Identifying Pathways Enriched with Differential Dependencies using LASSO

Recent advances in high-throughput transcriptomic profiling have given rise to several statistical methods for analyzing Gene Regulatory Networks (GRNs). More recently, the context-specificity of GRNs, mainly their topology, has garnered significant interest. Evaluation of Differential Dependency (EDDY) is one approach to analyzing these Differential Dependency Networks (DDNs), using Bayesian inference and network likelihood distributions to identify pathways enriched with differential dependency relationships. Bayesian inference, however, requires that expression data be quantized prior to analysis, leading to a possible loss of information. We propose an alternate approach for identifying differential dependency relationships based on using LASSO regression for GRN reconstruction. To assess differentiality, we cross-calculate Coefficients of Determinance and find the Hamming Distances between the GRNs. Statistical significance of uncovered differentiality is assessed asymptotically by approximating a null distribution via a beta distribution for which model parameters are estimated from initial permutations. In tests on data generated synthetically from models of increasing hamming distance, the metrics matched expected trends and showed high GRN inference accuracy. From CCLE data, the software called 7 Biocarta pathways as enriched with differential dependencies. Thus, we have successfully removed the quantization limitations of Bayesian EDDY and created a new network-level analysis of differential dependency.

1 Introduction

The advancement of high-throughput transcriptomic profiling has introduced a wealth of information that requires computational processing to interpret. Specifically, cDNA microarrays are leading to large databases of gene expression records. Of particular interest to biologists is the potential to use this information to model gene-gene interactions for both clinical care and drug discovery. Interpreting differential computational models of Gene Regulatory Networks (GRNs), specifically the topology of their context-specific forms, has been the subject of various statistical tools, as identifying pathways that differ significantly between cellular contexts could help further focus research efforts.

Network-level analysis of differentialy expressed genes ("Network differentiality") has been previously explored by [1]. This approach, called Evaluation of Differential Dependency (EDDY), seeks to identify pathways enriched with differential dependencies by using Bayesian network inference to translate quantized expression data from two phenotypic variants of cells, such as cancerous and wild-type, into Gene Regulatory Networks. Leave One Out (LOO) permutations generate a network likelihood distribution for each class, on which Jensen-Shannon Divergence can be applied to calculate differentiality. A modified permutation test, which utilizes a beta approximation to estimate a null distribution, is then used to assess the significance of the detected differentality.

A significant limitation of EDDY is the dependency on quantized data necessary for Bayesian Inference. Quantization is the practice of converting continuous values to discrete levels through binning, for example, representing up-regulation as $\{1\}$, neutrality as $\{0\}$, and down-regulation as $\{-1\}$. Such binning is a necessary for a computationally convergent solution, but the continuous-to-discrete transition risks information loss. The goal of this research is to identify pathways that are enriched with differential dependency relationships using continuous expression data values. More simply, we seek to remove the quantization dependency of the current EDDY implementation.

Several alternate approaches exist to assess differential dependency, including the Single Gene Test, Gene Set Enrichment Analysis (GSEA), and Gene Set Coexpression Analysis (GSCA). In this paper, we explore an approach to evaluating differential dependency based on a linear model. The advantage of using a linear model for GRN inference is that expression data does not need to be quantized prior to analysis. In effect, we propose substituting the Bayesian Inference method of the current implementation of EDDY with a linear model to infer the GRN and using an approach more consistent with the linear model to assess differentiality.

[4] describes the linear approach to GRN reconstruction proposed by Meinshausen and Bühlman. This method uses an iterative regression approach to formulate a system of equations, whereby the union of all nonzero coefficients can be thought of as the edges of the GRN. To contruct the system, [4] uses the Least Absolute Shrinkage and Selection Operator, or LASSO, which is outlined by Tibshirani in [2].

The LASSO is an L_1 penalized linear regression technique defined by

$$\boldsymbol{\beta} = \arg\min_{\beta} \left\{ \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 \right\} \text{ subject to } \|\boldsymbol{\beta}\|_1 \leq t \tag{1}$$

where t is a user-defined tuning parameter, $\vec{Y}$ is the target matrix, and X is the data matrix. Equation (1) can be rewritten with a Lagrangian multiplier λ as

$$\boldsymbol{\beta} = \arg\min_{\beta} \left\{ \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \right\} \tag{2}$$

The LASSO observes two distinct advantages over other regression techniques. Like the ridge regression, LASSO is not susceptible to unbounded $\|\boldsymbol{\beta}\|_1$ increase when predictor variables are correlated. However, unlike ridge regression, LASSO will also shrink coefficients to zero depending on their correlation to the target variable and the Lagrangian λ or tuning parameter t. [2] dis-

cusses, for instance, that the selection of $t = t_0/2$ will effectively be limiting the variable subset to size $p/2$. These two features are important for the unique considerations of GRN reconstruction. Feature selection is necessary for creating the appropriately sparse adjacency matrix. Additionally, constrained $\vec{\beta}$ ensure that correlated predictors that may have edges among themselves in the GRN model do not affect the regression. Given these characteristics, Meinshausen and Bühlmann suggest representing a high-dimensional graph as a set of p local neighborhoods, where a neighborhood of node i is defined as all nodes $k \in \Gamma$ such that all $j \in \Gamma \backslash k$ are conditionally independent of i, $0 < i \leq p$. By solving (2) for each node in a network, [4] posits that a sparse inverse covariance adjacency matrix can be created that estimates $G(\Gamma, E)$, where the nonzero elements in the $\vec{\beta}$ from LASSO regression become edges. Because it is convex, (2) can be approximated via the coordinate descent approach specified in [3], the Interior-Point Method or another Quadratic Programming solving technique. In response to [4], [7] presents a graphical adaptation of LASSO for estimating sparse inverse covariance matrices. It partitions the estimated covariance matrix and the empirical covariance matrix, iteratively solving the LASSO problem and progressively refining the estimated covariance matrix. However, graphical LASSO results in a symmetric adjacency matrix and thus is unfavorable as a model for Gene Regulatory Networks as it does not allow for representation of directed relationships.

Zhang *et al.* (2009) introduced a LASSO-based GRN model for assessing differential dependency built off the concept of p-iterated LASSO regression presented in [4]. Like [1], [5] intends to compare two classes of GRNs to identify strong statistical differentiation, but unlike [1], [5] uses a LASSO/LARS approach to infer the GRN model, however, both still permute data to determine statistical significance. A revision to [5], [6] incorporates prior biological knowledge by using a prior-knowledge bias term on a prior-knowledge covariance matrix. One notable limitation of [5, 6] is that neither offers an assessment on network-level differentiality. Rather, they focus on the significant differentiality of individual genes in the GRN, which is not as effective as studying the entire GRN in its context.

We call our software LASSO-based Evaluation of Differential Dependency, or LASSO EDDY. We present our approach to creating the GRN adjacency matrix with the LASSO in section 2. In section 3, we describe how the LASSO GRNs are compared to indicate differentiality. In section 4, we present our approach to assessing the significance of the detected differentiality. In section 5, we discuss additional features to modify the LASSO algorithm for proper GRN reconstruction. Sections 6 and 7 present testing procedures and data for LASSO EDDY. We offer commentary on our approach in sections 8 and 9.

2 LASSO Sparse Adjacency Matrix Estimation for Gene Regulatory Networks

To infer sparse adjacency matrices from gene expression data, we use an approach similar to [4]. Suppose we have a network of p nodes with n i.i.d observations, $O(n)$. Then suppose there exists a gene set $\Gamma_g \in \Gamma$. Our algorithm first isolates this gene set from the expression file and then follows algorithm 1.

Algorithm 1 LASSO GRN Adjacency Matrix Estimation

Require: $p, \Gamma_g, O(n)$
  for all $p \in \Gamma_g$ **do**
    $\lambda_{max} \leftarrow 1$
    $\lambda \leftarrow \lambda_{max}$
    while $\lambda > 0$ **do**
      while $k < 10$ **do**
        $f(x) \leftarrow LASSO(O(n), \lambda))$
        $error \leftarrow predict(f(x))$
        $k \leftarrow k + 1$
      end while
      $\text{error}_\lambda \leftarrow \overline{error}$
      $\lambda \leftarrow \lambda - step$
    end while
    $\lambda \leftarrow \arg\min_\lambda \text{error}$
    if $\lambda = \lambda_{max}$ **then**
      $\lambda_{max} \leftarrow 2\lambda_{max}$
      Repeat
    end if
    $\lambda = \text{doBinarySearch}\{\lambda_{max}, 0\}$
    $f(x) \leftarrow LASSO(O(n), \lambda))$
  end for

Expression data is divided according to user-defined class labels. Algorithm 1 is conducted for each set of data. A regularization parameter λ is initialized and the bounds of λ are first established using a linear search, the final λ specified with a binary search. In both cases, and for each λ, a LASSO solution using Block Coordinate Descent is found. Ten-fold cross validation is used to find the error for each λ. At the end of the cross validation, the λ with the lowest error is recorded. In the linear search, the upper bound is initialized arbitrarily, but is raised if the suggested λ is the upper bound. This is to fit a condition in [3] that the upper bound of λ is the smallest value for which $\vec{\beta} = \vec{0}$. Once the upper bound is set, the same process is repeated in a binary search to maximize accuracy. Both searches are done on a log scale for computational efficiency. Once the true λ is set, the LASSO is performed once again using the entire dataset to construct the equation relating the target genes' expression to that of the other genes.

Once the gene's LASSO solution is found, the final weight array W is updated. After all LASSO

solutions are found, the weight array represents the sparse inverse covariance matrix. However, when creating an edgelist, it may be useful to denote a sensitivity parameter ε. If $|W_{ij}| < \varepsilon$, the edge between nodes i and j is negligible and not included. Such thresholding prevents overfitting and reduces the number of false positives. By default, $\varepsilon = 10^{-4}$, but can be changed based on the desired certainty of a suggested relationship. It is worth noting, however, that setting the ε leads to a trade off between sensitivity and specificity. By selecting low values of ε, one can minimize the likelihood of Type I errors, but by selecting larger values, one can minimize the likelihood of Type II errors.

After the weight matrix is found, an adjacency matrix is created by (3)

$$A_{ij} = \begin{cases} 1, & \text{if } \left|W_{ij}\right| > \varepsilon \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

3 Detecting Differentiality with EDDY LASSO

We next investigate how to use LASSO GRNs to detect and quantify differentiality between two cell variants and give indication as to the significance of the detected differentiality. To assess differentiality we propose two distinct but complementary approaches. For a topological perspective on the differentiality, we use Hamming Distance to compare network topologies. For a purely statistical metric, we cross-calculate Coefficients of Determination to investigate the statistical value of differentiality.

To calculate a Hamming distance, we first find the union of all edges in the adjacency matrices $A^{(1)}$ and $A^{(2)}$ by equation (4).

$$e = \sum_i \sum_{j \neq i} |A_{ij}^{(1)}| + |A_{ij}^{(2)}| > 0 \tag{4}$$

Having found the total number of edges e, we then calculate the Hamming distance by (5), the L_1 norm of the difference in the adjacency matrices divided by two times the number of edges as calculated by (4).

$$\frac{\left\|A^{(1)} - A^{(2)}\right\|_1}{2e} \tag{5}$$

The Hamming distance is bounded between 0 and 1. The significance of this is explained in Section 4.

Hamming distances are topological measures of differentiality. For a more statistical perspective, we also include cross calculation of Coefficients of Determination. The COD is a measure of how well a linear model fits a set of data. For some target variable Θ with predicted values $\tilde{y}_i$ and actual values y_i, $0 < i \leq n$, the COD over a geneset of size p can be found using equation (6).

$$COD^{\Theta} = 1 - \frac{MSE_{\Theta}}{Var_{\Theta}} \tag{6}$$

where Var_{Θ} is defined as

$$\frac{\sum_i^n (y_i - \bar{y}_i)^2}{p} \tag{7}$$

and MSE_{Θ} as

$$\frac{\sum_i^n (y_i - \tilde{y}_i)^2}{p} \tag{8}$$

For each Θ in the LASSO GRN, we calculate (6) for both its own data and that of the other variant, and calculate total ΔCOD by

$$\Delta COD^{\Theta} = \frac{1}{2}\left(COD^{\Theta}_{f_1 d_1} - COD^{\Theta}_{f_1 d_2} + COD^{\Theta}_{f_2 d_2} - COD^{\Theta}_{f_2 d_1}\right) \tag{9}$$

This creates a vector $\vec{COD}$ which represents the cross-calculated ΔCOD for each gene Θ in the geneset. We take the median ΔCOD as the representative ΔCOD for the geneset. Notice that ΔCOD approaching 1 represents poor modeling of GRN 1 for Data 2 and vice versa. This means that the

two models are not isomorphic and that the two GRNs are fundamentally noninterchangeble. Also note that, similar to Hamming Distance, the ΔCOD will be bounded between 0 and 1.

4 Evaluating the Significance of the Detected Differentiality

In this section, we describe a modified permutation test used to assess whether the detected differentiality from section 3 could have been due to random chance.

In a traditional permutation test, the two datasets used to create the original GRNs are merged into a common dataset. Monte-carlo sampling is performed to arbitrarily split the combined dataset into two classes. Then, the process for reconstructing the GRN and evaluating the differentiality are repeated, creating a new Hamming distance and median ΔCOD. The two datasets are merged again, thus completing a permutation. When this procedure is repeated k times, a distribution of Hamming Distances and median ΔCODs are created, from which a p-value describing the significance of the original differentiality can be extracted.

When conducting an ordinary permutation test, a large number of repetitions are necessary to ensure statistical significance, which could be computationally prohibitory. However, a beta approxmation can be used to estimate the null distribution, and a cumulative distribution function can retrive the p-value with fewer permutations. A beta approximation can only be used when the parameter is bounded between 0 and 1. However, as explained in section 3, both the Hamming Distance and the ΔCOD fall in this range. Therefore, either is applicable to approximate the beta distribution and extract the p-value. We use the statistics-based Coefficient of Determinance in this study. To find the beta approximation, 100 permutations are recommended, but statistical significance can be established with as low as 20. After completing the permutations, we calculate the mean and variance of the ΔCODs. To approximate the α and β parameters from the mean and

variance, we use (10) and (11)

$$\alpha = \left(\frac{1-\mu}{\sigma^2} - \frac{1}{\mu} \right) \mu^2 \tag{10}$$

$$\beta = \alpha \left(\frac{1}{\mu} - 1 \right) \tag{11}$$

Once the beta function is created, $p = 1 - CDF(x)$ gives the p-value for the original ΔCOD, x. If $p < 0.05$, it is generally accepted that the pathway from which the geneset was derived is enriched with differential dependency.

5 Additional Features of LASSO EDDY

Two optional features were introduced into LASSO EDDY to modify the accuracy of the modeling depending on the availability of related information.

5.1 Edge Controlling

To prevent overfitting, LASSO EDDY has the ability to sacrifice cross-validation error to encourage additional shrinkage. During λ approximation through cross-validation, the lambda with the minimum cross validation error is found. During edge controlling, the λ is increased by an arbitrary step size and the cross-validation error is recalculated. The λ continues to increase until the cross-validation error exceeds one standard deviation of the original cross validation error. When the λ increases, this is equivalent to increasing the penalty of the LASSO objective function, or artificially tightening the LASSO, which puts a harsher bias against nonzero β, so smaller β that would have otherwise been included are shrunk to zero. In section 6, we show that edge controlling improves the sensitivity and specificity of LASSO EDDY.

5.2 Prior Knowledge Integration

LASSO EDDY also allows for the integration of prior knowledge from human-annotated gene-gene interation databases. To incorporate prior knowledge, we propose the following modification

to the LASSO algorithm.

$$\beta = \arg\min_{\beta} \left\{ \|\mathbf{Y} - \mathbf{X}\beta\|_2^2 + \lambda \sum_p (1 - I_p W_p E_p)\, \beta_p \right\} \tag{12}$$

Where E_p is either 1, representing an edge, or a 0, representing a nonedge, I_p is a continuous value between 0 and 1 defining the overall influence of prior knowledge, and W_p is a continuous value between 0 and 1 that is assigned per relationship depending on the credibility of the source from which the information was derived.

6 Application and Testing of LASSO EDDY

This section describes one application LASSO EDDY to data from the Cancer Cell Line Encyclopedia (CCLE) and the procedures used to validate LASSO EDDY using synthetic data.

6.1 Applying LASSO EDDY to CCLE Data

The Broad Institute had previously recorded the expression data of more than 300 cell lines from The Cancer Cell Line Encyclopedia (CCLE). Researchers applied the drug BRD-K46556387 (CIL55) and classified each cell line as being either sensitive or unresponsive to the drug. Using the expression and sensitivity data from these cells, we now seek to use EDDY LASSO to identify pathways enriched with differential dependencies between the sensitive and unresponsive cells. We analyze 217 pathways from the Biocarta database, using 30 permutations to analyze each pathway and record the associated p-value.

6.2 Validation of LASSO EDDY

We also seek to validate LASSO EDDY through showing that the p-value, median ΔCOD and Hamming Distance follow their expectedd trends, that is, median ΔCOD and Hamming Distance increase and p-value decreases as the divergence between GRNs increases. To do this, a ground

truth model is generated from which graphs of increasing differentiality can be created by inverting entries in the associated adjancency matrix. The synthetic GRNs are based off the LASSO EDDY inferred model of the Biocarta ACE2 pathway, which has 13 genes and 29 offdiagonal non-zero entries. Graphs were generated with 2, 4, 6, 8, 10, 12, 14, and 16 edges inverted. A weight matrix and subsequent 300 expression data samples were synthetically generated for each GRN using a Multivariate Gaussian Distribution. Each modified GRN was compared using LASSO EDDY to the original GRN over 49 trials, the p-values, median ΔCODs and Hamming Distances recorded for each. To gain additional insight into the network recovery abilities of LASSO EDDY, we also created an adjacency matrix with 29 offdiagonal nonzero entries and compared the LASSO generated GRN to this ground truth over 67 trials. To test the efficacy of the edge controlling algorithm, the experiment was conducted both with and without edge controlling.

7 Results

This section describes the results for the application of LASSO EDDY to CCLE data and validation tests.

7.1 Application of LASSO EDDY to CCLE Data

LASSO EDDY found 7 pathways enriched with significant Differential Dependency Relationships. Of all the pathways analyzed, 14 were found to have a pvalue of less than 0.1. Results can be found in table 1. An example DDN generated with LASSO EDDY for the Biocarta IL5 pathway is shown in figure 1.

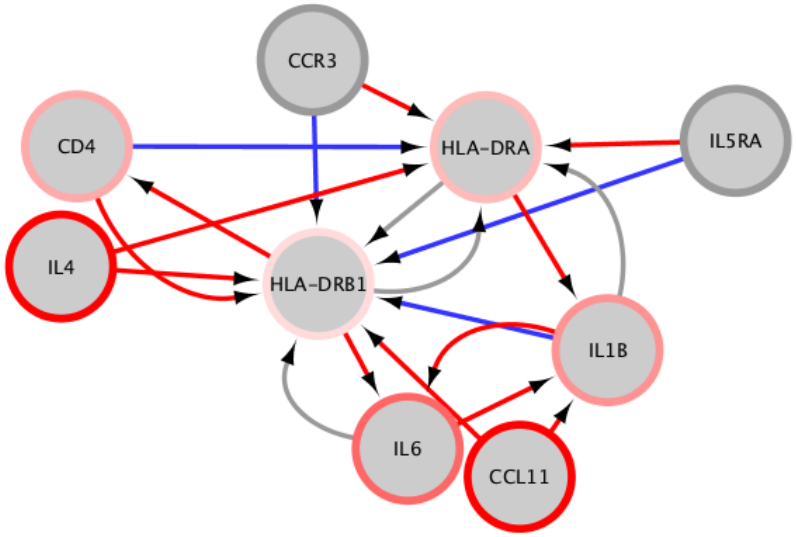


Figure 1: DDN of IL5 Pathway

Pathway	p-value
IL5 PATHWAY	0.007
EPONFKB PATHWAY	0.014
UCALPAIN PATHWAY	0.022
SPRY PATHWAY	0.031
RACCYCD PATHWAY	0.031
SRCRPTP PATHWAY	0.040
NKCELLS PATHWAY	0.0405
IL1R PATHWAY	0.051
SODD PATHWAY	0.062
TGFB PATHWAY	0.064
LONGEVITY PATHWAY	0.067
G2 PATHWAY	0.069
RHO PATHWAY	0.081
KREB PATHWAY	0.098

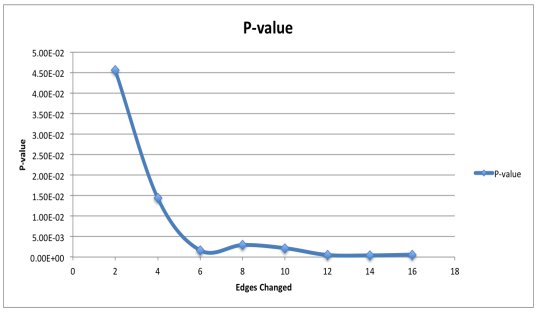
Table 1: Pathways with $p<0.1$

7.2 Validation of LASSO EDDY

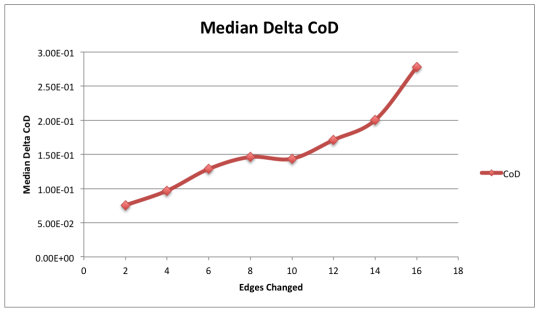
Table 2 shows the averages for p-value, median ΔCOD, and Hamming Distance over the 49 trials for each of the networks. The trends are shown in Figure 2.

Edges Changed	2	4	6	8	10	12	14	16
p-value	0.0456	0.0144	0.00164	0.00293	0.00217	0.000507	0.000409	0.000587
Median ΔCOD	0.0757	0.0970	0.129	0.146	0.144	0.171	0.200	0.278
Hamming Distance	0.214	0.226	0.253	0.277	0.303	0.310	0.338	0.341

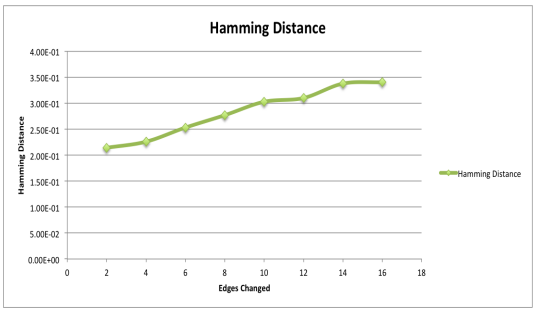
Table 2: LASSO EDDY Synthetic Data Test Results



(a) P-value



(b) Median Delta COD



(c) Hamming Distance

Figure 2: Simulation results

The results for calculating the accuracy of LASSO EDDY generated graphs is shown in the table below. The columns are true positive, false positive, true negative, false negative, accuracy, true positve rate, specificity, precision, and false positive rate, and the results represent averages over 67 trials.

	TP	FP	TN	FN	Accuracy	TPR	SPC	Precision	FPR
No Edge Control	25.104	36.463	90.537	3.896	74.129%	0.866	0.713	0.408	0.287
With Edge Control	25.209	27.612	99.388	3.791	79.870%	0.869	0.783	0.485	0.217

Table 3: LASSO EDDY network inference

8 Discussion

LASSO EDDY was able to identify 7 pathways from the Biocarta database as showing significant differential dependency relationships across the cell classes. Though the true number of significant pathways in the database remains unknown, that LASSO EDDY generates unique p-values that can fall on both sides of the significance threshold is encouraging. In further inspection of the DDNs, we noted that HLA-DRB1 seemed to play a significant role in mediating different responses of cell lines to the CIL55 compound along the IL5 pathways. We identify such critical mediators using a similar procedure to what is done in applications of [1]: analyzing which nodes in the GRN undergo the most significant change in betweenness centrality, which is a measure of the centrality or importance of a node in a network. This mathematical analysis of LASSO EDDY GRNs can be useful in identifying possible gene targets for drug therapy.

The true measure of the accuracy of LASSO EDDY is found in the test on synthetic data, for which the ground truth is known. The results showed that as LASSO compared graphs that were increasingly divergent, the p-values fell, while the median ΔCOD and Hamming Distance rose. This is consistent with expected trends, as median ΔCOD and Hamming Distance both measure the level of divergence between two networks.

Overall, LASSO EDDY experienced an average GRN inference accuracy of almost 80% with edge-controlling, with a range from the raw data of 75-95%. A key observation from table 3 is that edge controlling reduces the average number of false positives. Accuracy increases by about five percent, and specificity increases by seven percent. Therefore, we recommend using edge controlling in applications of LASSO EDDY.

An anomaly in the pattern occurs between 8 and 10 edge inversions, in that though the hamming distance increases, the median ΔCOD decreases. The cause of the outlier could have been that after the fourth round of edge inversions, the underlying inverse covariance matrices of the

two GRNs may have been more statistically similar than after the third and fifth sets of inversions. In effect, though the hamming distance between the models increased, the underlying statistical models derieved from the adjacency matrices were more similar. This points to a weakness in using median ΔCOD: it will not recognize the differentialiy of two structurally different graphs with similar modeling capabilities. The reverse is also true. Between two and four edge inversions, the Hamming Distance only marginally increased, but because this had a significant impact on the modeling capabilities of the GRN, it was reflected strongly in the median ΔCOD. Though a new metric could be defined that combines the hamming distance and median ΔCOD to create a more robust estimation of differentiality, the overall trends of the metrics show that the current implementation of LASSO EDDY has satisfied the requirements of the synthetic data validation.

9 Conclusion

We have successfully addressed the quantization limitations of Bayesian EDDY by substituting the Bayesian probabilistic model for the Gene Regulatory Network with the LASSO-generated linear model GRN, which allows for the use of continuous values for expression data. In doing so, we have also introduced a method of determining network-level differentiality of two GRNs, which advances current related linear model based DDN tools. To do this, we follow the iterative LASSO approach proposed by Meinshausen and Bühlman and cross calculate Coefficients of Determinance. To evaluate the statistical significance of this metric, we use a beta function to approximate a null distribution and extract a p-value. As a secondary statistic, we also calculate the hamming distance between the two GRNs. Synthetic data tests show that our metrics follow expected trends and show high GRN inference accuracy. The software developed through this research can now be used to analyze expression data retrieved from the lab bench. For future work, one consideration of LASSO EDDY is the generalization inherent in conforming the data to a linear model. In cases where the expression data is decisively nonlinear, a bayesian probabalistic model may even be more appropriate, and more research should be done to investigate the relationship between

Bayesian and Linear model GRN inference. Another informative future research could involve a correlative study on the accuracy of LASSO EDDY with respect to decreasing linearity of the expression data. In any case, we anticipate that the ability to use continuous gene expression data will provide more accurate and precise models for differential dependency networks.

References

[1] Jung, S. and Kim, S. (2014). EDDY: A Novel Statistical Gene Set Test Method to Detect Differential Genetic Dependencies. *Nucleic Acids Research*, **42**(7): e60, 13p.

[2] Tibshirani, R. (1996). Regression Shrinkage and Selection Via the LASSO. *Journal of the Royal Statistical Society. Series B (Methodological)*, **58**(1), 267-288

[3] Friedman, J., Hastie, T., Tibshirani, R. (2009). Regularization Paths for Generalized Linear Models via Coordinate Descent. Department of Statistics, Stanford University, 24p.

[4] Meinshausen, N. and Bühlmann, P. (2006). High-Dimensional Graphs and Variable Selection with the LASSO. *The Annals of Statistics*, **34**(3), 1436-62

[5] Zhang, B., Huai, L., Riggins, R. B., Zhan, M., Xuan, J., Zhang, Z., Hoffman, E. P., Clarke, R., and Wang, Y. (2008). Differential Dependency Network Analysis To Identify Condition-Specific Topological Changes in Biological Networks. *Bioinformatics*, **25**(4), 526-32

[6] Tian *et al.*: Knowledge-Fused Differential Dependency Network Models for Detecting Significant Rewiring in Biological Networks. *BMC Systems Biology* 2014. **8**:87.

[7] Friedman, J., Hastie, T. Tibshirani, R. (2007). Sparse Inverse Covariance Matrix Estimation with the Graphical LASSO. *Biostatistics*, **0**(0), 1-10

[8] Vignes M, Vandel J, Allouche D, Ramadan-Alban N, Cierco-Ayrolles C, et al. (2011) Gene Regulatory Network Reconstruction Using Bayesian Networks, the Dantzig Selector, the LASSO and Their Meta-Analysis. PLoS ONE 6(12): e29165. doi:10.1371/journal.pone.0029165