A Genome-Wide Association Study of Self-Rated Health

Miriam A. Mosing,[1,2] Karin J. H. Verweij,[1,2] Sarah E. Medland,[1,2] Jodie Painter,[1] Scott D. Gordon,[1] Andrew C. Heath,[3] Pamela A. Madden,[3] Grant W. Montgomery[1] and Nicholas G. Martin[1]

[1] Genetic Epidemiology, Molecular Epidemiology, and Queensland Statistical Genetics Laboratories, Queensland Institute of Medical Research, Brisbane, Queensland, Australia.
[2] School of Psychology, University of Queensland, Brisbane, Queensland, Australia.
[3] Department of Psychiatry, Washington University School of Medicine, St Louis, MO, United States of America

Self-rated health questions have been proven to be a highly reliable and valid measure of overall health as measured by other indicators in many population groups. It also has been shown to be a very good predictor of mortality, chronic or severe diseases, and the need for services, and is positively correlated with clinical assessments. Genetic factors have been estimated to account for 25–64% of the variance in the liability of self-rated health. The aim of the present study was to identify Single Nucleotide Polymorphisms (SNPs) underlying the heritability of self-rated health by conducting a genome-wide association analysis in a large sample of 6,706 Australian individuals aged 18–92. No genome wide significant SNPs associated with self-rated health could be identified, indicating that self-rated health may be influenced by a large number of SNPs with very small effect size. A very large sample will be needed to identify these SNPs.

Keywords: genome-wide association, genes, self-rated health, self-reported health, health

Self-rated health questions have been developed with the objective of quantifying an individual's perception of his or her overall health state. Even though these measures tend to be less sensitive to changes in specific disorders (Beaton & Schemitsch, 2003), it has been proven to be a highly reliable and valid measure of overall health as measured by means of other indicators in different population groups (Lundberg & Manderbacka, 1996). It has also been shown to be a very good predictor of mortality and the need for services (Leinonen et al., 2005), and is positively correlated with clinical assessments (Romeis et al., 2000). Furthermore, higher self-rated health has been associated with absence of chronic diseases, severe diseases, disabilities, functional limitations, and with higher physical activity, and better psychosocial wellbeing (Bryant et al., 2000; Idler, 1993; Leinonen et al., 2001; Rodin & McAvay, 1992). Typically, self-rated health is based on a single question asking the respondents to rate their current health status. Most individuals rate their health as moderate to good while few would rate their health as bad (Juerges, 2007).

Several twin studies have investigated the heritability of self-rated health (Christensen et al., 1999; Harris et al., 1992; Leinonen et al., 2005; Lichtenstein & Pedersen, 1995; Romeis et al., 2000; Silventoinen et al., 2007; Svedberg et al., 2001) estimating genetic factors to account for 25–64% of the variance in the liability of self-rated somatic health. A large longitudinal study of Finnish twins showed that the heritability of self-rated health was greatest at age 16, at 63% (95% CI: 0.56 – 0.67), declining steadily to age 25 with a heritability of 33% (CI: 0.25–0.41) (Silventoinen et al., 2006). The study revealed moderate correlations between the different health ratings at different life stages (r = 0.33–0.61), which were predominately due to genetic factors. The finding of decreasing heritability of self-rated health with age, however, is not confirmed by cross-sectional studies; for example, Mosing et al. (2009) found a heritability of 46% in an elderly twin sample (mean age = 61 ± 8.8). As self-rated health has been shown to be for a substantial part due to genetic factors, it would be interesting to explore the genetic variants underlying this trait. The aim of the present study was to identify SNPs underlying the heritability of self-rated health by conducting a genome-wide association (GWA) analysis on a large sample of Australian individuals who have previously rated their health status.

Methods

Participants and Measures

Self-rated health data were collected in four twin family studies conducted at the Queensland Institute of Medical Research (QIMR). The two earliest studies (Study 1 and Study 2) were conducted between 1993

and 1995. Study 3 was conducted between 1996 and 2000 and the most recent and largest study (Study 4), was collected between 2001 and 2005 (see Table 1a for more details). All four studies consisted of mailed-out questionnaires assessing health and lifestyle issues as well as demographic information and were approved by the QIMR Human Research Ethics Committee. Content and sampling methods of the four studies have been described in detail elsewhere (Bucholz et al., 1998; Hansell et al., 2008; Heath et al., 1997; Mosing et al., 2009a; 2009b). In Study 3 and 4 (more than 95% of the final sample), self-rated health was assessed with the following item: 'How would you describe your general physical health?', rated on a 4-point Likert scale, *Excellent* (1), *Good* (2), *Fair* (3), *Poor* (4). In the other two studies the self-rated health questions were worded slightly differently: 'In general, would you say that your physical health now is excellent, good, fair or poor?' (Study 1) and 'How would you describe your health at present?' — *Very good, Good, Fair, Poor, Very poor* (Study 2). As few individuals rated their health as poor or very poor and for consistency with the self-rated health questions of the other studies, which only had four instead of five categories, the categories poor and very poor of the self-rated health item in Study 2 were collapsed.

Self-rated health and genotype data were available for 6,706 individuals (3,710 females and 2,996 males) from 2,585 independent families aged between 18–92 years (Mean = 46; *SD* = 11). Of these, 1,403 (21%) individuals participated in more than one of the studies in which case the most recent rating was used. Test–retest Pearson correlations between the different self-rated health measures ranged between 0.48 and 0.65 and the correlation between the two identically worded items was not higher compared to the questions used in the other two studies. Table 1a shows the number of individuals derived from each study forming the final sample. Finally, in line with previous findings, most individuals rated their health as good while few reported poor health, resulting in a skewed distribution (Table 1b). Therefore, a square root transformation was applied to the final scores and the scores were treated as continuous. Previous behavior genetic analysis in a subsample of the present sample revealed heritability estimates of 46% with the remaining variance being due to non-shared environmental influences (Mosing et al., 2009b).

Genotyping, Quality Control, and Imputation Procedures

Over more than 20 years a wide range of phenotypic data and DNA samples have been collected as part of the different projects. The DNA samples were collected in accordance with standard protocols and were genotyped using the following Illumina SNP platforms: 317K, HumanCNV370-Quadv3, and Human610-Quad. Quality control (QC) procedures employed are discussed in full detail in Medland et al. (2009). Briefly, checks for ancestry outliers, Mendelian errors, Hardy Weinberg Equilibrium, and Minor Allele Frequency (MAF) were conducted separately for each of the projects and then again for the combined dataset. The final dataset consisted of 269,840 SNPs and was imputed by MACH (Abecasis, unpublished) using the data from the European HapMap 1+2, Release 22 Build 36. SNPs with an imputation quality score (r^2) greater than 0.3 were retained resulting in a total of 2,380,486 imputed SNPs. Finally, if only one individual from a monozygotic twin pair had been genotyped, the non-genotyped co-twin was assigned that genotype as well.

Statistical Analyses

The best guess genotype at each SNP was tested for association with self-rated health using the family-based association test in Merlin (Chen & Abecasis, 2007) accounting for family relationships. The additive genetic effect was computed by modeling the genotypic mean of the heterozygote (Aa) as the average of the two homozygotes (AA, aa). The generally accepted genome-wide significance level for the association between SNP and phenotype at $\alpha = 0.05$ is $7.2*10^{-8}$ or smaller, correcting for the total number of independent tests (Dudbridge & Gusnanto, 2008), and was also applied in the present study.

Table 1a

Raw Number and Percentage (in brackets) of Participants Drawn from Each Study

	Study 1[a]	Study 2[b]	Study 3[c]	Study 4[d]	Total
N (%)	43 (1%)	237 (4%)	158 (2%)	6268 (93%)	6706 (100%)
Age range	30–72	50–92	24–36	18–85	18–92

Note: [a](Heath, et al., 1997); [b](Bucholz, et al., 1998); [c](Mosing, Gordon, et al., 2009); [d](Hansell, et al., 2008).
The most recent score of participants who took part in more than one study was used.

Table 1b

Score Distribution of the Final Sample for Self-Rated Health

	Excellent	Good	Fair	Poor	Total
N (%)	1837 (27%)	3640 (54%)	1053 (16%)	176 (3%)	6706 (100%)

Additionally, a gene-based test (VEGAS), feasible for use with GWA data with related individuals (Liu et al., in press), was conducted to test whether there are any genes that harbor an excess of associated variants. Details of this procedure are summarized elsewhere (Liu et al., in press; Verweij et al., in press). In brief, this test explores association on a per-gene basis taking the p values of all SNPs within 50 kb of each gene, as well as linkage disequilibrium (LD) and number of SNPs per gene into account. A p value below $\alpha = 2.8*10^{-6}$ was considered to be significant as the gene-based association test included 17,585 genes (0.05/17,585).

Finally, power calculations were conducted by performing association tests on simulated datasets (based on our sample) in Merlin. The simulated datasets maintain the features of the original data in terms of marker informativeness, allele frequency, spacing, missing data patterns, and trait distribution despite replacing the phenotypic values and the genotypes for a randomly selected SNP with a minor allele frequency of 0.25. One thousand simulated data sets were generated on which association analyses were subsequently performed. Detailed information on the simulation procedure can be found on http://www.sph.umich.edu/csg/abecasis/Merlin/reference/simulation.html. The empirical power estimate is given by the proportion of genome-wide significant association tests detected in the 1000 association analyses. The present sample provides 99% and 50% power to detect SNPs explaining 1% and 0.5% of variance in self-rated health, respectively.

Results

We tested 2,380,486 SNPs for association with self-rated health correcting for age and sex. The Quantile-Quantile (QQ) plot (Figure 1) shows the association between the observed versus the expected (under the null-hypothesis of no association) p values of the autosomal associations.

Results of the association analysis (shown in Figure 2) indicate that there are no genome-wide significant ($\alpha = 7.2*10^{-8}$) association signals, with the smallest p-value of $2.3*10^{-7}$ obtained for a SNP (rs17043947) on chromosome 2p24.1. However, though not significant, we found two promising regions on chromosome two (2p24.1 and 2q14.3) with the top hits having p values of $2.3*10^{-7}$ and $7.6*10^{-7}$, respectively. Table 2 shows SNPs in the top 50 smallest p values for self-rated health. Redundant SNPs in high LD ($r^2 > .70$) with a more significant SNP were excluded.

The gene-based test did not reveal significant results ($\alpha = 2.8*10^{-6}$), with the smallest p-value being $9.0*10^{-5}$. Table 3 shows the five genes with the smallest p values.

Discussion

The present study is the first to perform a genome-wide association analysis on self-rated health. Despite the high power (99%) to detect SNPs accounting for 1% of the variance in self-rated health, no genome-wide significant SNPs were identified. However, though not significant, we found two promising regions on chromosome two. Also, some
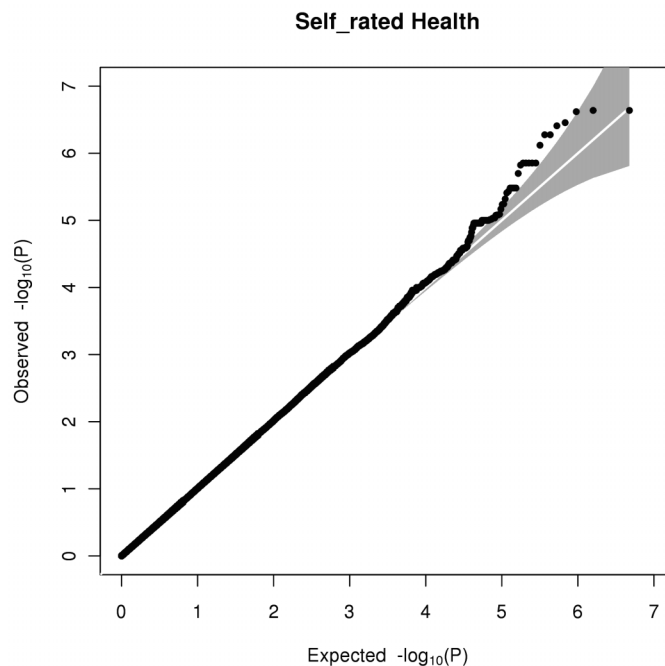
Self_rated Health

Figure 1

The Q-Q plot shows the association between the observed and expected $-\log_{10}$ (P-value) of the autosomal association between SNPs and self-rated health with the grey area representing the 95% confidence interval.
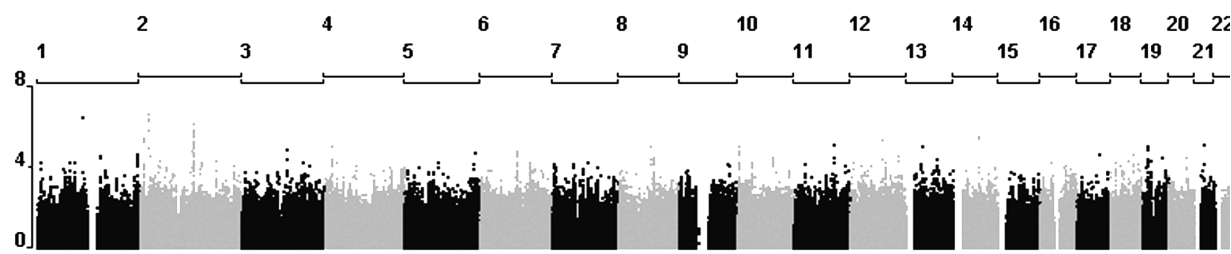
Figure 2

Manhattan plot showing the results of the genome-wide association analyses for self-rated health with the x-axis showing chromosome numbers and the y-axis the p value (–log10) of the association signals.

Table 2

Genetic Markers Showing Strongest Association With Self-Rated Health (Independent Markers Within Top 50 SNPs)

Chr	SNP	Base pair location	p value	SNPs in LD	Minor allele	MAF	Effect size	Closest gene	location
2	rs17043947	22736987	$2.3 * 10^{-7}$	8	T	.04	−0.07		
1	rs958798	110770223	$3.5 * 10^{-7}$		T	.17	0.03	KCNC4	Intronic
2	rs17043944	22724593	$5.3 * 10^{-7}$	7	T	.02	−0.10		
2	rs7567389	127982645	$1.5 * 10^{-6}$	2	A	.30	−0.03	CYP27C1	Upstream
2	rs6759460	127897011	$3.3 * 10^{-6}$	4	C	.24	0.03	AC110926	Intronic
14	rs6573416	62518348	$3.3 * 10^{-6}$	1	G	.13	0.03	SYT16	Intronic
2	rs2357266	9924387	$3.7 * 10^{-6}$	1	G	.35	−0.02		Intergenic
12	rs300489	78485994	$4.8 * 10^{-6}$		A	.23	0.03	NAV3	Intronic
21	rs7279441	24198815	$8.3 * 10^{-6}$	1	G	.14	0.03		Intergenic
4	rs17478107	16002288	$9.2 * 10^{-6}$		C	.28	−0.02	PROM1	Intronic
2	rs1799810	128176040	$9.2 * 10^{-6}$		T	.41	−0.02		
8	rs12680321	77148730	$9.8 * 10^{-6}$		T	.20	−0.03		Intergenic
13	rs9548119	38531581	$9.3 * 10^{-6}$	1	A	.22	0.03		Intergenic
13	rs9548119	38531581	$9.3 * 10^{-6}$	1	A	.22	0.03		Intergenic
2	rs1158867	128177377	$1.0 * 10^{-5}$		C	.41	−0.02		
10	rs11815041	1622424	$1.0 * 10^{-5}$	5	G	.49	0.02	ADARB2	Intronic
19	rs11085795	11988515	$1.0 * 10^{-5}$	1	A	.26	0.03	ZNF439	Intergenic
11	rs7120279	95720275	$4.9 * 10^{-5}$	1	C	.21	−0.03	MAML2	Intronic

Note: Independent markers: more than 500kb apart and in LD of $r^2 < 0.70$; SNPs in LD: the number of correlated SNPs that are in the top 50 (nonindependent groups of markers); Chr = Chromosome; MAF = Minor Allele Frequency; Closest gene = name of gene if the SNP is located in a known gene or within 50kb distance from a gene; Base pair locations: obtained from the HapMapI+II (b36r22) CEU legend files; Closest gene to the SNP: obtained from WGA Viewer release 57.

Table 3

Top Five Genes Showing the Strongest Association With Self-Rated Health

Gene	Chromosome	Start position	End position	Number of SNPs in gene	P value
ERCC3	2	127731335	127768222	74	$9.0 * 10^{-5}$
S100A5	1	151776246	151780865	30	$1.2 * 10^{-4}$
PROC	2	127892486	127903288	75	$1.5 * 10^{-4}$
S100A6	1	151773699	151775341	36	$1.7 * 10^{-4}$
S100A4	1	151782718	151784906	26	$1.8 * 10^{-4}$

of the top 50 SNPs were close to genes (e.g., *MAML2, PROM1, PROC*) broadly associated with a variety of health conditions, such as inflammation, coronary disease, cardio vascular disease, thrombosis, protein C deficiency etc (Reiner et al., 2008; Trynka, et al., 2009; Wu, et al., 2009). The Protein C (*PROC*) gene was also in the top 5 genes revealed by the gene-based test. Changes in these genes may have an effect on an individuals' self-rating of health. As no other study has explored the molecular genetic basis of self-rated health we cannot compare our findings.

Nevertheless, the fact that we did not find a genome-wide significant SNP is not totally unexpected considering that self-rated health is a very broad measure, influenced not only by several general somatic health factors but also by the health status presently experienced at the time of the rating. The concept of self-rated health has also been shown to be strongly associated with mental health, e.g. someone who is depressed may rate their health status as lower than someone who is in a good state of mind. Additionally, particular personality traits may play a role in how an individual rates his or her own health, for example a person scoring high in neuroticism would most likely rate their health slightly worse than a person very low in neuroticism. All these facts indicate that self-rated health is a very broad concept on a phenotypic level and may be genetically even more complicated. We suggest that very many rare variants of small effect size may influence self-rated health and are therefore difficult to detect. The fact that our Q-Q plot (Figure 1) lifts appreciable above the 95% confidence interval also hints at the highly polygenic nature of our trait, self-rated health. A recent paper by Yang et al. (2010) showed that 45% of the variance of human height could be explained considering all SNPs in a study (294,831), as opposed to 5% explained by SNPs detected by the conventional GWAS approach. This indicates that even in a very clear-cut and highly heritable phenotype such as human height, variance is explained by a large number of SNPs with very small effect; too small to be detected in a normal GWAS. The International Schizophrenia Consortium showed that by using the top-half (p value below 0.5) of the SNPs, they could quite consistently predict Schizophrenia and related disorders (e.g., bipolar disorder) in other samples, supporting the idea of a polygenic basis to the phenotype (Purcell et al., 2009). Another study on human height by Lango Allen et al. (submitted) also supports these findings: with a sample of almost 200,000 individuals they show that hundreds of genetic variants influence variance in adult height. This also shows that in order to find genes with such a small effect size a very large sample is needed. Aiming for this, a large consortium has been founded that plans to conduct a meta-analysis on self-rated health in the near future, combining several samples in order to possibly confirm the regions of interest found in the present study and find additional genetic variants underlying the variation of self-rated health and possibly even predict self-rated health and related health measures/indicators across different samples as in Purcell et al. (2009).

In summary, no genome-wide significant SNPs underlying self-rated health could be identified, indicating that the concept of self-rated health may be influenced by a large number of SNPs with very small effect size. In order to identify these, a very large sample would be needed which only can be accomplished by conducting a meta-analysis combining different samples.

References

Abecasis, G. R. (2010). *MACH*. Available at http://www.sph.umich.edu/csg/abecasis/MACH/index.html

Beaton, D. E., & Schemitsch, E. (2003). Measures of health-related quality of life and physical function. [Review]. *Clinical Orthopaedics and Related Research*, *413*, 90–105.

Bryant, L. L., Beck, A., & Fairclough, D. L. (2000). Factors that contribute to positive perceived health in an older population. *Journal of Aging and Health, 12*, 169–192.

Bucholz, K. K., Heath, A. C., Madden, P. A. F., Slutske, W. S., Statham, D. J., Dunne, M. P., & Martin, N. G. (1998). Drinking in an older population: Cross-sectional and longitudinal data from the Australian Twin Registry. In E. S. L. Gomberg, A. M. Hegedus & R. A. Zucker (Eds.), *Alcohol problems and aging* (pp. 41–62). Bethesda, MD: National Institutes of Health.

Chen, W. M., & Abecasis, G. R. (2007). Family-based association tests for genomewide association scans. *American Journal of Human Genetics, 81*, 913–926.

Christensen, K., Holm, N. V., McGue, M., Corder, L., & Vaupel, J. W. (1999). A Danish population-based twin study on general health in the elderly. *Journal of Aging and Health, 11*, 49–64.

Dudbridge, F., & Gusnanto, A. (2008). Estimation of significance thresholds for genomewide association scans. *Genetic Epidemiology, 32*, 227–234.

Hansell, N. K., Agrawal, A., Whitfield, J. B., Morley, K. I., Zhu, G., Lind, P. A., Pergadia, M. L., Madden, P. A., Todd, R. D., Heath, A. C., & Martin, N. G. (2008). Long-term stability and heritability of telephone interview measures of alcohol consumption and dependence. *Twin Research and Human Genetics, 11*, 287–305.

Harris, J. R., Pedersen, N. L., McClearn, G. E., Plomin, R., & Nesselroade, J. R. (1992). Age-differences in genetic and environmental influences for health from the Swedish Adoption Twin Study of Aging. *Journals of Gerontology, 47*, P213-P220.

Heath, A. C., Bucholz, K. K., Madden, P. A. F., Dinwiddie, S. H., Slutske, W. S., Bierut, L. J., Statham, D. J., Dunne, M. P., Whitfield, J. B., & Martin, N. G. (1997). Genetic and environmental contributions to alcohol dependence

risk in a national twin sample: Consistency of findings in women and men. *Psychological Medicine, 27,* 1381–1396.

Idler, E. L. (1993). Age-differences in self-assessments of health: Age-changes, cohort differences, or survivorship. *Journals of Gerontology, 48,* S289-S300.

Juerges, H. (2007). True health vs response styles: Exploring cross-country differences in self-reported health. *Health Economics, 16,* 163–178.

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F. et al. (submitted). Hundreds of variants influence human height and cluster within genomic loci and biological pathways. *Nature Genetics.*

Leinonen, R., Heikkinen, E., & Jylha, M. (2001). A pattern of long-term predictors of health ratings among older people. *Aging Clinical and Experimental Research, 13,* 454–464.

Leinonen, R., Kaprio, J., Jylha, M., Tolvanen, A., Koskenvuo, M., Heikkinen, E., & Ratanen, T. (2005). Genetic influences underlying self-rated health in older female twins. *Journal of the American Geriatrics Society, 53,* 1002–1007.

Lichtenstein, P., & Pedersen, N. L. (1995). Social relationships, stressful life events, and self-reported physical health — Genetic and environmental influences. *Psychology & Health, 10,* 295–319.

Liu, J. Z., McRae, A. F., Nyholt, D. R., Medland, S. E., Wray, N. R., Brown, K. M., Mann, G. J., Hayward, N. K., Montgomery, G. W., Visscher, P. M., Martin, N. G., & Macgregor, S. (2010). A versatile gene-based test for genome-wide association studies. *American Journal of Human Genetics, 87,* 139–145.

Lundberg, O., & Manderbacka, K. (1996). Assessing reliability of a measure of self-rated health. *Scandinavian Journal of Social Medicine, 24,* 218–224.

Medland, S. E., Nyholt, D. R., Painter, J. N., McEvoy, B. P., McRae, A. F., Zhu, G., Gordon, S. D., Ferreira, M. A., Wright, M. J., Henders, A. K., Campbell, M. J., Duffy, D. L., Hansell, N. K., Macgregor, S., Slutske, W. S., Heath, A. C., Montgomery, G. W., & Martin, N. G. (2009). Common variants in the trichohyalin gene are associated with straight hair in Europeans. *American Journal of Human Genetics, 85,* 750–755.

Mosing, M. A., Gordon, S. D., Medland, S. E., Statham, D. J., Nelson, E. C., Heath, A. C., Martin, N. G., & Wray, N. R. (2009a). Genetic and environmental influences on the co-morbidity between depression, panic disorder, agoraphobia, and social phobia: A twin study. *Depression and Anxiety, 26,* 1004–1011.

Mosing, M. A., Zietsch, B. P., Shekar, S. N., Wright, M. J., & Martin, N. G. (2009b). Genetic and environmental influences on optimism and its relationship to mental and self-rated health: A study of aging twins. *Behavior Genetics, 39,* 597–604.

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., & the International Schizophrenia Consortium (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature (London), 460,* 748–752.

Reiner, A. P., Carty, C. L., Jenny, N. S., Nievergelt, C., Cushman, M., Stearns-Kurosawa, D. J., Kurosawa, S., Kuller, L. H., & Lange, L. A. (2008). PROC, PROCR and PROS1 polymorphisms, plasma anticoagulant phenotypes, and risk of cardiovascular disease and mortality in older adults: The Cardiovascular Health Study. *Journal of Thrombosis and Haemostasis, 6,* 1625–1632.

Rodin, J., & McAvay, G. (1992). Determinants of change in perceived health in a longitudinal-study of older adults. *Journals of Gerontology, 47,* P373-P384.

Romeis, J. C., Scherrer, J. F., Xian, H., Eisen, S. A., Bucholz, K., Heath, A. C., Goldberg, J., Lyons, M. J., Henderson, W. G., & True, W. R. (2000). Heritability of self-reported health. *Health Services Research, 35,* 995–1010.

Silventoinen, K., Posthuma, D., Lahelma, E., Rose, R. J., & Kaprio, J. (2007). Genetic and environmental factors affecting self-rated health from age 16–25: A longitudinal study of Finnish twins. *Behavior Genetics, 37,* 326–333.

Svedberg, P., Lichtenstein, P., & Pedersen, N. L. (2001). Age and sex differences in genetic and environmental factors for self-rated health: A twin study. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences, 56,* S171–S178.

Trynka, G., Zhernakova, A., Romanos, J., Franke, L., Hunt, K. A., Turner, G., Bruinenberg, M., Heap, G. A., Platteel, M., Ryan, A. W., de Kovel, C., Holmes, G. K., Howdle, P. D., Walters, J. R., Sanders, D. S., Mulder, C. J., Mearin, M. L., Verbeek, W. H., Trimble, V., Stevens, F. M., Kelleher, D., Barisani, D., Bardella, M. T., McManus, R., van Heel, D. A., & Wijmenga, C. (2009). Coeliac disease-associated risk variants in TNFAIP3 and REL implicate altered NF-kappa B signalling. *Gut, 58,* 1078–1083.

Verweij, K. J. H., Zietsch, B. P., Medland, S. E., Gordon, S. D., Benyamin, B., Nyholt, D. R. et al. (in press). A genome-wide association study of Cloninger's Temperament scales: Implications for the evolutionary genetics of personality. *Biological Psychology.*

Wu, X. F., Spitz, M. R., Lee, J. J., Lippman, S. M., Ye, Y. Q., Yang, H. S., Khuri, F. R., Kim, E., Gu, J., Lotan, R., & Hong, W. K. (2009). Novel susceptibility loci for second primary tumors/recurrence in head and neck cancer patients: Large-scale evaluation of genetic variants. *Cancer Prevention Research, 2,* 617–624.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S. D., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E., & Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics.*