This thesis represents a genetic perspective on the association between exercise behavior and mental health. Longitudinal data (1991-2004) on regular exercise, symptoms of anxiety and depression, neuroticism, extraversion, sensation seeking, social problems and self-rated health were collected in twins and their family members registered at the Netherlands Twin Registry. The first part of this thesis described the population association between exercise behavior and symptoms of anxiety and depression and tested whether this association is causal or derives from a set of common genetic factors. The second part of this thesis aimed to further characterize the genetic basis of exercise behavior by applying genome-wide linkage and association techniques to exercise data from adults and by using advanced structural equation modeling of exercise data in parents and their adolescent offspring. This chapter provides a summary and discussion of the results, followed by some directions for future research.

Exercise behavior and symptoms of anxiety and depression

In chapters 2 to 4 it was reported that regular exercise in adults is associated with reduced symptoms of anxiety and depression, less neuroticism, more extraversion, more sensation seeking behaviors, fewer social problems and a better perceived health. These associations are small in size but hold for men and women of various ages (18 to 50 years). The results cor-

roborate previous findings on the relationship between exercise behavior and these variables (Allison et al., 2005; Arai & Hisamichi, 1998; Camacho et al., 1991; Farmer et al., 1988; Franken et al., 1994; Jack & Ronan, 1998; Potgieter & Bisschoff, 1990; Potgieter & Venter, 1995; Simonen et al., 2004; Weyerer, 1992).

To thoroughly test for causality in the association, I analyzed the cross-sectional and longitudinal association of exercise with symptoms of anxiety and depression for time intervals up to 11 years. A bivariate genetic model was used to test for causality in both the cross-sectional and longitudinal association. I argued that, if the association is fully explained by a causal effect of exercise on symptoms, *all* factors affecting exercise behavior should, indirectly and through the causal chain, also affect symptoms of anxiety and depression. In other words, under the causal hypothesis it is expected to find significant genetic and environmental correlations that together explain the association in a bivariate genetic model (since the causal chain is not explicitly modeled). Two additional methods were used to test for causality: the intra-pair differences method in MZ twins and longitudinal modeling of changes in exercise behavior and anxious and depressive symptoms in an individual over the years. All four methods did not point towards a causal effect of exercise on reduced symptoms of anxiety and depression. Rather, the association between exercise and these symptoms was best explained by a set of common genetic factors with opposite effects on exercise behavior and symptoms of anxiety, depression and neuroticism. Common genetic factors also explained the positive association between exercise behaviour and self-rated health (chapter 4), the latter reflecting different aspects of both physical and mental health (Eriksson et al., 2001; Idler & Benyamini, 1997).

Comparison with experimental studies on exercise, anxiety and depression

At first, these results seem at odds with findings from experimental studies demonstrating beneficial effects of exercise training on symptoms of anxiety and depression in clinical populations (Brosse et al., 2002; Dunn et al., 2005; Teychenne et al., 2008). There are several explanations for this discrepancy. First, population-based studies mainly include healthy subjects, whereas most experimental studies are conducted in subjects who suffer from an anxiety or depressive disorder. Clinical patients may respond differently to exercise and may have more room to improve than healthy subjects. Thus, there may be a floor effect of exercise on symptoms in the subjects who are included in population-based observational studies.

However, the effect of exercise on symptoms of anxiety and depression did not depend on the baseline levels of symptoms in our analyses, indicating that even in the subjects who initially score high on symptoms, an increase in exercise behavior does not reduce their symptoms.

A more relevant difference may be the setting of the exercise activities. In most experimental studies, exercise is monitored and part of a therapeutic program, whereas in this thesis the voluntary exercise behavior in people's every day life was studied. It may be that the anti-depressant effects only occur if exercise is prescribed and monitored within a therapeutic setting. In such a setting, the effects of exercising *per se* are confounded with positive feedback from supervisor, social interaction with other participants, and the often strong expectation of therapeutic efficacy of exercise (Blumenthal et al., 2007). The latter expectation is often (accidentally) amplified by the study recruitment procedure. It is reasonable to expect that a volunteer for "a study of exercise therapy for depression", if he or she is assigned to the exercise group, might tend to exhibit a favorable attitude toward exercise. Conversely, if a volunteer who signs up for "a study of exercise therapy for depression" is assigned to a non-exercise group (e.g., antidepressant medication, behavioral therapy, or a waiting list), it is reasonable to expect a certain degree of discontent and lack of belief in a good outcome.

A final difference between experimental and observational studies is that in experimental studies there may be a substantial selection bias, which is largely overcome in population-based observational studies. Experimental studies in clinical samples may only attract subjects who sought help for their problems and are willing to improve. In addition, exercise training studies may only attract subjects who are willing to engage in an exercise training program. These subjects may be the ones who have the experience that they are good at exercise, who like to exercise, or who have the strongest beliefs in the positive effects of exercise (Babyak et al., 2000). There may also be selective attrition; only those subjects who experience positive psychological effects complete the study.

Thus, the results from experimental studies cannot readily be generalized to the population at large. The results in this thesis show that the, often implicit, assumption in population-based studies and intervention programs that regular participation in exercise has beneficial psychological effects in *all* individuals in the population may not be valid. The results also imply that genetic factors need to be taken into account if we want to increase our understanding on the relationship between exercise and mental health. A major next step is to identify the genetic factors overlapping

between exercise behaviour, anxiety, and depression. Obvious areas to look for these pleiotropic genes are the neurobiological pathways in the brain that are likely to be involved in mood regulation, such as the norepinephrinergic, dopaminergic and serotonergic systems (Dunlop & Nemeroff, 2007; Hill et al., 1999; Lesch et al., 2003; Tremblay et al., 2002). Each of these pathways has been shown to be activated during exercise (Chaouloff, 1997; Dishman et al., 2006; Dishman, 1997).

One way to directly test for overlapping genes would be to test whether known genes for anxiety and depression are also associated with exercise behavior. This is however not an easy task. Although there have been numerous attempts to find genes for anxiety and depression, or related personality traits like neuroticism (Clement et al., 2002; Fullerton, 2006; Lopez-Leon et al., 2008; Middeldorp et al., 2008; Shifman et al., 2008; Sullivan et al., 2008), replication of association of genetic polymorphisms with these traits has been proven a difficult task. For example, a meta-analysis of gene-finding studies for major depressive disorder concludes that among the many susceptibility genes that have been studied, only six have been sufficiently replicated (the APOE, DRD4, GNB3, MTHFR, SLC6A3 and SLC6A4 genes) (Lopez-Leon et al., 2008). None of these genes showed a strong association to exercise behavior in the Dutch or American samples. Another strategy is to test for overlapping genes by first identifying the genes for exercise behavior, of which currently relatively little is known (Dishman et al., 2006; Rankinen et al., 2006a). Once identified, one could test in a next step whether these genes are also associated with mental health. In this thesis, I took this second approach by attempting to identify which genetic variants are associated with exercise behavior.

Genetics of exercise behavior

In chapters 5 to 7, several genomic regions and genetic variants were identified that are related to exercise behavior in adults. A genome-wide linkage scan showed that chromosomal region 19p13.3 was suggestively linked to exercise participation in Dutch adults (LOD=2.18). This region does not coincide with regions or genes found in previous smaller scaled linkage and association studies on exercise behavior and physical activity (Cai et al., 2006; Loos et al., 2005; Lorentzon et al., 2001; Salmen et al., 2003; Simonen et al., 2003a; Simonen et al., 2003b; Stefan et al., 2002; Winnicki et al., 2004), but it harbors a number of potentially interesting genes (the mus-

cle integrin-binding protein gene (MIBP), the thyroid receptor-interacting protein 10 gene (TRIP-10), the myosin IE gene (MYO1F), the endothelial differentiation G-protein coupled receptors 5 and 6 genes (EDG5 and EDG6), the thromboxane A2 receptor gene (TBXA2R) and the calponin-1 gene) that may be related to muscle performance and muscle blood flow.

A genome-wide linkage scan in British adult women showed that the chromosomal regions 3q22-q24 and 4q31-34 were linked to the maximum level of sports participation achieved. Suggestive linkages were found on chromosomes 3q22-q24 (at the sodium/hydrogen exchanger 9 (SLC9A9) gene) and 4q31-34 (near the fatty-acid binding protein 2 (FABP2) gene and the uncoupling protein 1 (UCP1) gene). The second region has been related to physical activity in a previous linkage study (Simonen et al., 2003b), an observation that is consistent with the hypothesis that some of the genetic influences on exercise behavior are mediated by exercise ability.

A genome-wide association analysis in Dutch and American adults revealed several novel SNPs in the SH3-domain GRB2-like (endophilin) interacting protein 1 (SGIP1) gene, the deoxyribonuclease II beta (DNASE2B) gene, the protease serine 16 (PRSS16) gene, the excision repair cross-complementing rodent repair deficiency, complementation group 2 (ERCC2) gene and the 3'-phosphoadenosine 5'-phosphosulfate synthase 2 (PAPSS2) gene that are associated with exercise participation. We replicated the association with two candidate genes for exercise behavior (Salmen et al., 2003; Stefan et al., 2002): the aromatase (CYP19A1) gene in the Dutch sample and the leptin receptor (LEPR) gene in the American sample. Associations of SNPs within several candidate linkage regions (Cai et al., 2006; De Moor et al., 2007a; Simonen et al., 2003b) were not replicated. Two of the genes found (SGIP1 and LEPR) are expressed in the hypothalamus and involved in the regulation of energy homeostasis (Park et al., 2006; Trevaskis et al., 2005). Their effects were independent of body mass index (BMI), suggesting a direct role of this pathway in the drive to exercise.

Our genome-wide association study, performed in two independent samples, constitutes the largest attempt in this field so far. However, overseeing the first wave of genome-wide association studies on complex traits and diseases, it is now clear that sample sizes must be increased much more (Craddock et al., 2008). For example, a recent genome-wide association analysis for adult height in 30,147 individuals identified 20 SNPs explaining only 3% of the variation in height (Weedon et al., 2008). This study illustrates that very large collaborative samples are needed to detect the small polygenic effects for complex traits such as height, and with-

out doubt also exercise behavior. Besides increased power to detect small effects, large collaborative samples have also shown to be useful to obtain more insight into the pathways through which genetic polymorphisms exert their influence on complex phenotypes (Craddock et al., 2008; McCarthy et al., 2008). An example of this is a recent genome-wide association study for type II diabetes combining data from multiple samples (Zeggini et al., 2007), in which association of the FTO gene was found in all samples except one in which the diabetes patients were selected based on low BMI. This showed that the effect of the FTO gene on type II diabetes was primarily mediated by adiposity (Frayling et al., 2007).

In addition to increases in sample size, an increase in the number and type of the genetic variants screened may also be required to fully characterize the heritability of exercise behavior. Many of the current genome-wide association studies are characterized by analyzing a set of common SNPs on the autosomal genome and then focusing on the most significant hits. Chapter 7 was no exception. There is good reason to suspect that additional analysis of the sex chromosomes and the mitochondrial DNA may yield a more complete picture. The genome-wide linkage scan for exercise participation reported in chapter 5 suggests that the genes affecting exercise in men and women are partly different. Analysis of sex chromosomes may shed more light on this sex heterogeneity in the genetic effects on exercise. Also, mitochondrial variants might well impact on exercise behavior. The mitochondrial DNA plays an important role in energy metabolism and variation at different sites of the mitochondria has already been related to several human diseases linked to exercise, such as type II diabetes (Lowell & Shulmanz, 2005; Saxena et al., 2006).

Other types of genomic variation such as insertion deletion polymorphisms, copy number variations (CNVs), methylation of the DNA and rare mutations should also be considered (Bodmer & Bonilla, 2008; McCarthy et al., 2008). For example, the 287-bp AluI repeat insertion-deletion polymorphism of the angiotensin converting enzyme (ACE) gene has been related to endurance performance, muscle efficiency and physical activity (Williams et al., 2000; Winnicki et al., 2004; Woods et al., 2000). The variation of SNPs in a genome-wide association study may not adequately capture the variation of other types of polymorphisms (Burgner et al., 2003) and therefore some important signals may be missed. Developments in technologies and methodologies to genotype and analyze these different types of genomic variation at a genome-wide basis will make it possible to extend the genetic analyses of exercise behavior and related traits.

Finally, new methodologies are needed that combine information

about genetic variation and known biological pathways and environmental factors relevant to exercise phenotypes. Methods have already been developed that incorporate information from previous linkage scans or candidate gene studies to prioritize specific regions or genes (Curtis et al., 2007; Li et al., 2008; Roeder et al., 2006). More recently a different approach was suggested to prioritize specific regions or genes by combining the p-values of individual SNPs for clusters of genes that have the same biological pathways (Wang et al., 2007). This method reduces the total number of tests and shifts the traditional focus on the most significant genes, to a focus on the most significant networks of genes that are involved in the same biological pathway or function. Of course, these methods all rely heavily on what is already known about the biology of the phenotype, but they may nevertheless provide a useful approach to genome-wide association analysis because it is more hypothesis-driven.

Hypotheses about genes for exercise

Pending the results from larger scaled GWA studies, with a more complete set of genetic markers, we can already speculate on the biological pathways that may prove important for exercise behavior. The first is homeostatic control over energy expenditure. Genes that affect the hypothalamic regulation of energy homeostasis have mainly been related to obesity and related metabolic diseases, such as type II diabetes mellitus (Liu et al., 2004; Park et al., 2006). The data reported in this thesis suggests that the regulation of energy homeostasis may also be related to the drive to spend energy and be physically active independently of BMI. This finding is consistent with the findings from two previous candidate gene association studies in which it was found that the LEPR and MC4R genes were related to physical activity independent of body composition phenotypes such as BMI (Loos et al., 2005; Stefan et al., 2002).

A second biological pathway through which genes may affect exercise behavior is through exercise ability. It is well-known that physical performance is genetically determined (Bouchard & Malina, 1998). Numerous twin studies show that different aspects of cardiorespiratory fitness and skeletomuscular strength and performance are heritable (Bouchard et al., 1998; Bouchard et al., 1999; De Mars et al., 2007; Perusse et al., 2001; Thomis et al., 1998), and some progress has been made in identifying the genetic variants that contribute to physical performance phenotypes (Rankinen et al., 2006a). These genetic variants help explain why one person is good at exercise and the other is not. The experience of being good at exercise, which probably starts early in life when children compare their

own ability and performance with that from friends and peers, may have a rewarding effect and this may stimulate an individual to continue to engage in exercise activities later in life. The experience of not being good at exercise and performing worse than the peer group may have punishing effects and may on the long term prevent an individual from regular participation in exercise.

A third biological pathway through which genes may affect exercise behavior is through the acute effects of exercise. Individuals who experience more rewarding than aversive effects after an acute bout of exercise are more likely to repeat their behavior and become regular lifetime exercisers. These rewarding effects may be physiological and psychological in nature (e.g., feeling 'energetic', less stressed, experiencing feelings of mastery and competence). In contrast, individuals for whom the rewarding effects do not outweigh the aversive effects will be less inclined to continue to exercise and are at higher risk to become and stay sedentary. A study in which persistent exercisers, recent adopters, fitness program dropouts and persistent sedentary individuals were extensively interviewed, suggested that regular exercisers differed from sedentary individuals mainly in that they enjoyed the exercise itself and felt that something was missing in their life when they did not regularly exercise (Gauvin, 1990). Similar to the genetic influences on changes in physical fitness as a result of exercise training (Bouchard et al., 1999; Perusse et al., 2001), it is hypothesized that the acute psychological effects of exercise training are also partly under genetic control.

A fourth pathway that could mediate the relationship between genes and exercise behavior is personality. It has been shown in chapter 2 that regular exercisers differ from non-exercisers on extraversion and sensation seeking. I have not yet tested whether the association with these traits reflects causal effects of personality on exercise behavior, but this is a distinct possibility. Individuals who are inclined to seek experience, thrills and adventure, or to enrich their social environment, may achieve this by participation in (competitive) sports. Sensation seeking behavior and extraversion are influenced by genetic factors (Johnson & Krueger, 2004; Stoel et al., 2006; Viken et al., 1994). Since it is well-known that the dopaminergic system in the brain is involved in experiencing pleasure and reward (Dunlop & Nemeroff, 2007) dopaminergic genes have been implied as an important source of this heritability. The finding that the observed decline of physical activity with age can be explained by depleted dopamine release and loss of dopamine receptors in the brain in human and non-human species (Ingram, 2000) is consistent with the hypothesis that

dopamine is involved in long-term exercise participation. Moreover, the dopamine 2 receptor gene is one of the few genes that has been implicated in physical activity in humans (Simonen et al., 2003a). Thus, one could speculate that genetic variants that are involved in the extraversion and sensation seeking are also involved in the drive to exercise.

Shared environment in adolescent exercise behavior

In contrast to the contribution of genetic factors to exercise behavior in adults (Beunen & Thomis, 1999; Eriksson et al., 2006; Kujala et al., 2002; Lauderdale et al., 1997; Stubbe et al., 2006a), exercise behavior in adolescents is largely influenced by shared environmental factors (Carlsson et al., 2006; Maia et al., 2002; Stubbe et al., 2005). In chapter 8, I used a parent-offspring design, including parents and siblings of adolescent twins, in order to determine whether the shared environment in adolescents is best explained by the influence of parental exercise behavior on their offspring's exercise behavior, by environmental factors specific to the adolescent generation, or by the effects of assortative mating. Since there are substantial differences in variance decomposition of exercise in both generations and also across sex, I extended the parent-offspring model to account for sex and generation differences in variance decomposition of exercise. Furthermore, data from adult twins and spouses were used to test for different causes of the spouse correlation (Heath & Eaves, 1985). It was found that the spouse correlation did not increase as a function of duration of the relationship. The spouse correlation was most likely because of phenotypic assortment.

The shared environment in adolescents mainly consists of generation-specific influences. There was little evidence for the influence of parental exercise behavior on their children's exercise behavior, except for the influence of fathers' behavior on sons' sports participation. The model tested the importance of parental exercise behavior on offspring exercise behavior. Parental influences that are unrelated to a parent's own exercise behavior, for example through social support mechanisms, cannot be ruled out. Future research needs to include measures of parental attitudes and social support towards children's exercise behavior to resolve whether there is cultural transmission of for example attitudes from parents to offspring. It also needs to focus on the generation specific environmental factors on adolescent exercise behavior (including peer behavior), as these, together

with additive genetic factors, appear to be the largest contributors to adolescent exercise behavior. It would further be of interest to examine the association between exercise behavior and well-being in adolescents and to test for causality. Since other factors contribute to regular exercise participation in adolescence than in adulthood, the mechanisms explaining the association between exercise and mental health may also be different.

Gene by exercise interaction

This thesis provides evidence for common genetic factors influencing exercise behavior and symptoms of anxiety and depression. The genetic influences on anxiety and depression may however also be *moderated* by exercise. Put differently, the effects of exercise on reduced symptoms of anxiety and depression may depend on an individual's genotype. This would mean that the causal effect of exercise on improved mood only holds for subgroups of the population. The possibility of gene by exercise interaction is consistent with the absence of a causal effect in the population at large and the presence of a causal effect in smaller subgroups of the population. It could also help explain the individual differences in the rewarding versus aversive psychological effects in response to exercise. As far as I know, there are no studies that investigated the possibility of gene by exercise interaction on reduced symptoms of anxiety and depression.

To test whether the genetic factors influencing mental health problems such as anxiety and depression can also be moderated by exercise, a research design in which twins participate in an exercise training program is needed. This design has already been successfully applied to the physiological effects of exercise training (Bouchard et al., 1999; Bouchard et al., 2000; Perusse et al., 2001), but these studies did not report psychological effects, such as changes in symptoms of anxiety and depression. A randomized controlled trial including extensive measurements at baseline and follow-up will provide more insight into the potential role of gene-exercise interaction in the association between exercise behavior and anxiety and depression. By including a wide range of measurements such as (changes in) cardiorespiratory fitness, muscle strength, anxiety, depression and attitudes towards exercise, it becomes possible to test the hypothesized mediating and moderating mechanisms between genetic variation at one hand and exercise and mental health at the other hand. Structural equation modeling techniques in twin pairs can for example be used to test whether the heritability of symptoms of anxiety and depression is a function of exercise

training (Purcell, 2002). Genotyping of twins will make it possible to study the interaction effect of measured genetic polymorphisms with exercise training on anxious and depressive symptoms (Fulker et al., 1999; van der Sluis et al., 2008a). Recent power studies show that a few hundred twin pairs would be needed to reliably detect such gene by exercise interaction effects (van der Sluis et al., 2008a; van der Sluis et al., 2008b).

Future perspective

The findings in this thesis have implications for population-based exercise intervention and prevention programs. Currently, findings on population associations between exercise and mental health are directly translated to these programs, assuming that the population association reflects a causal effect of exercise on well-being and assuming that exercise has the same beneficial effect on *all* participants. The results from this thesis do not support these two assumptions and urge for a change in perspective from less 'population-based' to more 'personalized' exercise intervention strategies. This requires an increased understanding of the pathways from genes to exercise behavior and of the differences in genetic sensitivity to the mental health benefits of exercise.