Predictive power of individual genetic and environmental factor scores

Martine A Thomis[1], Robert F Vlietinck[3], Hermine H Maes[5], Cameron J Blimkie[6], Marc van Leemputte[2], Albrecht L Claessens[1], Guy Marchal[4] and Gaston P Beunen[1]

[1]Center for Physical Development Research, and [2]Exercise Physiology and Biomechanics Laboratory, Faculty of Physical Education and Physiotherapy, Leuven
[3]Center for Human Genetics, and [4]Radiology Unit, Faculty of Medicine, Katholieke Universiteit Leuven, Belgium
[5]Virginia Institute for Psychiatric and Behavioral Genetics, Department of Human Genetics, Virginia Commonwealth University Virginia, Richmond, VA, USA
[6]Children's Sport and Exercise Science, The New Children's Hospital, Westmead, NSW, Australia

This study explores the use of an individual's genetic (IGFS) and environmental factor score (IEFS), constructed using genetic model fitting of a multivariate strength phenotype. Maximal isometric and dynamic strength measures, one maximal repetition load (1RM) and muscle cross-sectional area (MCSA) were measured in 25 monozygotic and 16 dizygotic twin pairs. The use of IGFS and IEFS in predicting the sensitivity to environmental stress was evaluated by the association of the scores with strength training gains after a 10-week high resistance strength training programme. Results show a high contribution of genetic factors to the covariation between maximal strength and muscle cross-sectional area (84–97%) at pre-training evaluation. Individual factor scores explained the largest part of the variation in 1RM and other strength measures at pre-training and post-training evaluation respectively. Genes that are switched on due to training stress (gene–environment interaction) could explain the decrease in explained variation over time. A negative correlation was found between IGFS and strength training gains (–0.24 to –0.51, $P < 0.05$); individuals with a high IGFS tend to gain less strength than individuals with low IGFS. Individual environmental factor scores have lower differential power. The predictive value of the IGFS has potential utility in identifying an individual's susceptibility to environmental stress in a variety of multifactorial characteristics, eg diseases and impairments, and for selection of sib pairs for QTL analyses. Twin Research (2000) 3, 99–108.

Keywords: genetic factor scores, twins, prediction, environmental stress, strength training, sib pair selection for QTLs

Introduction

Most common congenital malformations (eg clefting, spina bifida, pyloric stenosis, hip dislocation, club feet), adult diseases (eg ischemic heart disease, hypertension), and quantifiable biological traits (eg height, weight, blood pressure, intelligence) are multifactorial and are determined by both genetic and environmental factors. One of the major challenges in the field of human genetics is to identify individuals at high risk for a given disease or a favourable phenotype, and to predict an individual's outcome in an efficient prevention, training or educational programme. The importance of genetic and environmental effects in normal variation can be studied using data of genetically related subjects.[1,2]

Recent developments use the technologies of molecular biology to map gene loci explaining variation in quantitative traits (QTL).[3] Heritability studies estimate the importance of genetic factors at a population level. Statistical procedures are now available to estimate individual levels of genetic and environmental determination.[4] In this paper we use a path-analytic approach to construct individual genetic and environmental factor scores and to test whether these scores predict the susceptibility to environmental changes.

When multivariate observations are available from genetically related individuals, hypotheses can be tested about whether the same environmental and the same genetic factors have pleiotropic influences on phenotypically correlated measures.[5,6] The parameters of such a genetic factor analysis can be estimated, and individual genotypic and environmental factor scores (IGFS and IEFS) for each subject may be constructed by standard methods.[4] The major question in this study is:

Can individual genetic or environmental factor scores predict a subject's susceptibility to an environmental stress factor, specifically in this study an arm flexo training programme?

We hypothesise that, depending on the average heritability of the multivariate phenotype, the IGFS should predict the value of the individual phenotype at least at the pre-training level. The higher the heritability of the traits, the smaller the proportion of the variance explained by non-shared environmental factors, the greater the predictive power of the IGFS. Large variation in environmental scores, reflecting a high environmental determination, will lower the predictive power of the IGFS.

The predictive power of IGFS depends not only on the average heritable and environmental determination of the traits but also on the switching on of other genes in response to the environmental stress (genotype–environment interaction). The estimation of individual factor scores (IFS) in the pre-stress condition produces factor scores based on the genetic and environmental effects acting at this pre-stress level. If the stress activates new genes, then pre-stress IGFS will predict the post-stress phenotypes and changes in phenotypes less well.

The main purpose of this study is to test the predictive power of individual genetic and environmental factor scores in response to environmental stress in an empirical training study. Although feasible, it is ethically not desirable to do intervention studies in more relevant multifactorial diseases such as hypertension, obesity, cancer of behavioural disorders etc, therefore we decided to test the validity of the individual factor scores on muscular strength. Specific resistance training programmes are effective in increasing muscle strength and hypertrophy, and can, in a standardised manner, be used as an environmental stress factor in untrained subjects. We constructed individual genetic and environmental factor scores for isometric arm strength and muscle mass in 25 monozygotic and 16 dizygotic male, adult twin pairs. The environmental stress was a 10-week heavy-resistance training programme for the elbow flexors. Responses to this environmental stress factor were measured as absolute (post-training minus pre-training values) and relative increases (gain in strength expressed as a percentage of initial strength) in static and dynamic arm strength and muscle hypertrophy after the training programme. We hypothesise that subjects with higher individual genetic factor scores will be less responsive to the environmental stress (training) than subjects with smaller genetic factor scores, and will have correspondingly smaller strength gains following training.

Subjects and methods

Subjects

The sample for this study consisted of twin pairs from Flemish Brabant, Belgium. Male volunteer twins 17–30 years of age were included if both members of a twin pair had similar physical activity profiles and had not started, nor recently stopped, strength training during the preceding year. Forty-one twin pairs volunteered, their mean age was 22.4 ± 3.7 years. Subjects were fully informed of the measurement protocol before giving their written consent. The project was approved by the local medical ethics committee. Zygosity was determined by examination of the following genetic markers: ABO, Rhesus (D, C, C^w, c, E, e), MNSs and Duffy(a,b). The power to detect dizygotic (DZ) twins with this set of genetic markers was 91%. Differences in two genetic markers were used to establish dizygosity. The probability of monozygosity (MZ) of pairs with the same genetic markers was calculated.[7] All MZ pairs had a probability of monozygosity of at least 95%. Twenty-five pairs were classified as MZ and 16 as DZ.

Training protocol

Both members of the twin pair participated in a programme in which mainly the elbow flexors were trained. During 10 weeks, five sets of biceps curls were performed, 3 times a week on a training apparatus (Kettler Sport type 7408-150). Every week, the load of each set (with a precision of 0.5 kg) was adjusted to each subject's one repetition maximal value (1RM). The 1RM was defined as the maximal resistance that could be lifted a single time through the full range of motion. During each supervised training session, the first set was performed at 60% of 1RM with 14 repetitions (reps), the second set at 75% of 1RM with 12 reps, the third set at 80% of 1RM with 10 reps and 8 reps at 85% of 1RM for the fourth set. The fifth set at 65% of 1RM was performed until exhaustion.

Measurement protocol and variables

The estimation of IGFS and IEFS is based on a multivariate phenotype. Measurements that evaluated maximal isometric strength in the pre-training condition were chosen. These phenotypes consisted of the maximal static voluntary contraction at 140°, 110°, and 80° arm flexion (180° is the arm in full extension), and mean cross-sectional arm muscle area (cm^2). The evaluation of maximal static voluntary contraction was done after one week of adaptation to the training apparatus using low training

loads (50–70% 1RM) on a programmable dynamometer (Promett).[8,9] With this system, isometric, concentric, and eccentric contractions can be performed at different speeds and amplitudes imposed by the dynamometer. Subjects were asked to demonstrate maximal isometric strength and hold it for 3 seconds. The highest registered moment during this contraction was selected as the maximal isometric strength measurement expressed in Newton meter (Nm). Test–retest correlations ranged from 0.93 at the extreme angles to 0.97 at the middle angle (110°). The observer was able to evaluate each subject's maximal effort by visualised moment and electromyographic signals registered at M. biceps brachii, M. brachioradialis, M. brachialis, M. triceps brachii and M. triceps brachii. Computed tomography imaging scans were used to measure the mean cross-sectional arm muscle area.[10] Technical error of measurement for muscle area was $0.16\,cm^2$ with a reliability of 0.99. The mean muscle cross-sectional area (MCSA) of the four scans was used in the further analyses. The dependent phenotypes to evaluate the strength gain after training were the absolute and relative increases in 1RM, static strength at 110° flexion and strength at 140° flexion during maximal concentric contraction at a speed of 60°/s. Hypertrophic adaptations to the heavy-resistance strength programme were evaluated by absolute and relative increases in mean muscle cross-sectional area of the arm, measured by CT-imaging.

Genetic analyses

The causes of variation in muscle cross-sectional area and maximal static strength at the different elbow angles was first studied in a univariate way.[11] The significance of additive genetic variation, specific environmental factors and common environmental factors or dominance genetic variance was tested with model fitting.[2]

In order to construct the individual genetic and environmental factor scores, a common factor analytic model (Figure 1) was applied to the multivariate phenotype. The loading of common and variable-specific latent factors on the phenotypes was estimated using maximum likelihood estimation in Mx.[12] These loadings were then used in the estimation procedure for the IGFS and IEFS. This procedure is a regression method that minimises the differences between estimated and true factor scores.[13] It is the preferred method when the primary interest is the individual factor scores.[14] The Thurstone regression method for the estimation of factor scores[13] is preferred above the Bartlett estimator[15] because the correlations between true and estimated factor scores are higher and differences between simulated and predicted variances of the factor scores were somewhat smaller for the regression than for the Bartlett method.[4] More details on the model fitting procedure and the construction of IGFS and IEFS are given in Appendix A.
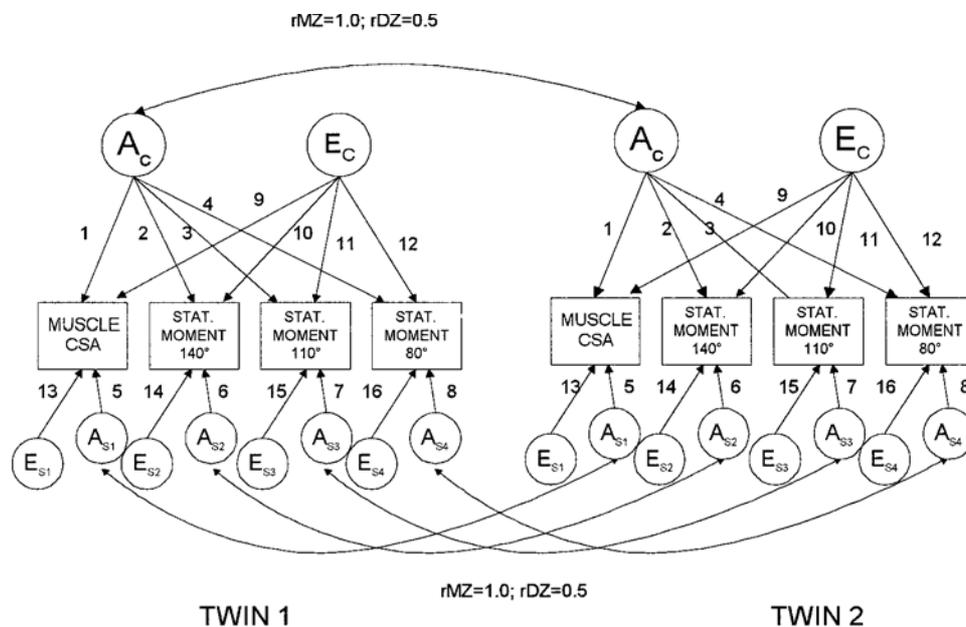

Figure 1 Path-diagram of the multivariate genetic analysis. Phenotypes are enclosed in squares, and latent factors are enclosed in circles. A_c and E_c are the additive genetic and non-shared environmental factors that are common to all phenotypes. A_{1-4} and E_{1-4} are additive genetic and non-shared environmental factors that are specific to each phenotype. The numbers at each causal uni-directional path indicate path coefficients. Double-headed arrows indicate correlations between latent factors (between additive genetic factors, 1 for MZ twins and 0.5 for DZ twins)

Twin Research

Analysis of the predictive value of individual factor scores

The predictive value of the individual genetic and environmental factor scores based on pre-training values was tested by the association of these scores with pre- and post-straining and absolute and relative training responses in 1RM, isometric strength at 110°, concentric strength at 140° flexion at 60°/sec and muscle cross-sectional area. These associations were tested by correlation coefficients. The distributions of all variables were tested for Gaussian normality using the Shapiro-Wilk test. We furthermore tested for birth order effects and differences in mean and variances between twin types with t tests and F tests, respectively. In all tests, the statistical significance level was chosen at $P < 0.05$.

Results

Univariate and multivariate genetic analysis

All data were normally distributed. Univariate genetic model fitting on pre-training phenotypes indicated that a model with additive genetic factors and unique environmental factors was the most parsimonious. For the mean muscle cross-sectional area (MCSA), there was evidence for a phenotypic interaction factor (one twin's larger MCSA going together with a smaller MCSA in his co-twin); however, this was due to a smaller total variance in MZ twins than in DZ twins. Univariate heritabilities were 0.92, 0.75, 0.78 and 0.66 for MCSA, and static moments at 140°, 110° and 80°, respectively.[11] The genetic contributions in this study correspond to other twin studies measuring maximal static or dynamic strength by arm pull, hand grip, pull-ups or combined strength scores (h^2 = 60–83%),[16–20] and to studies that estimate the heritability in regional arm musculature.[21–23]

The genetic common factor model (Figure 1, Table 1) fitted the data well (χ^2 = 61.04, df = 56, P = 0.30). The residual (co)variance matrix also showed small values. The common genetic factor (A_c) explained the largest part of the variation in each phenotype (64–76%), while phenotype-specific genetic and environmental factors were less important (0 to 20%). The common environmental factor only contributed 1% to 16% of the variation in each phenotype (Table 1A). However, leaving out this common factor worsened the fit of the model significantly. The high genetic correlations among the four phenotypes (> 0.85) indicated that the same genes influenced strength at different elbow angles as well as the muscle cross-sectional area. Non-shared environmental correlations were highest between the strength measures, but low between muscle mass and isometric strength (Table 1B).

Construction of genetic and environmental factor scores

The distribution of IGFS and IEFS was Gaussian. The mean value of all factor scores (Table 2) was not significantly different from zero, but the variation in both genetic and especially environmental factor scores was significantly smaller than the expected (1.0). Confidence intervals around the IGFS were smaller than those around the IEFS, and the confidence intervals around the factor scores tended to be larger in DZ than in MZ twins.

Table 1 (A) Proportion of explained variance in each phenotype by genetic and environmental factors. Legend and abbreviations as in Figure 1 (numbers in superscript give path coefficients as in Figure 1). (B) Bi-variate genetic and environmental correlations (above diagonal) and percentage of explained covariation explained by genetic and environmental factors (below diagonal)

A	Proportion of explained variance by genetic and environmental factors									
	Genetic variation					Environmental variation				
	A_c	A_1	A_2	A_3	A_4	E_c	E_1	E_2	E_3	E_4
MCSA[a]	0.64[(1)]	0.18[(5)]				0.01[(9)]	0.16[(13)]			
140°[b]	0.76[(2)]		0.00[(6)]			0.07[(10)]		0.17[(14)]		
110°[c]	0.72[(3)]			0.00[(7)]		0.16[(11)]			0.13[(15)]	
80°[d]	0.64[(4)]				0.05[(8)]	0.11[(12)]				0.20[(16)]

B	Genetic factors				Environmental factors			
	MCSA[a]	140°[b]	110°[c]	80°[d]	MCSA[a]	140°[b]	110°[c]	80°[d]
MCSA[a]	–	0.88	0.88	0.85	–	0.12	0.17	0.13
140°[b]	97	–	0.97	0.96	3	–	0.40	0.31
110°[c]	95	88	–	0.96	5	12	–	0.44
80°[d]	95	89	84	–	4	11	16	–

[a]MCSA: muscle cross-sectional area; [b]140°: maximal static moment at 140° of elbow flexion; [c]110°: maximal static moment at 110° of elbow flexion; [d]80°: maximal static moment at 80° of elbow flexion.

Table 2   Means, standard deviations and 95% confidence intervals (C.I.) of estimated factor scores

A	Total sample (n = 80)		MZ twins (n = 25)				DZ twins (n = 16)			
			Twin 1		Twin 2		Twin 1		Twin 2	
	IGFS	IEFS	IGFS	IEFS	IGFS	IEFS	IGFS	IEFS	IGFS	IEFS
Mean	0.006	0.015	0.000	0.000	0.000	0.000	0.027	0.086	0.006	−0.007
SD	0.89	0.54[b]	0.93	0.61[b]	0.93	0.55[b]	0.90	0.37[b]	0.82	0.58[a]
C.I.			± 0.60 (IGFS)		±1.5 (IEFS)		± 0.76 (IGFS)		± 1.72 (IEFS)	

[a]$P < 0.05$; [b]$P < 0.01$; significant difference from the expected standardised population parameters mean = 0 and SD = 1.

Strength training effects

The effect of strength training was first tested by analysis of variance for repeated measurements. The one repetition maximal strength increased significantly by 45.4% on average, maximal isometric strength by 23.3%, and maximal concentric strength by 24.9%. Hypertrophy of arm muscle area was smaller but significant (5.3%, $P < 0.01$). There was a significantly larger increase in MCSA in DZ twins than in MZ twins ($F = 4.7$, $P < 0.05$); however, no other zygosity interaction effects were found, indicating no difference in the muscular strength response to training between MZ and DZ twins. The variation in response to training between individuals was very large: the coefficient of variation varied between 34% and 142% for the absolute, and between 45% and 136% for relative changes in 1RM and maximal concentric strength respectively. Changes in 1RM scores, maximal isometric strength, and muscle hypertrophy were comparable to training effects in other strength training programmes.[24–28]

Predictability and differential power of individual genetic and environmental factor scores

Table 3 shows the correlations of IGFS and IEFS with the pre- and post-training and absolute and relative training responses in the four dependent phenotypes. IGFS was highly positively associated with both pre- and post-training phenotypes (0.67–0.86). Subjects with a high IGFS gained less absolute and relative strength as shown by the significant negative correlations with 1RM (−0.45—0.51), and relative change in isometric and concentric strength (−0.24). No significant correlation was found for the training effects on muscle mass (0.01, and −0.12). IEFS correlated moderately with pre-training isometric and concentric strength (0.44 and 0.30, respectively). No significant association was found between IEFS and training effects in the different phenotypes, except for a significant but low negative correlation with the increase in isometric strength (−0.24).

Figure 2 shows the power of how well a subject's baseline IFS predicts his observed baseline 1RM strength and his future strength gain after training. In this figure, both IFS and 1RM scores were categorised into quartile groups. Non-overlapping error bars indicate significant differences in number of individuals positioned in the phenotypical quartile groups by contrasting the individuals according to two IFS quartile groups (A, B: IGFS < P25 against IGFS ≥ P75; and C, D: IEFS < P25 against IEFS ≥ P75). Before training, subjects in either the lower or upper IGFS quartiles also had a high or low 1RM strength score (Figure 2A), while the extreme IEFS did not differentiate the subjects except somewhat in the middle (≥ P25–< P50) 1RM quartile (Figure 2C). For training responses, an inverse relationship was found. Subjects in the higher IGFS quartile gained the least strength (Figure 2B), whereas subjects in the lowest IGFS quartile were in the highest quartiles for their 1RM response. Contrasting the extremes of IEFS (Figure 2D) indicated that individuals with a low IEFS at baseline tended

Table 3   Correlations of IGFS and IEFS with the pre-, and post-training and training response phenotypes

		Phenotypes			
		1 RM[a]	isometric 110°	conc. 140° 60°/sec	MCSA[b]
IGFS with	pre-training	0.79[d]	0.86[d]	0.78[d]	0.84[d]
	post-training	0.72[d]	0.71[d]	0.67[d]	0.83[d]
	absolute change	−0.45[d]	−0.03	−0.15	0.01
	relative change	−0.51[d]	−0.24[c]	−0.24[c]	−0.12
IEFS with	pre-training	0.05	0.44[d]	0.30[c]	−0.11
	post-training	0.15	0.28[c]	0.22	−0.10
	absolute change	0.09	−0.14	−0.08	0.01
	relative change	0.07	−0.24[c]	−0.18	0.03

[a]1RM: one repetition maximal (kg); [b]MCSA: muscle cross-sectional area (cm^2); [c]$P < 0.05$; [d]$P < 0.001$.
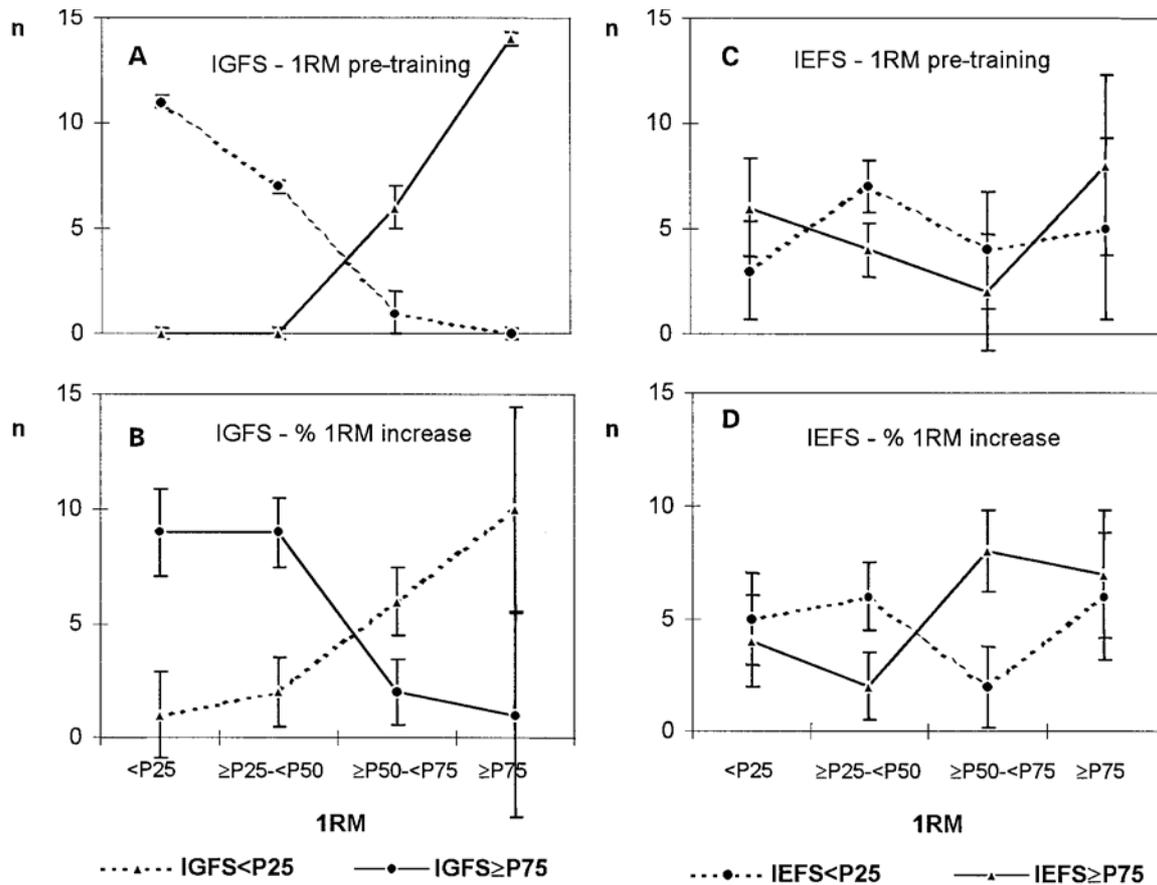
Figure 2 Differentiation in 1RM phenotype by IGFS and IEFS. Both IFS and 1RM scores are categorized in four quartile groups. Non-overlapping error bars indicate significant differences (*: $P < 0.05$) in number of individuals positioned in the phenotypical quartile group by contrasting two IFS percentile groups. A and B compare the extreme quartiles IGFS < P25 vs IGFS ≥ P75, and C and D the extreme quartiles IEFS < P25 vs IEFS ≥ P75. The Y-axis gives the number (n) of individuals in a given baseline IGFS or IEFS quartile who have 1RM strength scores in a certain quartile (X-axis)

to have a lower impact from the training than individuals with higher baseline IEFS scores.

Expressed as a relative risk, individuals having a low IGFS, had a 4.1 times increased chance of having a low baseline strength (RR = 4.1, CI_{95}: 1.9 to 8.79), while having a high IGFS increased the chance of having a high 1RM performance seven-fold (RR = 7.0, CI_{95}: 3.1 to 15.76). The relative risk in individuals with a low IEFS of a low strength performance was not significantly different from one, but having a high IEFS gave a two-fold increased chance of having a high baseline strength score (RR = 2.0, CI_{93}: 1.006 to 4.0). Individuals with a high IGFS had a six-fold significantly decreased chance (RR = 0.16, CI_{93}: 0.02 to 0.99) of a high training response; a similar relative risk was observed for individuals with a low IGFS to have a low training response. Relative risks to predict strength gains after training based on the IEFS scores were not significantly different from 1.

Discussion

To our knowledge, this is the first study that has investigated the use of individual genetic and environmental factor scores to predict the sensitivity of an individual to an environmental stress. Individual genetic and environmental factor scores were constructed from a multivariate common genetic factor model. This model gave a good explanation of the observed covariation (low χ^2); however, a more parsimonious model could be developed.[29]

Individual genotypic and environmental scores explained 60–74% of the variation in pre-training phenotypes (r^2 from Table 3). The proportion of explained variance decreased when post-training phenotypes and training effects were predicted. A first possible cause of this decrease could be environmental factors, other than the training programme, influencing the phenotypes during training. Although subjects were asked to maintain pre-

training physical activities during training, which was monitored by a 7-day recall questionnaire every week,[30] changes in physical activity, diet and other environmental factors could not be entirely controlled during the 10-week study period. Another important possibility, however, is that other, 'new' genetic factors, that do not contribute to the genetic variation in the pre-training strength, may be switched on during training (gene–environment interaction). The proportion of explained genetic variance by 'new' genetic factors in the post-training phenotype could not be explained by IGFS that are constructed from pre-training phenotypes. The importance of 'new' genetic factors affecting the post-training phenotype can be tested in a longitudinal model in which a specific genetic factor that only causes variation in the post-training phenotype but not in the pre-training phenotype is included. This specific gene–environment interaction was significant for the 1RM and maximal static moment at 110° flexion, and explained 19–21% of the variation in post-training strength but not for maximal torque in eccentric muscle work or concentric muscle work at lower velocities (30° and 60°/s).[31] Using a two-way analysis of variance method Thibault et al[32] found no evidence for a significant genotype–training interaction in peak torque output after 10 weeks of isokinetic knee flexion/extension training in five MZ twins. There was no evidence for specific genetic factors in our data to influence post-training muscle cross-sectional area.[31]

We are not aware of any method in the training literature that predicts individual responses to strength training. The prediction of individual training effects, based on the individual genetic factor scores in this study was more accurate for phenotypes with larger average training effects (1RM and relative changes). The observed negative relationship between pre-training genotype (IGFS) and strength increase, sometimes referred to as the law of initial values,[33] has also been reported in early strength training studies.[34] Stronger subjects gain less strength with a resistance-training programme than individuals with less strength.

The results indicated that IGFS not only classified subjects into high and low strength groups, but also, in spite of the large variability in training responses, predicted their low or high strength gain: of two individuals with similar strength at baseline, the one with the highest IGFS will gain less from training than the one with the lowest IGFS. The individual environmental factor scores, however, did not have the same power to categorise subjects, or to predict their future training gains. Environmental factors unique to individuals with a common effect on all measured phenotypes (E_c), such as previous training status or diet, only explained a small part of the

observed covariation. The environmental variation, unique for each phenotype (E_1–E_4), was more important. The predictive value of the IGFS and IEFS was, however, not significantly improved compared with the predictive values of the raw phenotypic scores (eg pre-MCSA scores predicting post-MCSA scores). Probably the high heritabilities of the phenotypes could explain these observations. In phenotypes with lower heritabilities like complex diseases or behavioural traits, the gain in predictive value of the IFS could be larger compared with the raw scores of the phenotypes.

This study demonstrated the feasibility of using multivariate genetic model fitting in the construction of individual genetic and environmental factor scores, and the use of model fitting in predicting the response to strength training. The same approach, however, has potential for applications in multifactorial diseases, such as hypertension or obesity. Quantifying the IGFS and IEFS of an individual with a high-risk phenotype (eg diastolic blood pressure above 90 mmHg or a systolic value above 140 mmHg) could indicate whether the cause of a high risk phenotype is mainly a genetic predisposition (high IGFS) or an environmental deviation (high IEFS).[4] Consequently therapeutic strategies may more efficiently concentrate on concrete actions on the regulatory mechanisms of hypertension in the case of a high genetic predisposition, or diminish the negative environmental stress factors, if subjects with hypertension express high environmental factor scores. Besides etiological classification, this approach might also predict the therapeutic outcome, and even guide the progress by monitoring the evolution of the IEFS. The weight matrix A (see Appendix 1, equation 3) can be calculated based on multivariate data from twins, or an extended twin and family design. This weight matrix is then multiplied by individual screening data to obtain IGFS and IEFS for each individual.

The results of this study should be interpreted in the context of the following limitations. Although the sample is one of the largest in an experimental strength-training design, genetic analyses usually require larger samples. The power of this study is sufficient to test for the significant contribution of genetic factors against a model with solely unique environmental contributions to the observed variation. The detection of a small proportion of additional familial environmental factors or genetic dominance would require much larger samples. In the multivariate case, however, power increases due to additional information, although the power to discriminate between different hypothesised models is still small. Further, results only apply to young adult men, who may not be representative of the general population.

IGFS and IEFS also improve the power of mapping Quantitative Trait Loci (QTL). Present strategies are based on identifying polymorphic marker alleles that are inherited identically by descent (IBD).[35] To increase the power of mapping QTLs, three strategies are suggested by Lander and Botstein:[36]

1) genotyping of sibs with extreme phenotypes;

2) multipoint interval mapping; and

3) reducing environmental variation and genetic variation not associated with the QTL.

Recent simulation studies have dealt with these issues[37–42] or with actual data.[3] Boomsma[41] reports a two-fold increase in power to detect linkage between a two-allele quantitative trait locus and a fully informative marker using the Haseman-Elston regression approach when using squared differences of individual genetic factor scores (based on a multivariate MZ and DZ twin analysis, including genetic variance that is not accounted for by the QTL and environmental variation) compared with squared differences of phenotypic scores between sibs. In a recent paper Boomsma and Dolan[42] performed power calculations (number of sib pairs to be studied to detect linkage) in which both sib pair selection and QTL analysis was based on an individual genetic factor score approach. The use of factor scores was shown to be universally more powerful than the use of just a multivariate or mean phenotypic data approach to detect linkage. The loss in power of using the same sample to both calculate the factor score regression matrix and to carry out the QTL analysis, outweighed the gain in power attributable to the use of factor scores.

In summary, this study explored the use of individual genetic and environmental factor scores in predicting an individual's susceptibility to environmental stress. The large proportion of explained variance in pre- and post-training strength as well as in strength increases by IGFS and IEFS leads to application of these scores in the development of individual strength training programmes, revalidation programmes and screening for elite athletes. Furthermore, the identification of genetic and environmental sources of deviation in individuals, has a major field of application in differentiating high-risk phenotypes in several multifactorial diseases. Also, selection of individuals based on discordant IGFS could increase the power to map quantitative trait loci.

Acknowledgements

References

1 Eaves L, Last K, Young P, Martin N. Model-fitting approaches to the analysis of human behavior. Heredity 1978; 41: 249–320.

2 Neale MC, Cardon LR. Methodology for Genetic Studies of Twins and Families. Kluwer Academic: Dordrecht, 1992.

3 Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC. Quantitative trait locus for reading disability on chromosome 6. Science 1994; 266: 276–279.

4 Boomsma DI, Molenaar PCM, Orlebeke JF. Estimation of individual genetic and environmental factor scores. Genet Epidemiol 1990; 7: 83–91.

5 Martin N, Eaves L. The genetical analysis of covariance structure. Heredity 1977; 38: 79–95.

6 McArdle JJ, Goldsmith HH. Alternative common-factor models for multivariate biometric analyses. Behav Genet 1990; 20: 569–608.

7 Meulepas, E, Vlietinck R, Van den Berghe H. The probability of dizygosity of phenotypically concordant twins. Am J Hum Genet 1988; 43: 817–826.

8 Van Leemputte M, Willems EJ. EMG quantification and its application to the analysis of human movements. Med Sports Sci 1987; 25: 177–194.

9 Vande Broek G, Van Leemputte M, Andries R, Willems EJ. Mechanical muscle properties after two types of plyometric training. In: Barabas A, Fabian GY (eds). Biometric in Sports XII. University Press: Sjókof, 1995, pp 98–101.

10 Thomis M, Claessens AL, Vlietinck R, Marchal G, Beunen G. Accuracy of anthropometric estimation of muscle-cross-sectional area of the arm in males. Am J Hum Biol 1997; 9: 73–86.

11 Thomis M, Beunen G, Van Leemputte M, Maes H, Blimkie CJ, Claessens AL, Marchal G, Willems E, Vlietinck R. Inheritance of static and dynamic arm strength and some of its determinants. Acta Physiol Scand 1998; 163: 59–72.

12 Neale MC. Mx: Statistical Modeling, 4th edn., Department of Psychiatry: Richmond, VA, 1997.

13 Thurstone LL. The Vectors of the Mind. University of Chicago Press: Chicago, 1935.

14 Saris WE, De Pijper M, Mulder J. Optimal procedures for estimation of factor scores. Sociol Meth Res 1978; 7: 85–106.

15 Bartlett MS. The statistical conception of mental factors. Br J Psychol 1937; 28: 97–104.

16 Weiss V. Die Heritabilitäten sportlicher Tests, berechnet aus den Leistungen zehnjährlichen Zwillingspaare. Leistungssport 1977; 9: 58–61.

17 Pirnay F, Crielaard JM. Influence de l'hérédité sur les performances physiques. Médic du Sport 1983; 57: 221–225.

18 Jones B, Klissouras V. Genetic variation in the strength–velocity relation of human muscle. In: Malina RM, Bouchard C (eds). Sport and Human Genetics. Human Kinetics: Champaign, IL, 1986, pp 155–163.

19 Reed T, Fabsitz R, Selby J, Carmelli D. Genetic influences and hand grip strength norms in the NLBI twin study male aged 59–69. Ann Hum Biol 1991; 18: 425–432.

20 Maes HHM, Beunen GP, Vlietinck RF, Neale MC, Thomis M, Vanden Eynde B, Lysens R, Simons J, Derom C, Derom R. Inheritance of physical fitness in 10-year-old twins and their parents. Med Sci Sports Exerc 1996; 28: 1479–1491.

21 Hoshi H, Ashizawa K, Kouchi M, Koyama C. On the intrapair similarity of Japanese monozygotic twins in some somatological traits. Okajimas Folia Anatomica Japonica 1982; 58: 675–686.

22 Maes HH, Beunen G, Vlietinck R. Heritability of health- and performance-related fitness. Data from the Leuven Longitudinal Twin Study. In: Duquet W, Day J (eds). Kinanthropometry IV. E & FN Spon: London, 1993, pp 140–149.

23 Loos R, Thomis M, Maes HH, Beunen G, Feys E, Derom C, Vlietinck R. Genetic analysis of muscularity in early adolescence: sex-specific regional changes of the genetic structure with age. J Appl Physiol 1997; 82: 1802–1810.

24 Davies J, Parker DF, Rutherford OM, Jones DA. Changes in strength and cross sectional area of the elbow flexors as a result of isometric strength training. Eur J Appl Physiol 1988; 57: 667–670.

25 Narci MV, Roi GS, Landoni L, Minetti AE, Cerretelli P. Changes in force, cross-sectional area and neural activation during strength training and detraining of the human quadriceps. Eur J Appl Physiol 1989; 59: 310–319.

26 Sale DG. Neural adaptation to strength training. In: Komi (ed.). Strength and Power in Sport: The Encyclopaedia of Sports Medicine. Blackwell Scientific Publications: Oxford, 1992, pp 249–265.

27 Kraemer WJ, Fleck SJ, Evans WJ. Strength and power training: physiological mechanisms of adaptation. In: Holloszy (ed.). Exercise and Sport Science Reviews, vol 24. Williams and Wilkins: Baltimore, 1996, pp 363–397.

28 Chambers RL, McDermott JC. Molecular basis of skeletal muscle regeneration. Can J Appl Physiol 1996; 21: 155–184.

29 Thomis M, Van Leemputte M, Maes H, Blimkie CJ, Claessens AL, Marchal G, Willems E, Vlietinck R, Beunen G. Multivariate genetic analysis of maximal isometric muscle force at different angles. J Appl Physiol 1997; 82: 959–967.

30 Blair SN, Haskell WL, Ho P et al. Assessment of habitual physical activity by a seven-day recall in a community survey and controlled experiments. Am J Epidemiol 1985; 122: 794–804.

31 Thomis M, Beunen G, Maes H, Blimkie CJ, Van Leemputte M, Claessens AL, Marchal G, Willems E, Vlietinck R. Strength training: importance of genetic factors. Med Sci Sports Exerc 1998; 30: 724–731.

32 Thibault MC, Simoneau J-A, Côté C, Boulay MR, Lagassé P, Marcotte M, Bouchard C. Inheritance of human muscle enzyme adaptation to isokinetic strength training. Hum Hered 1986; 36: 341–347.

33 Schutz RW. Analyzing change. In: Safrit MJ, Wood TM (eds). Measurement Concepts in Physical Education and Exercise Science. Human Kinetics: Champaign, IL, 1989; pp 207–228.

34 Hettinger TL, Muller EA. Muscular performance and training. In: Brown RC, Kenyon GS (eds). Classical Studies on Physical Activity. Prentice-Hall: New Jersey, 1968, pp 262–276.

35 Haseman JK, Elson RC. The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 1972; 2: 3–19.

36 Lander ES, Botstein D. Mapping mendalian factors underlying quantitative traits using RFLP linkage maps. Genet 1989; 121: 185–199.

37 Eaves L, Meyer J. Locating human quantitative trait loci: guidelines for the selection of sibling pairs for genotyping. Behav Genet 1994; 24: 443–455.

38 Risch N, Zhang H. Extreme discordant sib pairs for mapping quantitative trait loci in humans. Science 1995; 268: 1584–1589.

39 Risch N, Zhang H. Mapping quantitative trait loci with extreme discordant sib pairs: sampling considerations. Am J Hum Genet 1995; 58: 836–843.

40 Eaves L, Neale MC, Maes H. Multivariate multipoint linkage analysis of quantitative trait loci. Behav Genet 1996; 26: 519–525.

41 Boomsma DI. Using multivariate genetic modeling to detect pleiotropic quantitative trait loci. Behav Genet 1996; 26: 161–166.

42 Boomsma DI, Dolan CV. A comparison of power to detect a QTL in sib-pair data using multivariate phenotypes, mean phenotypes, and factor scores. Behav Genet 1998; 28: 329–340.

43 Sage AP, Melsa JP. Estimation Theory with Application to Communications and Control. McGraw-Hill: New York, 1971.

Appendix A

The multivariate phenotype (P) (see Figure 1) can be expressed by the following equation:

$$P_{(ij)} = h_c A_{c(j)} + e_c E_{c(j)} + h_s A_{(ij)} + e_s E_{(ij)} \quad (1),$$

where i represents the four different measured phenotypes (MCSA, STAT.MOMENT 140°, STAT. MOMENT 110°, STAT. MOMENT 80°) and j the examined individuals (j = 1,…82). The phenotype P of each individual is a function of his underlying genotype (A_c) that pleitropically influences all four phenotypes, and environmental factors (E_c) that are not shared in families, and therefore unique to each individual, but also influence all four phenotypes. The factor loadings of the measured phenotypes on the latent factor A_c and E_c are indicated by h_c and e_c. The residual variable-specific variance is also partitioned in genetic factors (A_1 and A_4) and unique environmental factors (E_1 to E_4), path coefficients are indicated by h_s and e_s.

The loading of each common and specific factor is estimated by maximum likelihood (ML) in Mx (Neale[12]). The following structural equation is solved:

$$S \Rightarrow \Sigma = \Lambda\ \Psi\ \Lambda' + \Theta \quad (2),$$

where

S = observed 2p × 2p (p = number of phenotypes = 4) covariance matrix of observations in twin 1 and twin 2 (expressed in deviations from the group mean);

Σ = predicted 2p × 2p covariance matrix of twin 1 and twin 2;

Λ = 2p × 2m matrix, where m = 2 is the number of common latent factors, containing the estimated loadings of the common latent factors on the four phenotypes of both twins (path coefficients 1–4 and 9–12 in Figure 1); the loadings are constrained to be equal for twin 1 and twin 2 and for MZ and DZ twins;

Ψ = 2m × 2m matrix of correlations between the latent factors; the correlation between A_c is 1 in MZ twins, 0.5 in DZ twins, the correlation

between the E_c factors is 0 for both MZ and DZ twins;

Θ = $2p \times 2p$ symmetric matrix of estimates of variable-specific unique environmental and genetic variances that are equated for both members of the twin pair and between MZ and DZ twins (path coefficients 5–8 and 13–16 in Figure 1). Within twin pairs, the unique genetic factors are correlated 1.0 in MZ twins and 0.5 in DZ twins.

The construction of individual genetic and environmental factor scores in the multivariate genetic analysis was performed by the Thurstone regression technique.[4] The following linear expression was used to obtain the weight matrix A, by minimising the sum of squares of the difference between estimated and true factor scores:

$$A = \Psi \Lambda' (\Lambda \Psi \Lambda' + \Theta)^{-1} \qquad (3),$$

where

Λ = matrix of loadings from multivariate phenotypes on common factor A_c and E_c;

Ψ = matrix of correlations between latent factors;

Θ = diagonal matrix of unique genetic and environmental variation.

This weight matrix was used to compute factor scores for each subject by multiplying the weight matrix by both the subject's multivariate phenotypic scores and his co-twin's phenotypic scores.

$$FSC = AP' \qquad (4),$$

where

Fsc = $[IGFS_1, IEFS_1, IGFS_2, IEFS_2]$, is the vector of individual factor scores, IGFS = individual genetic factor score, IEFS = individual environmental factor score, subscripts 1 and 2 indicate twin 1 and twin 2;

P = the measured multivariate phenotype of observations of twin 1 and twin 2, expressed in Z-scores $(n \times 2p)$;

A = $(2m \times 2p)$ weight matrix, derived from equation (3).

Two-sided, 95% confidence intervals of the factor scores were calculated as $IGFS \pm 1.96 \times SE_{IGFS}$ and $IEFS \pm 1.96 \times SE_{IEFS}$ for both MZ and DZ twins. SE_{IGFS} and SE_{IEFS} are the square root of the diagonal elements of the matrix from equation 5, which is a $2m \times 2m$ matrix of the sampling distribution of constructed factor scores. Matrix V is calculated based on the factor loadings in matrix Λ, the correlations between the common factors in matrix Ψ, and the estimated covariance matrix Σ of respectively MZ and DZ twins. This follows from standard Kalman filtering techniques:[43]

$$V = \Psi[\Psi^{-1} - \Lambda' \Sigma^{-1} \Lambda]\Psi \qquad (5).$$

The confidence intervals only depend on the factor loadings and the amount of variable-specific unique variance (in Σ) and will increase if the proportion of unique variance increases.