Identification of novel genetic variants associated with cardiorespiratory fitness

Running head: Genetic variants and VO_{2max}

Anja Bye[1,2], Marie Klevjer[1,2], Einar Ryeng[3], Gustavo JJ Silva[4], Jose Bianco Nascimento Moreira[1,2], Dorthe Stensvold[1], Ulrik Wisløff[1,5]

1. Cardiac Exercise Research Group (CERG) at Dept. of Circulation and Medical Imaging, Faculty of Medicine, Norwegian University of Science and Technology (NTNU), Norway
2. Department of Cardiology, St. Olavs Hospital, Trondheim, Norway
3. Genomics Core Facility (GCF), Norwegian University of Science and Technology (NTNU), Norway
4. Institute for Experimental Medical Research, Oslo University Hospital and University of Oslo, Oslo, Norway
5. School of Human Movement & Nutrition Sciences, University of Queensland, Australia

Address for correspondence: Anja Bye, Department of Circulation and Medical Imaging, PO box 8905, Medical Technical Research Center, 7491 Trondheim, Norway. Telephone: +47 72828023. E-mail: Anja.Bye@ntnu.no.

Word count: 6950 (including references and figure captions)

Key words: Maximal oxygen uptake, genetics, cardiovascular disease risk

Abbreviations

ACTN3:	Alpha-actinin-3
ADRB3:	Beta-3 adrenergic receptor
APOA1:	Apolipoprotein A1
APOER:	Apolipoprotein E receptor 2
BAHD1:	Bromo adjacent homology domain containing 1
BMI:	Body mass index
CRP:	C-reactive protein
CVD:	Cardiovascular disease
DNA:	Deoxyribonucleic acid
EDN1:	Endothelin 1
ERα:	Estrogen receptor alpha
GWAS:	Genome-wide association studies
HDL:	High-density lipoprotein
HUNT:	Nord-Trøndelag Health Study
IGF2:	Insulin-like growth factor 2
KCNQ1:	Potassium voltage-gated channel subfamily Q member 1
KCNQ1:	Potassium voltage-gated channel subfamily Q member 1
LDL:	Low-density lipoprotein
LDLR:	Low density lipoprotein receptor
LXR:	Liver X receptor
MAF:	Minor allele frequency
MET:	Metabolic equivalent
MYLIP:	Myosin regulatory light chain interacting protein
MYOCD:	Myocardin
PROX1:	Prospero homeobox protein 1
SNP:	Single-nucleotide polymorphism
VIP:	Vasoactive intestinal peptide
VIPR2:	Vasoactive intestinal peptide receptor 2
VLDLR:	Very low density lipoprotein receptor
VO_{2max}:	Maximal oxygen uptake

Abstract

Introduction: Low maximal oxygen uptake (VO_{2max}) is a strong and independent risk factor for all-cause and cardiovascular disease (CVD) mortality. For other CVD risk factors, numerous genetic association studies have been performed, revealing promising risk markers and new therapeutic targets. However, large genomic association studies on VO_{2max} are still lacking, despite the fact that VO_{2max} has a large genetic component. **Methods:** We performed a genetic association study on 123.545 single-nucleotide polymorphisms (SNPs) and directly measured VO_{2max} in 3470 individuals (exploration cohort). Candidate SNPs from the exploration cohort were analyzed in a validation cohort of 718 individuals, in addition to 7 wild-card SNPs. Sub-analyses were performed for each gender. Validated SNPs were used to create a genetic score for VO_{2max}. *In silico* analysis and genotype-phenotype databases were used to predict physiological function of the SNPs. **Results:** In the exploration cohort, 41 SNPs were associated with VO_{2max} ($p<5.0*10^{-4}$). Six of the candidate SNPs were associated with VO_{2max} also in the validation cohort, in addition to three wild-card SNPs ($p<0.05$, in men, women or both). The cumulative number of high-VO_{2max} SNPs correlated negatively with CVD risk factors, e.g. waist-circumference, visceral fat, fat %, cholesterol levels and BMI. *In silico* analysis indicated that several of the VO_{2max}-SNPs influence gene expression in adipose tissue, skeletal muscle and heart. **Conclusion:** We discovered and validated new SNPs associated with VO_{2max} and proposed possible links between VO_{2max} and CVD. Studies combining several large cohorts with directly measured VO_{2max} are needed to identify more SNPs associated with this phenotype.

Introduction

Low aerobic fitness, quantified as maximal oxygen uptake (VO_{2max}), is a strong and independent predictor of all-cause and cardiovascular mortality in healthy individuals and in patients with cardiovascular disease (CVD).[1-4] VO_{2max} is determined by a combination of genetic and environmental factors, and the genetic contribution is suggested to be ~50 %.[5, 6] Identification of genes and genomic variations associated with VO_{2max} would lead to a better understanding of this complex trait, and provide possible links between VO_{2max} and CVD. Previously, a few genes and genomic loci have been associated with VO_{2max}.[7-9] However, most studies are limited in size and employ the conventional hypothesis-driven approach of searching for pre-specified genomic associations, which limits the discovery of new genetic loci. Hence, the scientific community call for a large-scale systematic screening of genetic variants associated with directly measured VO_{2max} in a large well-characterized population.[10]

By taking advantage of one of the world's largest database of objectively measured VO_{2max}, we report the first large-scale systematic screening for genetic variants associated with VO_{2max}. Furthermore, we explore the association between VO_{2max}-related SNPs and CVD risk factors, and their potential biological implications by using *in silico* tools and genotype-phenotypes databases.

Material and Methods

Study participants

The Nord-Trøndelag Health Study (HUNT) is one of the largest health studies ever performed. It includes a unique database of questionnaire data, clinical measurements and biological samples. During the past 35 years, 120.000 individuals have contributed throughout four waves of the HUNT study (HUNT1 in 1984-86, HUNT2 in 1995-97, HUNT3 in 2006-08 and HUNT4 in 2017-19) in Norway. Participants in the present study attended a sub project during the third wave of HUNT (*HUNT3 Fitness Study*) designed to directly measure maximal oxygen uptake (VO_{2max}) in a healthy adult population.[11] Exclusion criteria for the *HUNT3 Fitness Study* were present or previous heart disease, stroke, angina, lung disease (asthma, chronic bronchitis, chronic obstructive pulmonary disease, and sarcoidosis), cancer, current pregnancy, orthopedic limitations and use of hypertensive medication. In total, 3470 participants that reached a true VO_{2max} were selected for genotyping after excluding first- and second-degree relatives (siblings, parents, children, grandparents, aunts, uncles or grandchildren). Close relatives were excluded both by using data from Statistics Norway, and by searching for segmental sharing using PLINK.[12] In the validation cohort, DNA-samples were analyzed from 718 participants from the *Generation 100 Study*.[13] This cohort includes both men and women, aged 70-77 years, which reached a true VO_{2max} using the same criteria as the *HUNT3 Fitness Study*. All participants were free from heart- or lung-disease, never had cancer, and did not have any other medical contraindication or orthopedic limitation to exercise. First- and second-degree relatives were also excluded leaving 718 individuals for genetic association study. This study was approved by the Regional committee for medical research ethics (4.2008.2792), the Nord-Trøndelag Health Study, the Norwegian Data Inspectorate, and by the National Directorate of Health. The study was in conformity with Norwegian laws and the Helsinki declaration, and a signed informed consent was obtained from all participants.

Clinical measurements

Weight and height were measured on a combined scale (Model DS-102, Arctic Heating AS, Nøtterøy, Norway), and body mass index (BMI) was calculated as weight divided by height squared (kg/m^2). Fat, muscle percentage and visceral fat were obtained using the InBody 720 scale (Biospace, Seoul, Korea).

Testing maximal oxygen uptake (VO_{2max})

An individualized protocol was applied to measure $VO_{2max.}$[14] Each test-subject was familiarized with treadmill walking during the warm-up of 8–10 minutes, also to ensure safety and avoid handrail grasp when this was not absolutely necessary. Oxygen uptake kinetics were measured directly by a portable mixing chamber gas-analyzer (Cortex MetaMax II, Cortex, Leipzig, Germany) with the participants wearing a tight face mask (Hans Rudolph, Germany) connected to the MetaMax II device. The system has previously been found reliable and valid in our laboratory. Heart rate was measured by radio telemetry (Polar S610i, Polar Electro Oy, Kempele, Finland). From the warm-up pace, the load was regularly increased. When the participants reached an oxygen consumption that was stable over 30 seconds, treadmill inclination (1–2% each step) or velocity (0.5–1 km/h) were increased until the participants were exhausted. A maximal test was achieved when the respiratory quotient reached >1.05 or when the oxygen uptake did not increase >2 ml/kg/min despite increased workload. VO_{2max} was measured as liters of oxygen per minute (l/min), and subsequently calculated as VO_{2max} relative to body mass (ml/kg/min) and VO_{2max} scaled ($ml/kg^{0.75}/min$).

Questionnaire-based information

Physical activity is likely to be the most important behavioral factor influencing VO_{2max}, and is therefore an important confounder to adjust for when we need to isolate the genetic contribution to the phenotype. Physical activity was registered based on the responses to a self-administered questionnaire.[15] The questionnaires included three questions and each participant's response to the questions (i.e. numbers in brackets) were multiplied to calculate a physical activity index score:

Question 1: "How frequently do you exercise?", with the response options "Never" (0), "Less than once a week" (0), "Once a week" (1), "2-3 times per week" (2.5) and "Almost every day" (5). *Question 2*: "If you exercise as frequently as once or more times a week: How hard do you push yourself?" with the response options: "I take it easy without breaking a sweat or losing my breath" (1), "I push myself so hard that I lose my breath and break into sweat" (2) and "I push myself to near exhaustion" (3). *Question 3*: "How long does each session last?", with the response options: "Less than 15 minutes" (0.1), "16-30 minutes" (0.38), "30 minutes to 1 hour" (0.75) and "More than 1 hour" (1.0). As the second and third question only addressed people who exercised at least once a week, both "Never" and "Less than once a week" yielded an index score of zero. Participants with a zero score were categorized as inactive, 0.05-1.5 as low activity, 1.51-3.75 as medium activity, and 3.76-15.0 as high activity.

Blood analysis

Standard biochemical analysis were performed on fresh venous non-fasting blood samples at Levanger Hospital, Norway. Non-fasting glucose was analyzed by hexokinase/G-G-PDH methodology reagent kit 3L82-20/3L82-40 Glucose (Abbott Diagnostics, Illinois, US), high-density lipoprotein (HDL) cholesterol by the Accelerator selective detergent methodology reagent kit 3K33-20 Ultra HDL (Abbott Diagnostics), total cholesterol by enzymatic cholesterol esterase methodology reagent kit 7D62-20 Cholesterol (Abbott Diagnostics), triglycerides by Glycerol Phosphate Oxidase methodology reagent kit 7D74 Triglyceride (Abbott Diagnostics) and C-reactive protein (CRP) by the Areoset CRP Vario kit (Abbott Diagnostics). Triglycerides and CRP were measured in only 80 % of the HUNT population. Low-density lipoprotein (LDL) cholesterol was calculated based on information on total cholesterol, HDL-cholesterol and triglycerides.

Genotyping of exploration cohort

DNA was extracted from blood samples stored in the *HUNT* biobank as described elsewhere.[16] DNA samples were analyzed by the custom-made Cardio-Metabochip including approximately 210.000

SNPs (Illumina, CA, US). The annotation on the chip is based on Genome build 36.3. The Cardio-Metabochip was designed by representatives of the following genome-wide association studies (GWAS) meta-analysis consortia: CARDIoGRAM (coronary artery disease), DIAGRAM (type 2 diabetes), GIANT (height and weight), MAGIC (glycemic traits), Lipids (lipids), ICBP-GWAS (blood pressure), and QT-IGC (QT interval). The candidate SNPs were selected according to five sets of criteria: (I) individual SNPs displaying evidence for association in GWAS meta-analysis to diseases and traits relevant to metabolic and atherosclerotic-cardiovascular endpoints, (II) detailed fine mapping of loci validated at genome-wide significance from these meta-analyses, (III) all SNPs associated at genome-wide significance with any human trait, (IV) "wildcards" selected by each Consortium for Consortium-specific purposes, and (V) other useful content, including SNPs that tag common copy number polymorphisms, SNPs in the human leukocyte antigen region, SNPs marking the X and Y chromosomes and mitochondrial DNA, and for sample fingerprinting.

The study was designed as a quantitative trait approach with VO_{2max} as a continuous variable, as this provides the best statistical power. The genotyping raw data was subjected to systematic quality control using the statistical software PLINK.[12] Individuals with low genotype call rate (less than 90 %) were excluded. SNPs with a genotype call rate less than 95 % or a minor allele frequency less than 1 % were also excluded. Furthermore, SNPs that clearly deviate from the expected Hardy-Weinberg Equilibrium were excluded ($p<10^{-7}$). Individuals who showed gender discrepancies based on the heterozygosity rate from chromosome X were also excluded.

Genotyping of validation cohort

Candidate SNPs from the exploration cohort, as well as a 7 wild-card SNPs not included on the Cardio-Metabo chip, were genotyped using the Agena Biosciences MassARRAY® platform (formerly Sequenom). SNP multiplexes were designed using Assay Design Suite v.1.0 software (Agena Bioscience, San Diego, CA, US). Genotyping was performed according to the manufacturer's protocol using IPLEX Gold assay (Agena Bioscience, San Diego, CA, US) and analyzed using the

MassARRAY Analyzer 4 platform. Mass signals for the different alleles were captured with high accuracy by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). Genotype clustering and individual sample genotype calls were generated using Sequenom TyperAnalyzer v.4.0 software (Agena Bioscience, CA, US).

In silico analysis of transcription starting sites

To determine if candidate SNPs were located in transcription factors binding sites, we performed *in silico* analysis of predicted transcription factor binding sites using the software PROMO.[17]

Genotype-Tissue Expression (GTEx) database

To explore the relationship between SNPs and gene expression in different human tissues/organs, we used the GTEx database. The database includes ~900 post-mortem donors and opens the possibility for studying the effects of genetic variation in multiple human reference tissues.[18]

BXD mouse database

The BXD database is an open-access web service for systems genetics (www.genenetwork.org) to explore the genetic control of multiple phenotypes. [19] The database includes more than 2000 phenotypes across a large panel of isogenic but diverse strains of mice (BXD type) Among other phenotypes, the database contains phenotypes such as heart rate and oxygen consumption and blood parameters, such as hematocrit and iron levels, highly relevant for exploring the functional importance of VO_{2max}-related genes. We tested potential correlations between expression levels of candidate VO_{2max}-related genes and relevant phenotypes in mouse on both chow diet and high fat diet, independently.

Statistical analyses

The association between the final 123.545 variants and VO_{2max} were analyzed by linear regression using PLINK. The main covariates for the VO_{2max} phenotype were gender, age (years) and physical activity level (Kurtze score). The cut-off for significance were set to ($p<5.0*10^{-4}$) in the exploration

cohort, as findings reaching the traditional genome-wide significance were considered unlikely due to the low number of available cases and that VO_{2max} is a complex trait. To overcome the issue of using a moderately stringent p-value, validation of the findings in a separate cohort was necessary. In the validation cohort, associations between VO_{2max} and candidate SNPs was tested using the same statistical analyses as in the exploration cohort. Nominal p-value was considered significant ($p<0.05$). A genetic score was created using a combination of 9 SNPs associated with VO_{2max}. Each participant was scored according to the sum of high VO_{2max} genotypes carried. The differences in VO_{2max} between participants with increasing numbers of favorable genotypes were calculated by one-way ANOVA using the LSD post hoc test.

Results

Characteristics of the participants in the exploration cohort (*HUNT3 Fitness Study*) and the validation cohort (*Generation 100 Study*) are shown in **Table 1**.

After filtration of genotyping data, 123.545 SNPs were tested for their association with VO_{2max}. 41 SNPs were significantly associated with VO_{2max} in the exploration cohort after adjusting for age, gender and physical activity level ($p<5.0*10^{-4}$). Relevant locus zoom plots can be found in Supplementary Figure 1. The candidate SNPs were subsequently genotyped in a validation cohort, in addition to 7 wild-card SNPs not included on the chip used for the exploration cohort. The association between VO_{2max} and six novel SNPs were replicated in the validation cohort ($p<0.05$, **Table 2**). The SNP in the promoter region of the Myosin Regulatory Light Chain Interacting Protein (MYLIP) (rs3757354) did not pass the significance threshold in the validation cohort, however sub analyses for each gender showed a highly significant association in women, and the SNP was therefore included in Table 2. Three of the 7 wild-card SNPs in the genes beta-3 adrenergic receptor (*ADRB3*), alpha-actinin-3 (*ACTN3*) and endothelin 1 (*EDN1*) were associated with VO_{2max} in men, women or both genders ($p<0.05$, **Table 2**). Candidate SNPs that failed to be replicated in the validation cohort can be found in Supplementary Table 1.

Considering that VO_{2max} is a complex trait influenced by multiple genetic factors,[20] we assessed whether a cumulative effect existed between the number of favorable genotypes and VO_{2max}. By using a combination of the 9 SNPs from **Table 2**, and scoring the high VO_{2max}-associated genotypes 1 and low VO_{2max} genotypes 0, we calculated a genetic score for each participant estimating inborn VO_{2max}. In the validation cohort, the variations in VO_{2max} ranged from 63 ml/kg$^{0.75}$/min to 98 ml/kg$^{0.75}$/min, for participants scoring 1 or 7, respectively (**Figure 1A**). This corresponded to unscaled VO_{2max}-values ranging from 22.3 ml/kg/min to 32.7 ml/kg/min, for participants scoring 1 or 7, respectively. To

illustrate that the power of this allele combination was independent of physical activity levels, we split the participants into two subgroups, participants below (inactive) and above (active) the median physical activity level. Interestingly, the proposed score appears to be robust even with the reduced sample power of this sub-analysis (**Figure 1B**).

Using the same SNPs as basis, we also found a cumulative effect on the number of favorable SNPs and the decline in several risk factors associated with CVD, e.g. waist circumference, visceral fat, fat %, cholesterol and BMI (**Figure 2**). In addition, among the participants with 1-4 favorable SNPs, 36 % were on treatment for hypertension, compared to 23 % in those with more than 4 favorable SNPs ($p<0.05$). Among participants reporting little or no physical activity (Kurtze score <3.75, n=235) those with 1-4 favorable SNPs had higher fat percentage (+ 2 %), visceral fat (+9 %), total cholesterol (+ 5 %) and LDL-cholesterol (+ 6 %) compared those with a more than 4 favorable SNPs ($p<0.05$).

To explore and predict physiological consequences of the VO_{2max}-SNPs, we used *in silico* tools and genotype-phenotype databases. The non-synonymous SNP rs3803357, located in the first exon of the Bromo adjacent homology domain containing 1 (*BAHD1*) gene, cause a shift from the amino acid glycine to lysine. The group of participants homozygote for the rs3803357 minor allele (TT) (24 %) had a 3 $ml/kg^{0.75}/min$ lower VO_{2max} than the group carrying the heterozygote allele (GT) (50 %) or the common allele homozygotes (GG) 26 % (**Figure 3A**). In the validation cohort, the group of participant's homozygote for the rs3803357 minor allele (TT) (24 %) had 4 % and 7 % lower VO_{2max} compared to those harboring the (GT) and (GG) variants, respectively (**Figure 3B**). SNPs located outside the promoter region or within introns and exons may influence transcription of proximal genes. Using the Genotype-Tissue Expression (GTEx) database, rs3803357 was found to be associated with differential expression *BAHD1* in the left ventricle (p=9.0e-9) (**Figure 3C**). By using the BXD mice population, we found significant negative correlations between cardiac expression of *Bahd1* and basal VO_2 (in an untrained state), as well as with myocardial mass (**Figure 3D**).

Another SNP that was found to be associated with VO_{2max} in women, rs3757354, was located within the 2-kb upstream region of *MYLIP*. Women homozygote for the rs3757354 common allele (GG) (56 %) had a 3 $ml/kg^{0.75}$/min higher VO_{2max} than the group carrying the heterozygote allele (AG) (37 %) or the minor allele homozygotes (AA) (7 %) (**Figure 4A**). To determine if rs3757354 could interfere with transcription factor binding, we performed *in silico* analysis to discover possible transcription factor binding sites. The analysis predicted that having the A allele at rs3757354 creates a perfect binding site for the estrogen receptor alpha (ER-α) targeting the sequence TGACC, whereas having the G allele at rs3757354 is likely to disable the binding of ER-α, potentially reducing estrogen-induced expression of *MYLIP* (**Figure 4B**). Using the GTEx database, we found that rs3757354 was associated with differential expression of *MYLIP* in the adipose tissue, skeletal muscle and the heart ($p<0.05$). By using the BXD mice population, we found significant negative correlations between cardiac expression of *Mylip* and heart mass (**Figure 4C**). Participants harboring the high-VO_{2max} genotype (GG) had significantly lower waist, BMI, visceral fat, fat percentage, CRP-levels, as well as significantly higher HDL-cholesterol as compared to the low-VO_{2max} genotypes (AA) and (AG) (**Figure 4D**). Furthermore, among those with the low-VO_{2max} genotypes (AA) and (AG) significantly more of the participants were on treatment for hypercholesterolemia (10 %) compared to those with the high-VO_{2max} genotype (GG) (3 %) (**Figure 4E**).

Discussion

Here we report the first large-scale screening for genetic variants associated with maximal oxygen uptake (VO_{2max}). So far, the lack of large studies directly measuring VO_{2max} has limited the possibilities for large genetic association studies for this phenotype. In this present study, we validated 6 new SNPs associated with VO_{2max}, and replicate associations with 3 SNPs previously associated with fitness-related traits. [10, 21, 22] Based on these nine SNPs we proposed a genetic score for each participant reflecting inborn VO_{2max}. The mean difference in VO_{2max} between those with 1 favorable SNP compared to those with 7 favorable SNPs was 10.4 ml/kg/min, which is equal to a difference in 3 METs (as 1 MET≈3.5 ml/kg/min). In other prospective studies, it has been suggested that a decrease of 1 MET (3.5 $mL \cdot kg^{-1} \cdot min^{-1}$) is associated with increased risk of diabetes, hypertension and the metabolic syndrome,[23-25] whereas a corresponding increase has been associated with lower risk of all-cause and CVD mortality.[2, 26] Interestingly, the number of favorable VO_{2max}-SNPs carried correlated negatively with several CVD risk factors, like waist circumference, BMI, visceral fat, fat percentage, total cholesterol and LDL-cholesterol. Furthermore, among the participants with 1-4 favorable SNPs, significantly more participants were on treatment for hypertension, compared to those with 5 or more SNPs ($p<0.05$). Furthermore, sedentary participants with 1-4 favorable SNPs had higher fat percentage, more visceral fat, higher total cholesterol and higher LDL-cholesterol compared those with 5 or more favorable SNPs. This indicated that inborn high VO_{2max} is associated with decreased CVD risk.

Since VO_{2max} is a strong predictor of cardiovascular health,[2, 3, 24, 27] SNPs associated with VO_{2max} may provide physiological explanation for the link between VO_{2max} and CVD. In this present study, a significant association was found between VO_{2max} and a missense mutation in the exon of *BAHD1* (rs3803357), which involves transcription of different amino acids depending on genotype. Potentially this SNP may influence BAHD1 protein function, or alter transcription levels of BAHD1 or other

proximal genes. According to data from the GTEx database, rs3803357 is associated with differences in BAHD1 gene expression in adipose tissue, skeletal muscle and left ventricle, but also with differential expression of proximal genes in different tissues. Furthermore, the BXD mouse database indicates that cardiac *Bahd1* levels correlates with basal VO_2 and heart mass in mice. A previous study has shown that BAHD1 act as a transcription repressor that, among other things, is involved in epigenetic repression of different cardiac growth factors.[28] BAHD1 is known to repress insulin growth factor 2 (*IGF2*) expression by binding to its promoter and recruiting heterochromatin proteins[28]. Interestingly, we have previously shown that *Igf2* is one of the most significantly upregulated genes in the left ventricle of rats with inherited high $VO_{2max.}$[29] In addition to the links to cardiac phenotype, we also found trends toward lower fat percentage and total cholesterol levels in participants with the high-VO_{2max} genotypes. Further studies are needed to explore the links between these genomic loci and VO_{2max}.

Another interesting SNP found to be associated with VO_{2max} was located in an intron of the vasoactive intestinal peptide receptor 2 (*VIPR2*). *VIPR2* encodes a neuropeptide receptor that is expressed in the heart and the coronary arteries.[30] In the heart, VIPR2 regulates cardiomyocyte contractility in response to binding of vasoactive intestinal peptide (VIP).[31] The release of VIP also increases coronary artery vasodilatation.[30] Interestingly, several studies have shown that physical activity induces the release of VIP, hence VIPR2 is likely to be important for cardiovascular adaptions during exercise.[32, 33] Furthermore, in rats and humans with cardiomyopathy, the levels of VIPR2 are reduced both in heart and serum, suggesting also a link between VIPR2 and CVD.[30, 34]

One of the validated SNPs (rs3757354) was located in the promoter region of *MYLIP*, potentially interfering with transcription factor binding sites. SNPs in promoter regions may cause loss of transcription factor binding sites or formation of a novel binding sites, which may influence how the gene is transcribed upon different stimuli.[35] The SNP in the promoter region of *MYLIP* was

significantly associated with VO_{2max} in women in the validation cohort, but not in men, indicating that this genotype influence VO_{2max} in a gender-specific manner. Interestingly, *in silico* analysis using the software PROMO indicated that rs3757354 is located in the transcription factor binding site of the estrogen receptor alpha (ER-α). In fact, having the G allele in that locus is predicted to disable the binding of ER-α, thus abolishing estrogen-induced expression of *MYLIP*. For participants carrying the high VO_{2max} genotype GG at this locus, the *in silico* analysis predicted that ER-α is unable to bind and induce expression of *MYLIP*. In contrast, participants with the low VO_{2max} genotype AA are predicted to harbor intact binding sites for ER-α targeting the sequence TGACC. As ER-α is activated by estrogen, this may explain why this SNP is only important for VO_{2max} in women. This was further supported by evidence from the GTEx-database, showing that rs3757354 was associated with differential expression of *MYLIP* in the adipose tissue, skeletal muscle and the heart ($p<0.05$). Interestingly, using the BXD mouse database, we found significant negative correlations between cardiac expression of *Mylip* and myocardial mass. Furthermore, a previous transcriptome characterization of estrogen-treated human myocardium identified MYLIP as a sex-specific element influencing contractile function, more specifically showing a negative correlation between cardiac expression of *Mylip* and contractile function.[36] In line with our data, several other studies have reported gender-specific associations with *MYLIP* genotypes.[36, 37] For instance, Yan *et al* report that G allele-carrying women from the *Bai Ku Yao population* had higher levels of HDL-cholesterol than the non-carriers. Furthermore, G allele-carrying women from the *Han* population had decreased levels of total cholesterol and apolipoprotein A1 (ApoA1) compared to non-carriers. None of these associations were seen in men.[37] In our study, rs3757354 was also found to be significantly associated with HDL-cholesterol, and several other CVD risk factors like waist circumference, BMI, visceral fat, fat percentage and high-sensitivity CRP-levels. Furthermore, G-allele homozygotes were less likely to be on cholesterol treatment, suggesting that these women are less prone to hypercholesterolemia. Studies in mice show that increased liver expression of *Mylip* promotes degradation of the LDL-receptors (LDLR) and thereby circulating LDL-cholesterol.[38] Induction of *MYLIP* expression by the liver X

receptors (LXRs) transcription factors is important for cholesterol homeostasis.[38] Upon stimulation by LXRs or LXR agonists, MYLIP degrades the LDLR, apolipoprotein E receptor 2 (ApoER2) and the very low-density lipoprotein receptor (VLDLR) thereby raising circulating LDL-cholesterol. Furthermore, cells lacking *Mylip* exhibit markedly elevated levels of LDLR and increased rates of LDL-uptake.[39] Overall, the literature provides compelling evidence suggesting important physiological consequences of *MYLIP* genetic variation. Based on the gender-specific associations of rs3757354, and the previous reported associations with longevity and CVD, [37, 40, 41] these findings may shed new light on the gender-differences in CVD and the influence of sex-specific hormones.[42] Furthermore, the location of rs3757354 in a potential transcription factor binding site that is under the control of estrogen encourages this hypothesis.

The potassium voltage-gated channel, KQT-like subfamily, member 1 (KCNQ1) is a well-characterized gene involved in potassium handling in cardiomyocytes. The rs2074238 located in an intron of *KCNQ1* was associated with VO_{2max} both in the exploration and validation cohort. In previous meta-analysis, the minor allele T of this SNP has been associated with a shortening of the QT interval, a measure of myocardial repolarization time.[43] Prolongation of the QT interval duration, is a risk factor for drug-induced arrhythmias and sudden cardiac death.[44] Other studies have also reported that this particular SNP affects QT interval in healthy Europeans.[45, 46] In our study, participants harboring the genotype previously associated with prolonged QT interval, had a significantly higher $VO_{2max,}$ compared to the other genotypes. This may shed new information on the U-shaped association between risk of arrhythmias and VO_{2max}.[47] However, only mechanistic studies will be able to identify the true functional consequences of rs2074238.

Due to large differences in human physiology between men and women, and that gender is a major determinant of VO_{2max},[48] it is likely that some genetic variants have more effect in one gender compared to the other. In our study, the lack of similar dependency among men and women for several

of the reported SNPs indicates that they may affect VO_{2max} in a gender-specific manner. A previous study suggests that androgenic hormones are likely to make a significant contribution to VO_{2max} in men, hence, the relative effect of the VO_{2max}-related SNPs may be lower in men than in women.[49] Since our approach only covers a part of the genome, we do not have sufficient evidence to fully evaluate the genetic contribution to VO_{2max} in men compared to women. As DNA-sequencing technology becomes more accessible, future studies will hopefully be able to explain gender differences with greater confidence.

Limitations

There are some limitations related to this study not discussed previously. First, the age distribution of the validation cohort is different from the exploration cohort, hence, we may fail to validate some of the SNPs from the exploration cohort due to their importance in different stages of life. Next, as this study only includes individuals with Caucasian decent, the results are not necessary valid for other ethnicities, and would have to be validated in other cohorts. Furthermore, estimation of physical activity level is an important source of bias, as this parameter is included as a covariate in the genetic association analyses. Nevertheless, as regular physical activity has large influence on VO_{2max}, this was considered a necessary covariate despite the use of self-reported data.

Conclusion

This is the first large genetic association study on directly measured VO_{2max}. We discovered and validated new genetic loci associated with VO_{2max} and explored their physiological importance using genotype-phenotype databases and *in silico* tools. We proposed a genetic signature of inborn VO_{2max} consisting of 9 SNPs that could distinguish high vs. low fitness individuals based on simultaneous carriage of multiple favorable alleles. Interestingly, the number of favorable SNPs correlated negatively with the presence of several CVD risk factors. Future studies combining several large cohorts with directly measured VO_{2max} are needed to identify more SNPs associated with this complex phenotype.

Acknowledgement and funding

This study was founded by grants from the K.G. Jebsen Foundation, the Norwegian Health Association, the Liaison Committee between the Central Norway Regional Health Authority (RHA) and the Norwegian University of Science and Technology (NTNU), St. Olavs Hospital and the Faculty of Medicine and Health Sciences at NTNU and Foundation for Cardiovascular Research at St. Olav's Hospital and NTNU. The chip genotyping was provided by the Genomics Core Facility (GCF), NTNU. GCF is funded by the Faculty of Medicine and Health Sciences at NTNU and Central Norway RHA. Genotyping of candidate SNPs was provided by the Centre for Integrative Genetics (CIGENE) at the Norwegian University of Life Sciences. The Nord-Trøndelag Health Study (HUNT) is a collaboration between HUNT Research Center, Faculty of Medicine and Health Sciences, NTNU, Nord-Trøndelag County Council, Central Norway RHA, and the Norwegian Institute of Public Health. Professor Johan Auwerx and researcher Hao Li at the École Polytechnique Fédérale in Lausanne, Switzerland contributed with important data and analyses from the BXD database. Testing of VO_{2max} in *Generation 100* was provided by NeXt Move core facility at NTNU. NeXt Move is funded by the Faculty of Medicine and Health Sciences, NTNU and Central Norway RHA. The results of the study are presented clearly, honestly, and without fabrication, falsification or inappropriate data manipulation.

The authors have declared that no conflict of interest exists.

References

1. Kavanagh T, Mertens DJ, Hamm LF, et al. Prediction of long-term prognosis in 12 169 men referred for cardiac rehabilitation. *Circulation.* 2002;106:666-671.

2. Myers J, Prakash M, Froelicher V, Do D, Partington S, Atwood JE. Exercise capacity and mortality among men referred for exercise testing. *The New England journal of medicine.* 2002;346:793-801.

3. Ozemek C, Laddu DR, Lavie CJ, et al. An Update on the Role of Cardiorespiratory Fitness, Structured Exercise and Lifestyle Physical Activity in Preventing Cardiovascular Disease and Health Risk. *Prog Cardiovasc Dis.* 2018;61:484-490.

4. Imboden MT, Harber MP, Whaley MH, et al. The Association between the Change in Directly Measured Cardiorespiratory Fitness across Time and Mortality Risk. *Prog Cardiovasc Dis.* 2019;62:157-162.

5. Bouchard C, Daw EW, Rice T, et al. Familial resemblance for VO2max in the sedentary state: the HERITAGE family study. *Medicine and science in sports and exercise.* 1998;30:252-258.

6. Schutte NM, Nederend I, Hudziak JJ, Bartels M, de Geus EJ. Twin-sibling study and meta-analysis on the heritability of maximal oxygen consumption. *Physiological genomics.* 2016;48:210-219.

7. Bouchard C, Rankinen T, Chagnon YC, et al. Genomic scan for maximal oxygen uptake and its response to training in the HERITAGE Family Study. *J Appl Physiol.* 2000;88:551-559.

8. Rico-Sanz J, Rankinen T, Rice T, et al. Quantitative trait loci for maximal exercise capacity phenotypes and their responses to training in the HERITAGE Family Study. *Physiological genomics.* 2004;16:256-260.

9. Ahmetov I, Kulemin N, Popov D, et al. Genome-wide association study identifies three novel genetic markers associated with elite endurance performance. *Biol Sport.* 2015;32:3-9.

10. Bray MS, Hagberg JM, Perusse L, et al. The human gene map for performance and health-related fitness phenotypes: the 2006-2007 update. *Medicine and science in sports and exercise.* 2009;41:35-73.

11. Aspenes ST, Nilsen TI, Skaug EA, et al. Peak Oxygen Uptake and Cardiovascular Risk Factors in 4631 Healthy Women and Men. *Medicine and science in sports and exercise.* 2011;43:1465-1473.

12. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics.* 2007;81:559-575.

13. Stensvold D, Viken H, Rognmo O, et al. A randomised controlled study of the long-term effects of exercise training on mortality in elderly people: study protocol for the Generation 100 study. *BMJ open.* 2015;5:e007519.

14. Rognmo Ø, Hetland E, Helgerud J, Hoff J, Slørdahl SA. High intensity aerobic interval exercise is superior to moderate intensity exercise for increasing aerobic capacity in patients with coronary artery disease. *Eur J Cardiovasc Prev Rehabil.* 2004;11:216-222.

15. Kurtze N, Rangul V, Hustvedt BE, Flanders WD. Reliability and validity of self-reported physical activity in the Nord-Trøndelag Health Study: HUNT 1. *Scandinavian journal of public health.* 2008;36:52-61.

16. Moses EK, Johnson MP, Tommerdal L, et al. Genetic association of preeclampsia to the inflammatory response gene SEPS1. *Am J Obstet Gynecol.* 2008;198:336 e331-335.

17. Messeguer X, Escudero R, Farre D, Nunez O, Martinez J, Alba MM. PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics (Oxford, England).* 2002;18:333-334.

18. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nature genetics.* 2013;45:580-585.

19. Andreux PA, Williams EG, Koutnikova H, et al. Systems genetics of metabolism: the use of the BXD murine reference panel for multiscalar integration of traits. *Cell.* 2012;150:1287-1299.

20. Wilmore JH, Leon AS, Rao DC, Skinner JS, Gagnon J, Bouchard C. Genetics, response to exercise, and risk factors: the HERITAGE Family Study. *World Rev Nutr Diet.* 1997;81:72-83.

21. Rankinen T, Perusse L, Rauramaa R, Rivera MA, Wolfarth B, Bouchard C. The human gene map for performance and health-related fitness phenotypes. *Medicine and science in sports and exercise.* 2001;33:855-867.

22. Rankinen T, Bray MS, Hagberg JM, et al. The Human Gene Map for Performance and Health-Related Fitness Phenotypes: The 2005 Update. *Medicine and science in sports and exercise.* 2006;38:1863-1888.

23. Carnethon MR, Gulati M, Greenland P. Prevalence and cardiovascular disease correlates of low cardiorespiratory fitness in adolescents and adults. *Jama.* 2005;294:2981-2988.

24. Carnethon MR, Gidding SS, Nehgme R, Sidney S, Jacobs DR, Jr., Liu K. Cardiorespiratory fitness in young adulthood and the development of cardiovascular disease risk factors. *JAMA.* 2003;290:3092-3100.

25. Ozemek C, Lavie CJ, Rognmo O. Global physical activity levels - Need for intervention. *Prog Cardiovasc Dis.* 2019;62:102-107.

26. Kodama S, Saito K, Tanaka S, et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. *Jama.* 2009;301:2024-2035.

27. Kaminsky LA, Arena R, Ellingsen O, et al. Cardiorespiratory fitness and cardiovascular disease - The past, present, and future. *Prog Cardiovasc Dis.* 2019;62:86-93.

28. Bierne H, Tham TN, Batsche E, et al. Human BAHD1 promotes heterochromatic gene silencing. *Proceedings of the National Academy of Sciences of the United States of America.* 2009;106:13826-13831.

29. Bye A, Langaas M, Hoydal MA, et al. Aerobic capacity-dependent differences in cardiac gene expression. *Physiological genomics.* 2008;33:100-109.

30. Dvorakova MC. Cardioprotective role of the VIP signaling system. *Drug news & perspectives.* 2005;18:387-391.

31. Henning RJ, Sawmiller DR. Vasoactive intestinal peptide: cardiovascular effects. *Cardiovascular research.* 2001;49:27-37.

32. Galbo H, Hilsted J, Fahrenkrug J, Schaffalitzky De Muckadell OB. Fasting and prolonged exercise increase vasoactive intestinal polypeptide (VIP) in plasma. *Acta physiologica Scandinavica.* 1979;105:374-377.

33. Oktedalen O, Opstad PK, Schaffalitzky de Muckadell OB. The plasma concentrations of secretin and vasoactive intestinal polypeptide (VIP) after long-term, strenuous exercise. *Eur J Appl Physiol Occup Physiol.* 1983;52:5-8.

34. Dvorakova MC, Pfeil U, Kuncova J, et al. Down-regulation of vasoactive intestinal peptide and altered expression of its receptors in rat diabetic cardiomyopathy. *Cell Tissue Res.* 2006;323:383-393.

35. Sinnett D, Beaulieu P, Belanger H, et al. Detection and characterization of DNA variants in the promoter regions of hundreds of human disease candidate genes. *Genomics.* 2006;87:704-710.

36. Kararigas G, Bito V, Tinel H, et al. Transcriptome characterization of estrogen-treated human myocardium identifies myosin regulatory light chain interacting protein as a sex-specific element influencing contractile function. *Journal of the American College of Cardiology.* 2012;59:410-417.

37. Yan TT, Yin RX, Li Q, et al. Association of MYLIP rs3757354 SNP and several environmental factors with serum lipid levels in the Guangxi Bai Ku Yao and Han populations. *Lipids in health and disease.* 2012;11:141.

38. Zelcer N, Hong C, Boyadjian R, Tontonoz P. LXR regulates cholesterol uptake through Idol-dependent ubiquitination of the LDL receptor. *Science.* 2009;325:100-104.

39. Scotti E, Hong C, Yoshinaga Y, et al. Targeted disruption of the idol gene alters cellular regulation of the low-density lipoprotein receptor by sterols and liver x receptor agonists. *Molecular and cellular biology.* 2011;31:1885-1893.
40. Weissglas-Volkov D, Calkin AC, Tusie-Luna T, et al. The N342S MYLIP polymorphism is associated with high total cholesterol and increased LDL receptor degradation in humans. *The Journal of clinical investigation.* 2011;121:3062-3071.
41. Santos PC, Morgan AC, Jannes CE, Krieger JE, Santos RD, Pereira AC. The MYLIP p.N342S polymorphism is associated with response to lipid-lowering therapy in Brazilian patients with familial hypercholesterolemia. *Pharmacogenetics and genomics.* 2014;24:548-555.
42. Regitz-Zagrosek V, Kararigas G. Mechanistic Pathways of Sex Differences in Cardiovascular Disease. *Physiological reviews.* 2017;97:1-37.
43. Newton-Cheh C, Eijgelsheim M, Rice KM, et al. Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nature genetics.* 2009;41:399-406.
44. Straus SM, Kors JA, De Bruin ML, et al. Prolonged QTc interval and risk of sudden cardiac death in a population of older adults. *Journal of the American College of Cardiology.* 2006;47:362-367.
45. Noseworthy PA, Havulinna AS, Porthan K, et al. Common genetic variants, QT interval, and sudden cardiac death in a Finnish population-based study. *Circ Cardiovasc Genet.* 2011;4:305-311.
46. Holm H, Gudbjartsson DF, Arnar DO, et al. Several common variants modulate heart rate, PR interval and QRS duration. *Nature genetics.* 2010;42:117-122.
47. Andersen K, Rasmussen F, Held C, Neovius M, Tynelius P, Sundstrom J. Exercise capacity and muscle strength and risk of vascular disease and arrhythmia in 1.1 million young Swedish men: cohort study. *Bmj.* 2015;351:h4543.
48. Zeiher J, Ombrellaro KJ, Perumal N, Keil T, Mensink GBM, Finger JD. Correlates and Determinants of Cardiorespiratory Fitness in Adults: a Systematic Review. *Sports Med Open.* 2019;5:39.
49. Yang N, MacArthur DG, Gulbin JP, et al. ACTN3 genotype is associated with human elite athletic performance. *American journal of human genetics.* 2003;73:627-631.

Figure captions

Figure 1: The distribution of SNPs favoring high VO_{2max} in participants from the validation cohort. A; Mean VO_{2max} among participants in the validation cohort grouped according to number of favorable VO_{2max}-SNPs. **B**; Stratified analysis of inactive and active participants in the exploration cohort (based on a Kurtze-score, lower or higher than the median of the cohort). Group differences are calculated using one-way ANOVA with post hoc tests. Data are shown as mean and SEM. The number of individuals in each category is displayed inside the columns. SNPs: Single nucleotide polymorphisms, VO_{2max}: Maximal oxygen uptake, SEM: Standard error of mean.

Figure 2: The number of favorable SNPs in each participant associated with cardiovascular risk factors in the validation cohort. Data are shown as mean and SEM. SNPs: Single nucleotide polymorphisms, SEM: Standard error of mean.

Figure 3: *BAHD1* missense mutation at rs3803357. A: Distribution of VO_{2max} according to rs3803357 genotype in the *HUNT* cohort (n=2944 participants, 1445 men and 1500 women) displayed as mean and SE, *p<0.0005 between genotype TT and the two other genotypes, **B**: Distribution of VO_{2max} according to rs3803357 in the *Generation 100* cohort (n=718 participants, 328 men and 390 women) displayed as mean and SE, * p<0.005 between all three genotypes, **C**: Impact on rs3803357 on left ventricle gene expression in humans retrieved from the GTEx database (Homo ref=GG, Het=GT, Homo Alt=TT), **D**: Correlations between cardiac expression of BAHD1 in BDX mice strains and basal VO_2 (in an untrained state) and heart mass. *BAHD1*: Bromo adjacent homology domain containing 1, SNP: Single nucleotide polymorphism, VO_{2max}: Maximal oxygen uptake, SE: Standard error.

Figure 4: *MYLIP* upstream variant at rs3757354. A: Distribution of VO_{2max} among women according to rs3803357 genotype in the *Generation 100* cohort (n= 390) displayed as mean and SE, without adjustments for age and physical activity level (p<0.002), **B**: DNA sequence showing the A allele at rs3757354, which creates a perfect binding site for the estrogen receptor alpha (ER-α), **C**: Correlations between cardiac expression of MYLIP in BDX mice strains and heart mass. **D**: Waist circumference (cm), BMI, visceral fat (cm^2), fat percentage, HDL-cholesterol (mmol/l) and high-sensitivity CRP (mg/l) among women according to rs3757354 in the *Generation 100* cohort (n= 390) displayed as mean and SE, E: Percentage of women on cholesterol-lowering drugs according to rs3757354 in the *Generation 100* cohort (n= 390)

displayed as mean and SE. MYLIP: Myosin Regulatory Light Chain Interacting Protein, VO_{2max}: Maximal oxygen uptake, BMI: Body mass index, HDL: High-density lipoprotein, CRP: C-reactive protein.

Table 1. Participant characteristics

	HUNT3 Fitness **Exploration cohort**	***Generation 100*** **Validation cohort**
N	3470	718
Females, No. (%)	1563 (45 %)	390 (54.3 %)
Males, No. (%)	1907 (55 %)	328 (45.7 %)
Age, years	47 (19-84)	73 (70-77)
Height, cm	173 (148-200)	171 (147-195)
Weight, kg	78 (39-135)	75 (42-132)
Body mass index, kg/m^2	26 (17-44)	25 (17.5-42)
Waist circumference, cm	90 (57-134)	93 (68-129)
Total body fat (%)	-	29.1 (10.5-50.1)
Visceral fat (cm^2)	-	111 (26-270)
Systolic blood pressure, mmHg	126 (79-190)	135 (90-203)
Diastolic blood pressure, mmHg	72 (36-117)	74 (47-104)
Diabetes, No. (%)	48 (1.4 %)	17 (2.4 %)
Smoking status, No. (%)		
Never	1800 (51.9 %)	359 (50 %)
Current	458 (13.2 %)	51 (7.1 %)
Former	1173 (33.8 %)	269 (37.5 %)
Unknown	39 (1.1 %)	39 (5.4 %)
Biochemical data		
Total cholesterol, mmol/l	5.4 (2.3-10.0)	5.6 (2.6-9.9)
LDL cholesterol, mmol/l	-	3.4 (0.95-5.98)
HDL cholesterol, mmol/l	1.4 (0.5-3.4)	1.7 (0.6-4.5)
Triglycerides, mmol/l	1.8 (0.2-7.5)	1.1 (0.34-3.9)
Glucose, mmol/l	5.4 (2.6-12.4)	5.6 (3.1-10.5)
High-sensitivity CRP, mg/l	-	2.0 (0.1-36.6)
Treadmill data		
VO_{2max}, ml/kg$^{0.75}$/min	123 (56-222)	94 (42-161)
Resting heart rate, beats/min	59 (34-120)	64 (40-101)
Peak heart rate, beats/min	181 (107-231)	159 (96-203)
Physical activity level, No. (%)		
Inactive	421 (12.1 %)	52 (7.2 %)
Low	761 (21.9 %)	110 (15.3 %)
Medium	1274 (36.7 %)	350 (48.8 %)
High	991 (28.6 %)	193 (26.9 %)
Unknown	23 (0.7 %)	13 (1.8 %)

Data is shown as mean (min-max) or as mean (percentage of the participants). LDL: Low-density lipoprotein, HDL: High-density lipoprotein, CRP: C-reactive protein, VO_{2max}: Maximal oxygen uptake.

Table 2: SNPs associated with VO_{2max} both in the exploration cohort (HUNT) and in the validation cohort (Generation 100)

						Exploration cohort				Validation cohort					
SNP	**Chr.**	**Proximal gene**	**SNP location**	**High VO_{2max} genotype**	**Minor allele**	**MAF**	**BETA**	**STAT**	**p-value**	**MAF All**	**BETA All**	**STAT All**	**p-value All**	**p-value Women**	**p-value Men**
rs10494973	1	*PROX1*	Intron	GG	C	0.03	-1.01	-1.97	$2.6 *10^{-4}$	0.03	-1.91	-1.97	0.04	NS	NS
rs5370	6	*EDN1*	Missense	TT/TG	T					0.21	0.77	1.90	NS	0.001	NS
rs3757354	6	*MYLIP*	Upstream 2KB	GG	A	0.23	-0.74	-1.33	$2.8*10^{-4}$	0.25	-0.13	-0.33	NS	0.001	NS
rs6950857	7	*VIPR2*	Intron	GG	A	0.03	-4.68	-3.48	$7.2*10^{-5}$	0.04	-1.47	-1.61	0.002	NS	NS
rs4994	8	*ADRB3*	Missense	CC/CT	C					0.07	1.32	1.97	0.04	NS	0.005
rs1815739	11	*ACTN3*	NC transcript	CC/CT	C					0.45	-0.73	-2.14	0.03	NS	NS
rs2074238	11	*KCNQ1*	Intron	TT	T	0.07	1.75	1.938	$2.0*10^{-4}$	0.07	1.51	2.35	0.02	NS	0.03
rs3803357	15	*BAHD1*	Missense	GG/GT	T	0.48	-1.53	-3.30	$1.7*10^{-5}$	0.49	-0.90	-2.61	0.002	NS	0.03
rs16946588	17	*MYOCD*	Intron	AA	G	0.09	-1.55	-1.90	$4.8*10^{-5}$	0.09	-0.43	-0.71	0.04	NS	NS

SNPs: Single-nucleotide polymorphisms, VO_{2max}: Maximal oxygen uptake, Chr.: Chromosome, MAF: Minor allele frequency, PROX1: Prospero homeobox protein 1, EDN1: Endothelin 1, MYLIP: Myosin regulatory light chain interacting protein, VIPR2: Vasoactive intestinal peptide receptor 2, ADRB3: beta-3 adrenergic receptor, ACTN3: Alpha-actinin-3, KCNQ1: potassium voltage-gated channel subfamily Q member 1, BAHD1: Bromo adjacent homology domain containing 1, MYOCD: myocardin, NS: Not significant

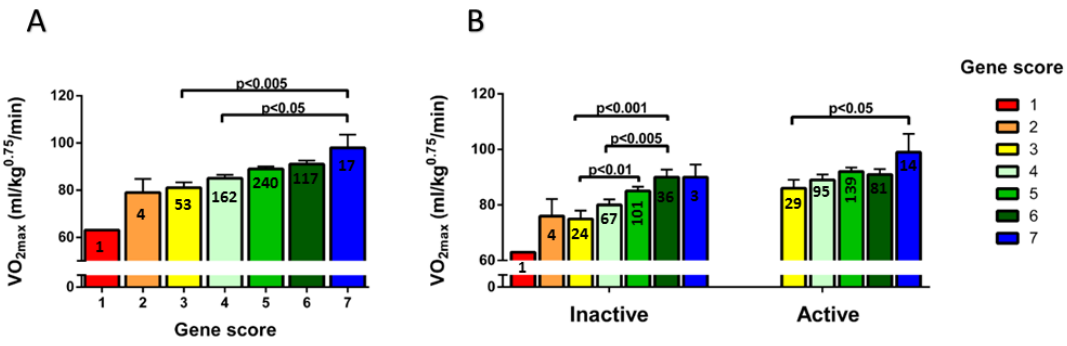
A
B
p<0.005
p<0.05
p<0.001
p<0.005
p<0.01
p<0.05
VO2max (ml/kg0.75/min)
Gene score
Inactive
Active
1
4
53
162
240
117
17
1
4
24
67
101
36
3
29
95
139
81
14

Figure 1

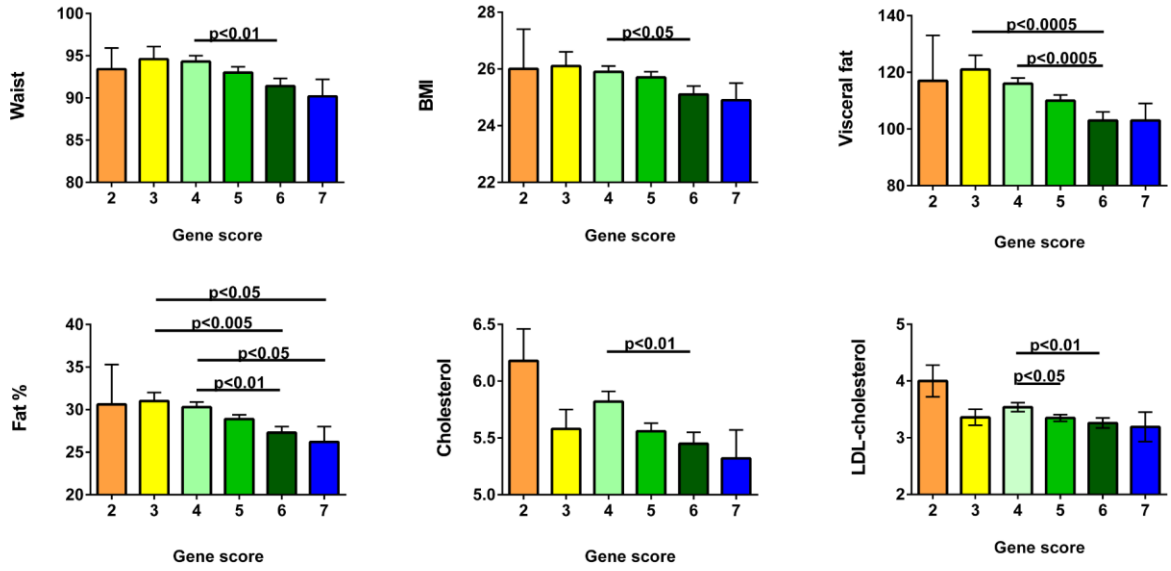
Waist
p<0.01
Gene score
BMI
p<0.05
Gene score
Visceral fat
p<0.0005
p<0.0005
Gene score
Fat %
p<0.05
p<0.005
p<0.05
p<0.01
Gene score
Cholesterol
p<0.01
Gene score
LDL-cholesterol
p<0.01
p<0.05
Gene score

Figure 2

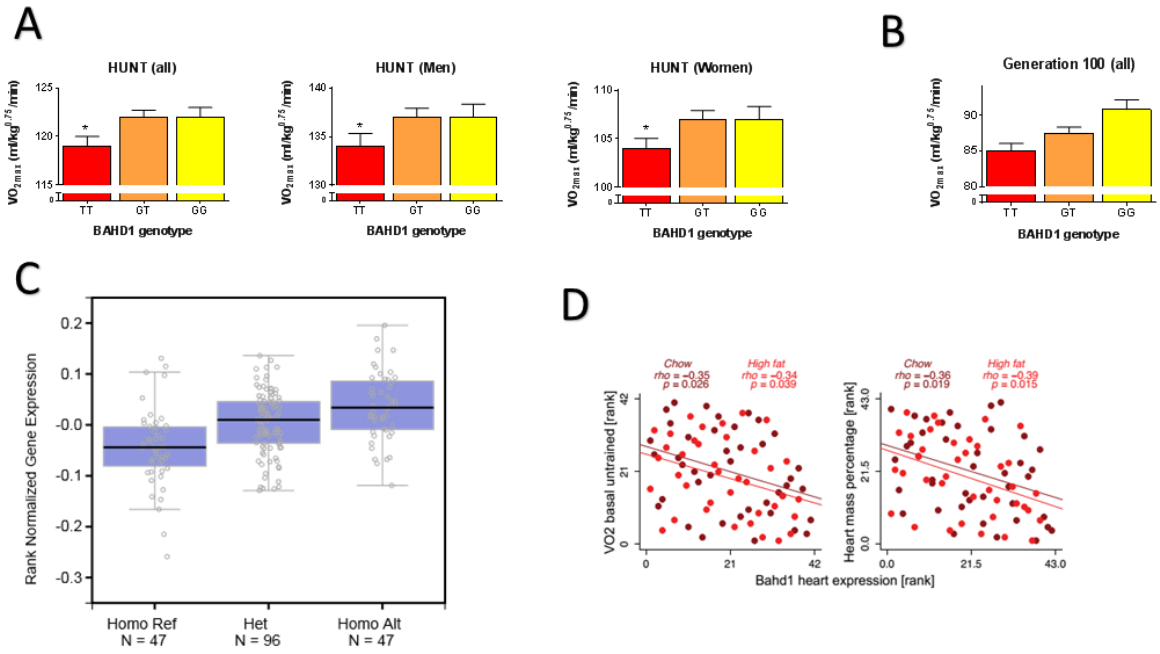
A
HUNT (all)
HUNT (Men)
HUNT (Women)
BAHD1 genotype
TT
GT
GG
B
Generation 100 (all)
C
Rank Normalized Gene Expression
Homo Ref
N = 47
Het
N = 96
Homo Alt
N = 47
D
Chow
rho = -0.35
p = 0.026
High fat
rho = -0.34
p = 0.039
VO2 basal untrained [rank]
Chow
rho = -0.36
p = 0.019
High fat
rho = -0.39
p = 0.015
Heart mass percentage [rank]
Bahd1 heart expression [rank]

Figure 3

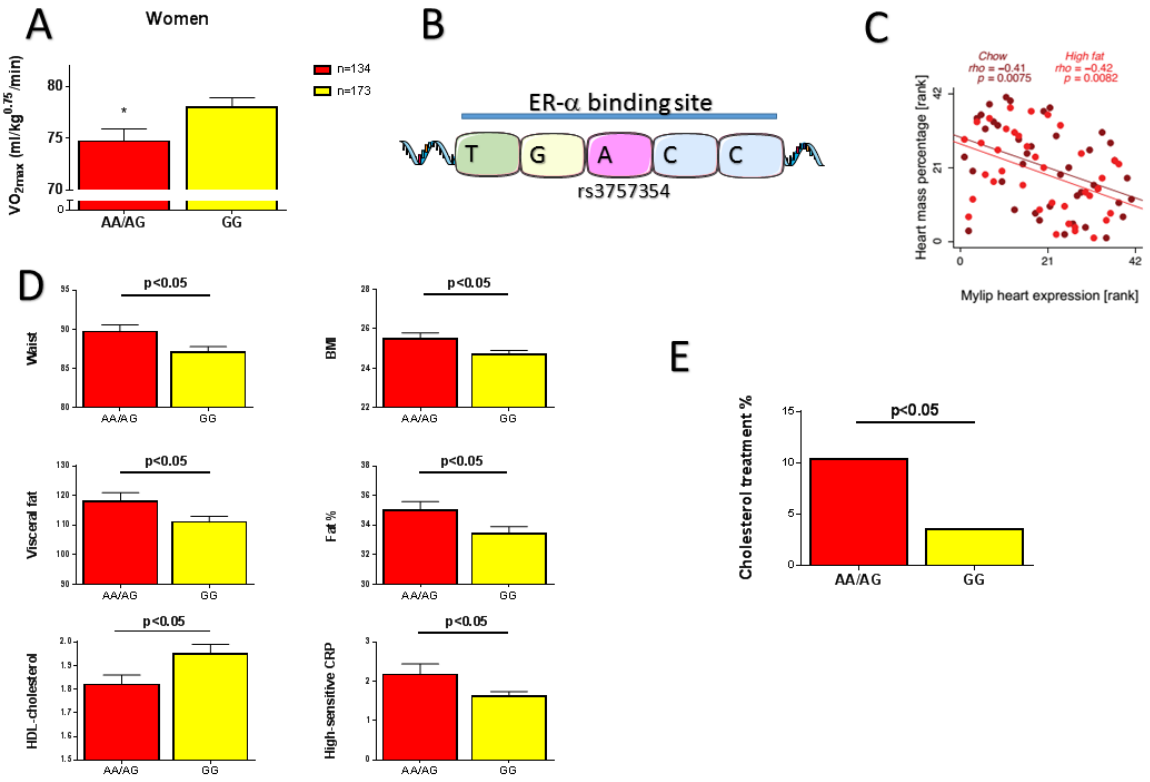
A
Women
n=134
n=173
AA/AG
GG
B
ER-α binding site
T G A C C
rs3757354
C
Chow
rho = -0.41
p = 0.0075
High fat
rho = -0.42
p = 0.0082
Heart mass percentage [rank]
Mylip heart expression [rank]
D
Waist
BMI
Visceral fat
Fat %
HDL-cholesterol
High-sensitive CRP
p<0.05
E
Cholesterol treatment %

Figure 4