Clinical Mass Spectrometry 18 (2020) 13–26

Contents lists available at ScienceDirect

Clinical Mass Spectrometry

journal homepage: www.elsevier.com/locate/clinms

A proposal for score assignment to characterize biological processes from mass spectral analysis of serum

Joanna Roder [a,*], Lelia Net [a], Carlos Oliveira [a], Krista Meyer [a], Senait Asmellash [a], Sabine Kasimir-Bauer [b], Harvey Pass [c], Jeffrey Weber [d], Heinrich Roder [a], Julia Grigorieva [a]

[a] Biodesix, Inc, 2970 Wilderness Place, Boulder, CO 80301, USA
[b] Department of Gynecology and Obstetrics, University Hospital of Essen, Hufelandstrasse 55, 45147 Essen, Germany
[c] Department of Cardiothoracic Surgery, New York University Langone Medical Center, 550 1st Ave, New York, NY 10016, USA
[d] Perlmutter Cancer Center at NYU Langone Medical Center, 550 1st Ave, New York, NY 10016, USA

ARTICLE INFO

Article history:
Received 3 September 2019
Received in revised form 1 September 2020
Accepted 2 September 2020
Available online 9 September 2020

Keywords:
Proteomics
Mass spectrometry
Serum proteome
Biomarker
Set enrichment analysis
Biological scores

ABSTRACT

Introduction: Most diseases involve a complex interplay between multiple biological processes at the cellular, tissue, organ, and systemic levels. Clinical tests and biomarkers based on the measurement of a single or few analytes may not be able to capture the complexity of a patient's disease. Novel approaches for comprehensively assessing biological processes from easily obtained samples could help in the monitoring, treatment, and understanding of many conditions.
Objectives: We propose a method of creating scores associated with specific biological processes from mass spectral analysis of serum samples.
Methods: A score for a process of interest is created by: (i) identifying mass spectral features associated with the process using set enrichment analysis methods, and (ii) combining these features into a score using a principal component analysis-based approach. We investigate the creation of scores using cohorts of patients with non-small cell lung cancer, melanoma, and ovarian cancer. Since the circulating proteome is amenable to the study of immune responses, which play a critical role in cancer development and progression, we focus on functions related to the host response to disease.
Results: We demonstrate the feasibility of generating scores, their reproducibility, and their associations with clinical outcomes. Once the scores are constructed, only 3 μL of serum is required for the assessment of multiple biological functions from the circulating proteome.
Conclusion: These mass spectrometry-based scores could be useful for future multivariate biomarker or test development studies for informing treatment, disease monitoring and improving understanding of the roles of various biological functions in multiple disease settings.

1. Introduction

Complex diseases, such as cancer, are characterized by numerous gene-environment interactions and perturbations on genetic, epigenetic and physiological levels, involving affected tissues and the organism as a whole. A method that can assess levels of, and changes in, related biological functions from simple blood-based measurements, and that can be repeated over time, would be of great utility for understanding human disease.

Abbreviations: AIR, acute inflammatory response; ANG, angiogenesis; APR, acute phase reaction; ALK, anaplastic lymphoma kinase; BRCA1/2, Breast Cancer Gene 1, Breast Cancer Gene 2; CA, complement activation; CI, confidence interval; CV, coefficient of variation; CPH, Cox proportional hazards; ECM, extracellular matrix organization; EGFR, epidermal growth factor receptor; FDA, US Food and Drug Administration; GLY, glycolysis; HbA1c, hemoglobin A1c; HR, hazard ratio; IFN1, interferon type 1 signaling and response; IFNg, Interferon γ signaling and response; IRn, type n immune response; IT, immune tolerance; LC MS-MS, liquid chromatography with tandem mass spectrometry; MALDI ToF, matrix-assisted laser desorption/ionization time of flight; MRM, multiple reaction monitoring; MS, mass spectral; m/Z, mass/charge; NSCLC, non-small cell lung cancer; OS, overall survival; PC, principal component; PCA, principal component analysis; PCn, principal component n; PD-1, programmed cell death protein 1; PD-L1, programmed death-ligand 1; QC, quality control; WH, wound healing.

* Corresponding author.
E-mail addresses: joanna.roder@biodesix.com (J. Roder), lelia.net@biodesix.com (L. Net), senait.asmellash@biodesix.com (S. Asmellash), sabine.kasimir-bauer@uk-essen.de (S. Kasimir-Bauer), Harvey.pass@nyulangone.org (H. Pass), Jeffrey.weber@nyumc.org (J. Weber), heinrich.roder@biodesix.com (H. Roder).

https://doi.org/10.1016/j.clinms.2020.09.001

Measuring univariate biomarkers, or a linear combination of a limited number of analytes, to characterize a disease trait or a particular aspect of the physiological state of a patient has shown significant utility: several gene alterations predicting predisposition to a disease (e.g., *BRCA1* and *BRCA2* mutations for breast and ovarian cancers [1]) or response to targeted therapy (e.g., *EGFR* mutations or *ALK* rearrangements in lung cancer [2], or *BRAF* mutations in melanoma [3]) are broadly used in clinical practice. There are also a number of protein biomarkers, such as glycated hemoglobin (HbA1c), fructosamine, and glycated albumin for detecting and monitoring pre-diabetes [4], or expression of PD-L1, which is associated with response to PD-1 blockade in various cancers [5] that have demonstrated important clinical value. However, there is a growing understanding that combinations of biomarkers can contain more information than a single biomarker [6,7]. For complex diseases, such as cancer, where outcomes are defined by the interplay between immune response and tumor biology, the likelihood that a single biomarker will be sufficient to predict clinical outcomes in response to immune-targeted therapy is low [8].

The application of modern data analysis methods to multiple measurements, simultaneously, may result in the creation of qualitatively different, more powerful, biomarkers for disease prognosis and treatment optimization. Despite these expectations, this approach has thus far seen limited success: only a small number of -omics tests have passed rigorous independent validation and demonstrated satisfactory levels of reproducibility [9]. Examples of the few validated multivariate biomarkers in clinical use include: the VeriStrat® proteomic test (Biodesix, Boulder, CO) for lung cancer [10], the FDA_cleared 70-gene MammaPrint® assay (Agendia, Irvine, CA) for breast cancer [11], and the OVA1® test (Vermillion, Austin, TX) to predict ovarian malignancy [12]. These tests provide binary results: VeriStrat Good or Poor, related to prognosis in NSCLC; MammaPrint High Risk or Low Risk, associated with breast cancer recurrence, and high or low risk of ovarian cancer for OVA1. While these binary classifications are useful in the specific indications in which they are employed, a more granular approach that could be extended across multiple disease settings would be of use when measuring levels of fundamental biological processes. Taking into account potentially broad applications of assessing one or more processes of interest in an individual sample, we sought to explore the possibility of development of related biological scores.

Availability of a biological sample is of paramount importance in clinical practice. Analysis of blood samples remains a mainstay of medical testing and provides ready access to the circulating proteome – a rich source of biological information. The abundance of proteins in blood-based samples spans more than 10 orders of magnitude and can reflect various aspects of the physiological state of an organism [13]. Measurement of the circulating proteome is especially suitable for analyzing interactions between host and disease on a systemic level, as it can assess both inflammation and immune responses [14,15]. Matrix-Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-TOF) mass spectrometry is a highly reproducible method for the high-throughput measurement of the protein content of serum or plasma that requires small sample volumes [16,17]. Although the utility of conventional MALDI mass spectrometry for biomarker research was limited by its low sensitivity, the Deep MALDI® method has enabled a dramatic increase in the number of mass spectral (MS) features (peaks) that can be quantified [17]. This advance has extended the information content of MALDI spectra and permitted the development and validation of multiple MALDI mass spectrometry-based serum proteomic tests [18,19].

Several methods have been used to identify associations between multiple MS features and biological processes [20–22]. These approaches are based on the principles of set enrichment analysis [23,24]. Instead of analyzing expression differences in a univariate manner, one protein or gene at a time, enrichment analysis assesses expression differences consistent across a set of genes or proteins related to a specific biological function. Hence, it is possible to associate the measurement of a feature in a mass spectrum not just with a specific protein, but directly with a particular biological process. In prior work, we have shown that MS features measured in MALDI mass spectra are associated with many biological processes of interest, such as immune response and inflammation [22]. Furthermore, we demonstrated that for many processes we could identify multiple associated MS features that could contain information about the activation of the specific process. Here, we further develop these ideas to describe a method which allows MS data to be combined into scores that can quantify the level of biological processes. These scores could be used to characterize the level of specific biological processes, possibly as a function of time, for an individual patient or used in the development of univariate or multivariate tests to classify patients into different molecular phenotypes. Taking cancer as the disease of interest, we generate scores associated with biological functions related to cancer development and prognosis using data derived from the serum of cancer patients via MALDI mass spectrometry. We then analyze the reproducibility of these scores and explore their potential biological relevance and potential association with clinical variables using serum samples collected from multiple cohorts of cancer patients. We demonstrate the potential viability and utility of this biological score approach for future multivariate biomarker and molecular diagnostic test development studies in complex disease settings when the serum proteome is expected to contain information relevant for diagnosis, prognosis, and treatment decisions.

2. Methods

2.1. Samples

All serum samples used in this study were collected under Ethics Committee/Institutional Review Board-approved protocols according to the requirements of the relevant commercial or academic biobank, clinical trial, or observational study.

2.1.1. Sample sets for score development and validation

Four sample sets were used for the development and validation of the scores.

a. Reference Set 1 consisted of 100 serum samples from patients diagnosed with non-small cell lung cancer (NSCLC); 46 patients were female, 54 male. Samples were obtained from the commercial biobanks Conversant Bio (Huntsville, AL) and Oncology Metrics (Fort Worth, TX).
b. Reference Set 2 consisted of 49 serum samples from lung cancer and cancer-free patients. They were purchased from commercial biobanks, including Conversant Bio (Huntsville, AL), AdeptBio (Memphis, TN), and ProMedDx (Norton, MA).
c. The Score Development Set consisted of 85 serum samples collected prior to second-line therapy from patients with advanced NSCLC who were subsequently treated with erlotinib as part of a randomized clinical trial [10].
d. The Score Validation Set consisted of 123 serum samples collected prior to second-line therapy from patients with advanced NSCLC who were subsequently treated with chemotherapy as part of the same clinical trial as the patients whose samples constituted the Score Development Set [10]. Patients from that trial were randomly allocated to

receive either chemotherapy or erlotinib, and, hence, the patient populations associated with the Score Development Set and the Score Validation Set should be similar, although serum samples were available for only 208 of the 263 patients enrolled in the study.

2.1.2. Sample sets for demonstration of score utility

The following sample sets were used for demonstration of score utility.

a. The Melanoma Set consisted of pretreatment serum samples from 118 patients treated with nivolumab as second or later line therapy in a phase II clinical study, NCT01176461 [18].
b. The Ovarian Cancer Set consisted of 102 serum samples collected at the time of surgery, prior to subsequent chemotherapy, from patients with ovarian cancer [25].
c. The Early Stage Lung Cancer Set consisted of serum samples collected prior to surgery from 117 patients with Stage I NSCLC.

The Score Validation Set was also used in this setting.

2.1.3. Sample sets for reproducibility assessment

a. The Quality Control Reference Sample, created by pooling serum for five healthy subjects (purchased from ProMedDx LLC (Norton, MA, USA)) for quality control and batch correction during spectral preprocessing, was used to assess within-batch reproducibility.
b. The Machine Qualification Set of 40 samples collected from patients with lung cancer or colorectal cancer (purchased from Oncology Metrics (Fort Worth, TX, USA)), created for use in mass spectrometer qualification, was used to assess between-batch reproducibility.

The Early Stage Lung Cancer Set was also used as a population-representative cohort for the assessment of score reproducibility.

2.2. Data acquisition

2.2.1. Protein expression

An aptamer-based approach was used to obtain protein expression measurements for the two Reference Sets of samples [26]. The 1.3 k SOMAscan® assay (Somalogic, Boulder, CO) was used to assess 1305 proteins in the serum samples of Reference Set 1, while 1129 analytes had been measured using a prior version of the assay (the 1.1 k SOMAscan assay). The list of proteins contained in each SOMAscan panel is provided in the supplementary materials.

2.2.2. MS acquisition

Serum samples were processed using a standard operating procedure that has been described previously [18]. Mass spectra were obtained using a MALDI-TOF mass spectrometer (SimulTOF Systems, Marlborough, MA). The Deep MALDI® methodology [17] was used to generate spectra from 3 μL of serum. The Deep MALDI approach allows the detection and quantitation of MS features in serum across a wider range of abundances than standard MALDI. This is achieved by exposing samples to 400,000 MALDI laser "shots" rather than the several thousand "shots" used in standard MALDI applications.

Independent sample preparation and spectral acquisition were performed twice for each serum sample in Reference Set 1. Sample volume only allowed one collection of mass spectra from Reference Set 2. Sample preparation and spectral acquisition was carried out three times for the Early Stage Lung Cancer Set. The Quality Control Reference Sample was prepared 14 times and spotted onto one MALDI plate (i.e., run in one batch) to allow assessment of within-batch reproducibility. The Machine Qualification Set of 40 samples were prepared and spotted onto one MALDI plate. This process was repeated nine times for the evaluation of between-batch reproducibility. The Score Development Set was run in three batches, separately from the three batches of the Score Validation Set.

2.3. Spectral processing

The spectra were processed to render them comparable between samples. Spectral processing, including background estimation and subtraction, normalization, alignment, and batch correction to reference spectra, followed methods previously described [18]. Parameters specific to processing for this study are provided in the supplementary materials. A pre-defined set of 274 MS features, with mass(m)/charge(Z) ratios between 3 kDa and 30 kDa, were used in this study. These features were selected because they are commonly found in spectra generated from human serum and are known to be reproducible. Each MS feature was defined as an m/Z range defined by two bounding m/Z values. The feature value of a feature for a particular spectrum was defined as the sum of intensities of the processed spectrum within the m/Z range of the feature. The definitions of the 274 features can be found in the supplementary materials.

2.4. Construction of a biological score

The method used to construct a score for a particular biological process of interest is outlined in Fig. 1 and described in detail in the sections below.

2.4.1. Definition of the protein set

A set of proteins related to a biological process of interest was generated by querying the GeneOntology database [27,49] using AmiGO [50] and EMBL-EBI QuickGO [51] web applications. Evidence was filtered to exclude automatic electronic annotations and allow all manual assignment codes. This resulted in a list of relevant gene products. The intersection of this list with the proteins measured in the SOMAscan panels defined the Protein Set for the specific biological process. For use with the Reference Set 1, we took the intersection of this list with the larger SOMAscan panel; for use with Reference Set 2, we took the intersection of this list with the proteins common to the two SOMAscan panels, (see Fig. 1). The Protein Sets associated with each process included in this study are available in the supplementary materials.

When selecting biological processes for this study, we focused on those related to hallmarks of cancer [28] that have prominent molecular representation in the circulating proteome. In particular, we considered various aspects of inflammation, interferon signaling and response, complement activation, and wound healing [28–30]. We also explored extracellular matrix organization, angiogenesis, and glycolysis.

2.4.2. Identification of MS features associated with the biological process using set enrichment analysis

MS features associated with the biological process of interest were identified using set enrichment analysis methods. The specific method used was based on the approach of Subramanian et al. [23], but included an extension designed to increase its power to detect associations in larger sample sets [31]. This approach has previously been successfully employed in this setting [22]. In addition, this method provided a natural framework for the combination of set enrichment analysis data from both Reference Set 1 and Reference Set 2. (Raw protein expression data from Reference

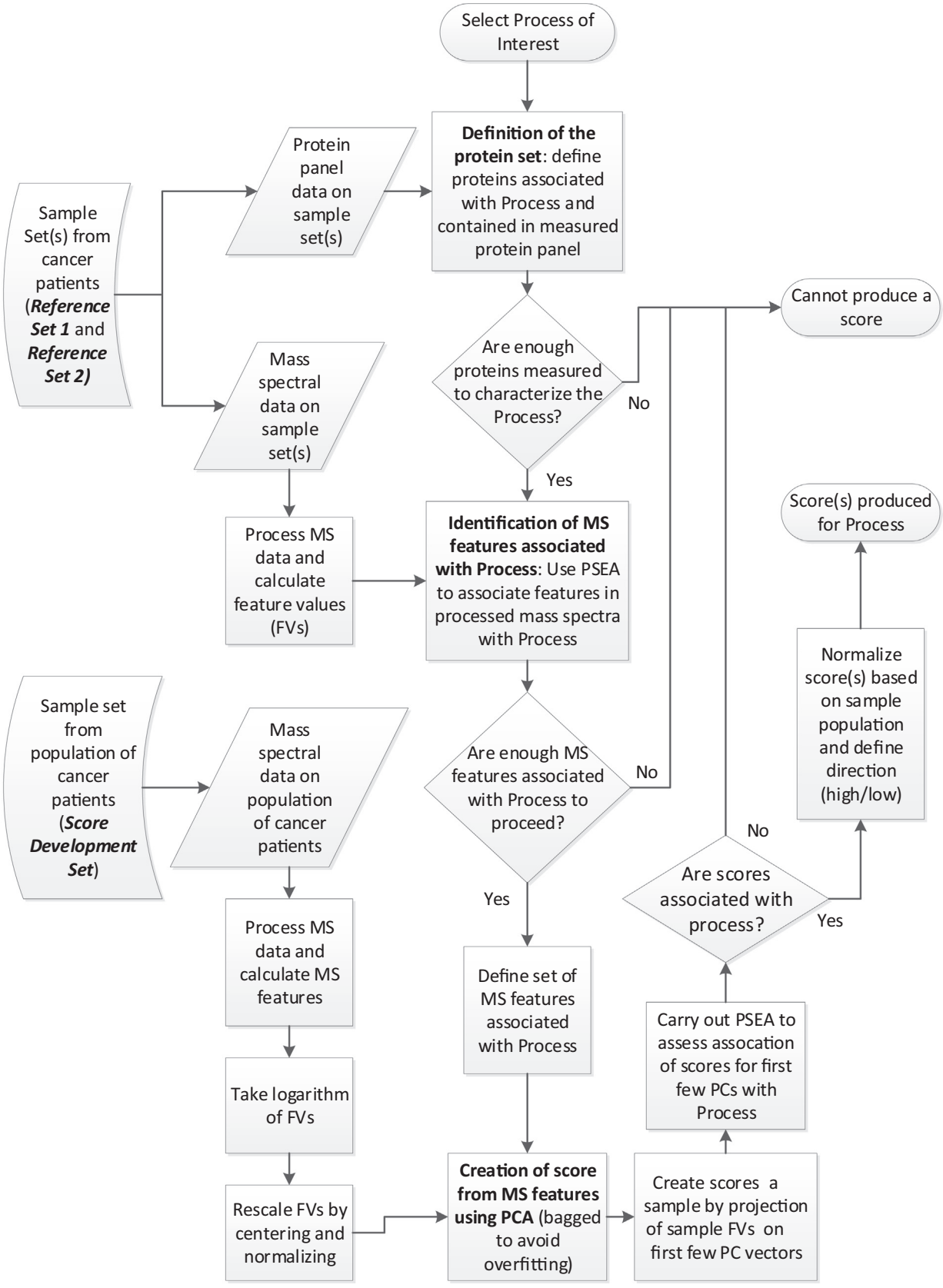

Fig. 1. Flowchart of the score creation process.

Set 1 and 2 were not combined directly due to batch effects between the two datasets and differences in the proteins contained in the two panels.)

In detail, the standard method was applied to Reference Set 2, due to its relatively small size [23]. The alternative method, which used an alternate enrichment metric averaged over multiple splits

of the dataset into halves [31], was applied to the larger Reference Set 1. The two metrics assessing enrichment generated for Reference Set 1 and Reference Set 2 were then normalized, to ensure they were comparable [23,52], and averaged to yield an assessment of association of a specific MS feature with the biological process in question across both Reference Sets. The significance of this association was evaluated by comparing it with the null distribution created via random permutation of the MS feature values over the samples. This approach generated a p value of association for each MS feature with the biological process of interest.

This entire process was carried out independently for each of the two MS acquisitions for Reference Set 1. A MS feature was deemed associated with the biological process of interest if the p value of association was below a chosen threshold for the analyses for both independent acquisitions. In general, a threshold of 0.05 was taken. However, for biological processes where many MS features were associated at this level of significance, a lower threshold was applied to allow a more reliable identification of associated features. Table 1 summarizes the significance level used and the number of associated features identified for the biological processes considered, as well as the number of proteins contained in the corresponding Protein Sets. We define the number of MS features associated with the biological process q to be the integer, K_q. Lists of the MS features identified as associated with each process are contained in the supplementary materials. In this study, we only considered constructing biological scores for processes for which we identified at least eight associated MS features. However, this threshold is arbitrary and, in principle, could be reduced to allow the combination of smaller numbers of MS features.

2.4.3. Creation of the score from the MS feature values using principal component analysis

The score for the biological process of interest was created using MS data from the Score Development Set. Feature values for each of the 274 features were available for each of the 85 samples. As described below, an ensemble-averaged [32] or bagged ("**b**ootstrap **agg**regat**ed**") version [33] of principal component analysis (PCA) [34] was carried out to minimize the tendency of PCA to overfit when there are more attributes (MS features) than instances (samples).

As many of the MS features have tails in their distributions, the logarithm of the feature values was taken to bring the distributions closer to normality. The mean and standard deviation of each MS feature was calculated for the Score Development Set (N = 85). These values were then used to center and scale each feature in a simple z-score normalization.

Fifty-six samples from the dataset were randomly selected; this subset constitutes one bag. PCA was performed on this subset of samples using the MS data for the features that had been determined as associated with the biological process. This process generated a number of principal component (PC) directions, N_{PC}, equal to the minimum of the number of features used, N_f, and the number of samples in the subset (i.e., $N_{PC} = 56$ if $N_f > 56$ and $N_{PC} = N_f$ if $N_f \leq 56$). This was repeated many (2^{17}) times for different randomly selected subsets of 56 samples (bags) from the Score Development Set. The set of PC directions was averaged pairwise over the multiple subset realizations, these combined pairwise, and so on, until all had been averaged together to yield ensemble-averaged PC directions. To check that the ensemble average was producing a meaningful average, we inspected the distribution of the angles between the ensemble-averaged PC vector and the PC vector calculated for each subset realization [35]. Histograms of these distributions are shown in the Supplementary Materials, together with similar plots showing the distributions of angles between the ensemble-averaged first PC vector and higher PC vectors per bag, the distributions of angles between the ensemble-average second PC vector and the third PC vector per bag, for processes in which the third PC was used for a score, and the distribution of the proportion of variation explained by PCs used for scores as determined for each bag. (Supplementary Figs. 1a-c, 2, and 3). The distributions of angle between the ensemble-averaged PC and the PCs calculated per bag were generally quite narrow (95th percentile <12° for 10 of the scores, Supplementary Table 8) and there was little overlap of the angular distributions between one PC and another relative to the ensemble averaged first PC.

Only the first three ensemble-averaged PC directions (vectors), $PC1^{(q)}$, $PC2^{(q)}$, $PC3^{(q)}$, were considered for score creation for each biological process q, although higher PCs could be considered in future studies. Each PC vector for biological process q has $K_q = N_f$ components, $PCm_i^{(q)}$, where $m = 1,2,3$ and $1 \leq i \leq K_q$. The three candidate scores for a sample were defined as the projections of the feature value vector of the sample on to these PC vectors. More precisely, if the sample had feature values FV_i ($1 \leq i \leq K_q$) for the K_q features associated with biological process q, we defined 3 candidate scores, $S1^{(q)}$, $S2^{(q)}$, $S3^{(q)}$ for process q for the sample via

$Sm^{(q)} = \Sigma_{i=1,Kq} PCm_i^{(q)} FV_i$ for $m = 1,2,3$.

To determine that the scores created with this method were useful as assessments of the particular biological processes, they had to satisfy several criteria. First, the score for process q had to be associated with process q. Even though the scores $Sm^{(q)}$ were generated as linear combinations of feature values of features associated with the biological process q, it does not necessarily follow that the scores themselves were associated with this biological process. Hence, to investigate which of $Sm^{(q)}$ had utility in assessing the biological process q, set enrichment analysis was performed for each score. Scores were generated for each sample in Reference Set 1 and Reference Set 2 and the association of each of the candidate scores with the process was assessed using the set enrichment analysis process. Scores that were not found to be associated with the biological process considered in their creation

Table 1
Biological processes investigated with the number of associated MS features, the threshold used to determine association, and the number of associated proteins in the 1.3 k SOMAscan assay. * For angiogenesis, the number of MS features identified was 18, but only 15 were used in score generation to provide a more angiogenesis-specific score (see details in Section 2.4.3).

Biological processes	Significance Level	Number of associated MS features identified	Number of proteins in the Protein Set
Acute inflammatory response	1.E-04	57	13
Acute phase reaction	1.E-04	91	14
Angiogenesis	0.05	15*	109
Complement activation	1.E-04	68	35
Extracellular matrix organization (ECM)	0.05	11	97
Glycolysis	0.05	14	20
Immune tolerance	0.01	50	10
Interferon type 1 signaling/ response (IFN type 1)	0.01	41	31
Interferon γ signaling/ response (IFN γ)	0.05	32	68
Type 1 immune response	0.05	32	24
Type 17 immune response	0.05	34	13
Type 2 immune response	0.05	9	11
Wound healing	1.E-05	10	49

were discarded. Second, we further demanded that the score or scores created for a particular process were more strongly associated with that process than scores created using sets of MS features deemed as associated with other biological processes, i.e., the score(s) made using the sets of MS features deemed associated with a biological process q had to be more strongly associated with process q than scores made using sets of MS features for processes other than q. In one instance (Angiogenesis), to produce a score satisfying this criterion, we discarded three of the 18 features, which were associated not just with this function, but also with IFN γ signaling and response.

The scores meeting these criteria were generated for all samples in the Score Development Set and the Score Validation Set. The distributions of the scores were compared between Development Set and Validation Set to check for any overfitting to the Score Development Set. Good generalization of the score distribution was found between the sets, as shown in Fig. 2. (The corresponding plots for the other biological process/principal component scores considered are included in the supplementary materials. In addition, Mann-Whitney p values for comparison of medians and Brown-Forsythe p values for comparison of variance about the median are provided for all scores.)

The scores were then standardized by centering the median at 0 and scaling by the difference between the 84th and 16th percentile (so that the variance of the distribution was around 1). This was carried out on the pooled Score Development and Score Validation Set population. In addition, the "direction" of the score was determined, i.e., the standardized score was multiplied by 1 or −1 so that higher score values corresponded to higher levels or higher activation of the biological process. This latter step was necessary as PCA determines the PCs to within an arbitrary multiplicative factor of 1 or −1; without this adjustment, higher or lower scores could correspond to higher levels of activation of the biological process. The direction was determined by examining the correlation of each score with the protein expression values within Reference Set 1 for proteins known to be relevant to the biological process in question. For example, relevant proteins measured in the panel known to be elevated for high levels of the biological process had to be associated with high scores. If correlations between proteins important for a process and the score for that process were very weak or gave conflicting information on score direction, the score was rejected; unambiguous determination of the score direction was a third criterion for a score to be considered useful.

In summary, each retained candidate score is completely defined by: the feature value normalization factors, its PC vector, $PCm^{(q)}$, the score normalization factors, and any required direction correction. Once these parameters have been determined, the score can be generated for any sample for which processed mass spectra are available.

3. Results

3.1. Reproducibility of scores

The reproducibility of the scores was assessed within and between batches and studied separately for a population-representative cohort.

The Quality Control Reference Sample was prepared from scratch 14 times and each preparation spotted onto a single MALDI plate. Biological scores were generated from each of the 14 resulting Deep MALDI spectra. The standard deviation across the 14 preparations were calculated to assess within batch reproducibility. The standard deviations for the Complement Activation Score, Glycolysis Score, Wound Healing Score and Type 17 Immune Response Score (from third principal component) were 0.12, 0.14, 0.19, and 0.16, respectively. Across all 17 scores created, the standard deviations ranged from 0.09 for the ECM score created from the third principal component to 0.22 for the third principal component score for Interferon γ, with a median of 0.16. (Note that each score was scaled so that its distribution had a standard deviation across a clinical population of 1.)

Between batch reproducibility was assessed using the Deep MALDI spectra obtained from nine runs of the Machine Qualification Set. This set of 40 serum samples can be spotted onto a single MALDI plate. This process was repeated nine times across a time span of 31 months; each time all samples were prepared from scratch. The standard deviation across batches for Complement Activation Score ranged from 0.07 to 0.30 (median 0.14), depending on sample. The corresponding results for Glycolysis Score, Wound Healing Score, and Type 17 Immune Response Score (PC3) were 0.11–0.50 (median 0.28), 0.13–0.60 (median 0.29), and 0.14–0.43 (median 0.29). Plots of the mean score against the standard deviation of the score across batches for all 40 samples and histograms of the standard deviations across batches for all 40 samples are provided for all 17 scores in the supplementary materials. While score reproducibility was clearly sample dependent, the standard deviation across batches did not show any marked dependence on the mean score value of the sample.

The Machine Qualification Set had been assembled to provide a good coverage of MS feature values within the 40 sample set. It was not designed to be representative of a clinical patient population. In order to provide an assessment of score reproducibility on a clinical population representative sample set, the scores were generated for multiple, independent, MS generations from the Early Stage Lung Cancer Set. Spectra were acquired for each of the 117 samples in this set from scratch three times over a period of more than two months. Fig. 3 illustrates reproducibility by showing concordance plots of Complement Activation Score, Glycolysis Score, Wound Healing Score (all from PC1) and Type 17 Immune Response Score (from PC3) for the three sets of spectra generated.

Corresponding plots for the other scores are shown in the Supplementary Materials. Coefficients of determination (r^2) for concordance between Run 2 and Run 1 or Run 3 and Run 1 varied between 0.79 and 0.97 across all 17 scores, with median being 0.87. Slopes of the least squared fit lines for the concordance plots varied between 0.87 and 1.03 across all 17 scores, with median 0.94. The reasonable concordance demonstrated between the scores produced by repeat measurements on the same samples is a necessary condition for practical application of the scores.

3.2. Correlations between scores for different biological processes

The biological processes included in this study have varying levels of correlations with each other. Some aspects of host response to cancer are intrinsically related. For example, reaction of the innate immune system to the disease includes various features of acute response and inflammation [36,37], while activation of the complement cascade impacts both adaptive and innate aspects of the cancer immunity cycle [38]. Other processes, such as glycolysis, changes in angiogenesis, or alterations of extracellular matrix, are related to different hallmarks of cancer [28], and they may be less correlated. These relationships are reflected in the heatmaps of the correlation matrix between scores shown in Fig. 4.

Fig. 4 confirms that the scores display the inter-process correlations that are expected. For example, the scores for the closely related processes of acute inflammatory response, complement activation, and acute phase reactions, show strong correlations with each other in all four cohorts. Scores for processes that would be considered as less related, such as acute inflammatory response and extracellular matrix organization, show low levels of correla-

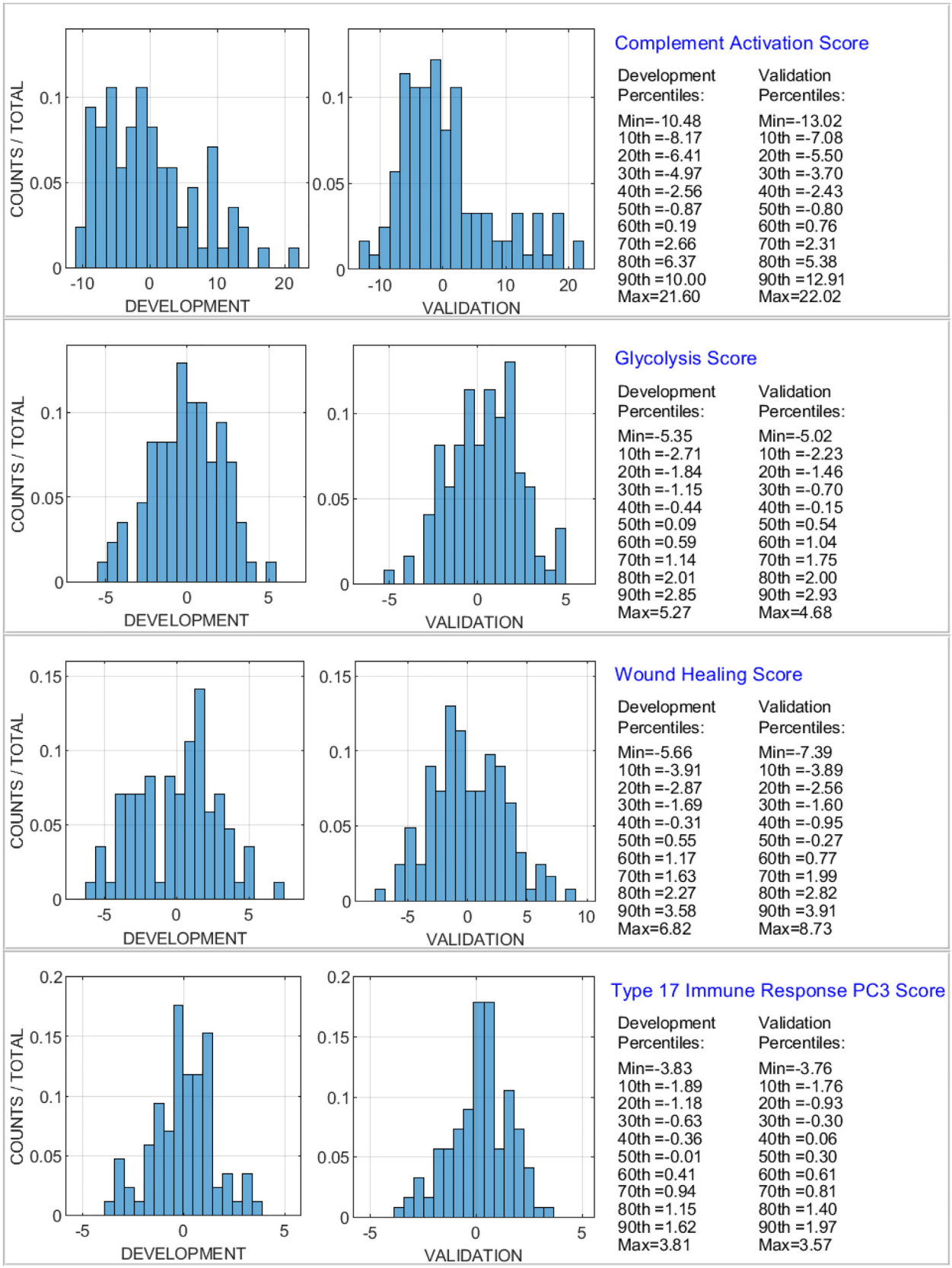

Fig. 2. Histograms showing score distributions for the Score Development Set ("DEVELOPMENT") and the Score Validation Set ("VALIDATION"), with inset percentiles, for the biological processes: complement activation, glycolysis, wound healing (all from PC1) and type 17 immune response (from PC3).

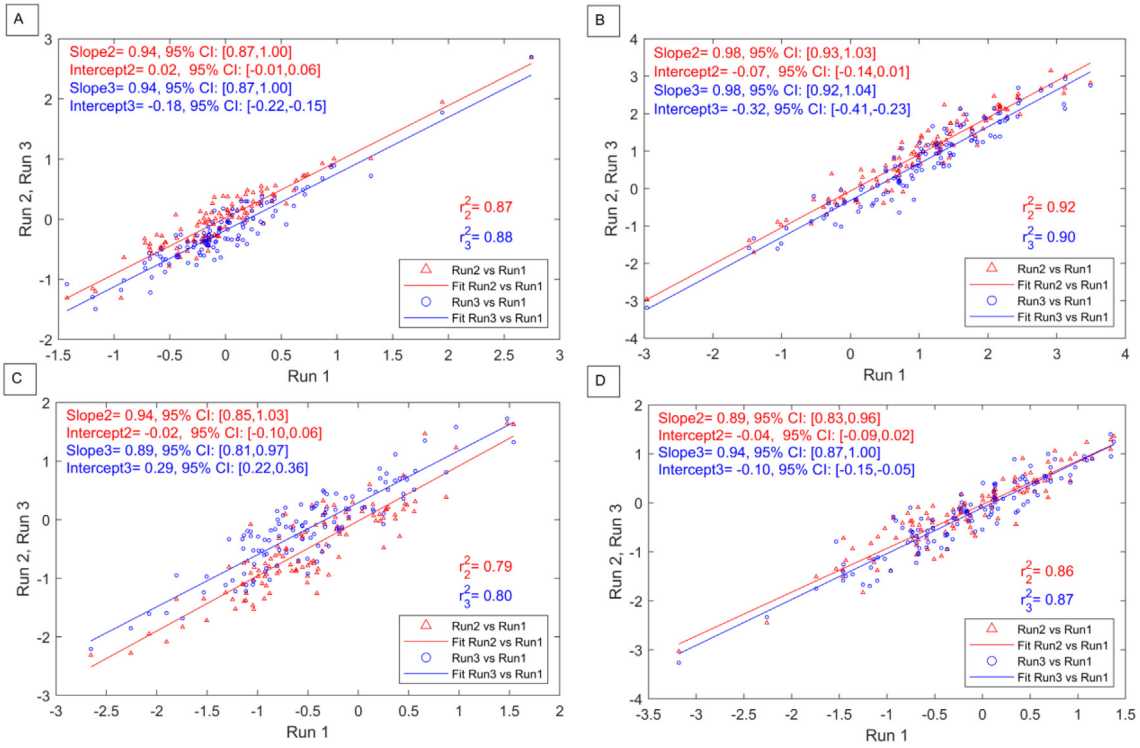

Fig. 3. Concordance plots showing the reproducibility of the scores generated from three independent spectral acquisitions (Run 1, Run 2, and Run 3) for samples in the Early Stage Lung Cancer Set. Scores shown are A – Complement Activation, B – Glycolysis, C – Wound Healing, (all from first principal component) and D – Type 17 Immune Response (from third principal component (PC3)). The corresponding least squares regression lines and statistics are shown for Run 2 vs Run 1 (in red) and Run 3 vs Run 1 (in blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

tion in all four cohorts. While overall the correlation structure is very similar between the four cohorts, the Early Stage Lung Cancer Set is somewhat different from the other three cohorts, with slightly less correlation between both IFN type 1 and IFN γ signaling and response and most of the other processes.

3.3. Applications of scores

3.3.1. Differences in score distribution across populations

Fig. 5 shows the distributions of scores for complement activation, type 17 immune response, wound healing, and glycolysis in three different tumor types (ovarian cancer, lung cancer, and melanoma) and two stages of lung cancer: early stage, pre-surgery and late stage, metastatic. The results illustrate that the distribution of each score can differ depending on the clinical indication and that the nature of differences in distribution across indications depends on which score one is interested in. As an example of differences across indication, the center of the Glycolysis Score distribution varies with indication. The Mann-Whitney test p values for ovarian vs advanced stage lung, advanced stage lung vs melanoma, and melanoma vs early stage lung are <0.001, 0.016, and <0.001, respectively. Interestingly, the center of the distribution differs between early and late stage lung cohorts for both Glycolysis Score ($p < 0.001$) and Wound Healing Score ($p < 0.001$), while it is similar for the Complement Activation score ($p = 0.432$). While the center of the Complement Activation Score distribution is similar for both lung cancer cohorts, its width is clearly broader in the advanced stage lung cancer cohort than in the early stage lung cancer cohort (Brown-Forsythe $p < 0.001$). Mann-Whitney and Brown-Forsythe p values are provided for all six pairwise comparisons of the scores in the Supplementary Materials (Supplementary Tables 8 and 9). Note that these kinds of differences were not observed when the score distribution was compared between the Score Development Set and Score Validation Set, which were drawn from the same patient population (see Fig. 2 and Supplementary Materials).

Tumor development and progression are associated with complex pathophysiological changes that are reflected in the circulating proteome. Specific characteristics of these changes depend on tumor site and cell origin, tumor microenvironment, host age, hormonal, and immune status, and reflect the immense heterogeneity of cancer [39]. Our observations comparing score distributions across indications support the notion that proteomic-based scores can reflect nuanced distinctions between disease states. Further validation in additional cohorts will be required to confirm the observed differences in score distribution.

3.3.2. Scores as predictors of outcome

One possible utility of the scores in medical applications is related to their associations with clinical outcomes. We evaluated these relationships in two ways using the scores created for the 13 biological processes in Table 1 that passed the score creation criteria. First, the association of each score, as a continuous variable, with overall survival (OS) was assessed within a Cox proportional hazard (CPH) model. Calculations were performed for advanced stage lung cancer patients treated with single agent chemotherapy (Score Validation Set, N = 123), ovarian cancer patients pre-surgery who subsequently received adjuvant chemotherapy (Ovarian Cancer Set, N = 102) and advanced melanoma patients who received immunotherapy (Melanoma Set, N = 118). The results are shown in Table 2. Second, each cohort was stratified into two subgroups by score above the cohort median ("score high") or below the cohort median ("score low"). Overall survival was then compared between the "score high" and "score low" subgroups using the Kaplan-Meier method and CPH models. Plots for four scores for the three patient cohorts analyzed are shown in Fig. 6.

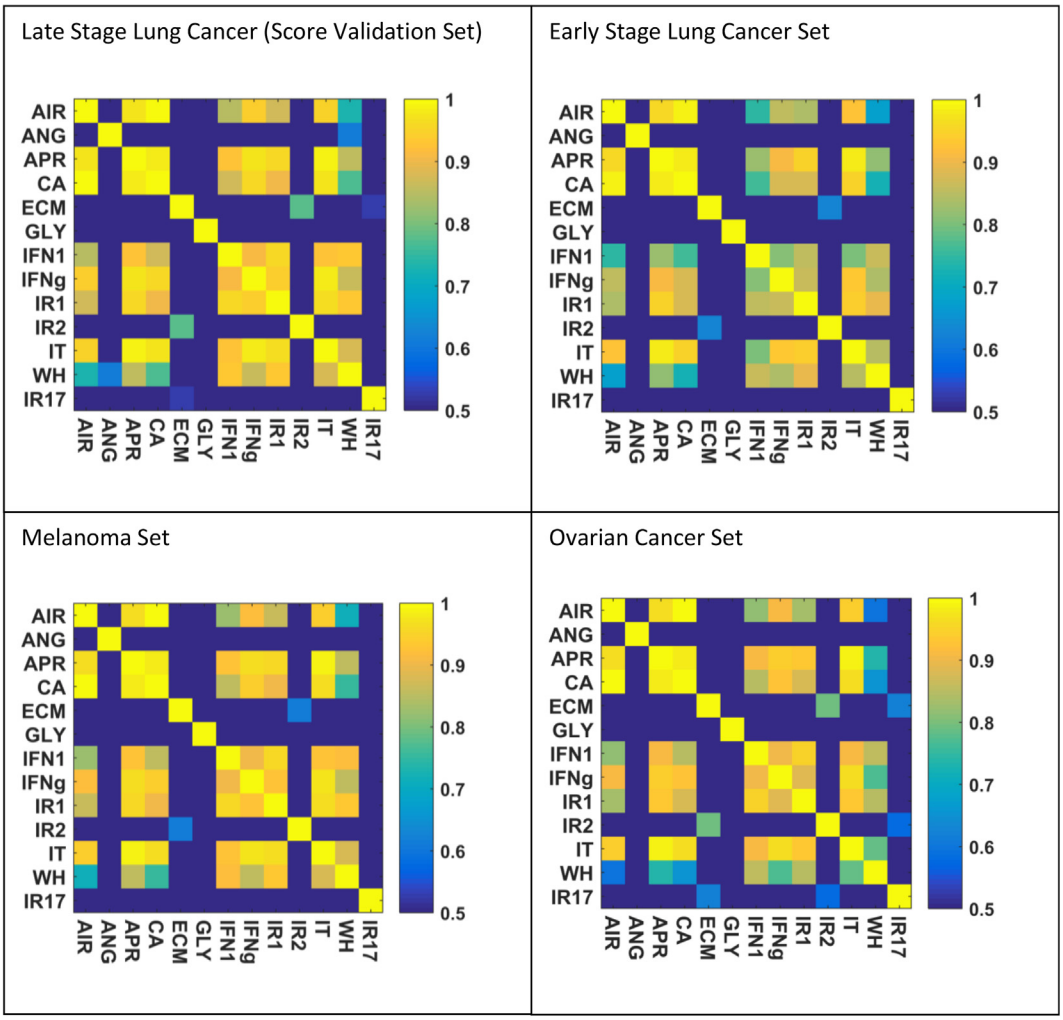

Fig. 4. Heatmaps of correlation matrix between pairs of scores for different biological processes across samples from different cohorts: Late Stage Lung Cancer (Score Validation Set), Early Stage Lung Cancer Set, Melanoma Set, Ovarian Cancer Set. The biological processes are abbreviated as AIR: Acute inflammatory response, ANG: Angiogenesis, APR: Acute phase reaction, CA: Complement activation, ECM: Extracellular matrix organization, GLY: Glycolysis, IFNg: Interferon γ signaling and response, IFN1: Interferon type 1, IT: Immune tolerance, WH: Wound healing, IR1: Type 1 immune response, IR2: Type 2 immune response, IR17: Type 17 immune response. Correlation matrix elements <0.5 are shown in dark blue. All scores are derived from the first principal component vector, apart from IR17, which is from the third principal component vector. Only one score per process (lowest PC) is illustrated. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

From Table 2 it is clear that many scores were associated with overall survival in all evaluated indications (e.g., Acute Inflammatory Response Score, Complement Activation Score, and Immune Tolerance Score), while other scores, such as that for angiogenesis, showed no sign of association with survival. Further, some processes (e.g., glycolysis, extracellular matrix organization, or type 17 immune response) seemed to impact outcome only in one indication (ovarian cancer) and not in the others. These observations are also reflected in Fig. 6, where impact of scores was studied based on threshold definitions.

The results are of interest in the context of comparative properties of cancers of different origins. All three tumor types demonstrated consistent correlations between survival and various aspects of innate immunity and inflammation (acute immune response, activation of complement system, acute phase reaction, interferon gamma-signaling and responses). This observation is in line with the general understanding of inflammation as an enabling characteristic of cancer [29,40] and supports previous results on association of prognostic and predictive multivariate proteomic tests with systemic host response to the disease [22,18,19,41]; this behavior should be expected if the scores provide information on levels of inflammation and the state of the innate immune system. However, differences were also observed between tumor types. The Ovarian Cancer Set demonstrated associations of OS with Extracellular Matrix Organization Score and Glycolysis Score, as well as a correlation of OS with one specific type of immune response (Type 17 Immune Response Score), which were not observed in the other two indications. Additionally, patients with melanoma or lung cancer, but not ovarian cancer, demonstrated association of OS with type 1 immune response, signaling and response to IFN type 1, and a correlation of OS (in melanoma) or a trend to correlation (lung cancer) with wound healing.

Investigation of association of scores with outcome can be useful to assess the information that is contained in multiple scores constructed using different principal component vectors for a particular biological process. For example, we were able to create scores for the first and third principal components for IFN γ signaling and response. While the score from the first principal component showed a significant univariate association with survival for patients with melanoma ($p = 0.001$), the score from the third component had $p = 0.225$. However, when incorporated simultane-

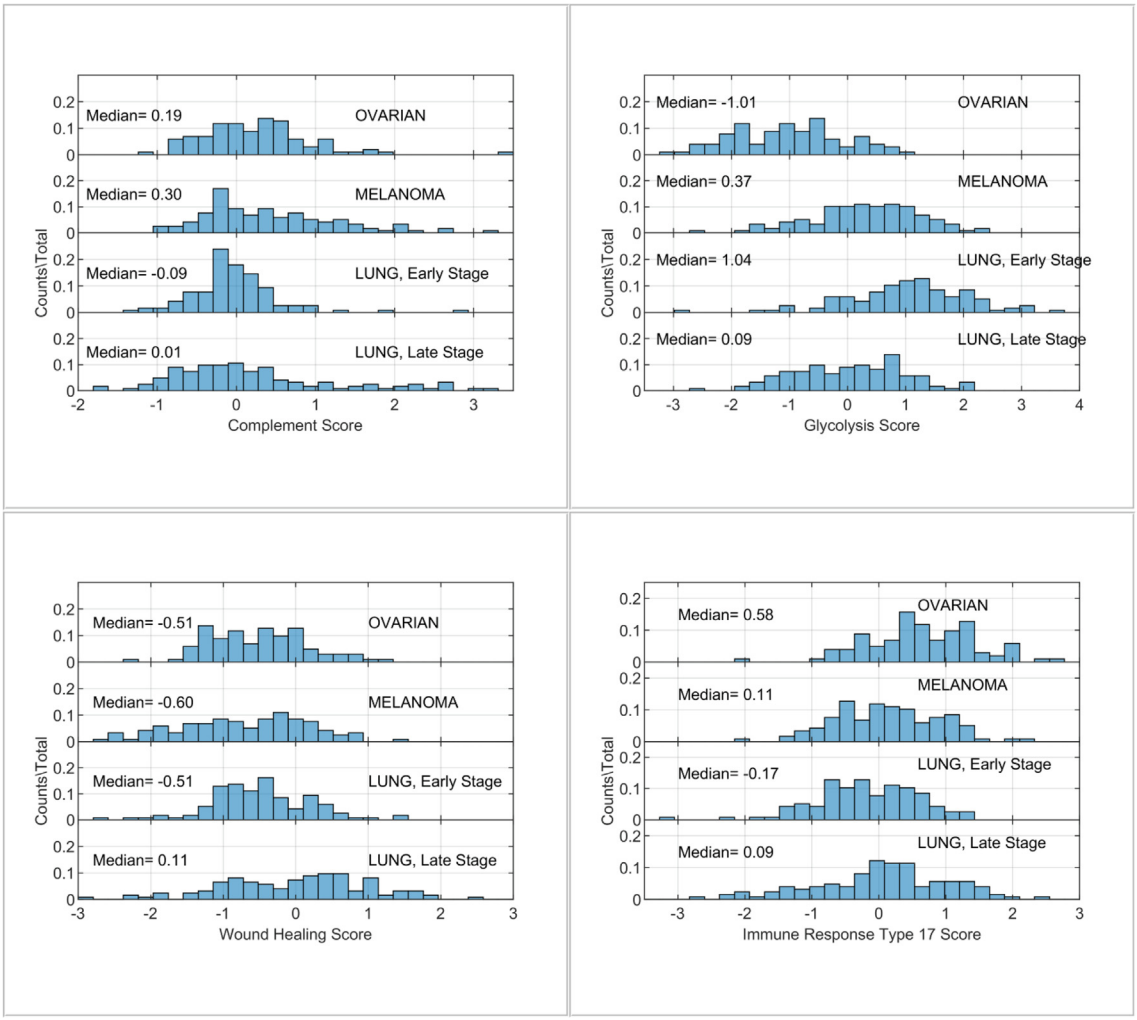

Fig. 5. Histograms of distributions of Complement Activation Score, Glycolysis Score, Wound Healing Score, and Type 17 Immune Response Score for four indications: ovarian cancer patients (Ovarian Cancer Set), advanced melanoma patients (Melanoma Set), early stage lung cancer patients (Early Stage Lung Cancer Set), late stage lung cancer patients (Score Validation Set).

Table 2
Univariate Cox proportional hazard ratios and p values for the scores built for the 13 biological processes of Table 1 for the Score Validation Set, the Ovarian Cancer Set and the Melanoma Set. Scores are created from the first principal component, unless indicated by the PCn suffix (n = 2 or 3 for second or third principal component, respectively). P values have not been corrected for multiple comparisons.

	Ovarian Cancer Set		Melanoma Set		Score Validation Set (Late Stage Lung Cancer)	
Scores	P	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)
Acute inflammatory response	0.031	1.53 (1.04–2.26)	<0.001	1.78 (1.33–2.38)	0.002	1.31 (1.10–1.56)
Complement activation	0.034	1.52 (1.03–2.24)	<0.001	1.75 (1.31–2.34)	0.004	1.30 (1.09–1.55)
Acute phase reaction	0.045	1.51 (1.01–2.26)	<0.001	1.76 (1.30–2.39)	0.007	1.29 (1.07–1.54)
IFN γ	0.022	1.56 (1.07–2.28)	0.001	1.71 (1.25–2.33)	0.008	1.27 (1.06–1.51)
Immune tolerance	0.047	1.47 (1.01–2.16)	<0.001	1.71 (1.27–2.31)	0.011	1.26 (1.05–1.51)
IFN type 1	0.086	1.42 (0.95–2.12)	0.001	1.67 (1.25–2.22)	0.023	1.23 (1.03–1.47)
Type 1 immune response	0.212	1.23 (0.86–1.96)	0.001	1.63 (1.22–2.17)	0.029	1.23 (1.02–1.47)
Wound healing	0.195	0.74 (0.47–1.17)	0.013	0.69 (0.52–0.93)	0.078	1.19 (0.98–1.44)
ECM_PC3	0.019	0.77 (0.61–0.96)	0.623	1.06 (0.85–1.32)	0.208	0.89 (0.74–1.07)
ECM_PC2	0.068	1.36 (0.98–1.88)	0.382	0.89 (0.69–1.15)	0.478	1.07 (0.89–1.29)
Type 2 immune response	0.036	1.40 (1.02–1.92)	0.249	0.87 (0.68–1.11)	0.482	1.07 (0.88–1.31)
ECM	0.006	1.48 (1.12–1.96)	0.578	1.07 (0.84–1.38)	0.524	1.07 (0.87–1.30)
Angiogenesis	0.696	1.09 (0.72–1.64)	0.220	1.20 (0.90–1.59)	0.857	1.02 (0.84–1.24)
Glycolysis	0.045	0.72 (0.52–0.99)	0.539	0.92 (0.72–1.19)	0.857	1.02 (0.83–1.25)
Type 17 immune response_PC3	0.038	1.49 (1.02–2.18)	0.464	1.13 (0.82–1.56)	0.897	1.01 (0.83–1.25)
Glycolysis_PC3	0.099	0.74 (0.52–1.06)	0.698	1.07 (0.77–1.49)	0.957	0.99 (0.80–1.23)
IFN γ _PC3	0.691	1.07 (0.77–1.50)	0.225	0.84 (0.64–1.11)	0.989	1.00 (0.82–1.23)

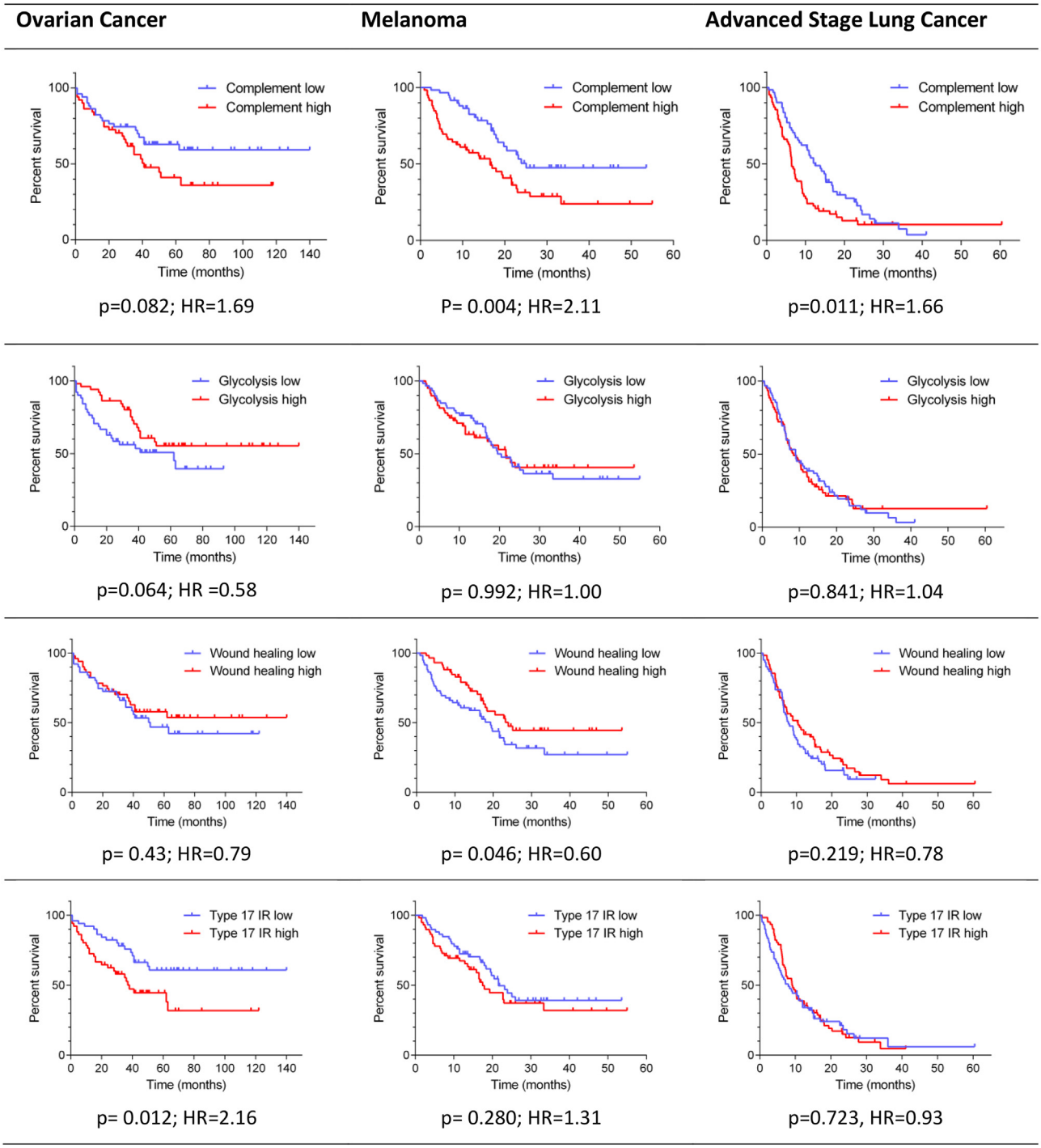

Fig. 6. Kaplan-Meier plots of overall survival for the Ovarian Cancer Set, the Melanoma Set, and the Score Validation Set (advanced stage NSCLC) stratified by score high (above median for cohort) and score low (below median for cohort) for Complement Activation Score, Glycolysis Score, Wound Healing Score (all from first PC) and Type 17 Immune Response (Type 17 IR) Score (from third PC).

ously into a multivariate Cox model, both scores were independent predictors of survival ($p < 0.001$ and 0.034 for first and third PC scores, respectively). Hence, they provided independent information about the survival of melanoma patients treated with immune checkpoint inhibitors. While both scores were associated with IFN γ and were constructed from MS features associated with IFN γ, the information that they provided was complementary, possibly relating to different aspects of IFN γ signaling. Elucidation of the biological significance of the ability of scores to predict outcome and the differences between indications (tumor types and therapies) requires further investigation and validation in independent patient cohorts. However, these observations are consistent with the notion that scores are related to relevant biological processes and may reflect different aspects of the disease state. The results also indicate that the scores contain information with potential utility in clinical and research settings, and that they may assist in monitoring and better understanding of the biological processes.

3.3.3. Scores as engineered features

In addition to using the scores as continuous measures of biological processes of interest, the scores can be thought of as a method of feature reduction or feature engineering [42]. Modern

machine learning methods are now able to efficiently and robustly combine many attributes to generate tests able to predict patient outcomes [18,19,43]. However, in some settings it may still be advantageous to reduce the number of variables used for test development or to focus on certain processes of interest.

To demonstrate the potential utility of the biological scores in this setting, we developed a test able to stratify patients with melanoma into two groups with better and worse outcomes when treated with an immune checkpoint inhibitor. For this purpose, we used all 17 scores presented in this manuscript. The Melanoma Set was used as the classifier development set. We used a hierarchical, dropout-regularized classifier architecture and an approach in which training class labels and classifier are iteratively refined to reveal the molecular structure consistent with the endpoint of interest [44]. More details are provided in the Supplementary Materials. A test was created for which patients classified as Good Prognosis had better overall survival compared with those classified as Poor Prognosis: the hazard ratio (HR) for survival between Good and Poor prognosis groups (95% confidence interval (CI)) was 0.60 (95% confidence inteval (CI) 0.35–1.04). The test validated well on an independent cohort of patients with advanced melanoma treated with checkpoint inhibition (see [19] for validation cohort details, HR (95% CI) = 0.60 (0.33–1.09)). This example illustrates the potential of the scores as engineered or meta-features for multivariate test development.

4. Discussion

We have demonstrated that, after using set enrichment analysis methods to identify MS features associated with biological processes of interest, it is possible to combine these features into a score characterizing a specific biological process. Such scores allowed the assessment of various biological processes within a patient based on the evaluation of the patient's serum proteome by mass spectrometry. We showed that these scores have potential for both monitoring of levels of specific biological processes, possibly longitudinally, for an individual patient and for use in multivariate tests that classify patients according to their molecular phenotype.

This current study has some limitations. We can only make scores for a particular biological process if the protein panels used to assess the Reference Sets contain proteins relevant for this process. If the protein panel does not include enough relevant proteins, it will not be possible to identify associated MS features. It would be of interest to extend protein expression measurements to larger panels to provide better coverage of biological processes of interest and increase the statistical power with which they can be investigated. The accuracy and reproducibility of the protein expression measurements are also key – with improved measurement precision providing greater power to identify associations. For similar reasons, the quality of MS data is very important. Here, we have used the Deep MALDI mass spectrometry method at a level of 400,000 laser shots. This provides us with a median CV across these MS features of around 5% [17]. Improving spectral quality by averaging over more laser shots could reduce CVs for these features and allow consideration of more MS features. Increasing the reproducibility of our MS acquisition would potentially enable us to identify more features associated with some processes than we can detect currently. In addition, increased precision of measurement could improve the reproducibility profile of the generated scores. We have employed multiple levels of quality control (QC) in our data acquisition and processing procedures to maintain data quality and reproducibility. In particular, we have implemented multiple automated QC checks in our spectral processing to detect degraded and oxidized samples, contaminated samples, and samples yielding spectra with MS peaks inconsistent with those typically found in human serum. In addition, all batches undergo batch QC checks using spectra from reference samples that are plated on each batch of experimental samples. As for other data acquisition modalities, a multilayered QC system is essential for ensuring data integrity and reproducibility.

In this study we have used MALDI mass spectrometry to survey the serum proteome. MALDI mass spectrometry has the advantages of being a high throughput, reproducible method which requires only small volumes (3 μL) of serum. The recognized disadvantages of MALDI include its semi-quantitative nature, its inability to detect the lowest abundance components of a complex mixture such as serum, and the fact that some proteins are not amenable to ionization under its soft ionization method. However, the approach that we have taken to score development, which does not rely on identification of the proteins constituting individual MS peaks or the ability to measure abundance (absolute or relative) of specific named proteins, largely circumvents these limitations. It is still possible, though, that the inability of MALDI to assess proteins with the lowest serum abundances and those not undergoing soft ionization could prevent creation of scores if these factors lead to the lack of detectable MS features associated with a biological process of interest.

It should be possible to create biological scores from serum proteomic data obtained from other platforms. For example, from an antibody-based protein panel, one has abundance measurements of many known proteins. It would theoretically be possible to use PCA or a different bioinformatics tool to combine the abundances of proteins known to be relevant for a particular biological process to create a score. Association of the resulting score with that biological process could again be verified using set enrichment analysis methods. Alternatively, data from other types of mass spectrometry, such as LC MS-MS or MRM could be employed. For methods where the (relative) abundances of known proteins are obtained, these could be combined as for the antibody-based panel data; for methods where the measured MS features are not known to directly represent (relative) abundances of known protein entities, a method parallel to the one we have presented here could be applied. Some of these methods require larger volumes of serum than we needed for our Deep MALDI sample characterization (e.g., SOMAscan requires at least 150 μL per sample and LC MS-MS studies typically report volume requirements of around 80 μL to allow for the depletion or fractionation of serum samples) and we were not able to test other methods due to sample volume limitations for our cohorts. MRM methods could be viable, as it has been reported that it is possible to measure 80 relatively high abundance proteins from 5 μL of serum [45]. However, these proteins would have to be prespecified and relevant to the biological process for which a score is required. If employing a different protein measurement technique, it would be necessary to adopt appropriate quality control and batch correction processes and to carry out a separate assessment of the reproducibility of the scores, once they had been generated and passed the relevant checks in the score creation workflow (Fig. 1).

We used ensemble-averaged PCA to combine the MS features associated with a particular biological process into a score. We then applied a series of checks (test of association of score with the process of interest, identification of direction of score from protein expression data, and validation of score distribution in a separate sample set) to test that the PCA process had produced a score with minimal required properties. This method had the advantage of being easy to implement and interpret and not requiring any choice of additional parameters or functions. The latter is an important consideration when working in high dimensional feature space with relatively few samples, where some algorithmic

approaches fail. For some processes where we could identify associated MS features, we were unable to make scores that met our criteria. One reason for this could be the limitations of standard linear PCA, which can only identify directions of variability in the data set that are rotations of our feature space axes. It would be of interest to investigate other methods of combining the MS features that might produce useful scores where linear PCA fails or could produce scores more strongly associated with the process of interest than those from linear PCA. Kernel PCA [46] is one obvious extension that could be explored.

There may also be some limitations to the use and applicability of the scores based on blood-based testing. While, theoretically, the serum proteome contains molecules representative of all processes in the organism, in practice some biological processes, such as those happening at the cellular level within small tumors, may be hard to assess. We have concentrated on the cancer disease state and relevant biological processes, which, as we anticipated from previous work, would be amenable to study from serum. While we expect that many other disease states, including those where inflammation and the immune system play key roles, will be good candidates for use of scores derived from mass spectrometry of blood-based samples, such scores may not be useful for some other biological processes and conditions.

The preliminary applications of scores, described herein, have demonstrated consistency with expected behavior in terms of association with outcomes, correlation between processes, and correlations across indications. We have also shown that scores can be used as engineered features for multivariate test development. It should be noted that these observations and related hypotheses are exploratory in nature and require independent validation. We expect that this biological score approach may be potentially useful in other applications, for example, monitoring of changes in biological processes during the course of disease or during treatment via longitudinal serum sampling. For example, recent work has used biological scores to investigate the mechanisms of early progression and immune-related adverse events for patients with non-small cell lung cancer treated with immunotherapy [47,48]. Complement, interferon gamma, and immune tolerance scores after three weeks on therapy, but not before treatment initiation, were associated with early disease progression. These scores, as well as extracellular matrix organization score, at three weeks were also associated with adverse events, which all occurred later, with median onset time of 105 days after therapy initiation. These observations provide hypotheses of underlying biological mechanisms that require validation in future studies and confirmation with alternative molecular approaches.

5. Conclusions

In summary, we have shown that it is possible to create scores to characterize biological processes using MS data generated from serum samples. The advantage of this approach lies in its ability to provide information on complex biological processes using MALDI-TOF mass spectrometry of easily available blood samples, which do not require invasive procedures and require only small sample volumes. The scores demonstrated acceptable levels of reproducibility between independent spectral acquisitions and have the potential for clinical relevance across different types of cancer. Serum samples assess properties of the whole organism, not just tumor, so the application of these scores could extend well beyond oncology. In particular, it would be of interest to evaluate the utility of this approach in other inflammatory and immune-related diseases and in settings, such as longitudinal monitoring, where the ease of sample collection would be important.

Declaration of Competing Interest

Joanna Roder, Heinrich Roder, Julia Grigorieva, Lelia Net, Senait Asmellash, Carlos Oliveira, and Krista Meyer are or were employees of Biodesix, Inc. and have or had stock or stock options in Biodesix, Inc. Joanna Roder, Heinrich Roder and Carlos Oliveria are inventors on related patents assigned to Biodesix, Inc. Sabine Kasimir-Bauer, Harvey Pass and Jeffrey Weber declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The collection of the Early Lung Cancer patient cohort was achieved from an ongoing U01 NCI Early Detection Research Network Grant (5U01CA214195, Dr. Harvey Pass). We acknowledge collaborations on the clinical application of mass spectrometry-based tests and utility of biological scores with Dr. P Ascierto and colleagues at the Instituto Nazionale Tumori, IRCCS Fondazione “G. Pascale”, Naples, Italy and Dr. Y Chae, Dr. A. Davis and colleagues at Northwestern University Feinberg School of Medicine, Chicago, USA.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.clinms.2020.09.001.

References

[1] A. Paul, S. Paul, The breast cancer susceptibility genes (BRCA) in breast and ovarian cancers, Front. Biosci. (Landmark Ed.) 19 (2014) 605–618.
[2] D. Morgensztern, M.J. Campo, S.E. Dahlberg, R.C. Doebele, E. Garon, D.E. Gerber, S.B. Goldberg, P.S. Hammerman, R.S. Heist, T. Hensing, et al., Molecularly targeted therapies in non-small-cell lung cancer annual update 2014, J. Thoracic Oncol. 10 (1 suppl. 1) (2015) S1–S63.
[3] L. Cheng, A. Lopez-Beltran, F. Massari, G.T. MacLennan, R. Montironi, Molecular testing for BRAF mutations to inform melanoma treatment decisions: a move toward precision medicine, Modern Pathol. 31 (1) (2018) 24–38.
[4] B. Dorcely, K. Katz, R. Jagannathan, S.S. Chiang, B. Oluwadare, I.J. Goldberg, M. Bergman, Novel biomarkers for prediabetes, diabetes, and associated complications, Diabetes, Metab. Syndrome Obes.: Targets Ther. 10 (2017) 345–361.
[5] S.P. Patel, R. Kurzrock, PD-L1 expression as a predictive biomarker in cancer immunotherapy, Mol. Cancer Ther. 14 (4) (2015) 847–856.
[6] B. Zethelius, L. Berglund, J. Sundström, E. Ingelsson, S. Basu, A. Larsson, P. Venge, J. Arnlöv, Use of multiple biomarkers to improve the prediction of death from cardiovascular causes, New Engl. J. Med. 358 (20) (2008) 2107–2116.
[7] C.A. Borrebaeck, Precision diagnostics: moving towards protein biomarker signatures of clinical utility in cancer, Nat. Rev. Cancer 17 (3) (2017) 199–204.
[8] G.V. Masucci, Cesano, R. Hawtin, S. Janetzki, J. Zhang, I. Kirsch, K.K. Dobbin, J. Alvarez, P.B. Robbins, S.R. Selvan, H.Z. Streicher, L.H. Butterfield, M. Thurin, Validation of biomarkers to predict response to immunotherapy in cancer: Volume I – pre-analytical and analytical validation, J. Immunother. Cancer 4 (76) (2016), eCollectiom 2016. Review.
[9] A.J. Vargas, C.C. Harris, Biomarker development in the precision medicine era: lung cancer as a case study, Nat. Rev. Cancer 16 (2016) 525.
[10] V. Gregorc, S. Novello, C. Lazzari, S. Barni, M. Aieta, M. Mencoboni, F. Grossi, T. De Pas, F. de Marinis, A. Bearz, et al., Predictive value of a proteomic signature in patients with non-small-cell lung cancer treated with second-line erlotinib or chemotherapy (PROSE): a biomarker-stratified, randomised phase 3 trial, Lancet Oncol. 15 (7) (2014) 713–721.
[11] F. Cardoso, L.J. van't Veer, J. Bogaerts, L. Slaets, G. Viale, S. Delaloge, J.Y. Pierga, E. Brain, S. Causeret, M. DeLorenzi, et al., 70-Gene signature as an aid to treatment decisions in early-stage breast cancer, N. Engl. J. Med 375 (8) (2016) 717–729.
[12] R.E. Bristow, A. Smith, Z. Zhang, D.W. Chan, G. Crutcher, E.T. Fung, D.G. Munroe, Ovarian malignancy risk stratification of the adnexal mass using a multivariate index assay, Gynecol. Oncol. 128 (2) (2013) 252–259.
[13] N.L. Anderson, N.G. Anderson, The human plasma proteome: history, character, and diagnostic prospects, Mol. Cell. Proteomics 1 (11) (2002) 845–867.

[14] S. Schenk, G.J. Schoenhals, G. de Souza, M. Mann, A high confidence, manually validated human blood plasma protein reference set, BMC Med. Genomics 1 (2008) 41.
[15] P. Gautam, S.C. Nair, K. Ramamoorthy, C.V. Swamy, R. Nagaraj, Analysis of human blood plasma proteome from ten healthy volunteers from Indian population, PLoS ONE 8 (8) (2013) e72584.
[16] G.L. Hortin, The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome, Clin. Chem. 52 (7) (2006) 1223–1237.
[17] M. Tsypin, S. Asmellash, K. Meyer, B. Touchet, H. Roder, Extending the information content of the MALDI analysis of biological fluids via multi-million shot analysis, PLoS ONE 14 (12) (2019).
[18] J.S. Weber, M. Sznol, R.J. Sullivan, S. Blackmon, G. Boland, H.M. Kluger, R. Halaban, A. Bacchiocchi, P.A. Ascierto, M. Capone, et al., A Serum protein signature associated with outcome after anti-PD-1 therapy in metastatic melanoma, Cancer Immunol. Res. 6 (1) (2018) 79–86.
[19] P.A. Ascierto, M. Capone, A.M. Grimaldi, D. Mallardo, E. Simeone, G. Madonna, H. Roder, K. Meyer, S. Asmellash, C. Oliveira, J. Roder, J. Grigorieva, Proteomic test for anti-PD-1 checkpoint blockade treatment of metastatic melanoma with and without BRAF mutations, J. Immunother. Cancer 7 (1) (2019) 91.
[20] M. Bessarabova, A. Ishkin, L. JeBailey, T. Nikolskaya, Y. Nikolsky, Knowledge-based analysis of proteomics data, BMC Bioinf. 13 (16) (2012) S13.
[21] S. Cha, M.B. Imielinski, T. Rejtar, E.A. Richardson, D. Thakur, D.C. Sgroi, B.L. Karger, In situ proteomic analysis of human breast cancer epithelial cells using laser capture microdissection (LCM)-LC/MS: annotation by protein set enrichment analysis (PSEA) and gene ontology (GO), Mol. Cell. Proteomics (2010).
[22] J. Grigorieva, S. Asmellash, C. Oliveira, H. Roder, L. Net, J. Roder, Application of protein set enrichment analysis to correlation of protein functional sets with mass spectral features and multivariate proteomic tests, Clin. Mass Spectrom. 15 (2020) 44–53, https://doi.org/10.1016/j.clinms.2019.09.001.
[23] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U. S. A. 102 (43) (2005) 15545–15550.
[24] M. Ackermann, K. Strimmer, A general modular framework for gene set enrichment analysis, BMC Bioinf. 10 (2009) 47.
[25] H. Roder, S. Kasimir-Bauer, R. Kimmig, P. Wimberger, R. Zeillinger, D.C. Castillo-Tong, J. Sehouli, I. Vergote, S. Mahner, J. Roder, et al., Serum mass spectrometry analysis in primary ovarian cancer (OC) treated with surgery and adjuvant chemotherapy (CT), J. Clin. Oncol. 31 (suppl.) (2013), abstr 5575.
[26] L. Gold, D. Ayers, J. Bertino, C. Bock, A. Bock, E.N. Brody, J. Carter, A.B. Dalby, B.E. Eaton, T. Fitzwater, et al., Aptamer-based multiplexed proteomic technology for biomarker discovery, PLoS ONE 5 (12) (2010) e15004.
[27] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, et al., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, Nat. Genet 25 (1) (2000) 25–29.
[28] D. Hanahan, R.A. Weinberg, Hallmarks of cancer: the next generation, Cell 144 (5) (2011) 646–674.
[29] F. Cavallo, C. De Giovanni, P. Nanni, G. Forni, P.L. Lollini, 2011: the immune hallmarks of cancer, Cancer Immunol. Immunother. 60 (3) (2011) 319–326.
[30] F.R. Balkwill, A. Mantovani, Cancer-related inflammation: common themes and therapeutic opportunities, Semin. Cancer Biol. 22 (1) (2012) 33–40.
[31] J. Roder, B. Linstid, C. Oliveira, Improving the power of gene set enrichment analyses, BMC Bioinf. 20 (1) (2019) 257.
[32] Z. Zhou, Ensemble Methods: Foundations and Algorithms. Chapman & Hall/CRC Machine Learning and Pattern Recognition Series, CRC Press, Boca Raton, FL, 2012.
[33] L. Breiman, Bagging predictors, Mach. Learning 24 (1996) 123–140.
[34] T. Hastie, R. Tibshirani, J.H. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, New York, 2001.
[35] R. Beran, M.S. Srivastava, Bootstrap tests and confidence regions for functions of a covariance matrix, Ann. Stat. 13 (1) (1985) 95–115.
[36] L.M. Coussens, Z. Werb, Inflammation and cancer, Nature 420 (6917) (2002) 860–867.
[37] P. Berraondo, L. Minute, D. Ajona, L. Corrales, I. Melero, R. Pio, Innate immune mediators in cancer: between defense and resistance, Immunol. Rev. 274 (1) (2016) 290–306.
[38] R. Pio, D. Ajona, S. Ortiz-Espinosa, A. Mantovani, J.D. Lambris, Complementing the cancer-immunity cycle, Front. Immunol. 10 (2019) 774.
[39] G.H. Heppner, M. Shekhar, Tumor heterogeneity is fundamental to the tumor ecosystem, Oncology (Williston Park) 28 (9) (2014) 780–781.
[40] Y.A. Fouad, C. Aanei, Revisiting the hallmarks of cancer, Am. J. Cancer Res. 7 (5) (2017) 1016–1036.
[41] M.J. Fidler, C.L. Fhied, J. Roder, S. Basu, S. Sayidine, I. Fughhi, M. Pool, M. Batus, P. Bonomi, J.A. Borgia, The serum-based VeriStrat® test is associated with proinflammatory reactants and clinical outcome in non-small cell lung cancer patients, BMC Cancer 18 (1) (2018) 310.
[42] A. Zheng, A. Casari, Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists, O'Reilly Media, Sebastopol, CA, USA, 2018.
[43] M. Müller, K. Hummelink, D. Hurkmans, A. Niemeijer, K. Monkhorst, J. Roder, et al., A serum protein classifier identifying patients with advanced non-small cell lung cancer who derive clinical benefit from treatment with immune checkpoint inhibitors, Clin. Can. Res. (in press).
[44] H. Roder, C. Oliveira, L. Net, B. Linstid, M. Tsypin, J. Roder, Robust identification of molecular phenotypes using semi-supervised learning, BMC Bioinf. 20 (1) (2019) 273.
[45] S. Ozcan, J.D. Cooper, S.G. Lago, D. Kenny, N. Rustogi, P. Stocki, S. Bahn, Towards reproducible MRM based biomarker discovery using dried blood spots, Sci. Rep. 7 (2017) 45178.
[46] B. Schölkopf, A. Smola, K. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (1998) 1299–1319.
[47] A. Davis, J. Park, L. Kim, G. Gim, W. Iams, M. Oh, et al., Serum proteomic scores for understanding the mechanisms of immune-related adverse events (irAEs) in non-small cell lung cancer. Abstract 5527, AACR Annual Meeting, 2020.
[48] A. Davis, J. Park, W. Iams, M. Oh, R. Lentz, J. Roder, et al., Serum proteomic scores for understanding response and mechanisms for resistance to immune checkpoint inhibitors in non-small cell lung cancer. Abstract 5526, AACR Annual Meeting, 2020.

Web references

[49] Gene Ontology <http://www.geneontology.org> (last accessed January 2018).
[50] AmiGO <http://amigo.geneontology.org/amigo> (last accessed January 2018).
[51] EMBL-EBI QuickGO <https://www.ebi.ac.uk/QuickGO/> (last accessed January 2018).
[52] GSEA User Guide, <http://software.broadinstitute.org/gsea/doc/GSEAUserGuideFrame.html> (last accessed May 2019).