## Supplementary Material
## Cohort Description

The following longitudinal twin cohorts were included:

**BrainSCALE:** Brain Structure and Cognition: an Adolescent Longitudinal Twin Study into genetic etiology — University Medical Center Utrecht and Netherlands Twin Register, Amsterdam.

*Subjects:* The sample comprised 59 monozygotic (MZ) and 68 dizygotic (DZ) twins (69 male/58 female from 75 families; 52 complete pairs) from the BrainSCALE cohort [van Soelen et al., 2012b] who participated at mean age 9.2 (0.1) years and again at mean age 12.2 (0.3) years. Mean scanning interval was 2.9 (0.2) years. Exclusion criteria consisted of having any metal material in the head, having a pacemaker, a known history of any major medical condition or psychiatric illness. Zygosity was determined based on DNA polymorphisms and confirmed by genome-wide single nucleotide polymorphism (SNP) data. All subjects and their parents signed informed assents and informed consents, respectively. The study was approved by the Central Committee on Research involving Human Subjects of the Netherlands (CCMO) and was in agreement with the Declaration of Helsinki (Edinburgh amendments).

*Imaging and processing:* MRI data were collected on a 1.5 Philips Achieva MR scanner at both occasions, using the same acquisition parameters [Peper et al., 2009]. Three-dimensional T1-weighted coronal spoiled-gradient echo scans of the whole head (256 × 256 matrix, TE = 4.6 ms, TR = 30 ms, flip angle = 30°, 160–180 contiguous slices; $1 \times 1 \times 1.2$ mm$^3$ voxels, Field-of-View = 256 mm/70%) were acquired for volumetric analysis. Global and subcortical volumes were obtained from the longitudinal FreeSurfer pipeline, version 5.3. Heritabilities of change in global structures in this cohort have been reported in [van Soelen et al., 2013]. Heritability of change in subcortical structures in this cohort has been reported in [Swagerman et al., 2014].

**QTIM: Queensland Twin IMaging Study** — Queensland Brain Institute, University of Queensland, and QIMR Berghofer Medical Research Institute, Brisbane, Australia

*Subjects:* The full sample of 22 complete twin pairs, and 5 unpaired cotwins, included at two targeted baseline ages (12 and 16 years). The younger group comprised 9 pairs (4 MZ, 5 DZ; 11 females/7 males from 9 families) with a mean age of 12.5 (0.2) years and a mean scanning interval of 3.5 (.4) years. The older group comprised 13 pairs (6 MZ, 7 DZ) and 5 unpaired twins (21 females/10 males from 18 families) with a mean age of 16.5 (0.4) years and a mean scanning interval of 3.5 (0.3) years. Twins were assessed for their suitability for imaging and screened by self-report for significant medical, psychiatric or neurological conditions, including head injuries. All were right-handed (based on 12 items from Annett's Handedness Questionnaire [Annett 1970]. Zygosity was determined objectively by typing nine independent DNA microsatellite polymorphisms using standard PCR methods and later confirmed by genome-wide SNP genotyping (Illumina 610K chip). The study was approved by the Human Research Ethics Committees of the

Queensland Institute of Medical Research, University of Queensland, and Uniting Health Care. Written informed consent was obtained from each participant and a parent or guardian.

*Imaging and processing:* MRI data were collected on a 4T Bruker Medspec scanner at both occasions. Structural T1-weighted 3D images were acquired (TR = 1500 ms, TE = 3.35 ms, TI = 700 ms, 240 mm FOV, 0.9 mm slice thickness, 0.9375 x 0.9375 x 0.9 mm$^3$ voxels). Images were acquired in 256 slices (coronal acquisition) at the first time point, and then, in 240 or 256 slices depending on acquisition orientation at the second time point (65% coronal (256 slices), 35% sagittal (240 slices)). Global and subcortical volumes were extracted through the longitudinal processing stream of FreeSurfer 5.3.

**Utwins1** — University Medical Center, Utrecht and Netherlands Twin Register, Amsterdam

*Subjects:* 160 twins (77 MZ/83 DZ; 95 male/55 female) were recruited from the twin-pair cohort at the University Medical Centre Utrecht and from the Netherlands Twin Registry, VU University Amsterdam, as described in [Brans et al., 2010]. Average age was 29.7 (7.8) years at baseline. Follow-up duration was 5.3 (0.7) years. DNA testing using polymorphic markers determined zygosity. Except for one twin pair, all twins and their siblings were reared together. All participants gave written informed consent. The study was carried out according to the directives of the Declaration of Helsinki (amendment of Edinburgh, 2000) and was approved by the medical ethics committee for research in humans (METC) of the University Medical Centre Utrecht, the Netherlands.

*Imaging and processing:* MRI data were collected on a 1.5 Philips Achieva NT scanner at both occasions, using the same acquisition parameters. T1-weighted three-dimensional fast field echo (3D-FFE) scans with 160–180 contiguous coronal slices (TE = 4.6 ms, TR = 30 ms, flip angle = 30°, 1 × 1 × 1.2 mm$^3$ voxels) were acquired for volumetric analysis. Global and subcortical volumes were obtained from the longitudinal FreeSurfer pipeline, version 5.3. Heritability of change in global and cortical structures in this cohort has been reported in [Brans et al., 2010; Brouwer et al., 2014].

**VETSA** — Vietnam Era Twin Study of Aging, University of California, San Diego and Boston University, U.S.

*Subjects:* The sample comprised 331 male twins (150 MZ/106 DZ paired twins/75 unpaired) who were randomly recruited from the Vietnam Era Twin Registry and had imaging data at two time points. Average age at baseline was 56.3 (2.6) years. Follow-up duration was 5.5 (0.5) years. All participants were in some branch of U.S. military service at some time between 1965 and 1975. Nearly 80% reported no combat experience; health and lifestyle characteristics are similar to American men in their age range [Kremen et al., 2006, 2013]. Zygosity was determined by 25

microsatellite markers.  The study was approved by the Institutional Review Boards at participating institutions, and all participants gave written informed consent.

*Imaging and processing:* At time 1, T1-weighted images were acquired on Siemens 1.5 T scanners in San Diego and Boston.  Acquisition parameters were:  TI = 1000 ms, TE=  3.31 ms, TR = 2730 ms, flip angle = 7°, slice thickness = 1.33 mm, voxel size 1.3 x 1.0 x 1.3 mm$^3$.  Global and subcortical volumes were obtained with FreeSurfer 3. 3.01b. Data were reviewed for quality, registered, and averaged to improve signal-to-noise.  Heritability of cortical and subcortical structures has been reported elsewhere [Kremen et al., 2010; Panizzon et al., 2009; Eyler et al., 2012].

At time 2, T1-weighted images obtained in San Diego were acquired on a GE 3T Discovery 750x scanner with the following acquisition parameters:  TE= 3.164  ms, TR = 8.084 ms, TI = 600 ms, flip angle = 8°, pixel bandwidth = 244.141, FOV = 24 cm, frequency = 256, phase = 192, number of slices = 172, slice thickness = 1.2 mm).  T1-weighted images obtained in Boston were acquired on a Siemens 3T Tim Trio scanner with the following acquisition parameters:  TE = 4.33 ms, TR = 2170 ms, TI = 1100 ms, flip angle = 7°, pixel bandwidth = 140, number of slices = 160, slice thickness = 1.2 mm.  Image processing was conducted at UCSD using FreeSurfer for global and subcortical structures.  Time 2 imaging and processing has been described elsewhere [McEvoy et al., 2015].


**OATS** — Older Australian Twins Study, Australia

*Subjects:* The study cohort was drawn from Wave 1 and Wave 2 of the longitudinal Older Australian Twins Study (OATS), a study of twins aged 65 years or older at baseline living in the three Eastern states of Australia (New South Wales, Victoria and Queensland) and registered with the Australian Twin Registry (ATR). Twins are followed-up approximately every two years. The zygosity of each twin pair had been confirmed previously by genotyping with high-density SNP arrays. The methodology of OATS has previously been described in detail [Sachdev et al., 2009]. The sample used for this study was comprised of 196 (130 female/66 male) non-demented Caucasian participants with scanning data available at Waves 1 and 2.  There were 57 monozygotic and 41 dizygotic (20 opposite sex) twin pairs with a mean age of 69.27 (4.67) years. The mean scanning interval was 2.31 (0.71) years. Informed written consent was obtained from all participants, and the study had the appropriate institutional ethics committees' approvals.

*Imaging and processing*: At Wave 1, MRI data were obtained on three 1.5 Tesla scanners and a 3 Tesla scanner owing to the multi-site nature of this study. Siemens Magnetom Avanto and Sonata scanners (Siemens Medical Solutions, Malvern PA, USA) with similar years of manufacture and upgrade were used in centres 2 (Victoria) (79 participants) and 3 (Queensland) (46 participants), respectively. In centre 1 (New South Wales), a 1.5 T Philips Gyroscan scanner (Philips Medical Systems, Best, Netherlands) (55 participants) was used initially, followed by a 3 Tesla Philips Achieva Quasar Dual scanner (16 participants). At Wave 2, the only change was that all centre 1

participants were scanned by the same 3 Tesla Philips Achieva Quasar Dual scanner that was used in Wave 1. The acquisition protocols and parameters were tested and matched between the centres through standardization of spatial resolution and slice thickness, using a 3D phantom to correct geometric distortions, and using five volunteers who were scanned on the four scanners [Sachdev et al., 2009]. Twin pairs were always scanned on the same scanner and were scanned either on the same day or within a few weeks of each other.

The 3D T1-weighted MRIs scans were used for computing the neuroimaging phenotypes for cerebral cortex and subcortical structures. 3D T1-weighted volumetric sequence was performed using a similar protocol for the 1.5 Tesla scanners in the three centres with in-plane resolution = 1 × 1 mm$^2$, slice thickness = 1.5 mm, slice number = 144, TR = 1530 ms, TE = 3.24 ms, TI = 780 ms, and flip angle = 8°. The acquisition parameters for the 3 Tesla Philips scanner in centre 1 were: TR/TE = 6.39/2.9 ms, in-plane resolution = 1 × 1 mm$^2$, slice thickness = 1 mm, slice number = 190, resulting isotropic voxels of 1 × 1 × 1 mm$^3$. Two 3D T1-weighted scans were acquired for each participant for an increased signal-to-noise ratio. MRI measures were obtained from the longitudinal FreeSurfer pipeline, version 5.3.

## Supplementary Material
## Statistical Modeling

For models 1,3 and 4, left and right volumes of global volumes (cortical grey matter, cortical white matter, cerebellum grey matter, cerebellum white matter, lateral ventricles) and subcortical structures (thalamus, caudate nucleus, putamen, pallidum, hippocampus, amygdala, nucleus accumbens) were added to increase statistical power. For models 1 and 3, change rates for all volumes were created by subtracting volume at baseline volume from volume at follow-up and dividing by the scanning interval.

Model 1) The univariate ACE model - figure S1A
In this model we included the change rate per year as the only phenotype.
We computed heritability of the change rate in this model as $a_1^2 / (a_1^2 + c_1^2 + e_1^2)$. The contributions of common environmental influences and unique environmental influences were computed similarly.

Model 2) Latent change ACE model - figure S1B
In this model we included left and right volume at baseline and follow-up for each subject. Two latent factors modeling level and change were included. This model was based on the model proposed by McArdle [McArdle et al., 2009], with the extension of modeling genetic and environmental effects on the phenotypes [Panizzon et al., 2015]. The latent factor model uses all available information from left and right volumes, allowing different path loadings for left and right and baseline and follow-up volume (factor loadings $f_1$ and $f_2$). Heritability of change in this model was computed as $(a_{LC}^2 + a_C^2)/ (a_{LC}^2 + a_C^2 + c_{LC}^2 + c_C^2 + e_{LC}^2 + e_C^2)$. The contributions of common environmental influences and unique environmental influences were computed similarly. This model, like model 3 below, allows for testing whether the genetic factors influencing volume and volume change overlap, by computing the genetic correlation $R_g$ between the two latent factors. If this correlation is different from 1 and -1, it can be concluded that there are genetic influences on change that are not shared with the genetic influences on baseline volume.

Model 3) Bivariate ACE model - figure S1C
In this model we included the baseline volume and change rate per year for each subject. Heritability of the change rate in this model was computed as $(a_{12}^2 + a_{22}^2)/ (a_{12}^2 + a_{22}^2 + c_{12}^2 + c_{22}^2 + e_{12}^2 + e_{22}^2)$. The contributions of common environmental influences and unique environmental influences were computed similarly. In this model, we can obtain an estimate for the genetic correlation $R_g = (a_{12} \, a_{11})/(\sqrt{(a_{11}^2(a_{12}^2 + a_{22}^2))})$.

Model 4) Bivariate ACE model - figure S1D
In this model we included baseline and follow-up volume for each subject. The factor $A_2$, $C_2$ and $E_2$ model residual variance at follow-up that is not captured by the factors $A_1$, $C_1$ and $E_1$ influencing baseline volume. Genetic variance of volumetric change was computed according to: $a_{11}^2 + a_{12}^2 +$

$a_{22}^2 - 2a_{11}a_{21}$ and divided by the total variance of change $(a_{11}^2 + c_{11}^2 + e_{11}^2) + (a_{12}^2 + a_{22}^2 + c_{12}^2 + c_{22}^2 + e_{12}^2 + e_{22}^2) - 2[(a_{11}a_{12} + c_{11}c_{12} + e_{11}e_{12})]$ to obtain an estimate of heritability. The contributions of common environmental influences and unique environmental influences were computed similarly.