Multivariate Genome-Wide and Integrated Transcriptome and Epigenome-Wide Analyses of the Well-being Spectrum.

B.M.L. Baselmans[1,3], R. Jansen[2], J. van Dongen[1,3], Y. Bao[5], M. Smart[5], M. Kumari[5], A. Abdellaoui[1,3], M. P. van de Weijer[1], G. Willemsen[1,3,4], J.J. Hottenga[1,3,4], BIOS consortium*, Social Science Genetic Association Consortium*, E.J.C. de Geus[1,3,4], D.I. Boomsma[1,3,4], M.G. Nivard[1,3,6], M. Bartels[1,3,4,6]

Affiliations

[1] Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, 1081 BT, the Netherlands

[2] Department of Psychiatry, VU University Medical Center, Amsterdam, 1081 HL, the Netherlands

[3] Amsterdam Public Health Research Institute, Amsterdam, 1081 BT, the Netherlands

[4] Amsterdam Neuroscience, Amsterdam, 1081 HV, the Netherlands

[5] Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, UK

[6] These authors jointly supervised the work and are the corresponding authors

* These consortia are acknowledged as banner-co-authors for the key role their previous work played. Please find a detailed description of their role and membership at the end of the manuscript

Acknowledgements:

We would like to thank all the twins and family members for their participation. This work was supported by the Netherlands Organization for Scientific Research (NWO: MagW/ZonMW grants 904-61-090, 985-10-002,904-61-193,480-04-004, 400-05-717, NWO-bilateral agreement 463-06-001, NWO-VENI 451-04-034, Addiction-31160008, Middelgroot-911-09-032, Spinozapremie 56-464-14192), Biobanking and Biomolecular Resources Research Infrastructure (BBMRI –NL, 184.021.007), the VU University's Institute for Health and Care Research (EMGO+) and Neuroscience Campus Amsterdam (NCA), the European Science Council (ERC Advanced, 230374), the Avera Institute for Human Genetics, Sioux Falls, South Dakota (USA) and the National Institutes of Health (NIH, R01D0042157-01A). Part of the genotyping was funded by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health (NIMH, MH081802) and by the Grand Opportunity grants 1RC2MH089951-01 and 1RC2 MH089995-01 from the NIMH. Part of the analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org), which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003), the Dutch Brain Foundation, and the department

of Psychology and Education of the VU University Amsterdam. M.Bartels is/was financially supported by a senior fellowship of the (EMGO+) Institute for Health and Care and a VU University Research Chair position. MGN is supported by Royal Netherlands Academy of Science Professor Award (PAH/6635) awarded to DIB and supported by Biobanking and Biomolecular Resources Research Infrastructure (BBMRI–NL, 184.021.007.).

Understanding Society is an initiative funded by the Economic and Social Research Council (ES/H029745/1) and various Government Departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public. The research data are distributed by the UK Data Service. The genome-wide scan data were analysed and deposited by the Wellcome Trust Sanger Institute. Information on how to access the data can be found on the Understanding Society website https://www.understandingsociety.ac.uk/. Genotype-phenotype data access for UKHLS is available by application to Metadac: www.metadac.ac.uk.

COMPETING FINANCIAL INTEREST: The authors declare not competing financial interest

CORRESPONDING AUTHORS: Meike Bartels (m.bartels@vu.nl) and Michel Nivard (m.g.nivard@vu.nl)

Several phenotypes related to well-being (e.g., life satisfaction, positive affect, neuroticism, and depressive symptoms), are genetically highly correlated ($|r_g| > .75$). Multivariate analyses of these traits, collectively referred to as the well-being spectrum, reveals 24 genome-wide significant loci. We integrated the genetic findings with large human transcriptome and epigenome datasets. Integrated analyses implicate gene expression at 48 additional loci and CpG methylation at 28 additional loci in the etiology of well-being.

Well-being plays an important role in economics, psychology, and medicine[1,2]. Well-being owes it interdisciplinary prominence to its associations with physical and mental health, and its role as a desired socio-economic outcome and index of economic development[3]. Most existing research on behavioral traits is characterized by a focus on a single phenotype, despite the strong correlations between related traits. The high genetic correlations ($|r_g| > .75$)[4] between life satisfaction, positive affect, neuroticism, and depressive symptoms suggest a common liability. We performed a multivariate genome-wide meta-analysis (multivariate GWAMA) (N= 584, 302) of these four phenotypes to increase the power to identify associated genetic variants (**Supplementary Table 1)**.

Our analyses leveraged publically available univariate GWAMA[4] of life satisfaction (N = 71,650), positive affect (N = 180,281), neuroticism (N = 170,911), and depressive symptoms (N = 161,460). For the purpose of the multivariate GWAMA, we reversed the estimated SNP effects on neuroticism and depressive symptoms to ensure a positive correlation with life satisfaction and positive affect. The dependence between effect sizes (error correlation), induced by sample overlap, was estimated from genome-wide summary statistics obtained from univariate GWAMA (http://www.thessgac.org) using LD score regression (see online methods). Knowledge of the error correlation between univariate meta-analyses allowed otherwise unused (dependent) samples to be meta-analyzed, providing a gain in power while guarding against inflated type I error rates (see online methods).

We recognize that the measures included in the well-being spectrum are not necessarily interchangeable. Therefore, we performed two types of multivariate GWAMA; 1) N-weighted multivariate GWAMA, which assumes a single underlying construct (see online methods); 2) A model-average GWAMA, where we relaxed the assumption of a unitary effect of the SNP on all traits. For the latter analyses, we performed eight model-based GWAMA´s, capturing various combinations of SNP effects (see online methods). To account for within and between model variability, we took a weighted average of the effect size and standard error for each SNP, across the eight models. For each of the eight models we computed an AICc weight to ensure that all models are weighted properly[5]. The two multivariate analysis methods, we developed, can cope with multiple traits, repeated measures, and can be applied to cohort level and meta-analysis summary statistics. The model average GWAMA can further accommodate existing multivariate GWAMA techniques and models. Together these techniques are excellent future proof tools

for analysis in a field where sample overlap is complex to track and large existing GWAMA projects need to be integrated with massive novel data sources.

In our N-weighted multivariate GWAMA*, we identified 16 significant loci (**Fig. 1A**). Ten of these loci have not previously been associated with any of the traits[4]. Twelve loci, which were identified in previous univariate analyses of these traits[4], were not significant in the N-weighted multivariate GWAMA (**Supplementary Table 2-3,8**). Our model averaging GWAMA* (**Fig. 1B-E**) detected 20 significant loci, of which 12 have not previously been associated with these traits (**Supplementary Table 2, 4-7)**. Eight loci identified in univariate analyses of these traits were not significant in the model averaging GWAMA. An overview of all genetic variants, previously associated with the broader well-being spectrum and the ones reported in the current manuscript, are presented in **Supplementary Table 8**. For both multivariate strategies we observed, in comparison to univariate GWAMA, higher median test statistics (N-weighted λ_{GC} = 1.297, average model-based λ_{GC} = 1.217), but almost no increase in LD score intercept (N-weighted =1.01, average model-based = 0.94). The low LD score intercept confirmed that the inflation is due to an increase in power to detect causal polygenic signal, rather than population stratification or inaccurate accounting for sample overlap (see online methods, **Supplementary Table 9, Supplementary Fig. 1**).

* We compared our GWAMA results with the previous result as reported in Okbay et al. (2016) Note that this is a conservative choice, because in our analyses we use the publically available results (www.thessgac.org), in which a large cohort (23andMe) specifically from the LS results, is omitted, resulting in a unavoidable loss of power.

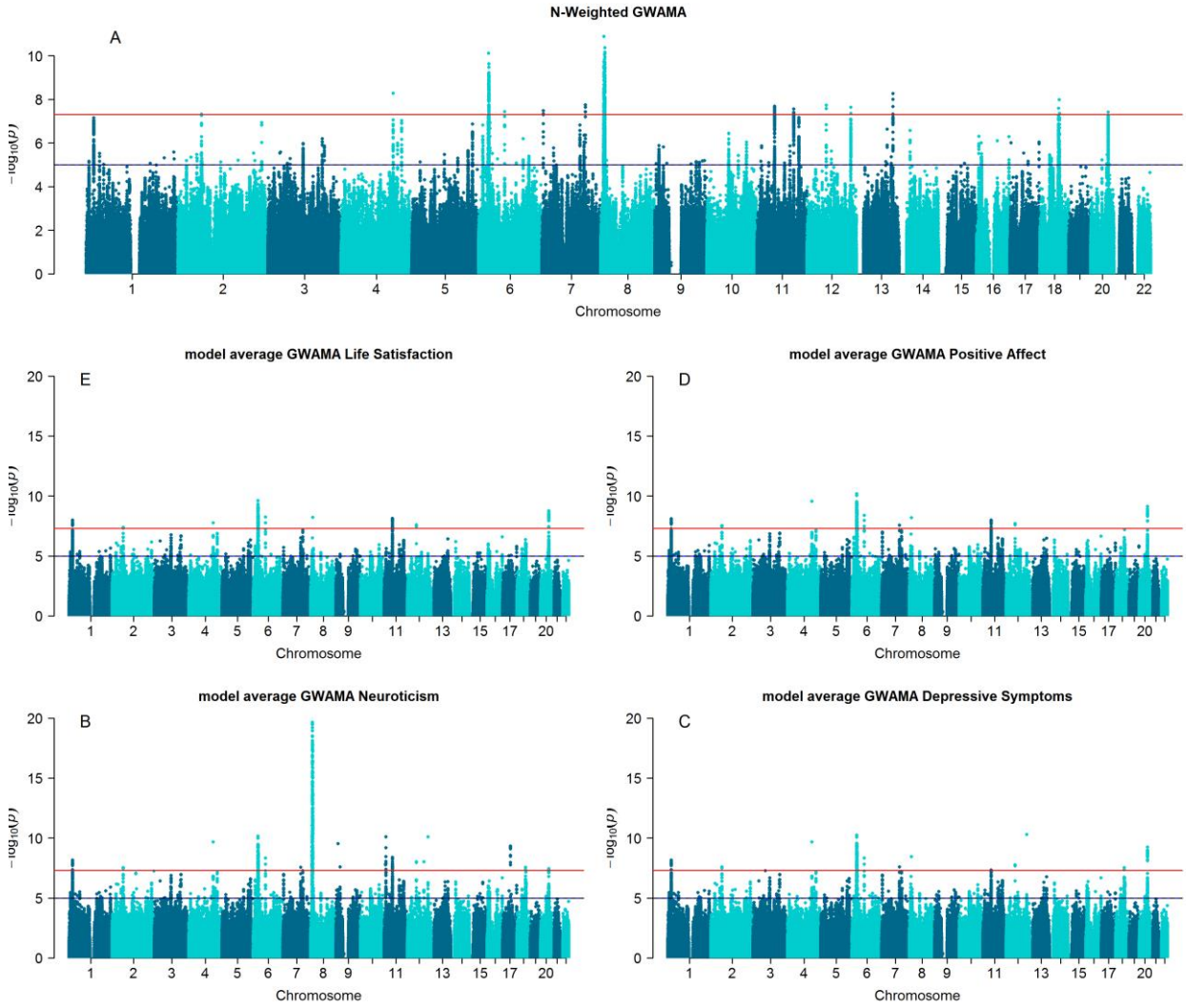

Figure 1. Panel A contains a Manhattan plot where the $-\log_{10}$(P-value) obtained from our N-weighted GWAMA of a SNP effect on life satisfaction, positive affect, neuroticism, and depressive symptoms is plotted against the genomic location for the SNP. Panel B contains four Manhattan plots, these are obtained from our model averaging GWAMA, where a series of models is fit to GWAMA summary statistics for life satisfaction, positive affect, neuroticism, and depressive symptoms, and the SNP effect is evaluated with respect to each of the four outcomes.

To confirm the gain in power of our multivariate GWAMA results, we performed polygenic risk score prediction (PRS) in two independent samples (online methods). We predicted the phenotypes in the well-being spectrum (life satisfaction, positive affect, neuroticism, and depressive symptoms). The combined PRS based on the N-weighted GWAMA improved prediction with an average increase in R^2 of 0.36 (1.5-fold increase), ranging from a 0.08 to a 0.60 R^2 increase (i.e., a 0.08 to 4.6-fold increase). PRS based on

model-average GWAMA improved with an average R^2 of 0.37 (i.e., a 1.6-fold increase), ranging from a 0.05 to a 0.62 R^2 increase (i.e., 0.05 to 4.8-fold increase, **Fig. 2, Supplementary Table 11**).

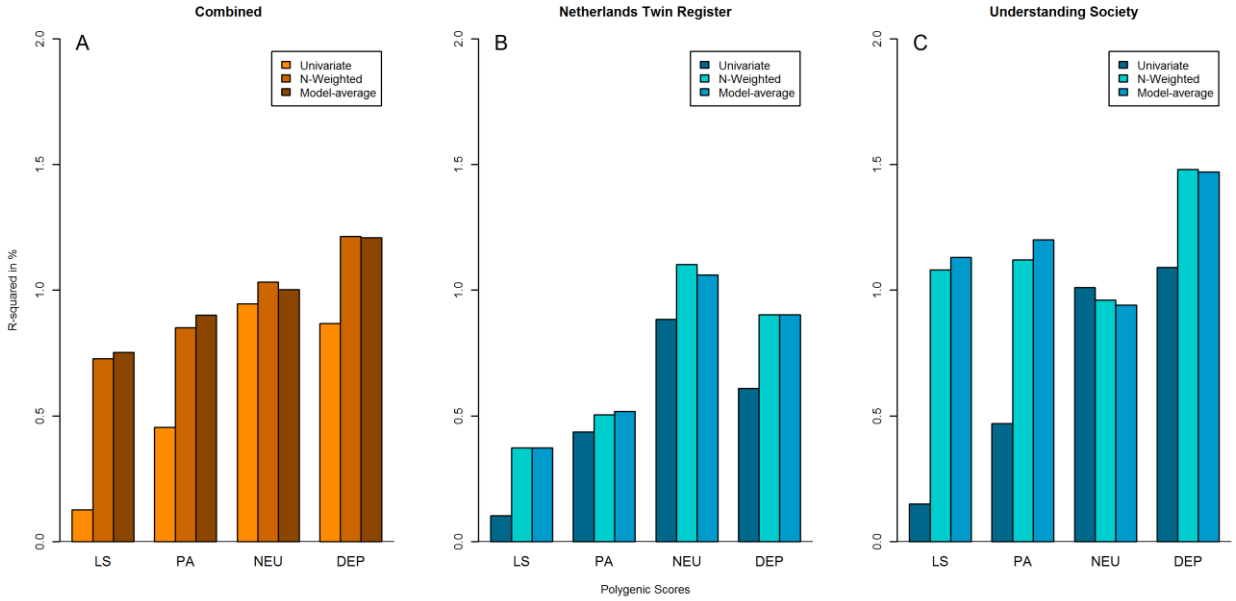

Figure 2. The result of polygenic risk prediction based on univariate discovery GWAMA, N-weighted discovery GWAMA or model averaging discovery GWAMA. The unit on the Y-axis is the squared standardized beta obtained from a regression of the trait on the PRS, age, sex and 10 principle components. A displays the combined N-weighted polygenic score results. B displays the polygenic prediction results from the Netherlands Twin Register and C displays the polygenic results from Understanding Society.

We searched the HGRI-EBI catalog[6] of published genome-wide association studies (GWAS catalog) for genes located near loci associated with well-being in order to identify pleiotropic effects **(Supplementary Fig. 2)**. In total, we found 113 genes within a 250kb distance of the independent loci, of which 47 genes were previously reported in other studies. Of these 47 genes, a substantial number (40.4%) is previously associated with neuropsychiatric traits, followed by immune disease (12.7%), and drug response associations (12.7%). For a complete overview, see **Supplementary Table 10**.

Both multivariate GWAMAs aggregated the effect of a single SNP across multiple traits based on prior knowledge of the genetic correlation between these traits. In a similar fashion, we proceeded to aggregate the effect across multiple SNPs based on prior knowledge that these SNPs collectively influence the expression level of a gene transcript or the methylation level at a CpG site**.** Given the equal performance

of both multivariate GWAMAs and to avoid multiple testing, we used the results of the N-weighted GWAMA. Aggregation of SNP effect, which have a common effect on gene expression or CpG methylation, lowers the multiple testing burden, strengthens the signal, and identifies specific gene transcripts or CpG methylation sites.

In this procedure, known as summary- based transcriptome-wide and methylome-wide association analyses (TWAS and MWAS)[7,8], we imputed the effect of changes in gene expression or CpG site methylation on the well-being spectrum. Information on the relation between SNP and CpG methylation or gene expression were obtained from the summarized results of cis-expression (e)QTL and cis-methylation (m)QTL studies in whole blood[9,10]. Since there is evidence that cis-regulation of gene expression is highly consistent across tissues (genetic correlation brain-blood = .66)[11], we are confident in using eQTL results obtained based on gene expression measure in blood as a proxy for expression in brain tissue. We considered TWAS and MWAS loci significant at a Benjamini Hockberg adjusted p-value < 0.05[12]. Our analyses uncovered 60 transcript-trait associations, of which 48 are distinct loci found more than 500kb from a significant GWAMA locus (**Supplementary Table 12**). We uncovered 195 CpG methylation well-being associations, mapping to 28 distinct novel loci more than 500kb from a significant GWAMA locus (**Supplementary Table 13**). TWAS or MWAS imply an effect of gene expression or CpG site on the trait[7]. Reverse causation, where variation in well-being influences gene expression or CpG site methylation, needs to be tested. To this end, we computed 22 PRS based on the N weighted GWAMA, consecutively omitting one of the 22 chromosomes (see online methods). With these PRS, we predicted the CpG site methylation and gene expression on the omitted chromosome. Neither gene expression nor methylation level were predicted by the PRS (**Supplementary Table 14-15**). This is inconsistent with the presence of reversed causality (i.e., the model in which well-being causes expression and methylation).

We searched the GWAS catalog[6] for both our TWAS and MWAS hits that did not reach genome-wide significance in the GWAMA analyses (>500kb distance, **Supplementary Fig. 2**). Based on TWAS, we found 48 genes, of which 28 genes were previously associated with other traits. Of these 28 genes, 39% were associated with biochemical markers, 32% were associated with neuropsychiatric traits, and 28.6% with anthropometric features. Based on MWAS, 28 genes were located near 44 CpG sites. Of these 28 genes, 18 were previously associated with other traits, 66% of these were associated with neuropsychiatric or neuropsychiatric drug response, 50% were associated without otherwise specified diseases or traits, and 38.9% were associated with biochemical markers. The strong pleiotropy with neuropsychiatric traits, strengthens our confidence that the loci found in TWAS and/or MWAS results are valid well-being spectrum loci. A complete overview is provided in **Supplementary Table 16-17**.

Four loci are associated with well-being in all three GWAMA, TWAS and MWAS analyses, which suggests a role for both gene expression and CpG methylation at these loci in well-being, or possible mediation of the relation between gene expression and well-being by CpG methylation (**Supplementary Fig. 3-4)**. The locus at chromosome 6 within the MHC region is of special interest. Recent work has identified 3 individual signals related to schizophrenia in the MHC region, one of which is linked to complement 4 (C4A) gene expression and synapse elimination during puberty[13]. The genome-wide significant signal for well-being in the MHC region is not in strong LD with lead eQTL's for C4A gene expression. Rather a second independent signal tagged by rs13194504 is associated with both schizophrenia and well-being. TWAS and MWAS results for the MHC region implicate the expression of ZKSCAN4 and methylation of cg08798685 in the etiology of well-being.

We performed further biological annotation using LD score regression (**Supplementary Table 18)**. Our analyses revealed significant enrichment of SNP effects in regions of the genome characterized by active histone marks in the following brain tissues: fetal brain tissue, the germinal matrix (a highly cellular and highly vascularized region in the brain from which cells migrate out during brain development), the mid-frontal lobe, inferior temporal lobe, angular gyrus, middle hippocampus, cingulate gyrus, and anterior caudate (**Supplementary Fig. 5**). We found further enrichment of SNP effects in a single endocrine tissue, specifically the Thymus. Note that endocrine tissue that is more frequently implicated in depression (e.g., the thyroid and pituitary) were not tested. Further analysis using LD score regression, where the LD scores are based on evolutionary markers[14], revealed that SNPs of very recent origin (i.e. a low allelic age) explained substantially more variation than ancient SNPs (**Supplementary Table 19 and Supplementary Fig. 6)**. These findings indicate the presence of negative selection shaping variation in well-being through recent evolution **(**see online methods**).**

In summary, previous univariate analyses of phenotypes in the well-being spectrum were moderately successful. Our analysis gained power through multivariate GWAMA and we identified novel loci associated with well-being. Model averaging GWAMA identified additional loci, associated with some but not all traits in the well-being spectrum, and provided flexibility in terms of model specification. Model averaging can in fact incorporate *any* multivariate GWAMA or GWAS model for which the per SNP model fit can be expressed in terms of an AICc fit statistic. The averaging procedure is done per locus, allowing for heterogeneity across phenotype and loci. The model averaging procedure is not limited to analyses of multiple GWAMA but can in fact be applied to tackle heterogeneity between individual cohorts in GWAMA. We found that TWAS and MWAS can yet further increase the pool of loci related to variation in complex traits, like well-being. These analyses would not have been possible were it not for good scientific practice of openly sharing summary level data. We foresee multivariate GWAMA will

play an important role in the ongoing discussion on the boundaries of the set of phenotypes, which define "well-being", or more narrowly in studying heterogeneity in the psychiatrically relevant subset of traits or symptoms, which pertain to "depression". We propose model averaging as an agnostic approach, which enables the use of a very broad phenotype definition and gives researchers the explicit choice of considering multiple multivariate models, maximizing power for locus discovery while retaining the ability to detect trait specific effects.

Online Methods

N-weighted multivariate GWAMA.

We obtained summary statistics from previous analyses in Okbay et al.[4] , where multiple cohorts contributed to the univariate GWAMAs of life satisfaction, positive affect, neuroticism and depressive symptoms http://www.thessgac.org/. To quantify the dependence between the univariate GWAMAs, we estimated the cross trait LD score intercept (CTI)[15,16].

$$CTI = \frac{N_s * r_p}{\sqrt{N_1 N_2}}$$

Where N_s equals the sample overlap, N_1 the sample size for trait one and N_2 the sample size for trait two, r_p equals the phenotypic correlation between trait one and two. The CTI is approximately equal to the covariance between the test statistics obtained in a GWAMA of trait 1 and trait 2. We assume the estimated CTI is equal to the true CTI, though note the uncertainty in the estimated CTI is generally low. Given the estimated covariance between effect sizes we can meta-analyse the four dependent GWAMAs and obtain a multivariate test statistic per SNP:

$$Z_k = \frac{\sum_{i=1}^{4} w_{ik} * Z_{ik}}{\sqrt{\sum_{i=1}^{4} w_{ik} * V_{ik} + \sum_{i=1}^{4}\sum_{j=1}^{4} \sqrt{w_{ik} * w_{jk}} * C_{i,j,k}\ (j \neq i)}}$$

Where w_{ik} is the square root of the sample size for SNP k in the GWAMA of trait i, Z_{ik} is the test statistics of SNP k in the GWAMA of trait i; V_{ik} is the variance of the test statistic for SNP k in the GWAMA of trait i (i.e 1 given that Z is a standardized test statistic) and $C_{i,j,k}$ is the covariance between test statistics for SNP k between GWAMA of trait i and trait j (where C equals CTI obtained from cross trait LD score regression between trait i and trait j). The multivariate test statistic Z_k, is a standardized sum of tests statistics all of which follow a normal distribution under their respective null distributions, the statistic Z_k follows a standard normal distribution under the null hypothesis of no effect.

Model averaging GWAMA

Consider the following model:

$$\beta = MVN(\gamma X + e, V)$$

Where β (1xn) is a multivariate normal vector of effect sizes obtained from the regression of n standardized phenotypes on a standardized genotype (SNP). The matrix V (nxn) is the variance-covariance matrix of effect sizes, matrix X a design matrix (pxn), and γ the corresponding vector of parameters (1xp). The indexed p denotes the number of variables included in the means model of the response vector β.

In this context, a regular GWAMA restricts the design matrix X to a unit vector (i.e. we model a single genetic effect, which is assumed identical across cohorts, and any observed variation is attributed to sample fluctuation). Generally, matrix V is diagonal, and contains the squared standard errors of elements in β. A regular GWAMA is the most restricted model one can consider. However, when considering multivariate GWAMA (i.e. the elements in β reflect SNP effects on separate yet correlated phenotypes) this model might be too restrictive even when traits have a substantial genetic correlation, not all genetic effects need to be shared between traits or be identical in magnitude. The least restrictive model is to consider the SNP effects in β independent (i.e. run univariate GWAMA of the correlated phenotypes). In between the most restrictive and least restrictive model, a manifold of models can be specified, equating the effects in y across combinations of traits, while allowing it to differ between other combinations of traits. These models can be specified by ways of the design matrix X.

One could consider a manifold (z) of models (m), each with a different design matrix X.

$$\beta_1 = MVN(\gamma_1 X_1 + e, V)$$

$$\beta_2 = MVN(\gamma_2 X_2 + e, V)$$

…

$$\beta_z = MVN(\gamma_z X_z + e, V)$$

When considering i correlated phenotypes, a simple expansion of X is to allow for 2 vectors (p =2), a unit vector and a second vector which is coded dichotomously (0,1) where the coding varies over each of the m models. Other codings, based on analysis of the genetic correlation between traits (i.e. PCA or Cholesky decomposition), can be applied to summary statistics and included in the average. Practically, this allows for the existence of 2 distinct genetic effect. This procedure results in $.5 * k^2$ models. The 1df model with a unit vector for X and $.5 * k^2 - 1$ 2-df models with a unit vector and a second vector which codes for all possible combinations of pairs of k traits. However, simply considering m models for all SNPs across the genome results in a prohibitive increase of the already substantial multiple testing burden. Given m possible models, each of which predict a different vector γ, and uncertainty for the predicted elements in γ, a possible way forward is to average the model predictions. The models are weighted by the relative proportion of evidence for each model. Specifically, the weights can be based on the AICc[17] information criteria. The AICc for model m equals:

$$AICc_m = -\ln(LogLik_m) + 2k_m + \frac{2k_m(k_m+1)}{n-k_m-1}$$

For each AICc we compute the delta (Δ_m) to the best (i.e lowest) AICc value, and from these we compute the model weights (g) for the k models as:

$$g_m = \frac{\exp(-\frac{1}{2}\Delta_m)}{\sum_{m=1}^{z}\exp(-\frac{1}{2}\Delta_m)}$$

We predict the vector β using each of the models

$$\hat{\beta}_m = \gamma_m X_m$$

One can aggregate the prediction over all models as:

$$\beta_a = \frac{\hat{\beta}_m * g_m}{\sum_1^z g_m}$$

And we aggregate the uncertainty within and between models to obtain $var\ (\beta_a)$:

$$var(\beta_a) = \left[\sum_{m=1}^{z} g_m \sqrt{\mathrm{var}(\hat{\beta}_m) + (\hat{\beta}_m - \hat{\bar{\beta}})^2} \right]^2$$

The resulting vector $\beta_a$contains the model averaged effect sizes for the effect of a particular SNP on the phenotypes subjected to multivariate analysis. Note how the variance estimate contains a variance component which reflects within model variability ($\mathrm{var}(\hat{\beta}_m)$) which equals the square of the standard error, and a variance component between model variability ($(\hat{\beta}_m - \hat{\bar{\beta}})^2$) in estimate, which ensures no overfitting occurs.

Our procedure boosts power if the SNP effect is concordant between traits, while retaining strongly discordant SNP effects if the model favors these. Model averaging offers several avenues for extension. One can constrain the SNP effects across multiple SNPs based on biological knowledge of the relation between the SNPs and gene expression, or CpG methylation (analog to TWAS). Alternative it might be beneficial to average the AICc weights across regions of the genome. Model averaging can in principle accommodate *any model* for which the AICc information criterion can be expressed. These models should result in a vector of SNPS effects (β) and an asymptotic variance for the SNP effects. In the current application, models per SNP are estimated in R using the "metafor" package and models are averaged using the "AICcmodavg" package[18,19].

Polygenic Risk Prediction

To confirm the gain in power of our multivariate GWAMA results, we perform polygenic risk score prediction (PRS) in two independent samples; 1) the Netherlands Twin Register (NTR)[20,21] and Understanding Society (UKHLS)[22]. We predict the phenotypes in the well-being spectrum (life satisfaction, positive affect, neuroticism, and depressive symptoms). In NTR, LS and PA data are available in 9,143 and 6,836 genotyped participants. LS is measured longitudinally using the Satisfaction with Life Scale consisting of five items (e.g., "My life is going more or less as I wanted") with responses

given on a seven-point scale, resulting in a minimum score of five and a maximum score of 35[23]. PA is also measured longitudinally using four questions that were adapted from the Subjective Happiness Scale[24] (e.g., "On the whole, I am a happy person") with responses on a seven-point scale, resulting in a minimum score of four and a maximum score of 28. Neuroticism data are available for 8,527 genotyped participants. The Big Five personality traits (including neuroticism) were measured by using the NEO-FFI[25], a sixty-item personality questionnaire consisting of five subscales: neuroticism, extraversion, openness, agreeableness and conscientiousness. The responses were given on a five-point scale (0-4). Subscale scores are constructed for each time point by taking the sum across the twelve subscale-specific items (after recoding opposite-stated items), and are set to missing if ten or more items of the total scale are unanswered. When subjects have fewer than ten missing items, missing items are scored at two (which is neutral given the 0-4 scale). Depressive symptoms are obtained from the DSM-oriented Depression subscale of the age-appropriate survey from the ASEBA taxonomy[26] and are available in 7,898 participants. To measure depressive symptoms, fourteen questions are used (e.g., "Enjoys little ") and responses were given on a three-point scale ranging from zero ("not true") to two ("very true"). The DSM-oriented subscale is constructed for each time point by taking the sum across the fourteen subscale-specific items and is set to missing if more than twenty percent of the total survey items were unanswered. When less than twenty percent of items are missing for a participant, the missing items are replaced by the participant's mean score.

In UKHLS data are available in 9,944 participants genotyped on the Illumina Human Core Exome Beadchip. LS was measured longitudinally (waves 1-6). Participants were asked how satisfied they were "with life overall" with responses given on a seven-point scale, resulting in a minimum score of one and a maximum score of seven. PA is also measured longitudinally (waves 1 and 4 only) using The Warwick-Edinburgh Mental Well-being scale (WEMWBS). SWEMWBS is a shortened version of WEMWBS. This 7-item short version (see Tennant et al., 2007) is scored on a 5-point Likert scale, from "none of the time" to "all of the time", and summed to give a total score, ranging from 7 to 35 Neuroticism data are available for 8,198 genotyped participants from wave 3. The Big Five personality traits (including neuroticism) were measured using The Big Five Inventory (BFI), a 44-item personality questionnaire consisting of five subscales: neuroticism, extraversion, openness, agreeableness and conscientiousness. The responses were given on a seven-point scale (1-7). The neuroticism score combines three items on the neuroticism subdomain. Component scores were calculated as the average item response if no more than one of the three input responses was missing. Depressive symptoms (DS) were measured longitudinally (waves 1-6) and obtained from The General Health Questionnaire (GHQ) available in 9,203 participants. The 12 question GHQ was used containing questions relating to concentration, loss of sleep and general happiness. The 12 questions are scored on a four-point scale (1-4). Valid answers to the 12 questions of

the GHQ-12 were converted to a single scale by recoding 1 and 2 values on individual variables to 0, and 3 and 4 values to 1, and then summing, giving a scale running from 0 (the least distressed) to 12 (the most distressed).

The weights used for the polygenic scores are based on our two flavors of multivariate GWAMAs as well as the four univariate GWAMAs. Scores are based on the intersection of SNPs available in any of these GWAMAs. In NTR, SNPs were imputed to a common reference SNPs with MAF < 0.005, Hardy-Weinberg Equilibrium (HWE) with $p < 10^{-12}$ and call rate < 0.95 were removed. Individuals are excluded from the analyses if the genotyping call rate < 0.90, the inbreeding coefficient as computed in PLINK[48] (F) was < -0.075 or > 0.075), if the Affymetrix Contrast QC metric is< .40, if the Mendelian error rate > 5 standard deviations (SDs) from the mean, or if the gender and Identity-by-State (IBS) status does not agree with known relationship status and genotypic assessment. In UKHLS, SNPs were imputed to a common reference (1000 Genomes project March 2012 version 3). SNPs with MAF < 0.01, HWE $p < 10^{-4}$ and call rate < 0.98 were removed, individuals with A B and C were removed In NTR 1,224,793 SNPs passed QC and were used to construct polygenic scores and in UKHLS 955,441 SNPs passed QC and were used to construct polygenic scores. The phenotypes were regressed on sex, age as well as principal components which were included to correct for ancestry and the polygenic scores. Results can be found at **Supplemental Table 11.**

Summary-Based transcriptome wide (TWAS) and methylome wide (MWAS) association studies

We used the tool DIST[27] to impute the HapMap reference based results for the N-weighted GWAMA to the 1000Genomes Phase1 reference. We aggregate SNP effects informed by their common effect on expression level of gene or CpG methylation, as was proposed by Gusev et al.[7] We used the BIOS eQTL resource as eQTL reference set to build imputation models to predict gene expression using multiple eQTL SNPs[9]. Models are built per gene (gene models) by identifying independent eQTL SNPs based on stepwise conditional regression.[9] The z-score for each eQTL SNP is used in TWAS as a weight (q).The eQTLs used are available at http://genenetwork.nl/biosqtlbrowser/. Based on the gene models, N-weighted GWAMA summary statistics and LD based on the GONL reference[28], TWAS is performed. That is, for each gene- prediction-model containing eQTLs S_1- S_N with weights $q=q_1,q_2,...,q_n$, the corresponding GWAMA z-scores $z=z_1,z_2,...,z_n$ and LD is an n-by-n correlation matrix for eQTLs S_1- S_N, were used to construct a test statistic:

$$Z_{twas} = \frac{\sum_{i=1}^{n} q_i z_i}{\sqrt{q * LD * q}}$$

MWAS was performed following the same procedure to build imputation models to predict CpG site methylation of the DNA strand using multiple mQTL SNPs. The methylation site specific weights were obtained from the BIOS mQTL study[10].

Reverse Causation testing

TWAS or MWAS imply an effect of gene expression or CpG site on the trait.[7] To rule out reverse causation where variation in the well-being spectrum influence gene expression or CpG site methylation, we apply the following algorithm. We compute 22 PRS based on the N-Weighted GWAMA using NTR as holdout cohort. In each of these 22 PRS we omit one of the 22 chromosomes and predict the CpG site and gene transcript located on that particular chromosome. For instance, at chromosome 1, our MWAS analyses yielded out 8 CpG sites. Then with our PRS omitting chromosome 1, we try to predict these 8 CpG sites. If this prediction results in significant beta coefficients, reverse causation cannot be ruled out. For our prediction protocol, we standardize the PRS as well as the TWAS counts and CpG sites. To account for relatedness in NTR, generalized estimation equation (GEE) is used.[29] In TWAS prediction, we use sex, age at blood sampling, smoking, height, weight, plate and appropriate PCs as covariates. In MWAS prediction, we use sex, age at DNA blood sampling, array row number, smoking, eosinophils, -monocytes and neutrophils percentages as well as the appropriate PCs. To overcome the multiple testing burden, we use Bonferroni correction. For MWAS, significance was set at $p < 0.00026$ and for TWAS at $p < 0.001$. Results are displayed in **Supplementary Tables 14-15.**

GWAS Catalog lookup

We search the NHGI GWAS catalog[6] to determine which of our genome-wide significant GWAMA, MWAS, and TWAS markers have been previously reported (database searched on 18-02-2017). We apply two strategies; 1) for our GWAMA analyses, we identify genes that are within a 250kb distance from a genome wide significant SNPs For our TWAS and MWAS analyses the focus is specifically on markers that were not significant in GWAMA analyses. Therefore, we rule out any genes that we identify in TWAS or near MWAS hits that are within 500kb distance from a GWAMA lead SNP. Gene transcripts and genes near CpG sites which conform to the criteria outlined above are looked up in the GWAS catalog. Results can be found in **Supplementary Table 10, 16-17**.

Stratified LD score regression

To determine whether specific genomic regions are enriched for genetic effects on the well-being spectrum phenotypes, we used LD Score regression[15,16]. We are specifically interested in regions of the genome which are histone modified in a specific tissue. For example, regions of the genome which are histone modified in the prefrontal cortex, can be transcribed more frequently in prefrontal tissue. The enrichment of these genomic regions in their effect on well-being suggest the involvement of processes in the prefrontal cortex in the etiology of wellbeing.

LD Score regression is based on the relationship between the observed chi-square of a SNP and the degree of LD between a SNP and its neighbor. SNPs in strong LD are more likely to tag causal effects on complex traits and therefore have a higher expected chi-square. The procedure can be extended to *stratified* LD score regression where multiple LD scores are created, each of which captures the LD for a SNP with other SNPs of a specific category of interest, for example SNPs in a histone modified region of the genome.

We follow the exact procedure described by Finucane et al.[30] We estimated stratified LD Score regression for the "baseline" model which contains 53 categories. The model consists of a category containing all SNPs, 24 categories corresponding to main annotations of interest, 24 categories corresponding to 500-bp windows around the main annotations, and categories corresponding to 100-bp windows around ChIP-seq peaks (i.e. regions that are Sensitive to DNase1 or associated with histones bearing the modification marks H3K4me1, H3K4me3, H3K27ac or H3K9ac). In addition to the analysis of the baseline model, we performed analyses using cell type-specific annotations for the four histone marks, which correspond to specific chemical modifications to the histone protein which in turn package and orders the DNA molecule. Epigenetic modifications of histones, specifically histones bearing the marks H3K4me1, H3K4me3, H3K27ac or H3K9ac, all of which are associated with increased transcription of DNA into RNA. Each cell type-specific annotation corresponds to a histone mark in a specific cell obtained from distinct human tissue, for example H3K27ac in Fetal Brain cells, generating 220 combinations of histone modification by tissue. When generating estimates of enrichment for the 220 Histone mark by tissue annotations, we control for overlap with the functional categories in the full baseline model, but not for overlap with the 219 other cell type specific annotations. Then for our well-being phenotype, we ran LD Score regression on each of the 220 models (one for each histone by tissue combination) and ranked the histone by tissue annotations by P-value derived from the Z-values of the coefficient. Results are displayed in **Supplementary Table 18**.

Stratified LD score regression to detect negative selection

Using LD score regression, where the LD scores are based on evolutionary markers, we test whether negative selection is present in shaping variation in the well-being spectrum through recent evolution. We use the same procedure as extensively described by Gazal et al.[14] and report the proportion of heritability for each LD-related annotation of the baseline model subdivided in 5 quantiles ranging from 20% common SNPs with youngest allelic age to 20% of common SNPs with oldest allelic age. These proportions are computed based on a joint fit of the baseline-LD model, but measure the heritability explained by each quantile of each annotation while including the effect of other annotations. Results are displayed in **Supplementary Table 19.** The effect of nucleotide diversity, recombination rate, background selection, CpG-content predicted allelic age, and Levels of LD in Africa all are consistent with those found in Gazal et al., which Gazal et al show by ways of forward simulation to be consistent with the presence of negative selection[14].

References

1. Steptoe, A., Deaton, A. & Stone, A. a *Lancet* **385**, 640–648 (2014).

2. Diener, E., Oishi, S. & Lucas, R.E. *Am. Psychol.* **70**, 234–242 (2015).

3. Helliwell, J., Layard, R. & Sachs, J. (2015).

4. Okbay, A. et al. *Nat. Genet.* **48**, 624–633 (2016).

5. Wagenmakers, E.-J. & Farrell, S. *Psychon. Bull. Rev.* **11**, 192–196 (2004).

6. Welter, D. et al. *Nucleic Acids Res.* **42**, 1001–1006 (2014).

7. Gusev, A. et al. *Nat. Genet.* **48**, 245–52 (2016).

8. Gamazon, E.R. et al. *Nat. Genet.* **47**, 1091–1098 (2015).

9. Zhernakova, D. V et al. *Nat. Genet.* **49**, (2016).

10. Bonder, M.J. et al. *Nat. Genet.* **49**, 131–138 (2017).

11. Liu, X. et al. *bioRxiv* (2016).

12. Benjamini, Y. & Hochberg, Y. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).

13. Sekar, A. et al. *Nature* **530**, 177–83 (2016).

14. Gazal, S. et al. *bioRxiv* (2016).

15. Bulik-Sullivan, B.K. et al. *Nat. Genet.* **47**, 291–295 (2015).

16. Bulik-Sullivan, B. et al. *Nat. Genet.* **47**, 1236–1241 (2015).

17. Akaike, H. *Biometrika* **66**, 237–242 (1979).

18. Viechtbauer, W. *J. Stat. Softw.* **36**, 1–48 (2009).

19. Mazerolle, M.J. *R Packag. version* **1**, 1–15 (2011).

20. van Beijsterveldt, C.E.M. et al. *Twin Res. Hum. Genet.* **16**, 252–67 (2013).

21. Willemsen, G. et al. *Twin Res. Hum. Genet.* **16**, 271–81 (2013).

22. University of Essex. Institute for Social and Economic Research, N.S.R. and K.P. (2016).

23. Diener, E.D., Emmons, R.A., Larsen, R.J. & Griffin, S. *J. Pers. Assess.* **49**, 71–75 (1985).

24. Lyubomirsky, S. & Lepper, H.S. *Soc. Indic. Res.* **46**, 137–155 (1999).

25. Costa, P.T. & McCrae, R.R. (Psychological Assessment Resources Inc., Odessa: 1992).

26. Achenbach, T.M. & Rescorla, L. (2003).

27. Lee, D., Bigdeli, T.B., Riley, B.P., Fanous, A.H. & Bacanu, S.A. *Bioinformatics* **29**, 2925–2927 (2013).

28. Genome of the Netherlands Consortium *Nat. Genet.* **46**, 1–95 (2014).

29. Zorn, C.J. *Am. J. Pol. Sci.* 470–490 (2001).

30. Finucane, H.K. et al. *Nat. Genet.* **47**, 1228–1235 (2015).

Figure legends

Supplementary Figure 1. Quantile-quantile plots for the N-weighted and 4 average GWAMAs. Panel A is the N-weighted GWAMA. Panel B is model-average Neuroticism, Panel C is model-average Depressive symptoms, Panel D is model-average Positive Affect and Panel E is model-average Life Satisfaction.

Supplementary Figure 2. GWAS catalog lookup for the GWAMA, TWAS, and MWAS analyses. Traits found in the GWAS catalog were categorized in 18 categories. Each category is expressed as percentage of the number of genes previously reported in the GWAS catalog search. Panel A displays the GWAMA lookup results, panel B displays the MWAS lookup results, and panel C displays the TWAS lookup results.

Supplementary Figure 3. Local association in the MHC region. Panel A provides a local Manhattan plot for the MHC region with interposed on top the LD with a strong eQTL for the C4 gene linked to neuronal pruning in adolescence and schizophrenia by Sekar et al[13]. Panel B is a scatter plot for the $-\log_{10}(p)$ against the R2 with the C4 eQTL. Panel C provides a local Manhattan plot for the MHC region with interposed on top the LD with SNP rs13194504, the strongest MHC signal found for schizophrenia. Panel D is a scatter plot of the $-\log_{10}(p)$ against the R2 with rs13194504. Round symbols represent SNPs, square symbols represent gene transcripts and triangle symbols represent CpG sites.

Supplementary Figure 4. Locus specific signal on chromosome 20 and 13. On chromosome 13, TWAS implicates FARP1 and MWAS implicated cg02183491 in well-being, while on chromosome 20, TWAS implicates CSE1L while MWAS implicated cg17657707. Red-yellow color gradient reflects the LD with the lead SNP. Round symbols represent SNPs, square symbols represent gene transcripts and triangle symbols represent CpG sites.

Supplementary Figure 5. 220 Cell specific histone modified region enrichment. The bar plot is reflecting the FDR adjusted p-value for tissue specific histone modified regions of the genome, as estimated using partitioned LD-score regression.

Supplementary Figure 6. Proportion of heritability explained by quintiles of each LD-related annotation. The red line indicates the proportion of heritability when there is no enrichment (20% of SNPs explain 20% of heritability.

MEMBERS OF SOCIAL SCIENCE GENETIC ASSOCIATION CONSORTIUM

The following people, who are not listed as coauthors on this manuscript, contributed to the original GWAMA on Subjective Well-being, Depressive Symptoms, and Neuroticism, on which the present paper is based. The views presented in the present paper may not reflect the opinions of the individuals listed below.

We thank: Aysu Okbay, Jan-Emmanuel De Neve, Patrick Turley, Mark Alan Fontana, S Fleur W Meddens, Richard Karlsson Linnér, Cornelius A Rietveld, Jaime Derringer, Jacob Gratten, James J Lee, Jimmy Z Liu, Ronald de Vlaming, Tarunveer S Ahluwalia, Jadwiga Buchwald, Alana Cavadino, Alexis C Frazier-Wood, Nicholas A Furlotte, Victoria Garfield, Marie Henrike Geisel, Juan R Gonzalez, Saskia Haitjema, Robert Karlsson, Sander W van der Laan, Karl-Heinz Ladwig, Jari Lahti, Sven J van der Lee, Penelope A Lind, Tian Liu, Lindsay Matteson, Evelin Mihailov, Michael B Miller, Camelia C Minica, Ilja M Nolte, Dennis Mook-Kanamori, Peter J van der Most, Christopher Oldmeadow, Yong Qian, Olli Raitakari, Rajesh Rawal, Anu Realo, Rico Rueedi, Börge Schmidt, Albert V Smith, Evie Stergiakouli, Toshiko Tanaka, Kent Taylor, Gudmar Thorleifsson, Juho Wedenoja, Juergen Wellmann, Harm-Jan Westra, Sara M Willems, Wei Zhao, LifeLines Cohort Study, Najaf Amin, Andrew Bakshi, Sven Bergmann, Gyda Bjornsdottir, Patricia A Boyle, Samantha Cherney, Simon R Cox, Gail Davies, Oliver S P Davis, Jun Ding, Nese Direk, Peter Eibich, Rebecca T Emeny, Ghazaleh Fatemifar, Jessica D Faul, Luigi Ferrucci, Andreas J Forstner, Christian Gieger, Richa Gupta, Tamara B Harris, Juliette M Harris, Elizabeth G Holliday, Jouke-Jan Hottenga, Philip L De Jager, Marika A Kaakinen, Eero Kajantie, Ville Karhunen, Ivana Kolcic, Meena Kumari, Lenore J Launer, Lude Franke, Ruifang Li-Gao, David C Liewald, Marisa Koini, Anu Loukola, Pedro Marques-Vidal, Grant W Montgomery, Miriam A Mosing, Lavinia Paternoster, Alison Pattie, Katja E Petrovic, Laura Pulkki-Råback, Lydia Quaye, Katri Räikkönen, Igor Rudan, Rodney J Scott, Jennifer A Smith, Angelina R Sutin, Maciej Trzaskowski, Anna E Vinkhuyzen, Lei Yu, Delilah Zabaneh, John R Attia, David A Bennett, Klaus Berger, Lars Bertram, Dorret I Boomsma, Harold Snieder, Shun-Chiao Chang, Francesco Cucca, Ian J Deary, Cornelia M van Duijn, Johan G Eriksson, Ute Bültmann, Eco J C de Geus, Patrick J F Groenen, Vilmundur Gudnason, Torben Hansen, Catharine A Hartman, Claire M A Haworth, Caroline Hayward, Andrew C Heath, David A Hinds, Elina Hyppönen, William G Iacono, Marjo-Riitta Järvelin, Karl-Heinz Jöckel, Jaakko Kaprio, Sharon L R Kardia, Liisa Keltikangas-Järvinen, Peter Kraft, Laura D Kubzansky, Terho Lehtimäki, Patrik K E Magnusson, Nicholas G Martin, Matt McGue, Andres Metspalu, Melinda Mills, Renée de Mutsert, Albertine J Oldehinkel, Gerard Pasterkamp, Nancy L Pedersen, Robert Plomin, Ozren Polasek, Christine Power, Stephen S Rich, Frits R Rosendaal, Hester M den Ruijter, David Schlessinger, Helena Schmidt, Rauli Svento, Reinhold Schmidt, Behrooz Z Alizadeh, Thorkild I A Sørensen, Tim D Spector, John M Starr, Kari Stefansson, Andrew Steptoe, Antonio Terracciano, Unnur Thorsteinsdottir, A Roy Thurik, Nicholas J Timpson, Henning Tiemeier, André G Uitterlinden, Peter Vollenweider, Gert G Wagner, David R Weir, Jian Yang, Dalton C Conley, George Davey Smith, Albert Hofman, Magnus Johannesson, David I Laibson, Sarah E Medland, Michelle N Meyer, Joseph K Pickrell, Tõnu Esko, Robert F Krueger, Jonathan P Beauchamp, Philipp D Koellinger, Daniel J Benjamin & David Cesarini

MEMBERS OF THE BIOS CONSORTIUM

The following people, who are not listed as coauthors on this manuscript, contributed to the collection and analysis of the expression and methylation data on which the present paper is based. The views presented in the present paper may not reflect the opinions of the individuals listed below.

Management Team Bastiaan T. Heijmans (chair), Peter A.C. 't Hoen, Joyce van Meurs, Aaron Isaacs, Rick Jansen, Lude Franke

Cohort collection Dorret I. Boomsma, René Pool, Jenny van Dongen, Jouke J. Hottenga (Netherlands Twin Register); Marleen MJ van Greevenbroek, Coen D.A. Stehouwer, Carla J.H. van der Kallen, Casper G. Schalkwijk (Cohort study on Diabetes and Atherosclerosis Maastricht); Cisca Wijmenga, Lude Franke, Sasha Zhernakova, Ettje F. Tigchelaar (LifeLines Deep); P. Eline Slagboom, Marian Beekman, Joris Deelen, Diana van Heemst (Leiden Longevity Study); Jan H. Veldink, Leonard H. van den Berg (Prospective ALS Study Netherlands); Cornelia M. van Duijn, Bert A. Hofman, Aaron Isaacs, André G. Uitterlinden (Rotterdam Study).

Data Generation Joyce van Meurs (Chair), P. Mila Jhamai, Michael Verbiest, H. Eka D. Suchiman, Marijn Verkerk, Ruud van der Breggen, Jeroen van Rooij, Nico Lakenberg.

Data management and computational infrastructure Hailiang Mei (Chair), Maarten van Iterson, Michiel van Galen, Jan Bot, Dasha V. Zhernakova, Rick Jansen, Peter van 't Hof, Patrick Deelen, Irene Nooren, Peter A.C. 't Hoen, Bastiaan T. Heijmans, Matthijs Moed.

Data Analysis Group Lude Franke (Co-Chair), Martijn Vermaat, Dasha V. Zhernakova, René Luijk, Marc Jan Bonder, Maarten van Iterson, Patrick Deelen, Freerk van Dijk, Michiel van Galen, Wibowo Arindrarto, Szymon M. Kielbasa, Morris A. Swertz, Erik. W van Zwet, Rick Jansen, Peter-Bram 't Hoen (Co-Chair), Bastiaan T. Heijmans (Co-Chair).