesis declaring that complex C is not relevant to the difference between

classes X and Y ; the t-statistic on a set of samples Z , $f_{PFSNet}(C, X, Y, Z)$, is calculated as the mean of the list $\{\delta(C, s_j, X, Y) \mid s_j \text{ is a sample in } Z\}$ divided by its standard error:

$$f_{PFSNet}(C, X, Y, Z) = \frac{\text{mean}(C, X, Y, Z)}{\text{sd}(C, X, Y, Z)}$$

2.6 Analysis of variance across intra-patient and inter-patient samples

Assuming limited technical noise, variation within a patient's sample is expected to be lower than between patients. Similarly, the variations for normal samples are not expected to be similar to those for cancer samples, or between normal and cancer samples.

For every type of pairing of punches, the respective distribution of the squared differences for the hit-rate (qPSP) and delta score (PFSNet) of every complex is examined. For intra-sample comparison (Int, or internal), the pair of punches must be from the same sample. For inter-sample comparison (Ext, or external), the pair of punches must be from different samples but from the same class. Mix comparisons consider pairs from normal and cancer. An internal mixed comparison takes mixed pairs from the same patient. An external mixed comparison considers mixed pairs from different patients (the normal and cancer component must come from different patients).

The distributions of squared differences are represented as side-by-side boxplots, where the median and confidence intervals are marked and shown.

2.7 Network-based approach versus single-protein approach's robustness and stability using leave-one-patient-out cross-validation

We propose that network-based approaches such as qPSP and PFSNet are more robust and tolerant to noise/heterogeneity than the use of single proteins. We investigate this using leave-one-patient-out cross validation (**Supplementary Figure 2**). Here, all data of one patient (both normal and tumour punches) are set aside as the test set. A naïve Bayes classifier is then trained on significant qPSP or PFSNet complexes or proteins selected on the basis of the data of the two remaining patients (the training set). This procedure is iterated on each patient, each time using that patient's data as the test set and the other patients' data as the training set. The complexes and proteins were selected fresh in each iteration based on the same significant criteria (t-test with a p-value cut-off at 0.05). Because the selection was performed on a different training set of patients each time, the set of significant proteins or complexes might differ across each iteration.

The average accuracy of the cross-validation above based on correct assignment of classes (normal and cancer) was used as the assessment criteria.

3. Results

3.1 Differences exists across all patients but broadly, the normal and cancer classes are distinguishable

At the protein level, if we looked at the samples individually (Figure 1A rows 1 and 2) or as a whole (Figure 1A row 3), we would find generally strong differences between normal and cancer tissue such that even without statistical feature selection, the samples would naturally separate (using an unsupervised hierarchical clustering approach). A similar stratification was also achieved by qPSP (Figure 1B) and PFSNet (Figure 1C) respectively.

The three patients did in themselves exhibit interesting differences: P1 and P2 present very strong differential segregation where the AU scores (indicating the stability of the cluster) were well above 95. However, this is less true for P3, where the normal and cancer classes were less well resolved (although the samples do broadly split into normal and cancer branches), possibly because P3 harbors a more heterogeneous tumor.

Moreover, when we combine the data from all three patients together, although the samples do still separate broadly into normal and cancer clusters, the AU scores drop appreciably. The red boxes (Figure 1 row 3), indicating stable and confidently identified clusters (for normal and cancer phenotype classes) no longer exist. One possible explanation is because the different patients are inherently heterogeneous, such that their protein or network-based scoring profiles are highly variable (suggesting

heterogeneity). It is also noteworthy that the perceived heterogeneity is greater for proteins (Figure 1A) and qPSP (Figure 1B), but less so for PFSNet (Figure 1C).

At the patient level, applying a p-value cutoff of 0.05 allowed identification of differential proteins and complexes. The intersections are analyzed using Venn diagrams in Figure 2A/B/C [31]. Indeed, for proteins and qPSP, the intersection of significant proteins/complexes between patients is unsurprisingly low (Figure 2A/B).

Amongst differential proteins, only 2 are common to all patients, an abysmally low number, which also highlights how different each patient's tissues appear to be (Figure 2A). In qPSP, only 6 complexes were common to all patients examined, but the patients now appeared more similar (Figure 2B).

Closer examination of the Venn diagrams in Figure 2A/B reveals some insights on patient cancer similarity. Using Figure 2B (qPSP) to explain, P1 and P3 have considerable overlap ($6+3+21+23 = 53$). P1 and P2 have little in common ($6+1=7$). P2 and P3 also have relatively little in common ($6+8+10+1=25$). P3 has two different tumor types that we considered separately in order to discern their differences as well. Expectedly, the overlaps between P3A and P3D are deep since they come from the same individual ($6+21+10+20=57$). What is particularly interesting is that P3A has a substantial unique component of 21 complexes whereas P3D only has 7 (which is similar in size to the unique components of P1 and P2 of 7 and 5 respectively).

Even though heterogeneity exists amongst the different patients, the significant complexes in each patient should be sufficiently informative to confidently cluster the

other patients since after all, they have the same disease phenotype. To prove this point, we performed reciprocal qPSP clustering using complexes from each patient --- i.e., we used the significant complexes found in one patient for clustering all the other two patients iteratively (Supplementary [Figure 3](#)). The results fall in line with expectation: Generally, complexes from each patient can split all others well. P2, however, exhibits some weakness in this regard, which is not surprising given that it has less in common with the other patients.

PFSNet presents a rather different viewpoint on heterogeneity. While protein-level and qPSP analyses retain apparent patient-specific idiosyncracies, PFSNet obviates these differences. In contrast to protein-level analysis (where 46% of the differential proteins are differential in exactly one patient) and qPSP (where 30% of the differential protein complexes are differential in exactly one patient), PFSNet reports 7 protein complexes as differential in exactly one patient amongst 136 protein complexes differential in at least one patient (i.e., 5%). Amongst 133 qPSP and 136 PFSNet significant complexes, 96 are common. There is no obvious enrichment if we considered complexes that are shared by at least 2 patients: Out of 93 qPSP and 129 PFSNet complexes, 69 are common. Since PFSNet and qPSP seem to provide different perspectives on heterogeneity, we checked the following (with reference to [Figure 2A](#)):

In qPSP, there are 40 “single-patient” complexes. We found that 20 of these 40 complexes (i.e., 50%) are reported to be significant by PFSNet in at least two patients. This indicates that the high heterogeneity suggested by qPSP may be an artefact (the

samples are really more similar than they appear to be). Also, we note a caveat that some qPSP complexes significant in exactly 1 patient might be noise [18]; obviously, these have a higher likelihood to be noise than those that are significant in multiple patients.

While it may appear that the 83 complexes in the perfect intersection of PFSNet (Figure 2C) is only drawing adjoining peripherally shared complexes, it is actually not entirely so: 14 out of 20+10 complexes shared between P3A/P3D (Figure 2B) moved to the 83. And out of 93 qPSP-significant complexes shared by at least 2 patients, 45 of these shifted towards the PFSNet center (of 83 complexes) ($45/93 = 0.480$). In contrast, 15 out of 40 ($=0.375$) complexes unique to each patient shifted. There is thus a 1.28x enrichment ($0.48/0.375$) for shared complexes to be centrally shifted according to PFSNet. Moreover, all 7 PFSNet single-patient complexes are reported by qPSP as significant in at least two patients; thus, even PFSNet single-patient complexes are likely capturing common characteristics shared amongst patient samples. Together, it seems PFSNet more likely captures common characteristics shared amongst patient samples rather than those which are truly differentiating at the individual level.

For proteins, the mutually supporting information captured is sparse. For example, out of 104 unique P1 proteins (Figure 2A), only one of these are found to be involved in the 7 unique complexes significant for qPSP (89 proteins) and none for PFSNet (7 proteins). This suggests protein-level feature selection has higher noise and poorer sensitivity than complex-level feature selection.

The phenomenon that 37.5% of “single-patient” qPSP complexes and 48% of “double-/triple-patient” qPSP complexes shifted to the PFSNet center can be explained by PFSNet’s higher sensitivity (as PFSNet is able to report these complexes as significant in more patients). The increased sensitivity of PFSNet is likely attributable to its reciprocal class-specific weighting (and, less likely, its one-sided statistical test procedures) [18]. To understand the situation from a functional perspective, we looked at the six core qPSP complexes (also predicted by PFSNet as significant in every patient) (Figure 2D) that affect DNA replication, protein synthesis and signaling. While these may appear to be core cancer-related functions, we should also add that unique patient-specific complexes --- i.e., those significant qPSP complexes that are found only in individuals and are not found amongst PFSNet complexes shared amongst at least 2 patients. (Supplementary Table 2), are no less important, underpinning the widespread upheavals typically associated with cancer.

What is certain is that patient-specific analysis based on multiple tissue punches can reveal much about the unique etiology of one’s cancer. And that each person’s cancer has a different fingerprint that may be useful for personalized diagnosis/prognosis. On the other hand, if we just want to know whether a person has a cancer or not, the broad-scale differences between normal and cancer tissue appear to be strong enough for easy detection. However, this knowledge (and its caveats) is already known to some degree in earlier work [3-5].

While protein expression may be used for diagnosis, network-based scoring based on qPSP hit-rates and PFSNet deltas are useful data summarizations with clearly distinguishable effect sizes (see Figure 2E and Supplementary Figures 4 and 5). Moreover, protein expression can be messy and difficult to interpret, especially when trying to establish differences amongst different individuals (Supplementary Figure 6); network-based scores provide a more elegant summarization (Supplementary Figures 4 and 5).

Moreover, hierarchical clustering based on the GFS-normalized proteins (grouped by complexes) confirms that the use of protein information and normalization alone is not enough, and that the heterogeneity at the protein level converges at the protein-complex level, i.e., there is heterogeneity within shared complexes, beyond “single-patient” complexes (Supplementary Figure 7).

3.2 Leave-one-patient-out cross-validation reveals network-based approaches are superior to protein-based feature selection

In the previous section, despite a lack of concordance, we demonstrated that features predicted in one patient, can be used in another and achieve a fair degree of phenotype class separation (Supplementary Figure 3).

Although protein-based feature selection can achieve class separation (Figure 1A), the signal from individual proteins are thought to be noisy and less confident than considering complexes. This view is supported somewhat by considering the intersection of significant proteins relative to significant complexes (Figure 2A/B/C).

Although there are many more reported proteins than complexes (3000 against 380), the intersection between differential proteins is much lower (2) than those amongst differential complexes.

Achieving general class segregation may not be enough for diagnosis. Since diagnosis is essentially an exercise in prediction, we evaluate the ability to predict class labels using a supervised leave-one-out cross-validation approach by training a naïve Bayes learner iteratively (See Materials and Methods/[Supplementary Figure 2](#)).

Both protein and network-based features are evaluated by the same standard t-test given two different significance levels ([Table 1](#)). In both scenarios involving a stringent ($P < 0.01$) and less stringent p-value ($P < 0.05$) cutoff, the network-based methods by far outperform protein-based features. Amongst network-based approaches, PFSNet unsurprisingly, outperforms qPSP and can make accurate predictions on the particularly difficult/elusive P2. Earlier, we already know that there are few overlaps on proteins and qPSP complexes between P1/P3 with P2 (whereas PFSNet is able to identify a much larger pool of “common” features amongst all 3 patients). These results are consistent with expectation.

But what if we just use features shared by all patients? This is examined by the $P < 0.05$ significant intersection scenario ([Table 1](#)). For proteins, qPSP and PFSNet, this is 2 proteins, 6 qPSP complexes and 83 PFSNet complexes respectively. All approaches did better in this scenario, underpinning the importance of meta-analysis [3-5]. This result also clearly demonstrates that network-based approaches, particularly PFSNet, is

robust against apparent heterogeneity, and can select powerful features that can make accurate predictions on new samples.

Although PFSNet seems to perform well, its performance is based on 83 complexes, as opposed to only 6 in qPSP. What is interesting is that if we used the 6 qPSP complexes with PFSNet's delta scores, we obtain similar performance as with qPSP's hit-rates. While this may suggest that the different effect-size calculations may not matter, what does matter is that network-based contextualization has emergent effects not found at the level of single proteins: When we isolated the proteins found in these 6 complexes and used their expression values for training, prediction accuracy falls drastically (Table 1, rightmost column). To take into account that the good performance of qPSP/PFSNet is not solely due to GFS and that there are benefits for complex-level scoring (hit-rates and deltas), we checked the prediction accuracy based on the GFS-normalized data matrix; this demonstrated that while GFS normalization on the protein matrix greatly boosts prediction accuracy, it is inferior to both qPSP hit-rates and PFSNet deltas, confirming that network-based contextualization adds emergent information that is otherwise non-inferable from the individual proteins (Supplementary Table 3).

To look further into heterogeneity, we looked at qPSP predictions since it suggests that there is greater ITH present (and that there are more features to work with). We find that in this comparison, more mistakes were made in predicting 3D than 3A. That is, when using P1/P2 to predict P3D and P3A separately, the classifier makes more

428 mistakes on P3D (0.75 accuracy) than P3A (0.83 accuracy). This is not surprising as
429 ductal cancers represent a different prostate-cancer region, and therefore should have a
430 somewhat different profile. To further demonstrate the robustness of the predictions
431 based on this analysis technique, we trained a Naïve Bayes model based on all patient
432 samples using the hit-rates from the six intersecting complexes and validated
433 independently in another set of three prostate-cancer patients (3 normals, 3 cancers; 2
434 technical replicates for each set) [32]. The accuracy in this case was 83.3%, which
435 means that out of 12, two misclassifications were made. Interestingly, both
436 misclassifications were from the normal class instead of the cancer class. Moreover,
437 these two misclassified samples are actually technical replicates from the same patient.
438 Using the constituent proteins within these significant complexes instead for training
439 the classifier, the accuracy falls to 67.7% (four misclassifications were made).
440 Clearly, dealing with heterogeneity appropriately through the use of powerful network
441 models can lead to identification of informative features that can be used to build robust
442 and accurate models. For comparison, the model built using the two shared proteins had
443 an accuracy of only 0.5 on the external data set, which means the significant common
444 protein features have no reproducible predictive power.

445 So far, we have shown that normal and cancerous tissues are broadly different from
446 each other, such that differential features found in one patient can work on another. We
447 also note that while network-based approaches are generally better at prediction than

the use of protein (expression), the network-based methods, qPSP and PFSNet paint different perspectives on ITH.

To gain a deeper perspective into how these network-based measures (qPSP hit-rates and PFSNet deltas) are informative when we compile information from different tumor sites (within the same patient and between patients; and within one tissue type and across normal versus cancer tissues), we investigated the changes in the variability (distribution of the squared differences) of qPSP hit-rates and PFSNet deltas from taking random pairs from the following scenarios --- C_int and N_int refers to cancer pairs and normal pairs from within the same patient. C_ext and N_ext refers to pairs taken from cancers and normal respectively from different patients. CN_int and CN_ext refers to mix pairs (cancer and normal) but from within same patient and different patients respectively.

Suppose we are in the biomarker discovery phase, where the markers are not known (or we do not know what to look for), and therefore all features (protein or complexes) have to be considered. We observed that the differences in variability are low between the different test scenarios (Figure 3A/B top column). The similarity of the distributions over all features across C_int/ C_ext/ N_int/ N_ext/ CN_int/ CN_ext, is because most features are irrelevant and thus they dominate these distributions and render these distributions similar across C_int/ C_ext/ N_int/ N_ext/ CN_int/ CN_ext. Fortunately, for distinguishing tumor and normal (i.e. biomarker discovery), it is standard practice look for useful/ informative features at the extreme end of the CN_int

or CN_ext distributions; cf. qPSP and PFSNet. So while the distributions over all features look similar across C_int/ C_ext/ N_int/ N_ext/ CN_int/ CN_ext, this similarity does not impede biomarker discovery.

In contrast, when the biomarkers are already known (diagnosis phase), taking one representative sample per class to make a pairwise comparison is very discriminative (strongly differential) (Figure 3A/B bottom column). The 6 qPSP significant complexes from the intersection of all patients could be considered viable biomarkers since they are conserved features. Reconsidering the same paired scenarios as before, the differences between internal and external pairs for normal and cancer become more discernible (Figure 3A). There is increased range of variability for both internal and external scenarios. The mixed scenarios (CN_int and CN_ext) in particular, exhibited the greatest change in variability. For the internal mix scenario (CN_int), the increased variability is expected because the pairs originate from different classes. The use of the 6 conserved features markedly shows this increased variability. For the external mix scenario (CN_ext), the variability is even greater, due to inter-patient heterogeneity. Because CN_int and CN_ext pairs demonstrate higher changes in variation, this suggests that for diagnosis purpose (where the features are known), one punch per sample class is sufficient.

Repeating the same analysis, but using PFSNet deltas (Figure 3B), a similar scenario is observed. Interestingly, PFSNet shows that cancer tissues are more varied in general (C_int, C_ext, CN_int and CN_ext), and suppresses variation within normal

tissues (N_{int} and N_{ext}). This supports our observation that PFSNet amplifies inter-class differences.

PFSNet tends to pick many more complexes than qPSP that end up within the intersection. Furthermore, we find that using qPSP-selected complexes but swapping in PFSNet deltas also gives good prediction performance (Table 1). To see how this impacts variation analysis, we take only qPSP-selected complexes but swapped out the qPSP hit-rate in favor of the PFSNet delta (Figure 3C). This results in preservation of the heterogeneity between same/different-patient tumor sites while also demonstrating that non-cancer tissue is more homogeneous. However, we also note some rather unusual observations, particularly in that cross comparison of normal to cancer tissue both in the same patient, and between patients, do not result in markedly higher variation than internal comparisons of cancer tissue (within and between patients) (Figure 3C). To verify that this is not an artefact, and to dissect deeper, we looked at the distribution of proteins expressions, qPSP hit-rates and PFSNet deltas, with concomitant variation at the level of the individual (Supplementary Figure 8A/B) which demonstrated that internal comparisons of cancer tissue within the same patient tended to generate not only higher variation, but also of very different magnitudes between individuals (Supplementary Figure 8A/B, third column and 5th column respectively). This accounts for cross comparisons of cancer tissue between individuals resulting in widely diverging variation (Figure 3C). Comparing the actual network scores between normal and cancer tissue within the same patient (P3D), and between

511 P3D and P2, we can also see that merging them would result only in increase of
512 divergence of variation, but not necessarily higher variation (Supplementary Figure
513 8C).

514 4. Discussions

515 4.1 Shared features are rare according to qPSP but informative according to
516 PFSNet: Which is correct?

517 The reciprocal clustering (Supplementary Figure 3) alongside the intersections
518 amongst common differential features (Figure 2A/B/C) reveals that a small number of
519 shared differential features (genes that may have potential involvement in cancer
520 causation) is sufficient to separate the patient samples into cancer and benign (i.e.,
521 seemingly normal). We also observe that most features are not shared according to
522 qPSP (only 6 complexes out of 133 are significant in every patient), but not according
523 to PFSNet (83 complexes out of 136 are significant in every patient). Moreover, half of
524 qPSP non-shared features are actually shared according to PFSNet.

525 So how to resolve this? Given our earlier dissection, we suspect that
526 qPSP/protein-based analysis are not sensitive enough to detect in each patient the other
527 77 (83 – 6) shared complexes of PFSNet. It is possible that the 6 qPSP complexes are
528 the most obvious or significant ones among the 83 PFSNet complexes. And they may
529 indeed be the convergence points (along with the other 77 complexes) since 1/ all
530 patients manifest the same disease, and 2/ cancer hallmarks are conserved, which also

implies that ultimately, the network regions that are affected would also be similar and conserved. However, given our analysis, we can only claim that the 6 complexes are the most differentially expressed but cannot claim that the six complexes necessarily correspond to triggers (causal factors) and are not merely correlated to the cancer phenotype [3, 4]. The small intersection amongst all patients (for proteins and qPSP complexes as seen in [Figure 2A/B](#)) is an artefact. Our analysis showed that half of qPSP “single-patient” complexes to be shared amongst patients as reported by PFSNet.

PFSNet presents a rather different viewpoint on heterogeneity. As PFSNet and qPSP both share the GFS normalization procedure upstream, its differential performance stems possibly from the class-weighting step (which amplifies class-specific differences while ensuring each class-specific feature is shared by many members of its class). We do not think the single-sided statistical testing procedure in PFSNet is an explanatory factor for the increased sensitivity PFSNet appears to exhibit. That the single-sided statistical testing procedure boosts sensitivity of PFSNet is only a valid claim if most PFSNet-significant complexes have p-value between 0.025 and 0.05 (these are complexes that would become insignificant when a two-tail test was used instead). In previous work [18] and as also observed here, PFSNet-significant complexes usually have p-values way below 0.025 (and thus, rendering the argument that one-sided tests are more sensitive a moot point anyway).

Therefore, we trust that the large intersection detected amongst patients by PFSNet is more likely to be the correct picture ([Figure 2C](#)): 60 out of 83 complexes in

PFSNet center is qPSP-significant in at least 1 patient. Moreover, since PFSNet boosts inter-class effects, it is useful for diagnosis purpose.

But what about the heterogeneity that exist in each individual? We know earlier that some half of qPSP single-patient (and thus presumably heterogeneity-associated) complexes are PFSNet significant in more than one patient (i.e. half of the presumed heterogeneity-associated complexes due to qPSP are artefacts given qPSP's poorer sensitivity). Moreover, we are also aware that qPSP is more susceptible to noise relative to other subsequently developed methods like PFSNet [18]. Nonetheless, we do not rule out that some qPSP-only single-patient complexes truly reflect heterogeneity.

As for protein expression, these are even more likely dominated by noise (only 2 common proteins out of three thousand significant proteins seems unfathomable). Although protein expression-based perspective is important, given these results, it seems difficult to offer advice to say "the use of protein-based expression is also acceptable". Instead, we advocate taking a conservative and cautious approach with the use of absolute protein expression data. If not, at least, use a good normalization method such as GFS [28].

Regardless, the use of network-based scores, be it qPSP or PFSNet, provides discriminative signal (allows us to tell apart samples) if we look at complexes at the extreme ends (i.e., significant p-value) (Figure 3). In this respect, there is no big difference between qPSP and PFSNet (the 6 qPSP complexes are also found in PFSNet

and thus are picked out automatically when we look at complexes at the extreme end of the PFSNet p-value distribution).

Although network-based analysis can afford us better statistical reproducibility and deeper insight, they do not all work the same way and should not just be applied blindly. There are useful insights to be derived from a strong mechanistic understanding of how these network-based methods work, and when they do not agree, to systematically reason and dissect which one is more likely correct instead of simply taking the intersection [33].

4.2 Multiple punches are necessary for biomarker discovery but only one is needed for diagnosis

A critical question that is often asked is whether a single punch (or tissue sample from one site) contains enough information that can summarize a patient's entire tumor. An intuitive answer would be no, given that a tumor contains subpopulations with different mutations. But is this really true?

So far, we have used all punches of a patient to build models, and show the models give good predictions. We also know that when we pair punches on known features, there is appreciable change in variation. However, we have not yet explicitly shown that using just one cancer and one normal punch from some patients allows us to build accurate prediction models. To test this briefly, we repeated the leave-one-out cross-validation using 1, 2 and 3 tissue punches (we trained the classifier on a randomly selected set of tissue punches in the training set, but only validated it on 1 punch of a

randomly selected patient sample in the test set; both technical replicates per tissue sample are maintained for this analysis) (Supplementary Table 4). The analysis method matters: For qPSP and PFSNet, 1 punch is essentially sufficient for most scenarios while protein-based analysis appears to slowly gain in power with punch increment.

We also need to be wary of technical variation in SWATH: Taking either the first or second replicate can give rise to markedly different prediction accuracies (Supplementary Table 5). We suspect the presence of batch effects: For PFSNet and qPSP, it does not really matter which of either technical replicates are used (the accuracies are in general high except for P2). However, when looking at either technical replicates for proteins, accuracies are relatively higher when the data is separated by technical replicates, but when combined, accuracies drop dramatically. We know from previous experiments that network-based methods are resistant to batch effects (of which technical variation is also a contributing factor) whereas protein-based features are not [34-36]. It is interesting to see how the pervasive batch effect rears its head in unexpected situations.

Our results suggest that trying to identify reliable markers requires capturing as much variability as possible within patients, which in turn, leads to better identification of intra-patient-specific complexes (Figure 3). The nature of the method for calculating effect sizes and performing statistical feature selection also matters: qPSP suggests that only a few complexes are reliable, and are common to all samples. The use of these 6 complexes for predictive modelling leads to good models and may be explained off as

614 convergence events (towards manifestation of the cancer hallmarks). And so, these
615 complexes identified as markers need only be used on a single punch in future tests for
616 diagnosis. However, we also need to consider PFSNet, where more than half of the
617 predicted complexes are in the intersection, and appreciably greater than qPSP or single
618 proteins. The use of these PFSNet complexes do not lead to worse models. Perhaps, for
619 methods such as the protein-based t-test or qPSP, instead of detecting heterogeneity,
620 what is being picked up could be simply a lack of statistical power, as is often the case
621 with small sample sizes and high variability.

622 4.3 Complex-based is superior to protein-based approaches as a hedge 623 against heterogeneity

624 Single-protein measurements are known to be less reliable than group-based
625 measures such as complexes. Group-based measures have an improved signal-to-noise
626 ratio.

627 It is therefore expected that network-based methods such as qPSP and PFSNet
628 would outperform proteins. There are three major advantages to using network-based
629 approaches: Firstly, they build better, more accurate models from previous data. This is
630 important as biomarkers need to be reliable and consistent. Secondly, they are able to
631 identify shared features amongst patients better. Note that only 2 proteins were found to
632 be shared amongst over 3000 while 6 and 83 out of 380 complexes were shared (for
633 qPSP and PFSNet respectively). It is also noteworthy that the 6 qPSP shared complexes
634 are highly informative, raising the prediction accuracy to high levels regardless of

whether qPSP hit-rates or PFSNet deltas were used. Thirdly, coverage is increased indirectly --- not all proteins are identified in SWATH, the predicted complexes offers possibilities for deep mining and validation of the undetected proteins for more marker/target candidates.

Unfortunately, the creators of the prostate cancer data did not generate a corresponding RNA-seq dataset. It would be interesting to establish if similar results are obtainable at the gene expression level as well.

4.4 Implications for biomarker analysis

While the t-test and qPSP suggest the different punches are heterogeneous, this heterogeneity disappears when the data is viewed at a higher perspective (PFSNET complexes). A plausible explanation is that, the protein-/gene-level heterogeneity leads to a common end point reflected by the homogeneity of PFSNET scores at the level of complexes. This is a plausible perspective because all the punches are the same cancer after all. This is a boon with regards to diagnosis.

However, the protein/gene-level heterogeneity, despite the complex-level homogeneity, does impact treat strategy/response. To see this, suppose there are say 5 distinct start points A/B/C/D/E that converge on an end point X to turn on a cancer pathway. A drug that blocks X will block cancer development, while a drug that blocks say A will block cancer with A as starting point (but not those with B/C/D/E as starting points). So it is still important to figure out what the protein/gene-level heterogeneity is, despite complex-level homogeneity. We suspect the heterogeneity of protein/gene

expression within significant complexes has more bearing than those outside significant complexes, although this hypothesis may not be proven satisfactorily given this dataset.

There are also other levels of details that alter drug response. For example, while A is a starting point of a patient's cancer, and there is a drug that normally blocks A. However, if the target site on A is mutated in this patient, he may not respond to the drug. Or, the patient has different drug metabolism (e.g. he has extra pumps that pump the drug out of his cells) that changes his response to the drug. These may not be easily detected in the data used in this manuscript (i.e., drug response data is needed as augmentation). Furthermore, without the presence of other supporting data e.g. from other proteomics platforms (to estimate technical bias), RNA-sequencing for concordance analysis (i.e., the gene expression profiles support the proteomics analysis), it is difficult to paint a complete mechanistic picture. However, even in spite of the endemic coverage, consistency and reproducibility issues in modern-day proteomics, interesting perspectives and insights may be derived by a careful and thorough cross-examination of the data using a variety of analytical techniques. Furthermore, we advocate the use of network-based analysis as it has useful noise-resistant properties.

5. Conclusions

Cancer is a heterogeneous disease, and causal factor analysis has not been well considered or incorporated in most early comparative studies (e.g. normal vs cancer). Through network-based approaches such as PFSNet and qPSP, we find that despite the perceived overwhelming amount of heterogeneity amongst individual patients, the low overlaps may stem from a lack of sensitivity of the analytical technique used. In identifying common causes, obtaining multiple punches/samplings per patient is useful for identifying conserved yet discriminative features. But once these features have been verified, only a single punch/sampling is needed subsequently for validation studies or actual diagnostic use.

The third and most important finding in this study is that network-based approaches are superior to traditional protein-based analytical strategies. And, when properly applied, is possibly a more reliable means for biomarker/target discovery.

References

- [1] D. Hanahan, R.A. Weinberg, Hallmarks of cancer: the next generation, *Cell* 144(5) (2011) 646-74.
- [2] D. Hanahan, R.A. Weinberg, The hallmarks of cancer, *Cell* 100(1) (2000) 57-70.
- [3] W.W.B. Goh, L. Wong, Why breast cancer signatures are no better than random signatures explained, *Drug Discov Today* 23(11) (2018) 1818-23.
- [4] W.W.B. Goh, L. Wong, Turning straw into gold: building robustness into gene signature inference, *Drug Discov Today* 24(1) (2019) 31-36.
- [5] D. Venet, J.E. Dumont, V. Detours, Most random gene expression signatures are significantly associated with breast cancer outcome, *PLoS Comput Biol* 7(10) (2011) e1002240.
- [6] M.L. Whitfield, L.K. George, G.D. Grant, C.M. Perou, Common markers of proliferation, *Nature reviews* 6(2) (2006) 99-106.
- [7] J.D. Rowley, Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining, *Nature* 243(5405) (1973) 290-3.
- [8] M. Verma, Personalized medicine and cancer, *J Pers Med* 2(1) (2012) 1-14.
- [9] P.M. Iannaccone, W.C. Weinberg, F.D. Deamant, On the clonal origin of tumors: a review of experimental models, *International Journal of Cancer Journal International Du Cancer* 39(6) (1987) 778-84.

709 [10] J.J. Cunningham, J.S. Brown, T.L. Vincent, R.A. Gatenby, Divergent and
710 convergent evolution in metastases suggest treatment strategies based on
711 specific metastatic sites, *Evol Med Public Health* 2015(1) (2015) 76-87.

712 [11] E.I. Dumbrava, F. Meric-Bernstam, Personalized cancer therapy-leveraging a
713 knowledge base for clinical decision-making, *Cold Spring Harb Mol Case Stud*
714 4(2) (2017) mcs.a001578.

715 [12] J. Zhang, J. Fujimoto, J. Zhang, D.C. Wedge, X. Song, J. Zhang, S. Seth, C.W.
716 Chow, Y. Cao, C. Gumbs, K.A. Gold, N. Kalhor, L. Little, H. Mahadeshwar, C. Moran,
717 A. Protopopov, H. Sun, J. Tang, X. Wu, Y. Ye, W.N. William, J.J. Lee, J.V. Heymach,
718 W.K. Hong, S. Swisher, Wistuba, II, P.A. Futreal, Intratumor heterogeneity in
719 localized lung adenocarcinomas delineated by multiregion sequencing, *Science*
720 346(6206) (2014) 256-9.

721 [13] E.C. de Bruin, N. McGranahan, R. Mitter, M. Salm, D.C. Wedge, L. Yates, M.
722 Jamal-Hanjani, S. Shafi, N. Murugaesu, A.J. Rowan, E. Gronroos, M.A. Muhammad,
723 S. Horswell, M. Gerlinger, I. Varela, D. Jones, J. Marshall, T. Voet, P. Van Loo, D.M.
724 Rassel, R.C. Rintoul, S.M. Janes, S.M. Lee, M. Forster, T. Ahmad, D. Lawrence, M.
725 Falzon, A. Capitanio, T.T. Harkins, C.C. Lee, W. Tom, E. Teeffe, S.C. Chen, S. Begum,
726 A. Rabinowitz, B. Phillimore, B. Spencer-Dene, G. Stamp, Z. Szallasi, N. Matthews,
727 A. Stewart, P. Campbell, C. Swanton, Spatial and temporal diversity in genomic
728 instability processes defines lung cancer evolution, *Science* 346(6206) (2014)
729 251-6.

730 [14] W.W. Goh, Y.H. Lee, M. Chung, L. Wong, How advancement in biological
731 network analysis methods empowers proteomics, *Proteomics* 12(4-5) (2012)
732 550-63.

733 [15] T. Guo, P. Kouvonen, C.C. Koh, L.C. Gillet, W.E. Wolski, H.L. Rost, G.
734 Rosenberger, B.C. Collins, L.C. Blum, S. Gillessen, M. Joerger, W. Jochum, R.
735 Aebersold, Rapid mass spectrometric conversion of tissue biopsy samples into
736 permanent quantitative digital proteome maps, *Nature medicine* 21(4) (2015)
737 407-13.

738 [16] W.W. Goh, T. Guo, R. Aebersold, L. Wong, Quantitative proteomics signature
739 profiling based on network contextualization, *Biology direct* 10(1) (2015) 71.

740 [17] W.W.B. Goh, L. Wong, Advanced bioinformatics methods for practical
741 applications in proteomics, *Briefings in Bioinformatics* (2017) bbx128-bbx128.

742 [18] W.W.B. Goh, L. Wong, Advancing clinical proteomics via analysis based on
743 biological complexes: A tale of five paradigms, *Journal of Proteome Research*
744 15(9) (2016) 3167-79.

745 [19] W.W.B. Goh, L. Wong, Evaluating feature-selection stability in
746 next-generation proteomics, *Journal of bioinformatics and computational biology*
747 14(5) (2016) 1650029.

748 [20] K. Lim, L. Wong, Finding consistent disease subnetworks using PFSNet,
749 *Bioinformatics* 30(2) (2014) 189-96.

750 [21] T.L. Guo, L.; Zhong, Q.; Rupp, N. J.; Charmpi, K.; Wong, C. E.; Wagner, U.;
751 Rueschoff, J. H.; Jochum, W.; Fankhauser, C. D.; Saba, K.; Poyet, C.; Wild, P. J.;
752 Aebersold, R.; Beyer, A., Multi-region proteome analysis quantifies spatial
753 heterogeneity of prostate tissue biomarkers, *Life Sci Alliance* 1(2) (2018)
754 e201800042.

755 [22] S. Srihari, C.H. Yong, A. Patil, L. Wong, Methods for protein complex
756 prediction and their contributions towards understanding the organisation,
757 function and dynamics of complexes, *FEBS letters* 589(19) (2015) 2590-2602.

758 [23] C.H. Yong, L. Wong, From the static interactome to dynamic protein
759 complexes: Three challenges, *Journal of bioinformatics and computational*
760 *biology* 13(2) (2015) 1571001.

761 [24] H.B. Fraser, J.B. Plotkin, Using protein complexes to predict phenotypic
762 effects of gene mutation, *Genome biology* 8(11) (2007) R252.

763 [25] W.W. Goh, M.J. Sergot, J.C. Sng, L. Wong, Comparative network-based
764 recovery analysis and proteomic profiling of neurological changes in valproic
765 Acid-treated mice, *J Proteome Res* 12(5) (2013) 2116-27.

766 [26] A. Ruepp, B. Brauner, I. Dunger-Kaltenbach, G. Frishman, C. Montrone, M.
767 Stransky, B. Waagele, T. Schmidt, O.N. Doudieu, V. Stumpflen, H.W. Mewes,
768 CORUM: the comprehensive resource of mammalian protein complexes, *Nucleic*
769 *acids research* 36(Database issue) (2008) D646-50.

770 [27] W.W.B. Goh, L. Wong, NetProt: Complex-based Feature Selection, J Proteome
771 Res 16(8) (2017) 3102-12.

772 [28] A. Belorkar, L. Wong, GFS: Fuzzy preprocessing for effective gene expression
773 analysis, BMC Bioinformatics 23;17(Suppl 17) (2016) 540.

774 [29] R. Suzuki, H. Shimodaira, Pvclust: an R package for assessing the uncertainty
775 in hierarchical clustering, Bioinformatics 22(12) (2006) 1540-2.

776 [30] J.A. Hanley, The statistical legacy of William Sealy Gosset ("Student"),
777 Community Dent Health 25(4) (2008) 194-5.

778 [31] J. Oliveros, VENNY. An interactive tool for comparing lists with Venn
779 diagrams, BioinfoGP, CNB-CSIC.

780 [32] Q. Zhong, T. Guo, N. Toussaint, U. Wagner, K. Charmpi, L. Calzone, A. Beyer, R.
781 Aebersold, P.J. Wild, Q. Zhong, Abstract 5565: Multi-omic profiling of prostate
782 cancer evolution in 39 patients, Cancer Research 77(13 Supplement) (2017)
783 5565.

784 [33] W.W. Goh, L. Wong, Integrating Networks and Proteomics: Moving Forward,
785 Trends in biotechnology 34(12) (2016) 951-9.

786 [34] W.W. Goh, W. Wang, L. Wong, Why batch effects matter in omics data, and
787 how to avoid them, Trends in biotechnology 35(6) (2017) 498-507.

788 [35] W.W. Goh, L. Wong, Protein complex-based analysis is resistant to the
789 obfuscating consequences of batch effects --- A case study in clinical proteomics,
790 BMC genomics 4;18(Suppl 2) (2016) 142.

791 [36] J. Gregori, L. Villarreal, O. Mendez, A. Sanchez, J. Baselga, J. Villanueva, Batch
792 effects correction improves the sensitivity of significance tests in spectral
793 counting-based comparative discovery proteomics, J Proteomics 75(13) (2012)
794 3938-51.

795

796

Table 1 The performances of the models based on leave-one-patient-out cross-validation shows that network-based approaches like qPSP and PFSNet outperform protein-based differential analysis.

Training<- validation	p < 0.05 t test			p < 0.01 t test			P < 0.05 significant intersection			Using qPSP significant intersection	
	PFSNet	qPSP	Proteins	PFSNet	qPSP	Proteins	PFSNet	qPSP	Proteins	PFSNet	Proteins
P1 P2<-P3	0.86	0.83	0.56	0.86	0.88	0.61	0.83	0.94	0.83	0.94	0.5
P1 P3<-P2	0.83	0.5	0.67	0.83	0.5	0.33	0.92	0.83	0.5	0.83	0.58
P2 P3<-P1	0.92	1	0.58	0.92	0.92	0.83	1	1	0.75	1	0.41