Biostatistics statement

Data processing

Raw microarray data were normalized using the Robust Multichip Array (RMA) method. The expression data were then re-normalized using a modified z-score described previously.[12-15] We calculated the log base 2 of normalized values above, computed Z scores for each array, multiplied the Z scores by 2, and added an offset of 8 units to each value. The reason for this transformation is to produce a set of Z-like scores for each array that have a mean of 8 and standard deviation of 2. The advantage of this modified Z score is that a two-fold difference in expression corresponds approximately to a 1-unit change.

QTL (eQTL) Mapping

eQTL mapping was performed at gene and exon levels through the WebQTL module on GeneNetwork using our published methods [12-14]. This methodology uses regression analysis to determine the association between variability in a trait vs. variability in alleles at markers across the genome. Simple interval mapping was performed to identify potential eQTLs that regulate *Mypn* expression levels and estimate the significance at each location using known genotype data for those sites. Composite interval mapping was also performed to control for genetic variance associated with major eQTLs and therefore identify any secondary eQTLs that may have been otherwise masked. Each of these analyses produce a likelihood ratio statistic (LRS), providing us with a quantitative measure of confidence of linkage between the observed phenotype, in this case variation in expression level of *Mypn*, and known genetic markers. The genome-wide significance for each eQTL was established using a permutation test that compared the LRS of our novel site with the LRS values for 1000~10,000 genetic permutations [16].

Identification of upstream candidate genes

To identify upstream gene of *Mypn*, we determined the 1.5-LOD location of the significant eQTL of *Mypn*. All genes in this eQTL region were used for candidate gene analysis. The following criteria were used to identify the most likely candidates: 1) the gene is highly expressed in the heart; 2) the gene is significant ($p < 0.05$) correlated with *Mypn* expression in the heart; 3) the gene has non-synonymous SNP, missense SNP or indel in coding regions of the gene, or the gene has significant cis-eQTL.[14]

Genetic correlation and partial correlation analysis

We computed Pearson product-moment correlations between expression of *Mypn* and expression of all other probe sets across the genome to produce sets of genetically correlated genes. After that, in order to identify biologically relevant correlates of *Mypn*, we also performed partial correlation analyses to remove linkage disequilibrium by controlling for cis-regulated genes near *Mypn*.[14] Both genetic correlation and partial correlation can be computed using the tools on GeneNetwork.

Gene set enrichment analysis

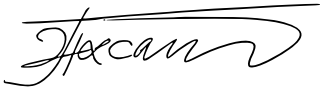
The genes that have both significant genetic correlation and partial correlation with *Mypn* were selected for gene set enrichment analysis. After removing Riken clones, intergenic sequences, predicted genes, and probes not associated with functional mouse genes, the remaining list of correlates with mean expression levels above baseline in the heart were uploaded to Webgestalt (http://bioinfo.vanderbilt.edu/webgestalt/) for gene enrichment analysis.[17] The *p* values generated from the hypergeometric test were automatically adjusted to account for multiple comparisons using the Benjamini and Hochberg correction.[18] The categories with an adjusted p value (adj P) of <0.05 indicated that the set of submitted genes are significantly over-represented in that categories.

Gene network construction

The gene network was constructed and visualized using Cytoscape utility through "Gene-set Cohesion Analysis Tool (GCAT)" (http://binf1.memphis.edu/gcat/index.py). The nodes in the network represent genes and the edge between two nodes represent cosine score of Latent Semantic Indexing (LSI) that determines the functional coherence of gene sets is larger than 0.6. The significance of the functional cohesion is evaluated by the observed number of gene relationships above a cosine threshold of 0.6 in the LSI model. The literature p-value (LP) is calculated using Fisher's exact test by comparing the cohesion of the given gene set to a random one.[19]

MALDI-TOF (MS) and TOF/TOF (tandem MS/MS) analysis

Both the resulting peptide mass and the associated fragmentation spectra were submitted to GPS Explorer version 3.5 equipped with MASCOT search engine (Matrix science) to search the database of National Center for Biotechnology Information non-redundant (NCBInr). Searches were performed without constraining protein molecular weight or isoelectric point, with variable carbamidomethylation of cysteine and oxidation of methionine residues, and with one missed cleavage allowed in the search parameters. Candidates with either protein score C.I.% or Ion C.I.% greater than 95 were considered significant.



02/08/2017