



OPEN

Genomics and transcriptomics landscapes associated to changes in insulin sensitivity in response to endurance exercise training

Louise Y. Takeshita^{1,9}, Peter K. Davidsen^{1,9}, John M. Herbert¹, Philipp Antczak^{1,8}, Matthijs K. C. Hesselink², Patrick Schrauwen², S. John Weisnagel³, Jeremy M. Robbins⁴, Robert E. Gerszten⁴, Sujoy Ghosh⁵, Mark A. Sarzynski⁶, Claude Bouchard⁷ & Francesco Falciani¹✉

Despite good adherence to supervised endurance exercise training (EET), some individuals experience no or little improvement in peripheral insulin sensitivity. The genetic and molecular mechanisms underlying this phenomenon are currently not understood. By investigating genome-wide variants associated with baseline and exercise-induced changes (Δ) in insulin sensitivity index (S_i) in healthy volunteers, we have identified novel candidate genes whose mouse knockouts phenotypes were consistent with a causative effect on S_i . An integrative analysis of functional genomic and transcriptomic profiles suggests genetic variants have an aggregate effect on baseline S_i and ΔS_i , focused around cholinergic signalling, including downstream calcium and chemokine signalling. The identification of calcium regulated MEF2A transcription factor as the most statistically significant candidate driving the transcriptional signature associated to ΔS_i further strengthens the relevance of calcium signalling in EET mediated S_i response.

Regular endurance exercise is a strong physiological stimulus that plays an important role in skeletal muscle homeostasis. It leads to a multitude of functional improvements when performed regularly (i.e. exercise training) and is considered a cornerstone in the prevention of type 2 diabetes^{1,2} by increasing tissue responsiveness to circulating insulin. Skeletal muscle contraction and peripheral insulin action are highly inter-twined³. In fact, up to 80% of the in vivo insulin-mediated glucose disposal in the postprandial state occurs in skeletal muscle⁴, making it quantitatively the most important tissue for systemic glucose disposal. However, we and others have demonstrated that healthy individuals are highly heterogeneous in their ability to improve peripheral insulin sensitivity (S_i) in response to endurance exercise training (EET)⁵⁻⁷. Notably, despite good adherence to the EET program, a significant percentage of individuals (up to ~20%) show no changes in S_i and some even demonstrate decreases in S_i values⁵⁻⁷. Furthermore, we have previously shown that such phenomenon is likely to include a substantial genetic component⁸ and that healthy individuals with high and low S_i responses to EET have different skeletal muscle gene expression patterns at baseline⁶.

A number of studies have been performed to understand the molecular basis of insulin resistance (IR), a pathological alteration in insulin sensitivity linked to many metabolic disorders, such as type 2 diabetes. Although the exact underlying cause of IR has not been fully elucidated, a number of major mechanisms, including oxidative stress, inflammation, insulin receptor mutations, endoplasmic reticulum stress, and mitochondrial

¹Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK. ²Department of Nutrition and Movement Sciences, NUTRIM School for Nutrition and Translational Research in Metabolism, Maastricht University Centre, Maastricht, The Netherlands. ³Diabetes Research Unit, Endocrinology and Nephrology Axis, CRCHU de Québec, Université Laval, Québec City, Canada. ⁴Division of Cardiovascular Medicine, and Cardiovascular Research Center, Beth Israel Deaconess Medical Center, Boston, MA 02215, USA. ⁵Centre for Computational Biology and Program in Cardiovascular and Metabolic Disorders, Duke-NUS Medical School, Singapore, Singapore. ⁶Department of Exercise Science, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA. ⁷Human Genomics Laboratory, Pennington Biomedical Research Center, Baton Rouge, LA, USA. ⁸Center for Molecular Medicine Cologne, University Hospital Cologne, 50931 Cologne, Germany. ⁹The authors contributed equally: Louise Y. Takeshita and Peter K. Davidsen. ✉email: falcianf@liverpool.ac.uk

dysfunction have been suggested⁹. Overall, however, the molecular mechanisms underlying variation in S_i in a healthy population and the heterogeneous ability to improve S_i through EET are currently not well understood.

Here we address this important question by computational analysis of genome-wide association study (GWAS) and skeletal muscle gene expression datasets derived from the HERITAGE Family Study. Our analysis identified several candidate genes linked to mechanisms of baseline S_i , as well as training-induced changes in S_i (ΔS_i). Homozygous mouse knockouts of four of these candidates show alterations in glucose disposal and other relevant phenotypes, suggesting that our approach is likely to have identified genes causally linked to S_i . Furthermore, analysis of both GWAS and skeletal muscle transcriptomics data shows that a molecular signature linked to calcium-regulated cholinergic signalling may be an important component of the observed variation in S_i in a healthy population and predicts exercise-induced changes in S_i both in HERITAGE and an independent clinical exercise study.

Methods

HERITAGE family study. The sample, study design, and EET protocol of the HERITAGE study have been described elsewhere¹⁰. Briefly, for the Caucasian sample of HERITAGE, 479 sedentary adults (233 males) from 99 nuclear families composed of parents (≤ 65 years old) and offspring (≥ 17 years old) were defined as completers ($> 95\%$ of all exercise session requirements) following exposure to a standardized and fully monitored progressive 20-week EET program (frequency of cycle ergometer sessions was three times per week). Participants were all sedentary, but healthy at baseline and not taking medications for hypertension, diabetes, or dyslipidemia. A detailed description of the study design and methodology (including a table with demographic data) can be found in the Supplementary Information section.

Intravenous glucose tolerance test (IVGTT) protocol. A frequently-sampled IVGTT was performed after an overnight fast (12 h), at baseline and post-intervention (24–36 h after the last exercise bout) following the protocol described in¹¹. In premenopausal women, the test was scheduled to coincide with the follicular phase of the menstrual cycle. The S_i index ($\text{mU}/[\text{L} \times \text{min}]$), which measures the ability of an increment in plasma insulin to enhance the net disappearance of glucose from plasma was derived using the MINMOD Millennium software¹². Changes in S_i (ΔS_i) were calculated as post-training S_i minus baseline S_i .

GWAS genotype data processing. Single nucleotide polymorphism (SNP) genotyping ($\sim 325,000$ SNPs, Illumina Human CNV370-Quad v3.0 BeadChips) on genomic DNA from lymphoblastoid cells was performed and subjected to extensive quality control as previously described¹³. SNPs excluded from association analyses were filtered according to the following criteria: (a) minor allele frequency $< 5\%$, (b) violated Hardy–Weinberg equilibrium ($p < 1 \times 10^{-6}$), and (c) missing values in $> 10\%$ of individuals. SNPs are based on dbSNP build 151 with genomic coordinates for GRCh38 (hg38) assembly. To estimate linkage disequilibrium (LD), r^2 correlation values between SNPs were calculated using default parameters in PLINK v1.9 (www.cog-genomics.org/plink/1.9/).

GWAS analysis. Baseline S_i was adjusted for sex, age, log-transformed BMI, and weight-adjusted $\text{VO}_{2\text{max}}$ and ΔS_i was adjusted for log-transformed baseline S_i , sex, age, log-transformed BMI, and weight-adjusted $\text{VO}_{2\text{max}}$. Associations between the normalized trait residuals and SNP genotypes were investigated using additive linear mixed effect (LME) models that accounted for within family correlations (function `lme` of the ‘nlme’ R package v3.1¹⁴). Significance thresholds were calculated using the *SimpleM* method¹⁵ implemented in R programming language¹⁶. Conventional Bonferroni correction is overly conservative in genome-wide analyses due to high LD observed in genetic variants. The *SimpleM* method uses principal component analysis to calculate the effective number of independent tests, which resulted in 199,278 tests. A Bonferroni adjustment on this number results in a significance threshold of $p < 2.51 \times 10^{-7}$ ($0.05/199,278$). A suggestive significance threshold was set at $p < 1 \times 10^{-5}$. All statistical analyses were performed using R version 3.5.1. SNPs were mapped to genes based either on their position (located within a 20 kb window upstream and downstream of the gene) or if they have been identified as eQTL of a gene expressed in skeletal muscle tissue. Positional mapping was performed using MAGMA v1.07b and eQTL associations were retrieved from GTEx Portal release V8¹⁷.

Candidate genes validation. We assessed the potential relevance of the candidate genes identified by the GWAS analysis by using a publicly available dataset from the International Mouse Phenotyping Consortium (IMPC) database¹⁸. We selected a panel of physiological measurements of relevance to S_i . Details of the protocols are available from the database web site (<https://www.mousephenotype.org/>). Briefly, the ability to metabolize glucose has been assessed using three parameters derived from an intra-peritoneal glucose tolerance test (IPGTT). These were: (1) initial response to glucose challenge, (2) fasting blood glucose concentration and (3) the area of glucose response under the curve. Body composition was assessed by DEXA scan. Further characterization included a panel of blood measurements including insulin, cholesterol, glucose, glycerol, free fatty acids, and creatinine. In addition, respiratory exchange ratio was also available.

Functional GWAS. To test whether genes within specific biological pathways are enriched by genetic associations with lower p-values than expected by chance, we applied GLOSSI¹⁹ from *cpvSNP* R package (v 1.18.0)²⁰. First, GWAS results were pruned to keep only independent SNPs ($r^2 > 0.8$) resulting in 249,035 SNPs. After positionally mapping the remaining SNPs to genes (± 20 kb window), their p-values were used as input to compute the estimate of enrichment within a given biological pathway. Resulting p-values were corrected for mul-

multiple testing using Bonferroni correction. Gene set collections used KEGG pathways (c2.cp.kegg.v7.1.entrez) retrieved from MSigDB v7.1^{21,22} and manually curated functional modules representing genes required for normal skeletal muscle activity²³. Threshold for significantly enriched biological pathways was defined as $p_{\text{adj}} < 0.05$. In order to further investigate the most important pathways, we selected the most significant SNPs associations ($\alpha < 0.05$) within the pathways identified by GLOSSI and remapped these on KEGG pathways using the web-based tool DAVID (version 6.8). Threshold for significant biological pathways was defined as $p_{\text{adj}} < 0.05$, provided in DAVID with Benjamini–Hochberg adjustment.

RNA extraction and global gene expression profiling. Vastus lateralis muscle biopsies were also obtained in a subsample of the SNP-genotyped participants ($n = 41$) before and after (~96 h after final exercise session) the intervention using a percutaneous needle. Each biopsy was immediately frozen in liquid nitrogen and stored at -80°C until RNA preparation. RNA extraction as well as reverse transcription were done as previously described²⁴. Affymetrix U133 + 2 arrays were used to quantitate global mRNA expression levels. The raw microarray CEL files are deposited in the public Gene Expression Omnibus (GEO) database²⁵ under accession number GSE117070.

Microarray data processing. Raw CEL files were Robust Multichip Average (RMA) normalized following removal of probes termed ‘absent’ in more than 80% of the samples by the MAS5 algorithm inside the ‘affy’ package (26,151 probesets discarded)²⁶. Quality control plots of the polyA-control RNAs (spike-ins added right after RNA purification) highlighted a batch issue that was resolved by applying the COMBAT software²⁷. The JetSet package was used to select a single ‘optimal’ probeset to represent each gene based on specificity, robustness against mRNA degradation, and MAS5 present call rate²⁸. As the most representative probeset for each gene is selected, they have high splice isoform coverage.

Gene set enrichment analysis (GSEA). The entire skeletal muscle transcriptome was ranked by individually regressing pre-training mRNA expression levels against baseline S_i and ΔS_i , using linear mixed effect models that accounted for within family correlations. Both outcome variables were adjusted by age, sex, log-transformed BMI, weight-adjusted $\text{VO}_{2\text{max}}$, and type I fibre percentage (see²⁹ for details on the fibre typing), with ΔS_i also being adjusted for baseline S_i . Based on this ranking (Student t -statistic), we performed a pre-ranked GSEA using the default parameters in *clusterProfiler* v3.14.3 R package³⁰ to identify candidate biological pathways significantly enriched in genes that are associated with baseline S_i and ΔS_i (either top or bottom of the distribution). Gene set collections used were transcription factor targets and KEGG biological pathways (c3.tft.v7.1.entrez.gmt and c2.cp.kegg.v7.1.entrez, respectively) retrieved from MSigDB v7.1²¹ and manually curated functional modules representing genes required for normal skeletal muscle activity²³. Threshold for significant sets was defined as false discovery rate (FDR) < 0.05 .

Transcriptomics-based model to predict ΔS_i . To test if baseline expression of MEF2A target genes is predictive of exercise-induced ΔS_i we applied a regression-based modelling approach allowing for pairwise interactions (function *lm* of the ‘stats’ R package¹⁶) between baseline expression of three genes. Only genes translating to proteins that interact with MEF2A were included and consisted of a total of 50 experimentally validated interactors with high confidence score (> 0.8) identified in the STRING database³¹ (Supplementary Table S1). All possible linear regression models based on all possible combinations of three-genes sets were examined (a total of 19,600 models). More precisely, we define:

$$\Delta S_i = a\theta_1 + b\theta_2 + c\theta_3 + d\theta_1\theta_2 + e\theta_1\theta_3 + f\theta_2\theta_3 + \text{sex} + \text{VO}_{2\text{max}}/\text{kg} + \varepsilon$$

where mRNA abundance is represented by θ and the noise model component by ε . Weight-adjusted $\text{VO}_{2\text{max}}$ and sex were included as covariates.

Ethics declaration. The study protocol was approved by the Institutional Review Boards at each of the five participating centers of the HERITAGE Family Study consortium (Indiana University, Laval University, University of Minnesota, Texas A&M University, and Washington University at St. Louis). Written informed consent was obtained from each study participant. The subjects who were under 18 years, one of the parents gave consent in addition to the participant. This was an easy procedure to follow as we were recruiting whole nuclear families. All research was performed in accordance with the Declaration of Helsinki.

Results

Overview of the analysis strategy. The overarching goal of this study is to investigate the genetic and molecular basis of the variation in S_i and ΔS_i following EET, across a healthy population. We address this by integrating a traditional GWAS approach with the analysis of skeletal muscle transcriptomics data within HERITAGE, one of the largest studies to evaluate the response of several physiological measurements to EET. The strategy, described in a schematic format in Fig. 1, involved:

1. The identification of genetic variants linked to S_i and ΔS_i (Fig. 1A) and the validation of the corresponding gene candidates in a mouse knock down experiment database (Fig. 1B).
2. The identification of transcriptional signatures linked to baseline S_i and ΔS_i (Fig. 1C).
3. The identification of transcription factors that may be driving the transcriptional signatures identified above (Fig. 1E).

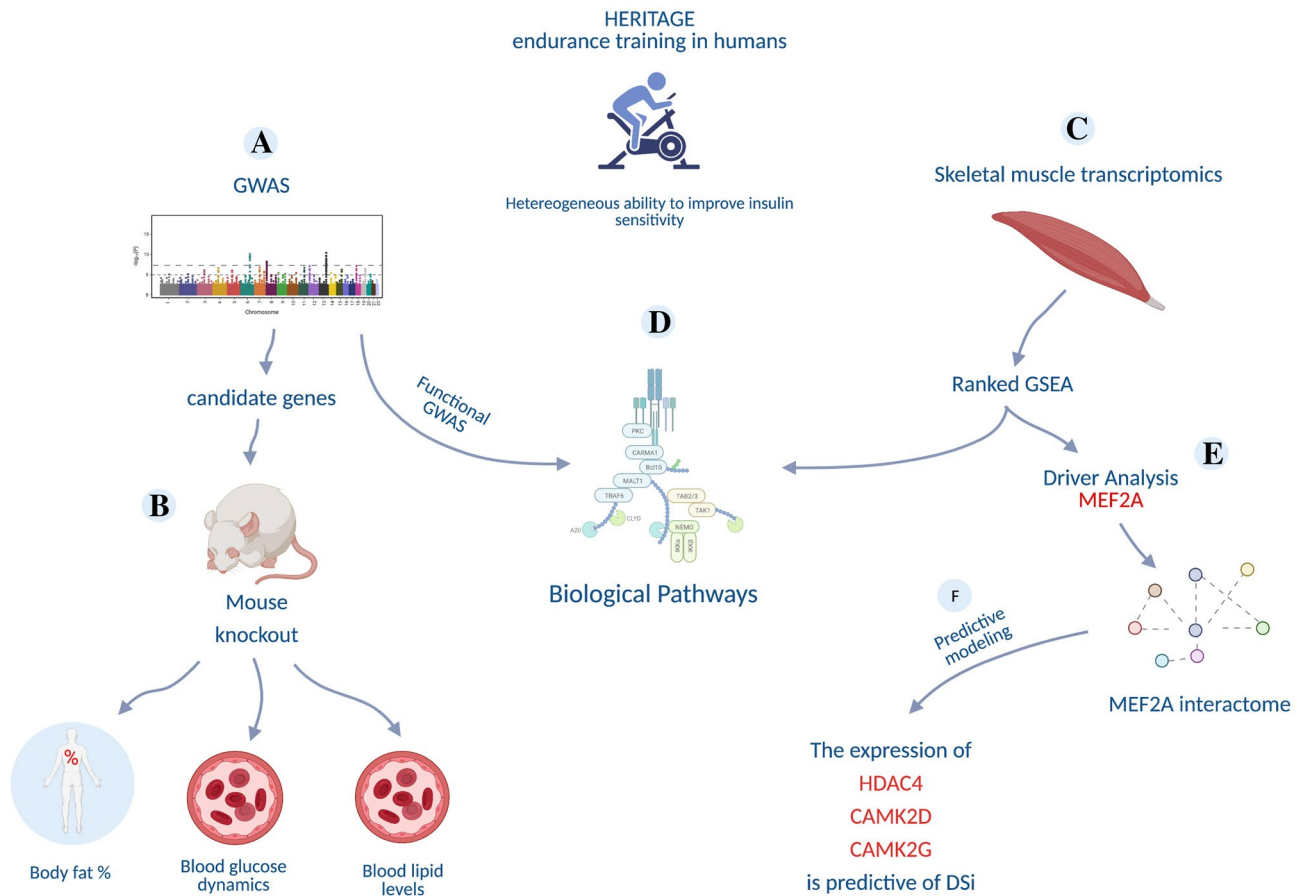


Figure 1. Overview of the study design, consisting of five interconnected steps: (A) identification of genetic variants linked to insulin sensitivity index (S_i), (B) validation of the corresponding gene candidates in a mouse knock down experiment database, (C) identification of transcriptional signatures in skeletal muscles correlating with baseline S_i and ΔS_i , (D) integration of genetics and transcriptomics signatures, (E) identification of transcription factors likely to drive the transcriptional signatures linked to ΔS_i which led to the identification of MEF2A transcription factor as main driver of the transcriptional profile, and (F) development of a statistical model that can predict ΔS_i from the transcriptional state of MEF2A interacting genes.

4. Testing whether baseline expression of genes linked to TF drivers are good predictors of ΔS_i (Fig. 1F).

In addition, by mapping gene candidates identified by GWAS and the transcriptional signatures we tested the hypothesis that genetic variation may be linked to downstream changes in gene expression (Fig. 1D).

GWAS analysis identifies putative loci linked to baseline and exercise-induced changes in S_i . Investigation of genetic variants linked to inter-individual heterogeneity of ΔS_i is based on the assumption that its underlying molecular mechanisms have a genetic component. Our analysis revealed that this assumption is indeed likely to be correct as 29% of the variance in ΔS_i is accounted for by family membership (Supplementary Fig. S1). Moreover, from an ANOVA, there is 40% more variance between families than within families ($p = 0.02$), providing additional suggestive evidence that the changes in S_i in response to exercise training are characterized by a significant heritable component.

Therefore, we set to identify specific genetic variants linked to S_i and ΔS_i by GWAS. As a first step we tested whether relevant physiological variables may be confounding factors that should be considered in the analysis. We discovered that variation in baseline S_i was significantly linked to BMI, VO_{2max}/kg , sex and age (Supplementary Fig. S2A). We also discovered that ΔS_i was significantly linked to baseline S_i and sex (Supplementary Fig. S2B), with 27% of the variance in ΔS_i explained by baseline S_i (Supplementary Fig. S2C). A similar amount of variance in post-training S_i was also explained by baseline S_i (Supplementary Fig. S2D). Based on this evidence, we performed a GWAS with the objective to identify SNPs linked to baseline S_i and ΔS_i , where both traits were adjusted for potential confounding variables.

The GWAS analysis identified one SNP significantly associated to baseline S_i (*DNAL1* rs11622678, $p = 3.79 \times 10^{-8}$) plus seven SNPs with suggestive association ($p < 10^{-5}$), and ten SNPs with suggestive association ($p < 10^{-5}$) with ΔS_i (Fig. 2). Positional and eQTL mapping revealed genes that are located within or near these SNPs (± 20 kb window) and/or have their expression correlated with them (Table 1).

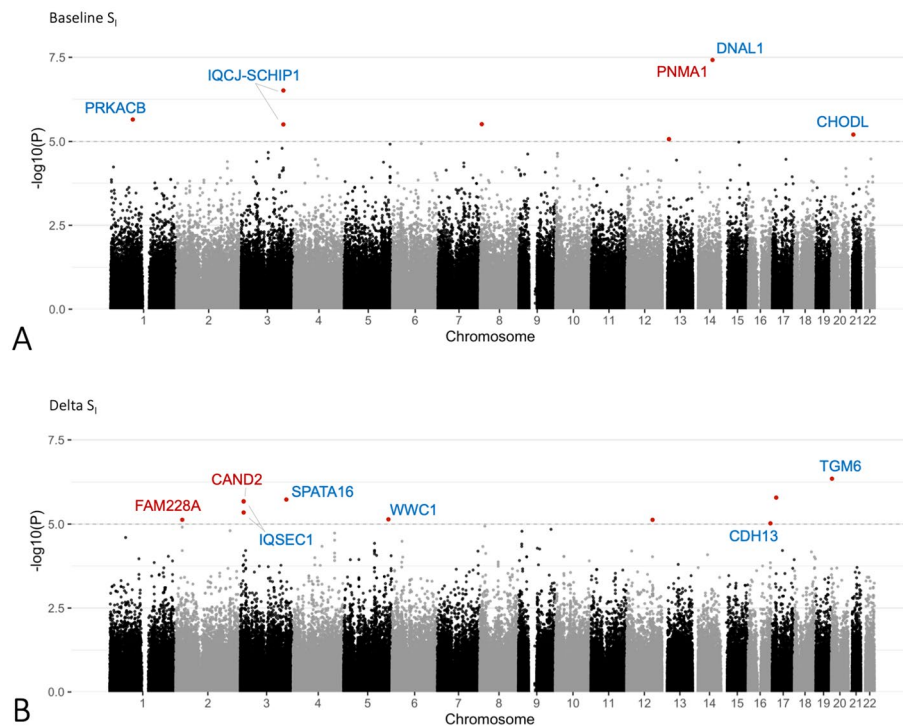


Figure 2. Manhattan plots showing genetic loci associated with baseline and ΔS_i and their mapped genes. Genetic loci reaching at least suggestive association are represented by points highlighted in red. Genes mapped to these loci are annotated, where genes mapped by position are highlighted in blue, and genes mapped by eQTL are highlighted in red. Highlighted loci showing no gene annotation did not map to any genes according to the criteria we used. For additional details, see Table 1.

We next tested whether these candidate genes may be causally linked to insulin-regulated physiology. We used a publicly available dataset from the International Mouse Phenotyping Consortium database¹⁸. This database is the result of an international effort involving 19 assessment centres and reporting a wide range of physiological measurements from a collection of 6440 gene knockouts. Here we selected a panel of physiological measurements of relevance to S_i from mouse knockouts in our putative gene lists (see methods section and Fig. 3 for a full list). Eight out of our twelve genes identified by the GWAS (67%) were represented in this database and had a full set of the selected measurements. These genes are PNMA1, DNAL1, IQCJ, CHODL for baseline S_i and TGM6, SPATA16, CAND2, CDH13 for ΔS_i . We found that *DNAL1* and *IQCJ* (linked to baseline S_i), and *TGM6* and *CAND2* (linked to ΔS_i) had significant associations with one or more selected mouse phenotypes (Fig. 3).

The *CAND2* knockout is characterized by lower glucose levels, improved glucose tolerance, lower levels of total and HDL cholesterol, increased levels of glycerol, free fatty acids and plasma creatinine, increased lean mass, and decreased fat mass. Interestingly, phenotypes associated with *TGM6* knockout were found to be in accordance with those found associated with *CAND2*, leading also to lower levels of plasma glucose, improved glucose tolerance, lean mass increase, and fat mass decrease. It is worth noting that *CDH13*, *CHODL*, *PNMA1*, and *SPATA16*, although present in the database and tested for relevant phenotypes did not show significant associations with phenotypes relevant to S_i .

Functional analysis of the genomic landscape associates the calcium signalling and cholinergic synapse pathways to S_i . Complex traits are often characterised by a relatively small contribution of multiple genetic variants that all together contribute to the phenotype of interest. Traditional univariate GWAS analysis, which is often underpowered, may fail to identify these complex interactions. Functional GWAS analysis addresses this issue by testing whether genetic variants tend to cluster within given biological pathways. With this in mind, we analysed our GWAS results with GLOSSI, one such approach. After pruning GWAS results to select independent SNPs ($R^2 < 0.8$) and positionally mapping those to genes (± 20 kb window), we identified significantly enriched biological pathways (KEGG pathways and curated skeletal muscle pathways²²) for both baseline S_i and ΔS_i . This analysis initially identified 17 significant pathways linked to baseline S_i and 8 pathways specifically linked to ΔS_i (Supplementary Fig. S3). The 17 pathways linked to baseline S_i included 470 genes linked to genetic variants with a lower nominal p value ($p < 0.05$). Interestingly, we saw that the four most significantly enriched pathways were sufficient to represent half of the 470 most significant genes. These pathways were *calcium signalling* (77 genes), *axon guidance* (63 genes), *chemokine signalling* (66 genes) and *cholinergic synapse* (47 genes) (Fig. 4, Supplementary Fig. S4 and Supplementary Table S2).

SNP	Chr	Location (bp)	P	Gene symbol	Gene mapping	Gene location (bp)
Delta S_i						
rs4815227	20	2441096	4.50E-07	TGM6	Positional	2380908–2432753
rs12449918	17	15392862	1.63E-06	–	–	–
rs6799845	3	172873569	1.86E-06	SPATA16	Positional	172889357–173141268
rs9211 ^a	3	12898047	2.10E-06	IQSEC1	Positional	12897043–13282998
–	–	–	–	LOC105376955	Positional	12882360–12885099
–	–	–	–	RP11-767C1.2	Skeletal muscle eQTL	12832219–12832728
–	–	–	–	CAND2	Skeletal muscle eQTL	12796472–12871916
rs14191 ^a	3	12898847	2.13E-06	RP11-767C1.2	Skeletal muscle eQTL	12832219–12832728
–	–	–	–	CAND2	Skeletal muscle eQTL	12796472–12871916
–	–	–	–	IQSEC1	Positional	12897043–13282998
–	–	–	–	LOC105376955	Positional	12882360–12885099
rs2600330	3	12926604	4.52E-06	IQSEC1	Positional	12897043–13282998
rs7722673	5	168417231	7.23E-06	WWC1	Positional	168292060–168472303
rs7595372	2	24409253	7.45E-06	FAM228A	Skeletal muscle eQTL	24175069–24200849
rs969863	12	97202907	7.48E-06	–	–	–
rs422455	16	83729855	9.51E-06	CDH13	Positional	82626794–83796610
Baseline S_i						
rs11622678 [*]	14	73655058	3.79E-08	PNMA1	Skeletal muscle eQTL	73711783–73714372
–	–	–	–	RP3-414A15.10	Skeletal muscle eQTL	73616700–73633941
–	–	–	–	DNAL1	Positional	73644875–73703728
–	–	–	–	RP3-414A15.2	Skeletal muscle eQTL	73530152–73530610
rs10936174	3	159198318	3.07E-07	IQCJ	Positional	159069252–159266307
–	–	–	–	IQCJ-SCHIP1	Positional	159069252–159897366
rs7524898	1	84167162	2.24E-06	PRKACB	Positional	84077975–84238498
rs10107799	8	6102222	3.07E-06	–	–	–
rs7653174	3	159120808	3.13E-06	IQCJ	Positional	159069252–159266307
–	–	–	–	IQCJ-SCHIP1	Positional	159069252–159897366
rs205666	21	18120899	6.27E-06	CHODL	Positional	17819595–18267373
rs9511351 ^b	13	24575137	8.56E-06	LOC101927375	Positional	24541663–24567321
rs9553347 ^b	13	24557233	8.56E-06	LOC105370295	Positional	24538602–24541913
–	–	–	–	LOC101927375	Positional	24541663–24567321

Table 1. Genomic loci showing, at a minimum, suggestive association with ΔS_i and baseline S_i . ^{*}Statistically significant association. ^{a,b} $r^2 > 0.8$.

The same analysis performed with the top-most significant genes within the pathway enrichment set for ΔS_i identified only 31 genes. While this limited number of genes preclude a systematic pathway enrichment analysis, 21 genes could be mapped within the *adrenergic signalling in cardiomyocyte* pathway. Importantly, five of these genes also mapped within the *calcium signalling* pathway (*RYR2*, *SLC8A1*, *CACNA1C*, *CACNA1D* and *CACNA1S*), providing a link with the analysis performed on the baseline S_i . These included subunits of the ATPase NA^+/K^+ transporter (ATPA2/A4/B1/B3) and additional calcium voltage-gated channels. In addition, there were four subunits of the cytochrome C oxidase enzyme (*COX4I2*, *COX6B1*, *COX7A1*, *COX7A2L*).

Transcription factor driver analysis identifies the calcium dependent transcription factor MEF2A as the most significant driver of the ΔS_i transcriptional signature. The results of the functional GWAS suggest a role of skeletal muscle in insulin dependent glucose uptake and the effects of exercise in remodelling this tissue. Therefore, we set to investigate the transcriptomic profile of skeletal muscles in a subset of the HERITAGE individuals. We wanted to identify baseline transcriptional signatures that correlate to S_i and ΔS_i as well as the transcription factors that may drive such signatures. More specifically, by using a GSEA based approach we searched for enrichment in transcription factor binding sites in the list of genes correlated to S_i and ΔS_i .

Only when including fibre type composition in the models we have identified gene sets (a total of 45) mapped to known transcription factor binding sites significantly enriched by genes whose skeletal muscle expression correlated to ΔS_i (Supplementary Table S3). Fibre type composition has been previously linked to insulin-dependent glucose uptake in skeletal muscle^{32–34} and its addition to the ΔS_i model also led to a larger number of significantly enriched biological pathways. These encompassed a variety of biological functions such as signalling, energy and amino acid metabolism, tissue homeostasis, protein degradation, immune system, and translation (Supplementary Fig. S5). Interestingly, the KEGG pathways chemokine signalling, neuroactive ligand receptor

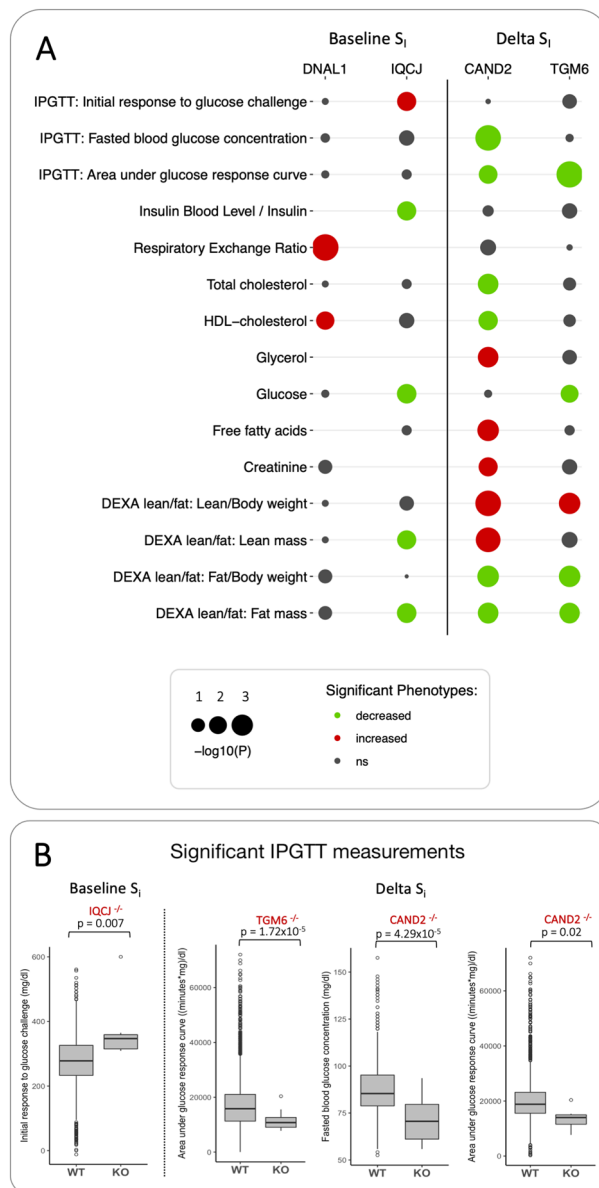


Figure 3. Validation of gene candidates from GWAS analysis in a mouse knock down experiment database. **(A)** Mouse knockout of DNAL1 and IQCJ (associated with baseline S_1) and CAND2 and TGM6 (associated with ΔS_1) lead to changes related to metabolism and skeletal muscle development ($P_{adj} < 0.05$, ns = not significant). Details of each experiment can be found in the IMPC database¹⁸. **(B)** Boxplots detailing significant associations of IQCJ, CAND2 and TGM6 knockouts with measurements from IPGTT (intra-peritoneal glucose tolerance test) experiment in mice, which is analogous to intravenous glucose tolerance test (IVGTT) method used to measure S_1 in study participants.

interaction, and the functional term calcium dynamics/homeostasis required for excitation–contraction coupling were reminiscent of the results for the functional GWAS analysis (Supplementary Table S2).

Remarkably, when we examined which transcription factor may be able to explain the expression of genes in the S_1 transcriptional signatures, we found that the top candidate gene was the calcium dependent transcription factor MEF2A (Fig. 5). The hypothesis that MEF2A drives a significant part of the transcriptional signature linked to ΔS_1 is supported by the observation that the global transcriptional signature associated to MEF2A knockdown in C2C12 can recapitulate the transcriptional signature correlated to ΔS_1 in the HERITAGE cohort (Fig. 1, Supplementary Information and Supplementary Table S4). This result is consistent with the findings from the functional GWAS that suggested a key role of calcium signalling in baseline S_1 . The linkage between genetic variation and the transcriptomics signatures in skeletal muscle associated with S_1 emerged through the linkage between calcium signalling and MEF2A.

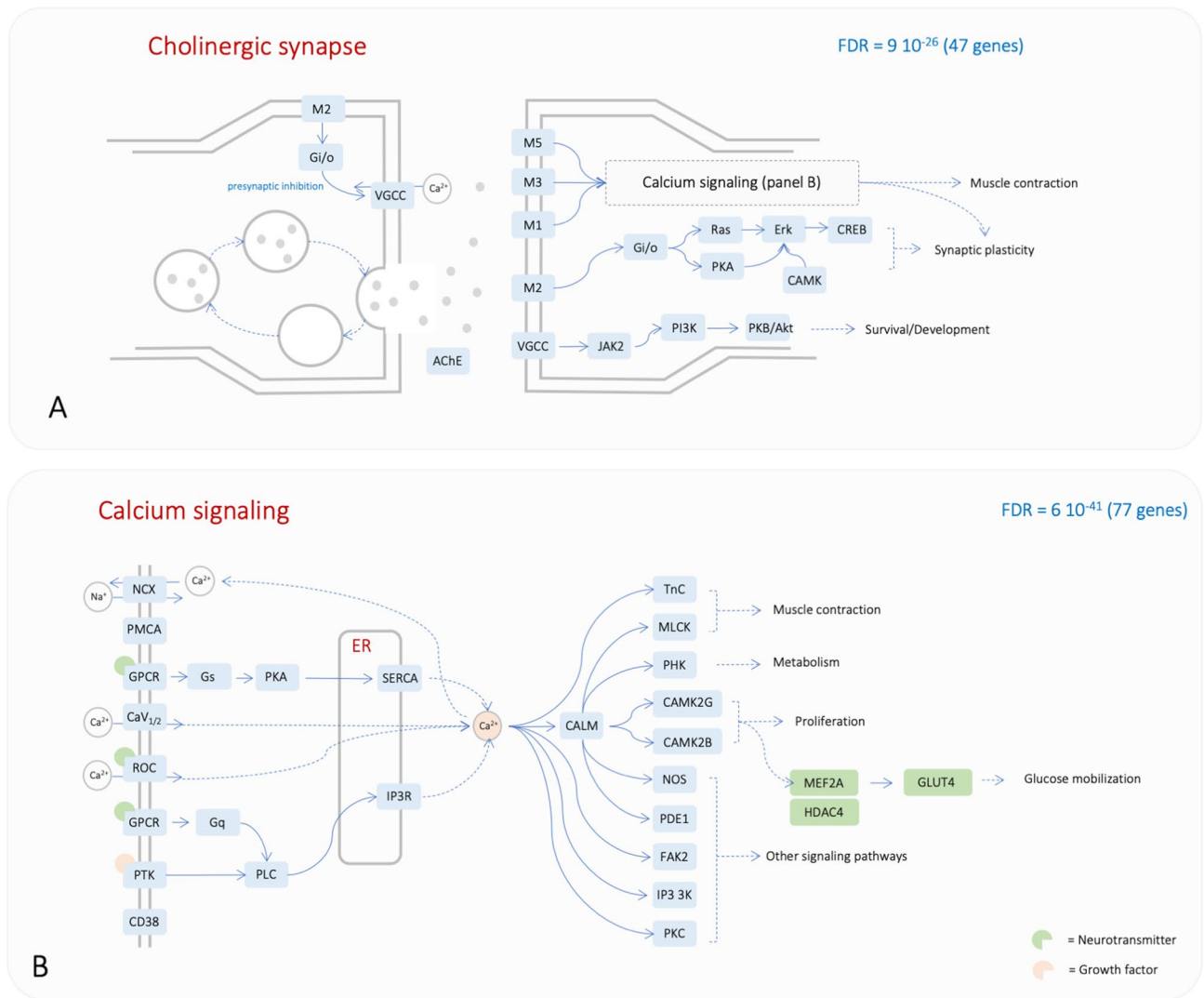


Figure 4. Most significant pathways identified by functional GWAS. **(A)** cholinergic synapse pathway and **(B)** calcium signalling pathway. The two KEGG pathways are interconnected **(B)**. The KEGG pathway calcium signalling has been edited to add three additional genes of relevance (green rectangles). Each node can represent multiple genes. Symbols in the two panels are the same as in the original KEGG pathway map (NCX: Na/Ca exchanger; PMCA: ATPase plasma membrane Ca²⁺ transporting; CaV12: calcium voltage-gated channel subunit alpha1 A and C; ROC: nicotinic acetylcholine receptor alpha-7; GPCR: cysteinyl leukotriene receptor 1; PTK: epidermal growth factor receptor; CD38: ADP-ribosyl cyclase 1; PLC: phosphatidylinositol phospholipase C; Gs: guanine nucleotide-binding protein G(s) subunit alpha; Gq: guanine nucleotide-binding protein subunit alpha-11; SERCA: P-type Ca²⁺ transporter type 2A; IP3R: inositol 1,4,5-triphosphate receptor type 1; TnC: troponin C; MLCK: myosin-light-chain kinase; PHK: phosphorylase kinase alpha/beta subunit; NOS: nitric-oxide synthase; PDE1: calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase; FAK2: focal adhesion kinase 2; IP3 3K: 1D-myo-inositol-triphosphate 3-kinase; M1: muscarinic acetylcholine receptor M1; M2: muscarinic acetylcholine receptor M2; M3: muscarinic acetylcholine receptor M3; M5: muscarinic acetylcholine receptor M5; VGCC: calcium voltage-gated channel subunit alpha1 C; AChE: acetylcholinesterase; Gi/o: G protein subunit beta 5; CAMK: CAMK2G and CAMK2B; CREB: cAMP responsive element binding protein 3; PKB/Akt: AKT serine/threonine kinase 3). Genes shown in this figure map to SNPs with *p* value < 0.05.

The development and validation of a baseline MEF2 transcriptional signature predictive of ΔS_i . Ranked GSEA analysis revealed that genes with baseline expression levels showing higher (positive) or lower (inverse) correlation with ΔS_i were found to be enriched by MEF2A targets (Supplementary Fig. S6 and Supplementary Table S5). We also show that the transcription factor MEF2A may be the main actor driving that transcriptional signature. We therefore set to develop a predictor of ΔS_i based on the baseline expression of genes that are part of the MEF2A interactome. We focused on 50 experimentally validated interactors with a high confidence score (identified in the STRING database, see Supplementary Table S1 for an exhaustive list) and we performed a comprehensive analysis of all possible linear regression models based on gene expression levels of

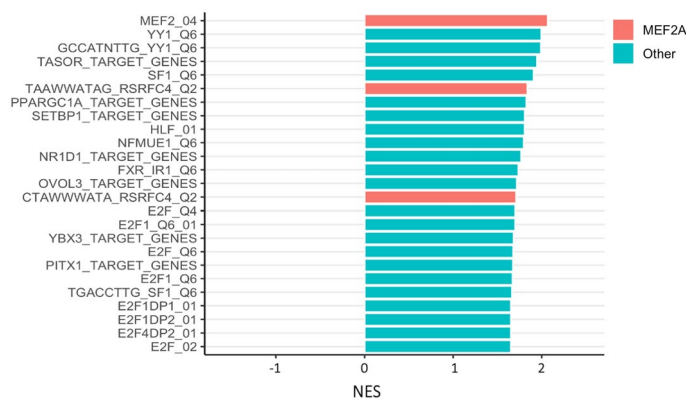


Figure 5. Top 25 transcription factor target gene sets significantly enriched in GSEA ($P_{\text{adj}} < 0.05$) ranked by NES (normalised enrichment score). Gene sets with a positive NES are enriched by genes with skeletal muscle baseline expression positively correlated to ΔS_i while genes with a negative NES are enriched by genes with skeletal muscle baseline expression negatively correlated to ΔS_i . Sets corresponding to MEF2A target genes are highlighted in red. A full list of significant transcription factor target gene sets is available on Supplementary Table S3.

Variable	Regression coefficient	Standard error	t-value	p value
VO _{2max}	-0.03	0.004	-0.68	0.50
Sex	-2.15	0.83	-2.59	0.013
CAMK2D mRNA	40.73	27.98	1.46	0.15
CAMK2G mRNA	74.78	32.04	2.33	0.02
HDAC4 mRNA	-34.94	22.37	-1.56	0.13
CAMK2D:CAMK2G	-12.08	3.41	-3.54	0.001
CAMK2D:HDAC4	6.94	1.63	4.27	<0.001
CAMK2G:HDAC4	-0.40	2.90	-0.14	0.89

Table 2. Result of the mRNA-based multivariate regression model for S_i training response among the subset of White HERITAGE participants for which global gene expression data before and after the exercise program are available ($n = 47$). Importantly, the response variable (i.e. ΔS_i) spanned a broad range (-79 to $+120\%$) among the participants included in the current analysis. Sex and VO_{2max} adjusted for body size were included as covariates in the model. Notably, the abundance of these transcripts was not responsive to the training intervention, but rather higher basal expression levels were associated with greater gains in S_i . CAMK2D: 228555_at; CAMK2G: 212757_s_at; HDAC4: 228813_at.

all possible combinations of three-genes sets (a total of 19,600 models). Remarkably, the most predictive statistical model ($R^2 = 48\%$, F-value = 6.4, $p < 0.001$; Table 2) included the direct activators of the MEF2 transcription factor (*HDAC4*, *CAMK2D*, and *CAMK2G*, Fig. 4B). In comparison, a linear regression model comprising only of sex and VO_{2max}/kg as predictors showed poor predictive power ($R^2 = -0.01\%$, F-value = 0.98 and $p = 0.39$), reinforcing the large contribution of these genes on predicting ΔS_i . Consistent with the established heteromultimeric nature of CaMKII, the model did show a highly significant interaction between the two isoforms *CAMK2D* and *CAMK2G*. Regression diagnostics confirmed conformity of the residuals to the assumptions of normality, linearity and homoscedasticity.

To test the general applicability of the HERITAGE predictive gene signature, we took advantage of a previously published Affymetrix muscle gene expression dataset from a smaller independent mixed exercise training cohort³⁵. Importantly, this cohort also spanned a broad range in terms of the training-induced change in the rate of peripheral glucose disappearance (Rd) (ΔR_d ranging from -20 to $+145\%$). Intriguingly, the baseline multi-gene RNA signature was able to explain 30% of the training-induced change in Rd among the healthy middle-aged male participants ($n = 14$) in the replication cohort (see Supplementary Fig. S7, a value close to the estimated ΔS_i variance in the HERITAGE cohort when family membership has been accounted for).

Discussion

Here, we have shown that variation in insulin sensitivity across a normal healthy population and its modulation by EET is a complex trait where combined variation in genes linked to the KEGG pathways *cholinergic signalling*, *calcium signalling*, *axon guidance* and *chemokine signalling* is likely to be an important component. Despite such

complexity, we have been able to identify genes that are causally linked to glucose disposal and other relevant phenotypic traits in mouse knockouts.

GWAS candidate genes associated to baseline and ΔS_i . Our study is the first to investigate genome-wide associations with changes in S_i in response to exercise training. Despite the SNPs we identified in the present study as significantly associated to baseline S_i have not been previously associated to insulin sensitivity according to GWAS Catalogue database³⁶, some of them have been previously linked to insulin resistance or related phenotypes. Importantly, mouse knockouts for four of these candidates (*TGM6*, *CAND2*, *IQSEC1* and *DNAL1*) showed a relevant phenotype, suggesting a causal link with S_i . Here we review the evidence in the literature that is consistent with our findings.

The function of *TGM6* product, transglutaminase 6 (TG6), has not been studied extensively and is not yet well understood. However, transglutaminase 2 (TG2) has been implicated in glucose metabolism^{37,38} and glucose tolerance³⁹. Transglutaminases catalyse serotonylation⁴⁰, a process involved in the modulation of insulin secretion in pancreatic beta-cells⁴¹. The results of the *TGM6* knockout experiment and the function of TG2 suggest a role for TG6 in S_i and warrants further investigation.

There are several pieces of evidence that are consistent with a role of *CAND2* in insulin sensitivity. *CAND2* is mostly expressed in muscle tissues^{17,42}. Beyond its role in myogenic differentiation^{43–45}, *CAND2* interacts with insulin receptor substrate 1 (IRS1) and is stimulated by insulin in type 2 diabetes patients, but not in non-diabetic controls⁴⁶. *CAND2* acts by modulating the assembly of ubiquitin–proteasome related complexes, such as E3 ligases^{47,48}, which also have been implicated in insulin resistance and diabetes and are known to target key insulin signalling molecules⁴⁹. Moreover, *CAND2* has been shown to be upregulated during a 3 h hyperinsulinemic euglycemic clamp in vastus lateralis muscle of healthy subjects⁵⁰. *CAND2* has been previously mapped to SNPs associated with related phenotypes such as waist-hip ratio⁵¹ and waist circumference adjusted for BMI^{52,53}.

SNP rs11622678 located in chromosome 14 reached a statistically significant association with baseline S_i and was positionally mapped to *DNAL1*. Mutations in this gene cause primary ciliary dyskinesia as this gene affects movement of cilia and flagella⁵⁴, and have been associated with respiratory diseases and lung function^{18,51}. Interestingly, knockout of this gene led to increased respiratory exchange ratio (RER), with higher values indicating that carbohydrates are the main source of substrates being oxidized (Fig. 3).

Two SNPs ($r^2 = 0.49$) have been positionally mapped to the fusion transcript *IQCJ-SCHIP1* spanning two adjacent independent genes. Although the functions of *IQCJ-SCHIP1* are still poorly understood, genetic variants tagging *IQCJ* have been associated to modulation of blood lipid levels in multiple independent studies^{55–58}, while *SCHIP1* has been implicated in axon guidance^{59–61} and was upregulated in differentiated myotubes compared to undifferentiated⁶² (see Supplementary Material for additional discussion on PNMA1, CHODL, SPATA16 and CDH13 which are either not present in the IMPC database or no relevant traits showed significant changes following knockout).

Furthermore, the observation that baseline S_i negatively correlates with ΔS_i (Supplementary Fig. S2C) is intriguing and suggestive of the existence of an upper limit for S_i . This would result in a lower margin for improvement in individuals with a high S_i value.

Is calcium mobilization triggered by muscle contraction potentially responsible for changes in S_i ? In addition to the genes discussed above, we identified a consistent accumulation of SNPs correlated to baseline S_i and ΔS_i in *calcium signalling* and *cholinergic synapse* pathways (Supplementary Table S2). These findings suggest that cholinergic signalling via mobilization of calcium in skeletal muscle may mechanistically link muscle contraction to insulin sensitivity. A study investigating correlation between baseline gene expression and exercise-induced $\% \Delta S_i$ found that several significantly correlated genes in a validation cohort mapped to Ca^{2+} signalling, including *CACNA1S* and *CAMK2D*⁶³.

There is considerable evidence that insulin signalling and muscle contraction are linked processes that activate multiple signalling cascades leading to glucose uptake^{64–67}. Glucose uptake by skeletal muscle tissue is mediated by GLUT4, which upon stimulation by either insulin and/or contraction is translocated to the plasma membrane from vesicles. In insulin-mediated glucose uptake, insulin binding triggers a cascade of molecular reactions that lead to GLUT4 translocation, also triggering a transient Ca^{2+} influx in muscle cells. This process is suggested to enhance GLUT4 translocation and docking in the plasma membrane. Meanwhile, muscle contraction initiated by membrane depolarisation and increased concentrations in intracellular Ca^{2+} leads to activation of Ca^{2+} sensors such as Ca^{2+} -calmodulin-dependent kinase II (CAMKII), which are key molecules in contraction-stimulated glucose transport. Activated CAMKII promote dissociation of HDAC4 from MEF2A transcription factor, leading to its activation and increased GLUT4 transcription^{68,69}. Therefore, insulin- and contraction-mediated signalling pathways related to skeletal muscle glucose uptake are inter-twined, where GLUT4 increased expression and availability entrained by muscle contraction would also contribute to improved insulin-mediated glucose uptake.

It is therefore conceivable that variation in genes controlling muscle contraction (represented in the cholinergic synapse and calcium signalling pathways) could result in lower levels of intracellular Ca^{2+} , leading to lower activation of the CAMKII and consequently reduced GLUT4 availability (Fig. 4B). Diminished localization of GLUT4 to the membrane, which is also Ca^{2+} dependent, could also contribute to glucose uptake impairment in skeletal muscle. Interestingly, none of the analyses shown here directly linked *SCL2A4* (gene encoding for GLUT4) to either baseline or ΔS_i . It is possible that our study, is not sufficiently well powered to capture *SCL2A4* effect size, or that other mechanisms affecting GLUT4 regulation, such as post-translation modifications⁷⁰, modulate insulin sensitivity. Additionally, none of the publicly available GWAS studies on insulin sensitivity (GWAS Catalog database³⁶) reported associations with *SCL2A4*.

Genetic variation may not be the only mechanism controlling glucose mobilization and ultimately S_i . A study investigating epigenetic patterns associated to type 2 diabetes has shown that first-degree relatives of patients with diabetes have differential DNA methylation patterns in genes related to insulin and Ca^{2+} signalling pathways compared to healthy individuals with no family history of the disease. Intriguingly, DNA methylation of genes involved in Ca^{2+} signalling pathways including *MEF2A*, which we also have identified in our approach, decreased after exercise⁷¹. The role of genetic and epigenetic variation in Ca^{2+} signalling in modulating inter-individual variability in insulin sensitivity warrants further investigation.

Other potential mechanisms linking genetic variation to S_i . Our analyses suggest that genetic variation affecting other biological mechanisms could also be modulating S_i . The chemokine signalling pathway has been identified in both GWAS and transcriptomics functional analyses, suggesting that variation in genes within this pathway could be affecting their expression and contributing to variation in exercised-induced S_i response.

Several studies have suggested a role of chemokines and chemokine receptors in the development of insulin resistance, which is attributed in part to a state of low-grade inflammation due to elevated blood glucose and lipid levels induced by diet and excess adiposity^{72,73}. This leads to induction of pro-inflammatory mediators such as chemokines that interfere with insulin signalling pathways. In the present study, we have identified gene expression profiles associated to inflammation-related pathways correlated with S_i response, suggesting genetic variation affecting chemokine signalling could affect the inflammatory processes that naturally occur with exercise. Additionally, the emerging research field of ‘immunometabolism’⁷⁴ has generated data indicating that a cross-talk between immune- and metabolic-related molecules is essential to normal skeletal muscle physiology⁷⁵.

Cell adhesion molecules pathway, which include key molecules involved in modulating ECM integrity, was also identified in both GWAS and transcriptomics functional analyses. There is some evidence linking insulin resistance to ECM remodelling⁷⁵ with mechanisms attributed to physical impairment of insulin access to its receptor by increased ECM deposition, or to the roles of integrins in mediating insulin signalling⁷⁶. Deletion of the muscle-specific integrin $\beta 1$ (expressed by *ITGB1*) results in decreased insulin sensitivity, whereas the ECM of insulin resistant human muscle were reported to be associated to decreased abundance α -actinin 2 (expressed by *ACTN2*). Interestingly, in our functional investigation encompassing ΔS_i associated genetic variants, two SNPs mapping to *ACTN2* and *ITGB1* genes are second and third top significant SNPs ($p < 0.001$), but several other top associated SNPs map to cell adhesion molecules, suggesting that mutations in ECM-related molecules could be influencing the S_i -related traits at rest or in response to regular exercise.

Conclusions

The relatively large proportion of individuals who fail to improve metabolic fitness traits justify the importance of developing evidence-based personalized exercise prescription to maximize the health-promoting benefits of a physically active lifestyle. To develop such individualized recommendations for exercise, it is vital that the molecular basis driving phenotypic response variation be understood. Our multi-omics approach is a step in this direction as it provides evidence of a genetic component affecting calcium signalling that might be responsible for the large heterogeneity in ΔS_i following a fully supervised EET program. The predictive RNA signature can potentially be used to stratify individuals before any intervention has taken place. Further studies are needed to test whether our signature could be predictive of response in different training protocols or whether ΔS_i non-responders could benefit from different training regimes (e.g. high-intensity interval training or resistance exercise). This is important as skeletal muscle insulin resistance is one of the earliest hallmarks of the development of type 2 diabetes and other metabolic complications. Promisingly, muscle RNA abundance can now be more easily quantified due to the development of less invasive micro-needle biopsy sampling. Further, one-step multiplex real-time RT-PCR assays could offer a rapid, sensitive and cheap diagnostic option if a molecular predictor could be validated and replicated in multiple cohorts.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 14 January 2021; Accepted: 25 August 2021

Published online: 02 December 2021

References

1. Knowler, W. C. *et al.* Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N. Engl. J. Med.* **346**, 393–403 (2002).
2. Tuomilehto, J. *et al.* Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N. Engl. J. Med.* **344**, 1343–1350 (2001).
3. Koval, J. A. *et al.* Effects of exercise and insulin on insulin signaling proteins in human skeletal muscle. *Med. Sci. Sports Exerc.* **31**, 998–1004 (1999).
4. Ferrannini, E. *et al.* The disposal of an oral glucose load in patients with non-insulin-dependent diabetes. *Metab. Clin. Exp.* **37**, 79–85 (1988).
5. Boulé, N. G. *et al.* Effects of exercise training on glucose homeostasis: The HERITAGE family study. *Diabetes Care* **28**, 108–114 (2005).
6. Teran-Garcia, M., Rankinen, T., Koza, R. A., Rao, D. C. & Bouchard, C. Endurance training-induced changes in insulin sensitivity and gene expression. *Am. J. Physiol. Endocrinol. Metab.* **288**, E1168–E1178 (2005).
7. Huffman, K. M. *et al.* Metabolite signatures of exercise training in human skeletal muscle relate to mitochondrial remodelling and cardiometabolic fitness. *Diabetologia* **57**, 2282–2295 (2014).
8. Bouchard, C. Genomic predictors of trainability. *Exp. Physiol.* **97**, 347–352 (2012).

9. Yaribeygi, H., Farrokhi, F. R., Butler, A. E. & Sahebkar, A. Insulin resistance: Review of the underlying molecular mechanisms. *J. Cell Physiol.* **234**, 8152–8161 (2019).
10. Bouchard, C. *et al.* Aims, design, and measurement protocol. *Med. Sci. Sports Exerc.* **27**, 721–729 (1995).
11. Walton, C., Godsland, I. F., Proudder, A. J., Felton, C. & Wynn, V. Evaluation of four mathematical models of glucose and insulin dynamics with analysis of effects of age and obesity. *Am. J. Physiol. Endocrinol. Metab.* **262**, E755–E762 (1992).
12. Boston, R. C. *et al.* MINMOD Millennium: A computer program to calculate glucose effectiveness and insulin sensitivity from the frequently sampled intravenous glucose tolerance test. *Diabetes Technol. Ther.* **5**, 1003–1015 (2003).
13. Bouchard, C. *et al.* Genomic predictors of the maximal O₂ uptake response to standardized exercise training programs. *J. Appl. Physiol.* **1985**(110), 1160–1170 (2011).
14. Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models.* (2020).
15. Gao, X., Starmer, J. & Martin, E. R. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet. Epidemiol.* **32**, 361–369 (2008).
16. R Core Team. *R: A Language and Environment for Statistical Computing.* (R Foundation for Statistical Computing, 2018).
17. GTEx Consortium *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
18. Dickinson, M. E. *et al.* High-throughput discovery of novel developmental phenotypes. *Nature* **537**, 508–514 (2016).
19. Chai, H.-S. *et al.* GLOSSI: A method to assess the association of genetic loci-sets with complex diseases. *BMC Bioinform.* **10**, 102 (2009).
20. McHugh, C., Larson, J. & Hackney, J. *cpvSNP: Gene Set Analysis Methods for SNP Association p-Values that Lie in Genes in Given Gene Sets.* (Bioconductor version: Release (3.11), 2020). <https://doi.org/10.18129/B9.bioc.cpvSNP>.
21. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).
22. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551 (2021).
23. Mukund, K. & Subramanian, S. Co-expression network approach reveals functional similarities among diseases affecting human skeletal muscle. *Front. Physiol.* **8**, 980 (2017).
24. Phillips, B. E. *et al.* Molecular networks of human muscle adaptation to exercise and age. *PLoS Genet.* **9**, e1003389 (2013).
25. Edgar, R. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
26. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy—Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics (Oxford, England)* **20**, 307–315 (2004).
27. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Bioinformatics (Oxford, England)* **8**, 118–127 (2007).
28. Li, Q., Birkbak, N. J., Györfy, B., Szallasi, Z. & Eklund, A. C. Jetset: Selecting the optimal microarray probe set to represent a gene. *BMC Bioinform.* **12**, 474 (2011).
29. Rico-Sanz, J. *et al.* Familial resemblance for muscle phenotypes in the HERITAGE Family Study. *Med. Sci. Sports Exerc.* **35**, 1360–1366 (2003).
30. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
31. von Mering, C. *et al.* STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* **33**, D433–437 (2005).
32. Helge, J. W., Kriketos, A. D. & Storlien, L. H. Insulin sensitivity, muscle fibre types, and membrane lipids. *Adv. Exp. Med. Biol.* **441**, 129–138 (1998).
33. Fisher, G. *et al.* Associations of human skeletal muscle fiber type and insulin sensitivity, blood lipids, and vascular hemodynamics in a cohort of premenopausal women. *Eur. J. Appl. Physiol.* **117**, 1413–1422 (2017).
34. Pataky, M. W. *et al.* Skeletal muscle fiber type-selective effects of acute exercise on insulin-stimulated glucose uptake in insulin-resistant, high-fat-fed rats. *Am. J. Physiol. Endocrinol. Metab.* **316**, E695–E706 (2019).
35. Meex, R. C. R. *et al.* Restoration of muscle mitochondrial function and metabolic flexibility in type 2 diabetes by exercise training is paralleled by increased myocellular fat storage and improved insulin sensitivity. *Diabetes* **59**, 572–579 (2010).
36. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
37. Rossin, F. *et al.* Transglutaminase 2 ablation leads to mitophagy impairment associated with a metabolic shift towards aerobic glycolysis. *Cell Death Differ.* **22**, 408–418 (2015).
38. Kumar, S., Donti, T. R., Agnihotri, N. & Mehta, K. Transglutaminase 2 reprogramming of glucose metabolism in mammary epithelial cells via activation of inflammatory signaling pathways. *Int. J. Cancer* **134**, 2798–2807 (2014).
39. Bernassola, F. *et al.* Role of transglutaminase 2 in glucose tolerance: knockout mice studies and a putative mutation in a MODY patient. *FASEB J.* **16**, 1371–1378 (2002).
40. Bader, M. Serotonylation: Serotonin signaling and epigenetics. *Front. Mol. Neurosci.* **12**, 288 (2019).
41. Paulmann, N. *et al.* Intracellular serotonin modulates insulin secretion from pancreatic beta-cells by protein serotonylation. *PLoS Biol.* **7**, e1000229 (2009).
42. Aoki, T. *et al.* TIP120B: A novel TIP120-family protein that is expressed specifically in muscle tissues. *Biochem. Biophys. Res. Commun.* **261**, 911–916 (1999).
43. Aoki, T., Okada, N., Wakamatsu, T. & Tamura, T. TBP-interacting protein 120B, which is induced in relation to myogenesis, binds to NOT3. *Biochem. Biophys. Res. Commun.* **296**, 1097–1103 (2002).
44. Shiraishi, S. *et al.* TBP-interacting protein 120B (TIP120B)/cullin-associated and neddylation-dissociated 2 (CAND2) inhibits SCF-dependent ubiquitination of myogenin and accelerates myogenic differentiation. *J. Biol. Chem.* **282**, 9017–9028 (2007).
45. Suzuki, H., Suzuki, A., Maekawa, Y., Shiraishi, S. & Tamura, T. Interplay between two myogenesis-related proteins: TBP-interacting protein 120B and MyoD. *Gene* **504**, 213–219 (2012).
46. Caruso, M. *et al.* Increased interaction with insulin receptor substrate 1, a novel abnormality in insulin resistance and type 2 diabetes. *Diabetes* **63**, 1933–1947 (2014).
47. You, J., Wang, M., Aoki, T., Tamura, T. & Pickart, C. M. Proteolytic targeting of transcriptional regulator TIP120B by a HECT domain E3 ligase. *J. Biol. Chem.* **278**, 23369–23375 (2003).
48. Liu, X. *et al.* Cand1-mediated adaptive exchange mechanism enables variation in F-box protein expression. *Mol. Cell* **69**, 773–786 (2018).
49. Yang, X.-D., Xiang, D.-X. & Yang, Y.-Y. Role of E3 ubiquitin ligases in insulin resistance. *Diabetes Obes. Metab.* **18**, 747–754 (2016).
50. Rome, S. *et al.* Microarray profiling of human skeletal muscle reveals that insulin regulates approximately 800 genes during a hyperinsulinemic clamp. *J. Biol. Chem.* **278**, 18063–18068 (2003).
51. Kichaev, G. *et al.* Leveraging polygenic functional enrichment to improve GWAS power. *Am. J. Hum. Genet.* **104**, 65–75 (2019).
52. Tachmazidou, I. *et al.* Whole-genome sequencing coupled to imputation discovers genetic signals for anthropometric traits. *Am. J. Hum. Genet.* **100**, 865–884 (2017).
53. Zhu, Z. *et al.* Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *J. Allergy Clin. Immunol.* **145**, 537–549 (2020).

54. Mazor, M. *et al.* Primary ciliary dyskinesia caused by homozygous mutation in DNAL1, encoding dynein light chain 1. *Am. J. Hum. Genet.* **88**, 599–607 (2011).
55. Bandesh, K. *et al.* Genome-wide association study of blood lipids in Indians confirms universality of established variants. *J. Hum. Genet.* **64**, 573–587 (2019).
56. Vallée Marcotte, B. *et al.* Plasma triglyceride levels may be modulated by gene expression of IQCJ, NXP1, PHF17 and MYB in humans. *Int. J. Mol. Sci.* **18**, 257 (2017).
57. Vallée Marcotte, B. *et al.* Novel genetic loci associated with the plasma triglyceride response to an omega-3 fatty acid supplementation. *J. Nutrigenet. Nutrigenom.* **9**, 1–11 (2016).
58. Rudkowska, I. *et al.* Genome-wide association study of the plasma triglyceride response to an n-3 polyunsaturated fatty acid supplementation. *J. Lipid Res.* **55**, 1245–1253 (2014).
59. Kwaśnicka-Crawford, D. A., Carson, A. R. & Scherer, S. W. IQCJ-SCHIP1, a novel fusion transcript encoding a calmodulin-binding IQ motif protein. *Biochem. Biophys. Res. Commun.* **350**, 890–899 (2006).
60. Klingler, E. *et al.* The cytoskeleton-associated protein SCHIP1 is involved in axon guidance, and is required for piriform cortex and anterior commissure development. *Development* **142**, 2026–2036 (2015).
61. Martin, P.-M. *et al.* Schwannomin-interacting protein 1 isoform IQCJ-SCHIP1 is a multipartner ankyrin- and spectrin-binding protein involved in the organization of nodes of Ranvier. *J. Biol. Chem.* **292**, 2441–2456 (2017).
62. Tsumagari, K. *et al.* Gene expression during normal and FSHD myogenesis. *BMC Med. Genom.* **4**, 67 (2011).
63. Barberio, M. D. *et al.* PDP1 gene expression correlates with exercise-training insulin sensitivity changes. *Med. Sci. Sports Exerc.* **48**, 2387–2397 (2016).
64. Zierath, J. R., Krook, A. & Wallberg-Henriksson, H. Insulin action and insulin resistance in human skeletal muscle. *Diabetologia* **43**, 821–835 (2000).
65. Lanner, J. T., Bruton, J. D., Katz, A. & Westerblad, H. Ca²⁺ and insulin-mediated glucose uptake. *Curr. Opin. Pharmacol.* **8**, 339–345 (2008).
66. Kjøbsted, R. *et al.* Enhanced muscle insulin sensitivity after contraction/exercise is mediated by AMPK. *Diabetes* **66**, 598–612 (2017).
67. Sylow, L., Kleinert, M., Richter, E. A. & Jensen, T. E. Exercise-stimulated glucose uptake—regulation and implications for glycaemic control. *Nat. Rev. Endocrinol.* **13**, 133–148 (2017).
68. Ojuka, E. O., Goyaram, V. & Smith, J. A. H. The role of CaMKII in regulating GLUT4 expression in skeletal muscle. *Am. J. Physiol. Endocrinol. Metab.* **303**, E322–E331 (2012).
69. Niu, Y. *et al.* Exercise-induced GLUT4 transcription via inactivation of HDAC4/5 in mouse skeletal muscle in an AMPKα2-dependent manner. *Biochimica et Biophysica Acta (BBA) Mol. Basis Dis.* **1863**, 2372–2381 (2017).
70. Sadler, J. B. A., Bryant, N. J., Gould, G. W. & Welburn, C. R. Posttranslational modifications of GLUT4 affect its subcellular localization and translocation. *Int. J. Mol. Sci.* **14**, 9963–9978 (2013).
71. Nitert, M. D. *et al.* Impact of an exercise intervention on DNA methylation in skeletal muscle from first-degree relatives of patients with type 2 diabetes. *Diabetes* **61**, 3322–3332 (2012).
72. Yao, L., Herlea-Pana, O., Heuser-Baker, J., Chen, Y. & Barlic-Dicen, J. Roles of the chemokine system in development of obesity, insulin resistance, and cardiovascular disease. *J. Immunol. Res.* **2014**, e181450 (2014).
73. Xu, L., Kitade, H., Ni, Y. & Ota, T. Roles of chemokines and chemokine receptors in obesity-associated insulin resistance and nonalcoholic fatty liver disease. *Biomolecules* **5**, 1563–1579 (2015).
74. Hotamisligil, G. S. Foundations of immunometabolism and implications for metabolic health and disease. *Immunity* **47**, 406–420 (2017).
75. Mukund, K. & Subramaniam, S. Skeletal muscle: A review of molecular structure and function, in health and disease. *WIREs Syst. Biol. Med.* **12**, e1462 (2020).
76. Williams, A. S., Kang, L. & Wasserman, D. H. The extracellular matrix and insulin resistance. *Trends Endocrinol. Metab.* **26**, 357–366 (2015).

Acknowledgements

We thank Drs. Arthur S. Leon, D.C. Rao, James S. Skinner, Tuomo Rankinen, Jacques Gagnon, Treva Rice and the late Jack H. Wilmore for contributions to the planning, data collection, and conduct of the HERITAGE project, and for the data management of the IVGTT. Special thanks are also given to Richard S. Bergman from the Cedars-Sinai Medical Centre for his expertise in the analyses of the IVGTT data. This research was partially funded by National Heart, Lung, and Blood Institute Grants HL-45670, HL-47317, HL-47321, HL-47323, and HL-47327, all in support of the HERITAGE Family Study). L.Y. Takeshita has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under the TransBioLine project with grant agreement No. 821283. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. This communication reflects the author's view and neither IMI nor the European Union or EFPIA are responsible for any use that may be made of the information contained therein. P. K. Davidsen was supported by a PhD studentship funded by the Birmingham MRC-ARUK Centre for Musculoskeletal Ageing Research. C. Bouchard is partially funded by the John W. Barton Sr. Chair in Genetics and Nutrition. Z. S. Ghosh and C. Bouchard are partially supported by the National Institute of General Medical Sciences (NIGMS)-funded COBRE Grant 8-P30-GM-118430-01. S. Ghosh is supported in part by NIGMS Grant 2-U54-GM-104940, which funds the Louisiana Clinical and Translational Science Center and by the National Medical Research Council, Ministry of Health, Singapore (WBS R913200076263). M. A. Sarzynski is partially supported by NHLBI Grant R01HL146462 and NIGMS Grant P20GM103499, which funds the South Carolina IDeA Network of Biomedical Research Excellence.

Author contributions

M.K.C.H., P.S. and C.B. participated in the exercise training study design. L.Y.T., P.K.D. and F.F. formulated the data analysis plan. L.Y.T. and P.K.D. performed data analysis. J.M.H., P.A. and F.F. contributed to overall data analysis. M.A.S., S.G. and C.B. contributed to the GWAS analysis. C.B. contributed to the quality control and modelling of the glucose and insulin data. L.Y.T., P.K.D. and F.F. drafted the manuscripts. F.F. produced Fig. 4 and L.Y.T. produced all other manuscript figures. S.J.W., J.M.R. and R.E.G. substantially contributed to manuscript revision providing expert knowledge on physiological aspects involving insulin sensitivity. All co-authors contributed to the critical review and editing of the manuscript. F.F. is the guarantor of this work and, as such,

had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-98792-1>.

Correspondence and requests for materials should be addressed to F.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021