Genomic studies in ageing research: the need to integrate genetic and gene expression approaches

W. M. Passtoors1, M. Beekman1, D. Gunn2, J. M. Boer3, B. T. Heijmans1, R. G. J. Westendorp4, B. J. Zwaan5 & P. E. Slagboom1

From the 1Section of Molecular Epidemiology, Leiden University Medical Center, Leiden, the Netherlands; 2Unilever Corporate Research, Colworth Laboratory, Sharnbrook, Bedford, UK; 3Center for Human and Clinical Genetics; 4Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands; and 5Section of Evolutionary Biology, Institute for Biology, Leiden University, Leiden, the Netherlands


Genome-wide and hypothesis-based approaches to the study of ageing and longevity have been dominated by genetic investigations. To identify essential mechanisms of a complex trait such as ageing in higher species, a holistic understanding of interacting pathways is required. More information on such interactions is expected to be obtained from global gene expression analysis if combined with genetic studies. Genetic sequence variation often provides a functional gene marker for the trait, whereas a gene expression profile may provide a quantitative biomarker representing complex cellular pathway interactions contributing to the trait. Thus far, gene expression studies have associated multiple pathways to ageing including mitochondrial electron transport and the oxidative stress response. However, most of the studies are underpowered to detect small age-changes. A systematic survey of gene expression changes as a function of age in human individuals and animal models is lacking. Well designed gene expression studies, especially at the level of biological processes, will provide hypotheses on gene–environmental interactions determining biological ageing rate. Cross-sectional studies monitoring the profile as a chronological marker of ageing must be integrated with prospective studies indicating which profiles represent biomarkers preceding and predicting physiological decline and mortality. New study designs such as the Leiden Longevity Study, including two generations of subjects from longevity families, aim to achieve these combined approaches.

Keywords: ageing, expression, genomics, longevity, transcriptomics.

Introduction

The most prevalent diseases in Western societies have in common that age is a major risk factor. The process of ageing across different species is marked by the occurrence of molecular and physiological changes in all cells and tissues. It is not clear which of these changes are causal to ageing, but the accumulation of damaged molecules and loss of homeostasis is generally considered to be central to the rate of the ageing process, generating lifespan variation even in genetically identical individuals in a common environment [1, 2]. Although different age-related diseases may arise by different pathophysiological mechanisms, it is not unlikely that common mechanisms of ageing influence the onset of different age-dependent diseases and the lifespan. This may be explained if these mechanisms affect the accumulation of damaged molecules and eventually the homeostasis of the organism. Indications that such mechanisms exist come from the observation that the occurrence of multiple age-dependent disease conditions is
simultaneously postponed in caloric restricted laboratory rodents [3] and from the offspring of human centenarians. These individuals are expected to have inherited longevity assurance mechanisms that attenuate their ageing rate and protect from disease [4]. For more than two decades there is an ongoing search for mechanisms driving the ageing process in different species. Phenotypic features of ageing that are usually investigated include the mean and maximum population lifespan and the occurrence of age-dependent disease. True intermediate phenotypes, or biomarkers which mark the rate of biological ageing and the mortality risk of individuals in a preclinical phase of life, have hardly been identified. The availability of such biomarkers, comparable with the way bone mineral density is a biomarker predicting the risk for osteoporosis [5], would help in the search for mechanisms underpinning ageing.

Successfully applied genetic approaches in ageing research mainly involve the identification of genetic variation influencing intra-species differences in lifespan. These genetic studies revealed that some mechanisms of lifespan regulation, such as the insulin/insulin-like growth factor I signalling (IIS) pathway, are common to species ranging from nematodes to mice (see [3, 6] for reviews). Genetically modified ageing mutants in these studies have increased life spans by more than twofold. Other mechanisms so far identified by genetic research in model systems include mitochondrial function and histone modification control [7]. Genetic association studies of insulin signalling genes and other candidate loci with human longevity have given contradictory results (see [8, 9] for reviews). The gene most consistently associating with human longevity in all studies is the apoE locus [10], which is poorly investigated in lifespan studies in animals. In general, the population based cross-sectional analysis of gene frequencies with age has thus far provided few consistent results. This may be explained by the fact that the mean life expectancy in western societies has increased dramatically in the last 100 years, and that the current variation in lifespan in such populations, therefore, is mainly determined by nongenetic factors (75% [11, 12]). Genetic influences on human longevity are much higher in families containing a clustering of long-lived individuals [13–15] and these observations have resulted in new genetic studies based on familial longevity.

Far less effort than on genetics has been put into obtaining an understanding of the gene expression profiles that are associated with lifespan and how this relates to the genetic variation associated with longevity. To achieve this goal, a systematic survey of gene expression changes that occur in somatic tissues throughout the lifetime of individuals is required. Such profiles may reflect the interaction of the genome with environmental challenges that occur during the ageing process. Increased power in genetic studies must be accomplished, to uncover information on the interactive effects of multiple loci in ageing of higher species, which is expected to be essential in complex traits. It is thus the combination of approaches that promises most success in unravelling ageing as a complex trait.

In transcriptomic research, changes in gene expression profiles in various tissues are being explored mainly as comparisons of diseased and healthy individuals. Gene expression studies have been performed on a large scale in research into cancer [16, 17], stroke [18, 19], immune diseases [20–23] and other age-related diseases [24, 25]. In cancer research, expression profiles were indicated to be able to identify specific types of cancer, classify tumour stage or severity, disease progression and the response to certain treatments [26–31]. The basic design in most cancer research is relatively simple compared with other disease research, because target and healthy tissue from patients can directly be compared and the changes in gene expression are often large. For a complex trait like human ageing, the primary affected cells or tissues are unknown and the gene expression changes are probably small and thus more difficult to find. Therefore, this paper serves two functions: (i) to review and discuss the gene expression studies performed in the ageing field and the insights into the biological process involved in ageing and longevity they have revealed, and (ii) to draft up future recommendations for this research field.
Transcriptomics in human ageing

Until now, transcriptomics studies into human ageing utilize candidate gene-based and explorative approaches to identify which mRNA expression levels are regulated in an age-dependent fashion. Study designs mainly consist of comparisons between young and elderly individuals, sometimes including centenarians, or a comparison of RNA expression profiles in patients with progeroid syndromes to young and elderly individuals. Transcriptomic studies into human cancers require several tens of samples to identify significant expression patterns changing at least twofold between cancer and healthy tissue from the same patient [32]. The expected differences because of ageing are much smaller and more difficult to detect given the large inter-individual genomic variation in humans that is not because of ageing [33]. Large experimental groups as well as good study designs are required to enable reproducible conclusions to be made from studies of gene expression and ageing.

Candidate gene-based studies

Research based on candidate genes is generally performed in a cross-sectional design using cultured skin fibroblasts [34, 35], lymphocytes [36] and peripheral blood mononuclear cells (PBMC) [37]. The majority of the studies were performed by quantitative PCR methods [34, 35, 37], others use Northern blot analysis [36] to investigate specific candidate genes, such as heat shock protein HSP70 [36], osteoclastogenesis inhibitory factor [34], several proteasome subunits [35] and caspases [37].

Some statistically significant results were obtained from candidate gene-based studies [34–36] investigating a small number of genes in only two–five healthy females per group. Centenarians appeared to have an expression level that resembles that of young (<40 years) individuals more than that of elderly individuals (65–80 years) [34–36]. The authors interpreted this result as an indication that the expression levels of these genes of long-lived individuals as well as in younger ones reflects the healthier condition of these individuals as compared with elderly individuals.

In a relatively large study of 246 individuals aged 10–102 years old, Lacelle et al. [37] investigated gene expression of four caspases as a function of chronological age. In general, their results indicated a larger variation within an age category than across categories. Caspase-1 expression is increased in elderly (70–89 years) and long-lived (90+ years) individuals compared with younger (≤69 years) individuals. Caspase-8 expression was increased in the elderly only and caspase-3 in the long-lived individuals only, when compared with the younger age group. Expression of caspase-10 was similar across all age groups. In this cohort, a high level of caspase-8 is the biosignature for elderly individuals, whilst high levels of caspase-3 and low levels of caspase-8 mRNA expressions were the hallmark biosignature for nonagenarians and centenarians. With regard to both caspase-1 and -3, population profiles revealed a slow shift with age towards higher levels of caspase mRNA expression. The authors hypothesize that these age-related shifts are important, as they may indicate that a subgroup of individuals within the younger population, expressing high levels of caspase-3 mRNA in their PBMC, is favoured to attain longevity. In the cross-sectional design of this study, however, such hypotheses cannot be tested, as it is not possible to establish whether the changes observed were the cause or the consequence of an older age. A large prospective study of individuals followed from middle age onwards for medical history and lifestyle until the date of death would be able to reveal whether expression changes predict a decline of health and mortality risk. Such large human epidemiological studies do exist, and their follow-up period is 15 to over 50 years [38, 39].

Genome-wide studies

The introduction of the whole genome microarrays for gene expression research provided the opportunity to investigate the expression of the whole human transcriptome. Transcriptomic studies with an explorative approach could provide useful information about correlated genes and biological processes involved in ageing, including those that are not yet considered as candidates. Explorative expression studies have been performed on home-spotted microarrays containing...
∼4 000–6 000 transcripts [40–42] and on commercial Affymetrix microarrays, from 12 000 [43, 44] to ∼54 000 [45–49] transcripts on each array. The design commonly used was the cross-sectional comparison of young and elderly healthy individuals, about eight individuals maximum per group or an analysis of individuals across an age-range. Various tissues were investigated in these studies, such as cultured lymphocytes [40], fibroblast cell lines [41, 42, 46], brain [43, 44], kidney [48] and skeletal muscle [45, 47, 49].

Several pathways were highlighted by differentially expressed genes between young (19–29 years) and elderly (65–85 years) individuals [40], including energy metabolism, cell cycle, signal transduction and DNA repair (see Table 1 for a summary) [45, 49]. Other studies additionally included the comparison of individuals with progeria, which are genetic disorders that cause segmental premature ageing such as Werner syndrome [41] and Hutchinson-Gilford progeria syndrome [42], to represent ageing mechanisms in the general population. These studies showed that genes from important pathways act similarly in old (87–96 years) individuals and progeria patients (7–37 years) suggesting that some but not all normal ageing mechanisms are accelerated in progeroid syndromes. There is still debate as to whether progeria is a good model for human ageing [50].

Only four large transcriptomic studies were performed investigating age ranges from 13 to 106 years [43, 44, 47, 48], using sample sizes from 30 to 81 individuals. Applying Affymetrix microarrays with ∼12 000–54 000 transcripts per array, these groups found 250–985 transcripts that differ significantly in gene expression as a function of age in several tissues of healthy donors, with fold changes below two [43, 44]. Biological pathways found to be changed with age are listed in Table 1 and includes genes involved in the mitochondrial electron transport chain, cell cycle and extracellular matrix.

In addition to limited sample sizes, a major problem in gene expression studies of healthy humans is the restricted availability of relevant human tissue. Zahn et al. [47] performed a meta-analysis on data of multiple studies investigating RNA from muscle, kidney and brain in a cross-sectional design. There was almost no correlation between the expression levels, or changes with age of individual genes in the different tissues, meaning that in each tissue different individual genes change their expression with age. Importantly, using pathway analysis, gene expression changes influencing six pathways were consistently found in all three investigated tissues (the first six pathways that are listed in Table 1). These results indicate that these pathways, but not particularly individual genes, are common elements in the age-related expression changes of different tissues. This may implicate that in addition to tissue specific effects, a common ageing signature may be found in any tissue that reflects the age of the whole organism. This would give major implications for human epidemiological studies, for which frequently only blood is available.

A striking observation in the genome-wide studies was that individuals younger than 30 years showed very similar expression patterns [43]. Also, the age group older than 90 years of age showed very similar expression patterns, although this homogenous pattern is different from that of the age group younger than 30 years. However, the individuals aged between 30 and 90 showed highly variable expression patterns of which some are more similar to the young age group and others to the old age group. This may coincide with individuals diverging in their rates of ageing as they go through middle age and later their diversity in age at death. The heterogeneity in gene expression that becomes detectable from the age of 30 onwards may be an intermediate phenotype or biomarker, on the condition that it would represent biological differences in the ageing process of individuals that precede and contribute to disease and mortality. Two studies provided proof that this indeed was the case, by showing that the gene expression profiles of human muscle and kidney corresponded to the measured physiological functionality (the ‘biological age’) of the tissue rather than to the chronological age of the subject [47, 48].
<table>
<thead>
<tr>
<th>Pathways</th>
<th>Age range studies</th>
<th>N total</th>
<th>Young versus old studies</th>
<th>N case/N control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electron transport chain</td>
<td>Lu et al. [43]</td>
<td>30</td>
<td>Visala Rao et al. [40]</td>
<td>2/2</td>
</tr>
<tr>
<td></td>
<td>Erraji-Benckekroun et al. [44]</td>
<td>39</td>
<td>Kyng et al. [41]</td>
<td>5/6</td>
</tr>
<tr>
<td></td>
<td>Rodwell et al. [48]</td>
<td>74</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zahn et al. [47]</td>
<td>81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell cycle/cell growth</td>
<td>Lu et al. [43]</td>
<td>30</td>
<td>Visala Rao et al. [40]</td>
<td>2/2</td>
</tr>
<tr>
<td></td>
<td>Erraji-Benckekroun et al. [44]</td>
<td>39</td>
<td>Ly et al. [42]</td>
<td>3/2</td>
</tr>
<tr>
<td></td>
<td>Rodwell et al. [48]</td>
<td>74</td>
<td>Kyng et al. [41]</td>
<td>5/6</td>
</tr>
<tr>
<td></td>
<td>Zahn et al. [47]</td>
<td>81</td>
<td>Welle et al. [49]</td>
<td>8/8</td>
</tr>
<tr>
<td>Extracellular matrix</td>
<td>Lu et al. [43]</td>
<td>30</td>
<td>Visala Rao et al. [40]</td>
<td>2/2</td>
</tr>
<tr>
<td></td>
<td>Rodwell et al. [48]</td>
<td>74</td>
<td>Ly et al. [42]</td>
<td>3/2</td>
</tr>
<tr>
<td></td>
<td>Zahn et al. [47]</td>
<td>81</td>
<td>Csoka et al. [46]</td>
<td>4/3</td>
</tr>
<tr>
<td>Chloride transport</td>
<td>Lu et al. [43]</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rodwell et al. [48]</td>
<td>74</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zahn et al. [47]</td>
<td>81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complement activation</td>
<td>Lu et al. [43]</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rodwell et al. [48]</td>
<td>74</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zahn et al. [47]</td>
<td>81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ribosomes</td>
<td>Lu et al. [43]</td>
<td>30</td>
<td>Visala Rao et al. [40]</td>
<td>2/2</td>
</tr>
<tr>
<td></td>
<td>Rodwell et al. [48]</td>
<td>74</td>
<td>Visala Rao et al. [40]</td>
<td>2/2</td>
</tr>
<tr>
<td></td>
<td>Zahn et al. [47]</td>
<td>81</td>
<td>Welle et al. [49]</td>
<td>8/8</td>
</tr>
<tr>
<td>Inflammation</td>
<td>Lu et al. [43]</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Erraji-Benckekroun et al. [44]</td>
<td>39</td>
<td>Visala Rao et al. [40]</td>
<td>2/2</td>
</tr>
<tr>
<td></td>
<td>Rodwell et al. [48]</td>
<td>74</td>
<td>Visala Rao et al. [40]</td>
<td>2/2</td>
</tr>
<tr>
<td></td>
<td>Zahn et al. [47]</td>
<td>81</td>
<td>Kyng et al. [41]</td>
<td>5/6</td>
</tr>
<tr>
<td></td>
<td>Welle et al. [49]</td>
<td>8/8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Signal transduction</td>
<td>Erraji-Benckekroun et al. [44]</td>
<td>39</td>
<td>Visala Rao et al. [40]</td>
<td>2/2</td>
</tr>
<tr>
<td></td>
<td>Rodwell et al. [48]</td>
<td>74</td>
<td>Visala Rao et al. [40]</td>
<td>2/2</td>
</tr>
<tr>
<td></td>
<td>Zahn et al. [47]</td>
<td>81</td>
<td>Welle et al. [49]</td>
<td>8/8</td>
</tr>
<tr>
<td>Protein metabolism</td>
<td>Lu et al. [43]</td>
<td>30</td>
<td>Ly et al. [42]</td>
<td>3/2</td>
</tr>
<tr>
<td></td>
<td>Erraji-Benckekroun et al. [44]</td>
<td>39</td>
<td>Welle et al. [49]</td>
<td>8/8</td>
</tr>
<tr>
<td>Transcription</td>
<td>Lu et al. [43]</td>
<td>30</td>
<td>Csoka et al. [46]</td>
<td>4/3</td>
</tr>
<tr>
<td></td>
<td>Rodwell et al. [48]</td>
<td>74</td>
<td>Welle et al. [49]</td>
<td>8/8</td>
</tr>
<tr>
<td>Lipid metabolism</td>
<td>Lu et al. [43]</td>
<td>30</td>
<td>Visala Rao et al. [40]</td>
<td>2/2</td>
</tr>
<tr>
<td></td>
<td>Erraji-Benckekroun et al. [44]</td>
<td>39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ca homeostasis/signalling</td>
<td>Lu et al. [43]</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAP kinase/PKC family</td>
<td>Lu et al. [43]</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Erraji-Benckekroun et al. [44]</td>
<td>39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxidative stress</td>
<td>Lu et al. [43]</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Erraji-Benckekroun et al. [44]</td>
<td>39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stress response/DNA repair</td>
<td>Lu et al. [43]</td>
<td>30</td>
<td>Visala Rao et al. [40]</td>
<td>2/2</td>
</tr>
<tr>
<td></td>
<td>Ly et al. [42]</td>
<td>3/2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kyng et al. [41]</td>
<td>5/6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Welle et al. [45]</td>
<td>8/7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Welle et al. [49]</td>
<td>8/8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Animal ageing models

A large number of studies have investigated the changes in gene expression in relation to ageing and longevity in laboratory animals using commercially available microarrays. The most frequently used animal model for this research is the mouse [51–54], but nematodes [55], flies [55–57], rats [58] and monkeys [59, 60] have also been investigated. A range of tissues were studied such as liver, heart, muscle, aorta and brain. Across all species and in most experimental designs there is a gender effect on ageing features and gene expression [59, 61].

The main design for measuring chronological age-changes, as in humans, was a comparison of young and old individuals, using on average three animals per group. As such studies are usually performed in inbred strains, the variation between individual animals is smaller than amongst human individuals. A meta-analysis showed that, amongst many observed effects that were not comparable between species, the mitochondrial electron transport chain genes showed a consistent overall decrease in expression with chronological age in humans, mice and flies, but not in nematodes [47]. Interestingly, this pathway also consistently changed with age in different human tissues as mentioned above [47]. Another meta-analysis found that genes involved in the oxidative stress response were similarly changed with age in human, mouse and rat brain as well as in nematodes [43].

Together these studies will eventually provide insight in the gene expression changes just associated with chronological age, those with biological contribution to the ageing process of each species and those pointing at evolutionary conserved mechanisms of longevity. Some short-lived mutants are expected to represent a model for human progeria, whereas long-lived mutants and caloric restricted animals may represent familial and sporadic (nonfamilial) human longevity, respectively. Caloric Restriction (CR) refers to a diet with fewer calories than is typical, but which contains all the necessary nutrients and vitamins to support life. The comparison of caloric restricted rodents and littermates that did not receive CR showed that CR resulted in a partial inhibition of age-related changes in gene expression [51–54, 58]. This indicates that CR, as it prolongs lifespan, also influences the age-related changes in expression profiles in a way that may point to a decreased rate of biological ageing. These results were consistent amongst mice, flies and nematodes but could not be confirmed in monkeys [60].

The question is whether pathways affected by CR were also affected by mutations inducing increased lifespan in these species. A recent study compared gene expression profiles associated with mutations inducing longevity by reduced IIS signalling in nematodes, flies and mice [62]. Conservation of the IIS related longevity regulation was tested at the level of individual orthologous genes, for paralogous gene sets and at the level of biological processes reflected by gene classes. There was hardly any conservation of IIS regulation at the gene level between the species. However, the process analysis revealed that the conserved element in IIS regulation of the long-lived mutants associated with up-regulation of sugar catabolism and energy generation, with cellular detoxification pathways and with down-regulation of protein biosynthesis and translation. The authors suggested

Table 1 (Continued)

<table>
<thead>
<tr>
<th>Pathways</th>
<th>Age range studies</th>
<th>N total</th>
<th>Young versus old studies</th>
<th>N case/N control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apoptosis</td>
<td>Erraji-Benckekroun et al. [44]</td>
<td>39</td>
<td>Kyng et al. [41]</td>
<td>5/6</td>
</tr>
<tr>
<td>Energy metabolism</td>
<td>Erraji-Benckekroun et al. [44]</td>
<td>39</td>
<td>Welle et al. [49]</td>
<td>8/8</td>
</tr>
<tr>
<td>Hormonal</td>
<td>Lu et al. [43]</td>
<td>30</td>
<td>Welle et al. [49]</td>
<td>8/8</td>
</tr>
<tr>
<td>Synaptic transmission</td>
<td>Lu et al. [43]</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vesicular transport</td>
<td>Lu et al. [43]</td>
<td>30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
that of these IIS regulated processes those that most likely reflect the longevity regulation, because of findings in other ageing studies, are lowered protein biosynthesis and increased cellular detoxification reflected by glutathione-S-transferase (GST) activity.

**Future directions for human research**

Clearly, genomic studies in animal models generate many hypotheses that are worth testing in humans, where the possibilities for investigating healthy individuals are very limited. Here we will discuss what improvements may be accomplished.

**Sample size**

As for genetic association studies, the sample size of a gene expression study in humans greatly determines the credibility of the results, especially concerning the multiple testing performed in genome-wide explorative analyses. A review from Allison et al. [63] nicely summarized the main statistical issues in expression studies. Minimal sample size calculations depend on design, the total background variation within the data, the magnitude of the expression change that is to be detected, the desired statistical power or probability to detect differentially expressed genes, the specified error rate and the statistical method being used to detect the changes in gene expression. They advocated that the minimal, and certainly not optimal, sample size to be used in expression studies is five biological cases per group. This amount is only useable when testing for differential expression with fold changes of three and up [64], and then the reproducibility of the results is still questionable. As fold changes in differential gene expression are expected to be below 2 in human ageing, it has been suggested that a minimum of 25 individuals per group is required in the analysis of a case control design [65, 66]. Large sample sizes are thus required. As a consequence, the primary studies aimed at the identification of biological processes relevant for human ageing by gene expression approaches are bound to be performed on easily available tissue from healthy individuals such as blood, buccal cells from the mouth and skin fibroblasts.

**The value of studying whole blood in transcriptomics**

Because blood is an easily available tissue, several studies have investigated gene expression profiles in whole human blood. Whole blood mRNA was used for studies into Huntington disease, Down syndrome, multiple sclerosis, psoriatic arthritis, Alzheimer’s disease and epilepsy [67–70]. In a candidate gene study for schizophrenia, 21 out of the 31 candidate genes expressed in the brain were similarly expressed in whole blood [67]. Gene expression in blood is also shown to mirror disease state in multiple sclerosis [20]. In general, the expression levels of genes as investigated on a 34K Affymetrix microarray show an average correlation of 0.5 when comparing blood and other tissues [67]. Blood has also been successfully used for classification of disease subgroups, for progression prediction and to investigate the effects of medical treatment [71–74]. It is expected that blood reflects a limited proportion of the biological processes relevant for ageing and longevity. However, ageing does not affect one specific tissue and blood both initiates and reflects processes that affect the whole body. Therefore, studying gene expression in blood cells is a good start for finding the relevant intermediate phenotype that emerges from genetic variation for ageing and longevity. Gene expression profiles in blood were found to be highly similar in monozygotic and moderately similar in dizygotic twin pairs, with a heritability of 0.33 [75–77]. Variation in gene expression in healthy unrelated individuals is found to be limited and detected as a result of blood cell subsets, gender, age, time of day and individual differences [78–80].

As an illustration to the latter point, we compared gene expression profiles of mRNA isolated from whole blood samples obtained from three healthy married volunteer couples (61–76 years). Blood was taken on four occasions; on two different days (3 weeks apart) and in the morning (10 a.m.) as well as in the afternoon (2 p.m.). Gene expression was measured using CodeLink whole genome microarrays (~54 000 transcripts). Figure 1 provides an overview of the 1548 genes most variable between the individuals, which showed very similar expression patterns at
the four time points within an individual, suggesting a stable expression profile over this time interval. In addition, statistical analysis of this pilot comparison revealed that the expression of very few genes was influenced by the time of day, or day at which the blood was collected. Most differentially expressed genes were associated with individual variation, including gender.

All in all, current data justifies the use of whole blood in the study of human ageing. However, as will be the case for including only one tissue, much of the relevant biological context will be missed. Therefore, such investigations should be linked to parallel studies in animal models. These models can show what interactions between biological processes with each other and with environmental stimuli are reflected in blood. As has been shown in animal studies, the human studies will become more informative if study designs allow for a combined analysis of genetic variation, gene expression and environmental variation.

**Study design**

Cross-sectional designs comparing young and elderly individuals are thus far most frequently used in gene expression studies of human ageing. In these designs there is no possibility to distinguish causal changes...
that are of major importance in determining biological ageing, mortality and longevity of individuals, and chronological changes that reflect effects thereof. Prospective study designs are available in large epidemiological human cohort studies of 20–30 years to follow up, that may indicate which RNA profiles (if cells are available in the study) precede and predict changes in physiological health and mortality risk. Twin studies will add vital information as to how genetic variation and environmental factors relevant for human ageing influence gene expression profiles throughout life. Comparable with long-lived strains in animal models, human lifespan variation is studied also in families of exceptional longevity. Especially the offspring of long-lived individuals may reflect which RNA profiles associate to the familial longevity trait at middle age [13]. The middle aged offspring of centenarians do express part of the beneficial phenotype already in the sense that they have a life-long decreased risk of mortality from coronary heart disease, diabetes and cancer. This can be viewed as secondary to a decreased rate of biological ageing [4] occurring in longevity families.

As an illustration of how new designs can be used in genomic investigations of human ageing and longevity we will now discuss the concerted genetic and gene expression studies in the Leiden Longevity Study. The aim of this research is to identify genomic and biochemical parameters associated with human longevity by combining research in familial long-lived individuals and their family members and in sporadic long-lived individuals from the general population. The Leiden Longevity Study (Fig. 2) consists of 420 Caucasian families with at least two long-lived siblings (men aged 89 years or above; women aged 91 years or above, \(n = 960\)), the middle aged offspring (\(n = 1764\)) and the partners of this offspring (\(n = 757\)). In 2001, <0.5% of the Dutch population fulfilled these sex-specific criteria [14]. The survival benefit of the long-lived families is marked by a 30% decreased mortality risk observed in the survival analysis of three generations, the parents of the sibling pairs, their deceased siblings and their offspring [14]. The familial survival advantage thereby exceeds the increased lifespan expectancy in the last generations of western societies because of nongenetic factors (improved nutrition, hygiene and health care). Therefore, this selected population is enriched for heritable familial longevity. An ongoing genome-wide association study using Affymetrix 500k SNP arrays should reveal which genetic variation is associated with familial longevity and which individuals carry this variation.

For the purpose of biomarker identification, including gene expression profiling, a cross-sectional comparison of young controls (partners of offspring) and elderly cases (long-lived siblings) may be performed. Chronological age changes will also be observed in an analysis of the age range of each subgroup: long-lived siblings (89–103 years), offspring (34–80 years) and partners (30–79 years). Biologically relevant age changes in biomarkers will emerge from a comparison of offspring and partners. The offspring can be viewed as cases with a higher susceptibility to become long-lived and their partners, representatives of the general population, as controls [14]. The key advantage of this design is that the offspring and their partners are of similar age and have a similar environment. Therefore, differentially expressed genes in the comparison of these two groups could represent processes associating to familial longevity detectable at middle age and thereby preceding the long-lived

Fig. 2 Design of the Leiden Longevity study. Long-lived families were recruited if they included a long-lived sibling pair with males aged 89 years or older and women aged 91 years or older.
phenotype. Some of these genes will become even more marked in their differential expression when studied in a comparison of the elderly long-lived cases and controls. This methodology has revealed metabolic markers [81, 82] as potential biomarkers and is now being used for gene expression studies. Whole blood RNA samples from 150 individuals of the Leiden Longevity study are being analysed (50 long-lived siblings, 50 offspring and 50 partners). At the level of the individual, the expression profile or expression level of specific genes can be linked to the genetic variation that was found to associate with familial longevity in the genetic studies. This analysis is expected to reveal causal changes in biological processes associated with longevity. As stated before, to obtain an understanding of the biology of ageing, it is highly relevant to investigate interactions of genomic (genetic and gene expression) variation with the environmental component. For this purpose data on nutrition, lifestyle and socio-economic status are being collected. Although in human studies the biological depth of such information is limited, those are the factors that may be beneficially manipulated to improve health.

The proof of the pudding for a parameter to be a biomarker of biological ageing will be established only if it predicts decline of physiological health and mortality risk in a prospective survey of the middle aged offspring and partners. A study of such a marker in a prospective population based cohort study, such as the Leiden 85+ study [83], will reveal whether it marks ageing mechanisms relevant only for familial longevity or also for population wide longevity. To illustrate the value of this methodology, we have recently observed that low concentrations of small Low Density Lipoprotein particles is a feature of familial longevity, detectable at middle age and is expressed especially in healthy familial nonagenarians (health based on scores on the Activity of Daily Living questionnaire), but also in sporadic nonagenarians in the population [82].

**Concluding remarks**

Genetic and transcriptomic approaches have both provided insight into the biological processes involved in ageing. Transcriptomic studies identified a number of promising pathways that may be affected by ageing in different species, including humans, and across different tissues within a species, amongst which the electron transport chain and oxidative stress response. These pathways are connected because when synthesizing ATP, the mitochondrial electron transport chain releases electrons which can reduce oxygen, forming reactive oxygen species (ROS) such as superoxide. These ROS can induce oxidative stress and may contribute to the decline in mitochondrial function associated with the ageing process [84]. Caloric restriction has been found to not only increase the life-span of rodents, but also to reduce ROS production [85]. This can be explained when high amounts of energy are available, mitochondria do not operate very efficiently and generate more ROS, whilst when calorie restricted, energy is conserved and there is less free radical generation. Glutathione-S-transferases are involved in cellular detoxification and have recently been linked to longevity across species [62]. Glutathione-S-transferase can act as an antioxidative enzyme by protecting against damaging ROS and thereby decreasing oxidative stress in cells [86].

Gene expression studies in animal models have identified various biological processes that can serve as hypotheses in human studies. However, for a true testing of such hypotheses and the identification of human pathways, improvements in methodology are vital. These include larger sample sizes, the concerted application of population/patients based designs and familial designs that can distinguish chronological age effects from biological ageing. Highly relevant is also the combined analysis of genetic variation, gene expression variation and environmental factors. Figure 3 represents a scheme depicting how genomic studies may be integrated in research aimed at disentangling the complex biology of ageing. Any input study of lifespan variation whether associated with naturally occurring genetic variation (human cohorts, or different animal strains), induced by mutations or by modulation of the environment (such as in CR) may be studied for gene expression changes.
Gene expression studies may directly reveal leads to relevant mechanisms in a species, but especially the integration of gene expression studies at the level of biological processes (such as in [62]) will provide hypotheses as to whether and which gene-environmental interactions associate with common processes determining biological ageing rate. It is also necessary to integrate cross-sectional studies monitoring the profile as a chronological marker of ageing, and prospective studies indicating which of these is a biomarker preceding and predicting physiological decline and mortality. The integration of such information with knowledge emerging from gene-environmental interactions will reveal causal mechanisms for ageing that can be tested simultaneously in animal models and in humans.

Human studies should include larger sample sizes to increase the power of the analyses and reliability of the results, better systematic surveys of gene expression changes over a lifetime and designs that allow a good integration with genetic studies. Understanding the aetiology of a complex trait such as longevity will require more than just transcriptomics. There will be a need to combine genetics, transcriptomics, proteomics and metabolomics to gain further knowledge about biological processes involved in ageing and longevity.

Conflict of interest statement
No conflict of interest was declared.

Acknowledgements
We acknowledge Unilever Colworth for their help in arraying the blood samples.

The efforts in the Leiden Longevity Study are supported by a grant from Innovation Oriented research Program IOP on Genomics (SenterNovem IGE5007), by the stimulation grant from the Netherlands Genomics Initiative/Netherlands Organisation for Scientific Research (NGI/NWO) (grant number 05040202; Netherlands Initiative into Healthy Ageing) and by the Centre for Medical Systems Biology (CMSB).

References
et al. Apolipoprotein E genotype determines survival in the oldest old (85 years or older) who have good cognition. Arch Neurol 1996; 53: 418–22.


Comparing genomic microarray analyses of genetic dissection of evolution and conserva-


84 Huang H, Manton KG. The role of oxidative damage in mitochondria during aging: a review. *Front Biosci* 2004; 9: 1100–17.


Correspondence: Willemijn Passtoors, MSc, Section of Molecular Epidemiology, Leiden University Medical Center, PO Box 9600, Zone S5-P, 2300 RC Leiden, the Netherlands.

(fax: +31 071 526 8280; e-mail: w.m.passtoors@lumc.nl)