Current Biology, Vol. 12, 712–723, April 30, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S0960-9822(02)00808-4

Genome-Wide Transcript Profiles in Aging and Calorically Restricted *Drosophila melanogaster*

Scott D. Pletcher,[1,4] Stuart J. Macdonald,[1] Richard Marguerie,[1] Ulrich Certa,[3] Stephen C. Stearns,[2] David B. Goldstein,[1] and Linda Partridge[1]
[1]Department of Biology
University College London
The Darwin Building
Gower Street
London, WC1E 6BT
United Kingdom
[2]Department of Ecology and Evolutionary Biology
Box 208106
Yale University
New Haven, Connecticut 06520-8106
[3]F. Hoffmann La Roche & Company
Roche Genet
CH 4070 Basel
Switzerland

Summary

Background: We characterized RNA transcript levels for the whole *Drosophila* genome during normal aging. We compared age-dependent profiles from animals aged under full-nutrient conditions with profiles obtained from animals maintained on a low-calorie medium to determine if caloric restriction slows the aging process. Specific biological functions impacted by caloric restriction were identified using the Gene Ontology annotation. We used the global patterns of expression profiles to test if particular genomic regions contribute differentially to changes in transcript profiles with age and if global disregulation of gene expression occurs during aging.

Results: Whole-genome transcript profiles contained a statistically powerful genetic signature of normal aging. Nearly 23% of the genome changed in transcript representation with age. Caloric restriction was accompanied by a slowing of the progression of normal, age-related changes in transcript levels. Many genes, including those associated with stress response and oogenesis, showed age-dependent transcript representation. Caloric restriction resulted in the downregulation of genes primarily involved in cell growth, metabolism, and reproduction. We found no evidence that age-dependent changes in transcription level were confined to genes localized to specific regions of the genome and found no support for widespread disregulation of gene expression with age.

Conclusions: Aging is characterized by highly dynamic changes in the expression of many genes, which provides a powerful molecular description of the normal aging process. Caloric restriction extends life span by slowing down the rate of normal aging. Transcription levels of genes from a wide variety of biological functions and processes are impacted by age and dietary conditions.

Background

Aging is characterized by an age-dependent increase in the probability of death and concomitant decline in reproductive output. A powerful experimental approach for investigating mechanisms of aging is provided by interventions that slow down or accelerate the normal process. Manipulations that increase life span, such as mutants in single genes [1], are strong candidates for such interventions. More problematic are progerias, in which aspects of normal aging appear to be accelerated. Difficulties in interpretation arise because interventions that shorten life span could do so, at least in part, by introducing novel pathologies. The lack of reliable, quantitative descriptions of the normal aging process [2] has frustrated efforts to develop a direct test of the effect of interventions that alter life span by changing the rate of normal aging. A case in point is caloric restriction (CR), in which nutrient intake is restricted to about 65% of voluntary levels. CR has been shown to increase life span in taxonomically diverse organisms including yeast [3], nematode worms [4], fruit flies [5], and rodents [6]. CR has also been shown to reduce reproductive output, to retard the development of age-related diseases such as cancer, and to slow down some normal, age-dependent changes in gene expression [7]. However, a detailed, genome-wide assessment of the consequences that CR has on normal transcriptional changes that occur with age has not so far been possible.

To ascertain detailed age-dependent profiles of transcript representation in normally aging organisms and to determine the impact that CR has on these transcriptional changes, we measured age-dependent gene expression profiles in large cohorts of female *Drosophila melanogaster* maintained on either full (control) or diluted food medium. The diluted food medium imposes a regimen of caloric restriction that delays the onset of senescent mortality and increases mean and maximum life span by roughly 2-fold [5] (Figure 1). A key goal of this project was to ensure enough statistical power to detect modest or complicated age dependencies in gene expression. We employed a powerful experimental design, with 5-fold replicated measurements at six and eight different age points in the control and CR cohorts, respectively. In addition, two replicate measurements were allocated to very young flies, which were maintained on medium with intermediate yeast/glucose content for 3 days prior to transfer to treatment medium. Our experiment allowed for the separation of biological sources of variation in expression from experimental error, and we developed an analysis of variance framework that took full advantage of the age-dependent trends in the data (see the Experimental Procedures).

We described and characterized the biology that underlies age- and CR-induced changes in transcript represen-

[4]Correspondence: s.pletcher@ucl.ac.uk

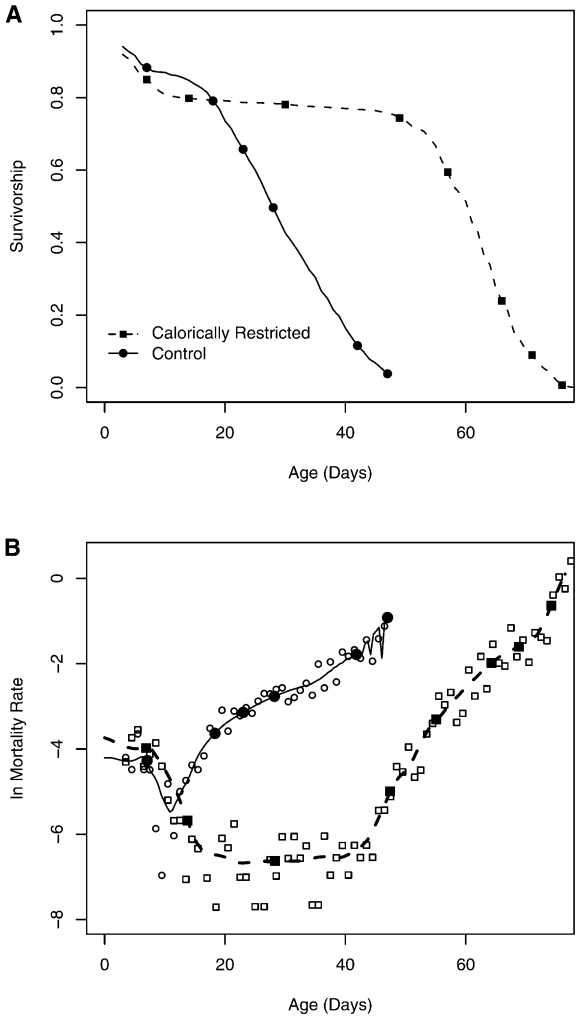

Figure 1. Survival Statistics for Control and Calorically Restricted Cohorts of *Drosophila*

(A and B) (A) Survivorship and (B) age-specific mortality for flies maintained on either control (enriched) medium or caloric restriction (CR) medium, which contains 66% less yeast and glucose. The large filled symbols represent days at which samples of 50 live flies were taken for expression analysis. Survivorship was calculated using a Kaplan-Meier estimate treating sampled flies as censored observations (log-rank test of control versus CR flies; $p = 0$). Mortality rates are product limit-based Kaplan-Meier hazard estimates (open symbols) smoothed using a kernel estimator (solid and dashed lines) described in Mueller and Wang [23]. Initial cohort sizes were 2500 (control) and 3000 females (CR). The mean (SE) life span of the control and CR flies was 25.4 (0.46) and 46.2 (0.24) days, respectively. The maximum life span in CR conditions was 78 days, and the last 50 flies alive on enriched media were sampled on day 47.

tation using the annotation project directed by the Gene Ontology (GO) Consortium (http://www.gene-ontology.org). Three separate ontologies, comprising biological process, molecular function, and cellular component, define a set of well-defined terms and relationships through which we interpreted the role of a particular gene, gene product, or gene-product group. Although rapidly evolving, the three ontologies contained information on approximately 53% (7,389/14,028) of the probe sets (genes and potential splice variants) present on our array at the time this article was formulated. The GO is designed to emphasize the transfer of gene/protein function information among organisms, and it serves well as a foundation for unifying the growing amount of expression information related to aging in a variety of model systems [8]. Data involving the ontologies and age- and diet-dependent expression profiles, as well as software for browsing the data, are available from the corresponding author.

Results and Discussion

Age- and Diet-Dependent Changes in Expression Are Prevalent and Complex

Transcript representation of a high percentage of the fly genome changed significantly with age. Despite requiring strict statistical correction for multiple testing, we identified 1264 probe sets (9% of all probe sets) representing 1203 genes (some genes were represented by 2 or more different probe sets) with age-related changes in either or both of the control and CR cohorts (Figure 2; top). Under this strict criterion, 885 and 827 probe sets exhibited age-dependent profiles in the control and CR treatments, respectively, and 448 probe sets were common to both (additional probe sets with changes in both treatments are seen when the criterion is relaxed). A total of 1782 age-dependent changes are significant at a level that allows for a single false positive, and if the statistical stringency is relaxed further for comparison to published results, the number rises to roughly 3200 (nearly 23% of the available probe sets). These findings contrast with previous genome analyses of aging in the fly [9] and mouse [7, 10], which both found age-related changes in roughly 2% of the genes examined, and in rhesus monkeys [11], where roughly 6% of the available genes were up- or downregulated with age. Part of this discrepancy likely results from changes specific to female reproduction (Zou et al. [9] examined males, while we used once-mated females), and our analysis of individual genes bears this out (Tables 1 and 2). However, most of the age-dependent changes in transcription continue well beyond the age when egg laying has stopped (S.D.P. and S.J.M., unpublished data) and are not obviously connected with reproduction.

Closer examination of the 1203 genes with age-dependent patterns revealed a variety of complex age-related changes in transcript representation (Figure 2; top). At least 18 genes showed strong quadratic tendencies (Figures 2H and 2O), and a subset of approximately 176 genes exhibited expression trajectories that were highly correlated with mortality rates (Figures 2I and 2J). Many genes showed age-dependent expression trajectories whose rate of change were reduced by CR and changed at a rate roughly proportional to the physiological age of the cohorts (Figure 2; bottom). A small number of genes, however, exhibited trajectories that were predominantly determined by chronological time and showed similar rates of change in both treatments. The transcript levels of these genes may be useful for markers of chronological age, a phenotypic surrogate of which has not been discovered in flies (Figure 2N) [12].

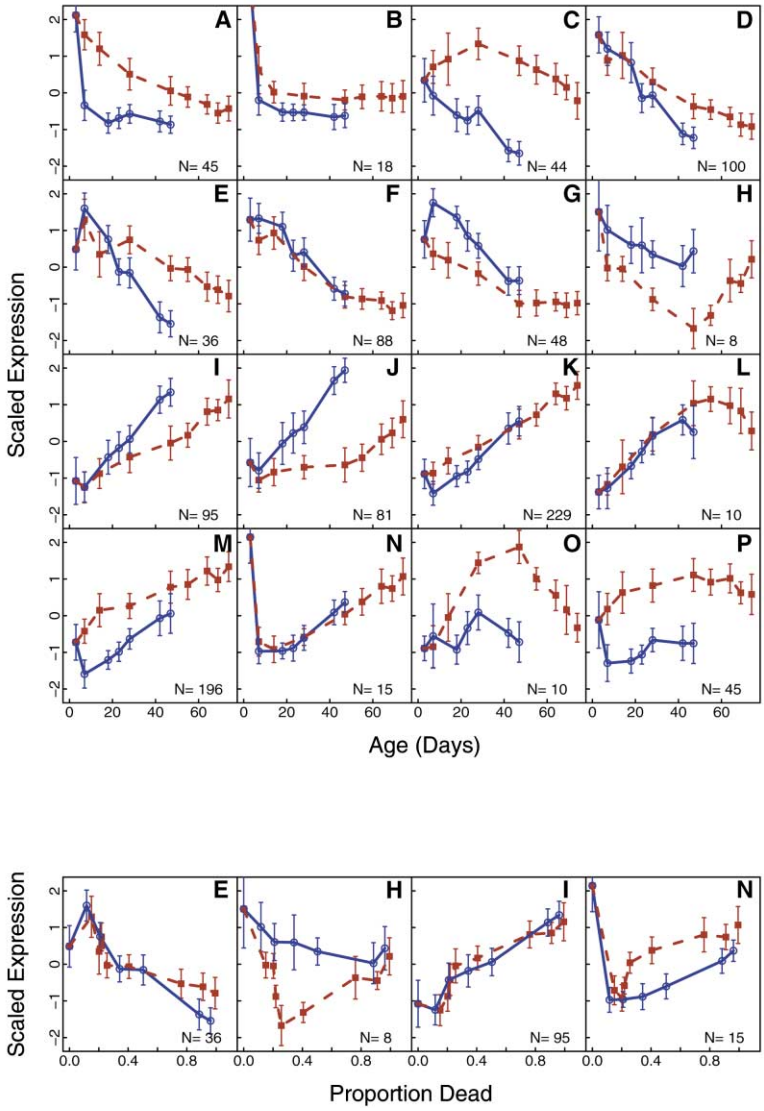

Figure 2. A Sample of Trajectories of Transcript Representation Found to Change Significantly with Age in Control or Calorically Restricted Conditions

(A–P) Groupings were trimmed from a hierarchically clustered tree. Blue lines indicate expression trajectories, averaged over the genes in that grouping, under control food conditions, and red lines indicate profiles under CR conditions. "N" is the number of expression profiles contributing to the observed pattern, and error bars represent the standard deviation of the mean expression. Before clustering, expression measures were normalized to have a mean of 0 and a standard deviation of 1 across both treatments. Normalized trajectories were then plotted as scaled expression values. Top: expression profiles are presented as a function of chronological time. Bottom: a select group of expression profiles are presented as a function of cohort survivorship.

Although many of the age-dependent profiles were nonlinear, a large subset of changes could be described as tending to increase or decrease with age. In this subset, most age-dependent changes are less than 4-fold, although several (18 in CR and 42 in the control) exhibit age-dependent changes greater than 16-fold (Figure 3). The majority of changes in transcript representation were increases in both treatments (62% of the transcripts were upregulated with age in at least one treatment, while 38% were downregulated) (Figure 4). In most cases, an increase (decrease) in one treatment was associated with an increase (decrease) in the other. A large number of age-dependent changes in the control treatment were ameliorated under caloric restriction (130 probe sets show upregulation in the control and no change in CR, while 90 show downregulation in control and no change in CR). However, an equally large number of changes in CR were not seen in the control (Figure 4).

We identified another 2188 probe sets representing 2079 genes that did not change with age but were either up- or downregulated in the CR treatment. Transcript representation of these genes responded immediately (within 3 days) to low-calorie medium, and the difference persisted throughout the life span. The average magnitude of these changes was much smaller than those observed as a function of age, with no significant changes over 4-fold (Figure 5). The changes are distributed roughly equally between up- and downregulation (44% of the total are upregulated). Nearly all (98%) of the significantly downregulated probe sets show less than a 1.5-fold change, while roughly 40% of the significantly upregulated genes show a 1.5-fold increase or greater (Figure 5).

Age-Dependent Expression Profiles Carry Signatures of Senescence

The qualitative similarity of expression profiles in the control and CR treatments (Figures 2; bottom, and 4) suggested that CR extends life span by ameliorating many of the normal transcriptional changes that occur

Table 1. Molecular Function Ontology of Probe Sets with Significant Age-Dependent Transcript Representation

Functional Ontology	GO Code	Probe Sets					
		Chip	Signature	Down		Up	
Defense/immunity protein	GO:0003793	45	13*	0	(0)	8	(62)
Antimicrobial peptide	GO:0003795	38	12*	0	(0)	8	(67)
Antibacterial peptide	GO:0003797	21	8*	0	(0)	4	(50)
Enzyme	GO:0003824	2937	375**	81	(22)	153	(41)
Hydrolase	GO:0016787	1403	180**	43	(24)	71	(39)
Acting on glycosyl bonds	GO:0016798	96	23**	1	(4)	17	(74)
Hydrolyzing O-glycosyl compounds	GO:0004553	92	23**	1	(4)	17	(74)
Chitinase	GO:0004568	18	7*	0	(0)	6	(86)
Glucosidase	GO:0015926	19	10**	0	(0)	8	(80)
Alpha-glucosidase	GO:0004558	9	8**	0	(0)	7	(88)
Peptidase	GO:0008233	476	73**	22	(30)	25	(34)
Serine-type peptidase	GO:0008236	216	38*	14	(37)	7	(18)
Serine-type endopeptidase							
Chymotrypsin	GO:0004263	9	6*	4	(67)	0	(0)
Oxidoreductase	GO:0016491	365	82**	14	(17)	27	(33)
Acting on diphenols and related substances as donors							
With cytochrome as acceptor	GO:0016681	8	5*	3	(60)	0	(0)
Ubiquinol-cytochrome-c reductase	GO:0008121	8	5*	3	(60)	0	(0)
Acting on CH-OH group of donors	GO:0016614	107	26**	3	(12)	11	(42)
With NAD or NADP as acceptor	GO:0016616	77	20**	2	(10)	7	(35)
Acting on NADH or NADPH	GO:0016651	45	13*	2	(15)	3	(24)
Transferase							
Transferring alkyl or aryl groups other than methyl group	GO:0016766	56	15*	3	(20)	10	(67)
Transferring pentosyl groups	GO:0016763	21	8*	2	(25)	6	(75)
Purine-nucleoside phosphorylase	GO:0004731	5	4*	0	(0)	4	(100)
Enzyme regulator	GO:0030234	144	27*	0	(0)	21	(78)
Enzyme inhibitor	GO:0004857	79	20**	0	(0)	17	(85)
Proteinase inhibitor	GO:0004866	52	17**	0	(0)	16	(94)
Serine protease inhibitor	GO:0004867	37	11*	0	(0)	10	(91)
Ligand binding or carrier	GO:0005488	1126	142*	24	(17)	61	(43)
Nucleic acid binding	GO:0003676	1042	52**	25	(48)	6	(12)
DNA binding	GO:0003677	689	26**	9	(35)	6	(23)
Transcription factor	GO:0003700	524	20**	6	(30)	6	(30)
RNA polymerase II transcription factor	GO:0003702	307	12*	5	(42)	2	(17)
Signal transducer							
Receptor							
Peptidoglycan recognition	GO:0016019	14	7*	0	(0)	5	(71)
Structural protein							
Structural protein of chorion (sensu Insecta)	GO:0005213	7	6**	5	(83)	0	(0)
Transporter	GO:0005215	826	112**	25	(22)	40	(36)
Carrier	GO:0005386	434	66*	20	(30)	12	(18)
Primary active transporter	GO:0015399	215	37*	14	(38)	5	(13)
Ubiquinol-cytochrome-c reductase	GO:0008121	8	5*	3	(60)	0	(0)
Ion transporter	GO:0015075	179	33*	9	(27)	5	(15)
Cation transporter	GO:0008324	142	29*	8	(28)	5	(17)
Monovalent inorganic cation transporter	GO:0015077	74	20**	8	(40)	0	(0)
Hydrogen transporter	GO:0015078	74	20**	8	(40)	0	(0)
Ubiquinol-cytochrome-c reductase	GO:0008121	8	5*	3	(60)	0	(0)
Oxygen transporter	GO:0005344	90	25**	2	(8)	17	(68)
Cytochrome P450	GO:0015034	90	25**	2	(8)	17	(68)

Gene ontology (GO) numbers and short descriptions are provided for groups that show significant overrepresentation (bold) or underrepresentation (italics) in the 1264 probe sets representing the 1203 signature genes. Data comprise the number of probe sets in each category present on the entire chip; the number present in the set of signature genes; and the number of signature genes in that category that are strongly down- or upregulated with age in both experimental treatments. Percentages of total are in parentheses. A double asterisk indicates a highly significant association between the functional group and age-dependent changes (i.e., $p < 1.9 \times 10^{-5}$, Bonferonni correction). A single asterisk indicates a significant association between the functional group and age-dependent transcriptional changes ($p < 3.7 \times 10^{-4}$, a level giving one false positive).

with age. Using the expression profiles of the 1203 genes that changed expression significantly with age in one or the other treatment, we constructed a hierarchical clustering with respect to age and treatment. Despite a nearly 2-fold increase in average life expectancy in the CR conditions, transcript levels were more similar among *relatively* young or old flies (i.e., when age is expressed as a fraction of mean life span) than they were among flies of similar chronological age or similar environmental conditions (Figure 6A). The expression levels of these genes were strongly indicative of the physiological age of the cohort, and therefore the majority of these genes carried signatures of senescence in their age-dependent patterns of transcript representa-

Table 2. Biological Process Ontology of Probe Sets with Significant Age-Dependent Transcript Representation

Process Ontology	GO Code	Probe Sets					
		Chip	Signature	Down		Up	
Cell growth and maintenance							
Metabolism							
Protein metabolism and modification	GO:0006411	757	42*	14	(33)	9	(21)
Transport							
Protein transport	GO:0015031	170	4*	1	(25)	2	(50)
Stress response	GO:0006950	110	24*	0	(0)	15	(63)
Response to pest/pathogen/parasite	GO:0009613	69	20**	0	(0)	13	(65)
Humoral defense mechanism	GO:0006959	54	15*	0	(0)	9	(60)
Humoral defense mechanism (sensu Invertebrata)	GO:0016065	54	15*	0	(0)	9	(60)
Antimicrobial humoral response	GO:0006960	51	15*	0	(0)	9	(60)
Antibacterial humoral response	GO:0006961	29	12**	0	(0)	6	(50)
Cell communication							
Response to external stimulus							
Response to biotic stimulus	GO:0009607	154	35**	0	(0)	22	(63)
Defense response	GO:0006952	139	33**	0	(0)	20	(61)
Immune response	GO:0006955	69	23**	0	(0)	15	(65)
Perception of biotic stimulus	GO:0009595	6	5*	0	(0)	4	(80)
Perception of pest/pathogen/parasite	GO:0009596	6	5*	0	(0)	4	(80)
Perception of bacteria	GO:0016045	6	5*	0	(0)	4	(80)
Response to bacteria	GO:0009617	35	17**	0	(0)	10	(59)
Developmental processes							
Gametogenesis							
Oogenesis							
Eggshell formation	GO:0007304	8	6**	6	(100)	0	(0)
Insect chorion formation	GO:0007306	7	6**	6	(100)	0	(0)

Presentation is as in Table 1. A double asterisk indicates a highly significant association between the functional group and age-dependent changes (i.e., $p < 2.6 \times 10^{-5}$, Bonferonni correction). A single asterisk indicates significant association between the functional group and age-dependent transcriptional change ($p < 5.2 \times 10^{-4}$, a level giving one false positive).

tion. That not all genes carried such signatures is evident when the clustering involves the 2079 genes that were significantly up- or downregulated in CR conditions but did not change with age (Figure 6B) or when all genes are considered (data not shown). In the former, expression patterns were characteristic of dietary conditions, while the entire structure of the tree is unstable and variable across algorithms when all genes are included.

Genes with Age-Dependent Transcription

The sheer number of potentially important age-dependent changes means that generating an efficient and detailed description of the changes on a gene-by-gene basis is extremely difficult, if not impossible. The prevalence of transcriptional modifications, however, allows a classification of the changes into reasonably sized groups, which are based on the biological relationships among the genes and their products. Comparing the distribution of age- and diet-dependent changes across those groups to the distribution across the genome as a whole allows the identification of significant patterns in a statistically rigorous way (see the Experimental Procedures). Once these patterns are identified, closer examination of potentially interesting genes becomes feasible.

Using the Gene Ontology, we identified several molecular functions significantly overrepresented in the set of 1264 signature probe sets that were identified as changing with age in either the control or calorically restricted treatments (Table 1). Some changes are expected to result from known processes of physiological degeneration such as cessation of egg production, while others are likely to be adaptive responses to aging-induced damage and deterioration. Nearly 86% (6/7) of the probe sets involved in chorion formation (GO:0005213) were found to change with age, and of those that changed with age, 5/6 declined in both treatments. None were seen to increase with age. This clearly reflects reproductive senescence in females. Three of the chorion protein genes (*CP18*, *CP36*, and *CP38*) declined more slowly in CR females, while *CP16* and *CP19* exhibited similar dynamics in both treatments. A large group of enzyme inhibitors, primarily serine protease inhibitors (GO:0004867), were found to increase in transcript representation with age. Of those in the signature class, 85% were found to increase significantly in both control and CR treatments. The functions of cytochrome P450s and antibacterial peptides are associated with the aging process, as a significant number of these genes increase transcript representation with age. Surprisingly, several transcription factors, especially verified RNA polymerase II transcription factors (GO:0003702), were significantly *underrepresented* in the signature class, which strongly suggests that the transcriptional representation of these genes does not change with age and is actively maintained at fixed levels throughout life.

Although the biological process ontology is less well developed and contains fewer annotated genes, there are several striking patterns that emerge from our analysis (Table 2). First, nearly all forms of stress response (GO:0006950), including response to pathogens (primarily antibacterial), are overrepresented in the signature class. Furthermore, these probe sets exclusively increase in transcript representation with age. Similar behavior is seen in the genes involved in the perception of

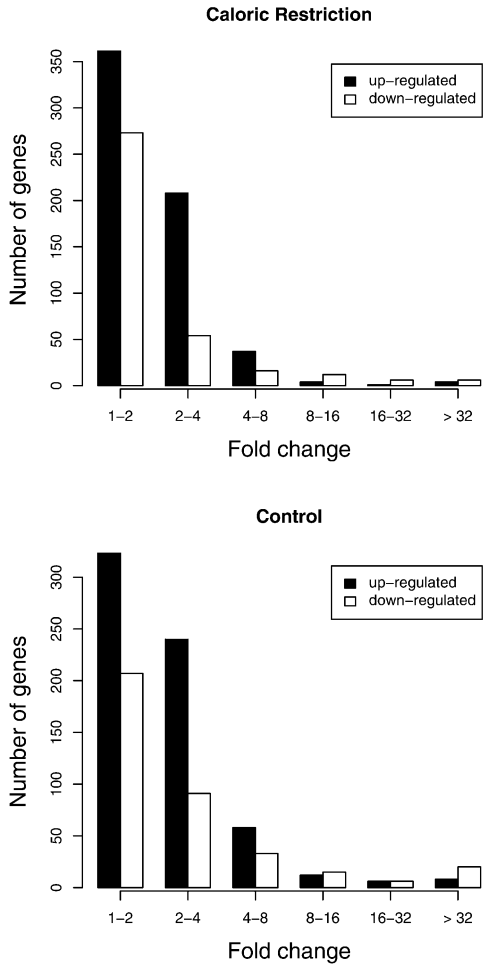

Figure 3. Fold-Change Data for Genes with Age-Dependent Transcript Representation

Distribution of fold-change measures across age for genes that tend to increase or decrease with age in the caloric restriction (top) and control (bottom) conditions. Fold change was calculated using the least squares estimated regression line describing the rate of change in transcriptional representation with age.

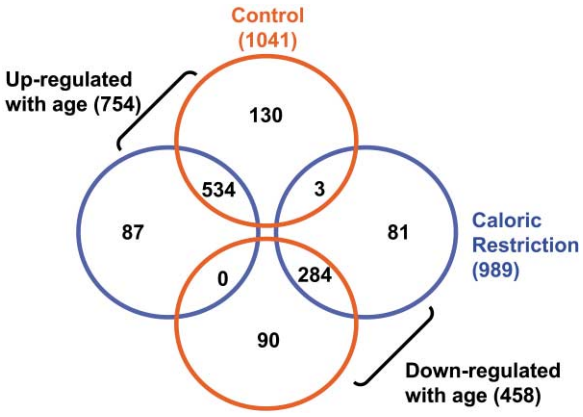

Figure 4. Relationships among Sets of Genes with Age-Dependent Transcript Representation

A Venn diagram describing the numbers of probe sets that are upregulated or downregulated with age in the caloric restriction (blue) and control (red) treatments. Because probe sets were previously established as exhibiting age-dependent transcript representation, significant increases or decreases in expression were based on a p value reflecting only the 1264 age-dependent probe sets.

biotic stimulus (GO:0009595), such as bacteria. Several peptidoglycan recognition proteins (PGRPs) show highly significant, 2- to 3-fold increases with age in both treatments. In addition to perception of an external stimulus, transcript levels of genes involved in the response to external stimulus, especially response to bacteria (GO:0009617), are highly influenced by age. Again, some of the PGRPs are also present in this group, along with well-known immune defense genes: Cecropins, Attackins, *Defensin*, and *Relish*. Nearly all genes involved in protein metabolism and modification (GO:0006411) maintained relatively constant levels of transcription through the life span. These may have their transcription strongly regulated.

There are also significant nonrandom associations between age-dependent changes in transcript representation and genes whose product is localized to a particular cellular structure or gene product group (see Table S1 in the Supplementary Material available with this article online). The most significant is the overrepresentation of genes whose product is localized to the mitochondrion, especially those found on the inner membrane and involved in electron transport (GO:0005746). This highly represented group, which includes several ubiquinol-cytochrome-c reductases, shows strong declines in expression with age; as do genes whose products localize to the nucleolus (GO:0005730).

Genes with Diet-Dependent Transcription

A total of 2188 probe sets exhibit significant up- or downregulation within 3 days of transfer to calorically restricted medium, followed by relatively constant expression thereafter. Several trends are apparent in the GO analysis (Table 3, and Tables S2 and S3 in the Supplementary Material).

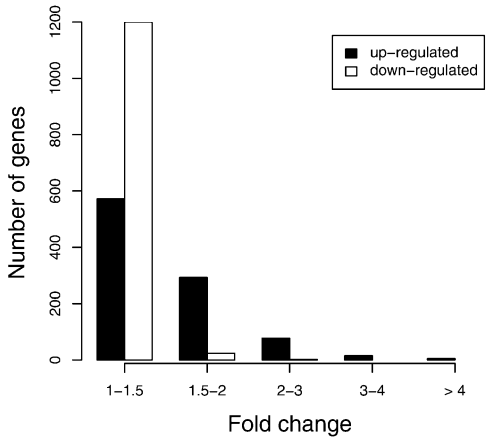

Figure 5. Fold-Change Data for Genes with Diet-Dependent Transcript Representation

Distribution of fold-change measures for the 2188 probe sets that do not change with age but show significant up- or downregulation in the caloric restriction environment.

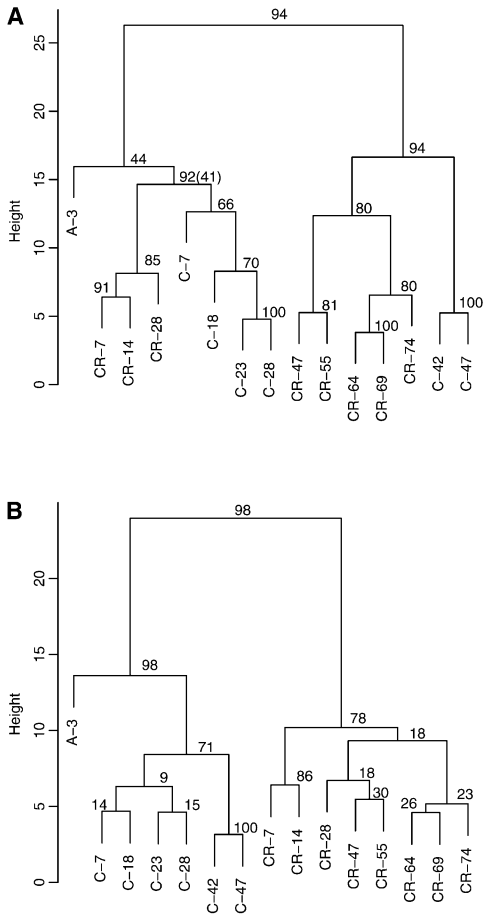

Figure 6. Age-Dependent Transcript Profiles Distinguish Young Flies from Old Flies and Provide Signatures of Senescence

(A and B) Results of experiment-based, hierarchical clustering obtained from a divisive algorithm (28). Age 3 (A-3), control (C), and calorically restricted (CR) treatments are presented for each of the sampled ages in the experiment. Terminal branches are labeled with the treatment (C or CR) followed by the age (in days) of the sample used for expression analysis. With the exception of age 3 days (A-3) and age 14 days in CR conditions (CR-14), each combination of age and treatment includes data from at least 5 replicate chips. Numbers are bootstrap confidence estimates of the given node (given as the percentage of 5000 bootstrapped hierarchical clusters in which the given node is intact [24]). Expression data used for clustering included either: (A) The 1203 signature genes that were found to change with age in either the control or CR treatments, or (B) the 2079 genes that were significantly up- or downregulated in CR conditions but did not change with age. In (A), the nodes supporting division into relatively old and young flies (i.e., when age is scaled to mean life span) in the two treatments are highly supported (94% and 92%), indicating that signatures of senescence are robust. The least stable node in the tree is the separation of age 3 days and other early-age samples, which remained intact in 44% of the bootstrapped samples. This is due to the similarity between this age (A-3) and the early-age control branches (C-7 to C-28), which misleadingly reduced confidence in the branch separating early-age CR flies from controls (to 41%, as indicated in parentheses). When age 3 was removed from the data set, the node separating early-age CR and control branches was highly supported (92% confidence). The failure to achieve near 100% stability in those nodes is indicative of a subset of genes that scale with chronological time (e.g., Figure 3N).

Genes classified as DNA repair proteins (GO:0006281) and DNA replication factors (GO:0006260), along with genes whose products are involved in cell cycle control, DNA replication, chromosome condensation, chromosome segregation, and other cell cycle processes (GO:0007049), are significantly downregulated in reduced-calorie conditions. Localization of products from downregulated genes strongly identifies the nucleus as a primary site of downregulation, involving genes associated with the formation of the replication fork (GO:0005657) and the chromatin assembly complex (GO:0005678). Many developmental processes, including nearly all aspects of oogenesis (GO:0007292) and the candidate aging gene, *methuselah*, are significantly repressed in the reduced-calorie environment (Table 3). Protein metabolism and ubiquitin-dependent protein degradation (GO:0006511) is downregulated, despite protein biosynthesis (GO:0006412) being significantly unchanged.

Although the number of genes upregulated under caloric restriction is roughly equivalent to the number that are downregulated, the *distribution* of those changes is very different. Downregulation is primarily directed to groups of genes involved in similar biological processes (Table 3). Upregulation, however, is much less focused, with a small number of genes involved in response to external stimulus showing significant diet-mediated expression. This observation implies that the large numbers of upregulated genes are essentially randomly distributed with respect to function. Because, under the Affymetrix protocol, expression measures are proportional to the total mRNA pool, these results are consistent with a cellular response to a low-calorie environment that primary involves massive downregulation of a large number of genes related to cell growth and maintenance or to developmental processes (Table 3). As a consequence of the targeted downregulation and its impact on the mRNA pool, most genes whose absolute transcript abundance remains unchanged by the treatment will appear to be upregulated, perhaps to a large extent (e.g., Figure 5). Targeted measures of particular mRNA transcripts in individual flies, using sensitive techniques such as real time PCR, are needed to examine this possibility.

Genomic Hypotheses of Aging

We used our microarray data to evaluate two hypotheses that are related to aging and that are posed explicitly at the level of whole-genome expression patterns. The first is that age-dependent expression patterns are mediated by changes in transcriptional silencing through modification of chromosomal architecture and organization, and the second is that aging and death result from a large-scale break down in homeostasis and concomitant disregulation of gene expression in old individuals.

Chromosomal Position of Signature Genes Is Random

In eukaryotes, transcriptional modification of gene expression is influenced by specific proteins acting at individual promoters as well as by more general phenomena including regional activation and inactivation of chromosomal domains [13] and availability of proteins involved

Table 3. Biological Process Ontology of Probe Sets with Significant Up- or Downregulation under Calorically Restricted Conditions

Process Ontology	GO Code	Probe Sets	
		Chip	CR Class
Upregulated with CR			
Cell growth and maintenance	GO:0008151	1890	89**
Cell cycle	GO:0007049	157	2*
Metabolism			
Biosynthesis			
Protein biosynthesis	GO:0006412	240	2**
Nucleobase, nucleoside, nucleotide, nucleic acid metabolism	GO:0006139	390	7**
RNA metabolism	GO:0016070	130	1*
RNA processing	GO:0006396	127	1*
Protein metabolism and modification	GO:0006411	757	31**
Cell communication			
Perception of light	GO:0009583	66	14*
Response to abiotic stimulus			
Radiation response	GO:0009314	69	14*
Downregulated with CR			
Cell communication			
Sensory perception	GO:0007600	96	1*
Signal transduction			
G protein-coupled receptor protein signaling pathway	GO:0007186	133	2*
Cell growth and maintenance	GO:0008151	1890	302**
Cell cycle	GO:0007049	157	79**
Cell cycle control	GO:0000074	30	16**
Cell cycle checkpoint	GO:0000075	15	9**
DNA replication and chromosome cycle	GO:0000067	78	47**
Chromosome condensation	GO:0030261	19	13**
Chromosome segregation	GO:0007059	25	13**
M phase	GO:0000279	98	47**
M phase-specific microtubule process	GO:0000072	16	8*
Nuclear division	GO:0000280	73	38**
Meiosis	GO:0007126	35	16**
Female meiosis	GO:0007143	23	12**
Mitosis	GO:0007067	51	29**
Control of mitosis	GO:0007088	11	8**
Mitotic anaphase	GO:000090	5	4*
Mitotic prophase	GO:0000088	15	11**
Mitotic chromosome condensation	GO:0007076	15	11**
Sister chromatid cohesion	GO:0007062	9	6*
Mitotic cell cycle	GO:0000278	94	56**
M phase of mitotic cell cycle	GO:0000087	51	29**
S phase of mitotic cell cycle	GO:0000084	37	26**
Cell organization and biogenesis			
Microtubule-based process	GO:0007017	151	29*
Metabolism	GO:0008152	1293	228**
Protein degredation	GO:0030163	83	19*
Proteolysis and peptidolysis	GO:0006508	83	19*
Ubiquitin-dependent protein degradation	GO:0006511	57	18**
Nucleobase, nucleoside, and nucleic acid metabolism	GO:0006139	390	113**
DNA metabolism	GO:0006259	121	52**
DNA packaging	GO:0006323	38	14**
Establishment/maintenance chromatin architecture	GO:0006325	37	14**
Chromatin assembly/disassembly	GO:0006333	17	8*
DNA repair	GO:0006281	46	15**
DNA replication	GO:0006260	37	26**
Transcription	GO:0006350	118	24*
RNA metabolism	GO:0016070	130	35**
RNA processing	GO:0006396	127	35**
mRNA processing			
RNA splicing	GO:0008380	95	26**
Phosphate metabolism	GO:0006796	376	67**
Phosphorylation	GO:0016310	274	50**
Protein phosphorylation	GO:0006468	237	44**
Protein metabolism and modification	GO:0006411	757	106*
Protein modification	GO:0006464	353	65**

(*continued*)

Table 3. Continued

Process Ontology	GO Code	Probe Sets	
		Chip	CR Class
Developmental processes	GO:0007275	390	69**
Embryogenesis and morphogenesis			
Mitotic cycle control	GO:0007346	12	8**
Gametogenesis	GO:0007276	114	27**
Oogenesis	GO:0007292	96	23*
Nurse cell/oocyte transport	GO:0007300	8	5*
Oocyte construction	GO:0007308	53	15*
Axis determination	GO:0007309	53	15*
Dorsal/ventral axis determination	GO:0007310	27	12**

Gene ontology (GO) numbers and short descriptions are provided for groups that show significant overrepresentation (bold) or underrepresentation (italics) in the group of genes that are either upregulated under calorically restricted conditions (top) or downregulated under those conditions (bottom). Data comprise the number of probe sets in each category present on the entire chip and the number of probe sets significantly affected by caloric restriction. A double asterisk indicates a highly significant association between the functional group and age-dependent changes (i.e., $p < 2.6 \times 10^{-5}$, Bonferonni correction). A single asterisk indicates a signficant association between the functional group and age-dependent transcriptioanl changes ($p < 5.2 \times 10^{-4}$, a level giving one false positive). Some intermediate entries in the ontology have been omitted. The complete set of significant entries is available from the corresponding author.

in nucleosome formation [14]. Such factors have also been associated with the aging process through theories implicating telomere shortening [15] and global deacetylation of histones [16]. These lines of evidence suggest the possibility that changes in gene expression with age are mediated through modifications in chromosomal structure and organization. We examined whether genes with age-dependent transcript representation are clustered together, perhaps located preferentially near telomeres, or in some other way nonrandomly distributed throughout the genome. Importantly, a nonrandom distribution would have two possible explanations. It could result from a nonrandom distribution through the genome of gene family members that show age-specific behavior, or it could represent age-dependent structural changes that impact gene expression. We calculated the chromosomal distribution of the 1203 signature genes and tested this distribution against that obtained from the set of genes available from the complete annotated fly sequence (available from FlyBase/GadFly). With the exception of a moderately higher representation of genes in the cytological region 66D10–66F6 (roughly base pairs 8581000– 9044840) and underrepresentation of genes near region 64A2–64B2 (roughly base pairs 3942625– 4174544) on chromosome arm 3L, there was no evidence for age-dependent changes in expression being associated with any particular chromosome arm or domain (such as telomeres) (Figure 7). The overrepresentation in regions 66D10–66F6 is intriguing. However, similar patterns were observed when separately examining genes with expression profiles that either increase or decrease with age (ruling out regional repression or derepression with age), and signature genes in this region were not found to have similar expression profiles.

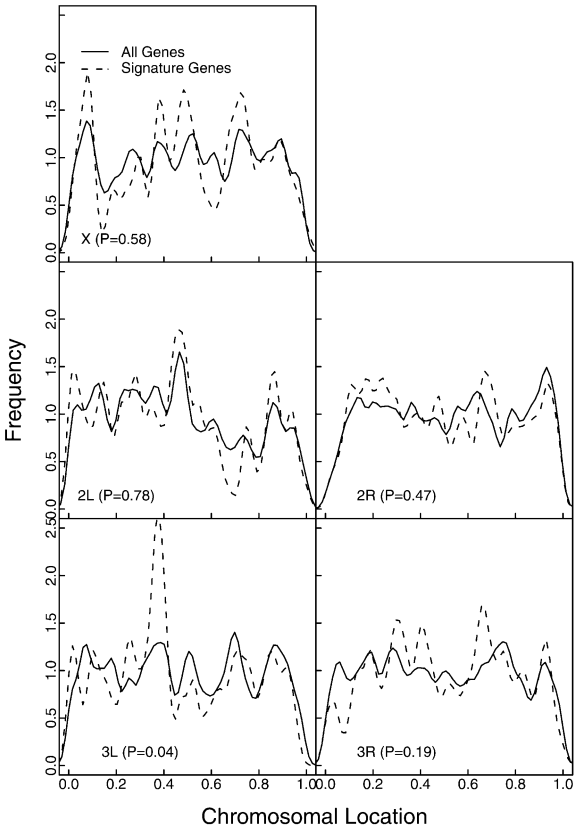

Figure 7. Genomic Distribution of Genes with Age-Dependent Transcript Representation Is Random

Chromosomal arm distributions of gene location from the whole genome sequence (solid lines) and from the location of the 1203 signature genes (dashed lines). Gene location is relative to euchromatin length and was calculated by obtaining the number of the center base in the gene sequence and dividing by the total number of sequenced base pairs on the chromosome arm in question. Sequence data was obtained from GadFly. Distributions were obtained by density estimation via kernel smoothing with a window size of 0.2% of euchromatin length [25]. p values are from a two-sided Kolmogorov-Smirnov test of equality of distributions using the raw (unsmoothed) data.

No Evidence for Genome-Wide Disregulation of Expression

Age-dependent deterioration of genome integrity in various tissues through an increase in the frequency of somatic mutations, oxidative DNA damage, and chromosomal aberrations is thought to precipitate a progressive and global loss of homeostasis and gene regulation

in aging organisms [17]. If true, global disregulation might manifest itself on a gene-by-gene basis as an age-dependent increase in the among-individual variance of expression. Our experimental design provides a test of this hypothesis by allowing an estimate of the rate of change with age in biologically derived variance in gene expression for each of the genes on the oligonucleotide array. If system-level gene disregulation is occurring, the distribution of rates of change would be predominantly positive, as the average gene would be expected to experience an increased variability in expression among individuals as a function of age. In both control and CR flies, however, the distribution of rates of change is centered directly on zero, with half of the genes experiencing an increase in among-individual variance and half experiencing a decrease (data not shown). This is true whether the analysis includes all probe sets or just those representing the signature genes. We should emphasize that, if disregulation were occurring in a cell-autonomous fashion, then we would probably not detect it from our whole-organism analysis, as random cell-to-cell variation would combine to greatly reduce the detectability of any increase in variance in transcript representation with age. To the extent of our analysis, these results provide genome level support to the conclusions of Rogina et al. [18], who found no increase in variance in six genes expressed in the *Drosophila* antennae.

Conclusions

Aging involves a highly dynamic system, in which alterations in gene expression are widespread and complicated. Despite their complexity, the many and varied age-dependent expression patterns contain robust signatures of senescence and provide a powerful framework for a molecular classification of aging. A detailed comparison of expression profiles confirms that, at the level of the transcriptome, calorically restricted treatments have many of the properties of normal aging, with the exception that they occur over a longer period of time. In the same way that we were able to show that caloric restriction acts predominantly through a slowing of many of the physiological changes that normally occur with age [6], the signatures of senescence provided by expression profiling could be used to discriminate between an increase in the rate of normal aging, i.e., a progeria, and interventions that make the animals short lived for other reasons, thereby identifying legitimate short-lived model systems. Further consideration of the statistical patterns of age-dependent changes in transcript representation enabled a refutation of two common hypotheses concerning aging: genomic localization and genome disregulation. Based on the chromosomal distribution of genes carrying signatures of senescence, we find no evidence relating age-dependent changes in transcript level to genes located in specific regions of the chromosome, such as telomeres, and we find no support for systemic gene disregulation with advancing age.

Our functional analysis of genes with age- and diet-dependent transcript representation identified several well-known metabolic pathways impacted by aging and the response to caloric restriction, and it suggests several novel directions for future research. Response to general stresses is evident in aging flies, as is upregulation of several families of cytochrome P450s, which catalyze a variety of reactions in eukaryotes, including those involved in drug metabolism. Age-dependent declines in the transcription of genes expressed in germ line tissue clearly illustrate reproductive senescence. Caloric restriction results in rapid downregulation of large numbers of genes whose product is localized to the nucleus and that are involved in cell growth and maintenance. Surprisingly, one of the most dramatic transcriptional changes in aging flies is associated with response to microbial infection and involves highly increased expression of several antimicrobial genes. This defense response exhibits a component that is correlated with the health and physiological age of the animals and a component that is independent of these factors, and it suggests the intriguing possibility that infection is a primary cause of death in laboratory *Drosophila* cultures.

Experimental Procedures

Aging Assay

Wild-type Dahomey flies were used for the experiment. Since their inception in 1970, source populations have been maintained in population cages with overlapping generations, and they exhibit similar life spans to newly caught, wild populations [19]. Experimental flies were collected as virgins using light CO_2 anesthesia over a period of approximately 18 hr, following two generations of controlled breeding under constant larval density. Following collection, males and females were placed together on 10% yeast/glucose medium and allowed to mate for 36 hr. After mating, males were removed, and females were placed on either 15% yeast/glucose medium (control conditions) or 5% yeast/glucose medium (CR conditions) in standard vials at 10 females/vial. Initial cohort sizes were 2500 and 3000 females for the control and CR treatments, respectively. Deaths were scored daily, and flies were transferred to fresh medium and were randomly redistributed to 10 flies/vial three times per week. For expression analysis, samples of 50 flies were randomly chosen, flash frozen in liquid nitrogen in groups of 5, and transferred to $-80^{\circ}C$. Samples were taken at the following ages (in days): 7, 18, 23, 28, 42, and 47 (control); 7, 14, 28, 47, 55, 64, 69, and 74 (CR). In a second, identical experiment, a sample was taken just after mating and before transfer to treatment media (3 days of age). Fly media recipes are available upon request.

Experimental Design

With the exception of age 3 and CR age 14, at least 5 replicate Affymetrix GeneChips were hybridized for each sample point using standard Affymetrix procedures. These five replicates were derived from three independent RNA extractions. Two of the three RNA extractions provided total RNA for two independent cDNA and cRNA reactions, and each cRNA reaction provided the material for a single chip. The third RNA extraction provided material for a single cDNA and cRNA synthesis and a single chip. The order of all experimental protocols was randomized such that two chips from the same age and treatment were not more likely to experience correlation in gene expression through laboratory procedures executed at the same time. Technical concerns resulted in one additional replicate added to ages 42 and 47 in the control treatment, and only 2 chips for age 3 days and for age 14 days in the CR treatment. The total number of gene chips in the experiment was therefore 71.

Statistical Analyses

The Affymetrix roDROMEGa array contained 14,028 *Drosophila* probe sets representing 12,975 genes (many genes were represented by more than one probe set). Expression measures were normalized and calculated using dCHIP (http://www.dchip.org) [20], which incorporates a linear model-based analysis.

Following the calculation of expression measures, the data were exposed to a series of weighted regression (including linear, quadratic, and cubic regressions) and ANOVA models. The models were nested to insure proper inference from the three independent RNA extractions, which led to the five replicate chips. Since expression values for many genes were very different at age 3 days than they were at other ages, statistical analyses were done including and excluding this observation point. Expression values were weighted by the inverse of their standard deviations. Statistical significance for individual probe sets was determined via a step-down Bonferonni correction and permutation analysis to maintain an experimentwise $p = 0.05$ [21]. Genes that do not change with age but are significantly different between control and CR conditions were identified using nested and weighted ANOVA with treatment as a fixed effect.

Clustering was carried out using a divisive clustering algorithm ("diana") implemented in the *cluster* package of the free statistical programming language R (http://www.r-project.org). The topology of the tree was robust to the algorithm used (i.e., agglomerative methods and divisive methods produced the same structure). For each age and treatment, we calculated the weighted average expression for each gene of interest (weights were the inverse of the standard deviation of expression obtained from the linear model fitting), and the resulting average expression profiles were input into the clustering algorithms.

Estimates of the rate of increase in biologically based variance in expression with age were obtained through the use of statistical techniques designed for the genetic analysis of function-valued characters [22]. In short, linear variance functions were fitted, using maximum likelihood, to the expression data from each gene in each treatment. The between-RNA-extraction variance function is proportional to the variance in expression among individual flies (scaled by the number of flies in the extraction). Therefore, a significant age-dependent change in between-RNA-extraction variance is due to change in the among-individual variance of expression. The significance of the slope term in the variance was determined using a likelihood ratio test against a model with no change in variance with age.

Gene Ontology Analyses

We annotated every probe set on our GeneChip by integrating the information (as of January 28, 2002) on the gene ontology (GO) web site (http://www.geneontology.org) with the information available from FlyBase (http://www.fruitfly.org). We first associated each probe set on the custom array with its corresponding current curated gene entry in FlyBase. Next, each gene was queried for available GO annotations (i.e., GO numbers) for molecular function, biological process, and cellular component. We excluded annotations derived exclusively from electronic annotation (evidence code IEA).

The GO is technically a directed acyclic graph, which resembles a hierarchical structure in which an individual entry may have two or more parents. We appealed to this structure to determine the number of genes that were annotated as belonging to a particular GO term and the number of genes belonging to that term and its children. For example, on our chip, the GO term *antibacterial peptide* (GO:0003797) in the function hierarchy contained 21 probe sets annotated at that level or below, which includes the terms *gram-negative antibacterial peptide* (GO:0008225), *gram-positive antibacterial peptide* (GO:0008224), and *male-specific antibacterial peptide* (GO:0003798). For a target group phenomenon (e.g., expression levels that change significantly with age), we first isolated the probe sets belonging to that group (call this a focus group). In the case of genes with age-dependent expression, the number of probe sets was 1264, which is approximately 9% of the probe sets on the chip. We then compared the number of probe sets associated at and below a specific GO term in the focus group with the analogous number for the entire chip. If the focus group comprises 9% of the whole genome, then under the null hypothesis of no relationship between age-dependent expression and biological function, we expect (for each GO term) 9/100 as many probe sets in the focus group as there are on the entire chip. To the extent that the observed number of probe sets in the focus group is significantly greater than the expected number, the genes with that GO term are considered to be highly associated with the target phenomenon. Continuing with the example of *antibacterial peptide*, we find 21 probe sets on the chip, and 8 represented in the signature class (Table 1). The p value associated with the null hypothesis of no association is obtained from the binomial distribution with 21 tries, 0.090 probability of success, and ≥ 8 successes ($p = 0.0003$). Recognizing that the hypothesis tests are not independent, we considered two corrections for multiple testing: a strict Bonferonni, and a second correction coinciding with the expectation of one false positive.

Supplementary Material

Supplementary Material including results of the cellular component ontology analyses for genes with age- and diet-dependent transcript representation and results of the molecular function ontology analysis for genes with diet-dependent transcript representation is available at http://images.cellpress.com/supmat/supmatin.htm.

Acknowledgments

We thank Martin Strahm for his efforts in the design of the custom array. L.P acknowledges the National Environmental Research Council (UK) for support. Three anonymous reviewers provided valuable comments. This research was supported by Biotechnology and Biological Sciences Research Council grant 31/G12125 to D.B.G. and L.P. D.B.G. is a Royal Society Wolfson Research Merit Award Holder.

Received: February 24, 2002
Revised: February 7, 2002
Accepted: March 1, 2002
Published: April 30, 2002

References

1. Guarente, L., and Kenyon, C. (2000). Genetic pathways that regulate ageing in model organisms. Nature *408*, 255–262.
2. Miller, R.A. (2001). Biomarkers and the genetics of aging in mice. In Cells and Surveys, C.E. Finch, J.W. Vauple, and K. Kinsella, eds. (Washington, D.C.: National Academy Press), pp. 180–213.
3. Jiang, J.C., Jaruga, E., Repnevskaya, M.V., and Jazwinski, S.M. (2000). An intervention resembling caloric restriction prolongs life span and retards aging in yeast. FASEB J. *14*, 2135–2137.
4. Braeckman, B.P., Houthoofd, K., and Vanfleteren, J.R. (2001). Insulin-like signaling, metabolism, stress resistance and aging in *Caenorhabditis elegans*. Mech. Ageing Dev. *122*, 673–693.
5. Chapman, T., and Partridge, L. (1996). Female fitness in *Drosophila melanogaster*: an interaction between the effect of nutrition and of encounter rates with males. Proc. R. Soc. Lond. B Biol. Sci. *263*, 755–759.
6. Yu, B.P., Masaro, E., and Mcmahan, C.A. (1985). Nutritional influences on aging of Fischer 344 rats: physical, metabolic and longevity characteristics. J. Gerontol. *40*, 657–670.
7. Cao, S.X., Dhahbi, J.M., Mote, P.L., and Spindler, S.R. (2001). Genomic profiling of short- and long-term caloric restriction effects in the liver of aging mice. Proc. Natl. Acad. Sci. USA *98*, 10630–10635.
8. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29.
9. Zou, S., Meadows, S., Sharp, L., Jan, L.Y., and Jan, Y.N. (2000). Genome-wide study of aging and oxidative stress response in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA *97*, 13726–13731.
10. Lee, C.-K., Weindruch, R., and Prolla, T.A. (2000). Gene-expression profile of the ageing brain in mice. Nat. Genet. *25*, 294–297.
11. Kayo, T., Allison, D.B., Weindruch, R., and Prolla, T.A. (2001). Influences of aging and caloric restriction on the transcriptional profile of skeletal muscle from rhesus monkeys. Proc. Natl. Acad. Sci. USA *98*, 5093–5098.
12. Rogina, B., and Helfand, S.L. (1995). Regulation of gene expression is linked to life span in adult *Drosophila*. Genetics *141*, 1043–1048.
13. Cohen, B.A., Mitra, R.D., Hughes, J.D., and Church, G.M. (2000).

A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nat. Genet. *26*, 183–186.

14. Vogelauer, M., Wu, J., Suka, N., and Grunstein, M. (2000). Global histone acetylation and deacetylation in yeast. Nature *408*, 495–498.
15. Rudolph, K.L., Chang, S., Lee, H.W., Blasco, M., Gottlieb, G.J., Greider, C., and Depinho, R.A. (1999). Longevity, stress response, and cancer in aging telomerase-deficient mice. Cell *96*, 701–712.
16. Imai, S., Armstrong, C.M., Kaeberlein, M., and Guarente, L. (2000). Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. Nature *403*, 795–800.
17. Finch, C. (1990). Longevity, Senescence and the Genome (Chicago: University of Chicago Press).
18. Rogina, B., Vaupel, J.W., Partridge, L., and Helfand, S.L. (1998). Regulation of gene expression is preserved in aging *Drosophila melanogaster*. Curr. Biol. *8*, 475–478.
19. Sgro, C.M., and Partridge, L. (1999). A delayed wave of death from reproduction in *Drosophila*. Science *286*, 2521–2524.
20. Li, C., and Wong, W.H. (2001). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc. Natl. Acad. Sci. USA *98*, 31–36.
21. Dudoit, S., Yang, Y.H., Speed, T.P., and Callow, M.J. (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Statistica Sinica, in press.
22. Pletcher, S.D., and Geyer, C.J. (1999). The genetic analysis of age-dependent traits: modeling the character process. Genetics *153*, 825–833.
23. Mueller, H.G., and Wang, J.L. (1994). Hazard rates estimation under random censoring with varying kernel and bandwidths. Biometrics *60*, 61–76.
24. Efron, B., Halloran, E., and Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. Proc. Natl. Acad. Sci. USA *93*, 13429–13434.
25. Bowman, A., and Azzalini, A. (1997). Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations (Oxford: Oxford University Press).