Multi-omics approaches for better understanding of the downstream effects of genetic risk factors
Zhernakova, Daria

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date: 2016

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
CHAPTER 7

An integrative systems genetics approach reveals potential causal genes and pathways related to obesity
An integrative systems genetics approach reveals potential causal genes and pathways related to obesity

_Lisette J.A. Kogelman¹, Daria V Zhernakova², Harm-Jan Westra³⁴⁵, Susanna Cirera¹, Merete Fredholm¹, Lude Franke², and Haja N. Kadarmideen¹*

1. Department of Veterinary Clinical and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Grønnegårdsvej 7, 1870 Frederiksberg C, Denmark
2. University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands
3. Divisions of Genetics and Rheumatology, Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA.
4. Partners Center for Personalized Genetic Medicine, Boston, MA, USA.
5. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA.

*Corresponding author.
Identification of potential causal genes and pathways related to obesity

Abstract

Background

Obesity is a multi-factorial health problem in which genetic factors play an important role. Limited results have been obtained by single-gene studies, using either genomic or transcriptomic data. RNA-Sequencing technology has shown its potential in gaining deeper and accurate knowledge about the transcriptome, and may reveal novel genes affecting complex diseases. Integration of genomic and transcriptomic variation (eQTL mapping) has identified causal variants that affect complex diseases. We integrated transcriptomic data of adipose tissue and genomic data from a porcine model to gain insight in the mechanisms involved in obesity using a systems genetics approach.

Methods

Based on a selective gene expression profiling approach, we selected 36 animals based on a previously created genomic Obesity Index (OI) for RNA-Sequencing of subcutaneous adipose tissue. Differentially expression (DE) analysis was performed using the OI as a continuous variable in a linear model. Subsequently, eQTL mapping was performed to integrate both 60K Porcine SNP Chip data with the RNA-Sequencing data. Results were restricted based on genome-wide significant SNPs, detected DE genes and previously detected co-expressed gene modules. Further data integration was perceived by detecting co-expression patterns among eQTLs and integration with protein data.

Results

DE analysis of RNA-Sequencing data revealed 458 DE genes. The eQTL mapping resulted in 987 cis-eQTLs and 73 trans-eQTLs (FDR < 0.05), of which the cis-eQTLs were associated with metabolic pathways. We reduced the eQTL search space by focusing on DE- and co-expressed genes and disease-associated SNPs, resulting in the detection of obesity-related genes and pathways. Building a co-expression network using eQTLs resulted in the detection of a module having a strong association with lipid pathways. Furthermore, we detected several obesity candidate genes, e.g. ENPP1, CTSL and ABHD12B.
Conclusions

To our knowledge, this is the first study which performed an integrated genomics and transcriptomics (eQTL) study jointly using and modeling genomic and subcutaneous adipose tissue RNA-Sequencing data on obesity in a porcine model. We detected several pathways and potential causal genes for obesity. Further validation and investigation may reveal their exact function and association with obesity.

Background

Obesity is characterized by an excessive amount of body adipose tissue. As adipose tissue has many endocrine functions, obesity is a very complex condition and is associated with several severe diseases such as Type 2 diabetes, metabolic syndrome and several types of cancer. The prevalence of obesity is exponentially rising world-wide [1] and its enormous consequences for the quality of life and life expectancy have led to the need for a better understanding of the molecular pathology involved. To date, Genome-Wide Association Studies (GWAS) have identified many different loci associated with obesity and obesity-related phenotypes [2, 3]. However, they explain a limited amount of the 40–70% predicted genetic variation [2] and provide limited insight into the biological pathways and molecular mechanisms involved.

Transcriptomic analysis can further elucidate the molecular mechanisms, as gene expression provides a link between genetic variations and their corresponding phenotypic alterations [4]. Commonly, transcriptomic data is used to gain biological insight by detecting differences in expression levels [5] between two conditions (i.e., healthy and diseased). Many studies have applied DE analysis to obesity using different tissues, and have detected many different biologically relevant genes [6-10]. However, all these studies have used a microarray or qPCR platform. The high-throughput RNA-Sequencing (RNA-Seq) technology has demonstrated advantages beyond microarray technology (Wang et al, 2009): the transcriptome is measured more fully and it facilitates the discovery of novel genes. It has also been shown that in DE analysis RNA-Seq data outperforms microarray data.
Identification of potential causal genes and pathways related to obesity in terms of the accuracy of measuring gene expression levels and therefore, can potentially detect more DE genes [11].

Systems genetics approaches via the integration, joint modeling and analyses of various high-throughput -omics data that represent different levels of biological organization, are becoming more popular in genetic studies [12-14]. The integration of genomic and transcriptomic data can be achieved using expression Quantitative Trait Loci (eQTL) studies, whereby genetic variants that underpin differences in expression levels are mapped [15, 13]. It has been shown that eQTLs are highly heritable [16, 17] and have the potential to provide more biological insight into GWAS findings [15]. Several studies have investigated obesity and obesity-related diseases using an eQTL approach, but mainly using microarray expression data [18-20]. Moreover, integration of these data with those from other publically available databases, such as those containing protein–protein interactions (PPI), might also provide further insight into the biological mechanisms behind complex diseases [21, 22].

Here, we present the analysis of RNA-Seq data from subcutaneous adipose tissue from a porcine model specifically created to study obesity. The pig has similar metabolic, physiological and genetic features to humans, that are all more similar to human than rodents, and has shown great potential as a medical model [23, 24]. Extensive discussions on why pigs are better model are given in our paper describing the pig resource population [25]. We previously showed that the F2 population created for obesity studies demonstrates a high heritability for obesity-related phenotypes [25] and we have detected several obesity-related genes and pathways using network approaches on the genotype [26] and RNA-Seq data [27]. Furthermore, we previously created the Obesity Index: an aggregate additive genetic value for obesity, by combining nine different obesity-related phenotypes [26]. Based on a selective expression profiling design, we selected 36 animals in three groups (lean, intermediate and obese) for RNA-Sequencing, and showed previously that those animals show differences in metabolic features [27]. In this study, we performed an integrative systems genetics approach to identify causal genes and associated pathways for obesity. This was achieved by integrating information on genetic variants from GWAS
(based on high-throughput genotype data), with information from co-expression networks (based on RNA-Seq data), and differential gene expression (based on RNA-Seq data), using an eQTL mapping approach [28, 29]. Furthermore, biologically interesting co-expression modules were integrated with the known protein–protein interaction networks to identify key transcriptional and other protein coding factors underlying the causation of obesity. Thus, the integration of multi-omic biological datasets led to the detection of several obesity-related genes and molecular pathways.

### Material and Methods

#### The pig population

The animal resource used in this study was established using breeds that diverged with respect to obesity traits in the parental generation, resulting in a F2 population that was highly divergent with respect to obesity and obesity-related traits. All F2 animals were fed and housed equally. Animal care and maintenance have been conducted according to the Danish “Animal Maintenance Act” (Act 432 dated 09/06/2004) and biological samples were collected according to the Danish “Veterinary Procedures Act” (Act 433 dated 09/06/2004). For a detailed description of the pedigree and the phenotypes, see [25]. In this study, we only used the animals from the Duroc × Göttingen Minipig intercross, which has been previously described [27]. In short, 279 F2 animals resulting from an intercross between the Gottingen minipig (predisposed for obesity) and Duroc production pig (selected for lean meat content), were intensively phenotyped (e.g. weight, conformation traits, Dual-energy X-ray absorptiometry (DXA) scanning, and slaughter characteristics) and genotyped using the Illumina 60k Porcine SNP Chip. We previously

<table>
<thead>
<tr>
<th>Group</th>
<th># Males</th>
<th># Females</th>
<th># unique grandsires</th>
<th>Mean OI</th>
<th>St. dev. OI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lean</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>-2.47</td>
<td>0.75</td>
</tr>
<tr>
<td>Intermediate</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>0.07</td>
<td>0.11</td>
</tr>
<tr>
<td>Obese</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>2.4</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Table 1. Phenotypic characteristics of selected animals.
Identification of potential causal genes and pathways related to obesity

created and described the Obesity Index [26] – a response variable which is an aggregate genotypic value representing the degree of obesity based on the selection index theory [30]. The Obesity Index follows a normal distribution; therefore, animals for RNA-Sequencing of the subcutaneous adipose tissue were selected using a selective gene expression profiling approach [31, 32]. This approach was carefully chosen to ensure sufficient power to detect eQTLs with a relatively small number of animals. In total, we selected 36 animals across three groups: 12 lean, 12 intermediate, and 12 obese animals (Table 1). The family structure (maximizing the number of full siblings to three) and sex distribution was taken into account in the selection of the animals.

**RNA-Sequencing**

RNA-Sequencing (RNA-Seq) was performed as previously described [27]. In short, total RNA was isolated from porcine subcutaneous adipose tissue using the RNeasy Lipid Tissue Mini kit (Qiagen, Germany) following the manufacturer’s recommendations. The RNA quantity and quality were assessed by a Nanodrop ND-1000 spectrophotometer and the integrity of the isolated RNA was visually inspected by gel electrophoresis and by measuring the RNA quality indicator (RQI) value on an Experion™ system (BioRad) using the Eukaryote Total RNA StdSens kit (BioRad). Libraries were subsequently constructed using 400 ng total RNA and a TruSeq RNA Sample Prep (Illumina) with Poly-A pull down rRNA depletion, following the manufacturer’s recommendations. Samples were sequenced on the HiSeq2500 platform, by dividing the 36 samples over four lanes and using 100 bp paired-end reads. Before alignment, the quality of the reads was checked and the adapters were detected using FastQC. The reads were mapped to the genome assembly SSscrofa10.2.72 in STAR aligner using default parameters [33], whereby detected adapters were removed. On average, 20,390 protein-coding genes were detected among the mapped reads. Read counts were estimated using HTSeq [34]. Because transcripts with extremely low expression levels are less reliable [35], transcripts with expression levels equal to or fewer than five counts were removed from the dataset, resulting in 12,253 transcripts per sample. The between-sample bias was removed by estimating the library size factor using the estimateSizeFactor() function in DESeq [36]. Normalization
was then performed using the voom() variance-stabilization function in the R-package Limma [37], and samples were corrected for sex and transformed to log2-counts per million to approach normality.

**Association of gene expression with degree of obesity**

Differentially expressed (DE) genes were detected using a linear model in the R-package Limma [37], taking the obesity index (OI) as continuous variable and sex as factor into account. Firstly, a linear model was fitted for each gene using the function lmFit() in Limma:

\[ y_{ij} = \beta_{j,\text{OI}} \text{OI}_i + \beta_{j,\text{sex}} \text{sex}_i + \epsilon_{ij} \]

where \( y_{ij} \) is the measured expression level of individual \( i \) for gene \( j \), \( \beta_{j,\text{OI}} \) is the estimated regression coefficient from regressing gene expression value of the \( i \)th individual in its Obesity Index (\( \text{OI}_i \)) for the \( j \)th gene, \( \beta_{j,\text{sex}} \) is the estimated regression coefficient of sex, and \( \epsilon_{ij} \) is the error component. Genes were called to be DE-genes in case the \( \beta_{j,\text{OI}} \) was significantly different from 0. Moderated t-statistics (ratio of the log2-fold change to its standard error) and log-odds of differential expression by empirical Bayes moderation of the standard errors towards a common value were then calculated using the function eBayes() in Limma. While moderated t-test statistics account for differences in variance of gene expression of a gene across replicates, the log-odds represents the odds of a gene being DE with an OI different from 0. The resulting p-value obtained from the moderated t-statistic were corrected for multiple testing using the Benjamini-Hochberg procedure [38]. Genes were identified as DE genes when the adjusted p-value (FDR) was below 0.05. The corresponding gene symbols were reported using BioMart [39].

**eQTL mapping**

The integration of the SNP genotype and RNA-Seq data was obtained using an eQTL study approach [28, 12], following the method as described in Westra et al. [29]. The raw expression data obtained by RNA-Seq was quantile-normalized, log2-normalized and centered around zero on a gene level, and finally, a z-transformation was performed on the sample level. Genes were excluded when their expression was below five read counts for any of the samples, because lowly expressed genes potentially introduce a bias. Furthermore, previous studies have shown that the first expression
Identification of potential causal genes and pathways related to obesity

principal components capture experimental variation such as technical and batch effects [40]. Therefore, several eQTL studies were performed on the complete dataset, when different principal components were removed. In total, four principal components were removed, as this resulted in the highest number of detected eQTLs. The SNPs were included in the genotype data when they had a call rate above 0.95, a minor allele frequency (MAF) above 0.05 and were in Hardy-Weinberg equilibrium (P-value > 10^{-4}). The SNPs were mapped onto the genome using genome assembly SSCrofa10.2.74.

Both cis- and trans-eQTL mapping were performed, where eQTLs were considered to be cis-acting when the distance between the gene and SNP was less than 1 Mb and were considered to act in trans when the distance was greater than 1 Mb or when the eQTL was located on another chromosome. The relatively large flanking distance of 1 Mb was chosen, because haplotype blocks in pigs are larger than in humans and are even larger in F2 populations. Other eQTL pig studies adapted an even larger flanking distance, e.g. 10 Mb for assigning cis-eQTLs [41, 18]. To correct for multiple testing, we created a null distribution of p-values by permuting expression phenotypes relative to genotypes 10 times, and then compared the real eQTL p-value distribution to the null distribution. eQTLs were only considered to be significant if the FDR was below 0.05. In case several eQTLs were detected per gene, the strongest effect for each gene was presented.

The eQTL mapping was first performed on all SNPs and genes that exceeded the QC thresholds (52,004 SNPs and 12,253 genes). Furthermore, the analyses were restricted to genes that were differentially expressed in this study (458 genes) and to SNPs that were significantly associated with the Obesity Index, according to our previously published analysis (366 SNPs) [26].

The detected eQTLs were further investigated according to their physical location in the genome and their effects. The distance from the expression SNP (eSNP) to the affected gene was calculated as the distance from the location of the SNP to the location of the Transcription Start Site of the affected gene. Furthermore, the location and the effect of the eSNPs were assigned using the Variant Effect Predictor [42] with the SSCrofa10.2.74 assembly, where the results were restricted to the most severe annotation of the SNP variant.
Integration of eQTLs with co-expression network analysis

We previously conducted a weighted gene co-expression network analysis (WGCNA) on the same RNA-Seq data, and recently published the method and results [27]. To integrate the results from the previous study with those of this study, we evaluated how many eQTLs in the WGCNA modules (clusters of highly interconnected genes) could be detected. We focused solely on the modules that were previously detected as potential biologically interesting: the Blue Module, the Brown Module, the Light-yellow Module, the Black Module and the Green-yellow Module [27]. Because we are interested in potential causal genes within these modules, we extracted only those genes from the modules that were identified as cis-eQTLs.

Supervised gene co-expression network analysis and the integration of protein–protein interactions

All detected genes in the eQTL mapping and DE analysis were used for supervised weighted gene co-expression network analysis (sWGCNA). The network was constructed using the framework of Langfelder and Horvath [43], which is also described in Kogelman et al. (2014a). Briefly, the adjacency matrix was created by calculating the Pearson’s correlations between the selected genes, and was raised by a power $\beta$, to reach a scale-free topology index (R2) of at least 0.90. The Topological Overlap Measure (TOM), for the degree of shared neighbors between pairs of genes, was calculated based on the adjacency matrix and was used as input for the gene dendrogram (1-TOM). Following this step, modules were detected and assigned to a color as branches of the gene dendrogram using the DynamicTreeCut algorithm [44], using a minimum module size of 25 genes per module. The module eigengene, the first principal component of each module, represents the module’s expression and was used to detect biologically relevant modules. The Module–Trait Relationship was calculated as the correlation between the module eigengene and traits of interest, and modules with a significant correlation $>0.5$ were selected for functional annotation. Genes in the module were retained when their intra-modular connectivity was $>0.6$, and when their intra-modular connectivity with other modules was $>0.6$. The module hubgene was detected as the gene in the module with the highest connectivity, or
Identification of potential causal genes and pathways related to obesity

Based on a high intra-modular connectivity (>0.8) and high correlation between the individual gene and the trait of interest (>0.6).

Potentially relevant biological modules were visualized in Cytoscape [45] and were integrated with the known protein–protein interactions from IntAct [46], where corresponding UniProt accession IDs were extracted using BioMart [39]. The networks of the selected module and protein–protein interaction were merged and network analysis within Cytoscape was performed, to gain insight into the network topology. Subsequently, the community clustering algorithm GLay [47] (in the ClusterMaker app) was used to detect clusters in the merged co-expression and protein–protein interaction (PPI) network.

**Functional annotation**

Over-represented Gene Ontology (GO) terms and KEGG pathways among the DE genes were detected using the software GoSeq [48], because GoSeq corrects for length bias in RNA-Seq data. Firstly, a Probability Weighting Function (PWF) for all genes was calculated, based on a given set
of biased data for gene length with the function nullp() in GoSeq. Secondly, a selection-unbiased testing was performed for GO or KEGG enrichment amongst the DE genes. The p-values were corrected for multiple testing using the Benjamini-Hochberg method and GO terms and KEGG pathways were detected as being significantly over-represented using a p-value of 0.05. The over-represented GO terms and KEGG pathways were separately detected for all DE genes, including the up- and downregulated DE genes.

Functional annotation analysis of the detected eQTLs was performed using GeneNetwork (http://www.genenetwork.nl), which detected over-represented GO terms, KEGG pathways, associated phenotypes and tissues [49]. GeneNetwork is based on expression datasets from humans, mice and rats, and predicts functions of genes against known pathways in various biological databases. The over-representation of GO-terms and pathways was tested using the Mann-Whitney U test, and p-values were corrected for multiple testing using the Bonferroni correction.

The complete workflow used in this study is presented in Figure 1.

**Results and discussion**

**Association of gene expression with degree of obesity**

The detection of DE genes can lead to a better understanding of genetic and biological differences between two different conditions and to the detection of predictive biomarkers [50, 51]. In this study, we have used the level of obesity as a continuous variable, whereby we correct for the effect of sex. In total, we found 458 DE genes (FDR < 0.05), with a $\beta_{j,oi}$ ranging from -0.42 to 0.48. All DE results are presented in Additional File 1.

The heatmap of the top 100 DE genes (Figure 2) shows that mostly those from the lean and obese animals cluster together within each group. However, those from intermediate animals cluster with both the obese and lean groups. This is possibly due to the obesity index (OI) values of the intermediate group being at the borderline between the two groups. The genes are partly clustered based on the direction of regulation (up or down). The top 10 DE genes include *TAS1R3, CSGALNACT1, MAML3, ROM1, LRRC16B, EML5, RPS12*
Identification of potential causal genes and pathways related to obesity

(all downregulated in obese animals) and ADAM9 (upregulated in obese animals). We discuss here only selected genes from this group. The TAS1R3 gene encodes a taste receptor, which has been associated with sweetness responsiveness in mice [52]. Taste receptors influence the perception of food [53], and therefore affect eating behavior [54], which might be associated with obesity [55, 56]. The CSGALNACT1 (chondroitin sulfate N-acetylgalactosaminyltransferase 1) gene plays a role in the process of glucuronidation: the addition of glucuronic acid to a substrate. This gene has been associated with Bell’s palsy (dysfunctioning of the facial nerve, resulting in facial paralysis) and Morquio Syndrome b (a metabolic disease, characterized by the inability to breakdown glycosaminoglycans). However, this gene also has an important function in the development of osteoarthritis [57], a bone remodeling disease that has been associated with obesity. The 40S ribosomal protein S12 is encoded by RPS12, which is associated with diabetic nephropathy in African Americans [58]. The final gene, ADAM9, encodes a transmembrane glycoprotein that functions in integrin binding and SH3 domain binding. The ADAM9 gene is involved in the formation of multinuclear cells, and consequently, with osteoclast fusion [59] and because of its downregulation might result in reduced bone mass. Bone mineral density is closely related to obesity: it has been suggested that the increased body load on joints resulting from obesity might stimulate the formation of bone mass, even though data from studies remain controversial [60]. However, body adipose tissue is also closely associated with the immune system and bone remodeling (decreased bone mineral density), by the secretion of adipocytes [61]. In a previous study, we identified several genes that might play a role in these associations [27]. The downregulation of these genes in obese animals might correspond to the decreased bone mass via the secretion of adipocytes, but does not correspond to an increased obesity induced bone mass.

Out of the 458 DE genes, 249 were downregulated and 209 were upregulated in obese animals. Functional annotation analysis was performed for the down- and upregulated genes separately. Surprisingly, almost all over-represented pathways are the result of the presence of upregulated genes. The over-represented pathways are mostly associated with the immune system (e.g., Ribosome, $P = 4.54 \times 10^{-12}$ and leukocyte trans-endothelial migration, $P$
Chapter 7

\[ = 4.32 \times 10^{-4} \], except for starch and sucrose metabolism \((P = 6.67 \times 10^{-4})\) and osteoclast differentiation \((P = 6.67 \times 10^{-4})\). However, osteoclasts are derived from macrophages and therefore, are also related to the immune system [62], which was shown previously [27]. Within the gene ontology (GO) terms, DE genes showed a striking relationship, mainly among the upregulated genes, with the immune system, \(\text{(e.g. immune system process, } P = 2.79 \times 10^{-20} \text{ and immune response, } P = 7.92 \times 10^{-18})\). The downregulated genes were mainly involved with functions related to ribosomes and the translational process \(\text{(e.g. translational termination, } P = 6.56 \times 10^{-17} \text{ and cytosolic ribosome, } P = 1.33 \times 10^{-15})\), which cannot be linked directly to obesity itself.

**The location and annotation of eQTLs**

The eQTL mapping led to the detection of 1,070 eQTLs: 987 cis-eQTLs

![Figure 2. Heatmap of the top 100 differentially expressed genes.](image)
and 73 trans-eQTLs (FDR < 0.05). All detected eQTLs are presented in Additional File 2. SNPs used for eQTL mapping were filtered based on a high minor allele frequency (MAF < 0.05). To ensure that this has not affected results due to the low number of animals, we have investigated the MAF of the eSNPs. From those results (not presented) it was evident that most SNPs that were detected as eQTL had a high MAF. More specifically, only 14 SNPs of the detected eQTLs had a MAF between 0.05 and 0.10. Based on those results and the statistical models used to estimate the eQTL effects, followed by appropriate significance testing, we believe that the results are reliable.

The detected eQTLs were annotated as cis, when the distance between the eSNP and gene was within 1Mb. This is confirmed by the results, which show that most of the cis-eQTLs are located close to the Transcription Start Site (TSS), as expected, because this is the region where most transcriptional regulation occurs (Figure 3a). Further investigation of the effect of the eQTLs (Figure 3b) showed that most were located in intergenic (46%) and intronic regions (37%). In comparison with the complete set of SNPs that passed quality control, we found that the eQTLs were located more often in intronic regions (>8%) and less often in intergenic regions (<10%); the frequencies in all other regions were not altered. These percentages agree with the results of previous studies, which showed that disease-associated SNPs are mainly located within non-coding regions (~90%) [63-65]. We found comparable frequencies of intergenic and intronic SNPs between cis- and trans-eQTLs, where SNPs from cis-eQTLS were also located in exonic regions (3’UTR, 5’UTR, mis-sense, and splice regions). These intergenic and intronic SNPs might be in linkage disequilibrium with a causative SNP, but might also have a regulatory function on the gene expression of the eQTL.

Of the detected trans-eQTLs, 10 SNPs were located on a different chromosome than the target gene. We further investigated the distance between the SNP and target gene of the trans-eQTLs that were located on similar chromosome, showing that a large part of the trans-eQTLs would be detected as cis by increasing the window size several Mb’s (Figure 4a). Furthermore, we investigated the number of target genes per detected trans-eQTL and showed that most SNPs were targeting up to five genes (Figure 4b).
Figure 3. **a)** Histogram depicting the distance from the eSNP to the affected transcript of all **cis**-eQTLs, with the Transcription Start Site as a base (distance = 0 kb); **b)** Visualization of the variant effects of the eSNPs (both **cis** and **trans**) as a percentage.
Identification of potential causal genes and pathways related to obesity

Functional annotation of eQTLs

Functional annotation analysis using all detected cis-eQTLs (987) revealed many metabolism-related GO terms and pathways (Table 2). We also found that several adipose tissues (e.g. adipocytes and subcutaneous adipose tissue) were highly significantly associated with the cis-eQTLs.

In order to identify potential causal genes, we limited the number of eQTL association tests, by confining the eQTL mapping to DE genes (458 genes) and secondly, to obesity-associated SNPs resulting from the previously conducted GWAS on the Obesity Index (366 SNPs) [26]. Using the restriction of DE genes, we found a total of 36 eQTLs, among which different GO terms and pathways related to cholesterol transport and other lipid process were represented, e.g. protein–lipid complex (P = 2 x 10^{-4}) and lipid digestion, mobilization, and transport (P = 2 x 10^{-5}). Restriction to GWAS SNPs resulted in the detection of 24 eQTLs. Functional annotation showed that these genes were mainly expressed in adipose tissue (subcutaneous fat, abdominal; P = 4 x 10^{-3}) and associated phenotypes also showed a link with obesity-related characteristics (e.g. abnormal triglyceride level, P = 1 x 10^{-3}). Other GO terms and pathways were related to transcription (e.g. viral transcription, P = 3 x 10^{-4}), metabolism (tyrosine metabolism, P = 1 x 10^{-3}), or immunity (e.g. influenza infection, P = 2 x 10^{-4}). Of those 24 eQTLs, the expression of one target gene was significantly associated (P < 0.05) with the OI: C15orf26. However, this gene does not seem to have any previously discovered association with obesity or obesity-related diseases. Two other genes tend to be significant: RAB11A and USP36 (P < 0.1) Previ-
### Table 2. GeneNetwork results over all cis-eQTLs.

<table>
<thead>
<tr>
<th>Pathway or process</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO: BP</td>
<td></td>
</tr>
<tr>
<td>co-factor biosynthetic process</td>
<td>$1 \times 10^{-18}$</td>
</tr>
<tr>
<td>co-factor metabolic process</td>
<td>$6 \times 10^{-18}$</td>
</tr>
<tr>
<td>water-soluble vitamin metabolic process</td>
<td>$3 \times 10^{-17}$</td>
</tr>
<tr>
<td>GO: CC</td>
<td></td>
</tr>
<tr>
<td>mitochondrial matrix</td>
<td>$5 \times 10^{-13}$</td>
</tr>
<tr>
<td>organelle outer membrane</td>
<td>$1 \times 10^{-12}$</td>
</tr>
<tr>
<td>outer membrane</td>
<td>$2 \times 10^{-12}$</td>
</tr>
<tr>
<td>GO: MF</td>
<td></td>
</tr>
<tr>
<td>transferase activity, transferring alkyl or aryl (other than methyl) groups</td>
<td>$2 \times 10^{-17}$</td>
</tr>
<tr>
<td>co-enzyme binding</td>
<td>$3 \times 10^{-14}$</td>
</tr>
<tr>
<td>co-factor binding</td>
<td>$3 \times 10^{-13}$</td>
</tr>
<tr>
<td>KEGG</td>
<td></td>
</tr>
<tr>
<td>amino sugar and nucleotide sugar metabolism</td>
<td>$1 \times 10^{-12}$</td>
</tr>
<tr>
<td>pyrimidine metabolism</td>
<td>$4 \times 10^{-10}$</td>
</tr>
<tr>
<td>glutathione metabolism</td>
<td>$4 \times 10^{-9}$</td>
</tr>
<tr>
<td>Reactome</td>
<td></td>
</tr>
<tr>
<td>metabolism of water-soluble vitamins and co-factors</td>
<td>$3 \times 10^{-16}$</td>
</tr>
<tr>
<td>metabolism of vitamins and co-factors</td>
<td>$3 \times 10^{-16}$</td>
</tr>
<tr>
<td>mitochondrial tRNA amino-acylation</td>
<td>$6 \times 10^{-16}$</td>
</tr>
<tr>
<td>MGI</td>
<td></td>
</tr>
<tr>
<td>abnormal inhibitory postsynaptic currents</td>
<td>$3 \times 10^{-11}$</td>
</tr>
<tr>
<td>abnormal cell morphology</td>
<td>$7 \times 10^{-11}$</td>
</tr>
<tr>
<td>abnormal circulating amino acid level</td>
<td>$9 \times 10^{-10}$</td>
</tr>
<tr>
<td>Tissues and cells</td>
<td></td>
</tr>
<tr>
<td>Adipocytes</td>
<td>$1 \times 10^{-48}$</td>
</tr>
<tr>
<td>Choroid</td>
<td>$3 \times 10^{-48}$</td>
</tr>
<tr>
<td>aortic valve</td>
<td>$8 \times 10^{-47}$</td>
</tr>
<tr>
<td>subcutaneous fat</td>
<td>$10 \times 10^{-46}$</td>
</tr>
<tr>
<td>subcutaneous fat, abdominal</td>
<td>$2 \times 10^{-45}$</td>
</tr>
</tbody>
</table>
Identification of potential causal genes and pathways related to obesity

ously, *RAB11A* has been showed to be an element in the GLUT4 trafficking machinery [66] and has been associated with glucose metabolism [67]. To our knowledge, *USP36* has not been associated before with obesity or obesity-related diseases. Unfortunately, we did not identify any overlap of eQTLs between the restricted analyses of the DE-subset and GWAS-subset.

From the 36 eQTLs detected among the subset of DE genes, 35 were also present in the eQTLs using no restrictions. One of these eQTLs involves the *ENPP1* (ectonucleotide pyrophosphatase/phosphodiesterase 1) gene (Figure 5a) that encodes a type II transmembrane glycoprotein (*P* = 1.43 x 10^-4). This glycoprotein plays a role in the regulation of pyrophosphate levels, and several other functions are known: e.g., bone mineralization and the regulation of purinergic signaling. More importantly, this gene is highly expressed in adipose tissue and has been associated with obesity, type-2 diabetes, and insulin resistance [68, 69] and also with eating behavior in a pig population [56]. Moreover, expression of this gene inhibits insulin-signaling via reduced insulin-receptor tyrosine kinase activity [70]. The data here showed a large difference between the AA and GG phenotypes, whereby the AA animals were lean and GG animals (with an increased expression of *ENPP1*) were obese, which agrees with data in the literature. For example, the AA animals (*n* = 8) weighed 0.695 kg (SD = 0.13) at birth and showed a mean gain of 0.37 kg/day (SD = 0.10), compared with GG animals (*n* = 10), which weighed 0.923 kg (SD = 0.12) at birth and gained 0.535 kg/day (SD = 0.12). At slaughter, the AA animals contained 24.88 mm of backfat (SD = 7.59) whereas the GG animals had 39.25 mm of backfat (SD = 7.19). However, we found no difference in the fasting glucose levels between the different genotypes.

Another eQTL that has been associated with obesity and insulin resistance is *CTSL* (Cathepsin L), a lysosomal cysteine proteinase (Figure 5b, *P* = 5.55 x 10^-6). Several studies have investigated the role of *CTSL* and have shown, for example, that inhibition of *CTSL* results in limited adipogenesis or lipid accumulation [71] by reducing the levels of pivotal transcriptional mediators of adipogenesis. Moreover, the pharmacological inhibition of *CTSL* resulted in reduced body weight gain and levels of *CTSL* were elevated in obese and diabetic patients [72]. Our results confirm these findings;
Figure 4. Barplots of the selected eQTLs with one selected phenotype per eQTL (with p-value representing the level of significance for difference in phenotype between the two homozygous genotypes), where a, b and c are cis-acting eQTLs and d represents a trans-acting eQTL. DXA fat is fat estimated using the dual-energy X-ray absorptiometry (DXA) scan. Genes that are marked with * are significant DE genes and the SNP that is marked with ** is genome-wide significantly associated with obesity.
Identification of potential causal genes and pathways related to obesity

for example, the AA animals (n = 11) weighed 0.93 kg (SD = 0.16) at birth and showed a mean weight gain of 0.44 kg/day (SD = 0.10), whereas the GG animals (n = 2) weighed 0.64 kg (SD = 0.15) at birth and gained 0.32 kg/day (SD = 0.04). Furthermore, the AA animals contained 2,730.30 g (SD = 898.83) of fat at dual-energy X-ray absorptiometry (DXA) scanning and 2.48 kg (SD = 1.30) of leaf fat at slaughter, whereas the GG animals had 1,602.00 g (SD = 694.84) of fat at DXA scanning and 2.20 kg (SD = 1.63) of leaf fat at slaughter. The gene function prediction of GeneNetwork also showed a clear role of CTSL in the regulation of cholesterol and lipids, e.g. regulation of plasma lipoprotein particle levels (P = 2.98 x 10^{-16}) and lipid storage (P = 3.8 x 10^{-14}).

Another detected eQTL was CIDE-C (cell death-inducing DFFA-like effector c), also called Fat Specific Protein 27 (FSP27), which was more highly expressed in the TT genotype than the CC genotype (Figure 5c, P = 2.23 x 10^{-4}). This gene promotes triglyceride (lipid droplet) formation and has a negative regulatory effect on adipocyte apoptosis [73, 74]. A CIDE-C knockout model in mice resulted in smaller lipid droplets [75]. Furthermore, CIDE-C is regulated by insulin via the Akt1/2- and JNK2-dependent pathways in adipocytes [76]. Animals with the TT genotype (n = 8) were heavier (e.g. 8.29 kg vs. 17.91 kg at DXA scanning) and showed a higher mean daily gain (0.31 vs. 0.51 kg/day) than CC animals (n = 8) and had a considerably higher amount of fat than in CC animals: 1,562.00 g vs. 3,203.51 g estimated by DXA scanning and 2.01 kg vs. 3.01 kg at slaughter (weight of leaf fat). These results differ from findings in other studies, where CIDE-C mRNA levels were lower in obese subjects and were negatively correlated with BMI and percentage fat mass, but increased in obese patients after weight loss [77]. However, the Gene Expression Omnibus (GEO) database also contained studies that showed a lower CIDE-C expression for high weight gainers vs. low weight gainers (Accession number: GDS2319) and a higher expression in the adipose stem cells of morbidly obese individuals vs. non-obese individuals (Accession number: GDS5056). GeneNetwork identified many adipose-related GO terms and pathways for CIDE-C using the predicted function, e.g. the GO Biological Process triglyceride metabolic process (P = 1.29 x 10^{-76}) and GO Cellular Component lipid particle (P = 1.27 x 10^{-88}).
In addition to the cis-eQTLs, we detected several (73) trans-eQTLs and the functional annotation of this group of genes using GeneNetwork resulted in the detection of (subcutaneous) adipose tissue as associated over-represented tissue ($P = 6 \times 10^{-4}$). Furthermore, a wide variety of significant GO terms and pathways were detected, which were not all directly linked to obesity, e.g. excitatory synapse ($P = 10 \times 10^{-5}$). In general, trans-eQTLs provide a fundamental understanding of potential gene-to-gene regulatory architecture of complex traits and diseases and can also be used to predict transcription factor binding sites [78]. For the trans-eQTLs, genes and SNPs were restricted to DE genes and GWAS SNPs. Among the trans-eQTLs, only two genes overlapped between all trans-eQTLs and those among the DE genes (GFRα3 and MYH3), and only one gene overlapped between all trans-eQTLs and the trans-eQTLs among the GWAS-significant SNPs (ABHD12B). The GFRα3 gene encodes the artemin receptor, which is a neurophin with various functions, e.g. nerve regeneration and tumor-cell migration [79], although no direct link has previously been found between either GFRα3 or MYH3 (which encodes myosin, heavy chain 3, skeletal muscle, embryonic) and obesity. Myosin converts chemical energy into mechanical energy via ATP hydrolysis. Growth characteristics in cattle and pigs have been associated with MYH3 expression, in addition to a difference in muscle growth between and lean and obese pigs, which suggests an association between MYH3 and adiposity. However, we observed no difference in obesity-related phenotypes according to MYH3 expression. The third gene, ABHD12B (α/β-hydrolase domain containing 12B), has also not been previously associated with obesity (Figure 5d, $P = 2.36 \times 10^{-9}$). However, it plays a role in lysophosphatidylserine (LPS) metabolism and ABHD12B knockout mice have a deregulated accumulation of proinflammatory lipids [80]. Furthermore, it has been shown that LPS stimulates glucose transport in adipocytes [81]. In this study, animals with the AA genotype on ALGA0006476 ($n = 12$) were more obese than animals with the CC genotype ($n = 8$) and showed a mean daily gain of 0.47 kg/day (SD = 0.11) and 0.38 kg/day (SD = 0.12) and a weight of leaf fat at slaughter of 3.06 kg (SD = 1.36) and 1.81 kg (SD = 1.23), respectively. The expression of ABHD12B was higher in CC, suggesting that upregulation/activation of
Identification of potential causal genes and pathways related to obesity

this gene results in leaner animals. To our knowledge, no other studies have shown significant effects for the expression of \textit{ABHD12B}.

\textbf{Integration of eQTL results with gene co-expression network analysis}

Previously, we investigated the RNA-Seq data from this study using a gene co-expression network approach (Weighted Gene Co-expression Network Analysis, WGCNA) \cite{27} and we detected five modules that were potentially biologically associated with obesity-related characteristics. We hypothesized that modules containing more eQTL genes than expected by chance would pinpoint modules that are more likely to be causal for the trait under study. Therefore, we investigated how many eQTLs (out of the 987 detected \textit{cis}-eQTLs) were present in each of the five modules detected using WGCNA. We found five eQTLs in the Green-yellow module (47 genes), five eQTLs in the Brown Module (86 genes), two eQTLs in the Blue module (69 genes) and one eQTL in the Black module (36 genes) and no eQTLs in the Light-yellow module. This represents 10.64\%, 5.81\%, 2.90\%, 2.78\% and 0\% of the number of genes in that particular module, respectively, which is unfortunately not higher than expected by chance (hypergeometric test). The eQTL in the Black module is a novel gene (uncharacterized protein), with no known orthologues.

The Green-yellow module was strongly associated with obesity-related phenotypes, but in our previous study, the functional annotation did not identify a relationship with obesity. We now found five eQTLs in this module: \textit{ALDH1L2}, \textit{GGTA1}, \textit{KRR1}, \textit{ME3}, and \textit{OPTN}. The \textit{OPTN} gene encodes optineurin, a protein that has been investigated intensively in the neuroscience field, and is associated with primary open-angle glaucoma and amyotrophic lateral sclerosis \cite{82}. Notably, it also plays a role in adipogenesis, and modulates the developmental switch into brown preadipocytes \cite{83} GeneNetwork predicts the adipocytokine signaling pathway ($P = 1.5 \times 10^{-7}$) as the most likely associated KEGG pathway. For the other genes, no obvious association with obesity or obesity-related phenotypes was found.

In the Brown module, we found five eQTLs, representing four unique genes (two eQTLs were located in one gene: \textit{RIN2}): \textit{ARF6}, \textit{PMVK}, \textit{MSRB2},
and RIN2. The ARF6 gene encodes a small GTP-binding protein that regulates vesicular trafficking actin cytoskeletal dynamics [84]. The PMVK gene encodes an enzyme that functions in the cholesterol biosynthesis pathway, which converts mevalonic acid-5P to mevalonic acid 5-pyrophosphate. Furthermore, it has been shown to be critical in the regulation of the secretion of insulin in pancreatic β cells [85]. The other genes were not related to obesity or obesity-related phenotypes.

In our previous study, the Blue module revealed a potential genetic association between obesity, the immune system, and bone remodeling (osteoporosis). Therefore, we would expect that the eQTLs in this module (LAT2 and IGSF6) have a more causal role in this genetic association. Both genes play a role in the immune system, but have not been previously shown to be directly related to obesity.

**Supervised gene co-expression network and integration with protein–protein interactions**

Both DE genes and cis-eQTL genes were used as input for supervised Weighted Gene Co-expression Network Analysis (sWGCNA), with the aim of focusing on potential causal genes for obesity. The Pearson’s correlations among 1,408 unique genes (987 cis-eQTLs and 458 DE genes) were calculated and raised to a power $\beta$, of three, reaching a scale-free topology index ($R^2$) of 0.93. Using the topological overlap measure (TOM), we detected eight modules of at least 25 genes per module, three of which showed strong significant correlations with the Obesity Index (MTR$_{OI}$) and other obesity-related traits: the YellowsWGCNA Module (MTR$_{OI}$ = -0.74), BluesWGCNA Module (MTR$_{OI}$ = -0.71), and TurquoisesWGCNA Module (MTR$_{OI}$ = 0.69). The functional annotation of those modules showed only strongly significant GO-terms and pathways for the TurquoisesWGCNA module, which were related to obesity and the biological processes lipid localization ($P_{adj} = 5.35 \times 10^{-12}$) and lipid transport ($P_{adj} = 2.39 \times 10^{-10}$) were most significantly over-represented. Based solely on the connectivity of the genes in this module, ITGB2 (β 2-Integrin) is the hubgene (highest connectivity) of this module, however, it is not included in the module after gene selection based on the gene-trait correlation (correlation of 0.59 with the OI). According to the gene network
Identification of potential causal genes and pathways related to obesity

prediction of GeneNetwork, ITGB2 is co-expressed with many other genes, which all possess functions within the immune system (e.g., activation of immune response). This gene has also previously been associated with obesity: mutations in ITGB2 have been associated with obesity in mice [86] and fat oxidation and insulin metabolism were impaired in knockout mice [87]. The ITGB2 gene has also been associated with obesity in humans: a polymorphism in this gene is associated with obesity among Japanese individuals living in the US (westernized diet) [88]. Based on other measures than only connectivity, such as intra-modular connectivity and gene-trait correlation, other important genes were detected within this module: NCKAP1L, S100A10, VSIG4 and CD68, which have all been associated with immune-related processes. The NCKAP1L gene is also strongly co-expressed with many other genes but is not expressed in adipose tissue, and functions in immune-related processes. S100A10 encodes the protein p11, which functions in cellular processes such as exo- and endocytosis, and has been associated with serotonergic signaling and consequently, in depression [89]. VSIG4 encodes a B7 family-related protein and negatively regulates T cell activation and IL-2 production [90]. CD68 encodes a glycoprotein that is mainly expressed by monocytes and tissue macrophages [91] and binds to oxidized LDL on the cell surface and might consequently play a role in atherosclerotic lesions [92]. We further analyzed and visualized this module in Cytoscape and merged the turquoise network with the known protein–protein interactions from IntAct. The resulting network consisted of 419 nodes and 3,015 edges (Figure 6a). We calculated the network statistics and clustered the genes/proteins within Cytoscape (Figure 6b). The largest cluster contained 137 proteins that interacted with a single gene: CALCOCO2 (Calcium binding and coiled–coil domain 2). This gene is significantly up-regulated in adipose tissue (P = 0.0003) according to HumanMine (http://humanmine.org) and has been shown to be negatively correlated with the level of triglycerides in muscle tissue in a mouse model for human obesity [93]. Two out of the four genes that were detected as potentially important/hub genes (VSIG4 and CD68), were located within the second largest module. The third gene (NCKAP1L) formed a cluster together with five other genes. The fourth gene (S100A10), formed a cluster by itself, with ten other proteins.
Figure 5. Visualization of the Turquoises WGCNA Module in Cytoscape; a) The complete network and b) Clustering of the network based on GLay community clustering algorithm. Genes in the Turquoise module are colored turquoise and interacting proteins are colored orange. The size of the nodes is dependent on the Betweenness Centrality of the node. Edges are colored green to represent a gene–gene interaction (the darkness depends on the weight of the correlation) and gray represents protein–protein and gene–protein interactions.
Identification of potential causal genes and pathways related to obesity

Conclusion

In this study, we examined the transcriptome and genome of 36 lean, intermediate and obese pigs using a variety of multi-omic systems genetics approaches, with the aim of detecting potential causal genes and regulatory networks for human obesity using the pig as a model. We performed differential expression analyses, weighted gene co-expression network analyses and integrative systems genetics analyses of obesity by integrating and jointly analyzing the genome and transcriptome using an eQTL approach. Furthermore, we generated networks using the identified eQTLs, to provide causal networks and to identify more biologically relevant causal genes.

We successfully identified many differentially expressed genes and the associated pathways showed several immune-related pathways and gene ontology terms, mainly among the upregulated genes. Furthermore, we conducted eQTL mapping, a systems genetics approach, to detect which genetic variants affect the expression levels of obesity-related genes. We detected many cis- and trans-acting eQTLs, mostly located in intronic and intragenic regions, which were further analyzed by pathway analyses and we detected many different metabolic pathways using GeneNetwork. To limit our eQTL search to the most promising and potential causal genes, we focused on and restricted to DE genes and to SNPs that exceeded the genome-wide significance threshold in a GWA study with the Obesity Index scores. Furthermore, we evaluated how many of the detected eQTLs were present in clusters of highly interconnected genes, which were detected in our previous study, because those clusters containing many eQTLs might represent a causal function of the module leading to obesity. The restriction of the data in these ways led to a subset of eQTLs that were further analyzed using gene ontology and pathway analysis, which resulted in several adiposity-related terms and pathways. The detected trans-eQTLs could not be directly linked to obesity, but provided insights into complex trans-regulatory mechanisms. Finally, we performed a supervised WGCNA (sWGCNA) on all detected cis-eQTLs and identified several modules that highly correlated with obesity-phenotypes. One of these modules showed a strong association with lipid pathways. However, integration with known protein–protein interactions from a publically available database did not provide further
insight into important underlying mechanisms.

For years, DNA markers have been studied in association with complex traits, for example obesity, which has led to the detection of several associated genes. Similarly, gene expression has been studied in detail to detect associated genes. However, the combination of DNA markers and gene expression data leads to a better understanding of the mechanisms behind the translation from DNA marker via transcription towards complex disease, and therefore targets more potentially causal genes. In this study, we detected several eQTLs that revealed genes that may be potentially causal genes for obesity, due to the combined association of DNA marker and transcription with obesity. These genes have been previously associated with obesity-related traits, but have not all been associated with obesity. In this study, we identified for example, the genes ENPP1, CTSL, CIDE-C and ABHD12B as potential causal genes of obesity, and the further validation (e.g. by qPCR in a large human population) and investigation of these genes might lead to biomarkers for obesity. However, in our study we selected the strongest eQTL effect per DNA marker/gene target, but due to LD this might not always be the true causal gene. Therefore, other strategies are needed to prove causality of our detected potentially causal genes, for which several integrative approaches are available, e.g. proposed by Schadt et al. [94].

In conclusion, this systems genetics study (integrating RNA-Seq transcriptomic and genomic information) revealed potential causal genes, as well as the genetic and regulatory architecture of obesity pathways using a porcine model. Furthermore, systems biological mechanisms of obesity, including several relevant gene ontology terms and molecular pathways related to obesity are presented here. To the best of our knowledge, this is the first study to report integrated transcriptomic and genomics data in a porcine model for obesity.

Availability of supporting data

The RNA-Seq expression data in this publication has been deposited in NCBI’s Gene Expression Omnibus and is available through accession number GSE61271.
Identification of potential causal genes and pathways related to obesity

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

HNK was the overall project leader who conceived this systems genetics study and supervised LJAK in the genetics, bioinformatics and systems biology analyses. LJAK analyzed all the data and wrote the first draft of the manuscript. MF established the pig population used in the study, collected and provided the phenotypic measurements and adipose tissue for RNA Sequencing. SC performed the RNA isolation, and quality control of RNA samples before RNA Sequencing. The initial eQTL mapping was carried out at the University Medical Centre Groningen (The Netherlands) where DVZ and HJW helped LJAK with the analysis, under the supervision of LF. Later eQTL analyses were extended to include gene networks at the University of Copenhagen by LJAK under the supervision of HNK. All authors wrote, read, and approved the final version of the manuscript.

Additional data

The following additional data are available with the online version of this paper.

Additional data file 1 is a table listing the differentially expressed genes between the different obesity levels.

Additional data file 2 contains the detected cis- and trans-eQTLs in the complete dataset.

Acknowledgements

The project is supported by a grant (Nr. 0603-00457B) from the Danish Council for Strategic Research (BioChild Project: www.biochild.ku.dk) and from the EU-FP7 Marie Curie Actions – Career Integration Grant (CIG-293511), both granted to Haja N. Kadarmideen, from a Ph.D. stipend awarded to Lisette J.A. Kogelman from the University of Copenhagen, and
from the Danish Ministry of Science and Technology (the “UNIK Project for Food Fitness and Pharma for Health”) to Merete Fredholm.

References


12. Kadarmideen H, von Rohr P, Janss L. From genetical genomics to sys-
Identification of potential causal genes and pathways related to obesity

tems genetics: potential applications in quantitative genomics and animal breeding. Mammalian Genome. 2006;17(6):548-64. doi:10.1007/s00335-005-0169-x.
27. Kogelman LJA, Cirera S, Zhernakova D, Fredholm M, Franke L, Kadars-


41. Ponsuksili S, Du Y, Murani E, Schwerin M, Wimmers K. Elucidating Molecular Networks That Either Affect or Respond to Plasma Cortisol Concentration in
Identification of potential causal genes and pathways related to obesity


56. Do DN, Strathe AB, Ostersen T, Jensen J, Mark T, Kadarmideen HN. Genome-Wide Association Study Reveals Genetic Architecture of Eating Behavior


8. doi:10.1038/ncb1623.


