

# Container-aided Integrative QTL and RNA-seq Analysis of Collaborative Cross Mice Supports Distinct Sex-Oriented Molecular Modes of Response in Obesity

Ilona Binenbaum (✉ [ilonab@upatras.gr](mailto:ilonab@upatras.gr))

University of Patras <https://orcid.org/0000-0002-1045-5725>

Hanifa Abu-Toamih Atamni

Tel Aviv University Sackler Faculty of Medicine

Georgios Fotakis

e-NIOS PC

Georgia Kontogianni

Biomedical Research Foundation of the Academy of Athens

Theodoros Koutsandreas

e-NIOS PC

Eleftherios Pilalis

e-NIOS PC

Richard Mott

University College London

Heinz Himmelbauer

Universitat fur Bodenkultur Wien

Fuad A. Iraqi

Tel Aviv University Sackler Faculty of Medicine

Aristotelis Chatziioannou

Biomedical Research Foundation of the Academy of Athens

---

## Research article

**Keywords:** Collaborative cross, obesity, sex-differences, high-fat diet, QTL, RNAseq

**Posted Date:** September 15th, 2020

**DOI:** <https://doi.org/10.21203/rs.2.13505/v2>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on November 3rd, 2020. See the published version at <https://doi.org/10.1186/s12864-020-07173-x>.

# Abstract

**Background:** The CC mouse population is a valuable resource to study the genetic basis of complex traits, such as obesity. Although the development of obesity is influenced by environmental factors, the underlying genetic mechanisms play a crucial role in the response to these factors. The interplay between the genetic background and the gene expression pattern can provide further insight into this response, but we lack robust and easily reproducible workflows to integrate genomic and transcriptomic information in the CC mouse population.

**Results:** We established an automated and reproducible integrative workflow to analyse complex traits in the CC mouse genetic reference panel at the genomic and transcriptomic levels. We implemented the analytical workflow to assess the underlying genetic mechanisms of host susceptibility to diet induced obesity and integrate these results with diet induced changes in the hepatic gene expression of susceptible and resistant mice. Hepatic gene expression differs significantly between obese and non-obese mice, with a significant sex effect, where male and female mice exhibit different responses and coping mechanisms.

**Conclusion:** Integration of the data showed that different genes but similar pathways are involved in the genetic susceptibility and disturbed in diet induced obesity. Genetic mechanisms underlying susceptibility to high-fat diet induced obesity differ in female and male mice. The clear distinction we observe in the systemic response to the high-fat diet challenge and to obesity between male and female mice points to the need for further research into distinct sex-related mechanisms in metabolic disease.

## Background

The Collaborative Cross (CC) is a multiparent genetic reference panel (GRP) of recombinant inbred lines (RIL) of mice derived from eight different founder strains. The CC resource was developed to facilitate the study of the genetic basis of complex traits, and serve as a uniquely powerful resource for the mapping and integration of various phenotypes and genotypes data [1]. CC lines have been shown to be highly variable for traits related to both normal physiology and disease and have been used successfully in multiple system genetics studies. The CC panel has been the catalyst for the development of a variety of bioinformatics tools for haplotype inference and reconstruction, and genetic mapping [2,3,4].

Based on a recent report, people with overweight or obese phenotype account for almost two thirds of the population in the U.S.A. Based on the latest estimates in European Union countries, overweight affects 30-70% and obesity affects 10-30% of adults. Some medical associations classify obesity (defined as BMI  $\geq$  30) as a disease. The development of obesity is affected by various environmental factors such as excess high fat/carbohydrate enriched food consumption and sedentary lifestyle. However, underlying genetic mechanisms are involved in determining the host response to these factors, with the rate of heritability of body mass index (BMI) ranging from 40 to 70% in various studies. The CC panel is an experimental population that aids with the dissection of the genetic mechanisms underlying

susceptibility to complex traits, such as obesity. Chromosomal regions that are involved in the susceptibility or resistance to the trait can be mapped as quantitative trait loci (QTL) with high precision. QTL mapping results can be strengthened and enriched through integration of RNA-seq data to identify gene expression differences in susceptible versus resistant individuals.

High-throughput sequencing technologies produce millions of reads in relatively short time, overcoming the limitations of previous technologies and unravelling previously inaccessible complexities in the transcriptome. However, the overall high complexity of the produced datasets, due to their large size and low signal-to-noise ratio, hinders the interpretation of the underlying information. Most of the recent analytical methods depend on individual tools that the user must download, install and use on their physical drives. The process of deciding which bioinformatic tool better suits the needs of the researcher (depending on the experimental approach, the scientific question of interest, as well as the computational needs) might be time consuming and requires expertise.

Our main aim in this paper, was to develop an easy-to-use, scalable and cost-effective workflow for the integrative analysis of genotyping and RNA-seq data from CC mice, offering cross-platform portability to different high-end computing configurations. We then aimed to use this workflow to assess the underlying QTL that influence genomic susceptibility to high-fat diet induced obesity in mice and the hepatic gene expression response of the mice to high-fat diet and obesity.

The tools used in each step of the developed workflow are connected through Python scripts that can be found in our public GitHub repository and can be fully modified. The Python scripts are constructed in a user-friendly manner, so that casual users are required to input only the basic parameters (such as the paths to input files, indexes and annotation files) and the whole process will be executed automatically, assuming default parameters. More advanced users can apply different sets of input commands or fully modify the scripts according to their requirements.

All bioinformatic tools comprising this workflow have been containerised using Docker containers, resulting in a portable infrastructure that can be executed on physical drives or cloud servers. The advantage of the containerisation is that it replaces the installation of numerous pieces of software, with complex dependencies, by downloading a pre-built and ready-to-run image file, containing all necessary software and their required dependencies. Another advantage, incurred by the utilization of containers, is that the applications run in an isolated and sanitised container environment, where all dependencies are configured for optimal performance, preventing conflicts with other installed programs, in the hosting environment. The containerisation ensures the standard operation and performance of the applications, that is not affected from system updates or programming errors from the host end, making the process transparent to the end user and fully reproducible every time. Furthermore, each Docker container can be used as a stand-alone version of the tool making the workflow scalable and adaptable to the individual needs of the researcher.

A question that arises is to what extent the start-up, deployment and instantiation of the containers performed by the Docker daemon affect the overall performance and the computational cost of a

workflow. An IBM research study suggests that the overhead for CPU and memory performance introduced by Docker containers is negligible, and that containerised applications perform equally or better when compared to virtual machines technology [5]. As far as the containerisation of genomic pipelines, it has been suggested that Docker containerisation has a negligible impact on the execution performance of common genomic pipelines, especially when the tasks (such as the alignment of reads to a reference genome) are generally very time consuming [6].

## Methods

### Experiments

**Breeding and housing:** The Collaborative Cross mouse lines are novel and highly genetically diverse mouse resource population derived from a genetically diverse set of eight founder mouse strains (A/J, C57BL/6J, .129 Sv/ImJ, NOD, NZO, CAST, PWK, WSB), designed specifically for complex trait analysis, with the aim to overcome the limitations of previously available resources [63]. A cohort of CC lines were developed and currently are housed at conventional environmental conditions at the small animal facility of Tel-Aviv University (TAU) between generations of G27 to G64 of inbreeding by full-sib mating (Iraqi et al., 2012). The CC lines are maintained and bred by our team at TAU, and the study was approved by the Institutional Animal Care and Use Committee (IACUC), with the approval numbers M-12-025 and M-14-007.

Mice were weaned at age of 3 weeks old, housed separately by sex, maximum five mice per cage, with standard rodents' chow diet (TD.2018SC, Teklad Global, Harlan Inc., Madison, WI, USA, containing % Kcal from Fat 18%, Protein 24%, and Carbohydrates 58%) and water ad libitum. All animals housed in TAU animal facility at conventional open environment conditions, in clean polycarbonate cages with stainless metal covers, and bedded with wood shavings. A Light: dark cycles of 12:12 hours, and constant room temperature of 22<sup>0</sup> c (±2). Due to genetic variations between the CC lines, breeding rate, number and sex of litters in each cycle might vary. Therefore, the CC lines were assessed based on litters' availability, while making our best efforts to scan as many as possible of the CC lines with representation of both sexes.

**Study cohort.** The study cohort consisted of 540 mice, from 60 different CC lines that were generated, maintained and studied at TAU small animal facility. 43 of lines had presentation of both sexes. A subgroup of 84 mice from 43 CC lines, with 20 lines with presentation of both sexes, was selected for RNA sequencing.

**Dietary challenge:** At the age of 8 weeks old, baseline body weight was recorded (average body weight (±SE) of females is 18.92 gr (±0.27) and males gr 22.92 (±0.24); Abu-Toamih Atamni et al. 2016a). Thereafter, mice were switched to high-fat diet (HFD) starting the dietary challenge (TD 88137 Harlan Teklad, Madison, WI, USA; containing 42% of calories from fat and 34.1% from carbohydrate, primarily sucrose) for a period of 12 weeks with free access to food and water. During the experiment, mice welfare

and health status were monitored daily. The experiment was stopped (euthanasia) for mice that showed deterioration in health, which is manifested in phenomena such as limited movement, heavy weight loss (about 10% between the weekly weigh measures and over 20% of the initial body weight), apathy and lack of physical activity, interrupted equilibrium (physical instability), breathing difficulty, exceptional behaviour (high aggressiveness / loneliness), extremely high glucose levels (> 400 mg/dL).

**Phenotyping.** Mice were weighed at the beginning of the high-fat diet challenge and bi-weekly for the following 12 weeks. Mice that gained less than 10 gr to their initial body weight over the dietary challenge period were considered Normal, while those that gained over 10 gr of body weight were considered Obese. The median weights at the beginning of the dietary challenge and the median body weight gain for each CC-line are presented in Supplementary table 1. At the end of 12 weeks dietary challenge, an intraperitoneal glucose tolerance test (IPGTT) was performed after 6 hours fasting to evaluate the diabetic stage of the mice. By using the method of IP injection of glucose in this study, the gut effect bypassed and the glucose stimulated insulin secretion is lower, compared to Oral administration of glucose. The IPGTT lasted 180 minutes, with glucose levels measured at different time points before and after the glucose administration. Glucose tolerance was calculated as the total area under the curve between the initial and end time point of the test, for each CC line, separately for females and males. Mice that reached the end time point of IPGTT with glucose levels under 400 mg/dL were considered Non-diabetic, while mice with glucose levels over 400 mg/dL were labelled as Diabetic. The complete CC line, sex, baseline body weight at the start of the dietary challenge, body weight gain after 12 weeks of dietary challenge, and blood glucose levels at 180 minutes after the IPGTT challenge data included in the analysis are presented in Supplementary table 2.

**RNA extraction:** After overnight recovery from the IPGTT, mice were sacrificed by cervical dislocation protocol, and their livers were collected to liquid nitrogen and thereafter stored in a freezer (-80°C). RNA extraction was performed using a QIAGEN commercial kit (Cat.No.73404). Quality control of the RNA samples was performed with 2100 BioAnalyzer (Agilent). The RNA Integrity Number (RIN) software was used to estimate the integrity of the total RNA sample, samples with RIN above 7.0 passed the quality control test.

**RNA-seq libraries.** RNA-seq libraries were prepared using the TruSeq Stranded mRNA library preparation kit (Illumina). Libraries were pooled and sequenced on the Illumina HiSeq 2000 and 2500 sequencers with Illumina v3 sequencing chemistry. Paired-end sequencing was performed by reading 50 bases at each end of a fragment. Overall, each sample consisted of 24M to 37.5M RNA-sequencing fragments with an average of 31.5 M fragments.

**RNA-seq analysis:** The base images for the Docker containers were pulled from the repositories of Biocontainers and Rocker respectively [7]. For the purposes of Differential Expression testing and QTL analysis we developed Docker containers with integrated R v.3.4.2 and all the required packages installed (such as Bioconductor, EdgeR and HAPPY), the Docker images also contain ready-to-run R scripts set on default parameters so that users unfamiliar with R programming can perform the analyses using R

packages otherwise unavailable as stand-alone versions (EdgeR and HAPPY), while advanced users can modify the scripts in a sanitised Docker container environment. In order to run the workflow we developed, the user is required to install the docker engine locally and then pull the images from the Docker Hub repository. This task is relatively fast and requests minimal programming knowledge in comparison to the skills needed for downloading and installing a combination of several pieces of third-party software, while configuring their implicit dependencies and libraries.

**Quality Control:** For quality assessment of the reads we used the FastQC tool [8], which is the golden standard for quality control workflows. We did not incorporate a trimming tool in our workflow as aggressive trimming of reads has been suggested to alter RNA-seq expression estimates and the soft-clipping performed by HISAT2 makes read trimming not strictly necessary [9]. However, we provide a docker container with Trim Galore!, which is a tool that makes use of the publicly available adapter trimming tool Cutadapt [10] and FastQC for optional quality control [11].

**Mapping of reads:** Mapping of the reads was performed using the HISAT2 tool, which is a fast and sensitive alignment program for mapping RNA-seq reads [12]. The high efficiency of HISAT2 is based on the indexing scheme it utilises (employing two types of indexes based on Burrows - Wheeler transform and the Ferragina-Manzini index), allowing the tool to perform alignments very fast and with equal or better accuracy than any other method currently available. HISAT2 supports genomes of any size and has low memory requirements (approximately 4.3 GB of RAM for the human genome).

**Assigning sequence reads to genomic features:** For the purpose of counting reads to genomic features (such as genes, exons, promoters and genomic bins) we used the software program FeatureCounts (v1.5.0-p1), from the Subread stand-alone package [13]. FeatureCounts is considerably faster than existing methods and has exceptionally low memory requirements, while being one of the top-ranking software in accuracy.

**DE Testing:** The DE testing was performed with the R package EdgeR [14] using a modified Rscript from Su [15]. The Rscript is run through a Docker container that employs R version 3.3.0 and has been built with Bioconductor version 3.4 and all the required packages installed. The normalisation (for library size, gene length and sequencing depth) was performed on the raw count matrix produced by FeatureCounts using the Trimmed Mean of M-values or TMM [16], which is set as the default normalisation method. A negative binomial generalized linear model (GLM) was fitted to the data and the testing procedure for determining differential expression was performed using quasi-likelihood (QL) F-test. The P-value adjustment was performed using the Benjamini-Hochberg method and in order to restrict the false discovery rate (FDR) to 0.05, all the genes with adjusted P-values less than 0.05 were selected. The filtering of the genes list (threshold default values: adjusted P-value 0.05 and  $|\log_2FC| \geq 1$ ) was performed with a custom python script.

**Functional Analysis:** The functional pathway analysis was performed with the BioInfoMiner platform [17]. BioInfoMiner exploits biological hierarchical vocabularies through statistical and network analysis to detect and rank significantly altered processes and the hub linker genes involved. For our analysis we

utilised Gene Ontology (GO) [18] and MGI Mammalian Phenotype Ontology [19]. BioInfoMiner maps the genes on a genomic network created from semantic data and prioritises them based on topological properties after minimising the impact of semantic noise (bias) through different types of correction. This analysis accomplishes the systemic interpretation of the complex cellular mechanisms described in the input gene lists, while at the same time it prioritises genes with central functional and regulatory roles in important cellular processes, underlying the studied phenotype. The correction for potential semantic inconsistencies on the selected vocabulary is implemented by linking the annotation of each gene with the ancestors of every direct correlated ontological term, consequently restoring the sound structure of an ontological tree. The BioInfoMiner platform is available online at the website <https://bioinforminer.com>.

## Genetic mapping

**Genotyping:** All CC lines were genotyped with high-density SNP markers using the MDA (620K SNP markers), MUGA (7.7K markers) and MegaMUGA (77K markers) genotyping arrays based on the Illumina Infinium platform. All SNPs with heterozygous or missing genotypes in the 8 CC founders or not common between the arrays were filtered out, leaving 170,935 SNPs. The SNPs were mapped onto build 37 of the mouse genome. A descent probability distribution was computed using the HAPPY HMM for each of the 170,935 SNPs intervals. The genotype status of the CC lines using the MUGA SNP was presented in CTC 2012 report [20, 21].

**QTL mapping.** The QTL mapping was based on the haplotype mosaic reconstruction with HAPPY. We developed two automated containerised R scripts using the mapping methodology proposed in [22]. The first script uses the probability distribution of descent as calculated by the HAPPY algorithm to test for association between the reconstructed haplotype for each CC line at each locus and the median body weight gain at different time points.

The second script estimates confidence intervals (CI) for each QTL through simulation of a QTL with a similar logP and strain effects in the neighbourhood of the observed QTL peak. We used the SNPtools R package to extract the genes inside the 95% CI for each QTL [23].

# Results

## Data stratification

It is well documented that obesity and related health complications are affected by gender, and that sex-specific differences have a genetic basis and cannot be attributed only to differences in hormonal regulation [24]. Gender-specific differences in adiposity as well as fat distribution, in addition to the distinctive genetic basis and hormonal regulation of men and women, may result in sex-specific patterns.

In order to assess the presence of confounding factors, we performed a multi-dimensional scaling analysis, using the EdgeR R package. According to the MDS plot (Figure 1) the points separate into two groups, but not on the basis of the differences between the two conditions (normal vs obese). This

finding led to the conclusion that the observed separation in the principal component analysis performed, was the result of gender-specific differences, meaning that a straightforward approach would be prone to confounding biases.

In order to partition the samples into non-overlapping groups, we performed a stratified sampling dividing the total sample population into four strata, considering the sex and diabetes status of the mice. Of the total 85 samples, 21 samples were categorised in the Male-Nondiabetic group, 21 samples were categorised in the Male-Diabetic group, 36 samples were categorised in the Female-Nondiabetic group and 5 samples in the Female-Diabetic group. One sample (DT 123) was not used in the analysis due to missing data.

## QTL analysis

We performed haplotype association analysis, using the median value of the weight gained in the twelve weeks of the high-fat diet challenge. We mapped one QTL on chr3 for female mice and one QTL on chr5 for male mice, designated as ObFL and ObML for obesity female locus and obesity male locus, respectively. Details on the position and size of each QTL are given in Table 1. We performed functional analysis with BioInfoMiner on the genes extracted from each of the QTLs using the Gene Ontology and MGI Mammalian Phenotype Ontology vocabularies.

***Female mice body weight gain:*** In female mice, prioritised genes included *Sgms2* and *Hadh* that were found to be involved in various processes, such as increased energy expenditure (MP:0004889), decreased susceptibility to diet-induced obesity (MP:0005659) and increased circulating free fatty acid level (MP:0001554). Interestingly, *Sgms2* deficiency in mice increases insulin sensitivity and ameliorates high-fat diet-induced obesity and *Hadh*<sup>-/-</sup> mice, while having a disrupted  $\beta$ -oxidation pathway, are also protected from diet-induced obesity [25, 26]. Other prioritised genes are *Lef1*, *Dkk2* and *Egf*, which are all involved in the Wnt signaling pathway (GO:0016055). Noncanonical Wnt signaling has been shown to contribute to obesity-associated metabolic dysfunction by increasing adipose tissue inflammation [27].

***Male mice body weight gain:*** In male mice the gene that is most highly prioritised is *Ppargc1a*, which is found to be involved in various enriched processes both in GO and in MGI Mammalian Phenotype Ontology, including decreased muscle weight (MP:0004232), regulation of muscle tissue development (GO:1901863) and lipid modification (GO:0030258). *Ppargc1a* is a transcriptional coactivator that regulates genes involved in energy metabolism through its interaction with *Ppar $\gamma$* . In human GWAS analysis a single nucleotide variant of *Ppargc1a* (rs8192678) has been associated with susceptibility to obesity and insulin resistance [28]. Another important gene, prioritised by both vocabularies, was *Cckar*, involved in processes such as abnormal small intestinal transit time (MP:0006002) and abnormal intestinal cholesterol absorption (MP:0002645). Rats with a naturally occurring mutation in *Cckar* (Otsuka Long-Evans Tokushima Fatty (OLETF) rat) develop diabetes and obesity [29]. Prioritised gene *Sod3* is involved along with *Ppargc1a* in: response to reactive oxygen species (GO:0000302), increased susceptibility to injury (MP:0005165), and abnormal cytokine secretion (MP:0003009). Over-expression of *Sod3* in high-fat diet fed mice has been shown to block diet induced obesity [30]. *Med28* is involved in the

regulation of muscle cell differentiation (GO:0051147). Med28 is one of the subunits of the Mediator complex, which acts as a transcription factor coactivator and plays an important role in muscle metabolism by enhancing the transcriptional activity of Ppargc1a and Ppara [31]. Lastly, Slit2 is also prioritised and also shares enriched terms with Ppargc1a, such as regulation of smooth muscle cell migration (GO:0014910) and response to organonitrogen compound (GO:0010243). Slit2 has been shown to regulate metabolic function and thermogenic activity and improve glucose homeostasis in diet-induced obese mice, mainly through Ucp1 [32], which has been found highly under-expressed in male obese mice, in comparison to female obese mice, in our analysis.

## RNA-seq analysis

DE analysis was performed on the Male-Nondiabetic and Female-Nondiabetic groups separately. In order to perform the statistical differential expression tests we divided the samples into two subgroups, 10 samples were categorised in the Male-Nondiabetic-Obese group (the male case group), 11 samples were categorised in the Male-Nondiabetic-Normal (the male control group) 12 samples were categorised in the Female-Nondiabetic-Obese group (the female case group) and 24 samples in the Female-Nondiabetic-Normal group (the female control group). Differences in library size were addressed through the TMM normalisation.

***Female non-diabetic normal vs female non-diabetic obese:*** As a first step we tried to assess the DE genes within the two genders, for this reason we performed the DE tests between control and case groups that belonged to the same gender. The Female-Nondiabetic-Normal samples (control group) versus the Female-Nondiabetic-Obese samples (case group) DE genes list consisted of 1382 DE genes. Using BioInfoMiner for the enrichment analysis we obtained two prioritised gene list of 23 DE genes from GO and 21 DE genes from MGI Mammalian Phenotype. Seven genes were common in both gene lists. Processes that were highly enriched in DE genes in female mice include digestion (GO:0007586), regulation of insulin secretion (GO:0050796), response to lipid (GO:0033993), fatty acid biosynthetic process (GO:0006633), increased energy expenditure (MP:0004889), decreased susceptibility to diet-induced obesity (MP:0005659), abnormal glucose tolerance (MP:0005291) and decreased circulating leptin level (MP:0005668). Top prioritised, linker genes in female mice include Lepr and Ppargc1a that were under-expressed in obese females and Pnlip, Pyy, and Ins2 that were over-expressed in obese females. Lepr functions as a receptor for leptin, an adipose secreted hormone that regulates energy expenditure, satiety, lipid and glucose metabolism and immune system activation. Pancreatic lipase is normally secreted from the pancreas and is efficient in the digestion of dietary fats. Inhibition of Pnlip may prevent high-fat diet-induced obesity in mice. Orlistat, an inhibitor of Pnlip was the first FDA-approved anti-obesity therapeutic target for the treatment of diet-induced obesity [33]. Extra pancreatic expression of Pnlip has been observed in mice induced by fasting via the PPAR $\alpha$ -FGF21 signalling pathway [34]. PYY is synthesised and released from specialised cells found predominantly within the distal gastrointestinal tract and regulates appetite. Transgenic mice with increased circulating PYY are resistant to diet-induced obesity [35]. Ins2 ectopic expression in the liver has been observed before in

mice subjected to high-fat diet [36]. We hypothesise that Pyy may be expressed ectopically in the liver in the same way as Ins2 in response to the high-fat diet.

**Male non-diabetic normal vs male non-diabetic obese:** The DE genes list for the Male-Nondiabetic-Normal samples (control group) versus the Male-Nondiabetic-Obese samples (case group) consisted of 1589 DE genes. We performed functional analysis on this list and we obtained two prioritised gene lists of 24 DE genes from GO and 22 genes from MGI Mammalian Phenotype. Ten of the genes from the two prioritised gene lists were common in both. In the case of male mice enriched pathways are overwhelmingly related to muscle and cardiac muscle processes: impaired muscle contractility (MP:0000738), abnormal skeletal muscle mass (MP:0004817), myopathy (MP:0000751), cardiac muscle hypertrophy (GO:0003300). A second category of enriched processes involves ion transport (GO:0006811). Top prioritised linker genes in male mice include Nos1, Ryr1, Des, Ttn all under-expressed in obese males while Ryr2 is over-expressed. Ryr1 is mainly expressed in skeletal muscle. The encoded protein functions as a calcium release channel in the sarcoplasmic reticulum. However, there is a number of studies suggesting that RyRs are widely expressed and have been linked with inositol 1,4,5-trisphosphate receptors in hepatocytes [37]. Down-regulation of Ryr1 could lead to reduced release of  $Ca^{2+}$  from the sarcoplasmic (muscle cells) and the endoplasmic reticulum (hepatic cells) into the cytoplasm and therefore hinder the triggering of muscle contraction as well as the regulation of mitochondrial metabolism and glycogen degradation [38, 39]. Des encodes a muscle-specific class III intermediate filament, which is important to help maintain the structure of sarcomeres and has been linked with hepatic fibrosis due to obesity [40]. Titin is an essential component of skeletal and cardiac muscles, but it has been suggested in recent studies that titin isoforms are expressed in non-muscle tissues including the liver, with an essential role in maintaining cellular organisation and contributing to signal transduction [41]. Ryr2 is primarily expressed in cardiac muscle. Ryr2 channels are associated with mitochondrial metabolism, gene expression and cell survival, in addition to their role in cardiomyocyte contraction. A recent study links Ryr2 to insulin release and glucose homeostasis, suggesting that the upregulation of Ryr2 might be a coping mechanism in response to stress [42]. Nos1 produces nitric oxide (NO), which has multiple biological functions. Downregulation of Nos1 in obesity and diabetes is largely attributed to insulin resistance. [43].

**Male vs female comparisons:** Finally, in order to assess the DE genes between the two genders we performed DE tests between groups that belonged to different genders. The DE genes list for the Male-Nondiabetic-Normal samples versus the Female-Nondiabetic-Normal samples comparison, consisted of 1923 genes. Functional analysis produced two prioritised gene lists consisting of 28 genes from GO and 25 genes from MGI Mammalian Phenotype, with 9 common genes. As for the DE genes for the Male-Nondiabetic-Obese samples versus the Female-Nondiabetic-Obese samples, the genes list consisted of 1940 DE genes. Functional analysis produced two prioritised gene lists consisting of 40 genes from GO and 22 genes from MGI, with 8 common genes. Between the two groups 482 DE genes are shared while 1441 genes are uniquely DE in the Male-Nondiabetic-Normal versus the Female-Nondiabetic-Normal samples and 1458 genes are uniquely DE in the Male-Nondiabetic-Obese versus the Female-Nondiabetic-Obese samples.

The 482 DE genes that are common in the two comparisons are enriched in genes that are involved in processes such as the epoxygenase P450 pathway (GO:0019373), long-chain fatty acid metabolic process (GO:0001676), negative regulation of gluconeogenesis (GO:0045721), increased circulating gonadotropin level (MP:0003362), and regulation of NF- $\kappa$ B import into nucleus (GO:0042345). Functional analysis of these 482 DE expressed genes produced 21 prioritised genes from GO and 20 prioritised genes from MGI Mammalian Phenotype. Of these 8 genes were common in both lists. Cav1, Egfr, Gstp1, Gcg, and Nox4 are consistently over-expressed in male versus female mice both in the non-obese and obese comparisons. Caveolin-1 regulates hepatic lipid accumulation and glucose metabolism and plays an important role in metabolic adaptation [44]. Hepatic glucagon action has been associated with elevated fatty acid oxidation and might act protectively against non-alcoholic fatty liver disease (NAFLD) [45]. Esr1, Il1b and Ptgs2 are consistently over-expressed in female versus male mice in both comparisons. Ptgs2 is involved in inflammation in fat and drives obesity-linked insulin resistance and fatty liver [46]. Tph1 is under-expressed in non-obese males and over-expressed in obese males in comparison to females. In contrast, Ryr2 is over-expressed in non-obese males and under-expressed in obese males in comparison to females. Tph1 is involved in the synthesis of serotonin, which is known to modulate appetite, energy expenditure and thermogenesis. Tph1-deficient mice fed a high-fat diet are protected from obesity, insulin resistance and NAFLD [47].

Functional analysis of the 1441 genes that were uniquely DE in the Male-Nondiabetic-Normal versus the Female-Nondiabetic-Normal comparison produced enriched processes like abnormal bone volume (MP:0010874), increased mesenteric fat pad weight (MP:0009298), unsaturated fatty acid and lipid metabolic process (GO:0033559, GO:0006629), regulation of hormone levels (GO:0010817), and a number of terms related to inflammatory processes, such as abnormal IgG2a level (MP:0020176), increased susceptibility to autoimmune diabetes (MP:0004803) and small intestinal inflammation (MP:0003306). The list of prioritised genes consisted of 21 genes from GO and 23 genes from MGI, with 9 overlapping genes. Genes involved in inflammatory processes, such as Il2, Il4, Il10, and Il1r1, were predominantly over-expressed in female mice. Vdr and Lepr were also over-expressed in females. Males had over-expressed Mrap2 and Oprm1. Mrap2 encodes a protein that modulates melanocortin receptor signalling. Mice deficient in Mrap2 exhibit severe obesity and a mutation in Mrap2 may be associated with severe obesity in human patients [48].

Finally, functional analysis of the 1458 unique DE genes from the Male-Nondiabetic-Obese versus the Female-Nondiabetic-Obese comparison resulted again in enriched immune response related processes (GO:0006955) and positive regulation of inflammatory response (GO:0050729) with related genes also predominantly over-expressed in obese females. The list of prioritised genes consisted of 34 genes from GO and 20 genes from MGI Mammalian Phenotype, with 4 overlapping genes. Genes related to immunological processes include Ifng, Il6, and Tnf. Other enriched processes were muscle system processes (GO:0003012) and abnormal muscle contractility (MP:0005620), abnormal adrenaline level (MP:0003962), abnormal blood pH regulation (MP:0003027), digestion (GO:0007586) and decreased glucagon secretion (MP:0002711). Genes involved in these processes were predominantly under-expressed in male mice, for example, the Mc2r and Pyy. Notably, both Ins1 and Ins2 are also highly over-

expressed in obese female mice compared to almost zero levels of expression in male mice. Mc2r belongs to the MCR family that plays an important role in appetite and energy regulation via leptin signalling. MCRs expression has been found to increase in the liver in response to muscle tissue damage and potentially exert a protective effect through metabolic regulation [49]. We have used the common prioritised genes using both GO and MGI Mammalian Phenotype ontologies to produce two heatmaps of male versus female comparisons which are presented in Figure 2.

## Discussion

Obesity encompasses a complex array of traits closely related to the development of type 2 diabetes, metabolic syndrome and associated with comorbidities such as cardiovascular disease, hypertension, atherosclerosis and various cancers. In this study we have used a systems biology approach to examine genetic susceptibility and hepatic gene response to obesity in CC mice. The CC panel is a unique resource with wide genetic diversity that enables complex trait genomic mapping with very high resolution. Moreover, it provides the opportunity to study differences in genetic expression in a diverse genetic reference population under controlled environmental conditions.

To facilitate this analysis, we developed unified, automated workflow using Docker containers that can be simultaneously deployed on different computation infrastructures and can facilitate cross-platform collaboration by creating an identical insulated and stable operating environment that can be precisely controlled resulting in consistency overtime, without the worry of conflicting dependencies, UNIX compatibility issues and system updates. Docker containers can be used effectively to address some of the major setbacks of microarray and next generation sequencing technologies. The fast start-up time of Docker containers opens up the ability to chain together the execution of multiple containers and the development of sophisticated workflows for genomic analyses.

This work takes into consideration the overall analytical performance of the workflow as well as the computational cost. Our aim was to develop an automated, accurate, and computationally efficient analytical process, while keeping the computational requirements at a minimum. Exploiting the minimal performance loss introduced by the Docker engine we developed a modular architecture for the integrative analysis of QTL and RNA-seq data that can be run on a personal computer (with approximately 4 GB of RAM or higher). For the aforementioned purpose the bioinformatic tools have been picked according to their overall effectiveness and their computational cost as described in the methods. The workflow we propose requires no more than 4.3 GB of RAM, 10 GB of physical drive and can run a top-down analysis of 20 GB of input data in less than 3 hours (these values may change accordingly if multiprocessing or multi-threading are used).

Using our workflow we mapped two QTLs related to the body weight gain phenotype, one on chromosome 5 for male mice and one on chromosome 3 for female mice. The QTL on chromosome 5 has already been described as significant in relation to % of body fat, in response to an atherogenic diet by Su *et al.* [50], who in their study have also suggested Ppargc1a as a candidate gene through

comparative genomics and haplotype analysis. There has also been evidence that suggests Ppargc1a activity may be significantly influenced by sex, although results to that direction have been contradictory with the rs8192678 allele influencing the risk of developing obesity in men but not in women, while in female mice Ppargc1a expression is estrogen regulated and has a protective effect against obesity-induced oxidative damage [51, 52]. In our analysis Ppargc1a is under-expressed in obese versus normal females. Additional genes involved in pathways closely related to the obesity phenotype highlighted by our analysis as potential candidates that play a role in male mice susceptibility to obesity are Cckar, Sod3, Med28, and Slit2.

QTL on chromosome 3 have been previously described to be involved in % of body fat and heat loss, without the introduction of a high fat diet [53, 54]. To our knowledge, this is the first time that the two mapped QTL are described in a sex specific context. It is notable that while the QTL and DE do not overlap significantly, they converge at the functional level, pointing to potentially common regulatory mechanisms. For example, in female mice there is a single differentially expressed gene located inside the QTL on chr3, Cyp2u1, a hydroxylase that is involved in the metabolism of long chain fatty acids. However, the two gene lists derived from the QTL and differential expression analysis share a high number of common relevant enriched pathways, such as increased energy expenditure (MP:0004889), decreased susceptibility to diet-induced obesity (MP:0005659) and increased circulating free fatty acid level (MP:0001554).

Comparison of male and female DE results (presented in supplementary tables 3 and 4) showed that male and female mice have completely different mechanisms of response to high-fat diet and diet induced obesity. Sex differences in glucose metabolism have been previously described in mouse strains. Males have been found to be more prone to developing insulin resistance and obesity than females after being fed a high-fat / high carbohydrate diet. In addition, the same study showed that estrogen contributes to insulin sensitivity in females, and testosterone exacerbates insulin resistance in C57BL/6J mice [55].

In our results, ectopic expression of Pyy and Ins2 in the liver appear to be coping mechanisms utilised by female obese mice but not males. We also observed extrapancreatic expression of Pnlip in female obese mice. A hepatic to pancreatic switch that occurs in a compensatory mode under stress conditions has been previously described [56]. Under-expression of Lepr indicates liver resistance to Leptin, which leads to impaired hepatic insulin sensitivity, regulation of lipid metabolism and glucose homeostasis in female obese mice [57, 58]. Our findings regarding the cases of male non-diabetic obese mice when compared with male non-diabetic normal mice indicate that obesity in the male population heavily affects genes related to the skeletal and cardiac tissues, which is in agreement with previous studies connecting obesity with impaired muscular structure and function, as well as abnormal regulation of insulin secretion and glucose homeostasis [59, 60, 42]. Diet induced obesity has been shown to alter skeletal muscle fiber types in male but not in female mice [61].

A direct comparison of DE genes in female versus male mice that have been subjected to high-fat diet (the complete lists of DE genes of Male-Nondiabetic-Normal vs Female-Nondiabetic-Normal and Male-Nondiabetic-Obese vs Female-Nondiabetic-Obese, are available in Supplementary tables 5 and 6, respectively) confirms that the responses of both resistant and susceptible individuals are different in the two sexes. Male mice appear to undergo and respond to oxidative stress from the high-fat diet by over-expressing antioxidant enzymes such as Gstp1 and Nox4 (Figure 3). They also respond to the metabolic stress they undergo by over-expressing metabolism modulating genes, such as Cav1 (Figure 3). Additionally, as can be seen in Figure 3, female mice over-express genes involved in the regulation of the inflammatory response, such as Il1b, in comparison to male mice. A previous study of the sex dependent inflammatory response of high-fat diet fed mice has shown that male mice tend to develop low-grade systemic inflammation, while female mice are protected through expansion of their population of anti-inflammatory T lymphocytes [62]. As systemic low-grade inflammation caused by obesity is believed to play a role in insulin resistance, it is of great interest to further look into sex dependent inflammatory responses and coping mechanisms.

Our work addresses data integration from analytical platforms at the genomic and transcriptomic level, but we expect that our findings can be extended to accommodate broader integration schemes and result in a variety of computational genomic analysis pipelines and will highlight the potential of the Docker technologies for the development of prototype analytical workflows that suit the individual needs of different research objectives. As a future step in this project, we plan to develop a fully automated, dockerised workflow to perform eQTL analysis on Collaborative Cross mice, which is a significantly more computationally demanding task and will greatly benefit from the possibility of fast and reliable remote deployment.

## Conclusions

We observed that the genetic mechanisms, which underlie susceptibility and response to high-fat diet (HFD) induced obesity differ in female and male mice. This clear distinction in the systemic response to the high-fat diet challenge and to obesity between male and female mice points to the need for further research into distinct sex-related mechanisms in metabolic disease in humans as well.

Moreover, the integration of the data using ontological functional analysis, which showed that different genes but similar pathways are involved in the genetic susceptibility and are disturbed in diet induced obesity, show the need of setting the level of integration of genomic and transcriptomic data at the level of the pathway and not the gene. This way the collective background biological knowledge is meaningfully utilised to systemically interpret results at the genomic scale.

## Abbreviations

CC: Collaborative Cross

GRP: Genetic Reference Panel HFD: High-Fat Diet

RIL: Recombinant Inbred Lines

BMI: Body Mass Index

QTL: Quantitative Trait Loci

CPU: Central Processing Unit

TAU: Tel-Aviv University

SE: Standard Error

IPGTT: intraperitoneal glucose tolerance test

IP: intraperitoneal

RIN: RNA Integrity Number

log<sub>2</sub>FC: log<sub>2</sub> Fold Change

GO: Gene Ontology

MGI: Mouse Genome Informatics

CI: Confidence Interval

MDS: Multidimensional scaling

DE: Differential Expression / Differentially Expressed

GWAS: Genome-wide Association Study

Sgms2: sphingomyelin synthase 2

Hadh: hydroxyacyl-Coenzyme A dehydrogenase

Lef1: lymphoid enhancer binding factor 1

Dkk2: dickkopf WNT signaling pathway inhibitor 2

Egf: epidermal growth factor

Ppargc1a: peroxisome proliferative activated receptor, gamma, coactivator 1 alpha

Cckar: cholecystokinin A receptor

Sod3: superoxide dismutase 3

Med28: mediator complex subunit 28

Slit2: slit homolog 2

Ucp1:

Lepr: leptin receptor

Pnlip: pancreatic lipase

Pyy: peptide YY

Ins1: insulin I

Ins2: insulin II

Nos1: nitric oxide synthase 1

Ryr1: ryanodine receptor 1

Ryr2: ryanodine receptor 2

Des: desmin

Ttn: titin

Cav1: caveolin 1

Egfr: epidermal growth factor receptor

Gstp: glutathione S-transferase

Gcg: glucagon

Nox4: NADPH oxidase 4

NAFLD: non-alcoholic fatty liver disease

Esr1: estrogen receptor 1

Il1b: interleukin 1 beta

Ptgs2: prostaglandin-endoperoxide synthase 2

Tph1: tryptophan hydroxylase 1

Il2: interleukin 2

Il4: interleukin 4

Il6: interleukin 6

Il10: interleukin 10

Il1r: interleukin 1 receptor

Vdr: vitamin D receptor

Mrap2: melanocortin 2 receptor accessory protein 2

Oprm1: opioid receptor, mu 1

Ifng: interferon gamma

Tnf: tumour necrosis factor

Mc2r: melanocortin 2 receptor

MCR: melanocortin receptor

Cyp2u1: cytochrome P450 family 2 subfamily U member 1

## Declarations

### Ethics approval and consent to participate

The Institutional Animal Care approved all experiments in this study and Use Committee (IACUC) at TAU, which adheres to the Israeli guidelines, which follow the NIH/USA animal care and use protocols (approved experiment number M-12-025 and M-14-007).

### Consent for publication

Not applicable

### Availability of data and material

The dataset(s) supporting the conclusions of this article is(are) available in the GEO (and SRA) repositories, under the accession number GSE126490 at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE126490>.

All docker images used in this analysis are freely available on Docker Hub at enios/rnaseq-qt1 repository, <https://hub.docker.com/r/enios/rnaseq-qt1/>. Accompanying Python scripts and documentation are

available on Github at e-nios/RNA-seq-QTL repository, <https://github.com/e-nios/RNA-seq-QTL>.

The genotype data of the CC mouse lines are available on the UNC Systems Genetics website, <http://csbio.unc.edu/CCstatus/index.py>.

### **Competing interests**

Dr. Himmelbauer is an Associate Editor of BMC Genomics, and we would like to ask for excluding him from being associated with reviewing and decision making on this manuscript.

### **Funding**

This work was funded by the Operational Program "Competitiveness, Entrepreneurship and Innovation 2014-2020" (co-funded by the European Regional Development Fund) and managed by the General Secretariat of Research and Technology, Ministry Of Education, Research & Religious Affairs, under the project "Innovative Nanopharmaceuticals: Targeting Breast Cancer Stem Cells by a Novel Combination of Epigenetic and Anticancer Drugs with Gene Therapy (INNOCENT)" (7th Joint Translational Call - 2016, European Innovative Research and Technological Development Projects In Nanomedicine) of the ERA-NET EuroNanoMed II for the development of the bioinformatic analysis pipeline, Hendrech and Eiran Gotwert Fund for studying diabetes, Wellcome Trust grants 085906/Z/08/Z, 075491/Z/04 and 090532/Z/09/Z; for financially supporting the initiating, breeding, genotyping and maintaining the CC mouse colony at TAU, core funding by Tel-Aviv University (TAU) for supporting staff at Iraqi's lab, Israeli Science foundation (ISF) grants 429/09 for performing the mouse assessment for T2D and technical staff, and European Sequencing and Genotyping Infrastructure (ESGI) consortium grant for conducting the RNAseq analysis.

Ilona Binenbaum's PhD thesis was supported by a scholarship from the State Scholarship Foundation in Greece (IKY) (Operational Program "Human Resources Development - Education and Lifelong Learning" Partnership Agreement (PA) 2014-2020).

### **Authors' contributions**

IB was involved with data analysis and writing the MS; HJAA was involved with project design, data collection and analysis, and writing the MS; GF and GK were involved with data analysis and writing the MS; TK and EP were involved with data analysis; RM and HH were involved with project design; FAI is the co-corresponding author, involved with project design, data analysis and writing the MS. AC was involved with data analysis and writing the MS. All authors have read and approved the manuscript.

### **Acknowledgements**

Not applicable

## **References**

1. Churchill, G. A., Airey, D. C., Allayee, H., et al., and Consortium, C. T. (2004). The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nature genetics*, 36, 1133–1137.
2. Mott, R., Talbot, C. J., Turri, M. G., Collins, A. C., and Flint, J. (2000). A method for fine mapping quantitative trait loci in outbred animal stocks. *Proceedings of the National Academy of Sciences*, 97(23), 12649–12654.
3. Valdar, W., Holmes, C. C., Mott, R., and Flint, J. (2009). Mapping in structured populations by resample model averaging. *Genetics*, 182(4), 1263–1277.
4. Fu, C.-P., Welsh, C. E., de Villena, F. P.-M., and McMillan, L. (2012). Inferring ancestry in admixed populations using microarray probe intensities. In *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine - BCB '12*. ACM Press.
5. Felter, W., Ferreira, A., Rajamony, R., and Rubio, J. (2014). An updated performance comparison of virtual machines and linux containers. Technical report, RC25482 (AUS1407-001), IBM Research Division.
6. Tommaso, P. D., Palumbo, E., Chatzou, M., Prieto, P., Heuer, M. L., and Notredame, C. (2015). The impact of Docker containers on the performance of genomic pipelines. *PeerJ*, 3, e1273.
7. Boettiger, C. and Eddelbuettel, D. (2017). An introduction to Rocker: Docker containers for R.
8. Andrews, S. (2010). Fastqc: a quality control tool for high throughput sequence data. available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
9. Williams, C. R., Baccarella, A., Parrish, J. Z., and Kim, C. C. (2016). Trimming of sequence reads alters RNA-seq gene expression estimates. *BMC Bioinformatics*, 17(1).
10. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1), 10.
11. Krueger, F. (2014). Trim Galore: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type libraries. available online at [http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/).
12. Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, 12(4), 357–360.
13. Liao, Y., Smyth, G. K., and Shi, W. (2013). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7), 923–930.
14. Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2009). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139–140.
15. Su, S. (2014). EdgeR.R script available at <https://github.com/galaxyproject/tools-iuc/blob/master/tools/edger/edger.R>.
16. Robinson, M. D. and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 11(3), R25.
17. Koutsandreas, T., Binenbaum, I., Pilalis, E., Valavanis, I., Papadodima, O., and Chatziioannou, A. (2016). Analyzing and visualizing genomic complexity for the derivation of the emergent molecular

- networks. *International Journal of Monitoring and Surveillance Technologies Research*, 4(2), 30–49.
18. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25(1), 25–29.
  19. Smith, C. L. and Eppig, J. T. (2012). The Mammalian Phenotype Ontology as a unifying standard for experimental and high-throughput phenotyping data. *Mammalian Genome*, 23(9-10), 653–668.
  20. Iraqi, F. A., Mahajne, M., Salaymah, Y., Sandovski, H., Tayem, H., Vered, K., Balmer, L., Hall, M., Manship, G., Morahan, G., Pettit, K., Scholten, J., Tweedie, K., Wallace, A., Weerasekera, L., Cleak, J., Durrant, C., Goodstadt, L., Mott, R., and Yalcin, B. (2012). The genome architecture of the Collaborative Cross mouse genetic reference population. *Genetics*, 190(2), 389–401.
  21. Yang, H., Ding, Y., Hutchins, L. N., Szatkiewicz, J., Bell, T. A., Paigen, B. J., Graber, H., de Villena, F. P.-M., and Churchill, G. A. (2009). A customized and versatile high-density genotyping array for the mouse. *Nature Methods*, 6(9), 663–666.
  22. Durrant, C., Tayem, H., Yalcin, B., Cleak, J., Goodstadt, L., de Villena, F. P.-M., Mott, R., and Iraqi, F. A. (2011). Collaborative Cross mice and their power to map host susceptibility to *Aspergillus fumigatus* infection. *Genome Research*, 21(8), 1239–1248.
  23. Gatti, D. (2015). SNPtools: accessing, subsetting and plotting mouse SNPs. <https://rdrr.io/cran/snptools/man/snptools-package.html>.
  24. Link, J. C. and Reue, K. (2017). Genetic basis for sex differences in obesity and lipid metabolism. *Annual Review of Nutrition*, 37(1), 225–245.
  25. Li, Z., Zhang, H., Liu, J., Liang, C.-P., Li, Y., Li, Y., Teitelman, G., Beyer, T., Bui, H. H., Peake, D. A., Zhang, Y., Sanders, P. E., Kuo, M.-S., Park, T.-S., Cao, G., and Jiang, X.-C. (2011). Reducing plasma membrane sphingomyelin increases insulin sensitivity. *Molecular and Cellular Biology*, 31(20), 4205–4218.
  26. Schulz, N., Himmelbauer, H., Rath, M., van Weeghel, M., Houten, S., Kulik, W., Suhre, K., Scherneck, S., Vogel, H., Kluge, R., Wiedmer, P., Joost, H.-G., and Schürmann, A. (2011). Role of medium- and short-chain l-3-hydroxyacyl-CoA dehydrogenase in the regulation of body weight and thermogenesis. *Endocrinology*, 152(12), 4641–4651.
  27. Fuster, J. J., Zuriaga, M. A., Ngo, D. T.-M., Farb, M. G., Aprahamian, T., Yamaguchi, T. P., Gokce, N., and Walsh, K. (2014). Noncanonical Wnt signaling promotes obesity-induced adipose tissue inflammation and metabolic dysfunction independent of adipose tissue expansion. *Diabetes*, 64(4), 1235–1248.
  28. Vandenbeek, R., Khan, N. P., and Estall, J. L. (2017). Linking metabolic disease with the PGC-1 $\alpha$  Gly482Ser polymorphism. *Endocrinology*, 159(2), 853–865.
  29. Funakoshi, A., Miyasaka, K., Jimi, A., Kawanai, T., Takata, Y., and Kono, A. (1994). Little or no expression of the cholecystokinin-a receptor gene in the pancreas of diabetic rats (Otsuka Long-Evans Tokushima Fatty = OLETF rats). *Biochemical and Biophysical Research Communications*, 199(2), 482–488.

30. Cui, R., Gao, M., Qu, S., and Liu, D. (2014). Overexpression of superoxide dismutase 3 gene blocks high-fat diet-induced obesity, fatty liver and insulin resistance. *Gene Therapy*, 21(9), 840–848.
31. Youn, D. Y., Xiaoli, A. M., Pessin, J. E., and Yang, F. (2016). Regulation of metabolism by the Mediator complex. *Biophysics Reports*, 2(2-4), 69–77.
32. Svensson, K. J., Long, J. Z., Jedrychowski, M. P., Cohen, P., Lo, J. C., Serag, S., Kir, S., Shinoda, K., Tartaglia, J. A., Rao, R. R., Chédotal, A., Kajimura, S., Gygi, S. P., and Spiegelman, B. M. (2016). A secreted Slit2 fragment regulates adipose tissue thermogenesis and metabolic function. *Cell Metabolism*, 23(3), 454–466.
33. Weibel, E. K., Hadvary, P., Hochuli, E., Kupfer, E., and Lengsfeld, H. (1987). Lipstatin, an inhibitor of pancreatic lipase, produced by *Streptomyces toxytricini*. Producing organism, fermentation, isolation and biological activity. *The Journal of antibiotics*, 40, 1081–1085.
34. Inagaki, T., Dutchak, P., Zhao, G., Ding, X., Gautron, L., Parameswara, V., Li, Y., Goetz, R., Mohammadi, M., Esser, V., Elmquist, J. K., Gerard, R. D., Burgess, S. C., Hammer, R. E., Mangelsdorf, D. J., and Kliewer, S. A. (2007). Endocrine regulation of the fasting response by PPAR $\alpha$ -mediated induction of fibroblast growth factor 21. *Cell Metabolism*, 5(6), 415–425.
35. Boey, D., Lin, S., Enriquez, R. F., Lee, N. J., Slack, K., Couzens, M., Baldock, P. A., Herzog, H., and Sainsbury, A. (2008). PYY transgenic mice are protected against diet-induced and genetic obesity. *Neuropeptides*, 42(1), 19–30.
36. Kojima, H., Fujimiya, M., Matsumura, K., Nakahara, T., Hara, M., and Chan, L. (2004). Extrapancreatic insulin-producing cells in multiple organs in diabetes. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 2458–2463.
37. Pierobon, N., Renard-Rooney, D. C., Gaspers, L. D., and Thomas, A. P. (2006). Ryanodine receptors in liver. *Journal of Biological Chemistry*, 281(45), 34086– 34095.
38. Hajnóczky, G., Robb-Gaspers, L. D., Seitz, M. B., and Thomas, A. P. (1995). Decoding of cytosolic calcium oscillations in the mitochondria. *Cell*, 82, 415–424.
39. Assimacopoulos-Jeannet, F. D., Blackmore, P. F., and Exton, J. H. (1977). Studies on alpha-adrenergic activation of hepatic glucose output. studies on role of calcium in alpha-adrenergic activation of phosphorylase. *The Journal of biological chemistry*, 252, 2662–2669.
40. Siersbæk, M., Varticovski, L., Yang, S., Baek, S., Nielsen, R., Mandrup, S., Hager, G. L., Chung, J. H., and Grøntved, L. (2017). High fat diet-induced changes of mouse hepatic transcription and enhancer activity can be reversed by subsequent weight loss. *Scientific Reports*, 7, 40220.
41. Ackermann, M. A., Shriver, M., Perry, N. A., Hu, L.-Y. R., and Kontrogianni-Konstantopoulos, A. (2014). Obscurins: Goliaths and dauids take over non-muscle tissues. *PLoS ONE*, 9(2), e88162.
42. Santulli, G., Pagano, G., Sardu, C., Xie, W., Reiken, S., D'Ascia, S. L., Cannone, M., Marziliano, N., Trimarco, B., Guise, T. A., Lacampagne, A., and Marks, A. R. (2015). Calcium release channel RyR2 regulates insulin release and glucose homeostasis. *Journal of Clinical Investigation*, 125(5), 1968–1978.

43. Sansbury, B. E. and Hill, B. G. (2014). Regulation of obesity and insulin resistance by nitric oxide. *Free Radical Biology and Medicine*, 73, 383–399.
44. Fernandez-Rojo, M. A. and Ramm, G. A. (2016). Caveolin-1 function in liver physiology and disease. *Trends in Molecular Medicine*, 22(10), 889–904.
45. Berglund, E. D., Lustig, D. G., Baheza, R. A., Hasenour, C. M., Lee-Young, R. S., Donahue, E. P., Lynes, S. E., Swift, L. L., Charron, M. J., Damon, B. M., and Wasserman, D. H. (2011). Hepatic glucagon action is essential for exercise-induced reversal of mouse fatty liver. *Diabetes*, 60(11), 2720–2729.
46. Hsieh, P.-S., Jin, J.-S., Chiang, C.-F., Chan, P.-C., Chen, C.-H., and Shih, K.-C. (2009). COX mediated inflammation in fat is crucial for obesity-linked insulin resistance and fatty liver. *Obesity*.
47. Crane, J. D., Palanivel, R., Mottillo, E. P., Bujak, A. L., Wang, H., Ford, R. J., Collins, A., Blümer, R. M., Fullerton, M. D., Yabut, J. M., Kim, J. J., Ghia, J.-E., Hamza, S. M., Morrison, K. M., Schertzer, J. D., Dyck, J. R. B., Khan, W. I., and Steinberg, G. R. (2014). Inhibiting peripheral serotonin synthesis reduces obesity and metabolic dysfunction by promoting brown adipose tissue thermogenesis. *Nature Medicine*, 21(2), 166–172.
48. Asai, M., Ramachandrappa, S., Joachim, M., Shen, Y., Zhang, R., Nuthalapati, N., Ramanathan, V., Strohlic, D. E., Ferket, P., Linhart, K., Ho, C., Novoselova, T. V., Garg, S., Ridderstrale, M., Marcus, C., Hirschhorn, J. N., Keogh, J. M., O'Rahilly, S., Chan, L. F., Clark, A. J., Farooqi, I. S., and Majzoub, J. A. (2013). Loss of function of the melanocortin 2 receptor accessory protein 2 is associated with mammalian obesity. *Science*, 341(6143), 275–278.
49. Malik, I. A., Triebel, J., Posselt, J., Khan, S., Ramadori, P., Raddatz, D., and Ramadori, G. (2011). Melanocortin receptors in rat liver cells: change of gene expression and intracellular localization during acute-phase response. *Histochemistry and Cell Biology*, 137(3), 279–291.
50. Su, Z., Korstanje, R., Tsaih, S.-W., and Paigen, B. (2008). Candidate genes for obesity revealed from a C57BL/6J × 129S1/SvImJ intercross. *International Journal of Obesity*, 32(7), 1180–1189.
51. Ridderstråle, M., Johansson, L. E., Rastam, L., and Lindblad, U. (2006). Increased risk of obesity associated with the variant allele of the PPARGC1a Gly482Ser polymorphism in physically inactive elderly men. *Diabetologia*, 49(3), 496–500.
52. Besse-Patin, A., Lèveillé, M., Oropeza, D., Nguyen, B. N., Prat, A., and Estall, J. L. (2017). Estrogen signals through peroxisome proliferator-activated receptor-γ coactivator 1α to reduce oxidative damage associated with diet-induced fatty liver disease. *Gastroenterology*, 152(1), 243–256.
53. Moody, D. E., Pomp, D., Nielsen, M. K., and Van Vleck, L. D. (1999). Identification of quantitative trait loci influencing traits related to energy balance in selection and inbred lines of mice. *Genetics*, 152, 699–711.
54. Brockmann, G. A., Haley, C. S., Renne, U., Knott, S. A., and Schwerin, M. (1998). Quantitative trait loci affecting body weight and fatness from a mouse line selected for extreme high growth. *Genetics*, 150, 369–381.
55. Parks, B. W., Sallam, T., Mehrabian, M., Psychogios, N., Hui, S. T., Norheim, F., Castellani, L. W., Rau, C. D., Pan, C., Phun, J., Zhou, Z., Yang, W.-P., Neuhaus, I., Gargalovic, P. S., Kirchgessner, T. G., Graham,

- M., Lee, R., Tontonoz, P., Gerszten, R. E., Hevener, A. L., and Lusis, A. J. (2015). Genetic architecture of insulin resistance in the mouse. *Cell Metabolism*, 21(2), 334–347.
56. Shanmukhappa, K., Mourya, R., Sabla, G. E., Degen, J. L., and Bezerra, J. A. (2005). Hepatic to pancreatic switch defines a role for hemostatic factors in cellular plasticity in mice. *Proceedings of the National Academy of Sciences*, 102(29), 10182–10187.
57. Amitani, M., Asakawa, A., Amitani, H., and Inui, A. (2013). The role of leptin in the control of insulin-glucose axis. *Frontiers in Neuroscience*, 7.
58. Huynh, F. K., Neumann, U. H., Wang, Y., Rodrigues, B., Kieffer, T. J., and Covey, S. D. (2012). A role for hepatic leptin signaling in lipid metabolism via altered very low density lipoprotein composition and liver lipase activity in mice. *Hepatology*, 57(2), 543–554.
59. Tomlinson, D. J., Erskine, R. M., Morse, C. I., Winwood, K., and Onambélé-Pearson, G. (2015). The impact of obesity on skeletal muscle strength and structure through adolescence to old age. *Biogerontology*, 17(3), 467–483.
60. Choi, K. M. (2016). Sarcopenia and sarcopenic obesity. *The Korean Journal of Internal Medicine*, 31(6), 1054–1060.
61. DeNies, M. S., Johnson, J., Maliphol, A. B., Bruno, M., Kim, A., Rizvi, A., Rustici, K., and Medler, S. (2014). Diet-induced obesity alters skeletal muscle fiber types of male but not female mice. *Physiological Reports*, 2(1), e00204.
62. Pettersson, U. S., Waldén, T. B., Carlsson, P.-O., Jansson, L., and Phillipson, M. (2012). Female mice are protected against high-fat diet induced metabolic syndrome and increase the regulatory T cell population in adipose tissue. *PLoS ONE*, 7(9), e46057.
63. Soller, M. & Iraqi, F. (2014). The Collaborative Cross - A next generation mouse genetic resource population for high resolution genomic analysis of complex traits. *Livest. Sci.* 166, 19-25.

## Table

Due to technical limitations, Table 1 is only available as a download in the supplemental files section.

## Figures



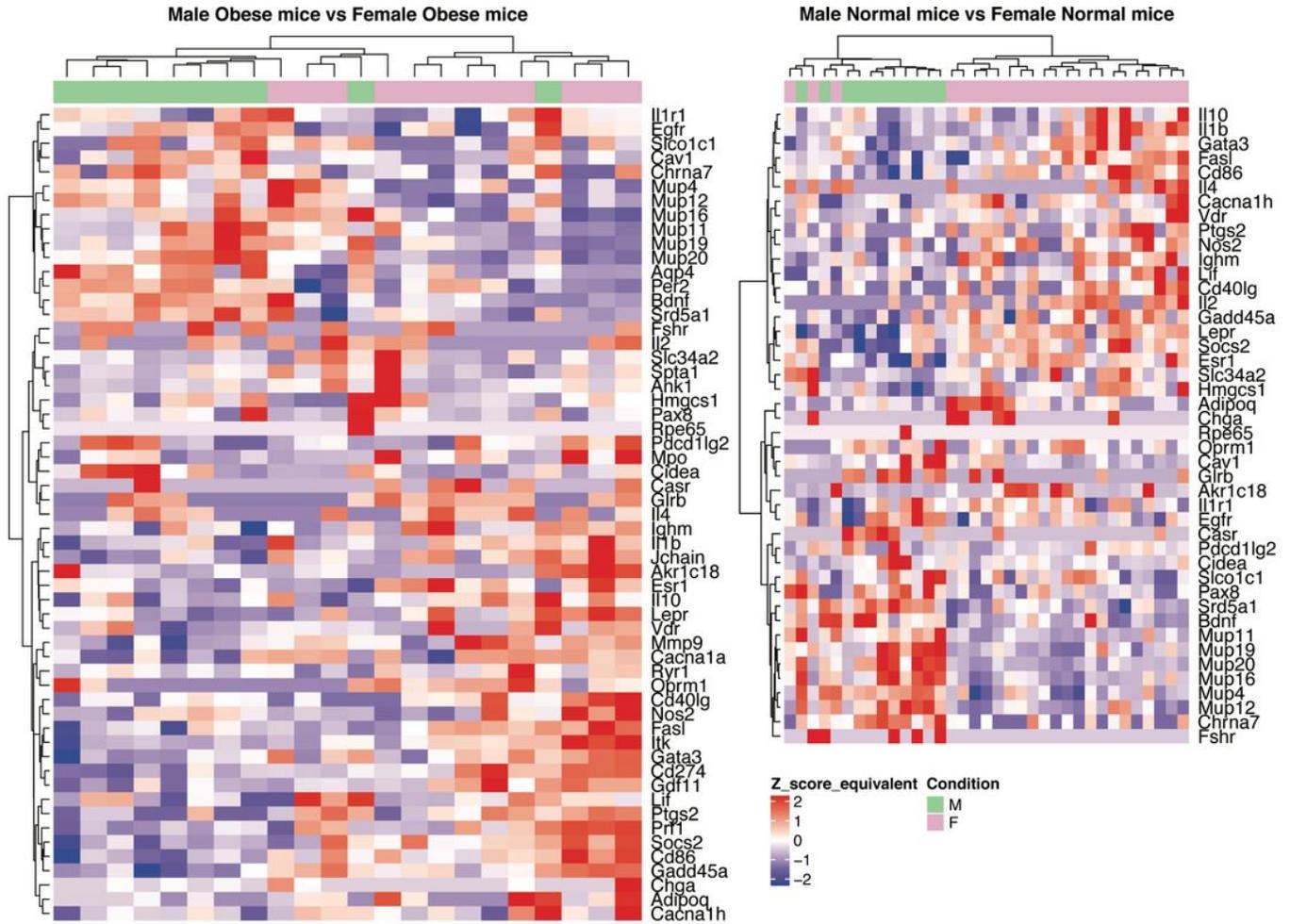


Figure 2

Heatmap graphical representation of Male Obese mice versus Female Obese (right). The genes (y axis) are derived from the union of the prioritised lists produced.

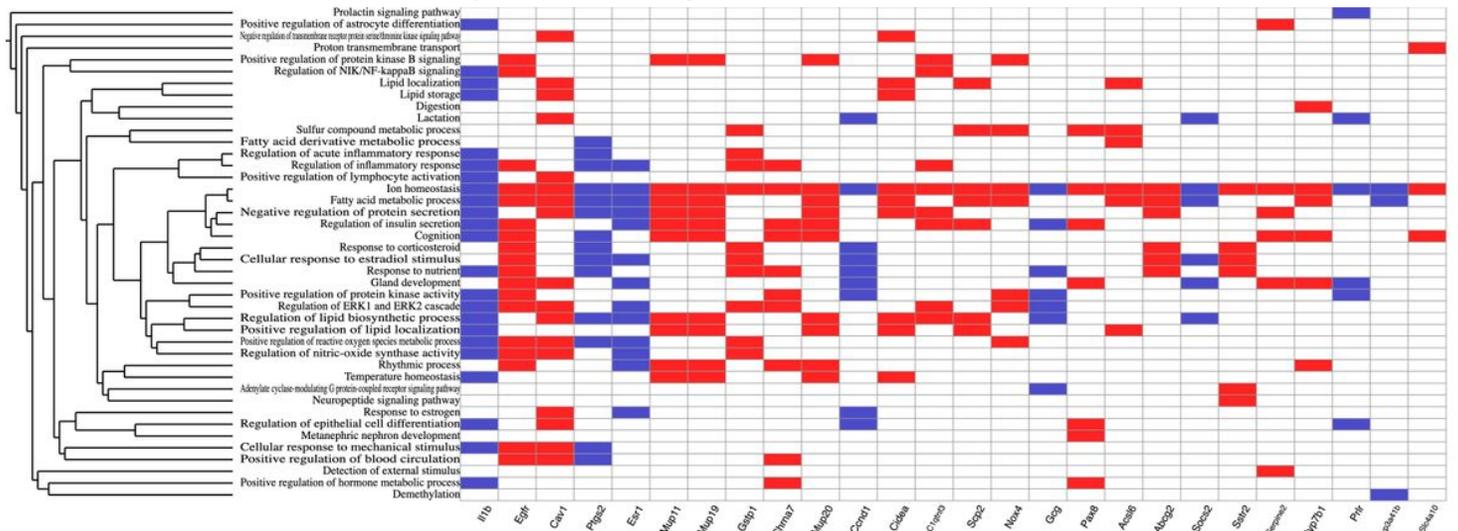


Figure 3

Heatmap graphical representation of the most highly prioritised genes that are common in Male vs Female comparisons and their respective ontologies based on GO produced by BioInfoMiner. Genes marked blue are overexpressed in Female mice while genes marked red are overexpressed in Male mice.

## Supplementary Files

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

- [Table1.png](#)
- [Suptable1.csv](#)
- [Suptable2.csv](#)
- [Suptable4.csv](#)
- [Suptable3.csv](#)
- [Suptable5.csv](#)
- [Suptable6.csv](#)
- [NC3RsARRIVEGuidelinesChecklist.pdf](#)