APPENDIX

Fitness, Physical Activity, and Cardiovascular Disease: Longitudinal and Genetic Analyses in the UK Biobank Study

Emmi Tikkanen, Stefan Gustafsson, Erik Ingelsson

Table of Contents

Methods

Study sample

In 2006-2010, over 500,000 individuals aged 40-69 years were enrolled into the UK Biobank, a longitudinal cohort study based in the UK. Participants have undergone a range of physical measurements, detailed assessments about health-related factors, and sampling of blood, urine and saliva. The participants have also agreed to have their future health, including disease events, monitored. In our study, we utilized the data collected at the UK Biobank assessment centers at baseline, combined with the information on incident disease events from the hospital and death registry. After excluding individuals who had withdrawn consent at the time of the study and prevalent CVD events (N=17,717), 484,918 individuals remained in our study sample for observational analyses of CVD. In addition, 2,531 individuals reported too high or low reported values for physical activity variables according to data cleaning rules of IPAQ data[1], and these were removed in analyses involving physical activity data. For analyses of CRF, we utilized a subset of 66,652 individuals free from CVD at the baseline that underwent a submaximal exercise test on a treadmill. In addition, we also analyzed a subset of 100,843 individuals with objectively measured physical activity with a wrist-worn accelerometer. To evaluate the gene-environment interaction effects of fitness and physical activity on disease incidence, we used 146,541 individuals with genome-wide genetic data available (5,705 prevalent CVD cases removed). Finally, we performed genome-wide association studies (GWAS) of grip strength and physical activity in a subset of 120,285 European individuals with genetic data. The UK Biobank study was approved by the North West Multi-Centre Research Ethics Committee and all participants provided written informed consent to participate in the UK Biobank study. The study protocol is available online[2].

Baseline data

In this study, the exposures of interest were different measures of fitness and physical activity (grip strength, total physical activity and cardiorespiratory fitness [CRF]). Grip strength was measured in a sitting position using a Jamar J00105 hydraulic hand dynamometer. The participants were asked to squeeze the device as hard as they could for three seconds, and the maximum value that was reached during that time was recorded. Both hands were measured in turn (UK Biobank field ID 46 for left and 47 for right hand). In line with prior studies[3, 4], to adjust for confounding of strength by body mass, we calculated relative grip strength as an average of measurements of right and left hand divided by weight (ID 21002). Physical activity was assessed with a short form IPAQ questionnaire[1], which includes six questions of frequency (IDs 864, 884, and 904) and duration (IDs 874, 894 and 914) of walking, moderate-intensity and vigorous exercise. The answer "Unable to walk" in 864 was recoded to 0 and "Prefer not to answer" and "Do not know" in all six variables were set missing. Objective assessment of physical activity was measured for a 7-day period using Axivity AX3 wrist-worn triaxial accelerometer. The non-wear time was detected and imputed by the expert working group and total physical activity was calculated by averaging all worn and imputed values[5]. CRF was assessed with net oxygen consumption (VO_2), calculated from individuals' body weight and maximum workload (ID 6032) during the cycle ergometry on a stationary bike (eBike, Firmware v1.7), with using the equation VO2=7+10.8(workload)/weight[6].

In addition, we used information of potential confounders, specifically age (field ID 21022), sex (ID 31), region of the UK biobank assessment center (ID 54; recoded to three groups: UK, Scotland and Wales), ethnicity (ID 21000; recoded to four groups: white, black, Asian, mixed), Townsend index reflecting socioeconomic status (ID 189), smoking status (ID 20116; current, former, never), body mass index (ID 21001), diabetes (ID 2443), lipid medication (ID 20003; including following medications: simvastatin, pravastatin, fluvastatin, atorvastatin, rosuvastatin, ezetimibe, nicotinic acid product or fenofibrate), systolic blood pressure (ID 4080, but if missing ID 93), and height (ID 50) as covariates in our models. The details of these measurements can be found in the study protocol[2].

Genomic data

Genome-wide genotyping was conducted with the UK BiLEVE and UK Biobank Axiom arrays including 820,967 genetic markers in Affymetrix Research Services Laboratory, Santa Clara, California, USA. The quality control (QC) and imputation of the data were carried out centrally and has been described in the study protocol [2]. In short, the data consisted of 152,736 samples and 806,466 single nucleotide polymorphisms (SNPs) after sample and marker exclusions in the QC. Imputation was performed with IMPUTE2 by using 1000 Genomes Phase 3 merged with the UK10K haplotype reference panel. In our analyses, we included unrelated individuals with self-reported British descent

and European/Caucasian ethnicity based on principal component analysis (N=120,285). We included genetic markers with MAC ≥ 20 and PLINK info > 0.8, calculated in the ~120k sample.

Outcomes and follow-up

The disease outcomes were defined as primary events using in-patient hospital and death registry data that have been linked to the UK Biobank. CHD was defined as International Classification of Diseases (ICD) edition 9 codes 410-411, edition 10 codes I20.0, I21, and I22, and surgical codes for percutaneous transluminal coronary angioplasty and coronary artery bypass graft (codes K40-K46, K49-K50, and K75). Stroke was defined as ischemic (ICD-9: 433-434, ICD-10: I63) or hemorrhagic stroke (ICD-9: 430-432, ICD-10: I60-I62). Heart failure was defined as ICD-9 code 428 and ICD-10 code I50. Atrial fibrillation was defined as ICD-9 code 427.3, ICD-10 code I48, and surgical codes K50.1, K62.2-K62.4. The hospital registry -based follow-up ended on March 31, 2015 in England, August 31, 2014 in Scotland, and 28 February, 2015 in Wales. Individuals were censored on these dates, time of event in question or the time of death, whichever occured first. Death due to CVD was defined using the same ICD-10 codes for different endpoints from the death registry. Death registry included all deaths that occured before January 31, 2016 in England and Wales, and November 30, 2015 in Scotland.

Statistical methods

Imputation

Missing values of the baseline data were imputed with multivariate imputation by chained equation (MICE) by using predictive mean matching[7]. By using all variables of the final analysis model (frequency and duration of exercise, grip strength, BMI, smoking, lipid medication, systolic blood pressure, diabetes, height and Townsend index), Nelson-Aalen estimate of cumulative hazard, and the event indicator as the input, we selected predictors for each variable with missing values by using *quickpred* function from *mice* package in R. This function computes predictor matrix for each variable based on: 1) correlations between observed values of the variable of interest and other variables; and 2) correlations between an indicator of missingness of the variable of interest and other variables. We performed five repetitions of imputations. The imputed values were compared with the observed values to evaluate the performance of the imputation. We then performed data quality control for frequency and duration variables and calculated total physical activity ("IPAQ-PA") as MET-hours per week according to IPAQ scoring protocol[1]. We did not perform imputation for the CRF and acceleration variables.

Observational analysis

Associations between measures of fitness and physical activity, and CVD events were analyzed using Cox proportional hazards models. The distributions of subjective (IPAQ) and objective (accelerometer) measures of physical activity were skewed, whereas the distributions of grip strength and CRF were approximately normal (**Figure S1**). Thus, to facilitate comparison between effects of different measures, physical activity measures were first rank transformed and then, all measures were scaled to standard normal distribution. Analyses were conducted separately for CHD, ischemic and hemorrhagic stroke, heart failure and atrial fibrillation (AF), as well as for combined CVD events. In secondary analyses, we also analyzed associations with all-cause death. Accelerometer data was used for all-cause death analysis only, due to short follow up (data was collected from May 2013 until Dec 2015). For each endpoint, we ran three sets of multivariable-adjusted models: a) adjusting for age, sex and region of the UK Biobank assessment center; b) additional adjustment for possible confounders[8] including ethnicity, BMI, smoking, lipid medication, systolic blood pressure, diabetes, height and Townsend index; and c) adjusting for IPAQ-PA and/or grip strength in addition to those in b). Proportional hazards assumption was assessed using Schoenfeld's test, and when not fulfilled ($P \leq 0.001$), we added interaction terms with time for those covariates for which proportional hazards assumption was not met. In addition, we stratified all models by region to allow different baseline hazard function for each stratum. All analyses were conducted separately for five imputed datasets and results were pooled with Rubin's rule[7].

Analyses of interactions between fitness and physical activity and genetic determinants of CHD

Next, we evaluated the risk-modifying effects of fitness and physical activity in individuals with different genetic risk load for CHD and AF. First, we calculated a genetic risk scores (GRSs) for CHD and AF representing joint effects of individual and independent genetic markers. The genetic markers were selected from the largest published GWAS for CHD[9] and AF[10], and the GRS was calculated as the weighted sum of the risk alleles by using effect sizes from the reference GWAS[9, 10] as

weights. The GRS was then divided into tertiles to stratify individuals into high, intermediate and low genetic risk category. Similarly, we stratified grip strength, IPAQ-PA and CRF into tertiles to compare hazard ratios for subjects in different groups. Further, to evaluate whether there was an interaction between exercise traits and genetic risk of CHD, we added interaction terms between the measures of fitness and physical activity, and the GRS. The models were adjusted for age, sex, ethnicity, genotype array (ID 22000; two levels UK BiLEVE and Axiom) and 15 principal components (ID 22009) and stratified by region of the UK Biobank assessment center.

Genome-wide association analysis (GWAS)

The discovery analyses of genetic variants associated with grip strength and IPAQ-PA were conducted in 80,000 individuals, randomly sampled from the genomics dataset. We did not perform GWAS of CRF or accelerometer data due to limited sample size with both the phenotype and genotypes. Grip strength was analyzed as continuous trait and IPAQ-PA was analyzed as rank-transformed, continuous trait. Analyses were conducted with PLINK[11] (version 1.9) by using linear regression assuming additive model for association between phenotypes and genotype dosages. Age, sex, genotype array and 15 principal components were included as covariates. The remaining genomics data of 40,285 European descent individuals were used to replicate SNPs with genome-wide significance from the discovery analysis. Finally, we conducted a pooled analysis of discovery and replication datasets to discover additional genetic variants with suggestive evidence of association with fitness and physical activity, and for creation of instrumental variables for Mendelian randomization (MR) analyses.

Functional analysis

To provide some initial evidence regarding relevant pathways, tissues and causal genes involved in fitness and physical activity, we used DEPICT[12]. To optimize balance between specificity and power, DEPICT analysis was performed by including loci at a lower significance level ($P \leq 10^{-5}$) in the pooled analysis after pruning variants in high linkage disequilibrium ($r^2 > 0.05$). Further, we used GTEx portal[13] to correlate our genome-wide significant loci with tissue-specific gene expression levels. We searched for variants significantly associated with transcript levels in the surrounding region (cis-eQTLs) across all tissues, but with higher *a priori* interest in skeletal muscle, cardiovascular and adipose tissue.

Mendelian randomization

We performed two-sample MR, which estimates the causal effect by contrasting the SNP effects on the exposure with the SNP effects on the outcome in independent datasets. Genetic variants from the GWAS of grip strength were used as instrumental variables (IVs) and publicly available GWAS data for CHD[9] as an outcome. If the IV SNPs were not available in the outcome GWAS, we used proxies in high LD with the lead variants ($r^2 \geq 0.8$) defined using 1000 Genomes European sample data. The effect sizes of IV SNPs were standardized and the alleles from the exposure and outcome GWAS were harmonized to match the same effect allele. We used three methods to estimate causal effects; standard inverse-variance weighted (IVW) regression, as well as two robust regression methods, the median-based method, and Egger regression[14]. Consistency of the causal estimates across all SNPs was evaluated with heterogeneity statistics and Egger regression were used to assess horizontal pleiotropy. To evaluate potential mediating mechanisms, we clustered SNPs based on their associations with other traits by using hierarchical clustering with Euclidean distance and Ward method[15, 16], and applied MR in these subgroups of SNPs. Analyses were conducted with R-package *TwoSampleMR*[17]. Power for MR analyses was estimated with an online tool created by Burgess[18].

Analyses were conducted with R (version 3.3.0).

References

1. International Physical Activity Questionnaire. https://sites.google.com/site/theipaq/home (accessed Jan 10 2017).
2. UK Biobank. https://http://www.ukbiobank.ac.uk/ (accessed Jan 10 2017).
3. Choquette S, Bouchard DR, Doyon CY, Sénéchal M, Brochu M, Dionne I. Relative strength as a determinant of mobility in elders 67–84 years of age. A NuAge study: Nutrition as a determinant of successful aging. *J Nutr Health Aging* 2009: 1-6.
4. Lawman HG, Troiano RP, Perna FM, Wang CY, Fryar CD, Ogden CL. Associations of Relative Handgrip Strength and Cardiovascular Disease Biomarkers in U.S. Adults, 2011-2012. *American journal of preventive medicine* 2016; **50**(6): 677-83.
5. Doherty A, Jackson D, Hammerla N, et al. Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study. *PloS one* 2017; **12**(2): e0169649.
6. Swain DP. Energy cost calculations for exercise prescription: an update. *Sports Med* 2000; **30**(1): 17-22.
7. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC medical research methodology* 2009; **9**: 57.
8. 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.
9. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature genetics* 2015; **47**(10): 1121-30.
10. AFGen Consortium, Metastroke Consortium of the ISGC, Neurology Working Group of the Charge Consortium. Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nature genetics* 2017.
11. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 2015; **4**: 7.
12. Pers TH, Karjalainen JM, Chan Y, et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nature communications* 2015; **6**: 5890.
13. GTEx Portal. http://www.gtexportal.org/home/ (accessed Jan 17 2017).
14. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. *Epidemiology* 2017; **28**(1): 30-42.
15. Fall T, Xie W, Poon W, et al. Using Genetic Variants to Assess the Relationship Between Circulating Lipids and Type 2 Diabetes. *Diabetes* 2015; **64**(7): 2676-84.
16. Yaghootkar H, Scott RA, White CC, et al. Genetic evidence for a normal-weight "metabolically obese" phenotype linking insulin resistance, hypertension, coronary artery disease and type 2 diabetes. *Diabetes* 2014; **63**(12): 4369-77.
17. Gibran Hemani, Jie Zheng, Kaitlin H Wade, et al. MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations. (accessed.
18. Online sample size and power calculator for Mendelian randomization with a binary outcome. https://sb452.shinyapps.io/power/ (accessed Jan 10 2017).

Figures

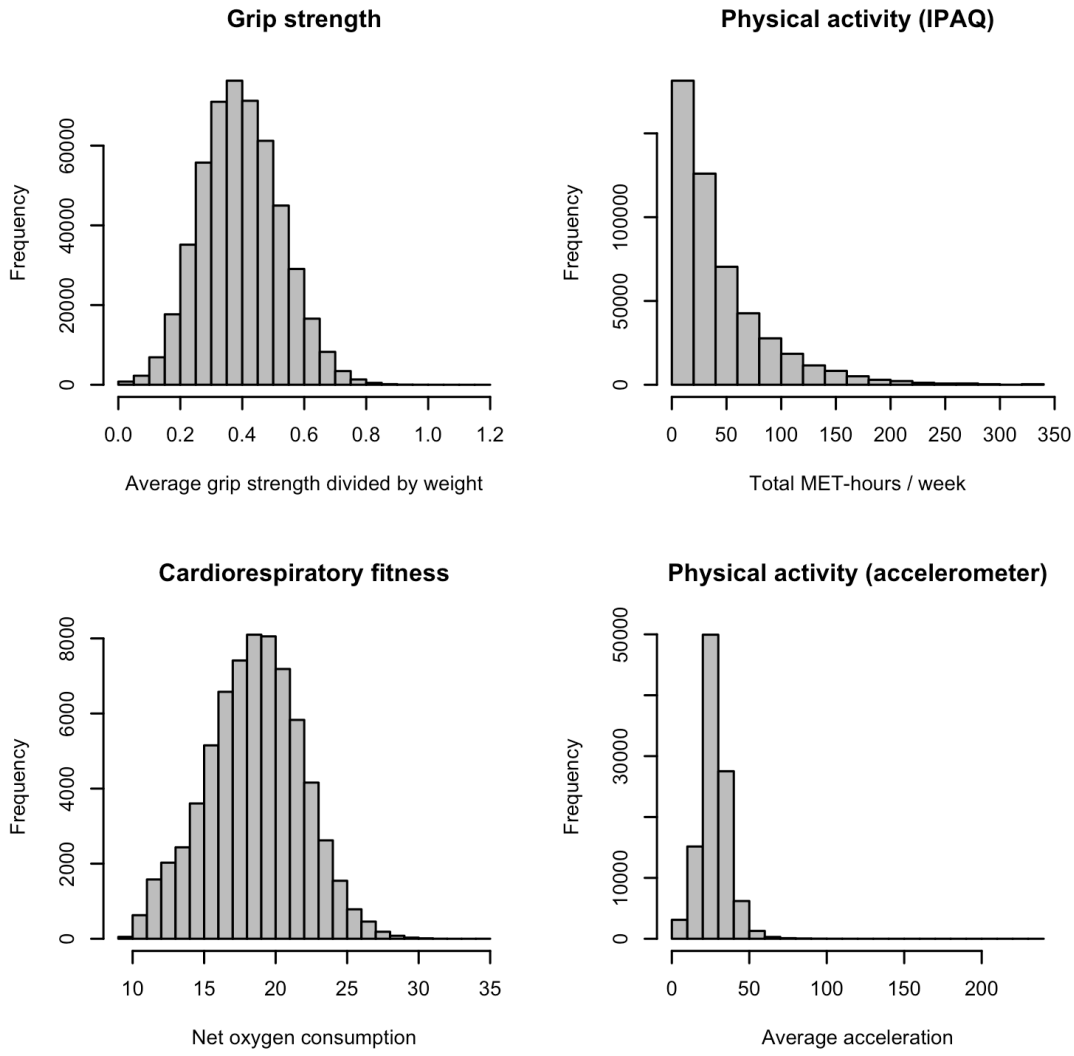

Figure S1. Distributions of fitness and physical activity variables.

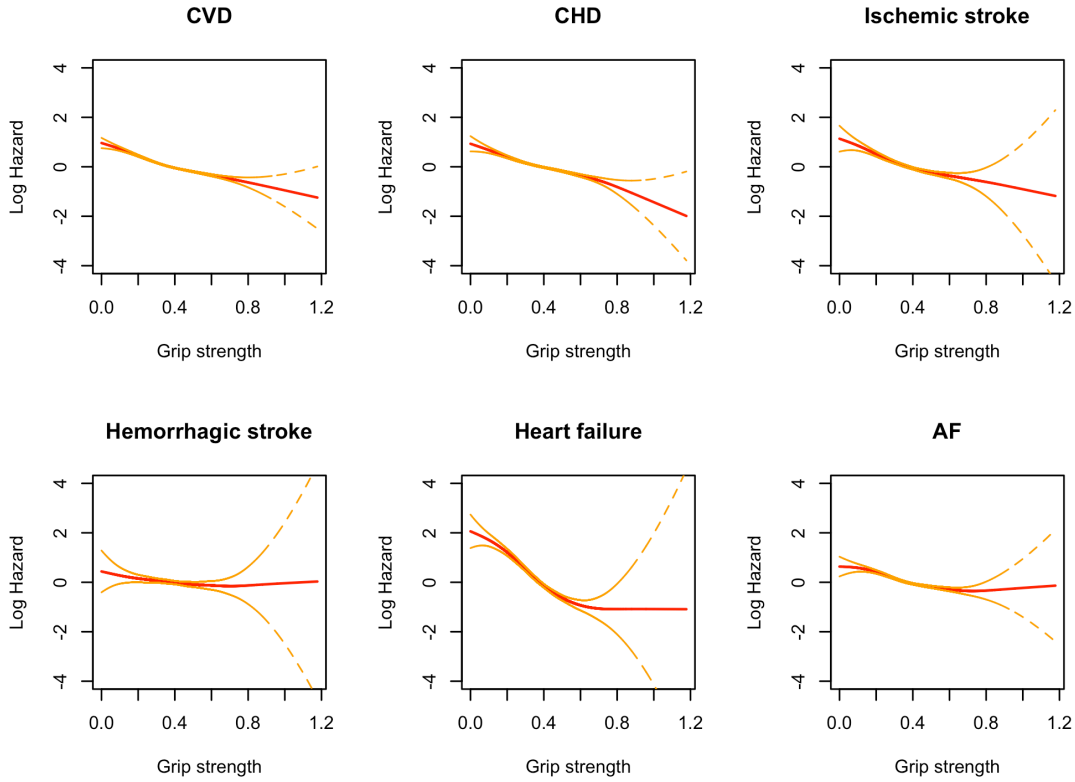

Figure S2. Relations of grip strength and cardiovascular disease (CVD) events. Lines are based on a regression spline of Cox proportional hazards. CHD, coronary heart disease; AF, atrial fibrillation.

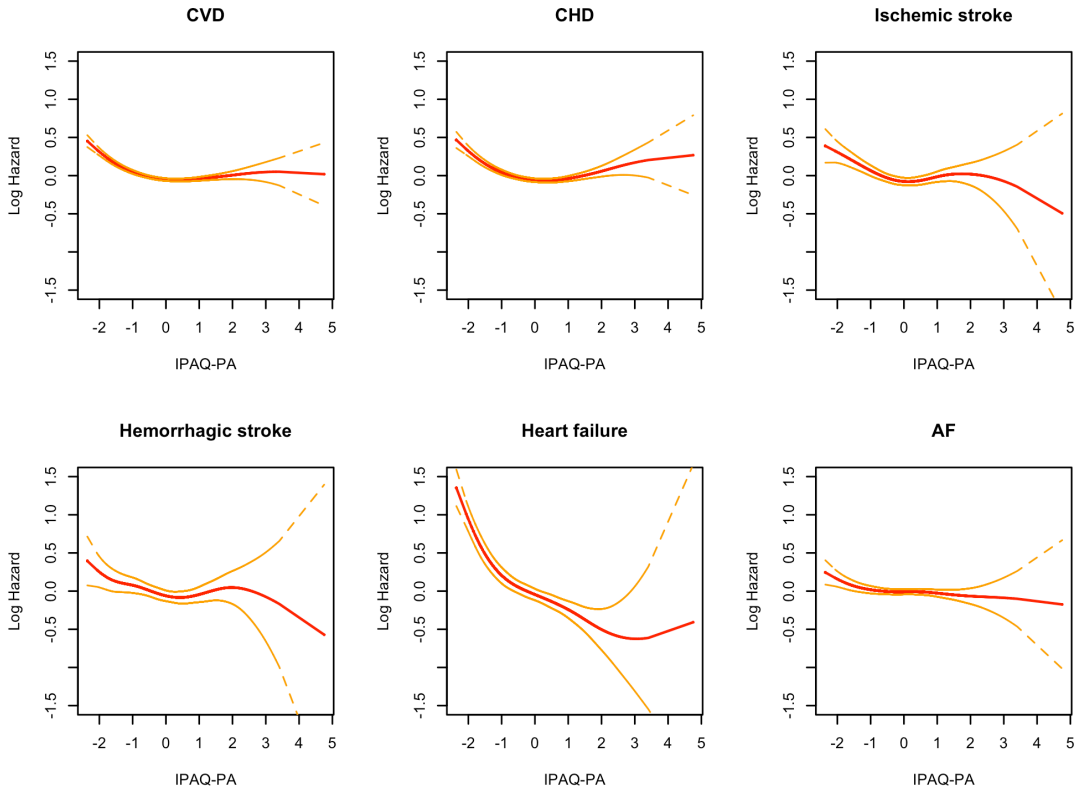

Figure S3. Relations of questionnaire-based physical activity (IPAQ-PA) and cardiovascular disease (CVD) events. Lines are based on a regression spline of Cox proportional hazards. CHD, coronary heart disease; AF, atrial fibrillation.

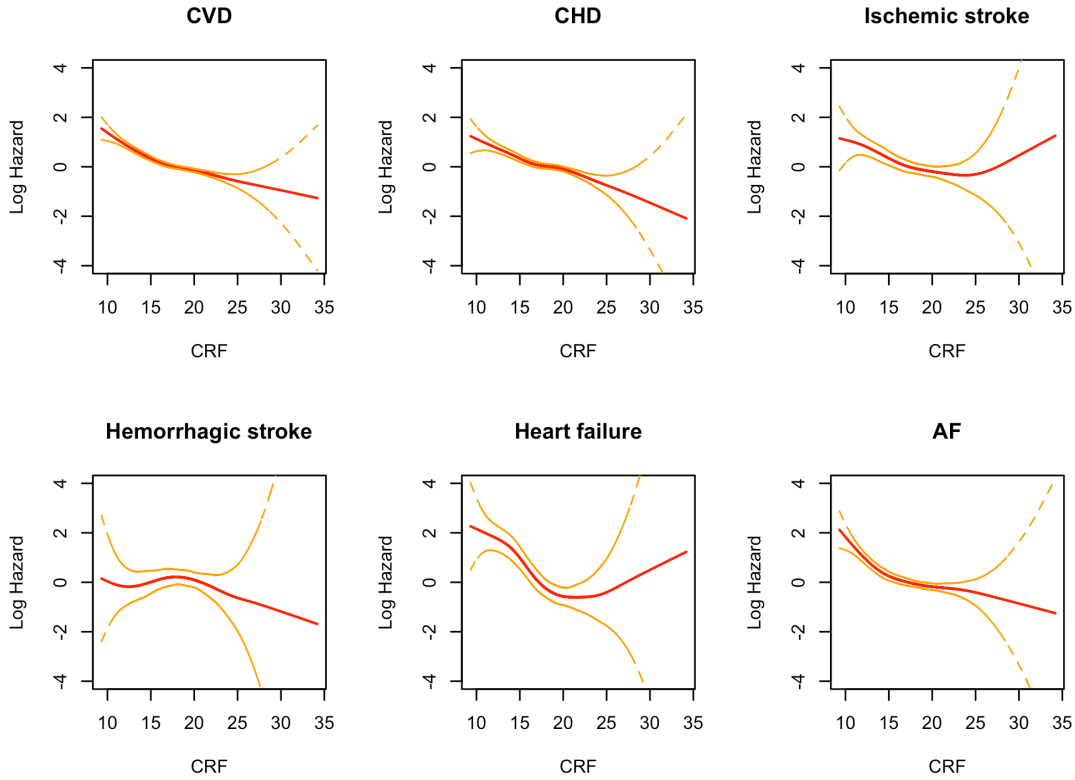

Figure S4. Relations of cardiorespiratory fitness (CRF) and cardiovascular disease (CVD) events. Lines are based on a regression spline of Cox proportional hazards. CHD, coronary heart disease; AF, atrial fibrillation.

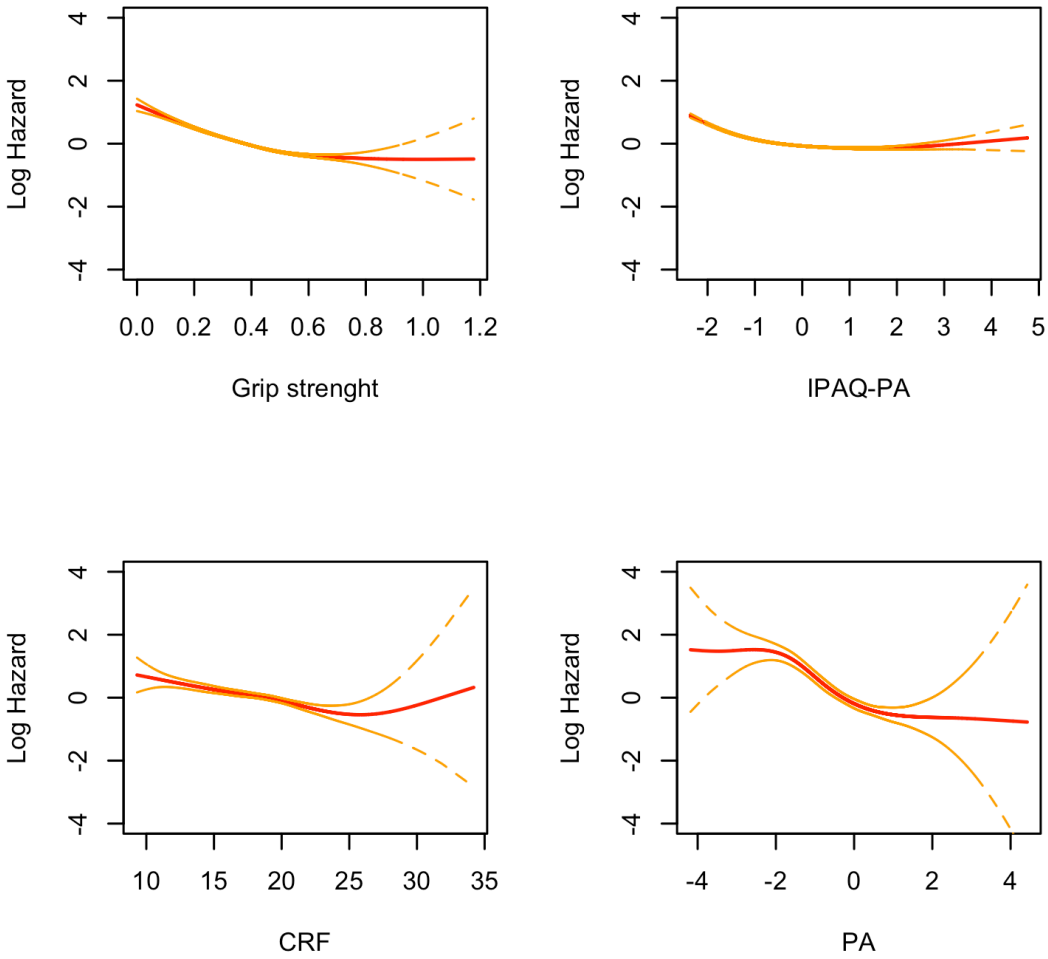

Figure S5. Relations of grip strength, questionnaire-based (IPAQ-PA) and objective (PA) physical activity, and cardiorespiratory fitness (CRF), and all-cause death. Lines are based on a regression spline of Cox proportional hazards.

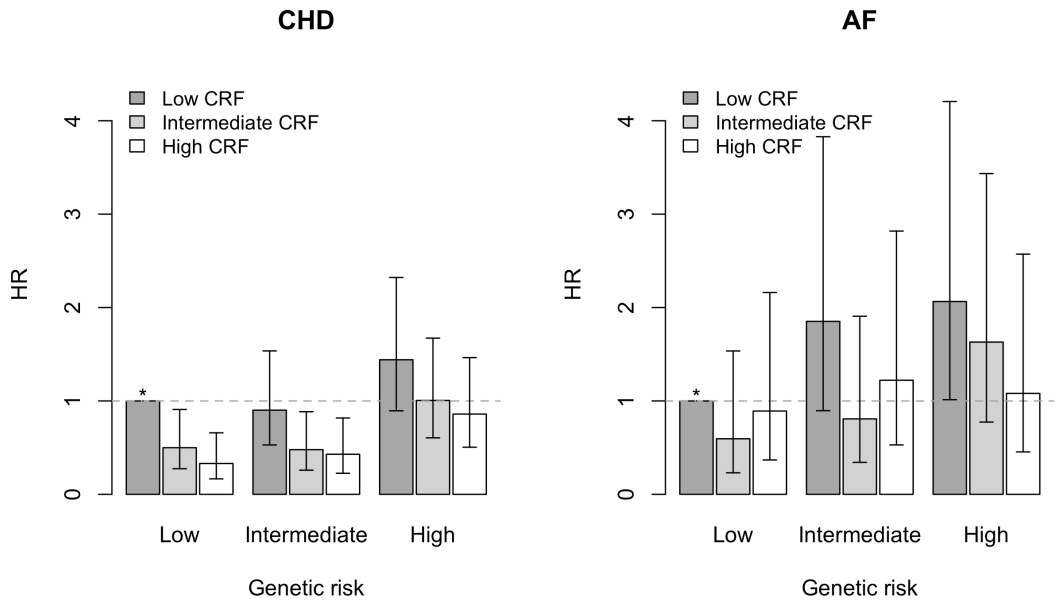

Figure S6. Hazard ratios with 95% confidence intervals for coronary heart disease (CHD) and atrial fibrillation (AF) according to tertiles of genetic risk and cardiorespiratory fitness (CRF).

*** Reference group.**

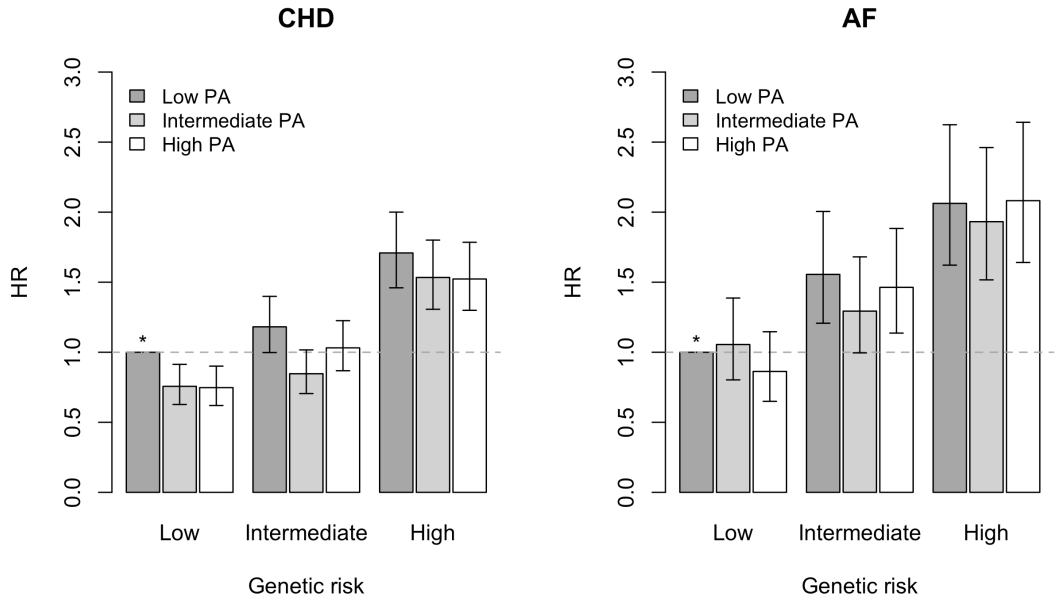

Figure S7. Hazard ratios with 95% confidence intervals for coronary heart disease (CHD) and atrial fibrillation (AF) according to tertiles of genetic risk and physical activity (PA).

*** Reference group.**

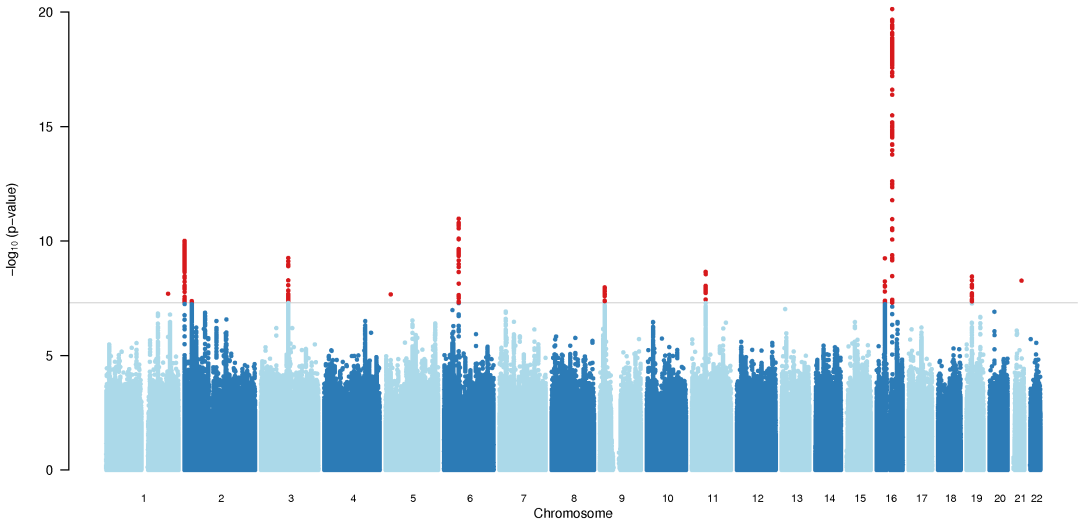

Figure S8. Manhattan plot for grip strength loci in discovery analysis.

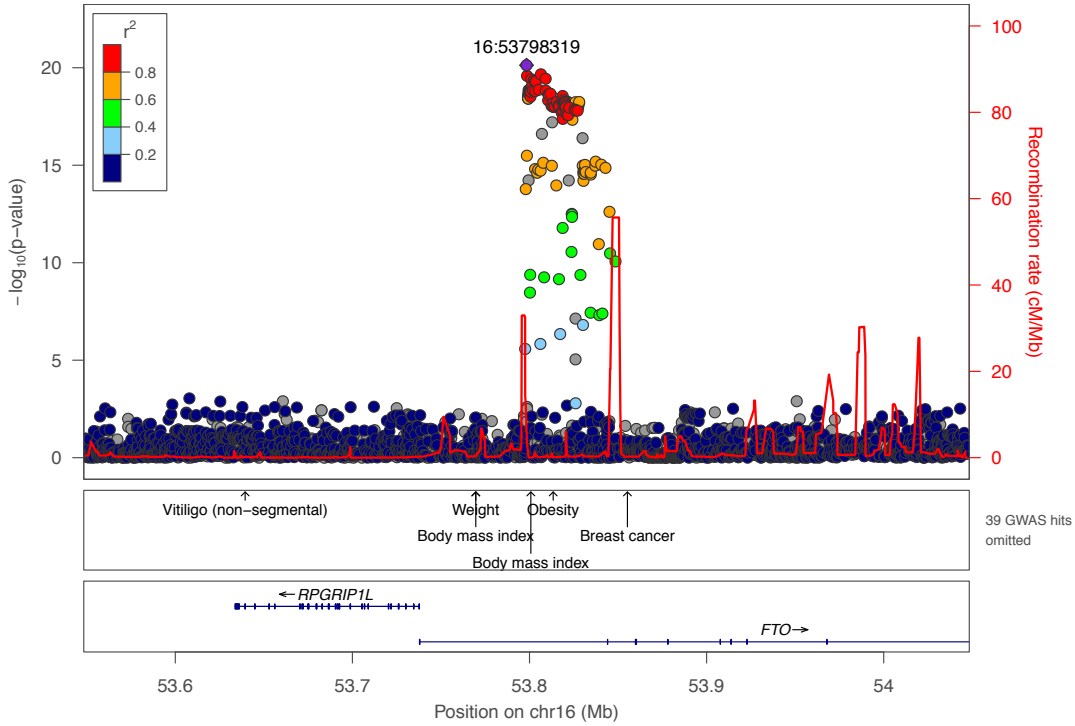

Figure S9. Locuszoom plot for FTO locus (top variant rs28429148).

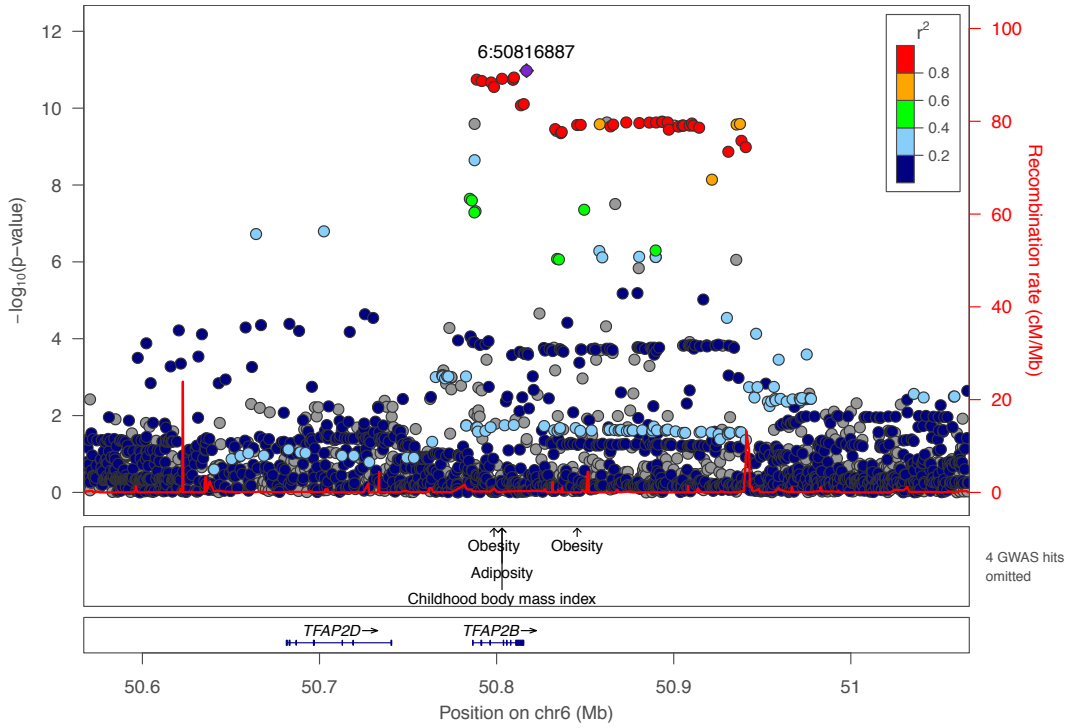

Figure S10. Locuszoom plot for TFAP2B locus (top variant rs72892910).

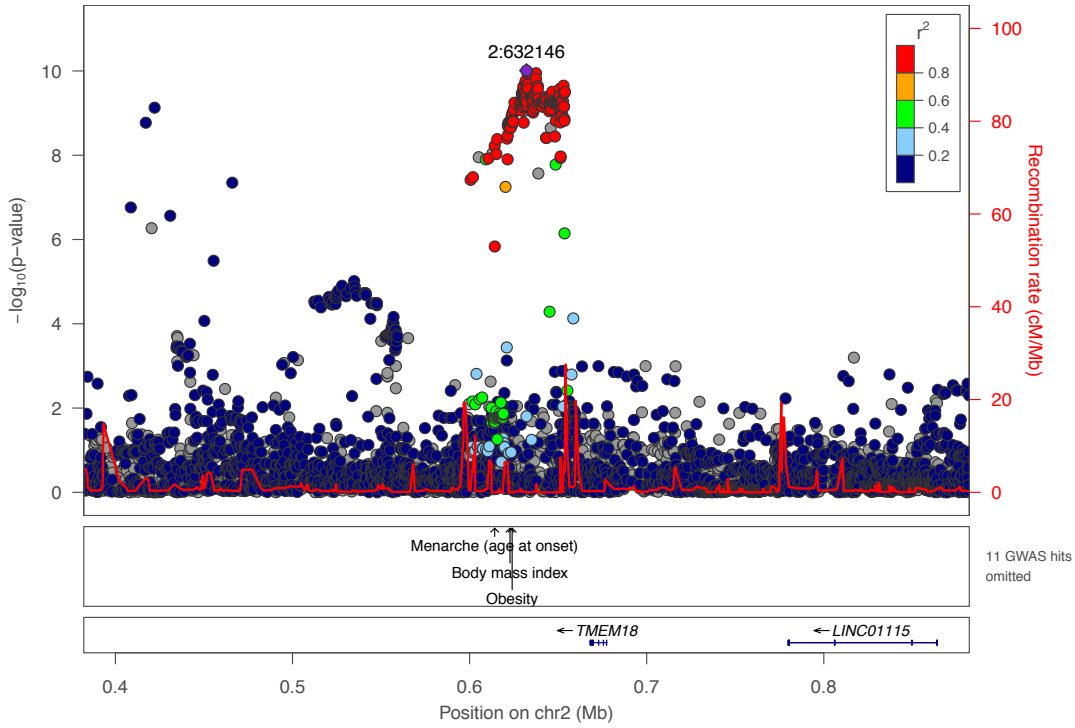

Figure S11. Locuszoom plot for TMEM18/ FAM150B locus (top variants rs12623218 and rs62107261).

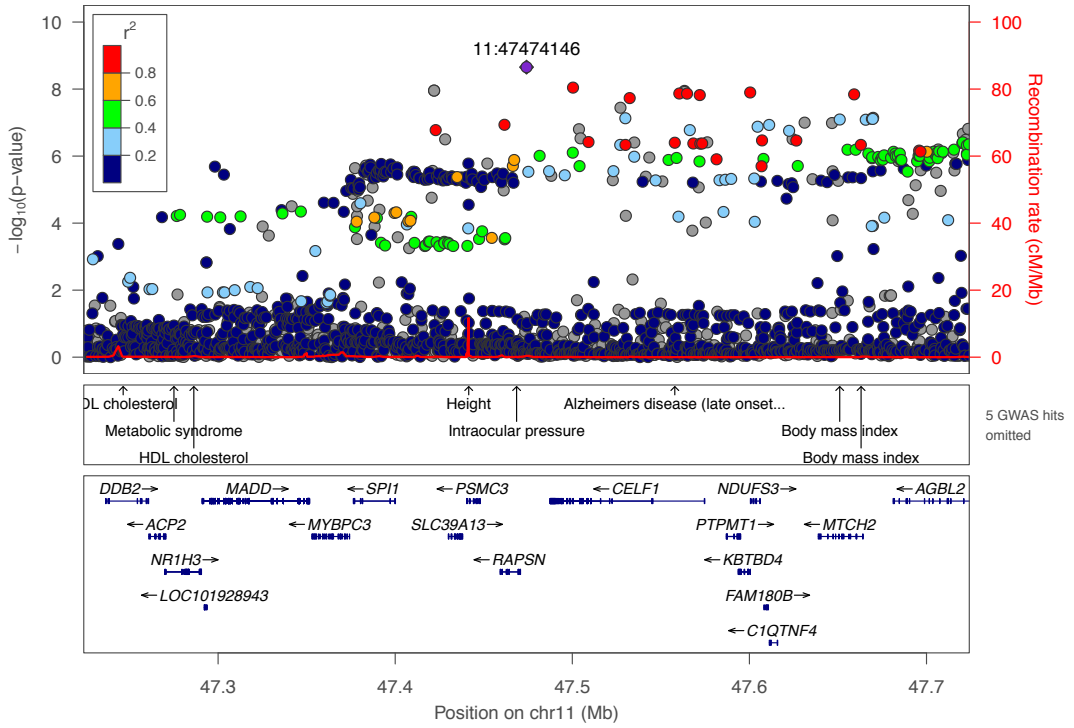

Figure S12. Locuszoom plot for RAPSN locus (top variant rs12361415).

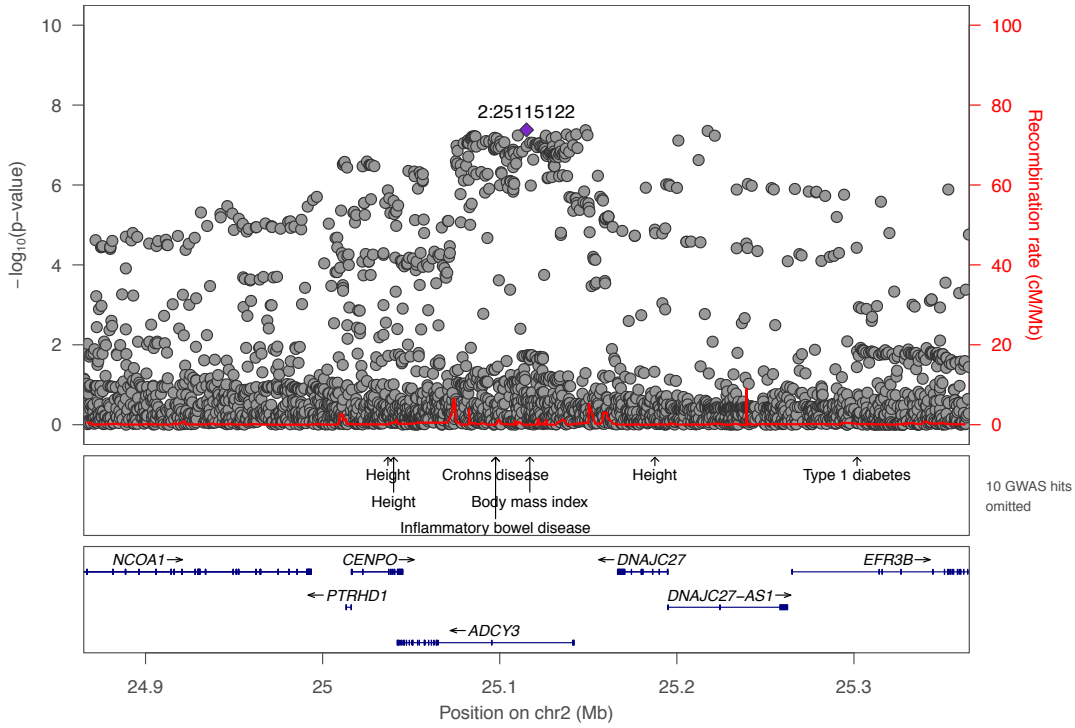

Figure S13. Locuszoom plot for ADCY3 locus (top variant rs556981345).

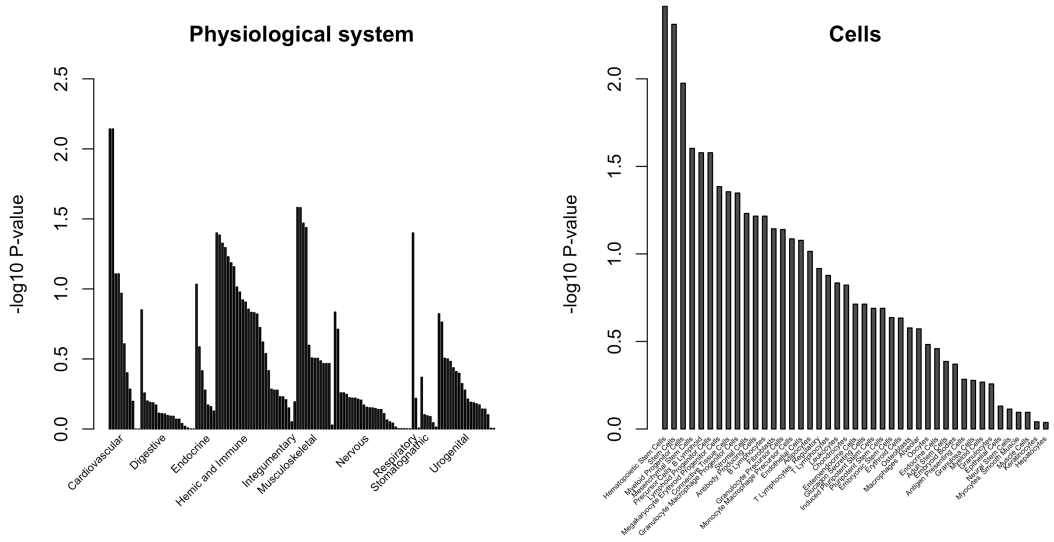

Figure S14. Tissue and cell enrichment for grip strength loci ($P \leq 1 \times 10^{-5}$).

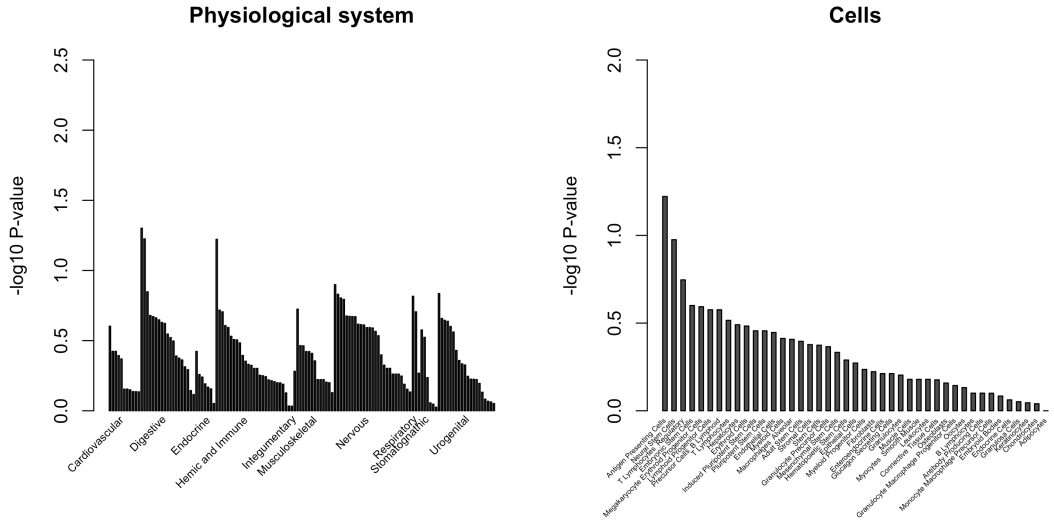

Figure S15. Tissue and cell enrichment for physical activity loci ($P \leq 1 \times 10^{-5}$).

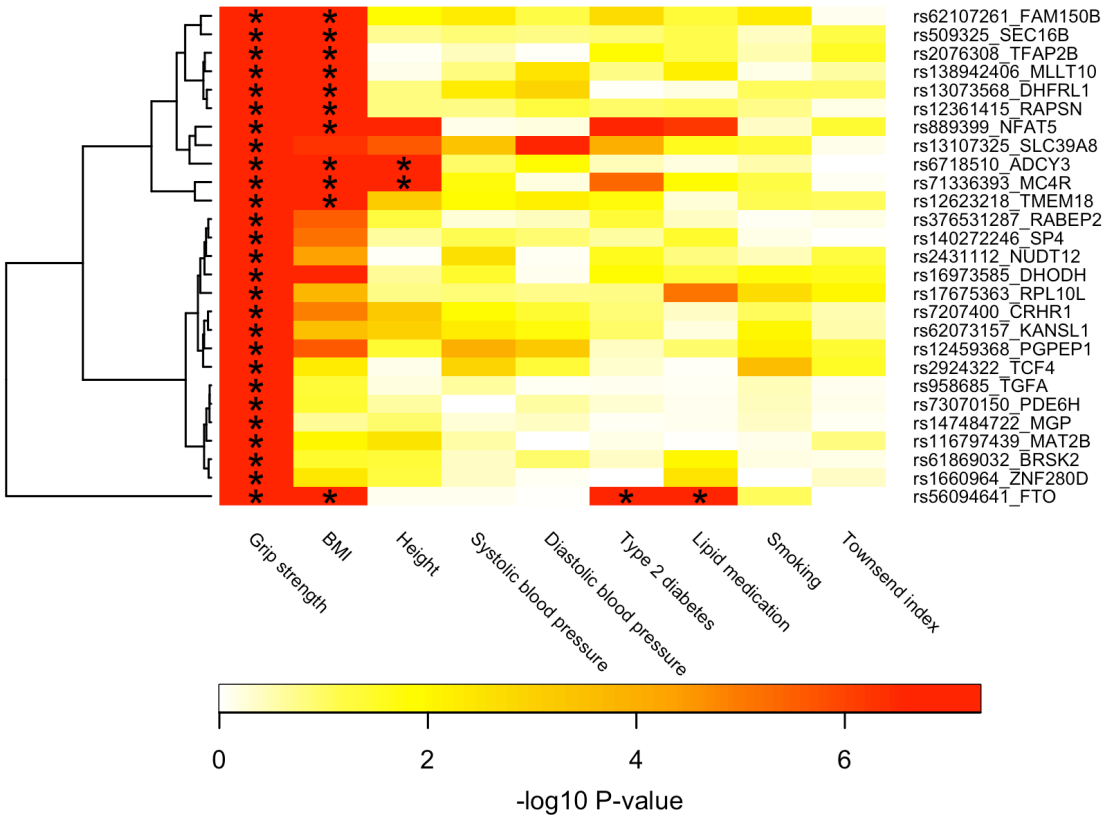

Figure S16. Associations of grip strength loci with other traits.

*** denotes genome-wide significant ($P \leq 5 \times 10^{-8}$) association.**

Tables

Table S1. Incidence and hazard rations for all-cause mortality by physical activity and fitness traits.

		All-cause death		
	Model	**N events**	**HR (95 % CI)**	**P-value**
Grip strength	1	14,419	0.75 (0.73 , 0.76)	<0.001
	2	14,419	0.76 (0.74 , 0.78)	<0.001
	3	14,350	0.78 (0.76 , 0.79)	<0.001
IPAQ-PA	1	14,350	0.83 (0.82 , 0.84)	<0.001
	2	14,350	0.86 (0.84 , 0.87)	<0.001
	3	14,350	0.87 (0.86 , 0.89)	<0.001
CRF	1	1,162	0.78 (0.72 , 0.83)	<0.001
	2	1,162	0.75 (0.69 , 0.81)	<0.001
	3	1,157	0.76 (0.70 , 0.83)	<0.001
PA	1	348	0.52 (0.46 , 0.58)	<0.001
	2	348	0.56 (0.50 , 0.63)	<0.001
	3	347	0.56 (0.50 , 0.63)	<0.001

The effects are in SD-units in fitness and physical activity traits. Model adjustments: 1. Age, sex and region. 2. Age, sex, region, diabetes, smoking, systolic blood pressure, body mass index, lipid medication, height, ethnicity and Townsend index. 3. All in 2) plus IPAQ-PA (grip strength analyses) or grip strength (IPAQ-PA analyses). Analyses of CRF and PA were adjusted for both IPAQ-PA and grip strength.
Abbreviations: IPAQ-PA, physical activity assessed by international physical activity questionnaire; CRF, cardiorespiratory fitness; PA, Physical activity assessed by wrist-worn accelerometer; HR, hazard ratio; CI, confidence interval.

Table S2. Genetic loci associated with fitness and physical activity in pooled analysis.										
SNP	**CHR**	**Position**	**Gene**	**EA**	**OA**	**EAF**	**Beta**	**Se**	**P-value**	**Info**
Grip strength										
rs56094641	16	53806453	FTO	A	G	0.5972	0.035	0.003	3.8×10^{-24}	0.998
rs138942406	10	21954234	MLLT10	G	A	0.7118	0.028	0.004	1.9×10^{-14}	0.993
rs62107261	2	422144	FAM150B	T	C	0.9521	-0.058	0.008	1.4×10^{-13}	0.994
rs2076308	6	50791640	TFAP2B	G	C	0.8211	0.032	0.005	1.9×10^{-13}	0.995
rs12623218	2	632146	TMEM18	T	A	0.1726	0.032	0.005	2.3×10^{-13}	0.984
rs12361415	11	47474146	RAPSN	T	G	0.7131	-0.025	0.004	7.4×10^{-12}	0.998
rs6718510	2	25122323	ADCY3	A	C	0.5500	0.023	0.003	9.7×10^{-12}	0.990
rs509325	1	177894591	SEC16B	T	G	0.7908	0.027	0.004	2.7×10^{-11}	0.996
rs13073568	3	94031839	DHFRL1	G	T	0.4847	-0.022	0.003	4.3×10^{-11}	0.989
rs16973585	16	72032231	DHODH	T	A	0.6827	-0.023	0.004	1.6×10^{-10}	0.996
rs71336393	18	57855319	MC4R	G	GA	0.7627	0.025	0.004	4.1×10^{-10}	0.985
rs2924322	18	53244414	TCF4	A	T	0.1194	0.035	0.005	6.2×10^{-10}	0.837
rs12459368	19	18459377	PGPEP1	A	G	0.7334	-0.023	0.004	1.0×10^{-9}	1.004
rs7207400	17	43824360	CRHR1	T	C	0.7032	0.022	0.004	1.2×10^{-9}	1.088
rs1660964	15	57158987	ZNF280D	T	A	0.2159	-0.025	0.004	1.3×10^{-9}	0.987
rs13107325	4	103188709	SLC39A8	C	T	0.9255	0.038	0.006	1.5×10^{-9}	0.995
rs116797439	5	163225051	MAT2B	T	A	0.9636	-0.054	0.009	6.3×10^{-9}	0.910
rs73070150	12	15162893	PDE6H	G	A	0.8453	-0.027	0.005	6.8×10^{-9}	0.979
rs62073157	17	44334052	KANSL1	G	A	0.7757	0.023	0.004	9.1×10^{-9}	1.010
rs140272246	7	21480486	SP4	ACAACAACAACAACAACAACAG	A	0.3669	-0.020	0.004	9.8×10^{-9}	0.983
rs61869032	11	1477797	BRSK2	C	T	0.7005	0.021	0.004	1.1×10^{-8}	0.995
rs889399	16	69556583	NFAT5	C	G	0.5901	-0.019	0.003	1.6×10^{-8}	0.980
rs376531287	16	28918483	RABEP2	G	GT	0.4612	-0.020	0.004	1.8×10^{-8}	0.883

SNP	CHR	Position	Gene	EA	OA	EAF	Beta	Se	P-value	Info
rs17675363	14	47100492	RPL10L	T	G	0.5958	-0.019	0.003	2.7×10^{-8}	0.973
rs958685	2	70703847	TGFA	C	A	0.4845	-0.018	0.003	3.0×10^{-8}	0.998
rs2431112	5	103931707	NUDT12	G	A	0.5603	0.018	0.003	4.1×10^{-8}	0.992
rs147484722	12	15060253	MGP	A	AAAAAT	0.6045	0.019	0.003	4.4×10^{-8}	0.962
IPAQ-PA										
rs9316077	13	44808665	SMIM2	A	C	0.4159	0.023	0.004	1.4×10^{-8}	0.986
rs146370962	9	81253918	PSAT1	G	A	0.9797	0.086	0.016	3.2×10^{-8}	0.813

Abbreviations: SNP, single-nucleotide polymorphism; CHR, chromosome; EA, effect allele, OA, other allele; EAF, effect allele frequency; Se, standard error; IPAQ-PA, physical activity assessed by international physical activity questionnaire.

Table S3. Genetic loci ($P \leq 1 \times 10^{-5}$) used for DEPICT analysis.
Grip Strength
rs6584479 rs11197871 rs138942406 rs6480792 rs10749903 rs78799917 rs10750224 rs7943757 rs61869032 rs34389751 rs11039204 rs12361415 rs147313654 rs142737000 rs149682893 rs145669855 rs75763126 rs76929617 rs111377653 rs10444491 rs4980987 rs11059675 rs147730268 rs6488898 rs11060385 rs11147063 rs147484722 rs73070150 rs17122812 rs12367809 rs78362716 rs117474961 rs11117062 rs695980 rs1413119 rs73358723 rs61992671 rs11160627 rs17675363 rs66939314 rs2046166 rs2402280 rs8035136 rs75072088 rs1660964 rs11636381 rs2870111 rs371750930 rs376531287 rs4548895 rs56094641 rs889399 rs12926961 rs16973585 rs144134261 rs75819168 rs3815156 rs74251309 rs17629022 rs7207400 rs62073157 rs916888 rs56895823 rs60856912 rs113679679 rs2924322 rs71336393 rs9675886 rs9947450 rs12959157 rs8084121 rs7409148 rs55675854 rs8109936 rs12459368 rs10404726 rs1532127 rs11669079 rs145617179 rs67441433 rs12403795 rs509325 rs10920460 rs1690822 rs765751 rs4927011 rs12079982 rs12757124 rs10641489 rs17391694 rs13040470 rs11908637 rs4810954 rs2869414 rs10485622 rs1736157 rs2826236 rs80145984 rs12463633 rs11678029 rs56262274 rs79245693 rs6730196 rs2164694 rs11677892 rs145647888 rs16861098 rs77625597 rs35533458 rs6718510 rs6545975 rs935166 rs4952491 rs62107261 rs10189584 rs72866952 rs11686591 rs1376406 rs2864823 rs72818529 rs12623218 rs958685 rs72933788 rs147685579 rs62265762 rs4306833 rs4680147 rs13096478 rs80000244 rs73197745 rs4336063 rs62242853 rs7646275 rs247411 rs55932154 rs71324800 rs13073568 rs13107325 rs35319653 rs77136985 rs13118227 rs148736505 rs2446802 rs79030393 rs143127438 rs11133338 rs17659715 rs71602496 rs74692061 rs2431112 rs341339 rs3822742 rs2926836 rs116797439 rs39784 rs1812554 rs13361710 rs4704463 rs10070734 rs34518 rs9383940 rs2237147 rs147925111 rs3956845 rs858982 rs2394520 rs368338926 rs2260051 rs1794514 rs2492933 rs78648104 rs2076308 rs114382070 rs9294260 rs10224575 rs6977081 rs140272246 rs2711111 rs76027953 rs4314553 rs117514101 rs43002 rs144692302 rs1494908 rs12386857 rs36120599 rs17716502 rs2517257 rs150331294 rs13252670 rs117855046 rs10097854 rs7470818 rs2722798 rs116923056 rs10993937 rs145962136 rs73583505 rs16934842 rs1998705
IPAQ-PA
rs114443370 rs10882725 rs72811051 rs3993141 rs138310178 rs35429668 rs75808010 rs9316077 rs61999442 rs149259806 rs8007614 rs1022724 rs148072704 rs62005607 rs55681820 rs75016602 rs7203956 rs906181 rs7245004 rs916695 rs1235337 rs184531897 rs79366036 rs13022622 rs80241253 rs75340524 rs116251592 rs10186318 rs75538549 rs145747951 rs877483 rs13076445 rs72918135 rs16895216 rs146859965 rs192824833 rs36075243 rs1805353 rs1081158 rs74839724 rs147135068 rs60110552 rs10046488 rs200603971 rs1877264 rs73366608 rs7460106 rs7821708 rs4871839 rs4878772 rs146370962
Abbreviations: IPAQ-PA, physical activity assessed by international physical activity questionnaire.

Table S4. Functional annotation for grip strength and IPAQ-PA loci.						
SNP(s)	**CHR**	**Position**	**Closest gene**	**Pooled p-value**	**Gene prioritization (DEPICT)**	**eQTL**
Grip strength						
rs56094641, rs28429148	16	53806453	FTO	3.8×10^{-24}		
rs138942406	10	21954234	MLLT10	1.9×10^{-14}		
rs62107261	2	422144	FAM150B	1.4×10^{-13}		
rs2076308, rs72892910	6	50791640	TFAP2B	1.9×10^{-13}		TFAP2B (Lung)
rs12623218	2	632146	TMEM18	2.3×10^{-13}	TMEM18	
rs12361415	11	47474146	RAPSN	7.4×10^{-12}		C1QTNF4 (Adipose - Subcutaneous), MADD (Esophagus - Muscularis), CELF1 (Nerve - Tibial), FNBP4 (Cells - Transformed fibroblasts), RAPSN (Nerve - Tibial)*
rs6718510, rs556981345	2	25122323	ADCY3	9.7×10^{-12}		ADCY3 (Whole Blood), CENPO (Whole Blood), DNAJC27 (Whole Blood)*
rs509325	1	177894591	SEC16B	2.7×10^{-11}		
rs13073568	3	94031839	DHFRL1	4.3×10^{-11}		STX19 (Artery - Aorta, Nerve - Tibial), PROS1 (Heart - Atrial Appendage)
rs16973585	16	72032231	DHODH	1.6×10^{-10}	DHODH	DHODH (Artery - Tibial)*
rs71336393	18	57855319	MC4R	4.1×10^{-10}		
rs2924322	18	53244414	TCF4	6.2×10^{-10}		RP11-397A16.2 (Cells - Transformed fibroblasts)
rs12459368	19	18459377	PGPEP1	1.0×10^{-9}		SUGP2 (Esophagus - Muscularis)
rs7207400	17	43824360	CRHR1	1.2×10^{-9}		
rs1660964	15	57158987	ZNF280D	1.3×10^{-9}	ZNF280D, TCF12	LINC00926 (Skin - Sun exposed [lower leg])
rs13107325	4	103188709	SLC39A8	1.5×10^{-9}	SLC39A8	
rs116797439	5	163225051	MAT2B	6.3×10^{-9}		
rs73070150	12	15162893	PDE6H	6.8×10^{-9}		

rs62073157	17	44334052	KANSL1	9.1×10^{-9}		
rs140272246	7	21480486	SP4	9.8×10^{-9}		
rs61869032	11	1477797	BRSK2	1.1×10^{-8}		KRTAP5-AS1 (Brain - Caudate [basal ganglia])
rs889399	16	69556583	NFAT5	1.6×10^{-8}	WWP2, NFAT5, NQ01	RP11-419C5.2 (Skin - Sun exposed [lower leg]), CLEC18A (Whole Blood), RP11-296I10.6 (Skin - Sun exposed [lower leg]), NFAT5 (Thyroid)*
rs376531287	16	28918483	RABEP2	1.8×10^{-8}		
rs17675363	14	47100492	RPL10L	2.7×10^{-8}	RPL10L	
rs958685	2	70703847	TGFA	3.0×10^{-8}	TGFA	TGFA (Testis)
rs2431112	5	103931707	NUDT12	4.1×10^{-8}		RP11-6N13.1 (Testis)
rs147484722	12	15060253	MGP	4.4×10^{-8}		
IPAQ-PA						
rs9316077	13	44808665	SMIM2	1.4×10^{-8}		
rs146370962	9	81253918	PSAT1	3.2×10^{-8}		

* More than five eQTLs found, showing up to five genes with most significant tissue.
Abbreviations: SNP, single-nucleotide polymorphism; CHR, chromosome; eQTL, expression quantitative trait loci.

Table S5. Association between genetic risk score (GRS) of grip strength and IPAQ-PA with other fitness and physical activity traits.

	SNPs with $P \leq 5x10^{-8}$			SNPs with $P \leq 1x10^{-6}$		
	Beta*	**Se**	**P-value**	**Beta***	**Se**	**P-value**
Grip strength GRS						
IPAQ-PA	0.0047	0.0029	0.11	0.0126	0.0029	1.4×10^{-5}
CRF	0.1170	0.0223	1.5×10^{-7}	0.1825	0.0220	1.1×10^{-16}
PA	0.0137	0.0061	0.03	0.0308	0.0061	5.2×10^{-7}
IPAQ-PA GRS						
Grip strength	0.0002	0.0003	0.42	0.0005	0.0003	0.07
CRF	0.0041	0.0221	0.85	-0.0234	0.0222	0.29
PA	0.0100	0.0061	0.10	0.0205	0.0061	7.6×10^{-4}

* Per SD in GRS. Models were adjusted for age, sex, region, genotype array and principal components. Two GRSs were calculated; first based on independent SNPs with $P \leq 5x10^{-8}$, and the second based on SNPs with $P \leq 1x10^{-6}$ in pooled association analysis.

Abbreviations: SNP, single-nucleotide polymorphism; IPAQ-PA, physical activity assessed by international physical activity questionnaire; CRF, cardiorespiratory fitness; PA, Physical activity assessed by wrist-worn accelerometer.

Table S6. Results from two-sample Mendelian Randomization analysis for grip strength and CHD				
Method	**OR (95% CI)**	**P-value**	**Heterogeneity P-value**	**Pleiotropy P-value**
Fixed effects meta analysis (simple SE)	0.56 (0.46 , 0.68)	9.1×10^{-10}	NA	0.8424
Fixed effects meta analysis (delta method)	0.57 (0.47 , 0.71)	1.8×10^{-8}	0.23	
Random effects meta analysis (delta method)	0.57 (0.46 , 0.72)	5.0×10^{-7}	0.23	
Maximum likelihood	0.56 (0.46 , 0.69)	4.1×10^{-9}	0.15	
Inverse variance weighted	0.56 (0.44 , 0.70)	1.4×10^{-7}	0.17	
Weighted median	0.69 (0.51 , 0.95)	0.01	NA	
MR Egger	0.51 (0.19 , 1.36)	0.17	0.10	
Abbreviations: CHD, coronary heart disease; OR, odds ratio, NA, not applicable.				