Published in final edited form as:
Mech Ageing Dev. 2012 ; 133(9-10): 581–590. doi:10.1016/j.mad.2012.07.005.

Genotype × age interaction in human transcriptional ageing

Jack W. Kent Jr[1,*], **Harald H. H. Göring**[1], **Jac C. Charlesworth**[2], **Eugene Drigalenko**[1], **Vincent P. Diego**[1], **Joanne E. Curran**[1], **Matthew P. Johnson**[1], **Thomas D. Dyer**[1], **Shelley A. Cole**[1], **Jeremy B. M. Jowett**[3], **Michael C. Mahaney**[1], **Anthony G. Comuzzie**[1], **Laura Almasy**[1], **Eric K. Moses**[1,4], **John Blangero**[1], and **Sarah Williams-Blangero**[1]

[1]Department of Genetics, Texas Biomedical Research Institute, San Antonio, Texas 78245, USA

[2]Menzies Research Institute Tasmania, Hobart, Tasmania 7000, Australia

[3]Baker IDI Heart and Diabetes Institute, Melbourne, Victoria 3004, Australia

[4]Centre for Genetic Epidemiology and Biostatistics, University of Western Australia, Perth, Australia

Abstract

Individual differences in biological ageing (*i.e.*, the rate of physiological response to the passage of time) may be due in part to genotype-specific variation in gene action. However, the sources of heritable variation in human age-related gene expression profiles are largely unknown. We have profiled genome-wide expression in peripheral blood mononuclear cells from 1,240 individuals in large families and found 4,472 human autosomal transcripts, representing ~4,349 genes, significantly correlated with age. We identified 623 transcripts that show genotype by age interaction in addition to a main effect of age, defining a large set of novel candidates for characterization of the mechanisms of differential biological ageing. We applied a novel SNP genotype×age interaction test to one of these candidates, the ubiquilin-like gene *UBQLNL*, and found evidence of joint *cis*-association and genotype by age interaction as well as *trans*-genotype by age interaction for *UBQLNL* expression. Both *UBQLNL* expression levels at recruitment and *cis* genotype are associated with longitudinal cancer risk in our study cohort.

Keywords

Transcriptional ageing; genotype by age interaction; ubiquitins; *UBQLNL*; cancer risk gene

[*]Corresponding author: Jack W. Kent Jr., Dept. of Genetics, Texas Biomedical Research Institute, P.O. Box 760549, San Antonio, Texas 78245-0549, jkent@txbiomedgenetics.org, (v) 01.210.258.9570 (f) 01.210.258.9444.

Supplementary Information

A Supplementary Table S1 is provided that lists all 4,472 age-correlated transcripts ranked by $\log_{10}$(P-value) for their respective best-fit regression models vs. the null assumption of no correlation. Also noted is evidence of polygenic correlation-type G×AI at 1% and 5% FDR. Supplementary Table S2 provides information on all 10 transcripts showing significant or suggestive evidence of *trans*-interacting association. Supplementary Table S3 lists 83 transcripts associated by correlated expression or locus-specific genotype with *UBQLNL* expression.

Accession

ArrayExpress (www.ebi.ac.uk/arrayexpress/) accession number E-TABM-305 includes raw expression values of all transcripts on the microarray, normalized expression values of all 19,648 analysed autosomal transcripts, and information on subject sex and age.

1. Introduction

1.1. Transcriptional ageing

While there is on-going theoretical debate whether biological ageing is intrinsically 'programmed' or incidental to cumulative environmental effects (Medawar, 1952; Charlesworth et al., 2000; Holliday, 2006), both poles of the debate are consistent with individual variation in the rate at which ageing occurs (as variation in the developmental program on one hand, or as robustness to environmental insult on the other). Recent technological advances have enabled direct investigations of differences in global gene expression with age (Ly et al., 2000; Hawse et al., 2000; Welle et al., 2003, 2004; Lu et al., 2004; Rodwell et al., 2004; Melk et al., 2005; Storey et al., 2005; Zahn et al., 2006). To date, most human studies have compared gene expression between age classes of individuals (e.g., young vs. old), but such categorical comparisons do not reveal the trajectories of age-related change. Moreover, while many genes are expected to show age-related changes in expression, much of the individual variation in biological ageing may result from genotype×age interaction (G×AI) effects on a smaller number of genes. Indeed, interaction between specific genetic variants and environmental exposures (age included) are a possible source of the 'missing heritability' (Maher, 2006; Manolio et al., 2009) not yet accounted for in standard genome-wide association studies.

Here we examine the effects of age and GxAI on genome-wide transcriptional profiles of peripheral blood mononuclear cells (PBMCs) collected at the time of original recruitment from 1,240 members of extended families participating in the San Antonio Family Heart Study (SAFHS); (data collection described in Göring et al., 2007). Of 47,289 targets queried, 19,648 transcripts representing ~18,519 autosome-encoded genes were expressed at levels greater than background at a 5% false discovery rate (FDR; Benjamini & Hochberg, 1995). Approximately 85% of these phenotypes exhibited heritable variation in expression level at 5% FDR.

At the time of blood collection the study cohort had a median age = 37.6y (range 15.4–94.2y, interquartile range 25.1–49.6y). The broad range of ages represented by members of the pedigreed sample allowed us to assess age effects in the cross-sectional gene expression data. Our goal was to identify transcripts that change significantly with age and to test hypotheses about the genetic basis of differential transcriptional ageing, including G×AI.

1.2. Genotype × age interaction

Many genes are expected to show changes in expression with age, but (presumably) only a subset of these contribute to individual variation in transcriptional response to ageing. In classical studies of genetically identical model organisms (e.g., *Drosophila*, Dobzhansky & Spassky, 1944; pure-strain agricultural cultivars, Finlay & Wilkinson, 1963; algae, Bell, 1990) reared in different conditions, genotype × environment interaction revealed itself in divergent phenotype means (*norms of reaction*) across discrete environments. "Genotype" in these experiments typically referred to the genomic "type" as a whole. More recently, the capacity for transcriptional profiling has permitted identification of specific loci exhibiting genotype × environment interaction – e.g., by comparing gene expression in different culture conditions in yeast (Landry et al., 2006) and *C. elegans* (Li et al., 2006).

Extending these studies to uncontrolled environmental effects on outbred populations – specifically, humans – poses technical and analytical challenges that have only recently become tractable. A flurry of recent studies report the effects of observed environmental factors on allelotypes of specific candidate genes, including *ADH1B* genotype × alcohol use interaction effects on breast cancer risk (McCarty et al., 2012) and effects on behaviour of dopamine D4 receptor (*DRD4*) copy-number polymorphism interactions with measured

environmental factors (alcohol use, Creswell et al., 2012; socioeconomic status, Schweitzer et al., 2012).

In the present study, we undertook an agnostic search of all age-correlated expression phenotypes by performing formal tests of polygenic G×AI. Theory suggests that this interaction may take two forms (Bell, 1990; Blangero, 1993). The genetic variance of a given transcript (σ^2_g) may be age-dependent, suggesting differences in the *scale* of gene action (Fig. 1b). Alternatively (or simultaneously), the genetic correlation (ρ_g) between expression levels at different ages may be age-dependent, suggesting that the *relative effect sizes* of the genes contributing to a given trait vary with age (Fig. 1c). For our cross-sectional data, we extended the covariance-decomposition mixed models to express both σ^2_g and ρ_g as continuous functions of the age difference between individuals (Blangero & Konigsberg, 1991; Blangero, 1993; Almasy et al., 2001; Diego et al., 2003). The two tests are that σ^2_g is constant across all ages (the null for variance-type interaction) and that ρ_g between measures of the trait at different ages = +1 (the null for correlation-type interaction). In principle, of course, either, both, or neither null may be rejected, depending on the genes contributing to the trait of interest and their responses to age.

An analogous dichotomy of G×AI should be observable at the level of individual genes and genetic variants. Supposing the regulation of gene expression to be a polygenic process, the gene-specific equivalent of interaction σ^2_g could be reflected in an age-related change in variance attributable to a *cis*-acting variant (that is, a regulatory variant that is proximal to the physical location of a gene of interest), while the gene-specific equivalent of interaction ρ_g could be seen as a *trans*-acting variant – another regulatory site whose effect on expression of the gene of interest changes with age. We have developed a novel statistical test for distinguishing such effects and applied it to our polygenic G×AI candidates.

2. Materials and methods

2.1. Study cohort

Recruitment of the Mexican American families in SAFHS began in 1991 with ascertainment on family size rather than any disease state, although the cohort reflects the elevated risk of this ethnic stratum for Type 2 diabetes (15.3% at recruitment) and other cardiovascular risk factors (Mitchell et al., 1996). Subjects have been recalled up to three times to provide a wealth of genetic and phenotypic data. All procedures have been performed with the informed consent of the participants and with approval of the Institutional Review Board of the University of Texas Health Science Center – San Antonio.

The 1,240 SAFHS participants with gene expression data represent 46 extended families ranging in size from 3–87 phenotyped relatives. The complex family structure provided information on 12,548 pairwise relationships distributed across multiple households and over a very broad range of age difference (Table 1). All 1,240 individuals, including 107 marry-in spouses without phenotyped children, contributed to the estimation of the effect of SNP genotype.

2.2. Gene expression phenotypes

Collection, extraction, and standardization of the expression phenotypes are described in Göring et al. (2007). Briefly: RNA was extracted from PBMCs collected at recruitment and quantified using Illumina Sentrix Human Whole Genome Series I BeadChip microarrays. Average gene expression levels obtained from BeadStudio analysis of microarray output were standardized as Z-scores, adjusted for individual variation in overall signal, and normalized by inverse-Gaussian (rank-normal) transformation.

2.3. Genotyping

A panel of ~550K haplotype-tagging SNP genotypes (Illumina Bead Station 500 GX) was available for 1,189 of the SAFHS participants with gene expression phenotypes. The experimental error rate (based on duplicates) was 2 per 100,000 genotypes, and the average call rate per sample was 97%. Specific SNPs were removed from analysis if they had call rates <95% (about 4,000 SNPs) or deviated from Hardy-Weinberg equilibrium genotype frequencies at a 5% FDR (12 SNPs). SNP genotypes were checked for Mendelian consistency using the program SimWalk2 (Sobel & Lange, 1996); approximately 1 per 1,000 genotypes was blanked due to Mendelian errors. Maximum likelihood techniques that account for pedigree structure were used to estimate allelic frequencies. Missing genotypes were imputed with MERLIN (Burdick *et al.*, 2006). Association tests were performed using genotype scores that represented the number of copies of each SNP's minor allele (0,1,2, or a weighted fractional score for imputed genotypes).

2.4. Statistical and quantitative genetic analyses

The statistical programming language R (R Development Core Team, 2011) was used for descriptive statistics and graphics as well as the survival analyses (described in section 2.5).

Quantitative genetic analyses were conducted in SOLAR (Almasy & Blangero, 1998) with extensions written by the present authors. Because of the complex pedigree structure of the extended families, individual measures of gene expression could not be treated as statistically independent observations. Consequently, we analysed mixed models that included the random effect of kinship as well as the fixed effects of covariates. Briefly:

Observe that an individual's phenotype y_i can be decomposed as

$$y_i=\mu+\mathbf{x}_i\boldsymbol{\beta}+g_i+e_i \quad \text{(Eqn. 1)}$$

where μ is the phenotype mean, $\mathbf{x}_i$ a vector of covariate measures (including SNP genotype scores if desired), $\boldsymbol{\beta}$ a vector of regression beta coefficients, g_i the deviation from the mean due to additive genetic effects, and e_i an error term. After regressing out the covariate effects, the covariance of residual phenotypes between any two individuals i, j can be decomposed as

$$\mathrm{cov}_{i,j}=k_{i,j}\sigma^2 g+l_{i,j}\sigma^2 e \quad \text{(Eqn. 2)}$$

where $k_{i,j}$ is the expected proportion of alleles shared identical by descent (0.5 for parent-child or full sibs, 0.25 for 2nd degree relatives, etc.); σ^2_g and σ^2_e are, respectively, the additive genetic and residual components of total phenotypic variance; and $I_{i,j}$ is an indicator variable that is 1 if i and j are the same individual and 0 otherwise (Almasy & Blangero, 1998). The standardized additive genetic variance is called the *heritability* (h^2) – the proportion of total variance attributable to additive genetic effects. In practice, each system of pairwise covariance equations was fitted simultaneously to estimates of all parameters (regression and covariance decomposition alike) until maximum likelihood estimates were obtained. Note that estimating the SNP covariate betas (when present in the association tests) jointly with the other parameters provided an appropriate correction for the non-independent observations.

The linear decomposition of the phenotypic covariance is extremely flexible. In particular, the variance components in Eqn. 2 can be expressed as:

$$\sigma^2_g=\exp(\alpha_g+\beta_g\delta_i)\exp(\alpha_g+\beta_g\delta_j)\exp[-\lambda(\Delta\mathrm{age}_{i,j})] \quad \text{(Eqn. 3)}$$

$$\sigma^2_e = \exp(\alpha_e + \beta_e \delta_i)\exp(\alpha_e + \beta_e \delta_j) \quad \text{(Eqn. 4)}$$

such that σ^2_g and σ^2_e are replaced by exponential functions of each individual's deviation in age (δ_i, δ_j) from the sample mean. The additive genetic term also includes a genetic correlation between measures of the phenotype in i, j that depends on their absolute difference in age ($\Delta age_{i,j}$) – that correlation being 1 if $\Delta age_{i,j} = 0$. This expanded statement models the two theoretical types of G×AI (Blangero & Konigsberg, 1991; Almasy et al., 2001; Diego et al., 2003) with respect to the expected genome-wide allele sharing among relatives. We used this *polygenic* G×AI model to identify candidate gene expression phenotypes for more focussed SNP-based association tests.

SNP association tests were performed by including genotype scores as covariates in mixed models that also included covariates sex and age and the random effect of kinship. The first four principal components from analysis of genotype correlations were also included as covariates to guard against spurious association due to population structure (Price et al., 2006). We extended the association model to include a multiplicative age × SNP genotype interaction term – the SNP-specific equivalent of variance-type G×AI. To model correlation-type G×AI, we further tested this saturated model against a null model in which the regression slopes for *both* SNP genotype and the interaction term were constrained to 0 (a test with 2 degrees of freedom, df) to search for *trans*-acting variants whose association with the phenotype was detectable only contingent on interaction with age.

Genome-wide significance for the association tests was assessed based on an empirical threshold obtained from the distribution of *P*-values in 10,000 simulated null genome-wide association scans using the SAFHS SNP genotypes and pedigree structure. A test was declared 'significant' at $P<1.3\times10^{-7}$, the cut-off for the lower 5% tail of the empirical distribution, or 'suggestive' at a P-value not expected to occur more than once per genome scan ($P<1.9\times10^{-6}$). The same thresholds were used for the 2df tests of SNP G×AI (joint test of association with SNP genotype and age) although these criteria may be conservative (see Discussion).

2.5. Survival analysis

Mortality in the SAFHS cohort was assessed as of a reference date (31 October 2009) as follows: Copies of death certificates were obtained, where possible, for participants whose death had been reported by study recruiters. Individuals with death certificates were right-censored at date of death, and ICD10 codes for causes of death were supplied by a professional nosologist based on medical examiners. notes from de-identified death certificates. All other individuals were right-censored as 'alive' at the last date of contact with study recruiters. Incident cancer, diabetes, and cardiovascular events (heart attack or heart surgery) were recorded based on self-reports at SAFHS clinic visits. Cox proportional hazard analyses were performed using the R routine coxme which, like SOLAR, incorporates the random effect of kinship (Pankratz et al., 2005; Therneau, 2011).

3. Results

Our exploration of age-related gene expression proceeded as a stepwise prioritization of candidate genes: starting with all heritable, autosomal gene expression phenotypes, we filtered these by (a) association with age; (b) polygenic G×AI; (c) prior linkage evidence (Göring et al., 2007) for *cis*-regulation; (d) corresponding evidence of *cis*-association; (e) evidence of *trans*-interacting SNP G×AI. Finally, we present a detailed characterization of an interesting example, *UBQLNL*.

3.1. Age-related expression profiles

Every quantitative genetic analysis in SOLAR included age as covariate (see Methods). To examine trajectories of gene expression with age, we modified the age dependence of expression levels using four general regression functions: linear (the default), log-linear, quadratic, and a linear change-point function in which expression level reaches a plateau after a critical age. Sex was included as a covariate in each model and, as in all models, a residual genetic variance was estimated to account for non-independence among individuals due to kinship. The resulting mixed models for each transcript were ranked by the Bayesian information criterion (BIC; Raftery, 1995; Kass & Raftery, 1995), which took into account the additional parameters estimated in the quadratic and change-point models.

Of the 16,678 autosomal transcripts with heritable variation in expression, 4,472 (26.8%) were significantly correlated with age at FDR = 5% (*P*.0134). The best-fit regression model was linear for 54.8% of the transcripts and log-linear for 42.1%, consistent with an established tenet of gerontology that physiological processes relevant to senescence tend to decline in a linear fashion (Kohn, 1985; Shock, 1985; Sehl & Yates, 2001). The more complex change-point and quadratic regression models gave the best fit for 123 and 13 transcripts, respectively (Supplementary Table S1). For the transcripts whose profiles were best described by the linear change-point model, the estimated median change-point age (that is, the age at which linear change with age ceased) was 25.9y (range 19.3–65.7y). The SAFHS sample did not include juveniles, so processes specific to early development could not captured in this study.

The direction of change of expression with age for each transcript was defined by calculating the expected value of the expression phenotype at ages 25 and 70y using the best-fit regression model for the transcript. Significantly more transcripts showed decrease in expression level with age than increase (2,550 vs. 1,922; Fisher's exact test: odds ratio=1.327, $P=3.17\times10^{-11}$).

Pathway analysis by Ingenuity Pathway Analysis v4.1 software was employed to assign the age-correlated transcripts to broad categories of gene function and to look for over-representations of age-dependent transcripts. A right-tailed Fisher's exact test was performed for each category to compare the number of age-correlated genes in the category to the total number of occurrences of these genes in all categories of annotation. Functional assignments were obtained for 1,748 (37.3%) of the transcripts and 448 (9.6%) could be characterized further by canonical pathway (Fig. 1).

Several broad patterns emerge from pathway analysis. Among the top 12 functional assignments, there was an excess of up-regulation with age for transcripts related to immune response and inflammation, cell compromise, and cell death, as found in some earlier studies (Ly et al., 2000; Hawse et al., 2004; Lu et al., 2004). Eleven of the top 12 canonical pathway assignments exhibited an excess of up-regulation with age. The prominence of canonical pathway assignments related to immune response, oxidative stress, and cellular damage may reflect expression patterns specific to the lymphocyte source tissue as well as age effects. The functional patterns of age-related change in expression are suggestive of both diminished cellular maintenance and accumulation of cellular damage with age, consistent with theoretical expectation (Vijg & Suh, 2005). These results are broadly in agreement with a meta-analysis of our data by Hong *et al.* (2008), although their analysis did not account for the non-independence of family members and used different criteria for correcting for multiple tests. In particular, the meta-analysis noted the preponderance of down- vs up-regulated transcripts and the enrichment for inflammation-related genes in the PBMC-based expression data.

3.2. Genotype × age interaction: polygenic

Of the 4,472 transcripts that were significantly correlated with age, none exhibited variance-type G×AI after correction for multiple testing; however, 623 (13.94%) exhibited significant correlation-type G×AI at 5% FDR ($P \leq 0.007$; Supplementary Table S1). These results suggest that, in this sample, individual variation in the response of gene expression to ageing primarily reflects age-differential contributions of multiple loci. The functional assignments for this subset are shown in Fig. 2.

3.2. Genotype × age interaction: localized

To begin detailed exploration of the molecular basis of transcriptional ageing, we focused our attention on a very small subset of 17 G×AI transcripts that had prior evidence of *cis*-regulation (*i.e.*, the primary signal in a genome-wide linkage scan was localized at or near the structural locus of the expressed gene; Göring et al., 2007) confirmed by association tests including tests for *cis*- and *trans*-acting SNP G×AI (see Methods).

Of the 17 expression phenotypes analysed, three (*UBQLNL*, *VSTM1*, *ZNF638*) gave significant evidence ($P < 1.3\times10^{-7}$), based on the stringent 2 df test, of correlation-type SNP G×AI at genomic positions other than the primary *cis*-association loci (Supplementary Table S2). Seven (*ABCC3*, *SNRNP25*, *GRHPR*, *HNRNPL*, *MT2A*, *PIGB*, and *TMEM8A*) also showed suggestive evidence ($P < 1.9\times10^{-6}$) of correlation-type G×AI at one or more *trans* loci (Supplementary Table S2).

3.4 *UBQLNL*

NM_145053, a transcript of the ubiquilin-like gene *UBQLNL*, attracted our attention as a member of the ubiquitin gene family whose biological functions are not yet well understood. Two SNPs within 1Mb of the chromosome 11 *UBQLNL* structural gene locus were associated with *UBQLNL* expression at genome-wide significance, consistent with the evidence for *cis*-linkage (Table 2; Fig 3, Panel A). SNPs rs7939159 ($P = 3.7\times10^{-10}$, df=1) and rs7129909 ($P = 1.1\times10^{-8}$, df=1) are located within a tripartite motif-containing protein gene cluster (*TRIM34* and *TRIM22*, respectively) and are in low linkage disequilibrium (LD; $r = 0.061$) in our sample.

In addition, three SNPs at a *trans*-locus near *PPP1R3C* on chromosome 10 showed significant evidence by the 2 df test for association contingent on G×AI (Table 2; Fig. 3, Panel B). These SNPs were in near complete LD ($r > 0.99$) with one another but not with the *cis*-acting variants ($|r| < 0.05$). For subsequent analyses, rs11597974 ($P = 2.8\times10^{-8}$, df=2) was taken as representative of the set. While physically located near *PPP1R3C*, these SNPs were not associated with its expression, and *PPP1R3C* expression (which was not heritable) was not correlated with that of *UBQLNL*. Thus, the *trans*-association of these SNPs with *UBQLNL* does not appear to be mediated by a *cis*-effect on the nearby gene.

We applied a Bayesian test (Blangero et al., 2009) of all possible combinations of these three SNPs and their interactions with age. The best model includes as covariates the *cis*-acting SNPs and the *trans*-acting G×AI term, each with posterior probability = 1. Cumulatively, these three covariates account for 11.6% of the total phenotypic variance, with 2.5% of total variance (= 21.6% of the covariate effect) attributable to *trans*-G×AI (Table 3).

G×AI can be visualized by *joint regression* (Finlay & Wilkinson, 1963) of the phenotypic values of specified genotypes on the environment of interest. In the absence of G×AI, the norms of reaction of each genotype should be similar in slope but (assuming association of genotype with trait) different in intercept. The left panels of Figure 4 show the norms of

reaction of *UBQLNL* expression on age for the two *cis*-acting SNPs. The intercepts at mean age are distinct and consistent with an additive effect of the alleles, while the overall trend for all *cis*-genotypes is age-decreasing, with decrease in variance with age (reflected in the convergence of genotype means). The remaining columns in the figure, which stratify the results by *trans*-genotype, reveal a more complex response: the *trans* minor allele is evidently associated with increased *UBQLNL* expression with age, countering the overall trend. The contrast between the major- and minor-allele *trans* homozygotes, for which the slopes of the norms of reaction are generally opposite in sign, is a classic indicator of correlation-type G×AI (Bell, 1990). An exception to this pattern is the rare allele of rs7939159, which shows less variation in relation to *trans*-genotype.

3.5. Relationship to clinical outcomes

As noted, our gene expression phenotypes were obtained from PMBCs collected at baseline, and evidence for age correlation and G×AI is based on cross-sectional data. In the ~16y since initial recruitment, SAFHS participants have been recalled up to 3 times for follow-up and companion studies, providing longitudinal data on the same cohort (Fig. 5). 158 participant deaths have been identified during recalls; of 123 decedents for whom death certificates were available, cancer was primary or contributing cause of death for 36, the second most common cause after cardiovascular disease (N=73). In addition, medical histories taken at each clinic visit included self-reports of incident cancer for 51 surviving participants. We employed a mixed-model Cox proportional hazards analysis implemented in R routine coxme (Pankratz et al., 2005) that, like our SNP association method, includes a random effect of kinship. Both *UBQLNL* expression level at baseline and rs7939159 genotype had a nominally significant effect (P<0.05) on all-cause and cancer mortality and significantly affected combined cancer outcomes (expression, P=0.0016; rs7939159 genotype, P=0.00007; Table 4). Neither *UBQLNL* expression nor rs7939159 genotype was a significant hazard for death by cardiovascular disease. While these results should be treated with caution, given the small number of cancer cases and the heterogeneity of cancer type, they do suggest a possible physiological role for *UBQLNL* that should be addressed in future studies.

In a massive whole-transcriptome correlation analysis of the SAFHS expression data (H.H.H.G., E.I.D., unpublished), *UBQLNL* expression is significantly correlated with 80 other validated autosomal transcripts of genes involved in mitochondrial function and cell cycle regulation (Supplementary Table S3). In addition, the panel of *cis*-acting and *trans*-interacting SNPs is associated (at either genome-wide significance or suggestive evidence) with expression of three of these genes, including a key responder to DNA damage, the ataxia-telangiectasia mutated gene *ATM* (Supplementary Table S3). These relationships, combined with the Cox proportional hazards data, suggest that variation in *UBQLNL* expression may reflect genotype-specific differences in cellular response to age-related accumulation of DNA damage.

4. Discussion

Our analyses of transcriptional ageing are based on one of the largest samples studied to date. Due to the relative ease of obtaining PBMCs, we were able to assess genome-wide expression profiles in a sample of more than 1,200 individuals. This unique dataset allowed us to characterize trajectories of change in expression with age, revealing a diversity of trajectories consistent with observations on physiological traits (Arking, 2010). Use of familial information enabled us to identify a substantial number of genes whose level of expression is both age-correlated and heritable, thus defining a novel set of biomarkers of ageing. These new biomarkers are by definition directly proximal to gene action, providing a high-resolution resource for characterizing pathways of physiological ageing. Importantly,

the genes showing polygenic G×AI effects represent high-priority targets like *UBQLNL* for molecular dissection to determine the mechanisms of differential biological ageing.

As noted in the Introduction, there has been limited discovery to date of specific genetic variants in humans that clearly exhibit genotype × environment interaction, and published studies typically focus on candidate genes and/or categorical environments. Here we present a method for screening large numbers of genetic variants in a *continuous* environment (age) for associations that are detectable only through interaction. Our threshold ($P<1.3\times10^{-7}$) for detection of either association or G×AI is based on a large simulation of GWAS under the null, with a stringent 2df test for G×AI. This test may be conservative: for the three transcripts exhibiting *trans*-G×AI significant at our threshold (*UBQLNL*, *VSTM1*, *ZNF638*), the levels of evidence were well below 5% FDR ($q_i = 0.007, 0.004, 0.002$, respectively). Thus, the evidence for these discoveries is strong; further work is needed to fully characterize the null distribution of this novel test.

Little is currently known about the biological function of *UBQLNL*, although its sequence homology with other members of the ubiquitin gene family suggests that it may share a role in protein turnover and regulation of cell cycle progression (http://www.ncbi.nlm.nih.gov/gene/143630; accessed 8/15/2011). *UBQLNL* is a positional candidate gene for type 2 diabetes in at least one genome-wide association study in Mexican Americans (Hayes et al., 2007), although its expression is not significantly correlated with diabetes status in our sample (data not shown). It is also one of many genes included in a proposed genetic risk score for response to nicotine-patch therapy for cigarette addiction (Rose et al., 2010; Uhl et al., 2010), although it was not identified in our profile of smoking-related gene expression in SAFHS (Charlesworth et al., 2010). Our panel of *cis*-acting/*trans*-interacting SNPs is not associated with smoking in SAFHS (data not shown).

Additional molecular work is needed to clarify the functions of *UBQLNL* and the *trans*-interacting variants on chromosome 10. The published data suggest that *UBQLNL* may function in response to a variety of environmental threats to cellular or genomic integrity, including oxidative stress associated with diabetes and toxic effects of smoking. As noted, rs7939159 genotype was a significant hazard in the mixed-model Cox proportional hazards analysis, with the odds ratio suggesting increased risk of cancer outcomes associated with the minor allele (Table 4) – the same allele that showed a paradoxical response to *trans*-G×AI in the joint regression (Fig. 4).

Conclusion

We have identified a large set of human genes exhibiting G×AI in expression levels and have implemented a method for distinguishing *cis*- and *trans*-acting variants contributing to localized patterns of G×AI. This work sets the stage for detailed exploration of the genetic processes contributing to differential biological ageing, exemplified by our application of these methods to a candidate ubiquitin-family gene, *UBQLNL*. We have found that *UBQLNL* expression levels as well as variation in an age-interacting associated SNP are associated with subsequent cancer in our sample, demonstrating the utility for gene discovery of PBMC-derived gene expression phenotypes in general and analysis of age effects in particular.

Individual genes are never expressed in isolation, as exemplified by the extensive genetic correlation of *UBQLNL* with other transcripts. Development of analytical tools for incorporating high-dimensional relationships and multiple types of genomic information is on-going (see, e.g., Zhu et al., 2012). A most interesting challenge will be to apply G×AI

analytical techniques not only to individual gene expression phenotypes but also to the networks that comprise them.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the San Antonio Family Heart Study subjects for their on-going participation in genetic research. Our work was supported by US National Institutes of Health grants AG031277 (transcriptional ageing), HL045522 (SAFHS), MH059490 (computational methods development), and RR013556 and RR017515 (facilities). SNP genotyping and PBMC expression profiling were funded by the Azar/Shepperd families, San Antonio, Texas, with additional support of expression profiling from ChemGenex Pharmaceuticals, Ltd., Australia. The survival survey was supported by a pilot grant to J.W.K by the Max and Minnie Tomerlin Voelcker Foundation, San Antonio, Texas. The supercomputing facilities used for this work were funded in part by the AT&T Foundation.

Abbreviations

df	degrees of freedom
FDR	false discovery rate
GxAI	genotype by age interaction
LD	linkage equilibrium
PBMC	peripheral blood mononuclear cells
SAFHS	San Antonio Family Heart Study

References

Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. Am. J. Hum. Genet. 1998; 62:1198–1211. [PubMed: 9545414]

Almasy L, Towne B, Peterson C, Blangero J. Detecting genotype×age interaction. Genet. Epidemiol. 2001; 21(suppl. 1):S819–S824. [PubMed: 11793786]

Arking, R. The Biology of Aging: Observations and Principles. 3rd ed.. New York: Oxford University Press; 2006.

Bell G. The ecology and genetics of fitness in *Chlamydomonas* I. Genotype-by-environment interaction among pure strains. Proc. R. Soc. Lond. B. 1990; 240(1298):295–321.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. Royal Stat. Soc. Ser. B. 1995; 57(1):289–300.

Blangero J. Statistical genetic approaches to human adaptability. Hum. Biol. 1993; 65(6):941–966. [PubMed: 8300087]

Blangero J, Konigsberg LW. Multivariate segregation analysis using the mixed model. Genet. Epidemiol. 1991; 8:299–316. [PubMed: 1761203]

Blangero J, Göring HHH, Kent JW, Williams JT, Peterson CP, Almasy L, Dyer TD. Quantitative trait nucleotide analysis using Bayesian model selection. Hum. Biol. 2009; 81(5–6):829–847. [Republished from Hum. Biol. 77 (5), 541-559, 2005.]. [PubMed: 20504200]

Burdick JT, Chen W-M, Abecasis GR, Cheung VG. In silico method for inferring missing genotypes in pedigrees. Nat Genet. 2006; 38:1002–1004. [PubMed: 16921375]

Charlesworth B. Fisher, Medawar, Hamilton and the evolution of aging. Genetics. 2000; 156(3):927–931. [PubMed: 11063673]

Charlesworth JC, Curran JE, Johnson MP, Göring HHH, Dyer TC, Diego VP, Kent JWJr, Mahaney MC, Almasy L, MacCluer JW, Moses EK, Blangero J. Transcriptomic epidemiology of smoking:

The effect of smoking on gene expression in lymphocytes. BMC Med. Genomics. 2010; 3:29. [PubMed: 20633249]

Creswell KG, Sayette MA, Manuck SB, Ferrell RE, Hill SY, Dimoff JD. *DRD4*polymorphism moderates the effect of alcohol consumption on social bonding. PLoS One. 2012; 7(2):e28914. [PubMed: 22347363]

Diego VP, Almasy L, Dyer TD, Soler JMP, Blangero J. Strategy and model building in the fourth dimension: A null model for genotype×age interaction as a Gaussian stationary stochastic process. BMC Genet. 2003; 4(Suppl. 1):S34. [PubMed: 14975102]

Dobzhansky T, Spassky B. Genetics of natural populations. XI. Manifestation of genetic variants in *Drosophila pseudo-obscura* in different environments. Genetics. 1944; 29:270–290. [PubMed: 17247121]

Finlay KW, Wilkinson GN. The analysis of adaptation in a plant-breeding programme. Aust.J. Agric. Res. 1963; 14:742–754.

Göring HHH, Curran JE, Johnson MP, Dyer TD, Charlesworth JC, Cole SA, Jowett JB, Abraham LJ, Rainwater DL, Comuzzie AG, Mahaney MC, Almasy L, MacCluer JW, Kissebah AH, Collier GR, Moses EK, Blangero J. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. Nat. Genet. 2007; 39(10):1208–1216. [PubMed: 17873875]

Hawse JR, Hejtmancik JF, Horwitz J, Kantorow M. Identification and functional clustering of global gene expression differences between age-related cataract and clear human lenses and aged human lenses. Exp. Eye Res. 2004; 79(6):935–940. [PubMed: 15642332]

Hayes MG, Pluzhnikov A, Miyake K, Sun Y, Ng MC, Roe CA, Below JE, Nicolae RI, Konkashbaev A, Bell GI, Cox NJ, Hanis CL. Identification of type-2 diabetes genes in Mexican Americans through genome-wide association studies. Diabetes. 2007; 56(12):3033–3044. [PubMed: 17846124]

Holliday R. Aging is no longer an unsolved problem in biology. Ann. NY Acad. Sci. 2006; 1067:1–9. [PubMed: 16803964]

Hong M-G, Myers AJ, Magnusson PKE, Prince JA. Transcriptome-wide assessment of human brain and lymphocyte senescence. PLoS One. 2008; 3(8):e3024. [PubMed: 18714388]

Kass RE, Raftery AE. Bayes factors. J. Am. Stat. Assoc. 1995; 90():773–795.

Kohn RR. Aging and age-related diseases: Normal processes. Aging. 1985; 28:1–44. 1985.

Landry CR, Oh J, Hartl DL, Cavalieri D. Genome-wide scan reveals that genetic variation for transcriptional plasticity in yeast is biased toward multi-copy and dispensable genes. Gene. 2006; 366:343–351. [PubMed: 16427747]

Li Y, Álvarez OA, Gutteling EW, Tijsterman M, Fu J, Riksen JAG, Hazendonk E, Prins P, Plasterk RHA, Jansen RC, Breitling R, Kammenga JE. Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. PLoS Genet. 2006; 2(12):e222. [PubMed: 17196041]

Lu T, Pan Y, Kao SY, Li C, Kohane I, Chan J, Yankner BA. Gene regulation and DNA damage in the ageing human brain. Nature. 2004; 429(6994):883–891. [PubMed: 15190254]

Ly DH, Lockhart DJ, Lerner RA, Schultz PG. Mitotic misregulation and human aging. Science. 2000; 287(5462):2486–2492. [PubMed: 10741968]

Maher B. Personal genomes: The case of the missing heritability. Nature. 2008; 456(7218):18–21. [PubMed: 18987709]

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009; 461(7265):747–753. [PubMed: 19812666]

Mitchell BD, Kammerer CM, Blangero J, Mahaney MC, Rainwater DL, Dyke B, Hixson JE, Henkel RD, Sharp RM, Comuzzie AG, VandeBerg JL, Stern MP, MacCluer JW. Genetic and environmental contributions to cardiovascular risk factors in Mexican Americans: The San Antonio Family Heart Study. Circulation. 1996; 94(9):2159–2170. [PubMed: 8901667]

Medawar, PB. An Unsolved Problem of Biology. London: HK Lewis & Co.; 1952.

Melk A, Mansfield ES, Hsieh SC, Hernandez-Boussard T, Grimm P, Rayner DC, Halloran PF, Sarwal MM. Transcriptional analysis of the molecular basis of human kidney aging using cDNA microarray profiling. Kidney Int. 2005; 68(6):2667–2679. [PubMed: 16316342]

Pankratz VS, de Andrade M, Therneau TM. Random-effects Cox proportional hazards model: General variance components methods for time-to-event data. Genet. Epidemiol. 2005; 28(2):97–109. [PubMed: 15532036]

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 2006; 38(8): 904–909. [PubMed: 16862161]

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria: ISBN 3-900051-07-0; 2011. URL: http://www.R-project.org/.

Raftery, AE. Bayesian model selection in social research. In: Marsden, PV., editor. Sociological Methodology. Vol. 1995. Oxford: Blackwell; 1995. p. 111-195.

Rodwell GEJ, Sonu R, Zahn JM, Lund J, Wilhelmy J, Wang L, Xiao W, Mindrinos M, Crane E, Segal E, Myers BD, Brooks JD, Davis RW, Higgins J, Owen AB, Kim SK. A transcriptional profile of aging in the human kidney. PLoS Biol. 2004; 2(12):e427. [PubMed: 15562319]

Rose JE, Behm FM, Drgon T, Johnson C, Uhl GR. Personalized smoking cessation: Interactions between nicotine dose, dependence and quit-success genotype score. Mol. Med. 2010; 16(7–8): 247–253. [PubMed: 20379614]

Sehl ME, Yates FE. Kinetics of human aging: I. Rates of senescence between ages 30 and 70 years in healthy people. J. Gerontol Biol. Sci. 2001; 56(5):B198–B208.

Shock, NW. The physiological basis of aging. In: Morin, RJ.; Bing, RJ., editors. Frontiers in Medicine: Implications for the Future. New York: Human Sciences Press, Inc.; 1985. p. 300-312.

Storey JD, Xiao W, Leek JT, Tompkins RG, Davis RW. Significance analysis of time course microarray experiments. Proc. Nat. Acad. Sci. USA. 2005; 102(36):12837–12842. [PubMed: 16141318]

Therneau, T. coxme: Mixed effects Cox models. R package version 2.1-2. 2011. URL: http://CRAN.R-project.org/package=coxme.

Uhl GR, Drgon T, Johnson C, Ramoni MF, Behm FM, Rose JE. Genome-wide association for smoking cessation success in a trial of precessation nicotine replacement. Mol. Med. 2010; 16(11–12):513–526. [PubMed: 20811658]

Vijg J, Suh Y. Genetics of longevity and aging. Ann. Rev. Med. 2005; 56:193–212. [PubMed: 15660509]

Welle S, Brooks AI, Delehanty JM, Needler N, Thornton CA. Gene expression profile of aging in human muscle. Physiol. Genomics. 2003; 14(2):149–159. [PubMed: 12783983]

Welle S, Brooks AI, Delehanty JM, Needler N, Bhatt K, Shah B, Thornton CA. Skeletal muscle gene expression profiles in 20–29 year old and 65–71 year old women. Exp. Gerontol. 2004; 39(3):369–377. [PubMed: 15036396]

Zahn JM, Sonu R, Vogel H, Crane E, Mazan-Mamczarz K, Rabkin R, Davis RW, Becker KG, Owen AB, Kim SK. Transcriptional profiling of aging in human muscle reveals a common aging signature. PLoS Genet. 2006; 2(7):e115. [PubMed: 16789832]

Zhu J, Soval P, Xu Q, Dombeck KM, Xu EY, Vu H, Tu Z, Brem RB, Bumgarner RE, Schadt EE. Stitching together multiple data dimensions reveals interacting metabolomics and Transcriptomic networks that modulate cell regulation. PLoS Biol. 2012; 10(4):e1001301. [PubMed: 22509135]

Highlights

> At least 4300 human genes show age-related changes in expression. > At least 623 of these show genotype by age interaction (GxAI) effects on expression. > *UBQLNL* expression exhibits both *cis*- and *trans*-acting GxAI effects. > A *cis*-regulatory SNP in *UBQLNL* is associated with lifetime risk of cancer.

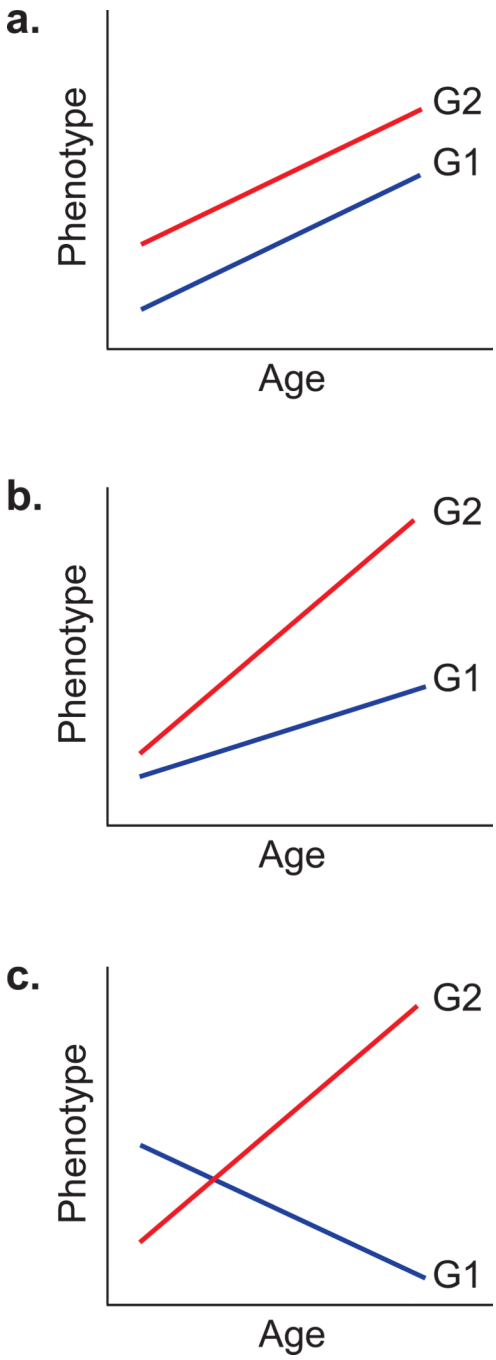
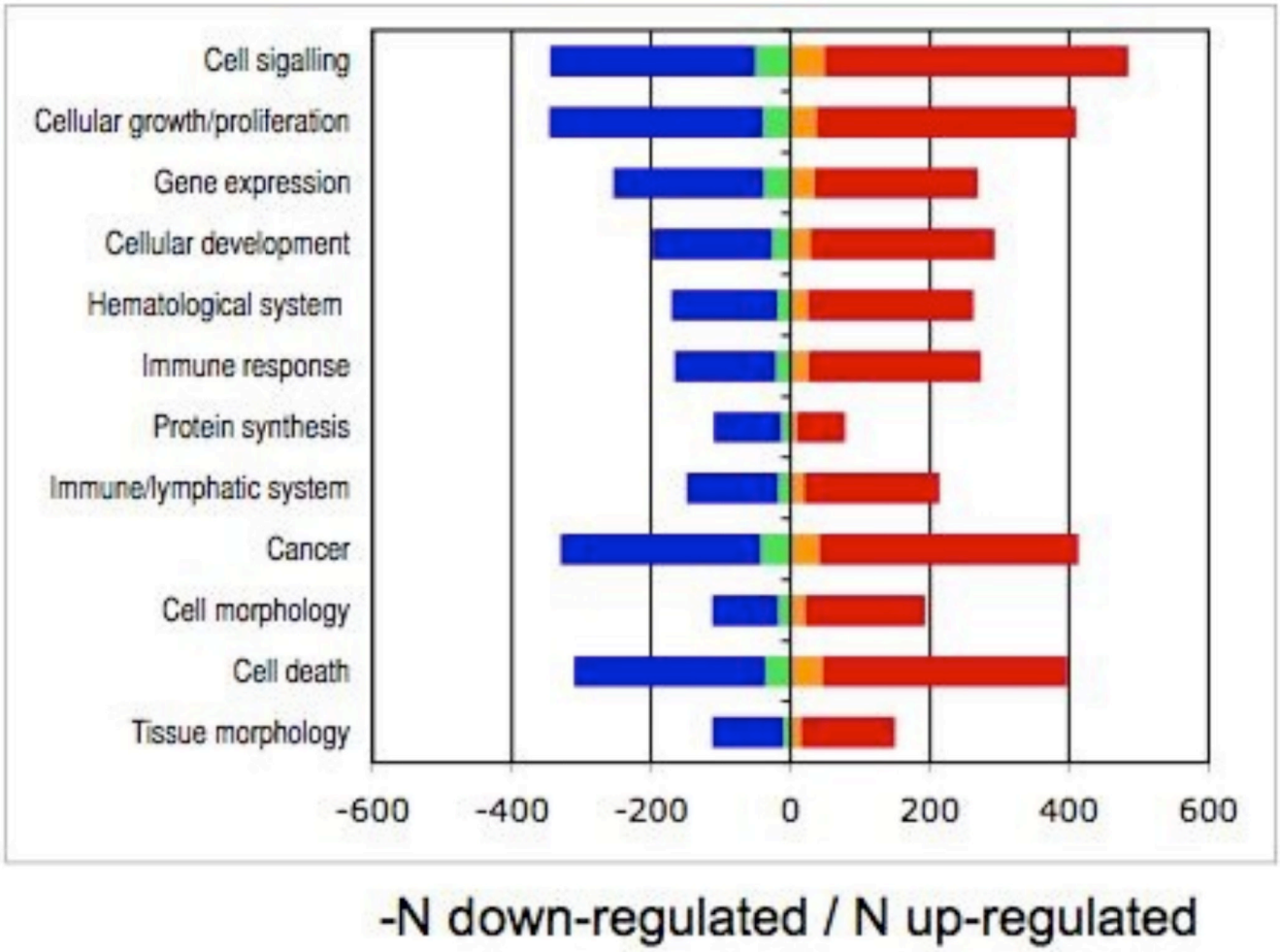


Figure 1. Hypothetical norms of reaction

mean expression levels of a hypothetical gene for two genotypes G1, G2 measured at different ages (after Bell (1990)). **a**, no interaction: G2 is more highly expressed at all ages, so genotype responses to age have correlation=+1; trait variance does not change with age. **b**, variance-type interaction: G2 is always more highly expressed (correlation=+1) but genotype means diverge with age (and consequently variance increases with age). **c**, correlation-type interaction: responses of G1 and G2 do not have correlation=+1 since ranks of G1, G2 change with age; trait variance does not change monotonically with age.



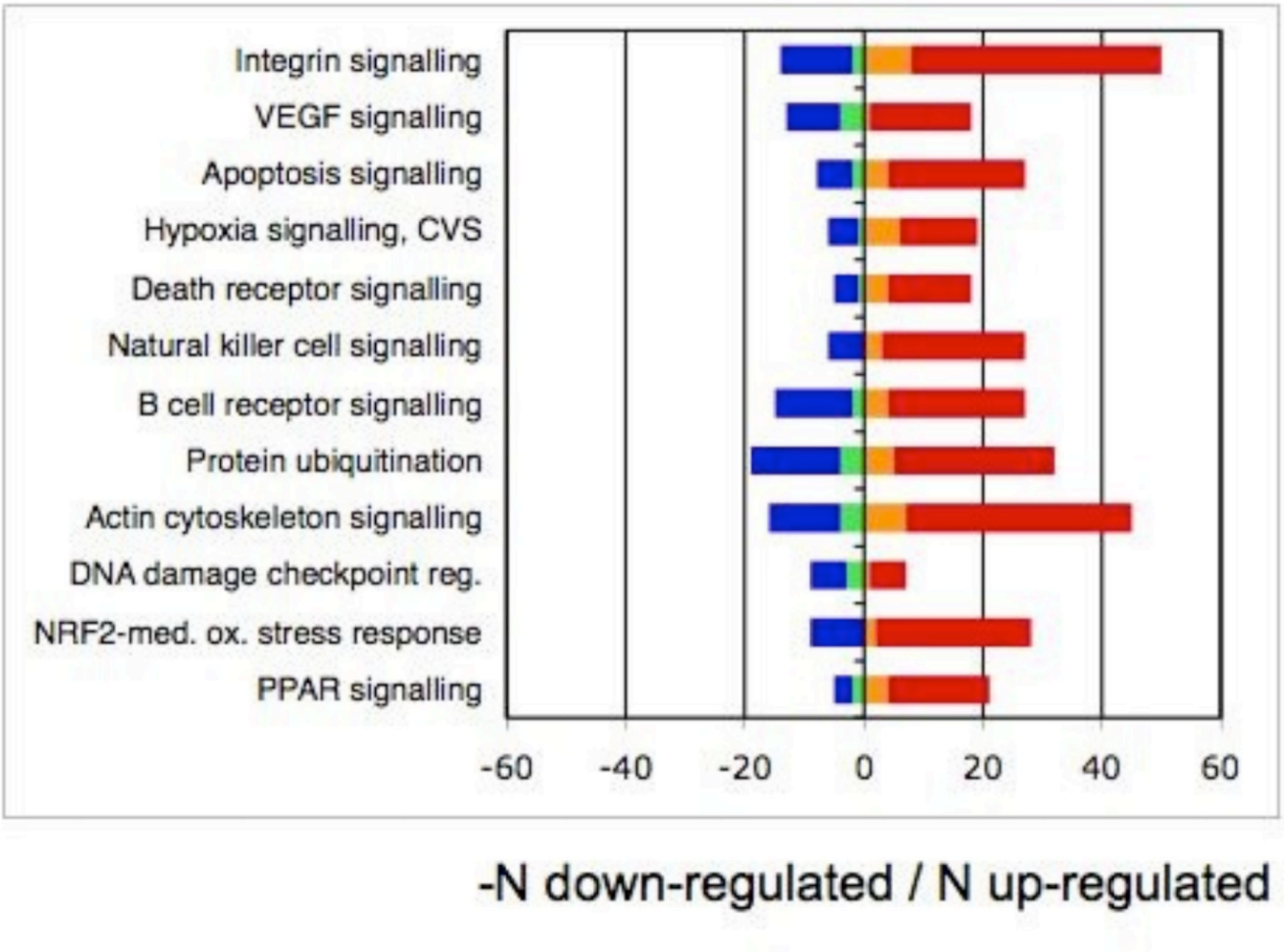
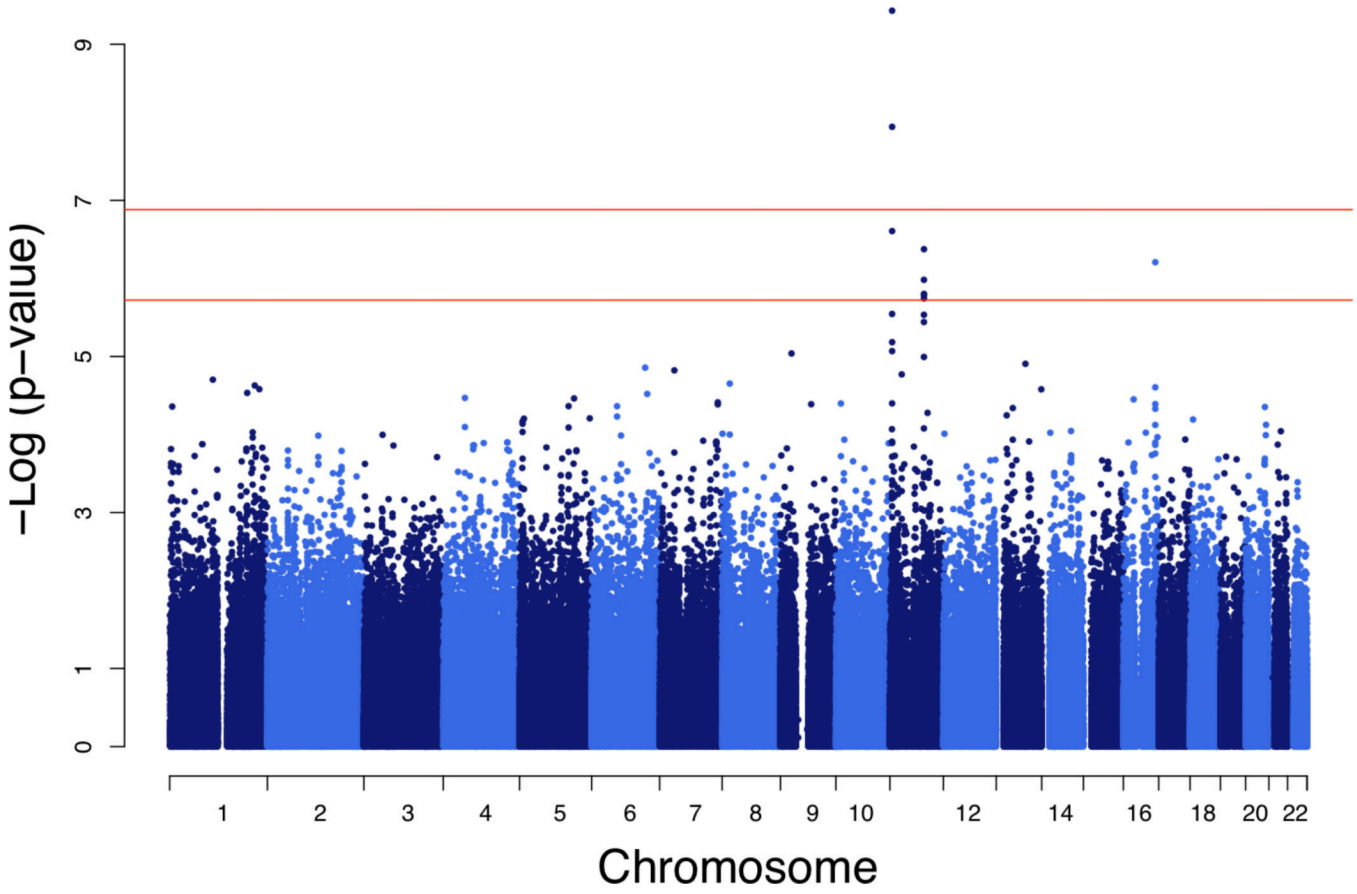


Figure 2. Functional characterization of age-correlated transcripts

a, top 12 assignments to functional class ranked by decreasing significance (range: $P=2.3\times10^{-29}$, $P=1.6\times10^{-12}$). **b**, top 12 assignments to canonical pathway (range: $P=6.2\times10^{-9}$, $P=1.1\times10^{-4}$). Legend: down-regulated with age, non-G×A: blue; down-regulated G×A: green; up-regulated G×A: orange; up-regulated, non-G×A: red.



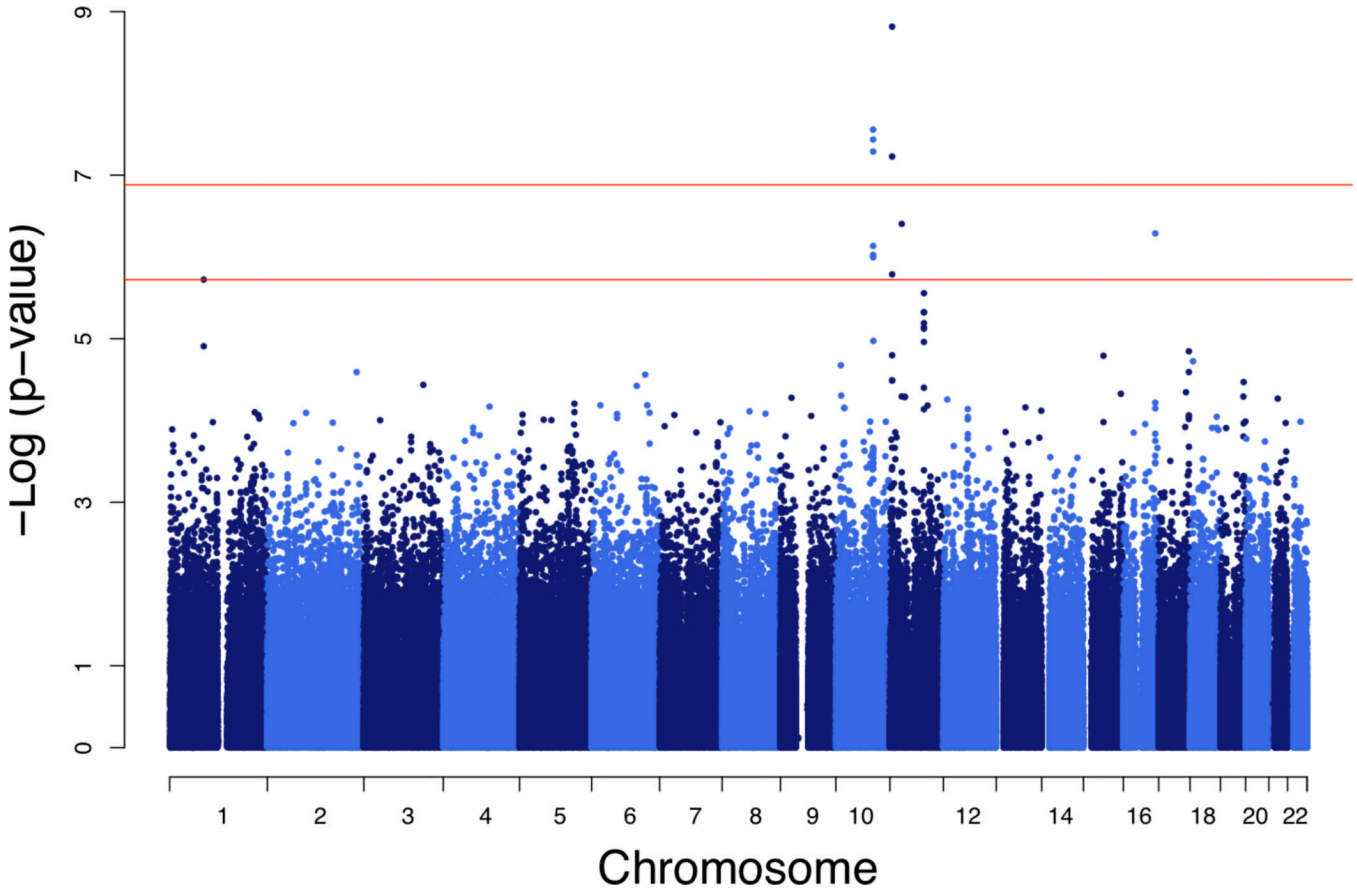
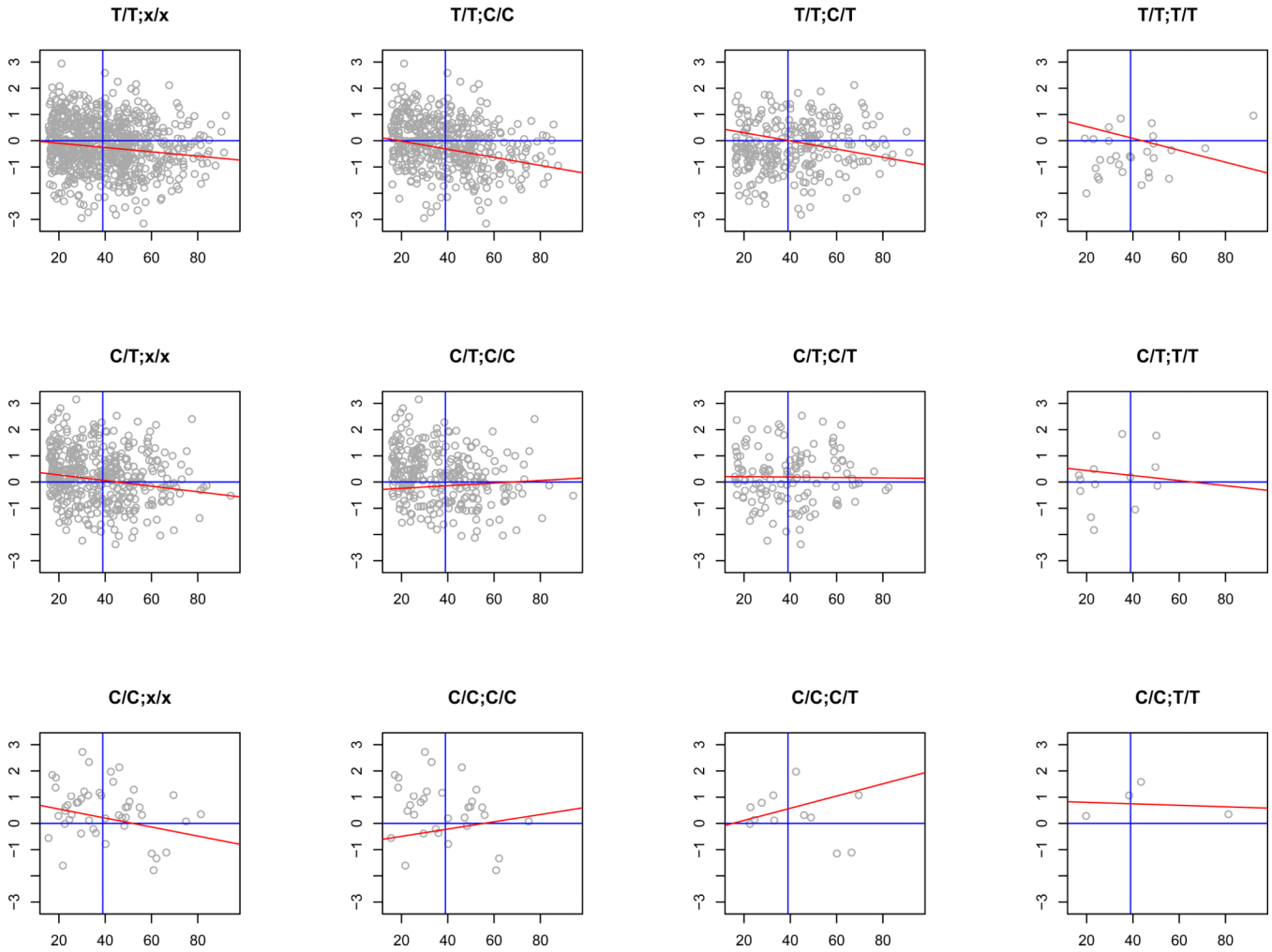


Figure 3. Genome-wide plots for SNP association with UBQLNL expression

a, evidence for association with 536,821 haplotype-tagging SNPs (1 df test). **b**, 2 df test for association contingent on SNP genotype × age interaction. Red horizontal lines represent empirical thresholds for suggestive evidence of association (occurring once per genome scan by chance, $P=1.9\times10^{-6}$) and genome-wide $P=0.05$ ($P=1.3\times10^{-7}$). Empirical thresholds were obtained from 10,000 replicate GWAS of a simulated non-associated trait.



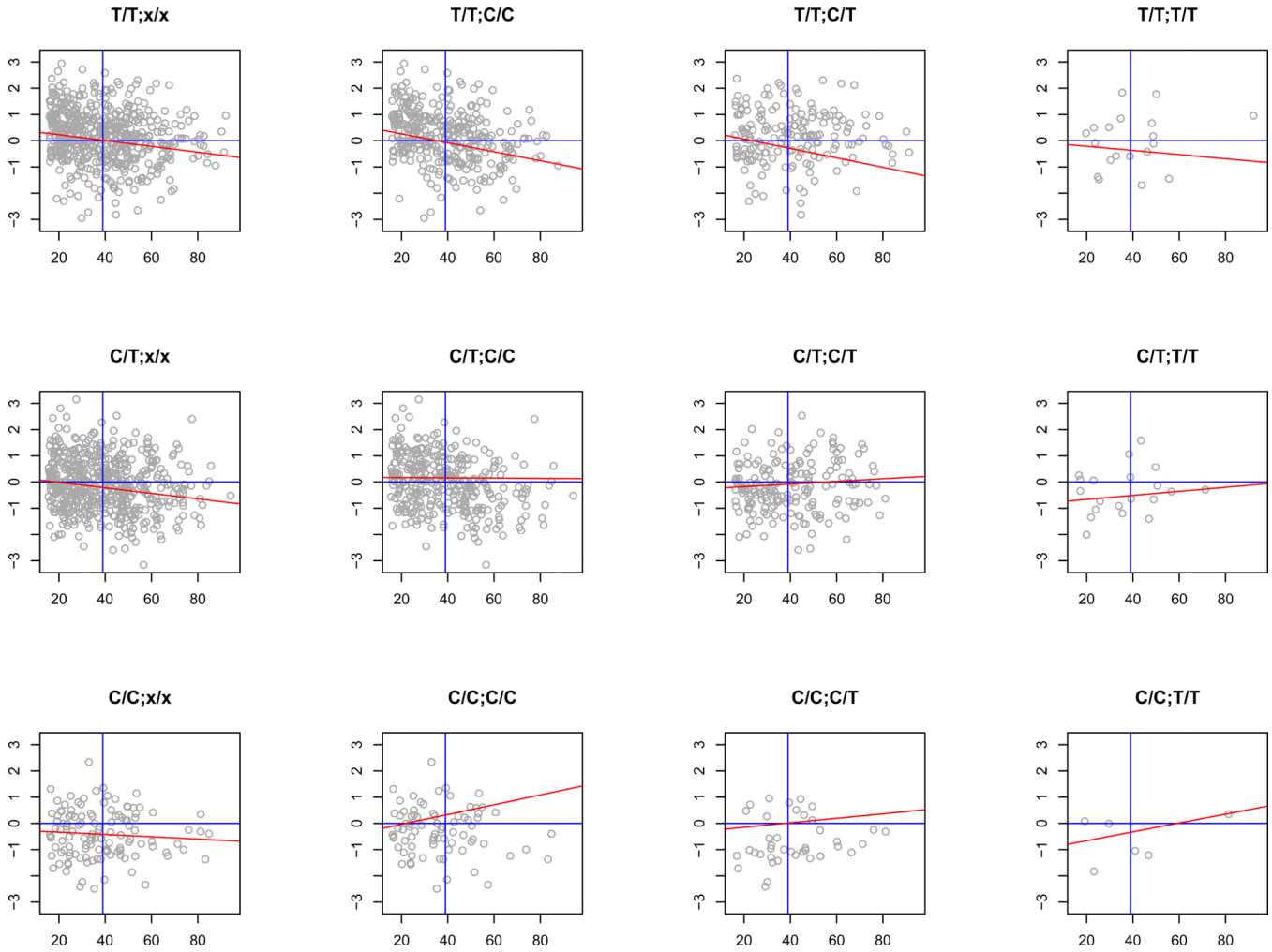


Figure 4. Norms of reaction for *cis*- and *trans*-regulatory variants

Genotype-specific regressions of *UBQLNL* mRNA levels on age. Panel A, *cis*-SNP = rs7939159; Panel B, *cis*-SNP = rs7129909; *trans*-SNP = rs11591635 in each case. Plot labels = *cis*-/*trans* genotype. Left column: norms of reaction for *cis*-acting SNP genotypes without stratification by *trans*-interacting genotype show an overall reduction of expression level with age (red lines reflect maximum-likelihood estimates of regression slope and intercept at cohort mean age). Remaining columns: *cis*-genotypes stratified by *trans*-genotype reveal patterns of *cis*-*trans* G×AI (see text). Grey axes represent mean values for *UBQLNL* expression (vertical axis, horizontal line) and age (horizontal axis, vertical line).

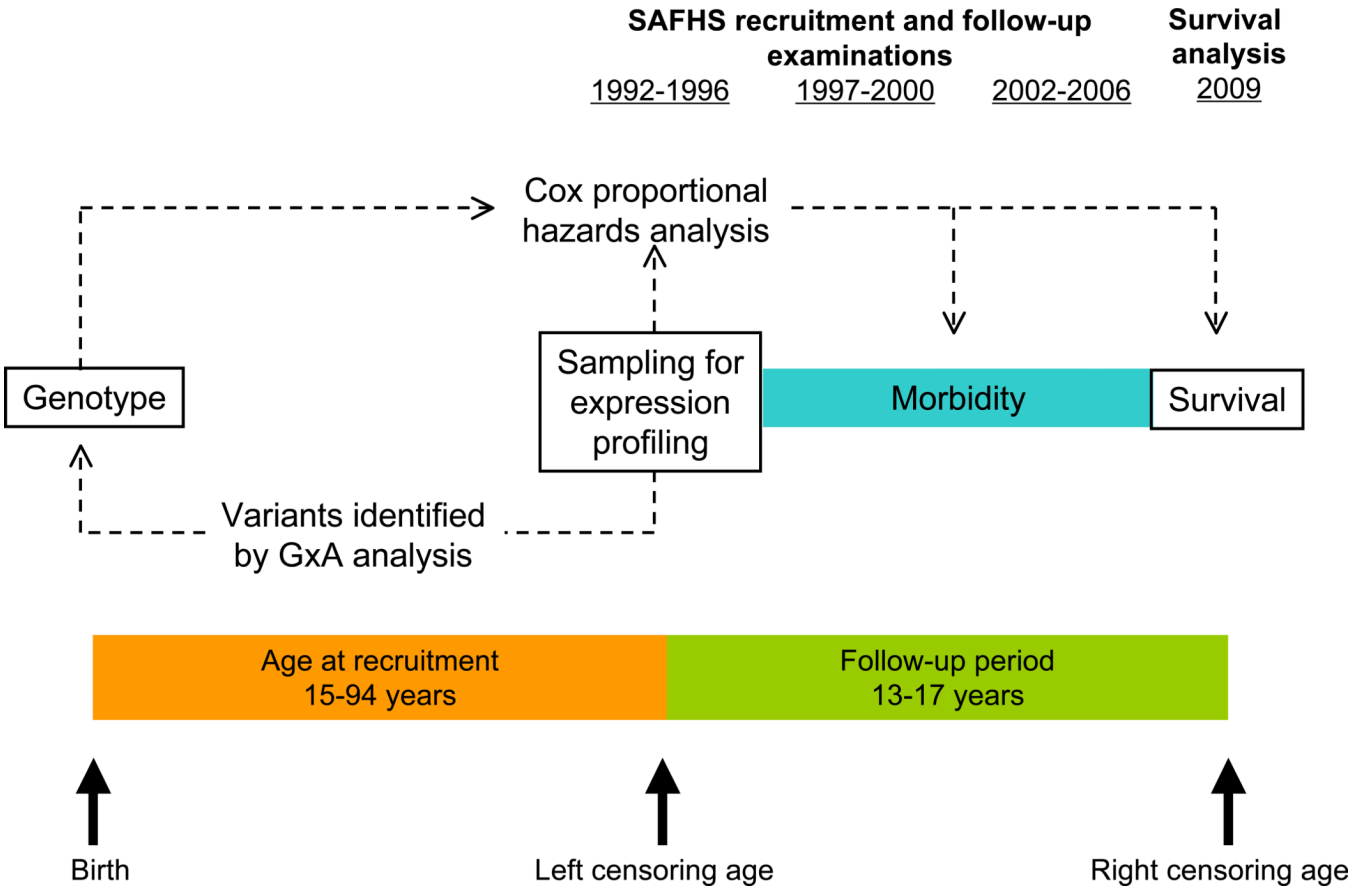


Figure 5. Experimental design of survival analysis

Expression phenotypes were measured in blood samples drawn at recruitment. G×AI analyses were based on cross-sectional data for this time point, using the range of ages at recruitment (15–94ya). Longitudinal disease incidence and mortality data were collected for the cohort over the subsequent ~16y follow-up period. Cox proportional hazards analysis was performed for hazards including the prospective effect of baseline expression levels of *UBQLNL* and genotype for associated SNPs.

Table 1

Pairwise relationships among phenotyped individuals

Relationship	N pairs	Median Δage, y	Range of Δage, y	Degree	%Shared household
Monozygotic twins	3	0.00	0.00	0	100.00
Parent-child	1,007	25.79	11.25–56.61	1	46.28
Full sibs	1,144	5.56	0.00–24.77	1	15.47
Grandparent-grandchild	341	49.73	31.50–77.31	2	10.85
Avuncular	2,275	22.48	0.07–58.19	2	2.42
Half sibs	168	10.00	0.08–39.85	2	11.31
Double 1st cousins	11	2.68	0.06–8.69	2	0.00
1st cousins & half-1st cousins	2	2.54	1.27–3.82	2.415	0.00
Greatgrand-parent-greatgrandchild	16	62.88	50.99–68.26	3	0.19
Grandavuncular	646	41.38	18.96–75.13	3	0.46
Half-avuncular	379	17.80	0.18–59.79	3	0.79
1st cousins	2,467	7.04	0.00–39.16	3	0.85
Double 1st cousins, once removed	55	23.13	9.44–37.11	3	0.00
Half-1st cousins & 2nd cousins	6	3.96	0.05–7.87	3.415	0.00
Great-grandavuncular	28	61.61	51.37–75.73	4	0.00
Half-grandavuncular	18	14.46	4.92–46.00	4	0.00
1st cousins, once removed	2,299	16.60	0.01–55.49	4	0.04
Half-1st cousins	379	7.06	0.10–34.10	4	0.00
Double 2nd cousins	45	4.69	0.03–14.30	4	0.00
5th -degree relatives	1,107	9.23	0.01–60.44	5	0.00
6th -degree relatives	240	10.54	0.01–29.50	6	0.00
TOTAL	**12,548**	**15.45**	**0.00–77.31**	**--**	**6.26**

Counts of pairwise relationships among the 1,240 phenotyped subjects and distribution of their absolute differences in age (Δage). Degree of relationship may be non-integral for individuals related via multiple loops. Household sharing is based on domicile at time of recruitment.

Table 2

SNPs associated with *UBQLNL* expression

SNP	*P*-value, association (1 df test)	P-value, SNP G×AI (2 df test)	Chromosomal location	Chromosomal position, bp	Minor allele	Minor allele frequenc
rs7939159	3.7×10^{-10} *	1.5×10^{-9} *	11p15	5,598,577	C	0.211
rs7129909	1.1×10^{-8} *	5.9×10^{-8} *	11p15	5,667,753	C	0.316
rs11591635	0.1933 NS	3.7×10^{-8} *	10q23-q24	93,350,282	G	0.206
rs11597974	0.1679 NS	2.8×10^{-8} *	10q23-q24	93,350,310	T	0.206
rs1418159	0.1896 NS	5.1×10^{-8} *	10q23-q24	93,352,607	G	0.205

* $P <$ empirical genome-wide significance threshold = 1.3×10^{-7}. Chromosomal positions are based on Human Genome build 36.3.

Table 3

Bayesian quantitative trait nucleotide analysis of *UBQLNL* expression

Parameter	Average point Estimate	Standard error	Posterior Probability	Proportion of variance explained
heritability	0.2467	0.0582	1	0.2467
rs7939159	0.3566	0.0513	1	0.0498
rs7129909	−0.2849	0.0435	1	0.0412
rs11591635	0	0	0	0
rs7939159 × age	0	0	0	0
rs7129909 × age	0	0	0	0
rs11591635 × age	0.0168	0.0028	1	0.0254

Table 4

Cox proportional hazards analysis

Hazard: *UBQLNL* expression level (*N*=1,232)a						
		Odds ratios				
Outcome	***N*, outcome**	**Sex=male**	**BMI[b]**	**Diabetes**	**Smoking**	***UBQLNL***
Mortality, all causes	**104**	**1.812[*]**	**1.014**	**1.234**	**1.316**	**1.304[*]**
Mortality, cancer	30	1.408	1.047	1.028	0.832	1.547[*]
Mortality, CVD	62	2.086[*]	1.053[*]	1.848	1.603	1.062
Cancer, mortality and morbidity	73	0.759	1.032	0.974	1.166	1.450[§]
Hazard: rs7939159 genotype (*N*=1,196)[a]						
		Odds ratios				
Outcome	*N*, outcome	Sex=male	BMI[b]	Diabetes	Smoking	SNP[c]
Mortality, all causes	98	1.950[§]	1.024	1.237	1.380	1.514[*]
Mortality, cancer	32	1.278	1.048	1.263	1.110	1.894[*]
Mortality, CVD	62	2.868[§]	1.072[§]	1.679	1.581	0.964
Cancer, mortality and morbidity	77	0.734	1.032	1.082	1.252	2.080[†]

[a] Some participants excluded due to missing data.

[b] Average BMI across all clinic visits.

[c] Other SNPs not significant hazards (data not shown).

[*] $P<0.05$

[§] $P<0.01$

[†] $P<0.0001$ No superscript=not significant.