Levels of histidine-rich glycoprotein variants in human blood are associated to chronological age and predict mortality

*Mun-Gwan Hong[1)], Tea Dodig-Crnković[1)], Xu Chen[2)], Kimi Drobin[1)], Woojoo Lee[2,3)], Yunzhang Wang[2)], Ronald Sjöberg[1)], Jacob Odeberg[1,4)], Anders Hamsten[5)], Angela Silveira[5)], Per Hall[2,6)], Peter Nilsson[1)], Yudi Pawitan[2)], Sara Hägg[2)], Mathias Uhlén[1)], Nancy L. Pedersen[2)], Patrik K. E. Magnusson[2)] and Jochen M. Schwenk[1),]**

[1)] Science for Life Laboratory, Department of Protein Science, KTH - Royal Institute of Technology, Box 1031, SE-171 21 Solna, Sweden

[2)] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Nobels väg 12A, SE-171 77 Stockholm, Sweden

[3)] Department of Statistics, Inha University, 235 Yonghyun-dong, Nam-gu, Incheon, 402-751, Korea

[4)] Department of Medicine Solna, Karolinska Institutet and Karolinska University Hospital, SE-171 76 Solna, Sweden

[5)] Cardiovascular Medicine Unit, Department of Medicine Solna, Karolinska Institutet, Akademiska Stråket 1, SE-171 64 Solna, Sweden.

[6)] Department of Oncology, Södersjukhuset, Sjukhusbacken 10, SE-118 83 Stockholm, Sweden

* Corresponding author

jochen.schwenk@scilifelab.se

1. SUMMARY

An important aspect of age-related research is to find proteins in human blood that can be used to track physiological processes of aging. Here, we have used a multiplexed affinity proteomics approach to search for the presence of age-associated protein levels in human body fluids. First, serum samples from 156 subjects aged 50-92 years were explored using a comprehensive bead array assay including 7,258 antibodies. We identified 16 age-associated profiles (adjusted $P < 0.05$) and followed up on the most significantly age-associated profiles in eight additional study sets (n = 4,044 individuals) analyzing both serum and plasma. Meta-analysis highlighted a consistent increase with age ($P = 5.37 \times 10^{-6}$) for variants of histidine-rich glycoprotein (HRG), which we confirmed by antibody validation assays and genome wide association studies. Higher levels of HRG, which is a plasma glycoprotein produced by the liver, also increased the risk of mortality during about 8.5 years follow-up (interquartile range = 7.7-9.3) after blood sampling at a hazard ratio = 1.25 per standard deviation ($P = 6.45 \times 10^{-5}$). Our multi-cohort affinity proteomics analysis found that blood levels of the multi-purpose HRG variants were associated with age and all-cause mortality. This combination suggests that elevated HRG levels could serve as an accessible molecular indicator for biological aging.

2. INTRODUCTION

Aging is the single most dominant risk factor of common diseases of elderly and death in the human population (López-Otín, Blasco, Partridge, Serrano, & Kroemer, 2013). Molecular insights into aging could enable direct identification of future treatments for various diseases and would increase our understanding of longevity and related mechanisms. However, many of the underlying molecular processes and changes in humans still remain poorly understood (López-Otín et al., 2013). Aging is most often studied using animal models or cell lines (López-Otín et al., 2013), despite the vast differences in lifespans (2 weeks to 100 years) (Gorbunova, Seluanov, Zhang, Gladyshev, & Vijg, 2014). Findings from these model organisms should preferably be translated into studies on humans. Biological age of humans or the prediction of mortality has already been investigated via DNA methylation, telomere length, proteomic studies, mining of clinical records (Ganna & Ingelsson, 2015; Jylhävä, Pedersen, & Hägg, 2017), and some studies showed several candidates for these traits (Barron, Lara, White, & Mathers, 2015; Ganna & Ingelsson, 2015; Marioni et al., 2015). Recently rejuvenation factors were found in mouse blood (Katsimpardi et al., 2014), which suggest the potential to find aging governing molecules in human blood. In this study, we aim to search for novel proteins aging predictor in blood associated with both age and mortality. This poses a challenge to the analytical methods in terms of sample complexity and availability.

To study proteins at a wider scale, there are currently two major technological concepts available for measuring the proteome: mass spectrometry (MS) and affinity-based proteomics (Ayoglu et al., 2011). Both approaches have been used to study the plasma proteome (Schwenk et al., 2017), offering a unique window into human health and diseases. Even though affinity proteomics has suffered from a lack of binding reagents to the proteome (Stoevesandt & Taussig, 2012), antibody resources such as the Human Protein Atlas (HPA) (Uhlén et al., 2015) or aptamer-based platforms (Emilsson et al., 2018), now offer the possibility to apply affinity proteomics for broader discovery projects. The capacities to conduct near population-based studies implementing also genome-wide association studies (GWAS) is very attractive, which has been demonstrated using different plasma proteomics assays (Melzer et al., 2008; Suhre et al., 2017). Several interesting connections between proteins and genetic variants in humans have recently been identified in large scale cohort studies (Emilsson et al., 2018; Sun et al., 2018). The two-omics approach have indeed provided novel insights into the links between the distant molecular systems, but also indirect validation to mitigate uncertainty in the molecular assays.

Utilizing the in-house developed antibody assays based on the suspension bead arrays (Schwenk, Igel, Neiman, et al., 2010) we aimed to profile serum and plasma from a large number

of individuals, studying the changes in age-related plasma protein levels. Our strategy was to explore, filter and rank plasma profiles associated with age in extended sets of samples and to confirm antibody selectivity by applying different validation assays (Supporting Figure 1).

3. RESULTS

We analyzed human serum and plasma samples of 4200 subjects for blood proteins associated with age. Using affinity proteomic assays initially based on a large set of antibody reagents, protein profiling was conducted with nine sets of samples to determine the most consistent age associations. Our aim was to describe proteins present in serum and plasma that could serve as indicators of biological age to increase our understanding of age-related phenotypes.

Screening for age-associated profiles

Age-associated protein profiles were first investigated in a set of 156 human subjects selected in 5-year age intervals from a Swedish twin cohort using our proteomic assays (Table 1). The sex-matched samples from the study set included 30 monozygotic (MZ) twin pairs (age 50-70). The average intraclass correlation within pairs of antibody profiles was small (ICC = 0.13) so twins were treated as unrelated. Minimal effects of the twin relationship were corroborated by a linear mixed model that considered the dependency.

Assays using a total of 7,258 HPA antibodies were applied to profile age-associated proteins in serum. For this screening, target inclusion criteria were purely dependent on availability of antibodies but not due to their target antigens (Byström et al., 2014). This set of antibodies comprised targets from 6,370 protein-encoding genes (about 32% of the non-redundant human proteome) and profiles were obtained using antibody suspension bead arrays (Drobin, Nilsson, & Schwenk, 2013), which provided up to 384 profiles on 384 samples per batch. The acquired data was preprocessed through quality control including outlier removal and normalizations to account for experimental variation across individual samples, assay plates and data batches (details in Experimental procedures). Linear regression models were then used to determine the protein profiles that changed monotonically with increasing age. The models revealed 16 out of 7258 (adjusted $P < 0.05$) protein profiles that were age associated when screening sera of individuals at the age of 50 to 92.

Study sets to replicate and confirm the discovered age-associations

Next, we aimed at following up on this screening phase and focused on validating the most significant indications, accepting that other previously published age-associated proteins will not be included. Because the age range of set 1 covered life expectancy in Sweden (84 years for women and 80 years for men in 2015 (Statistics Sweden, 2016)) we considered it more likely that findings from this set might be related with accelerated ageing and mortality. Concentrating on the three most significant findings from set 1 (Table 2), we investigated eight additional sample sets (set 2-9) (Table 1). Out of 4,044 subjects in total, 829 subjects were from non-diseased control groups.

The entire set of subjects were from 3 to 93 years old at blood draw. Blood samples had been prepared either as serum or plasma (Table 1, Supporting Figure 2), hence differences originating from these preparation types were likely. Twelve sera in the otherwise plasma sample set 2 were not included in the meta analyses. Sample sets 6 to 9 (729 subjects) were from four independent studies (Table 1)(Baldassarre et al., 2010; Gabrielson et al., 2017; Odeberg et al., 2014; Samnegård et al., 2005). Two other sample sets (set 4 and 5) included 100 subjects that were selected for cancer-related studies and derived from the same twin cohort as set 1 (Lichtenstein et al., 2002; Magnusson et al., 2013). Besides one single subject, there was no overlap between these and the individuals analyzed during the discovery. The sample set 3 (2999 subjects) was again chosen from the population-based twin cohort (Lichtenstein et al., 2002; Magnusson et al., 2013), in which disease status was not considered during recruitment. Almost all (98.5%, all except forty-four) subjects were not included in sample set 1.

Discovery and confirmation of age-association of HRG in serum and plasma

Protein profiles were generated by using antibody bead arrays, as this platform allows to combine antibodies towards different targets whilst consuming only minute amounts of samples. Our analysis revealed consistent age-associated trends for HPA045005 across the eight sample sets for replication (Supporting Figure 3). Accounting for differences in age ranges and spectra, the combined effect of age on HPA045005 in the 9 sample sets was estimated using a random effects model and showed consistent association across sample sets (meta-analysis, $P = 5.37 \times 10^{-6}$, Figure 1). Since the observed trends two other candidate profiles, generated by HPA039928 and HPA029931, could not be replicated to the same degree in all study sets as for HPA045005 (Supporting Figure 3), we chose to focus our efforts on HPA045005.

Focusing on HPA045005, of which the profile was most significantly associated and replicated, we investigated if the described protein levels in the circulation were controlled by genetic components. Employing GWAS with sample set 3 (N = 2308), a locus in chromosome 3q27.3 was solely found to be associated with the antibody profile ($P < 1.13 \times 10^{-9} = 0.01 / 8{,}833{,}947$) among ~8.8M genetic variants imputed from or genotyped by Illumina BeadChip for >700K single nucleotide variants (Figure 2A). The locus spans two genes, *FETUB* and *HRG*, in the human genome (Figure 2B). The most significantly associated genetic variant in the locus was the single nucleotide polymorphism (SNP) of rs9898 ($P = 2.35 \times 10^{-97}$, minor allele frequency (MAF) = 0.32, increment per number of minor allele (β) = 0.15), which is a missense variant that induces the amino-acid sequence change from proline to serine in the histidine rich glycoprotein (HRG). Among the top 99 genome-wide significant SNPs ($P < 1.13 \times 10^{-9}$) having RefSNPs (rs) number, four SNPs (rs1042464, rs2228243, rs10770) including rs9898 are non-

synonymous and two SNPs (rs3890864 and rs56376528) are located near (<2kbp) to transcription start site (Supporting Table 1). All are located in exons or upstream of *HRG*. This GWAS result indicates that the antibody describes the protein levels of HRG, a protein secreted by the liver and found in abundance in blood (Morgan, Koskelo, Koenig, & Conway, 1978). Associations of plasma HRG levels to SNPs have also been found in previous plasma profiling studies (Suhre et al., 2017).

Next, we confirmed the binding of HPA045005 to HRG by using a sandwich immunoassay. Beads with HPA045005, an additional anti-HRG antibody (HPA054598), as well as negative controls were combined to detect HRG. When analyzing a full-length recombinant HRG in serial dilution experiments, we found that pairing both HPA045005 and HPA054598 with a biotinylated version of HPA054598 allowed us to detect HRG in a concentration dependent manner (Supporting Figure 4A). Here, the signals obtained from both sandwich assay pairs was substantially higher than the internal negative controls. To further elucidate the selectivity of HPA045005 over other proteins, we used a protein microarray (Sjoberg et al., 2012). Besides the antigen used to generate HPA045005, the array contained 12,412 other protein fragments. In this analysis HPA045005 exclusively bound to its corresponding antigen (Supporting Figure 4B). This indicated that the antibody does not generally cross-react in an unspecific manner, which points at a selective recognition of abundant HRG in serum and plasma. This molecular analysis supports the GWAS findings that there is an affinity of HPA045005 to the secreted liver HRG, when using the antibody in bead-based assays for the analysis of serum and plasma samples.

In addition to HPA045005, the GWAS with sample set 3 included several other antibodies of which one monoclonal binder (BSI0137) targeted the HRG protein. For BSI0137 there was one locus in the gene *HRG* which was strongly associated with the antibody's profile ($P < 1 \times 10^{-300}$, Supporting Figure 5). Interestingly, the identified locus included all four non-synonymous SNPs previously observed to associate with HPA045005. However, the most significant SNPs was not rs9898 but rs1042464, and the slopes of correlating BSI0137 with these SNPs were indeed opposed to those for HPA045005 (Figure 2C). Applying probabilistic identification of causal SNPs (PICS) method (Farh et al., 2015), we confirmed that it was highly unlikely to observe rs9898 as the most significantly associated SNP with HPA045005 when rs1042464 was the causal SNP (none in 100000 permutations), while the significance of rs1042464 was within possible range assuming rs9898 was causal (Supporting Figure 6). With the facts that rs9898 causes HRG to contain either Pro186 or Ser186 and that HPA045005 levels increased with the number of allele C in dosage dependent manner, the PICS analysis revealed a differential HRG affinity of HPA045005: HPA045005 has a preferred affinity to HRG with Pro186.

Mortality association and prediction

Finding the age-association levels of HRG variants lead us to study further aging-related aspect. We therefore accessed the Swedish death registry that listed information whether the subjects were still alive or not with a follow-up time of ~8.5 years after donating the serum samples. We chose the largest sample set of the subjects at mid to old ages that spanned the average life expectancy in Sweden (sample set 3, N=2973, 48-93 years old) in order to gain statistical power needed for the analysis of the death, which was a relatively rare event. To test for all-cause mortality, a Cox proportional hazards model with age as the time scale was used. The Cox model was further adjusted for the effects of sex and revealed that the HRG variant levels obtained by HPA045005 were significantly associated with the mortality during follow up (inter-quartile range (IQR) = 7.7-9.3; $P = 1.13 \times 10^{-4}$), whereas HRG levels determined by BSI0137, with no correlation with HPA045005 ($R^2 = 0.006$), were not associated to mortality (P = 0.57) and age (P = 0.43). As protein variant levels increase with age, the HRG value of HPA045005 was standardized using a linear regression model for age and age squared for each sex, followed by scaling in order to account for the linear and quadratic effects of age and to let the hazard ratio (HR) quantifiable that was estimated by the survival model. The hazards model using the standardized HRG value affirmed the association (number of deaths=362, $P = 6.45 \times 10^{-5}$), estimating that the risk of all-cause mortality increased 1.25 times per standard deviation (SD) of the HRG values compared to persons at the same age and sex. In the model accounting for potential genetic effect of the associated SNP, rs9898, estimated HR of HRG was even higher to 1.31 per SD with barely changed significance ($P = 7.75 \times 10^{-5}$), despite smaller number of samples (N = 2307). Noteworthy, the SNP itself was not associated to mortality (P = 0.69). No evident difference was observed when stratifying by the genotype of the SNP on HR of HRG, while linear association of age with non-standardized HRG was stronger as the number of T allele increased (Supporting Table 2). A Cox model stratified by sex suggested stronger association in women (N = 1602, deaths = 160, $P = 2.13 \times 10^{-5}$, HR = 1.35 per SD) than in men (N = 1371, deaths = 202, P = 0.059, HR = 1.15 per SD). The comparison of the extreme subsets with standardized HRG levels of the upper and lower quartile demonstrated that the difference of median age at death was 1.8 years in favor of the bottom quarter ($P = 3.87 \times 10^{-3}$, HR = 1.54, Figure 3). The difference was 1.9 years in men (86.9 years vs. 85.0), while 0.6 years in women (89.6 vs. 89.0; Supporting Figure 7). The difference of life expectancy at the age of 45 between the two extreme quarters was 3.7 years in women (87.3 vs. 91.0), while 2.8 years in men (83.9 vs 86.7), assuming age-at-death follows Weibull distribution (Supporting Table 3). Potential influence of general inflammation on survival were also tested by models including clinically measured C-reactive protein (CRP) values. As for HRG, two Cox models were fitted to 1) CRP and 2) age-adjusted CRP. The latter was obtained using same linear model as HRG including the adjustment for same covariate. The outcome of

P = 0.024 (N = 2971, HR = 1.07 per SD) and P = 0.023 (HR = 1.01) was far less significant than for HRG. Next, we included CRP as a covariate in the Cox model for HRG in order to determine if inflammation in general would have an influence of HRG-related mortality. The resulting CRP-adjusted HRG association reduce the significance only from $P = 6.45 \times 10^{-5}$ to $P = 1.05 \times 10^{-4}$. (HR = from 1.25 to 1.24 per SD) We also confirmed that none of the hazard models violated the proportionality assumption of the Cox model. The predictive power of HRG levels together with age for all-cause mortality was tested using a Cox model with the time from sample collection as the time scale. The Harrell's C-index of the model was 0.766.

4. DISCUSSION

HRG, a multi-functioning protein in plasma

We analyzed the age-association of proteins with antibody-based assays in blood prepared as serum or plasma, and found increasing levels of HRG to be consistently associated with age. GWAS and sandwich assays confirmed binding of HRG. Using another anti-HRG antibody we demonstrated a differential affinity to variants of HRG, of which one was the protein captured by the antibody in the exploratory discovery antibody assay.

According to mRNA sequencing data of human tissues, HRG is exclusively expressed in liver (Uhlén et al., 2015). HRG has been known and described as an abundant protein in human blood plasma (Morgan et al., 1978; Poon, Patel, Davis, Parish, & Hulett, 2011). It has been characterized to interact with diverse molecules including heparin, immunoglobulin G (IgG), Zn^{2+}, and complement components (Poon et al., 2011). HRG in plasma is known be involved not only in immune response toward foreign substances and clearance of dead cells, but also in vascular biology including anti-coagulation (Poon et al., 2011). HRG has functional similarities with CRP, such as coagulation and inflammation (Poon et al., 2011), which is another indicator of aging and mortality (Barron et al., 2015). HRG levels have previously been correlated and linked to blood ABO type and age (Drasin & Sahud, 1996). It has been also found as a biomarker of preeclampsia, which entails angiogenic imbalance and defective coagulation control (Bolin, Akerud, Hansson, & Akerud, 2011). But, because of its molecular composition and abundance, HRG has also been assigned to many different biological processes. Hence it is yet difficult to pinpoint and postulate the most plausible mechanism of increasing HRG levels in the process of aging.

In the gene of *HRG*, there are at least 4 genetic variants relatively abundant (MAF > 10%) and these lead to different amino acid at these 4 positions. Dealing with variants generated by different combinations of those genetic variants, our GWAS analysis revealed the associations of both antibodies' profiles with those non-synonymous SNPs. The PICS analysis provided indications about which amino acid residue of the HRG may affect the antibody recognition. Using this novel approach, we found that HPA045005 preferred the HRG with Pro186 over the variant with Ser186, which is located on the N2 region of HRG (Poon et al., 2011). We know some individuals can express one protein variant only. No single genetic variant around *HRG* reached genome-wide significance for mortality risk in a study including the TwinGene cohort (Ganna et al., 2013). Thus, we could postulate that detecting the abundance of HRG with HPA045005 revealed changes related to biological age that were linked to the residue Pro186. Interestingly, two of the neighboring residues are amino acids with functional groups: Asp184 is a possible N-

glycosylation site and a glutathione modification has been implicated for Cys185 (Kassaar et al., 2014). In rabbits, the latter modification and possible plasmin cleavage sites at Arg298 and Arg414 haven been linked angiogenesis. Hence, it is plausible that changes occurring on these positions may also be dependent whether the variant carries a Pro186 or Ser186. Proline is classified as a non-polar amino acid, known to disrupt the formation of alpha helices and beta sheets but often found in loops and turns. Serine, on the other hand, is a polar amino acid, know to form hydrogen bonds with other residues and a possible site for O-glycosylation. While further investigations of the molecular details will be needed, we speculate that the different physiochemical properties of proline and serine could also indicate that the HRG variants interact at different affinity and selectivity with other proteins. This could possibly point at that the variants participation in different physiological processes.

Limitation and variation

Our study was cross-sectional and the ages of the participants covered those age ranges of average lifespan in many of the profiled sample sets, including both serum and plasma. There is still limited information available to interpret the trend of increasing of HRG levels as advanced age reflected longitudinal change within individual subjects. On the other hand, those gradual alterations were repeatedly observed in multiple independent study sets derived from different Swedish cohorts, which provides strong indications that the association of HRG with age was confirmed. As we also found that the elevated level of HRG comparing to same-aged peers was correlated with higher risk of mortality, the age-dependent diversity may imply a time-wise transition along individual ages, possibly biological ages. We observed variation in the degree of the age-dependent transition, which is visible in Figure 2. To some extent, the variation can be explained by the shift of signal range in each assay, which was primarily developed to screen for possible associations and not standardized to determine absolute abundance levels. Seeing that the estimated slopes from sample sets 2, 3, and 9 were relatively lower than the values from the other sets, some parts of the variation might originate from the difference in age range and sample source, collection and selection procedures. For example, the individuals in the sample sets 2, 7 and 9 were substantially younger (median age 40, 52 and 54 years, respectively) compared to all others (~65 years old). The sample sets 2, 3, and 9 were near population-based, while the others were healthy individuals except those in the sample set 1, in which older women and men were overrepresented due to same number selection per age group.

Significance as a potential predictor of mortality

Several indicators in blood were found predictive for mortality risk in previous studies and Barron et al. (2015) showed three markers, CRP, N-terminal pro brain natriuretic peptide (NT-

proBNP), and white blood cell (WBC) count, were statistically significant in meta-analyses (Barron et al., 2015). The HR estimate of HRG in this study (1.54 between top and bottom quarters) was comparable with the combined estimate in the meta-analysis (1.42 for CRP, 1.43 for NP-proBNP, and 1.36 for WBC count). In a previous study, CRP was linked to mortality risk in a follow-up duration (median 8.9 years) similar to our study. HR per SD of CRP was estimated to 1.18 (Schnabel et al., 2013), which is slightly lower than our estimate of 1.25 from HRG. Our HRG estimate was also slightly higher than the HR of DNA methylation (1.09 - 1.21) (Marioni et al., 2015). Comparing with the questionnaire-derived measures examined in Ganna and Ingelsson's study, HRG (C-index = 0.766 with age) marginally outperformed the top predictors (max C-index = 0.74 including age) in the extensive population-based mortality study (Ganna & Ingelsson, 2015).

Other recent affinity proteomics approaches have also shown age related signatures of aging, highlighting GDF15 as well as other proteins of coagulation system (Tanaka et al., 2018). While that study acknowledged the need for further validation, we have conducted extensive effort to confirm our observations across many different cohorts. Our strategy, on the other hand, was not to include previously known age-related proteins, hence did not shortlist antibodies that could be useful for screening plasma for these markers by using our method.

In conclusion, increased blood levels of HRG variants in older humans was discovered and adequately replicated in multiple sample sets by affinity proteomics. Appropriate molecular approaches were employed to characterize the identity of the protein to pave the way to develop targeted assays for expanding the analysis of our primary data. The supporting evidence of HRG serving as a predictive indicator for all-cause mortality within 8.5 years after blood draw suggests that levels of HRG variants in blood could be used as an aging indicator.

5. EXPERIMENTAL PROCEDURES

Cohort design and sample selection

a) Sample set 1 from TwinGene

A population wide collection of blood from 12,614 twins born between 1911-1958 has been undertaken in a project called TwinGene. The primary aim of the TwinGene project has been to systematically transform the oldest cohorts of the Swedish Twin Registry (STR) into a molecular-genetic resource (Magnusson et al., 2013). From 2004 to 2008, a total of 21,500 twins (~200 twin pairs per month) were contacted by the invitation to the study containing information of it and its purpose, also consent forms and health questionnaire. The study population was limited to those participating in the Screening Across the Lifespan Twin Study (SALT) which was a telephone interview study conducted in 1998-2002 (Lichtenstein et al., 2002). Other inclusion criteria were that both twins in the pair had to be alive and living in Sweden. Subjects were excluded from the study who had declined to participate in future studies or been enrolled in other STR DNA sampling projects. When the signed consent forms returned, blood-sampling equipment was sent to the subjects, who were asked to visit local health-care facilities on the morning, after fasting from 20:00 the previous night, from Monday to Thursday and not the day prior to a national holiday. This was to ensure that the sample tube would be delivered to the Karolinska Institutet (KI) Biobank by the following morning by overnight mail. After arrival, the serum was stored in liquid nitrogen.

The contribution for sample set 1 of serum samples from the TwinGene study consisted of: A) samples from 96 unrelated twins distributed in groups of 12 subjects (6 males and 6 females) in each age strata 50, 55, 60, 65, 70, 75, 80 and 85 years of age. The width of the age intervals was approximately +/- 3 months, and B) samples from 60 MZ twins (30 complete pairs) distributed in groups of 12 (3 male pairs, 3 female pairs) in each age strata of 50, 55, 60, 65 and 70 years of age. The width of the age intervals was approximately +/- 3 months.

b) Sample set 2 from LifeGene

Life Gene is a prospective cohort study that includes collection of plasma and serum, tests of physical performance, as well as questionnaire responses regarding a wide range of lifestyle factors, health behaviors and symptoms (Almqvist et al., 2011). Participants respond to a web-based questionnaire and book time for a visit to a LifeGene test center, at which blood samples are taken. EDTA plasma was processed at the test center as follows: the EDTA tube with a gel plug was centrifuged, put into -20ºC prior to shipment in a cold chain. All samples were sent to KI

Biobank for further separation into aliquots in REMP plates and frozen at -70ºC. All participants or, in the case of children under the age of 11, their guardians, provided signed consent.

The sample set 2 cohort consisted of 5 male and 5 female samples randomly chosen from each of the ages <5, 10, 15, 20, 25, 30, 35, 40, 45, 50 and 55 (+/- 3 months). For 12 participants, serum was also available.

c) Sample sets 3, 4, and 5 from TwinGene

Sample sets 3, 4, and 5 were selected from the same cohort, TwinGene (Magnusson et al., 2013), as for sample set 1 (described above). Out of 132 microtiter 96-well plates for storage of TwinGene samples, the twelve plates having the largest age span (>20 years) among samples in a plate and another randomly chosen twenty plates comprising a sufficient number of samples (>91) were selected. Sample set 3 consisted of the three thousand samples in the selected 32 storage places. The data of one individual was removed in the analyses because age of the subject is missing. Independently from the sample selection, sample sets 4 and 5 were age and gender matched controls for breast and prostate cancer studies, respectively. The mortality data was obtained by connecting individuals in TwinGene to the data in the Swedish tax authorities by personal identification number. The data was updated on 2015-01-10. Clinical blood chemistry assessments of hs-CRP of the samples in TwinGene was performed using Syncron LX System (Beckham Coulter).

d) Sample sets 6 to 9

The sample sets 6 to 9 are described in Supporting Text.

e) Ethics

All the studies were approved by the Ethics Board of the correspondent hospital or institution, and conducted in agreement with the Declaration of Helsinki. The ethical approval document numbers are 2007/644-31/2 for TwinGene, 2009/615-31/1 for LifeGene, 03-115 and 2017/404-32 for IMPROVE, 95-397 and 02-091 for SCARF, EPN 2009/762 and LU 298-91 for CHAPS, and 2010/958-31/1 for Karma. All subjects, or their guardians, provided their informed consent to participation in individual studies.

Data acquisition - Assay design and SBA procedure

All 372 samples from sample sets 1 and 2 together were randomly allocated into wells in four 96-well plates. One sample from each sample set was loaded into two more wells as a repeated control within a plate. Another sample in each cohort was transferred to two more wells of two different plates as a control to examine inter-plate variation. The data of each of those 4 samples

was combined by taking mean of three measures. All the human materials were biotinylated together with four negative controls that contained only buffer. For the entire 19 assays for discovery stage, the samples were labeled two times.

The selected antibodies were divided into collections of 384 antibodies including positive and negative controls, anti-albumin and no antibody, respectively. These antibodies were then coupled onto beads and used to create a suspension bead array (SBA). For discovery, the selection of the affinity binders for one SBA was determined by technical reasons such as the available amount. Every antibody in an SBA was coupled with beads with a different colour code as detailed together with the assay procedure in the Supporting Information and as described earlier (Drobin et al., 2013).

Quality control and preprocessing

All values in an assay (with 384 antibodies) of the samples seemingly failed were removed. The criteria were 1) median bead counts < 20, 2) median MFIs < median of buffer only sample, and 3) outlier(s) detected by robust PCA using 'rrcov' R package (version 1.3-4)(Hubert, Rousseeuw, & Branden, 2005). Because serum and plasma showed considerable dissimilarity, as expected (Supporting Figure 2)(Schwenk, Igel, Kato, et al., 2010), the obtained data was split by the blood preparation type. Probabilistic quotient normalization (PQN) was applied to minimize sample-wise fluctuation (Dieterle, Ross, Schlotterbeck, & Senn, 2006). The effects of the 96-well plates were minimized by Multi-MA method (Hong, Lee, Nilsson, Pawitan, & Schwenk, 2016). More details are available in Supporting Text.

Data acquisition of replication sample sets

Data of other replication samples (sample sets 3-9) were acquired using the same protocol with a few variations. For each original study for sample sets 4-8, the samples were distributed into plates together with patient samples. The 383 other antibodies selected for each of the intended studies were included in the assays. Experiment and data preprocessing were conducted together with those additional samples and antibodies. Data of disease-free controls and for HPA045005 were extracted from the processed full data sets. Likewise, we obtained the data of the other top 2 candidates shown in Table 2.

Other experimental details

Experimental details on antibody selection, bead array assays, genome-wide association study, sandwich immunoassays, and protein microarray analysis are available in Supporting Text.

Statistical analysis

The preprocessed intensity data was log-transformed ahead of following analyses. To control family-wise error rate, Bonferroni method was employed in adjusting P-values unless otherwise specified. The linear association of an antibody signal level with age was tested with ordinary linear regression using R. The meta-analysis was conducted using the inverse variance method with between-study variance estimated by DerSimonian-Laird model (DerSimonian & Laird, 1986), which was implemented in the R-package "meta". We used a linear mixed model to address the correlation between twins where the response variable was the normalized antibody measurement and age was a fixed covariate. This model was performed using the R-package "lme4". For the association test for mortality, Cox proportional hazards model was fitted to the survival data with age as the time-scale and right censoring at the age on the updated date of death information (Thiébaut & Bénichou, 2004). In the survival analysis for two group comparison, the subjects in sample set 3 were divided into two groups, top and bottom quarters by the standardized HRG values, which were the scaled residuals of linear model where the normalized MFIs of HRG were regressed on age and age squared for women and for men, separately. The hazard models were adjusted for sex if applicable and for CRP as described above. The proportionality assumption of the models was tested using Schoenfeld residuals (Grambsch & Therneau, 1994). Survival analyses including computation of Harrell's C-index (Harrell, Califf, Pryor, Lee, & Rosati, 1982) were conducted using the R packages "survival" and "eha".

6. ACKNOWLEDGMENTS

We like to thank Camilla Björk and Jens Mattsson from MEB at the Karolinska Institutet, everyone in the Affinity Proteomics group at SciLifeLab, and especially Claudia Fredolini, MariaJesus Iglesias, Matilda Dale, Sanna Byström, Martin Zwahlen, Björn Forsström, Björn Winckler, and Philippa Pettingill for supporting this work. We also thank the entire staff of the Human Protein Atlas for their efforts, Hanna Tegel and Johan Rockberg and their team for providing the recombinant HRG protein. This work was supported by ProNova VINN Excellence Centre for Protein Technology (VINNOVA, Swedish Governmental Agency for Innovation Systems), the Knut and Alice Wallenberg Foundation, and Science for Life Laboratory. We acknowledge The Swedish Twin Registry for access to samples and data. The Swedish Twin Registry is managed by Karolinska Institutet and receives funding through the Swedish Research Council under the grant no 2017-00641. LifeGene was supported by grants from the Swedish Research Council, Torsten and Ragnar Söderbergs Foundation, Stockholm County Council, and AFA Försäkringar.

The authors declare no conflict of interest.

Researchers interested in using STR data must obtain approval from a Swedish Ethical Review Board and from the Steering Committee of the Swedish Twin Registry. Researchers using the data are required to follow the terms of an agreement containing a number of clauses designed to ensure protection of privacy and compliance with relevant laws. For further information, contact Patrik Magnusson (Patrik.Magnusson@ki.se).

7. AUTHOR CONTRIBUTIONS

M-G.H. and J.M.S. designed the study. T.D-C., K.D., and R.S. acquired the proteomic data and performed the analyses of the data. M-G.H., X.C., and Y.W. analyzed GWAS data. M-G.H., T.D-C., W.L., Y.P., S.H., P.K.E.M., and J.M.S. performed statistical analyses. J.O., A.H., A.S., P.H., N.L.P., and P.K.E.M. generated phenotypic data and contributed the samples for present study. All authors were involved in writing and reviewing the manuscript.

8. REFERENCES

Almqvist, C., Adami, H.-O., Franks, P. W., Groop, L., Ingelsson, E., Kere, J., . . . Pedersen, N. L. (2011). LifeGene--a large prospective population-based study of global relevance. *European journal of epidemiology, 26*(1), 67-77. doi:10.1007/s10654-010-9521-x

Ayoglu, B., Häggmark, A., Neiman, M., Igel, U., Uhlén, M., Schwenk, J. M., & Nilsson, P. (2011). Systematic antibody and antigen-based proteomic profiling with microarrays. *Expert review of molecular diagnostics, 11*(2), 219-234. doi:10.1586/erm.10.110

Baldassarre, D., Nyyssönen, K., Rauramaa, R., de Faire, U., Hamsten, A., Smit, A. J., . . . group, I. s. (2010). Cross-sectional analysis of baseline data to identify the major determinants of carotid intima-media thickness in a European population: the IMPROVE study. *European heart journal, 31*(5), 614-622. doi:10.1093/eurheartj/ehp496

Barron, E., Lara, J., White, M., & Mathers, J. C. (2015). Blood-borne biomarkers of mortality risk: systematic review of cohort studies. *PloS One, 10*(6), e0127550. doi:10.1371/journal.pone.0127550

Bolin, M., Akerud, P., Hansson, A., & Akerud, H. (2011). Histidine-rich glycoprotein as an early biomarker of preeclampsia. *American Journal of Hypertension, 24*(4), 496-501. doi:10.1038/ajh.2010.264

Byström, S., Ayoglu, B., Häggmark, A., Mitsios, N., Hong, M. G., Drobin, K., . . . Schwenk, J. M. (2014). Affinity proteomic profiling of plasma, cerebrospinal fluid, and brain tissue within multiple sclerosis. *Journal of Proteome Research, 13*(11), 4607-4619. doi:10.1021/pr500609e

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*(3), 177-188.

Dieterle, F., Ross, A., Schlotterbeck, G., & Senn, H. (2006). Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Analytical chemistry, 78*(13), 4281-4290. doi:10.1021/ac051632c

Drasin, T., & Sahud, M. (1996). Blood-type and age affect human plasma levels of histidine-rich glycoprotein in a large population. *Thrombosis Research, 84*(3), 179-188.

Drobin, K., Nilsson, P., & Schwenk, J. M. (2013). Highly multiplexed antibody suspension bead arrays for plasma protein profiling. *Methods in Molecular Biology, 1023*, 137-145. doi:10.1007/978-1-4614-7209-4_8

Emilsson, V., Ilkov, M., Lamb, J. R., Finkel, N., Gudmundsson, E. F., Pitts, R., . . . Gudnason, V. (2018). Co-regulatory networks of human serum proteins link genetics to disease. *Science (New York, N.Y.), 48*, eaaq1327. doi:10.1126/science.aaq1327

Farh, K. K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W. J., Beik, S., . . . Bernstein, B. E. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature, 518*(7539), 337-343. doi:10.1038/nature13835

Gabrielson, M., Eriksson, M., Hammarstrom, M., Borgquist, S., Leifland, K., Czene, K., & Hall, P. (2017). Cohort Profile: The Karolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA). *International Journal of Epidemiology, 46*(6), 1740-1741g. doi:10.1093/ije/dyw357

Ganna, A., & Ingelsson, E. (2015). 5 year mortality predictors in 498 103 UK Biobank participants: a prospective population-based study. *Lancet, 386*(9993), 533-540. doi:10.1016/S0140-6736(15)60175-1

Ganna, A., Rivadeneira, F., Hofman, A., Uitterlinden, A. G., Magnusson, P. K., Pedersen, N. L., . . . Tiemeier, H. (2013). Genetic determinants of mortality. Can findings from genome-wide

association studies explain variation in human mortality? *Human Genetics, 132*(5), 553-561. doi:10.1007/s00439-013-1267-6

Gorbunova, V., Seluanov, A., Zhang, Z., Gladyshev, V. N., & Vijg, J. (2014). Comparative genetics of longevity and cancer: insights from long-lived rodents. *Nature reviews. Genetics, 15*(8), 531-540. doi:10.1038/nrg3728

Grambsch, P. M., & Therneau, T. M. (1994). Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. *Biometrika, 81*(3), 515-526.

Harrell, F. E., Jr., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. *JAMA, 247*(18), 2543-2546.

Hong, M.-G., Lee, W., Nilsson, P., Pawitan, Y., & Schwenk, J. M. (2016). Multidimensional Normalization to Minimize Plate Effects of Suspension Bead Array Data. *Journal of Proteome Research, 15*(10), 3473-3480. doi:10.1021/acs.jproteome.5b01131

Hubert, M., Rousseeuw, P. J., & Branden, K. V. (2005). ROBPCA: A New Approach to Robust Principal Component Analysis. *Technometrics, 47*(1), 64-79. doi:10.1098/004017004000000563

Jylhävä, J., Pedersen, N. L., & Hägg, S. (2017). Biological Age Predictors. *EBioMedicine, 21*, 29-36. doi:10.1016/j.ebiom.2017.03.046

Kassaar, O., McMahon, S. A., Thompson, R., Botting, C. H., Naismith, J. H., & Stewart, A. J. (2014). Crystal structure of histidine-rich glycoprotein N2 domain reveals redox activity at an interdomain disulfide bridge: implications for angiogenic regulation. *Blood, 123*(12), 1948-1955. doi:10.1182/blood-2013-11-535963

Katsimpardi, L., Litterman, N. K., Schein, P. A., Miller, C. M., Loffredo, F. S., Wojtkiewicz, G. R., . . . Rubin, L. L. (2014). Vascular and neurogenic rejuvenation of the aging mouse brain by young systemic factors. *Science, 344*(6184), 630-634. doi:10.1126/science.1251141

Lichtenstein, P., De Faire, U., Floderus, B., Svartengren, M., Svedberg, P., & Pedersen, N. L. (2002). The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. *Journal of internal medicine, 252*(3), 184-205.

López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The Hallmarks of Aging. *Cell, 153*(6), 1194-1217. doi:10.1016/j.cell.2013.05.039

Magnusson, P. K. E., Almqvist, C., Rahman, I., Ganna, A., Viktorin, A., Walum, H., . . . Lichtenstein, P. (2013). The Swedish Twin Registry: establishment of a biobank and other recent developments. *Twin research and human genetics, 16*(1), 317-329. doi:10.1017/thg.2012.104

Marioni, R. E., Shah, S., McRae, A. F., Chen, B. H., Colicino, E., Harris, S. E., . . . Deary, I. J. (2015). DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biology, 16*, 25. doi:10.1186/s13059-015-0584-6

Melzer, D., Perry, J. R., Hernandez, D., Corsi, A. M., Stevens, K., Rafferty, I., . . . Ferrucci, L. (2008). A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genetics, 4*(5), e1000072. doi:10.1371/journal.pgen.1000072

Morgan, W. T., Koskelo, P., Koenig, H., & Conway, T. P. (1978). Human Histidine-Rich Glycoprotein. II. Serum Levels in Adults, Pregnant Women and Neonates. *Proceedings of the Society for Experimental Biology and Medicine, 158*(4), 647-651. doi:10.3181/00379727-158-40265

Odeberg, J., Freitag, M., Forssell, H., Vaara, I., Persson, M. L., Odeberg, H., . . . Lindblad, U. (2014). The influence of smoking and impaired glucose homoeostasis on the outcome in

patients presenting with an acute coronary syndrome: a cross-sectional study. *BMJ Open, 4*(7), e005077. doi:10.1136/bmjopen-2014-005077

Poon, I. K. H., Patel, K. K., Davis, D. S., Parish, C. R., & Hulett, M. D. (2011). Histidine-rich glycoprotein: the Swiss Army knife of mammalian plasma. *Blood, 117*(7), 2093-2101. doi:10.1182/blood-2010-09-303842

Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., . . . Willer, C. J. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics, 26*(18), 2336-2337. doi:10.1093/bioinformatics/btq419

Samnegård, A., Silveira, A., Lundman, P., Boquist, S., Odeberg, J., Hulthe, J., . . . Eriksson, P. (2005). Serum matrix metalloproteinase-3 concentration is influenced by MMP-3 -1612 5A/6A promoter genotype and associated with myocardial infarction. *Journal of internal medicine, 258*(5), 411-419. doi:10.1111/j.1365-2796.2005.01561.x

Schnabel, R. B., Yin, X., Larson, M. G., Yamamoto, J. F., Fontes, J. D., Kathiresan, S., . . . Benjamin, E. J. (2013). Multiple inflammatory biomarkers in relation to cardiovascular events and mortality in the community. *Arteriosclerosis, Thrombosis, and Vascular Biology, 33*(7), 1728-1733. doi:10.1161/ATVBAHA.112.301174

Schwenk, J. M., Igel, U., Kato, B. S., Nicholson, G., Karpe, F., Uhlén, M., & Nilsson, P. (2010). Comparative protein profiling of serum and plasma using an antibody suspension bead array approach. *Expert Review of Proteomics, 10*, 532-540.

Schwenk, J. M., Igel, U., Neiman, M., Langen, H., Becker, C., Bjartell, A., . . . Uhlén, M. (2010). Toward next generation plasma profiling via heat-induced epitope retrieval and array-based assays. *Molecular & Cellular Proteomics, 9*(11), 2497-2507. doi:10.1074/mcp.M110.001560

Schwenk, J. M., Omenn, G. S., Sun, Z., Campbell, D. S., Baker, M. S., Overall, C. M., . . . Deutsch, E. W. (2017). The Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays. *Journal of Proteome Research, 16*(12), 4299-4310. doi:10.1021/acs.jproteome.7b00467

Sjoberg, R., Sundberg, M., Gundberg, A., Sivertsson, A., Schwenk, J. M., Uhlen, M., & Nilsson, P. (2012). Validation of affinity reagents using antigen microarrays. *N Biotechnol, 29*(5), 555-563. doi:10.1016/j.nbt.2011.11.009

Statistics Sweden. (2016). *Life expectancy in Sweden 2011–2015. Life tables at national and county level*.

Stoevesandt, O., & Taussig, M. J. (2012). Affinity proteomics: the role of specific binding reagents in human proteome analysis. *Expert Rev Proteomics, 9*(4), 401-414. doi:10.1586/epr.12.34

Suhre, K., Arnold, M., Bhagwat, A. M., Cotton, R. J., Engelke, R., Raffler, J., . . . Graumann, J. (2017). Connecting genetic risk to disease end points through the human blood plasma proteome. *Nature communications, 8*, 14357. doi:10.1038/ncomms14357

Sun, B. B., Maranville, J. C., Peters, J. E., Stacey, D., Staley, J. R., Blackshaw, J., . . . Butterworth, A. S. (2018). Genomic atlas of the human plasma proteome. *Nature, 558*(7708), 73-79. doi:10.1038/s41586-018-0175-2

Tanaka, T., Biancotto, A., Moaddel, R., Moore, A. Z., Gonzalez-Freire, M., Aon, M. A., . . . Ferrucci, L. (2018). Plasma proteomic signature of age in healthy humans. *Aging cell*, e12799. doi:10.1111/acel.12799

Thiébaut, A. C. M., & Bénichou, J. (2004). Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. *Statistics in Medicine, 23*(24), 3803-3820. doi:10.1002/sim.2098

Uhlén, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., . . . Ponten, F. (2015). Proteomics. Tissue-based map of the human proteome. *Science, 347*(6220), 1260419. doi:10.1126/science.1260419

9. SUPPORTING INFORMATION

Supporting Text - Materials and methods

10. TABLES

Table 1. Sample sets

Study set	Age [yr]	Sex (F : M)	Sample Type	Indication*	Cohort name	References
Set 1	50 – 92	78 : 78	Serum	Population	TwinGene	(Lichtenstein et al., 2002; Magnusson et al., 2013)
Set 2	3 – 6	6 : 6	Serum	Population	LifeGene	(Almqvist et al., 2011)
	9 – 63	102 : 102	Plasma	Population		
Set 3	48 – 93	1613 : 1386	Serum	Population	TwinGene	(Lichtenstein et al., 2002; Magnusson et al., 2013)
Set 4	51 – 86	50 : 0		Breast cancer		
Set 5	56 – 75	0 : 50		Prostate cancer		
Set 6	55 – 78	16 : 27	Plasma	Cardiovascular disease	IMPROVE	(Baldassarre et al., 2010)
Set 7	41 – 60	12 : 31	Plasma	Myocardial infarction	SCARF	(Samnegård et al., 2005)
Set 8	48 – 73	20 : 23	Plasma	Acute coronary heart syndrome	CHAPS	(Odeberg et al., 2014)
Set 9	40 – 73	600 : 0	Plasma	Mammography	KARMA	(Gabrielson et al., 2017)

* Subjects included in the presented study did not include subjects diagnosed with the disease of the indication area, but the subjects assigned as controls for the different disease cohorts.

Table 2. Antibody profiles with strongest association to age

HPA ID	FDR	FDR – Twin†	Within-pair correlation‡
HPA045005	3.92×10^{-11}	4.07×10^{-10}	0.428
HPA039928	6.83×10^{-10}	2.15×10^{-9}	0.312
HPA029931	1.00×10^{-5}	5.30×10^{-5}	0.404

† The false discovery rate (FDR) of the linear mixed effects model with random effects of twin pair.
‡Intraclass correlation coefficient between within-pair of 30 twin pairs.

11. FIGURE LEGENDS

Figure 1. Meta-analysis from 9 different sample sets

In the forest plot, the numbers in parenthesis indicate the age range of the included subjects. For each sample set, the estimated effect from the linear regression model, 95% confidence interval of it, and study weight in the meta-analysis are shown as a tick, a line, and a grey box, respectively, in the middle. The numeric value of the effect is clarified at the right side.

Figure 2. GWAS results of the age-associated plasma profile

(A) Manhattan plot. The significance of association between genotypes and HPA045005 profiles is presented vertically. The dashed guide line marks the stringent threshold of P-value for GWAS, which is P=0.01 after Bonferroni correction. One peak in chromosome 3 indicates strong association of a locus with the molecular phenotype. (B) LocusZoom (Pruim et al., 2010) on associated locus. The illustration shows the elements of chromosome 3 associated with HPA045005 profiles. Zooming in on the peak of the Manhattan plot in (A), the genes around the locus are presented together with the associated SNPs. (C) Box plots to show the association between genotypes of rs9898 and two antibody profiles, HPA045005 and BSI0137. The trends were opposite.

Figure 3. Survival analysis comparing upper and lower quarters of HRG levels

The individuals of sample set 3 were divided into four subsets by the quartiles of HRG levels. Differential mortality across follow-up time is illustrated by the survival curves. Some detailed statistics related to this survival analysis are presented in Supporting Table 3.

12. FIGURES

Figure 1.

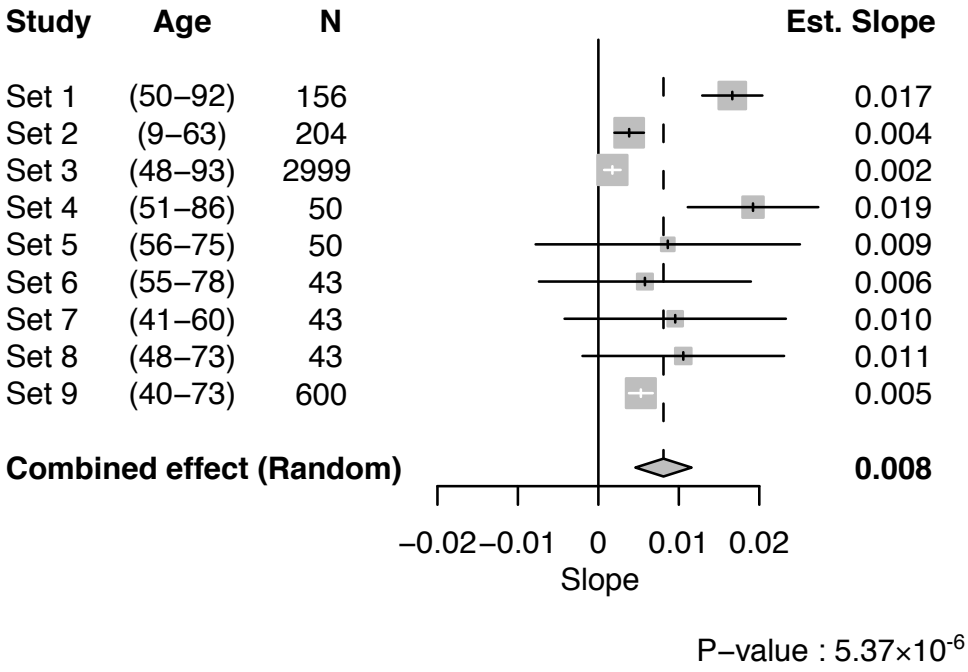

Figure 2.

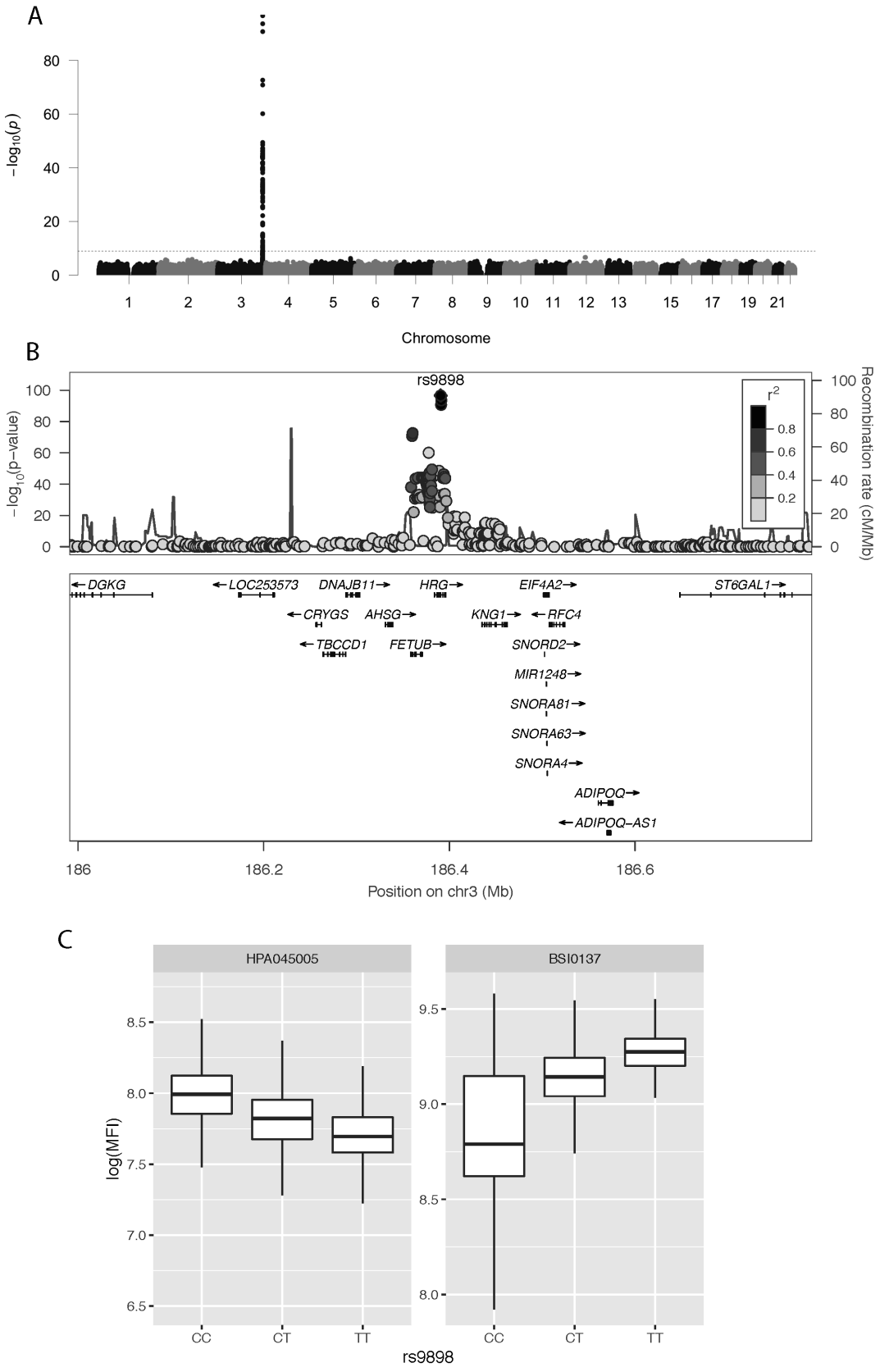

Figure 3.