Genetic evidence for common pathways in human age-related diseases

Simon C. Johnson,[1,]* Xiao Dong,[1,]* Jan Vijg[1,2] and Yousin Suh[1,3]

[1]Department of Genetics, Albert Einstein College of Medicine, Bronx, NY
[2]Department of Ophthalmology and Visual Sciences, Albert Einstein College of Medicine, Bronx, NY
[3]Department of Medicine Endocrinology, Albert Einstein College of Medicine, Bronx, NY, USA

Summary

Aging is the single largest risk factor for chronic disease. Studies in model organisms have identified conserved pathways that modulate aging rate and the onset and progression of multiple age-related diseases, suggesting that common pathways of aging may influence age-related diseases in humans as well. To determine whether there is genetic evidence supporting the notion of common pathways underlying age-related diseases, we analyzed the genes and pathways found to be associated with five major categories of age-related disease using a total of 410 genomewide association studies (GWAS). While only a small number of genes are shared among all five disease categories, those found in at least three of the five major age-related disease categories are highly enriched for apoliprotein metabolism genes. We found that a more substantial number of gene ontology (GO) terms are shared among the 5 age-related disease categories and shared GO terms include canonical aging pathways identified in model organisms, such as nutrient-sensing signaling, translation, proteostasis, stress responses, and genome maintenance. Taking advantage of the vast amount of genetic data from the GWAS, our findings provide the first direct evidence that conserved pathways of aging simultaneously influence multiple age-related diseases in humans as has been demonstrated in model organisms.
Key words: ageing; aging; genetics; gerontogenes; human; inflammation; Ink4a; longevity gene.

Introduction

Aging is a complex process of progressive functional decline influenced by environmental, genetic, and stochastic factors (Lopez-Otin *et al.*, 2013). Genetic approaches to studying aging using yeast, nematode, fly, and rodent models have identified conserved genetic factors that modulate aging (Kenyon, 2010; Vijg & Suh, 2013). These include

Correspondence

Jan Vijg and Yousin Suh, Department of Genetics, Albert Einstein College of Medicine, 1301 Morris Park Ave, Bronx, NY 10461, USA. Tel.: +1 718 678 1151; fax: +1 718 678 1016; e-mail: jan.vijg@einstein.yu.edu (J.V.); Tel.: 718-678-1111; fax: 718-678-1113; e-mail: yousin.suh@einstein.yu.edu *(Y.S.)*
*These authors contributed equally to this work.

genome and epigenome maintenance; nutrient-sensing signaling through insulin/IGF-1, mTOR, and AMP kinase; regulation of proteostasis through cellular degradation pathways and activity of the unfolded protein responses; inflammation and senescence pathways; and key transcriptional regulators such as the Foxo transcription factors. In model organisms, alterations to these conserved genetic modifiers of aging can impact both survival (lifespan) and health span, the percentage of life spent free from significant pathology. Alterations to genome maintenance, inflammation, or proteostasis can result in shortened lifespan and symptoms of rapid aging (Merkwirth *et al.*, 2012; Schleit *et al.*, 2013; Jurk *et al.*, 2014), while genetic, dietary, and pharmacological interventions that reduce growth signaling through IIS, mTOR, and AMPK tend to increase lifespan and delay and reduce the frequency and severity of aging-related pathologies. Thus, interventions designed to target the underlying mechanisms of aging are expected to provide great benefit to human health by attenuating a broad range of pathologies (Fontana *et al.*, 2010; Berry & Cirulli, 2013; Johnson *et al.*, 2013; Torgovnick *et al.*, 2013; Sikora, 2014). While the notion that age-related diseases are driven by common underlying pathways of aging is supported by model organism studies (Fig. 1), whether it remains true in human aging is unclear.

Genomewide association studies (GWAS) have provided unprecedented opportunities to identify genes and genomic regions associated with complex traits in human populations, including disease risk. At the time of this study, 1738 GWAS had been conducted for 816 human traits (including human diseases), together reporting 11 533 trait-associated single-nucleotide polymorphisms (SNPs) reaching a suggestive association threshold of $P < 1 \times 10^{-5}$ (see the Web resource of GWAS catalog: http://www.genome.gov/gwastudies). These results have greatly contributed to our understanding of biological pathways involved in individual human traits. In contrast, little insights have been gained from GWAS regarding the genetic pathways regulating human longevity. To date, 5 GWAS of longevity, involving long-lived individuals compared to control populations, have been reported in the catalog of published GWAS. Among these studies, the APOE gene has emerged as the only consistently replicated locus associated with human lifespan. Human health span has not been studied by GWAS, and human age-related diseases have been studied only as independent traits. Here, we have taken a novel approach to jointly analyzing a large set of GWAS data to directly address the question of whether human age-related diseases are linked by common underlying pathways of aging as has been demonstrated in model organisms.

Results

To approach the question of whether age-related diseases are influenced by common genetic pathways, we downloaded all available GWAS data from the complete (November 9, 2013) release of the NHGRI GWAS database (Welter *et al.*, 2014). To focus our analysis on traits that have been sufficiently studied, as well as to address computational limitations in downstream analyses, we chose to restrict our analyses to traits that have been examined in a minimum of five independent GWAS, resulting in a total of 410 GWAS in the 5 age-related disease categories. Furthermore, in each GWAS, we only considered SNPs with *P*-value less
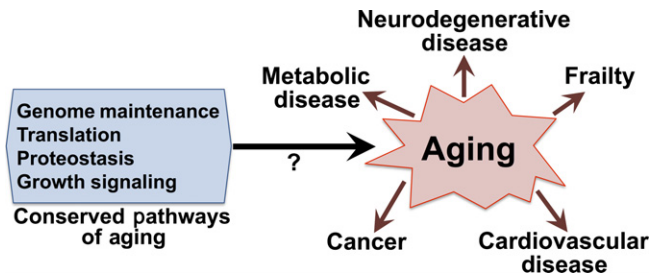
Fig. 1 Common pathways in aging and age-related disease. Aging is the single greatest risk factor for age-related disease, and it is well established that conserved genetic pathways of aging impact multiple age-related pathologies in model organisms, but the role of conserved pathways of aging in human age-related disease is unclear.

than 1×10^{-5} as trait associated. The combination of these two criteria ensures that only significant SNPs from well-studied traits are represented in the work presented here. We defined each trait meeting these criteria as either age-associated (Table 1) or non-age-associated (Table S1, Supporting information). Age-associated traits were then assigned to one of five broad and well-established age-related disease categories: cardiovascular disease, metabolic disease, neurodegenerative disease, cancer, and collectively other age-related traits [see (Perez-Lopez *et al.*, 2009; Walter *et al.*, 2011; Martin, 2012; Johnson *et al.*, 2013; Brunet & Berger, 2014)]. The other age-related trait category includes traits conventionally associated with aging, such as telomere length and inflammation, as well as related age-associated diseases including rheumatoid arthritis and age-related macular degeneration. GWAS traits, properties, and assignments are indicated in Table 1 (age-associated traits).

A gene-based assessment of age-related diseases

GWAS SNPs were assigned to genes (see Methods) resulting in a total number of 1975 unique genes among all five categories (Fig. 2). Only three genes (0.15% of total genes) were shared among all the five age-related diseases categories (Fig. 2A, Table 2). All three shared genes fell in the MHC locus, a highly variable region over-represented in GWAS studies. Clustering of age-related disease groups by gene list similarity (fraction of genes shared vs. total) indicates that cardiovascular disease, metabolic disease, and other age-related trait categories cluster more closely, while cancer and neurodegenerative diseases represent the most distinct group with the least overlap between the two (Fig. 2B). Metabolic disease and other age-related diseases show the highest similarity between two groups at an overlap of 0.083 (fraction of all genes), whereas metabolic disease and neurodegenerative disease show the least overlap at 0.029 (fraction of total genes).

The low number of shared genes among all age-related disease groups is likely in part a result of the stringent statistical cut-offs of GWAS. To broaden our analysis to include genes that are shared in at least the majority of age-related disease categories, we assessed the genes shared by at least three of the five age-related categories resulting in a total of 50 genes (2.5% of unique genes, listed in Table 2), including 12 MHC genes as well as a number of genes involved in inflammation, cell cycle regulation, and cholesterol/apolipoprotein metabolism (Fig. 2C, Table 2). Notably, both TOMM40/APOE, the highly reproducible longevity-associated locus, and p16INK4a, a cell cycle/senescence regulator gene locus previously identified to be associated with multiple

Table 1 Genomewide association studies diseases used in this study

Disease	Number of GWAS	Significant SNPs	Disease category
Adiponectin levels	9	44	Metabolic
Age-related macular degeneration	9	66	Other
Alzheimer's disease	18	64	Neurodegenerative
Alzheimer's disease (late onset)	9	35	Neurodegenerative
Amyotrophic lateral sclerosis	12	54	Neurodegenerative
Atrial fibrillation	5	20	Cardiovascular
Blood pressure	5	57	Cardiovascular
Body mass index	17	124	Other
Bone mineral density	12	97	Other
Breast cancer	26	122	Cancer
C-reactive protein	8	46	Other
Colorectal cancer	14	47	Cancer
Coronary heart disease	17	133	Cardiovascular
Endometriosis	5	26	Cancer
Fasting plasma glucose	8	11	Metabolic
Glaucoma (primary open-angle)	5	12	Other
HDL cholesterol	12	100	Metabolic
Hypertension	8	33	Cardiovascular
LDL cholesterol	12	76	Metabolic
Longevity	*5*	*43*	** (longevity)*
Lung cancer	10	24	Cancer
Mean platelet volume	5	51	Other
Melanoma	7	18	Cancer
Metabolite levels	8	98	Metabolic
Multiple sclerosis	16	162	Neurodegenerative
Myopia (pathological)	7	76	Other
Obesity	7	77	Metabolic
Osteoarthritis	5	8	Other
Pancreatic cancer	5	39	Cancer
Parkinson's disease	13	67	Neurodegenerative
Prostate cancer	19	104	Cancer
Pulmonary function	5	55	Cardiovascular
QT interval	8	76	Cardiovascular
Rheumatoid arthritis	14	78	Other
Telomere length	7	16	Other
Triglycerides	12	79	Metabolic
Type 2 diabetes	32	141	Metabolic
Urate levels	9	68	Metabolic
Uric acid levels	5	34	Metabolic

Diseases/traits that met the criteria for inclusion in this study are indicated with the number of GWAS studies, number of GWAS significant SNPs, and disease category assignment as indicated. The trait longevity, while meeting our criteria for inclusion, was not included.

diseases (Jeck *et al.*, 2012), are represented in this gene set. A Gene Ontology (GO) analysis of these 50 genes against the baseline set of the total 1975 unique genes revealed a significant enrichment of GO terms in the apolipoprotein metabolism pathway (Fig. 2D, Table S2, Supporting information). Strikingly, apolipoprotein metabolism is the only biological process significantly enriched in the gene-based overlap between the age-related disease categories defined in this work.

A pathway approach to age-related disease

To test whether age-related diseases share common underlying genetic mechanisms and pathways, we next utilized a novel approach of assessing the pathway-based overlap between disease categories. For
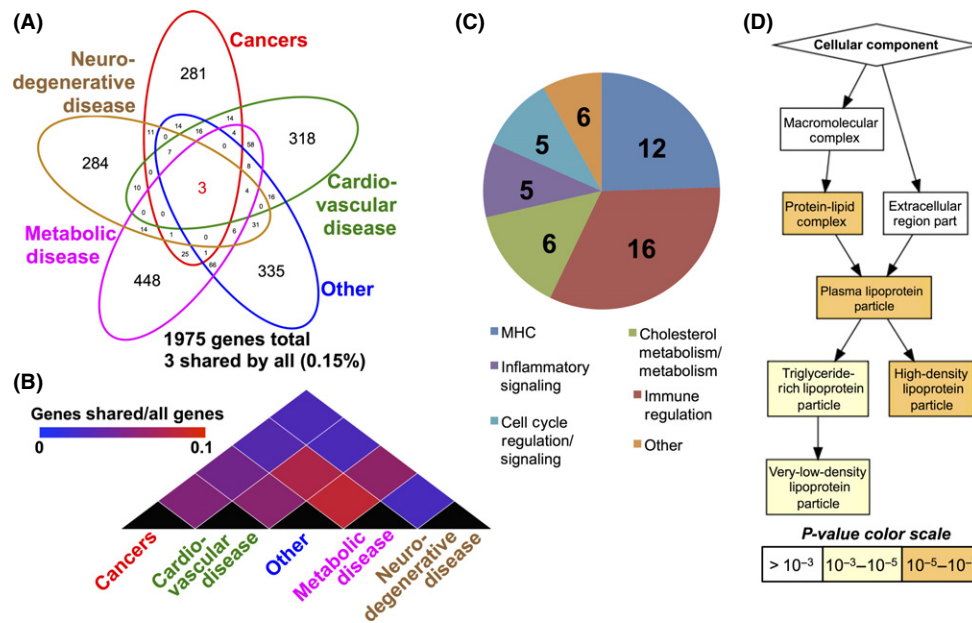
Fig. 2 Genomewide association studies (GWAS) significant genes shared among age-associated diseases. Age-related disease categories show very little overlap of GWAS significant genes (A), with only 3 (0.15%) of genes identified shared among all groups, while the majority of genes appear to be detected in only one group. (B) Clustering of age-related disease categories by similarity shows that cardiovascular disease, metabolic disease, and other age-related pathologies are closely related, while cancers and neurodegenerative diseases represent the most distinct groups. (C) Number and category of genes shared among at least 3 age-related disease categories represent those predominately involved in immunity, inflammation, cell cycle regulation, and cholesterol metabolism. (D) GO analysis of these genes compared to the background set of all 1975 GWAS detected genes revealed that apolipoprotein metabolism is significantly and specifically enriched in this gene set.

this purpose, we compiled GO terms for each identified gene in each group and assessed the overlap of GO terms between disease groups (Fig. 3A). Assignment of GO terms resulted in a total list of 2734 unique GO terms among all 5 age-related disease genes (Fig. 3B). Of these, 209 GO terms (7.6% of total) were shared among all five groups. Clustering of disease groups by similarity resulted in a pattern identical to that observed using gene-based clustering, indicating that the use of GO terms did not alter the qualitative comparisons between diseases. As with a gene-based approach, we found cardiovascular disease and metabolic disease are more closely related while cancer and neurodegenerative disease are the most distantly related (Fig. 3C). Among the 7.6% of GO terms overlapped among all five age-related disease categories, there is a significantly higher than random similarity between terms based on the background ($P < 1 \times 10^{-15}$, one-sided Wilcoxon rank-sum test, see Methods for more details). Among the overlapping GO terms, we found clusters of terms related to nutrient-sensing signaling, translation, genome maintenance, proteostasis, oxidative stress responses, inflammation, and, as in the gene-based assessment, lipoprotein metabolism, and most of them are known to be canonical pathways of aging (Tables S3 and S4, Supporting information, Fig. 4). A principle component analysis of these shared GO terms using REViGO term clustering and analysis further indicated the presence of shared canonical aging pathways in the 5 age-related disease categories (Fig. S2, Supporting information). Considering the stringent criteria for inclusion of individual GWAS traits and the use of only genes with significant trait-associated SNPs, the significant enrichment of canonical pathways of aging among the pathways shared in all 5 age-related disease categories is striking.

For reference, we examined the genes associated with the GWAS trait longevity at $P < 1 \times 10^{-5}$ and the GO terms associated with this gene set (Tables S5 and S6, Supporting information). Lipoprotein metabolism pathways reach nominal significance for enrichment among this gene set, although no GO term is significant after multiple testing correction.

Statistical assessment

Our approach of identifying genes associated with a significance level of $P < 1 \times 10^{-5}$ in multiple independent GWAS traits is amenable to the Fisher's combined probability test (Li et al., 2014). This test indicates that the combined P-value for 3 independent tests with a cut-off of $P < 1 \times 10^{-5}$ each is equal to 6.32×10^{-13}, while the combined P-value for 5 independent tests each with a cut-off of $P < 1 \times 10^{-5}$ is 4.9×10^{-20}. Furthermore, a comparison of genes or GO terms identified using the GWAS cut-off of $P < 1 \times 10^{-5}$ vs. a GWAS cut-off of $P < 1 \times 10^{-7}$ indicates that using a more stringent GWAS threshold prevents the identification of any overlapping genes but has little impact on the number of identified GO terms (Fig. S3, Table S7, Supporting information).

Discussion

It is widely accepted among gerontologists that common processes mechanistically underlie both aging and the pathogenesis of multiple age-related diseases and that targeting common factors in aging will have a significant benefit to human health (Berry & Cirulli, 2013; Torgovnick et al., 2013; Fontana et al., 2014; Sikora, 2014). A wealth of experimental data from lower organism studies supports this concept, and human progeroid syndromes indicate that disruption of key biological processes can result in the premature onset of multiple age-related pathologies (Ghosh & Zhou, 2014). There has, however, been

Table 2 Genomewide association studies associated genes shared among 3 or more age-related disease categories

Gene ID	Gene name	Function
3118	Major histocompatibility complex, class II, DQ alpha 2	MHC
3119*	Major histocompatibility complex, class II, DQ beta 1	MHC
3120	Major histocompatibility complex, class II, DQ beta 2	MHC
3122*	Major histocompatibility complex, class II, DR alpha	MHC
3123	Major histocompatibility complex, class II, DR beta 4	MHC
100507714	HLA class II histocompatibility antigen, DQ beta 1 chain-like	MHC
100507709	HLA class II histocompatibility antigen, DRB1-7 beta chain-like	MHC
101060835*	HLA class II histocompatibility antigen, DQ beta 1 chain-like	MHC
100509457	HLA class II histocompatibility antigen, DQ alpha 1 chain-like	MHC
3117	Major histocompatibility complex, class II, DQ alpha 1	MHC
56244	Butyrophilin-like 2 (MHC class II associated)	MHC
84166	NLR family, CARD domain containing 5	MHC Regulation
7940	Leukocyte specific transcript 1	Immune Regulation
80740	Lymphocyte antigen 6 complex, locus G6C	Immune Regulation
80741	Lymphocyte antigen 6 complex, locus G5C	Immune Regulation
1379	Complement component (3b/4b) receptor 1-like	Immune Regulation
1460	Lymphocyte antigen 6 complex, locus G5B	Immune Regulation
199	Allograft inflammatory factor 1	Immune Regulation
259197	Natural cytotoxicity triggering receptor 3	Immune Regulation
259215	Lymphocyte antigen 6 complex, locus G6F;	Immune Regulation
5819	Poliovirus receptor-related 2	Immune Regulation
58496	Lymphocyte antigen 6 complex, locus G5B	Immune Regulation
58530	Lymphocyte antigen 6 complex, locus G6F	Immune Regulation
6934	Transcription factor 7-like 2 (T-cell specific)	Immune Regulation
10665	Erythroid Differentiation-Related Factor 1	Immune Regulation
28	ABO blood group	Immune Regulation
5089	Pre-B-cell leukemia homeobox 2	Immune Regulation
63940	G-protein signaling modulator 3	Immune Regulation
177	Advanced glycosylation end product specific receptor	Inflammation
4050	Lymphotoxin beta (TNF superfamily, member 3)	Inflammation
10019	SH2B adaptor protein 3	Inflammation

(Continued)

Table 2 (Contd.)

Gene ID	Gene name	Function
10554	Lysophosphatidic acid acyltransferase, alpha	Inflammation/Lipid Signaling
64116	Solute carrier family 39 (zinc transporter), member 8	Other
1029[†]	Cyclin-dependent kinase inhibitor 2A (p16)	Cell Cycle Regulation (p16 *INK4a*)
1030	Cyclin-dependent kinase inhibitor 2B (p15)	Cell Cycle Regulation
2262	Glypican 5	Cell Signaling
4855	Notch homolog 4	Cell Signaling
6311	Ataxin 2	Cell Signaling/ Intracellular Trafficking
57827	Apoliprotein M	Cholesterol Metabolism/ Cell Signaling
60526	Apolipoprotein B	Cholesterol Metabolism/ Other
1071	Cholesteryl ester transfer protein, plasma	Cholesterol Metabolism
341	Apolipoprotein C-I	Cholesterol Metabolism
348	Apolipoprotein E	Cholesterol Metabolism
55937	Apolipoprotein M	Cholesterol Metabolism
217	Aldehyde dehydrogenase 2 family	Metabolism
3990	Lipase, hepatic	Metabolism
6048	Ring finger protein 5	Histone modification
6838	Surfeit 6	Nucleolar Protein/Ribosome Biogenesis
79068	Fat mass and obesity associated (FTO)	mRNA Regulation
10452[†]	TOM40	Mitochondrial Protein Import (APOE LOCUS)

Gene ID, name, and generic function, based on NCBI gene description, are indicated. *—Genes associated with all five disease categories. [†]—Genes previously reported to associate with multiple age-related diseases.

little direct evidence that this is true in normal human aging and age-related disease, and the role of canonical aging pathways in human age-related pathologies has not been established.

Genomewide association studies, a common approach to identifying genetic loci of importance to human complex traits in large populations, have led to the discovery of key genes in a variety of individual age-related diseases such as Alzheimer's disease (Kim *et al.*, 2014), cardiovascular disease (Sayols-Baixeras *et al.*, 2014), and a variety of age-associated cancers (Monteiro & Freedman, 2013; Barbieri & Tomlins, 2014; Veron *et al.*, 2014), but GWAS meta-analyses have also yielded little in regards to identifying shared genes and pathways, identifying only APOE, 9P21, and the HLA loci (Jeck *et al.*, 2012). The failure to identify canonical aging pathways by GWAS and GWAS meta-analyses has called into question whether these pathways are important to human longevity and health span. Determining whether age-related pathologies share common pathways is important given that current strategies aimed at developing interventions against age-related disease are based on the model that targeting underlying processes of aging can impact multiple diseases simultaneously.

Fig. 3 A pathway-based analysis of age-related diseases. (A) Age-related disease groups were analyzed using a pathway-based approach by examining the GO terms associated with genomewide association studies identified genes rather than the genes themselves. (B) A comparison between age-related disease categories shows a more significant similarity between ontology terms than was observed using individual genes (7.6% vs. 0.15%). (C) Although the relative percentage of overlapping terms is greater the relative similarities between disease categories are unchanged as determined by unsupervised clustering.
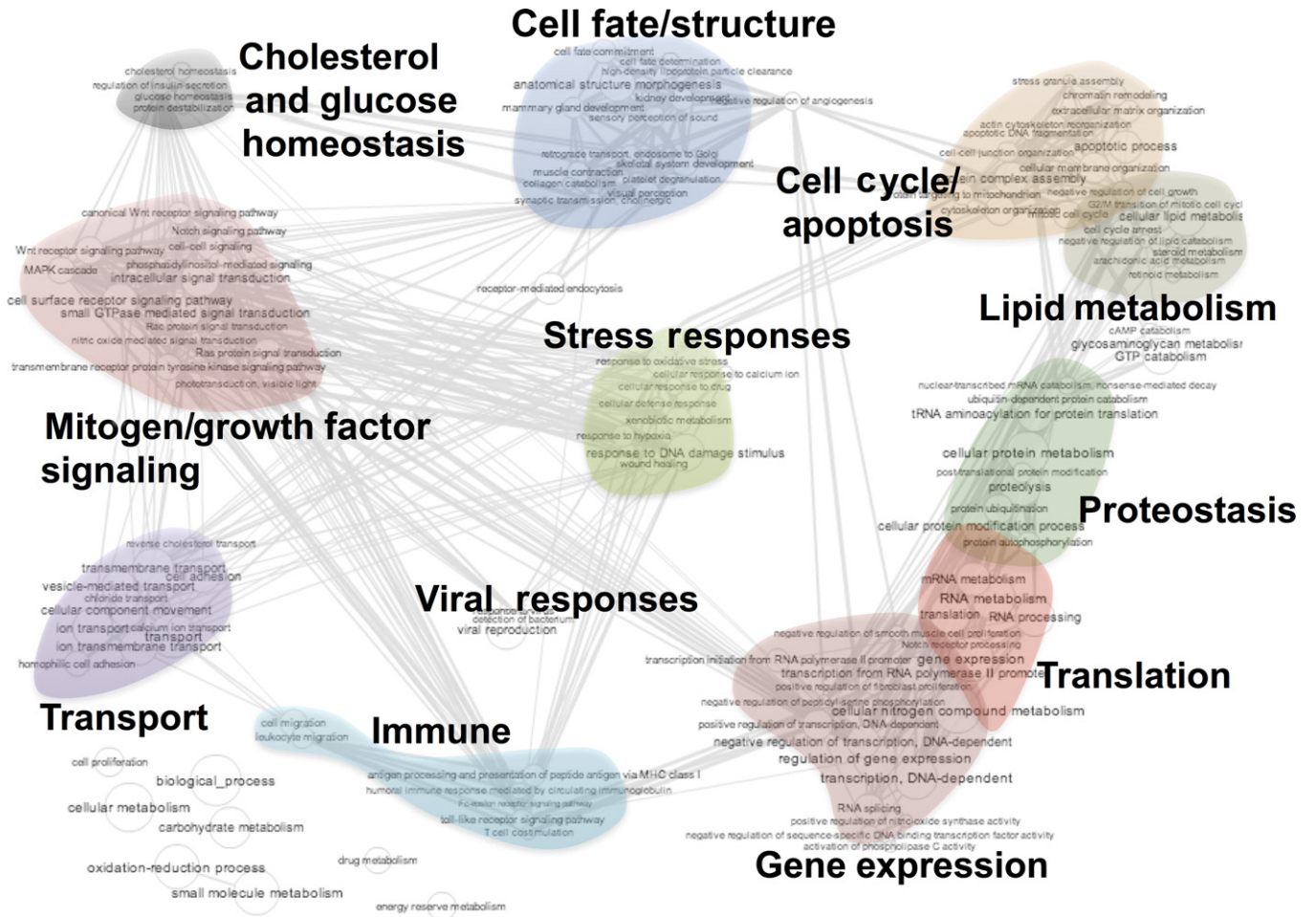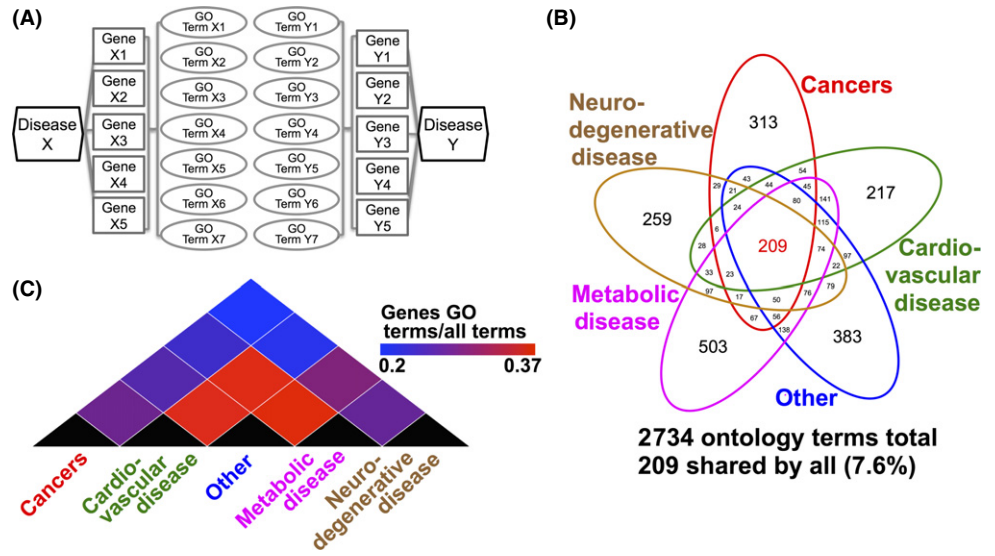

Fig. 4 Pathway analysis identified shared pathways among age-related diseases. Visualization of the GO terms shared by all five age-related disease categories reveals common pathways in human age-related disease. These include many canonical aging pathways identified in model organisms, such as nutrient-sensing signaling, proteostasis, and stress responses, as well as cholesterol metabolism, as was identified in the gene-based analysis. GO terms shared among age-related diseases are significantly enriched for similar terms compared to the background (P-value $< 10^{-15}$, one-sided Wilcoxon rank-sum test, see Methods).

Genomewide association studies approach has proven invaluable in identifying genes and regions of putative importance to human phenotypes and pathologies, but attempts to use GWAS to identify genetic variation that influences human longevity and health span have largely been unsuccessful (Newman *et al.*, 2010) (Beekman *et al.*, 2010) (Deelen *et al.*, 2014). Among known genetic modifiers of human aging, only the APOE locus has been reproducibly associated with both lifespan and health span by GWAS (Deelen *et al.*, 2011, 2014). The APOE locus is also strongly associated with neurological disease (particularly Alzheimer's disease) (Naj *et al.*, 2014; Wang *et al.*, 2014), cardiac, metabolic, and vascular diseases (Maxwell *et al.*, 2013; Hellwege *et al.*, 2014), plasma C-reactive protein levels (Ellis *et al.*, 2014; Schick *et al.*, 2014), and, in a recent study, nonpathological cognitive aging (Davies *et al.*, 2014). Thus, while GWAS have generally had difficulty in identifying genetic modulators of aging, apolipoprotein metabolism is an exception and appears to be a genuine GWAS detectable common pathway in human age-related disease. A more recent GWAS meta-analysis took a unique approach to studying aging by considering overall 'wellness', a surrogate for health span, based on the concept that successful aging would be related to broad resistance to disease (Jeck *et al.*, 2012). These authors assessed bins of GWAS SNPs, by chromosome location, according to the number of unique diseases they are associated with in an attempt to identify genes common to many diseases and, by inference, general health span although, importantly, it should be noted that the analysis was inclusive of all disease-associated SNPs rather than considering only age-related diseases. This approach identified a set of regions associated with immune function, including the ubiquitous MHC locus, as well as the 9p21 senescence-associated locus, both also identified in our study (Fig. 2), but failed to provide evidence supporting the role of conserved pathways such as IIS, mTOR, oxidative stress response, or genome maintenance.

Our gene-based findings suggest that while inflammation, immune regulation, and cholesterol metabolism are all broadly important in human aging, cholesterol metabolism genes alone are strikingly enriched among multiple age-related diseases. Multiple apolipoproteins have been associated with disease, and APOE is a particularly notable genetic loci in human health, as discussed. Consistent with these prior findings, our data suggest that apolipoprotein metabolism is a key underlying pathway in multiple human age-related diseases. Apolipoprotein genes have been associated with GWAS significance to a remarkable array of age-related pathologies including chronic renal disease, cardiovascular disease, inflammation, metabolic disease, hepatic dysfunction, alzheimer's, dementia, and cognitive decline (Melegh *et al.*, 2012; Schmidt *et al.*, 2012; Wasser *et al.*, 2012; Imes & Austin, 2013; Tosto & Reitz, 2013). Our findings suggest that apolipoprotein metabolism may represent a mammalian-specific underlying pathway in aging and age-related disease, supporting the notion that interventions in lipoprotein metabolism will provide significant benefits to human health. Epidemiological studies already support the adoption of earlier and more widespread statin use, and least one study has suggested that statins broadly affect the aging process (Boccardi *et al.*, 2013; Robinson, 2014). Clearly, apolipoprotein metabolism warrants continued attention as a safe and efficacious clinical target in aging.

In addition to providing further evidence supporting the critical importance of apolipoprotein metabolism in human age-related disease, here, we provide evidence supporting for the model that common, evolutionarily conserved pathways influence many age-related diseases. The data presented here provide new evidence supporting the continued

pursuit of interventions designed to combat age-related disease based on genetic pathways of aging discovered in lower organisms. While many of these pathways, such as genome maintenance and IIS/mTOR signaling, have already been implicated in human health, our study provides the first evidence that GWAS of age-related diseases show a signature of conserved pathways of aging. Finally, while our study focused on age-related disease, our novel pathway-based approach using gene ontology terms for comparison provides a new method for identifying shared pathways of disease. We anticipate that this approach can be applied to traits that are mechanistically poorly defined to provide novel insight into the pathogenesis of human diseases.

While our findings are supported by the previously published literature described our approach is not without limitations. The primary limitation of this study is the GWAS catalog itself—only SNPs identified in published GWAS at the significance level cut-off of $P < 10^{-5}$ are included in this analysis. Therefore, our results only represent the portion of all genes important for age-related diseases that have been successfully identified by GWAS. An additional caveat of our approach is that the initial step of assigning GWAS traits to disease and age-related vs. non-age-related categories introduces some interpretation in the analysis. While each subsequent step is completely unbiased differences in trait assignment may slightly affect the result. Fortunately, given the scarcity of overlapping genes and the weak impact that decreasing the P-value cut-off to $P < 10^{-7}$ has on the identified overlapping GO terms, it appears that minor alterations to trait assignment will have no major impact on the outcome using this approach. Finally, while our approach involved a GWAS P-value cut-off in the initial stage of analysis, it may also be possible to examine all GWAS SNPs for association with multiple age-related disease categories and rank them by Fisher's combined probability test P-value. This approach would likely identify additional associations.

Methods

GWAS data and associated genes

The GWAS catalog was downloaded from the National Human Genome Research Institute (http://www.genome.gov/26525384) (Welter *et al.*, 2014). At the time of download on November 9, 2013, the catalog contained 1738 GWAS, reporting 11 533 SNPs with assigned 'rs' numbers associated with 816 diseases/traits. To prevent bias due to limited numbers of studies for some diseases/traits, we limited our analyses to those with at least 5 independent GWAS reports resulting in 39 age-related diseases. This dataset includes a total of 410 independent GWAS (Table 1). SNPs were assigned to genes based if they are located on a gene or they are in high linkage disequilibrium (LD) with SNPs on a gene. The LD information is calculated using SCAN database (http://www.scandb.org/newinterface/about.html) (Gamazon *et al.*, 2010). These assignments produced 1975 protein-coding genes in high or complete LD or overlap with the diseases/trait-associated SNPs. Individual traits were assigned to one of five age-related disease categories, or determined to be non-age-associated, based on established criteria (see Results and (Perez-Lopez *et al.*, 2009; Martin, 2012; Johnson *et al.*, 2013; Brunet & Berger, 2014)).

GO enrichment analysis and visualization

Gene ontology (GO) analysis was performed using the Gene Ontology enRIchment anaLysis and visuaLizAtion tool (GORILLA) (http://cbl-gorilla.cs.technion.ac.il/) (Eden *et al.*, 2009), with graphical representations of gene ontology trees and enrichment P-values produced using

the linked REViGO tool (http://revigo.irb.hr/) (Supek *et al.*, 2011). Genes appearing 3 or more categories were analyzed against the background set of the entire gene list using the two unranked lists of genes analysis mode. Visualizations of GO terms found in the overlap between all age-related disease groups (Figs 4 and S2) were produced using REViGO.

GO term similarity measurement

Given that standard gene ontology enrichment analyses cannot be performed on this data, as there is no standardized approach to determining pathway enrichment from a GO term set (GO term enrichment is typically measured against a gene set), we considered the statistical value of this result using an *ad hoc* assessment. If the GO terms appearing in the overlapping term set represents an enrichment of specific processes, then we would expect the GO terms to be closely related as defined by their proximity in an ontology tree (Wang *et al.*, 2007). Briefly, we (i) aggregated the contributions of each GO term based on ancestor terms in the gene ontology tree to give a semantic value for the term; and (ii) we calculated the semantic similarity of each pair of GO terms based on these values (detailed method in (Wang *et al.*, 2007)). Adopting this method, pairwise similarities of GO terms in the overlap between the five disease groups, and in the background set, were estimated separately using an R package 'GOSemSim' based on GO biological process tree (Yu *et al.*, 2010). The significance of the mean value difference between the GO terms in the disease group overlap compared to the background set was determined using one-sided Wilcoxon rank-sum test.

P-value calculation in pathway-based approach

We calculated the pairwise similarity of this GO term group and compared it to the baseline set. The calculated similarity for the overlapping set was determined to be 0.15 ± 0.13, with the background set similarity calculated as 0.13 ± 0.1 (P-value is $< 1E-15$, one-sided Wilcoxon rank-sum test), indicating that the relative similarity of GO terms in the overlapping set could not be achieved by random selection.

Funding

Conflict of Interest

None declared.

References

Barbieri CE, Tomlins SA (2014) The prostate cancer genome: perspectives and potential. *Urol. Oncol.* **32**, 53 e15–22.

Beekman M, Nederstigt C, Suchiman HE, Kremer D, van der Breggen R, Lakenberg N, Alemayehu WG, de Craen AJ, Westendorp RG, Boomsma DI, de Geus EJ, Houwing-Duistermaat JJ, Heijmans BT, Slagboom PE (2010) Genome-wide association study (GWAS)-identified disease risk alleles do not compromise human longevity. *Proc. Natl Acad. Sci. USA* **107**, 18046–18049.

Berry A, Cirulli F (2013) The p66(Shc) gene paves the way for healthspan: evolutionary and mechanistic perspectives. *Neurosci. Biobehav. Rev.* **37**, 790–802.

Boccardi V, Barbieri M, Rizzo MR, Marfella R, Esposito A, Marano L, Paolisso G (2013) A new pleiotropic effect of statins in elderly: modulation of telomerase activity. *FASEB J.* **27**, 3879–3885.

Brunet A, Berger SL (2014) Epigenetics of aging and aging-related disease. *J. Gerontol.* **69**(Suppl 1), S17–S20.

Davies G, Harris SE, Reynolds CA, Payton A, Knight HM, Liewald DC, Lopez LM, Luciano M, Gow AJ, Corley J, Henderson R, Murray C, Pattie A, Fox HC, Redmond P, Lutz MW, Chiba-Falek O, Linnertz C, Saith S, Haggarty P, McNeill G, Ke X, Ollier W, Horan M, Roses AD, Ponting CP, Porteous DJ, Tenesa A, Pickles A, Starr JM, Whalley LJ, Pedersen NL, Pendleton N, Visscher PM, Deary IJ (2014) A genome-wide association study implicates the APOE locus in nonpathological cognitive ageing. *Mol. Psychiatry* **19**, 76–87.

Deelen J, Beekman M, Uh HW, Helmer Q, Kuningas M, Christiansen L, Kremer D, van der Breggen R, Suchiman HE, Lakenberg N, van den Akker EB, Passtoors WM, Tiemeier H, van Heemst D, de Craen AJ, Rivadeneira F, de Geus EJ, Perola M, van der Ouderaa FJ, Gunn DA, Boomsma DI, Uitterlinden AG, Christensen K, van Duijn CM, Heijmans BT, Houwing-Duistermaat JJ, Westendorp RG, Slagboom PE (2011) Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell* **10**, 686–698.

Deelen J, Beekman M, Uh HW, Broer L, Ayers KL, Tan Q, Kamatani Y, Bennet AM, Tamm R, Trompet S, Guethbjartsson DF, Flachsbart F, Rose G, Viktorin A, Fischer K, Nygaard M, Cordell HJ, Crocco P, van den Akker EB, Bohringer S, Helmer Q, Nelson CP, Saunders GI, Alver M, Andersen-Ranberg K, Breen ME, van der Breggen R, Caliebe A, Capri M, Cevenini E, Collerton JC, Dato S, Davies K, Ford I, Gampe J, Garagnani P, de Geus EJ, Harrow J, van Heemst D, Heijmans BT, Heinsen FA, Hottenga JJ, Hofman A, Jeune B, Jonsson PV, Lathrop M, Lechner D, Martin-Ruiz C, McNerlan SE, Mihailov E, Montesanto A, Mooijaart SP, Murphy A, Nohr EA, Paternoster L, Postmus I, Rivadeneira F, Ross OA, Salvioli S, Sattar N, Schreiber S, Stefansson H, Stott DJ, Tiemeier H, Uitterlinden AG, Westendorp RG, Willemsen G, Samani NJ, Galan P, Sorensen TI, Boomsma DI, Jukema JW, Rea IM, Passarino G, de Craen AJ, Christensen K, Nebel A, Stefansson K, Metspalu A, Magnusson P, Blanche H, Christiansen L, Kirkwood TB, van Duijn CM, Franceschi C, Houwing-Duistermaat JJ, Slagboom PE (2014) Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum. Mol. Genet.* **23**, 4420–4432.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48.

Ellis J, Lange EM, Li J, Dupuis J, Baumert J, Walston JD, Keating BJ, Durda P, Fox ER, Palmer CD, Meng YA, Young T, Farlow DN, Schnabel RB, Marzi CS, Larkin E, Martin LW, Bis JC, Auer P, Ramachandran VS, Gabriel SB, Willis MS, Pankow JS, Papanicolaou GJ, Rotter JI, Ballantyne CM, Gross MD, Lettre G, Wilson JG, Peters U, Koenig W, Tracy RP, Redline S, Reiner AP, Benjamin EJ, Lange LA (2014) Large multiethnic Candidate Gene Study for C-reactive protein levels: identification of a novel association at CD36 in African Americans. *Hum. Genet.* **133**, 985–995.

Fontana L, Partridge L, Longo VD (2010) Extending healthy life span–from yeast to humans. *Science* **328**, 321–326.

Fontana L, Kennedy BK, Longo VD, Seals D, Melov S (2014) Medical research: treat ageing. *Nature* **511**, 405–407.

Gamazon ER, Zhang W, Konkashbaev A, Duan S, Kistner EO, Nicolae DL, Dolan ME, Cox NJ (2010) SCAN: SNP and copy number annotation. *Bioinformatics* **26**, 259–262.

Ghosh S, Zhou Z (2014) Genetics of aging, progeria and lamin disorders. *Curr. Opin. Genet. Dev.* **26C**, 41–46.

Hellwege JN, Palmer ND, Raffield LM, Ng MC, Hawkins GA, Long J, Lorenzo C, Norris JM, Ida Chen YD, Speliotes EK, Rotter JI, Langefeld CD, Wagenknecht LE, Bowden DW (2014) Genome-wide family-based linkage analysis of exome chip variants and cardiometabolic risk. *Genet. Epidemiol.* **38**, 345–352.

Imes CC, Austin MA (2013) Low-density lipoprotein cholesterol, apolipoprotein B, and risk of coronary heart disease: from familial hyperlipidemia to genomics. *Biol. Res. Nurs.* **15**, 292–308.

Jeck WR, Siebold AP, Sharpless NE (2012) Review: a meta-analysis of GWAS and age-associated diseases. *Aging Cell* **11**, 727–731.

Johnson SC, Rabinovitch PS, Kaeberlein M (2013) mTOR is a key modulator of ageing and age-related disease. *Nature* **493**, 338–345.

Jurk D, Wilson C, Passos JF, Oakley F, Correia-Melo C, Greaves L, Saretzki G, Fox C, Lawless C, Anderson R, Hewitt G, Pender SL, Fullard N, Nelson G, Mann J, van de Sluis B, Mann DA, von Zglinicki T (2014) Chronic inflammation induces telomere dysfunction and accelerates ageing in mice. *Nat. Commun.* **2**, 4172.

Kenyon CJ (2010) The genetics of ageing. *Nature* **464**, 504–512.

Kim DH, Yeo SH, Park JM, Choi JY, Lee TH, Park SY, Ock MS, Eo J, Kim HS, Cha HJ (2014) Genetic markers for diagnosis and pathogenesis of Alzheimer's disease. *Gene* **545**, 185–193.

Li Q, Hu J, Ding J, Zheng G (2014) Fisher's method of combining dependent statistics using generalizations of the gamma distribution with applications to genetic pleiotropic associations. *Biostatistics* **15**, 284–295.

Lopez-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G (2013) The hallmarks of aging. *Cell* **153**, 1194–1217.

Martin GM (2012) Stochastic modulations of the pace and patterns of ageing: impacts on quasi-stochastic distributions of multiple geriatric pathologies. *Mech. Ageing Dev.* **133**, 107–111.

Maxwell TJ, Ballantyne CM, Cheverud JM, Guild CS, Ndumele CE, Boerwinkle E (2013) APOE modulates the correlation between triglycerides, cholesterol, and CHD through pleiotropy, and gene-by-gene interactions. *Genetics* **195**, 1397–1405.

Melegh BI, Duga B, Sumegi K, Kisfali P, Maasz A, Komlosi K, Hadzsiev K, Komoly S, Kosztolanyi G, Melegh B (2012) Mutations of the apolipoprotein A5 gene with inherited hypertriglyceridaemia: review of the current literature. *Curr. Med. Chem.* **19**, 6163–6170.

Merkwirth C, Martinelli P, Korwitz A, Morbin M, Bronneke HS, Jordan SD, Rugarli EI, Langer T (2012) Loss of prohibitin membrane scaffolds impairs mitochondrial architecture and leads to tau hyperphosphorylation and neurodegeneration. *PLoS Genet.* **8**, e1003021.

Monteiro AN, Freedman ML (2013) Lessons from postgenome-wide association studies: functional analysis of cancer predisposition loci. *J. Intern. Med.* **274**, 414–424.

Naj AC, Jun G, Reitz C, Kunkle BW, Perry W, Park YS, Beecham GW, Rajbhandary RA, Hamilton-Nelson KL, Wang LS, Kauwe JS, Huentelman MJ, Myers AJ, Bird TD, Boeve BF, Baldwin CT, Jarvik GP, Crane PK, Rogaeva E, Barmada MM, Demirci FY, Cruchaga C, Kramer PL, Ertekin-Taner N, Hardy J, Graff-Radford NR, Green RC, Larson EB, St George-Hyslop PH, Buxbaum JD, Evans DA, Schneider JA, Lunetta KL, Kamboh MI, Saykin AJ, Reiman EM, De Jager PL, Bennett DA, Morris JC, Montine TJ, Goate AM, Blacker D, Tsuang DW, Hakonarson H, Kukull WA, Foroud TM, Martin ER, Haines JL, Mayeux RP, Farrer LA, Schellenberg GD, Pericak-Vance MA, the Alzheimer Disease Genetics C (2014). Effects of multiple genetic loci on age at onset in late-onset Alzheimer disease: a genome-wide association study. *JAMA Neurol.* **11**, 1394–1404.

Newman AB, Walter S, Lunetta KL, Garcia ME, Slagboom PE, Christensen K, Arnold AM, Aspelund T, Aulchenko YS, Benjamin EJ, Christiansen L, D'Agostino RB Sr, Fitzpatrick AL, Franceschini N, Glazer NL, Gudnason V, Hofman A, Kaplan R, Karasik D, Kelly-Hayes M, Kiel DP, Launer LJ, Marciante KD, Massaro JM, Miljkovic I, Nalls MA, Hernandez D, Psaty BM, Rivadeneira F, Rotter J, Seshadri S, Smith AV, Taylor KD, Tiemeier H, Uh HW, Uitterlinden AG, Vaupel JW, Walston J, Westendorp RG, Harris TB, Lumley T, van Duijn CM, Murabito JM (2010) A meta-analysis of four genome-wide association studies of survival to age 90 years or older: the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium. *J. Gerontol.* **65**, 478–487.

Perez-Lopez FR, Chedraui P, Haya J, Cuadros JL (2009) Effects of the Mediterranean diet on longevity and age-related morbid conditions. *Maturitas* **64**, 67–79.

Robinson JG (2014) Starting primary prevention earlier with statins. *Am. J. Cardiol.* **114**, 1437–1442.

Sayols-Baixeras S, Lluis-Ganella C, Lucas G, Elosua R (2014) Pathogenesis of coronary artery disease: focus on genetic risk factors and identification of genetic variants. *Appl. Clin. Genet.* **7**, 15–32.

Schick UM, Auer PL, Bis JC, Lin H, Wei P, Pankratz N, Lange LA, Brody J, Stitziel NO, Kim DS, Carlson CS, Fornage M, Haessler J, Hsu L, Jackson RD, Kooperberg C, Leal SM, Psaty BM, Boerwinkle E, Tracy R, Ardissino D, Shah S, Willer C, Loos R, Melander O, McPherson R, Hovingh K, Reilly M, Watkins H, Girelli D, Fontanillas P, Chasman DI, Gabriel SB, Gibbs R, Nickerson DA, Kathiresan S, Peters U, Dupuis J, Wilson JG, Rich SS, Morrison AC, Benjamin EJ, Gross MD, Reiner AP, on Behalf of the Cohorts for H, Aging Research in Genomic E, the National Heart L, Blood Institute GOESP (2014) Association of exome sequences with plasma C-reactive protein levels in >9000 participants. *Hum. Mol. Genet.* **24**, 559–571.

Schleit J, Johnson SC, Bennett CF, Simko M, Trongtham N, Castanza A, Hsieh EJ, Moller RM, Wasko BM, Delaney JR, Sutphin GL, Carr D, Murakami CJ, Tocchi A, Xian B, Chen W, Yu T, Goswami S, Higgins S, Holmberg M, Jeong KS, Kim JR, Klum S, Liao E, Lin MS, Lo W, Miller H, Olsen B, Peng ZJ, Pollard T, Pradeep P, Pruett D, Rai D, Ros V, Singh M, Spector BL, Vander Wende H, An EH, Fletcher M, Jelic M, Rabinovitch PS, MacCoss MJ, Han JD, Kennedy BK, Kaeberlein M (2013) Molecular mechanisms underlying genotype-dependent responses to dietary restriction. *Aging Cell* **12**, 1050–1061.

Schmidt H, Freudenberger P, Seiler S, Schmidt R (2012) Genetics of subcortical vascular dementia. *Exp. Gerontol.* **47**, 873–877.

Sikora E (2014) Aging and longevity. *Postepy Biochem.* **60**, 125–137.

Supek F, Bosnjak M, Skunca N, Smuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800.

Torgovnick A, Schiavi A, Maglioni S, Ventura N (2013) Healthy aging: what can we learn from Caenorhabditis elegans? *Z. Gerontol. Geriatr.* **46**, 623–628.

Tosto G, Reitz C (2013) Genome-wide association studies in Alzheimer's disease: a review. *Curr. Neurol. Neurosci. Rep.* **13**, 381.

Veron A, Blein S, Cox DG (2014) Genome-wide association studies and the clinic: a focus on breast cancer. *Biomark. Med.* **8**, 287–296.

Vijg J, Suh Y (2013) Genome instability and aging. *Annu. Rev. Physiol.* **75**, 645–668.

Walter S, Atzmon G, Demerath EW, Garcia ME, Kaplan RC, Kumari M, Lunetta KL, Milaneschi Y, Tanaka T, Tranah GJ, Volker U, Yu L, Arnold A, Benjamin EJ, Biffar R, Buchman AS, Boerwinkle E, Couper D, De Jager PL, Evans DA, Harris TB, Hoffmann W, Hofman A, Karasik D, Kiel DP, Kocher T, Kuningas M, Launer LJ, Lohman KK, Lutsey PL, Mackenbach J, Marciante K, Psaty BM, Reiman EM, Rotter JI, Seshadri S, Shardell MD, Smith AV, van Duijn C, Walston J, Zillikens MC, Bandinelli S, Baumeister SE, Bennett DA, Ferrucci L, Gudnason V, Kivimaki M, Liu Y, Murabito JM, Newman AB, Tiemeier H, Franceschini N (2011) A genome-wide association study of aging. *Neurobiol. Aging* **32**, 2109 e2115–2128.

Wang JZ, Du Z, Payattakool R, Yu PS, Chen C-F (2007) A new method to measure the semantic similarity of GO terms. *Bioinformatics* **23**, 1274–1281.

Wang X, Lopez OL, Sweet RA, Becker JT, DeKosky ST, Barmada MM, Demirci FY, Kamboh MI (2014) Genetic determinants of disease progression in Alzheimer's disease. *J. Alzheimers Dis.* **43**, 649–655.

Wasser WG, Tzur S, Wolday D, Adu D, Baumstein D, Rosset S, Skorecki K (2012) Population genetics of chronic kidney disease: the evolving story of APOL1. *J. Nephrol.* **25**, 603–618.

Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson H (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–D1006.

Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* **26**, 976–978.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

Fig. S1 Enriched Biological Process GO terms in genes associated with multiple age-related diseases. Gene ontology analysis revealed cholesterol metabolism ontology terms as highly and specifically enriched among genes associated with multiple age-related disease categories by GWAS (cellular component terms in Fig. 2).

Fig. S2 Principal component analysis of gene ontology terms appearing in all 5 age-related disease categories. A principal component analysis plotted using REViGO. Color indicates term uniqueness while size represents the log size of the term group in the gene ontology tree. Key parent terms are indicated.

Fig. S3 Venn Diagrams of Genes and SNPs in Age-Related Disease Categories using an Initial GWAS Cut-off of $P < 10^{-7}$.

Table S1 GWAS traits not meeting our criteria for study number and statistical strength but not included in any of the 5 age-related disease categories.

Table S2 Enriched Gene Ontology Terms In Genes Shared by 3 or More Age-related Disease Groups. False discovery *P*-value is calculated using the Benjamani–Hochberg method.

Table S3 Select GO terms appearing in all 5 age-related disease categories. A full list of the 209 shared terms is provided in Table S4.

Table S4 A full list of enriched GO terms found in the overlapping term set.

Table S5 Genes Associated with the GWAS Trait Longevity at $P < 1 \times 10^{-5}$. See Table 2 for comparison with genes associated with multiple age-related diseases. Only APOE and APOCI are common between these two gene sets.

Table S6 Biological Process GO Terms Associated with Longevity GWAS Trait. Only GO terms with nominal *P*-value < 0.05 listed here. No GO terms remain significant following Benjamini multiple testing correction. GO terms associated with apolipoprotein metabolism indicated in bold. See Table S2 and S3 for comparison with GO terms found in all 5 age-related disease groups.

Table S7 Comparison of SNP, Gene, and GO Term Numbers Identified by GWAS Cut-off