Chapter 10

Inferring Molecular Interactions Pathways from eQTL Data

Imran Rashid, Jason McDermott, and Ram Samudrala

Abstract

Analysis of expression quantitative trait loci (eQTL) helps elucidate the connection between genotype, gene expression levels, and phenotype. However, standard statistical genetics can only attribute the changes in expression levels to loci on the genome, not specific genes. Each locus can contain many genes, making it very difficult to discover which gene is controlling the expression levels of other genes. Furthermore, it is even more difficult to find a pathway of molecular interactions responsible for controlling the expression levels. Here we describe a series of techniques for finding explanatory pathways by exploring the graphs of molecular interactions. We show several simple methods can find complete pathways that explain the mechanism of differential expression in eQTL data.

Key words: eQTL, pathway inference, gene regulation, signaling pathways.

1. Introduction

Recent studies on expression quantitative trait loci (eQTL) have revealed that differential gene expression is sometimes tightly linked to variation in specific chromosomal locations *(1, 2)*. When gene expression is also associated with phenotypes such as disease, there is great interest in discovering the pathway connecting genetic variation and differential expression. However, this remains a difficult task. The chromosomal locations generally include many candidate causative genes. Genetic markers are able to narrow the genetic variation down to a region of roughly ten genes. When the expression level of a gene is strongly linked to one locus, genetic variation in one of these genes is presumably causative for the differential expression. However, linkage only cannot tell us which gene out of all the genes in the locus is causative.

Jason McDermott et al. (eds.), *Computational Systems Biology, vol. 541*
© Humana Press, a part of Springer Science+Business Media, LLC 2009
DOI 10.1007/978-1-59745-243-4_10

Furthermore, even when the causative gene is known, it is very difficult to predict all the molecules involved in the regulatory pathway. In some situations, the differentially expressed gene falls within the locus it is linked to – in this case (*cis*-regulation), we assume that changes in the gene's promoter, enhancer, or other *cis*-regulatory sites effect mRNA expression levels. In many cases, however, the differentially expressed gene is physically very distant from the linked locus. We assume the locus contains genes that regulate the differentially expressed gene, potentially through a long and intricate pathway (*see* **Fig. 10.1**). However, uncovering the precise pathway that regulates transcription remains difficult. This is because genes, proteins, and other biological molecules form highly context-dependent interaction networks, allowing for many possible paths from a causative gene in the linked locus to the differentially expressed gene. Here, we consider several approaches to solving this problem by searching graphs of known molecular interactions. Given the broad scope of this problem, we focus on differentially expressed genes that are strongly linked to exactly one locus. The methods section describes the techniques for determining linked loci, various methods to find pathways in interaction graphs, and several approaches to evaluating these methods. Finally, we discuss an evaluation of these methods and the pathways they discover, along with possible future extensions to these methods.

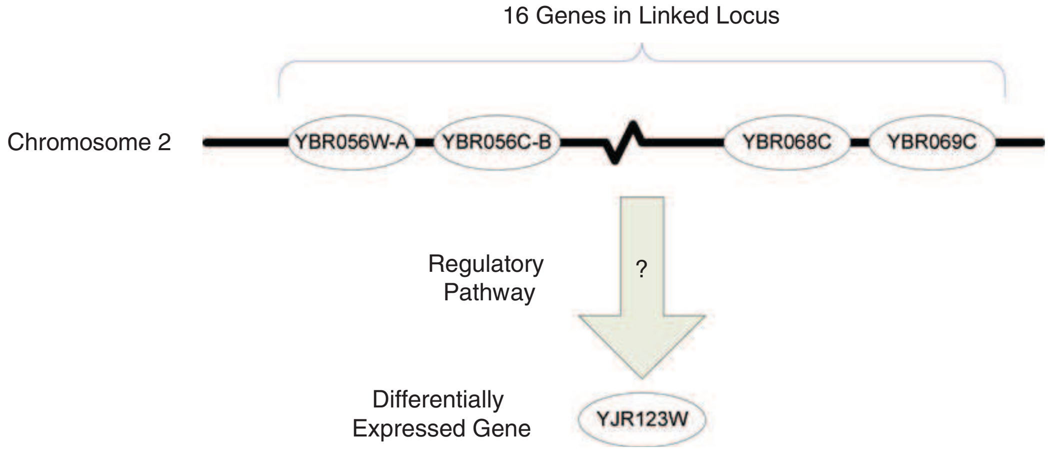

Fig. 10.1. **Finding regulatory pathways.** eQTL studies lead us to believe that genetic variation within the linked locus is responsible for the differential expression of another gene. However, we do not know which gene in the locus is responsible for regulating the expression. Furthermore, even if we did know which gene in the locus was responsible, we would still not know the regulatory pathway responsible. Here, we survey several methods for finding potential regulatory pathways from databases of known molecular interactions. This example is from eQTL studies in yeast *(1)*, where the differentially expressed gene YJR123W has been linked to a locus with 16 genes in it.

2. Methods

Given a set of differentially expressed genes and their linked loci, we wish to find pathways from each locus to each differentially expressed gene, such that the differential expression can be explained by the pathway. This involves both discovering which gene in the locus is causing the differential expression as well as finding a coherent pathway from that gene to the differentially expressed gene, i.e., the mechanistic basis of the wiring diagram that explains the gene expression.

2.1. Finding Linked Loci

The first step in analyzing eQTL data is determining linkage. There has already been a great deal of work to find linked loci through quantitative methods. Those methods are not the focus of this chapter; here we highlight a few techniques to direct further exploration of the reader. All of these methods involve assessing the effect each locus has on the expression level of every gene. The most direct approaches use a Wilcoxon ranksum test to assign statistical significance to each locus *(1, 3)*. More sophisticated techniques include calculating linkage by simultaneously considering multiple markers and intervals, as well as additional corrections for multiple testing *(1, 4–6)*. Though we only focus on analysis once linkage information is in hand, we want to emphasize the strong dependence of all methods on the assessment of linkage; clearly, it is very important to determine linkage very carefully and rigorously (*see* **Note 1**).

2.2. Building a Molecular Interaction Graph

We wish to find pathways where each interaction is a known interaction between two molecules. This requires assembling a catalog of all interactions, to facilitate searching. Thus far, we have focused on protein–protein, transcription regulation, and phosphorylation interactions. There are several publicly available databases of this information, including many high-throughput experiments *(7–10)*. We have also included predicted protein–protein interactions from Bioverse (http://bioverse.compbio.washington.edu) *(11)*, determined through the interolog method. Note that this does not represent all known interactions; further work must explore the inclusion of metabolites as well as the effect of non-coding RNA (*see* **Note 2**). Related chapters discuss the techniques for reconstructing interaction graphs.

The most natural way to store these interactions is in the form of a graph, with nodes for each molecule and edges for each interaction. Due to the sparseness of these graphs, an adjacency list is the most efficient way to store the edges of the graph. The directionality of each edge should correspond to the type of interaction, e.g., physical interactions are bidirectional while transcription regulation interactions are directed from transcription factors to their targets.

2.3. Finding Pathways Between Differentially Expressed Genes and Linked Loci

With a set of differentially expressed genes and their linked loci, as well as an interaction graph, we can formulate our problem of finding plausible causative pathways. All of these methods attempt to find the causative gene and pathway in the same step. We will search all paths from the differentially expressed gene to all genes in the linked locus; the most plausible causative path should lead to the correct causative gene. In the next few sections, we will examine several methods for determining which pathways are the most plausible.

2.3.1. Finding Linear Paths Through Directed Graph Searching

The simplest formulation of a pathway is a linear sequence of interactions, from the causative gene to the differentially expressed gene. Our only obstacle is the choice of method for evaluating pathways and choosing the best one. Searching a graph for linear paths is already a well-studied problem. We explain a variety of different evaluating pathways, beginning with very simple approaches and building up to more complex procedures. The most straightforward method of evaluating a pathway is by its length. Clearly, a shorter pathway is more believable than a very long one. We can find the shortest path from genes in the linked locus to the differentially expressed gene through a standard breadth-first search (BFS). For efficiency, instead of searching for pathways from each gene in the linked locus to the differentially expressed gene, we can reverse the direction of all edges in the graph and search for paths from the differentially expressed gene to each gene in the linked locus. BFS is guaranteed to find the shortest path to every node, and is very fast and efficient.

Unfortunately, path length alone is an insufficient criterion for determining the most likely causative pathway. Often there will be many pathways of the same length. Furthermore, in many cases BFS will find pathways involving genes with uncorrelated expression levels, while a slightly longer pathway exists, which involves highly correlated genes. Using this idea, we can favor pathways with highly correlated genes, by assigning a cost to each edge and searching for the lowest cost path. Let corr(i, j) be the correlation coefficient of the expression levels for genes i and j. For each edge (i, j) in the interaction graph, we compute a cost $C(i, j)$:

$$C(i, j) = -\log|\mathrm{corr}(i, j)|$$

The absolute value of the correlation coefficient is taken because both positive and negative correlations are believable in biological pathways, corresponding to activators and repressors. The negative log is taken because our graph search finds the lowest cost paths, while we wish to favor the interactions between highly correlated genes (*see* **Note 3**). Finding lowest cost paths can be implemented very efficiently using a uniform cost search. While BFS uses a first-in first-out queue to order the exploration of nodes in the graph, with a uniform cost search we explore nodes with the shortest pathway. (This is done efficiently through use of the priority queue data structure.)

Though this finds pathways with highly correlated genes, many of the pathways are still not biologically plausible. In particular, as these pathways are intended to explain gene expression level, we expect the differentially expressed gene to be immediately controlled by some transcription factor. We can make simple modifications to our graph searching methods to take advantage of the extra biological information we have on each interaction. For example, it is easy to modify a uniform cost search to ensure the first step is always to a transcription factor. Furthermore, the interaction graph could be extended to include other types of interaction data. For example, if microRNA and their targets were included in the interaction graph, we might expect microRNA to be involved in regulating expression in a manner similar to transcription factors.

2.3.2. Random Walks Across an Interaction Graph

One major shortcoming of all of the previous approaches has been that they find only linear pathways. True biological pathways involve many non-linear components, such as parallel pathways, which give biological systems added robustness, or feedback loops, which allow large responses to small stimuli. In a recent paper, Tu et al. proposed using a random walk across an interaction graph to find causative pathways *(3)*. Their method involves finding pathways in the reverse direction, so it begins by reversing the direction of all edges in the interaction graph. Then they perform the following steps:

1. In the reversed interaction graph, start at the differentially expressed gene.
2. Randomly choose a neighbor of the current node, and go to that node.
3. If the new node is a gene in the locus, then stop the search and record the end gene.

   If not, return to **step 2**.
4. Repeat the random walk 10,000 times. Choose the causative gene as the gene in the linked locus that was visited most frequently.

When choosing a neighbor to visit, nodes with highly correlated expression levels are favored. The probability of choosing a neighbor for the next step is proportional to the correlation coefficient of the expression levels. Furthermore, to prevent arbitrarily long paths, the walk is terminated after ten steps if it still has not reached a gene in the linked locus.

The choice of a causative pathway is not as clear in this scenario. Many of the nodes that were visited were along spurious paths. However, we do not only want to find the most visited nodes along a linear path between the differentially expressed gene and the causative gene; this approach was chosen for its ability to find non-linear pathways. Thus, we count how many times all nodes were visited during the random walks. If the causative gene has been visited *c* times, we include all genes that have been visited some fraction *c/k* times (for example, *c/*3 times), and which are on a path from the differentially expressed gene to the causative gene.

Here we explain one small optimization of this approach. Instead of actually performing a random walk, we can replace it with a closed form equivalent; we will compute the probability the random walk will end at each gene in the linked locus. This improves efficiency, and more importantly it is able to exactly model the random walk without the need for many repeated trials.

If we number the nodes from 1 to *n*, let us denote the probability of being at node *i* after *t* steps as P*t*[*i*]. Let us denote the start node. We will denote the probability of transitioning to node *j* from node *i* to be *T*[*i, j*]. Then we can precisely compute P*t*[*i*]:

$$P_0[s] = 1 \text{ and } P_0[i] = 0 \text{ for } i \neq s.$$

$$Pt[\mathrm{j}] = !\mathrm{n\,i} = 1Pt - 1[i]T[i,j].$$

To get to node *j* at time *t*, the random walk must be in some neighbor node *i* at time *t* − 1, and then it must move from node *i* to node *j*.

Using these equations, the probability distribution can be efficiently calculated to any arbitrary time *t* by storing only two arrays of size *n*, one for the probabilities at time *t* and one for time *t* − 1.

While this approach is straightforward and has been used to discover pathways from eQTL data, several variants of this method still need to be explored. Several other distance metrics using random walks across a graph have been proposed – further testing is required to determine which method is optimal (see *(12)* for a review of other distance metrics). Furthermore, the interaction graphs in higher organisms are significantly larger; for large enough graphs, these methods will be impractical due to memory constraints.

2.4. Evaluation Using Gene Knockouts

It is currently difficult to assess the accuracy of inferred pathways from eQTL data because there is no good data set for validation. In virtually all cases, the true pathway is unknown. Furthermore, we cannot expect to find all previously known biological pathways – the nature of the eQTL data relies on natural genetic variation, and this variation may not perturb known pathways. Thus, there is no good "gold standard" data set (*see* **Note 4**).

Tu et al. proposed a solution to this problem by using gene knockouts to create a fake "gold standard" eQTL data set *(3)*. The Rosetta compendium of gene knockouts in yeast *(13)* provides expression levels in over 200 gene knockout experiments. This gives the same information as in eQTL data: expression levels combined with genetic variation. In the knockout experiments, we know precisely which gene has been removed, while genetic markers from eQTL studies narrow the region down to a locus with ~5–50 genes. To simulate eQTL data, we create a locus of ten genes around the true knockout. Each of the proposed methods can then be tested with the expression data and these created loci. If the causative gene in the inferred pathway is the gene knockout, then we conclude the method was successful. We are assuming that if the predicted pathway includes the correct causative gene, then the entire pathway is correct; this assumption will not always be true.

It is very difficult to determine which genes are truly differentially expressed due to the gene knockouts. Expression data is extremely noisy; to decide which genes are differentially expressed, we must set a threshold. Also, as proposed by Tu et al., we can cluster genes by common transcription factors, and only take large clusters. Both of these methods require subjective thresholds.

These methods greatly simplify our test cases. We have found the accuracy of these methods is highly dependent on the subjective decisions for gene expression. Nonetheless, though these test cases are far from perfect, this is a good first evaluation before turning to much more expensive and time-consuming verification with laboratory experiments.

3. Discussion

We have outlined several techniques for finding pathways from eQTL data. In our experiments on gene knockouts, we have found that all of the methods above perform significantly better than random choice. For example, these methods suggested an interesting feedback cycle in the MAPKKK pathway. The pathway starts from the pheromone response element Ste2 and after many interactions it regulates Dig1; we find that Dig1 in turn regulates the transcription of Ste2 (*see* **Fig. 10.2**). This feedback cycle has been reported before through direct experimentation; we simply highlight this as one noteworthy pathway uncovered by these methods *(14–17)*.

We also consider one example from using true eQTL from yeast *(1)*. Consider the differentially expressed gene YJR123W and the set of 16 genes that are in the linked locus, as in **Fig. 10.1**. All

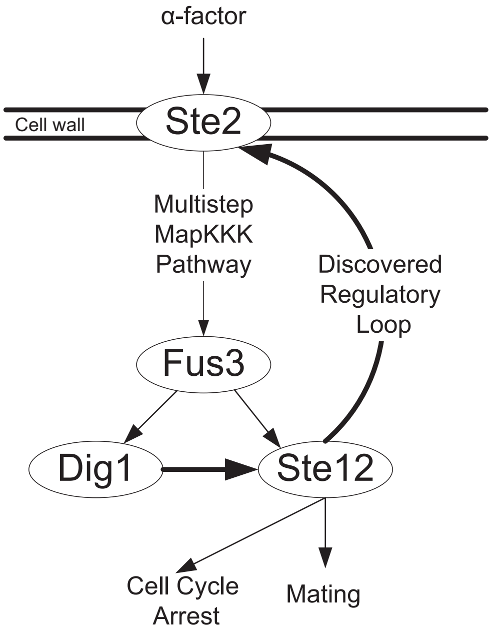

Fig. 10.2. **An example pathway found in gene knockout tests.** The MAPKKK pathway is known to start when the pheremone receptor Ste2 interacts with α-factor. After many steps, the pathway regulates Dig1 and Ste12. From knockout experiments, we found that Dig1 regulates Ste2, demonstrating a feedback cycle in the MapKKK pathway. This feedback cycle was already known, but it serves to demonstrate the types of pathways that can be uncovered through these methods.

of the methods discussed above consistently highly rank a pathway from YBR059C to YJR123W (*see* **Fig. 10.3**). We note that the roles we predict are consistent with the annotations in CYGD *(8)* and previous literature: YBR059C is a well-known kinase, and has been implicated as an important protein in signaling pathways *(18)*; YKR092C is known to be phosphorylated, and it is believed that it may bind DNA or act as a cofactor *(19)*; YDR174W is known to bind DNA or act as a cofactor for transcriptional regulation *(20, 21)*. As the true pathway is unknown in this case, we cannot be certain of these results without direct experimental validation. However, the agreement between this proposed pathway and the available literature data is very promising. In total, we predict 411 explanatory pathways for eQTL in yeast, and expect many of these will be correct given the accuracy in gene knockout experiments.

In general, we have found the accuracy of all the methods are extremely similar (*see* **Fig. 10.4**). There is also a high degree of overlap between the pathways found by each method (*see* **Fig. 10.5**). However, these results are highly dependent on the set of test conditions used (the *p*-value cutoffs for defining differentially expressed genes from the knockout data, as well as the

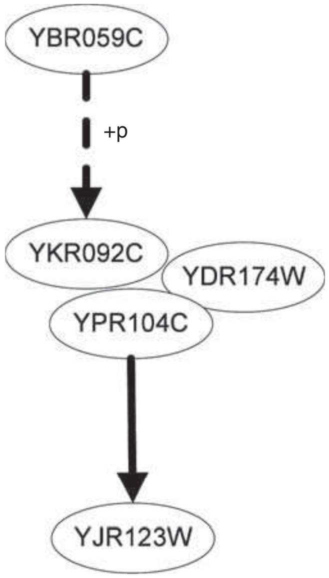

Fig. 10.3. A pathway explaining the differential expression of YJR123W in yeast eQTL experiments (see Fig. **10.1**). Here the *dashed line* represents the phosphorylation of YKR092C by the kinase YBR059C, the grouped proteins are a physical complex formation, and the final *solid line* is a transcriptional regulation interaction by transcription factor YPR104C to its target YJR123W.

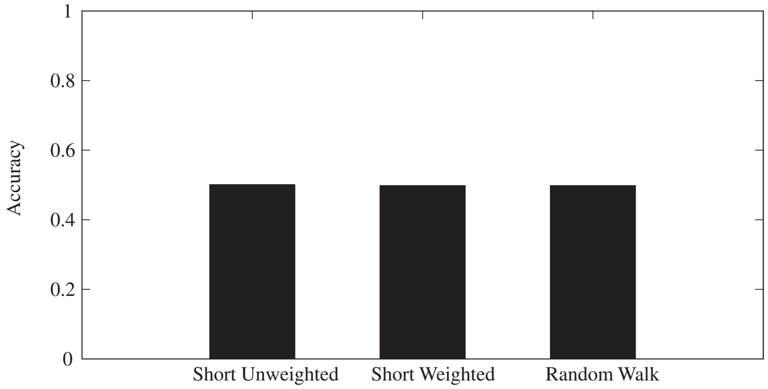

Fig. 10.4. **A comparison of the accuracy of the three methods to map pathways from eQTL**. All the methods have nearly identical accuracy on the gene knockout tests.

p-value cutoffs used in defining the interaction graph). We have found that under some conditions simply taking the shortest unweighted path outperforms other methods, while in other conditions the random walk method is the best. We are still attempting to uncover patterns that explain the differences in each method. However, we can clearly see that each method is capable of producing explanatory pathways, as seen in the examples above.

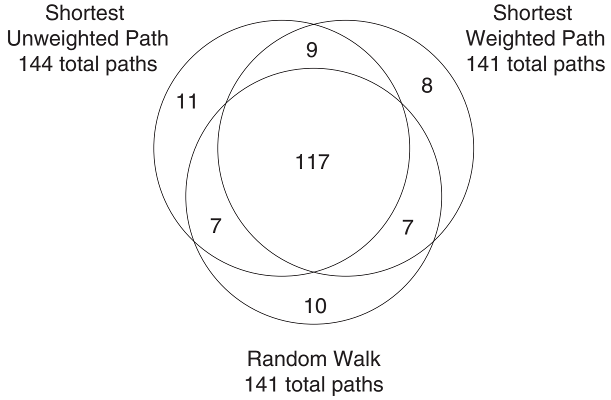

Fig. 10.5. **Overlap of pathways found by each of the three methods evaluated on the known gene knockout test cases**. Aside from minor differences, the methods are correct for the same test cases.

Our work highlights the strong dependence all of these methods have on the interaction graph. If even one interaction from the true pathway is missing from the interaction graph, none of these methods will be able to deduce the correct pathway. Furthermore, if the interaction graph contains many spurious edges, then all of these methods can easily be misled to deduce the wrong pathway, by finding a "short-circuited" pathway. We hope that this work will motivate continuing work to improve the accuracy and coverage of current molecular interaction databases (*see* **Note 2**).

Though these approaches have been somewhat successful, they are far from complete. They do not achieve perfect accuracy, and they cannot find explanatory pathways in many cases. Indeed, we have also eliminated a large portion of the data by considering only highly differentially expressed genes and genes that are clustered by transcription factor. Still, we are able to find a large number of novel pathways with these methods.

The approach described here also serves as an extensible framework to try additional methodology. Each of the variants described here requires only very small changes to the underlying code. With minimal effort, additional approaches can be added and compared to the existing methodology. For example, one could easily test what effect changing the interaction graph has on the predictions. This framework greatly facilitates the development of techniques for generating new candidate pathways.

We wish to emphasize that these approaches only serve for hypothesis generation. Verification of the proposed pathways requires additional work in the laboratory. Though the accuracies of these methods are not ideal, they do narrow a nearly infinite space of possibilities down to a tractable number. We cannot and do not expect these methods to be perfect, given the errors in our interaction graph, the missing molecular interactions (e.g., metabolites), the noise in the gene expression levels, and the numerous other factors that are not modeled at all (e.g., chromatin remodeling).

One major drawback of all these approaches is that they are designed to find pathways when a gene shows strong linkage to only a single locus. This accounts for a small fraction (less than 20%) of real eQTL data – as shown by the studies of Brem and Kruglyak *(1)*. Several studies have proposed methods for analyzing eQTL data by simultaneously considering all genes and all loci *(22, 2)*. These approaches attempt to construct a complete network of all genes and their relationships using machine-learning techniques. However, these approaches only indicate the influence of some genes on the expression level of other genes; they do not deduce pathways of molecular interactions, which provide a mechanistic basis for the wiring diagram for gene expression.

The techniques evaluated in this manuscript and the techniques for network inference complement each other and should be integrated together to provide the most coherent explanation possible. As a first step, once the network reconstruction algorithms have uncovered the most important relationships, the techniques described here could be used to find the pathway of molecular interactions that are responsible. For example, if the network reconstruction algorithms suggest that gene A influences gene B (through some potentially long and indirect method), we could search for a pathway in the interaction graph connecting gene A to gene B. Furthermore, we feel that the interaction graph should be directly incorporated into the network reconstruction algorithms, through the use of structure priors, i.e., the network reconstruction algorithms should favor inferring relationships along pathways with very good explanations. We are currently exploring these approaches to enable the inference of a greater number of more accurate pathways.

4. Notes

1. *Importance of determining linkage.* As noted previously, all subsequent analysis is based on first using statistical tests to determine to which loci each gene is linked. There are several techniques that differ in their sophistication.
2. *Importance of the interaction graph.* All of the methods mentioned here are highly sensitive to the edges, which are in the interaction graph. If even one edge in the true pathway is missing from the graph of molecular interactions, then it will

be impossible to discover the true pathway. If the graph contains false edges, all of the graph searching methods can very easily be misled. Many of the interactions in our network come from high-throughput experiments, such as yeast two-hybrid or ChIP-chip experiments. Converting this data to a graph requires the use of subjective *p*-value cutoffs. This means the interaction graph will always contain some spurious interactions, and it will be missing many true interactions. We are still exploring the robustness of these methods to errors in the interaction graph.

3. *Converting correlations to distances.* When converting expression correlations to distances, large correlations must be converted into small distances and small correlations into large distances. We chose the distance to be $1-|\mathrm{corr}(i,j)|$; however, this is simply a convenient heuristic. Several other transformations could be applied; we have experimented with $-\log(|\mathrm{corr}(i,\ j)|)$ and $1/(|\mathrm{corr}(i,\ j)|)$, and found they all obtain similar results.
4. *No gold standard data set for testing.* As there is no eQTL data set for which all of the true pathways are known, it is very difficult to determine if any new methods are successful or not, without direct experimentation. The best test set that can be used currently is the gene knockout data. However, we must still decide to use some subjective criteria to decide which genes are differentially expressed. Furthermore, we do not know the true pathway between the knockout and the differentially expressed genes.

Acknowledgments

This work was supported by NSF Graduate Research Fellowship DGE0203031.

References

1. Brem RB, Storey JD, Whittle J, Kruglyak L. Genetic interactions between polymorphisms that affect gene expression in yeast. Nature 2005, 436:701–03.
2. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH. Genetics of gene expression surveyed in maize, mouse and man. Nature 2003, 422:297–302.
3. Tu Z, Wang L, Arbeitman MN, Chen T, Sun F. An integrative approach for causal gene identification and gene regulatory pathway inference. Bioinformatics 2006, 22:e489–96.
4. Doerge RW, Zeng ZB, Weir BS. Statistical issues in the search for genes affecting quantitative traits in experimental populations. Stat Sci 1997, 123:195–219.
5. Storey JM, Akey JM, Kruglyak L. Multiple locus linkage analysis of genomewide expression in yeast. Plos Biol 2005, 3:e267.

6. Chen L, Storey JD. Relaxed significance criteria for linkage analysis. Genetics 2006, 173:2371–81.
7. Lee S, Pe'er D, Dudley AM, Church GM, Koller D. Identifying regulatory mechanisms using individual variation reveals key role for chromatin modification. PNAS 2006, 103:14062–67.
8. Gldener U, Mnsterktter M, Kastenmller G, Strack N, van Helden J, Lemer C, Richelles J, Wodak SJ, Garca-Martnez J, Prez-Ortn JE, Michael H, Kaps A, Talla E, Dujon B, Andr V, Souciet JL, De Montigny J, Bon E, Gaillardin C, Mewes HW. Cygd: The comprehensive yeast genome database. Nucl Acids Res 2005, 33(Database Issue): D364–68.
9. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne J, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok M, Kellis DK, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA. Transcriptional regulatory code of a eukaryotic genome. Nature 2004, 431:99–104.
10. Ptacek J, Devgan G, Michaud G, Zhu H, Zhu X, Fasolo J, Guo H, Jona G, Breitkreutz A, Sopko R, McCartney RR, Schmidt MC, Rachidi N, Lee S, Mah S, Meng L, Stark MJR, Stern DF, De Virgilio C, Tyers M, Andrews B, Gerstein M, Schweitzer B, Predki PF, Snyder M. Global analysis of protein phosphorylation in yeast. Nature 2005, 438:679–84.
11. McDermott J, Bumgarner RE, Samudrala R. Functional annotation from predicted protein interaction networks. Bioinformatics 2005, 21:3217–26.
12. Chandra AK, Raghavan P, Ruzzo WL, Smolensky R. The electrical resistance of a graph captures its commute and cover times. In Proceedings of the Twenty First Annual ACM Symposium on Theory of Computing, 1989, pp. 574–86.
13. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, Kidd MJ, King AM, Meyer MR, Slade D, Lum PY, Stepaniants SB, Shoemaker DD, Gachotte D, Chakraburtty K, Simon J, Bard M, and Friend SF. Functional discovery via a compendium of expression profiles. Cell 2005, 102:109–26.
14. Fields S, Herskowitz I. The yeast ste12 product is required for expression of two sets of cell-type-specific genes. Cell 1985, 42: 923–30.
15. Song O, Dolan J, Yuan Y, Fields S. Pheromone-dependent phosphorylation of the yeast ste12 protein correlates with transcriptional activation. Genes Dev 1991, 5: 741–50.
16. Chen Q, Konopka BJ. Regulation of the g-protein-coupled alpha-factor pheromone receptor by phosphorylation. Mol Cell Biol 1996, 161:247–57.
17. Hartig A, Holly J, Saari G, MacKay VL. Multiple regulation of ste2, a mating type-specific gene of saccharomyces cerevisiae. Mol Cell Biol 1986, 6:2106–14.
18. Cope MJ, Yang S, Shang C, Drubin DG. Novel protein kinases ark1p and prk1p associate with and regulate the cortical actin cytoskeleton in budding yeast. J Cell Biol 1999, 144:1203–18.
19. Meier UT. Comparison of the rat nucleolar protein nopp140 with its yeast homolog srp40. Differential phosphorylation in vertebrates and yeast. J Biol Chem 1996, 271:19376–84.
20. Lu J, Kobayashi E, Brill SJ. Characterization of a high mobility group ½ homolog in yeast. J Biol Chem 1996, 271:33678–85.
21. Gadal O, Labarre S, Boschiero C, Thuriaux P. Hmo1, an hmg-box protein, belongs to the yeast ribosomal DNA transcription system. EMBO J 2002, 21:5498–507.
22. Lee SI, Pe'er D, Dudley A, Church G, Koller D. Identifying regulatory mechanisms using individual variation reveals key role for chromatin modification. Proc Natl Acad Sci USA 2006, 103:14062–67.