

Gaussian Graphical Modeling Reveals Specific Lipid Correlations in Glioblastoma Cells

Nikola S. Mueller^a, Jan Krumsiek^b, Fabian J. Theis^b, Christian Böhm^c
and Anke Meyer-Baese^d

^aMax Planck Institute of Biochemistry, Martinsried, Germany;

^bHelmholz Zentrum, Neuherberg, Germany;

^cLudwig-Maximilians Universität, Munich, Germany;

^dFlorida State University, Tallahassee, USA

ABSTRACT

Advances in high-throughput measurements of biological specimens necessitate the development of biologically driven computational techniques. To understand the molecular level of many human diseases, such as cancer, lipid quantifications have been shown to offer an excellent opportunity to reveal disease-specific regulations. The data analysis of the cell lipidome, however, remains a challenging task and cannot be accomplished solely based on intuitive reasoning. We have developed a method to identify a lipid correlation network which is entirely disease-specific. A powerful method to correlate experimentally measured lipid levels across the various samples is a Gaussian Graphical Model (GGM), which is based on partial correlation coefficients. In contrast to regular Pearson correlations, partial correlations aim to identify only direct correlations while eliminating indirect associations. Conventional GGM calculations on the entire dataset can, however, not provide information on whether a correlation is truly disease-specific with respect to the disease samples and not a correlation of control samples. Thus, we implemented a novel differential GGM approach unraveling only the disease-specific correlations, and applied it to the lipidome of immortal Glioblastoma tumor cells. A large set of lipid species were measured by mass spectrometry in order to evaluate lipid remodeling as a result to a combination of perturbation of cells inducing programmed cell death, while the other perturbations served solely as biological controls. With the differential GGM, we were able to reveal Glioblastoma-specific lipid correlations to advance biomedical research on novel gene therapies.

Keywords: Correlation Networks, Partial Correlations, Gaussian Graphical Models, Lipidomics, Glioblastoma

1. INTRODUCTION

Despite recent progress in therapy and surgical intervention, Glioblastoma multiforms, malignant primary brain tumors, are nearly always fatal. The *in vitro* model of human Glioblastoma brain tumors is the U87 cell line, the major characteristic of which is its resistance to apoptosis (programmed cell death). Recent studies showed that the combined perturbation of gene transfection with the p53 tumor suppressor gene prior to chemotherapy with SN-38 triggers cell death in the (otherwise immortal) Glioblastoma cell line.^{1,2} At first a proteomic study showed a down-regulation of Galectin-1 in response to the combined perturbation,¹ which motivated the elucidation of lipid regulations.² In order to measure the lipidome, a specialized mass spectrometry (MS) technique was developed.³ On an organism-wide scale, changes in complex polar lipid levels were reliably identified. The set of all commonly regulated lipids might reveal dysregulations of e.g. metabolic pathways or functionally similar proteins. However, the molecular details of the perturbation-affected lipid coregulations still remain to be elucidated.

We aimed to identify partial correlations of lipid concentrations while accounting for the biological interpretation of the perturbation. To that end, we used Gaussian Graphical Models (GGMs), which are statistical graph models based on partial correlation coefficients. We chose to use a GGM over simple Pearson correlations since correlations are only detected for direct but not indirect dependencies.⁴ Beyond conventional GGM analysis,

Further author information: (Send correspondence to A.M-B.: E-mail: amb@eng.fsu.edu)

Independent Component Analyses, Wavelets, Neural Networks, Biosystems, and Nanoengineering IX,
edited by Harold Szu, Liyi Dai, Proc. of SPIE Vol. 8058, 805819 · © 2011 SPIE
CCC code: 0277-786X/11/\$18 · doi: 10.1117/12.884196

where one GGM is calculated for the entire data set, we introduce a disease-driven GGM calculation. With this here introduced differential GGM approach, we can now address the question whether a correlation in the GGM is biologically relevant or not. In general, not every identified correlation on the entire dataset is equally relevant to the disease, especially if the majority of the dataset are control measurements. While identifying only those lipids that respond to the biologically relevant perturbations but not to control perturbations, we answered the key question: Which lipids or lipid classes are co-affected by the perturbation by wild-type (wt) p53 transfection prior to SN-38 chemotherapy triggering apoptosis of the brain tumor cell lines?

2. GLIOBLASTOMA AND ITS LIPIDOME

U87 cells transfected with wt tumor suppressor gene p53 prior to treatment with the chemotherapeutic drug SN-38 underwent modest apoptosis and cell cycle arrest in G2, while chemotherapy alone did not trigger the same phenotype.¹ The reverse order of SN-38 treatment prior to p53 transfection results in almost complete apoptosis and complete G2 arrest. To analyze the lipid variations as a response to the effective perturbation, high-throughput MS/MS experiments were conducted as follows. Cell lysates of all perturbed cell lines were analyzed for variations of lipid levels (Fig. 1a).^{2,3} A specialized Fourier-Transform Ion-Cyclotron-Resonance (FT-ICR) MS/MS technique was developed to separate complex lipids.³ With the FT-ICR MS/MS, polar lipids, such as phospholipids, as well as complex glycolipids, such as gangliosides were reliably identified. Quantitative analysis of relative abundance profiles of polar lipids were obtained from cell lysates, whereby lipid levels were measured across six different perturbations and wt (without perturbation) with two technical replicates. Out of the large set of lipids, 167 polar lipids were measured with FT-ICR MS/MS across six lipid classes (varying primarily in their respective head groups). While lipid head groups can uniquely be identified with MS/MS, the associated fatty acid side chains cannot be independently resolved. An example for a complex lipid with ambiguous fatty acid side chains is PS(C36:4) that could have e.g. C18:2/C18:2 fatty acids incorporated, but also C16:0/C20:4 or C16:2/C20:2, etc.. Note, that some lipid classes, like gangliosides, have one variable and one fixed fatty acid side chain, thus, both side chains can unambiguously inferred. The MS/MS result – the matrix to be analyzed in this study – holds concentrations of lipids for each cell line for all perturbations.

Only the combined perturbation of p53 adenoviral transfection prior to SN-38 chemotherapy is biologically relevant for this study. In order to identify those lipids that specifically respond to the combined perturbation a

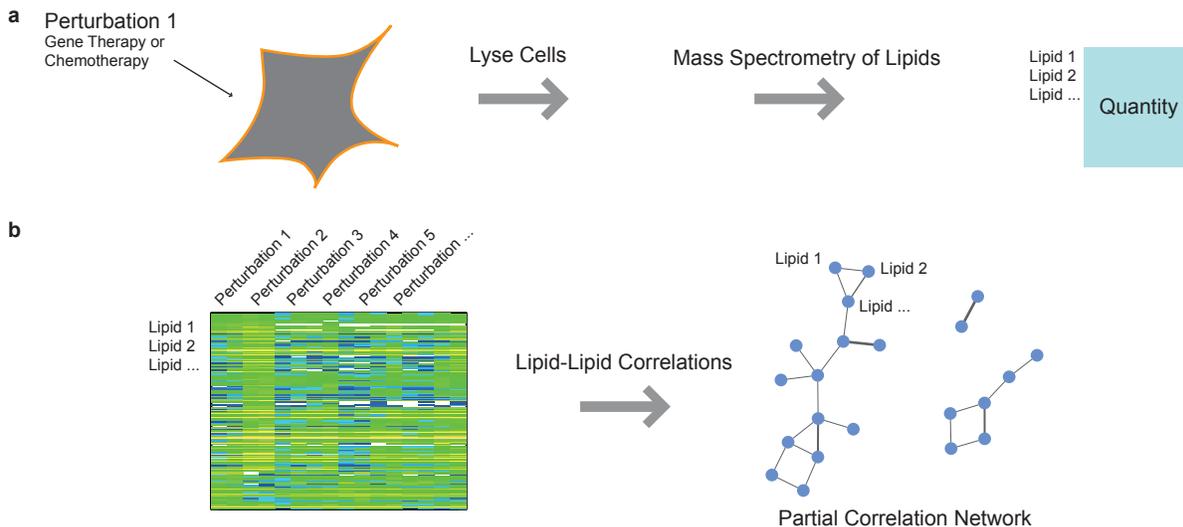


Figure 1. **From Cells to Lipid Correlations.** **a.** U87 cell lines were perturbed and subsequently lysed prior to MS analysis. Subsequently, lipid concentrations of 167 polar lipid species were obtained. **b.** The raw data of this study holds lipid quantifications for various perturbations (the samples). Pairwise correlations of lipids result in an undirected graph of lipid-to-lipid interactions holding the partial correlation values. Only statistically significant correlations are included in the resulting network. Edge widths indicate correlation strengths.

series of control experiments were conducted, which were permutations of the single perturbations as described previously,^{2,5} e.g. SN-38 alone, empty virus transfection or empty virus transfection prior to chemotherapy. In order to unravel the lipid remodeling that effected or was affected by apoptosis of U87 cells, the comparison of wt cell lines with the p53 plus SN-38 perturbations is not sufficient. For example, lipid remodeling can be the result of singular effects, like the transfection of the empty adenovirus, only the wt p53 adenovirus or solely the SN-38 chemotherapy. Only the entire dataset with all perturbations and wt allows to statistically exploit the wealth of all perturbation effects, which might not be feasible by comparing only two biologically relevant perturbations.

3. GAUSSIAN GRAPHICAL MODELS

Traditionally, correlation networks have been used to obtain information on coregulations of variables $L = (l_1, \dots, l_p)$, $|L| = p$ measured across all samples $S = (s_1, \dots, s_n)$, $|S| = n$; with $X = (x_{ls})$ the raw data matrix used for calculations. In case of the present metabolite data, a correlation coefficient will provide information on the degree of dependence between the measured variables. This pairwise correlation is thereby calculated based on the measurements across all samples – the cell lines with various perturbations (Fig. 1b).

The standard measure of pairwise correlations are Pearson product-moment correlation coefficients $P = (\rho_{ij})$, which quantify the linear dependency between two variables l_i and l_j . A common problem of Pearson correlation coefficients are indirect effects giving rise to a plethora of unspecifically high correlation coefficients throughout *omics* datasets.⁴ GGMs attempt to estimate conditional dependencies between measured variables over all samples rather than marginal dependencies, thereby eliminating such indirect correlations. The derivation of partial correlation coefficients can also be explained by linear regression: The partial correlation between the lipids l_1 and l_2 is the correlation of the residuals that result from linearly regressing l_1 and l_2 against the remaining lipids (l_3, \dots, l_p) .⁶ In our study, the partial correlation ζ_{ij} provides information on the coregulation of two lipids l_i, l_j .

To generate a GGM, the number of samples with respect to the number of variables determine the approach used for the calculation. If the number of samples n exceeds the number of variables p , full-order partial correlations $Z = (\zeta_{ij})$ can be calculated in a straight-forward manner from the inverse of the covariance matrix P as

$$\Omega = (\omega_{ij}) = P^{-1}$$

$$Z = (\zeta_{ij}) = -\omega_{ij} / \sqrt{\omega_{ii}\omega_{jj}}.$$

Statistical tests are next applied to determine whether a partial correlation ζ_{ij} is significantly different from zero ζ_{ij}^* (we mark a significant partial correlation with an asterisk) resulting in the GGM Z^* . Of the partial correlation matrix Z we construct Z^* as

$$Z^* = (\zeta_{ij}^*) = \begin{cases} \zeta_{ij} & \text{if } \zeta_{ij} \text{ is significant} \\ 0 & \text{else} \end{cases}$$

and we denote $\exists \zeta_{ij}^*$ for $\zeta_{ij}^* > 0$. A GGM is an undirected graph obtained by partial correlation calculation with subsequent statistical testing for edge significance (Fig. 2a). The graph nodes represent the measured variables whereas the edge weights correspond to significant partial correlation coefficients. If the number of samples is smaller than the number of variables ($n < p$) the straight-forward GGM calculation cannot be applied but a regularization and a likelihood estimation step have to be included. For $n < p$ the covariance matrix is rank-deficient,⁷⁻⁹ as a consequence the covariance matrix is not positive definite and can, thus, not be inverted. In the case of the present lipidomics data, we indeed have $n < p$ with $p = 157$ lipids and $n = 8^*$ samples. To estimate the GGM for $n < p$, Strimmer and colleagues¹⁰ introduced an all-in-one approach. One estimation step is a shrinkage approach and is applied^{9,10} to obtain the true correlation matrix \hat{P} . The other estimation step distinguishes actually existing edges from “null” edges in the GGM by fitting a statistical model assuming these two population of edges. The GGM is finally build by adjusting for local false-discovery rates (FDR).^{9,10} This method of regularized GGMs was already applied to transcriptomics datasets^{6,11} and will here be applied to our lipidomics dataset.

*Eight samples were measured with two technical replicates. Analyses were performed on the raw data including the replicates.

When calculating the GGM, all samples are assumed to be independent,⁷ but inspection of the present lipidome dataset showed a strong correlation between all samples. Although correlations between the technical replicates were higher than between perturbations, the overall correlation of disease and control samples was very high ($> .95$). In case of dependent samples the covariance estimates are no longer optimal: its standard deviation monotonically increases with larger correlation coefficients of samples.⁸ Note that the result of the strong correlation between all samples already indicates that the successful perturbation of cells transfected with wt p53 prior to SN-38 chemotherapy has strong effects only on few lipids and not the lipid levels in general. To account for the high dependencies between samples, we calculated the GGM mimicking that all samples are replicates of one another. Since seven of the eight samples are only measured as controls (which were introduced as control replicates with respect to the one perturbation of interest), this approach is reasonable for our study.

4. DIFFERENTIAL GGM

To identify those partial correlations of lipids only resulting from the biologically relevant perturbation and not from side effects of one perturbation, we implemented the following concept of disease-specificity. For simplicity, we name the biologically relevant perturbation “disease” in contrast to the “controls” in the following, although this combination of perturbation is the one inhibiting tumor cell growth. Let S be the set of n samples composed of control and one disease sample $S = (s_1, \dots, s_n) = (s_D, s_{C_1}, \dots, s_{C_{n-1}}) = (s_D, s_C)$ with the disease sample s_D and the union of all control samples s_C . Imagine $\zeta^*(S)$ to be a significant correlation on the entire dataset S . It may then be a result of a perfect correlation of controls not substantially affected by the disease samples or be a result where primarily the disease samples induce a correlation on the entire dataset (controls alone are not correlated). In other words: if a correlation has no specific relevance to the disease, we would still detect a correlation when using a truncated dataset with solely control samples. These correlations, which are mostly a result of strong control sample correlation, can be considered “false positive” with respect to true disease relevance. In order to gather all truly disease relevant correlation, we also have to account for the reverse case, equivalently the “false negatives” with respect to disease relevance. If a correlation exists on the control samples s_C but is suppressed on the entire dataset S , the disease samples do not follow the correlation of the controls, wherein the correlation is again relevant with respect to the disease. This reverse case corresponds to the concept of suppressed variables, which denotes a variable to be a suppressor if it suppresses the correlation between some other variable to the remaining variables.^{12,13}

All disease relevant partial correlations were assessed in an approach inspired by jackknife resampling.¹⁴ GGMs are calculated by leaving out one sample from the dataset ($Z_{S \setminus s_i}^*$) during each iteration, resulting in a set of partial correlation coefficients for each lipid pair (l_i, l_j) of $\{\zeta^*(S), \zeta^*(S \setminus s_D), \zeta^*(S \setminus s_{C_1}), \dots, \zeta^*(S \setminus s_{C_{n-1}})\}$ for all existing significant partial correlations. Figure 2b illustrates the approach to build a differential GGM by evaluating the set of leave-one-out GGMs with respect to the criterion of disease-specificity. A pseudo-code formalizes the differential GGM approach:

```

ggm <- empty set of GGMs
ggm(0) <- result of GGM with S
for (i = 1:n){
  ggm(i) <- result of GGM with S \ Si
}

dGGM <- empty set of differential GGM edges
for (e=(li,lj) : all possible edges){
  if (e fulfills IAij w.r.t. ggm) {
    dGGM -> add e between nodes li and lj
  }
}
return dGGM

```

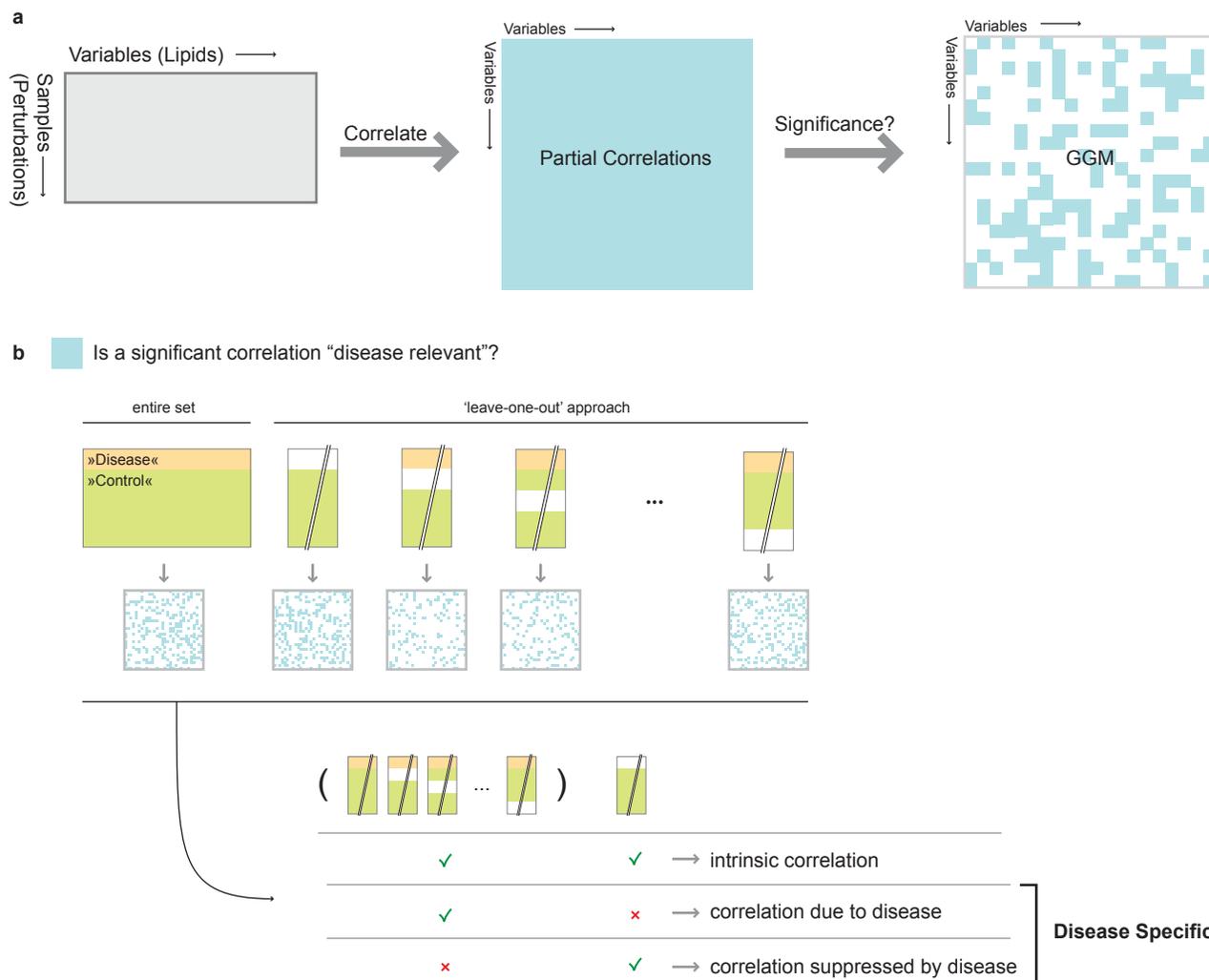


Figure 2. Raw data transformed to disease specific correlations. **a.** The lipidome raw data is a matrix of samples over variables. The samples are the individual perturbations which are grouped into control samples and the sample(s) of interest to the study, here simply called ‘disease’. Partial correlations of all variables are obtained and later evaluated with respect to statistical significance. **b.** To investigate whether a significant partial correlation is specific for the disease sample, partial correlations (as in **a**) were calculated for the entire dataset as well as for datasets where each one sample was left out. Unless a correlation is significant in all GGMs, it is considered disease-specific.

In detail, we extract those interactions IA_{ij} of (l_i, l_j) which fulfill the criterion to be disease relevant by comparing all GGMs with respect to the disease sample s_D as

$$IA_{ij} = [\neg \exists \zeta^*(S \setminus s_D) \wedge \forall_{s_i \in \{S, S \setminus s_C\}} \exists \zeta^*(s_i)] \vee [\exists \zeta^*(S \setminus s_D) \wedge \forall_{s_i \in \{S, S \setminus s_C\}} \neg \exists \zeta^*(s_i)].$$

In other words, we consider an edge disease-specific if it fulfills either one of two criteria: (1) The edge is not significant in the GGM of $S \setminus s_D$, the dataset S without the disease sample s_D , while it is significant in the GGM constructed from the entire dataset S as well as in all GGMs of $S \setminus s_C$ where each one control sample was left out for the calculation. (2) The reverse case holds if the edge is significant on the dataset without the disease sample ($S \setminus s_D$) – equivalent to a correlation of control samples – while the edge is not significant if the disease sample is present in the dataset (that are the datasets of S and any $S \setminus s_C$). As a result, we obtained one differential GGM of only direct lipid-lipid correlations resulting from the combination of wt p53 transfection prior to SN-38 chemotherapy for the Glioblastoma lipidome.

5. RESULTS

We generated GGMs for all perturbation combinations of the Glioblastoma lipidome according to out jackknife-inspired approach. The FDR cutoff value was set to $q = 0.01$. Compared to conventional GGM applications (analysis only of the entire dataset), we can break down each significant correlation with respect to the contribution of each sample. If we examine the lipidome solely from the perspective of conventional GGM calculations, we would obtain 256 significant lipid-lipid correlations. Thereof 25 correlations are disease relevant with respect to the perturbation of p53 gene therapy prior to SN-38 chemotherapy (Fig. 3a). Surprisingly, less than 10% of all significant interactions of a GGM from the entire dataset were actually disease-specific, or figuratively speaking true positive. Drawing any biological conclusions from correlations on the entire data set may therefore be misleading. In addition, we identified 9 significant lipid-lipid correlations which are suppressed by the disease relevant sample. Subsequently, we were able to identify 34 lipid-lipid interactions on the Glioblastoma lipidome which are significantly correlated upon p53 gene therapy prior to SN-38 chemotherapy.

The resulting disease-specific, differential GGM is depicted in Figure 3b. Since we obtained correlations across all six lipid species, our results are more comprehensive than the results of previous analyses^{2,5} where lipid species were always handled separately. Closer inspection of the 45 lipids involved in the disease relevant differential

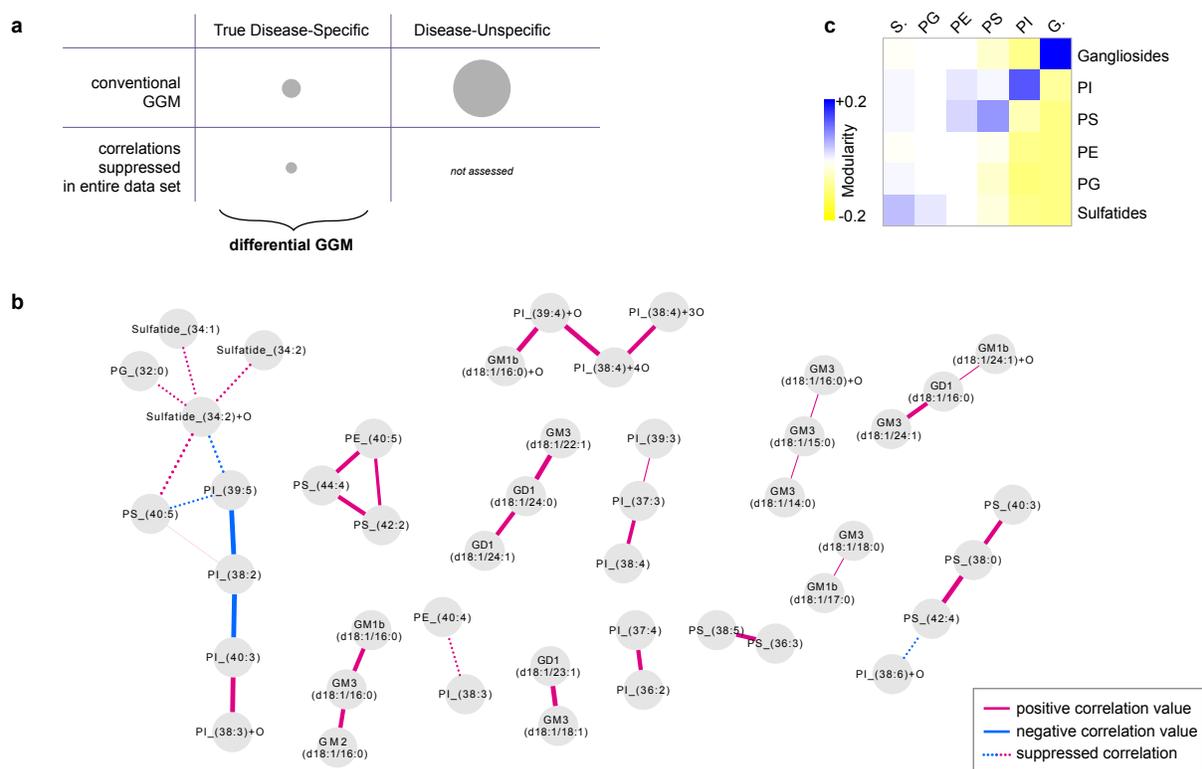


Figure 3. Lipids specifically regulated when Glioblastoma cell lines were effectively perturbed. a. Relative number of disease specific and unspecific lipid-lipid partial correlations in the GGM. Analysis of the entire dataset is named “conventional” GGM with respect to disease specificity. **b.** Disease relevant GGM which is associated with the combined perturbation of p53 adenoviral transfection prior to SN-38 chemotherapy in U87 Glioblastoma cell lines. Edges between lipid nodes are drawn if a significant correlation exists. Positive and negative correlations were color-coded in pink and blue, respectively; Suppressed correlations drawn with dotted lines. Edge line widths indicate degree of dependencies (absolute partial correlation value). The numbers C:D indicates the number of carbon atoms (C) and double bonds (D) of the fatty acid side chain(s). **c.** Modularity matrix was calculated by using lipid classes as cluster label for the GGM shown in **b**. Modularity values were color-coded between -0.2 and $+0.2$ from yellow to blue, respectively. Modularity values close to 1 indicate strong inner-cluster connectivity and little links outside its cluster.

GGM revealed an overrepresentation of specific lipid classes. Sulfatides are glycosphingolipids with two variable ceramide tails. Out of five measured sulfatides, three (60%) were differentially correlated. The three C31:1, C34:2 and C34:2+O are all short chain ceramides with increased levels for the p53 plus SN-38 perturbation.² We can assign the C34:2+O sulfatide a more important role with respect to the disease, as it has a prominent role in the differential GGM with five edges. Note, that we revealed the sulfatide regulation only by inspecting the suppressed correlations, which would have been overlooked by conventional GGM analysis. Gangliosides are glycosphingolipids where one of the two side chains is fixed to a C18:1 fatty acid. They additionally vary in their number of salic acid residues (mono, di or tri). In general, 17 out of 32 (53%) measured gangliosides were coregulated in the disease-specific GGM. Of the the major gangliosides found in adult brain (GM3/GD3),¹⁵ only one was measured by MS. Interestingly, the GM3 was found to be overrepresented with 61% in the GGM (8 out of 13 measured). As previously shown to have decreased level for the p53 plus SN-38 perturbation,² the long chain gangliosides GD1 and GM1b were also found to be overrepresented in the GGM by 50% (4 of 8) and 66% (4 of 6), respectively. Besides the two lipid classes which are overrepresented by more than a half of the measured lipids, another interesting lipid class were Phosphoinositols (PIs). PIs are phospholipids with two esterified fatty acyl residues and inositol as the polar head group. One fourth of the PI were found to be enriched in the GGM (14 of 55). In the original study, the phosphatidylglycerols (PGs) were used as a generic example to show the increased levels of all four phospholipids subclasses.² Nevertheless, we detected an overrepresentation of PIs. A more detailed biological analysis of the PI may reveal the affected mechanisms.

Finally, we aimed to analyze the extend to which the lipid classes were interlinked with each other in the disease-specific GGM. We calculated the modularity [†] by considering each lipid class as the node class label (Fig. 3c). We assume the lipid classes with little or no links to other classes to have a disease relevant regulation based on their molecular characteristics and not due to e.g. fatty acid remodeling. The sulfatides show the most prominent inner-group linkage, indicating that this class was specifically affected by the p53 plus SN-38 perturbation. The gangliosides and all four phospholipids classes were generally interlinked, indicating that a disease relevant mechanism is rather linked to common fatty acid side chains than their unique characteristic head groups.

6. CONCLUSION

We have developed a biologically driven technique to analyze high-throughput measurements. The novel method of a differential GGM is inspired by the experimental design of the biological study to reveal disease relevant information. The differential GGM was applied to the influence of p53 gene therapy prior to SN-38 chemotherapy on U87 Glioblastoma cell lines. We identified only those lipid correlations which are solely induced by the combined perturbation and not just by a single perturbation. Beyond prior studies of quantification histograms and lipid profiles on single lipid classes, we succeeded in analyzing lipids across their classes for the Glioblastoma lipidome which is easily comprehensibly. The disease-specific correlations will advance the understanding of primary brain tumors and their mechanism to immortality.

REFERENCES

- [1] Puchades, M., Nilsson, C. L., Emmett, M. R., Aldape, K. D., Ji, Y., Lang, F. F., Liu, T.-J., and Conrad, C. A., "Proteomic investigation of glioblastoma cell lines treated with wild-type p53 and cytotoxic chemotherapy demonstrates an association between galectin-1 and p53 expression.," *J Proteome Res* **6**, 869–875 (Feb 2007).
- [2] He, H., Nilsson, C. L., Emmett, M. R., Ji, Y., Marshall, A. G., Kroes, R. A., Moskal, J. R., Colman, H., Lang, F. F., and Conrad, C. A., "Polar lipid remodeling and increased sulfatide expression are associated with the glioma therapeutic candidates, wild type p53 elevation and the topoisomerase-1 inhibitor, irinotecan.," *Glycoconj J* **27**, 27–38 (Jan 2010).

[†]Modularity was introduced as a measure of community structure.¹⁶ It measures how modular a set of nodes is compared to a random network model with identical connectivity. The fraction of edges between cluster i and cluster j is corrected for the connectivity of cluster i . The fraction $e_{ij} = \frac{l_{ij}}{L}$ is determined by the number of edges l_{ij} between vertices of i and j divided by the number of edges in the graph L . The connectivity of the cluster a_i is defined by the sum of degrees d_i of vertices in i , or $a_i = \sum_j e_{ij}$. Modularity is then $M_{ij} = (e_{ij} - a_i^2)$

- [3] He, H., Conrad, C. A., Nilsson, C. L., Ji, Y., Schaub, T. M., Marshall, A. G., and Emmett, M. R., "Method for lipidomic analysis: p53 expression modulation of sulfatide, ganglioside, and phospholipid composition of u87 mg glioblastoma cells.," *Anal Chem* **79**, 8423–8430 (Nov 2007).
- [4] Krumsiek, J., Suhre, K., Illig, T., Adamski, J., and Theis, F. J., "Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data.," *BMC Syst Biol* **5**, 21 (Jan 2011).
- [5] Görke, R., Meyer-Bäse, A., Wagner, D., He, H., Emmett, M. R., and Conrad, C. A., "Determining and interpreting correlations in lipidomic networks found in glioblastoma cells.," *BMC Syst Biol* **4**, 126 (2010).
- [6] Schäfer, J. and Strimmer, K., "Learning large-scale graphical gaussian models from genomic data.," in [*In Science of Complex Networks: From Biology to the Internet and WWW*], (2005).
- [7] Opgen-Rhein, R. and Strimmer, K., "Using regularized dynamic correlation to infer gene dependency networks from time-series microarray data.," *4th International Workshop on Computational Systems Biology, WCSB 2006*, 73–76 (June 2006).
- [8] Monakov, A. A., "Estimation of the covariance matrix for dependent signal samples: polarization diversity systems," **30**(2), 484–492 (1994).
- [9] Schäfer, J. and Strimmer, K., "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics.," *Stat Appl Genet Mol Biol* **4**, Article32 (2005).
- [10] Opgen-Rhein, R. and Strimmer, K., "From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data.," *BMC Syst Biol* **1**, 37 (2007).
- [11] de la Fuente, A., Bing, N., Hoeschele, I., and Mendes, P., "Discovery of meaningful associations in genomic data using partial correlation coefficients.," *Bioinformatics* **20**, 3565–3574 (Dec 2004).
- [12] Velicer, W. F., "Suppressor Variables and the Semipartial Correlation Coefficient," *Educational and Psychological Measurement* **38**(4), 953–958 (1978).
- [13] Das, A. and Kempe, D., "Algorithms for subset selection in linear regression," in [*Proceedings of the 40th annual ACM symposium on Theory of computing*], *STOC '08*, 45–54, ACM, New York, NY, USA (2008).
- [14] Miller, R. G., "The jackknife-a review," *Biometrika* **61**(1), 1–15 (1974).
- [15] Ando, S. and Yu, R. K., "Fatty acid and long-chain base composition of gangliosides isolated from adult human brain.," *J Neurosci Res* **12**(2-3), 205–211 (1984).
- [16] Newman, M. E. J. and Girvan, M., "Finding and evaluating community structure in networks.," *Phys Rev E Stat Nonlin Soft Matter Phys* **69**, 026113 (Feb 2004).