Journal of proteome research

pubs.acs.org/jpr

Technical Note

Computationally Efficient Bayesian Estimation of Graphical Networks for Omics Data

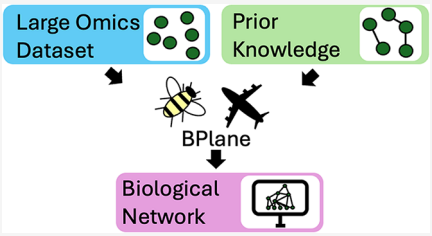
Daniel W. Adrian, Erik D. VonKaenel, Moses Y. Obiri, David J. Degnan, Amy C. Sims, Kristie Oxford, and Lisa M. Bramer*

Cite This: *J. Proteome Res.* 2026, 25, 4306−4314

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Graphical networks are useful, widely used modeling approaches to represent complex biological processes with biological measurements generated by platforms such as mass spectrometry. Bayesian analyses of graphical networks for omics data have several advantages over their frequentist counterparts such as the inclusion of prior knowledge in the estimation of models. However, Bayesian approaches to date have only been feasible for data with a couple of hundred biomolecules due to prohibitive computational time, but omics data often contain tens of thousands of biomolecules. Here, we present and illustrate a more computationally efficient approach named BPlane (Bayesian PseudoLikelihood-based Algorithm for Network Estimation) to extend Bayesian modeling capabilities for larger-sized data sets, such as most untargeted proteomics data. Via simulation, we demonstrate that BPlane produces substantial computational savings over a current state-of-the-art Bayesian algorithm while maintaining competitive edge detection accuracy. We further demonstrate the computational benefit of BPlane on a SARS-CoV-2 proteomics data set with 7000 proteins.



KEYWORDS: *Bayesian statistical modeling, Gaussian graphical model, omics data, pseudolikelihood, EM algorithm*

INTRODUCTION

The omics revolution[1] has led to a data explosion, enabling the measurement of thousands to tens of thousands of biomolecules at a time. Biological relational networks are a popular way to analyze these data. In these networks or graphs, nodes represent individual biomolecules such as genes, proteins, RNA transcripts, or metabolites, and edges reflect a relationship between node pairs such as interactions. This informs network medicine,[2] which follows the hypothesis that a disease is rarely a consequence of an abnormality in a single gene but reflects various processes that interact in a complex network. A better understanding of this interactivity has led to the identification of disease genes[3] and disease pathways, which may inform targets for drug development.[4,5]

Gaussian graphical model (GGM) estimation is one approach to estimating these networks.[6] This model assumes that each sample, consisting of thousands of variables, one for each node, follows a multivariate Gaussian distribution after the application of preprocessing strategies.[7,8] Each edge estimated by GGMs corresponds to an estimated nonzero partial correlation between two biomolecule variables, where partial correlation is defined as the correlation between two variables conditioned on all others; alternatively, zero partial correlations correspond to the absence of an edge. A fundamental result in GGMs is that two variables will have zero partial correlation when the corresponding entry of the inverse covariance matrix (or precision matrix) is zero.[9] Thus, estimating the graphical network reduces to the task of estimating the precision matrix.

The maximum-likelihood estimate of the precision matrix is the inverse of the sample covariance matrix. However, it is well-known that this approach is impractical for several reasons.[6] First, estimation requires computing the inverse of the sample covariance matrix, which is not possible when the number of samples (n) is smaller than the number of observed variables (p). This is almost always the case for omics data, where the number of samples is often substantially less than the number of variables when data are captured from techniques such as mass spectroscopy. Second, even if direct inversion of the sample covariance matrix were possible, the estimated precision matrix obtained from this method rarely has partial correlation values equal to zero, which corresponds to completely connected and therefore uninformative graphs.

To determine which pairs of biomolecules are connected, based on observed data, it is common to introduce more sparsity (i.e., zeros) in the estimated precision matrix via alternative modeling strategies. A common approach utilizes LASSO regression.[10] Among the first to follow this approach

Received: December 25, 2025
Revised: June 7, 2026
Accepted: June 18, 2026
Published: June 30, 2026

was a neighborhood selection method[11] that estimated the edges attached to each node separately. However, this approach suffers from inconsistency; an edge, say between nodes 1 and 2, may be estimated to be present within the neighborhood of node 1 but not within the neighborhood of node 2. In a seminal paper, the graphical LASSO (or GLASSO)[12] solves this problem through penalized likelihood estimation of the precision matrix via the LASSO penalty. The GLASSO algorithm has had many refinements, improving its stability[13] and speed,[14] and generalizing it to multiple graphs.[15]

Further refinements to GLASSO have utilized the pseudolikelihood[16] as an approximation of the exact Gaussian likelihood utilized by GLASSO.[17−20] The main advantage of these pseudolikelihood-based algorithms is speed of computation. By reducing the computational complexity and number of convergence checks required, these methods result in significant computational savings in the common case when $n \ll p$. In particular, CONCORD[20] is a pseudolikelihood-based algorithm with favorable convergence properties due to the convexity of its objective function, a feature lacking in the SPACE,[18] symmetric LASSO,[19] and SPLICE algorithms.[17]

Bayesian approaches to estimating GGMs have several advantages over their frequentist counterparts. One is that the Bayesian approach outputs posterior edge probabilities, which are more readily interpretable than frequentist-based binary edge indicators or p-values for edge inference. Another advantage is that Bayesian methods allow for the incorporation of prior information on edge probabilities obtained through databases such as KEGG[21] or UniProt[22] or through text mining of literature.[23] Still another advantage is that the Bayesian approach provides a framework for the combination of information across multiple experiments of omics data, when direct combination of omics data can result in biases due to batch effects.[24]

Unfortunately, a disadvantage of Bayesian approaches is their high computation time compared with frequentist approaches. Recent advances, including the BAGUS algorithm,[25,26] have improved the computation time by utilizing an EM algorithm[27] in place of more computationally expensive Markov chain Monte Carlo (MCMC) methods.[28] However, as we will show, even this algorithm needs to be sped up to be viable for estimating omics-data-based graphs with thousands of nodes.

In this paper, we present a faster alternative to existing methods, which we name BPlane, short for Bayesian PseudoLikelihood-based Algorithm for Network Estimation. BPlane leaves the Bayesian framework of the BAGUS algorithm intact but updates the optimization strategy used in the M-step of the EM algorithm, replacing BAGUS's utilization of the GLASSO algorithm with the CONCORD algorithm.

Here, we illustrate the BPlane algorithm and its methodological underpinnings. We present simulation experiments that quantify the extent of BPlane's computational improvements and support the assertion that its edge detection accuracy can be considered comparable to the BAGUS algorithm. Next, we apply BPlane to proteomics data from a SARS-Cov-2 study to demonstrate its practical utility and gains in computational efficiency. Finally, we close with a brief discussion of our findings.

METHODS

Please note that the Supporting Information provides a more technical explanation of the statistical methods discussed in this section.

Bayesian Model Framework

A typical omics data set consists of n samples of p biomolecule measurements. The Gaussian graphical model (GGM) assumes that the n samples are independent, and each follows a multivariate normal distribution with mean equal to the zero vector of length p and covariance matrix Σ. The partial correlation between two biomolecule measurements is zero if and only if the corresponding entry of the inverse covariance matrix $\Theta = \Sigma^{-1}$ is zero.[9] Thus, estimation of Θ produces the biomolecular graph, where edges are drawn (or not) between biomolecules i and j if the corresponding entry of Θ, θ_{ij}, is estimated to be nonzero (or zero).

BPlane adopts the BAGUS algorithm's Bayesian model framework,[25] which uses a "spike-and-slab" prior[29] to model the off-diagonal entries of Θ. The "spike" represents a distribution sharply concentrated around zero, representing the nonedges with $\theta_{ij} = 0$; on the other hand, the "slab" distribution is more diffuse, representing the edges with $\theta_{ij} \neq 0$. The Bayesian analysis estimates whether the "spike" or the "slab" is the true distribution of θ_{ij} and outputs the posterior probability of an edge between biomolecules i and j. The Bayesian model also allows the incorporation of prior edge probabilities, which may be obtained from biological databases[21,22] or literature.[23] While the BAGUS algorithm is designed for data set samples coming from a single group, BPlane can be extended to the GemBag[26] multigroup Bayesian framework as well. We focus on the single group case for ease of exposition.

The BPlane Algorithm

The BAGUS algorithm[25] utilizes the EM algorithm to calculate the estimates of Θ and the posterior edge probabilities. The EM algorithm has two steps to each iteration: the E (expectation) step and the M (maximization) step. The innovation in the BPlane algorithm lies in modifying BAGUS's M-step to notably increase computational efficiency. While BAGUS applies GLASSO[13] in the M-step, BPlane utilizes the CONCORD[20] algorithm instead.

BPlane has a faster M-step than BAGUS for two reasons. First, while GLASSO maximizes the multivariate normal log-likelihood through block-wise coordinate descent, BPlane maximizes a modified pseudolikelihood through simple coordinate descent. This means that GLASSO updates of the Θ matrix elements within each column, repeatedly until convergence, before proceeding to the next column. BPlane, on the other hand, updates each element of Θ only once per M-step. A second advantage of BPlane is that the computational complexity of the M-step is reduced from $O(p^3)$ to $O(np^2)$ when $n < p$. This results in considerable computational savings in omics data, where the number of samples n is often substantially less than the number of biomolecules p.

Practical Considerations

To implement the BPlane algorithm in practice, we must define criteria to determine convergence of the edge probabilities and how to draw edges in the estimated graph from these probabilities. We check the convergence of the probability of an edge between biomolecules i and j after EM iteration t, denoted by $\omega_{ij}^{(t)}$, by comparing to their previous iteration's value and controlling for the maximum change over all edges. In symbols, we compare $\max_{i,j} |\omega_{ij}^{(t)} - \omega_{ij}^{(t-1)}|$ after each EM iteration to a specified threshold, here set as 0.001, for a maximum number of iterations. Though the threshold of 0.001 is somewhat arbitrary, we argue that it is quite small for the maximum change over a large probability matrix.

The supplement provides more details about the BPlane methodology. The R and C++ code for BPlane can be found here: https://github.com/PNNL-Predictive-Phenomics/BPlane.

Simulation Experiments

Our simulation experiments compared BPlane with the BAGUS and GLASSO algorithms as well as the G-Wishart weighted proposal

https://doi.org/10.1021/acs.jproteome.5c01277
J. Proteome Res. 2026, 25, 4306−4314

algorithm (WWA)[30] for Bayesian MCMC sampling, which has shown reduced computing time relative to competing MCMC algorithms. We evaluated these algorithms in terms of computation time and edge detection performance over simulations meant to represent omics data.

The GGM model-based data were generated using the `bdgraph.sim` function from the R package `BDgraph`[31] as follows. The graphs followed a power-law (aka scale-free) degree distribution, which biological networks have been shown to follow 32 where the number of edges is one less than the number of nodes. Given a network, the precision matrix was generated from the G-Wishart distribution, and the biomolecule measurements were generated from a multivariate normal distribution. We generated data with n = 50 and p varying in powers of two from 10 to 8000. These values of n and p were chosen so that the largest simulated data set had similar dimensions to the SARS-CoV-2 data set, which has n = 45 and p = 6968.

While our simulations focus on sparse scale-free networks due to their relevance for biological systems,[31,33] we note that the underlying Bayesian spike-and-slab graphical modeling framework has previously been evaluated across a wide range of network structures, including star, AR(2), circle, random, nearest-neighbor, and scale-free graphs.[25,26] These studies demonstrated consistent performance in edge recovery across structurally diverse settings.[25,26] Since BPlane preserves the statistical formulation of BAGUS and differs only in the optimization strategy used in the M-step, we expect similar robustness across network structures beyond the scale-free structures considered here.

The previous simulation study assumed noninformative prior information to be fair to the frequentist GLASSO method, setting prior edge probabilities equal to 0.5 for all edges. However, as utilizing the prior biological information in omics data is a primary motivator for this work, we also performed a study that utilized various amounts of prior information. For simulated GGM data with n = 50 and p = 2000, the proportion of edges in the true graph supplied informative prior probabilities of 0.9 varied across 0.00, 0.25, 0.50, 0.75, and 1.00. Other edges in the graph, including the remaining true edges and all true nonedges, are given noninformative prior probabilities of 0.5. We fit the BPlane algorithm for 500 replicate data sets at each setting.

To run the competing algorithms, we used the `huge`[34] R package for GLASSO, C++ code from the GemBag[26] GitHub repository for BAGUS, and associated code from a Bayesian GGM review for WWA.[35]

Computation Time

We compared the computation time needed for all four algorithms to estimate graphs for the simulated data sets. All algorithms ran within R[36] on a 64-bit Windows laptop utilizing a 2.10 GHz CPU and 16 GB RAM. All timed processes used a single core. The computation time for BPlane, BAGUS, and GLASSO included estimation of tuning parameters; WWA ran 30,000 MCMC iterations.

Edge Detection Performance

We evaluated the edge detection utility of the four algorithms with three metrics based on the posterior edge probabilities for the Bayesian methods and the estimated graph (i.e., zeros or ones) for GLASSO. The primary metric is the area under the Receiver Operating Characteristic (ROC) curve, or AUC, which reflects the degree to which the edge probabilities for edges in the true graph are higher than those of nonedges in the true graph. It ranges from 0 (worst) to 1 (best), where an AUC of 0.5 means the edge probability matrix is as good as a random guess. We also add two measures of the magnitude of the edge probabilities, Pr^{+} and Pr^{-}, which are average edge probabilities over true edges and true nonedges, respectively.[35]

SARS-CoV-2 Dataset

The SARS-CoV-2 proteomics data set was used to demonstrate the computational utility of BPlane in estimating strain-specific protein networks. We further evaluated these networks and the differences between strains to show examples of SARS-CoV-2 biomarkers noted in previous studies.

Primary Cell Culture

Primary human airway epithelial (HAE) lung cells were isolated from distal lung tissue and processed as described previously.[37] The lung cells were obtained under protocol 03−1396, which was approved by the University of North Carolina at Chapel Hill Institutional Review Board. HAE cultures were plated under BEGM growth media and allowed to reach confluence. Within 7 days of plating, cultures were grown under an air−liquid interface (ALI) for 6 weeks to allow full maturation and development of cilia. Cultures were then shipped to PNNL and allowed to acclimate for at least 7 days prior to infection.

Sample Processing for Proteomics

Prior to infection, HAE cultures were washed twice for 30 min with 1× sterile phosphate-buffered saline to remove mucous from the apical surface. HAE were mock-infected or infected (multiplicity of infection (MOI) 5) with SARS-CoV-2 Washington Strain (SARS-Related Coronavirus 2, Isolate USA-WA1/2020 NR-52281; which was deposited by the Centers for Disease Control and Prevention and obtained through BEI Resources, NIAID, NIH) or SARS-CoV-2 Italy Strain (SARS-Related Coronavirus 2, Isolate Italy-INMI1 NR-52284; which was deposited by Dr. Maria Capobianchi for distribution through BEI Resources, NIAID, NIH) and harvested at 48, 60, and 72 hpi. HAE cultures were processed for proteomic analysis using the MPLEx method.[38−40] Briefly, 5 volumes of cold (−20 °C) 2:1 (v:v) chloroform:methanol solution were used to extract proteins, lipids, and metabolites from samples. The upper (enriched in lipids) and lower (enriched in hydrophilic metabolites) phases were collected and stored for downstream analysis. The protein disc was washed with cold 100% methanol, centrifuged at 16,000g for 10 min at 4 °C, and the supernatant was removed and air-dried. Proteins were dissolved and denatured in 8 M urea prepared in 50 mM ammonium bicarbonate, and protein concentrations were quantified using the BCA assay (ThermoFisher, Waltham, MA). Proteins were transferred to 96-well plates and reduced by adding 10 mM dithiothreitol and incubated for 1 h at 37 °C with 1200 rpm shaking. Thiol groups were alkylated by adding 40 mM iodoacetamide and incubating for 1 h at 37 °C with 1200 rpm shaking. Samples were diluted 8-fold with 50 mM ammonium bicarbonate, with 1 mM calcium chloride added. Proteins were digested with 1:50 (m:m) trypsin (Promega):protein for 6 h at 37 °C with 1200 rpm shaking. After trypsinization, peptides were cleaned up using C18 solid phase extraction as previously described.[41] The digestion and solid phase extraction were performed on an EpMotion liquid-handling robot (Eppendorf).

TMT16 Plex labeling reactions were prepared, quenched, dried, and desalted over C18 columns as previously described.[42] For each TMT set, one channel was assigned to a pooled mix of all samples, which was used as a reference for normalization and quality control. Samples were fractionated into 96 fractions by reverse phase and concatenated into 12 fractions to increase the depth of peptide identification.[43] Fractions were concentrated and frozen prior to LC-MS. Proteomics samples were analyzed using a Thermo Scientific UltiMate 3000 series high-performance liquid chromatography system and Q-Exactive Plus Orbitrap mass spectrometer (Thermo Scientific, San Jose, CA). Samples were loaded into a trap column (150 μm i.d., 4 cm length, packed in-lab with Jupiter C18 packing material, 300 Å pore size, 5 μm particle size; Phenomenex, Torrance, CA, USA) using mobile phase A (0.1% formic acid in water). Peptides were then reverse eluted off the trap onto a 75 μm i.d., 25 cm column with integrated emitter (New Objective, Littleton, MA), packed in-lab with Waters BEH C18 1.7 μm packing material (Milford, MA), at a flow rate of 200 nL/min and column temperature of 45C for 120 min. The MS1 spectra were collected at a scan range of 300−1800 m/z, a resolution of 70,000, an automatic gain control (AGC) target of 3 × 106, and a maximum injection time of 20 ms. MS2 spectra were collected at a resolution of 35,000, a maximum AGC target of 1×10^5, and a maximum ion injection time of 100 ms.

Mass spectral files were processed with mzRefinery to recalibrate masses and extract peak lists.[44] Peptides were identified by searching the human SwissProt sequences of Uniprot Knowledgebase (downloaded November 5, 2019) using MS-GF+.[45] The identification

https://doi.org/10.1021/acs.jproteome.5c01277
J. Proteome Res. 2026, 25, 4306−4314

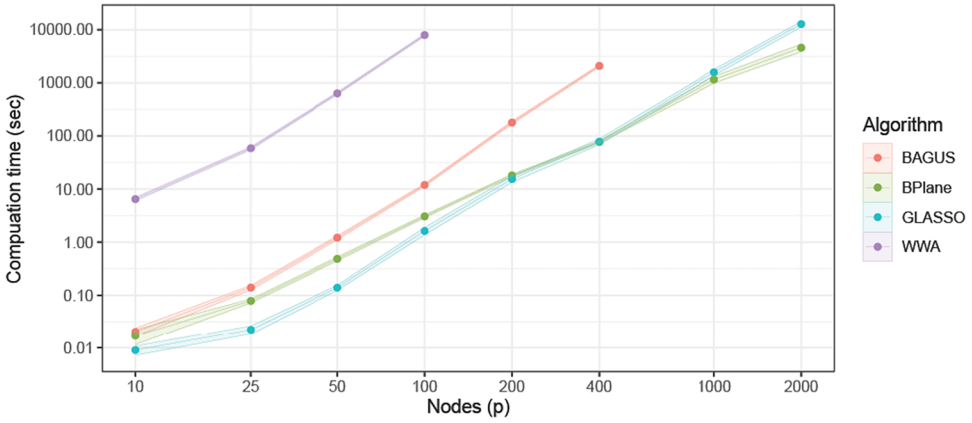


Figure 1. Computation times (mean plus/minus 2 standard errors) for the BAGUS, BPlane, GLASSO, and WWA algorithms. Note the logarithmic scales.

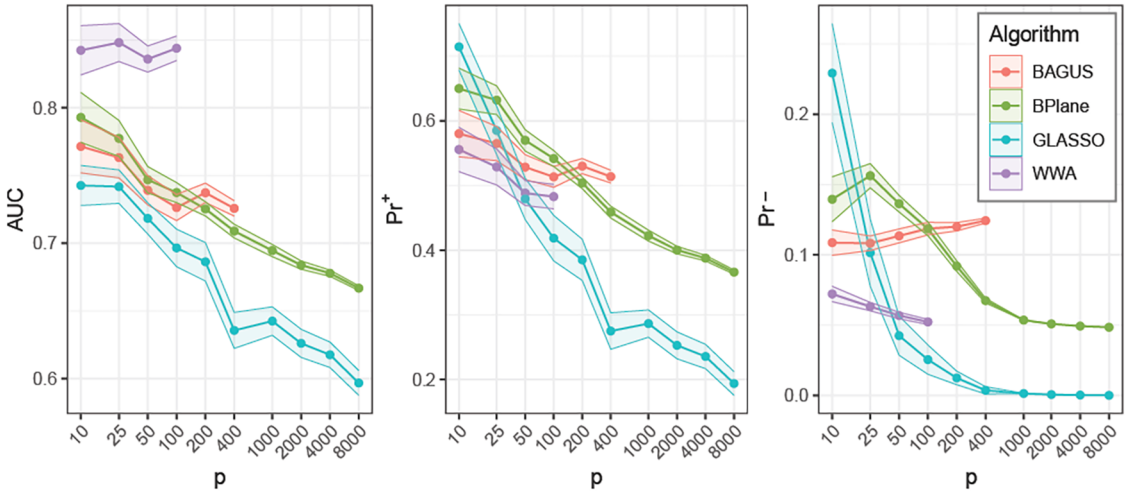


Figure 2. Comparison of edge detection scores using AUC, Pr^+, and Pr^-.

parameters included (a) parent ion mass tolerance of ±6 ppm, (b) at least partial tryptic digestion with 2 missed cleaved sites allowed, (c) static modification of cysteine carbamidomethylation (+57.0215 Da) and N-terminal/lysine TMT labeling (+229.1629 Da), and (d) variable modifications of methionine oxidation (+15.9949 Da). Peptide-spectrum matches, peptides, and proteins were filtered based on MS-GF probabilities to a false-discovery rate of <1% at all levels. Quantitative information was extracted from the TMT reporter ions using MASIC.[46]

Peptide-level data were quantified to the protein level using the `pmartR` package,[7,8] which yielded 12,375 proteins. As is common in LC-MS proteomic data sets, a substantial fraction of data was missing.[47] Prior proteomics studies have addressed missingness in data sets through complete-case filtering i.e., restricting downstream analyses to proteins quantified across all samples, in order to avoid bias introduced by imputation methods.[48−52] As such, we limited our analysis to the 6968 proteins containing no missing values across samples. We used the BPlane algorithm to estimate the precision matrix of order 6968 for each of the three infection groups.

Comparison to IntAct Database

The top "rewired" proteins (ones with different edges between the mock and infection conditions) were then extracted for the mock vs Washington and mock vs Italy comparisons. To investigate a tractable subset of proteins, the top 100 rewired proteins from each comparison were selected. All proteins involved in the human SARS-CoV-2 infection were extracted from the IntAct[53] database and matched to the top proteins to flag known proteins implicated in COVID-19 infection. This analysis serves as a qualitative assessment of the biologically plausibility of the network. IntAct[53] annotations are used solely to identify the overlap between proteins, and absence from IntAct[53] is not taken as evidence of biological irrelevance.

RESULTS

Simulation Studies

Figure 1 displays the mean computation time per graph, plus or minus two standard errors, based on 100 replications. Note that we did not run WWA for $p \geq 200$ and BAGUS for $p \geq 1000$ as they became too computationally expensive. Comparing the computation times, it is evident that WWA takes the longest, followed by BAGUS, and BPlane and GLASSO take the least time. WWA is hampered by sampling from the *G*-Wishart distribution, which has remained computationally expensive.[54] The BAGUS algorithm is more computationally expensive than BPlane, and the gap between the two increases as the number of nodes increases. These results agree with the computational complexities of BPlane and BAGUS being $O(np^2)$ to $O(p^3)$, respectively, when $n < p$. Somewhat surprisingly, the computation times of BPlane and GLASSO are competitive, but this supports the argument that "the idea that Bayesian algorithms lack the computational advantage of their frequentist counterparts is outdated."[35]

We used AUC, Pr^+, and Pr^- to evaluate the performance of each method's ability to recover simulated, ground-truth edges.

https://doi.org/10.1021/acs.jproteome.5c01277
J. Proteome Res. 2026, 25, 4306−4314

Figure 2 displays the means of each measure, plus or minus two standard errors, based on 100 replications. Comparing the AUC of WWA to the other algorithms, it can be argued that the computational expense of the WWA method pays off. However, WWA is not a computationally feasible option for estimating graphs for omics data sets with thousands of biomolecules. The other three methods perform more similarly in terms of AUC, with AUC decreasing with the size of the graphs. More detailed comparisons between the AUCs of BPlane and the other methods, including confidence intervals and p-values, are provided in the supplement. The Pr^+ values decrease with the size of the graphs as well and demonstrate that the Bayesian methods could benefit from the incorporation of prior information in $n \ll p$ cases. Other metrics, including the sensitivity, specificity, Matthews correlation coefficient, and F1 score, are given in the Supplement.

These simulations suggest that BPlane provides a Bayesian approach that scales computationally to large graphs better than WWA and BAGUS and scales as well as GLASSO while remaining competitive with BAGUS and GLASSO in terms of edge detection performance.

The results of the simulation study evaluating the effect of prior information are summarized in Table 1. Using BPlane-based posterior edge probabilities, we calculated the AUC for each of 500 simulated data sets when different proportions of true edges were assigned informative prior probabilities of 0.9. Results indicate that even when the number of samples is small relative to the number of biomolecules, such as for the simulated data with $n = 50$ and $p = 2000$, prior information can greatly increase the predictive ability of BPlane.

Table 1. Means and Standard Deviations of AUCs for the 500 Simulated Datasets when Different Proportions of True Edges Possess Informative Prior Probabilities

	AUC	
informative proportion	mean	SD
0.00	0.685	0.016
0.25	0.758	0.012
0.50	0.835	0.008
0.75	0.915	0.004
1.00	0.998	0.0002

Sars-Cov-2 Dataset

We tested the BPlane algorithm on the SARS-CoV-2 proteomics data set described previously, comparing data from the strains arising in Washington state (USA) and Italy with a mock virus. Three donors contributed five replicate HAE cells to be measured at each of six time points—that is, with different replicates measured at each time point—every 12 h postinfection for 72 h. Thus, the collected data set consisted of 90 samples from each virus. Analysis was limited to the last three time points (at 48, 60, and 72 h postinfection)—that is, 45 samples for each condition—as a compromise between sample size and consistency of the viral time conditions. To ensure no model assumptions would be violated, we analyzed the correlation of protein profiles between samples from the same donor and samples from different donors and found no significant difference in correlation values. Thus, no adjustment to the data or models was needed to account for a donor effect.

The BPlane algorithm took between 176, 139, and 162 EM iterations to achieve convergence for the Washington, Italy, and mock samples, respectively, taking approximately 66 min of total computation time. The same computational settings as in the simulation studies were used with no parallel processing. When we tried the BAGUS algorithm on the same data for comparison, the computation progress was remarkably slow: it took three times as long to complete only the first EM iteration on the Washington data as BPlane took for convergence of all three data sets.

We estimated the proteomic network for each infection group by determining an edge present when the posterior edge probability was greater than 0.5. Figure 3 summarizes the overlap of the edges for the three graphs using an UpSet plot. We note that the graphs for the two SARS-CoV-2 strains have more edges in common than either has with the graph for the mock virus, as expected. This data application demonstrates that the BPlane algorithm allows for computationally feasible Bayesian analysis of a large proteomics data set that would take significantly longer time using BAGUS. To address the stability

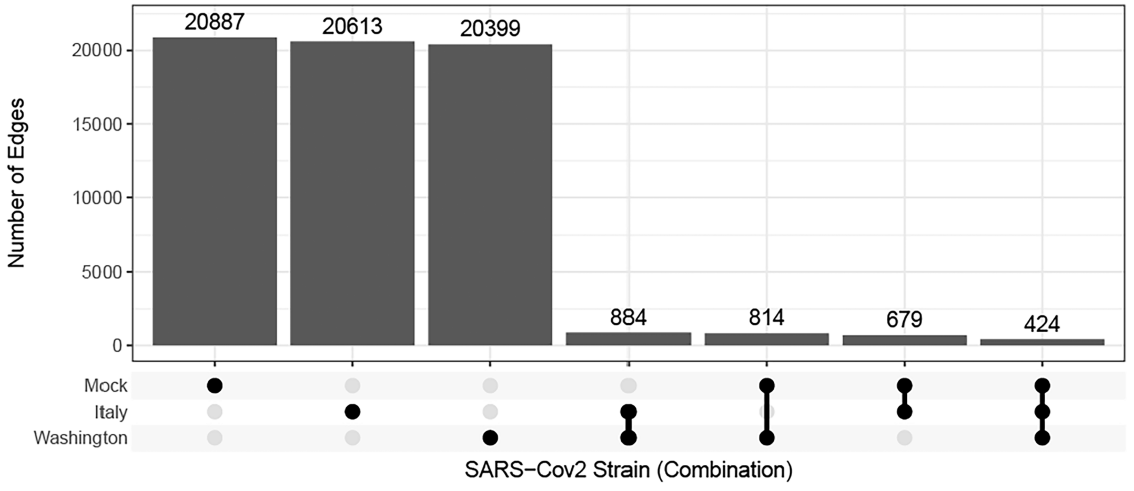


Figure 3. Edge overlap distribution from the three proteomics networks estimated by BPlane associated with the Washington, Italy, and mock SARS-CoV-2 strains.

https://doi.org/10.1021/acs.jproteome.5c01277
J. Proteome Res. 2026, 25, 4306−4314

Table 2. Subset of the Top 100 Most Changed Proteins between Mock and Infection (either Washington or Italy) with a Citation in the IntAct Database Supporting Protein−Protein Interaction Network Rewiring Following Human SARS-CoV-2 Infection

human protein (UniProt ID | gene name)	# changes	comparison	interacting SARS-CoV-2 Protein(s)	PubMed ID(s)
P07954 | FUMH	6	Washington vs Mock	P0DTD8	33845483
P30153 | 2AAA	6	Washington vs Mock	P0DTC7, P0DTD3	32838362
Q68DH5 | LMBD2	6	Washington vs Mock	P0DTC5	33845483
Q6NXT6 | TAPT1	6	Washington vs Mock	P0DTD3	32353859, 33060197
Q8IZQ1 | WDFY3	6	Washington vs Mock	P0DTC5	33845483
O95613 | PCNT	5	Washington vs Mock	P0DTC6	32838362
P05023 | AT1A1	5	Washington vs Mock	A0A663DJA2, P0DTC3, P0DTC4, P0DTC5, P0DTC6, P0DTC7, P0DTC8	32838362, 34901782, 36217030
P05161 | ISG15	5	Washington vs Mock	P0DTD1	32428392, 34270554, 32726803
P09132 | SRP19	5	Washington vs Mock	P0DTC9	34578187
P10599 | THIO	5	Washington vs Mock	P0DTC9	34901782
P13667 | PDIA4	5	Washington vs Mock	P0DTC3	32838362
P29317 | EPHA2	5	Washington vs Mock	P0DTC3, P0DTD8	33845483
P46379 | BAG6	5	Washington vs Mock	P0DTC3, P0DTC7, P0DTC8, P0DTD3	32838362
P60033 | CD81	5	Washington vs Mock	P0DTD8	33845483
P67809 | YBOX1	5	Washington vs Mock	P0DTC9	34232536
P60033 | CD81	7	Italy vs Mock	P0DTD8	33845483
O43324 | MCA3	6	Italy vs Mock	P0DTD2	36217030
P07954 | FUMH	6	Italy vs Mock	P0DTD8	33845483
P23468 | PTPRD	6	Italy vs Mock	P0DTC3	32838362
P61006 | RAB8A	6	Italy vs Mock	P0DTC3, P0DTD1	33845483, 36217030
Q13535 | ATR	6	Italy vs Mock	P0DTC7	32838362, 33845483
Q15036 | SNX17	6	Italy vs Mock	P0DTC3, P0DTD8	33845483
Q6NT16 | S18B1	6	Italy vs Mock	P0DTC5	33845483
Q8WVV9 | HNRLL	6	Italy vs Mock	P0DTD1	34159380
Q9UBQ0 | VPS29	6	Italy vs Mock	P0DTC2	34504087
O14828 | SCAM3	5	Italy vs Mock	P0DTD8	33845483
O43156 | TTI1	5	Italy vs Mock	P0DTC7	33845483
P09669 | COX6C	5	Italy vs Mock	P0DTC9	34578187
P22087 | FBRL	5	Italy vs Mock	P0DTC9, P0DTD2	34029587, 34578187, 34232536, 34901782, 36217030
P26045 | PTN3	5	Italy vs Mock	P0DTC4	36115835
P26641 | EF1G	5	Italy vs Mock	P0DTC9	34901782
P28300 | LYOX	5	Italy vs Mock	P0DTC8	32353859, 33060197
P29317 | EPHA2	5	Italy vs Mock	P0DTC3, P0DTD8	33845483
P31040 | SDHA	5	Italy vs Mock	P0DTF1	36217030
P31431 | SDC4	5	Italy vs Mock	P0DTC2	34069441
P40925 | MDHC	5	Italy vs Mock	P0DTC3	33845483
P46782 | RS5	5	Italy vs Mock	P0DTC9, P0DTD3	34029587, 32838362
P51114| FXR1	5	Italy vs Mock	P0DTC9	34029587
P51570| GALK1	5	Italy vs Mock	P0DTD3	32838362
P58107| EPIPL	5	Italy vs Mock	P0DTC2, P0DTD3	36217030
P68431 | H31	5	Italy vs Mock	P0DTD2	36217030
P84090 | ERH	5	Italy vs Mock	P0DTC9	34901782
Q10469 | MGAT2	5	Italy vs Mock	P0DTC3	32838362
Q13596 | SNX1	5	Italy vs Mock	P0DTC3	33845483
Q13823 | NOG2	5	Italy vs Mock	P0DTC9	34578187, 36217030

https://doi.org/10.1021/acs.jproteome.5c01277
J. Proteome Res. 2026, 25, 4306−4314

of the inferred graphs, we present a bootstrap-based analysis in the Supporting Information.

Next, to assess highly rewired proteins (those with the most changed edges) between the mock vs Washington strain and the mock vs Italy strain, IntAct[53] was queried for experimentally supported interactions between human proteins and SARS-CoV-2. Multiple top-rank proteins have been independently reported to interact with SARS-CoV-2 proteins[55−70] (Table 2). The overlap between our most rewired nodes and curated SARS-CoV-2 host interactors supports that the inferred network rewiring highlights proteins already implicated in infection.

CONCLUSION

Bayesian analyses of graphical networks for omics data have several advantages over frequentist analysis, including production of readily interpretable edge probabilities, incorporation of prior biological information, and a framework for combining data from multiple experiments. Yet Bayesian approaches have previously been generally limited to networks with hundreds of nodes due to prohibitive computational time, when omics data often consists of tens of thousands of biomolecules. To bridge this gap, we illustrated the BPlane algorithm in this paper. Adopting the Bayesian spike-and-slab GGM fit through the BAGUS algorithm,[25] BPlane speeds up BAGUS by implementing the modified log pseudolikelihood introduced in the CONCORD algorithm.[20] This reduces the computational complexity of each EM iteration from $O(p^3)$ to $O(np^2)$ when $n < p$, resulting in substantial savings for omics data sets, where often $n \ll p$.

In simulation experiments, we demonstrated that substantial computational savings of BPlane over two Bayesian algorithms and competitive computational time with GLASSO. Meanwhile, BPlane produces comparable edge detection performance with these other algorithms. When applied to LC-MS proteomics data with three groups of 45 samples on 6968 proteins, the BPlane algorithm converged after only 66 min of run time on a personal computer, while BAGUS took three times as long to complete only a single EM iteration on one of the groups. These real data analyses offer compelling validation that BPlane can extend the capabilities of Bayesian analysis for large omics data sets past what was previously possible.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.5c01277.

Mathematical details for the Bayesian model framework, the BPlane algorithm, and practical considerations (Section S1); additional details for simulation methods (Section S2); additional results for simulation studies (Section S3) (PDF)

AUTHOR INFORMATION

Corresponding Author

Lisa M. Bramer − *Pacific Northwest National Laboratory, Richland, Washington 99354, United States;* orcid.org/0000-0002-8384-1926; Email: lisa.bramer@pnnl.gov

Authors

Daniel W. Adrian − *Grand Valley State University, Allendale, Michigan 49401, United States;* orcid.org/0009-0005-4390-8670

Erik D. VonKaenel − *Pacific Northwest National Laboratory, Richland, Washington 99354, United States;* orcid.org/0000-0002-8933-7413

Moses Y. Obiri − *Pacific Northwest National Laboratory, Richland, Washington 99354, United States*

David J. Degnan − *Pacific Northwest National Laboratory, Richland, Washington 99354, United States;* orcid.org/0000-0001-5737-7173

Amy C. Sims − *Pacific Northwest National Laboratory, Richland, Washington 99354, United States*

Kristie Oxford − *Pacific Northwest National Laboratory, Richland, Washington 99354, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.5c01277

Author Contributions

M.Y.O and E.D.V provided statistical support. D.J.D. provided biological support. L.M.B. conceptualized and supervised the project. A.C.S. and K.O. provided the dataset and its technical description. D.W.A. wrote the manuscript with significant contributions from all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the PNNL Laboratory Directed Research and Development Program and is a contribution of the Predictive Phenomics Initiative. PNNL is operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract No. DEAC05-76RL01830.

ABBREVIATIONS

AUC, area under the receiver operating characteristic (ROC) curve; BAGUS, Bayesian regularization for graphical models with unequal shrinkage; BPLANE, Bayesian pseudolikelihood-based algorithm for network estimation; CONCORD, convex correlation selection method; EM, expectation-maximization; GEMBAG, group estimation of multiple Bayesian graphical models; GGM, Gaussian graphical model; GLASSO, graphical LASSO; KEGG, Kyoto encyclopedia of genes and genomes; LASSO, least absolute shrinkage and selection operator; LC-MS, liquid chromatography−mass spectrometry; MCMC, Markov chain Monte Carlo; PNNL, Pacific Northwest National Laboratory; SPACE, sparse partial correlation estimation; SPLICE, sparse pseudolikelihood inverse covariance estimation

REFERENCES

(1) Mathé, E.; John, L. H.; Daniel, G. S.; James, L. C. The Omics Revolution Continues: The Maturation of High-Throughput Biological Data Sources. *Yearbook Med. Inf.* **2018**, *27* (1), 211−222.

(2) Barabási, A.-L.; Gulbahce, N.; Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **2011**, *12* (1), 56−68.

(3) Goh, K.-I.; Michael, E. C.; David, V.; Barton, C.; Marc, V.; Albert-László, B. The Human Disease Network. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (21), 8685−8690.

(4) Hopkins, A. L. Predicting promiscuity. *Nature* **2009**, *462*, 167−168.
(5) Schadt, E. E.; Stephen, H. F.; David, A. S. A network view of disease and compound screening. *Nat. Rev. Drug Discovery* **2009**, *8*, 286−295.
(6) Shutta, K. H.; De, V. R.; Denise, M. S.; Raji, B. Gaussian graphical models with applications to omics analyses. *Stat. Med.* **2022**, *41*, 5150−5187.
(7) Degnan, D. J.; Kelly, G. S.; Rachel, R.; Daniel, C.; Evan, A. M.; Nathan, A. J.; Damon, L.; Bobbie-Jo, M. W.-R.; Lisa, M. B. pmartR 2.0: A Quality Control, Visualization, and Statistics Pipeline for Multiple Omics Datatypes. *J. Proteome Res.* **2023**, *22* (2), 570−576.
(8) Stratton, K. G.; Kelly, G. S.; Webb-Robertson, B. J.; Bobbie-Jo, M. W.-R.; McCue, L. A.; Ann, M. L.; Stanfill, B.; Bryan, S.; Claborne, D.; Daniel, C.; Godinez, I.; Iobani, G.; Johansen, T.; Thomas, J.; Thompson, A. M.; Allison, M. T.; Burnum-Johnson, K. E.; Kristin, E. B.-J.; Waters, K. M.; Katrina, M. W. pmartR: Quality Control and Statistics for Mass Spectrometry-Based Biological Data. *J. Proteome Res.* **2019**, *18*, 1418−1425.
(9) Steffen, L. L. *Graphical Models*; Oxford University Press, 1996.
(10) Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58* (1), 267−288.
(11) Nicolai, M.; Peter, B. High-dimensional graphs and variable selection with the lasso. *Ann. Stat.* **2006**, *34* (3), 1436−1462.
(12) Friedman, J.; Trevor, H.; Robert, T. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **2008**, *9* (3), 432−441.
(13) Rahul, M.; Trevor, H. The graphical lasso: New insights and alternatives. *Electron. J. Stat.* **2012**, *6*, 2125−2149.
(14) Witten, D. M.; Jerome, H. F.; Noah, S. New Insights and Faster Computations for the Graphical Lasso. *J. Comput. Graphical Stat.* **2011**, *20* (4), 892−900.
(15) Danaher, P.; Pei, W.; Daniela, M. W. The joint graphical lasso for inverse covariance estimation across multiple classes. *J. R. Stat. Soc., Ser. B* **2014**, *76* (2), 373−397.
(16) Besag, J. Statiatical analysis of non-lattice data. *J. R. Stat. Soc., Ser. D* **1975**, *24* (3), 179−195.
(17) Guilherme, V. R.; Peng, Z.; Bin, Y. *A Path Following Algorithm for Sparse Pseudo-Likelihood Inverse Covariance Estimation*; SPLICE, 2008.
(18) Peng, J.; Pei, W.; Nengfeng, Z.; Ji, Z. Partial Correlation Estimation by Joint Sparse Regression Models. *J. Am. Stat. Assoc.* **2009**, *104* (486), 735−746.
(19) Jerome, F.; Trevor, H.; Robert, T. *Applications of the Lasso and Grouped Lasso to the Estimation of Sparse Graphical Models*; Stanford University, 2010.
(20) Khare, K.; Sang-Yun, O.; Bala, R. A convex pseudolikelihood framework for high dimensional partial correlation estimation with convergence guarantees. *J. R. Stat. Soc., Ser. B* **2015**, *77* (4), 803−825.
(21) Kanehisa, M.; Miho, F.; Yoko, S.; Yuriko, M.; Mari, I.-W. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **2025**, *53* (D1), D672−D677.
(22) Bateman, A.; Martin, M. J.; Orchard, S.; The UniProt Consortium; et al. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **2023**, *51* (D1), D523−D531.
(23) Degnan, D. J.; Clayton, W. S.; Moses, Y. O.; Erik, D. V.; Grace, S. K.; James, D. K.; David, L. N.; Karl, T. L. P.; Lisa, M. B. Protein-Protein Interaction Networks Derived from Classical and Machine Learning-Based Natural Language Processing Tools. *J. Proteome Res.* **2024**, *23*, 5395−5404.
(24) Leach, D. T.; Kelly, G. S.; Jan, I.; Rachel, R.; Bobbie-Jo, M. W.-R.; Lisa, M. B. malbacR: A Package for Standardized Implementation of Batch Correction Methods for Omics Data. *Anal. Chem.* **2023**, *95*, 12195−12199.
(25) Lingrui, G.; Naveen, N. N.; Feng, L. Bayesian Regularization for Graphical Models With Unequal Shrinkage. *J. Am. Stat. Assoc.* **2019**, *114* (527), 1218−1231.
(26) Xinming, Y.; Lingrui, G.; Naveen, N. N.; Feng, L. GemBag: Group Estimation of Multiple Bayesian Graphical Models. *J. Mach. Learn. Res.* **2021**, *22*, 1−48.
(27) Geoffrey, J. M.; Thriyambakam, K. *The EM Algorithm and Extensions*; Wiley Interscience, 2007.
(28) Vogels, L.; Reza, M.; Marit, S.; Ş. İlker, B. Bayesian Structure Learning in Undirected Gaussian Graphical Models: Literature Review with Empirical Comparison. *J. Am. Stat. Assoc.* **2024**, *119* (548), 3164−3182.
(29) Veronika, R.; Edward, I. G. The Spike-and-Slab LASSO. *J. Am. Stat. Assoc.* **2018**, *113*, 431−444, DOI: 10.1080/01621459.2016.1260469.
(30) van den Boom, W.; Beskos, A.; De Iorio, M. The G-Wishart weighted proposal algorithm: Efficient posterior computation for Gaussian graphical models. *J. Comput. Graphical Stat.* **2022**, *31* (4), 1215−1224.
(31) Mohammadi, R.; Wit, E. C. BDgraph: An R package for Bayesian structure learning in graphical models. *J. Stat Software* **2019**, *89*, 1−30.
(32) Newman, M. E. J. The Structure and Function of Complex Networks. *SIAM Rev.* **2003**, *45* (2), 167−256.
(33) Albert, R. Scale-free networks in cell biology. *J. Cell Sci.* **2005**, *118* (21), 4947−4957.
(34) Zhao, T.; Liu, H.; Roeder, K.; Lafferty, J.; Wasserman, L. High-dimensional undirected graph estimation *Version R package version* 2011; Vol. *1* 5.
(35) Vogels, L.; Mohammadi, R.; Schoonhoven, M.; Birbil, Ş. İ. Bayesian structure learning in undirected Gaussian graphical models: Literature review with empirical comparison. *J. Am. Stat. Assoc.* **2024**, *119* (548), 3164−3182.
(36) Team, R. D. C. *R: A Language and Environment for Statistical Computing*. 2025.
(37) Fulcher, M. L.; Gabriel, S.; Burns, K. A.; Yankaskas, J. R.; Randell, S. H. Well-differentiated human airway epithelial cell cultures. *Methods Mol. Med.* **2005**, *107*, 183−206.
(38) Nakayasu, E. S.; Nicora, C. D.; Sims, A. C.; Burnum-Johnson, K. E.; Kim, Y. M.; Kyle, J. E.; Matzke, M. M.; Shukla, A. K.; Chu, R. K.; Schepmoes, A. A.; et al. MPLEx: a Robust and Universal Protocol for Single-Sample Integrative Proteomic, Metabolomic, and Lipidomic Analyses. *mSystems* **2016**, *1* (3), No. e00043-16, DOI: 10.1128/mSystems.00043-16.
(39) Burnum-Johnson, K. E.; Kyle, J. E.; Eisfeld, A. J.; Casey, C. P.; Stratton, K. G.; Gonzalez, J. F.; Habyarimana, F.; Negretti, N. M.; Sims, A. C.; Chauhan, S.; et al. MPLEx: a method for simultaneous pathogen inactivation and extraction of samples for multi-omics profiling. *Analyst* **2017**, *142* (3), 442−448.
(40) Nicora, C. D.; Sims, A. C.; Bloodsworth, K. J.; Kim, Y. M.; Moore, R. J.; Kyle, J. E.; Nakayasu, E. S.; Metz, T. O. Metabolite, Protein, and Lipid Extraction (MPLEx): A Method that Simultaneously Inactivates Middle East Respiratory Syndrome Coronavirus and Allows Analysis of Multiple Host Cell Components Following Infection. *Methods Mol. Biol.* **2020**, *2099*, 173−194.
(41) Piehowski, P. D.; Petyuk, V. A.; Orton, D. J.; Xie, F.; Moore, R. J.; Ramirez-Restrepo, M.; Engel, A.; Lieberman, A. P.; Albin, R. L.; Camp, D. G.; et al. Sources of technical variability in quantitative LC-MS proteomics: human brain tissue sample analysis. *J. Proteome Res.* **2013**, *12* (5), 2128−2137.
(42) Hutchinson-Bunch, C.; Sanford, J. A.; Hansen, J. R.; Gritsenko, M. A.; Rodland, K. D.; Piehowski, P. D.; Qian, W. J.; Adkins, J. N. Assessment of TMT Labeling Efficiency in Large-Scale Quantitative Proteomics: The Critical Effect of Sample pH. *ACS Omega* **2021**, *6* (19), 12660−12666.
(43) Kwan, Y. W.; To, K. W.; Lau, W. M.; Tsang, S. H. Comparison of the vascular relaxant effects of ATP-dependent K+ channel openers on aorta and pulmonary artery isolated from spontaneously hypertensive and Wistar-Kyoto rats. *Eur. J. Pharmacol.* **1999**, *365* (2−3), 241−251.
(44) Gibbons, B. C.; Chambers, M. C.; Monroe, M. E.; Tabb, D. L.; Payne, S. H. Correcting systematic bias and instrument measurement drift with mzRefinery. *Bioinformatics* **2015**, *31* (23), 3838−3840.

https://doi.org/10.1021/acs.jproteome.5c01277
J. Proteome Res. 2026, 25, 4306−4314

(45) Kim, S.; Pevzner, P. A. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat. Commun.* **2014**, *5*, No. 5277.
(46) Monroe, M. E.; Shaw, J. L.; Daly, D. S.; Adkins, J. N.; Smith, R. D. MASIC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS) features. *Comput. Biol. Chem.* **2008**, *32* (3), 215−217.
(47) Webb-Robertson, B.-J.; Bobbie-Jo, M. W.-R.; Wiberg, H. K.; Holli, K. W.; Matzke, M. M.; Melissa, M. M.; Brown, J. N.; Joseph, N. B.; Wang, J.; Jing, W.; McDermott, J. E.; Jason, E. M.; Smith, R. D.; Richard, D. S.; Rodland, K. D.; Karin, D. R.; Metz, T. O.; Thomas, O. M.; Pounds, J. G.; Joel, G. P. Review, Evaluation, and Discussion of the Challenges of Missing Value Imputation for Mass Spectrometry-Based Label-Free Global Proteomics. *J. Proteome Res.* **2015**, *14*, 1993−2001.
(48) Chang, H. Y.; Deng, Y.; Li, R.; Avtonomov, D.; Wen, B.; Haynes, S. E.; Leprevost, F. d. V.; Zhang, B.; Yu, F.; Nesvizhskii, A. I. Analysis of isobaric quantitative proteomic data using TMT-Integrator and FragPipe computational platform. *Nat. Commun.* **2026**, *17*, No. 410, DOI: 10.1038/s41467-026-70118-7.
(49) Kong, W.; Hui, H. W. H.; Peng, H.; Goh, W. W. B. Dealing with missing values in proteomics data. *Proteomics* **2022**, *22* (23−24), No. e2200092.
(50) Liu, Q.; Zhang, J.; Guo, C.; Wang, M.; Wang, C.; Yan, Y.; Sun, L.; Wang, D.; Zhang, L.; Yu, H.; et al. Proteogenomic characterization of small cell lung cancer identifies biological insights and subtype-specific therapeutic strategies. *Cell* **2024**, *187* (1), 184−203 e28.
(51) Saei, A. A.; Gullberg, H.; Sabatier, P.; Beusch, C. M.; Johansson, K.; Lundgren, B.; Arvidsson, P. I.; Arner, E. S. J.; Zubarev, R. A. Comprehensive chemical proteomics for target deconvolution of the redox active drug auranofin. *Redox Biol.* **2020**, *32*, No. 101491.
(52) Tabb, D. L.; Kaniyar, M. H.; Bringas, O. G. R.; Shin, H.; Di Stefano, L.; Taylor, M. S.; Xie, S.; Yilmaz, O. H.; LaCava, J. Interrogating data-independent acquisition LC-MS/MS for affinity proteomics. *J. Proteins Proteom* **2024**, *15* (3), 281−298.
(53) Del Toro, N.; Shrivastava, A.; Ragueneau, E.; Meldal, B.; Combe, C.; Barrera, E.; Perfetto, L.; How, K.; Ratan, P.; Shirodkar, G.; et al. The IntAct database: efficient access to fine-grained molecular interaction data. *Nucleic Acids Res.* **2022**, *50* (D1), D648−D653.
(54) Wang, H.; Li, S. Z. Efficient Gaussian graphical model determination under G-Wishart prior distributions. *Electron. J. Statist.* **2012**, *6*, 168−198, DOI: 10.1214/12-ejs669.
(55) Armstrong, L. A.; Lange, S. M.; Cesare, V. D.; Matthews, S. P.; Nirujogi, R. S.; Cole, I.; Hope, A.; Cunningham, F.; Toth, R.; Mukherjee, R.; et al. Biochemical characterization of protease activity of Nsp3 from SARS-CoV-2 and its inhibition by nanobodies. *PLoS One* **2021**, *16* (7), No. e0253364.
(56) Cai, T.; Yu, Z.; Wang, Z.; Liang, C.; Richard, S. Arginine methylation of SARS-Cov-2 nucleocapsid protein regulates RNA binding, its ability to suppress stress granule formation, and viral replication. *J. Biol. Chem.* **2021**, *297* (1), No. 100821.
(57) Cattin-Ortolá, J.; Welch, L. G.; Maslen, S. L.; Papa, G.; James, L. C.; Munro, S. Sequences in the cytoplasmic tail of SARS-CoV-2 Spike facilitate expression at the cell surface and syncytia formation. *Nat. Commun.* **2021**, *12* (1), No. 5333.
(58) Chen, Z.; Wang, C.; Feng, X.; Nie, L.; Tang, M.; Zhang, H.; Xiong, Y.; Swisher, S. K.; Srivastava, M.; Chen, J. Interactomes of SARS-CoV-2 and human coronaviruses reveal host factors potentially affecting pathogenesis. *EMBO J.* **2021**, *40* (17), No. EMBJ2021107776.
(59) Freitas, B. T.; Durie, I. A.; Murray, J.; Longo, J. E.; Miller, H. C.; Crich, D.; Hogan, R. J.; Tripp, R. A.; Pegan, S. D. Characterization and noncovalent inhibition of the deubiquitinase and deISGylase activity of SARS-CoV-2 papain-like protease. *ACS Infect. Dis.* **2020**, *6* (8), 2099−2109.
(60) Gogl, G.; Zambo, B.; Kostmann, C.; Cousido-Siah, A.; Morlet, B.; Durbesson, F.; Negroni, L.; Eberling, P.; Jané, P.; Nominé, Y.; et al. Quantitative fragmentomics allow affinity mapping of interactomes. *Nat. Commun.* **2022**, *13* (1), No. 5472.
(61) Gordon, D. E.; Hiatt, J.; Bouhaddou, M.; Rezelj, V. V.; Ulferts, S.; Braberg, H.; Jureka, A. S.; Obernier, K.; Guo, J. Z.; Batra, J.; et al. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* **2020**, *370* (6521), No. eabe9403.
(62) Gordon, D. E.; Jang, G. M.; Bouhaddou, M.; Xu, J.; Obernier, K.; White, K. M.; O'Meara, M. J.; Rezelj, V. V.; Guo, J. Z.; Swaney, D. L.; et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **2020**, *583* (7816), 459−468.
(63) Hudák, A.; Letoha, A.; Szilak, L.; Letoha, T. Contribution of syndecans to the cellular entry of SARS-CoV-2. *Int. J. Mol. Sci.* **2021**, *22* (10), No. 5336.
(64) Li, J.; Guo, M.; Tian, X.; Wang, X.; Yang, X.; Wu, P.; Liu, C.; Xiao, Z.; Qu, Y.; Yin, Y.; et al. Virus-host interactome and proteomic survey reveal potential virulence factors influencing SARS-CoV-2 pathogenesis. *Med* **2021**, *2* (1), 99−112. e7.
(65) Nabeel-Shah, S.; Lee, H.; Ahmed, N.; Burke, G. L.; Farhangmehr, S.; Ashraf, K.; Pu, S.; Braunschweig, U.; Zhong, G.; Wei, H.; et al. SARS-CoV-2 nucleocapsid protein binds host mRNAs and attenuates stress granules to impair host stress response. *iScience* **2022**, *25* (1), No. 103562, DOI: 10.1016/j.isci.2021.103562.
(66) Shin, D.; Mukherjee, R.; Grewe, D.; Bojkova, D.; Baek, K.; Bhattacharya, A.; Schulz, L.; Widera, M.; Mehdipour, A. R.; Tascher, G.; et al. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **2020**, *587* (7835), 657−662.
(67) Stukalov, A.; Girault, V.; Grass, V.; Karayel, O.; Bergant, V.; Urban, C.; Haas, D. A.; Huang, Y.; Oubraham, L.; Wang, A.; et al. Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* **2021**, *594* (7862), 246−252.
(68) Zheng, X.; Sun, Z.; Yu, L.; Shi, D.; Zhu, M.; Yao, H.; Li, L. Interactome analysis of the nucleocapsid protein of SARS-CoV-2 virus. *Pathogens* **2021**, *10* (9), No. 1155.
(69) Zheng, Y.-X.; Wang, L.; Kong, W.-S.; Chen, H.; Wang, X.-N.; Meng, Q.; Zhang, H.-N.; Guo, S.-J.; Jiang, H.-W.; Tao, S.-C. Nsp2 has the potential to be a drug target revealed by global identification of SARS-CoV-2 Nsp2-interacting proteins. *Acta Biochim. Biophys. Sin.* **2021**, *53* (9), 1134−1141.
(70) Zhou, Y.; Liu, Y.; Gupta, S.; Paramo, M. I.; Hou, Y.; Mao, C.; Luo, Y.; Judd, J.; Wierbowski, S.; Bertolotti, M.; et al. A comprehensive SARS-CoV-2−human protein−protein interactome reveals COVID-19 pathobiology and potential host therapeutic targets. *Nat. Biotechnol.* **2023**, *41* (1), 128−139.

https://doi.org/10.1021/acs.jproteome.5c01277
J. Proteome Res. 2026, 25, 4306−4314