The Journal of Experimental Biology 210, 1607-1612
Published by The Company of Biologists 2007
doi:10.1242/jeb.004887

Common aging pathways in worms, flies, mice and humans

Stuart K. Kim

Department of Developmental Biology and Genetics, Stanford University Medical Center, Stanford, CA 94305-5329, USA

e-mail: kim@cmgm.stanford.edu

Accepted 8 March 2007

Summary

Development of functional genomics tools has made it possible to define the aging process by performing genome-wide scans for transcriptional differences between the young and the old. Global screens for age regulation have been performed for worms and flies, as well as many tissues in mice and humans. Recent work has begun to analyze the similarities and differences in transcriptional changes in aging among different species. Most age-related expression changes are specific for a given species, but genes in one pathway (the electron transport chain pathway) show common age regulation in species from worms to humans. Evolutionary theories of aging provide a basis to understand how age regulation of a genetic pathway might be preserved between distantly related species.

Glossary available online at
http://jeb.biologists.org/cgi/content/full/210/9/1607/DC1

Key words: age regulation, gene expression profile, electron transport chain, gene set enrichment analysis, aging, evolution, genomics.

Introduction

Aging is a complex process driven by diverse molecular pathways and biochemical events. The gradual decline in cellular functions associated with aging is not caused by changes in the expression or activity of just a few individual genes, but rather by the cumulative changes from many genes. To elucidate molecular differences associated with aging, an attractive approach is to use DNA microarrays to scan the entire genome for genes that change expression with age. The set of age-regulated genes provides a comprehensive and unbiased view of molecular changes associated with age. The identities of particular age-regulated genes may suggest specific mechanisms for aging; e.g. decreased expression of the electron transport chain genes in old age suggests changes in energy generation and oxidative damage at the end of life. Furthermore, analysis of the entire set of age-regulated genes (a gene expression profile) can reveal emergent themes about the aging process; e.g. evaluation of all age-regulated genes in normal and calorically restricted mice shows that caloric restriction may slow down age-related expression changes (Park and Prolla, 2005).

Nearly all organisms age, and yet lifespan can be very different between species. For example, among major model organisms, the worm *Caenorhabditis elegans* lives for 2 weeks, the fly *Drosophila melanogaster* lives for 2 months, the mouse *Mus musculus* lives for 2 years and humans lives for ~80 years. A great deal might be learned by comparing the aging process in different species, revealing why humans age so slowly compared with worms. For example, one could ask whether human cells are exceptionally well protected against mitochondrial oxidative damage, DNA damage or telomerase shortening compared with worm cells. Several recent papers have compiled gene expression profiles for aging from multiple species and then compared them to each other to distinguish aspects of aging that are species specific and those that are shared.

There is a rich set of literature on using DNA microarray experiments to profile gene expression differences for aging in worms, flies, mice and humans (Kim, in press). Transcriptional profiles for aging contain quantitative data on age-related changes in expression for a large fraction of the genome. However, relatively few studies have integrated aging transcriptional profiles from different studies in a systematic way to find similarities and differences in aging among different species. Genes that show age-related transcriptional differences in multiple species are exceptionally interesting as biomarkers for age. Their age-related decline scales with lifespan, such that age-related changes occur relatively quickly in short-lived animals but slowly in long-lived ones. By contrast, genes that show age regulation in mice but not humans may help identify pathways and mechanisms that account for much longer lifespan in humans.

Early work by McCarroll and colleagues compared transcriptional changes in flies and worms (McCarroll et al., 2004). This work introduced the concept of treating quantitative changes in gene expression as a molecular phenotype and then comparing different expression profiles to each other to reveal overlaps in expression changes between two experiments. One of the results from this early work was an apparent similarity between aging in flies and worms. However, there were statistical flaws used in the calculation, and it is unclear whether the overlap between flies and worms was statistically significant (Melov and Hubbard, 2004). Furthermore, the work did not really measure aging, as the greatest changes in gene expression occurred in young adulthood and not in old age (McCarroll et al., 2004). Nevertheless, this paper introduced key concepts about using gene expression profiles as molecular phenotypes and set the stage for later papers to generate expression data on aging and statistical methods to analyze the data.

Fraser and colleagues compared the effects of aging on gene expression in the brains of humans and chimpanzees (Fraser et al., 2005). They analyzed previously published data on aging in the human brain, including five different areas of the cerebral cortex (Evans et al., 2003; Khaitovich et al., 2004; Lu et al., 2004). All five areas of the cortex showed similar patterns of age-related expression changes. Next, they measured changes in expression in the brain cortex as a function of age in chimpanzees. They found no correlation between age-related changes in the cortex of humans and chimpanzees, indicating that aging in humans and chimpanzees is very different.

A series of recent papers has compared age-related expression profiles in worms, mice, flies and humans. For worms and flies, DNA microarrays have been used on whole animals over the entire lifespan to profile transcriptional changes of aging (Landis et al., 2004; Lund et al., 2002; Pletcher et al., 2002). For mice, Zahn et al. used data from AGEMAP, which is a large database of expression changes as a function of age in 16 mouse tissues (J. Zahn, unpublished data). For humans, Zahn et al. measured age-related transcriptional changes in muscle and compared them with aging changes in the kidney and the brain (Lu et al., 2004; Rodwell et al., 2004; Zahn et al., 2006).

To compare transcriptional profiles between these four species, Zahn et al. developed new statistical methods for gene set enrichment analysis and empirical meta-analysis (Zahn et al., 2006). These new statistical methods were needed to overcome a pervasive methodological challenge in genomics studies called multiple hypothesis testing. The main strength of DNA microarrays – simultaneous readings of expression from thousands or tens of thousands of genes – is also a major statistical hurdle because the thousands of gene expression measurements make it possible for random events to occur that are very rare (i.e. the problem of multiple hypothesis testing). In a standard experiment testing only one gene, a P value of 0.05 is convincing because there is only a 1 in 19 chance that the result could occur by chance. However, with 1000 genes, a P value threshold of 0.05 is not convincing because about 50 of the 1000 genes can meet this threshold by chance. Hence, DNA microarray experiments nearly always use more stringent P values (e.g. <0.001) in order to screen out most of the data that can occur by chance.

One approach to overcome the statistical hurdles posed by multiple hypothesis testing is to use gene set enrichment analysis (Subramanian et al., 2005; Zahn et al., 2006). To use this method, one first categorizes all of the genes into pathways or gene sets; typically, one uses the gene sets defined by the Gene Ontology consortium (http://www.geneontology.org/), such as the set of ribosomal genes or the set of genes that encode components of the electron transport chain. Then, one looks at the expression of every gene in the gene set to see if there is an overall trend in expression. Do most of the genes increase or decrease with age? This method is both powerful and sensitive because: (1) there are only about ~600 gene sets *versus* ~20 000 genes in the genome, which reduces the problem of multiple hypothesis testing and (2) small changes in expression in many genes in a pathway can become statistically significant when considered as a group. For example, there are 95 genes that encode components of the electron transport pathway. Using gene set enrichment analysis, Zahn et al. showed that there is an overall trend of decreasing expression with age for this set of electron transport genes in worms, flies, mice and humans (Zahn et al., 2006), even though only a small number of these genes show very strong age-related changes by themselves (Fig. 1).

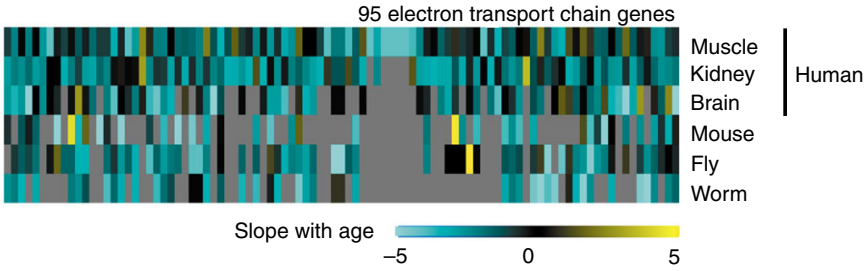

Fig. 1. The electron transport chain decreases expression with age in humans, mice, flies and worms. Rows represent either human tissues or model organisms. Columns correspond to 95 genes in the electron transport chain gene set. Scale represents the normalized slope of the change in $\log_2$ expression level with age. Grey indicates that genes were not present in that species. Original data from Zahn et al. (Zahn et al., 2006).

Empirical meta-analysis is a method to combine gene expression results from different experiments (Zahn et al., 2006). Zahn et al. compiled data on age-related changes in expression from worms, flies, mouse kidneys and human brain, kidney and muscle. The experimental designs used in each of these studies were quite different. The human study used patients with varying ages whereas the worm, fly and mouse studies used staged animals at discrete ages. The human and fly experiments used Affymetrix GeneChips (Santa Clara, CA, USA), the mouse experiment used spotted cDNA filters, and the worm experiment used DNA microarrays. Each gene expression study used different numbers of samples (e.g. 26 worm samples, 40 mouse samples, 81 human muscle samples). Despite these differences in experimental design, Zahn et al. were able to use empirical meta-analysis to search for gene sets that showed a general increase or decrease in expression with age in different tissues and species. From each experiment, one can calculate a *P* value that a gene or gene set changes expression with age. Empirical meta-analysis combines the *P* values from each individual experiment and then calculates an overall *P* value for an age-related trend in multiple experiments.

Only one gene group, the electron transport chain gene set, was similarly age regulated in worms, flies, mice and humans (Fig. 1) (Zahn et al., 2006). The overall level of expression in this set of genes decreased about twofold in old age for all four species. The electron transport chain genes are located in the nuclear DNA and encode components of a mitochondrial enzyme complex that is the primary source of generation of free radicals in the cell. Free radicals are highly reactive side-products that non-specifically damage cell components such as proteins and DNA. Oxidative damage from the mitochondrial free radicals may accumulate with time and thereby decrease overall cell function and ultimately limit organismal lifespan (Golden et al., 2002). The lifespan of worms and humans differ by ~2000-fold (2 weeks *versus* 80 years), and the slope of age-related changes in expression for this pathway scales with lifespan such that old worms and old humans showed a similar overall decrease in expression levels. Because this genetic pathway showed similar age regulation in diverse species, it may be an exceptionally good biomarker of age.

Do changes in the electron transport chain genes in old age promote longevity or hasten senescence? In mammals, the functional significance of age-related changes in this pathway is unclear. However, in *C. elegans*, reduction of gene activity of the electron transport chain genes by RNAi has a strong effect on extending lifespan (Lee et al., 2003). This observation suggests that decreased expression of genes in the electron transport chain pathway in old age may help prolong life in worms, and possibly other species as well.

Genes that encode proteins in the lysosome show a common increasing trend in expression in humans, mice and flies but not worms (J. Zahn, unpublished data). The lysosome is responsible for degeneration of cell surface receptors, and increased expression of lysosomal genes may mark increased receptor turnover in old age.

Aging – a universal process that falls outside the force of natural selection

These results show clear evidence for shared age-related transcriptional changes in diverse species. To understand how these age-related transcriptional changes might be preserved in diverse species, it is important to consider current theories of the evolutionary basis of aging. There is an extensive and sophisticated literature on evolutionary theories of aging that provides key concepts to understand and interpret inter-species aging differences (Campisi, 2005; Golden et al., 2002; Kenyon, 2005; Kirkwood, 2005; Kirkwood and Austad, 2000; Martin, 2002).

A key consideration is that old animals do not normally constitute a significant fraction of the population in the wild. Wild populations die from predation and disease, and thus the principle determinant of natural longevity is extrinsic mortality (Fig. 2) (Kirkwood and Austad, 2000). For example, mice live 2–3 years under laboratory conditions, but 90% of mice die in their first year in the wild (Phelan and Austad, 1989). During human evolutionary history, the effective human population size was a mere 10 000 individuals, and life expectancy was around 20 years. Human life expectancy increased to 28 years in ancient Greece and Rome and to 46 years by the year 1900 (Martin, 2002). Currently, life expectancy is ~77 years in North America. Thus, only in modern human history do old individuals constitute a significant fraction of human society. In ancient times, most individuals died before the onset of white hair, wrinkled skin, muscle deterioration, loss of joint fluidity and other signs of old age. The effects of aging such as these are processes observed only under laboratory conditions (for animals) or in modern society (for humans).

Because senescent individuals are rare in the wild, changes that occur in old animals or people do not either help or hurt the fitness of their populations. Extrinsic mortality from predation and disease, not aging, has the strongest effect on overall fitness of a population. This realization led Medawar to develop a powerful theory of aging called the mutation accumulation theory. This theory states that old age is not under selective pressure *per se*, and there is no evolutionary mechanism to rid a population of mutations that cause detrimental effects only in old animals (Medawar, 1952).

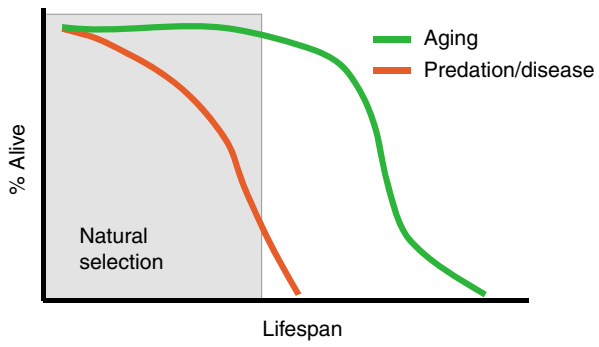

Fig. 2. Animals in natural populations die of predation and disease, not old age.

Mutations with late-acting phenotypes accumulate in the population over evolutionary time, and the cumulative effect of all of these late-acting mutations causes organismal senescence and limits lifespan.

Evolution would select for aging only insofar as that aging must occur more slowly than extrinsic mortality (i.e. one year for mice and 20 years for humans). For animals that lack predators, such as humans, bats and birds, the aging process is relatively slow because these animals survive many years in the wild. For animals that die rapidly from predation, such as worms, flies and mice, the aging process is correspondingly rapid as there is no evolutionary pressure to select for long life. In a classic experiment on the evolution of aging, Austad assessed the effect on aging due to a rapid decline in predation (Austad, 1993). He found that mainland opossums have a high rate of predation from carnivores. He found a small section of land that had become isolated from the mainland by a river, and which also happened to lack predators of the opossum. The opossums on the island had a significantly longer lifespan than those on the mainland, indicating rapid evolution of a longer lifespan. This result suggests that the lack of predation of island opossums made it possible for those individuals with slower rates of aging to come under natural selection and for the population to then evolve a longer lifespan. By analogy to the opossum example, the human genome is capable of supporting a longer lifespan than the mouse genome because there is less death from predation and disease in primitive humans than in mice.

Three categories of aging

What sorts of causes could explain declining organismal function in old age, and would one expect to see similar or different aging expression patterns in different species? The mutation accumulation theory suggests that mechanisms that limit lifespan may fall into three categories (Fig. 3). The first category includes all aging mechanisms that involve a gradual deterioration of cellular and metabolic processes with age. There is a strong selection for optimal function of cellular and biochemical pathways until young adulthood. After that, these pathways are not well maintained due to lack of selective pressure, and cellular function gradually declines, eventually resulting in cellular dysfunction and organismal senescence. Examples of aging mechanisms in this class include oxidative damage (Harmon, 1972), DNA damage (Hasty et al., 2003), telomere shortening (Shay and Wright, 2001), protein glycation (Schmidt et al., 1994) and transcriptional noise (Bahar et al., 2006), which are all mechanisms that could lead to cellular degradation and dysfunction in old age.

According to this view, aging represents the detrimental effect of cellular pathways that are not well maintained after young adulthood. The specific pathways that degenerate may be different for each species. For example, DNA damage may accumulate in humans but not worms because humans have about 10^{14} cells and *C. elegans* has only 959 cells (Sulston and Horvitz, 1977; Sulston et al., 1983). Telomere shortening may have a stronger effect on lifespan of humans than mice because mice have much longer telomeres than humans (Hemann et al., 2000). It could be that mutations that have accumulated in mice and that lead to rapid decline in three years may be entirely different from those that have accumulated in humans and that result in slow decline over 80 years. For mechanisms that fall in this category, genetic causes of aging are likely to be specific for each species.

The second category of aging mechanism involves pathways that are linked to a process that occurs during development or young adulthood. Some genetic pathways may exhibit antagonistic pleiotropy, which means that the same pathway is beneficial in young adulthood but detrimental in old age (Williams, 1957). For example, cell senescence is a mechanism to limit cell division that may be beneficial early in life to help prevent cancer but may be detrimental late in life because it limits cell divisions and eventually prevents cellular regeneration in aging tissues (Campisi, 2005). Caloric restriction falls into this category of aging because it can extend life in old age but is clearly linked to beneficial effects in young adults. Caloric restriction extends lifespan in yeast, worms, flies, mice and probably primates (Kenyon, 2005). All animals

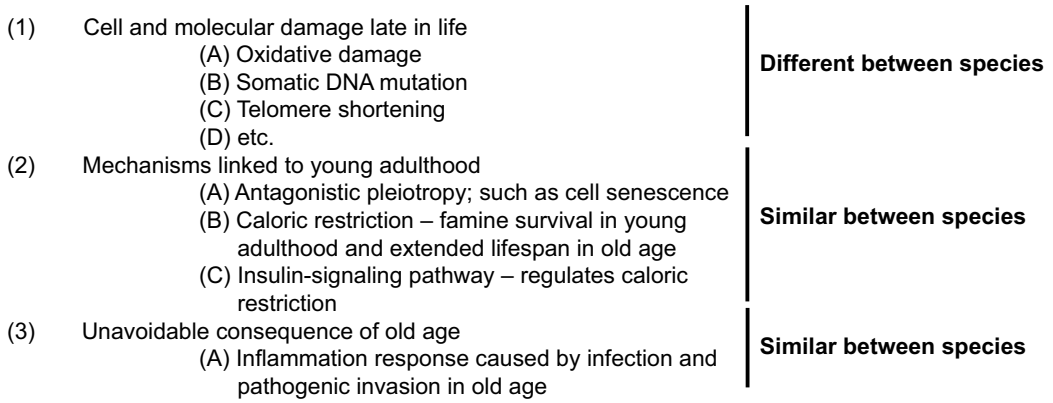

Fig. 3. Three categories of aging predicted by the mutation accumulation theory.

face times of starvation and drought, and a common strategy to improve fitness during harsh times is to extend lifespan and fertility in order to propagate the species when plentiful times return.

Because cell senescence and famine survival are key processes affecting young animals, they are under strong selective pressure and are evolutionarily conserved. The effects of cell senescence and caloric restriction on aging in late life may occur repeatedly for different animals, not because their effects on aging are conserved but rather because these aging mechanisms are tightly linked to functions in young animals.

Caloric restriction appears to be under control of the insulin-like growth factor signaling pathway (Guarente and Kenyon, 2000). Like caloric restriction, this signaling pathway has a role in famine survival and lifespan extension in many species. Specifically, mutations in the insulin-like signaling pathway have been shown to increase lifespan in worms, flies and possibly mice (Guarente and Kenyon, 2000).

The third category of aging includes external mechanisms that are unavoidably associated with old age. One example might be infection and inflammation that occurs in old age (Franceschi et al., 2000). As any animal grows old and begins to show physiological decline, there is an increased opportunity for pathogenic invasion. Thus, pathogenic invasion from external sources may occur in all species as they grow old, not because it is evolutionarily conserved but rather because it is an inescapable condition of old age.

What do the transcriptional profiles of aging in multiple species inform us about these three categories of aging? First, the vast majority of age-related transcriptional changes are different between different species (Fraser et al., 2005; Zahn et al., 2006). This observation suggests that most of the aging process falls under the first category – species-specific degeneration of cellular and metabolic pathways in old age.

Second, age regulation of only one pathway, the electron transport pathway, was observed in worms, flies, mice and humans (Zahn et al., 2006). The mechanisms responsible for decreased expression of the electron transport chain in old age are not known. It is unlikely that this pathway is commonly age regulated because of evolutionary selective pressure *per se*. Rather, common age regulation is likely the result of an unavoidable consequence of old age (aging category three). One possibility is that cellular damage from mitochondria in old age is unavoidable, which would suggest that oxidative damage may be a ubiquitous source of damage that limits lifespan in all animals. However, this possibility seems unlikely because mitochondrial function is highly conserved in all metazoans, and it seems unlikely that mitochondria in mice would cause extensive oxidative damage that severely limits life span (2 years) whereas mitochondria in humans would cause minimal damage and permit a long lifespan (80 years). Second, it is not obvious why mitochondria in lower animals would not function as efficiently as human mitochondria, and hence permit lifespans approaching that of humans. Another possibility is that expression of the electron transport chain scales with overall metabolic activity of the cell and that aging lowers cellular metabolic activity in all animals.

In summary, comparison of aging transcriptional profiles in worms, flies, mice and humans provides a quantitative, global view of the overall relatedness of the aging process across different species. These results provide a view of the relative proportion of the aging process that is specific to humans (private) rather than shared across animals (public). The vast majority of age-related transcriptional changes are private to humans, and these are likely the result of cell degeneration pathways that are species specific (aging category one) (Fig. 3). The emerging view from the genomics experiments is that the aging process is quite different in mice and humans, emphasizing the need for research using human samples to uncover aging mechanisms relevant to human longevity. A small amount of age-related changes in expression are public across species. These public aging pathways may be linked to functions in young adults (aging category two) or may be unavoidable consequences of growing old (aging category three). Identification of these public pathways is key because they highlight specific aging pathways that can be dissected apart in model organisms to elucidate general principles of aging.

This work was funded by grants from the Ellison Medical Foundation and the NIH.

References

Austad, S. N. (1993). Retarded senescence in an insular-population of opossums. *J. Zool.* **152**, 695-708.

Bahar, R., Hartmann, C. H., Rodriguez, K. A., Denny, A. D., Busuttil, R. A., Dolle, M. E., Calder, R. B., Chisholm, G. B., Pollock, B. H., Klein, C. A. et al. (2006). Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature* **441**, 1011-1014.

Campisi, J. (2005). Aging, tumor suppression and cancer: high wire-act! *Mech. Ageing Dev.* **126**, 51-58.

Evans, S. J., Choudary, P. V., Vawter, M. P., Li, J., Meador-Woodruff, J. H., Lopez, J. F., Burke, S. M., Thompson, R. C., Myers, R. M., Jones, E. G. et al. (2003). DNA microarray analysis of functionally discrete human brain regions reveals divergent transcriptional profiles. *Neurobiol. Dis.* **14**, 240-250.

Franceschi, C., Bonafe, M., Valensin, S., Olivieri, F., De Luca, M., Ottaviani, E. and De Benedictis, G. (2000). Inflamm-aging. An evolutionary perspective on immunosenescence. *Ann. N. Y. Acad. Sci.* **908**, 244-254.

Fraser, H. B., Khaitovich, P., Plotkin, J. B., Paabo, S. and Eisen, M. B. (2005). Aging and gene expression in the primate brain. *PLoS Biol.* **3**, e274.

Golden, T. R., Hinerfeld, D. A. and Melov, S. (2002). Oxidative stress and aging: beyond correlation. *Aging Cell* **1**, 117-123.

Guarente, L. and Kenyon, C. (2000). Genetic pathways that regulate ageing in model organisms. *Nature* **408**, 255-262.

Harmon, D. (1972). The biologic clock: the mitochondria? *J. Am. Geriatr. Soc.* **20**, 145-147.

Hasty, P., Campisi, J., Hoeijmakers, J., van Steeg, H. and Vijg, J. (2003). Aging and genome maintenance: lessons from the mouse? *Science* **299**, 1355-1359.

Hemann, M. T., Hackett, J., IJpma, A. and Greider, C. W. (2000). Telomere length, telomere-binding proteins, and DNA damage signaling. *Cold Spring Harb. Symp. Quant. Biol.* **65**, 275-279.

Kenyon, C. (2005). The plasticity of aging: insights from long-lived mutants. *Cell* **120**, 449-460.

Khaitovich, P., Muetzel, B., She, X., Lachmann, M., Hellmann, I., Dietzsch, J., Steigele, S., Do, H. H., Weiss, G., Enard, W. et al. (2004). Regional patterns of gene expression in human and chimpanzee brains. *Genome Res.* **14**, 1462-1473.

Kim, S. K. (in press). Genome-wide views of aging gene networks. In

Molecular Biology of Aging (ed. D. W. L. Guarente and L. Partridge). Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Kirkwood, T. B. (2005). Understanding the odd science of aging. *Cell* **120**, 437-447.

Kirkwood, T. B. and Austad, S. N. (2000). Why do we age? *Nature* **408**, 233-238.

Landis, G. N., Abdueva, D., Skvortsov, D., Yang, J., Rabin, B. E., Carrick, J., Tavare, S. and Tower, J. (2004). Similar gene expression patterns characterize aging and oxidative stress in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **101**, 7663-7668.

Lee, S. S., Lee, R. Y., Fraser, A. G., Kamath, R. S., Ahringer, J. and Ruvkun, G. (2003). A systematic RNAi screen identifies a critical role for mitochondria in *C. elegans* longevity. *Nat. Genet.* **33**, 40-48.

Lu, T., Pan, Y., Kao, S. Y., Li, C., Kohane, I., Chan, J. and Yankner, B. A. (2004). Gene regulation and DNA damage in the ageing human brain. *Nature* **429**, 883-891.

Lund, J., Tedesco, P., Duke, K., Wang, J., Kim, S. K. and Johnson, T. E. (2002). Transcriptional profile of aging in *C. elegans*. *Curr. Biol.* **12**, 1566-1573.

Martin, G. M. (2002). Gene action in the aging brain: an evolutionary biological perspective. *Neurobiol. Aging* **23**, 647-654.

McCarroll, S. A., Murphy, C. T., Zou, S., Pletcher, S. D., Chin, C. S., Jan, Y. N., Kenyon, C., Bargmann, C. I. and Li, H. (2004). Comparing genomic expression patterns across species identifies shared transcriptional profile in aging. *Nat. Genet.* **36**, 197-204.

Medawar, P. B. (1952). *An Unsolved Problem of Biology*. London: H. K. Lewis.

Melov, S. and Hubbard, A. (2004). Microarrays as a tool to investigate the biology of aging: a retrospective and a look to the future. *Sci. Aging Knowledge Environ.* **2004**, re7.

Park, S. K. and Prolla, T. A. (2005). Lessons learned from gene expression profile studies of aging and caloric restriction. *Ageing Res. Rev.* **4**, 55-65.

Phelan, J. P. and Austad, S. N. (1989). Natural selection, dietary restriction, and extended longevity. *Growth Dev. Aging* **53**, 4-6.

Pletcher, S. D., Macdonald, S. J., Marguerie, R., Certa, U., Stearns, S. C., Goldstein, D. B. and Partridge, L. (2002). Genome-wide transcript profiles in aging and calorically restricted *Drosophila melanogaster*. *Curr. Biol.* **12**, 712-723.

Rodwell, G. E., Sonu, R., Zahn, J. M., Lund, J., Wilhelmy, J., Wang, L., Xiao, W., Mindrinos, M., Crane, E., Segal, E. et al. (2004). A transcriptional profile of aging in the human kidney. *PLoS Biol.* **2**, e427.

Schmidt, A. M., Hori, O., Brett, J., Yan, S. D., Wautier, J. L. and Stern, D. (1994). Cellular receptors for advanced glycation end products. Implications for induction of oxidant stress and cellular dysfunction in the pathogenesis of vascular lesions. *Arterioscler. Thromb.* **14**, 1521-1528.

Shay, J. W. and Wright, W. E. (2001). Telomeres and telomerase: implications for cancer and aging. *Radiat. Res.* **155**, 188-193.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545-15550.

Sulston, J. E. and Horvitz, H. R. (1977). Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* **56**, 110-156.

Sulston, J. E., Schierenberg, E., White, J. G. and Thomson, J. N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64-119.

Williams, G. C. (1957). Pleiotropy, natural selection and the evolution of senescence. *Evolution* **11**, 398-411.

Zahn, J. M., Sonu, R., Vogel, H., Crane, E., Mazan-Mamczarz, K., Rabkin, R., Davis, R. W., Becker, K. G., Owen, A. B. and Kim, S. K. (2006). Transcriptional profiling of aging in human muscle reveals a common aging signature. *PLoS Genet.* **2**, e115.

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.