View Article Online
View Journal

Food & Function

Linking the chemistry and physics of food with health and nutrition

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: D. Milenkovic, R. Seth, M. G. Sweet, L. Iglesias Carres and A. P. Neilson, *Food Funct.*, 2026, DOI: 10.1039/D6FO00791K.

This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/food-function

View Article Online
DOI: 10.1039/D6FO00791K

Genetic Variants Modulate Effects of EGCG on Adiposity and Insulinemia in Diversity Outbred Mice

Dragan Milenkovic[1,2,*,#], Romit Seth[2,3,#], Michael G. Sweet[2], Lisard Iglesias-Carres[2], Andrew P Neilson[1,2,*]

1 Department of Food, Bioprocessing and Nutrition Sciences, North Carolina State University, Raleigh, North Carolina, United States.

2 Plants for Human Health Institute, North Carolina State University, Kannapolis, North Carolina, United States.

3 Department of Horticultural Science, North Carolina State University, Raleigh, North Carolina, United States.

[#]: equal contribution of the authors

[*]: Corresponding authors:

Dr Andrew P Neilson, email: aneilso@ncsu.edu

and

Dr Dragan Milenkovic, email: dmilenk@ncsu.edu

Plants for Human Health Institute, Department of Food, Bioprocessing and Nutrition Sciences, North Carolina State University, Kannapolis, NC, USA

View Article Online
DOI: 10.1039/D6FO00791K

Abstract

Dietary polyphenols have been studied for their potential to mitigate cardiometabolic disease risk; however, their metabolic effects show significant interindividual variability. While factors such as sex, age, health status or polyphenol metabolism have been proposed to explain this heterogeneity, the studies on host genetic variability are very limited. Moreover, reliance on genetically inbred mouse model limits the translational relevance of preclinical findings to humans. Emerging concepts in precision nutrition emphasize that responses to dietary bioactives are complex traits shaped by gene–environment interactions, suggesting that genetic background may critically influence bioactivity. In this study, we investigated gene-polyphenol interactions for (−)-epigallocatechin gallate (EGCG) using genetically diverse Diversity Outbred (DO) mice, which model human-like genetic heterogeneity. Male DO mice were fed a high-fat diet (HFD) followed by HFD supplemented with EGCG, and longitudinal changes in body weight, adiposity, insulin, and glucose were assessed. Genome-wide genotypes (>143,000 SNPs) were integrated with metabolic phenotypes using mixed-model genome-wide association analyses. EGCG supplementation elicited pronounced interindividual variability, including divergent effects on adiposity and glycemic regulation. Multiple genomic loci were associated with variation in insulin, body fat mass, and glucose responses, revealing both additive and dominant genetic effects, as well as shared loci influencing insulin regulation and adiposity. Candidate genes within associated regions were enriched for pathways related to insulin signaling, lipid metabolism, inflammation, and cytoskeletal regulation. These findings demonstrate that genetic background is a major determinant of EGCG bioactivity and underscores limitations of single inbred models for predicting phytochemical efficacy. Collectively, this study provides proof-of-concept that genetically diverse preclinical models can identify gene–polyphenol interactions

View Article Online
DOI: 10.1039/D6FO00791K

with translational relevance, supporting precision nutrition strategies to optimize dietary bioactive interventions.

Keywords: Epigallocatechin gallate (EGCG), polyphenols, genetic polymorphism, Genome-Wide Association Study (GWAS), gene-bioactive interaction

Introduction

Evidence from cohort studies and randomized controlled trials indicates that plant food bioactives, particularly polyphenols, may help to reduce the risk of chronic diseases, particularly of cardiometabolic diseases. However, an increasing number of studies have pointed out between subject variations in response to dietary interventions, suggesting that the consumption of particular foods or bioactive constituents may benefit some individuals more than others. Therefore, the recommendations for plant foods are promoted at a population level in a "one-size fits-all approach", which does not necessarily ensure that everyone is adequately exposed to and benefit from the protective constituents provided by these foods [1]. It has been suggested that different factors can influence the response to dietary interventions and thereby may influence the benefit/risk associated with consumption of a particular dietary constituent. Among these factors are gut microbiome, but also age, sex, lifestyle (diet, smoking, physical activity), ethnicity, pathophysiological status or medication [1, 2]. Another factor that can impact between-subject variations in response to bioactives is genetic polymorphism. However, only a very few studies have pointed out that genetic polymorphism can affect the biological response to or metabolism of polyphenols and polyphenol-rich foods. One of the first studies revealed that intake of coffee was associated with an increased risk of nonfatal myocardial infarction only

View Article Online
DOI: 10.1039/D6FO00791K

among individual carriers of the variant CYP1A2*1F of cytochrome P450 1A2 (CYP1A2) enzyme, individuals with slow caffeine metabolism [3]. The risks of albuminuria, hyperfiltration, and hypertension increased with heavy coffee intake only among those with the AC and CC genotypes of CYP1A2 at rs762551 associated with slow caffeine metabolism [4]. An association between the SNPs in SULT1A1, SULT1C4 and ABCC2 and the high excretion of phase II flavanone metabolites were observed [5] as well as association between SNPs in APOE, APOA1, and LPL with lipid levels and blood pressure [6]. Together, these preliminary results in the field of personalized and precision nutrition aim for a better prediction of individual responses rather than population-based averages.

Collectively, these observations highlight an urgent need to re-evaluate phytochemical efficacy through the lens of genetic diversity. Integrating genetically diverse preclinical models with genomic analyses offers a powerful approach to identify loci governing phytochemical responses, improve translational relevance, and ultimately enable precision nutrition strategies that move beyond population averages toward individualized dietary recommendations.

In response to these factors, the field is slowly shifting from a “one-size-fits-all” preclinical and clinical approach towards a pharmacogenomics approach [7-9], where the potential for individuals to exhibit hypo- or hyper-responsiveness to a given exposure is viewed in the context of their specific genetic attributes. The concept of significant gene environment (GxE) interactions is established in the nutrition field. For example, the genetic underpinnings of both spontaneous [10] and diet-induced obesity in mouse models was demonstrated [11, 12]. Indeed, researchers select the C57BL/6J mice for diet-induced obesity due to the Nnt mutation which drives obesity [12, 13]. It is inherently understood that genetics determines how mice respond to dietary stimulus, and that the response to the stimulus is a phenotype. However, the concept of gene-phytochemical

View Article Online
DOI: 10.1039/D6FO00791K

interactions is less widely studied. By extending this concept of GxE interaction to phytochemicals, we can hypothesize that the impact of a phytochemicals depends highly on the genetics of the consumer. However, due to lack of studies in this area, it remains largely unknown 1) how extensive the impact of genetics is on phytochemical efficacy, 2) whether different phytochemicals exhibit distinct GxE behavior, and 3) whether genetic variability in humans has limited translational success of phytochemicals from preclinical models with little genetic variability. In other words, we do not know whether preclinical data on phytochemical efficacy are experimental artifacts from a few inbred strains, or whether we should expect the activities to be widely observed across strains (and thus, in humans). These are key unknowns that must be addressed moving forward.

A key tool for precision nutrition is the availability of preclinical rodent models that possess genetic diversity mimicking humans. The impact of genetic variability on the efficacy of phytochemicals cannot be studied in single inbred rodent models with minimal genetic variation. While traditional outbred rodent populations [14] can overcome some of these limitations (and are often used for phytochemical bioavailability studies), the emergence of systematically designed diversity panels and populations such as the Collaborative Cross (CC) and Diversity Outbred (DO) [15, 16] mouse populations have revolutionized the study of GxE interactions, such as gene diet. These populations facilitate QTL mapping (DO) and heritability studies (DO and CC), which are essential to characterize the genetic underpinnings of gene x phytochemical interactions. Unlike traditional, genetically homogenous inbred strains (e.g., C57BL/6J), which often yield singular responses that fail to replicate in humans, the DO mouse population serves as a robust model for human genetic heterogeneity. DO mice are derived from eight highly divergent founder strains (including wild-derived strains), capturing over 45 million segregating

View Article Online
DOI: 10.1039/D6FO00791K

single-nucleotide polymorphisms (SNPs). This high level of genetic variation mirrors the complexity and diversity found in human populations far better than any single inbred strain. Because DO mice undergo continuous random outbreeding, their genomes are highly recombinant. This fine-grained genetic mosaic allows for high-resolution mapping of complex traits down to specific causative genes or narrow chromosomal intervals, which frequently have direct human orthologs. The DO mouse population has been used to dissect the genetic architecture of complex traits including physiological, behavioral, and molecular phenotypes [17]. Few studies of phytochemicals have utilized these tools. We previously demonstrated that the metabolic effects of quercetin were highly dependent on mouse strain and sex using CC founder strains [18]. Two studies also demonstrated that the metabolic effects of phytochemical-rich cranberry were highly strain-dependent in CC panel strains [19, 20]. Combined, these 3 studies provide provocative evidence that the bioactivities of phytochemicals are significantly influenced by the genetics of the consuming organisms.

The green tea flavanol (−)-epigallocatechin gallate (EGCG) is a useful case study on how genetics influence both the efficacy and safety of a phytochemical. EGCG is one of the most widely studied dietary phytochemicals. Significant preclinical rodent evidence suggests that EGCG or EGCG-rich green tea or its polyphenol extract can prevent or even reverse obesity by reducing body fat via multiple mechanisms including increased fatty acid oxidation and whole-body energy expenditure [21]. However, human clinical findings of the effects of EGCG or EGCG-rich product supplementation on obesity and lipid status have been variable and the effects appear to be modest [13, 22-26]. In the 2000's, it was discovered that polymorphisms in COMT, the gene which encodes catechol-O-methyl transferase, significantly affected both the bioavailability [27] and bioactivity [28, 29] of EGCG and green tea. In terms of safety, most rodent data suggests that

View Article Online
DOI: 10.1039/D6FO00791K

oral consumption of EGCG and green tea are safe at doses associated with diet and supplement use [30, 31], with the caveat that the dose and its distribution throughout the day can influence potential toxicity [32, 33]. However, rare but significant hepatotoxicity of green tea extract supplements has been reported in humans at doses typically regarded as safe (140-1000 mg/d) [34, 35]. Clinical studies showed no evidence of hepatotoxicity below 800 mg EGCG/day up to 12 months which is outside the range of the mean daily intake of EGCG (90–300 mg/day) resulting from the consumption of green tea infusions, however this level of exposure falls within the upper range (300–866 mg/day) of exposure by high-level consumers of green tea infusions [36]. Same study also reported evidence from interventional clinical trials that intake of 800 mg EGCG/day or above taken as a food supplement can induce an increase of serum transaminases in treated subjects compared to control, suggesting that the kinetics as well as the toxicity of green tea catechins could be modified by the matrix in which they are present.

While dosage, contaminants and interactions with other ingredients were identified as potential causes, genetic variability was also suggested as a potential factor [35]. An elegant and groundbreaking study by Church *et al.* [37] explored the impact of gene-EGCG interactions on hepatotoxicity in a large population of DO mice, which accounts for 95% of research mouse genetics and models human genetic variability. In this mouse cohort, a moderate EGCG dose resulted in hepatotoxicity in 16% of the mice. Genotyping and QTL mapping identified a genetic region with 49 genes with variants arising from a single founder mouse strain that were associated with increased probability of hepatotoxicity. Of those, there were 46 with human orthologues. The authors then genotyped individuals from the Drug-Induced Liver Injury Network (DILIN) and normal control participants. Variation in SNPs from the 46 human orthologues were analyzed for their association with hepatotoxicity, resulting in 3 hits (genes

View Article Online
DOI: 10.1039/D6FO00791K

with genotype variants showing differential probability of hepatotoxicity). This study illustrated two key points. First, the activity (adverse or beneficial) of a phytochemical can depend on genetic polymorphisms of the organism exposed to it. Second, rodent models with genetic diversity (such as DO mice) are useful and necessary for uncovering polyphenol GxE interactions with human clinical relevance, which cannot be fully appreciated using inbred mouse model.

The use of genetic polymorphisms to explain variability in efficacy, and then target interventions based on genetic predisposition, can significantly enhance the translatability of phytochemicals from preclinical evidence to human application. Given the fact that most preclinical data on the health benefits of phytochemicals have historically been generated in inbred rodent models, there is thus a significant and urgent need to revisit these data and revise our understanding of phytochemical efficacy in the context of genetic variation. There is significant potential to explore gene interactions in preclinical genetic diversity models, identify key genetic loci/genes governing phytochemical efficacy, and apply these to clinical translation either by recruiting pre-disposed populations or differentiating responders vs. non-responders. Based on the findings of Church *et al.* [37] that EGCG's bioactivity significantly varied due to genetic differences, we recently published a novel phenotyping study supporting the hypothesis that the impact of dietary EGCG supplementation on metabolic phenotypes was dependent on genetic background [38]. In that study, DO mice were fed a high-fat (HFD) diet (8 wk), followed by a HFD + EGCG (8 wk). As expected, all mice gained weight on the HF diet. All mice also gained weight on the HF+EGCG diet, but the average rate of weight gain slowed when EGCG was added. Furthermore, the individual rates of weight gain in the HF and HF+EGCG phases were not correlated. Provocatively, all mice gained body fat on the high-fat diet, but 22% of the

View Article Online
DOI: 10.1039/D6FO00791K

mice lost body fat when EGCG was introduced. Again, body fat changes between the two diet phases were poorly correlated. These data suggest that the genetic background of the mice significantly impacted both their response to a HF diet and the effect of EGCG in the context of a HF diet. The objective of the present study was to employ genotyping data from the DO mice used in our recent pilot study [38] to develop proof-of-concept and feasibility data for QTL mapping of gene EGCG interactions that influence the effects of EGCG on diet-induced obesity.

2 Materials and methods

2.1 Mice

Approval for the animal study was obtained from the Animal Care and Use Committee at the North Carolina Research Campus (Protocol 20-004). Three-week-old male DO mice (Stock #009376, Spring 2022, Generation G46, Litter 1, Non-sibling, $N = 50$) were obtained from Jackson Laboratory (Bar Harbor, ME, USA). Sample size was determined by the number of mice provided by the sponsor; power analysis was not conducted. Mice were cage randomized upon arrival and acclimatized to the facility for 11 days on a low-fat diet before administration of the high-fat (HF) diet. Mice were housed under standard specific pathogen-free conditions (12 h light/dark cycle, 30-70% relative humidity, 20 to 26°C, 5 mice/cage) on corncob bedding. Mice were separated for aggression/health reasons when necessary. Animals were allowed access to food and water *ad libitum* except during fasting periods (water only).

2.2 Diet and experimental design

Mice were initially maintained on a low-fat diet (10% kcal from fat, D12450J, Research Diets, New Brunswick, NJ, USA) prior to beginning the study. A high fat diet (HFD, 60% kcal from

View Article Online
DOI: 10.1039/D6FO00791K

fat, D12492, Research Diets) was administered to all mice for weeks 1 to 8 (the "−EGCG" period). EGCG (94%, Sunphenon) was obtained from Taiyo International, Inc. (Taiyo Kagaku Corp. Tokyo, Japan). Then for weeks 9-16 a HFD supplemented with 0.3% EGCG (the "+EGCG" period) was administered to all mice (60% kcal from fat, 0.3% w/w EGCG, D22022201, Research Diets). The compositions of the diets are shown in Table 1. These feeding intervals were selected based on our previous longitudinal study in DO founder strains, showing that 8 weeks of HF feeding was sufficient to induce obesity in many of these mice, and that 8 weeks of HF feeding + quercetin (a flavonoid, like EGCG) was sufficient to demonstrate anti-obesity effects in many mice as well. Based on an estimated average food intake of ~10% body weight/day in DO mice, the 0.3% EGCG diet was designed to provide ~300 mg EGCG/kg body weight/day. Based on a common body surface area conversions method, this mouse dose corresponds to ~ 24.3 mg/kg in an adult human (~1700 mg/70 kg). Dosing was selected based on previous studies demonstrating that within the DO population doses up to 50 mg/kg (i.p. injections) of EGCG show minimal hepatotoxicity, decreased bioavailability of orally administered EGCG vs i.p./i.v. (meaning that 300 mg/kg oral would not reach hepatotoxic concentrations), and data indicating that ~0.3% EGCG in the diet inhibits diet-induced obesity in male C57BL/6J mice. This was a comparatively high dose designed to produce a detectable effect in this phenotyping study.

2.3 Measurements

Blood glucose: Mice were fasted for 6h prior to all blood collection and glucose measurements. Blood was collected every 8 weeks (weeks 0, 8, 16) from the lateral saphenous vein using needle pricks and Microvette CB 300 Z capillary tubes (Starstedt AG & Co. KG, Nümbrecht,

View Article Online
DOI: 10.1039/D6FO00791K

Germany). Blood glucose was measured using glucose test strips (ReliOn Premier, Novo Nordisk, Bagsvaerd, Denmark) after blood collection. Blood samples were allowed to clot for 30 minutes then spun at 3500 x g for 10 minutes. The supernatant was collected and stored at −80°C.

Body mass and composition: Mice were weighed weekly. Body composition was measured at 3 times (weeks 0, 8, 16). EchoMRI Body Composition Analyzer (EchoMRI, Houston, TX, USA) was used to measure fat, lean muscle, free water, and total water mass.

Insulin: Insulin was measured at weeks 0, 8, and 16 using Ultrasensitive Insulin ELISA (MyBioSource, San Diego, CA, USA) per manufacturer instructions using fasted blood serum samples. Absorbance was measured using SpectraMax iD3 Microplate Reader (Molecular Devices, San Jose, CA, USA).

2.4 Mouse euthanasia

At the conclusion of week 16 all animals were euthanized by CO_2 per American Veterinary Medical Association guidelines. Cervical dislocation was used as a secondary method. Liver samples were flash frozen in liquid nitrogen and stored at −80°C for genotyping.

2.5 DNA genotyping

Mouse livers were genotyped using the Giga Mouse Universal Genotyping Array (GigaMUGA, Neogen, Lincoln, NE), which provides >143,000 SNPs built on the Illumina Infinium platform (https://pmc.ncbi.nlm.nih.gov/articles/PMC4751547/). These SNPs are distributed throughout the mouse genome, with special emphasis for markers that are informative in CC and DO populations. DNA extraction and genotyping were performed by Neogen.

View Article Online
DOI: 10.1039/D6FO00791K

2.6 Phenotype data analysis

Data analysis for phenotype data was performed as described previously [37]. Traits used for GWAS analysis were defined as % change in response between response to the high-fat diet + EGCG and the high-fat diet alone (the −EGCG period). This response-based phenotype was selected to capture inter-individual variation in responsiveness to EGCG supplementation while minimizing baseline differences among animals. Consequently, the GWAS was designed to identify genetic loci associated with variation in response to EGCG exposure rather than loci controlling absolute metabolic trait values. This approach enables the detection of genetic factors underlying differential responses to dietary intervention and potential gene × diet interactions.

2.7 Genotype data analysis

Genotype data were initially filtered to ensure suitability for genome-wide association studies. SNPs with more than 10% missing data or a minor allele frequency (MAF < 0.05) were removed. Marker-level quality metrics, including heterozygosity, allele frequency distribution, and genotype concordance (R^2), were assessed in GAPIT to confirm overall data quality. The genotype matrix (GD) and corresponding marker map (GM) were synchronized and validated to ensure consistent marker order across files. As the DO mice exhibit complex population structure and cryptic relatedness arising from their eight-founder ancestry, both population structure and kinship correction were incorporated into association models. Kinship matrices were estimated using the Zhang method implemented in GAPIT. Models containing 3, 5, and 10 PCs were evaluated using quantile–quantile plots, and three PCs (explaining 3.05%, 2.90%, and 2.80% of total genetic variation) were selected because they provided adequate control of genomic

View Article Online
DOI: 10.1039/D6FO00791K

inflation while minimizing overcorrection and loss of statistical power. Although R/qtl2 is commonly used for founder haplotype-based mapping in DO populations, GAPIT was selected because the primary objective of this study was SNP-level association mapping using multiple complementary statistical frameworks (GLM, MLM, MLMM, FarmCPU, and BLINK). This approach allowed direct estimation of additive SNP effects and facilitated cross-model validation of association signals.

2.8 Genome-Wide Association Studies (GWAS)

GWAS was performed using GAPIT v3 conducted in R (v4.4.1). Multiple models were employed to improve robustness of detection, including general linear model (GLM), mixed linear model (MLM), multi-locus mixed model (MLMM), Fixed and random model Circulating Probability Unification (FarmCPU), and Bayesian-information and linkage-disequilibrium iteratively nested keyway (BLINK). The analyses were run with PCA.total = 3 and SNP.MAF = 0.05, with results generated per trait. Genome-wide significance was assessed using a Bonferroni-corrected threshold calculated as $\alpha = 0.05$ divided by the number of analyzed SNPs. Because the exploratory nature of the study and limited sample size reduced power to detect modest-effect loci, a secondary suggestive threshold ($-\log10\ p > 4$) was used to identify candidate regions in Manhattan plot for further investigation. Candidate loci identified using the suggestive threshold were only considered biologically relevant when supported by multiple GWAS models and consistent allelic effect patterns. Additionally, Quantile–quantile (QQ) plots were used to assess inflation of test statistics under different model and PCA parameterizations. Manhattan plots were generated using GAPIT. To integrate results across multiple phenotypes, per-trait GWAS results were merged into a unified dataset and visualized using CMplot package

View Article Online
DOI: 10.1039/D6FO00791K

in R (https://github.com/YinLiLin/CMplot), which enabled multi-track and wide Manhattan plots, displaying associations across all traits simultaneously.

2.8 Candidate genes in the SNPs regions and functional analysis

For the SNPs identified as associated with the phenotypes, DNA sequences around the SNPs were blasted on the mice genome using ensembl BLAST. Once the 100% homologous regions were identified, genes +/- 200,000pb of the SNP region were searched. With the aim to identify what type of RNA the identified genes will give, we used ShinyGO tool [39]. To identify cellular pathway gene ontology in which these genes are involved, we used EnrichR online tool (https://maayanlab.cloud/Enrichr/) [40]. The alluvial plot and the circus plot were created using SRPlot (https://www.bioinformatics.com.cn/en) [41].

2.9 Statistical Analysis

All statistical analyses were performed in R version 4.4.1 (https://www.r-project.org). Genome-wide association analyses were conducted using GAPIT v3 with GLM, MLM, MLMM, FarmCPU, and BLINK models [42]. Population structure and relatedness were accounted for using principal component analysis (PCA) and kinship matrices as described above. Genome-wide significance was assessed using a Bonferroni-corrected threshold ($\alpha = 0.05$/number of SNPs), while a suggestive threshold ($-\log10\ P > 4$) was used to identify candidate loci for exploratory analyses. Model performance and control of false-positive associations were evaluated using quantile–quantile (QQ) plots. Manhattan plots were generated in GAPIT, and integrated multi-trait visualizations were produced using the CMplot package in R [43].

View Article Online
DOI: 10.1039/D6FO00791K

3 Results

3.1 Phenotyping

As we previously reported [38], we measured major biomarkers of obesity and glycemic control in a cohort of *N*=50 DO mice at baseline, after 8 weeks of a HFD alone, and after an additional 8 weeks of HFD + 0.3% EGCG (w/w). Body mass, body fat %, fasting blood insulin and fasting blood glucose are shown in Figure 1. Importantly, measures for individual mice were tracked over time to facilitate tracking of changes between diets and associate these changes with genotyping data.

Body mass data are shown in Figure 1A. Each of the 50 mice gained weight on HFD alone, with extreme variation (~16-109% increase from baseline). All mice also gained weight on HFD + EGCG, albeit with blunted weight gain and variability compared to the HFD alone (~4-52% increase from the end of the HFD period). Adiposity (body fat %) data are shown in Figure 1B. Like body mass, each mouse gained body fat on the HFD alone, with high variability (~5-640% increased from baseline, expressed as % change, not absolute % fat). The addition of EGCG to the HFD on average slowed body fat accretion, like body mass. However, unlike body mass, some mice gained body fat and others (11 out of 50) lost body fat when EGCG was added to the diet, with changes relative to the end of the HFD diet period of −52-+386% (again, expressed as % change, not absolute % fat). Fasting blood insulin data are shown in Figure 1C. For most, but not all mice, fasting insulin concentrations increased on the HFD alone (changes ranged from −94 to +1330%). When EGCG was added to the HFD, more extreme variation was observed (changes ranged from −82 to +3514%). Fasting blood glucose data are shown in Figure 1D. Fasting glucose data exhibited the least consistent trends between diet periods among the four

View Article Online
DOI: 10.1039/D6FO00791K

biomarkers. On the HFD alone, changes in fasting glucose were ~48% to +70%. On the HFD + EGCG diet, changes in blood glucose were −53% to +87%.

Unlike inbred mouse strains with little phenotypic variability, DO mice demonstrate wide phenotypic variability and distinct responses to dietary stimuli. These data suggest strong genetic control of obesity and glycemic control in response to HFD and EGCG. As we previously discussed [38], plateauing of weight and adiposity on a high-fat diet over time could explain some of the slowed body mass and fat accretion in the HFD + EGCG phase, but not the loss of body fat. Furthermore, although EGCG could cause loss of adiposity due to hepatotoxicity, we previously verified that hepatotoxicity was not present at the end of the HFD + EGCG phase by measuring serum alanine aminotransferase levels, a marker of liver damage.

3.2 Genotype Data QC, Population Structure and Kinship

From the initial set of 143,260 SNPs, obtained by genotyping DNA from mouse liver samples, a total of 108,885 high-quality markers were retained after filtering for missing data and applying a minor allele frequency (MAF) cutoff of 0.05. Both marker- and sample-level call rates were high, supporting the suitability of the dataset for genome-wide association analysis. Marker-level quality control revealed evenly distributed heterozygosity values without extreme outliers, a broad and balanced spectrum of allele frequencies, and uniformly high R^2 values across chromosomes, confirming reliable genotype calls (Fig. 2A).

Principal component analysis (PCA) indicated moderate population structure within the DO population, with PC1, PC2, and PC3 explaining 3.05%, 2.9%, and 2.8% of the variance, respectively (Fig. 2B). Based on QQ plot evaluation and genomic inflation patterns, the first three principal components were included as covariates in subsequent GWAS analyses to

View Article Online
DOI: 10.1039/D6FO00791K

account for population structure while maintaining statistical power. Kinship matrices were estimated internally in GAPIT using the Zhang method and incorporated into mixed linear models to account for relatedness among individuals.

The Linkage disequilibrium (LD) patterns were characterized across the genome (Fig. 3). Pairwise correlation coefficients (R) between markers (Fig. 3A) showed a wide distribution, with values ranging from –1 to 1 and a clear enrichment of positive correlations (Fig. 3B). The relationship between marker distance and LD indicated that most SNP pairs within short physical distances exhibited strong correlations, with a rapid decline in LD as the distance increased (Fig. 3C). Marker spacing was broadly distributed across the genome, with most inter-marker distances falling below 5 kb (Fig. 3D–E). LD decay analysis based on R^2 values (Fig. 3F) showed that LD decreased sharply within the first few kilobases, stabilizing at lower levels beyond ~5 kb. These results suggest that the filtered SNP dataset provided broad genome-wide coverage suitable for exploratory association mapping in the DO population. Although LD decayed rapidly across short physical distances, the effective mapping resolution is expected to be substantially larger than the observed LD decay because of marker density and limited sample size. Therefore, the identified loci should be considered candidate genomic regions requiring further fine-mapping and validation.

3.3 Genome-Wide Association Analysis (GWAS) for Insulin, Body Fat Mass and Glucose

To identify genomic regions associated with changes in the key metabolic traits following EGCG supplementation, that is from week 8 to week 16, genome-wide association studies (GWAS) were performed for insulin concentration, body fat mass (BFM), and glucose level using five statistical models: GLM, MLM, MLMM, FarmCPU, and BLINK were implemented in the

View Article Online
DOI: 10.1039/D6FO00791K

GAPIT framework. Circular Manhattan and QQ plots indicated adequate control of population structure and relatedness, with no evidence of substantial p-value inflation. The concordance of association signals across multiple GWAS models was used as supporting evidence for candidate locus identification. Loci repeatedly detected by independent analytical methods were considered higher-confidence candidate regions than loci identified by a single model.

3.3.1 Insulin-associated loci

A strong and consistent association was detected on chromosome 14, with additional moderate signals on chromosomes 10, 7, and 3, suggesting the presence of loci loci associated with variation in insulin concentration (Fig. 4A).The model-specific Manhattan plots (Fig. 4B) further supported this finding, with the most significant associations observed across both single-locus (GLM, MLM) and multi-locus (MLMM, FarmCPU, BLINK) models. The quantile–quantile (QQ) plots showed that most observed p-values aligned well with the expected null distribution, demonstrating adequate correction for population structure and kinship. Only the tail of the distribution deviated from expectation, suggesting the presence of true marker–trait associations rather than false positives (Fig. 4C).

The lead SNPs, UNC4953653, UNCHS019512, UNC24087882, UNCHS029548, and JAX00382029, exhibited significant associations ($-\log10\ p > 4$) with insulin concentration across multiple models. The allelic effect plots (Supplemental figure 1) revealed that genotypic classes coded as 0, 1, and 2 correspond to homozygous reference, heterozygous, and homozygous alternate genotypes, respectively. In each case, mice homozygous for the alternate allele (genotype 2) exhibited significantly altered insulin concentrations compared with the reference homozygotes. Most loci, including those on chromosomes 3, 10, and 7, showed strong

View Article Online
DOI: 10.1039/D6FO00791K

additive or dosage-dependent effects, where incremental increases in allele dosage corresponded to higher insulin levels. Among the identified loci, SNP JAX00382029 on chromosome 14 exhibited the largest estimated effect size (effect = 887.5) and was associated with a minor allele frequency of 0.09. This locus could explain a substantial proportion of phenotypic variance in the analyzed dataset; however, because low-frequency alleles can produce inflated effect estimates in small populations, these values should be interpreted cautiously until validated in larger cohorts.

For SNP UNCHS019512 (Figure 4D), statistical analysis of insulin changes stratified by genotype (G/G vs. T/G) revealed a significant genotype-dependent response to EGCG intake, with mice carrying the G/G genotype exhibiting a markedly lower change in insulin levels compared with those carrying the T/G genotype. The chromosome 10 locus (UNCHS029548) explained approximately 4.6% of the observed phenotypic variance and was consistently identified across MLMM, FarmCPU, and BLINK analyses, providing additional support for its biological relevance.

3.3.2 Body fat mass (BFM)–associated loci

The GWAS for BFM identified significant putative association peaks primarily on chromosomes 9 and 10, with SNPs UNCHS026712 and UNCHS029548 consistently surpassing the significance threshold across models (Fig. 5A, B and C). The allele effect plots indicated an allele dosage-dependent increase in fat mass, characteristic of an additive genetic effect, where incremental copies of the alternate allele were associated with higher fat accumulation (Supplemental figure 2). The co-localization of the chromosome 10 locus with the insulin-associated region suggests pleiotropic genetic effects or closely linked loci influencing both

View Article Online
DOI: 10.1039/D6FO00791K

adiposity and insulin regulation, reinforcing the hypothesis that shared genetic architecture governs multiple aspects of metabolic balance. Interestingly, for SNP UNCHS029548 (Figure 5D), stratification of mice by genotype (G/G vs. A/G) revealed a significant genotype-dependent response to EGCG intake, with mice carrying the G/G genotype exhibiting a markedly lower change in body fat mass levels compared with those carrying the A/G genotype.

3.3.3 Glucose-associated locus

In contrast to insulin and BFM, GWAS for glucose levels detected a single candidate locus on chromosome 17 (UNC28326389) (Fig. 6 A, and B). This locus may reflect a variant influencing glucose homeostasis through mechanisms partially independent of insulin-associated loci. The QQ plot indicated no substantial inflation, confirming that the observed signal reflects a true association rather than population artifact (Figure 6C). The effect plot revealed a pronounced allelic difference, where carriers of the heterozygous allele (A/C) exhibited higher glucose concentrations relative to reference homozygotes (C/C) (Figure 6D; Supplemental figure 3). The distinct pattern suggests a dominant or semi-dominant genetic effect, unlike the purely additive effects seen for insulin and fat mass.

3.3.4 Common Locus associated with insulin and BFM

Together, these results are suggestive of a polygenic and interconnected genetic basis of metabolic traits, with loci on chromosome 10 emerging as a shared hotspot influencing both insulin regulation and body fat deposition (Fig. 7). The dosage-dependent allelic effects observed for insulin and BFM point to additive gene action, while the glucose-associated locus suggests a potentially dominant regulatory effect. Such findings underscore how allelic variation at multiple

View Article Online
DOI: 10.1039/D6FO00791K

loci may cumulatively influence metabolic efficiency, energy storage, and glucose–insulin balance. The convergence of signals across traits also indicates pleiotropic genes or linked QTL that coordinately regulate energy metabolism, consistent with the complex genetic architecture of obesity and insulin resistance.

3.4 Identification and functional analyses of genes in SNPs regions

For the 27 SNPs associated with variability in the effect on insulin level, 172 genes were identified. Of these, 8 were identified as long non-coding RNAs, 12 pseudogenes, 41 protein coding genes, 1 gene coding for snoRNA and 110 unknown genes (Figure 8). It has been observed that 4 SNPs are localized within genes; UNCHS002441 is localized in the Kdm5b (lysine demethylase 5B) gene, UNC24184030 within GM31281, UNC24180391 within GM31227 and UNC24093641 within long non-coding GM35419. Functional analysis of the protein coding genes suggests that some of the identified genes are involved in the retinol metabolism, insulin secretion, insulin resistance, pancreas secretion or PI3K-Akt signaling pathway (Figure 9).

For the SNPs identified as associated with BFM, 84 genes were identified in the regions around the SNPs. Of the 84, 45 are protein coding genes and 1 gene coding for miRNA (Figure 8). Pathway analysis of these genes has shown that some of them are involved in the regulation of p53 signaling, IL3 and IL6 signaling, but also MAPK signaling or regulation of actin cytoskeleton. Moreover, comparison of identified genes against the NIH database of Genotypes and Phenotypes (dbGaP), we observed that variants within or near some of the genes were identified as involved in body fat distribution, triglycerides levels, or type 2 diabetes.

View Article Online
DOI: 10.1039/D6FO00791K

Global association analysis only showed association between 2 SNPs and variability in glucose level in mice following EGCG intake. Eleven genes were identified in the chromosomal regions of the 2 SNPs, of which 6 are protein coding genes and 5 unknown genes (Figure 8). Pathway analysis suggested that these protein coding genes can be involved in the regulation of glycerophospholipid metabolism or glycerolipid metabolism.

4 Discussion

The DO mouse model clearly revealed substantial inter-individual variability in metabolic phenotypes in response to diet-induced obesity and also in metabolic responses to EGCG in the context of obesity. This variability, which is not possible using inbred rodents, provides an opportunity to map QTL for response to EGCG supplementation. Precision nutrition is drawing significant interest, but genetic variability is surprisingly typically overlooked (or discussed as being important but ultimately not studied) when studying factors that explain interindividual variation. Given the availability and decreasing cost of genotyping platforms for both preclinical studies and human studies, dietary stimuli warrant considerable study in GxE models. Phytochemicals are but one of many dietary stimuli that may significantly differ in their effects based on genetic variation of the consuming organisms. The present data highlight the applicability of genotyping, phenotyping, and QTL mapping to identify genetic loci associated with the metabolic effects of dietary polyphenols. These data suggest that new breakthroughs may be made in personal nutrition and human translatability of polyphenol research by applying QTL mapping to characterize genetic underpinnings of interindividual response. This has the potential to advance precision nutrition applications of polyphenols more than any other type of study.

View Article Online
DOI: 10.1039/D6FO00791K

In this study, we performed a comprehensive genome-wide association analysis of insulin concentration, body fat mass (BFM), and glucose levels in mice using high-density SNP genotypes and multiple complementary statistical models implemented in GAPIT. Rigorous genotype quality control, explicit correction for population structure using principal components, and incorporation of kinship matrices ensured robust control of confounding effects. The concordance of association signals across single-locus and multi-locus models provides strong support for the biological relevance of the detected loci. GWAS for insulin levels revealed a complex but structured genetic architecture characterized by a major locus on chromosome 14 and additional moderate-effect loci on chromosomes 10, 7, and 3. The consistency of these signals across multiple models reduces the likelihood that these associations are not model-specific artifacts but might reflect true genetic determinants of insulin regulation.

Importantly, the strongest association signal was consistently detected on chromosome 14 across multiple GWAS models. The lead SNP at this locus (JAX00382029; Chr14:64.8 Mb) exhibited a relatively low minor allele frequency (MAF = 0.09) and the largest estimated phenotypic effect among all loci associated with insulin response. Additional loci on chromosomes 3, 7, and 10 showed more moderate effects and generally explained smaller proportions of phenotypic variance. The chromosome 10 locus (UNCHS029548), which was also associated with body fat mass, explained approximately 4.6% of the observed phenotypic variation in insulin response. Allelic effect plots demonstrated predominantly additive or dosage-dependent genetic effects at these loci, where increasing copies of the alternate allele were associated with progressively higher insulin concentrations. Such additive effects indicate a complex metabolic trait and imply that insulin regulation is modulated through cumulative contributions of alleles rather than

View Article Online
DOI: 10.1039/D6FO00791K

through simple dominant–recessive mechanisms. The observation that the largest estimated effect was associated with a relatively low-frequency allele is consistent with the general expectation that rare or less frequent variants may contribute disproportionately to phenotypic variation. However, because the present study was conducted in a relatively small cohort of DO mice (n = 50), effect size estimates should be interpreted cautiously. Limited sample sizes can result in upwardly biased estimates of PVE explained (commonly referred to as the Beavis effect), particularly for loci segregating at low frequencies. Consequently, the chromosome 14 locus should be considered a candidate region requiring validation in larger DO populations or independent cohorts before definitive conclusions regarding its contribution to insulin regulation can be drawn. This result is consistent with previous mouse and human studies showing that insulin resistance and hyperinsulinemia arise from the combined action of multiple loci with modest to moderate individual effects [44].

A key finding of this study is the identification of a shared locus on chromosome 10 that is significantly associated with both insulin level and body fat mass. The repeated detection of this locus across multiple GWAS models increases confidence that this signal reflects a biologically meaningful association rather than a model-specific artifact. The locus explained a moderate proportion of phenotypic variation and exhibited a clear dosage-dependent allelic effect, suggesting additive genetic regulation of multiple metabolic traits. The co-localization of association peaks for these two traits, together with concordant allele dosage dependent effects, strongly supports the presence of either pleiotropic genetic effects or tightly linked loci. Individuals carrying the alternate allele at this locus exhibited both elevated insulin levels and increased adiposity, suggesting that this genomic region influences energy storage and insulin

View Article Online
DOI: 10.1039/D6FO00791K

signaling in a coordinated manner. Pleiotropy between adiposity and insulin regulation is biologically plausible and well documented. Adipose tissue plays an active endocrine role, secreting adipokines that modulate insulin sensitivity, glucose uptake, and systemic inflammation [45, 46]. Genetic variants that alter adipose development, lipid storage, or adipokine signaling can therefore exert parallel effects on fat mass and insulin levels. The additive nature of the allelic effects observed here further suggests that gradual increases in adiposity driven by this locus may secondarily elevate insulin demand, potentially predisposing individuals to insulin resistance.

Unlike founder haplotype-based approaches implemented in R/qtl2, the SNP-based GWAS framework used here does not directly identify the founder strain(s) contributing to a significant association signal. Consequently, it was not possible to determine whether specific founder alleles, such as those originating from CAST/EiJ or PWK/PhJ, disproportionately contributed to the chromosome 14 or chromosome 10 associations. Future analyses incorporating founder haplotype reconstruction will be important for resolving the founder-specific genetic architecture underlying variability in metabolic responses to EGCG supplementation. Notably, while chromosome 9 also harbored BFM-associated loci, these regions did not overlap with insulin-associated signals, indicating that not all adiposity-related loci directly influence insulin regulation. This highlights the heterogeneity of fat mass genetics, where some loci act primarily on adipogenesis or lipid storage, whereas others exert systemic metabolic effects.

In contrast to insulin and BFM, glucose levels were associated with a single significant locus on chromosome 17. The allelic effect pattern at this locus differed markedly from the additive effects observed for insulin and BFM. Instead, carriers of the alternate allele showed a pronounced shift in glucose concentration, consistent with a dominant or semi-dominant mode of

View Article Online
DOI: 10.1039/D6FO00791K

inheritance. This distinct genetic architecture suggests that glucose homeostasis, at least under the conditions examined in this study, may be influenced by fewer loci with larger individual effects. Such loci may act upstream of insulin signaling, for example by affecting hepatic glucose production, glucose transport, or pancreatic β-cell glucose sensing. The lack of overlap between the glucose-associated locus and the insulin/BFM loci further implies partial genetic independence between glucose regulation and adiposity-driven insulin dynamics, a phenomenon previously observed in both mouse models and human populations [47, 48].

These GWAS results underscore the polygenic and interconnected nature of metabolic traits (and their response to EGCG administration). Nevertheless, given the exploratory nature of the study and the limited sample size, the identified loci should be regarded as candidate regions rather than definitive QTL. The observation that several loci were detected consistently across multiple statistical models provides support for their biological relevance, but independent replication and functional validation will be required to confirm causality. Insulin concentration and body fat mass share a partially overlapping genetic basis, with additive loci exerting pleiotropic effects on energy storage and endocrine regulation. In contrast, glucose levels appear to be influenced by more discrete genetic mechanisms, potentially reflecting regulatory checkpoints distinct from adiposity-driven insulin demand. The identification of additive, dosage-dependent loci is particularly relevant for understanding metabolic disease risk, as such loci can produce continuous phenotypic variation across populations. Small increases in allele dosage may cumulatively shift metabolic balance toward obesity and hyperinsulinemia, thereby potentially increasing susceptibility to insulin resistance and type 2 diabetes. The chromosome 10 locus identified here represents a promising candidate region for future fine-mapping, functional validation, and integration with transcriptomic or epigenomic data to identify causal genes and

View Article Online
DOI: 10.1039/D6FO00791K

pathways. Overall, this study highlights how integrative GWAS approaches using multiple statistical models can robustly dissect the shared and trait-specific genetic components of complex metabolic phenotypes, providing insights into the biological mechanisms underlying obesity, insulin regulation, and glucose homeostasis in relation with EGCG intake. These preliminary results demonstrate an interindividual variability in response to EGCG and allowed the identification of the genes and molecular pathways that drive this differential responsiveness. This work paves the way for identifications of the genetic determinants underlying variability in EGCG-mediated effects, enabling the development of personalized nutritional strategies tailored to individual genetic profiles. Ultimately, these insights will support precision nutrition approaches aimed at optimizing dietary interventions for improved vascular and metabolic health.

Among the genes identified in the regions of SNPs associated with the insulin level is TRIM13. It has been described that knock-out in this gene result in increased level of insulin in mice on high fat diet compared to normal mice, resulting in insulin resistance development [49]. Regarding another candidate genes, CNTFR, it has been observed that its plasma level and expression are correlated with insulin resistance [50]. Moreover, several studies have identified association between polymorphisms in TRPC1 (Transient Receptor Potential Canonical 1) gene and T2D [51], which that can inhibit the positive effect of exercise on type II diabetes under a HF diet-induced obesity [52]. Similarly, another region of the SNPs associated with variability in insulin level is Trex1, gene which inactivation has been identified as spontaneously develop of type 1 diabetes [53]. Other genes identified in our study has been described as involved in the development of diabetes, such as CCDC107 [54] or VCP [55].

View Article Online
DOI: 10.1039/D6FO00791K

Several genes identified in the regions of the SNPs associated with variability in body fat mass were described as involved in body adipose tissue development and obesity. For example, vinexin β-knockout knock-out mice were more obese and gained more obvious visceral fat accumulation than the wildtype mice fed with high fat diet [56]. Association between HTR1B (5-hydroxytryptamine receptor 1B) genetic polymorphism and BMI in women with bulimia nervosa was observed providing explanation why, in response to dieting, some women attain lower BMIs, whereas others have a natural limitation to their weight loss [57]. Among the genes identified is miRNA320, microRNA that exert posttranscriptional regulation of expression of genes, which has been identified as statistically associated with overweight and obesity [58]. Moreover, a link between BMP1 and the obesity was observed [59], as for SLC39A14 [60], or LRP1 [61].

This study presents few limitations. First, financial constraints limited us to N=50 DO mice. Future studies will need larger N (~250) to improve statistical power to obtain more accurate estimates of locus effect sizes and enable higher-confidence fine-mapping of associated regions. Previous studies in DO mice have demonstrated that several hundred animals may be required to reliably detect loci explaining modest proportions of phenotypic variance, whereas loci with smaller effects may require substantially larger sample sizes [62]. Second, due to the pilot nature of this study, we employed a single group longitudinal design to maximize the number of animals receiving EGCG. However, this design is confounded by the effect of time. To control for confounding effects of aging and time, a 2-arm design is needed in which only one arm receives EGCG. Combined with the need for a larger sample size, and ideal 2-arm design with sufficient animals per group could have up to 500 mice. Third, again due to the pilot nature of this study, we excluded sex as a variable. Future studies must include both sexes to improve generalizability

View Article Online
DOI: 10.1039/D6FO00791K

to humans. Lastly, the effects of lower doses of EGCG corresponding to 1-2 cups of green tea per day in humans need to be explored.

Taken together, this study represents one of the first genome-wide association analyses to identify genes and genetic regions involved in the interindividual variability of physiological responses to polyphenol intake. The candidate genes identified in this study represent biologically plausible targets involved in insulin signaling, adiposity, glucose homeostasis, and related metabolic pathways. However, the present findings should be considered preliminary and hypothesis-generating. Additional studies incorporating larger cohorts, founder haplotype reconstruction, fine-mapping, transcriptomic integration, and functional validation will be required to establish causal relationships between these loci and responsiveness to EGCG supplementation.

View Article Online
DOI: 10.1039/D6FO00791K

Acknowledgements

The authors wish to thank Drs. Dahlia M. Nielsen, David L. Aylor (Department of Biological Sciences, North Carolina State University, Raleigh, NC USA) and David W. Threadgill (Department of Biochemistry & Biophysics and School of Medicine, Texas A&M University, College Station, TX USA) for their contributions to the original experimental design. The authors also wish to thank Melissa VerHague and Jody Albright, (Nutrition Research Institute, University of North Carolina Chapel Hill, Kannapolis, NC USA) for assistance with body composition analyses. Funding and mice for this work were provided by the Jackson Laboratory (Bar Harbor, ME USA) via the Diversity Outbred Pilot Grant Program. EGCG (Sunphenon) was a generous gift from Taiyo International (Minneapolis, MN USA).

Funding

Support for APN was provided by the North Carolina Agricultural Research Service (NCARS) and the Hatch Program of the National Institute of Food and Agriculture (NIFA), U.S. Department of Agriculture. DM was supported by the United States Department of Agriculture, National Institute of Food and Agriculture (USDA-NIFA), Hatch project 7010153.

View Article Online
DOI: 10.1039/D6FO00791K

Author contributions

Dragan Milenkovic: Conceptualization; Methodology; Formal analysis; Investigation; Resources; Data Curation; Writing - Original Draft; Writing - Review & Editing; Visualization; Supervision; Project administration.

Romit Seth: Methodology; Formal analysis; Investigation; Writing - Original Draft; Writing - Review & Editing; Visualization.

Michael G. Sweet: Investigation; Writing - Review & Editing; Visualization.

Lisard Iglesias-Carres: Investigation; Writing - Review & Editing; Visualization.

Andrew P Neilson:Conceptualization; Methodology; Formal analysis; Investigation; Resources; Data Curation; Writing - Original Draft; Writing - Review & Editing; Visualization; Supervision; Project administration; Funding acquisition.

View Article Online
DOI: 10.1039/D6FO00791K

Table

Table 1. Nutritional and ingredient composition of mouse diets

	Low fat diet		High fat diet		High fat diet + EGCG	
Nutrient	**gram (%)**	**kcal (%)**	**gram (%)**	**kcal (%)**	**gram (%)**	**kcal (%)**
Protein	19	20	26	20	26	20
Carbohydrate	67	70	26	20	26	20
Fat	4	10	35	60	35	60
Total (kcal %)		100		100		100
Energy (kcal/g)	3.8		5.2		5.2	
Ingredient	**gram**	**kcal**	**gram**	**kcal**	**gram**	**kcal**
Casein	200	800	200	800	200	800
L-Cystine	3	12	3	12	3	12
Corn Starch	506.2	2025	0	0	0	0
Maltodextrin 10	125	500	125	500	125	500
Sucrose	68.8	275	68.8	275	68.8	275
Cellulose	50	0	50	0	50	0
Lard	20	180	245	2205	245	2205
Soybean Oil	25	225	25	225	25	225
Mineral Mix S10026	10	0	10	0	10	0
DiCalcium Phosphate	13	0	13	0	13	0
Calcium Carbonate	5.5	0	5.5	0	5.5	0
Potassium Citrate, 1 H_2O	16.5	0	16.5	0	16.5	0
Vitamin Mix V10001	10	40	10	40	10	40
Choline Bitartrate	2	0	2	0	2	0
EGCG[a] (94% pure)	0	0	0	0	2.5	0
FD&C Yellow Dye #5	0.04	0	0	0	0	0
FD&C Red Dye #40	0	0	0	00.05	0	
FD&C Blue Dye #1	0.01	0	0.05	0	0	0
Total	**1055.05**	**4057**	**773.85**	**4057**	**776.35**	**4057**
EGCG[a] (%)	0		0		0.3	

[a](−)-epigallocatechin gallate

View Article Online
DOI: 10.1039/D6FO00791K

References

1. E. R. Gibney, D. Milenkovic, E. Combet, T. Ruskovska, A. Greyling, A. Gonzalez-Sarrias, B. de Roos, F. Tomas-Barberan, C. Morand and A. Rodriguez-Mateos, Eur J Nutr, 2019, 58, 37-47.
2. D. Milenkovic, C. Morand, A. Cassidy, A. Konic-Ristic, F. Tomas-Barberan, J. M. Ordovas, P. Kroon, R. De Caterina and A. Rodriguez-Mateos, Adv Nutr, 2017, 8, 558-570.
3. M. C. Cornelis, A. El-Sohemy, E. K. Kabagambe and H. Campos, JAMA, 2006, 295, 1135-1141.
4. S. Mahdavi, P. Palatini and A. El-Sohemy, JAMA Netw Open, 2023, 6, e2247868.
5. L. N. Fraga, D. Milenkovic, C. P. Coutinho, A. C. Rozenbaum, F. M. Lajolo and N. M. A. Hassimotto, Mol Nutr Food Res, 2023, 67, e2200847.
6. L. N. Fraga, D. Milenkovic, F. M. Lajolo and N. M. A. Hassimotto, Nutrients, 2022, 14.
7. J. T. Oates and D. Lopez, Int J Biomed Investig, 2018, 1.
8. M. Pirmohamed, Nat Rev Genet, 2023, 24, 350-362.
9. D. T. Wake, N. Ilbawi, H. M. Dunnenberger and P. J. Hulick, Med Clin North Am, 2019, 103, 977-990.
10. D. L. Coleman, Diabetologia, 1978, 14, 141-148.
11. H. C. Freeman, A. Hugill, N. T. Dear, F. M. Ashcroft and R. D. Cox, Diabetes, 2006, 55, 2153-2156.
12. A. A. Toye, J. D. Lippiat, P. Proks, K. Shimomura, L. Bentley, A. Hugill, V. Mijat, M. Goldsworthy, L. Moir, A. Haynes, J. Quarterman, H. C. Freeman, F. M. Ashcroft and R. D. Cox, Diabetologia, 2005, 48, 675-686.

View Article Online
DOI: 10.1039/D6FO00791K

13. S. Lin, T. C. Thomas, L. H. Storlien and X. F. Huang, Int J Obes Relat Metab Disord, 2000, 24, 639-646.

14. R. Chia, F. Achilli, M. F. Festing and E. M. Fisher, Nat Genet, 2005, 37, 1181-1186.

15. G. A. Churchill, D. M. Gatti, S. C. Munger and K. L. Svenson, Mamm Genome, 2012, 23, 713-718.

16. D. W. Threadgill, D. R. Miller, G. A. Churchill and F. P. de Villena, ILAR J, 2011, 52, 24-31.

17. Z. Tatom, M. F. Miles and A. A. Palmer, Mamm Genome, 2026, 37, 28.

18. L. E. Griffin, L. Essenmacher, K. C. Racine, L. Iglesias-Carres, J. S. Tessem, S. M. Smith and A. P. Neilson, J Nutr Biochem, 2021, 87, 108521.

19. F. Amer-Sarsour, R. Abu Saleh, I. Ofek and F. A. Iraqi, Food Funct, 2021, 12, 4972-4982.

20. F. Amer-Sarsour, R. Tarabeih, I. Ofek and F. A. Iraqi, Animal Model Exp Med, 2023, 6, 196-210.

21. H. S. Moon, H. G. Lee, Y. J. Choi, T. G. Kim and C. S. Cho, Chem Biol Interact, 2007, 167, 85-98.

22. O. Asbaghi, M. Rezaei Kelishadi, D. A. Larky, R. Bagheri, N. Amirani, K. Goudarzi, F. Kargar, M. Ghanavati and M. Zamani, Br J Nutr, 2024, 131, 1125-1157.

23. R. Hursel, W. Viechtbauer and M. S. Westerterp-Plantenga, Int J Obes (Lond), 2009, 33, 956-961.

24. M. P. Kapoor, M. Sugita, Y. Fukuzawa and T. Okubo, J Nutr Biochem, 2017, 43, 1-10.

25. O. J. Phung, W. L. Baker, L. J. Matthews, M. Lanosa, A. Thorne and C. I. Coleman, Am J Clin Nutr, 2010, 91, 73-81.

View Article Online
DOI: 10.1039/D6FO00791K

26. Y. Wang, H. Xia, J. Yu, J. Sui, D. Pan, S. Wang, W. Liao, L. Yang and G. Sun, Heliyon, 2023, 9, e21228.

27. M. Inoue-Choi, J. M. Yuan, C. S. Yang, D. J. Van Den Berg, M. J. Lee, Y. T. Gao and M. C. Yu, Int J Mol Epidemiol Genet, 2010, 1, 114-123.

28. R. Hursel, P. L. Janssens, F. G. Bouwman, E. C. Mariman and M. S. Westerterp-Plantenga, PLoS One, 2014, 9, e106220.

29. R. J. Miller, K. G. Jackson, T. Dadd, A. E. Mayes, A. L. Brown, J. A. Lovegrove and A. M. Minihane, Mol Nutr Food Res, 2012, 56, 966-975.

30. R. A. Isbrucker, J. Bausch, J. A. Edwards and E. Wolz, Food Chem Toxicol, 2006, 44, 626-635.

31. R. A. Isbrucker, J. A. Edwards, E. Wolz, A. Davidovich and J. Bausch, Food Chem Toxicol, 2006, 44, 636-650.

32. K. D. James, S. C. Forester and J. D. Lambert, Food Chem Toxicol, 2015, 76, 103-108.

33. J. D. Lambert, M. J. Kennett, S. Sang, K. R. Reuhl, J. Ju and C. S. Yang, Food Chem Toxicol, 2010, 48, 409-416.

34. G. Mazzanti, A. Di Sotto and A. Vitalone, Arch Toxicol, 2015, 89, 1175-1191.

35. H. A. Oketch-Rabah, A. L. Roe, C. V. Rider, H. L. Bonkovsky, G. I. Giancaspro, V. Navarro, M. F. Paine, J. M. Betz, R. J. Marles, S. Casper, B. Gurley, S. A. Jordan, K. He, M. P. Kapoor, T. P. Rao, A. H. Sherker, R. J. Fontana, S. Rossi, R. Vuppalanchi, L. B. Seeff, A. Stolz, J. Ahmad, C. Koh, J. Serrano, T. Low Dog and R. Ko, Toxicol Rep, 2020, 7, 386-402.

36. E. P. o. F. Additives, F. Nutrient Sources added to, M. Younes, P. Aggett, F. Aguilar, R. Crebelli, B. Dusemund, M. Filipic, M. J. Frutos, P. Galtier, D. Gott, U. Gundert-Remy,

View Article Online
DOI: 10.1039/D6FO00791K

C. Lambre, J. C. Leblanc, I. T. Lillegaard, P. Moldeus, A. Mortensen, A. Oskarsson, I. Stankovic, I. Waalkens-Berendsen, R. A. Woutersen, R. J. Andrade, C. Fortes, P. Mosesso, P. Restani, D. Arcella, F. Pizzo, C. Smeraldi and M. Wright, EFSA J, 2018, 16, e05239.

37. R. J. Church, D. M. Gatti, T. J. Urban, N. Long, X. Yang, Q. Shi, J. S. Eaddy, M. Mosedale, S. Ballard, G. A. Churchill, V. Navarro, P. B. Watkins, D. W. Threadgill and A. H. Harrill, Food Chem Toxicol, 2015, 76, 19-26.
38. M. G. Sweet, L. Iglesias-Carres, P. N. Ellsworth, J. D. Carter, D. M. Nielsen, D. L. Aylor, J. S. Tessem and A. P. Neilson, Nutr Res, 2025, 133, 78-93.
39. S. X. Ge, D. Jung and R. Yao, Bioinformatics, 2020, 36, 2628-2629.
40. Z. Xie, A. Bailey, M. V. Kuleshov, D. J. B. Clarke, J. E. Evangelista, S. L. Jenkins, A. Lachmann, M. L. Wojciechowicz, E. Kropiwnicki, K. M. Jagodnik, M. Jeon and A. Ma'ayan, Curr Protoc, 2021, 1, e90.
41. D. Tang, M. Chen, X. Huang, G. Zhang, L. Zeng, G. Zhang, S. Wu and Y. Wang, PLoS One, 2023, 18, e0294236.
42. J. Wang and Z. Zhang, Genomics Proteomics Bioinformatics, 2021, 19, 629-640.
43. L. Yin, H. Zhang, Z. Tang, J. Xu, D. Yin, Z. Zhang, X. Yuan, M. Zhu, S. Zhao, X. Li and X. Liu, Genomics Proteomics Bioinformatics, 2021, 19, 619-628.
44. J. Flint and E. Eskin, Nat Rev Genet, 2012, 13, 807-817.
45. A. Guilherme, J. V. Virbasius, V. Puri and M. P. Czech, Nat Rev Mol Cell Biol, 2008, 9, 367-377.
46. S. E. Kahn, R. L. Hull and K. M. Utzschneider, Nature, 2006, 444, 840-846.

View Article Online
DOI: 10.1039/D6FO00791K

47. J. Dupuis, C. Langenberg, I. Prokopenko, R. Saxena, N. Soranzo, A. U. Jackson, E. Wheeler, N. L. Glazer, N. Bouatia-Naji, A. L. Gloyn, C. M. Lindgren, R. Magi, A. P. Morris, J. Randall, T. Johnson, P. Elliott, D. Rybin, G. Thorleifsson, V. Steinthorsdottir, P. Henneman, H. Grallert, A. Dehghan, J. J. Hottenga, C. S. Franklin, P. Navarro, K. Song, A. Goel, J. R. Perry, J. M. Egan, T. Lajunen, N. Grarup, T. Sparso, A. Doney, B. F. Voight, H. M. Stringham, M. Li, S. Kanoni, P. Shrader, C. Cavalcanti-Proenca, M. Kumari, L. Qi, N. J. Timpson, C. Gieger, C. Zabena, G. Rocheleau, E. Ingelsson, P. An, J. O'Connell, J. Luan, A. Elliott, S. A. McCarroll, F. Payne, R. M. Roccasecca, F. Pattou, P. Sethupathy, K. Ardlie, Y. Ariyurek, B. Balkau, P. Barter, J. P. Beilby, Y. Ben-Shlomo, R. Benediktsson, A. J. Bennett, S. Bergmann, M. Bochud, E. Boerwinkle, A. Bonnefond, L. L. Bonnycastle, K. Borch-Johnsen, Y. Bottcher, E. Brunner, S. J. Bumpstead, G. Charpentier, Y. D. Chen, P. Chines, R. Clarke, L. J. Coin, M. N. Cooper, M. Cornelis, G. Crawford, L. Crisponi, I. N. Day, E. J. de Geus, J. Delplanque, C. Dina, M. R. Erdos, A. C. Fedson, A. Fischer-Rosinsky, N. G. Forouhi, C. S. Fox, R. Frants, M. G. Franzosi, P. Galan, M. O. Goodarzi, J. Graessler, C. J. Groves, S. Grundy, R. Gwilliam, U. Gyllensten, S. Hadjadj, G. Hallmans, N. Hammond, X. Han, A. L. Hartikainen, N. Hassanali, C. Hayward, S. C. Heath, S. Hercberg, C. Herder, A. A. Hicks, D. R. Hillman, A. D. Hingorani, A. Hofman, J. Hui, J. Hung, B. Isomaa, P. R. Johnson, T. Jorgensen, A. Jula, M. Kaakinen, J. Kaprio, Y. A. Kesaniemi, M. Kivimaki, B. Knight, S. Koskinen, P. Kovacs, K. O. Kyvik, G. M. Lathrop, D. A. Lawlor, O. Le Bacquer, C. Lecoeur, Y. Li, V. Lyssenko, R. Mahley, M. Mangino, A. K. Manning, M. T. Martinez-Larrad, J. B. McAteer, L. J. McCulloch, R. McPherson, C. Meisinger, D. Melzer, D. Meyre, B. D. Mitchell, M. A. Morken, S. Mukherjee, S. Naitza, N. Narisu, M. J. Neville, B. A. Oostra,

View Article Online
DOI: 10.1039/D6FO00791K

M. Orru, R. Pakyz, C. N. Palmer, G. Paolisso, C. Pattaro, D. Pearson, J. F. Peden, N. L. Pedersen, M. Perola, A. F. Pfeiffer, I. Pichler, O. Polasek, D. Posthuma, S. C. Potter, A. Pouta, M. A. Province, B. M. Psaty, W. Rathmann, N. W. Rayner, K. Rice, S. Ripatti, F. Rivadeneira, M. Roden, O. Rolandsson, A. Sandbaek, M. Sandhu, S. Sanna, A. A. Sayer, P. Scheet, L. J. Scott, U. Seedorf, S. J. Sharp, B. Shields, G. Sigurethsson, E. J. Sijbrands, A. Silveira, L. Simpson, A. Singleton, N. L. Smith, U. Sovio, A. Swift, H. Syddall, A. C. Syvanen, T. Tanaka, B. Thorand, J. Tichet, A. Tonjes, T. Tuomi, A. G. Uitterlinden, K. W. van Dijk, M. van Hoek, D. Varma, S. Visvikis-Siest, V. Vitart, N. Vogelzangs, G. Waeber, P. J. Wagner, A. Walley, G. B. Walters, K. L. Ward, H. Watkins, M. N. Weedon, S. H. Wild, G. Willemsen, J. C. Witteman, J. W. Yarnell, E. Zeggini, D. Zelenika, B. Zethelius, G. Zhai, J. H. Zhao, M. C. Zillikens, D. Consortium, G. Consortium, B. C. Global, I. B. Borecki, R. J. Loos, P. Meneton, P. K. Magnusson, D. M. Nathan, G. H. Williams, A. T. Hattersley, K. Silander, V. Salomaa, G. D. Smith, S. R. Bornstein, P. Schwarz, J. Spranger, F. Karpe, A. R. Shuldiner, C. Cooper, G. V. Dedoussis, M. Serrano-Rios, A. D. Morris, L. Lind, L. J. Palmer, F. B. Hu, P. W. Franks, S. Ebrahim, M. Marmot, W. H. Kao, J. S. Pankow, M. J. Sampson, J. Kuusisto, M. Laakso, T. Hansen, O. Pedersen, P. P. Pramstaller, H. E. Wichmann, T. Illig, I. Rudan, A. F. Wright, M. Stumvoll, H. Campbell, J. F. Wilson, C. Anders Hamsten on behalf of Procardis, M. investigators, R. N. Bergman, T. A. Buchanan, F. S. Collins, K. L. Mohlke, J. Tuomilehto, T. T. Valle, D. Altshuler, J. I. Rotter, D. S. Siscovick, B. W. Penninx, D. I. Boomsma, P. Deloukas, T. D. Spector, T. M. Frayling, L. Ferrucci, A. Kong, U. Thorsteinsdottir, K. Stefansson, C. M. van Duijn, Y. S. Aulchenko, A. Cao, A. Scuteri, D. Schlessinger, M. Uda, A. Ruokonen, M. R. Jarvelin, D. M. Waterworth, P. Vollenweider,

View Article Online
DOI: 10.1039/D6FO00791K

L. Peltonen, V. Mooser, G. R. Abecasis, N. J. Wareham, R. Sladek, P. Froguel, R. M. Watanabe, J. B. Meigs, L. Groop, M. Boehnke, M. I. McCarthy, J. C. Florez and I. Barroso, Nat Genet, 2010, 42, 105-116.

48. R. A. Scott, V. Lagou, R. P. Welch, E. Wheeler, M. E. Montasser, J. Luan, R. Magi, R. J. Strawbridge, E. Rehnberg, S. Gustafsson, S. Kanoni, L. J. Rasmussen-Torvik, L. Yengo, C. Lecoeur, D. Shungin, S. Sanna, C. Sidore, P. C. Johnson, J. W. Jukema, T. Johnson, A. Mahajan, N. Verweij, G. Thorleifsson, J. J. Hottenga, S. Shah, A. V. Smith, B. Sennblad, C. Gieger, P. Salo, M. Perola, N. J. Timpson, D. M. Evans, B. S. Pourcain, Y. Wu, J. S. Andrews, J. Hui, L. F. Bielak, W. Zhao, M. Horikoshi, P. Navarro, A. Isaacs, J. R. O'Connell, K. Stirrups, V. Vitart, C. Hayward, T. Esko, E. Mihailov, R. M. Fraser, T. Fall, B. F. Voight, S. Raychaudhuri, H. Chen, C. M. Lindgren, A. P. Morris, N. W. Rayner, N. Robertson, D. Rybin, C. T. Liu, J. S. Beckmann, S. M. Willems, P. S. Chines, A. U. Jackson, H. M. Kang, H. M. Stringham, K. Song, T. Tanaka, J. F. Peden, A. Goel, A. A. Hicks, P. An, M. Muller-Nurasyid, A. Franco-Cereceda, L. Folkersen, L. Marullo, H. Jansen, A. J. Oldehinkel, M. Bruinenberg, J. S. Pankow, K. E. North, N. G. Forouhi, R. J. Loos, S. Edkins, T. V. Varga, G. Hallmans, H. Oksa, M. Antonella, R. Nagaraja, S. Trompet, I. Ford, S. J. Bakker, A. Kong, M. Kumari, B. Gigante, C. Herder, P. B. Munroe, M. Caulfield, J. Antti, M. Mangino, K. Small, I. Miljkovic, Y. Liu, M. Atalay, W. Kiess, A. L. James, F. Rivadeneira, A. G. Uitterlinden, C. N. Palmer, A. S. Doney, G. Willemsen, J. H. Smit, S. Campbell, O. Polasek, L. L. Bonnycastle, S. Hercberg, M. Dimitriou, J. L. Bolton, G. R. Fowkes, P. Kovacs, J. Lindstrom, T. Zemunik, S. Bandinelli, S. H. Wild, H. V. Basart, W. Rathmann, H. Grallert, D. I. G. Replication, C. Meta-analysis, W. Maerz, M. E. Kleber, B. O. Boehm, A. Peters, P. P. Pramstaller, M. A.

View Article Online
DOI: 10.1039/D6FO00791K

Province, I. B. Borecki, N. D. Hastie, I. Rudan, H. Campbell, H. Watkins, M. Farrall, M. Stumvoll, L. Ferrucci, D. M. Waterworth, R. N. Bergman, F. S. Collins, J. Tuomilehto, R. M. Watanabe, E. J. de Geus, B. W. Penninx, A. Hofman, B. A. Oostra, B. M. Psaty, P. Vollenweider, J. F. Wilson, A. F. Wright, G. K. Hovingh, A. Metspalu, M. Uusitupa, P. K. Magnusson, K. O. Kyvik, J. Kaprio, J. F. Price, G. V. Dedoussis, P. Deloukas, P. Meneton, L. Lind, M. Boehnke, A. R. Shuldiner, C. M. van Duijn, A. D. Morris, A. Toenjes, P. A. Peyser, J. P. Beilby, A. Korner, J. Kuusisto, M. Laakso, S. R. Bornstein, P. E. Schwarz, T. A. Lakka, R. Rauramaa, L. S. Adair, G. D. Smith, T. D. Spector, T. Illig, U. de Faire, A. Hamsten, V. Gudnason, M. Kivimaki, A. Hingorani, S. M. Keinanen-Kiukaanniemi, T. E. Saaristo, D. I. Boomsma, K. Stefansson, P. van der Harst, J. Dupuis, N. L. Pedersen, N. Sattar, T. B. Harris, F. Cucca, S. Ripatti, V. Salomaa, K. L. Mohlke, B. Balkau, P. Froguel, A. Pouta, M. R. Jarvelin, N. J. Wareham, N. Bouatia-Naji, M. I. McCarthy, P. W. Franks, J. B. Meigs, T. M. Teslovich, J. C. Florez, C. Langenberg, E. Ingelsson, I. Prokopenko and I. Barroso, Nat Genet, 2012, 44, 991-1005.

49. Y. Qian, G. Lei and L. Wen, Biochem Biophys Res Commun, 2020, 527, 138-145.

50. J. Perugini, E. Di Mercurio, A. Giuliani, J. Sabbatinelli, A. R. Bonfigli, E. Tortato, I. Severi, S. Cinti, F. Olivieri, C. W. le Roux, R. Gesuita and A. Giordano, Sci Rep, 2022, 12, 8331.

51. K. Chen, X. Jin, Q. Li, W. Wang, Y. Wang and J. Zhang, Endocr Res, 2013, 38, 59-68.

52. D. Krout, A. Schaar, Y. Sun, P. Sukumaran, J. N. Roemmich, B. B. Singh and K. J. Claycombe-Larson, J Biol Chem, 2017, 292, 20799-20807.

View Article Online
DOI: 10.1039/D6FO00791K

53. J. M. Zhao, Z. H. Su, Q. Y. Han, M. Wang, X. Liu, J. Li, S. Y. Huang, J. Chen, X. W. Li, X. Y. Chen, Z. L. Guo, S. Jiang, J. Pan, T. Li, W. Xue and T. Zhou, Nutr Metab (Lond), 2024, 21, 2.

54. Q. Huang, G. Deng, R. Wei, Q. Wang, D. Zou and J. Wei, Front Cardiovasc Med, 2020, 7, 580573.

55. H. Yu, Y. Gao, T. He, M. Li, Y. Zhang, J. Zheng, B. Jiang, C. Chen, D. Ke, Y. Liu and J. Z. Wang, Front Cell Dev Biol, 2022, 10, 818141.

56. R. Chen, X. Luo, X. Jiang and S. Deng, Biochem Biophys Res Commun, 2022, 596, 14-21.

57. R. D. Levitan, A. S. Kaplan, M. Masellis, V. S. Basile, M. L. Walker, N. Lipson, G. I. Siegel, D. B. Woodside, F. M. Macciardi, S. H. Kennedy and J. L. Kennedy, Biol Psychiatry, 2001, 50, 640-643.

58. E. Flowers, B. Stroebel, K. A. Lewis, B. E. Aouizerat, M. Gadgil, A. M. Kanaya, L. Zhang and X. Gong, Front Endocrinol (Lausanne), 2024, 15, 1419812.

59. E. Saglam, H. Karagedik, M. Dinc, D. Oke, P. Gun Atak, B. Karadeniz, G. Burul and U. Gormus Degrigo, Cureus, 2024, 16, e67025.

60. T. Maxel, K. Smidt, A. Larsen, M. Bennetzen, K. Cullberg, K. Fjeldborg, S. Lund, S. B. Pedersen and J. Rungby, BMC Obes, 2015, 2, 46.

61. T. Platek, A. Polus, J. Goralska, U. Razny, A. Gruca, B. Kiec-Wilk, P. Zabielski, M. Kapusta, K. Slowinska-Solnica, B. Solnica, M. Malczewska-Malec and A. Dembinska-Kiec, Mol Med, 2020, 26, 93.

View Article Online
DOI: 10.1039/D6FO00791K

62. D. M. Gatti, K. L. Svenson, A. Shabalin, L. Y. Wu, W. Valdar, P. Simecek, N. Goodwin, R. Cheng, D. Pomp, A. Palmer, E. J. Chesler, K. W. Broman and G. A. Churchill, G3 (Bethesda), 2014, 4, 1623-1633.

Figure legends:

Figure 1. Mouse body masses (**A**), body fat % (**B**), fasting blood insulin (**C**) and fasting blood glucose (**D**) in *N*=50 DO mice fed a low-fat diet prior to the experiment, a high-fat diet (HFD, weeks 1-8) and then a HFD + EGCG (weeks 9-16). Values (dots) represent individual mice. Lines link the same mice over time. Previously published data, adapted with permission from Elsevier.

Figure 2 (A) Marker-level quality control of the SNP dataset. Distributions of heterozygosity, minor allele frequency (MAF), and genotype concordance (R^2) are shown across 19 chromosomes. SNPs displayed balanced heterozygosity, broad allele frequency coverage consistent with the applied MAF threshold, and uniformly high R^2 values, confirming robust marker quality. **(B)** Principal component analysis (PCA) of the filtered genotype dataset. Variance explained by the first 10 PCs is shown. The first three PCs captured the major sources of genetic variation (PC1 = 3.05%, PC2 = 2.9%, PC3 = 2.8%) and were selected for use as covariates in GWAS models to control for population structure.

Figure 3. Linkage disequilibrium (LD) patterns across the genome. **(A)** Genome-wide LD (R) across markers. **(B)** Frequency distribution of pairwise R values. **(C)** Decay of LD (R) with physical distance (Kb). **(D)** Pairwise LD distances (Kb) across the genome. **(E)** Frequency

View Article Online
DOI: 10.1039/D6FO00791K

distribution of pairwise marker distances. **(F)** LD decay plot showing average R^2 declining with increasing inter-marker distance, with a gradual decay observed after ~2 Kb, consistent with moderate LD structure in the population.

Figure 4. Genome-wide association analysis of insulin concentration in mice. **(A)** Circular Manhattan plot showing genome-wide marker–trait associations across 19 autosomes using five GWAS models: GLM, MLM, MLMM, FarmCPU, and BLINK. Strong and consistent association signals are observed on chromosome 14. **(B)** Model-specific Manhattan plots display the distribution of $-\log_{10}(p)$ values for each model. **(C)** Quantile–quantile (QQ) plot showing the observed versus expected $-\log_{10}(p)$ values across the five models, indicating effective correction for population structure and kinship, with deviation in the upper tail representing true associations. (D) Allelic effect plots for one the top significant SNPs, UNCHS019512 according to genotype showing significant differences between genotypic classes for changes in insulin level.

Figure 5. Genome-wide association analysis of body fat mass (BFM) in mice. **(A)** Circular Manhattan plot showing genome-wide marker–trait associations across 19 autosomes using five GWAS models. Strong association signals are observed on chromosome 9 and 10. (B) Model-specific Manhattan plots display the distribution of $-\log_{10}(p)$ values for each model. **(C)** Quantile–quantile (QQ) plot showing the observed versus expected $-\log_{10}(p)$ values across the five models, indicating effective correction for population structure and kinship, with deviation in the upper tail representing true associations. (**D)** Allelic effect plots for one the top significant

View Article Online
DOI: 10.1039/D6FO00791K

SNPs, UNCHS029548, according to genotype showing significant differences between genotypic classes for changes in body fat mass.

Figure 6. Genome-wide association analysis of glucose content in mice. **(A)** Circular Manhattan plot showing genome-wide marker–trait associations across 19 autosomes using five GWAS models. Only association signals were observed on chromosome 17. **(B)** Model-specific Manhattan plots (right) display the distribution of $-\log_{10}(p)$ values for each model. **(C)** Quantile–quantile (QQ) plot showing the observed versus expected $-\log_{10}(p)$ values across the five models, indicating effective correction for population structure and kinship, with deviation in the upper tail representing true associations. **(D)** Allelic effect plots for the top significant SNP, UNC28326389, according to genotype showing significant differences between genotypic classes for changes in glucose level.

Figure 7. Genome-wide association analyses for glucose, insulin, body fat mass (BFM) and body weight (BW) using **(A)** FarmCPU, **(B)** BLINK and **(C)** MLMM revealed a shared major locus on chromosome 10 associated with BFM and insulin. The top SNP (UNCHS029548:10:127827082) showed a strong additive, dosage-dependent effect, where individuals carrying the alternate allele exhibited significantly higher insulin levels (D) and BFM (E), indicates a pleiotropic locus regulating both traits.

Figure 8. Classification of genes located near trait-associated SNPs. Histogram plots show the number and type of genes (protein-coding genes, snoRNA genes, pseudogenes, miRNAs,

View Article Online
DOI: 10.1039/D6FO00791K

lncRNAs, and unknown) identified in genomic regions surrounding SNPs associated with variability in insulin levels, body fat mass (BFM), and glucose levels.

Figure 9. Gene category and pathway analysis for genes in the regions of SNPs associated with metabolic traits. **(A)** Sankey diagram showing to SNPs associated with variability in insulin level, body fat mass, and glucose post-EGCG consumption, followed by genes located in proximity to SNPs and on the right functional categories of identified genes (protein-coding, non-coding, pseudogenes, and unknown). **(B)** circus plot illustrating the relationships between selected protein-coding genes and enriched biological pathways, highlighting shared involvement of key genes across insulin, adiposity, and glucose-related pathways.

Figure 1.

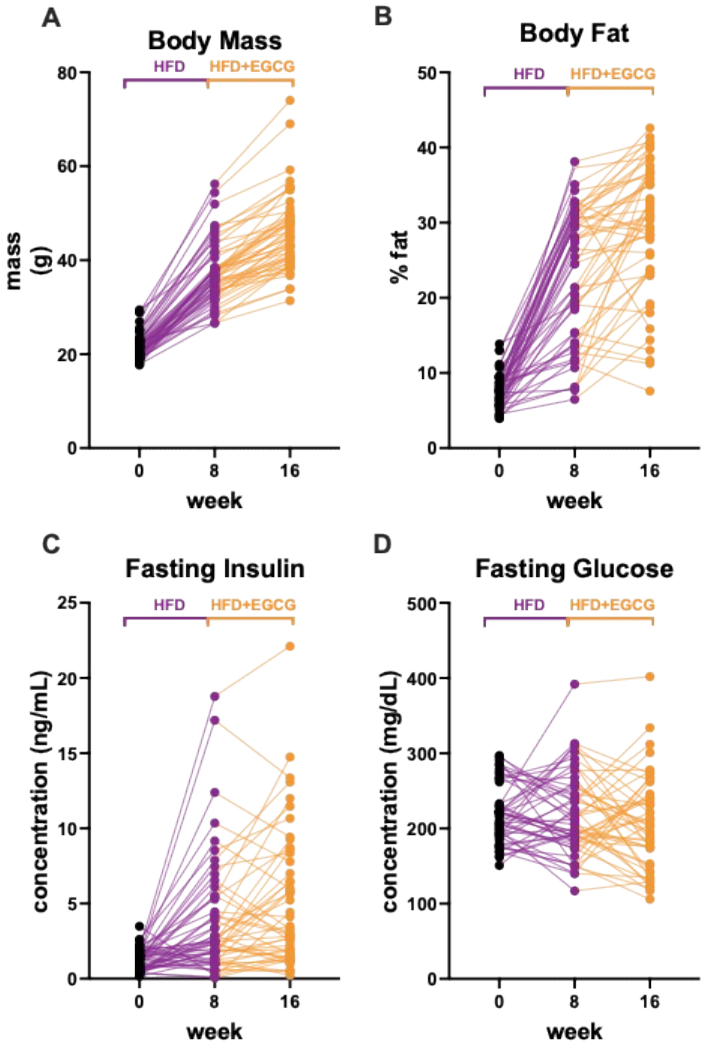


Figure 2

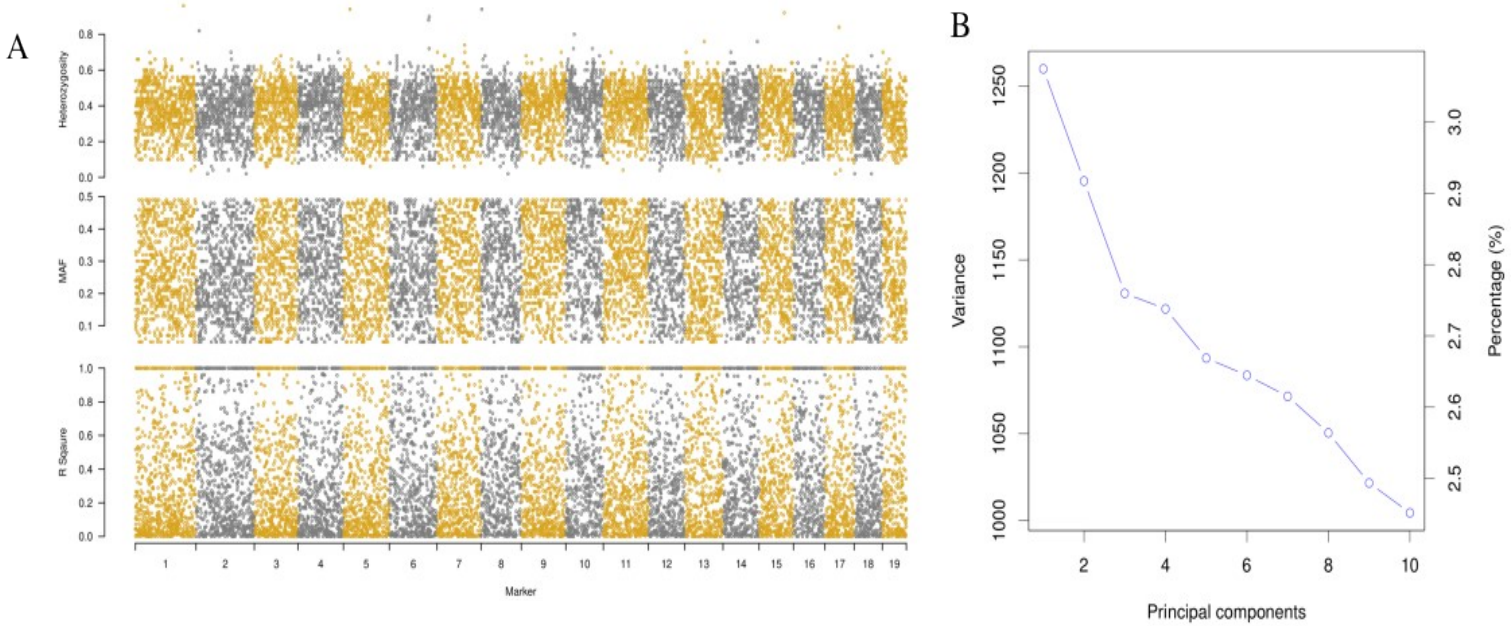


Figure 3

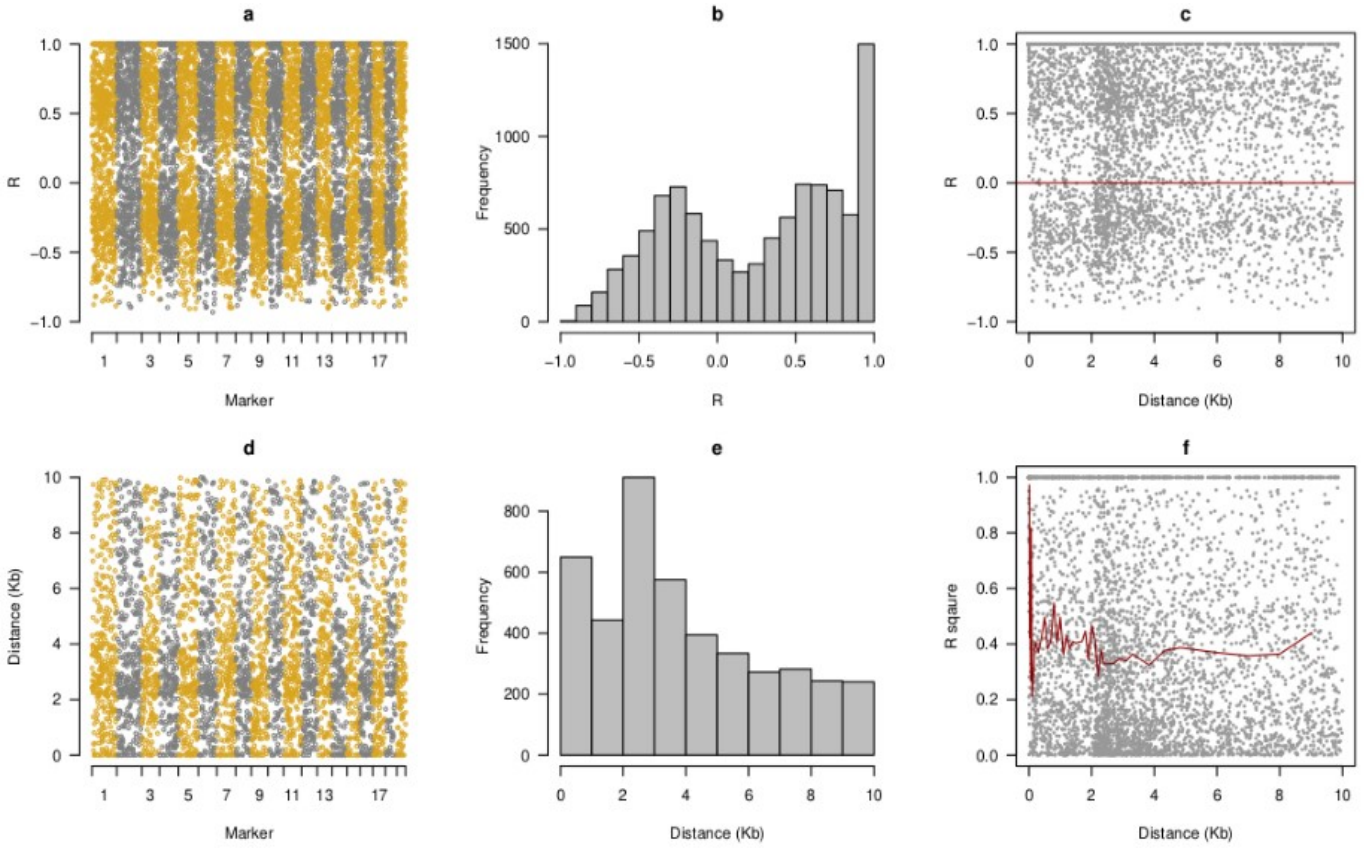
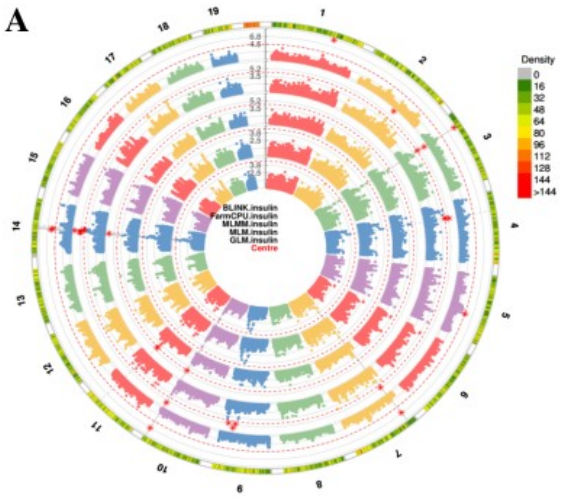
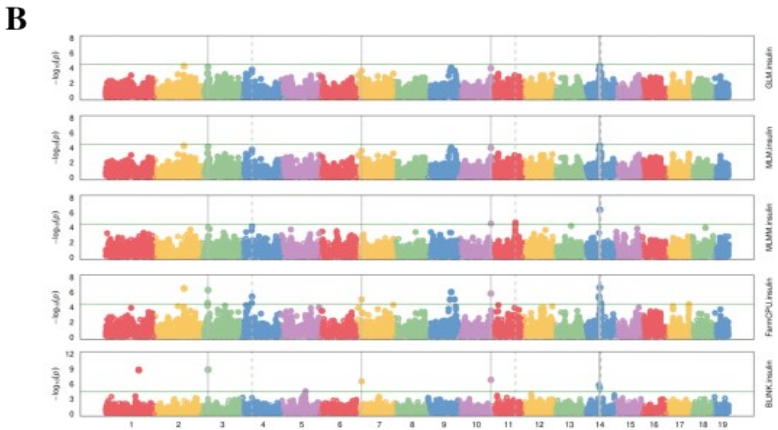
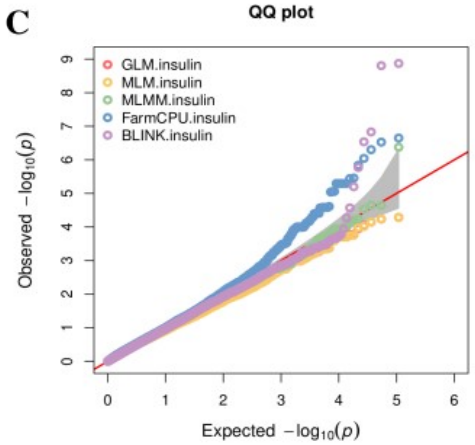


Figure 4







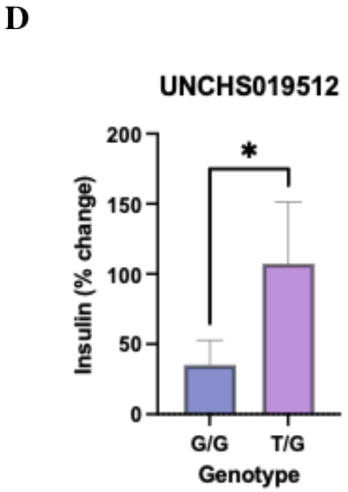


Figure 5

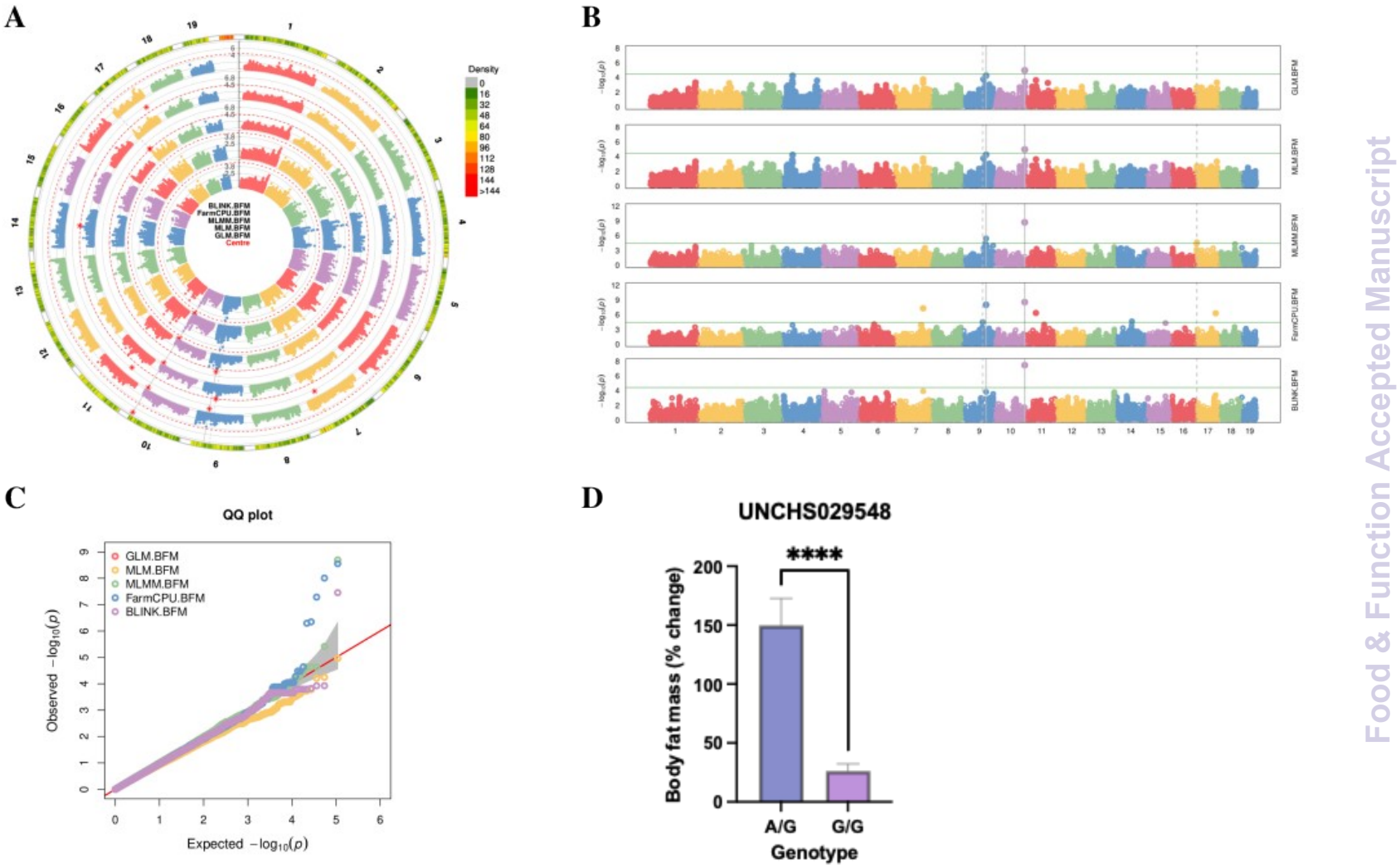


Figure 6

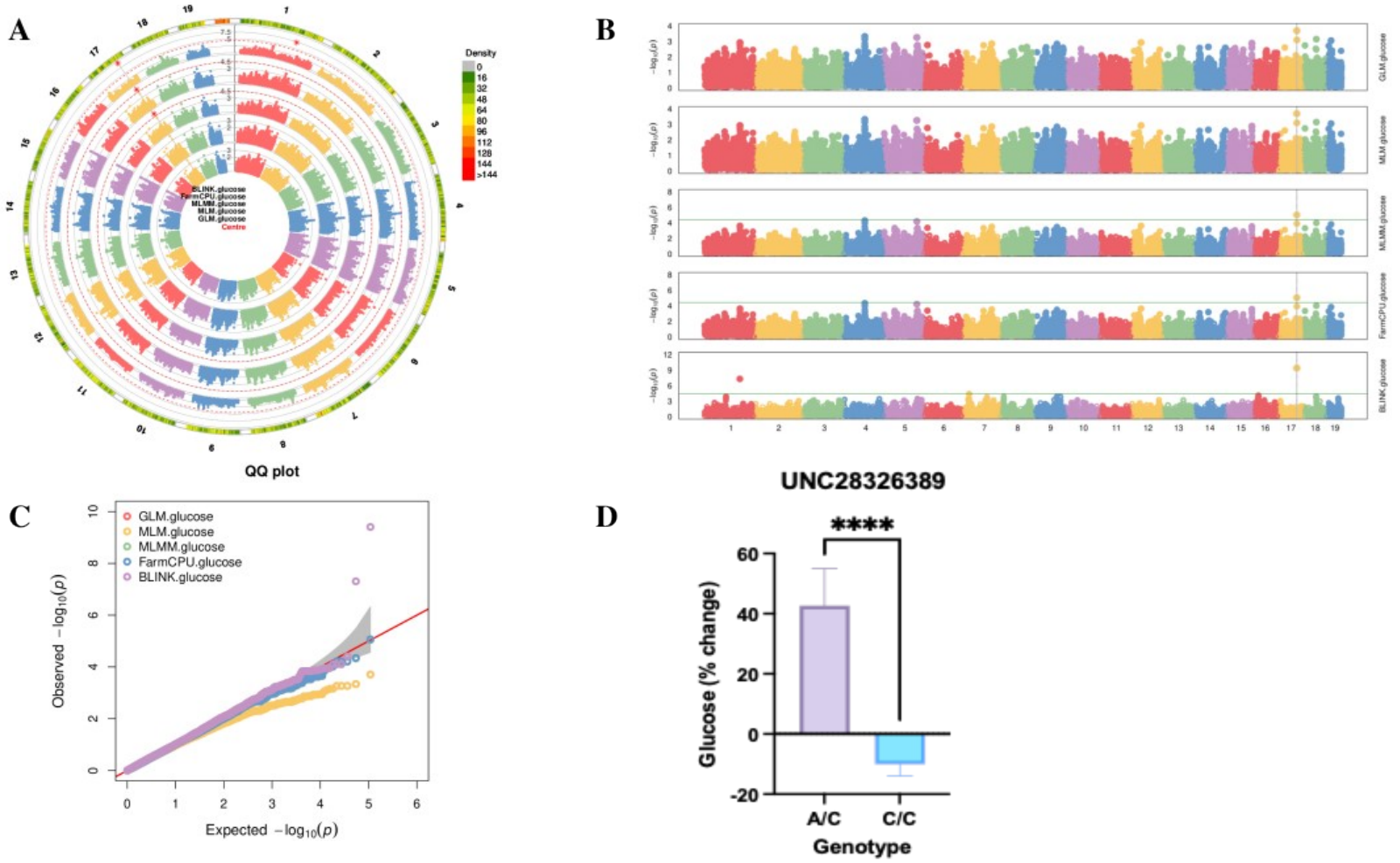


Figure 7

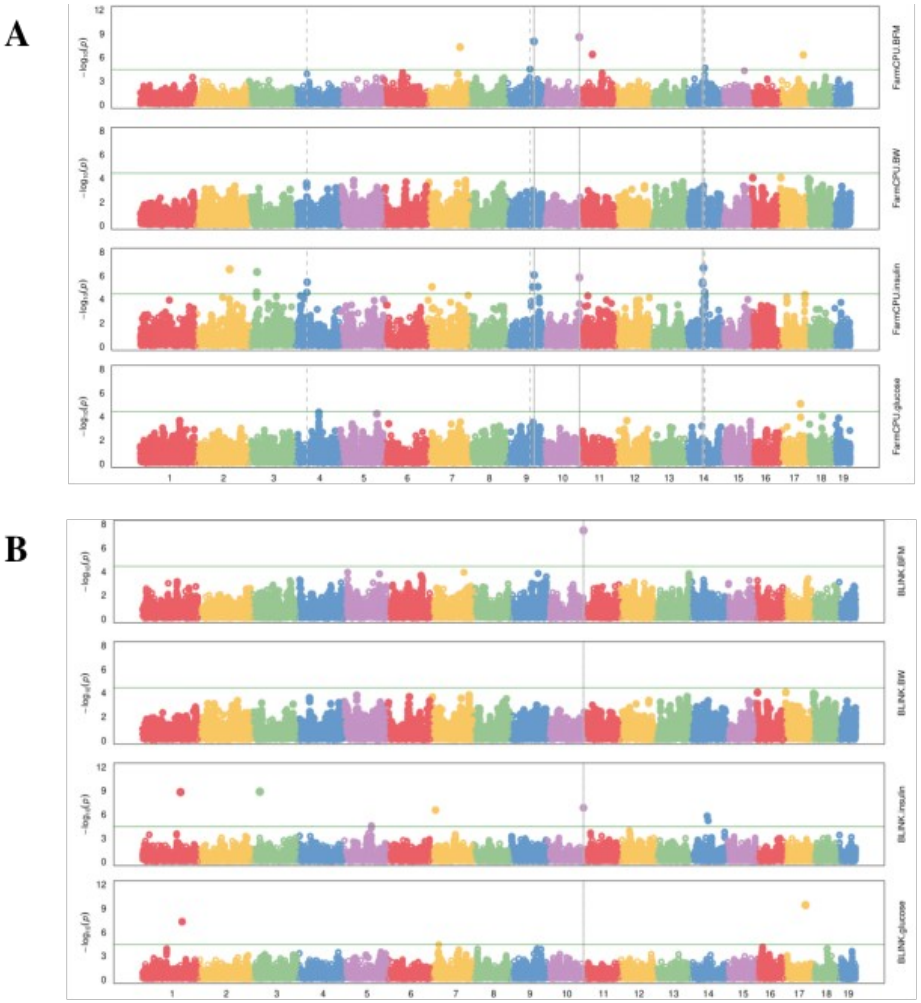


Fig 8

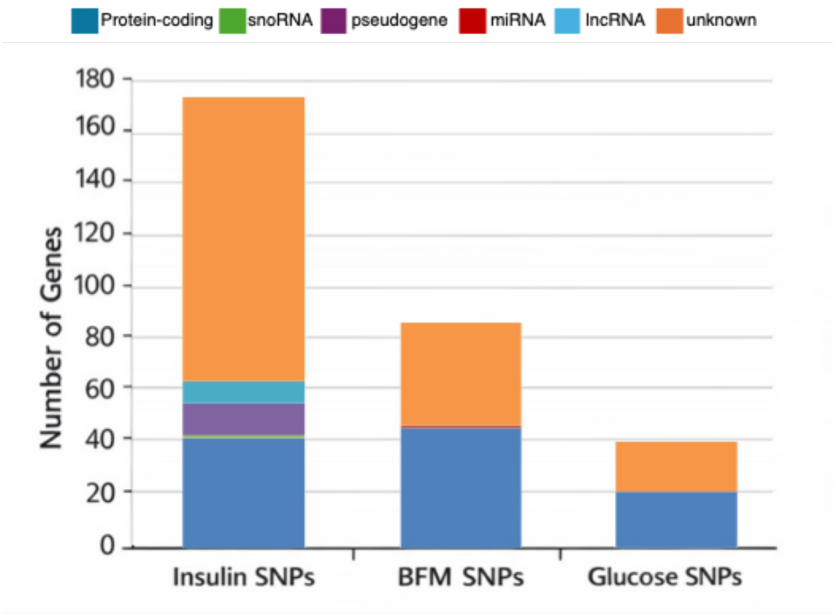
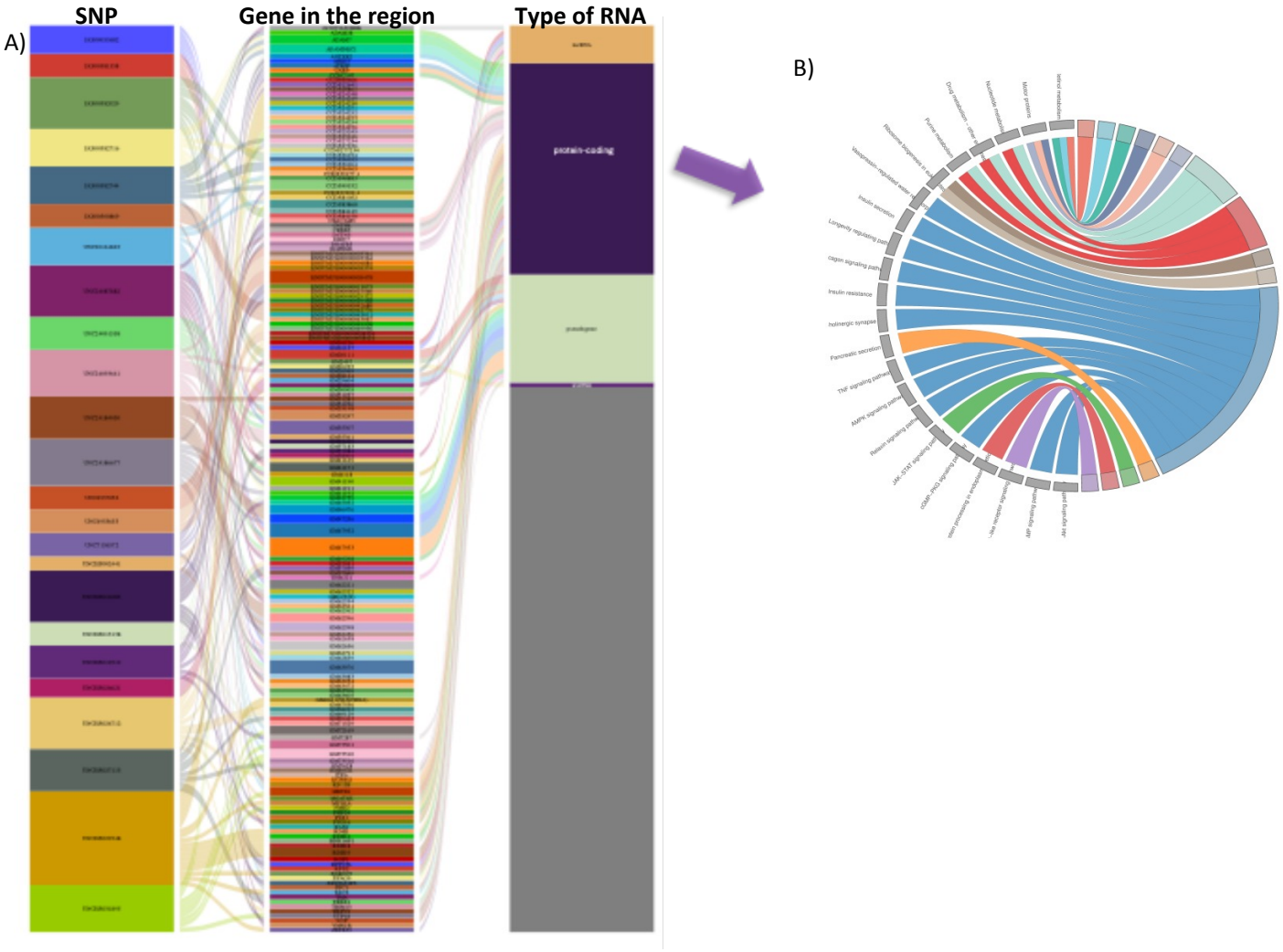


Fig 9



View Article Online
DOI: 10.1039/D6FO00791K

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author