The Journal of Experimental Biology 210, 1613-1621
Published by The Company of Biologists 2007
doi:10.1242/jeb.005454

The biological importance of measuring individual variation

Douglas L. Crawford* and Marjorie F. Oleksiak

Rosenstiel School of Marine and Atmospheric Sciences, Marine Biology and Fisheries, 4600 Rickenbacker Causeway, Miami, FL 33149, USA

*Author for correspondence (e-mail: dcrawford@rsmas.miami.edu)

Accepted 6 March 2007

Summary

Functional genomics research using *Fundulus heteroclitus* has focused on variation among individuals because of the evolutionary importance and value of *Fundulus* in explaining the human condition (why individual humans are different and are affected differently by stress, disease and drugs). Among different populations and species of *Fundulus*, there are evolutionarily adaptive differences in gene expression. This natural variation in gene expression seems to affect cardiac metabolism because up to 81% of the variation in glucose utilization observed in isolated heart ventricles is related to specific patterns of gene expression. The surprising result from this research is that among different groups of individuals, the expression of mRNA from different metabolic pathways explains substrate-specific metabolism. For example, variation in oxidative phosphorylation mRNAs explains glucose metabolism for one group of individuals but expression of glucose metabolism genes explains this metabolism in a different group of individuals. This variation among individuals has important implications for studies using inbred strains: conclusions based on one individual or one strain will not necessarily reflect a generalized conclusion for a population or species. Finally, there are surprisingly strong positive and negative correlations among metabolic genes, both within and between pathways. These data suggest that measures of mRNA expression are meaningful, yet there is a complexity in how gene expression is related to physiological processes.

Glossary available online at
http://jeb.biologists.org/cgi/content/full/210/9/1613/DC1

Supplementary material available online at
http://jeb.biologists.org/cgi/content/full/210/9/1613/DC2

Key words: functional genomics, cardiac metabolism, microarray, evolutionary selection, *Fundulus heteroclitus*, mRNA expression, gene expression, phenotypic variation.

Introduction

Microarrays measure the expression of hundreds or thousands of genes at a time, offering near global measurements at the level of the transcript. The importance of the patterns of expression revealed by this technology has been questioned (Feder and Walser, 2005). There are good reasons to wonder about the utility of quantifying the mRNA expression of specific genes, and most of these concerns have to do with the multiple biochemical and physiological steps that modulate gene expression other than at the level of the transcript (e.g. micro-RNAs, translation, protein turnover, covalent modifications of enzymes and protein–protein interactions). To better understand the validity of transcriptomic data and whether mRNA expression underpins phenotypic differences, we need to understand how gene expression varies within and among taxa, whether the variation in mRNA co-varies with a specific phenotypic character or biological process and whether these patterns of mRNA expression are adaptively significant. All three are necessary for defining the biological importance of gene expression because they each provide separate information to confirm or reject the importance of gene expression.

If mRNA expression regulates phenotypic variation, then we would expect the variation in gene expression to reflect the differences among populations and species. For example, one would expect that outbred, highly polymorphic species would have greater variance than inbred species. One would expect greater variance among taxa as the genetic distance is increased. This variation should also correlate with appropriate phenotypic variation. That is, with an increase in physiological performance, one would expect a change in the mRNA that affects this performance. Yet, and this is the crux of the problem, does the lack of correlation between mRNA and a physiological measure arise because one does not have the correct gene or because there

are epistatic interactions, both of which make the patterns difficult to discern? Finally, if the pattern of mRNA variation is to be shown as adaptive, then it must affect a phenotype that is selectively (biologically) important.

Results

Evolutionary variance in gene expression

Much progress has been made in understanding the variation in gene expression using microarrays in yeast, worms, fish, mice and humans (Cavalieri et al., 2000; Cheung et al., 2003a; Denver et al., 2005; Gibson and Weir, 2005; Jin et al., 2001; Oleksiak et al., 2002; Oleksiak et al., 2005; Pritchard et al., 2001; Schadt et al., 2003; Townsend et al., 2003; Whitehead and Crawford, 2005; Whitehead and Crawford, 2006a; Whitehead and Crawford, 2006b). Evolutionary analyses indicate that stabilizing selection affects much of gene expression (Denver et al., 2005; Lemos et al., 2005). Stabilizing selection distinguishes between individuals with phenotypes closer to the mean *versus* those that deviate from the mean and selects against individuals that deviate. Thus, the observation that stabilizing selection affects a majority of mRNA indicates that small changes in gene expression have biological effects.

Although many genes have significant differences in expression among individual *F. heteroclitus* within a population, the magnitude of these differences is small, typically less than 1.5-fold (Oleksiak et al., 2002; Oleksiak et al., 2005; Whitehead and Crawford, 2005; Whitehead and Crawford, 2006a; Whitehead and Crawford, 2006b). For *Fundulus*, it is the many small changes in metabolic gene expression that together appear to be responsible for the phenotypic variation in cardiac metabolism (Oleksiak et al., 2005). Additionally, although stabilizing selection eliminates many mutations affecting gene expression because this variation is slightly deleterious, there is significant additive genetic variation affecting gene expression that can be the source of adaptive change. In *Drosophila simulans* (Wayne et al., 2004), the additive heritable variation is distributed among chromosomes, with much of the effect acting in *trans* (Wayne et al., 2004). Among 14 large human families, approximately 29% of genes have significant additive variation due to both *cis* and *trans* loci (Morley et al., 2004). Among mice, 19% of all loci have significant additive variation (Cui et al., 2006). Among *Drosophila* (Nuzhdin et al., 2004) and primates (Caceres, 2003; Enard, 2002; Gilad et al., 2005; Gilad et al., 2006; Khaitovich, 2004a; Khaitovich, 2004b) much of this additive genetic variation for expression is evolving by natural selection.

In *Fundulus*, by applying evolutionary analyses to natural populations that have experienced the effects of selection, we were able to document patterns of expression affected by directional, stabilizing and balancing selection (Crawford et al., 1999a; Crawford et al., 1999b; Oleksiak et al., 2002; Pierce and Crawford, 1997a; Whitehead and Crawford, 2006a; Whitehead and Crawford, 2006b). Our data on *Fundulus*, as well as other investigators' data on *Caenorhabditis elegans* (Denver et al., 2005), *Drosophila* (Lemos et al., 2005; Nuzhdin et al., 2004) and humans (Caceres, 2003; Enard, 2002; Gilad et al., 2006; Gilad et al., 2005; Khaitovich, 2004a; Khaitovich, 2004b), suggest that the variation in gene expression is selectively important; thus, this variation is biologically important. This is only one, albeit an important, criterion for establishing the importance of gene expression. It is also important to define the heritability of gene expression and how it relates to important phenotypic differences.

Genetics of gene expression

Much of gene expression measured by microarrays is genetic; it differs between inbred lines, is associated with quantitative trait loci (QTLs) and has narrow sense heritability (h^2) greater than 30% [narrow sense heritability is due only to the additive genetic variation (V_a), or $h^2=V_a/V_p$, where V_p is phenotypic variation] (Cheung et al., 2003b; Gibson and Weir, 2005; Sharma et al., 2005; Tan et al., 2005). Among F1 generations from two inbred mice strains, approximately two-thirds of all loci have measurable h^2 with a quarter having an h^2 of >50% (Cui et al., 2006). In both humans and mice, the median h^2 is 34% among loci with measurable h^2 (Cui et al., 2006). The variation in regulatory processes affecting gene expression has been inferred by combining microarray and QTL studies [expressed QTLs (eQTL)]. These studies identify both *cis*- and *trans*-acting loci that are related to differences in gene expression in *Drosophila* (Wang et al., 2004; Wayne and McIntyre, 2002), yeast (Brem and Kruglyak, 2005; Brem et al., 2005; Ronald et al., 2005; Yvert et al., 2003), mice (Chesler et al., 2005; Doss et al., 2005; Ghazalpour et al., 2005; Schadt et al., 2003) and humans (Monks et al., 2004; Morley et al., 2004; Schadt et al., 2005). In general, 20–30% of differential expressions are due to a *cis*-eQTL (Doss et al., 2005; Ronald et al., 2005). Yet with more powerful analyses, gene expression becomes more complex, involving many loci with a few loci that affect the expression of many genes (Brem and Kruglyak, 2005; Brem et al., 2005; Gibson and Weir, 2005; Schadt et al., 2005; Stamatoyannopoulos, 2004). These data suggest a complex regulation of gene expression in which polymorphisms among several loci affect the variation in gene expression of a particular gene. It is important to realize that heritability in gene expression indicates that gene expression is stable between generations. This stability suggests that random biological variation or 'noise' is not the principal cause of variation in gene expression.

Although much progress has been made in understanding the variation in gene expression, we are unsure of its importance in affecting phenotypic variation. It is the phenotypic variation on which natural selection acts that defines human populations and humans' susceptibility to disease, drugs and stress, and thus is of scientific importance. To understand the importance of gene expression and its effect on phenotypic variation, more attention needs to be paid to the variation among individuals and whether there is variation in which genes are 'important' in effecting a change in phenotype. We present here a summary of microarray data from *Fundulus* that supports our contention that mRNA expression affects physiological performance and is thus evolutionarily important.

Cardiac performance

Cardiac metabolism in *Fundulus* is measured using isolated heart ventricles (Oleksiak et al., 2005; Podrabsky et al., 2000), and individual determinations were made by alternating between populations. Metabolic rates are measured in triplicate by determining oxygen consumption in a well-mixed chamber for each of three different substrates: 5 mmol l^{-1} glucose, 1 mmol l^{-1} fatty acid (FA; palmitic acid–bovine serum albumin) or LKA [5 mmol l^{-1} lactate, ketones (5 mmol l^{-1} each hydroxybutyrate and acetoacetate) and 0.1% ethanol]. Two inhibitors of glucose metabolism (20 mmol l^{-1} 2-deoxy-glucose and 10 mmol l^{-1} iodoacetic acid) were used when measuring FA and LKA metabolism. The addition of inhibitors reduces metabolism to less than 15% of the rate with glucose only. Adding FA or LKA caused a significant increase in metabolism compared with metabolic rates of hearts with inhibitors and glucose, indicating that much of glucose metabolism was inhibited when using both inhibitors. Although determination of metabolism for all three substrates took approximately 15 min, isolated heart ventricles from *Fundulus* are able to maintain stable metabolic rates for greater than 45 min. All metabolic rates from isolated heart ventricles were a function of body mass, which is highly correlated (r=0.7–0.85) with heart mass (see below). These data indicate that our measures of metabolism are substrate dependent.

Individual ventricles were splayed open for all metabolic measures. *Fundulus* hearts, like those of most small fish, lack coronary circulation, and thus oxygen is supplied by diffusion from the internal blood flow (Farrell, 1993). Dissecting or splaying open the heart provided greater access and more uniform interfaces for the Ringer solution and the internal surfaces of the heart.

When measuring oxygen consumption in isolated ventricles, we were primarily concerned with the variation between individuals that can arise from genetic, developmental or physiological mechanisms (e.g. due to acclimation to different temperatures). All individuals were acclimated to a common laboratory environment (temperature, salinity, light:dark cycle, feeding regime, etc.) and were assayed at similar times of day. Despite minimizing physiological sources of variation, there can be other biological differences. For example, a fish could be sick or stressed due to social interactions or 'unhappy' for unknown reasons. These sources of variation are difficult to ascertain, and their importance remains unknown. Technical variation can arise due to poor heart preparation, electrical interference with the oxygen electrode, poor mixing of Ringer solution or poor maintenance of the oxygen electrode. Multiple determinations provide some measure of consistency and thus estimates technical variance. However, because a single heart cannot be measured on separate days, the replicate measures are dependent on the status of the electrode, heart preparation and electronic noise during the 15–20 min measurement period. One can measure 'day effects' among individuals, and this was not significant (P>0.5). That is, among the 3 days that *F. heteroclitus* cardiac ventricles were measured, hearts had approximately the same mean. However, because there is much variation among individuals it would be difficult to detect a day effect. Thus, the lack of day effect does not eliminate this as a source of variation but only makes it less probable.

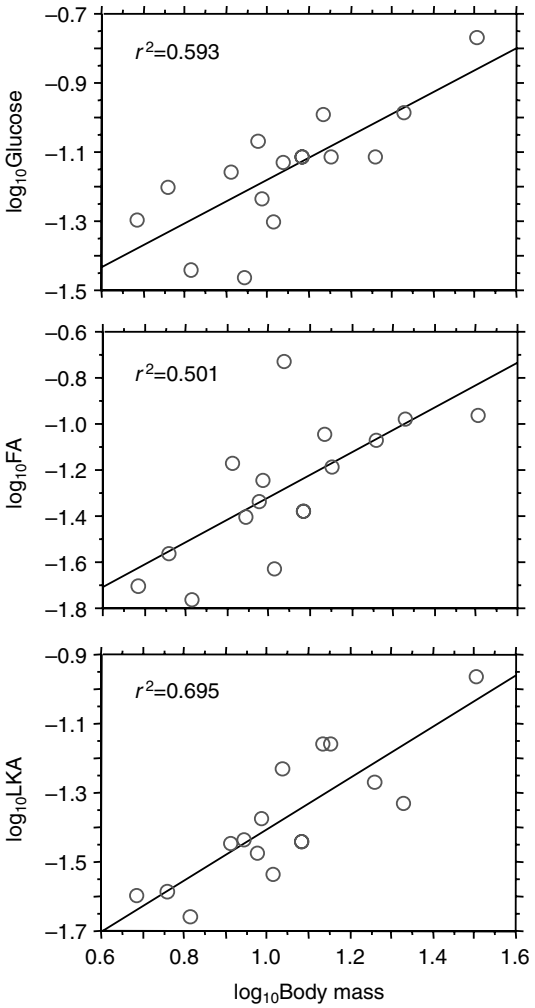

Fig. 1. Metabolism and body mass. $\log_{10}$ of substrate metabolism was regressed against $\log_{10}$ body mass. All regressions are significant (P<0.005). FA, fatty acid; LKA, lactate, ketones and alcohol.

Among 16 hearts measured (eight from a southern, Georgia population and eight from a northern, Maine population), more oxygen was consumed per time, with an average r^2 (explained variance) of 0.94, 0.86 and 0.84, for glucose, fatty acid or LKA, respectively. That is, there was little variance within a specific measure of oxygen consumption. Among the triplicate measures of oxygen consumption for each substrate, the coefficient of variation (CV; % standard deviation relative to the mean) was 12%, 31% or 21% for glucose, FA and LKA, respectively. This means that 95% of all measures of an individual heart will fall within 60% of the mean.

Among the 16 individuals, there were 4.9-, 15.9- and 4.8-fold differences between the highest and lowest metabolic rates for glucose, FA and LKA, respectively. These individuals differed in body mass, and body mass affected all three measures of metabolism (Fig. 1). Thus, this range of metabolism was due both to differences in body mass and to

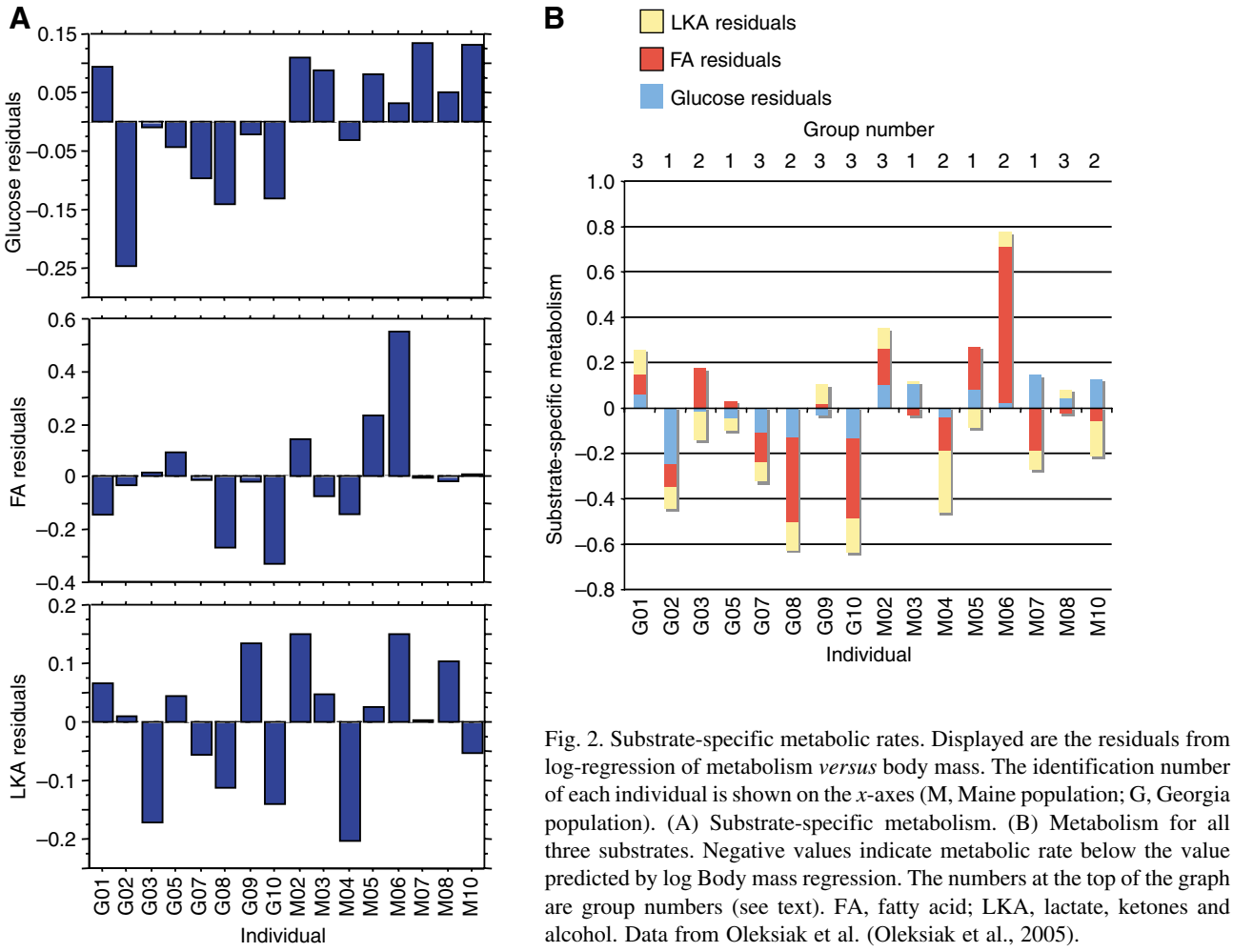

Fig. 2. Substrate-specific metabolic rates. Displayed are the residuals from log-regression of metabolism *versus* body mass. The identification number of each individual is shown on the *x*-axes (M, Maine population; G, Georgia population). (A) Substrate-specific metabolism. (B) Metabolism for all three substrates. Negative values indicate metabolic rate below the value predicted by log Body mass regression. The numbers at the top of the graph are group numbers (see text). FA, fatty acid; LKA, lactate, ketones and alcohol. Data from Oleksiak et al. (Oleksiak et al., 2005).

individual variation. The effect of body mass was removed by using the residual from the log regression, resulting in 2.0-, 11.2- and 2.5-fold differences between the highest and lowest metabolic rates for glucose, FA and LKA, respectively. With the caveats described above concerning technical variation, replicate measures enabled the statistical testing of inter-individual differences in substrate-specific metabolism among 16 individuals: $P<0.0001$, $P<0.005$ and $P<0.02$ for glucose-, FA- and LKA-dependent metabolism, respectively (Fig. 2). These data suggest that there was significant variation in metabolism among individuals. These differences also exist if one compares just the eight individuals within the Maine population ($P<0.05$ for all three measures) and glucose-specific metabolism in the Georgia population ($P<0.005$). Fig. 2 illustrates these differences, but more importantly is the variation in the relative use of each substrate within an individual (Fig. 2B). For most individuals, metabolism fuelled by FA was greater than with either glucose or LKA (Fig. 2). For example, for individual ME06, FA metabolism was more than 10-fold greater than the other two substrates. Yet, in other individuals (ME10 and ME03), glucose metabolism was greater than FA metabolism. These measures of substrate-specific metabolism within an individual are unlikely to be due to technical variation because they are measured within a short time (less than 20 min) on the same heart preparation.

Fundulus gene expression and metabolism

For *F. heteroclitus*, we have demonstrated that gene expression explained the variation in cardiac metabolism among 16 male individuals from natural outbred populations raised in a common environment (Fig. 3) (r^2=0.65–0.81) (Oleksiak et al., 2005). But the relationship was complex, a vast majority of genes were different between individuals and these expression patterns cluster the 16 individuals into three groups. It is within these three groups that one can show an association between gene expression and metabolic rates.

For mRNA expression, 94% of the genes had expression levels that were significantly different among individuals within a population ($P<0.01$) (Oleksiak et al., 2005). Using a very conservative multiple correction (Bonferroni's or F_{max} permutations), 84% were significantly different. These differences were not due to one or a few individuals: all possible permutations of six out of eight individuals within each population had on average 78.9% of genes that were significantly different among individuals ($P<0.01$). This high frequency of differences was among male individuals raised in

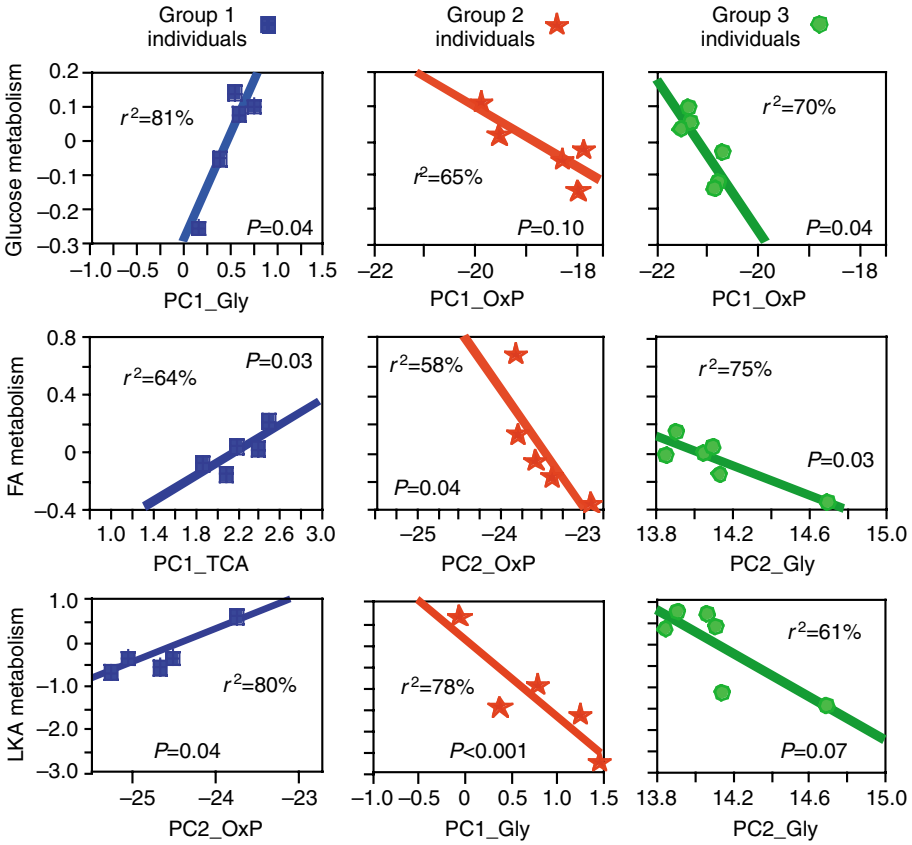

Fig. 3. Gene expression and cardiac metabolism. Principal components (PC) summarize metabolic pathway gene expression. The nine graphs illustrate which metabolic pathway explains the most variation in substrate-specific metabolism for the three groups. FA, fatty acid; LKA, lactate, ketones and alcohol; Gly, glycolysis; OxP, oxidative phosphorylation; TCA, tricarboxylic acid. Modified from Oleksiak et al. (Oleksiak et al., 2005).

a common environment and was unrelated to body mass or any other obvious physiological or experimental condition (Oleksiak et al., 2005). What is apparent is that the three groups of individuals share different patterns of expression (Oleksiak et al., 2005).

Among the 16 individuals examined, shared gene expression patterns cluster the 16 individuals into three groups of 5, 5 and 6 individuals. However, there was no difference in the mean or variation in substrate-specific metabolism among these three groups ($P>0.4$ for glucose and FA metabolism, and $P>0.05$ for LKA metabolisms) (Fig. 2). Yet, gene expression differences among the three groups were statistically robust. Differences in gene expression within each of the three groups were less than those found within a single population (62–73% *vs* 94% of genes are statistically significant) and were significantly less ($P<0.01$) than those found in the 2052 random combinations of five individuals (average 85%). The number of genes that were significantly different among the three groups (50 genes) was considerably more than the 12 genes that were different between populations and significantly greater than the average of one significant difference found among three random groups formed by 1000 random permutations. These data indicate that the three groups were functionally distinctive and the differences were robust.

To explore the relationship between metabolism and gene expression within the three groups, the variation in gene expression was reduced to the principal components (PCs) for glycolytic, tricarboxylic acid (TCA) and oxidative-phosphorylation metabolic pathways (Table 1) (see Table S1 in supplementary material for a list of genes). A PC is a linear equation that sums the measures of gene expression where each gene is multiplied by a coefficient. The weight of the coefficient is chosen to maximize the explained variation among individuals without, in this case, reference to metabolism. For example, the first PC for glucose-related enzymes explains 54% of the variation among these 15 genes. Additionally, there were both strong positive and negative weighting factors [e.g. *glyceraldehyde 3-phosphate dehydrogenase* (*GAPDH*) expression is multiplied by –0.63, and *aldehyde dehydrogenase* (*ALDH*) by 0.37] (Table 1).

These PCs summarize the metabolic pathway-dependent RNA expression and statistically explain significant proportions of the variation in cardiac metabolism, but only within the three defined groups of fish (Fig. 3). For glucose metabolism, 81% of the variation in group 1 individuals was explained by changes in expression of genes involved in glucose. However, in groups 2 and 3, these glycolytic genes had little power to explain the differences among individual metabolic rates. Instead, genes of the oxidative-phosphorylation pathway explained the variation in glucose-specific metabolic rates. Similarly, gene expression in different pathways explained FA- and LKA-specific metabolism. These patterns, where gene expression from a pathway explains substrate-specific metabolism, only occur if one examines these groups of individuals. Permutation analyses indicate that few other random sets of five individuals share a common

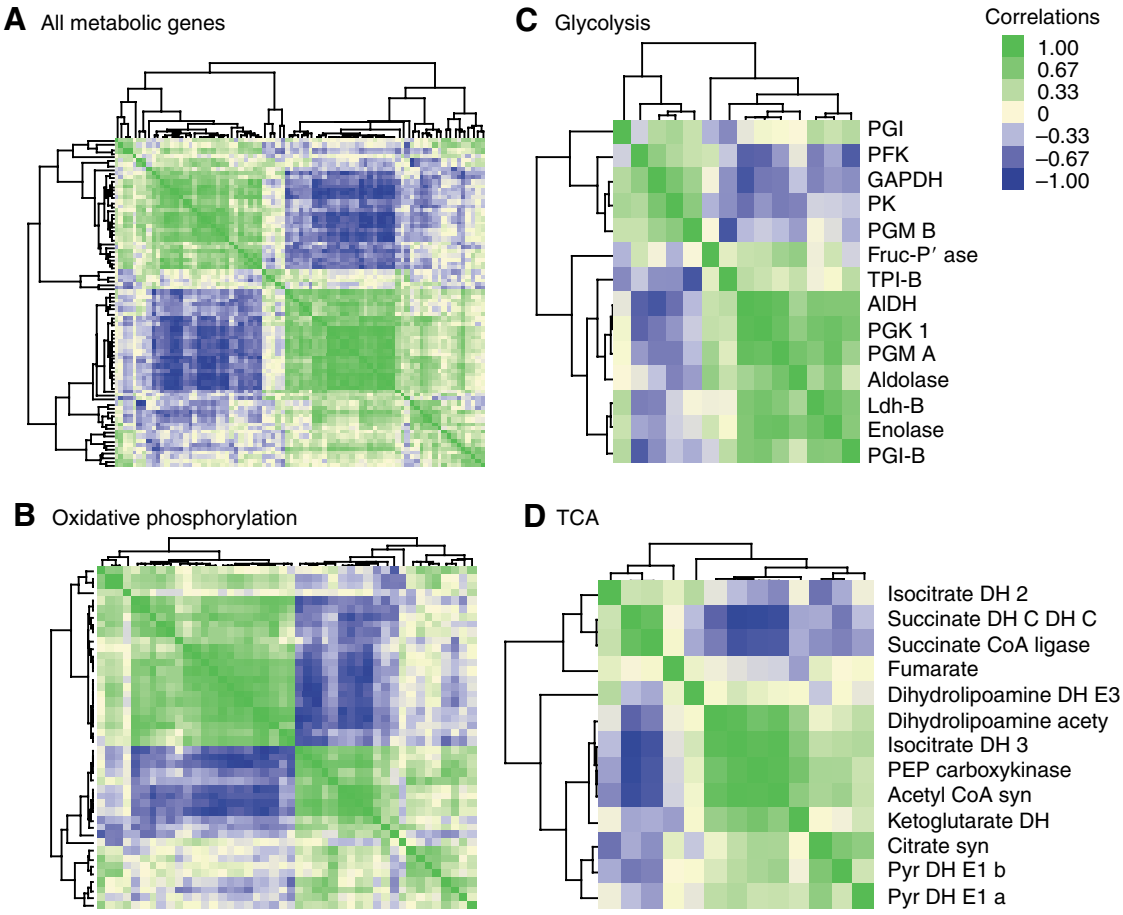

Fig. 4. Metabolic genes correlate expression. Correlations between the expressions of two enzymes; green boxes are significant positive correlations, blue is negative, and yellow is a lack of correlation. Correlation among (A) all metabolic genes; (B) oxidative phosphorylation genes; (C) glycolytic genes and (D) TCA genes. These are symmetric matrices, thus the order of genes across the top is the same as on the left side. Gene names are not given in A and B because of size constraints.

relationship between gene expression and phenotype. There is no relationship of one or more genes to metabolism among all 16 individuals. Nor do any of the PCs explain metabolism among all 16 individuals.

What these data suggest is that the genes that are important, which explain the variation in cardiac metabolism, differ among individuals. So, for example, altering glycolysis can affect glucose utilization but it will not do so in all individuals.

Correlated patterns of gene expression

Among the PCs, some genes have greater importance, as indicated by their larger coefficients. These coefficients are both negative and positive (Table 1). These coefficients reflect the strong correlation in expression among metabolic genes within and between pathways. In Fig. 4, the correlations between two genes among the 16 individuals are displayed as green and blue boxes in an all-against-all matrix. A significant positive correlation (a green cell) indicates that the expression for these two genes would be similar among all individuals (i.e. both would be high or low in the same individuals). Surprisingly, the expressions of most metabolic genes (Fig. 4) are either negatively or positively correlated. For example, an individual with a high expression of *phosphofructokinase* (*PFK*) also has high expression of *GAPDH* but low expression for *phosphoglucoisomerase B* (*PGI-B*) (Fig. 4C). These correlation patterns repeat themselves in each of the three pathways (Fig. 4B–D). In the linear pathway of glycolysis, an increase in *aldolase*, *Ldh-B* and enolase among 16 individuals is matched with a decrease in *PFK*, *GAPDH* and *pyruvate kinase*. These enzymes are interspersed along the pathway. What is more difficult to see are the correlations among proteins that form enzyme complexes or pathways. In the first enzyme complex of oxidative phosphorylation, NADH dehydrogenase, the subunits have significant positive or negative correlations for expression with each other (correlations are significant for 11 of the 20 NADH subunits). This pattern is repeated for complex 4 (cytochrome *c* oxidase) and complex 5 (ATP synthetase). These subunits have to form at a stoichiometric ratio for each enzyme, yet they have negative correlations for expression! Although these opposite patterns of gene expression present a biochemical conundrum, an obvious molecular reason is that many of these enzymes share common transcription factors.

Table 1. *Principal components for the three major metabolic pathways*

Glycolysis/gluconeogenesis genes (PC1, 54.12%; PC2, 15.55%)															
Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
E.C. #	1.1.1.27	1.2.1.12	1.2.1.3	2.4.1.1	2.7.1.11	2.7.1.40	2.7.2.3	3.1.3.11	4.1.2.13	4.2.1.11	5.3.1.1	5.3.1.9	5.3.1.9	5.4.2.1	5.4.2.2
Abbrev.	LDHB	GAPD	ALDH9A1	PYGM	PFKM	PKM2	PGK1	FBP1	ALDOA	ENO1	TPI1	GPI	GPI	PGAM1	PGM1
PC1	0.15	–0.63	0.37	0.01	–0.33	–0.26	0.26	0.09	0.12	0.18	0.1	–0.05	0.09	–0.17	0.3
PC2	–0.11	0.13	0.02	–0.02	0.45	–0.19	0.02	0.75	0.24	0.05	0.04	–0.2	–0.1	–0.04	0.22
TCA genes (PC1, 44.40%; PC2, 35.16%)															
Number	1	2	3	4	5	6	7	8	9	10	11	12	13		
E.C. #	1.1.1.41	1.1.1.41	1.2.4.1	1.2.4.1	1.2.4.2	1.3.5.1?	1.8.1.4	2.3.1.12	4.1.1.32	4.1.3.7	4.2.1.2	6.2.1.1	6.2.1.4		
Abbrev.	IDH2	IDH3A	PDHA1	PDHB	OGDH	SDHC	DLD	DLST	PCK2	CS	FH	ACAS2	SUCLG1		
PC1	–0.12	0.36	0.12	0.04	0.18	–0.29	0.01	0.25	0.33	0.07	–0.64	0.3	–0.23		
PC2	0.06	–0.28	–0.14	–0.06	0	0.34	–0.05	–0.21	–0.22	–0.15	–0.75	–0.2	0.23		
Oxidative phosphorylation genes (PC1, 48.76%; PC2, 16.30%)															
Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
E.C. #	1.10.2.2	1.10.2.2	1.10.2.2	1.10.2.2	1.6.5.3	1.6.5.3	1.6.5.3	1.6.5.3	1.6.5.3	1.6.5.3	1.6.5.3	1.6.5.3	1.6.5.3	1.6.5.3	1.6.5.3
Abbrev.	UQCRC2	UQCR10	UQCRC1	UQCRFS1	NDUFA4	NDUFA1	NDUFA10	NDUFA9	NDUFB6	MTND2	NDUFAB1	NDUFS4	NUFS1	NUOG	NDUFV2
PC1	0.12	–0.22	–0.07	0.2	–0.15	–0.11	0.01	–0.07	0	–0.2	0.14	0.11	0.17	0.31	–0.07
PC2	–0.04	0.01	–0.03	0.02	–0.03	–0.08	–0.07	–0.07	–0.08	0	–0.04	–0.14	–0.08	0.17	–0.04
Number	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
E.C. #	1.6.5.3	1.6.5.3	1.6.5.3	1.6.5.3	1.6.5.3	1.6.5.3	1.6.5.3	1.6.5.3	1.6.5.3	1.9.3.1	1.9.3.1	1.9.3.1	1.9.3.1	1.9.3.1	1.9.3.1
Abbrev.	NDUFA4	MTND5	NDUFS3	NDUFA8	NDUFB2	NDUFB8	NDUFB9	NDUFB1	NDUFB10	COXII	COX4I1	COX4I2	COX5A	COX6A1	COX6B
PC1	–0.22	–0.01	–0.07	–0.17	–0.24	–0.2	–0.21	–0.07	0.04	–0.13	0.06	0	0.03	–0.2	0.09
PC2	0.12	0	0.02	0.02	–0.29	0.04	0.05	–0.44	–0.09	0.08	–0.26	–0.11	–0.12	–0.06	0.01
Number	31	32	33	34	35	36	37	38	39	40	41	42	43	44	
E.C. #	1.9.3.1	1.9.3.1	1.9.3.1	1.9.3.1	3.6.1.34	3.6.1.34	3.6.1.34	3.6.1.34	3.6.1.34	3.6.1.34	3.6.1.34	3.6.1.34	3.6.1.34	3.6.1.35	
Abbrev.	COX7A2	COX8H	COX7C	COX8L	ATP5B	MTATP6	ATP5H	ATP5A2	ATP5D	ATP5C1	ATP5F1	ATP6V0C	ATP6V0D1	ATP6V1C1	
PC1	–0.06	–0.18	0.05	0.22	–0.25	–0.21	–0.15	0.15	0.01	–0.23	0.12	5.00%	8.00%	18.00%	
PC2	–0.07	–0.03	–0.11	0.15	0.02	0.25	–0.02	–0.14	–0.05	–0.07	–0.57				

Factor coefficients for each gene for the first two principal components are listed in parentheses. '%' is the percentage variance in gene expression explained by each PC. Many enzymes have more than one protein subunit, and numbers are used to identify these subunits. Full names of all genes are provided in Table S1 in supplementary material. Data from Oleksiak et al. (Oleksiak et al., 2005).

Discussion

Our microarray analyses of cardiac physiology demonstrated that >80% of the variation in substrate metabolism can be explained by variation in metabolic gene expression (Fig. 3) (Oleksiak et al., 2005). The strength of this study and our other cardiac microarray studies (Oleksiak et al., 2002; Oleksiak et al., 2005; Whitehead and Crawford, 2005; Whitehead and Crawford, 2006a; Whitehead and Crawford, 2006b) is that they used an outbred vertebrate population (such as humans) where the variation among individuals was substantial enough to statistically define meaningful patterns. The weakness of these studies is that they used outbred populations, which left unresolved the genetic contributions to the variation in gene expression that explain cardiac metabolism. For the studies provided here, only males were used, and these individuals were acclimated to the same environment. Thus, physiologically induced differences due to temperature, hypoxia or any other environmental parameter do not explain the observed differences.

Our conclusions that mRNA expression was statistically related to cardiac performance, that it is evolving by natural selection and that it varies in a predictable manner depending on genetic distance (Oleksiak et al., 2002; Oleksiak et al., 2005; Whitehead and Crawford, 2006a; Whitehead and Crawford, 2006b) suggest that microarray experiments can be meaningful. Thus, in contrast to Feder (Feder and Walser, 2005), we argue that there is much utility in measuring genome-wide patterns of gene expression. We suggest that these microarrays present patterns of mRNA expression that are both informative and provide unexpected relationships. However, they are not simple and thus are subject to misinterpretation.

Individual variation

We present several observations concerning individual variation: (1) there are significant differences in cardiac metabolism among individuals; (2) there are large differences in which substrate is preferentially used, (3) a vast majority of genes have significant differences in expression and (4) the genes that statistically explain the variation in substrate utilization differ among individuals. Our ability to explain these patterns depends on the delineation of three distinct groups among the 16 individuals, each group exhibiting a distinctive but consistent pattern of gene expression. The biological importance of these three groups is related to the

correlated patterns of gene expression among all 16 individuals (Fig. 4). Most genes were positively or negatively correlated. Thus, individuals with high expression of genes 'A' to 'E' also had low expression of genes 'W' to 'Z', and the opposite pattern happened for other individuals (high expression of W–Z and low expression of A–E). Thus, the groups (especially groups 2 and 3) represent individuals with opposite patterns of gene expression. The surprising observation is that these patterns occur within pathways. Thus, individuals in any one group share patterns of expression within a pathway that are different (opposite) from those in other groups. For example, among individuals in group 2, relative to group 3, there was greater expression of *GAPDH* and *PFK* but lower expression of *aldolase*, *aldehyde dehydrogenase*, *pyruvate kinase* and *PGI*.

What is the meaning of these large differences among individuals?

There were significant differences among individuals in (1) metabolism, (2) which metabolic substrate was preferentially used, (3) the pattern of metabolic gene expression and, consequently, (4) the relationship between gene expression and metabolism. The greater metabolic rate among some individuals was due to a greater utilization of glucose whilst in others it was the utilization of FA. The variation in specific substrates was explained by the expression of different metabolic pathways, but these explanations differ among groups of individuals. For example, fatty acid utilization can be explained by changes in oxidative, TCA or glycolytic enzyme expression depending on which group of individuals is examined.

These data have several implications. The first is that to examine one or a few inbred individuals could lead to misleading conclusions. Imagine having an inbred line from one *F. heteroclitus* individual. Cardiac metabolism in this imaginary strain could be primarily dependent on glucose and not on fatty acid utilization. Thus, one would conclude that glucose was the primary energy source for the heart. Yet, this is only correct for a subset of individuals. Similarly, the investigation of fatty acid utilization in cardiac tissue using one individual would reveal that gene expression in glycolytic enzymes regulates FA utilization as opposed to the TCA pathway, which would be found using another individual. Neither conclusion is incorrect, just misleading because it makes conclusions based on too few individuals.

The second implication from these data is that many genes or different pathways can affect substrate-specific metabolism. It is not surprising that the expression of glycolytic enzymes explains glucose metabolism (Fig. 3, group 1). However, the observation that the relative expression of genes in the oxidative phosphorylation pathway explained glucose metabolism among some individuals is somewhat surprising. More surprising was that glycolytic enzyme expression explained 75% of the variation in fatty acid utilization in group 3 individuals (Fig. 3). If these data are correct, it suggests that the activity of one pathway affects the flux through other pathways. Thus, we cannot measure one or a few enzymes and expect it to always explain metabolic variation among many different individuals.

The third implication is that it is unlikely that one gene, or a set of genes, is responsible for the phenotypic variation among all individuals. For example, if cardiac metabolism is related to health or fitness, then which genes affect the health of an individual is dependent on the status of other genes and pathways, an outcome that is entirely consistent with studies of metabolic epistasis, where metabolism is dependent on variation at other loci (Clark and Wang, 1997; Segrè et al., 2005). Thus, we should not expect a 'magic bullet' that will cure everyone of a specific disease. Instead, there will need to be different cures for different individuals, supporting the concept of personalized medicine that is currently driving many pharmaceutical research programs (Nadeau and Topol, 2006).

Finally, the lack of correlation between gene expression and phenotype is not necessarily due to the lack of importance of gene expression. Instead, if the importance of gene expression is context dependent then a significant relationship will only be discernable from within a specific context. Thus, the inability to accept an alternative hypothesis that there is a relationship between gene expression and phenotype does not support the null hypothesis. Instead, one has failed to address the proper hypothesis.

References

Brem, R. B. and Kruglyak, L. (2005). The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc. Natl. Acad. Sci. USA* **102**, 1572-1577.

Brem, R. B., Storey, J. D., Whittle, J. and Kruglyak, L. (2005). Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* **436**, 701-703.

Caceres, M. (2003). Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Natl Acad. Sci. USA* **100**, 13030.

Cavalieri, D., Townsend, J. P. and Hartl, D. L. (2000). Manifold anomalies in gene expression in a vineyard isolate of *Saccharomyces cerevisiae* revealed by DNA microarray analysis. *Proc. Natl. Acad. Sci. USA* **97**, 12369-12374.

Chesler, E. J., Lu, L., Shou, S., Qu, Y., Gu, J., Wang, J., Hsu, H. C., Mountz, J. D., Baldwin, N. E., Langston, M. A. et al. (2005). Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat. Genet.* **37**, 233-242.

Cheung, V. G., Conlin, L. K., Weber, T. M., Arcaro, M., Jen, K. Y., Morley, M. and Spielman, R. S. (2003a). Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat. Genet.* **33**, 422-425.

Cheung, V. G., Jen, K. Y., Weber, T., Morley, M., Devlin, J. L., Ewens, K. G. and Spielman, R. S. (2003b). Genetics of quantitative variation in human gene expression. *Cold Spring Harb. Symp. Quant. Biol.* **68**, 403-407.

Clark, A. G. and Wang, L. (1997). Epistasis in measured genotypes: *Drosophila* P-element insertions. *Genetics* **147**, 157-163.

Crawford, D. L., Pierce, V. A. and Segal, J. A. (1999a). Evolutionary physiology of closely related taxa: analyses of enzyme expression. *Am. Zool.* **39**, 389-400.

Crawford, D. L., Segal, J. A. and Barnett, J. L. (1999b). Evolutionary analysis of TATA-less proximal promoter function. *Mol. Biol. Evol.* **16**, 194-207.

Cui, X., Affourtit, J., Shockley, K. R., Woo, Y. and Churchill, G. A. (2006). Inheritance patterns of transcript levels in f1 hybrid mice. *Genetics* **174**, 627-637.

Denver, D. R., Morris, K., Streelman, J. T., Kim, S. K., Lynch, M. and Thomas, W. K. (2005). The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nat. Genet.* **37**, 544-548.

Doss, S., Schadt, E. E., Drake, T. A. and Lusis, A. J. (2005). *Cis*-acting expression quantitative trait loci in mice. *Genome Res.* **15**, 681-691.

Enard, W. (2002). Intra- and interspecific variation in primate gene expression patterns. *Science* **296**, 340.

Farrell, A. P. (1993). Cardiovascular system. In *The Physiology of Fishes* (ed. D. H. Evans), pp. 219-250. Boca Raton: CRC Press.

Feder, M. E. and Walser, J. C. (2005). The biological limitations of transcriptomics in elucidating stress and stress responses. *J. Evol. Biol.* **18**, 901-910.

Ghazalpour, A., Doss, S., Sheth, S. S., Ingram-Drake, L. A., Schadt, E. E., Lusis, A. J. and Drake, T. A. (2005). Genomic analysis of metabolic pathway gene expression in mice. *Genome Biol.* **6**, R59.

Gibson, G. and Weir, B. (2005). The quantitative genetics of transcription. *Trends Genet.* **21**, 616-623.

Gilad, Y., Rifkin, S. A., Bertone, P., Gerstein, M. and White, K. P. (2005). Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles. *Genome Res.* **15**, 674.

Gilad, Y., Oshlack, A., Smyth, G. K., Speed, T. P. and White, K. P. (2006). Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature* **440**, 242.

Jin, W., Riley, R. M., Wolfinger, R. D., White, K. P., Passador-Gurgel, G. and Gibson, G. (2001). The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat. Genet.* **29**, 389-395.

Khaitovich, P. (2004a). A neutral model of transcriptome evolution. *PLoS Biol.* **2**, e132.

Khaitovich, P. (2004b). Regional patterns of gene expression in human and chimpanzee brains. *Genome Res.* **14**, 1462.

Lemos, B., Meiklejohn, C. D., Caceres, M. and Hartl, D. L. (2005). Rates of divergence in gene expression profiles of primates, mice and flies: stabliizing selection and variability among functional categories. *Evolution* **59**, 126-137.

Monks, S. A., Leonardson, A., Zhu, H., Cundiff, P., Pietrusiak, P., Edwards, S., Phillips, J. W., Sachs, A. and Schadt, E. E. (2004). Genetic inheritance of gene expression in human cell lines. *Am. J. Hum. Genet.* **75**, 1094-1105.

Morley, M., Molony, C. M., Weber, T. M., Devlin, J. L., Ewens, K. G., Spielman, R. S. and Cheung, V. G. (2004). Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**, 743-747.

Nadeau, J. H. and Topol, E. J. (2006). The genetics of health. *Nat. Genet.* **38**, 1095-1098.

Nuzhdin, S. V., Wayne, M. L., Harmon, K. L. and McIntyre, L. M. (2004). Common pattern of evolution of gene expression level and protein sequence in *Drosophila*. *Mol. Biol. Evol.* **21**, 1308-1317.

Oleksiak, M. F., Churchill, G. A. and Crawford, D. L. (2002). Variation in gene expression within and among natural populations. *Nat. Genet.* **32**, 261-266.

Oleksiak, M. F., Roach, J. L. and Crawford, D. L. (2005). Natural variation in cardiac metabolism and gene expression in *Fundulus heteroclitus*. *Nat. Genet.* **37**, 67-72.

Pierce, V. A. and Crawford, D. L. (1997a). Phylogenetic analysis of glycolytic enzyme expression. *Science* **276**, 256-259.

Podrabsky, J. E., Javillonar, C., Hand Steven, C. and Crawford, D. L. (2000). Intraspecific variation in aerobic metabolism and glycolytic enzyme expression in heart ventricles. *Am. J. Physiol.* **279**, R2344-R2348.

Pritchard, C. C., Hsu, L., Delrow, J. and Nelson, P. S. (2001). Project normal: defining normal variance in mouse gene expression. *Proc. Natl. Acad. Sci. USA* **98**, 13266-13271.

Ronald, J., Brem, R. B., Whittle, J. and Kruglyak, L. (2005). Local Regulatory Variation in *Saccharomyces cerevisiae*. *PLoS Genet.* **1**, e25.

Schadt, E. E., Monks, S. A., Drake, T. A., Lusis, A. J., Che, N., Colinayo, V., Ruff, T. G., Milligan, S. B., Lamb, J. R., Cavet, G. et al. (2003). Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297-302.

Schadt, E. E., Lamb, J., Yang, X., Zhu, J., Edwards, S., Guhathakurta, D., Sieberts, S. K., Monks, S., Reitman, M., Zhang, C. et al. (2005). An integrative genomics approach to infer causal associations between gene expression and disease. *Nat. Genet.* **37**, 710-717.

Segrè, D., DeLuna, A., Church, G. M. and Kishony, R. (2005). Modular epistasis in yeast metabolism. *Nat. Genet.* **37**, 77-83.

Sharma, A., Sharma, V. K., Horn-Saban, S., Lancet, D., Ramachandran, S. and Brahmachari, S. K. (2005). Assessing natural variations in gene expression in humans by comparing with monozygotic twins using microarrays. *Physiol. Genomics* **21**, 117-123.

Stamatoyannopoulos, J. A. (2004). The genomics of gene expression. *Genomics* **84**, 449-457.

Tan, Q., Christensen, K., Christiansen, L., Frederiksen, H., Bathum, L., Dahlgaard, J. and Kruse, T. A. (2005). Genetic dissection of gene expression observed in whole blood samples of elderly Danish twins. *Hum. Genet.* **117**, 267-274.

Townsend, J. P., Cavalieri, D. and Hartl, D. L. (2003). Population genetic variation in genome-wide gene expression. *Mol. Biol. Evol.* **20**, 955-963.

Wang, M. H., Lazebny, O., Harshman, L. G. and Nuzhdin, S. V. (2004). Environment-dependent survival of *Drosophila melanogaster*: a quantitative genetic analysis. *Aging Cell* **3**, 133-140.

Wayne, M. L. and McIntyre, L. M. (2002). Combining mapping and arraying: an approach to candidate gene identification. *Proc. Natl. Acad. Sci. USA* **99**, 14903-14906.

Wayne, M. L., Pan, Y. J., Nuzhdin, S. V. and McIntyre, L. M. (2004). Additivity and trans-acting effects on gene expression in male *Drosophila simulans*. *Genetics* **168**, 1413-1420.

Whitehead, A. and Crawford, D. (2005). Variation in tissue-specific gene expression among natural populations. *Genome Biol.* **6**, R13.1-R13.14.

Whitehead, A. and Crawford, D. L. (2006a). Neutral and adaptive variation in gene expression. *Proc. Natl. Acad. Sci. USA* **103**, 5425-5430.

Whitehead, A. and Crawford, D. L. (2006b). Variation within and among species in gene expression: raw material for evolution. *Mol. Ecol.* **15**, 1197-1211.

Yvert, G., Brem, R. B., Whittle, J., Akey, J. M., Foss, E., Smith, E. N., Mackelprang, R. and Kruglyak, L. (2003). Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. *Nat. Genet.* **35**, 57-64.

The Journal of Experimental Biology 210, 1492-1496
Published by The Company of Biologists 2007
doi:10.1242/jeb.02783

Glossary of terms

This section is designed to help readers adapt to the complex terminology associated with contemporary molecular genetics, genomics and systems biology. Fuller descriptions of these terms are available at http://www.wikipedia.org/

Ab initio prediction	methods used to predict the potential genes encoded in the genome, which are trained on datasets made of known genes, and used computationally to predict coding regions out of genome without the aid of cDNA sequence. Although their performance is improving, these algorithms perform very poorly on non-protein coding genes.
Annotation	as applied to proteins, DNA sequences or genes. The storage of data describing these entities (protein/gene identities, DNA motifs, gene ontology categorisation, etc.) within a biological database. Active projects include FlyBase and WormBase. See Gene ontology.
Assembly	the process of aligning sequenced fragments of DNA into their correct positions within the chromosome or transcript.
cDNA	complementary DNA. This is DNA synthesised from a mature mRNA template by the enzyme reverse transcriptase. cDNA is frequently used as an early part of gene cloning procedures, since it is more robust and less subject to degradation than the mRNA itself.
ChIP	chromatin immunoprecipitation assay used to determine which segments of genomic DNA are bound to chromatin proteins, mainly including transcription factors.
Chip	see Microarray.
ChIP-on-chip	use of a DNA microarray to analyse the DNA generated from chromatin immunoprecipitation experiments (see ChIP).
cis-acting	a molecule is described as *cis*-acting when it affects other genes that are physically adjacent, on the same chromosome, or are genetically linked or in close proximity (for mRNA expression, typically a promoter).
Collision-induced dissociation	a mechanism by which molecules (e.g. proteins) are fragmented to form molecular ions in the gas phase. These fragments are then analysed within a mass spectrometer to provide mass determination.
Connectivity	a term from graph theory, which indicates the number of connections between nodes or vertices in a network. Greater connectedness between nodes is generally used as a measure of robustness of a network.
CpG islands	regions that show high density of ‘C followed by G’ dinucleotides and are generally associated with promoter elements; in particular, stretches of DNA of at least 200 bp with a C–G content of 50% and an observed CpG/expected CpG in excess of 0.6. The cytosine residues can be methylated, generally to repress transcription, while demethylated CpGs are a hallmark of transcription. CpG dinucleotides are under-represented outside regulatory regions, such as promoters, because methylated C mutates into T by deamination.
Edge	as in networks. Connects two nodes (or vertices) within a system. These concepts arise from graph theory.
Enhancer	a short segment of genomic DNA that may be located remotely and that, on binding particular proteins (*trans*-acting factors), increases the rate of transcription of a specific gene or gene cluster.
Epistasis	a phenomenon when the properties of one gene are modified by one or more genes at other loci. Otherwise known as a genetic interaction, but epistasis refers to the statistical properties of the phenomenon.

eQTL	the combination of conventional QTL analysis with gene expression profiling, typically using microarrays. eQTLs describe regulatory elements controlling the expression of genes involved in specific traits.
EST	expressed sequence tag. A short DNA sequence determined for a cloned cDNA representing portions of an expressed gene. The sequence is generally several hundred base pairs from one or both ends of the cloned insert.
Exaptation	a biological adaptation where the current function is not that which was originally evolved. Thus, the defining (derived) function might replace or persist with the earlier, evolved adaptation.
Exon	any region of DNA that is transcribed to the final (spliced) mRNA molecule. Exons interleave with segments of non-coding DNA (introns) that are removed (spliced out) during processing after transcription.
Gene forests	genomic regions for which RNA transcripts, produced from either DNA strand, have been identified without gaps (non-transcribed genomic regions). Conversely, regions in which no transcripts have ever been detected are called ‘gene deserts’.
Gene interaction network	a network of functional interactions between genes. Functional interactions can be inferred from many different data types, including protein–protein interactions, genetic interactions, co-expression relationships, the co-inheritance of genes across genomes and the arrangement of genes in bacterial genomes. The interactions can be represented using network diagrams, with lines connecting the interacting elements, and can be modelled using differential equations.
Gene ontology (GO)	an ontology is a controlled vocabulary of terms that have logical relationships with each other and that are amenable to computerised manipulation. The Gene Ontology project has devised terms in three domains: biological process, molecular function and cell compartment. Each gene or DNA sequence can be associated with these annotation terms from each domain, and this enables analysis of microarray data on groups of genes based on descriptive terms so provided. See http://www.geneontology.org
Gene set enrichment analysis	a computational method that determines whether a defined set of genes, usually based on their common involvement in a biological process, shows statistically significant differences in transcript expression between two biological states.
Gene silencing	the switching-off of a gene by an epigenetic mechanism at the transcriptional or post-transcriptional levels. Includes the mechanism of RNAi.
Genetic interaction (network)	a genetic interaction between two genes occurs when the phenotypic consequences of a mutation in one gene are modified by the mutational status at a second locus. Genetic interactions can be aggravating (enhancing) or alleviating (suppressing). To date, most high-throughput studies have focussed on systematically identifying synthetic lethal or sick (aggravating) interactions, which can then be visualised as a network of functional interactions (edges) between genes (nodes).
Genome	a portmanteau of gene and chromosome, the entire hereditary information for an organism that is embedded in the DNA (or, for some viruses, in RNA). Includes protein-coding and non-coding sequences.
Heritability	phenotypic variation within a population is attributable to the genetic variation between individuals and to environmental factors. Heritability is the proportion due to genetic variation usually expressed as a percentage.
Heterologous hybridization	the use of a cDNA or oligonucleotide microarray of probes designed for one species with target cRNA/cDNAs from a different species.
Homeotic	the transformation of one body part to another due to mutation of specific developmentally related genes, notably the *Hox* genes in animals and *MADS-box* genes in plants.
Hub	as in networks. A node with high connectivity, and thus which interacts with many other nodes in the network. A hub protein interacts with many other proteins in a cell.

Hybridisation	the process of joining (annealing) two complementary single-stranded DNAs into a single double-stranded molecule. In microarray analysis, the target RNA/DNA from the subject under investigation is denatured and hybridised to probes that are immobilised on a solid phase (i.e. glass microscope slide).
Hypomorph	in genetics, a loss-of-function mutation in a gene, but which shows only a partial reduction in the activity it influences rather than a complete loss (cf. hypermorph, antimorph, neomorph, etc).
Imprinting	a phenomenon where two inherited copies of a gene are regulated in opposite ways, one being expressed and the other being repressed.
Indel	insertion and deletion of DNA, referring to two types of genetic mutation. To be distinguished from a 'point mutation', which refers to the substitution of a single base.
Interactome	a more or less comprehensive set of interactions between elements within cells. Usually applied to genes or proteins as defined by transcriptomic, proteomic or protein–protein interaction data.
Intron	see Exon.
KEGG	The Kyoto Encyclopedia of Genes and Genomes is a database of metabolic and other pathways collected from a variety of organisms. See http://www.genome.jp/kegg
Metabolomics	the systematic qualitative and quantitative analysis of small chemical metabolite profiles. The metabolome represents the collection of metabolites within a biological sample.
Metagenomics	the application of genomic techniques to characterise complex communities of microbial organisms obtained directly from environmental samples. Typically, genomic tags are sequence characterised as markers of each species to inform on the range and abundance of species in the community.
Microarray	an arrayed set of probes for detecting molecularly specific analytes or targets. Typically, the probes are composed of DNA segments that are immobilised onto the solid surface, each of which can hybridise with a specific DNA present in the target preparation. DNA microarrays are used for profiling of gene transcripts.
Model species	a species used to study particular biological phenomena, the outcome offering insights into the workings of other species. Usually, the selection is based on experimental tractability, particularly ease of genetic manipulation. For the geneticist, it is an organism with inbred lines where sibs will be >98% identical (i.e. *Drosophila*, *Caenorhabditis elegans* and mice). For genomic science, it refers to a species for which the genomic DNA has been sequenced.
miRNA	a category of novel, very short, non-coding RNAs, generated by the cleavage of larger precursors (pri-miRNA). These short RNAs are included in the RNA-induced silencing complex (RISC) and pair to the 3′ ends of target RNA, blocking its translation into proteins (in animals) or promoting RNA cleavage and degradation (in plants).
mRNA	a protein-coding mRNA containing a protein-coding region (CDS), preceded by a 5′ and followed by a 3′ untranslated region (5′ UTR and 3′ UTR). The UTRs contain regulatory elements. A full-length cDNA contains the complete sequence of the original mRNA, including both UTRs. However, it is often difficult to assign the starting–termination positions for protein synthesis unambiguously. A cDNA containing the entire CDS is often considered acceptable for bioinformatic and experimental studies requiring full-length cDNAs.
ncRNA	non-coding RNA is any RNA molecule with no obvious protein-coding potential for at least 80 or 100 amino acids, as determined by scanning full-length cDNA sequences. It includes ribosomal (rRNA) and transfer RNAs (tRNA) and is now known to include various sub-classes of RNA, including snoRNA, siRNA and piRNA. Just like the coding mRNAs, a large proportion of ncRNAs are transcribed by RNA polymerase II and are large transcripts. A description of the many forms of ncRNA can be found at http://en.wikipedia.org/wiki/Non-coding_RNA.

Node	as in networks. Objects linked by edges to create a network.
PCR	polymerase chain reaction. A molecular biology technique for replicating DNA *in vitro*. The DNA is thus amplified, sometimes from very small amounts. PCR can be adapted to perform a wide variety of genetic manipulations.
piRNA	Piwi-interacting RNA. A class of RNA molecules (29–30 nt long) that complex with Piwi proteins (a class of the Argonaute family of proteins) and are involved in transcriptional gene silencing.
PMF	peptide mass fingerprinting. An analytical technique for protein identification in which a protein is fragmented using proteases. The resulting peptides are analysed by mass spectrometry and these masses compared against a database of predicted or measured masses to generate a protein identity.
Polyadenylation	the covalent addition of multiple A bases to the 3′ tail of an mRNA molecule. This occurs during the processing of transcripts to form the mature, spliced molecule and is important for regulation of turnover, trafficking and translation.
Post-source decay	in mass spectrometry. The fragmentation of precursor molecular ions as they accelerate away from the ionisation source of the mass spectrometer. All precursor ions leaving the ion source have approximately the same kinetic energy, but fragmentation results in smaller product ions that can be distinguished from precursor ions using a ‘reflectron’ by virtue of their lower kinetic energies.
Post-translational modification	the chemical modification of a protein after synthesis through translation. Some modifications, notably phosphorylation, affect the properties of the protein, offering a means of regulating function.
Principal component analysis (PCA)	a technique for simplifying complex, multi-dimensional datasets to a reduced number of dimensions, the principal components. This procedure retains those characteristics of the data that relate to its variance.
Promoter	a regulatory DNA sequence, generally lying upstream of an expressed gene, which in concert with other often distant regulatory elements directs the transcription of a given gene.
Proteome	the entire protein complement of an organism, tissue or cell culture at a given time.
Quantitative trait	inheritance of a phenotypic property or characteristic that varies continuously between extreme states and can be attributed to interactions between multiple genes and their environment.
qPCR	quantitative real-time PCR, sometimes called real-time PCR. A more quantitative form of RT-PCR in which the quantity of amplified product is estimated after each round of amplification.
QTL	quantitative trait loci. A region of DNA that contains those genes contributing to the trait under study.
RISC	RNA-induced silencing complex. A protein complex that mediates the double-stranded RNA-induced destruction of homologous mRNA.
RNAi	RNA interference or RNA-mediated interference. The process by which double-stranded RNA triggers the destruction of homologous mRNA in eukaryotic cells by the RISC.
RT-PCR	reverse transcription–polymerase chain reaction. A technique for amplifying a defined piece of RNA that has been converted to its complementary DNA form by the enzyme reverse transcriptase. See qPCR.
siRNA	small interfering RNA, or silencing RNA. A class of short (20–25 nt), double-stranded RNA molecules. It is involved in the RNA interference pathway, which alters RNA stability and thus affects RNA concentration and thereby suppresses the normal expression of specific genes. Widely used in biomedical research to ablate specific genes.

snoRNA	small nucleolar RNA. A sub-class of RNA molecules involved in guiding chemical modification of ribosomal RNA and other RNA genes as part of the regulation of gene expression.
SNP	single nucleotide polymorphism. A single base-pair mutation at a specific locus, usually consisting of two alleles. Because SNPs are conserved over evolution, they are frequently used in QTL analysis and in association studies in place of microsatellites, and in genetic fingerprinting analyses.
SSH	suppressive subtractive hybridisation. A powerful protocol for enriching cDNA libraries for genes that differ in representation between two or more conditions. It combines normalisation and subtraction in a single procedure and allows the detection of low-abundance, differentially expressed transcripts, such as those involved in signalling and signal transduction.
Structural RNAs	a class of non-coding RNA, long known to have a structural role (for instance, the ribosomal RNAs), transcribed by RNA polymerase I or III.
Systems biology	treatment of biological entities as systems composed of defined elements interacting in defined ways to enable the observed function and behaviour of that system. The properties of the systems are embedded in a quantitative model that guides further tests of systems behaviour.
TATA-boxes	sequences in promoter regions constituted by TATAAA, or similar variants, which were considered the hallmark of Promoters. Recent data show that they are present only in the minority of promoters, where they direct transcription at a single well-defined location some 30 bp downstream of this element.
trans-acting	a factor or gene that acts on another unlinked gene, a gene on a separate chromosome or genetically unlinked usually through some diffusible protein product (for mRNA expression, typically a transcription factor).
Transcript	an RNA product produced by the action of RNA polymerase reading the sequence of bases in the genomic DNA. Originally limited to protein-coding sequences with flanking UTRs but now known to include large numbers of products that do not code for a protein product.
Transcriptome	the full set of mRNA molecules (transcripts) produced by the system under observation. Whilst the genome is fixed for a given organism, the transcriptome varies with context (i.e. tissue source, ontogeny, external conditions or experimental treatment).
Transgene	a gene or genetic material that has been transferred between species or between organisms using one of several genetic engineering techniques.
Transinduction	generation of transcripts from intergenic regions. At least some such products do not relate to a definable promoter or transcriptional start site.
Transposon	sequences of DNA able to move to new positions within the genome of a single cell. This event might cause mutation at the site of insertion. Also called ‘mobile genetic elements’ or ‘jumping genes’.
Transvection	an epigenetic phenomenon arising from the interaction between one allele and the corresponding allele on the homologous chromosome, leading to gene regulation.
TUs	transcriptional units. Used to group all of the overlapping RNA transcripts that are transcribed from the same genomic strand and share exonic sequences.
UTR	untranslated region. Regions of the mRNA that lie at either the 3′ or 5′ flanking ends of the molecule (i.e. 3′ UTR and 5′ UTR). They bracket the protein-coding region and contain signals and binding sites that are important for the regulation of both protein translation and RNA degradation.

Table S1. *List of genes used for principal component analyses*

Number	E.C. #:	
Glycolysis/gluconeogenesis genes		
1	1.1.1.27	Lactate dehydrogenase B (LDHB)
2	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)
3	1.2.1.3	Aldehyde dehydrogenase 1 family member A
4	2.4.1.1	Glycogene phosphorylase muscle
5	2.7.1.11	6-Phosphofructokinase
6	2.7.1.40	Pyruvate kinase (muscle isozyme)
7	2.7.2.3	Phosphoglycerate kinase 1 (PGK)
8	3.1.3.11	Fructose bisphosphatase
9	4.1.2.13	Aldolase 1 A muscle
10	4.2.1.11	Enolase (alpha)
11	5.3.1.1	Triosephosphate isomerase chain B
12	5.3.1.9	Phosphoglucose isomerase
13	5.3.1.9	Glucose-6-phosphate isomerase
14	5.4.2.1	Phosphoglycerate mutase (PGM) type B
15	5.4.2.2	Phosphoglucomutase
TCA genes		
1	1.1.1.41	Isocitrate dehydrogenase 2 (mitochondrial IDH2)
2	1.1.1.41	Isocitrate dehydrogenase isozyme 3
3	1.2.4.1	Pyruvate dehydrogenase E1 beta subunit (lipoamide)
4	1.2.4.1	Pyruvate dehydrogenase E1 alpha subunit
5	1.2.4.2	Oxoglutarate (alpha ketoglutarate) dehydrogenase (lipoamide)
6	1.3.5.1?	Succinate dehydrogenase complex subunit C
7	1.8.1.4	Dihydrolipoamide dehydrogenase E3 component of pyruvate dehydrogenase complex
8	2.3.1.12	Dihydrolipoamide S acetyltransferase (E2 component of pyruvate dehydrogenase complex)
9	4.1.1.32	PEP carboxykinase phosphoenolpyruvate carboxykinase
10	4.1.3.7	Citrate synthase
11	4.2.1.2	Fumarate hydratase
12	6.2.1.1	Acetyl CoA synthetase
13	6.2.1.4	Succinate CoA ligase (GDP forming)
Oxidative phosphorylation genes		
1.10.2.2. = complex III cytochrome reductase		
1	1.10.2.2	Ubiquinol cytochrome *c* reductase core protein II
2	1.10.2.2	Ubiquinol cytochrome *c* reductase complex 7.2 kDa protein (cytochrome *c*1 nonheme 7 kDa protein) (complex III subunit x)
3	1.10.2.2	Ubiquinol cytochrome *c* reductase core protein I
4	1.10.2.2	Ubiquinol cytochrome *c* reductase iron sulfur subunit mitochondrial precursor (Rieske iron sulfur protein)
1.6.5.3 = complex 1 NADH dehydrogenase		
5	1.6.5.3	NADH dehydrogenase (ubiquinone)
6	1.6.5.3	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 1 (7.5 kDa MWFE)
7	1.6.5.3	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 10 (42 kDa)
8	1.6.5.3	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 9 (39 kDa)
9	1.6.5.3	NADH dehydrogenase (ubiquinone) 1 beta subcomplex 6 (17 kDa B17)
10	1.6.5.3	NADH dehydrogenase (ubiquinone) chain 2
11	1.6.5.3	NADH dehydrogenase (ubiquinone) acyl carrier chain (ACPM) alpha beta subcomplex 1 (8 kDa SDAP)
12	1.6.5.3	NADH dehydrogenase (ubiquinone) Fe S protein 4 (18 kDa) (NADH coenzyme Q reductase)
13	1.6.5.3	NADH dehydrogenase (ubiquinone) Fe S protein 1
14	1.6.5.3	NADH dehydrogenase (ubiquinone) Fe S protein 1 NADH ubiquinone oxidoreductase 75 kDa subunit precursor (complex I 75 kDa) (CI 75 kDa)
15	1.6.5.3	NADH dehydrogenase (ubiquinone) flavoprotein 2
16	1.6.5.3	NADH dehydrogenase (ubiquinone) MLRQ subunit (complex I MLRQ)
17	1.6.5.3	NADH dehydrogenase subunit 5
18	1.6.5.3	NADH dehydrogenase (ubiquinone) Fe S protein 3 (30 kDa) (NADH coenzyme Q reductase)
19	1.6.5.3	NADH ubiquinone oxidoreductase 19 kDa subunit (complex I 19 kDa) (CI 19 kDa) (complex I PGIV) (CI PGIV)
20	1.6.5.3	NADH ubiquinone oxidoreductase AGGG subunit precursor (complex I AGGG) (CI AGGG)
21	1.6.5.3	NADH ubiquinone oxidoreductase ASHI subunit precursor (complex I ASHI) (CI ASHI)
22	1.6.5.3	NADH ubiquinone oxidoreductase B22 subunit (complex I B22) (CI B22)

23	1.6.5.3	NADH ubiquinone oxidoreductase MNLL subunit (complex I MNLL) (CI MNLL)
24	1.6.5.3	NADH ubiquinone oxidoreductase PDSW subunit (complex I PDSW) (CI PDSW)
	1.9.3.1 = complex IV, cytochrome c oxidase	
25	1.9.3.1	Cytochrome *c* oxidase subunit II
26	1.9.3.1	Cytochrome *c* oxidase subunit IV isoform 1
27	1.9.3.1	Cytochrome *c* oxidase subunit IV isoform 2
28	1.9.3.1	Cytochrome *c* oxidase polypeptide Va
29	1.9.3.1	Cytochrome *c* oxidase polypeptide VIa precursor
30	1.9.3.1	Cytochrome *c* oxidase polypeptide VIb
31	1.9.3.1	Cytochrome *c* oxidase polypeptide VIIa liver heart mitochondrial precursor (cytochrome *c* oxidase subunit VIIa L)
32	1.9.3.1	Cytochrome *c* oxidase subunit VIIIb
33	1.9.3.1	CYTOCHROME C OXIDASE POLYPEPTIDE VIIC PRECURSOR (VIIIA)
34	1.9.3.1	cytochrome c oxidase subunit VIIIa
	3.6.1.34 = ATP synthetase	
35	3.6.1.34	ATP synthase subunit B
36	3.6.1.34	ATP synthase beta subunit
37	3.6.1.34	ATP synthase H^+ transporting mitochondrial F1 complex delta subunit
38	3.6.1.34	ATP synthase H^+ transporting mitochondrial F0 complex subunit d
39	3.6.1.34	ATP synthase H^+ transporting mitochondrial F1 complex alpha subunit isoform 1
40	3.6.1.34	ATP synthase H^+ transporting mitochondrial F1 complex gamma polypeptide 1
41	3.6.1.34	ATP synthase F0 subunit 6 ATPase 6
42	3.6.1.34	ATPase H^+ transporting lysosomal (vacuolar proton pump D subunit) 42 kDa
43	3.6.1.34	ATPase H^+ transporting lysosomal (vacuolar proton pump) 16 kDa (Atp6l)
44	3.6.1.35	Vacuolar ATP synthase subunit c (V ATPase c subunit) (vacuolar proton pump c subunit)