Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Exploring Rare Genetic Variation Underlying Metabolic Traits in the Fenland Cohort

John Perry

john.perry@ims.cam.ac.uk

University of Cambridge

Jack Murzynowski

University of Cambridge https://orcid.org/0009-0000-5341-2322

Nicola Kerrison

University of Cambridge

Felix Day

University of Cambridge

Jian'an Luan

University of Cambridge

Emily Morbey

University of Cambridge

Sam Lockhart

University of Cambridge

Søren Brage

University of Cambridge

Stephen O'Rahilly

MRC Metabolic Diseases Unit, Institute of Metabolic Science – Metabolic Research Laboratories, University of Cambridge, Cambridge, UK. https://orcid.org/0000-0003-2199-4449

Ken Ong

University of Cambridge

Nick Wareham

Article

Keywords:

Posted Date: August 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10473668/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Yes there is potential Competing Interest. J.R.B.P. is an employee/shareholder of Insmed Innovation UK / Insmed Inc., receives research funding from GSK and consultancy fees from Menwell Ltd and Ovartix Ltd. S.L. is currently a full-time employee of GSK outside of this research. The remaining authors declare no competing interests.

Exploring Rare Genetic Variation Underlying Metabolic Traits in the Fenland Cohort

Jack R.A. Murzynowski[1], Nicola D. Kerrison[1], Felix R. Day[1], Jian'an Luan[1], Emily J. Morbey[1], Sam M. Lockhart[2], Søren Brage[1], Stephen O'Rahilly[2], Ken K. Ong[1], John R.B. Perry[1*], Nicholas J. Wareham[1]

** Corresponding author*

Affiliations:

1. IMS Epidemiology, Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge, UK.
2. Metabolic Research Laboratory, Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge, UK.

Abstract

Recent large-scale whole-exome sequencing (WES) studies in population biobanks have transformed our understanding of the role of rare protein coding on human health. However, the size and design of these biobanks limit deep phenotyping and genotype-targeted follow-up. To address this, we present new WES in the Fenland study, a population-based cohort of 11,458 individuals with extended metabolic phenotyping. Using this resource, we performed the first large-scale rare variant association analysis of measured Resting Energy Expenditure (REE) and fasting/post-challenge measures of glucose and insulin. We identified damaging variants in *BRSK2, NID2* and *PC* influencing glycaemic traits; and in *COL8A1* and *CASQ1* reducing REE. Notably, the *CASQ1* association was sex- and menopause status-dependent, highlighting a potential estrogen-mediated mechanism by which menopause may contribute to declining metabolic health. Collectively, our findings demonstrate the utility of the Fenland study as a resource for genetic discovery.

Introduction

Whole-exome sequencing (WES) efforts in large-scale population biobanks such as UK Biobank (UKB) (1) and the All of Us Research Program (AoU) (2) have transformed our understanding of the role of rare genetic variation with disease and health-related traits. Unlike genotyping arrays, which assay common, largely non-coding, variants, WES enables assessment of rare protein-coding variation that can exert substantial effects on gene function and disease risk, and hence directly implicate causal genes (3). While whole-genome sequencing (WGS) extends coverage beyond coding regions, variation in non-coding sequence is generally more challenging to interpret and, at current cohort sizes, has yielded fewer associated variants with large effect sizes than protein-coding variation (4). Consequently, WES data remains a powerful approach, and to date hundreds of studies

have applied single-variant and gene-based burden approaches in these cohorts, identifying multiple gene-trait associations (5, 6).

Despite these advances, several challenges remain. The scale and design of large biobanks, while statistically powerful, often compromise on the depth of phenotyping. Many traits are estimated through self-report or single time-point measurements, and opportunities are limited for genotype-directed recall and detailed physiological follow-up. Rare variant association studies have also been limited by a lack of replication in independent datasets, reflecting challenges related to cohort size, phenotype availability, and variant sparsity (7-10). Well-characterised population-based cohorts with deep phenotyping therefore provide an important complement to large biobank resources, enabling more nuanced investigation of previously identified loci, particularly where consent for recall-by-genotype studies is in place. The Fenland study is one such cohort, comprising of ~12,000 individuals aged 29-64 years old at baseline with extensive metabolic phenotyping, including oral glucose tolerance testing (OGTT), detailed anthropometry, and physiological assessment (11). This enables interrogation of traits that are not routinely captured in larger studies, and prior genome-wide analyses using array-based genotyping in Fenland have identified common variant associations across these phenotypes (12-14). We now report the generation, quality control, and release of WES data in this cohort, extending genetic investigation to rare coding variation.

Using Fenland WES data, we develop a pipeline for performing rare variant gene-burden tests and apply it to disease-relevant metabolic traits that are not measured in existing large biobanks. In doing so we identify several new gene-associations such as *CASQ1* with resting energy expenditure (REE) and *BRSK2* with fasting glucose. Here we present the quality control, analysis and emerging results of the Fenland WES data release, which is available to researchers on request.

Results

Quality control of Whole Exome Sequencing Data

Generation and initial quality control (QC) of WES data in the Fenland study was performed by the Wellcome Sanger Institute (15). The QC pipeline included sample-level checks, ancestry assignment, and variant-level filtering using a random forest classifier, which achieved ~99.0% test accuracy when benchmarked against Genome-in-a-Bottle reference data (16). Variants were stratified into relaxed, medium, and stringent QC sets based on classifier output and genotype-level metrics (Supplementary Table 1). Only stringent variants showed expected allele counts (AC) in this sample that scaled with its sample size versus the 470K UK Biobank (UKB) WES dataset (38.7× higher in UKB; R^2 = 0.98), with consistent scaling also across gene-burden masks (Supplementary Figures 1-2). Consequently, only the stringent variant set was used. Following QC, 11,478 individuals and

3,379,158 variants remained for downstream analyses. Cohort level summary statistics are presented in Table 1.

Table 1: Baseline characteristics of the Fenland whole exome sequencing sample

Characteristic	Fenland WES cohort (N = 11478)
Demographics	
Age - years	48.6 ± 7.5 (29.4-64.7)
BMI	26.9 ± 4.8 (13.8-59.9)
Sex	
Female	6188 (53.9%)
Male	5289 (46.1%)
Ancestry	
European	11083 (96.6%)
South Asian	140 (1.2%)
African	119 (1.0%)
East Asian	102 (0.9%)
Admixed American	34 (0.3%)
Relatedness	
Unrelated	8923 (77.7%)
1st degree	1011 (8.8%)
2nd degree	285 (2.5%)
3rd degree	1259 (11.0%)

Continuous variables are presented as mean ± standard deviation, with the observed range in parentheses. Categorical variables are presented as number and percentage of participants. Genetic ancestry was assigned by projection onto principal components derived from the 1000 Genomes reference panel. Relatedness was estimated using KING and classified according to the closest identified relationship. BMI, body mass index; WES, whole-exome sequencing.

To further validate the new WES dataset, we used UKB as an external discovery cohort to test whether coding variant associations could be replicated in Fenland by performing exome-wide association studies (ExWAS) and gene-burden tests for height and BMI in both cohorts. For common coding variants, defined as minor allele frequency (MAF) >1%, UKB-associated variants showed strong directional concordance in Fenland, particularly at more stringent UKB significance thresholds (Supplementary Figure 3). For rare variant gene-burden tests (MAF ≤1%), concordance was strong for associations with height, with 82% of exome-wide significant genes in UKB showing concordant effect directions in Fenland. Concordance could not be reliably assessed for BMI, as too few genes met the AC ≥15 inclusion criterion in Fenland. Despite this, we replicated the top UKB signal for BMI (*MC4R*, p = 0.025) and multiple height-associated genes, including *ZFAT* ($p = 4.5\times10^{-6}$) in Fenland.

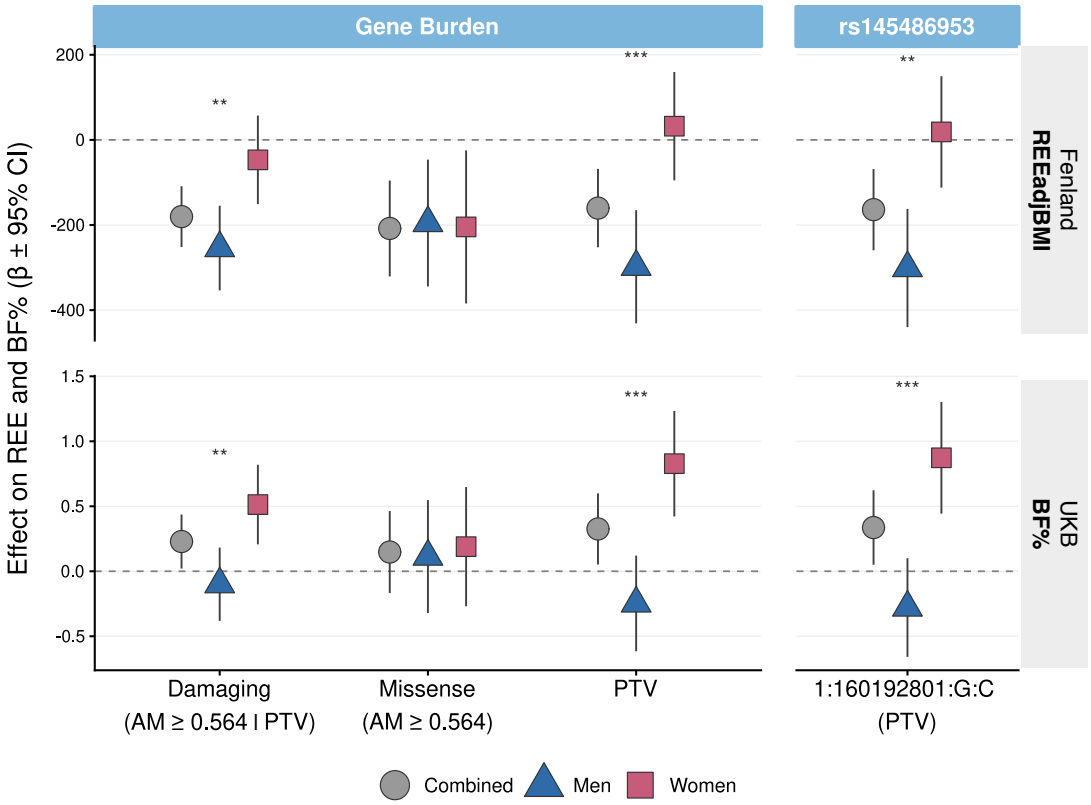
Investigation of Resting Energy Expenditure

Having validated the WES data, we leveraged the phenotypic depth of Fenland to analyse traits not available in large-scale biobanks with WES data, performing rare gene-burden tests for REE and glycaemic traits (fasting and 2-hour post-OGTT glucose and insulin). Analyses were restricted to samples of European ancestry and to genes with AC ≥15, using BOLT-LMM with significance thresholds corrected for the effective number of tests performed (Supplementary Table 2).

We measured REE in Fenland using indirect calorimetry, which estimates energy expenditure from respiratory gas exchange. Due to issues of cost feasibility, large biobanks do not include formal measures of REE but instead rely on estimated values calculated from anthropometry and bioelectrical impedance. However, these perform poorly; in Fenland, estimated REE and measured REE show only moderate concordance (R^2 = 0.54; Supplementary Figure 4), while in UKB estimated REE predominantly reflects fat-free mass (R^2 = 0.99). Leveraging indirect calorimetry in Fenland, we performed gene-burden tests (mask definitions in Supplementary Table 3) of REE adjusted for BMI (REEadjBMI) and identified two significant gene associations: *CASQ1* (Mask=DMG_AM056, AC=70, β=-180kJ/day, P=7.5×10^{-7}) and *COL8A1* (Mask=MISS_AM034, AC=27, β=-297kJ/day, P=3.9×10^{-7}). As a sensitivity analysis, we separately adjusted REE for correlated body composition traits in Fenland, including DEXA-derived lean mass (*CASQ1* β=-169kJ/day, P=5.7×10^{-6}; *COL8A1* β=-313kJ/day, P=1.1×10^{-7}), finding no meaningful attenuation of the associations for either gene (Supplementary Tables 4-5). To assess possible collider bias, we tested associations with BMI and body composition traits across Fenland and UKB. As DEXA data are limited in the UKB (~70,000 participants), we developed predictive models and imputed total lean and fat mass for the whole UKB European ancestry cohort (r^2 = 0.93 and 0.71; Supplementary Table 6, further details in Methods). No strong associations were observed in either study, with only a nominal signal for *CASQ1* with DEXA fat mass in the UKB (Supplementary Table 7).

Given known sex differences in energy expenditure and skeletal muscle biology, we tested for sex-specific effects. Qualifying variants in *CASQ1* conferred a male-specific reduction in REEadjBMI (β men = -254.16kJ/day, P men = 5.6×10-7, β women = -46.91kJ/day, P women = 0.38, P-het = 4.8×10^{-3}) consistent across alternative REE adjustments (Supplementary Table 5). This sex-specificity appeared driven by a single protein truncating variant (PTV), rs145486953 (AC men = 22, AC women = 17, P-het = 1.0×10^{-3}; Figure 1), although the gene-burden association with REE persisted after its exclusion (P = 2.3×10^{-4}). Notably, in UKB, rs145486953 showed female-specific associations with higher body fat percentage (BF%), BMI and total fat mass, with no corresponding associations observed in men (Supplementary Table 8). No sex-specificity was observed for *COL8A1*.

Figure 1: Rare variants in *CASQ1* have sex-specific associations with resting energy expenditure and adiposity

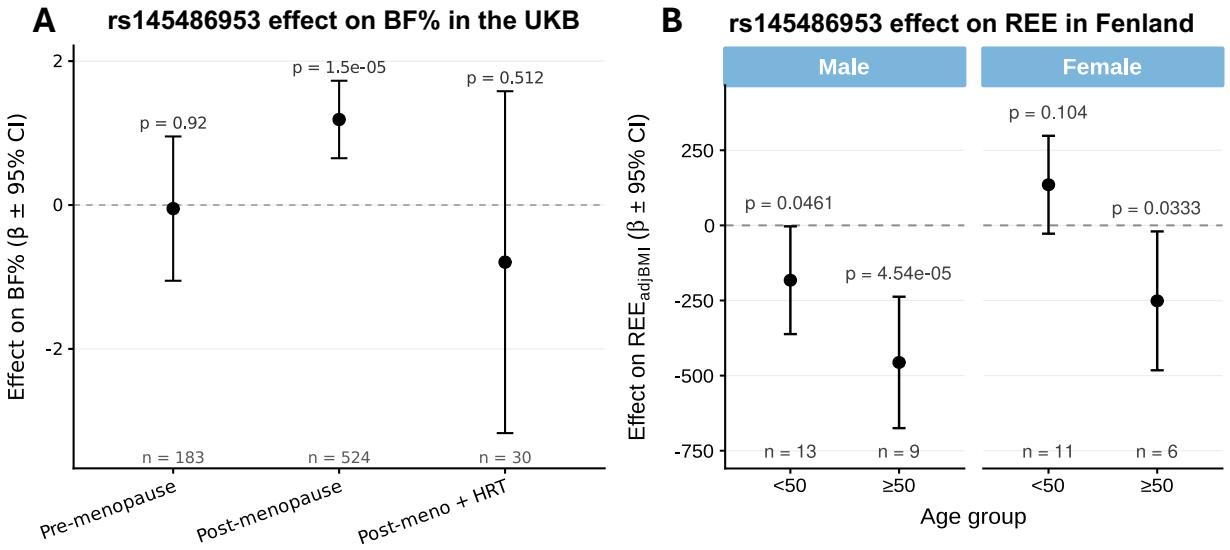


Effect estimates (β; points) and 95% confidence intervals (vertical bars) are shown for sex-combined analyses (grey circles), men (blue triangles) and women (pink squares). The upper panels show associations with resting energy expenditure (REE) measured by indirect calorimetry in the Fenland study and adjusted for body mass index (REEadjBMI; kJ/day). The lower panels show associations with bioimpedance-derived body fat percentage (BF%; percentage points) in UK Biobank (UKB). Gene-burden results are shown for three *CASQ1* masks: damaging variants, comprising protein-truncating variants (PTVs) or missense variants with an AlphaMissense (AM) pathogenicity score ≥0.564; missense variants with an AlphaMissense score ≥0.564; and PTVs alone. The right-hand panels show the corresponding single-variant association for the *CASQ1* PTV rs145486953 (GRCh38 chr1:160192801 G>C). The dashed horizontal line indicates no association. Stars indicate two-sided tests of heterogeneity between the male and female effect estimates: *P* <0.05 *, <0.01** and <0.001***.

To investigate the possible mechanism behind this sex-specificity, we examined whether the rs145486953-BF% association in women was modified by menopause status. In the UKB, rs145486953 was associated with higher BF% only in post-menopausal women (AC = 524, β=1.19%, P = 1.5×10^{-5}), with no effect in pre-menopausal women (AC = 183, β= -0.05%, P = 0.92) and little effect in post-menopausal women taking hormone replacement therapy (HRT) (AC = 30 β=-0.85, P=0.51; Figure 2A). Longitudinal analyses corroborated this finding, with carriers showing greater BF% increases than non-carriers as they transitioned through menopause (Carrier: N=21, $\overline{\Delta}$BF%=4.10%; Non-carrier: N=6,200, $\overline{\Delta}$BF%=2.39%; P-het = 0.049; Supplementary Figure 5), and interaction analyses confirming menopause status, but not age, as a significant BF% modifier (Supplementary Figure 6). In Fenland, where

menopause status was not reported, we used age 50 as a proxy and observed a similar pattern, with rs145486953 being significantly associated with REE in women aged ≥50 years (β = −251.20kJ/day, P = 0.033) but not in those aged <50 years (β = 135.28kJ/day, P = 0.10; Figure 2B).

Figure 2: Effects of rs145486953 on REE and BF% interact with menopause status or age



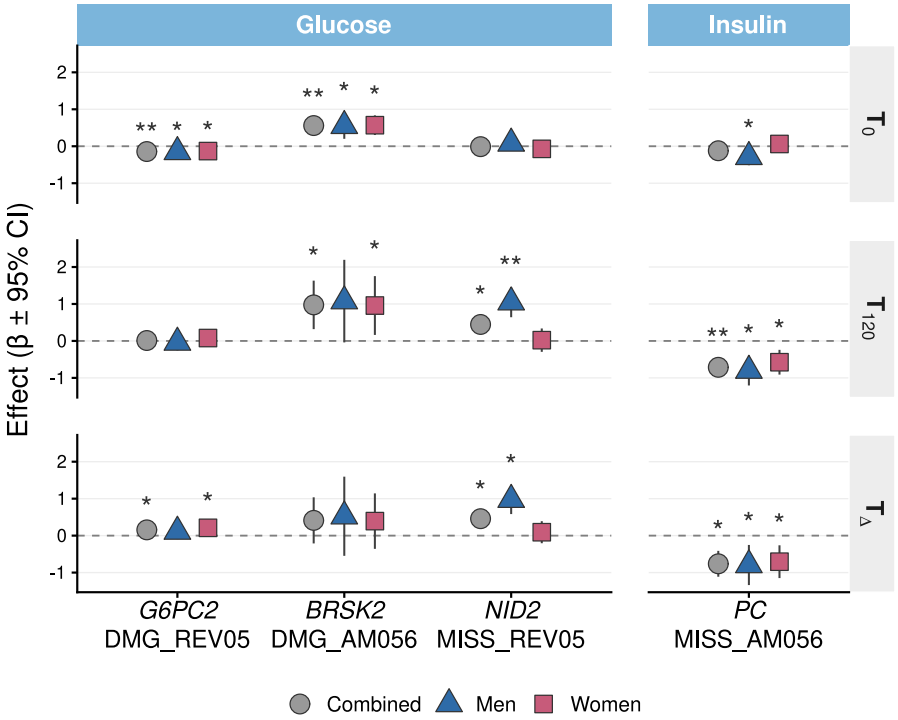
A) Association of rs145486953 with body fat percentage (BF%; percentage points) among pre-menopausal women, post-menopausal women not taking hormone replacement therapy (HRT), and post-menopausal women taking HRT in UK Biobank (UKB). **B)** Association of rs145486953 with body mass index-adjusted resting energy expenditure (REEadjBMI; kJ/day) in Fenland participants stratified by sex and age below or above 50 years. Points show effect estimates (β) and vertical bars show 95% confidence intervals; the dashed horizontal line indicates no association. Values beneath each estimate indicate the alternate allele count within that stratum.

To further explore the potential mechanisms influenced by *CASQ1*, we tested whether circulating sex hormone levels (oestradiol, SHBG, testosterone, and genetically instrumented FSH) might modify the effect of rs145486953 on BF% in UKB; however, we found no clear evidence (Supplementary Figure 7). Extending this to circulating proteomics, only PBXIP1 levels in men were significantly associated with rs145486953 carrier status after multiple testing correction (Men: AC=67, β=0.22, $P=6.1\times10^{-9}$; Women: AC=80, β=0.01, P=0.84, P-het = 3.4×10^{-5}). To assess potential gene dosage effects, we explored the effects of rs145486953 homozygosity across a range of anthropometric and physiological traits. Despite their small numbers in UKB (N=4), homozygotes showed larger reductions in predicted lean mass, consistent with a skeletal muscle role for *CASQ1* (Supplementary Figure 8).

Investigation of glycaemic traits

We investigated glycaemic traits in Fenland (N = 10,622 with OGTT data), performing sex-combined and sex-stratified gene-burden tests for fasting and 2-hour glucose and insulin, adjusting for BMI and restricting to non-diabetic individuals. We identified three novel associations (Figure 3). Rare variants in *BRSK2* were associated with increased fasting glucose (Mask=DMG_AM056, AC=16, β=0.56mmol/L, $P=2.3\times10^{-7}$); *PC* variants were associated with decreased 2-hour insulin (Mask=MISS_AM056, AC=34, β=-0.71log(pmol/L), $P=1.1\times10^{-7}$); and *NID2* variants were associated with increased 2-hour glucose but in men only (Mask=MISS_REV05; Men: AC=46, β=1.04mmol/L, $P=4.2\times10^{-7}$; Women: AC=64, β=0.02, P=0.90, P-het = 9.2×10^{-5}). We additionally recapitulated the known fasting glucose association at *G6PC2* (Mask=DMG_REV05, AC=348, β=-0.14, $P=1.2\times10^{-9}$) (17-20).

Figure 3: Significant gene-burden associations across glycaemic traits in Fenland



Effect estimates (β; points) and 95% confidence intervals (vertical bars) are shown for sex-combined analyses (grey circles), men (blue triangles) and women (pink squares). Associations are presented for fasting measurements (T_0), measurements obtained 2h after a 75g oral glucose tolerance test (T_{120}), and the change between these measurements ($T_\Delta = T_{120}-T_0$). Glucose effect estimates are expressed in mmol/L, while insulin measurements are shown on the log(pmol/L) scale. Results are shown for *G6PC2*, *BRSK2*, *NID2* and *PC* using the gene-burden mask producing the reported association. DMG_REV05 comprises protein truncating variants (PTVs) and missense variants with REVEL >0.5; DMG_AM056 comprises PTVs and missense variants with AlphaMissense ≥0.564; MISS_REV05 comprises missense variants with REVEL >0.5; and MISS_AM056 comprises missense variants with AlphaMissense ≥0.564. The dashed horizontal line indicates no association. A single star denotes a nominal two-sided association ($P < 0.05$), and two stars denote an association exceeding the relevant Bonferroni-corrected exome-wide significance threshold.

To assess independence of these rare variant associations from known common variant effects, we performed conditional analyses including polygenic risk scores for the corresponding trait derived from published MAGIC GWAS results (21). This did not materially attenuate associations for *BRSK2* or *NID2* (Supplementary Table 9), and neither mapped to previously reported MAGIC fasting or 2-hour glucose loci, nor to established type 2 diabetes loci (±1 Mb) (21, 22). No equivalent GWAS data exists for 2-hour insulin; however, *PC* lay outside known MAGIC insulin fold-change loci (±1 Mb), although the region was implicated in type 2 diabetes risk. We next sought replication in UKB using random glucose measurements in non-diabetic individuals. Discovery masks did not directly replicate, however alternative missense masks yielded nominal associations for *BRSK2* with fasting glucose (defined as random glucose restricted to participants reporting 8-24 hours fasting; Mask=MISS_AM034, AC=28, β=0.19, P=0.045) and *NID2* with random glucose, though without evidence of sex-specificity (Mask=MISS_AM034, AC=1331, β=0.03, P=0.032, P-het = 0.74). Insulin phenotypes were unavailable in the UKB, precluding replication of the *PC* signal.

Discussion

Validation of Fenland WES data

Quality control of the Fenland WES data was supported by both technical validation and the replication of known genetic signals. Established gene-trait relationships were observed across anthropometric and glycaemic traits, including *MC4R* for BMI, *ZFAT* for height, and *G6PC2* for fasting glucose, alongside strong directional concordance with UKB results. Together, these findings support the validity and biological consistency of the Fenland WES dataset despite its smaller sample size.

Novel insights into resting energy expenditure

Resting energy expenditure reflects the basal energetic cost of sustaining essential physiological processes and, in most individuals, is the largest component of total energy expenditure (~60-70%) (23, 24). Modest differences in basal energy expenditure can cumulatively influence weight trajectory, highlighting REE as a potential therapeutic target (25-27). However, accurate REE measurement requires calorimetry under controlled conditions, limiting scalability. Most biobanks therefore rely on proxy measures that predominantly reflect body composition and correlate only moderately with calorimetry-based values (22), creating a trade-off between scale and precision. To date, large common variant studies have relied on estimated phenotypes (28, 29), while calorimetry-based coding variant studies have remained small, with the largest previous gene-based burden test (n=262 mixed-ancestry participants) finding no significant associations (30), and the largest coding-variant study being limited to a sample of 373 American Indians (31). Here, we use indirect calorimetry-based REE measurements in ~10,000 Fenland participants to

conduct, to our knowledge, the largest rare variant burden analysis to date, identifying damaging variants in *CASQ1* and *COL8A1* associated with reduced REE.

CASQ1 encodes calsequestrin-1, the major sarcoplasmic reticulum (SR) Ca^{2+} buffer in fast-twitch skeletal muscle (32). Here, we find that damaging *CASQ1* alleles are associated with lower REE in men, and with adiposity and REE in post-menopausal women in UKB. This effect on REE is robust to adjustment for alternative DEXA-derived body composition measures and therefore is unlikely to be confounded by differences in body composition. A potential mechanism is reduced SERCA-dependent ATP turnover; SERCA1a is the dominant Ca^{2+} pump in fast muscle fibres, and Ca^{2+} cycling accounts for a substantial fraction (~40–50%) of muscle resting energy expenditure (33, 34). Accordingly, *CASQ1* loss of function is expected to reduce SR Ca^{2+} buffering, lowering Ca^{2+} cycling and SERCA-dependent ATP turnover at rest, and thereby REE. A role for CASQ1 in REE is supported in mouse models, where complete *Casq1* knockout results in increased REE (35, 36). This opposing direction of effect to our heterozygous effect in humans may reflect species-specific differences in calcium handling and energy metabolism, or differences due to allelic dosage.

The sex specificity of *CASQ1* is biologically plausible. Men have greater type II muscle fibre mass, and therefore a higher contribution of CASQ1-SERCA1a to whole-body energetics, whereas in pre-menopausal women the effect of *CASQ1* variation may be partially buffered by estrogen. Estrogen enhances SERCA activity and expression, and in *Casq1*-null mice markedly lowers heat-induced mortality by reducing SR Ca^{2+} leak and normalising calcium handling (37, 38). Loss of this buffering effect at menopause may therefore unmask the metabolic effects of *CASQ1* variation, consistent with the association with lower REE in men and post-menopausal women, but not pre-menopausal women. The accompanying association with increased adiposity in post-menopausal women may reflect the emergence of this REE deficit during the menopausal transition. By contrast, the absence of a corresponding adiposity signal in men despite lower REE may suggest compensatory physiological or behavioural adaptations, such as reduced energy intake, that offset the effects of reduced energy expenditure over time. This estrogen-mediated model is further supported by the male-specific association of rs145486953 with increased circulating levels of PBXIP1, an enhancer of estrogen receptor signalling (39). While the basis of this sex specificity remains unclear, it may reflect a genetic driver of estrogen sensitivity that is relevant only in a low-estrogen environment, whereas higher systemic estrogen levels in pre-menopausal women could saturate this response.

The possible mechanism underlying the observed association between *COL8A1* and REE is less clear. *COL8A1* encodes a short-chain extracellular matrix collagen implicated in vascular remodelling and angiogenesis (40), with experimental studies showing roles in cell migration and proliferative remodelling (41-43). One possible interpretation is that loss of

COL8A1 reduces the energetic cost of tissue maintenance by limiting extracellular-matrix turnover, though understanding the relevant tissue and mechanism requires further study.

Novel insights into glycaemic traits

We investigated glycaemic traits rarely captured in large-scale biobanks, performing rare gene-burden tests on fasting and post-challenge glucose and insulin, and identifying three novel associations. Variants in *BRSK2* were associated with higher fasting glucose in Fenland, with nominal replication in UK Biobank. This was possible as although UK Biobank only assessed random glucose at baseline, we were able to identify a subset of participants who report as fasted (8+ hours) at time of measurement (though the association was much less significant). *BRSK2* encodes an AMPK-related kinase specifically expressed in pancreas and brain that regulates glucose-stimulated insulin secretion, but experimental studies report opposing effects on β-cell function (44, 45). Our finding of higher fasting glucose in carriers of damaging variants in *BRSK2* support a role for *BRSK2* in linking nutrient sensing to basal insulin secretion.

Variants in *NID2* were associated with a male-specific decrease in 2-hour glucose in Fenland. Nominal replication was observed in UK Biobank using random glucose, a less specific proxy that does not capture the post-challenge, sex-specific effect. *NID2* encodes nidogen-2, a basement-membrane protein that has been shown to inhibit glucagon secretion and regulate glucose homeostasis (46). Notably, experimental overexpression causes increased adiposity and metabolic dysfunction specifically in male mice, supporting our sex-specific findings (47).

Missense variation in *PC* was associated with lower 2-hour insulin. *PC* encodes pyruvate carboxylase, a mitochondrial enzyme required for anaplerosis and the metabolic amplification of insulin secretion (48). Experimental studies consistently show that reduced *PC* impairs glucose-stimulated insulin secretion (49, 50). This finding extends that evidence into human genetics, where such associations have been limited by the lack of post-challenge insulin phenotypes.

Limitations and future research

A key limitation of our study is that replication was unavailable for some of our associations due to the lack of comparable phenotypes in large-scale biobanks. Where related traits were available, these were not always directly equivalent to Fenland measurements. For example, Fenland measured glucose from blood collected into fluoride tubes, which inhibit ex vivo glycolysis and reduce artefactual lowering of glucose after collection, whereas UK Biobank measured glucose in serum following sample separation after a delay of unknown duration, introducing potential for measurement error. Despite this we were able to replicate two of the glucose-associated signals identified, providing reassurance of the robust methodological approaches we have followed. We also note existing *in vitro* and *in vivo* work

supporting a biological role for other observed associations (51). Given the longitudinal design and supportive regulatory permissions of Fenland, recall-by-genotype studies represent a natural next step to directly characterise the biological mechanisms underlying these associations in variant carriers.

Conclusion

In this study, we present whole-exome sequencing data in the Fenland study and establish a rare variant analysis framework for deeply phenotyped metabolic traits. Although large-scale biobanks such as UK Biobank and All of Us have transformed rare variant discovery, their greatest strength is scale rather than physiological resolution. Fenland provides a complementary resource: smaller than those biobanks, but rich in metabolic and physiological phenotyping, with repeated oral glucose tolerance testing, indirect calorimetry, and participant consent for downstream recall-by-genotype. Our results show that this combination is valuable in two ways. First, Fenland is sufficiently well powered and technically robust to reproduce expected genetic architecture. Second, and more importantly, it enables rare variant discovery in phenotypes that are absent or crudely approximated in larger sequencing datasets, illustrated by the identification of a sex-specific *CASQ1* effect on REE and multiple novel gene associations for glycaemic traits.

Methods

Study description

The Fenland Study is a population-based cohort, recruited between 2005 and 2015 from general practice surgeries around Cambridgeshire, UK. Participants were aged between 29 and 64 (mean 48.6). Exclusion criteria included: terminal illness, clinically diagnosed diabetes mellitus, clinically diagnosed psychotic disorder, inability to walk unaided, pregnancy or lactation. The study was designed as a longitudinal cohort, with participants undergoing follow-up assessments in subsequent study phases. Written informed consent was attained from all participants, including that for future recall-by-genotype studies. The study was approved by the Cambridge Local Research Ethics Committee. Full study details have been described previously (11).

Whole exome sequencing and quality control

Quality control

Whole-exome sequencing (WES) data for the Fenland cohort were generated at the Wellcome Sanger Institute and underwent quality control (QC) using a standardised pipeline (52).

Sample contamination was assessed using VerifyBamID (v2.0), requiring FREEMIX ≤0.05. Concordance between Fenland WES and array genotypes was evaluated using *bcftools gtcheck* following liftover of array data to GRCh38, to identify potential sample mismatches or mapping errors (53). Sex concordance was assessed by comparing imputed genetic sex with reported sex. Genetic ancestry was inferred by projection onto principal components derived from the 1000 Genomes reference panel, with superpopulation labels assigned by proximity in principal component space (54). Within ancestry groups, samples were excluded if they were outliers (>±6 SD) for sequencing metrics including variant count, transition/transversion ratio, insertion/deletion ratio, heterozygosity rate, and heterozygote/homozygote ratio (Supplementary Figure 9). Relatedness was estimated using the KING algorithm implemented in PLINK (55).

Variant quality control used a random forest (RF) classifier to distinguish true variants from sequencing artefacts. Training data comprised high-confidence variants from the 1000 Genomes Project (54), HapMap3 (56), and Mills et.al. indels (57), alongside putative false positives defined using lenient hard filters (Quality by Depth <2, Fisher Strand >60, Mapping Quality <30). The classifier achieved 99.0% test accuracy, comparable to similar sequencing studies (e.g. BiB, ALSPAC), and showed consistent feature importance profiles across cohorts (Supplementary Table 1). Final filtering thresholds were defined using RF scores and genotype-level metrics, including depth (DP), genotype quality (GQ), allele balance, and missingness. Thresholds were calibrated separately for SNVs and indels against Genome in

a Bottle reference samples (58). Three QC levels (relaxed, medium, and stringent) were defined.

Variant annotation and masking

Variants were annotated using the Ensembl Variant Effect Predictor v114 (59) with additional plugins for REVEL (60), CADD (61), AlphaMissense (62) and LOFTEE (63) annotations. Variants were allowed to map to multiple genes where applicable. For each variant-gene pair, a single consequence was retained using a transcript prioritisation scheme favouring protein-coding transcripts, with preference given to MANE Select and canonical transcripts. Variants were grouped into gene-based masks based on predicted consequence and functional impact scores (Supplementary Table 3).

Validation against UKB

Minor allele counts for common variants with MAF >1% were compared between Fenland and UKB across random forest QC classes (relaxed, medium, stringent) using linear models to inform selection of the QC threshold. Gene-level carrier counts for rare variants (MAF ≤1%) across a random subset of 200 genes were then compared between datasets across four burden masks: high-confidence protein-truncating variants (HC_PTV), missense variants with REVEL >0.5, missense variants with CADD >25, and missense variants with AlphaMissense >0.564.

ExWAS and gene-based burden analyses for height and BMI were performed independently in the Fenland cohort and UKB using WES data. Analyses were restricted to variants with MAF ≤1% for burden tests and >1% for single-variant analyses. Gene-based burden tests were conducted using the same masks as listed, and analyses were restricted to genes with a minimum AC ≥15 within each cohort. For cross-cohort comparisons, effect estimates were aligned by allele, and directional concordance was assessed across UKB significance thresholds using binomial tests.

Phenotype Derivation

Resting energy expenditure in Fenland

Resting energy expenditure (REE) was derived at baseline assessment from indirect calorimetry using breath-by-breath measurements (Jaeger Oxycon Pro, CareFusion, CA) collected under standardised resting conditions in the Fenland cohort, as described previously (64). Participants were measured in a supine, awake state following a period of rest, with respiratory exchange recorded over a six-minute protocol. Breath-by-breath oxygen consumption (VO_2) and carbon dioxide production (VCO_2) were processed and extensively quality controlled. Briefly, physiologically implausible and unstable breaths were excluded, and analyses were restricted to the final 120 seconds of measurement to ensure steady-state respiratory exchange. Participants with evidence of non-steady-state

respiration or insufficient valid data were excluded. REE (kJ/day) was calculated from mean VO_2 and VCO_2 using the abbreviated Weir equation:

$$\mathrm{REE} = 6.025 \times (3.9 \times \mathrm{VO_2} + 1.1 \times \mathrm{VCO_2})$$

Values were regressed against REE device and room temperature to remove technical effects for subsequent analyses.

Anthropometric and physiological measures

In Fenland, anthropometric and physiological traits were derived from baseline measurements. Weight was measured to the nearest 100g and height was measured to the nearest 0.1cm (12). Body composition was assessed using whole-body dual-energy X-ray absorptiometry (DXA) in individuals weighing ≤140kg. Scans were performed using GE Healthcare Lunar Prodigy Advanced systems, with a minority acquired on iDEXA platforms. From these, total fat mass (TFM) and total lean mass (TLM) were obtained, and used to derive body fat percentage (BF%), fat mass index (FMI), and lean mass index (LMI).

In UKB, weight (field 23098), BMI (23104), BF% (field 23099) and predicted REE (field 23105) were obtained from baseline measurements using the Tanita BC-418MA device. Predicted total lean mass (TLM) and total fat mass (TFM) were derived for UK Biobank participants of European ancestry, adapting previously described DXA-calibrated bioimpedance prediction methods (65). Models were developed separately for men and women using DXA-measured lean and fat mass (fields 23280 and 23278) from the subset of 63,129 participants with available imaging. Candidate predictors included bioimpedance, anthropometric, and lifestyle measures collected at the same imaging visit, including height, weight, waist and hip circumference, whole-body and segmental (arm and leg) impedance, handgrip strength, physical activity, sleep duration, and smoking and alcohol status. Impedance measures from paired limbs were summed, and height-indexed impedance indices were derived. Handgrip strength was measured using a hydraulic hand dynamometer, with the maximum value across left and right hands (fields 46/47) retained. To avoid circularity, bioimpedance-derived body composition fields (fat-free mass, fat mass, water mass, BMR, BMI) were excluded from model development. All traits were winsorised at ±10 SD within each sex, and values below zero were set to missing. Within each sex, individuals with DXA measurements were split into training (95%) and validation (5%) sets. Predictor selection was performed using forward-backward stepwise regression with the Bayesian Information Criterion, yielding the final sex-specific models (Supplementary Table 6). Model performance was assessed in the held-out validation set by R^2 and RMSE, and the resulting models were subsequently applied to the full cohort to generate TLM and TFM estimates.

Hormonal and reproductive measures

Circulating sex hormones in UK Biobank were measured from baseline serum biochemistry samples. Total testosterone (field 30850), estradiol (field 30800) and sex hormone-binding globulin, SHBG (field 30830), were assayed centrally using Beckman Coulter DXI 800 chemiluminescent immunoassays. Testosterone and estradiol were measured using

competitive-binding immunoassays, while SHBG was measured using a two-step sandwich immunoassay. The manufacturer analytical ranges were 0.35–55.5 nmol/L for testosterone, 73–17,621 pmol/L for estradiol, and 0.33–226/242 nmol/L for SHBG. The UK Biobank observed lower reportable limits were 0.35 nmol/L for testosterone, 175 pmol/L for estradiol, and 0.33 nmol/L for SHBG. For full reference data see https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/serum_biochemistry.pdf. As FSH was not directly measured in UKB, we used the promoter variant rs10835638 at FSHB as a genetic proxy for circulating FSH levels (66). Hormonal contraceptive use (field 2784), hormone replacement therapy (HRT; field 2814), and menopause status (field 2724) were obtained from self-reported questionnaire data.

Glycaemic traits

In Fenland, participants completed a 2-hour 75 g oral glucose tolerance test (OGTT) following an overnight fast. Blood samples were collected at fasting (T_0) and 2 hours after glucose load (T_{120}). Plasma glucose (mmol/L) was assayed from fluoride oxalate tubes and measured using a Dade Behring Dimension RxL analyser. Insulin (pmol/L) was measured in plasma from EDTA tubes using a DELFIA immunoassay (PerkinElmer). Change measures were defined as $T_\Delta = T_{120} - T_0$ for both traits. All insulin traits were log-transformed.

In UKB, random glucose (field 30740) was derived from blood biochemistry measures taken at baseline visit. Fasting glucose was defined in ~15,000 individuals reporting 8-24 hours of fasting (field 74), which is not restrictive to an overnight fast.

Across studies/traits, outliers beyond ±10 standard deviations from the mean were removed. Diabetic individuals, defined as usage of diabetes medication or having fasting glucose ≥7 mmol/L, 2-hour glucose ≥11.1 mmol/L, or HbA1c ≥6.5%, or were excluded.

Statistical analysis framework

Association tests

Exome-wide association analyses were performed in Europeans ancestry individuals in both Fenland and the UKB using BOLT-LMM v2.4.1 (67) under the infinitesimal mixed model. The mixed model was fitted using an LD-pruned set of high-quality genotyped variants to account for population structure and relatedness. Association testing was conducted for both single variants and gene-based burden masks. For burden tests, variants were aggregated within each gene according to predefined masks, and tests were restricted to gene-mask combinations with a combined AC >15. Synonymous variants were analysed in parallel as a negative control, providing an internal null model for gene-based aggregation.

Unless otherwise stated, all analyses included standard covariates: age, age^2, sex (where applicable) and the first ten genetic principal components. Study-specific covariates (test site in Fenland and WES batch in the UKB) were also included.

Significance thresholding

Significance thresholds for gene-based burden tests were defined using a Bonferroni framework adjusted for correlation between burden masks and phenotypes (Supplementary Table 4). Burden test statistics were converted to signed Z-scores using the direction of effect and corresponding two-sided P-values. Correlation matrices were constructed between masks and between phenotypes using gene-level Z-scores, and used to estimate the number of independent tests using an eigenvalue-based approach, where

$$M_{eff} = \sum \min(\lambda_i, 1)$$

and λ_i are the eigenvalues of the correlation matrix (68).

For analyses of resting energy expenditure (REE), significance thresholds were defined as:

$$\frac{0.05}{N_{genes} \times M_{eff,mask}}$$

For analyses of related glycaemic traits (e.g. glucose at T_0, T_{120}, and ΔT), thresholds were defined as:

$$\frac{0.05}{N_{genes} \times M_{eff,mask} \times M_{eff,pheno}}$$

Sex-specificity

Where associations exceeded the significance threshold and passed DFBETA filtering, sex-stratified burden tests were performed separately in men and women, and sex-specificity was assessed by comparing sex-stratified effect sizes using the formula below (where f denotes female-, and m denotes male-estimates)(8):

$$z = \frac{\beta_f - \beta_m}{\sqrt{\mathrm{se}_f^2 + \mathrm{se}_m^2}}$$

Results were classed as sex-specific if P(z) < 0.05 and the association was not nominally significant in the opposite sex (P > 0.05).

Assessment of resting energy expenditure

Phenotype correlations

Phenotypic correlations between REE and anthropometric and body composition traits in Fenland were assessed using multivariable linear regression, adjusting for REE measurement device, room temperature, age and sex. Variance explained was summarised using R^2, and significance was derived from the corresponding model term.

Primary analysis

Gene-burden tests for REE in Fenland were performed in Europeans using BOLT-LMM. BMI, room temperature at REE measurement, and REE measurement device were included as additional covariates. To assess whether significant associations were confounded by body composition beyond BMI, primary models were re-run with additional adjustment for DEXA-derived measures. To evaluate potential collider bias, significant gene-mask combinations were tested for association with BMI and other body composition traits in both Fenland and UKB.

Menopause, age and hormone modification

In UKB, the effect of rs145486953 on BF% was assessed across baseline menopausal strata (pre-, post-, and post-menopausal on HRT) using BOLT-LMM. Longitudinal effects were evaluated in women with repeated BF% and menopause measures across instances, comparing differences between carriers and non-carriers using a two-sample t-test. In Fenland, women were stratified by age (<50 vs ≥50 years) as a proxy for menopausal transition, and associations with REEadjBMI were tested within strata using linear regression.

To assess modification by circulating sex hormones in UKB, interaction terms between rs145486953 carrier status and baseline levels of total testosterone, estradiol, and SHBG were tested in sex-stratified linear regression models of BF%. As FSH was not measured in UKB, genotype at the promoter variant rs10835638 in *FSHB* was used as a proxy for circulating FSH levels.

Proteomic associations

To explore downstream molecular effects of rs145486953, proteomic association analyses were performed in UKB individuals with available Olink plasma measurements (69). Olink protein measurements were processed by panel, and proteins were excluded if more than 25% of values were below the limit of detection. Individual measurements below the limit of detection were set to missing. Protein levels were tested for association with rs145486953 carrier status using linear regression, including sex-stratified analyses. Models included standard covariates with additional adjustment for Olink batch, processing start date and the number of proteins measured. Statistical significance was defined using Bonferroni correction for the number of proteins and sexes tested.

Assessment of glycaemic traits

Primary analysis

Gene-burden tests for glycaemic traits were performed in Fenland in non-diabetic Europeans using BOLT-LMM with BMI as a covariate. Analyses were conducted in both sex-combined and sex-stratified models.

Polygenic risk score adjustment

Polygenic risk scores (PRS) for glycaemic traits were constructed using pre-computed MAGIC weights for fasting glucose, fasting insulin, and 2-hour glucose (21). Variants were harmonised to hg38 and aligned by allele prior to score calculation using *plink2* (70). Scores were computed per chromosome on imputed genotype data and summed genome-wide for each individual. PRS were included as additional covariates in burden analyses of glucose traits and fasting insulin to account for common variant effects. No equivalent summary statistics were available for 2-hour insulin, so this trait was excluded from PRS adjustment.

Independence from common variant signals

To assess independence from known common-variant glycaemic loci, significant gene-mask associations were cross-referenced against MAGIC consortium index variants. As no GWAS data are available for 2-hour insulin, loci reported for insulin fold change were used as a proxy. For fasting glucose, fasting insulin, 2-hour glucose and insulin fold change, genome-wide significant variants ($P < 5\times10^{-8}$) were obtained from BMI-adjusted MAGIC summary statistics (including exome-wide analyses where available) and published European-ancestry GWAS lead-variant lists (14, 21, 71). Established type 2 diabetes loci were additionally obtained from large European-ancestry meta-analyses (22). Variants from summary statistics were pruned to approximately independent signals using distance-based clumping (±250 kb) and mapped to GRCh38 using Liftover (72). Gene-mask associations were considered to overlap known loci if an index variant lay within ±1 Mb.

Data availability

Permission to access the Fenland Study data for analyses through the IMS Epidemiology Trusted Research Environment can be requested by bona fide researchers for specified scientific purposes through an application process via the study website (https://studies.epi.ims.cam.ac.uk/fenland/data-sharing).

Acknowledgements

We are grateful to all Fenland volunteers and to the General Practitioners and practice staff for assistance with recruitment. We thank the Fenland Study Investigators, Fenland Study Co-ordination team and the Epidemiology Field, Data and Laboratory teams. We also acknowledge the participants and investigators of the UK Biobank Study.

Author Contributions Statement

J.R.A.M. performed all analyses and wrote the first draft of manuscript. N.D.K. and J.L. managed Fenland data and assisted with the initial cleaning of exome sequence data. F.R.D. and E.J.M. assisted with genetic analysis strategy. S.L. and S.O. assisted with interpretation of genes and mechanistic pathways. S.B. led the derivation and quality control of resting energy expenditure in Fenland. J.R.B.P. and K.O. jointly led and initially conceived this project. N.J.W. led the Fenland exome sequencing project and assisted with comparisons with other biobanks. All authors reviewed and edited the manuscript.

Funding

This work was conducted using the UK Biobank resource (application 9905). This work was funded by the Medical Research Council (unit programs: MC_UU_12015/2, MC_UU_00006/2). The Fenland Study (10.22025/2017.10.101.00001) is funded by the Medical Research Council (MRC) (MC_UU_12015/1). Exome Sequencing of the Fenland study was supported by the MRC (MC_PC_23015). S.L. is supported by the Academic Clinical Lectureship from Queen's University Belfast and the Northern Ireland Department of Health and a starter grant for clinical lecturers from the academy of medical sciences (SGL0331025).

Competing Interests Statement

J.R.B.P. is an employee/shareholder of Insmed Innovation UK / Insmed Inc., receives research funding from GSK and consultancy fees from Menwell Ltd and Ovartix Ltd. S.L. is currently a full-time employee of GSK outside of this research. The remaining authors declare no competing interests.

References

1. C. Sudlow *et al.*, UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
2. A. G. Bick *et al.*, Genomic data in the All of Us Research Program. *Nature* **627**, 340–346 (2024).
3. P. P. Fiziev *et al.*, Rare penetrant mutations confer severe risk of common diseases. *Science* **380**, eabo1131 (2023).
4. S. M. Gaynor *et al.*, Yield of genetic association signals from genomes, exomes and imputation in the UK Biobank. *Nature Genetics* **56**, 2345–2351 (2024).
5. Y. Zhao *et al.*, Protein-truncating variants in BSN are associated with severe adult-onset obesity, type 2 diabetes and fatty liver disease. *Nature Genetics* **56**, 579–584 (2024).
6. J. Flannick *et al.*, Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat Genet* **46**, 357–363 (2014).
7. E. J. Gardner *et al.*, Damaging missense variants in IGF1R implicate a role for IGF-1 resistance in the etiology of type 2 diabetes. *Cell Genom* **2**, None (2022).
8. L. R. Kaisinger *et al.*, Large-scale exome sequence analysis identifies sex- and age-specific determinants of obesity. *Cell Genom* **3**, 100362 (2023).
9. D. Curtis, Analysis of rare coding variants in 200,000 exome-sequenced subjects reveals novel genetic risk factors for type 2 diabetes. *Diabetes/Metabolism Research and Reviews* **38**, e3482 (2022).
10. D. Curtis, Analysis of 470,000 exome-sequenced cases and controls fails to identify any genes impacting risk of developing affective disorder. *Acta Neuropsychiatr* **37**, e69 (2025).
11. T. Lindsay *et al.*, Descriptive epidemiology of physical activity energy expenditure in UK adults (The Fenland study). *Int J Behav Nutr Phys Act* **16**, 126 (2019).
12. E. A. Clifton *et al.*, Associations between body mass index-related genetic variants and adult body composition: The Fenland cohort study. *Int J Obes (Lond)* **41**, 613–619 (2017).
13. J. S. Zheng *et al.*, Plasma Vitamin C and Type 2 Diabetes: Genome-Wide Association Study and Mendelian Randomization Analysis in European Populations. *Diabetes Care* **44**, 98–106 (2021).
14. A. Williamson *et al.*, Genome-wide association study and functional characterization identifies candidate genes for insulin-stimulated glucose uptake. *Nature Genetics* **55**, 973–983 (2023).
15. M. Koko *et al.*, Exome sequencing of UK birth cohorts [version 2; peer review: 3 approved, 1 approved with reservations]. *Wellcome Open Research* **9** (2024).
16. Anonymous, Genome in a bottle—a human DNA standard. *Nature Biotechnology* **33**, 675–675 (2015).
17. N. Bouatia-Naji *et al.*, A polymorphism within the G6PC2 gene is associated with fasting plasma glucose levels. *Science* **320**, 1085–1088 (2008).

18. D. A. Baerenwald *et al.*, Multiple functional polymorphisms in the G6PC2 gene contribute to the association with higher fasting plasma glucose levels. *Diabetologia* **56**, 1306–1316 (2013).
19. K. J. Karczewski *et al.*, Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK Biobank exomes. *Cell Genomics* **2**, 100168 (2022).
20. J. Wessel *et al.*, Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. *Nature Communications* **6**, 5897 (2015).
21. J. Chen *et al.*, The trans-ancestral genomic architecture of glycemic traits. *Nat Genet* **53**, 840–860 (2021).
22. A. Verma *et al.*, Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. *Science (New York, N.Y.)* **385**, eadj1182 (2024).
23. Y. Y. Lam, E. Ravussin, Analysis of energy metabolism in humans: A review of methodologies. *Mol Metab* **5**, 1057–1071 (2016).
24. R. K. Johnson, D. Coward-McKenzie, "CHAPTER 2 - Energy Requirement Methodology" in Nutrition in the Prevention and Treatment of Disease*,* A. M. Coulston, C. L. Rock, E. R. Monsen, Eds. (Academic Press, San Diego, 2001), https://doi.org/10.1016/B978-012193155-1/50004-0, pp. 31–42.
25. D. C. Fonseca *et al.*, Body weight control and energy expenditure. *Clinical Nutrition Experimental* **20**, 55–59 (2018).
26. K. D. Hall *et al.*, Quantification of the effect of energy imbalance on bodyweight. *Lancet* **378**, 826–837 (2011).
27. J. R. Speakman *et al.*, Total daily energy expenditure has declined over the past three decades due to declining basal expenditure, not reduced activity expenditure. *Nat Metab* **5**, 579–588 (2023).
28. L. Jiang, K. L. Penney, E. Giovannucci, P. Kraft, K. M. Wilson, A genome-wide association study of energy intake and expenditure. *PLoS One* **13**, e0201555 (2018).
29. P. R. Loh, G. Kichaev, S. Gazal, A. P. Schoech, A. L. Price, Mixed-model association for biobank-scale datasets. *Nat Genet* **50**, 906–908 (2018).
30. J. N. Hellwege *et al.*, Association of gene coding variation and resting metabolic rate in a multi-ethnic sample of children and adults. *BMC Obes* **4**, 12 (2017).
31. P. Piaggi *et al.*, Exome Sequencing Identifies A Nonsense Variant in DAO Associated With Reduced Energy Expenditure in American Indians. *J Clin Endocrinol Metab* **105**, e3989–4000 (2020).
32. J. S. Woo, S. Y. Jeong, J. H. Park, J. H. Choi, E. H. Lee, Calsequestrin: a well-known but curious protein in skeletal muscle. *Experimental & Molecular Medicine* **52**, 1908–1925 (2020).
33. I. C. Smith, E. Bombardier, C. Vigna, A. R. Tupling, ATP consumption by sarcoplasmic reticulum Ca^{2+} pumps accounts for 40-50% of resting metabolic rate in mouse fast and slow twitch skeletal muscle. *PLoS One* **8**, e68924 (2013).
34. C. J. Brandl, N. M. Green, B. Korczak, D. H. MacLennan, Two Ca2+ ATPase genes: homologies and mechanistic implications of deduced amino acid sequences. *Cell* **44**, 597–607 (1986).

35. M. Dainese *et al.*, Anesthetic- and heat-induced sudden death in calsequestrin-1-knockout mice. *Faseb j* **23**, 1710–1720 (2009).
36. A. Michelucci *et al.*, Antioxidants protect calsequestrin-1 knockout mice from halothane- and heat-induced sudden death. *Anesthesiology* **123**, 603–617 (2015).
37. A. Michelucci, S. Boncompagni, M. Canato, C. Reggiani, F. Protasi, Estrogens Protect Calsequestrin-1 Knockout Mice from Lethal Hyperthermic Episodes by Reducing Oxidative Stress in Muscle. *Oxid Med Cell Longev* **2017**, 6936897 (2017).
38. B. J. Hill, E. Muldrew, Oestrogen upregulates the sarcoplasmic reticulum Ca(2+) ATPase pump in coronary arteries. *Clin Exp Pharmacol Physiol* **41**, 430–436 (2014).
39. X. Wang *et al.*, The estrogen receptor-interacting protein HPIP increases estrogen-responsive gene expression through activation of MAPK and AKT. *Biochim Biophys Acta* **1783**, 1220–1228 (2008).
40. H. Shi, Y. Yu, K. Guo, R. He, Type VIII collagen: advances in matrix biology and translational promise. *Front Bioeng Biotechnol* **13**, 1732988 (2025).
41. G. Hou, D. Mulholland, M. A. Gronska, M. P. Bendeck, Type VIII collagen stimulates smooth muscle cell migration and matrix metalloproteinase synthesis after arterial injury. *Am J Pathol* **156**, 467–476 (2000).
42. A. Eser *et al.*, Migration and Growth Are Attenuated in Vascular Smooth Muscle Cells With Type VIII Collagen-Null Alleles. *Arteriosclerosis, Thrombosis, and Vascular Biology* **26**, 56–61 (2006).
43. U. Hopfer *et al.*, Lack of Type VIII Collagen in Mice Ameliorates Diabetic Nephropathy. *Diabetes* **58**, 1672–1681 (2009).
44. X. Y. Chen *et al.*, Brain-selective kinase 2 (BRSK2) phosphorylation on PCTAIRE1 negatively regulates glucose-stimulated insulin secretion in pancreatic β-cells. *J Biol Chem* **287**, 30368–30375 (2012).
45. J. Nie *et al.*, Synapses of amphids defective (SAD-A) kinase promotes glucose-stimulated insulin secretion through activation of p21-activated kinase (PAK1) in pancreatic β-Cells. *J Biol Chem* **287**, 26435–26444 (2012).
46. J. Lee *et al.*, Regulation of Type 1 Diabetes via Brown Adipocyte–Secreted Proteins and the Novel Glucagon Regulator Nidogen-2. *Diabetes* **74**, 907–920 (2025).
47. I. Kathuria *et al.*, Nidogen 2 Overexpression Promotes Hepatosteatosis and Atherosclerosis. *Int J Mol Sci* **25** (2024).
48. S. Farfari, V. Schulz, B. Corkey, M. Prentki, Glucose-regulated anaplerosis and cataplerosis in pancreatic beta-cells: possible implication of a pyruvate/citrate shuttle in insulin secretion. *Diabetes* **49**, 718–726 (2000).
49. J. Xu, J. Han, Y. S. Long, P. N. Epstein, Y. Q. Liu, The role of pyruvate carboxylase in insulin secretion and proliferation in rat pancreatic beta cells. *Diabetologia* **51**, 2022–2030 (2008).
50. Y. Yang *et al.*, The mitochondrial enzyme pyruvate carboxylase restricts pancreatic β-cell senescence by blocking p53 activation. *Proc Natl Acad Sci U S A* **121**, e2401218121 (2024).
51. K. Harris, S. A. E. Peters, M. Woodward, Sex hormones and the risk of myocardial infarction in women and men: a prospective cohort study in the UK Biobank. *Biol Sex Differ* **14**, 61 (2023).

52. G. Zakharov *et al.*, WxS-QC - a quality control pipeline for human Whole-Genome and Whole Exome sequencing cohorts. *bioRxiv* 10.64898/2025.12.15.694327, 2025.2012.2015.694327 (2025).
53. P. Danecek *et al.*, Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021).
54. A. Auton *et al.*, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
55. A. Manichaikul *et al.*, Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
56. Anonymous, The International HapMap Project. *Nature* **426**, 789–796 (2003).
57. R. E. Mills *et al.*, An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res* **16**, 1182–1190 (2006).
58. J. M. Zook *et al.*, Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Sci Data* **3**, 160025 (2016).
59. W. McLaren *et al.*, The Ensembl Variant Effect Predictor. *Genome Biology* **17**, 122 (2016).
60. N. M. Ioannidis *et al.*, REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* **99**, 877–885 (2016).
61. M. Kircher *et al.*, A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310–315 (2014).
62. J. Cheng *et al.*, Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **381**, eadg7492 (2023).
63. K. J. Karczewski *et al.*, The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
64. D. J. Wright (2018) Investigating the Relationship between Markers of Ageing and Cardiometabolic Disease. (Apollo - University of Cambridge Repository).
65. R. Powell *et al.*, Development and validation of total and regional body composition prediction equations from anthropometry and single frequency segmental bioelectrical impedance with DEXA. *medRxiv* 10.1101/2020.12.16.20248330, 2020.2012.2016.20248330 (2020).
66. K. S. Ruth *et al.*, Genetic evidence that lower circulating FSH levels lengthen menstrual cycle, increase age at menopause and impact female reproductive health. *Hum Reprod* **31**, 473–481 (2016).
67. P. R. Loh *et al.*, Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284–290 (2015).
68. J. Li, L. Ji, Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* **95**, 221–227 (2005).
69. B. B. Sun *et al.*, Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
70. C. C. Chang *et al.*, Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
71. S. M. Willems *et al.*, Large-scale exome array summary statistics resources for glycemic traits to aid effector gene prioritization. *Wellcome Open Res* **8**, 483 (2023).
72. A. S. Hinrichs *et al.*, The UCSC Genome Browser Database: update 2006. *Nucleic Acids Research* **34**, D590–D598 (2006).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FenlandWESsuppmaterials.pdf