Bioinformatics of genomic association mapping

Vaez Barzani, Ahmad

Chapter 2

Genetic and environmental influences on stability and change in baseline levels of C-reactive protein: a longitudinal twin study

Ahmad Vaez*, Arthur A. Sas*, Yalda Jamshidi, Ilja M. Nolte,

Tim D. Spector, Harold Snieder

*Equal contribution

Ready to be submitted

Abstract

Background: Cross-sectional twin and family studies report a moderate heritability of baseline levels of C-reactive protein (CRP) ranging from 0.10 to 0.65 for different age ranges. Here, we investigated the stability and relative impact of genetic and environmental factors underlying serum levels of CRP, using a longitudinal classical twin design.

Methods: A maximum of 6,201 female twins from the TwinsUK registry with up to three CRP measurements (i.e., visit 1 [V1], visit 2 [V2] and visit 3 [V3]) over a 10 year follow up period were included in this study. Structural equation modeling (SEM) was applied to dissect the observed phenotypic variance into its genetic and environmental components. To estimate the heritability of CRP as well as its genetic and environmental correlations across different time points, a trivariate model was used.

Results: Natural log (ln) CRP levels significantly increased from V1 to V2 ($p=4.4x10^{-25}$) and between V1 and V3 ($p=1.2x10^{-15}$), but not between V2 and V3. The median (IQR) follow-up time between V1 and V3 was 9.58 (8.00-10.46). Heritability estimates for CRP were around 50% and very constant over time (0.50-0.53). Adjustment for BMI slightly reduced heritability estimates (0.45–0.49). The genetic correlations between visits were significantly smaller than one, ranging from 0.66 to 0.85.

Conclusions: The present study provides evidence for stable heritability estimates of CRP with advancing age of around 50%. However, between-visit genetic correlations are significantly lower than 1 indicating emergence of new genetic effects on CRP levels with age.

Introduction

Ageing is known to be associated with a gradual elevation of inflammatory factors, attributed to the dysregulation of inflammatory pathways with ageing (1–4). Chronic low grade inflammation predisposes to many chronic, age-related diseases, such as those of the pulmonary and cardiovascular system (5–8). The genetic and environmental influences on baseline levels of inflammatory markers and the role of age as a moderator of these influences has been emphasized before (9).

An important, well established inflammatory marker is C-reactive protein (CRP). Its baseline levels are considered to reflect systemic inflammation. Considering the relationship of increased baseline CRP levels with a variety of disorders, including cancer (10), bipolar disorder (11), cardiovascular diseases (12–14), type 2 diabetes (15), and all-cause mortality (16), regulation of baseline CRP levels are of particular interest. In this context, baseline CRP levels have shown to be influenced by a variety of environmental and genetic factors. However, their relative importance and the exact extent to which these factors account for the total variance in CRP level remains unknown (17).

Heritability studies aim to estimate the relative influence of heritable and environmental factors on a trait (18). Twin and family studies in a wide variety of populations with different age ranges showed a moderate heritability of baseline CRP levels, with heritability estimates ranging from 0.10 to 0.65 (19–40).

Intra-individual CRP levels have been shown to be fairly stable over time. DeGoma et al. (41) analyzed serial CRP measures of 255 participants to evaluate the intraindividual variability of CRP over a median follow up period of 4.7 years. The multivariable-adjusted intraclass correlation coefficient (ICC) of CRP was estimated as 0.62. The intraindividual variability of CRP was also investigated by Wu et al. (42), using CRP levels of 56,218 Chinese adults over a two-year follow-up time. The ICC of CRP was reported as 0.55 for males and 0.60 for females. Interestingly, the stability of CRP gradually increased with age. However, twin and family studies mentioned above used single CRP measurement for their heritability calculation rather than longitudinal measurements. Limited by this cross-sectional design, heritability estimates for CRP as reported above only provide a snapshot at one particular point in time, potentially providing at least a partial explanation for the wide variety of heritability estimates reported in the literature (19–40).

To the best of our knowledge, no longitudinal twin studies on CRP levels have been conducted to date. The aim of this study was to evaluate the heritabilities

and the extent to which genetic and environmental influences contribute to the (in)stability of CRP over time in a large population of adult females using a classical twin design, including up to three CRP measurements over a ten year period.

Material and Methods

Subjects

The study was conducted in 6,201 females from the Twins UK registry. Details of the Twins UK registry have been published before (43). Zygosity was determined by questionnaire supplemented by DNA fingerprinting in cases with disputed or uncertain zygosity. CRP measurement follow-up was performed up to 3 times, giving 6,201 measurements in visit 1 (1,457 MZ-pairs, 1,584 DZ-pairs and 119 singletons), 2,251 measurements in visit 2 (452 MZ-pairs, 632 DZ-pairs and 83 singletons) and 528 measurements in visit 3 (139 MZ-pairs, 112 DZ-pairs and 26 singletons).

Sample analysis

High sensitive CRP was measured by latex-enhanced nephelometry on a Siemens (formally Behring) Prospec Nephelometer. The intra assay precision expressed as coefficient of variation (CV) of this method is around 3.5% CV at 1.5 mg/l and 3.1% at 12 mg/l and is expected to be <2% CV across the linear range of the assay.

Analytical approach

Natural log (ln) transformation was necessary for the CRP data in order to obtain a better approximation of the normal distribution. Secondly, the variables were adjusted for age. This is a common procedure in twin analyses because age can spuriously introduce a shared environmental effect if there is a significant correlation between the phenotype and age, because twins are always of the same age. Next, covariate analysis was performed, testing for: smoking, body mass index (BMI), current oral contraceptive (OC) use and current hormone replacement therapy (HRT). Since no significant contribution to CRP variance was found for smoking, OC and HRT, the covariate models used were: 1) Age and 2) Age + BMI. The residuals were used in the model fitting. Models were fitted to the raw data using normal theory maximum likelihood allowing inclusion of incomplete data, for example, when data were only available in one twin of a pair or in a limited number of visits.

Linear mixed model analysis was applied to determine whether lnCRP differed between visits while accounting for repeated measurements and twin relatedness by including the twin and family identification numbers as random effects in the model. Models with and without BMI as fixed effect were analyzed. The same approach was also used to test for differences in lnCRP levels between visits among those twins that returned for a second and/or a third visit. In order to evaluate potential selective drop out over the different visits, we used generalized estimating equations (GEE) to test the difference in age, BMI and lnCRP at baseline (i.e., visit1) between twins that returned for a second or third visit and those that did not return. GEE was also used to test for differences in baseline characteristics between MZ and DZ twins.

Model fitting

Structural equation modeling (SEM) was the primary method of analysis. SEM is based on the comparison of the variance-covariance matrices in MZ and DZ twin pairs and allows separation of the observed phenotypic variance into its genetic and environmental components: additive (A) or dominant (D) genetic components and common (C) or unique (E) environmental components, the latter also containing measurement error. The choice to start with either D or C in the full model depends on the relation between the MZ (rMZ) and DZ (rDZ) twin correlations. A D component is implied if 2xrDZ<rMZ whereas a C component is indicated if 2xrDZ>rMZ. Dividing each of these components by the total variance yields the different standardized components of variance. For example, the narrow sense heritability (h^2) can be defined as the proportion of the total variance attributable to additive genetic variation (18, 44).

For the longitudinal analysis, a trivariate SEM or path model (also known as a Cholesky decomposition, Figure 1) was used. With this model we can not only estimate the heritability of CRP at different times of measurement separately, but also the genetic (r_g) and environmental (r_e or r_c) correlations between different time points, giving an estimation of the (in)stability of genetic and environmental influences with advancing age. We can further test whether the genes influencing CRP are the same (i.e. $r_g=1$), partly the same (i.e. $0 < r_g <1$) or entirely different (i.e. $r_g=0$) at different times of measurement (and therefore different ages). If they are partly the same, this bivariate model allows quantification of the amount of overlap between genes influencing CRP at different ages by calculating the genetic correlation between the traits:

$r_g = COV_A$ (trait 1, trait 2)$/\sqrt{(V_A\text{trait1} * V_A\text{trait2})}$

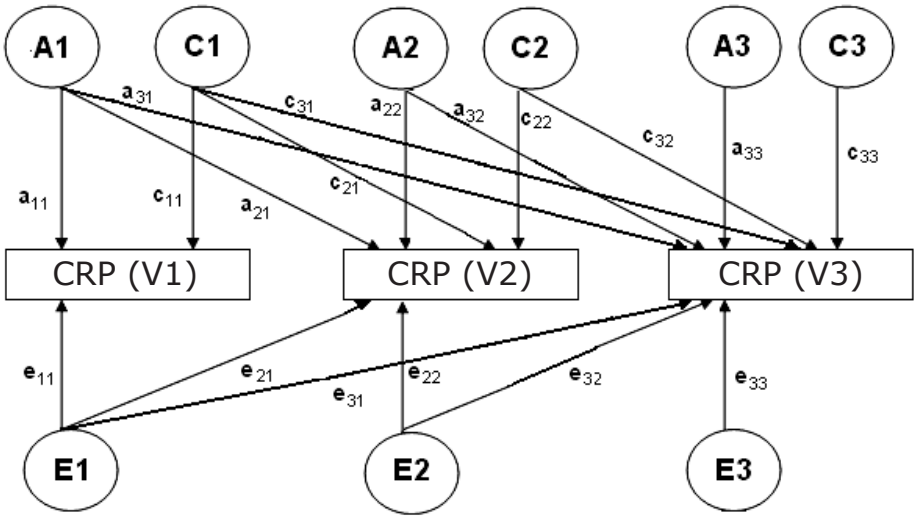

Figure 1 Path diagram for a bivariate model. For clarity, only one twin is depicted. A1, A2, A3 = Genetic variance components; C1, C2, C3 = common environmental variance components; E1, E2, E3 = unique environmental variance components; V1, V2, V3 = Visit 1, 2 and 3; a11 through a33 = genetic path coefficients (or factor loadings); c11 through c33 = common environmental path coefficients (or factor loadings); e11 through e33 = unique environmental path coefficients (or factor loadings).

Shared and unique environmental correlations can be calculated in a similar fashion (45, 46). In order to test for differences between twin 1 and twin 2, visits 1, 2 and 3 and differences between MZ and DZ twins, we tested whether the means could be set equal between different twins (twin 1 and twin 2), time points (visit 1, 2 and 3) and zygosity groups (MZ and DZ) without a decline in model fit. A significant decline indicates that means cannot be assumed to be equal.

Software

All data handling and preliminary analyses were done with STATA (version 10.1, Statacorp, TX, USA). Quantitative genetic modeling was carried out using Mx software package (47).

Results

Figure 2 shows the distributions of lnCRP at the three visits for all twins combined. lnCRP levels significantly increased from visit 1 (V1) to visit 2 (V2) ($p=4.4x10^{-25}$) and between V1 and visit 3 (V3) ($p=1.2x10^{-15}$), but not between V2 and V3 (p=0.69). Adjustment for BMI did not change these results. The median (IQR) follow-up time was 5.60 (2.87-7.56) years between V1 and V2, 6.17 (4.10-7.53) between V2 and V3 and 9.58 (8.00-10.46) between V1 and V3. lnCRP levels among the 2,251 "returners" significantly increased in the interval between V1 and V2 ($p=1.8x10^{-29}$) and between V1 and V3 (N= 528; $p=4.1x10^{-22}$), but not between V2 and V3 (N= 528; p=0.62) (Figure 3). Additionally adjusting lnCRP for BMI did not change these findings.

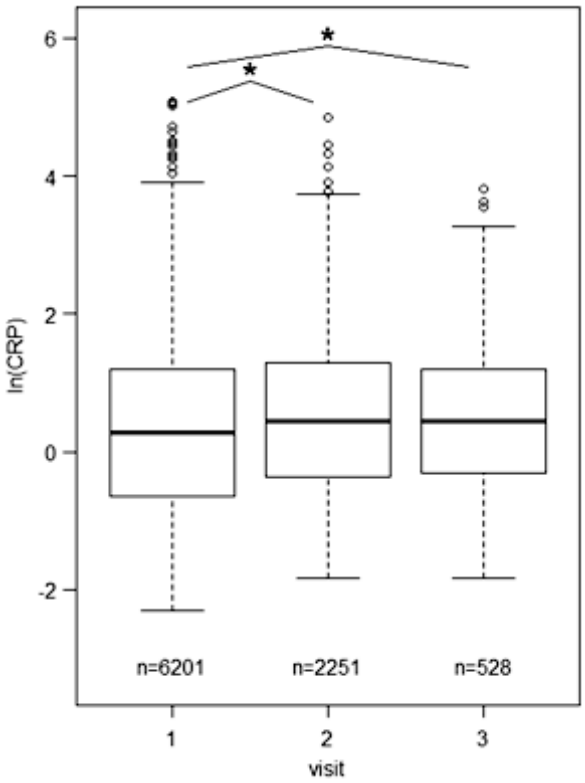

Figure 2 Distributions of lnCRP at the three visits. An asterisk means that there is a significant difference (p<0.05) in lnCRP between the respective visits.

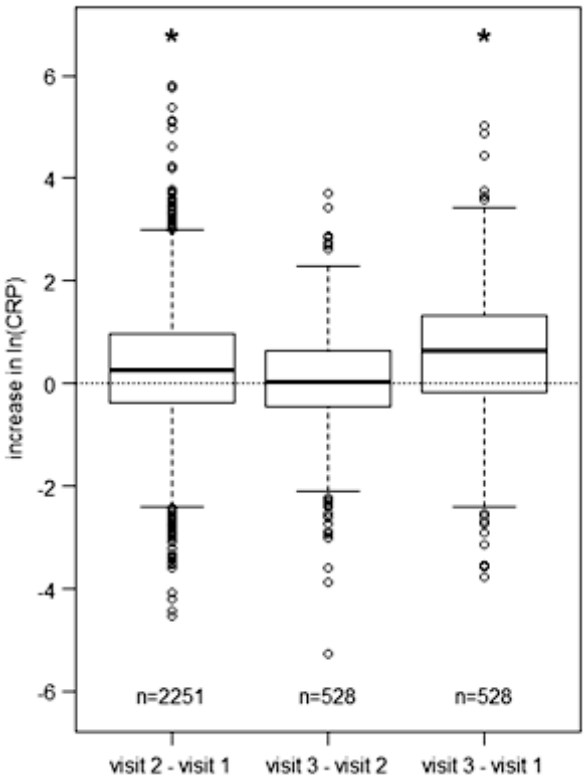

Figure 3 Distributions of paired differences in lnCRP between two visits. An asterisk means that the paired difference is significantly different from zero ($p<0.05$).

Baseline characteristics of MZ and DZ twins for the three visits are summarized in Table 1. Significant differences between MZ and DZ twins exist for age (Visit 2 and 3, $p<0.01$), BMI (Visit 2, $p<0.05$) and lnCRP levels (Visit 1 and 2, $p<0.05$). In our twin models we corrected for both age and BMI by including them as covariates.

Table 2 shows the intraclass twin correlations and results of the univariate SEM analysis of the two models for each of the three visits. For all three visits and both age adjusted and age plus BMI adjusted lnCRP values, MZ twin correlations were at least about twice as large as the DZ correlations clearly indicating the importance of genetic effects on lnCRP. In all models and visits, an AE-model was the best-fitting model. Heritabilities range from 0.50-0.53 (model 1) and 0.45-0.49 (model 2). The heritabilies remain relatively stable over time and their confidence intervals overlap for all visits and models.

Table 1 General characteristics of twins by zygosity and visit number.

		MZ		DZ	
	N	**Age (years)**	**N**	**Age (years)**	**P**
Visit 1	2,955	49.1±13.4	3,246	48.3±12.4	ns
Visit 2	934	57.9±10.1	1,317	56.0±10.3	<0.01
Visit 3	292	65.6±8.1	236	61.4±9.7	<0.01
	N	**BMI (kg/m²)**	**N**	**BMI (kg/m²)**	
Visit 1	2,955	25.4±4.6	3,246	25.6±4.7	ns
Visit 2	934	25.7±4.2	1,317	26.3±4.8	<0.05
Visit 3	292	26.1±4.2	236	26.3±4.4	ns
	N	**CRP (mg/L)**	**N**	**CRP (mg/L)**	
Visit 1	2,955	1.20 (0.48–3.15)	3,246	1.44 (0.58–3.47)	<0.05
Visit 2	934	1.45 (0.68–3.39)	1,317	1.61 (0.72–3.89)	<0.05
Visit 3	292	1.54 (0.73–3.18)	236	1.59 (0.73–3.80)	ns

Differences between MZ and DZ twins were tested using GEE with adjustment for age (for BMI) and age and BMI (for CRP). CRP was transformed by natural logarithm. Data are mean±SD for age and BMI, and median (IQR) for CRP. BMI indicates body mass index; MZ, monozygotic; DZ, dizygotic; ns, not significant; and CRP, C-reactive protein.

Table 3 shows the results of the longitudinal trivariate analysis (Cholesky decomposition). We first tested effects of twin, visit and zygosity on the means. For model 1, mean values of twin 1 and twin 2 could be set equal within MZ and within DZ twins, but could not be set equal across visits and zygosity groups. For model 2, in which CRP was adjusted for BMI, the means could additionally be set equal across all 3 visits, but remained different between MZ and DZ twins (see also Table 1). No evidence for a significant effect of genetic dominance was found as the AE model fitted best for both models. Heritability estimates for CRP were around 50% and very constant over time (0.50-0.53). Adjustment for BMI reduced heritabilities somewhat (0.45–0.49).

The genetic correlations (95% CI) between first and second (respectively second and third) follow-up visits were 0.82 (0.74-0.90) and 0.85 (0.71-0.97) (model

1), and 0.78 (0.70-0.87) and 0.77 (0.61-0.92) (model 2). These correlations are large, but significantly smaller than 1 based on the nonoverlapping 95% CIs indicating the emergence of new genetic effects with age. When comparing the first with the third visit, the genetic correlation dropped (0.66 for model 1 and 0.55 for model 2), indicating increasingly different genetic effects with age. Environmental correlations between first and second (respectively second and third) follow-up visits were much smaller than the genetic correlations with estimates of 0.16 and 0.27 (model 1), and 0.15 and 0.26 (model 2). When comparing the first with the third visit, the correlation remained the same (0.19).

Table 2 Intraclass correlations and parameter estimates of best fitting univariate models at the three visits.

Visit	Model	Intraclass correlations		Univariate Model Fitting		
		rMZ (95% CI)	rDZ (95% CI)	Best fitting model	A (95% CI)	E (95% CI)
1	**N, pairs**	**1457**	**1584**			
	1	0.54 (0.50-0.58)	0.24 (0.20-0.29)	AE	0.53 (0.50–0.56)	0.47 (0.44–0.50)
	2	0.48 (0.44-0.52)	0.20 (0.16-0.25)	AE	0.48 (0.44–0.51)	0.52 (0.49–0.56)
2	**N, pairs**	**452**	**632**			
	1	0.50 (0.43-0.57)	0.25 (0.18-0.33)	AE	0.50 (0.45–0.57)	0.50 (0.43–0.55)
	2	0.46 (0.38-0.53)	0.24 (0.17-0.31)	AE	0.45 (0.40–0.52)	0.55 (0.48–0.60)
3	**N, pairs**	**139**	**112**			
	1	0.54 (0.43-0.66)	0.13 (0.00-0.31)	AE	0.52 (0.39–0.62)	0.48 (0.38–0.61)
	2	0.51 (0.39-0.64)	0.15 (0.00-0.33)	AE	0.49 (0.36–0.59)	0.51 (0.41–0.64)

Model 1: adjusted for age
Model 2: adjusted for age and BMI

Table 3 Parameter estimates (95% CI) of best fitting trivariate models of CRP levels.

Model	Visit	1	2	3
1	1	**0.53 (0.50-0.56)**	0.16 (0.09-0.23)	0.19 (0.06–0.31)
	2	0.82 (0.74-0.90)	**0.50 (0.45-0.57)**	0.27 (0.12-0.40)
	3	0.66 (0.51-0.81)	0.85 (0.71-0.97)	**0.52 (0.39-0.62)**
2	1	**0.48 (0.44-0.51)**	0.15 (0.08-0.22)	0.19 (0.06-0.31)
	2	0.78 (0.70-0.87)	**0.45 (0.40-0.52)**	0.26 (0.12-0.39)
	3	0.55 (0.40-0.70)	0.77 (0.61-0.92)	**0.49 (0.36-0.59)**

The best fitting model for all analyses was the AE model. Genetic correlations [r_g (95% CI)] are given below the diagonal and environmental correlations [r_e (95% CI)] above the diagonal. Heritability (95% CI) estimates are given on the diagonal.
Model 1: adjusted for age
Model 2: adjusted for age and BMI

Discussion

The present study assessed the stability of genetic and environmental influences underlying baseline CRP levels, using a longitudinal classical twin design incorporating up to 3 follow-up measurements over a ten year period. We were able to demonstrate relative stable heritabilities with advancing age of around 50%, which are in the same range as previous studies (19–40). High genetic correlations of 0.66 to 0.85 between visits indicate that genes influencing CRP levels are mostly the same at different ages, whereas low environmental correlations of 0.16 to 0.27 show that environmental factors are largely different between visit. Genetic correlations were significantly different from 1, however, also indicating emergence of some new genetic effects on CRP with age.

The present study -to our knowledge- is the first to assess (and describe) the stability of genetic and environmental influences on baseline CRP levels in a longitudinal twin study. The longitudinal design with long follow up of up to 10 years and relatively large sample size provided more statistical power and methodological opportunities compared to previous smaller, cross-sectional studies.

We did not find evidence for genetic dominance however, in contrast to some previous cross-sectional twin studies that also had large sample sizes (38, 40).

A limitation of the present study however, is that our conclusions are not generalizable to men or subjects with diseases since only data on relatively healthy women was assessed. The benefit of this homogenous sample on the other hand, is that the results cannot be confounded by gender or disease since these covariates have previously been shown to have significant effects (48).

Even though we optimally made use of the available follow up measures of CRP over a ten year period, only subsamples of twins returned for the second and/or third visits. Those twins that returned for a second and/or third visit were not entirely representative of the whole sample as they were several years older, had lower BMI and lower levels of CRP at baseline (Supplementary Table 1). However, the Mx software package is capable of handling missing data by obtaining maximum likelihood estimates and takes advantage of including all available data rather than complete cases only (47). Given the similarity in the best fitting models and parameter estimates across the three visits, we believe it is unlikely this will have translated into major biases in our model fitting parameter estimates.

An interesting feature of our study, as mentioned above, is that we are the first to demonstrate relative stable heritabilities over time in a longitudinal design, even though the CRP levels itself do not seem stable (they significantly rise with advancing age). It has been hypothesized before that increased CRP levels with age may result from increases in "low grade, systemic, chronic inflammation" (due to atherosclerosis for example). One might have expected an increasingly important role for random (i.e., unique environmental) components reflecting reduced homeostatic control with age in this process. However, this was not supported by our findings.

The present study shows evidence of a substantial role for genetics in the regulation of baseline CRP levels. Heritabilities are stable with advancing age, and (more interestingly) the impact of environmental components remains relatively stable too (when comparing V1 – V3). Considering the genetic correlations were significantly smaller than 1 and reduced with follow up time, genes regulating CRP levels at younger ages must be partly different from those at more advanced ages. These results are in contrast with previous (cross-sectional) findings of other inflammatory markers, which indicate moderation of (changing) unique environmental factors with age in the regulation of IL-1β and TNF-α levels (9).

In conclusion, this study emphasizes the relatively stable role of genetics in regulation of baseline levels of CRP, emphasizing its potential as a biomarker of ageing over other (more biologically active) substances in the various immunological pathways. Still, the present study highlights the importance of a combination of both environmental factors and complex genetic pathways underlying the ageing process. Finally, even though the quantitative role of genetics in regulation of baseline CRP levels remained largely the same with age, the actual genes responsible for these effects were partly different at different ages. As such, future gene finding efforts need to take this into account, for example through investigating gene by age interaction effects.

References

1. Franceschi, C., Bonafè, M., Valensin, S., Olivieri, F., De Luca, M., Ottaviani, E. and De Benedictis, G. (2000) Inflamm-aging. An evolutionary perspective on immunosenescence. *Ann. N. Y. Acad. Sci.*, **908**, 244–254.

2. Franceschi, C., Capri, M., Monti, D., Giunta, S., Olivieri, F., Sevini, F., Panourgia, M.P., Invidia, L., Celani, L., Scurti, M., et al. (2007) Inflammaging and anti-inflammaging: a systemic perspective on aging and longevity emerged from studies in humans. *Mech. Ageing Dev.*, **128**, 92–105.

3. Vasto, S., Candore, G., Balistreri, C.R., Caruso, M., Colonna-Romano, G., Grimaldi, M.P., Listi, F., Nuzzo, D., Lio, D. and Caruso, C. (2007) Inflammatory networks in ageing, age-related diseases and longevity. *Mech. Ageing Dev.*, **128**, 83–91.

4. Pawelec, G., Effros, R.B., Caruso, C., Remarque, E., Barnett, Y. and Solana, R. (1999) T cells and aging (update february 1999). *Front. Biosci. J. Virtual Libr.*, **4**, D216–269.

5. Bruunsgaard, H., Pedersen, M. and Pedersen, B.K. (2001) Aging and proinflammatory cytokines. *Curr. Opin. Hematol.*, **8**, 131–136.

6. Schnabel, R.B., Lunetta, K.L., Larson, M.G., Dupuis, J., Lipinska, I., Rong, J., Chen, M.-H., Zhao, Z., Yamamoto, J.F., Meigs, J.B., et al. (2009) The relation of genetic and environmental factors to systemic inflammatory biomarker concentrations. *Circ. Cardiovasc. Genet.*, **2**, 229–237.

7. Danesh, J., Collins, R., Appleby, P. and Peto, R. (1998) Association of fibrinogen, C-reactive protein, albumin, or leukocyte count with coronary heart disease: meta-analyses of prospective studies. *JAMA*, **279**, 1477–1482.

8. Luc, G., Bard, J.-M., Juhan-Vague, I., Ferrieres, J., Evans, A., Amouyel, P., Arveiler, D., Fruchart, J.-C., Ducimetiere, P. and PRIME Study Group (2003)

C-reactive protein, interleukin-6, and fibrinogen as predictors of coronary heart disease: the PRIME Study. *Arterioscler. Thromb. Vasc. Biol.*, **23**, 1255–1261.

9. Sas, A.A., Jamshidi, Y., Zheng, D., Wu, T., Korf, J., Alizadeh, B.Z., Spector, T.D. and Snieder, H. (2012) The age-dependency of genetic and environmental influences on serum cytokine levels: A twin study. *Cytokine*, **60**, 108–113.

10. Allin, K.H., Bojesen, S.E. and Nordestgaard, B.G. (2009) Baseline C-Reactive Protein Is Associated With Incident Cancer and Survival in Patients With Cancer. *J. Clin. Oncol.*, **27**, 2217–2224.

11. De Berardis, D., Conti, C.M., Campanella, D., Carano, A., Scali, M., Valchera, A., Serroni, N., Pizzorno, A.M., D'Albenzio, A., Fulcheri, M., et al. (2008) Evaluation of C-reactive protein and total serum cholesterol in adult patients with bipolar disorder. *Int. J. Immunopathol. Pharmacol.*, **21**, 319–324.

12. Danesh, J., Wheeler, J.G., Hirschfield, G.M., Eda, S., Eiriksdottir, G., Rumley, A., Lowe, G.D.O., Pepys, M.B. and Gudnason, V. (2004) C-Reactive Protein and Other Circulating Markers of Inflammation in the Prediction of Coronary Heart Disease. *N. Engl. J. Med.*, **350**, 1387–1397.

13. Sesso, H.D., Buring, J.E., Rifai, N., Blake, G.J., Gaziano, J.M. and Ridker, P.M. (2003) C-reactive protein and the risk of developing hypertension. *JAMA J. Am. Med. Assoc.*, **290**, 2945–2951.

14. Kaptoge, S., Di Angelantonio, E., Lowe, G., Pepys, M.B., Thompson, S.G., Collins, R. and Danesh, J. (2010) C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: an individual participant meta-analysis. *Lancet*, **375**, 132–140.

15. Dehghan, A., Kardys, I., Maat, M.P.M. de, Uitterlinden, A.G., Sijbrands, E.J.G., Bootsma, A.H., Stijnen, T., Hofman, A., Schram, M.T. and Witteman, J.C.M. (2007) Genetic Variation, C-Reactive Protein Levels, and Incidence of Diabetes. *Diabetes*, **56**, 872–878.

16. Harris, T.B., Ferrucci, L., Tracy, R.P., Corti, M.C., Wacholder, S., Ettinger Jr, W.H., Heimovitz, H., Cohen, H.J. and Wallace, R. (1999) Associations of elevated Interleukin-6 and C-Reactive protein levels with mortality in the elderly. *Am. J. Med.*, **106**, 506–512.

17. Kathiresan, S., Larson, M.G., Vasan, R.S., Guo, C.-Y., Gona, P., Keaney, J.F., Wilson, P.W.F., Newton-Cheh, C., Musone, S.L., Camargo, A.L., et al. (2006) Contribution of Clinical Correlates and 13 C-Reactive Protein Gene Polymorphisms to Interindividual Variability in Serum C-Reactive Protein Level. *Circulation*, **113**, 1415–1423.

18. van Dongen, J., Slagboom, P.E., Draisma, H.H.M., Martin, N.G. and Boomsma, D.I. (2012) The continuing value of twin studies in the omics era. *Nat. Rev. Genet.*, **13**, 640–653.

19. Pankow, J.S., Folsom, A.R., Cushman, M., Borecki, I.B., Hopkins, P.N., Eckfeldt, J.H. and Tracy, R.P. (2001) Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. *Atherosclerosis*, **154**, 681–689.

20. Vickers, M.A., Green, F.R., Terry, C., Mayosi, B.M., Julier, C., Lathrop, M., Ratcliffe, P.J., Watkins, H.C. and Keavney, B. (2002) Genotype at a promoter polymorphism of the interleukin-6 gene is associated with baseline levels of plasma C-reactive protein. *Cardiovasc. Res.*, **53**, 1029–1034.

21. Retterstol, L., Eikvar, L. and Berg, K. (2003) A twin study of C-Reactive Protein compared to other risk factors for coronary heart disease. *Atherosclerosis*, **169**, 279–282.

22. Austin, M.A., Zhang, C., Humphries, S.E., Chandler, W.L., Talmud, P.J., Edwards, K.L., Leonetti, D.L., McNeely, M.J. and Fujimoto, W.Y. (2004) Heritability of C-reactive protein and association with apolipoprotein E genotypes in Japanese Americans. *Ann. Hum. Genet.*, **68**, 179–188.

23. Best, L.G., North, K.E., Tracy, R.P., Lee, E.T., Howard, B.V., Palmieri, V. and Maccluer, J.W. (2004) Genetic determination of acute phase reactant levels: the strong heart study. *Hum. Hered.*, **58**, 112–116.

24. de Maat, M.P.M., Bladbjerg, E.M., Hjelmborg, J. von B., Bathum, L., Jespersen, J. and Christensen, K. (2004) Genetic influence on inflammation variables in the elderly. *Arterioscler. Thromb. Vasc. Biol.*, **24**, 2168–2173.

25. MacGregor, A.J., Gallimore, J.R., Spector, T.D. and Pepys, M.B. (2004) Genetic effects on baseline values of C-reactive protein and serum amyloid a protein: a comparison of monozygotic and dizygotic twins. *Clin. Chem.*, **50**, 130–134.

26. Dupuis, J., Larson, M.G., Vasan, R.S., Massaro, J.M., Wilson, P.W.F., Lipinska, I., Corey, D., Vita, J.A., Keaney, J.F. and Benjamin, E.J. (2005) Genome scan of systemic biomarkers of vascular inflammation in the Framingham Heart Study: evidence for susceptibility loci on 1q. *Atherosclerosis*, **182**, 307–314.

27. Lange, L.A., Burdon, K., Langefeld, C.D., Liu, Y., Beck, S.R., Rich, S.S., Freedman, B.I., Brosnihan, K.B., Herrington, D.M., Wagenknecht, L.E., et al. (2006) Heritability and expression of C-reactive protein in type 2 diabetes in the Diabetes Heart Study. *Ann. Hum. Genet.*, **70**, 717–725.

28. Saunders, C.L. and Gulliford, M.C. (2006) Heritabilities and shared environmental effects were estimated from household clustering in national health survey data. *J. Clin. Epidemiol.*, **59**, 1191–1198.

29. Tang, W., Hong, Y., Province, M.A., Rich, S.S., Hopkins, P.N., Arnett, D.K., Pankow, J.S., Miller, M.B. and Eckfeldt, J.H. (2006) Familial clustering for features of the metabolic syndrome: the National Heart, Lung, and Blood Institute (NHLBI) Family Heart Study. *Diabetes Care*, **29**, 631–636.

30. Wörns, M.A., Victor, A., Galle, P.R. and Höhler, T. (2006) Genetic and environmental contributions to plasma C-reactive protein and interleukin-6 levels--a study in twins. *Genes Immun.*, **7**, 600–605.

31. Wessel, J., Moratorio, G., Rao, F., Mahata, M., Zhang, L., Greene, W., Rana, B.K., Kennedy, B.P., Khandrika, S., Huang, P., et al. (2007) C-reactive protein, an 'intermediate phenotype' for inflammation: human twin studies reveal heritability, association with blood pressure and the metabolic syndrome, and the influence of common polymorphism at catecholaminergic/beta-adrenergic pathway loci. *J. Hypertens.*, **25**, 329–343.

32. Fox, E.R., Benjamin, E.J., Sarpong, D.F., Rotimi, C.N., Wilson, J.G., Steffes, M.W., Chen, G., Adeyemo, A., Taylor, J.K., Samdarshi, T.E., et al. (2008) Epidemiology, heritability, and genetic linkage of C-reactive protein in African Americans (from the Jackson Heart Study). *Am. J. Cardiol.*, **102**, 835–841.

33. Su, S., Snieder, H., Miller, A.H., Ritchie, J., Bremner, J.D., Goldberg, J., Dai, J., Jones, L., Murrah, N.V., Zhao, J., et al. (2008) Genetic and environmental influences on systemic markers of inflammation in middle-aged male twins. *Atherosclerosis*, **200**, 213–220.

34. Schnabel, R.B., Lunetta, K.L., Larson, M.G., Dupuis, J., Lipinska, I., Rong, J., Chen, M.-H., Zhao, Z., Yamamoto, J.F., Meigs, J.B., et al. (2009) The relation of genetic and environmental factors to systemic inflammatory biomarker concentrations. *Circ. Cardiovasc. Genet.*, **2**, 229–237.

35. Su, S., Miller, A.H., Snieder, H., Bremner, J.D., Ritchie, J., Maisano, C., Jones, L., Murrah, N.V., Goldberg, J. and Vaccarino, V. (2009) Common Genetic Contributions to Depressive Symptoms and Inflammatory Markers in Middle-Aged Men: The Twins Heart Study. *Psychosom. Med.*, **71**, 152–158.

36. Wu, J., Pankow, J.S., Tracy, R.P., North, K.E., Myers, R.H., Feitosa, M.E., Province, M.A. and Borecki, I.B. (2009) A QTL on 12q influencing an inflammation marker and obesity in white women: the NHLBI Family Heart Study. *Obes. Silver Spring Md*, **17**, 525–531.

37. Jermendy, G., Horváth, T., Littvay, L., Steinbach, R., Jermendy, A.L., Tárnoki, A.D., Tárnoki, D.L., Métneki, J. and Osztovits, J. (2011) Effect of genetic and environmental influences on cardiometabolic risk factors: a twin study. *Cardiovasc. Diabetol.*, **10**, 96.

38. Rahman, I., Bennet, A.M., Pedersen, N.L., de Faire, U., Svensson, P. and Magnusson, P.K.E. (2009) Genetic Dominance Influences Blood Biomarker Levels in a Sample of 12,000 Swedish Elderly Twins. *Twin Res. Hum. Genet.*, **12**, 286–294.

39. Sas, A.A., Rijsdijk, F.V., Ormel, J., Snieder, H. and Riese, H. (2014) The relationship between neuroticism and inflammatory markers: a twin study. *Twin Res. Hum. Genet.*, **17**, 177–182.

40. Neijts, M., van Dongen, J., Kluft, C., Boomsma, D.I., Willemsen, G. and de Geus, E.J.C. (2013) Genetic architecture of the pro-inflammatory state in an extended twin-family design. *Twin Res. Hum. Genet.*, **16**, 931–940.

41. DeGoma, E.M., French, B., Dunbar, R.L., Allison, M.A., Mohler, E.R. and Budoff, M.J. (2012) Intraindividual variability of C-reactive protein: the Multi-Ethnic Study of Atherosclerosis. *Atherosclerosis*, **224**, 274–279.

42. Wu, S., Li, Y., Jin, C., Yang, P., Li, D., Li, H. and Shen, C. (2012) Intra-individual variability of high-sensitivity C-reactive protein in Chinese general population. *Int. J. Cardiol.*, **157**, 75–79.

43. Moayyeri, A., Hammond, C.J., Hart, D.J. and Spector, T.D. (2013) The UK Adult Twin Registry (TwinsUK Resource). *Twin Res. Hum. Genet.*, **16**, 144–149.

44. Snieder, H., Wang, X. and MacGregor, A.J. (2001) Twin Methodology. In *eLS*. John Wiley & Sons, Ltd.

45. Neale, M. and Cardon, L. (1992) Methodology for Genetic Studies of Twins and Families Springer Science & Business Media.

46. McCaffery, J.M., Snieder, H., Dong, Y. and de Geus, E. (2007) Genetics in psychosomatic medicine: research designs and statistical approaches. *Psychosom. Med.*, **69**, 206–216.

47. Neale, M.C., Boker, S.M., Xie, G. and Maes, H.H. (2003) Mx: Statistical Modeling. VCU Box 900126, Richmond, VA 23298: Department of Psychiatry.

48. Sutin, A.R., Terracciano, A., Deiana, B., Naitza, S., Ferrucci, L., Uda, M., Schlessinger, D. and Costa, P.T. (2010) High neuroticism and low conscientiousness are associated with interleukin-6. *Psychol. Med.*, **40**, 1485–1493.

Chapter 2 supplementary material

Supplementary Table 1 General characteristics at baseline (visit 1) for 'returning' subjects compared to 'non-returning' subjects.

	Zygosity	Non-returning	Returning V2 only	Returning V2+V3
Age (years)	MZ	47.3±14.0 (n=2.021)	51.6±11.5 (n=642) ***	56.4±7.7 (n=292) ***
	DZ	47.0±13.4 (n=1.929)	49.8±10.8 (n=1.081) ***	52.4±9.4 (n=236) ***
	Total	47.1±13.7 (n=3.950)	50.5±11.1 (n=1.723) ***	54.6±8.7 (n=528) ***
BMI (kg/m^2)	MZ	25.6±4.8 (n=2.021)	25.2±4.3 (n=642) *	24.7±4.0 (n=292) ***
	DZ	25.8±4.9 (n=1.929)	25.5±4.6 (n=1.081) *	24.8±3.4 (n=236) ***
	Total	25.7±4.8 (n=3.950)	25.4±4.5 (n=1.723) **	24.7±3.7 (n=528) ***
CRP (mg/L)	MZ	1.28 (0.56-3.37) (n=2.021)	1.16 (0.36-3.18) (n=642) **	0.80 (0.30-2.14) (n=292) ***
	DZ	1.53 (0.66-3.56) (n=1.929)	1.41 (0.56-3.52) (n=1.081)	1.09 (0.37-2.53) (n=236) ***
	Total	1.37 (0.60-3.43) (n=3.950)	1.30 (0.48-3.35) (n=1.723) *	0.90 (0.30-2.21) (n=528) ***

Differences between returning and non-returning (reference group) twins were tested using GEE with adjustment for age (for BMI) and age and BMI (for CRP). CRP was transformed by natural logarithm.
BMI indicates body mass index; CRP, C-reactive protein; MZ, monozygotic; DZ, dizygotic; and n, number of subjects.
Data are mean±SD for age and BMI and median (IQR) for CRP.
* p<0.05; ** p<0.01; *** p<0.001

Part 2 | How to perform a GWAS analysis