*Physiol Genomics* 17: 140–149, 2004.
First published January 27, 2004; 10.1152/physiolgenomics.00158.2003.

# An integrated reverse functional genomic and metabolic approach to understanding orotic acid-induced fatty liver

**Julian L. Griffin,[1] Stephanie A. Bonney,[2] Chris Mann,[2] Abdul M. Hebbachi,[3] Geoff F. Gibbons,[3] Jeremy K. Nicholson,[4] Carol C. Shoulders,[2] and James Scott[5]**

[1]*Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW;* [2]*MRC Clinical Sciences Centre, Faculty of Medicine, Imperial College London, Hammersmith Hospital, London W12 ONN;* [3]*Metabolic Research Laboratory, Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Headington, Oxford OX3 9JL;* [4]*Biomedical Sciences, Faculty of Medicine, Imperial College London, Exhibition Road, South Kensington, London; and* [5]*Genetics and Genomics Research Institute, Faculty of Medicine, Imperial College London, South Kensington, Armstrong Road, London SW7 2AZ, United Kingdom*

Submitted 23 September 2003; accepted in final form 12 January 2004

**Griffin, Julian L., Stephanie A. Bonney, Chris Mann, Abdul M. Hebbachi, Geoff F. Gibbons, Jeremy K. Nicholson, Carol C. Shoulders, and James Scott.** An integrated reverse functional genomic and metabolic approach to understanding orotic acid-induced fatty liver. *Physiol Genomics* 17: 140–149, 2004. First published January 27, 2004; 10.1152/physiolgenomics.00158.2003.—In functional genomics, DNA microarrays for gene expression profiling are increasingly being used to provide insights into biological function or pathology. To better understand the significance of the multiple transcriptional changes across a time period, the temporal changes in phenotype must be described. Orotic acid-induced fatty liver disease was investigated at the transcriptional and metabolic levels using microarrays and metabolic profiling in two strains of rats. High-resolution $^1$H-NMR spectroscopic analysis of liver tissue indicated that Kyoto rats compared with Wistar rats are predisposed to the insult. Metabolite analysis and gene expression profiling following orotic acid treatment identified perturbed metabolic pathways, including those involved in fatty acid, triglyceride, and phospholipid synthesis, β-oxidation, altered nucleotide, methyl donor, and carbohydrate metabolism, and stress responses. Multivariate analysis and statistical bootstrapping were used to investigate co-responses with transcripts involved in metabolism and stress responses. This reverse functional genomic strategy highlighted the relationship between changes in the transcription of stearoyl-CoA desaturase 1 and those of other lipid-related transcripts with changes in NMR-derived lipid profiles. The results suggest that the integration of $^1$H-NMR and gene expression data sets represents a robust method for identifying a focused line of research in a complex system.

transcriptomics; metabolomics/metabonomics; DNA microarray; system biology; bioinformatics

DNA MICROARRAYS are increasingly being used in functional genomic strategies for understanding gene function and pathology (5, 26, 27). To fully use the information generated by this technology, it is necessary to have a description of the changing phenotype so that gene expression can be understood in a biological context. To achieve this, there is recognition that the large-scale analysis of metabolites, as by proton nuclear magnetic resonance spectroscopy ($^1$H-NMR) or mass spectrometry, provides a final conduit to link differential mRNA responses with specific metabolic pathways by defining a metabolic phenotype (10, 12, 24).

The analysis of multivariate data sets that are generated within these different tiers of biological organization may potentially yield critical information about the mechanisms involved in biological transitions. However, enhanced understanding of complex biological systems requires the development of strategies for integration of the different tiers, which link information derived at each level in a systematic way (24, 34). In particular in complex biological systems where large data sets will be the routine, there is a need to integrate the different functional genomic approaches, taking into consideration the scaling and correlation problems associated with this integration.

In this study a simple perturbation of a biological system of relevance to our sphere of interest is investigated in terms of an integrative approach to explain the linkage between changes in gene expression and associated changes in metabolic pathways identified by NMR. Profound fatty liver can be induced in certain vertebrates by the feeding of orotic acid, a naturally occurring biological precursor of uridine (19, 25). This phenotype is maintained without obvious deleterious effect to the animal throughout extended periods of orotic acid feeding and is manifest over a period of several days in liver cells taken from orotic acid fed rodents and grown in primary culture in the absence of orotic acid (15, 20). Thus the hepatic phenotype should be tractable to functional genomic analysis.

We have used a two-tier strategy to define the metabolic and gene regulatory responses, by metabolic profiling (metabolomics/metabonomics) and DNA microarrays, created by administrating orotic acid in an inbred (Kyoto) and outbred (Wistar) strain of rat. As such, the system consisted of a simple perturbation but was subject to polygenic variation. The integration of the data from this two-tier system describes, in an understandable and biologically coherent direction, a series of relationships between nucleotide, fatty acid, glycerolipid, and cholesteryl ester biosynthesis, and of stress responses during orotic acid feeding. The approach also identified unexpected links, including an inverse correlation, between stearoyl-CoA desaturase 1 (SCD 1) mRNA levels and unsaturated fatty acids. We conclude that the integration of $^1$H-NMR and gene expression data can provide a robust approach to identify specific pathways and cellular responses underlying a changing pheno-

Article published online before print. See web site for date of publication (http://physiolgenomics.physiology.org).

Addresses for reprint requests and other correspondence: J. L. Griffin, Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, UK (E-mail: jlg40@mole.bio.cam.ac.uk); and J. Scott, Genetics and Genomics Research Institute, Faculty of Medicine, Imperial College London, South Kensington, Armstrong Road, London SW7 2AZ, UK (E-mail: j.scott@imperial.ac.uk).

type and, in addition, yield new insights into the mechanisms involved in the biological transition.

## MATERIALS AND METHODS

*Animal handling and sample collection.* All animal procedures conformed to Home Office, UK, guidelines for animal welfare. Rats (Charles River UK) were fed either standard laboratory diet or diet supplemented with 1% orotic acid (Sigma Aldrich UK) ad libitum. Orotic acid feeding is a well-characterized method for inducing reversible fatty liver disease in rodents (19, 25). Animals appeared to remain healthy throughout the feeding period. Rats were killed by cervical dislocation, and the left lateral lobe of the liver was excised. No physiological abnormalities were detected by visual inspection of the animals postmortem. Tissues were snap frozen and stored at −80°C.

*Metabolic profiles derived via $^1$H-NMR spectroscopy.* Solution-state $^1$H-NMR spectroscopy was performed on hepatic tissue extracts and blood plasma at 600.2 MHz using a Bruker AVANCE spectrometer interfaced with a 14.1-Tesla superconducting magnet and high-resolution inverse geometry $^1$H-NMR probe. Extracts of hepatic tissue ($n = 3$ for each time point and both strains) were prepared using an acetonitrile/water mixture as previously described (13) and reconstituted in $^2H_2O$ containing 4 mM trisilylpropionic acid (TSP). For the analysis of blood, plasma ($n = 3$ for each time point and both strains) was diluted twofold in a phosphate buffer containing $^2H_2O$ and 4 mM TSP. Solvent-suppressed spectra were acquired for tissue extracts and blood plasma into 32 k data points, averaged over 128 acquisitions. Spectral assignments were made with reference to previously published literature (2) and confirmed using two-dimensional pulse sequences.

Intact liver tissue (~10 mg; $n = 3$ for each time point and both strains) was examined using a high-resolution magic angle spinning (HRMAS) probe placed in the spectrometer described above. Solid-state spectra were acquired by spinning the samples at 5,000 Hz using a solvent-suppression pulse sequence to reduce the spectral intensity of the water resonance. Spectra were averaged over 256 acquisitions.

Spectra were processed using XWINNMR software, version 3.1 (Bruker, Germany). Following multiplication with a preexponential factor of 1 Hz and Fourier transformation, spectra were integrated across 0.04 ppm spectral regions between 0.4 and either 4.2 or 9.4 ppm for solid- and solution-state NMR, respectively, using the AMIX software package (Bruker). The different integral regions used were a result of the rolling baseline produced in some solid-state spectra caused by unresolved protein resonances. For solution-state spectra the region around the water resonance was excluded to minimize the effect of variable water suppression in the subsequent pattern recognition. Output vectors representing each spectrum were normalized across the integral regions so that each individual 0.04-ppm region was represented as a ratio to the entire spectral intensity, excluding the water resonance (13). While this produces data sets with inherent redundancy due to metabolites with multiple integral regions at different chemical shifts, this redundancy confirms spectral assignments. Metabolite ratios were derived by manual integration of the resonances specified.

*Lipid profiling by gas chromatography.* Livers ($n = 3$ for each time point and both strains) were homogenized using 5 vol (wt/vol) of 0.25 M sucrose. Following solvent extraction of the total lipid fraction, the triglyceride and cholesterol contents were determined using commercial kits (Boehringer-Mannheim, Lewes, UK). Fatty acid profiles of the total hepatic lipid content were obtained by gas chromatography following transesterification using a mixture of toluene/sulfuric acid in methanol (11).

*Gene expression profiling using microarrays.* Total RNA was extracted by an RNA isolation kit (Stratagene) from the livers of Wistar and Kyoto rats at *day 0* (Wistar $n = 3$, Kyoto $n = 2$), *days 1* and *3* (Wistar $n = 2$, Kyoto $n = 2$), and *day 14* (Wistar $n = 3$, Kyoto $n = 2$). cDNA was generated from total RNA using the SuperScript double-stranded cDNA synthesis kit (Invitrogen, Life Technologies), and cRNA was prepared by in vitro transcription (BioArray High Yield RNA transcript labeling kit; Affymetrix UK) and purified on an affinity resin column (RNeasy, Qiagen, UK). Biotinylated cRNA was fragmented and hybridized to RG-U34A Affymetrix GeneChips (Affymetrix, Santa Clara, CA; **http://www.affymetrix.com/support/technical/datasheets/rgu34_datasheet.pdf**). Data were analyzed with the Affymetrix Microarray Suite 5.0 algorithm to generate a "signal" value and a "detection" label. Prior to normalization, probes that generated an absent call in all cells were removed. Normalization was calculated using a global scaling method based on the entire data set. Globally scaled values <10 were rounded up to 10. The data set is deposited as a series at the Gene Expression Omnibus (**http://www.ncbi.nlm.nih.gov/geo**) and has the accession number GSE665.

*Pattern recognition and data processing.* All data are represented as means ± SD unless otherwise specified. The $^1$H-NMR spectra and transcriptional data sets were imported into the SIMCA package (Umetrics, Umea, Sweden) and preprocessed by three different scaling processes (9): *1*) mean centering by measuring the variance of each integral region about the mean; *2*) univariate using mean centering followed by scaling the variance by $1/s_k$, where $s$ is the standard deviation of the variable (integral region) $k$; and *3*) Pareto scaled to $(1/s_k)^{1/2}$. Pareto-scaled data sets produced the most robust models and are reported in this paper. Initially each data set was examined using principal component analysis (PCA) to examine trends and clusterings. Time progression of the data sets were interrogated by prediction to latent structures through partial least squares (PLS) using each integral region as an $X$-variable and time as the $Y$-variable. PLS calculates the linear relationship between a matrix $Y$ (in this case time) and a matrix $X$ (integral regions representing metabolites or transcripts). This may be represented as $Y = f(X) + E$, where $E$ represents the residuals not modeled.

Three types of preprocessing were applied to the $X$-matrix of integral values. Correlations were tested as part of the routine by sequentially leaving out every seventh sample and predicting its class membership. This goodness of fit algorithm was used to determine whether a correlation was significant ($Q^2 > 0.097$) and was derived from the predicted error sum of squares between the actual and predicted $Y$-matrix (PRESS) and the sum of squares proportion of the $X$-matrix explained by the model (SS) according to: $Q^2 = 1 - \text{PRESS/SS}$. While PLS models two data matrices in terms of linear regression, nonlinear responses may be modeled using a polynomial relationship or produces curvature in a linear model between PLS components representing the $X$- and $Y$-block data sets.

Transcriptional data was also ordered using hierarchical clustering with a variety of algorithms, including Euclidean distance, Standardized Euclidean distance (each coordinate in the sum of squares is inversely weighted by the sample variance of that coordinate), Mahalanobis distance, City Block/Manhattan distance, or Minkowski distance according to temporal and strain profiles within the program Matlab (**http://www.mathworks.com**). Hierarchical clustering plots were cross-checked using the cophenetic correlation coefficient between the original distance data and the clusterings formed as described in the statistical toolbox of Matlab.

*Genomic and metabolic correlations.* To correlate differentially expressed transcripts with changes in metabolites, two matrices were formed for PLS regression within SIMCA (Umetrics) with $^1$H-NMR spectra forming the $X$-block and transcriptional changes the $Y$-block. Both strains of rats were included in the same model, but as separate genomic/metabolic inputs. The PLS model mapped each observation in terms of metabolic and transcriptional variation to identify those animals where these changes were maximal and whether there was a trend in these changes in terms of time or rat strain. The variables most significant to the regression model were identified using the loadings plot derived from the PLS model, and these were used to

identify transcriptional and metabolic co-responses. For tissue taken from animals at *days 1*, *3*, and *14* with no corresponding gene array data, an average was taken for that rat strain and exposure period.

## RESULTS

*Characterization of metabolic phenotype prior to orotic acid feeding.* The baseline total hepatic triglyceride content of the outbred Wistar and inbred Kyoto rat strains were assayed biochemically and found to be comparable (Wistar, 56.4 ± 5.6 μg/mg of protein; Kyoto, 54.1 ± 19.25 μg/mg of protein). By contrast, the principal component analysis (PCA) of $^1$H-NMR spectra of intact liver tissue and of aqueous soluble tissue extracts identified distinct metabolic profiles for the two rat strains (Fig. 1, *A* and *B*). Liver tissue from Kyoto rats compared with the Wistar rats contained a higher proportion of mobile lipids, including both saturated ($CH_2$**$CH_2$**$CH_2$) and unsaturated (**CH═CH**) lipid moieties (bold type signifies proton resonance observed), although this did not produce an increase in cholesterol, and phosphatidylcholine. Liver from Kyoto rats also had less glucose and glycogen (Fig. 1, *A* and *B*). These HRMAS $^1$H-NMR observable resonances derive mainly from mobile cytoplasmic lipid droplets, rather than membrane lipids, and are detected because they have a high level of isotropic mobility to produce narrow NMR signals (14). Cytoplasmic lipid droplets are usually mainly stored triglyceride. Spectra of tissue extracts also demonstrated that hepatic tissue from Kyoto rats contained more of the choline oxidation products, betaine and trimethylamine oxide (TMAO). No strain differences were observed for the resonances associated with nucleosides including uridine, cytidine, or adenosine. Thus the Kyoto rat has higher baseline cytoplasmic triglyceride and methyl donors and lower glucose and glycogen and, compared with the Wistar rat, may be pharmacogenomically predisposed to develop fatty liver.

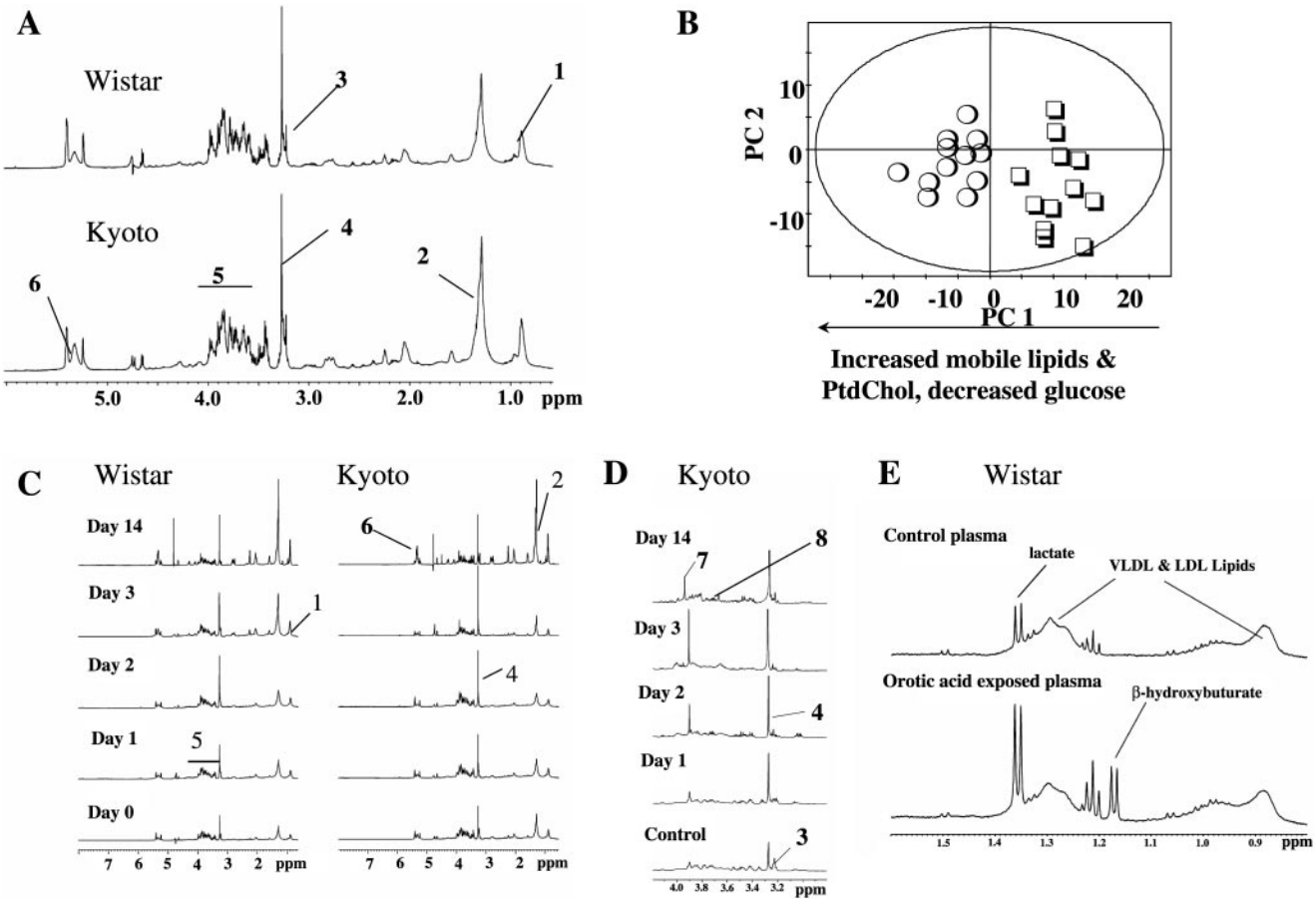


Fig. 1. Proton NMR-derived and principal component analysis (PCA)-derived metabolic profiles. *A*: high-resolution $^1$H-NMR spectra of hepatic tissue from Wistar and Kyoto rats fed normal laboratory diet. Kyoto rats had a higher proportion of mobile lipids including unsaturated fatty acids compared with Wistar rats (assignments made with reference to Ref. 2, according to line shape analysis and two-dimensional pulse sequences). *B*: PCA of hepatic tissue from Kyoto (○) and Wistar (□) rats fed a normal diet. The arrow demonstrates major metabolic differences along PC1. *C*: $^1$H-NMR spectra of hepatic tissue from Wistar and Kyoto rats exposed to orotic acid. Orotic acid initially induced a large increase in phosphatidylcholine and trimethylamine oxide (TMAO; co-resonant in intact tissues, but with phosphatidylcholine having a broader resonance). Hepatic tissue from both rat strains began to accumulate large amounts of mobile lipids between *days 3* and *14*, and this was more apparent in Kyoto rats. *D*: in aqueous extracts of hepatic tissue from Kyoto (shown) and Wistar rats, increases in TMAO, betaine (distinguished by $CH_2$ resonance at 3.90 ppm), and glycine, and a decrease in choline, were recorded. *E*: hepatic tissue changes were also accompanied by systemic metabolism changes, including an increase in plasma concentrations of β-hydroxybutyrate and lactate (Wistar rats shown). *1*) **$CH_3$**— lipid, *2*) $CH_2$**$CH_2$**$CH_2$ lipid, *3*) $N(CH_3)_3$ choline, *4*) $N(CH_3)_3$ phosphatidylcholine (broad resonance) and $N(CH_3)_3$ TMAO (minor contributions from betaine and phosphocholine), *5*) glucose and glycogen, *6*) **CH═CH** lipids, *7*) $CH_2$ betaine, and *8*) glycine. PtdChol, phosphatidylcholine.

*Metabolic changes induced by orotic acid.* The amount of triglyceride in livers from Wistar rats fed orotic acid for 14 days increased 8-fold relative to animals fed a diet without orotic acid supplementation, whereas the Kyoto strain demonstrated a massive 40-fold increase in total triglyceride content (Wistar, 438.1 ± 207.9 μg/mg of protein; Kyoto, 1,979.6 ± 189.1 μg/mg protein). Both strains of rats demonstrated a time-dependent accumulation of mobile lipids within their livers as detected by HRMAS NMR spectroscopy. The lipid moieties associated with **CH═CH**, $CH_2$**$CH_2$**$CH_2$, and terminal group ($CH_2$**$CH_3$**) resonances were particularly increased (Fig. 1*C*). By *day 14* this rise in lipids demonstrated a higher proportion of unsaturated lipids ($Integral_{CH=CH}/Integral_{CH2CH2CH2}$ Wistar control = 0.166 ± 0.001; Wistar *days 1–3* = 0.149 ± 0.005; Wistar *day 14* = 0.210 ± 0.004; Kyoto control = 0.166 ± 0.005; Kyoto *days 1–3* = 0.154 ± 0.012; Kyoto *day 14* = 0.216 ± 0.005; $P < 0.001$ for *day 14* being different from control and *days 1–3* for both strains). The resonance from the $C_{18}$ $CH_3$ group in cholesterol and its esters was clearly increased by *day 14* of orotic acid exposure in both rat strains (data not shown). Substantial increases in phosphocholine and phosphatidylcholine resonances were detected in both strains from *day 1* of exposure (Fig. 1*C*). This was accompanied by a 50% decrease in choline (Fig. 1*D*). PCA of intact liver spectra for both rat strains also revealed two distinct components associated with orotic acid feeding (results not shown). These were increases in TMAO, phosphocholine, and phosphatidylcholine over the first 3 days, and the subsequent rise in mobile lipids, accompanied by the decrease in glucose and glycogen from 3–14 days. The increase in mobile lipids was much more apparent in the Kyoto rats.

To substantiate the differences in the hepatic lipid profiles of rats fed orotic acid, we used gas chromatography analysis on liver tissue extracts. The total amount and proportion of lipid associated with 16:0, 16:1, 18:0, and 18:1 fatty acids increased with exposure to orotic acid. Importantly, there was a more marked increase in 16:1 and 18:1 fatty acid lipids, along with a proportional decreases in 18:0 fatty acids (Table 1). Thus the increased ratio of unsaturated to saturated lipids in the cytoplasm detected by HRMAS $^1$H-NMR spectroscopy was also reflected in the total lipid content.

As lipid resonances dominate the intact liver spectra, tissue extracts were also examined. Orotic acid was detected in the extracts for all animals receiving the dietary supplement and produced a large increase in uridine nucleotides (20- and 40-fold increases for Wistar and Kyoto rats, respectively; data not shown). A decrease in adenosine nucleotides was detected in Kyoto but not in Wistar rats. For both rat strains, PCA of the extract spectra separated the hepatic tissues exposed to orotic acid from control tissues (data not shown). The metabolites responsible for this separation included increases in uridine nucleotides, TMAO, betaine, and glycine and a decrease in choline (shown for Kyoto rats; Fig. 1*D*).

Table 1. *Liver fatty acid profiles following orotic acid feeding for 14 days as measured by GC-MS*

| | Wistar | | Kyoto | |
|---|---|---|---|---|
| | Std diet | Std diet + OA | Std diet | Std diet + OA |
| 16:0 | 21.9±0.7 | 23.6±0.8 | 23.3±0.9 | 27.5±4.4 |
| 16:1 | 0.74±0.23 | 1.7±0.3 | 0.81±0.06 | 1.88±0.15† |
| 18:0 | 21.6±0.5 | 15.1±0.6‡ | 20.0±1.1 | 11.6±1.2* |
| 18:1 | 6.6±0.3 | 17.4±1.5† | 7.2±0.1 | 22.6±1.7‡ |
| 16:1/16:0 | 0.033±0.008 | 0.072±0.008‡ | 0.035±0.002 | 0.071±0.01‡ |
| 18:1/18:0 | 0.31±0.02 | 1.17±0.15‡ | 0.36±0.01 | 1.951±0.07‡ |

Values are means ± SE of the percentage proportion of the total fatty acids extracted (n = 3). Std diet, standard laboratory diet; OA, orotic acid; GC-MS, gas chromatography-mass spectroscopy. *$P<0.05$, †$P<0.005$, and ‡$P<0.001$, according to Student's *t*-test for a significant difference between the means.

Orotic acid feeding also induced systemic changes in blood metabolite concentrations in both strains (Wistar rats are shown Fig. 1*E*). By TSP there were marked increases in plasma concentrations of β-hydroxybutyrate, which was undetectable in plasma from control rats. Lactate, acetate (not shown), and glucose (not shown) were also increased. There was also the expected decrease in the intensity of the LDL and VLDL signals (compound resonances at ~1.0 and 1.3 ppm for both $CH_2$ and $CH_3$) and a decrease in phosphatidylcholine.

A feature of the orotic acid fed rat is that the phenotype completely corrects with concomitant adenosine feeding (19). In the present study, the metabolic profiles of intact liver, liver extracts, and plasma from control animals and Wistar rats fed orotic acid plus adenosine for 14 days were indistinguishable (data not shown).

Overall orotic acid feeding increased hepatic cytoplasmic triglyceride accumulation, and this was more marked in the predisposed Kyoto rat strain than the Wistar strain. In both strains there were also prominent increases in total cholesterol, phosphocholine, phosphatidylcholine, and unsaturated fatty acids compared with saturated fatty acids in terms of the total lipid content of the tissue. The methyl donors, betaine and TMAO, and pyrimidine nucleotides were increased in both strains and were particularly marked in the predisposed Kyoto strain.

*Gene expression profiling during orotic acid feeding.* The gene expression profiles of the two rat strains at 1, 3, and 14 days were assessed relative to control animals using the Affymetrix RG-U34A GeneChip, containing 7,000 known transcripts and 1,000 expressed sequence tags (ESTs). Of these 8,000 transcripts, 4,132 expression elements were expressed in the microarray experiment.

We further investigated the gene microarray data set using a simple statistical filter to identify the principal genes involved in the orotic acid feeding perturbation in both strains of rats compared with the pre-dose animals. We argued that by combining the transcription data sets that genes changes common to both strains compared with control animals would represent those most affected by the orotic acid treatment. Although these transcript changes were at different levels in the different strains, combining the gene expression data from both strains increased the power for detecting genes affected by orotic acid feeding, while minimizing the number of false possibilities in the data set.

Using a threefold change in conjunction with statistical support of $P < 0.05$ for a change in transcription as a filter for the data, we identified the expression of 48 gene sequences as being altered by orotic acid feeding over the time course (Fig. 2*A*) (17). At a *P* value of <0.001, there was also support for orotic acid having an effect (i.e., between 1.5-fold and 3-fold expression change) on the expression of 12 further sequences. This analysis demonstrated that many of the transcripts identified by multivariate analysis did not pass through this more stringent reproducibility test, a problem associated with the

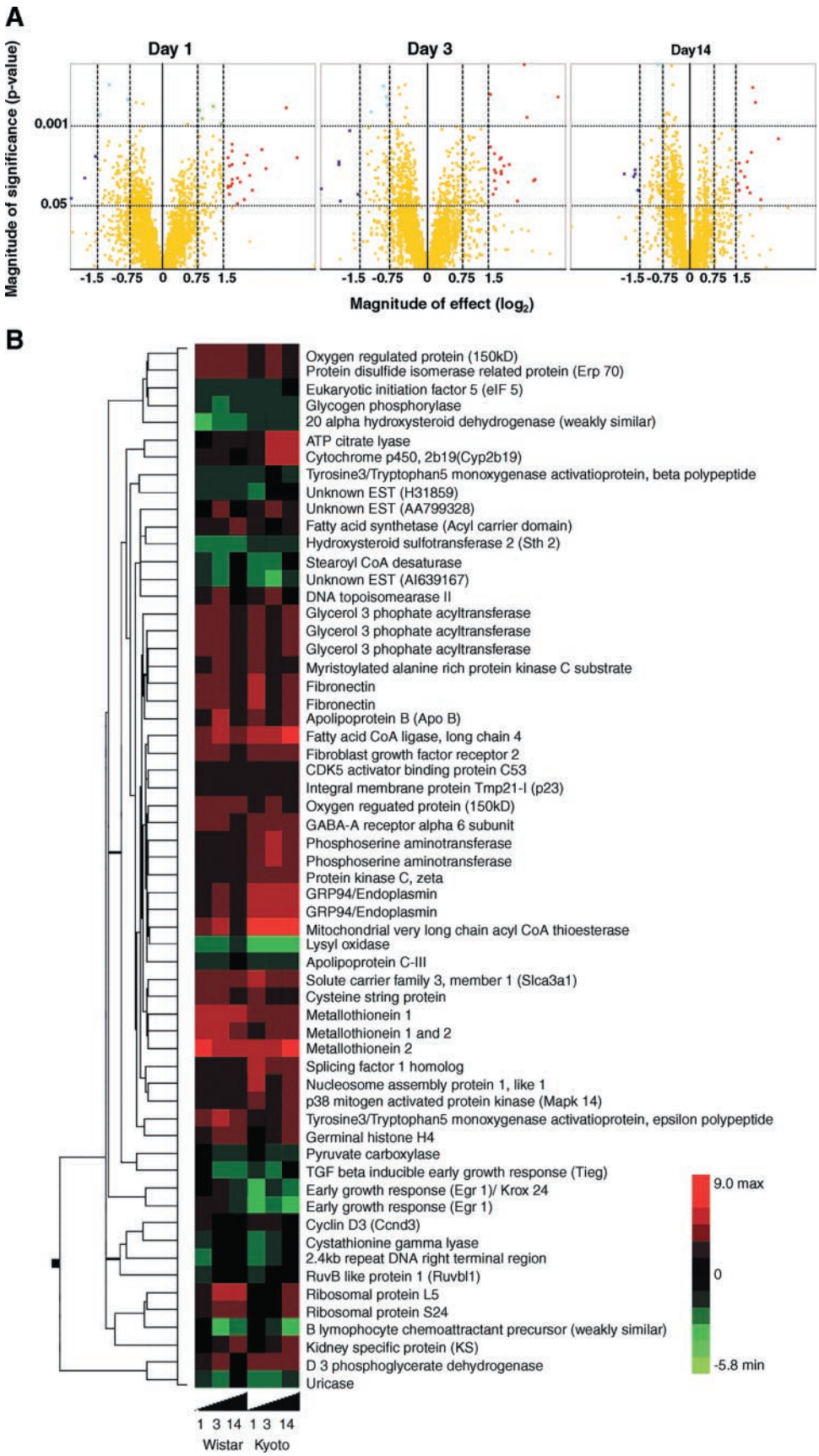


Fig. 2. *A*: a "volcano plot" of hepatic gene expression in animals fed orotic acid. Each point represents a gene from the microarray experiment. The magnitude of expression is shown on the *x*-axis, and the *P* value of the ANOVA significance test is on the *y*-axis. Red and dark blue indicate transcripts that are significantly ($P < 0.05$) overexpressed or underexpressed by 3-fold, respectively. Green and light blue indicate transcripts that are significantly ($P < 0.001$) overexpressed or underexpressed by 1.5-fold but less than 3-fold, respectively. *B*: hierarchical clustering of the 60 most significant gene changes. Clustering was performed using an average linkage algorithm with Manhattan distance method. Rows represent individual genes; columns represent Wistar and Kyoto rats at 1, 3, and 14 days of orotic acid feeding, respectively. Relative gene expression is indicated by color ranges from highly increased (intense red) to highly decreased (intense green). Yellow dots signify transcripts that showed no significant change in expression.

much larger transcriptional data sets compared with those derived from $^{1}$H-NMR spectra. Also, in the case of ATP citrate lyase only one of the four transcripts represented on the microarray GeneChip passed the filter as being significantly increased, highlighting the possibility of identifying false positives in microarray experiments. However, as microarray experiments are also affected by false negatives, ATP citrate lyase was kept in the data set for subsequent analyses. ATP citrate lyase was the only gene highlighted in this manner. This transcriptional data set of 60 transcripts was used in subsequent correlation analyses, but the different strains were considered independently to examine how these common transcriptional changes differed in the two rat strains.

The 60 transcripts identified included those involved in fatty acid biosynthesis and desaturation [ATP citrate lyase, fatty acid synthetase (FAS), SCD 1], fatty acid metabolism (cytochrome *P*-450, 2b19), triglyceride biosynthesis [glycerol-3-phosphate acyltransferase (GPAT)], VLDL secretion (apoB, apoCIII), fatty acid activation and the regulation of β-oxidation (fatty acid CoA ligase 4, mitochondrial very-long-chain acyl-CoA thioesterase), glycogen synthesis and gluconeogenesis (glycogen phosphorylase, pyruvate carboxylase), serine biosynthesis (phosphoserine aminotransferase, 3-D-phosphoglycerate dehydrogenase), endoplasmic reticulum and cytoplasmic stress responses [GRP94/endoplasmin, fibronectin, oxygen regulated protein (150 kDa), metallothionein 1 and 2, protein disulfide isomerase (PDI)-related protein, ERP70], inflammation [B lymphocyte chemoattractant precursor (weakly similar), p38 mitogen activated protein kinase], and growth control (TGFβ inducible early growth response, early growth response 1, cyclin D3).

Hierarchical clustering was used to organize the 60 differentially expressed gene transcripts according to temporal and strain profiles (Fig. 2*B* for Manhattan distance method). All of the plots were robust, as judged by cophenetic coefficient values >0.88, and as expected, transcripts from the same gene, or genes with similar function, clustered together. However, the majority the 60 transcripts could not be separated into distinctly regulated subgroups and thus provided minimal information. Despite this, the transcripts consistently perturbed in the two strains indicate the areas of metabolism affected by orotic acid and proteins on the metabolic pathways involved. These results are highly plausible biologically and consistent in their direction of change with the changes in metabolites. This is with the exception of the increase in unsaturated fatty acids induced by orotic acid and the decrease in the mRNA for SCD 1, which appear to move in counterintuitive directions. The metabolite and transcript changes are illustrated on Fig. 3.

*Integration of metabolic and gene expression profiles.* We elected to use supervised PLS multivariate analysis to integrate the metabolic and gene expression data sets. Although PLS is a linear approach, any nonlinear responses are generally disclosed as a curve in the data. By regressing the metabolic changes in the soluble extracts against transcriptional changes,

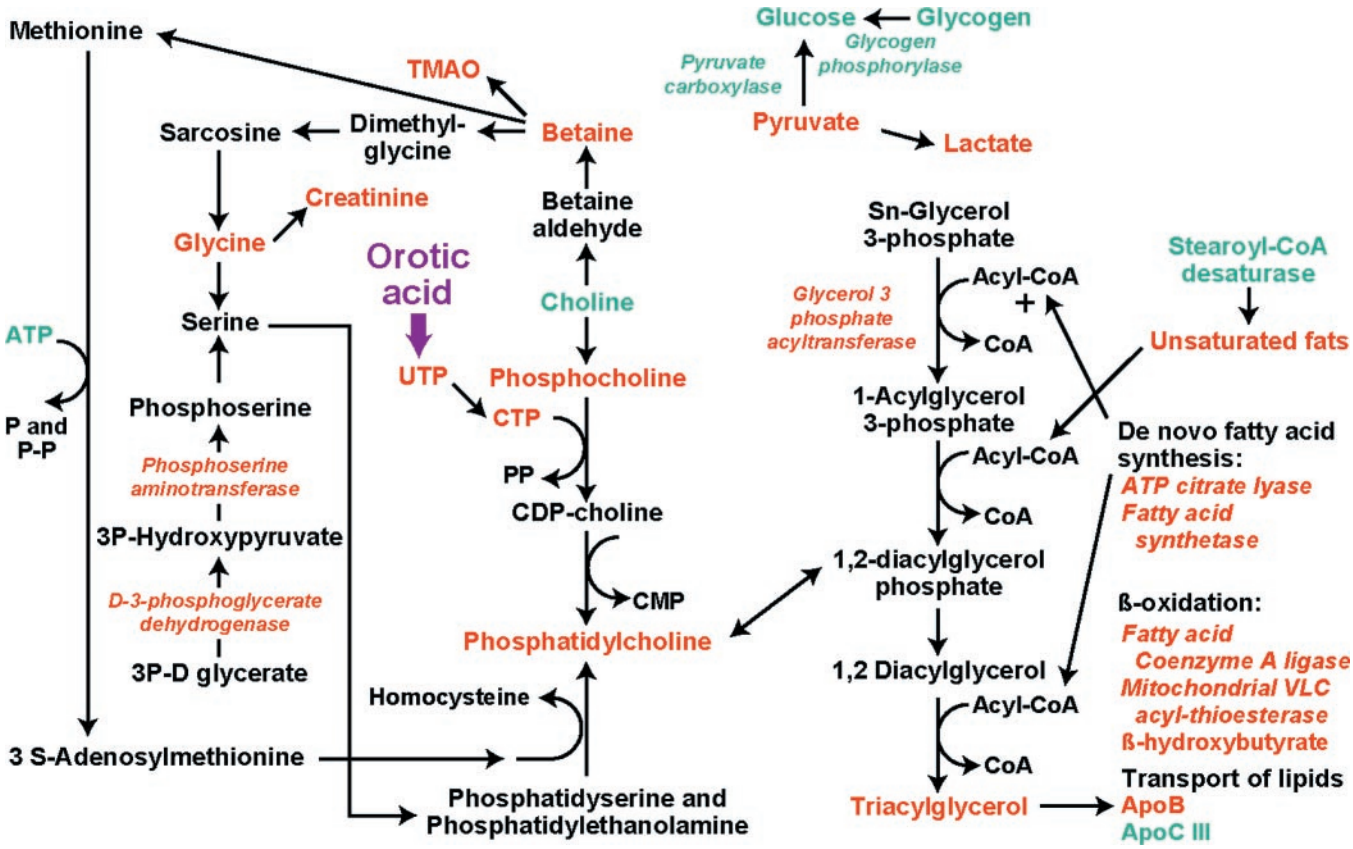


Fig. 3. Perturbed metabolic pathways identified by gene transcripts and metabolic profiles over 14 days of exposure. Pathways were traced using the Kyoto Encyclopedia of Genes and Genomes (KEGG), with those associated with triglyceride/phospholipid synthesis, choline metabolism, and methyl donor metabolism. Red = increased; green = decreased; italics signify enzymes with perturbed transcription.

a linear response was identified in Kyoto rats for the time period examined (Fig. 4*A*). In Wistar rats a trend was much less pronounced as had previously been observed in the PCA of the NMR spectra data. The PLS component correlated many of the differentially expressed transcripts with metabolites that have been previously detected by PCA of the metabolic data set and transcripts (Fig. 4, *B* and *C*). This included an increase in uridine nucleotides, phosphocholine, TMAO, and betaine and a decrease in glucose and choline (Table 2). These changes had positive co-responses with expression of metallothionein 1 and 2, GRP94 and cytochrome *P*-450 (2b19), FAS and the PDI-related protein, ERP70, and negative co-responses including SCD 1, apoCIII, glycogen phosphorylase, and pyruvate carboxylase.

The PLS model built using the intact tissue metabolic data set demonstrated two PLS components: the first described in the above model, and the second related to increased lipid triglyceride deposition in both strains (Fig. 4*D*). The data from the second PLS constraint identified transcripts that co-responded with an increase in $CH_2\mathbf{CH_2}CH_2$, $\mathbf{CH_3}CH_2$, and **CH═CH** lipid moieties and a decrease in phosphatidylcholine (Table 2). The positive and negative covariant transcripts were substantially overlapped with those for the water-soluble metabolites. This would seem to indicate common inductive mechanisms.

*Correlation of transcripts to metabolic changes using bootstrapping.* To further investigate the regression models of transcriptional and metabolic changes and test their robustness,

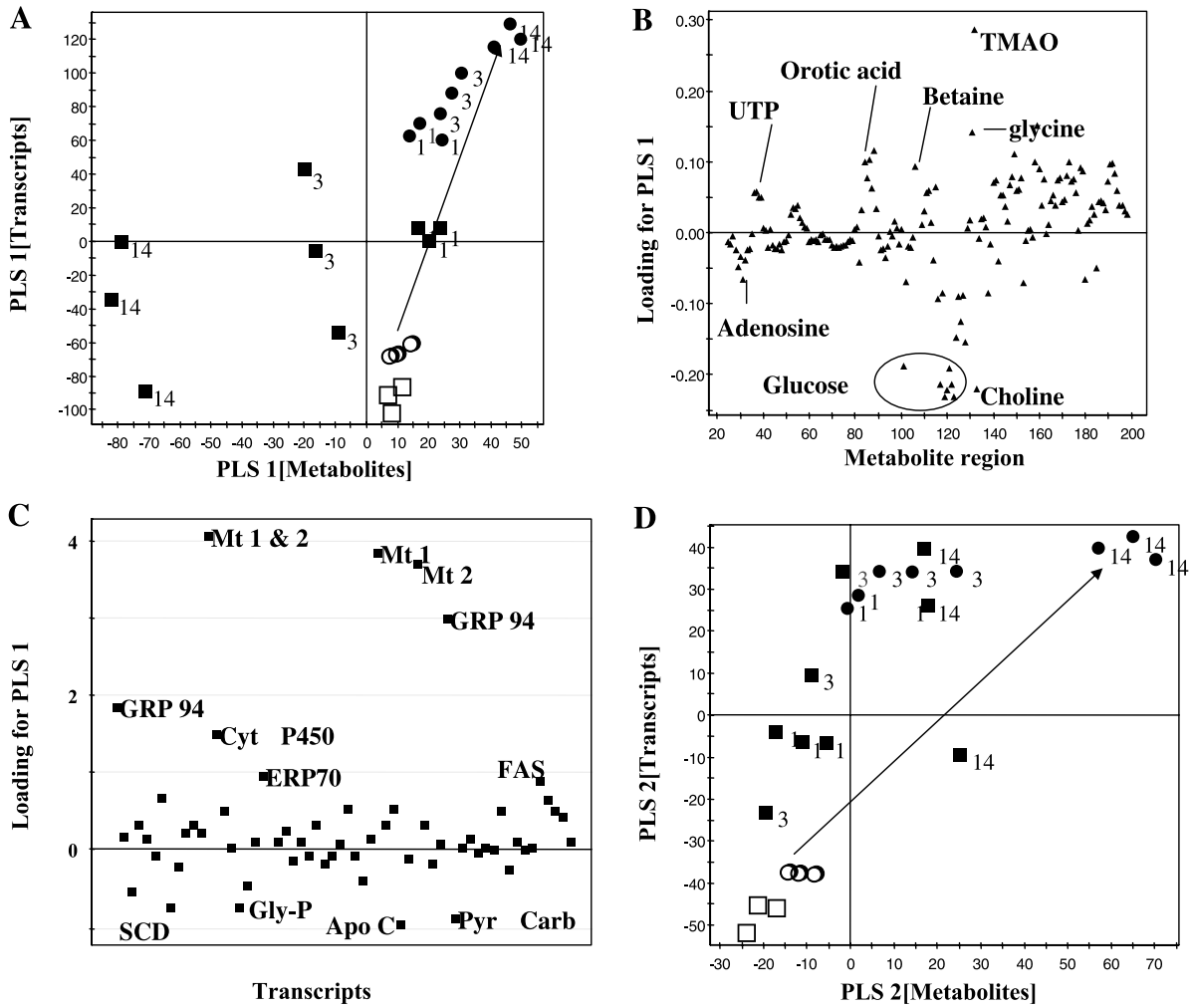


Fig. 4. Combined pattern recognition analyses of gene transcripts and metabolic changes. *A*: partial least squares (PLS) regression was used to correlate gene transcripts using the 60-transcript data set identified by the volcano plot filter against metabolic profile. Loadings plot displays regressed variance of Kyoto and Wistar rats exposed to orotic acid for 1, 3, and 14 days or control animals. A linear response across gene and metabolism changes from aqueous extracts was detected for Kyoto strain of rats, but the effect was less clear in Wistar rats. Metabolites (*B*) and genes (*C*) correlated by this regression were identified and referred to as significant if having a contribution of at least 10% of the maximum contribution for a gene or metabolite. Variance is plotted along the *y*-axis. The horizontal line symbolizes the point of no contribution. *D*: second PLS trend in the model of 60 gene transcripts and the intact liver metabolic data set. Open symbols: Kyoto rats (○) and Wistar rats (□) fed normal diet. Solid symbols: Kyoto (●) and Wistar rats (■) fed diet with orotic acid (subscript numbers on data points signify days of exposure). Mt, metallothionein; Cyt P450, cytochrome *P*-450; GRP 94, glucose regulated protein, 94 kDa; SCD, stearoyl-CoA desaturase 1; Gly-P, glycogen phosphorylase; ApoC, apolipoprotein C III; Pyr Carb, pyruvate carboxylase; ERP70, protein disulfide isomerase related protein; FAS, fatty acid synthetase.

Table 2. *PLS and PCA analysis of gene and metabolite changes for Wistar and Kyoto rats fed orotic acid for 1, 3, or 14 days derived from the pattern recognition models described*

| Pathway | Correlated Genes | Correlated Metabolites |
|---|---|---|
| Metabolism of choline production of uridine nucleotides | Increased: | Increased: |
| | Metallothionein-1 | Phosphocholine |
| | Metallothionein-2 | Glycine |
| | GRP94/endoplasmin | TMAO |
| | Cytochrome *P*-450 (Cyp2b19) | Betaine |
| | Fatty acid synthetase | UTP |
| | PDI-related protein ERP70 | |
| | Decreased: | Decreased: |
| | Stearyl-CoA desaturase | Glucose |
| | Pyruvate carboxylase | Glycogen |
| | ApoCIII | |
| Accumulation of lipids | Increased: | Increased: |
| | GRP94/endoplasmin | $CH_2CH_2CH_2$ |
| | Fibronectin | CH═CH |
| | PDI-related protein ERP70 | $CH_3CH_2$ |
| | Oxygen regulated protein | |
| | Glycerol 3-phosphate acyltransferase | |
| | ApoB | |
| | Fatty acid synthetase | |
| | Decreased: | |
| | Hydroxysteroid sulfotransferase | |
| | Stearyl-CoA desaturase | |
| | Weakly similar to 20 α-hydroxysteroid sulfoxide | |

Metabolites or transcribed genes were identified as being significantly correlated if their loadings contributions were greater than 10% of the metabolite or gene with the highest loadings contribution. TMAO, trimethylamine oxide; PLS, partial least squares; PCA, principal componant analysis. Supplementary data is available on request to the authors, with Excel spread sheets representing all the transcriptional and metabolic changes.

correlations between transcripts and key metabolic changes were investigated using a boot strapping routine (Matlab, Mathworks). Correlation coefficients were calculated for all 60 transcripts and the concentration of TMAO, the metabolite that demonstrated by far the largest change of all in the aqueous extracts, using bootstrapping with 10,000 iterations. This sampling with replacement routine allowed quantification of the reproducibility of these correlation coefficients for the transcripts.

The most positively correlated transcript with TMAO was GRP94, while the most negatively correlated was SCD 1 (Fig. 5). SCD 1 was also negatively correlated with $CH_2\mathbf{CH_2}CH_2$ and C**H**═C**H** lipids, mitochondrial very-long-chain acyl CoA thioesterase, and GPAT transcripts, and positively correlated with apoCIII. Accordingly, a negative correlation was observed between apoCIII and GPAT, which was positively correlated with fatty acid CoA ligase 4.

## DISCUSSION

In this study, we have defined the pathways perturbed by orotic acid feeding and integrated metabolic changes with gene expressions using multivariant approaches. A novel and unexplained finding is the marked accumulation of unsaturated fatty acids in both strains of rat in the face of decreased transcription of SCD 1.

The metabolic phenotypes derived by NMR spectroscopic analysis were used to define "normality" in the outbred genetically heterogeneous Wistar rat relative to the inbred Kyoto rat. The NMR hepatic lipid profile of Kyoto rats relative to Wistar rats prior to orotic acid exposure suggested that the strain might be predisposed to fatty liver. These differences were borne out in the dynamic changes induced by orotic acid in Kyoto rats, with a much more marked increase in mobile lipids. Thus the increased susceptibility of Kyoto rats to fatty liver disease becomes overt following the challenge imposed by orotic acid. The reduced levels of glucose and glycogen in the livers of the Kyoto rats also suggest that this strain may be susceptible to insulin resistance, sharing some of the genetic features of the insulin-resistant Kyoto-derived spontaneously hypertensive rat (1). This reverse functional genomic approach using NMR spectra may have broad implications for the identification of genetic predispositions to other pathological states in pharmacogenomics.

In light of the observation that adenosine supplementation in conjunction with orotic acid prevents and reverses fatty liver (19), a crucial finding is the increased NMR signal for phosphatidylcholine, the methyl donor betaine, and the choline oxidation product TMAO. Phosphatidylcholine can be produced by CTP- and ATP-dependent pathways. The ATP-dependent pathway involves the conversion of phosphatidylserine to phosphatidylethanolamine and in turn to phosphatidylcholine and may proceed through the activation of L-methionine with ATP via the activity of L-methionine adenosyltransferase. Methionine for this process is regenerated from homocysteine by methyl donation from betaine, suggesting a link between the changes detected in methyl donors and phospholipid fatty acid metabolism of rats fed orotic acid. These changes may also link into the reduced secretion of hepatic apoB-containing lipoproteins from these animals, since a previous study has reported that the overexpression of betaine homocysteine *S*-methyltransferase, which catalyzes the donation of methyl groups from betaine to homocysteine, enhances the secretion of apoB-containing lipoproteins from the liver cell line, McArdle RH-7777 (29). Conversely, the secretion of VLDL triglyceride is reduced in mice deficient for phosphatidylethanolamine *N*-methyltransferase (22). Furthermore, we also note betaine and TMAO are oxidation products of choline, and choline deficiency induces fatty liver disease through deficiency of methyl donors (6, 28). In addition, methionine adenosyltransferase 1A knockout mice are more susceptible than wild-type animals to choline-deficient diet-induced fatty liver disease and express many acute phase-response and inflammatory markers, including metallothionein and growth-related genes (18) as seen in this study.

Several genes in the lipogenic pathway are regulated by sterol regulatory element binding protein-1c (SREBP-1c) including FAS, GPAT, and SCD 1. The decrease in the expression of SCD 1 by orotic acid suggests decreased transcriptional activation by SREBP-1c and is also consistent with previous reports demonstrating reduced de novo fatty acid synthesis in rats fed orotic acid (30, 31). SREBP-1c-mediated downregulation of SCD 1 and of lipogenic flux could relate to the large increase in the concentration of 16:1 and especially 18:1 fatty acids in the liver following orotic acid feeding (16). Unsaturated fatty acids such as these have been shown to decrease SREBP-1c mRNA levels by two mechanisms. First, liver

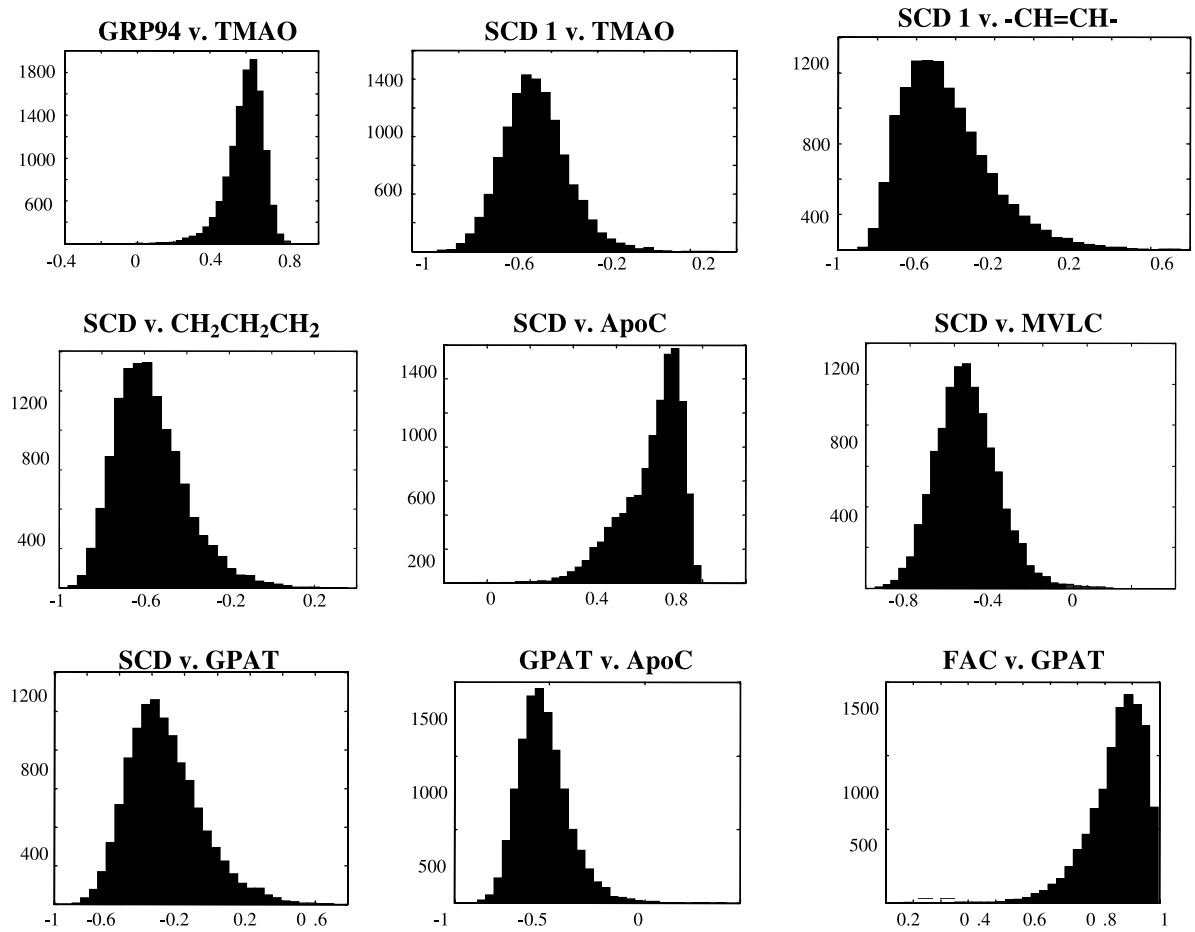


Fig. 5. Histograms of bootstraps for correlation coefficients between key metabolite regions and transcripts. The *x*-axis represents the correlation coefficients, while the *y*-axis represents the number of times this correlation was returned during 10,000 iterations. MVLC, mitochondrial very-long-chain acyl CoA thioesterase; GPAT, glycerol-3-phosphate acyltransferase; FAC, fatty acid CoA ligase 4; **CH═CH**, unsaturated lipid resonance; $CH_2\mathbf{CH_2}CH_2$, saturated lipid resonance.

X-receptor-α (LXRα) is a transcriptional regulator of SREBP-1c (3, 16. 23), and activation of LXRα by its natural ligands is antagonized by unsaturated fatty acids (32). Second, unsaturated fatty acids accelerate SREBP-1c transcript decay (32). Although these SREBP-1c-related mechanisms would explain the decreased SCD 1 expression and decreased lipogenic flux following orotic acid feeding, our data indicate that the expression of other SREBP-1c target genes such as FAS and GPAT are not decreased under these conditions. This is a paradoxical finding that suggests an orotic acid-mediated uncoupling of the normal coordination of the expression of genes involved in regulation of the lipogenic pathway.

An early and dramatic response to orotic acid feeding is the increased transcription of chaperone and stress response genes and of genes involved in growth and inflammation control. GRP94 and the PDI-related protein are endoplasmic reticulum (ER)-resident proteins involved in the folding of apoB during the assembly of the triglyceride-rich lipoproteins in VLDL (21, 33). Surprisingly, the apoB gene is consistently induced, perhaps as a response to the accumulation of lipid in ER vesicles with orotic acid feeding. The metallothioneins and fibronectin are important early response general stress genes, which respond to inflammation and oxidative stress (4, 7, 8). These changes are accompanied by decreased activity of growth control genes such as TGFβ inducible early growth response, early growth response 1, cyclin D3 and inflammation regulating genes like B lymphocyte chemoattractant precursor (weakly similar), and p38 mitogen activated protein kinase.

One caveat with interpreting the potential pathophysiological implications of the data generated is that the NMR and microarray analyses were done on whole liver extracts or whole tissue samples. Thus it was not possible to determine which cell types, for example, hepatocytes, macrophages, or stellate cells, were being influenced directly. Previous histopathological studies indicate that orotic acid largely acts directly on hepatocytes (19, 20, 25). However, improvements both in NMR spectroscopy miniaturization and amplification techniques for mRNA transcripts suggest that in the future such approaches may be coupled to laser-capture dissection of specific cell types.

The reverse functional genomics strategy described in the present paper demonstrates that metabolic phenotypes can be correlated with transcriptional data from DNA microarray experiments and used to define critical metabolic pathways, and importantly this approach can accommodate information from different animal strains across a time series. Although in the present study we have not evaluated posttranslational regulation, allosteric controls, or flux through pathways, we

have been able to obtain important new information on the orotic acid-induced fatty liver phenotype through the integration of transcriptional and metabolic changes. Multivariate analysis of the combined transcription and NMR spectroscopy data sets were well suited to the large matrices produced and could be used as a strategy for investigating toxicological insults and the implications of pharmacogenomic predisposition to therapeutic interventions.


## ACKNOWLEDGMENTS

We acknowledge the expert technical assistance from Alastair Edwards and Dr. Ned Mason, and we acknowledge the guidance from Hetal Patel, Drs. Helen Fields, and Helen Causton on the microarray experiments.

Present address of C. Mann: Genetix, Queensway, New Milton, Hampshire, BH25 5NN, UK.

## GRANTS

J. L. Griffin is grateful for the support of the Royal Society, UK. We acknowledge the British Medical Research Council, Imperial College London, and the British Heart Foundation for the funds for establishing the microarray facility.


## REFERENCES


1. **Aitman TJ, Glazier AM, Wallace CA, Cooper LD, Norsworthy PJ, Wahid FN, Al-Majali KM, Trembling PM, Mann CJ, Shoulders CC, Graf D, St Lezin E, Kurtz TW, Kren V, Pravenec M, Ibrahimi A, Abumrad NA, Stanton LW, and Scott J.** Identification of Cd36 (Fat) as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive rats. *Nat Genet* 21: 76–83, 1999.
2. **Bollard ME, Garrod S, Holmes E, Lindon JC, Humpfer E, Spraul M, and Nicholson JK.** High-resolution (1)H and (1)H-(13)C magic angle spinning NMR spectroscopy of rat liver. *Magn Reson Med* 44: 201–207, 2000.
3. **Brown MS and Goldstein JL.** The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell* 89: 331–340, 1997.
4. **Capdevila JH, Falck JR, and Harris RC.** Cytochrome P450 and arachidonic acid bioactivation. Molecular and functional properties of the arachidonate monooxygenase. *J Lipid Res* 41: 163–181, 2000.
5. **Chu S, DeRisl J, Eisen M, Mulholland J, Botstein D, Brown PO, and Herskowitz I.** The transcriptional program of sporulation in budding yeast. *Science* 282: 699–705, 1998.
6. **Clarke SD.** Nonalcoholic steatosis and steatohepatitis. I. Molecular mechanism for polyunsaturated fatty acid regulation of gene transcription. *Am J Physiol Gastrointest Liver Physiol* 281: G865–G869, 2001.
7. **Clement B, Grimaud JA, Campion JP, Deugnier Y, and Guillouzo A.** Cell types involved in collagen and fibronectin production in normal and fibrotic human liver. *Hepatology* 6: 225–234, 1986.
8. **Diehl AM.** Cytokine regulation of liver injury and repair. *Immunol Rev* 174: 160–171, 2000.
9. **Eriksson L, Johansson E, Kettaneh-Wold N, and Wold S.** *Introduction to Multi- and Megavariate Data Analysis using Projection Methods (PCA and PLS)*. Umea, Sweden: Umetrics, 1999.
10. **Fiehn O.** Metabolomics: the link between genotypes and phenotypes. *Plant Mol Biol* 48: 155–171, 2002.
11. **Fielding BA, Callow J, Owen RM, Samra JS, Matthews DR, and Frayn KN.** Postprandial lipemia: the origin of an early peak studied by specific dietary fatty acid intake during sequential meals. *Am J Clin Nutr* 63: 36–41, 1996.
12. **Gavaghan CL, Holmes E, Lenz E, Wilson ID, and Nicholson JK.** An NMR-based metabonomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk:ApfCD mouse. *FEBS Lett* 484: 169–174, 2000.
13. **Griffin JL, Williams HJ, Sang E, Clarke K, Rae C, and Nicholson JK.** Metabolic profiling of genetic disorders: a multitissue (1)H nuclear magnetic resonance spectroscopic and pattern recognition study into dystrophic tissue. *Anal Biochem* 293: 16–21, 2001.
14. **Griffin JL, Mann C, Scott J, Shoulders C, and Nicholson JK.** Choline containing metabolites during cell transfection: an insight into magnetic resonance spectroscopy detectable changes. *FEBS Lett* 509: 263–266, 2001.
15. **Hebbachi AM, Seelaender MC, Baker BW, and Gibbons GF.** Decreased secretion of very-low-density lipoprotein triacylglycerol and apolipoprotein B is associated with decreased intracellular triacylglycerol lipolysis in hepatocytes derived from rats fed orotic acid or n-3 fatty acids. *Biochem J* 325: 711–719, 1997.
16. **Horton JD, Goldstein JL, and Brown MS.** SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *J Clin Invest* 109: 1125–1131, 2002.
17. **Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, and Gibson G.** The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat Genet* 29: 389–395, 2001.
18. **Lu SC, Alvarez L, Huang ZZ, Chen L, An W, Corrales FJ, Avila MA, Kanel G, and Mato JM.** Methionine adenosyltransferase 1A knockout mice are predisposed to liver injury and exhibit increased expression of genes involved in proliferation. *Proc Natl Acad Sci USA* 98: 5560–5565, 2001.
19. **Marchetti M, Puddu P, and Caldarera CM.** Metabolic aspects of "orotic acid fatty liver" nucleotide control mechanisms of lipid metabolism. *Biochem J* 92: 46–51, 1964.
20. **Martin A, Biol MC, Raisonnier A, Infante R, Louisot P, and Richard M.** Impaired glycosylation in liver microsomes of orotic acid fed rats. *Biochim Biophys Acta* 718: 85–91, 1982.
21. **Murga C, Fukuhara S, and Gutkind S.** Novel molecular mechanisms in the pathways connecting G-protein-coupled receptors to MAP kinase cascades. *Trends Endocrinol Metab* 10: 122–127, 1999.
22. **Noga AA, Zhao Y, and Vance DE.** Plasma homocysteine is regulated by phospholipid methylation. *J Biol Chem* 277: 41358–42365, 2002.
23. **Ou J, Tu H, Shan B, Luk A, DeBose-Boyd RA, Bashmakov Y, Goldstein JL, and Brown MS.** Unsaturated fatty acids inhibit transcription of the sterol regulatory element-binding protein-1c (SREBP-1c) gene by antagonizing ligand-dependent activation of the LXR. *Proc Natl Acad Sci USA* 98: 6027–6032, 2001.
24. **Raamsdonk LM, Teusink B, Broadhurst D, Zhang N, Hayes A, Walsh MC, Berden JA, Brindle KM, Kell DB, Rowland JJ, Westerhoff HV, van Dam K, and Oliver SG.** A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat Biotechnol* 19: 45–50, 2001.
25. **Roheim PS, Switzer S, Girard A, and Eder HA.** The mechanism of inhibition of lipoprotein synthesis by orotic acid. *Biochem Biophys Res Commun* 20: 416–421, 1965.
26. **Schoolnik GK.** Microarray analysis of bacterial pathogenicity. *Adv Microb Physiol* 46: 1–45, 2002.
27. **Shalon D, Smith SJ, and Brown PO.** A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res* 6: 639–645, 1996.
28. **Simon JB, Scheig R, and Klatskin G.** Hepatic ATP and triglyceride levels in choline-deficient rats with and without dietary orotic acid supplementation. *J Nutr* 98: 188–192, 1969.
29. **Sowden MP, Collins HL, Smith HC, Garrow TA, Sparks JD, and Sparks CE.** Apolipoprotein B mRNA and lipoprotein secretion are increased in McArdle RH-7777 cells by expression of betaine-homocysteine S-methyltransferase. *Biochem J* 341: 639–645, 1999.
30. **Tokmakjian SD and Haines DS.** Early effects of dietary orotic acid upon liver lipid synthesis and bile cholesterol secretion in rats. *J Lipid Res* 26: 478–486, 1985.
31. **Windmueller HG and Spaeth AE.** Perfusion in situ with tritium oxide to measure hepatic lipogenesis and lipid secretion. Normal and orotic acid-fed rats. *J Biol Chem* 241: 2891–2899, 1966.
32. **Xu J, Teran-Garcia M, Park JH, Nakamura MT, and Clarke SD.** Polyunsaturated fatty acids suppress hepatic sterol regulatory element-binding protein-1 expression by accelerating transcript decay. *J Biol Chem* 276: 9800–9807, 2001.
33. **Yamauchi M, Mizuhara Y, and Toda G.** Serum vitronectin receptor in alcoholic liver disease: correlation with fibronectin receptor and morphological features. *Alcohol Alcohol Suppl* 1A: 37–43, 1993.
34. **Zhu H and Snyder M.** "Omic" approaches for unravelling signalling networks. *Curr Opin Cell Biol* 14: 173–179, 2002.