cancers

MDPI

Review

Analytical Considerations of Large-Scale Aptamer-Based Datasets for Translational Applications

Will Jiang [1], Jennifer C. Jones [2], Uma Shankavaram [1], Mary Sproull [1], Kevin Camphausen [1] and Andra V. Krauze [1,*]

[1] Radiation Oncology Branch, Center for Cancer Research, National Cancer Institute, NIH, 9000 Rockville Pike, Building 10, CRC, Bethesda, MD 20892, USA; will.jiang@nih.gov (W.J.); uma@mail.nih.gov (U.S.); sproullm@mail.nih.gov (M.S.); camphauk@mail.nih.gov (K.C.)

[2] Translational Nanobiology Section, Laboratory of Pathology, NIH/NCI/CCR, Bethesda, MD 20892, USA; jennifer.jones2@nih.gov

* Correspondence: andra.krauze@nih.gov

Simple Summary: Aptamers represent an emerging technology that enables researchers to screen biological matrices such as blood and urine for thousands of different proteins at a rapid pace with high precision and accuracy. However, the sheer data volume generated by this high-capacity screening technique also creates a fundamental challenge towards efficiently analyzing these complex datasets and translating findings for the clinic. We address the new analytical considerations brought forth by aptamers, explore the necessary statistical analysis needed, and create a baseline to analyze these large-scale databases more comprehensively. In addition, we explore how aptamers can co-exist with current proteomic platforms to produce more robust findings in an evolving, multi-faceted approach towards the field. Unlocking the underlying signals masquerading behind these large datasets will ultimately empower clinicians and researchers to better understand diseases of interest and to curate more robust findings for patient care.

Abstract: The development and advancement of aptamer technology has opened a new realm of possibilities for unlocking the biocomplexity available within proteomics. With ultra-high-throughput and multiplexing, alongside remarkable specificity and sensitivity, aptamers could represent a powerful tool in disease-specific research, such as supporting the discovery and validation of clinically relevant biomarkers. One of the fundamental challenges underlying past and current proteomic technology has been the difficulty of translating proteomic datasets into standards of practice. Aptamers provide the capacity to generate single panels that span over 7000 different proteins from a singular sample. However, as a recent technology, they also present unique challenges, as the field of translational aptamer-based proteomics still lacks a standardizing methodology for analyzing these large datasets and the novel considerations that must be made in response to the differentiation amongst current proteomic platforms and aptamers. We address these analytical considerations with respect to surveying initial data, deploying proper statistical methodologies to identify differential protein expressions, and applying datasets to discover multimarker and pathway-level findings. Additionally, we present aptamer datasets within the multi-omics landscape by exploring the intersectionality of aptamer-based proteomics amongst genomics, transcriptomics, and metabolomics, alongside pre-existing proteomic platforms. Understanding the broader applications of aptamer datasets will substantially enhance current efforts to generate translatable findings for the clinic.

Keywords: aptamers; biomarkers; proteomics; bioinformatics; translational

check for updates

Citation: Jiang, W.; Jones, J.C.; Shankavaram, U.; Sproull, M.; Camphausen, K.; Krauze, A.V. Analytical Considerations of Large-Scale Aptamer-Based Datasets for Translational Applications. *Cancers* **2022**, *14*, 2227. https://doi.org/10.3390/cancers14092227

Academic Editor: Maxim V. Berezovski

Received: 23 March 2022
Accepted: 18 April 2022
Published: 29 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Proteomics has continued to establish itself as a field of growing promise both diagnostically and therapeutically. Historically, the two primary methods in proteomic analysis were the

enzyme-linked immunosorbent assay (ELISA) and mass spectrometry (MS). ELISA falls short of high-throughput capabilities and faces many challenges in widespread clinical integration due to calibration and quantification, reagent stability/availability/cross-reactivity, biomarker validation, and a lack of validated algorithms for computational analysis [1]. Mass spectrometry emerged over a decade ago as a preferred method for proteomic analysis that could produce large datasets for biological application, predominantly liquid chromatography MS (LC–MS) [2,3]. The ability to analyze post-translational modifications (PTMs) of proteins [4] unleashed a powerful new tool to decode the complexity offered up by evolving MS techniques in extraction, fragmentation, analysis, and database reference. Next-generation high-throughput advances in MS, such as data-independent acquisition MS (DIA-MS), employ permanent digital proteome maps while offering exceptional reproducibility and avoiding the inconsistent precursor ion fragmentation present in large-scale datasets generated via data-dependent MS (DDA-MS) [5]. However, MS suffers from several fundamental drawbacks, including protein inferences in relation to isobaric amino acids, co-elution concerns over PTMs, and flaws in algorithmic database searches [6–8]. Established and emerging high-throughput screening approaches [9–23] (Table 1) offer a much more comprehensive dynamic range to accommodate the human proteome and produce a higher sensitivity with a much lower detection limit compared to traditional MS platforms [24,25]. Amongst proteomic-based platforms, aptamer-based approaches are growing in number and popularity, generating large-scale datasets (Table 1) [15,26,27]. Aptamers are single strands of oligonucleotides (either ssDNA or ssRNA) that bind with high affinity and selectivity [28]. Selection of aptamers occurs through an in vitro evolution process known as Systematic Evolution of Ligands by Exponential Enrichment (SELEX), in which oligonucleotide libraries undergo multiple automated rounds of positive and negative selection to identify strongly selective aptamers [29–31]. These aptamers offer both diagnostic and therapeutic utility. Their highly selective and stable nature offers the potential for robust reproducibility, allowing for the creation of large datasets across multiple experiments. Aptamers have been integrated through various point-of-care diagnostics, synergistic combinations with antibody-based assays, and high-throughput screenings [32–34]. However, while sharing the underlying application of aptamers, comparative studies must be undertaken with caution, as even panel-to-panel variance is a commonality. For instance, one of the leading aptamer-based proteomics platforms, SomaLogic, carries several different protein panels ranging from 1300 to over 7000 [9]. Although overlap in protein targets may exist between the various panels, such as the 1.3K [35] and 5K [36] panels, the custom analyte panels may challenge cross-study comparisons. This review addresses the cross-platform intersectionality challenges between aptamers and proteomic "gold standard" platforms such as immunoassay. Aptamer data open a new realm of possibilities in proteomics by overcoming the quantitative volume requirements of LC–MS and exceeding the analyte capacity of ELISA. In addition, aptamer-based technologies are also valuable for application across a wide range of mammalian species, which is helpful for regulatory approval of novel pharmaceuticals which fall under the Food and Drug Administration's (FDA) Animal Rule. Emerging assays are compatible with a wide range of mammalian species, facilitating simultaneous analysis of samples from human clinical sample repositories and research animal models [9]. Unilateral analyses in omics often fail to capture the complexity of biological biomarkers and struggle with reproducibility. One of the critical challenges in omics analysis is addressing the high dimensionality and cross-referencing of datasets required for independent validation. Omics has made considerable strides in aggregating large-scale databases and repositories for cross-comparisons in genomic and transcriptomic analyses in the last few decades. The establishment of high-quality proteomic data has already begun in LC–MS through the ProCan proteomic knowledgebase, where uniformity in study design, sample preparation, and data analysis is provided [37]. With aptamer-based platforms producing an even more immense amount of data per sample through high-throughput and high-multiplexing analysis, there is a growing need to address the computational analysis of large datasets arising from aptamer technology. Streamlining analysis of aptamer data may expedite the discovery of blood-based biomarkers to surpass the current FDA-approval rate of fewer than two biomarkers per year [38]. Here, we provide a comprehensive review of aptamer-generated data processing and analytical strategies for clinically relevant translation to establish aptamers as a powerful tool in the multi-omics landscape.

Table 1. Overview of Common Proteomic Platforms.

Analytical Technique	Category	Protein Sample Literature Values [L]	Accepted Biospecimen Types	Reported Dynamic Range ‡	CV [LL]	Protein Capacity (Multiplex)
Proximity Extension Assay (Olink) [11]	Antibody	1 µL	Plasma, tissue/cell, synovial fluid, CSF, plaque extract, and saliva	LLOQ = 0.25 pg/mL	7.8% (intra) and 10.6% (inter)	***
Reverse Phase Protein Arrays [12,39–42]	Antibody	5 µg (1.0 to 1.5 mg/mL protein)	Tissue/cell, plasma, serum, biopsies, body fluids	LOD = 0.55 fg/mL	<15%	***
Bio-Plex [13,43]	Antibody (bead)	12.5 µL (serum/plasma) 50 µL (cell culture)	Plasma, serum, tissue/cell	LOD = 0.6–6.4 pg/mL	2–15%	****
Simoa [14,44,45]	Antibody (bead)	25 µL	Plasma, serum, urine, tissue/cell, CSF, saliva	LOD = 0.005 pg/mL	<10%	***
Aptamer Group (Optmer) [15]	Aptamer	38 µL	Plasma (diagnostics and therapeutics), urine, tissue/cell, liquid matrices	LOD = 55 ng/mL	<5%	*
Base Pair Technologies [16,26]	Aptamer	5–100 µL	Plasma, serum, tissue/cell	LOD = 1 pg/mL		**
SOMAscan [9,27,46–49]	Aptamer	55–100 µL	Plasma, serum, CSF, urine, cell/tissue, synovial fluid, exosomes	LOD = 1.6 pg/mL	4.6%	******
Electrochemiluminescence Immunoassay (Meso Scale and Lumit) [17,50]	ECLIA	50 µL	Plasma, serum, tissue/cell, CSF, urine, blood spots, tears, synovial fluid, tissue extracts	LOD = fg/mL	5–10%	**
Multiplex ELISA [1,18,43,51,52]	ELISA	25–50 µL	Plasma, serum, tissue/cell, urine, saliva, CSF	LOD = 0.61 to 18.90 pg/mL	9.5–28.5% (inter/intra)	**
Singleplex ELISA [53]	ELISA	100 µL	Plasma, serum, tissue/cell, urine, saliva, CSF	LOD = pg/mL	1.6–6.4% (intra) and 3.8–7.1% (inter)	*
2D-PAGE [19]	Gel electrophoresis	~100 µg (15–50 µL)	Plasma, serum, tissue/cell, urine	LOD = 10 ng to 100 ng	<20%	******
DDA-MS [20]	MS	10 µL	Plasma, serum, tissue/cell	LOD = 157 ng/mL	5.7%	*****
SWATH-MS [21,54,55]	MS (DIA)	5–10 µg	Plasma, serum tissue/cell, platelets, monocytes/neutrophils	LOD = 1 fmol	13.7%	*****
iTRAQ [22,56]	MS (labeling in LC–MS–MS)	12 µg	Plasma, serum, tissue/cells, saliva	LOD = 1 fmol (50 µg/mL)	<0.53%	*****
SRM/MRM [23,57,58]	MS (LC–MS–MS)	15 µL	Plasma, tissue/cell, dried blood spots	LOD = µg/mL (no enrichment)	6.1% (intra) and 11% (inter)	**

[L] Actual requirements may vary depending on transit conditions, company selected, and number of panels desired. [LL] Values reflect reported literature values; technical specifications vary based upon instrument and sample conditions. Groupings: * = 1–10; ** = 10–100; *** = 100–500; **** = 500–1000; ***** = 1000–5000; ****** = 10,000+. ‡ Quantification subject to multifactorial variability secondary to data origin, methodology, assay type, and analytes tested.

2. Setting Up Aptamer Studies for Clinical Translation

Aptamer technology provides a means of achieving scalable wide-ranging protein analysis. Samples sent off for aptamer-based panels can consist of various biological matrices ranging from plasma to serum to urine. Requiring roughly around 50 µL of sample volume and possessing more streamlined sample processing workflow techniques [59], aptamer technology can generate highly multiplexed analyses from pre-existing biorepositories that cover more significant concentration gradients than current platforms (Figure 1). Recent studies in the literature back this idea, as archived plasma samples in longitudinal cohorts demonstrate a high protein stability of over 90% at the one-year mark [60]. Accessing existing serum or plasma samples from pre-existing cohorts is another convenient advantage of aptamer-based platforms that enable larger sample sizes. Customizing clinical study designs to cater to the advantages of the aptamer platform will play an essential role in guiding proper statistical analysis for protein effect sizes later. The advantages of utilizing aptamer-based platforms are also associated with unique challenges. While relatively small samples from pre-existing cohorts may be employed, this can result in multifactorial sources of potential sample variability (collection, storage, management of the patients included in the studies) and can result in the utilization of samples that were not necessarily carefully experimentally curated or classified for confounders.

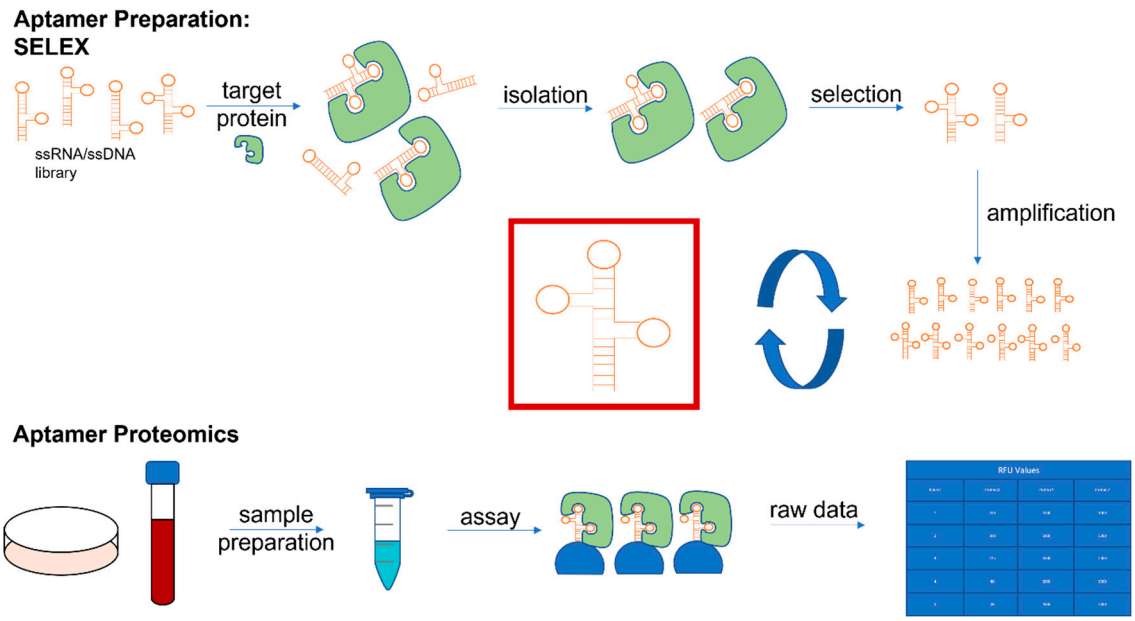

Figure 1. Generating proteomic datasets using aptamers.

2.1. Power, Power Analysis, and Protein Effect Size

Statistical power takes a requisite position at the forefront of study design considerations. Past studies on proteomic-based biomarker discovery reported less than 1% of reported biomarkers being incorporated into commercial assays, with a large number of published studies being underpowered [61]. One of the first National Institutes of Health-led workshops on proteomic biomarker pipelines emphasized the role of outlining basic statistical designs prior to study onset in order to maintain adherence to the FDA guidelines on biomarker translation [62]. A later workshop set forth a quantitative clinical criterion for proteomic biomarkers and a comprehensive statistical design with a requisite recommendation of 0.9 for both discovery and verification stage probability to achieve adequate power [63]. The challenges associated with the broad ranges of abundance in the proteome [64] and ranges of coefficient of variations (CVs) (including biological CVs) were all factored into Skates et al.'s statistical model for sample size. However, aptamer-based platforms may effectively reduce the cohort size requirements of MS-based proteomic platforms given their much lower values of CVs (~5%) [65] compared to technologies, such

as multiple reaction monitoring (up to 20%) [66]. Simulations for power analysis factor in CVs, the expected fold change needed, and desired power levels with a recommended four samples per group (CV = 20%, power = 0.8) [67]. Sensitivity and specificity are essential considerations during power analysis, and minimum targets vary according to target clinical use and commonly acceptable risk–benefit analysis [63]. Prospective-based studies striving for aptamer diagnostics require an intricate consideration for sample size and sensitivity and specificity targets in study design [68]. While larger sample sizes are best at increasing statistical power, study-dependent cost and practicality limitations may impede efforts to solicit more samples. Thus, while prospective power analysis is preferred, retrospectively applying power analysis is an option. Previous aptamer-based studies have also leveraged effect sizes (Spearman coefficients and Huber M-Values) to balance small sample sizes [69]. Pilot aptamer studies with limited cohort numbers have depended upon previous reports to estimate experimental models' area under the receiver operating characteristic curves to achieve requisite minimal power levels of 0.8 [70,71]. However, high-throughput strategies such as aptamer assays process an incredible multiplex of data in the discovery stage such that study design must consider the multiple hypothesis testing problems [72]. Addressing this requires proactive consideration of sample size and statistical analysis methodology to maintain high specificity and sensitivity. While analyses employing platforms with protein analysis capacity in the thousands, e.g., 1.3K, 4K, 7K protein panels [27,61,73], can reveal hundreds of biomarkers for a single disease of interest, the disconnect with translating these biomarkers further down the proteomic pipeline emphasizes the critical need for adequately powered study designs [27,74]. The disconnect is augmented by differential protein expression with respect to the number of samples affected by a significant signal and the extent of the signal itself across samples. These aspects will be discussed in the sections that follow.

2.2. Samples and Study Designs: Case–Control and Cohort Studies

Two of the most common choices for designing large-scale proteomic discovery studies are case–control and cohort studies. Case–control [75] enables researchers to access pre-existing biorepositories for clinical samples along with groups that match the disease of interest. However, one of the biggest hurdles facing case–control studies is the inability to compare the disease groups with healthy population controls matched for confounders, such as age, sex, disease-specific factors, and comorbidities. Matched (individual or frequency) case–control studies represent one route of generating such controls but can introduce new biases. The proteome is particularly vulnerable to conditions such as organ trauma, autoimmune disease and age, in particular, as pertaining to the analysis of samples originating from adults or a pediatric cohort as well as ethnicity and sex differences. While steps are being taken to employ reference normalization that accounts for some conditions and demographic aspects using robust reference normalization, this is an area of active evolution in aptamer datasets. Comparisons between matched and unmatched designs report minimal outcome distinctions aside from a slight decrease in statistical power for matched samples [76]. Care must be taken to address and reduce the selection bias that commonly occurs in single-institution case–control studies. For matched studies, conditional logistical regression is the mainstay analysis tool, but unconditional logistical regression has also been simulated to mirror these results [77] closely. A less common primary approach involves cohort studies that struggle with extended time frames, costs, and incomplete datasets due to patient dropouts. However, cohort studies may play a vital role later in aptamer-guided biomarker discovery due to their ability to access pre-existing longitudinal datasets for independent cohort cross-referencing [78]. Assay datasets can be optimized for analysis, ensuring normalization for inter- and intraplate variance with internal controls. Hybridization normalization, well-to-well standardization, signal calibration, and standard operating procedures support reproducibility. Thus, differential expression can then be determined based upon comparisons of relative fluorescence unit (RFU) ratios like multiplex

bead-based assay data. The following section highlights fundamental tools in transforming these large datasets into workable biological knowledge and translatable hypotheses.

3. Statistical Strategies for Analyzing Differential Expression

Aptamer-based platforms hold great promise within the field of shotgun proteomics and carry similar considerations as MS-based platforms in categorizing differential expression with the added advantage of high-throughput analysis of multiple samples simultaneously. Multimarker-based approaches usually tend to overfit, leading to conceptual errors. Statistical analyses should include an adjustment for multiple testing. Multiple biomarkers can be combined in a classifier that outperforms a single biomarker, and validation in an independent sample is imperative, which gives more confidence in the results. Failures of small studies to detect biomarkers often result from variability that interferes with determining effect sizes. Therefore, an increase in both samples and biomarkers contributes to developing multimarker classifiers with enhanced accuracy. Prospective-based studies striving for aptamer diagnostics require an intricate consideration for sample size, sensitivity, and specificity of the targets in study design [33]. A protein may exist in multiple forms within a cell or cell type. These protein isoforms originate from transcriptional, post-transcriptional, translational, post-translational, regulatory, and degrading and preserving processes that affect protein structure, localization, function, and turnover. The field has thus evolved to include a variety of methods for the separation of complex protein samples followed by identification using proteomics technology. It is inherently a systems science that considers protein abundances in a cell and the interplay of proteins, protein complexes, signaling pathways, and networks. In order to address the relevant challenges, the analytical tools can be categorized into four types: (1) quality control, (2) fundamental statistical analysis, (3) machine learning (ML) approaches, and (4) assignment of functional and biological information to describe and understand protein interaction networks [79].

3.1. Quality Control and Basic Statistics

Quality control (Table 2) is employed to observe the data variability, compare means between groups, and look for any anomaly that could cause a problem in the analysis. Quality control can identify significant areas of concern and flag samples as well as more specifically diagnose sources of potentially aberrant data signals; however, detailed capture and annotation is required at all levels from sample collection to analysis to ensure that robust data are obtained and analyzed for confounding features. Basic statistics are a critical first pass to identify the "low-hanging fruit" in the dataset. Methods such as the Student's *t*-test and its nonparametric equivalent, the Wilcoxon test, univariate or analysis of variance (ANOVA), or the nonparametric Kruskal–Wallis test are applied to identify the significant proteins. Due to inherent variability, statistics alone are often insufficient to discover the most biologically relevant information in a proteomic dataset but they are an essential first step in every analysis. Statistically significant results are helpful as seed data or bait in machine learning approaches.

3.2. Machine Learning Approaches

Classification by ML (Table 2) complements traditional statistics as it allows for consideration of many variables at once and removes much of the technical bias. Dataset complexity is reduced, as correlations and trends are identified that may not be visible or may be undetectable using traditional statistics, e.g., clustering using iterative subsampling. Given unbiased data inputs, ML classification has the potential to be unbiased by revealing patterns within data that may or may not relate to the original hypothesis. The researcher is then able to examine the clustering or classification results for new biological features that were not initially predicted. Thus, ML together with network tools enable hypothesis generation, as they uncover the real biology of the system in question. Swan et al. [94] discussed the benefit of ML methods for application to proteomic data and show that machine learning methods give an overall view of data and offer a large

potential for identifying relevant information among data. While ML approaches have been more extensively employed to advance aptamer discovery [95], the use of ML, deep learning (DL), and artificial intelligence (AI) applied to large-scale data originating from aptamer studies to arrive at clinically meaningful and relevant conclusions is still in its beginnings. Broadly, artificial intelligence methods can involve classic ML, using techniques such as support vector machines (SVM) and random forests (RF). Alternatively, DL with convolutional neural networks (CNNs) and hybrid techniques of both classical ML and DL may be employed [96]. Generally, AI methods may be supervised (where the model is told the outcome of interest) or unsupervised, where the model does not know the outcome [97]. Supervised approaches require the data to be divided into categories and training and testing sets, with the model trained on a portion of the data and tested on the remainder. As employed in DL, neural network models do not typically require annotation, but the process by which the model arrives at the results may be difficult to elucidate, i.e., a "black box." ML is currently evolving to analyze aptamer technology-generated data in conjunction with other data types, including clinical, imaging, pathology, and other omics data. A plethora of proteomic alterations are identified using aptamer technology with multiple and many as yet poorly understood signals; hence, aptamer data can lend itself to AI approaches to connect results to clinical meaning. Discussion is ongoing on the optimal means of analysis. Although specific examples of AI as applied specifically to aptamer technology-derived data are scant, the literature originating in MS and RNA sequencing offers more in-depth explorations [97], and in non-aptamer technology, proteomic data [98] parallels exist. Specific known protein alterations may be identified, analyzed, and then extrapolated to other related proteins using known or evolving signaling pathways. Alternatively, emphasis can be placed on filtering out pertinent signals using artificial intelligence approaches and then connecting these to known and unknown proteins and clinical data. Currently, there is a lack of ground truth in aptamer data, limiting the ability to validate findings and train DL methods that are traditionally data-hungry. There is also a lack of standard datasets to provide a reliable comparison for abnormal samples. Nonetheless, creating clinical connections using aptamer data is undergoing active progress across medical disciplines [35,99–101]. The goals of aptamer-based data intersecting with AI currently focus largely on diagnosis [35,99,100] evolving into response assessment [101], with few publications exploring ML to examine management or prognosis. In a diagnostic example, using urine samples, Dong et al. employed the SOMAscan platform to identify culture-positive urine samples in the urine of 16 children with urinary tract infections. ML with SVM based feature selection was performed in this study to determine the combination of urine biomarkers that optimized diagnostic accuracy [99]. The authors found that eight candidate urine protein biomarkers met filtering criteria resulting in area under the receiver operating characteristic curves (AUCs) ranging from 0.91 to 0.95, with the best prediction achieved by the SVMs with a radial basis function kernel [99]. In the context of arthritis, the serum proteome for patients with psoriatic arthritis and patients with rheumatoid arthritis was analyzed using nano-liquid chromatography–mass spectrometry (nano-LC–MS–MS), SOMAscan, and Luminex, and multivariate ML was employed on the data from all three platforms to separate patients with early-onset inflammatory arthritis to differentiate psoriatic and rheumatoid arthritis [100]. In the context of sleep apnea, Ambati et al. employed the Obstructive Apnea Hypopnea Index (OAHI), the Central Apnea Index (CAI), the 2% Oxygen Desaturation Index, and mean and minimum oxygen saturation indices during sleep to train a machine learning classifier using a SOMAscan 1.3K assay and achieved 76% validation accuracy [35]. Hewitson et al. also used SOMAscan and machine learning to identify nine proteins that were significantly different in autism spectrum disorder vs. typically developing boys, although the authors acknowledged that further verification with independent test sets is warranted [102]. In an example of aptamer data and AI applied towards response assessment, O'Neil et al. studied clinical remission in rheumatoid arthritis using 130 patient serum samples on a 1.3K SOMAscan platform. They employed unsupervised hierarchical clustering and supervised classifi-

cation to identify proteomic-driven clusters for model biomarkers associated with future disease flare after 12 months of follow-up and medication withdrawal. Network analysis was employed to define pathways that were enriched in proteomic datasets. The authors found that clustering did not predict future risk of flare, while the XGboost machine learning algorithm classified patients who relapsed with an AUC of 0.80 using only baseline serum proteomics [101]. Machine learning and AI methods are actively evolving, and their application to aptamer-based data is currently limited but growing as the technology is more widely applied. Most methods encountered currently involve classical ML and significant annotation. In the absence of robust controls and significant annotation, ML has been more often applied to aptamer technology-derived data in the context of diagnosis, as discussed above. Therapy-related questions require robust controls, which are difficult to obtain as patients undergo heterogenous management over time in most settings, which is expected to modify the proteome.

Table 2. Statistical methods for aptamer analysis.

Quality Control		
Student's *t*-test or nonparametric Wilcoxon	Mean difference between two groups	[80]
ANOVA or nonparametric Kruskal–Wallis	Variation between two or more groups	[81]
Visualization methods	Histogram, density plots, box and bar graphs	
Classification		
Principle Component Analysis (PCA)	Dimension reduction, separates groups based upon commonality	[82]
Independent Component Analysis (ICA)	Dimension reduction, separates groups based upon correlation	[83]
Partial Least Squares (PLS)	Discriminant analysis that separates groups by maximum covariation ranks the important features	[84]
Random Forest (RF)	Separates groups by similarity, ranks important features	[85]
Support-Vector Machine (SVM)	Classifies the sample by kernel function	[86]
Clustering		
K-means	Clustering of features or samples into user-specified numbers of clusters	[87]
Hierarchical	Unsupervised classification of features, samples, or any endpoint by dendrogram	[88]
Pathways		
Gene Set Enrichment Analysis (GSEA)	Pathway analysis and functional annotation	[89]
Ingenuity Pathway Analysis (IPA)	Pathway and functional annotation from curated databases	
Database for Annotation, Visualization, and Integrated Discovery (DAVID)	Pathway and functional annotation using Gene Ontology (GO)	[90]
Cytoscape	Network analysis visualization	[91]
Kyoto Encyclopedia of Genes and Genomes (KEGG)	Pathway analysis	[92]
Human Annotated and Predicted Protein Interactions (HAPPI)	Protein interactions	[93]

3.3. Pathway Analysis

Pathway analysis (Table 2) following statistical analysis, classification, and clustering can help organize a long list of proteins onto a short list of pathway knowledge maps, easing interpretation of the molecular interplay. The machine learning and clustering tools of omics data can be categorized into a supervised and unsupervised classification for seven popular types of machine learning: principal component analysis (PCA), independent component analysis (ICA), K-means, hierarchical clustering, partial least squares (PLS), random forests (RF), and SVMs. These methods are also summarized and compared in Table 2, which provides an overview of different classifications and clustering tools and how to select a method most likely to be effective for a specific dataset. The intersection of high-throughput, high-multiplex proteomic datasets, existing omics databases, and clinical features results in a rich systems biology analysis to better understand biological pathways and functional gene networks. Seated within systems biology is pathway analysis. Pathway analysis facilitates future hypothesis generation from high-throughput microarray data, localizing gene networks, and framing protein differential expression into

meaningful nodes and modules. Understanding how aptamer-identified proteins operate interconnectedly can support biomarker identification and further the identification of aberrant biological pathways in disease [103]. The underlying premise of pathway analysis involves preparing and standardizing protein differential expression from aptamer data (typically evaluated via fold-changes) [104], performing statistical analysis of relevant proteins, and applying a pathway database. There are three generations of pathway analysis: over-representation analysis (ORA), functional class scoring (FCS), and pathway topology (PT) [105]. ORA was the first and most simplistic form of pathway analysis. After initial statistical analysis, aptamer datasets reveal a list of statistically significant proteins that are either over- or under-expressed. Accompanying these proteins is a list of genes inputted into ORA to measure the most over-expressed genes via hypergeometric analysis, producing over- or under-represented pathways based upon a previously selected FDR. ORA was one of the first widely used pathway analysis strategies and has even been utilized in aptamer-based studies of the SARS-CoV-2 virus [46]. However, this inherent independent-based assumption of genes has generated high false-positive rates due to the correlations between genes that are ignored in ORA [106]. Furthermore, the arbitrary cutoff threshold can significantly impact the conclusions drawn [107]. Thus, while ORA presents a simple, cost-effective option for analyzing aptamer datasets, the fundamental limitations of single-set gene analysis restrict the technique's robustness. Functional class scoring, a second-generation approach, utilizes a three-step process of computing gene-level statistics, compiling them into pathway-level statistics, and assessing for statistical significance [105]. FCS overcomes the arbitrary thresholds utilized in ORA and further takes into consideration subtler changes and impacts of coordinated networks. However, recent studies have faulted FCS for its lack of specificity [108]. Gene set enrichment analysis (GSEA) is one of the most popular gene set analysis techniques under FCS and it has seen widespread use in aptamer studies, including discoveries in Duchene muscular dystrophy [109], myocardial infarction [47], and myeloid leukemia [110]. GSEA leveraged past research and was one of the first methodologies to focus analysis on deriving an understanding of gene sets [89]. GSEA has also given rise to parametric analysis of gene set enrichment (PAGE), which potentially offers a more sensitive analysis while avoiding the rigorous computational effort required in GSEA [111]. A third method known as generally applicable gene set enrichment (GAGE) has also emerged to tackle datasets of different sample sizes, which may be more applicable towards cross-validating aptamer studies of different experimental designs or methodologies [112]. Cross-comparisons of analysis techniques suggested that GAGE had the highest reproducibility and predicted the most relevant gene sets [113]. Finally, the third-generation approach known as pathway topology (PT) mimics the three-step process of FCS but applies pathway topology in computing gene-level statistics [105]. PT can consider the interactions between genes and avoid the independency assumptions about genes in FCS. A recent characterization of all pathway analysis methods gives PT-based approaches a slight edge over non-PT-based approaches in relation to real-world data [114]. Since the onset of pathway analysis, dedicated pathway analysis databases have emerged that support the distinct subcategorizations of pathway analysis described above. Popular databases are highlighted in Table 2, including Gene Set Enrichment Analysis (GSEA) [89], Ingenuity Pathway Analysis (IPA), the Database for Annotation, Visualization, and Integrated Discovery (DAVID) [90], Cytoscape [91], the Kyoto Encyclopedia of Genes and Genomes (KEGG) [92], and Human Annotated and Predicted Protein Interactions (HAPPI) [93]. A full comparison of current databases was recently conducted by Chowdhury and Sarkar [115]. One of the most comprehensive platforms currently available is QIAGEN's IPA program that features support for protein interactions, metabolic data, gene regulation, and sequencing [116]. IPA has facilitated extensive pathway analysis in aptamer datasets in atrial fibrillation [117], surgical procedures [48], and aging [118]. While pathway analysis is a fundamental step in deriving meaning from the differentially expressed proteins derived from aptamer technology, it is incumbent on researchers to understand this area's current challenges and limitations. First, pathway

analysis and databases rely on a body of published literature that continuously evolves and adapts. Second, meta-analyses of the published literature suggest that poor concordances exist based upon the type of pathway analysis selected and the database utilized [119,120]. As a result, care must be taken in selecting the best suitable methodology and database based upon the study design. The ability to create and validate pathway-specific reference panels for different clinical contexts while employing aptamer-based data will improve the ability to design studies as proteomic "controls" become more prevalent.

4. Multi-Omics Approaches and Verification

4.1. Proteomic Quantitative Trait Loci (pQTL)

While pathway analysis provides an excellent canonical interpretation of protein groups, the next step in verifying aptamer microarrays' proteomic data is to establish genetic anchorage for identified proteins. One of the emerging methods to accomplish this is through proteomic quantitative trait loci (pQTL), which evaluates the variance in proteins attributable to specific loci. pQTL has recently supplanted expression quantitative trait loci (eQTL) due to the latter's dependency on the poor relationship between mRNA expression and protein levels, as well as a reported disconnect of roughly 50% between pQTL and eQTL [121,122]. The incredible scalability of aptamer microarrays enables studies to connect large-scale proteomics data with pQTLs. Recent applications of aptamer microarrays have already been made in studying the proteome of patients infected with the recent SARS-CoV-2 virus [123]. In turn, pQTLs can substantiate cross-platform findings between aptamer platforms and others which can impact future translational capabilities of studies. Pietzner et al. reported several factors leading to aptamer-specific pQTL findings, such as lower observational correlations, lower binding affinities of aptamers, and extreme datapoints that merit consideration in the verification process [33]. pQTL takes on a fundamental role in verifying proteomic relationships hypothesized from discovery-stage analysis. After multi-fold reductions take place to isolate a small number of differentially expressed proteins, these proteins must be properly validated before being passed on as potentially viable biomarkers. One of the primary benefits of applying pQTL analysis is to establish an integrative understanding of causal networks and pathways by combining large-scale databases of both aptamers and genome-wide association studies data (GWAS) [124]. Ferkingstad et al.'s 4.9K SOMAscan protein assay identified nearly 18,000 associations of sequence variants and pQTLs and applied a multi-omics method of proteomics, transcriptomics, and genomics towards a large-scale aptamer database [125]. Furthermore, the authors emphasized associations of variants with high LD or *cis* (near the gene of interest) pQTLs in pursuit of developing drug targets [125]. A separate large-scale SOMAscan-based study (INTERVAL) also previously established the role of Mendelian randomization analysis for applying aptamer-based proteomic datasets towards discovering causal protein biomarkers [126]. Studies have also applied aptamer datasets and pQTLs with GWAS, mRNA, and eQTL datasets to link proteomic variability with genetic components in a cohort of irritable bowel syndrome patients [127]. Population samples have also been utilized in combining aptamer datasets with pQTLs to demonstrate utility and verification [128]. Aptamer platforms have also supported large-scale curated datasets which have been employed in "virtual proteomics" for clinical prediction and biomarker discovery [129].

4.2. GWAS and PWAS

Protein-wide association studies (PWAS) represent a newly emerging analysis tool in the multi-omics landscape that attempts to establish an understanding of combining GWAS with protein functions and phenotypes. Simulation testing suggests that PWAS is better for causal relationship analysis, has a reduced computational burden, is complementary to SKAT (GWAS), and can find PWAS-exclusive genes [130]. Applied to large-scale databases for Alzheimer's disease, PWAS uncovered additional AD genes of interest not found through traditional GWAS [131]. Summary data with GWAS can be performed using sum-

mary data-based Mendelian randomization and heterogeneity in dependent instruments (HEIDI) [132]. This growing integration between aptamer datasets, genomics, transcriptomics, and proteomics will continue to evolve as aptamer technology matures and scales in magnitude. Significant effort is directed at identifying translatable protein biomarkers while establishing a robust genetic anchorage for differentially expressed proteins, with the understanding that doing so is vital for clinical impact with eventual prospects for FDA approval [133].

5. Translational Challenges of Aptamer Proteomics

While the technical capabilities of aptamers with their flexibility, high multiplexing, and strong affinities open a wealth of opportunities both diagnostically [134] and therapeutically, only a minimal number of studies have seen success clinically. This review centers around examining data arising from aptamer-based assays used in biomarker and clinical diagnostics. Although SOMAscan remains one of the most popular platforms for high-throughput proteomic analysis of biological matrices, other proteomic aptamer platforms continue to be developed, such as ProtSeq [10], protein precipitation assays [135], and disease-specific panels [136,137]. However, aptamer-based proteomics has yet to be promoted to mainstream prominence as in the case of mass spectrometry platforms. We predict that the growing capabilities of aptamers will lead to a continual acceleration of aptamer platforms that have only recently been developed. As evidenced by mass spectrometry proteomics, aptamer-based proteomics will likely face similar challenges. Here, we cover challenges in proteomic biomarker translation and cross-platform consistency.

5.1. FDA Approval and Clinical Translation

Aptamer-based platforms embody the beginning stages of the traditional biomarker discovery pipeline (discovery, verification, and validation), sharing a position with mass spectrometry [138] (Figure 2). One of the common challenges in proteomic biomarker discovery is the poor translation into the clinic, with an extensive track record of biomarkers fizzling out during FDA approval [139]. While the high multiplexing capability of aptamers broadens the survey of the human proteome, the sheer number of proteins and their varying abundances make detecting disease-specific variations a challenging feat. As aptamer assays continue to broaden over time, a more encompassing picture may be drawn at the cost of higher false discovery rates. With the arduous task of achieving FDA approval [140], false discovery rates of inaccurate biomarkers come at a high cost. While adequately designed statistical analysis and sufficiently powered studies may address this, single protein-based biomarkers will merit a considerable effort to achieve approval. Multi-marker panels serve as a more robust substitute [141,142]. Two significant advances in proteomic biomarker discovery in the last decade have been the FDA approval of the OVA1 [143] and CKD273 [144] panels. The CKD273 panel was based upon capillary electrophoresis coupled to mass spectrometry (CE–MS) technology, which supports the throughput of thousands of peptides that strongly supported earlier stages of research [145]. Building large, comparative datasets that feature large sample sizes can greatly support discovery-stage proteomics. The high precision, throughput, and multiplex aspects of aptamers facilitate this goal. It is paramount for studies to consider analytical factors for effective translational hypothesis generation to follow. Current ambitions for blood-based diagnostics and liquid biopsies in cancer [146] will continue to fuel a strong interest in aptamer-identified biomarkers. Clinical translation, however, will depend on the ability to transfer findings between the clinic and bench research, which will require that subsets of large-scale aptamer-based data panels be transferrable between species to allow for findings to be replicated in laboratory animals and human samples to advance outcomes predicated on alteration in management.

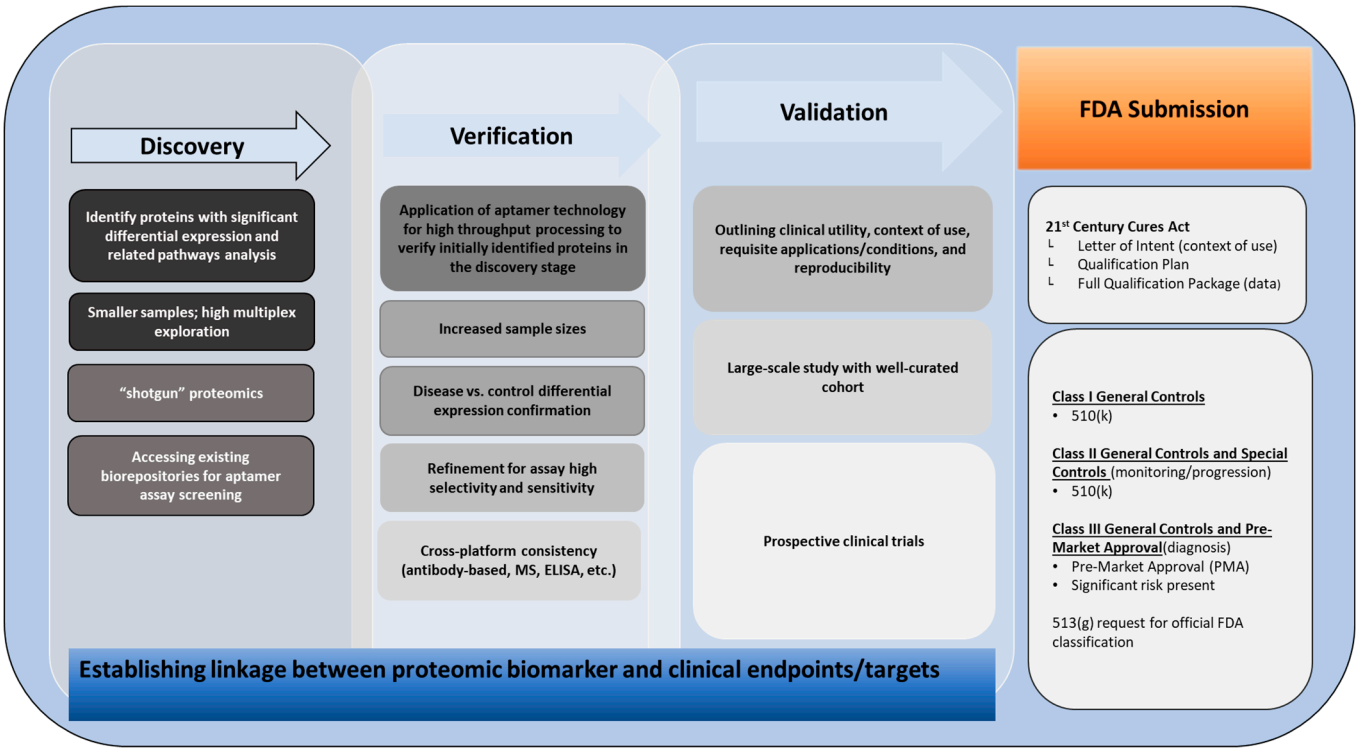

Figure 2. Process map advancing large-scale proteomic datasets from discovery to verification to validation towards identifying clinically meaningful biomarkers for FDA submission.

5.2. Cross-Platform Consistency

While aptamers can generate a large dataset based upon only a few clinical samples, these data may not substitute for a multi-omics approach. Single-omics analysis is an inherent limitation of aptamer assays, as protein expression levels may not be an accurate indicator with transcript levels and other omics data [147]. Previous studies have suggested a poor correlation between aptamer- and antibody-based platforms, which may limit the biomarker discovery potential of aptamer-based technology [148]. Cross-platform comparisons across the same cohorts have found wide-ranging correlations [149]. However, poor concordance may not suggest inefficacy. While inter- and intra-platform variations exist, a recent study in cardiac patients reported that all statistically significant proteins identified via aptamers were similarly identified in immunoassays [150]. Part of this variation may be attributable to aptamer technology's intrinsic ability to measure more extensive dynamic ranges not accounted for in techniques such as mass spectrometry [151]. Additionally, reproducibility studies with aptamers and ELISA have suggested a relatively high concordance rate between the two technologies [49]. Reports in the literature suggest that, despite some differences occurring between platforms, aptamers and immunoassays essentially point towards similar biomarker–disease associations, with higher biomarker concentrations leading to more robust findings [150]. While these studies are encouraging, cross-platform concordance remains a limitation of aptamer-based data. Thus, while care should be taken in drawing cross-platform comparisons and interpreting meta-analyses, aptamer technology should not be interpreted in isolation but rather seen as a complementary tool to current proteomic technology.

5.3. Intra-Platform Consistency

Intra-platform consistency should also be considered, for instance, as earlier versions of the SOMAscan assay specific for approximately 1300 or 4100 [73] protein targets may not be compared directly with the latest assay version specific for approximately 7000 protein targets. Though expression trends may be comparable, the raw numerical values for

each specific aptamer may change across assay versions due to technical variables, and each specific assay does not necessarily include the same cohort of aptamers, which has implications for efforts to leverage for comparison the findings from previous clinical studies which have utilized different versions of the assay. This phenomenon is also not limited to the SOMAscan technology, as other multiplexed proteomics platforms of all denominations (1.3K, 4K, 7K) [48,51,61,73] face the same challenges of compatibility across datasets as assay specifications change over time. Of note, a major Human Protein Atlas study tested over 9000 internally generated antibodies using immunoassays, finding that half failed to match literature results and only 7% demonstrated strong concordance with the literature [152]; similar challenges are anticipated in the context of aptamer-based data.

6. Conclusions

Aptamer technology generates extremely large datasets that are growing in number and popularity. The ability to harness a small amount of biospecimen and the potential for clinical applicability and aggregation with existing and evolving pathway analysis options make large scale aptamer-based datasets particularly attractive to researchers in all areas of medicine. Ongoing emphasis is being placed on study design and statistical considerations for analysis. However, unravelling the complexity of the human proteome will continue to pose a substantial challenge to translating large datasets into clinical utility. Practical guidance for planning aptamer-based studies includes:

(1) Study design The study design for aptamer-derived data is understandably contingent on the disease and clinical context. Maximal benefit is elicited if the study benefits from maximal data capture and annotation to ensure that potential confounders in the proteomic data signals can be addressed down the line. Controls are crucial for meaningful comparison. Controls may not necessarily represent a normal population but rather a population whose proteome in comparison to that of the intervention will allow the researcher to derive conclusions that will address the hypothesis being tested. Cohort studies may be employed with a "before" and "after" sample being obtained from the same patient, thus using the patient as their own control. Provided robust large-scale data are available and, following a thorough review, are sufficiently comparable to the study population, this is a reasonable option, with the study design in this context benefitting greatly from collaboration amongst researchers.
(2) Statistical analysis Statistical analysis for aptamer-derived data is actively evolving. Traditional approaches described here may be employed. ML approaches may be used with semi-supervised approaches likely to be the most successful and contingent on annotation of the data.
(3) An established team that combines researchers, statisticians, and clinicians will need to maintain a close relationship in the planned acquisition, curation, annotation and analysis of data to allow for meaningful translation into the clinic and advancement in relation to patient outcomes.

Author Contributions: Conceptualization, A.V.K. and W.J.; methodology, W.J. and A.V.K.; resources, J.C.J., M.S., U.S. and K.C.; writing—original draft preparation, W.J., M.S., U.S. and A.V.K.; writing—review and editing, J.C.J., K.C. and A.V.K.; visualization, W.J. and A.V.K.; supervision, A.V.K. and K.C.; project administration, A.V.K.; funding acquisition, A.V.K. and K.C. All authors have read and agreed to the published version of the manuscript.

Funding: Funding was provided in part by the NCI NIH intramural program (ZID BC 010990).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ANOVA	Analysis of Variance
AI	Artificial Intelligence
CV	Coefficient of Variations
CNN	Convolutional Neural Networks
DDA-MS	Data-Dependent Mass Spectrometry
DL	Deep Learning
DIA-MS	Data-Independent Acquisition Mass Spectrometry
ELISA	Enzyme-Linked Immunosorbent Assays
eQTL	Expression Quantitative Trait Loci
FDA	U.S. Food and Drug Administration
FDR	False Discovery Rate
FCS	Functional Class Scoring
GAGE	Generally Applicable Gene set Enrichment
GSEA	Gene Set Enrichment Analysis
GWAS	Genome-Wide Association Studies
HEIDI	Heterogeneity in Dependent Instruments
ICA	Independent Component Analysis
IPA	QIAGEN's Ingenuity Pathway Analysis
LC–MS	Liquid Chromatography–MS
ML	Machine Learning
MS	Mass Spectrometry
ORA	Over-Representation Analysis
PCA	Principal Component Analysis
PLS	Partial Least Squares
pQTL	Proteomic Quantitative Trait Loci
PT	Pathway Topology
PTM	Post-Translational Modification
PWAS	Protein-Wide Association Studies
RF	Random Forest
RFU	Relative Fluorescence Unit
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
SVM	Support Vector Machine

References

1. Tighe, P.J.; Ryder, R.R.; Todd, I.; Fairclough, L.C. ELISA in the multiplex era: Potentials and pitfalls. *Proteom. Clin. Appl.* **2015**, *9*, 406–422. [CrossRef]
2. Cravatt, B.F.; Simon, G.M.; Yates, J.R., 3rd. The biological impact of mass-spectrometry-based proteomics. *Nature* **2007**, *450*, 991–1000. [CrossRef]
3. Beer, I.; Barnea, E.; Ziv, T.; Admon, A. Improving large-scale proteomics by clustering of mass spectrometry data. *Proteomics* **2004**, *4*, 950–960. [CrossRef]
4. Mann, M.; Jensen, O.N. Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **2003**, *21*, 255–261. [CrossRef]
5. Krasny, L.; Huang, P.H. Data-independent acquisition mass spectrometry (DIA-MS) for proteomic applications in oncology. *Mol. Omics* **2021**, *17*, 29–42. [CrossRef]
6. Nesvizhskii, A.I.; Aebersold, R. Interpretation of shotgun proteomic data: The protein inference problem. *Mol. Cell Proteom.* **2005**, *4*, 1419–1440. [CrossRef]
7. Courcelles, M.; Bridon, G.; Lemieux, S.; Thibault, P. Occurrence and detection of phosphopeptide isomers in large-scale phosphoproteomics experiments. *J. Proteome Res.* **2012**, *11*, 3753–3765. [CrossRef]
8. Chick, J.M.; Kolippakkam, D.; Nusinow, D.P.; Zhai, B.; Rad, R.; Huttlin, E.L.; Gygi, S.P. A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. *Nat. Biotechnol.* **2015**, *33*, 743–749. [CrossRef]
9. SomaLogic. SOMAscan Proteomic Assay: Technical White Paper. Available online: https://www.somalogic.com/wp-content/uploads/2016/08/SSM-002-Rev-3-SOMAscan-Technical-White-Paper.pdf (accessed on 2 February 2022).
10. Hong, J.M.; Gibbons, M.; Bashir, A.; Wu, D.; Shao, S.; Cutts, Z.; Chavarha, M.; Chen, Y.; Schiff, L.; Foster, M.; et al. ProtSeq: Toward high-throughput, single-molecule protein sequencing via amino acid conversion into DNA barcodes. *iScience* **2021**, *25*, 103586. [CrossRef]

11. Wik, L.; Nordberg, N.; Broberg, J.; Björkesten, J.; Assarsson, E.; Henriksson, S.; Grundberg, I.; Pettersson, E.; Westerberg, C.; Liljeroth, E.; et al. Proximity Extension Assay in Combination with Next-Generation Sequencing for High-throughput Proteome-wide Analysis. *Mol. Cell Proteom.* **2021**, *20*, 100168. [CrossRef]
12. Coarfa, C.; Grimm, S.L.; Rajapakshe, K.; Perera, D.; Lu, H.-Y.; Wang, X.; Christensen, K.R.; Mo, Q.; Edwards, D.P.; Huang, S. Reverse-Phase Protein Array: Technology, Application, Data Processing, and Integration. *J. Biomol. Tech.* **2021**, *32*, 15–29. [CrossRef]
13. Gupta, V.; Zimmerman, R.; Zhan, T.; Hamilton, T.; Na, L.; Peng, J. Development and Validation of Bio-Plex Pro Human Chemokine Assays. 2014. Available online: https://www.bio-rad.com (accessed on 2 February 2022).
14. Wang, X.; Walt, D.R. Simultaneous detection of small molecules, proteins and microRNAs using single molecule arrays. *Chem. Sci.* **2020**, *11*, 7896–7903. [CrossRef]
15. Puscasu, A.; Zanchetta, M.; Posocco, B.; Bunka, D.; Tartaggia, S.; Toffoli, G. Development and validation of a selective SPR aptasensor for the detection of anticancer drug irinotecan in human plasma samples. *Anal. Bioanal. Chem.* **2021**, *413*, 1225–1236. [CrossRef]
16. Adhikari, M.; Strych, U.; Kim, J.; Goux, H.; Dhamane, S.; Poongavanam, M.-V.; Hagström, A.E.V.; Kourentzi, K.; Conrad, J.C.; Willson, R.C. Aptamer-Phage Reporters for Ultrasensitive Lateral Flow Assays. *Anal. Chem.* **2015**, *87*, 11660–11665. [CrossRef]
17. Hwang, B.B.; Engel, L.; Goueli, S.A.; Zegzouti, H. A homogeneous bioluminescent immunoassay to probe cellular signaling pathway regulation. *Commun. Biol.* **2020**, *3*, 8. [CrossRef]
18. Bastarache, J.A.; Koyama, T.; Wickersham, N.E.; Ware, L.B. Validation of a multiplex electrochemiluminescent immunoassay platform in human and mouse samples. *J. Immunol. Methods* **2014**, *408*, 13–23. [CrossRef]
19. Magdeldin, S.; Enany, S.; Yoshida, Y.; Xu, B.; Zhang, Y.; Zureena, Z.; Lokamani, I.; Yaoita, E.; Yamamoto, T. Basics and recent advances of two dimensional- polyacrylamide gel electrophoresis. *Clin. Proteom.* **2014**, *11*, 16. [CrossRef]
20. Whiteaker, J.R.; Lin, C.; Kennedy, J.; Hou, L.; Trute, M.; Sokal, I.; Yan, P.; Schoenherr, R.M.; Zhao, L.; Voytovich, U.J.; et al. A targeted proteomics–based pipeline for verification of biomarkers in plasma. *Nat. Biotechnol.* **2011**, *29*, 625–634. [CrossRef]
21. Collins, B.C.; Hunter, C.L.; Liu, Y.; Schilling, B.; Rosenberger, G.; Bader, S.L.; Chan, D.W.; Gibson, B.W.; Gingras, A.-C.; Held, J.M.; et al. Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *Nat. Commun.* **2017**, *8*, 291. [CrossRef]
22. Sandberg, A.; Branca, R.M.M.; Lehtiö, J.; Forshed, J. Quantitative accuracy in mass spectrometry based proteomics of complex samples: The impact of labeling and precursor interference. *J. Proteom.* **2014**, *96*, 133–144. [CrossRef]
23. Chambers, A.G.; Percy, A.J.; Yang, J.; Borchers, C.H. Multiple Reaction Monitoring Enables Precise Quantification of 97 Proteins in Dried Blood Spots. *Mol. Cell Proteom.* **2015**, *14*, 3094–3104. [CrossRef] [PubMed]
24. Tai, S.S.C.; Welch, M.J. Development and Evaluation of a Candidate Reference Method for the Determination of Total Cortisol in Human Serum Using Isotope Dilution Liquid Chromatography/Mass Spectrometry and Liquid Chromatography/Tandem Mass Spectrometry. *Anal. Chem.* **2004**, *76*, 1008–1014. [CrossRef] [PubMed]
25. Ozyurt, C.; Canbay, Z.C.; Dinckaya, E.; Evran, S. A highly sensitive DNA aptamer-based fluorescence assay for sarcosine detection down to picomolar levels. *Int. J. Biol. Macromol.* **2019**, *129*, 91–97. [CrossRef] [PubMed]
26. BasePair. Available online: https://www.basepairbio.com/ (accessed on 2 February 2022).
27. Singh, K.; Cheung, B.M.; Xu, A. Ultrasensitive detection of blood biomarkers of Alzheimer's and Parkinson's diseases: A systematic review. *Biomark. Med.* **2021**, *15*, 1693–1708. [CrossRef] [PubMed]
28. Ellington, A.D.; Szostak, J.W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **1990**, *346*, 818–822. [CrossRef]
29. Tuerk, C.; Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **1990**, *249*, 505–510. [CrossRef]
30. Liu, Q.; Zhang, W.; Chen, S.; Zhuang, Z.; Zhang, Y.; Jiang, L.; Lin, J.S. SELEX tool: A novel and convenient gel-based diffusion method for monitoring of aptamer-target binding. *J. Biol. Eng.* **2020**, *14*, 1. [CrossRef]
31. Brody, E.N.; Willis, M.C.; Smith, J.D.; Jayasena, S.; Zichi, D.; Gold, L. The use of aptamers in large arrays for molecular diagnostics. *Mol. Diagn.* **1999**, *4*, 381–388. [CrossRef]
32. Dhiman, A.; Kalra, P.; Bansal, V.; Bruno, J.G.; Sharma, T.K. Aptamer-based point-of-care diagnostic platforms. *Sens. Actuators B Chem.* **2017**, *246*, 535–553. [CrossRef]
33. Pietzner, M.; Wheeler, E.; Carrasco-Zanini, J.; Kerrison, N.D.; Oerton, E.; Koprulu, M.; Luan, J.; Hingorani, A.D.; Williams, S.A.; Wareham, N.J.; et al. Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. *Nat. Commun.* **2021**, *12*, 6822. [CrossRef]
34. Abatemarco, J.; Sarhan, M.F.; Wagner, J.M.; Lin, J.-L.; Liu, L.; Hassouneh, W.; Yuan, S.-F.; Alper, H.S.; Abate, A.R. RNA-aptamers-in-droplets (RAPID) high-throughput screening for secretory phenotypes. *Nat. Commun.* **2017**, *8*, 332. [CrossRef] [PubMed]
35. Ambati, A.; Ju, Y.E.; Lin, L.; Olesen, A.N.; Koch, H.; Hedou, J.J.; Leary, E.B.; Sempere, V.P.; Mignot, E.; Taheri, S. Proteomic biomarkers of sleep apnea. *Sleep* **2020**, *43*, zsaa086. [CrossRef] [PubMed]
36. Helms, L.; Marchiano, S.; Stanaway, I.B.; Hsiang, T.-Y.; Juliar, B.A.; Saini, S.; Zhao, Y.T.; Khanna, A.; Menon, R.; Alakwaa, F.; et al. Cross-validation of SARS-CoV-2 responses in kidney organoids and clinical populations. *JCI Insight* **2021**, *6*, 1–21. [CrossRef] [PubMed]

37. Tully, B.; Balleine, R.L.; Hains, P.G.; Zhong, Q.; Reddel, R.R.; Robinson, P.J. Addressing the Challenges of High-Throughput Cancer Tissue Proteomics for Clinical Application: ProCan. *Proteomics* **2019**, *19*, e1900109. [CrossRef]
38. Anderson, N.L. The clinical plasma proteome: A survey of clinical assays for proteins in plasma and serum. *Clin. Chem.* **2010**, *56*, 177–185. [CrossRef] [PubMed]
39. Boellner, S.; Becker, K.-F. Reverse Phase Protein Arrays-Quantitative Assessment of Multiple Biomarkers in Biopsies for Clinical Use. *Microarrays* **2015**, *4*, 98–114. [CrossRef] [PubMed]
40. Hennessy, B.T.; Lu, Y.; Gonzalez-Angulo, A.M.; Carey, M.S.; Myhre, S.; Ju, Z.; Davies, M.A.; Liu, W.; Coombes, K.; Meric-Bernstam, F.; et al. A Technical Assessment of the Utility of Reverse Phase Protein Arrays for the Study of the Functional Proteome in Non-microdissected Human Breast Cancers. *Clin. Proteom.* **2010**, *6*, 129–151. [CrossRef]
41. Byron, A. Reproducibility and Crossplatform Validation of Reverse-Phase Protein Array Data. In *Reverse Phase Protein Arrays: From Technical and Analytical Fundamentals to Applications*; Yamada, T., Nishizuka, S.S., Mills, G.B., Liotta, L.A., Eds.; Springer: Singapore, 2019; pp. 181–201. [CrossRef]
42. Brase, J.C.; Mannsperger, H.; Fröhlich, H.; Gade, S.; Schmidt, C.; Wiemann, S.; Beissbarth, T.; Schlomm, T.; Sültmann, H.; Korf, U. Increasing the sensitivity of reverse phase protein arrays by antibody-mediated signal amplification. *Proteome Sci.* **2010**, *8*, 36. [CrossRef]
43. Houser, B. Bio-Rad's Bio-Plex®suspension array system, xMAP technology overview. *Arch. Physiol. Biochem.* **2012**, *118*, 192–196. [CrossRef]
44. Wilson, D.H.; Rissin, D.M.; Kan, C.W.; Fournier, D.R.; Piech, T.; Campbell, T.G.; Meyer, R.E.; Fishburn, M.W.; Cabrera, C.; Patel, P.P.; et al. The Simoa HD-1 Analyzer:A Novel Fully Automated Digital Immunoassay Analyzer with Single-Molecule Sensitivity and Multiplexing. *J. Lab. Autom.* **2016**, *21*, 533–547. [CrossRef]
45. Schubert, S.M.; Arendt, L.M.; Zhou, W.; Baig, S.; Walter, S.R.; Buchsbaum, R.J.; Kuperwasser, C.; Walt, D.R. Ultra-sensitive protein detection via Single Molecule Arrays towards early stage cancer monitoring. *Sci. Rep.* **2015**, *5*, 11034. [CrossRef] [PubMed]
46. Chirinos, J.A.; Cohen, J.B.; Zhao, L.; Hanff, T.; Sweitzer, N.; Fang, J.; Corrales-Medina, V.; Ammar, R.; Morley, M.; Zamani, P.; et al. Clinical and Proteomic Correlates of Plasma ACE2 (Angiotensin-Converting Enzyme 2) in Human Heart Failure. *Hypertension* **2020**, *76*, 1526–1536. [CrossRef] [PubMed]
47. George, M.J.; Kleveland, O.; Garcia-Hernandez, J.; Palmen, J.; Lovering, R.; Wiseth, R.; Aukrust, P.; Engmann, J.; Damås, J.K.; Hingorani, A.D.; et al. Novel Insights Into the Effects of Interleukin 6 Antagonism in Non–ST-Segment&–Elevation Myocardial Infarction Employing the SOMAscan Proteomics Platform. *J. Am. Heart Assoc.* **2020**, *9*, e015628. [CrossRef] [PubMed]
48. Fong, T.G.; Chan, N.Y.; Dillon, S.T.; Zhou, W.; Tripp, B.; Ngo, L.H.; Otu, H.H.; Inouye, S.K.; Vasunilashorn, S.M.; Cooper, Z.; et al. Identification of Plasma Proteome Signatures Associated With Surgery Using SOMAscan. *Ann. Surg.* **2021**, *273*, 732–742. [CrossRef] [PubMed]
49. Han, Z.; Xiao, Z.; Kalantar-Zadeh, K.; Moradi, H.; Shafi, T.; Waikar, S.S.; Quarles, L.D.; Yu, Z.; Tin, A.; Coresh, J.; et al. Validation of a Novel Modified Aptamer-Based Array Proteomic Platform in Patients with End-Stage Renal Disease. *Diagnostics* **2018**, *8*, 71. [CrossRef] [PubMed]
50. MescoScale. Available online: https://www.bioxpedia.com/meso-scale-discovery-immunoassays/ (accessed on 2 February 2022).
51. Günther, A.; Becker, M.; Göpfert, J.; Joos, T.; Schneiderhan-Marra, N. Comparison of Bead-Based Fluorescence Versus Planar Electrochemiluminescence Multiplex Immunoassays for Measuring Cytokines in Human Plasma. *Front. Immunol.* **2020**, *11*. [CrossRef] [PubMed]
52. Overview of ELISA in Relation to Other Disciplines. In *The ELISA Guidebook*; Humana Press: Totowa, NJ, USA, 2009; pp. 1–8. [CrossRef]
53. Bastarache, J.A.; Koyama, T.; Wickersham, N.E.; Mitchell, D.B.; Mernaugh, R.L.; Ware, L.B. Accuracy and reproducibility of a multiplex immunoassay platform: A validation study. *J. Immunol. Methods* **2011**, *367*, 33–39. [CrossRef]
54. Gillet, L.C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteom.* **2012**, *11*, O111.016717. [CrossRef]
55. Wu, Y.; Williams, E.G.; Aebersold, R. Application of SWATH Proteomics to Mouse Biology. *Curr. Protoc. Mouse Biol.* **2017**, *7*, 130–143. [CrossRef]
56. Aggarwal, K.; Choe, L.H.; Lee, K.H. Shotgun proteomics using the iTRAQ isobaric tags. *Brief. Funct. Genom.* **2006**, *5*, 112–120. [CrossRef]
57. Picotti, P.; Rinner, O.; Stallmach, R.; Dautel, F.; Farrah, T.; Domon, B.; Wenschuh, H.; Aebersold, R. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat. Methods* **2010**, *7*, 43–46. [CrossRef] [PubMed]
58. Shi, T.; Su, D.; Liu, T.; Tang, K.; Camp, D.G., 2nd; Qian, W.-J.; Smith, R.D. Advancing the sensitivity of selected reaction monitoring-based targeted quantitative proteomics. *Proteomics* **2012**, *12*, 1074–1092. [CrossRef] [PubMed]
59. Collett, J.R.; Cho, E.J.; Ellington, A.D. Production and processing of aptamer microarrays. *Methods* **2005**, *37*, 4–15. [CrossRef] [PubMed]
60. Kim, C.H.; Tworoger, S.S.; Stampfer, M.J.; Dillon, S.T.; Gu, X.; Sawyer, S.J.; Chan, A.T.; Libermann, T.A.; Eliassen, A.H. Stability and reproducibility of proteomic profiles measured with an aptamer-based platform. *Sci. Rep.* **2018**, *8*, 8382. [CrossRef] [PubMed]
61. MacLean, E.; Broger, T.; Yerlikaya, S.; Fernandez-Carballo, B.L.; Pai, M.; Denkinger, C.M. A systematic review of biomarkers to detect active tuberculosis. *Nat. Microbiol.* **2019**, *4*, 748–758. [CrossRef]

62. Vidal, M.; Chan, D.W.; Gerstein, M.; Mann, M.; Omenn, G.S.; Tagle, D.; Sechi, S.; Workshop, P. The human proteome-a scientific opportunity for transforming diagnostics, therapeutics, and healthcare. *Clin. Proteom.* **2012**, *9*, 6. [CrossRef]
63. Skates, S.J.; Gillette, M.A.; LaBaer, J.; Carr, S.A.; Anderson, L.; Liebler, D.C.; Ransohoff, D.; Rifai, N.; Kondratovich, M.; Težak, Ž.; et al. Statistical design for biospecimen cohort size in proteomics-based biomarker discovery and verification studies. *J. Proteome Res.* **2013**, *12*, 5383–5394. [CrossRef]
64. Hortin, G.L.; Sviridov, D.; Anderson, N.L. High-abundance polypeptides of the human plasma proteome comprising the top 4 logs of polypeptide abundance. *Clin. Chem.* **2008**, *54*, 1608–1616. [CrossRef]
65. Gold, L.; Ayers, D.; Bertino, J.; Bock, C.; Bock, A.; Brody, E.N.; Carter, J.; Dalby, A.B.; Eaton, B.E.; Fitzwater, T.; et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS ONE* **2010**, *5*, e15004. [CrossRef]
66. Addona, T.A.; Abbatiello, S.E.; Schilling, B.; Skates, S.J.; Mani, D.R.; Bunk, D.M.; Spiegelman, C.H.; Zimmerman, L.J.; Ham, A.J.; Keshishian, H.; et al. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat. Biotechnol.* **2009**, *27*, 633–641. [CrossRef]
67. Levin, Y. The role of statistical power analysis in quantitative proteomics. *Proteomics* **2011**, *11*, 2565–2567. [CrossRef] [PubMed]
68. Tsim, S.; Kelly, C.; Alexander, L.; McCormick, C.; Thomson, F.; Woodward, R.; Foster, J.E.; Stobo, D.B.; Paul, J.; Maskell, N.A.; et al. Diagnostic and Prognostic Biomarkers in the Rational Assessment of Mesothelioma (DIAPHRAGM) study: Protocol of a prospective, multicentre, observational study. *BMJ Open* **2016**, *6*, e013324. [CrossRef] [PubMed]
69. Patel, V.; Dwivedi, A.K.; Deodhar, S.; Mishra, I.; Cistola, D.P. Aptamer-based search for correlates of plasma and serum water T2: Implications for early metabolic dysregulation and metabolic syndrome. *Biomark. Res.* **2018**, *6*, 28. [CrossRef] [PubMed]
70. Obuchowski, N.A.; McClish, D.K. Sample size determination for diagnostic accuracy studies involving binormal ROC curve indices. *Stat. Med.* **1997**, *16*, 1529–1542. [CrossRef]
71. Ngo, L.H.; Austin Argentieri, M.; Dillon, S.T.; Kent, B.V.; Kanaya, A.M.; Shields, A.E.; Libermann, T.A. Plasma protein expression profiles, cardiovascular disease, and religious struggles among South Asians in the MASALA study. *Sci. Rep.* **2021**, *11*, 961. [CrossRef] [PubMed]
72. Diz, A.P.; Carvajal-Rodríguez, A.; Skibinski, D.O.F. Multiple hypothesis testing in proteomics: A strategy for experimental work. *Mol. Cell Proteom.* **2011**, *10*, M110.004374. [CrossRef] [PubMed]
73. Emilsson, V.; Ilkov, M.; Lamb, J.R.; Finkel, N.; Gudmundsson, E.F.; Pitts, R.; Hoover, H.; Gudmundsdottir, V.; Horman, S.R.; Aspelund, T.; et al. Co-regulatory networks of human serum proteins link genetics to disease. *Science* **2018**, *361*, 769–773. [CrossRef]
74. Kiddle, S.J.; Sattlecker, M.; Proitsi, P.; Simmons, A.; Westman, E.; Bazenet, C.; Nelson, S.K.; Williams, S.; Hodges, A.; Johnston, C.; et al. Candidate blood proteome markers of Alzheimer's disease onset and progression: A systematic review and replication study. *J. Alzheimers Dis.* **2014**, *38*, 515–531. [CrossRef]
75. Mischak, H.; Critselis, E.; Hanash, S.; Gallagher, W.M.; Vlahou, A.; Ioannidis, J.P.A. Epidemiologic Design and Analysis for Proteomic Studies: A Primer on -Omic Technologies. *Am. J. Epidemiol.* **2015**, *181*, 635–647. [CrossRef]
76. Faresjö, T.; Faresjö, A. To match or not to match in epidemiological studies–same outcome but less power. *Int. J. Environ. Res. Public Health* **2010**, *7*, 325–332. [CrossRef]
77. Kuo, C.-L.; Duan, Y.; Grady, J. Unconditional or Conditional Logistic Regression Model for Age-Matched Case–Control Data? *Front. Public Health* **2018**, *6*, 57. [CrossRef] [PubMed]
78. Ngo, D.; Sinha, S.; Shen, D.; Kuhn, E.W.; Keyes, M.J.; Shi, X.; Benson, M.D.; O'Sullivan, J.F.; Keshishian, H.; Farrell, L.A.; et al. Aptamer-Based Proteomic Profiling Reveals Novel Candidate Biomarkers and Pathways in Cardiovascular Disease. *Circulation* **2016**, *134*, 270–285. [CrossRef] [PubMed]
79. Karimpour-Fard, A.; Epperson, L.E.; Hunter, L.E. A survey of computational tools for downstream analysis of proteomic and other omic datasets. *Hum. Genom.* **2015**, *9*, 28. [CrossRef] [PubMed]
80. Myers, R.H.; Myers, S.L.; Ye, K.; Walpole, R.E. *Probability and Statistics for Engineers and Scientists*, 7th ed.; Prentice Hall: London, UK, 2002.
81. Armstrong, R.A.; Slade, S.V.; Eperjesi, F. An introduction to analysis of variance (ANOVA) with special reference to data from clinical experiments in optometry. *Ophthalmic Physiol. Opt.* **2000**, *20*, 235–241. [CrossRef] [PubMed]
82. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **1933**, *24*, 417–441. [CrossRef]
83. Hyvärinen, A.; Oja, E. Independent component analysis: Algorithms and applications. *Neural. Netw.* **2000**, *13*, 411–430. [CrossRef]
84. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [CrossRef]
85. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
86. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
87. Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137. [CrossRef]
88. Rokach, L.; Maimon, O. *Clustering Methods*; Springer: New York, NY, USA, 2005; pp. 321–352. [CrossRef]
89. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [CrossRef] [PubMed]

90. Dennis, G.; Sherman, B.T.; Hosack, D.A.; Yang, J.; Gao, W.; Lane, H.C.; Lempicki, R.A. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **2003**, *4*, R60. [CrossRef]
91. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef] [PubMed]
92. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [CrossRef] [PubMed]
93. Chen, J.Y.; Mamidipalli, S.; Huan, T. HAPPI: An online database of comprehensive human annotated and predicted protein interactions. *BMC Genom.* **2009**, *10*, S16. [CrossRef] [PubMed]
94. Swan, A.L.; Mobasheri, A.; Allaway, D.; Liddell, S.; Bacardit, J. Application of machine learning to proteomics data: Classification and biomarker identification in postgenomics biology. *Omics* **2013**, *17*, 595–610. [CrossRef]
95. Bashir, A.; Yang, Q.; Wang, J.; Hoyer, S.; Chou, W.; McLean, C.; Davis, G.; Gong, Q.; Armstrong, Z.; Jang, J.; et al. Machine learning guided aptamer refinement and discovery. *Nat. Commun.* **2021**, *12*, 2366. [CrossRef]
96. Krauze, A.V.; Zhuge, Y.; Zhao, R.; Tasci, E.; Camphausen, K. AI-Driven Image Analysis in Central Nervous System Tumors-Traditional Machine Learning, Deep Learning and Hybrid Models. *J. Biotechnol. Biomed.* **2022**, *5*, 1–19. [CrossRef]
97. Odenkirk, M.T.; Reif, D.M.; Baker, E.S. Multiomic Big Data Analysis Challenges: Increasing Confidence in the Interpretation of Artificial Intelligence Assessments. *Anal. Chem.* **2021**, *93*, 7763–7773. [CrossRef]
98. Reel, P.S.; Reel, S.; Pearson, E.; Trucco, E.; Jefferson, E. Using machine learning approaches for multi-omics data analysis: A review. *Biotechnol. Adv.* **2021**, *49*, 107739. [CrossRef]
99. Dong, L.; Watson, J.; Cao, S.; Arregui, S.; Saxena, V.; Ketz, J.; Awol, A.K.; Cohen, D.M.; Caterino, J.M.; Hains, D.S.; et al. Aptamer based proteomic pilot study reveals a urine signature indicative of pediatric urinary tract infections. *PLoS ONE* **2020**, *15*, e0235328. [CrossRef] [PubMed]
100. Mc Ardle, A.; Kwasnik, A.; Szentpetery, A.; Hernandez, B.; Parnell, A.; de Jager, W.; de Roock, S.; FitzGerald, O.; Pennington, S.R. Identification and Evaluation of Serum Protein Biomarkers That Differentiate Psoriatic Arthritis From Rheumatoid Arthritis. *Arthritis Rheumatol.* **2022**, *74*, 81–91. [CrossRef] [PubMed]
101. O'Neil, L.J.; Hu, P.; Liu, Q.; Islam, M.M.; Spicer, V.; Rech, J.; Hueber, A.; Anaparti, V.; Smolik, I.; El-Gabalawy, H.S.; et al. Proteomic Approaches to Defining Remission and the Risk of Relapse in Rheumatoid Arthritis. *Front. Immunol.* **2021**, *12*. [CrossRef] [PubMed]
102. Hewitson, L.; Mathews, J.A.; Devlin, M.; Schutte, C.; Lee, J.; German, D.C. Blood biomarker discovery for autism spectrum disorder: A proteomic analysis. *PLoS ONE* **2021**, *16*, e0246581. [CrossRef] [PubMed]
103. O'Dwyer, D.N.; Norman, K.C.; Xia, M.; Huang, Y.; Gurczynski, S.J.; Ashley, S.L.; White, E.S.; Flaherty, K.R.; Martinez, F.J.; Murray, S.; et al. The peripheral blood proteome signature of idiopathic pulmonary fibrosis is distinct from normal and is associated with novel immunological processes. *Sci. Rep.* **2017**, *7*, 46560. [CrossRef]
104. Hung, J.-H.; Yang, T.-H.; Hu, Z.; Weng, Z.; DeLisi, C. Gene set enrichment analysis: Performance evaluation and usage guidelines. *Brief. Bioinform.* **2011**, *13*, 281–291. [CrossRef]
105. Khatri, P.; Sirota, M.; Butte, A.J. Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges. *PLOS Comput. Biol.* **2012**, *8*, e1002375. [CrossRef]
106. Gatti, D.M.; Barry, W.T.; Nobel, A.B.; Rusyn, I.; Wright, F.A. Heading down the wrong pathway: On the influence of correlation within gene sets. *BMC Genom.* **2010**, *11*, 574. [CrossRef]
107. Pan, K.-H.; Lih, C.-J.; Cohen, S.N. Effects of threshold choice on biological conclusions reached during analysis of gene expression by DNA microarrays. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 8961–8965. [CrossRef]
108. Maleki, F.; Ovens, K.; Hogan, D.J.; Kusalik, A.J. Gene Set Analysis: Challenges, Opportunities, and Future Research. *Front. Genet.* **2020**, *11*, 654. [CrossRef]
109. Parolo, S.; Marchetti, L.; Lauria, M.; Misselbeck, K.; Scott-Boyer, M.-P.; Caberlotto, L.; Priami, C. Combined use of protein biomarkers and network analysis unveils deregulated regulatory circuits in Duchenne muscular dystrophy. *PLoS ONE* **2018**, *13*, e0194225. [CrossRef] [PubMed]
110. Çelik, H.; Lindblad, K.E.; Popescu, B.; Gui, G.; Goswami, M.; Valdez, J.; DeStefano, C.; Lai, C.; Thompson, J.; Ghannam, J.Y.; et al. Highly multiplexed proteomic assessment of human bone marrow in acute myeloid leukemia. *Blood Adv.* **2020**, *4*, 367–379. [CrossRef] [PubMed]
111. Kim, S.Y.; Volsky, D.J. PAGE: Parametric analysis of gene set enrichment. *BMC Bioinform.* **2005**, *6*, 144. [CrossRef] [PubMed]
112. Luo, W.; Friedman, M.S.; Shedden, K.; Hankenson, K.D.; Woolf, P.J. GAGE: Generally applicable gene set enrichment for pathway analysis. *BMC Bioinform.* **2009**, *10*, 161. [CrossRef] [PubMed]
113. Maleki, F.; Ovens, K.L.; Hogan, D.J.; Rezaei, E.; Rosenberg, A.M.; Kusalik, A.J. Measuring consistency among gene set analysis methods: A systematic study. *J. Bioinform. Comput. Biol.* **2019**, *17*, 1940010. [CrossRef] [PubMed]
114. Nguyen, T.-M.; Shafi, A.; Nguyen, T.; Draghici, S. Identifying significantly impacted pathways: A comprehensive review and assessment. *Genome Biol.* **2019**, *20*, 203. [CrossRef] [PubMed]
115. Chowdhury, S.; Sarkar, R.R. Comparison of human cell signaling pathway databases—Evolution, drawbacks and challenges. *Database* **2015**, *2015*, 1–25. [CrossRef]
116. García-Campos, M.A.; Espinal-Enríquez, J.; Hernández-Lemus, E. Pathway Analysis: State of the Art. *Front. Physiol.* **2015**, *6*, 383. [CrossRef]

117. Norby, F.L.; Tang, W.; Pankow, J.S.; Lutsey, P.L.; Alonso, A.; Steffan, B.; Chen, L.Y.; Zhang, M.; Shippee, N.D.; Ballantyne, C.M.; et al. Proteomics and Risk of Atrial Fibrillation in Older Adults (From the Atherosclerosis Risk in Communities [ARIC] Study). *Am. J. Cardiol.* **2021**, *161*, 42–50. [CrossRef]
118. Sathyan, S.; Ayers, E.; Gao, T.; Weiss, E.F.; Milman, S.; Verghese, J.; Barzilai, N. Plasma proteomic profile of age, health span, and all-cause mortality in older adults. *Aging Cell* **2020**, *19*, e13250. [CrossRef]
119. Green, M.L.; Karp, P.D. The outcomes of pathway database computations depend on pathway ontology. *Nucleic Acids Res.* **2006**, *34*, 3687–3697. [CrossRef] [PubMed]
120. Karp, P.D.; Midford, P.E.; Caspi, R.; Khodursky, A. Pathway size matters: The influence of pathway granularity on over-representation (enrichment analysis) statistics. *BMC Genom.* **2021**, *22*, 191. [CrossRef] [PubMed]
121. Battle, A.; Khan, Z.; Wang, S.H.; Mitrano, A.; Ford, M.J.; Pritchard, J.K.; Gilad, Y. Genomic variation. Impact of regulatory variation from RNA to protein. *Science* **2015**, *347*, 664–667. [CrossRef] [PubMed]
122. Chick, J.M.; Munger, S.C.; Simecek, P.; Huttlin, E.L.; Choi, K.; Gatti, D.M.; Raghupathy, N.; Svenson, K.L.; Churchill, G.A.; Gygi, S.P. Defining the consequences of genetic variation on a proteome-wide scale. *Nature* **2016**, *534*, 500–505. [CrossRef] [PubMed]
123. Pietzner, M.; Wheeler, E.; Carrasco-Zanini, J.; Raffler, J.; Kerrison, N.D.; Oerton, E.; Auyeung, V.P.W.; Luan, J.a.; Finan, C.; Casas, J.P.; et al. Genetic architecture of host proteins involved in SARS-CoV-2 infection. *Nat. Commun.* **2020**, *11*, 6397. [CrossRef]
124. Yao, C.; Chen, G.; Song, C.; Keefe, J.; Mendelson, M.; Huan, T.; Sun, B.B.; Laser, A.; Maranville, J.C.; Wu, H.; et al. Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. *Nat. Commun.* **2018**, *9*, 3268. [CrossRef]
125. Ferkingstad, E.; Sulem, P.; Atlason, B.A.; Sveinbjornsson, G.; Magnusson, M.I.; Styrmisdottir, E.L.; Gunnarsdottir, K.; Helgason, A.; Oddsson, A.; Halldorsson, B.V.; et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* **2021**, *53*, 1712–1721. [CrossRef]
126. Sun, B.B.; Maranville, J.C.; Peters, J.E.; Stacey, D.; Staley, J.R.; Blackshaw, J.; Burgess, S.; Jiang, T.; Paige, E.; Surendran, P.; et al. Genomic atlas of the human plasma proteome. *Nature* **2018**, *558*, 73–79. [CrossRef]
127. Di Narzo, A.F.; Telesco, S.E.; Brodmerkel, C.; Argmann, C.; Peters, L.A.; Li, K.; Kidd, B.; Dudley, J.; Cho, J.; Schadt, E.E.; et al. High-Throughput Characterization of Blood Serum Proteomics of IBD Patients with Respect to Aging and Genetic Factors. *PLoS Genet.* **2017**, *13*, e1006565. [CrossRef]
128. Suhre, K.; Arnold, M.; Bhagwat, A.M.; Cotton, R.J.; Engelke, R.; Raffler, J.; Sarwath, H.; Thareja, G.; Wahl, A.; DeLisle, R.K.; et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* **2017**, *8*, 14357. [CrossRef]
129. Mosley, J.D.; Benson, M.D.; Smith, J.G.; Melander, O.; Ngo, D.; Shaffer, C.M.; Ferguson, J.F.; Herzig, M.S.; McCarty, C.A.; Chute, C.G.; et al. Probing the Virtual Proteome to Identify Novel Disease Biomarkers. *Circulation* **2018**, *138*, 2469–2481. [CrossRef] [PubMed]
130. Brandes, N.; Linial, N.; Linial, M. PWAS: Proteome-wide association study—linking genes and phenotypes by functional variation in proteins. *Genome Biol.* **2020**, *21*, 173. [CrossRef] [PubMed]
131. Wingo, A.P.; Liu, Y.; Gerasimov, E.S.; Gockley, J.; Logsdon, B.A.; Duong, D.M.; Dammer, E.B.; Robins, C.; Beach, T.G.; Reiman, E.M.; et al. Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis. *Nat. Genet.* **2021**, *53*, 143–146. [CrossRef]
132. Zhu, Z.; Zhang, F.; Hu, H.; Bakshi, A.; Robinson, M.R.; Powell, J.E.; Montgomery, G.W.; Goddard, M.E.; Wray, N.R.; Visscher, P.M.; et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **2016**, *48*, 481–487. [CrossRef] [PubMed]
133. King, E.A.; Davis, J.W.; Degner, J.F. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet.* **2019**, *15*, e1008489. [CrossRef]
134. Nguyen, V.-T.; Seo, H.B.; Kim, B.C.; Kim, S.K.; Song, C.-S.; Gu, M.B. Highly sensitive sandwich-type SPR based detection of whole H5Nx viruses using a pair of aptamers. *Biosens. Bioelectron.* **2016**, *86*, 293–300. [CrossRef]
135. Kim, K.; Lee, S.; Ryu, S.; Han, D. Efficient isolation and elution of cellular proteins using aptamer-mediated protein precipitation assay. *Biochem. Biophys. Res. Commun.* **2014**, *448*, 114–119. [CrossRef]
136. Jung, Y.J.; Katilius, E.; Ostroff, R.M.; Kim, Y.; Seok, M.; Lee, S.; Jang, S.; Kim, W.S.; Choi, C.M. Development of a Protein Biomarker Panel to Detect Non-Small-Cell Lung Cancer in Korea. *Clin. Lung. Cancer* **2017**, *18*, e99–e107. [CrossRef]
137. Müller, J.; Friedrich, M.; Becher, T.; Braunstein, J.; Kupper, T.; Berdel, P.; Gravius, S.; Rohrbach, F.; Oldenburg, J.; Mayer, G.; et al. Monitoring of plasma levels of activated protein C using a clinically applicable oligonucleotide-based enzyme capture assay. *J. Thromb. Haemost.* **2012**, *10*, 390–398. [CrossRef]
138. Nakayasu, E.S.; Gritsenko, M.; Piehowski, P.D.; Gao, Y.; Orton, D.J.; Schepmoes, A.A.; Fillmore, T.L.; Frohnert, B.I.; Rewers, M.; Krischer, J.P.; et al. Tutorial: Best practices and considerations for mass-spectrometry-based protein biomarker discovery and validation. *Nat. Protoc.* **2021**, *16*, 3737–3760. [CrossRef]
139. Füzéry, A.K.; Levin, J.; Chan, M.M.; Chan, D.W. Translation of proteomic biomarkers into FDA approved cancer diagnostics: Issues and challenges. *Clin Proteom.* **2013**, *10*, 13. [CrossRef] [PubMed]

140. Zhang, Z.; Chan, D.W. The road from discovery to clinical diagnostics: Lessons learned from the first FDA-cleared in vitro diagnostic multivariate index assay of proteomic biomarkers. *Cancer Epidemiol. Biomark. Prev.* **2010**, *19*, 2995–2999. [CrossRef] [PubMed]
141. Sabatine, M.S.; Morrow, D.A.; de Lemos, J.A.; Gibson, C.M.; Murphy, S.A.; Rifai, N.; McCabe, C.; Antman, E.M.; Cannon, C.P.; Braunwald, E. Multimarker approach to risk stratification in non-ST elevation acute coronary syndromes: Simultaneous assessment of troponin I, C-reactive protein, and B-type natriuretic peptide. *Circulation* **2002**, *105*, 1760–1763. [CrossRef] [PubMed]
142. Harlid, S.; Harbs, J.; Myte, R.; Brunius, C.; Gunter, M.J.; Palmqvist, R.; Liu, X.; Van Guelpen, B. A two-tiered targeted proteomics approach to identify pre-diagnostic biomarkers of colorectal cancer risk. *Sci. Rep.* **2021**, *11*, 5151. [CrossRef] [PubMed]
143. Ueland, F.R.; Desimone, C.P.; Seamon, L.G.; Miller, R.A.; Goodrich, S.; Podzielinski, I.; Sokoll, L.; Smith, A.; van Nagell, J.R., Jr.; Zhang, Z. Effectiveness of a multivariate index assay in the preoperative assessment of ovarian tumors. *Obs. Gynecol.* **2011**, *117*, 1289–1297. [CrossRef]
144. Argilés, Á.; Siwy, J.; Duranton, F.; Gayrard, N.; Dakna, M.; Lundin, U.; Osaba, L.; Delles, C.; Mourad, G.; Weinberger, K.M.; et al. CKD273, a new proteomics classifier assessing CKD and its prognosis. *PLoS ONE* **2013**, *8*, e62837. [CrossRef]
145. Nkuipou-Kenfack, E.; Zürbig, P.; Mischak, H. The long path towards implementation of clinical proteomics: Exemplified based on CKD273. *Proteomics–Clin. Appl.* **2017**, *11*, 1600104. [CrossRef]
146. Bratulic, S.; Gatto, F.; Nielsen, J. The Translational Status of Cancer Liquid Biopsies. *Regen. Eng. Transl. Med.* **2021**, *7*, 312–352. [CrossRef]
147. Gygi, S.P.; Rochon, Y.; Franza, B.R.; Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol. Cell Biol.* **1999**, *19*, 1720–1730. [CrossRef]
148. Lim, S.Y.; Lee, J.H.; Welsh, S.J.; Ahn, S.B.; Breen, E.; Khan, A.; Carlino, M.S.; Menzies, A.M.; Kefford, R.F.; Scolyer, R.A.; et al. Evaluation of two high-throughput proteomic technologies for plasma biomarker discovery in immunotherapy-treated melanoma patients. *Biomark. Res.* **2017**, *5*, 32. [CrossRef]
149. Raffield, L.M.; Dang, H.; Pratte, K.A.; Jacobson, S.; Gillenwater, L.A.; Ampleford, E.; Barjaktarevic, I.; Basta, P.; Clish, C.B.; Comellas, A.P.; et al. Comparison of Proteomic Assessment Methods in Multiple Cohort Studies. *Proteomics* **2020**, *20*, e1900278. [CrossRef] [PubMed]
150. Liu, R.X.; Thiessen-Philbrook, H.R.; Vasan, R.S.; Coresh, J.; Ganz, P.; Bonventre, J.V.; Kimmel, P.L.; Parikh, C.R. Comparison of proteomic methods in evaluating biomarker-AKI associations in cardiac surgery patients. *Transl. Res.* **2021**, *238*, 49–62. [CrossRef] [PubMed]
151. Billing, A.M.; Ben Hamidane, H.; Bhagwat, A.M.; Cotton, R.J.; Dib, S.S.; Kumar, P.; Hayat, S.; Goswami, N.; Suhre, K.; Rafii, A.; et al. Complementarity of SOMAscan to LC-MS/MS and RNA-seq for quantitative profiling of human embryonic and mesenchymal stem cells. *J. Proteom.* **2017**, *150*, 86–97. [CrossRef]
152. Berglund, L.; Bjorling, E.; Oksvold, P.; Fagerberg, L.; Asplund, A.; Szigyarto, C.A.; Persson, A.; Ottosson, J.; Wernerus, H.; Nilsson, P.; et al. A genecentric Human Protein Atlas for expression profiles based on antibodies. *Mol. Cell Proteom.* **2008**, *7*, 2019–2027. [CrossRef] [PubMed]